ESPOO 2006 VTT PUBLICATIONS 621

Matti Kutila

Methods for Machine Vision Based Driver Monitoring Applications

VTT PUBLICATIONS 621

Methods for Machine Vision Based Driver Monitoring Applications

Matti Kutila

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Department of Automation, for public examination and debate in Tietotalo Building, Auditorium TB104 at Tampere University of Technology on 8th of December, 2006, at 12 noon.

ISBN 951–38–6875–3 (soft back ed.)
ISSN 1235–0621 (soft back ed.)

ISBN 951–38–6876–1 (URL: http://www.vtt.fi/publications/index.jsp)
ISSN 1455–0849 (URL: http://www.vtt.fi/publications/index.jsp)

JULKAISIJA – UTGIVARE – PUBLISHER

VTT, Vuorimiehentie 3, PL 1000, 02044 VTT
puh. vaihde 020 722 111, faksi 020 722 4374

VTT, Bergsmansvägen 3, PB 1000, 02044 VTT
tel. växel 020 722 111, fax 020 722 4374

VTT Technical Research Centre of Finland, Vuorimiehentie 3, P.O.Box 1000, FI-02044 VTT, Finland
phone internat. +358 20 722 111, fax + 358 20 722 4374

VTT, Tekniikankatu 1, PL 1300, 33101 TAMPERE
puh. vaihde 020 722 111, faksi 020 722 8334

VTT, Teknikvägen 1, PB 1300, 33101 TAMMERFORS
tel. växel 020 722 111, fax 020 722 8334

VTT Technical Research Centre of Finland
Tekniikankatu 1, P.O. Box 1300, FI-33101 TAMPERE, Finland
phone internat. +358 20 722 111, fax +358 20 722 8334

Technical editing Anni Kääriäinen

Edita Prima Oy, Helsinki 2006

Kutila, Matti. Methods for Machine Vision Based Driver Monitoring Applications [Menetelmiä konenäköpohjaisiin kuljettajan monitorointisovelluksiin]. Espoo 2006. VTT Publications 621. 82 p. + app. 79 p.

Keywords driver monitoring, machine vision, distraction, fatigue, wavelets, SVM, neural networks, classification, cameras, traffic safety, vehicles, sensors, colour vision, alertness, gaze, eyes, head, workload, traffic safety and vigilance

Abstract

An increasing number of information and driver-assistive facilities—such as PDAs, mobile phones, and navigators—are a feature of today's road vehicles. Unfortunately, they occupy a vital part of the driver's attention and may overload him or her in critical moments when the driving situation requires full concentration. The automotive industry has shown a growing interest in capturing the driver's behaviour due to the necessity of adapting the vehicle's Human–Machine Interface (HMI), for example, by scheduling the information flow or providing warning messages when the driver's level of alertness degrades. The ultimate aim is to improve traffic safety and the comfort of the driving experience.

The scope of this thesis is to investigate the feasibility of techniques and methods, previously examined within the industry, for monitoring the driver's momentary distraction state and level of vigilance during a driving task. The study does not penetrate deeply into the fundamentals of the proposed methods but rather provides a multidisciplinary review by adopting new aspects and innovative approaches to state-of-art monitoring applications for adapting them to an in-vehicle environment. The hypotheses of this thesis states that detecting the level of distraction and/or fatigue of a driver can be performed by means of a set of image processing methods, enabling eye-based measurements to be fused with other safety-monitoring indicators such as lane-keeping performance or steering activity. The thesis includes five original publications that have proposed or examined image processing methods in industrial applications, as well as two experiment-based studies related to distraction detection in a heavy goods vehicle (HGV), complemented with some initial results from implementation in a passenger car.

The test experiments of the proposed methods are mainly described in the original publications. Therefore, the objective of the introduction section is to generate an overall picture of how the proposed methods can be successfully incorporated and what advantages they offer to driver-monitoring applications. The study begins by introducing the scope of this work, and continues by presenting data acquisition methods and image pre- and post-processing techniques for improving the quality of the input data. Furthermore, feature extraction from images and classification scheme for detecting the driver's state are outlined based in part on the author’s own experiments. Finally, conclusions are drawn based on the results obtained.

Kutila, Matti. Methods for Machine Vision Based Driver Monitoring Applications [Menetelmiä konenäköpohjaisiin kuljettajan monitorointisovelluksiin]. Espoo 2006. VTT Publications 621. 82 s. + liitt. 79 s.

Avainsanat: driver monitoring, machine vision, distraction, fatigue, wavelets, SVM, neural networks, classification, cameras, traffic safety, vehicles, sensors, colour vision, alertness, gaze, eyes, head, workload, traffic safety and vigilance

Tiivistelmä

Kuljettajan tukijärjestelmien määrä kasvaa tulevaisuudessa. Tämä helpottaa ajamista ja lisää ajomukavuutta, mutta tuo toisaalta mukanaan lieveilmiöitä. Muiden muassa matkapuhelimet, navigaattorit ja musiikkisoittimet kilpailevat yhä enenevässä määrin kuljettajan huomiokyvystä. Mikä pahinta, nämä laitteet saattavat haitata kuljettajan keskittymistä ja aiheuttaa onnettomuuden vaaran. Ajoneuvoteollisuus on tästä syystä osoittanut kasvavaa kiinnostusta kuljettajan tilaa monitoroivia järjestelmiä kohtaan. Nämä järjestelmät mahdollistaisivat kuljettajan ja ajotilanteen mukaan säätyvän älykkään käyttöliittymän kehittämisen. Tällainen käyttöympäristö voisi esimerkiksi viivyttää ei-kiireellisten ajoneuvon tilatietojen välittämistä, kuten tuulilasin pesunesteen loppumisesta varoittavaa viestiä, kunnes kuljettaja todetaan ”valmiiksi” vastaanottamaan informaatio. Tavoitteena on siis tehdä ajamisesta entistä mukavampaa ja mikä tärkeintä myös turvallisempaa, jottei kuljettajaa häirittäisi kriittisillä hetkillä.

Tämän väitöstyön tarkoituksena on tutkia teollisessa ympäristössä kokeellisesti hyväksi havaittujen menetelmien soveltuvuutta kuljettajan havaintokyvyn ja väsymystilan arviointiin. Työn tarkoitus ei ole tuottaa syvällistä analyysia ehdotetuista menetelmistä, vaan tarkastella asiaa poikkitieteellisesti. Tämä avartaa uusia näkökulmia ja innovatiivisia lähestymistapoja olemassa oleviin monitorointijärjestelmiin ja auttaa niiden sovittamisessa ajoneuvoympäristöön. Työssä testattava hypoteesi esittää, että kuljettajan häiriytyminen ja/tai väsymystila voidaan havaita kuvankäsittelymenetelmillä. Niiden avulla on mahdollista mitata häiriytymisaste kuljettajan silmistä ja yhdistää tätä tietoa muihin indikaattoreihin kuten kaistalla vaelteluun tai epätasaisiin ohjausliikkeisiin. Tämä työ koostuu viidestä alkuperäisjulkaisusta, joissa käsitellään ja testataan kuvankäsittelymenetelmiä teollisissa sovelluksissa, sekä kahdesta julkaisusta, joissa tutkitaan

kuljettajan häiriytymisen mittaamista kuorma-autossa. Näitä tuloksia on täydennetty alustavilla mittauksilla henkilöautoissa.

Tarkasteltujen menetelmien tulokset esitetään pääosin liitteenä olevissa alkuperäisjulkaisuissa. Johdanto-osan tarkoitus on luoda ajatus siitä, miten ehdotetut menetelmät tulisi yhdistää ja millaisia mahdollisuuksia ne avaavat kuljettajan monitorointisovelluksissa. Väitöskirjassa esitellään aluksi työn aihepiiri, sen jälkeen datan keruussa käytetyt laitteet, menetelmät ja kuvankäsittelytekniikat. Työ tarkastelee, kuinka tiedon luotettavuustaso paranee eri menetelmiä ja kuvankäsittelytekniikoita käyttämällä. Seuraavaksi kirja esittelee testituloksiin pohjautuen piirteiden irrotus- ja luokittelumenetelmiä kuljettajan tilan tunnistamiseksi. Lopuksi tarkastellaan saavutettuja tuloksia ja niiden merkittävyyttä.

Preface

The work of this thesis has been mainly initiated by the AIDE (Adaptive Integrated Driver-vehicle InterfacE, IST-1-507674-IP) project, which is funded by the European Commission in the 6th Framework Programme. Therefore, I am grateful to the whole consortium (28 partners) due to the fruitful discussions and guidance for working with in-vehicle systems. Specifically, I want highlight the significant contribution of Mr. Gustav Markkula from Volvo for the discussions and tips while designing and building the module for monitoring the distraction level of a driver. I am also more than happy to have had the chance to work with Dr. Trent Victor while I was preparing the journal article, which forms a crucial part of this thesis. To be honest, before starting to work in AIDE, I had no idea that this would be the framework for my dissertation.

My deepest gratitude goes to my supervisor Prof. Reijo Tuokko from Tampere University of Technology, for discussions, steering, and patiently waiting for the day when all was completed. I want to express my greatest thanks also to the pre-reviewers and the opponents Prof. Sukham Lee (SungKyunKwan University, South Korea), Prof. Ansgar Meroth (Heilbronn University, Germany) and Prof. Pasi Fränti (University of Joensuu, Finland) for the guidance and valuable advice given already in the early stages of finalising the work. Additionally, I give my deepest thanks to Prof. Ari Visa (Tampere University of Technology, Finland) who provided valuable recommendations when I took the first steps towards finalising the dissertation.

I am very grateful to the Nozone (An intelligent responsive pollution and odour abatement technology for cooking emission extraction systems, EVK4-CT-2002-30009) consortium for having the chance to work with them. Nozone was the past project funded by the European Commission in the 5th Framework Programme. I also want to express my gratefulness to Kuusakoski Oy and Mr. Antero Vattulainen, who has provided me with an opportunity to develop my expertise in the field of classification methods.

I would like to express my acknowledgement also to my colleagues Mr. Hannu Hakala (nowadays is employed by Elektrobit Group plc.), Dr. Tapani Mäkinen and Mr. Pertti Peussa in VTT Technical Research Centre of Finland for

organising funding and providing me with an opportunity to be involved in the driver-monitoring related topics and for access to European networks connected to the automotive industry.

I have had the honour of being a member of the Machine Vision team in VTT since it was inaugurated approximately seven years ago. All the people who have previously or are currently working in the team have contributed to the thoughts behind this work. Further, I want to express my deepest gratitude to two persons in particular in the team Prof. Jouko Viitanen and Mr. Juha Korpinen (who now works for Chip-man Technologies Ltd.) for their guidance and for initiating me into the world of machine vision technology. Of course, I am thankful also to Mr. Jukka Laitinen and Ms. Maria Jokela for their contributions and assistance. I would also like to thank my English language reviser Mr. Mark Phillips for assisting in the final preparation of the thesis.

I wish to express my gratitude to all my relatives and friends for their encouragement to continue until final end for finishing this study. In certain moments, when it was not clear whether this whole work would one day reach its fruition, at least two people always believed so, my parents Mrs. Liisa and Mr. Heikki Kutila. I am grateful forever for their support financially and mentally during this long educational journey.

Last but certainly not least, I want to express my deepest appreciation to my wife Mrs. Soile Kutila for spell checking, criticism, support, and those important moments when I had the chance to totally forget this work and recharge my batteries. A thousand thanks!

Tampere, November 2006

Matti Kutila

List of Original Publications

I. Kutila, M., Korpinen, J. & Viitanen, J. 2001. Camera Calibration in Machine Automation. *Human Friendly Mechatronics*. Selected papers of the International Conference on Machine Automation ICMA 2000. Osaka, Japan. 27–29 Sep 2000. Amsterdam: Elsevier. Pp. 211–216. ISBN: 0-444-50649-7.

II. Kutila, M. 2004. Calibration of the World Coordinate System with Neural Networks. *Proceedings of the 9th Mechatronics Forum International Conference Mechatronics 2004*. Culture & Convention Centre, METU, Ankara, Turkey. 30th Aug – 1st Sep 2004. Pp. 337–345. ISBN: 975-6707-13-5.

III. Kutila, M. & Viitanen, J. 2004. Parallel Image Compression and Analysis with Wavelets. *International Journal of Signal Processing*, Vol. 1, No. 1–4, pp. 65–68. ISSN: 1304-4478.

IV. Kutila, M. & Viitanen, J. 2005. Sensor Array for Multiple Emission Gas Measurements. *Proceedings of IEEE International Symposium on Circuits and Systems ISCAS 2005*. International Conference Center, Kobe, Japan. 23–26 May 2005. Pp. 1758–1761. ISBN: 0 7803 8834 8.

V. Kutila, M., Viitanen, J. & Vattulainen, A. 2005. Scrap Metal Sorting with Colour Vision and Inductive Sensor Array. *Proceedings of International Conference on Computational Intelligence for Modelling Control and Automation CIMCA 2005*. Vienna, Austria. 28–30 Nov 2005. Los Alamitos, CA: IEEE. Vol. 2, pp. 725–729. ISBN: 0 7695 2504 0.

VI. Markkula, G., Kutila, M., Engström, J., Victor, T. W. & Larsson, P. 2005. Online Detection of Driver Distraction – Preliminary Results from the AIDE Project. *Proceedings of the 2005 International Truck and Bus Safety and Security Symposium*. Washington, Alexandria, Virginia, U.S.A. 14–16 Nov 2005. Pp. 86–96.

VII. Kutila, M., Jokela, M., Mäkinen, T., Viitanen, J., Markkula, G. & Victor, T. W. Driver Cognitive Distraction Detection: Feature Estimation and Implementation. Submitted to *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering* on 11 Apr 2006. United Kingdom. ISSN: 0954-4070.

Contents

Abbreviations

ADAS	Advanced Driver Assistance Systems
AIDE	Adaptive Integrated Driver-vehicle InterfacE
AVI	Audio Video Interleave
AWAKE	System for Effective Assessment of Driver Vigilance and Warning According to Traffic Risk Estimation
CAA	Cockpit Activity Assessment
CAN	Controller Area Network
CCD	Charge-Coupled Devices
CMOS	Complementary Metal-Oxide Semiconductor
CPU	Central Processing Unit
DCT	Discrete Cosine Transform
EC	European Commission
EEG	Electroencephalogram
EOG	Electroculogram
EU	European Union
HGV	Heavy Goods Vehicle
HMI	Human-Machine Interface
IR	Infrared

IVIS	In-Vehicle Information System
JPEG	Joint Photographic Experts Group
MLP	Multi-Layer Perceptron
MOST	Media-Oriented Systems Transport
MPEG	Moving Pictures Experts Group
MP3	MPEG-1 Audio Layer-3
NIR	Near Infrared
NOZONE	An intelligent responsive pollution and odour abatement technology for cooking emission extraction systems
PCA	Principal Component Analysis
PDA	Personal Digital Assistant
PERCLOS	Percentage of eyelid closure over the pupil over time
PMD	Photonic Mixer Device
RBF	Radial Bases Function
RGB	Red-Gree-Blue colour space
SENSATION	Advanced Sensor Development for Attention, Stress, Vigilance & Sleep/Wakefulness Monitoring
SVM	Support Vector Machines
SMS	Small Vision System
TOF	Time Of Flight

VOC	Volatile Organic Compound
VTT	VTT Technical Research Centre of Finland
xPC	Industrial computer from The MathWorks
YUV	Luminance-Chrominance colour space

1. Introduction

1.1 Background

During 2004, VTT commissioned two market analyses to assess the national interest in implementing camera vision technology in the plastic and food industry: 132 responses were gathered from the plastic industry field and 146 from the food sector. More than half of the companies reported not yet employing vision techniques in their day-to-day work, which highlighted a growing market potential. However, an even more promising field for such technology is the vehicle industry, since future prospects are that more sophisticated In-Vehicle Information Systems (IVIS) and Advanced Driver Assistance Systems (ADAS) are needed to take account of driver's states and the actual driving environment. So far, the vision systems have been rarely adopted due to cost, lack of robustness and the large size of the equipment. Monitoring the driver's behaviour has received a lot of interest recently. However, it is not the only example in the traffic-safety field where a camera vision technique is generally applicable. Lane positioning, which is also an important driving performance descriptor (Publications VI and VII), is typically measured by an optical device (McCall & Trivedi 2006). Huber et al. (1998) present a camera implementation which uses polarisation planes to identify ice or water on the road, thus providing an opportunity for the driver to adapt speed and steering movements to reduce the chance of skidding. Hautiere et al. (2006) have explored a methodology for estimating the visible range in foggy conditions. The above examples indicate the potential for utilizing optical instrumentation in future vehicles and provides an understanding of why this topic is highly prominent in the automotive industry at the moment.

One of the major reasons for traffic accidents is the driver's own behaviour (e.g. in Figure 1) (Dingus et al. 2006, Neale et al. 2005, Klauer et al. 2006). According to French statistics, a lack of attention due to fatigue or sleepiness was a factor in one in three motorway accidents, while alcohol, drugs and distraction was a factor in one in five accidents in 2003 (Federation of French motorway and toll facility companies 2006). Moreover, Bellotti et al. (2005) and Tattegrain et al. (2005) recognised the necessity of adapting the information flow to the in-vehicle HMI by delaying non-urgent messages until the driver's dynamic behaviour returns to an unstressed traffic situation. It is anticipated that without smart information scheduling, the driver pays too much attention towards the entertainment facilities or the status of monitors in the vehicle. Similar rationalisations were earlier performed in designing the cockpits of aircrafts and fighters (Bruce et al. 1998). VTT has recently been commissioned by the European automotive industry in the field of monitoring a driver's momentary state. Driver monitoring is useful for many types application, including warnings when the driver's attention is impaired, providing possibility to reduce the effect of a distraction source or performing real-time HMI-adaptation (Arensberg 2004, Almén 2003, Claesson 2003, Larsson & Victor 2005, Victor 2000, Victor 2003). The activity, which has recently motivated the author, is a project called AIDE (Engström et al. 2006). The project aims to generate smart HMI technology for adapting the user interface of in-vehicle information systems (IVIS) and advanced driver assistance systems (ADAS) according to the driver's ability and available attention. The theme of this thesis is to discover the methods and the various technological features necessary in order to assess the driver's behaviour so as to enable HMI adaptation to critical traffic safety situations.

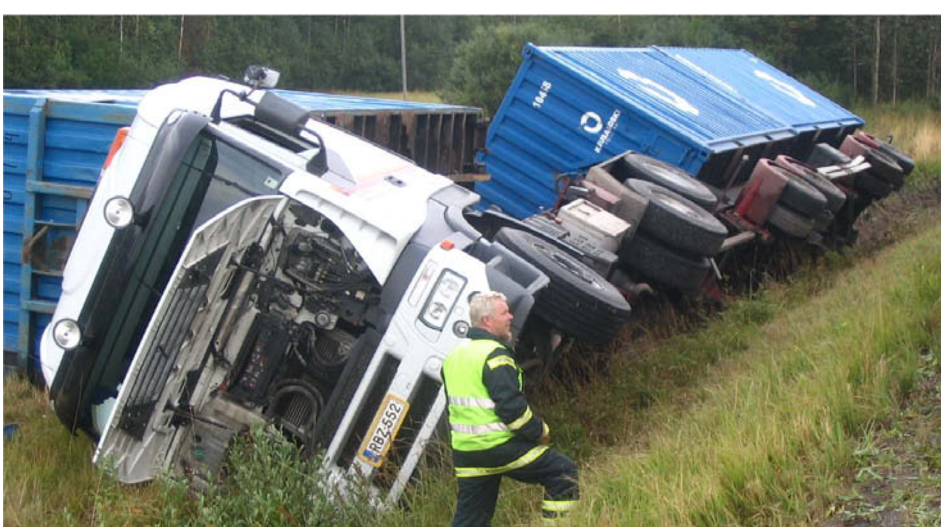


Figure 1. Heavy goods vehicle accident where the driver's attention was degraded due to an external event.

Publications I–V provide industrial experiments for constructing machine-vision facilities whereas Publications VI and VII explore methods and the results of experiments to detect the level of visual and cognitive distraction of a driver. The main effort in this thesis is applied to the principles of optical measurement. A number of studies exists that focus on creating a platform for monitoring-applications or which relate to fatigue detection. However, publications that merge these two topics are not commonly presented. Furthermore, for example, camera calibration techniques are considered rarely, although such is crucial in order to use low-cost camera components in stereo vision systems. Basically, the machine-vision principles (e.g. the steps needed to provide the classification result or difficulty of the varying lighting conditions) are equally important to vehicle systems and to industrial applications. An industrial aspect exists strongly in the background of this study. Thus, various methods that have been evaluated in industrial applications are proposed here to improve overall performance, and in particular, the robustness of driver-monitoring systems.

1.2 Hypothesis, objectives and constrains

The aim of this study is to provide guidelines for techniques and signal processing methods in order to monitor the driver's alertness and availability for driving (Figure 2).

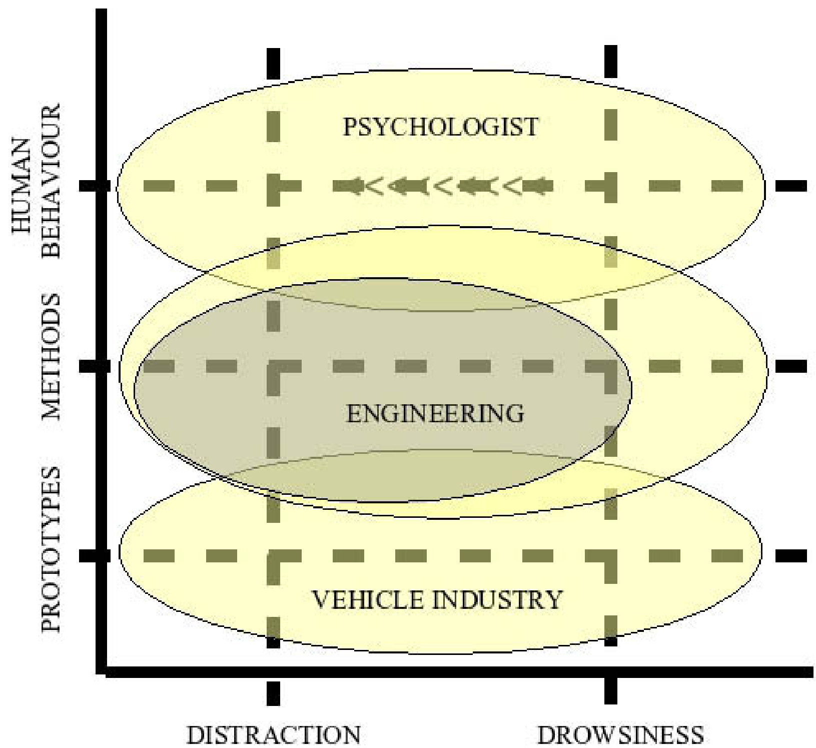


Figure 2. The ellipses describe by whom and where the activities of multi-disciplinary driver-monitoring are mostly performed. This study is related more to distraction detection (the dark grey ellipse) but provides some minor propositions for fatigue detection too.

The research hypothesis of this study is:

Image processing methods—such as data acquisition, camera calibration, attention mapping, feature extraction and classification—for performing eye-based measurements (movements, blinking, attention targets, etc.) in combination with other indicators (e.g. lane-keeping performance or steering wheel movements) enable the detection of the driver's momentary distraction or fatigue level.

This thesis combines 7 different applications: camera calibration (Publication I), object mapping (Publication II), neural networks (Publication II), wavelets (Publication III), data transmission (Publication III), data acquisition

(Publication IV), colour classification (Publication V), which are examined in the context of industrial applications. The two experimental studies in relation to driver monitoring (Publications VI and VII) generate the foundation for the arguments of this dissertation. To briefly summarise, Publications I–V are intended to provide a broad foundation to monitoring activities, whereas Publications VI and VII focus on the field experiments.

The major objectives of this thesis are:

- To present guidelines for generating the data flow, thus creating a platform for machine vision applications for driver monitoring
- To obtain experimental results for monitoring the driver's visual distraction level, which measures how much the driver’s eyes are directed to the road ahead
- To explore cognitive distraction detection in practice. Cognitive distraction refers to whether the driver’s thoughts are on the driving task or impaired by e.g. daydreaming, fatigue, deep thinking, etc. (Victor 2005).

Minor contributions are also focused on:

- Review requirements for data acquisition with camera vision equipment
- A feasibility discussion concerning the following topics: eliminating the optical errors of lenses, using wavelets for eye tracking, activity measures with colour analysis, using neural networks for distraction or fatigue analysis and automatic attention target mapping in a cockpit
- Description of the relevant parameters for assessing the state of a driver.

This thesis does not cover the following items:

- The experiments are restricted to vehicle drivers and are not directly applicable for human monitoring in other environments (e.g. aircrafts).
- No other sensing methods than machine vision sensing are explored (e.g. EEG/EOG analysis for detecting drowsiness of a driver). The literature review anticipates that eye analysis is the most appropriate

methodology to perform the activity analysis; moreover, an EEG measurement for example would require devices that have to be installed in a particular location and touch a human body and therefore, are not feasible for commercial monitoring equipment.

- An exhaustive analysis of compression techniques or a comparison of communications channels. The compression techniques have been researched exhaustively by a number of studies during the last two decades. The compression methods would generate a dissertation topic in itself and are, therefore, only mentioned in the text when relevant to discussing communication between in-vehicle information devices.
- Experimental results for utilising optical-error-removal, eye tracking or performing drowsiness/fatigue detection in a true traffic environment. The experimental results rely on the faceLAB system, which includes the above methods. Therefore, the topics are investigated in one sense but not directly. However, since the image processing is the main measurement principle, the methods needed to perform driver monitoring with a camera vision technique are explored individually step by step.
- The detection techniques are restricted to methods of supervised learning. Most of the time a human being behaves “normally”, being alert while driving. Therefore, unsupervised training methods would not presumably distinguish the abnormal states, since they may not appear under the variation of normal driving indicators.

1.3 Prior knowledge of driver monitoring

Monitoring of driver status can be divided into the two main branches: distraction detection and identifying sleepiness. However, they partially overlap since the context awareness of a driver is related to sleepiness and to cognitive distraction, which both represent mental occurrences in humans. Bergasa et al. (2006) have been motivated in his work to discussing distraction—also the main objective of Publications VI and VII—as the more severe problem due to an increasing number of Advanced Driver Assistance Systems (ADAS), PDAs, mp3 players, etc. in modern vehicles. However, in practice Bergasa et al. (2006) performed a hardware implementation for monitoring the level of a driver’s vigilance with Percent Eye Closure (PERCLOS) and with additional camera measurements: eye blinking frequency, gaze fixations and head movements.

Publication VI discusses visual distraction, which is the estimation of how much the driver pays attention to driving compared to non-driving related targets (e.g. radio, mobile phone, passenger, etc.). Dinges et al. (1998) have found that a relationship exists between eye closures and lapses of visual attention. Driver activity in the cockpit is also investigated by Wahlstrom et al. (2003), who built up a system for tracking the driver's eyes and detecting, for example, radio-directed activity. Their interest was in detecting the driver's momentary attention target, as detailed in Publication VI. Fletcher et al. (2001 and 2003) discussed how to estimate the driver state (fatigue, inattention due to traffic context) and fused these with the driving performance indicators (lane-keeping and obstacle detection). They utilised the faceLAB system (Seeing Machines 2006) in the experiments and produced some promising results. However, a more advanced evaluation of the results is needed before making further judgements on the method's performance.

The European Commission (EC) has announced two extensive activities for promoting the monitoring of driver fatigue: AWAKE and Sensation. Their objectives are to develop a technique that could be feasibly implemented in the vehicle to maintain the alertness of the driver. A major conclusion of the projects is that typically the fatigue measurement devices should not rely solely on detecting eye closures. The suggestion is also to adopt a behavioural analysis (e.g. limb, gaze or head movements, etc.) of the driver and also to utilise driving performance measures (e.g. lane-keeping or steering wheel reversal rate) (Boverie 2004). An example of such an idea is given by Grace et al. (1998), who have provided a methodology for fusing PERCLOS with behavioural measurements (use of acceleration, steering wheel movements, lane and head positions) so as to detect fatigue. It is envisaged that these indicators will later be merged with neural networks. However, probably the only commercial optical fatigue-detecting device that has been offered is that of Grace (2001). The device provides PERCLOS-based drowsiness assessment and is intended for use in HGVs.

Grauman et al. (2002) have proposed utilising PERCLOS alongside detection of a driver's head movements so as to improve the robustness of their monitoring scheme. The eye-closure measure that was adopted utilises principal component analysis (PCA) to decrease the effect of varying lighting conditions, that may occur, for example, as a result of tunnels, weather conditions, etc. Interestingly, the head nodding detection was implemented with an algorithm intended to increase total performance by making the eye-tracking more reliable by extracting glances towards a map or mirror checks. Eye-tracking was also the topic of Veeraraghavan and Papanikolopoulos (2001), who applied a method utilising skin colour for extracting eyes and thus, providing a platform for performing PERCLOS. Utilising colour information is the topic of Publication V and will also be discussed more in Chapter 4. Boverie et al. (2002) have examined how to detect driver vigilance by using vehicle speed, steering wheel movements, eye-blinking and vehicle lateral position. They model the driver statistically and generate a hypothesis that a large deviation compared to the "standard" model is the result of impaired vigilance. This is a practical approach, since training the classifier with non-vigilant data is almost impossible, i.e. the driver cannot be asked to drive drowsily for the first few kilometres. Santana Diaz et al. (2002) used the vehicle's lane-position variation, steering wheel movements and speed variation, transforming them into wavelets so as to monitor the driver's state. Bittner et al. (2000) have performed an experimental study measuring driver fatigue outside the laboratory and on a real road. Although a statistical analysis was not completed in the paper, the study reported effects in steering activity and lane keeping performance. However, it surprisingly reported sceptical results for an association between blinking frequency and fatigue, which does not concord with other related studies (Bergasa et al. 2006, Boverie et al. 2002, Dinges et al. 1998).

Pilutti and Ulsoy (1995) have investigated the possibility of using lane keeping performance and variations in steering movements to create and update on-the-fly the model of a driver. Their experiments support the assumption that the descriptors (lane position variation and steering activity) are relevant for monitoring the driver's state. On the other hand, the study of Rimini-Doering et al. (2001) explored the relationship between driver's eye movements and lane-keeping with the driver vigilance. Heitmann et al. (2001) analysed head and gaze position variance, pupillary changes, and eye blink rate to estimate driver alertness. They observed that all the tested input variables were influenced by the driver state. Moreover, they concluded that any of the single signals alone cannot provide a reliable indication of fatigue, so they instead promoted the utilisation of multiple descriptors in conjunction with a neural-fuzzy hybrid algorithm. Wang et al. (2003) have successfully examined the Gabor wavelet functions for eye measurements used in conjunction with MLP-type neural networks for detecting driver fatigue.

Thus far, all the presented monitoring applications have relied solely an actual data. Nevertheless, it is a fact that human sleepiness does not appear suddenly: the transition from a vigilant to sleepy state proceeds slowly, via a slight drowsiness phase. Zhu and Ji (2004) used a fusion of eyelid, gaze and head movement monitoring with facial expression to detect fatigue and complimented the descriptors with information on temperature and sleep history. The test results disclosed a very good and robust outcome for a number of different test subjects of different ages, genders and ethnic background.

Some driver monitoring techniques were investigated initially in the aviation industry some 20 years ago. Albery et al. (1987) published results that identified a correlation between the various human measures (visual evoked response, eye blinks, heart rates, arm muscle activities and blood pressure) and mental workload caused by noise in the cockpit of a fighter aircraft. Later, East et al. (2002) explored appropriate features and classification methods for detecting the mental workload of a fighter pilot. An EEG signals, accompanied by a subset of heart rate, breathing, and eye-blinking, were used to compare the capability of an MLP neural network and statistical classification methods. They concluded that the neural networks provided the better detection performance. O'Brien (1988) developed a hardware set-up for detecting the blinking frequency of the pilot's eyes. The hardware was verified by comparing the result to an EOG signal and they reported a 90% success rate for detecting eye blinks.

Lal et al. (2003) have created software to detect fatigue that utilises frequency analysis of the EEG signal. The promising results were based on tests performed on 10 test subjects. However, more exhaustive and naturalistic tests should be performed before strong conclusions can be drawn on the robustness and performance of the methodology. Gonzalez-Mendoza et al. (2003) used EEG/EOG with the support vector machines (SVM) to estimate driver vigilance. However, Bittner et al. (2000) reported unexpectedly in some of their experiments concerning the EEG/EOG signal's dependency on fatigue, even though other studies (Lal et al. 2003, Gonzalez-Mendoza et al. 2003) have suggested a relationship exists.

The general conclusion is that state-of-the-art driver monitoring techniques can be categorised into methods that measure the driver's actual state (e.g. eye-blinking, gaze movement, etc.) and those that assess the driver according to driving performance (e.g. lane-keeping, headway to front vehicle). The third and more advanced technique is a fusion of these first two methods.

2. Structure of the Thesis

The thesis begins (Chapter 1) by exploring the advantages of driver monitoring applications from the perspective of promoting traffic safety. Then the research hypothesis, aims and limitations of the work are declared. The final part of this first chapter consists of a review of the state-of-the-art techniques, implementations and pre-existing know-how for monitoring driver state, driver distraction or fatigue. Chapter 2 describes the structure of this thesis. In general, machine vision applications require a data flow chart which is illustrated in Figure 3. The data flow is also relevant in driver monitoring but the steps require different view points and adaptations in order to perform robustly and to be cost effective within in-vehicle systems. The driver monitoring technology workflow has not previously been considered comprehensively on scientific grounds despite the existence of solitary monitoring applications and their preferences. Therefore, a consideration of the whole chain is one of the innovations of this thesis. The monitoring data flow steps are discussed in more detail in Chapters 3, 4 and 5.

Chapter 3 presents methods for data acquisition and transmission. Optical sensing principles are the major focus but further aspects are considered by taking into account experiences from the gas sensor development project. Data transmission is significantly dealt with so as to explain the relevance of data compression when images or videos are transmitted.

Chapter 4 starts by providing guidelines for eliminating optical errors and performing feature extraction. The first part concentrates on the image post-processing stage which covers camera calibration. The feasibility of using wavelets and colour analysis in driver monitoring are discussed and additionally, the most relevant features are retrieved from the literature.

Chapter 5 is the crucial part of this thesis, discussing distraction and fatigue detection techniques and also outlining experimental results. The section covers classification methods for detecting visual and cognitive distraction. Additionally, the section discusses vigilance detection and automatic adaptation of the attention mapping. This section also contains an analysis of the proposed classification methods in practice.

Chapter 6 explains the relevance of the original publications to the objectives of this thesis and the author's contributions to each. Chapter 7 outlines the major achievements and considers the future development work necessary for building real commercial products that would likely be incorporated into vehicles by car manufacturers so as to monitor a driver's momentary state. The original publications are attached as appendices at the end of this thesis.

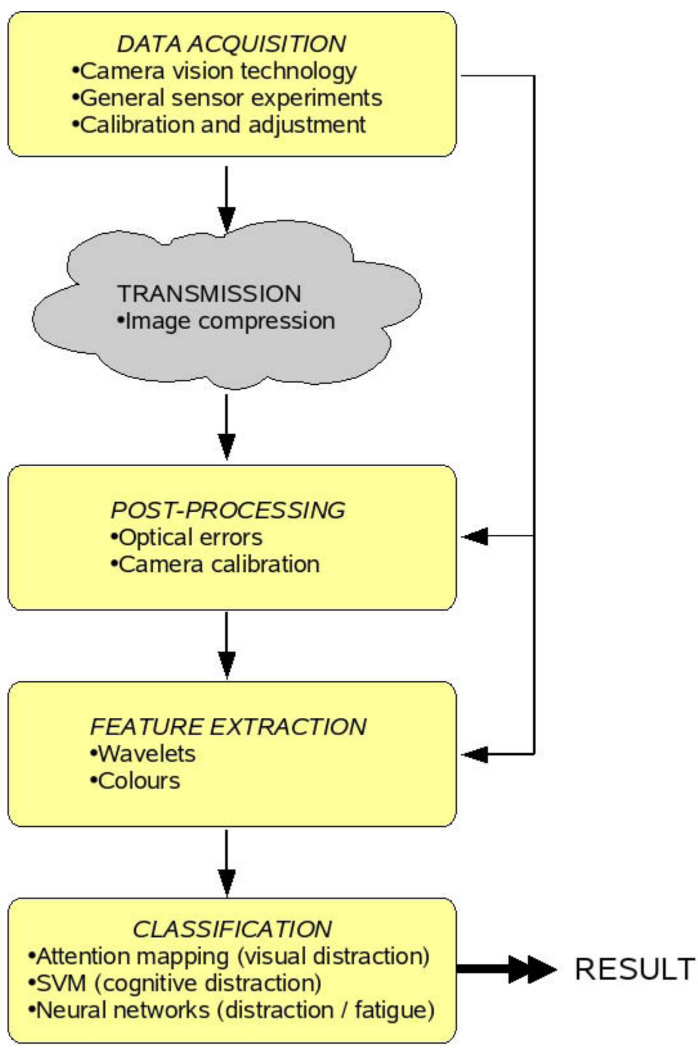


Figure 3. General data flow of machine vision systems, which is relevant also to driver monitoring applications. Items categorised in each step represent the topics and aspects discussed in this thesis.

3. Data Acquisition and Transmission

3.1 Overview

The topics of this chapter include two items: data acquisition and transmission, as detailed in top of the data flow diagram (Figure 3). The first topic is intended to describe the general requirements of data acquisition devices for monitoring driver behaviour. Since the major interests of this study are the principles of optical measurement, the main focus is on a stereo vision system, with a practical experiment of such a system detailed in Chapter 5 and Publication VII. However, data acquisition is also discussed at a more general level by taking into account the experiments of the gas sensor development process presented in Publication IV.

The second topic (data transmission) addresses the necessity of data compression when images or videos are transmitted, since the bandwidth of the vehicle buses (CAN: 20 kbit/s – 1 Mbit/s, MOST: 20–50 Mbit/s) are shared by multiple in-vehicle applications. The description of compression and reconstruction takes in the wider aspects of industrial experiences, which are examined in more detail in Publication III.

3.2 Data acquisition

In many cases, a simple on/off -type output is sufficient and more reliable, as in the gas sensing application (Publication IV). The gas sensor's (see Figure 4) correlation with a camera may sound awkward at first. However, both of them process an analogue signal that is converted to digital format for analysis in a computer. Thus the gas sensor is like an imaging element based on only one pixel and senses gas concentrations instead of light intensities. In general, the data source is not crucial for the purposes of this thesis. The main point is that the sensor provides the descriptors (i.e. features) that have an obvious dependency with the desired identification result. For example, the sensing system in Publication V could be an X-ray detector instead of a colour camera.

Figure 4. The developed data acquisition device for detecting gas concentrations in Publication IV.

Eye tracking sets high demands for data quality in the driver monitoring applications as the following examples anticipate. Eriksson and Papanikolopoulos (1997) presented an eye-tracking and also iris-finding technique by utilizing spatial complexity around the eyes. Perez et al. (2003) have developed the lightning arrangement for detecting a driver's pupils by using the corneal glint reflections. Ito et al. (2002) proposed motion-picture processing in which the peaks of the detected eye-closure shapes are utilised. Publication IV investigates the adaptation of the data acquisition (gas sensor) device to an operating environment. In the cases of driver monitoring, the machine vision system is utilised in outdoor conditions and therefore, a fast and easy sensor adaptation capability is desired. Publications VI and VII utilise a stereo vision system for gathering the input data to assess driver behaviour. The platform in both papers is the faceLAB (Seeing Machines 2006) stereo vision system, which provides 3D measurements concerning the driver's head and gaze movements and advanced eye analysis results (e.g. blinking frequency, eye closure, saccades, etc.). The low-cost Small Vision Systems (Konolige & Beymer 2006) may achieve the necessary inputs as well (see. Figure 5), but requires additional work for implementing the eye-tracking algorithm. Moreover, the Small Vision System requires further development before the data quality meets the requirements of analysing algorithms that are reported in Publications VI and VII. The faceLAB system provides an automatic calibration capability, which is the essential factor for a robust analysis of driver distraction.

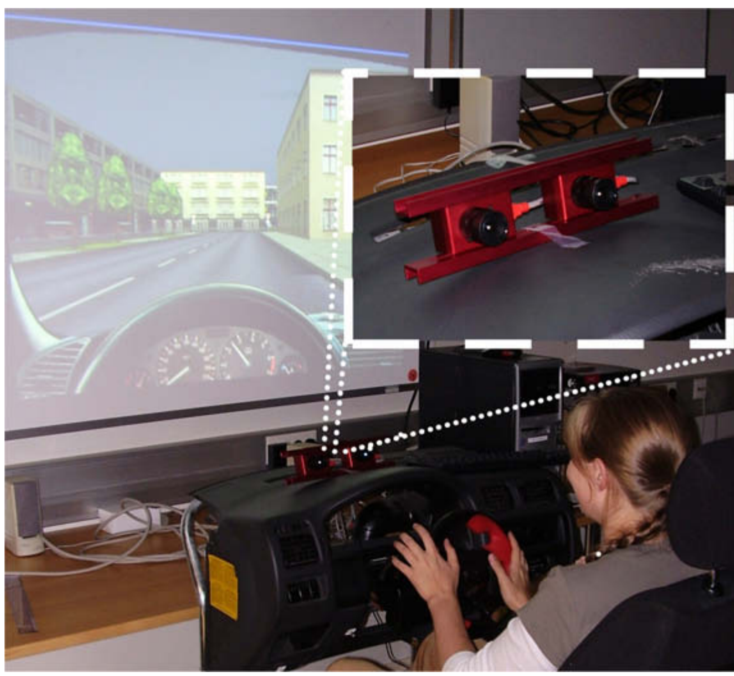


Figure 5. The stereo vision system installed on the driving simulator for testing and adapting the driver monitoring algorithms.

Noise due to a data source requires a proper filtering technique. Mostly, the noise patterns are predicted according to the preliminary known characteristics of the hardware. However, this is an important step since the consequence of an unstable signal makes it more difficult to create the proper feature vectors, which therefore may drastically decrease the overall target identification performance. In principle, two alternative processing or calibration techniques exist for eliminating anomalies in a raw sensor signal. The first option is to use hardware-based signal adaptation and the second is to convert the input signal to an appropriate format and then remove known errors. The hardware-based calibration methods are fast and provide in many cases a better result (e.g. the gas sensor in Publication IV), since the real signal is adapted and therefore information loss can be better controlled. The benefit of calibrating with software is that parameters can be changed on the fly and this is preferable in driver-monitoring applications where the environment is highly dynamic.

With proper calibration even a poor signal may provide sufficient results, but at a significantly lower cost than the selection of slightly better sensing elements (or cameras) as Publication IV indicates. It was initially predicted that typically the VOC or Ozone measurement devices would cost 1500 EUR each (Ho et al. 2001), but the current expected market price for the developed sensor is 800 EUR, which measures both gas types with adequate accuracy. The most demanding element in the development of the sensor was achieving proper calibration, which was initially successful in a laboratory. However, the example indicated that the final calibration had to be completed in a real environment where humidity, heat and dirt are realistic. An improved adaptation capability, which is discussed more on later, would improve the gas sensing application but on the other hand it would also increase the cost of the hardware. Nevertheless the same platform can be utilised for the sensor with an advanced adaptation technique if the price increase were acceptable.

Adaptation to a dynamic environment is an essential property since the identification performance rate is closely related to the appropriate features. In the faceLAB system, the cameras are adapted to the existing lighting conditions by automatically adjusting the gain, thus keeping the video signal at a sufficient level. Publication IV shows the application in which noise removal and gain control are also made at the hardware level. The gas sensor is calibrated internally by creating a lookup table, which maps the voltage output according to ozone and VOC levels. The sensor applies the gas levels without the need for further processing in a remote computing unit (i.e. the calibration has been performed internally in the data source). It should be noted that calibration is also discussed later in Chapter 4 but there it relates more to artificial image correction according to preliminary created formulas, i.e. it can be considered as a higher level correction, which is not typically implemented inside an embedded sensing device.

As mentioned, the test platform of the driver monitoring implementation of this thesis contains two Sony FCB-EX480A gray scale CCD cameras. The cameras are sensitive, guaranteeing sufficient operation also in dark lighting conditions (minimum required illumination is 0,7 lux). The cameras also included auto focus functionality and zooming capability. The cameras are high quality products intended for industrial surveillance and the drawback is the big size (50 x 52 x 88 cm) and the high price level (> 1000 EUR) when considering the in-vehicle products. The cameras are connected to the computer unit where the faceLAB (Seeing Machines 2006) software runs. The program tracks driver's eyes and performs eye based measurements (e.g. PERCLOS, saccades, etc.). Unfortunately, the program includes also multiple measures which consume computation power and are not needed by the distraction detection module. Thus, relevance of dedicated embedded sensing system is addressed when the development work progresses to a real product for minimising the size and the price of the module to the reasonable level for passenger cars.

3.3 Data transmission

One future scenario could be that even if driver monitoring is performed *inside* the vehicle, the result may be useful *outside* as well. Wireless sensor networks are becoming a reality in industrial installations and the same trend has obvious benefits in the traffic safety field. Future predictions anticipate that the road's infrastructure will include smart driver-assistant systems that will be able to communicate with the vehicles and also, the vehicles will include a capability to communicate with other vehicles. However, the reality is and will be also for the foreseeable future that wireless communication is limited by the available bandwidth. Therefore, it will be unlikely in the short term that they are capable of transmitting large data samples such as videos through the communication channel while the vehicles are moving. In some driver monitoring applications, a huge number of historical data is stored, which is impossible without signal compression (Ilic et al. 2004). Therefore, efficient compression algorithms are necessary especially in wireless communication, and this is apart from the bandwidth constraints of the in-vehicle buses (CAN: 20 kbit/s – 1 Mbit/s), which will however be increased to serve the development requirements of future multimedia devices (MOST: 25–50 Mbit/s).

An interesting study is Del Bue et al. (2002), which detail the development of a smart camera capable of efficiently compressing the background while maintaining tracked faces. Publication III discusses the wavelet-based compression method, which does not cause a blocking effect as does a DCT-transformation (Tan et al. 1995) of the JPEG format and additionally, may in the reconstruction phase, provide descriptors that are useful in driver monitoring. Image compression has been a topic of hundreds of articles, each proposing techniques dedicated to a certain application or condition. Therefore, techniques are not discussed deeply here but rather, some idea of the feasibility of the proposed methodology is given (Publication III). The relevance of the wavelet-based method will be discussed more in the chapter reviewing appropriate features.

In the prototype implementation (see Figure 6), the video monitoring unit is a separate computer. The monitoring computer is connected to the data logging unit which is capable of collecting synchronised data from the vehicle's CAN bus in order to utilise speed of a vehicle for the cognitive distraction detection. The logging unit also captures a video of the driver's face, which is synchronized with other gathered data for allowing offline analysis later. The videos are compressed to MPEG- or AVI-files since they are stored for debugging and supporting the tests. Thus they are not transmitted in the CAN buses due to the insufficient bandwidth. There are separate non-standard busses built for transmitting videos. The distraction monitoring application, to which the image processing unit transmits the necessary data, runs in an industrial real-time xPC computer of The Mathworks. The idea of using distributed computing units is practical also when considering the commercial implementations.

The aforementioned logging facility is an interesting feature since there have been discussions between European Union (EU) authorities that the vehicles at least those intended for professional driving should be equipped with a black box like aircrafts for storing the last moments before an accident. The driver's behaviour could be one of recorded things but this requires proper image compression for keeping size and price of the data storage unit low.

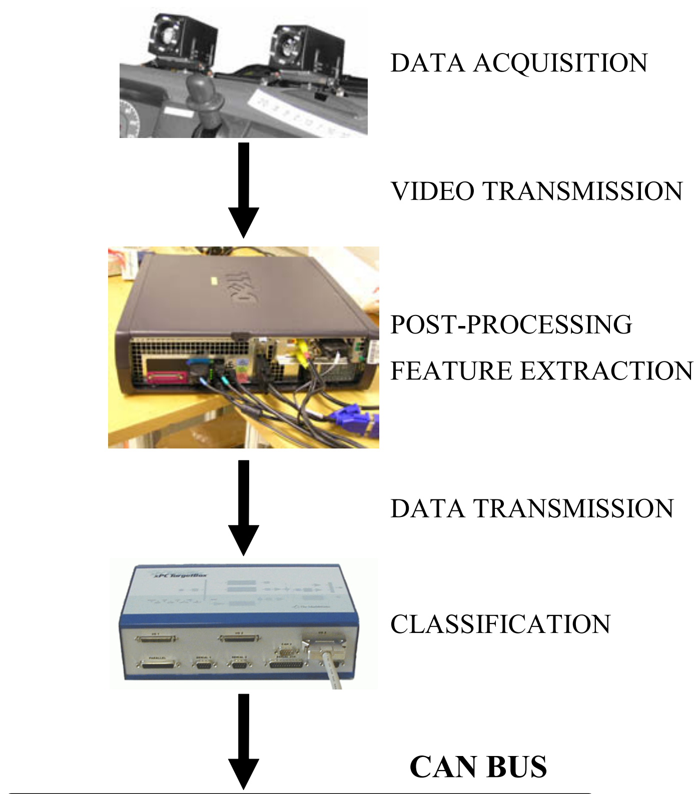


Figure 6. The hardware of the test systems.

4. Image Post-Processing and Feature Extraction

4.1 Overview

Since the price level of passenger vehicles are slightly decreasing despite the fact that an increasing number of in-vehicle electronics are being implemented, the costs of camera vision systems are required to be rather low (< 1500 EUR). Therefore, low-cost components are desired, which consequently promotes the importance of a software-based camera calibration routine. The distortion elimination procedure that can be applied to stereo vision to increase the robustness of the disparity calculation is described in begin of this chapter.

The second topic of this chapter focuses on feature extraction, as depicted in the data flow illustration (Figure 3). It reviews the feasibility of utilising wavelet descriptors and colour features in driver monitoring. They are implemented experimentally in the industrial applications described in Publications III and V. A more exhaustive treatment of the relevant features is applied in this chapter while utilising the literature review in Chapter 1.

4.2 Optical errors

Beymer (2000) introduces an application for counting the number of persons entering a shop. The system uses the Small Vision System (SVS) (Konolige & Beymer 2006), which, it was discovered, suffers high radial distortions, and therefore, they implemented the famous Tsai's method (Tsai 1987) so as to improve the robustness of the camera vision system. Eliminating optical errors is accentuated in a stereo vision application because the disparity calculation suffers or may even fail as a consequence of distortions. Low-cost stereo vision systems are expected to be incorporated into future passenger vehicles, thus removing optical errors will remain an important step in the development of driver-monitoring equipment.

Top-quality glass lenses do not cause severe errors and are usable in practical computer vision applications, but they are also many times more expensive than “traditional” optics. Plastics lenses, which are typically used in low-cost consumer devices like mobile phones, are the cheapest, but their imaging properties are poor and the captured pictures are mostly acceptable only for storing travelling memories. Car owners are ordinary people and they do not want to waste time for calibrating vehicle sensors. Thus, in addition to the low-price requirement, easy calibration is a crucial aspect for in-vehicle camera vision systems.

Ideal lenses refract light rays according to a pinhole model without influence from non-linear components in ray tracing (i.e. the rays are considered to pass the lens straightforwardly). However, lenses are made by grinding glass, which implicitly applies unique properties to each surface, thus making ray tracing more difficult. The quality of optics varies considerably depending on the material used and the manufacturing method, both of which reflect the main price factors. Each camera model is an approximation and the ideal camera model is impossible to formulate since all real imaging systems include some lens errors, generally called aberrations. The major errors are due to off-axis light rays when dealing with geometric optics. Dozens of different aberration types exist, some of them occurring independently and some having a mutual correlation.

The major aberration types (Hecht 1998):

- Distortion: pixels are mapped to incorrect locations (i.e. each image point is sharply focused but misplaced compared to ideal optics)
- Spherical aberration: the marginal light rays bend more than those which are nearby an optical axis, therefore, producing two separate image planes
- Coma: the rays which pass the lens in the periphery are focused closer to the optical axis than those tracing nearby the lens axis
- Astigmatism: the meridional and sagittal image planes occur at different distances from a lens
- Field curvature: the real image plane is rather curved than flat since all paraxial rays are converging via the single focal point
- Chromatic: a refraction index depends on the wavelength, thus bending colours of a light beam individually and consequently causing blurring.

All aberration types except the last one listed are classified as monochromatic, since they do not depend on the colour of a light beam. The monochromatic aberrations are more important in driver monitoring, since black and white cameras are mostly used so as to avoid chromatic aberrations.

The comprehensive modelling of imaging equipment requires highly complex differential formulas and in practice, it is convenient to focus on the major error sources. The error-removal methodology described in Publication I focuses specifically on removing distortion (Correia & Dinis 1998). Lenses with a short focal length provide more distortions than those with longer ones. An extreme case is a large view fish-eye lens (Shimizu et al. 1996). Shimizu et al. (1998) present a lens with a very large field-of-view, intended for robot navigation on a curved road. For this purpose, the resolution in the centre of the lens is sufficient but poor in the periphery due to high distortions. In driver monitoring applications, distortion removal is generally important due to the above-mentioned demands for a stereo vision system as well as due to the intention of using a large camera view to avoid the extra costs and complexity involved in implementing multiple cameras.

Distortion reforms the image in two ways (see Figure 7). Pincushion, also called negative distortion, expands the distance from the optical centre to the image's corner compared to the axial change. The effect of barrel (positive) distortion is the opposite to pincushion. There the horizontal and vertical locations are expanded respectively more than the pixels at the 45° angle. Distortion modifies the pixel locations around the optical axis. The problematic element is that the optical axis does not normally coincide with the centre of the lens. Therefore, the offset of the axis has to be solved before eliminating the distortion. Some methods propose mathematical formulas (Heikkilä 1997, Heikkilä, & Silven 1997, Zhuang & Roth 1996), which are added to the camera model. The alternative approach is to start by determining and compensating for the offset and then eliminating distortion.

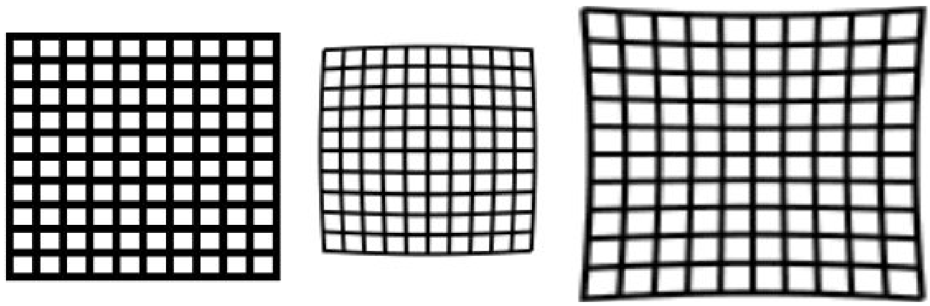


Figure 7. The left most is an undistorted image and on right are illustrated the effects of barrel and pincushion distortions.

An approximated linear correction algorithm can be created even if the exact error model is not known. In this context, linear correction means that the pixel locations are changed by determining the coefficient, the magnitude of which depends on the distance from the optical axis. That may work as a first aid for minimising the error but advanced calibration algorithms utilise high-order polynomial functions due to the non-linear nature of distortion. Probably the most famous calibration algorithm has been proposed by Tsai (Zhuang & Roth 1996). In that method, the calibration is performed in two consecutive steps, first solving the rotational and translational parameters and then the remaining ones. Weng (Zhuang & Roth 1996) also proposed calibration in multiple stages by first carrying out the rough parameter estimation and then secondly refining the result by using the first stage as an initial guess for the camera model. The same idea but with a different type of implementation is proposed by Heikkilä (1997).

Publication I presents a novel way of utilising Heikkilä's (1997) calibration proposition by exploring and varying the methodology and the application, which can adapt the calibration parameters by capturing only a single shot from the three-dimensional calibration object. The method gives the capability of removing distortion inaccuracies in the image before the stages of segmentation and target identification. Lens error removal is crucial, especially for the disparity calculation of the stereo vision system, since the quality of an image strongly affects the accuracy of the depth information. Furthermore, the optical errors deform the driver's facial features, impairing eye-tracking performance, which is an important aspect of driver monitoring.

The possibility of utilising neural networks for camera calibration was suggested by Sethuramasamyraja et al. (2003) concurrently with the work of Publication II. Both works annotated that the camera's internal parameters may cause severe errors when the image co-ordinates are mapped to a world frame. Therefore, the black box (i.e. the neural network as a camera model), which not only maps the camera co-ordinates to a global frame but also eliminates the effect of aberrations, has been investigated. Furthermore, Junghee and Choongwon (1999) have explored successfully distortion elimination with the "black-box" principle. The Sethuramasamyraja et al. (2003) system was used in guiding an autonomously moving robot. The same problems exist when faces are tracked with the camera vision technique in a driver-monitoring application. Ultimately, the idea of Publication II may also help to create an automatic calibration capability to adapt the vision system to the working environment and to camera setups.

The prototype system uses two Sony's high quality cameras where the optical errors are presumably small. Nevertheless, faceLAB includes internal calibration routine which according to the manuals fine tunes the focal length, thus representing the basics of camera calibration. The calibration method is not probably something which is pronounced by this thesis but the purpose and the idea are equivalent.

The calibration is an important thing when size and cost of the driver monitoring equipment are minimised. Discussion with the colleagues in the automotive industry has pointed out that in HGV the price level could be 1500 EUR but this is too much in a passenger car case. There the price of the whole monitoring facility should not exceed few hundred euros. Therefore, the used high quality cameras are not the optimal solution for the final implementation; rather small embedded cameras with plastic optics are preferred.

4.3 Wavelet features

Heisele's et al. (2002) study explored using a SVM classifier with the Haar wavelets to recognise the identity of a human. The method was discovered to work well in static conditions if the viewing angle was fixed (e.g. detecting people from a single image in a prior-known environment). However, the template matching of faces was found to be a better approach in a varying environment where a set of facial descriptors (e.g. eyes, nose, lips, etc.) were extracted and classified with a tree of SVM models.

Eye-tracking is the field in which wavelets have gained wide acceptance. Gu et al. (2002) have managed to increase the robustness of eye-tracking by using a Kalman filter for detecting large head movements and Gabor wavelets to achieve fast feature extraction. Retrieving the eyes from a grey-scale image is presented by D'Orazio et al. (2004), who explore a technique for tracking eyes with a combination of neural networks and wavelet descriptors.

The wavelet transformation divides the original information to low and high pass bands, thus providing an enriched number of uncorrelated attributes. The idea of using the wavelet transformation for eye tracking is attractive since it is a widely exploited method for compressing data in order to transmit images from a camera to a data processing unit. Additionally, wavelets provide an opportunity to perform tracking in parallel with the reconstruction phase of the original image (Publication III).

4.4 Facial feature extraction with colours

Naturally, the common feature to all gaze-based driver analysis techniques (Grauman et al. 2002, Bergasa et al. 2006, Grace et al. 1998, Rimini-Doering et al. 2001, Boverie et al. 2002) is the necessity to track a human's eyes. Singh and Papanikolopoulos (1999), Smith et al. (2000), Smith et al. (2003) and Wang et al. (2004) have shown how lip colour, eyes and the sides of the face can be used to track the orientation of the eyes. Lip detection is also important since it can reveal conversation with a passenger or on a mobile phone, thus signalling cognitive workload. The method was reported to provide a good tracking performance in daylight but difficulties were encountered in ambiguous lighting situations (such as night-driving). Publication V proposed a colour analysis technique that was developed for scrap-metal sorting originally but which is also applicable for recognising and tracking skin and therefore for resolving the face and eye-tracking problem. The proposed method is intended for a harsh and dirty industrial environment and adapts easily in varying lighting conditions.

Colour classification is typically sensitive to unstable lighting conditions (Publication V, Huber et al. 1998, Zrimec 2003). One option for minimizing the effect is to use a YUV colour space instead of RGB, and additionally, the errors due to non-ideal imaging devices can be reduced by using the size-invariant features (Gonzales & Woods 1993, Zrimec 2003, Stachowicz & Lemke 2002). Bagci et al. (2004) proposed Markov models for tracking and locating the driver's eyes, which were segmented by also utilising skin colour. They accentuated the method's resistance to scaling, translation and tilting of a human body. Glares due to light reflections from a road surface or eye glasses can also be minimised with the use of polarisers (Huber et al. 1998).

Another colour vision-based driver state measurement is presented by Veeraraghavan et al. (2005), who reported comparable results for implementing unsupervised (amount of body limb movements) and supervised learning processes (the Bayesian eigen-image analysis). In their experiments, the driver's activity was analysed by counting movements of the head and hands that were segmented according to skin colour.

4.5 Driver and driving-related parameters

Hoedemaeker et al. (2002) have identified that carmakers and research institutes interested in driver monitoring are doubtful whether non-intrusive measuring methods will succeed. Rather they prefer to estimate the workload according to the level of activity in using the vehicle controls and estimating their influence on driving (speed variation, headway to the front vehicle, etc.) or by generating a 'lookup table' in terms of factors like age, gender, road geometry, etc. Tattegrain et al. (2005) give comprehensive high-level guidelines for monitoring a driver and environment, including indications related to a driver's static characteristics (e.g. age, sex, etc.), dynamic behaviour and actual traffic context. This thesis neglects the driver's static parameters since distraction and fatigue—the key elements of this thesis—have a dynamic nature.

Table 1 summarises the review of prior knowledge on the subject, and is more detailed than that given in Chapter 1, Publication VI and Publication VII. As the table indicates, many different types of features exist and are being experimented with to detect distraction or fatigue in a driver or an aircraft pilot.

This thesis has selected the most appropriate features for the distraction detection experiments, including head and eye movements and lane-keeping analyses. Their relevance to the objectives of this study was explored in Publication VI.

Table 1. A review of the proposed driver state measures in the literature. The summary is divided into those addressing distraction and those related to fatigue/vigilance detection.

Parameter	Distraction detection	Vigilance detection
Lane keeping	Engström et al. 2005 Fletcher et al. 2001 and 2003 Horrey & Wickens 2004 McCall & Trivedi 2004 Östlund et al. 2004	Bittner et al. 2000 Boverie et al. 2002 Boverie 2004 Fletcher et al. 2001 and 2003 Grace et al. 1998 Pilutti & Ulsoy 1995 Rimini-Doering et al. 2001 Santana Diaz et al. 2002
Vehicle headway	McCall & Trivedi 2004 Östlund et al. 2004	
Vehicle speed	Engström et al. 2005 Östlund et al. 2004	Boverie et al. 2002 Santana Diaz et al. 2002
Accelerations		Grace et al. 1998
Steering wheel movements	McCall & Trivedi 2004	Bittner et al. 2000 Boverie 2004 Boverie et al. 2002 Pilutti & Ulsoy 1995 Santana Diaz et al. 2002
Pedal movements	McCall & Trivedi 2004	
PERCLOS	Dinges et al. 1998	Bergasa et al. 2006 Grace et al. 1998 Grauman et al. 2002
Eye-blinking frequency	Albery et al. 1987 East et al. 2002 O'Brien 1988	Bergasa et al. 2006 Boverie et al. 2002 Dinges et al. 1998 Heitmann et al. 2001

Eye movements	Engström et al. 2005 Hammel et al. 2002 Harbluk et al. 2002 Heitmann et al. 2001 Lee et al. 2004 Recarte & Numes 2003 Victor et al. 2005	Bergasa et al. 2006 Boverie 2004 Rimini-Doering et al. 2001 Wahlstrom et al. 2003 Wang et al. 2003 Zhu & Ji 2004
Head movements		Bergasa et al. 2006 Boverie 2004 Grace et al. 1998 Grauman et al. 2002 Heitmann et al. 2001 Zhu & Ji 2004
Limb movements	Albery et al. 1987	Boverie 2004
Pupillary changes		Heitmann et al. 2001 Wang et al. 2003
Heart rate	Albery et al. 1987 East et al. 2002 Östlund et al. 2004	
Blood pressure	Albery et al. 1987	
EEG / EOG	East et al. 2002 O'Brien 1988	Bittner et al. 2000 Gonzalez-Mendoza et al. 2003 Lal et al. 2003
Breathing	East et al. 2002	
Skin conductance	Östlund et al. 2004	
Temperature of environment		Zhu & Ji 2004
Sleep history		Zhu & Ji 2004
Driving environment	Fletcher et al. 2005	Fletcher et al. 2005

5. Classification Methods

5.1 Overview

This chapter proposes techniques for detecting a driver's momentary distraction level by using a syntactic, support vector machine or neural-network-type classifier. Details of the techniques are more comprehensively discussed in Publications II, V, VI and VII. The topic of Publication VII is the feasibility of a SVMlight algorithm (Joachims 1999) for detecting the cognitive distraction of a driver. Promising results for recognising artificially induced cognitive workload during real driving have been presented, which is scientifically revolutionary. Using an SVM-type pattern recognition method is an especially new idea in the field of optical driver monitoring.

Publications VI and VII describe the practical results of the monitoring experiments. This chapter provides the experiments of visual distraction detection with a syntactic classifier and the complementary results for detecting cognitive distraction in a passenger car. The last topic is a proposition of using neural networks for distraction/vigilance detection and semi-automatic attention mapping, which is important for detecting visual distraction. The method is based on earlier implementations of neural networks for automatic object mapping (Publication II).

5.2 Visual distraction detection with syntactic classifier

A rule-based 'keep it simple' idea in many cases works more robustly than the smart classification methods (smart referring to e.g. neural networks, Bayesian networks, SVM, etc.). Publication V describes the methodology for sorting scrap metals (copper, brass, aluminium) according to their colour attributes. The classifier used is syntactic and it addresses whether the features fit to the tolerances of the pre-defined colour models for the metals. The colour classification example has also promoted the importance of user-friendly tuning facilities to provide an optimal sorting capability. Publication VI discusses attention mapping and detecting visually distracting occurrences. In this case, the term visual distraction means whether the driver focuses his or her attention towards a road or other attraction (e.g. vehicle controls, a mobile phone, radio, etc.). Additionally, the term visual distraction includes in this case visual time-sharing, which is an indication that the driver is continuously making short glances off the road, hence sharing his/her attention between two targets (e.g. short glances towards mirrors). In the practical tests four different clusters were implemented in the prototype: road ahead, windscreen, left- and right mirror (see Figure 8). Optionally, the additional clusters (e.g. radio) could be implemented easily but were not necessary since the aim was focused on detecting eyes-off-road and time-sharing between mirrors and road. Figure 3 in Publication VII shows the architecture of the developed module that was used in testing the visual distraction detection.

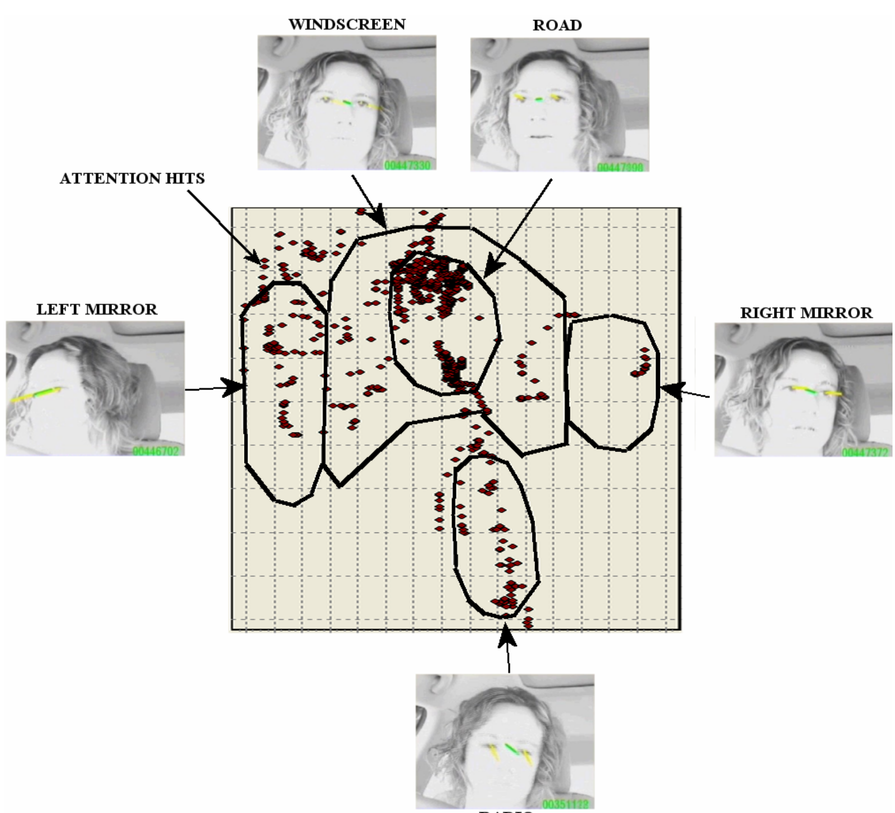


Figure 8. Attention mapping in a (SEAT) demonstration vehicle. In the tests, the radio cluster was not used, but it was captured for future needs. Note that the cockpit views are mirrored, thus when the driver looks to the left mirror it seems like a right mirror check and vice versa.

The determined clusters are results of iterative boundary tuning, which in the end is a compromise to take into account different driving habits. Therefore, the main benefit of using the classification algorithm presented in Publication V is more flexible adaptation (see Figure 9) compared to the one presented in Publication VI, where the clusters are estimated by using circles and counting the distribution of the driver's momentary glances. The idea of an adaptation facility is presented for the first time in Publication V and there is the national patent (Vattulainen et al. 2002) pending in connection with the proposed classifier adaptation method. The most innovative element is the user interface, which provides direct feedback for analysing false results.

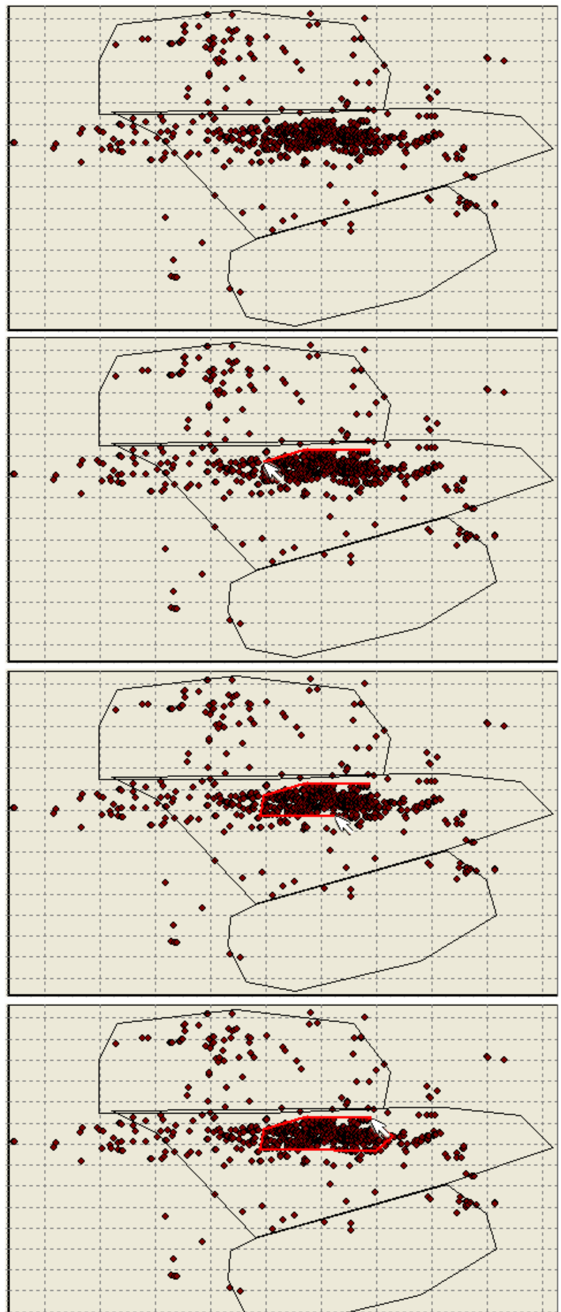


Figure 9. Definition of the road-ahead cluster for the rule-based (syntactic) classifier. Firstly the training examples are gathered and secondly the borders of the clusters are determined by dragging lines according to the hits.

Optimising the clusters requires good coverage of different drivers and the test environment, since each driver's behaviour is extremely individual. The generated clusters can be mathematically held as an average of the captured attention angles and therefore a lot of training data is needed to avoid statistical errors and consequently over-adaptation to a single driver. In practice, the adaptation is performed by extracting from the available data files a random sample of approx. 5000 hits per cluster. The test data were gathered in Sweden with a test HGV. The summary of the test conditions and subjects can be found in Table 2. The test drivers D1 and D2 are ignored when training and evaluating the visual distraction detection, since the eye tracking was not optimal due to an erroneous camera installation. The same test data, including also the drivers D1 and D2, have also been partially used in the SVM adaptation for the HGV case presented in the next chapter.

Table 2. Summary of the gathered test subjects for collecting the HGV data which are used in the evaluations of this thesis.

Driver ID	Date	Time	Sex	Age	Years with truck driving license	Years of work experience as professional truck driver	Normal freight type (Distribution, Long haul, Both)	Experienced with driving Volvo trucks	Experienced of driving with I-shift gear box	Route	Driving with semi trailer
D1	25.4.2005	12.00	M	57	33	30	L	x		1	
D2	25.4.2005	16.00	M	59	39	39	D	x	x	2	
D3	25.4.2005	19.30	F	37	3	4	D	x	x	2	
D4	26.4.2005	09.00	M	41	22	8	L	x		2	
D5	27.4.2005	09.00	M	27	4	4	B	x		1	X
D6	27.4.2005	19.30	M	24	3	3	D			2	X
D7	28.4.2005	19.30	M	44	18	19	D			1	X
D8	29.4.2005	09.00	M	57	37	20	B			2	X
D9	29.4.2005	16.00	M	45	18	18	L	x		2	X
D10	2.5.2005	13.00	M	45	27	26	L	x	x	2	X
D11	2.5.2005	17.00	M	22	4	3	D			2	X
D12	3.5.2005	18.30	M	21	3	2	L	x		1	X

The new classification scheme presented shortly by Publication VI is syntactic (i.e. rule based) since visual distraction is considered to have resulted whenever the driver's attention is outside the road ahead area. However, looking to the periphery (windscreen area but not on estimated road) and mirror checks are also detected by the developed algorithm. The test results announced in Publication VI are performed using an older version of the algorithm developed for the Volvo HGV. Outlines of the completed tests for the current algorithm are given in Table 3 since they are not reported in the original publications. Results are also compared between the test subjects in Table 4. The clusters are well detected except for the windscreen, which appears to be a problem. This is partially caused by the evaluation method being executed manually with counting hits and comparing the observations to the appropriate video. The glances towards the area between a mirror and road ahead are short and hence, hard to observe. The main conclusion is that the road ahead cluster (i.e. eyes-on-road) are well detected and the mirrors moderately so. Therefore, the total visual distraction detection (eyes-off-road and visual time-sharing) algorithms performs well. Table 4 also outlines the total hit rate as a reference for the eyes-off-road detection. There is not big difference between the drivers in performance if the eye tracking operates well. The improper rates of drivers 10 and 11 in Table 4 are mainly due to insufficient eye tracking rather than the distraction detection algorithm. However, the road ahead glances are very well detected (> 90%) in the HGV, which is very important for proper estimation of the visual workload (i.e. inattention to the road events).

Table 3. The performance of the current attention mapping algorithm. The driver refers to the test subject. During the tests the cockpit model was re-adapted, which improved discrimination of the road ahead and windscreen clusters.

DRIVER	TEST ID	ROAD AHEAD	LEFT MIRROR	RIGHT MIRROR	WINDSCREEN
D3	1	98%	32%	54%	8%
D3	2	100%	42%	67%	15%
D3	3	87%	31%	-	6%
D4	1	91%	21%	86%	7%
D4	2	91%	33%	46%	13%
D4	3	100%	26%	-	0%
D5	2	100%	21%	31%	7%
D5	3	100%	18%	29%	7%
D6	1	100%	71%	74%	2%
D6	2	98%	63%	76%	9%
D6	3	97%	68%	56%	0%
D7	1	85%	61%	14%	8%
D7	2	94%	51%	0%	16%
D7	3	100%	3%	-	12%
D8	1	100%	0%	-	21%
D8	2	98%	8%	53%	20%
D8	3	99%	6%	75%	0%
COCKPIT MODEL RE-ADAPTED					
D6	1	80%	51%	35%	40%
D6	2	91%	51%	13%	33%
D6	3	75%	61%	21%	30%
D9	1	100%	62%	53%	62%
D9	2	89%	36%	65%	61%
D9	3	74%	71%	-	22%
D10	1	44%	42%	34%	64%
D10	2	46%	19%	48%	46%
D10	3	59%	33%	31%	59%
D11	1	48%	62%	76%	43%
D12	1	69%	46%	10%	27%
D12	2	96%	60%	19%	45%
D12	3	63%	67%	44%	33%

Table 4. The average attention mapping hit rates per driver. The column "MODEL" refers to the fine-tuned SVM model while the Volvo HGV tests.

DRIVER	MODEL	ROAD AHEAD	TOTAL HIT RATE
D3	OLD	94,65%	49%
D4	OLD	93,88%	47%
D5	OLD	100,00%	41%
D6	NEW	82%	48%
D7	OLD	93%	48%
D8	OLD	99%	48%
D9	NEW	88%	63%
D10	NEW	50%	45%
D11	NEW	48%	57%
D12	NEW	76%	48%

The tests were also performed with data acquired by SEAT with a passenger car. Coverage of the different types of drivers is not as exhaustive as in the HGV case. The test data was gathered for three ordinary drivers who had approx. 5–10 years driving experience in Spain. The test data included motorway and city driving, undertaken during the day time. The test samples were shorter than in the case of the Volvo HGV, as they included only a few hundred samples for adapting each cluster.

Figure 10 and Figure 11 show differences between the SEAT car and the Volvo HGV cluster dimensions. Obviously, the driver's eye movements are larger in the HGV case as can be seen from the pictures. The horizontal and vertical axes of the graphs are the orientations around the X- and Y-axis respectively. The orientations of the axis and the imaging geometry of the test arrangements are shown in Figure 12. The experiments have showed that large gaze and head movements degrade the attention mapping performance. Thus, it is anticipated that this algorithm works even better in a passenger car, where glances towards mirrors are smaller and the driver's faces remain more in the camera view, which was one of the problems in the HGV tests. The initial test results with the passenger car support the assumption since they have provided slightly better results than the HGV evaluations. On the other hand, the problem according to the first experiments is the small mirror cluster in the car test compared to the HGV, which makes the sensitivity of the system more significant.

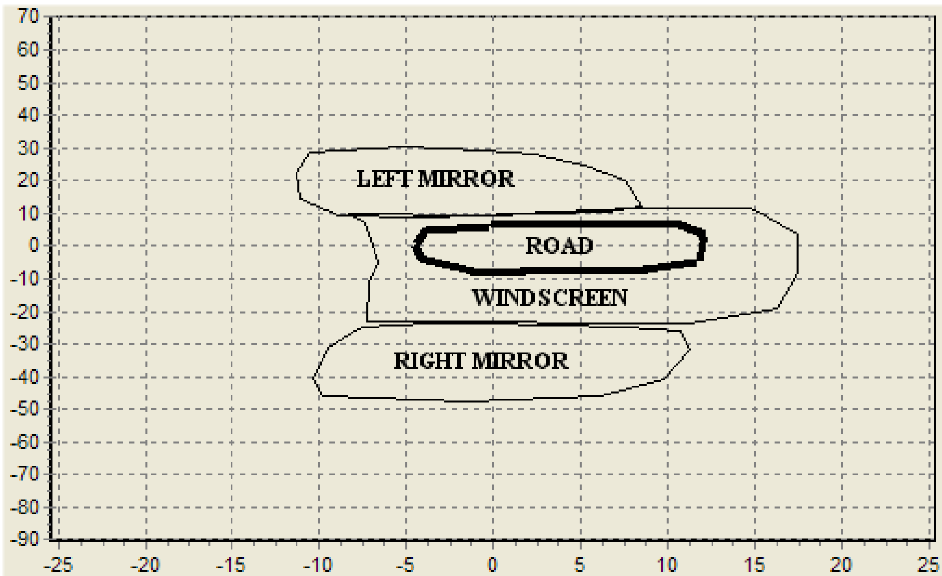


Figure 10. Attention clusters of the passenger car (SEAT Leon) implementation.

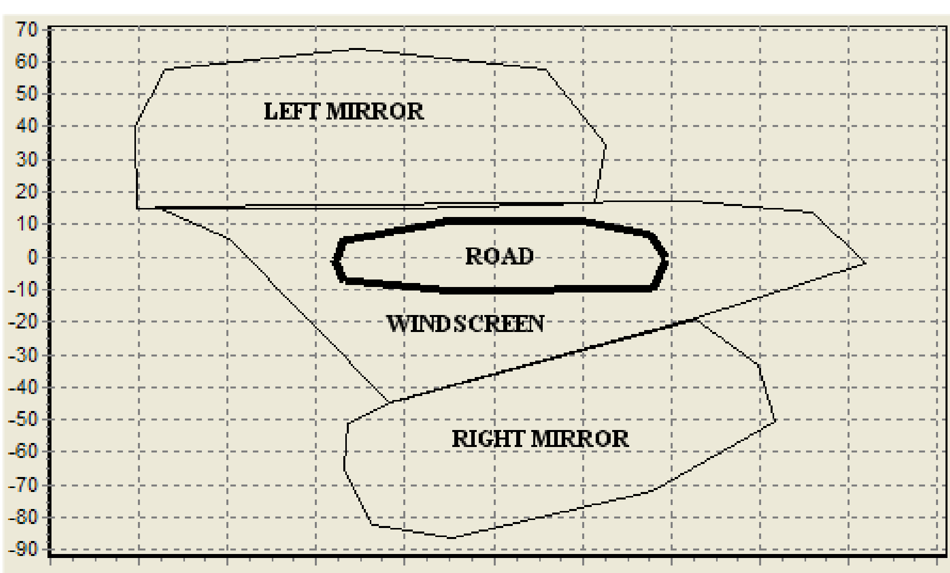


Figure 11. The attention clusters in the HGV prototype (Volvo FH12).

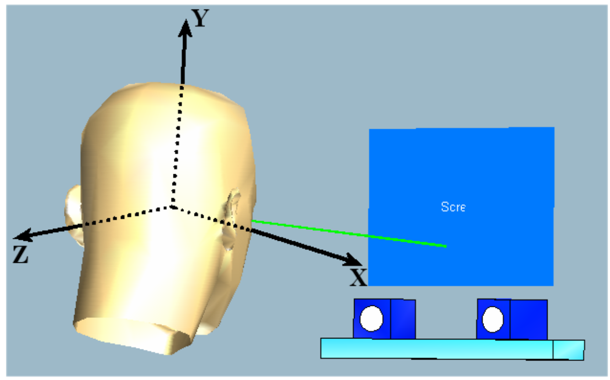


Figure 12. Co-ordination system of the stereo vision system.

The one aspect, which was also investigated when developing the attention mapping module, was opportunity to adjust the clusters on a fly. This work is still under construction since the first idea of adapting the location of the clusters according to road curve did not work as assumed. The test subjects did not react to the environmental changes equally and therefore the mapping performance unexpectedly degraded with 3% (Jokela 2006).

5.3 Cognitive distraction detection with SVM

Support vector machines have received a lot of attention recently due to their generalisation capability for processing different types of data. Therefore, multi-discipline exemplary applications have been introduced where SVM has provided good classification capability. Such examples are Sebastiani (2002), who has introduced comparative tests for classifying text to pre-defined categories with the use of different classifiers (Bayesian, SVM, neural networks, regression analysis, rule-based, model based and linear). The conclusion was that support vector machines, model-based methods and regression boundaries give better results than the currently favoured neural networks. Bellotti et al. (2004) have used the SVM classifier for detecting pedestrians on a road. Interestingly, the application utilised wavelet descriptors to distinguish whether a human exists in an image or not. Xia et al. (2003) have developed a methodology for eye-tracking in situations of low illumination by using the IR-band lighting and verifying the existence of eyes with SVM.

The SVM kernel functions can be divided to linear or non-linear ones (Burges 1998). The following is a simple canonical dot product and is formulated as:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) \tag{1}$$

Above $\mathbf{x}_i$ and $\mathbf{x}_j$ are pattern vectors and Φ mapping function for constructing linearly discriminated feature space. Since the kernel is rather simple, the boundary between two data sets is easy to determine. However, the linear kernel is not useful in the case of non-linear data. The non-linear kernel functions (sigmoid, RBF or polynomial) provide better adaptation properties for non-linearly separable data and are thus mostly preferred. The RBF kernel function is:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2}{2\sigma^2}\right) \tag{2}$$

where the parameter $\boldsymbol{\sigma}$ determines the width (i.e. coverage) of the kernel.

Due to the aforementioned arguments, an RBF kernel is selected for the cognitive distraction detection application presented in Publication VII, since the features are strongly non-linearly distributed according to the observation from Figure 13. On the right side of the image, the variations of gaze rotations and head rotations are shown in the 2D plane. The boundary between positive and negative results is circular rather than a straight line.

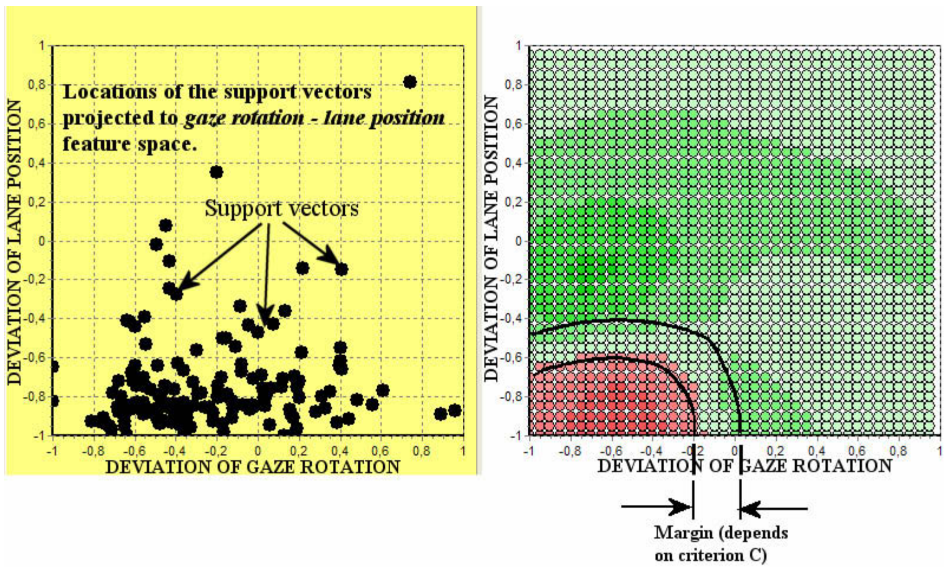


Figure 13. Capture from the SVM tuning tool. The graphs show an example for generating the SVM model for professional HGV drivers. The gaze rotation and lane position deviation features are projected to the average "level" of the other implemented features. The cognitive area is located at the left-bottom corner, where the lane keeping and gaze concentrations are high.

The idea of the SVM algorithm is to maximize the margin between two clusters. This is done by minimising the training set error in terms of α, which are the Lagrange multipliers (Bennett & Campbell 2000):

$$\min\left[\frac{1}{2}\sum_{i,j}^{m}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i,\mathbf{x}_j)-\sum_{i=1}^{m}\alpha_i\right] \tag{3}$$

subject to $\alpha_i \geq 0$, $i = 1, \ldots, m$

and $\sum_{i=1}^{m}\alpha_i y_i = 0$

Solving the above optimisation problem provides the border that maximises the distance between positive and negative clusters. The classification function is (Bennett & Campbell 2000):

$$f(\mathbf{x}) = sign\left(\sum_{i=1}^{m} y_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) - b\right) \quad (4)$$

where b is the primal threshold between negative and positive clusters.

The SVM algorithm is easy to understand especially when considering the 2-dimensional feature space. During the training phase, the border is adapted to meet the requirement of the margin between positive and negative training examples. As Publication VII introduces, with proper training tools, an adaptation of an algorithm is quick to perform and minimises the danger of ending up at the local minima's. The problem can be seen in the passenger-car tests (Table 5) where the gamma, which adjusts the "complexity" of the border and thus is the main reason for the over-fitting problem, is increased. With low gamma values the performance rates are fairly equal but with larger values there appears a disparity between the two tests. This anticipates a lack of generalisation due to too strong an adaptation within the training data.

Table 5. Effect of the gamma parameter for the behaviour of the classifier in the passenger-car tests. Bigger gamma increases the risk of over-fitting to the training data, which is observable also by exploring the deviation between the two tests.

	TEST 1		TEST 2		
GAMMA	NON-COGNITIVE	COGNITIVE	NON-COGNITIVE	COGNITIVE	DEVIATION
1,7	84	89	85	83	2,8
2,0	83	88	85	83	2,2
3,0	87	86	85	83	1,8
6,5	84	86	79	83	2,7
6,8	84	86	77	83	4,1

The intention of Publication VII was not to survey the principles of SVM deeply, but rather to focus on testing the feasibility of the classifier for driver activity monitoring. The classifier's advantages are apparent when several different features are used for recognising two different categories. The tests are performed by using lane-keeping performance and features related to gaze and head variations. Quality factors were also calculated, which addressed the coverage of the data in a time window. The low coverage is typically caused by large head movements, which on the one hand can be attributed to the non-cognitive state of a driver (i.e. active driving).

Since preparing Publication VII, more tests have been performed where the cognitive distraction detection algorithms were adapted to a passenger car, as Publication VII focused on tests performed with a HGV prototype. The test car did not include the lane tracker for measuring improved lane-keeping performance, which according to the studies McCall and Trivedi (2004), Fletcher et al. (2003) and Östlund et al. (2004) are due to increased cognitive workload. Table 6 provides workload detection performances in the passenger-car prototype for different configurations of the available features. The table still supports the assumption as explored for HGV in Publication VII that all the indicative parameters should be utilised when the cognitive distraction is detected.

Table 6. Feature selection in a passenger car.

Gaze Rotation	Gaze Rotation Quality	Head Rotation	Head Rotation Quality	Face Model Quality	Non-cognitive	Cognitive
X		X			47%	89%
X	X			X	59%	71%
		X	X	X	51%	84%
X	X	X	X		44%	89%
X	X	X		X	36%	93%
X	X	X	X	X	56%	79%

Further testing with the passenger car has progressed successfully and the results have been even better than in the HGV case, despite the absence of the lane tracker. With proper adaptation, an 87% detection performance was achieved for baseline driving and an 85% rate for detecting cognitive tasks. The test data included four different drivers who were driving in various conditions in the city and on motorways. The cognitive tasks were artificially induced. Part of the explanation for the improved results is that the eye tracking was better in the passenger car tests than in the HGV tests and another that the passenger car tests did not include for example tunnels driving or dark conditions at all.

The further field tests with HGV indicated that the cognitive workload is not worth of detection in cities since driver's alertness remains high due to a rapidly varying environment. Hence in the final implementation, the vehicle speed is read from the CAN bus and the cognitive distraction detection is paused when vehicle speed is below 50 km/h. This notion increased the performance of the algorithm significantly.

5.4 Discussion of neural networks for driver monitoring

5.4.1 Distraction / vigilance monitoring

Throughout the 1990s, laboratory work was widely carried out to adapt neural networks to practical applications. Today neural networks have been successfully examined in biotechnology, pattern recognition, speech analysis, and has also been experimentally applied in driver monitoring (Wang et al. 2003). Neural networks have similarities to the learning process of a human brain, which also constitutes linked neurons (Haykin 1999). Unfortunately, computers can only feasibly process simple mathematical formulas. Humans or animals have millions of neurons whereas in practical artificial networks only a few hundreds may exist, thus limiting the ability to adapt to a changing operating environment. Andreeva et al. (2004) introduced an implementation in which they recognised the drowsiness of a driver based on body posture measurements. The position of the human upper body was measured three dimensionally and the data was fed to a neural network. Fukumi (2005) presented a template-matching scheme with neural networks for detecting whether a driver's head existed in an image or not. The data source was a low resolution NIR camera (320 x 240 pixels). Additionally, a simplified PCA algorithm was implemented for capturing the driver's head orientations.

Information processing is the result of interaction between the neurons. A single neuron is an element, which includes an activation function for determining the response for the induced input. During the training phase, an error between the real outputs and the network's prediction is minimized by adjusting the weight factors and bias terms gradually. Different types of network model have been introduced in the literature and each of them has benefits and drawbacks depending on the target application. The Multilayer Perceptions (MLP) and Radial Bases Function (RBF) networks are the most widely used since they are a good composite of flexibility and adaptation capabilities (Publication II).

Prior studies (Andreeva et al. 2004, East et al. 2002, Fukumi 2005, Grace et al. 1998, Wang et al. 2003) have predicted the opportunities afforded by neural networks, especially in the field of drowsiness (or fatigue) detection. Fusing the data of multiple sources is the strength of neural networks, since in the driver monitoring case even the indicators of impaired vigilance may be weak (Publication VI). It is never desirable that a driver falls asleep; rather it is intended to perform wake-up stimuli when slight fatigue is present. Publication II discusses more about neural networks, which in that case however are pursued with the intention of automatically mapping the targets in the image frame co-ordinates to the world frame. Nevertheless, the same steps exist whenever neural networks are implemented: starting with selecting the proper network model, then training the network and finally validating performance.

Figure 14 illustrates the idea of using neural networks for detecting a cognitive (mental) distraction level or the fatigue of a driver. The method is quite similar to that presented in Publication VII, although there the support vector machine algorithm is examined. The benefit of the neural networks is that the fatigue and cognitive distraction analysis can be performed as parallel processes, while SVM on the other hand is a binary-type classifier, thus providing only one output. Of course, there might be multiple SVM models, but then the results would have to take into account whether the driving behaviour refers more to fatigue or to cognitive distraction, while neural networks could exclude the more unlikely result automatically.

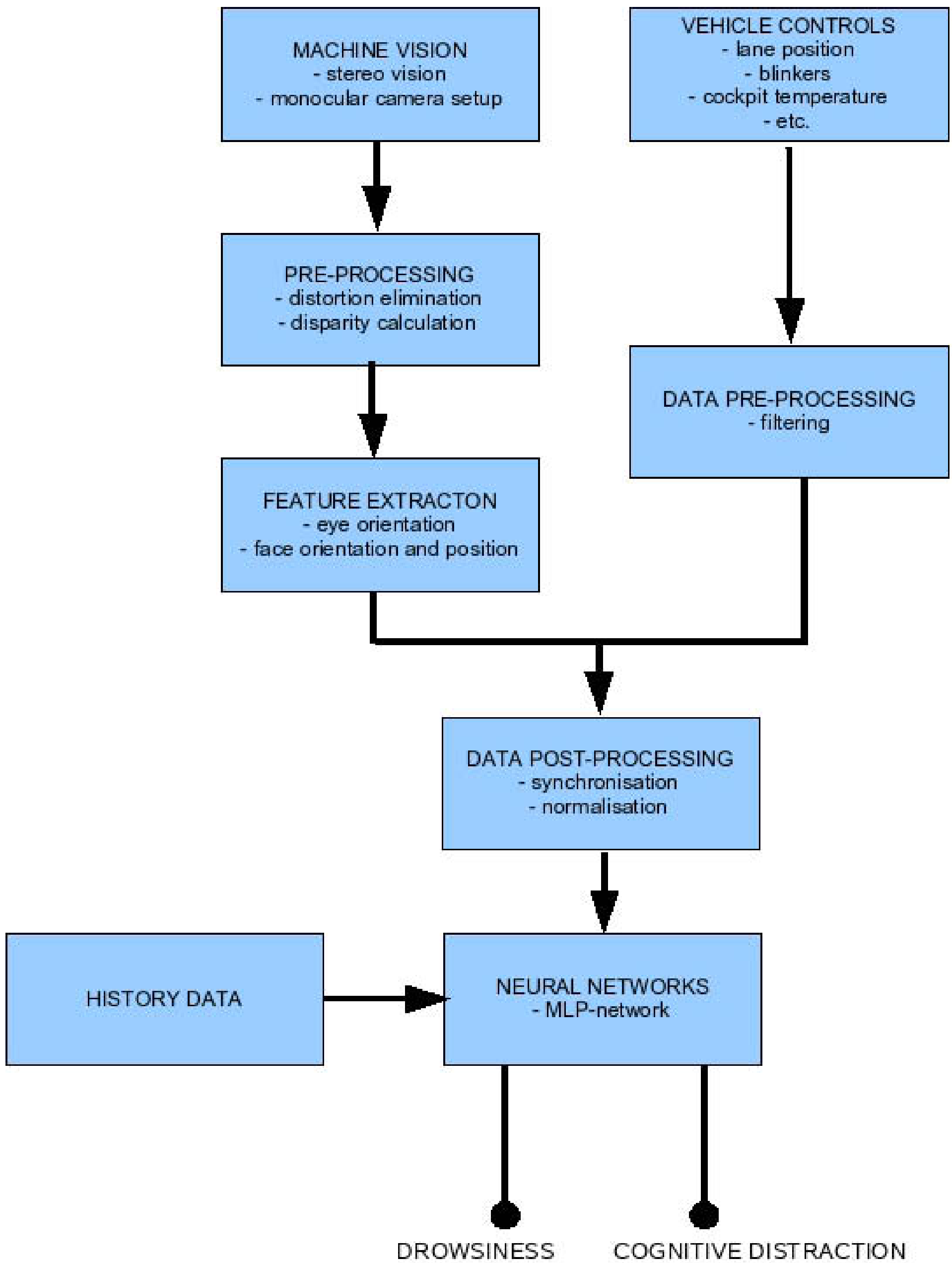


Figure 14. The proposed method for performing driver distraction or fatigue detection with neural networks.

5.4.2 Neural networks for attention mapping

Accurate mapping of the targets from an image frame to world-frame co-ordinates has always been one of the key elements in building a machine vision system. Luan et al. (2001) introduce a Photonic Mixer Device (PMD) with a time-of-light (TOF) principle for performing three-dimensional measurements. They used standard CMOS or CCD chips combined with optical phase-shifting elements and still managed to keep the price-level down (< 800 EUR) compared to conventional matrix-type 3D measurement applications (> 1500 EUR). Chiang and Huang (2004) proposed using MLP neural networks to detect the gaze angles of a car driver so as to determine whether the driver is looking upward, downward, left, right or straight ahead.

Chapter 5.2 (Visual distraction detection with syntactic classifier) and Publication VI discusses attention mapping in the cockpit of a vehicle in more detail, since identifying the momentary attention targets is important for detecting visual distraction. The proposed algorithm is based on pre-determined attention clusters, which eventually, could be dynamically adapted while driving. Ultimately, momentarily looking towards in-vehicle buttons is then captured by an application that monitors driver behaviour. An automatic adaptation capability is preferred since the driver’s behaviour is dynamic and an attribute of the driver’s characteristics. Publication II presents the method for mapping the world-frame co-ordinates from the image frame by means of a neural network. The same methodology could potentially be used for mapping the attention targets in a cockpit with a more advanced adaptation capability, thus minimising false mappings due to changes in driver behaviour. The procedure could progress such that the training is automatically restarted when the gaze or head orientations locate to a predefined large area of a mirror or a radio, then during fine-tuning, the area is minimised according to the driver’s orientation history. Modifying the clusters of the rule-based algorithm (Publication VI and Chapter 5) is somewhat time-consuming due to manual fine-tuning despite of the presented easy adaptation facility. Therefore, the method explored in Publication II would provide more on-the-fly automation to the attention mapping, which is essential when thinking about the reliability and robustness of an in-vehicle product.

6. Description of Original Publications and Author's Contributions

Publication I, Camera Calibration in Machine Automation

The publication presents a methodology and experimental results for eliminating distortions from an image caused by an imperfect optic. The calibration parameters are calculated by capturing a single shot from the compact calibration object. This is a major improvement developed by the author compared to existing calibration descriptions (Heikkilä 1997 and Zhuang & Roth 1996). The method can be generalised to be feasible for cameras operating in visible or near-infrared wavelengths. The distortion elimination method is necessary when low-cost camera vision systems are considered; for example, the system used in the tests of this thesis for examining the driver monitoring application costs approx. 35 000 EUR.

The author performed the tests presented in Publication I, designed the software and reviewed the feasibility of the Heikkilä's (1997) calibration method. The idea of utilising the presented calibration object was given by the co-authors.

Publication II, Calibration of the World Coordinate System with Neural Networks

The author has independently developed and tested the idea of using the MLP-type neural network for target mapping. The idea is to semi-automatically map image frame co-ordinates to a world frame without knowing the imaging geometry and ultimately, taking into account possible lens errors within the "black-box" camera model. The idea had been earlier suggested by Berthouze et al. (1996), while utilising two different artificial networks and interpolating the network models between them. The author of this study is not convinced that the models behave as straightforwardly as was suggested. Publication II proposes using only one network and to train that on multiple levels, which then adapt the parameters for different distances automatically.

In many environments, an accurate definition of the position of a camera is difficult to measure, as in a vehicle. However, mapping the attention objects (e.g. road, mirrors, radio, buttons), which are needed for estimating the visual distraction level, require easy and preferably automatic adaptation to cockpits. In this, the introduced methodology may assist in determining the locations of the attention clusters. However, the presented method was not tested experimentally in the driver monitoring applications directly.

Publication III, Parallel Image Compression and Analysis with Wavelets

The publication presents an image compression method that was applied for the first time to a confidential industrial application. Data compression is important whenever video is transmitted between computing or display units. The bandwidth of the in-vehicle buses is limited and therefore, the requirement for efficient compression is a prerequisite for transmitting images. Moreover, the paper discusses wavelets generally, which are interesting in relation to eye tracking (Heisele et al. 2002, Gu et al. 2002, D'Orazio et al. 2004) and in turn, is a basic requirement in optical fatigue or distraction detection.

The wavelet-based compression methodology and the use of low- and high- pass bands for feature extraction were designed in co-operation with the co-author Prof. Jouko Viitanen. The main author has partially built up the software and performed the experiments with the algorithm in a laboratory. The crucial improvement on the earlier work presented by Hilton et al. (1994), Manduca (1995), Welsted (1999) and Zixiang et al. (1999) is the computationally light wavelet-based compression implementation, which furthermore, is able to perform parallel image analysis (e.g. tracking eyes). While hundreds of studies dealing with wavelet-based image compression or analysis exist, not many deal with merging the two to provide parallel analysis and compression.

Publication IV, Sensor Array for Multiple Emission Gas Measurements

This publication presents a sensor that is intended for gas sensing in an industrial kitchen. The article assists to understand the basic requirements for a proper data acquisition device. The parallels with camera vision are not immediately obvious, but there are a lot of similarities, such as converting the analogue signal to digital, calibrating the sensor, and ensuring its adaptability to the existing environment. One crucial conclusion is that calibration has to be performed in the practical conditions, which was a feature of later experimentation in regard to the driver-monitoring applications (Publications VI and VII). Fine-tuning of the calibration is required for each new driver and therefore it should be easy and quick to execute, which in itself is an improvement on Marco et al. (1998). The calibration capability is even more important in achieving low-cost measuring devices, since many times the sufficient output needs to be achieved with a lower price if adaptation to the dynamic in-vehicle conditions is to be sensibly applied, as Publication IV shows.

The author has built the software, executed tests and calibrated the sensors. Prof. Jouko Viitanen mainly participated by designing the electronics.

Publication V, Scrap Metal Sorting with Colour Vision and Inductive Sensor Array

This paper presents scrap-metal sorting by means of a colour classification technique and is accomplished by a sensor fusion of colour information and the electrical properties of a metal. The experimental results were performed by means of a scrap-metal sorting device. The colour classification could potentially be used to track driver's limbs, face and eyes (Singh & Papanikolopoulos 1999, Smith et al. 2000, Smith et al. 2003, Wang et al. 2004, Veeraraghavan et al. 2005), thus providing an indication of the driver's activity level. However, the most essential input is the syntactic classification procedure, which is part of the experimentation in attention mapping to detect visual distraction. The interface—consistent with the classifier in Publication V—for performing flexible adaptation of the clusters is illustrated in Figure 9 and is the major achievement of the author in Publication V.

The author of this thesis has designed the lighting arrangement, built the programs, tuned the parameters, planned the sensor fusion implementation and assisted in the tests. The co-authors have participated in designing the system and contributing to the commercial aspects of the sorting device.

Publication VI, Online Detection of Driver Distraction – Preliminary Results from the AIDE Project

This publication presents some early results from the AIDE project, an overview of the data gathering, and an analysis of the relevant features for distraction detection. The visual distraction detection methodology introduced by the paper is developed with colleagues at the Volvo Technology Corporation. The presented algorithms were developed after publishing the paper; however, the earlier algorithms form the base for later developments made by the author. The results for the finalised visual distraction detection algorithms are presented in Chapter 5, along with an important statistical analysis of most of the features that were able to function as indicative parameters. The design of the distraction estimation procedure (the CAA module) has been compiled in co-operation with colleagues at Volvo and is a very new approach, that is to the authors' best knowledge not to be found elsewhere in the scientific literature. Utilising a support vector machine to detect cognitive distraction is mainly a design feature developed by the author that has also been implemented and partially field tested.

Publication VII, Driver Cognitive Distraction Detection: Feature Estimation and Implementation

This journal article is a multi-disciplinary review of the features and techniques for and effects of detecting cognitive distraction. Publication VI, Publication VII and this summary section bind all the other publications under the same rubric, the topic of which is driver monitoring. The article includes a literature review and tests for relevant features and describes the experiments performed for a support vector machine -type classifier. The results of the statistical analysis of Publication VI are utilised in the presented SVM implementation. The article also contains a fundamental analysis of how cognitive distraction affects human behaviour, which has mainly been contributed by colleagues at Volvo Technology Corporation. The presented technique for detecting cognitive distraction is new and is an improvement on the state-of-the-art work in this field. The author has planned the cognitive distraction detection technique, the SVM implementation, as well as performed the reported tests.

7. Conclusions and Future Work

The topics of this study cover all the items sketched in the data flow illustration in Figure 3. Thus this thesis provides a multi-disciplinary overview for building machine vision applications for driver monitoring. This thesis has outlined discoveries and put forward propositions and experimental results related to the following topics:

- Experiments for detecting cognitive distraction in a driver by utilising a support vector machine -type classification procedure
- Experimental results for detecting a driver's visual distraction by using a rule-based (syntactic) classification method
- Relevant features for detecting the driver's momentary state
- Requirements of data acquisition devices and data transmission techniques.

Additionally, this thesis includes discussion on the following topics that are applicable to driver-monitoring systems, but are not tested in the human monitoring field. They are intended more to open up novel aspects for detecting the driver's state.

- Neural networks for distraction or fatigue detection
- Feasibility of colour and wavelet descriptors for eye tracking and activity-level assessment of a driver
- Proposition for semi-automatic attention target mapping, which is needed in visual distraction detection.

Driver monitoring is a relatively young research field. The first studies were published at the end of the 1990s, but most activities have been carried out during the last 5 years. A lot of effort has been directed at fatigue detection, though few (~5) commercial products have been launched to date. Primarily, the optical measurement principal (PERCLOS) has been adopted to detect drowsiness. Distraction detection is a parallel sector, though it is not as widely investigated. Commercial products do not exist at the moment but a few patents have been applied for, which relate more to measurement arrangements than to methods. Therefore, the technical propositions of Publications VI and VII for visual and cognitive distraction detection are rather novel and innovative. Moreover, existing know-how focuses on a limited part of the monitoring system and do not fully consider the whole data processing concept (e.g. neglecting errors due to optics or insufficient eye-tracking): However, this concept is a fundamental premise of this thesis.

The developed algorithms already operate to a satisfactory level. The significance of the achieved results for the future is not easy to predict but at least the automotive industry, which is involved in the AIDE project, has expressed a wide interest in the proposed optical driver monitoring scheme of this thesis. My personal assumption is that this thesis will be published at an opportune moment. The last 5 years, the automotive industry has initiated their R&D departments to work in the optical driver-monitoring field and many breakthroughs have been achieved at laboratory level. It could be said that understanding a driver's behaviour has proceeded more intensively than the technical development of monitoring facilities that can measure and understand the driver's state and furthermore, take into account the size and price constrains of in-vehicle systems. I claim that this is one of the first attempts to bridge this gap between the explorations of psychologists and engineering science. Moreover, this dissertation is definitely among the first to adopt a holistic view of machine vision techniques to enable a computer to understand the driver's momentary state.

The experiments showed that even the very advanced faceLAB-vision system suffers some practical drawbacks. The test experiences indicated that robustness in varying lighting conditions is not at a sufficient level for a commercial monitoring product, since e.g. eye tracking is often lost in tunnels. The second drawback is the system's inability to preserve eye tracking during large head movements. However, the system's ability to adapt rapidly to a new driver—an extremely important aspect—was very pleasing. The test system included high-quality industrial cameras that automatically adjust their own iris and include an external zooming property. The vision system incorporated in the test arrangements costs more than 35 000 EUR, which exceeds the price of a typical passenger car. The price level is not even acceptable for a heavy goods vehicle, the average price of which is as much as 500 000 EUR. Therefore, lower cost components are required and can be achieved for example by installing low-cost optics, which however, raises the importance of camera calibration to preserve the stability of the stereo vision system.

However, in order to minimise the errors few more features could be adopted for making especially cognitive distraction detection more reliable. The most promising features are variation of speed which according to literature review is more stable when the driver is under the cognitive workload. Also PERCLOS, which is more commonly utilised for detecting fatigue, should be taken into account in the future scenarios. However, perhaps the most interesting way would be adopting accomplishing the cognitive distraction detection with fatigue detection and visual distraction detection when perhaps the combination would tell something much more advanced than can be imaged in the first sight. The alternative way is to adopt more historical data like sleep patterns (Zhu & Ji 2004) or braking patterns, etc.

The hypotheses of this thesis states that detecting the level of distraction or fatigue can be performed with a set of image processing methods, through the use of eye-based measurements and a fusion with the other indicators, such as lane-keeping performance or steering activity. Chapter 5 and Publication VII present experimental tests that report on the performance of visual and cognitive distraction detection implementations. Chapter 5 proposes and reviews also the relevant driver measures for monitoring driver behaviour. Publication VI presents the results of the statistical analysis, which were intended to indicate the relevance of the tested features and the scenarios for developing the algorithm. The faceLAB stereo vision (Seeing Machines 2006), which included internal eye-tracking capability, was utilised in the tests (see Figure 6). Visual distraction was detected by using the attention mapping algorithm, which corresponds with the classifier utilised in the scrap-metal sorting example of Publication V. The results show that the road-ahead cluster is found to have up to a 90% accuracy, which in turn reflects the performance level of the visual distraction detection. The algorithm is reliable, but requires some more work to adapt to new types of cockpit more easily. Cognitive distraction was detected with a support vector machine type classifier (Publication VII). The results indicated a 65% performance in detecting artificially induced cognitive tasks for HGV and even 80% for a passenger car. The results are competitive to the prior studies reviewed in Publication VII. Cognitive distraction detection progresses in the right direction, but more experiments are needed before making firm conclusions about how close the module is to a viable product, although selecting the support vector machine -type technique seems to be a sensible choice.

Thus, further work on the data processing algorithms and the platform are still needed in order to realise a product that meets the requirements for robustness and the viable price of a vehicle application. Market products for fatigue detection already exist, but not as an internal vehicle instrument. The first implementations will most probably be seen in HGVs. Distraction detection modules are anticipated to be implemented in the first vehicles somewhere around 2010–2012.

References

Albery, W. B., Repperger, D. W., Reid, G. B., Goodyear, C., Ramirez, L. E. & Roe, M. M. 1987. Effect of noise on a dual task: subjective and objective workload correlates [pilot performance in fighter aircraft cockpit]. IEEE National Aerospace and Electronics Conference: NAECON 1987. Dayton, OH, USA. 18–22 May 1987. Vol. 4, pp. 1457–1463.

Almén, L. 2003. Reducing the effects of driver distraction: A comparison of distraction alerts on driver attention. Licentiate thesis No. 1017. Linköping, Sweden: Linköping University.

Andreeva, E. O., Aarabi, P., Philiastides, M. G.., Mohajer, K. & Emami, M. 2004. Driver drowsiness detection using multi-modal sensor fusion. *Proceedings of the SPIE – The International Society for Optical Engineering*. Conference of Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications. Orlando, FL, USA. 14–15 Apr. 2004. Vol. 5434, Issue 1, pp. 380–390. ISSN: 0277-786X.

Arensberg, U. 2004. Tiredness and distraction are major sources of danger in traffic – the truck monitors how you are feeling. Press release 18 May, 2004. Volvo Truck Corporation.

Bagci, A. M., Ansari, R., Khokhar, A. & Cetin, E. 2004. Eye tracking using Markov models. *Proceedings of the 17th International Conference on Pattern Recognition*. Cambridge, UK. 23–26 Aug. 2004. Vol. 3, pp. 818–821. ISBN: 07695 2128 2.

Bennett, K. P. & Campbell, C. 2000. Support Vector Machines: Hype or Hallelujah? *SIGKDD Explorations*. Vol. 2, Issue 2, pp. 1–13.

Bellotti, C., Bellotti, F., De Gloria, A., Andreone, L. & Mariani, M. 2004. Developing a near infrared based night vision system. *IEEE Intelligent Vehicles Symposium 2004*. Parma, Italy. 14–17 June 2004. Pp. 686–691.

Bellotti, F., De Gloria, A., Montanari, R., Dosio, N. & Morreale, D. 2005. *COMUNICAR: designing a multimedia, context-aware human-machine interface for cars*. Cognition, Technology & Work, Vol. 7, No. 1, pp. 36–45.

Bergasa, L. M., Nuevo, J., Sotelo, M. A., Barea, R. & Lopez, M. E. 2006. Real-Time System for Monitoring Driver Vigilance. *IEEE Transaction on Intelligent Transportation Systems*. Vol. 7, Issue 1, pp. 63–77. ISSN: 1524-9050.

Berthouze, L., Chavand, F. & Barret, C. 1996. A camera neural model. *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*. Beijing, China. 14–17 Oct. 1996. Vol. 3, pp. 2124–2126. ISBN: 0 7803 3280 6.

Beymer, D. 2000. Person Counting Using Stereo. Workshop on Human Motion (HUMO'00). Texas, U.S.A. 7–8 Dec 2000. Pp. 127–134. ISBN: 0-7695-0939-8.

Bittner, R., Smrcka, P., Vysoky, P., Hana, K., Pousek, L. & Schreib, P. 2000. Detecting of fatigue states of a car driver. *Proceedings of the Medical Data Analysis. First International Symposium*, ISMDA 2000. Frankfurt, Germany. 29–30 Sep 2000. Pp. 260–273. ISBN: 3 540 41089 9.

Boverie, S. 2004. Driver fatigue monitoring technologies and future ideas. *Proceedings of the 2004 AWAKE Road Safety Workshop*. Balocco, Italy. 29 Sep 2004.

Boverie, S., Daurenjou, D., Esteve, D., Poulard, H. & Thomas, J. 2002. Driver Vigilance Monitoring – New Developments. IFAC, 15th Triennial World Congress. Barcelona, Spain. 21–26 July 2002.

Bruce, S., Rice, C. & Hepp, R. 1998. Design and test of military cockpits. *Proceedings of the IEEE Aerospace Applications Conference Proceedings*. Snowmass at Aspen, CO, USA. 21–28 Mar 1998. Vol. 3, pp. 5–14.

Burges, C. J. C. 1998. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, Vol. 2, Issue 2, pp. 121–167. ISSN: 1384-5810.

Chiang, C. & Huang, C. 2004. A neural-based surveillance system for detecting dangerous non-frontal gazes for car drivers. *IEICE Transactions on Information and Systems*, Vol. v E87-D, Issue 9, pp. 2229–2238. ISSN: 0916-8532.

Claesson, C. 2003. Volvo trucks researches driver distraction. Press release 5 May 2003. Volvo Truck Corporation. Available at: http://www.volvo.com/trucks/global/en-gb/newsmedia/press_releases/. Link checked 1 Jun 2006.

Correia, B. & Dinis, J. 1998. Camera model and calibration process for high-accuracy digital image metrology of inspection planes. *Proceedings of the SPIE Conference on Machine Vision Systems for Inspection and Metrology VII*. Boston, MA, USA. 4–5 Nov. 1998. Vol. 3521, pp. 42–53. ISSN: 0277-786X.

Del Bue, A., Comaniciu, D., Ramesh, V. & Regazzoni, C. 2002. Smart cameras with real-time video object generation. *Proceedings of ICIP 2002 International Conference on Image Processing*. Rochester, New York, USA. 22–25 Sep 2002. Vol. 3, pp. 429–432. ISSN: 1522-4880.

Dinges, D. F., Mallis, M. M., Maislin, G. & Powell, J. W. 1998. Evaluation of Techniques for Ocular Measurement as an Index of Fatigue and the Basis for Alertness Management. Report number: DOT-HS-808 762. Washington, DC. U.S.A.: National Highway Traffic Safety Administration. 118 p.

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z. R., Jermeland, J. & Knipling, R. R. 2006. The 100-Car Naturalistic Driving Study, Phase II – Results of the 100-Car Field Experiment. Performed by Virginia Tech Transportation Institute, Blacksburg, VA. Sponsored by National Highway Traffic Safety Administration, Washington, D.C. DOT HS 810 593 – April 2006.

D'Orazio, T., Leo, M., Spagnolo, P. & Guaragnella, C. 2004. A neural system for eye detection in a driver vigilance application. *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems*. Washington, WA, USA. 3–6 Oct. 2004. Pp. 320–325. ISBN: 0 7803 8500 4.

East, J. A., Bauer Jr., K. W. & Lanning, J. W. 2002. Feature selection for predicting pilot mental workload: A feasibility study. *International Journal of Smart Engineering System Design*, Vol. 4, Issue 3, Jul./Sep. 2002, pp. 183–193. ISSN: 1025-5818.

Engström, J., Johansson, E. & Östlund, J. 2005. Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 8, Issue 2, SPEC. ISS., pp. 97–120. ISSN: 1369-8478.

Engström, J., Arfwidsson, J., Amditis, A., Andreone, L., Bengler, K., Cacciabue, P. C., Janssen, W., Kussman, H. & Nathan, F. 2006. Towards the automotive HMI of the future: Mid-term results of the AIDE project. 10th International Forum on Advanced Microsystems for Automotive Applications (AMAA). Berlin, Germany. 25–27 Apr 2006.

Eriksson, M. & Papanikolopoulos, N. P. 1997. Eye-tracking for detection of driver fatigue. *Proceedings of the IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*. Boston, MA, USA. 9–12 Nov 1997. Pp. 314–319.

Federation of French motorway and toll facility companies. 2006. Web page. Available at: http://www.autoroutes.fr. Referred on 23 Apr 2006.

Fletcher, L., Apostoloff, N., Chen, J. & Zelinsky, A. 2001. Computer vision for vehicle monitoring and control. *Proceedings of the 2001 Australian Conference on Robotics and Automation*. Sydney, Australia. 14–15 Nov 2001. Pp. 67–72.

Fletcher, L., Apostoloff, N., Petersson, L. & Zelinsky, A. 2003. Vision in and out of Vehicles. *IEEE Intelligent Systems*, Vol. 8, Issue 3, pp. 12–17. ISSN: 1094–7167.

Fletcher, L., Petersson, L. & Zelinsky, A. 2005. Road scene monotony detection in a fatigue management driver assistance system. *Proceedings of IEEE Intelligent Vehicles Symposium*. Las Vegas, Nevada, USA. 6–8 June 2005. Pp. 484–489. ISBN: 0 7803 8961 1.

Fukumi, M. 2005. Driver face monitoring using a near-infrared camera. 7th IASTED International Conference on Signal & Image Processing. Honolulu, HI, USA. 15–17 Aug. 2005. Pp. 156–160. ISBN: 0 88986 516 7.

Gonzales, R. C. & Woods, R. E. 1993. Digital Image Processing. Addison-Wesley Publishing Company, USA. 716 p. ISBN: 0-201-60078-1.

Gonzalez-Mendoza, M., Santana-Diaz, A., Hernandez-Gress, N. & Titli, A. 2003. Driver vigilance monitoring, a new approach. *Proceedings on IEEE Intelligent Vehicle Symposium*. Versailles, France. 17–21 Jun 2002. Pp. 358–363. ISBN: 0 7803 7346 4.

Grace, R. 2001. Drowsy Driver Monitor and Warning System. International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. Aug. 2001.

Grace, R., Byrne, V. E., Bierman, D. M., Legrand, J.-M., Gricourt, D., Davis, B. K., Staszewski, J. J. & Carnahan, B. 1998. A. drowsy driver detection system for heavy vehicles. Proceedings of the 17th DASC. AIAA/IEEE/SAE. Digital Avionics Systems Conference. Bellevue, WA, USA. 31 Oct – 7 Nov 1998. Vol. 2, Issue 36, pp. 1–8. ISBN: 0 7803 5086 3.

Grauman, K., Morency, L.-P. & Darrell, T. 2002. Using Visual Cues to Sense Lapses in Driver Alertness. Presented in U.S. Department of Energy Computational Science Graduate Fellowship Conference. Washington, DC. 17–20 Jul 2002.

Gu, H., Ji, Q. & Zhu, Z. 2002. Active facial tracking for fatigue detection. *Proceedings Sixth IEEE Workshop on Applications of Computer Vision (WACV)*. Orlando, FL, USA. 3–4 Dec. 2002. Pp. 137–142. ISBN: 0 7695 1858 3.

Hammel, K. R., Fisher, D. L. & Pradhan, A. K. 2002. Verbal and spatial loading effects on eye movements in driving simulators: A comparison to real world driving. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*. Baltimore, MD, USA. 30 Sept. – 4 Oct. 2002. Pp. 2174–2178.

Harbluk, J. L., Noy, Y. I. & Eizenman, M. 2002. The Impact of Cognitive Distraction on Driver Visual Behaviour and Vehicle Control. *Transport Canada Technical Report*. Available at:
http://www.tc.gc.ca/roadsafety/tp/tp13889/en/menu.htm. Link checked on 2 Feb 2006.

Hautiere, N., Tarel, J.-P., Lavenant, J. & Aubert, D. 2006. Automatic fog detection and estimation of visibility distance through use of an onboard camera. *Machine Vision and Applications*, Vol. 17, Issue 1, pp. 8–20. ISSN: 0932-8092.

Haykin, S. 1999. Neural Networks. 2nd edition. Upper Saddle River, New Jersey: Prentice-Hall, Inc. 842 p. ISBN: 0132733501.

Hecht, E. 1998. Optics. 3rd ed. Addison Wesley. ISBN: 0201838877.

Heikkilä, J. 1997. Accurate camera calibration and feature based 3-D reconstruction from monocular image sequences. Dissertation thesis. Acta Universitas Ouluensis C Technica 108. Oulu, Finland: University of Oulu.

Heikkilä, J. & Silven, O. 1997. A four-step camera calibration procedure with implicit image correction. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Juan, Puerto Rico. 17–19 June 1997. Pp. 1106–1112. ISBN: 0 8186 7822 4.

Heisele, B., Verri, A. & Poggio, T. 2002. Learning and vision machines. *Published in Proceedings of the IEEE*, Vol. 90, Issue 7, pp. 1164–1177. ISSN: 0018-9219.

Heitmann, A., Guttkuhn, R., Aguirre, A., Trutschel, U. & Moore, E. M. 2001. Technologies for the Monitoring and Prevention of Driver Fatigue. *Proceedings of the First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.* Aspen, CO USA. 14–17 Aug. 2001. Pp. 81–86.

Hilton, M., Jawerth, B. & Sengupta, A. 1994. Compressing still and moving images with wavelets. *Multimedia Systems, Special issue on video compression*, Vol. 2, Issue 5, pp. 218–227. ISSN: 0942-4962.

Ho, C. K., Itamura, M. T., Kelley, M. J. & Hughes, R. C. 2001. Review of Chemical Sensors for In-Situ Monitoring of Volatile Contaminants. Report number: SAND2001-0643. 34 p.

Hoedemaeker, M., Ridder, S. N. & Janssen, W. H. 2002. Review of European human factors research on adaptive interface technologies for automobiles. Soesterberg, Netherlands: Human Factors Research Institute TNO. 60 p.

Horrey, W. J. & Wickens, C. D. 2004. The impact of cell phone conversations on driving: A meta-analytic approach. (Technical Report AHFD-04-2/GM-04-1.) Warren, MI: General Motors Corporation.

Huber, D. F., Denes, L., Hebert, M., Gottlieb, M., Kaminsky, B. & Metes, P. 1998. A spectro-polarimetric imager for intelligent transportation systems. *Proceedings of the SPIE – The International Society for Optical Engineering.* Vol. 3207. SPIE-Int. Soc. Opt. Eng. Pp. 94–102. ISSN: 0277-786X.

Ilic, S., Katupitiya, J. & Tordon, M. 2004. In-vehicle data logging system for fatigue analysis of drive shaft. International Workshop on Robot Sensing, 2004. ROSE 2004. Pp. 30–34.

Ito, T., Mita, S., Kozuka, K., Nakano, T. & Yamamoto, S. 2002. Driver blink measurement by the motion picture processing and its application to drowsiness detection. Proceeding of the IEEE 5th International Conference on Intelligent Transportation Systems. Singapore. 3–6 Sept. 2002. Pp. 168–173. ISBN: 0 7803 7389 8.

Joachims, T. 1999. Making large-Scale SVM Learning Practical. Advances in Kernel Methods. Published in Schölkopf, B., Burges, C. & Smola, A. (eds.) Support Vector Learning. MIT-Press. Available at: http://www-ai.cs.uni-dortmund.de/DOKUMENTE/joachims_99a.pdf. Link checked on 2 Feb 2006.

Jokela, M. 2006. Implementation and Analysis of The Cockpit Activity Assessment Module. Master of Science Thesis. Tampere, Finland: Tampere University of Technology. 61 p.

Junghee, J. & Choongwon, K. 1999. Robust camera calibration using neural network. *Proceedings of IEEE Region 10 Conference. TENCON 99. 'Multimedia Technology for Asia-Pacific Information Infrastructure'*. Cheju Island, South Korea. 15–17 Sept. 1999. Vol. 1, pp. 694–697. ISBN: 0 7803 5739 6.

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. & Ramsey, D. 2006.The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data. Performed by Virginia Tech Transportation Institute, Blacksburg, VA. Sponsored by National Highway Traffic Safety Administration, Washington, D.C. DOT HS 810 594 – April 2006.

Konolige, K. & Beymer, D. 2006. User's Manual, Software version 4.2. SRI International. Available at: http://www.videredesign.com. Link checked on 17 Apr 2006.

Lal, S. K. L., Craig, A., Boord, P., Kirkup, L. & Nguyen, H. 2003. Development of an algorithm for an EEG-based driver fatigue countermeasure. Journal of Safety Research, Vol. 34, Issue 3, pp. 321–328. ISSN: 0022-4375.

Larsson, P. & Victor, T. 2005. Method and arrangement for interpreting a subject's head and eye activity. United States Patent Application, US 2005/0073136 A1.

Lee, J., Reyes, M., Smyser, T., Liang, Y. & Thornburg, K. 2004. SAfety VEhicles using adaptive Interface Technology (Task 5) Final Report: Phase 1 of the SAVE-IT project. Available at: http://www.volpe.dot.gov/opsad/saveit. Link checked on 1 Jun 2006.

Luan, X., Schwarte, R., Zhang, Z., Xu, Z., Heinol, H. G., Buxbaum, B., Ringbeck, T. & Hess, H. 2001. 3D intelligent sensing based on the PMD technology. Proceedings of the SPIE – The International Society for Optical Engineering. Toulouse, France. 17–20 Sep. 2001. Vol. 4540, pp. 482–487. ISSN: 0277-786X.

Manduca, A. 1995. Compressing images with wavelet/subband coding. *IEEE Engineering in Medicine and Biology Magazine*, Vol. 14, Issue 5, pp. 639–646. ISSN: 0739-5175.

Marco, S., Ortega, A., Pardo, A. &Samitier, J. 1998. Gas identification with tin oxide sensor array and self-organizing maps: adaptive correction of sensor drifts. *IEEE Transactions on Instrumentation and Measurement*, Vol. 47, Issue 1, pp. 316–321. ISSN: 0018-9456.

McCall, J. C. & Trivedi, M. M. 2004. Visual context capture and analysis for driver attention monitoring. *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems*. Washington, D.C., USA. 3–6 Oct. 2004. Pp. 332–337.

McCall, J. C. & Trivedi, M. M. 2006. Video-Based Lane Estimation and Tracking for Driver Assistance: Survey, System, and Evaluation. *IEEE Transaction on Intelligent Transportation Systems*, Vol. 7, Issue 1, pp. 20–37. ISSN: 1524-9050.

Neale, V. L., Dingus, T. A., Klauer, S. G., Sudweeks, J. & Goodman, M. 2005. An overview of the 100-car naturalistic study and findings. *Paper presented at the 19th International Technical Conference on Enhanced Safety of Vehicles (ESV)*. Washington, DC. 6–9 June 2005.

O'Brien, P. M. 1988. Eyeblink monitoring as a means of measuring pilot physiological state. *Proceedings of the IEEE National Aerospace and Electronics Conference: NAECON 1988*. Dayton, OH, USA. 23–27 May 1988. Vol. 3, pp. 890–892.

Perez, A., Cordoba, M. L., Garcia, A., Mendez, R., Munoz, M. L., Pedraza, J. L. & Sanchez, F. 2003. A precise eye-gaze detection and tracking system. WSCG'2003: 11th International Conference in Central Europe on Computer Graphics, Visualization and Digital Interactive Media. Plzen, Czech Republic. 3–7 Feb. 2003. Pp. 105–108. ISSN: 1213-6972.

Pilutti, T. & Ulsoy, G. 1995. On-line identification of driver state for lane-keeping tasks. *Proceedings of 1995 American Control Conference – ACC'95*. Seattle, WA, USA. 21–23 June 1995. Vol. 1, pp. 678–681. ISBN: 0 7803 2445 5.

Recarte, M. A. & Nunes, L. M. 2003. Mental workload while driving: Effects on visual search, discrimination and decision making. *Journal of Experimental Psychology: Applied*, Vol. 9, Issue 2, pp. 119–137.

Rimini-Doering, M., Manstetten, D., Altmueller, T., Ladstaetter, U. & Mahler, M. 2001. Monitoring driver drowsiness and stress in a driving simulator. *Proceedings of First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*. Aspen, Colorado. 14–17 Aug. 2001. Pp. 58–63.

Santana Diaz, A., Jammes, B., Esteve, D. & Gonzalez Mendoza, M. 2002. Driver hypovigilance diagnosis using wavelets and statistical analysis. *Proceedings on IEEE 5th International Conference on Intelligent Transportation Systems*. Singapore. 3–6 Sep 2002. Pp. 162–167. ISBN: 0 7803 7389 8.

Seeing Machines. 2006. Web page. Available at: http://www.seeingmachines.com. Link checked 17 Apr 2006.

Sebastiani, F. 2002. Machine learning in automated text categorization. *ACM Computing Surveys (CSUR) archive*, Vol. 34, Issue 1, pp. 1–47. ISSN 0360-0300.

Sethuramasamyraja, B., Ghaffari, M. & Hall, E. 2003. Automatic Calibration and Neural Networks for Robot Guidance. *Proceedings of SPIE Intelligent Robots and Computer Vision XXI: Algorithms, Techniques, and Active Vision, 5267*. Pp. 137–144.

Shimizu, S., Suematsu, Y. & Hosono, H. 1996. A Method of Monocular Vision by Wide Angle Foveated Optical System – Object detection using Virtual Cylindrical Screen. *Proceedings of Fourth International Conference on Control, Automation, Robotics and Vision (ICARCV-96)*. Singapore. 3–6 Dec 1996.

Shimizu, S., Suematsu, Y. & Yahata, S. 1998. Wide-angle vision sensor with high-distortion lens. (Detection of camera location and gaze direction based on the two-parallel-line algorithm.) *JSME International Journal, Series C (Mechanical Systems, Machine Elements and Manufacturing)*, Vol. 41, Issue 4 pp. 893–900. ISSN: 1344-7653.

Singh, S. & Papanikolopoulos, N. P. 1999. Monitoring driver fatigue using facial analysis techniques. Proceedings of 1999 Intelligent Transportation Systems Conference. Tokyo, Japan. 5–8 Oct. 1999. Pp. 314–318. ISBN: 0 7803 4975 X.

Smith, P., Shah, M. & da Vitoria Lobo, N. 2000. Monitoring Head/Eye Motion for Driver Alertness with One Camera. *Proceedings of the Fifteenth IEEE International Conference on Pattern Recognition*. Barcelona, Spain. 3–7 Sep 2000. Vol. 4, pp. 636–642. ISBN: 0-7695-0750-6.

Smith, P., Shah, M. & da Vitoria Lobo, N. 2003. Determining driver visual attention with one camera. *IEEE Transactions on Intelligent Transportation Systems*. Vol. 4, Issue 4, pp. 205–218. ISSN: 1524-9050.

Stachowicz, M. S. & Lemke, D. 2002. Image segmentation and classification using color features. *Proceedings of 4th EURASIP – IEEE Region 8 International Symposium on Video/Image Processing and Multimedia Communications*. Zadar, Croatia. 16–19 June 2002. Pp. 57–64. ISBN: 953 7044 01 7.

Tan, Y., Wu, Y. & Chang, L. 1995. Wavelet transform coding for image sequence compression. *Proceedings of IEEE Singapore International Conference on Information Engineering*. 3–7 July 1995. Pp. 478–482.

Tattegrain Veste, H., Bellet, T., Boverie, S., Kutila, M., Viitanen, J., Bekiaris, E., Panou, M., Engström, J. & Angvall, A. 2005. Development of A Driver Situation Assessment Module In The Aide Project. *Proceedings of 16th IFAC WORLD CONGRESS*. Prague, Czech Republic. 4–8 Jul 2005.

Tsai, R. Y. 1987. A Versatile Camera Calibration Technique for High-Accuracy 30 Machine Vision Metrology Using Of-the-Shelf TV Cameras and Lenses. *IEEE Journal of Robotics and Automation*, pp. 323–344.

Vattulainen, A. Korpinen, J. Kutila, M. Pohjankoski, L. Reponen, S. Samela, V. & Viitanen, J. 2002. Menetelmä materiaalin lajitteluluokkien määrittämiseksi. Patent Application, FI20022246.

Veeraraghavan, H. & Papanikolopoulos, N. P. 2001. Detecting Driver Fatigue through the use of Advanced Face Monitoring Techniques. 48 p.

Veeraraghavan, H., Atev, S., Bird, N., Schrater, P. & Papanikolopoulos, N. 2005. Driver activity monitoring through supervised and unsupervised learning. IEEE Intelligent Transportation Systems Conference. Vienna, Austria. 13–15 Sep. 2005. Pp. 580–585. ISBN: 0 7803 9215 9.

Victor, T. 2000. On the need for attention support systems. *Journal of Traffic Medicine*, Vol. 28, Issue 2S.

Victor, T. 2003. System and method for monitoring and managing driver attention loads. United States Patent Application, US 2003/0181822 A1.

Victor, T. 2005. Keeping Eye and Mind on the Road. Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 9. Uppsala: Acta Universitatis Upsaliensis.

Victor, T. W., Harbluk, J. L. & Engstrom, J. A. 2005. Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 8, Issue 2, SPEC. ISS., pp. 167–190. ISSN: 1369-8478.

Wahlstrom, E., Masoud, O. & Papanikolopoulos, N. 2003. Vision-based methods for driver monitoring. Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems. Shanghai, China. 12–15 Oct. 2003. Vol. 2, pp. 903–908. ISBN: 0 7803 8125 4.

Wang, R., Guo, K., Shi, S. & Chu, J. 2003. A monitoring method of driver fatigue behavior based on machine vision. *Proceedings on IEEE IV2003 Intelligent Vehicles Symposium*. Columbus, OH, USA. 9–11 Jun 2003. Pp. 110–113. ISBN: 0 7803 7848 2.

Wang, R., Guo, L., Tong, B. & Jin, L. 2004. Monitoring mouth movement for driver fatigue or distraction with one camera. *Proceedings of the 7th International IEEE Conference on Intelligent Transportation System*s. Washington, WA, USA. 3–6 Oct. 2004. Pp. 314–319. ISBN: 0 7803 8500 4.

Welstead, S. T. 1999. Fractal and Wavelet Image Compression Techniques. SPIE Optical Engineering Press. 254 p. ISBN: 0819435031.

Xia, L., Fengliang, X. & Fujimura, K. 2003. Real-time eye detection and tracking for driver observation under various light conditions. Proceedings of IV'2002. IEEE Intelligent Vehicle Symposium. Versailles, France. 17–21 Jun 2002. Vol. 2, pp. 344–351. ISBN: 0 7803 7346 4.

Zhu, Z. & Ji, Q. 2004. Real time and non-intrusive driver fatigue monitoring. Proceedings – 7th International IEEE Conference on Intelligent Transportation Systems, ITSC 2004. Washington, DC, United States. 3–6 Oct 2004. Pp. 657–662.

Zhuang, H. & Roth, Z. 1996. Camera-Aided Robot Calibration. Florida Atlantic University, Florida: CRC Press.

Zixiang, X., Ramchandran, K. Orchard, M. T. & Ya-Qin, Z. 1999. A comparative study of DCT- and wavelet-based image coding. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 9, Issue 5, pp. 692–695. ISSN: 1051-8215.

Zrimec, T. 2003. Toward automatic colour calibration using machine learning. *IEEE International Conference on Industrial Technology*. Maribor, Slovenia. 10–12 Dec 2003. Vol. 1, Issue 1, pp. 227–232. ISBN: 0 7803 7852 0.

Östlund, J., Nilsson, L. et al. 2004. HASTE Deliverable 2: HMI and Safety-Related Driver Performance. EU funded project GRD1/2000/25361 S12.31962.

PUBLICATION I

Camera Calibration in Machine Automation

In: *Human Friendly Mechatronics*. Selected papers of the International Conference on Machine Automation ICMA 2000. Osaka, Japan. 27–29 September 2000. Pp. 211–216.

Camera calibration in machine automation

Matti Kutila, Juha Korpinen, and Jouko Viitanen

VTT Automation, Machine Automation
P.O. BOX 1302, FIN-33101 Tampere, Finland

This paper deals with utilization of a camera calibration method in machine automation. Currently available calibration methods are described shortly, and a new type of calibration object which is suitable for use in, e.g. small parts assembly cells, is presented. Also the method of searching the calibration points is given new consideration. Experimental results are presented on the tests using the calibration method with the novel target.

1 INTRODUCTION

Simplified camera models, so called pinhole models, normally assume that light rays go through the lens without bending. In real life light rays bend in the lens which causes geometric distortion (aberration) to the image. Aberrations are especially amplified in short focal length lenses. In the following, aberrations caused by the non-linearity of the lens are called radial distortion. Another type of distortion which is typically called tangential, is the result of the offset between the optical axis of the lens and the normal of the image plane, together with possible tilt of the image plane. Radial distortion is formed symmetrically around the location of the optical axis and therefore accurate correction of the radial distortion requires that tangential distortion is corrected first.

The purpose of the camera calibration is to define the accurate camera model, which also takes into account the distortions in the image. When the accurate camera model is known, it is possible to remove the effects of the distortions from the image and perform accurate geometric measurements based on it.

Several different kinds of calibration methods are proposed in literature. We have tested and investigated especially the method introduced by Heikkilä [2] which uses four calibration steps. We have further developed the method by adapting it to use a new kind of calibration object, which makes it easier to utilize the method in machine automation industry. Only one image of this object is needed for executing the calibration process. Other new developments are the modules of software for searching the calibration points and executing the calibration process. These were written in the ANSI C programming language. The software makes it easy to embed the calibration method in machine vision applications in e.g. automation cells.

The advantage of the new calibration method is that it is fully automatic, while previously several manual steps were needed, and several pictures of test images had to be taken. If the calibration object is located inside an automatic assembly cell, the camera parameters can even be calculated while the cell is running online.

2 CALIBRATION METHODS

2.1 Camera model

The purpose of the camera calibration is to build an accurate mathematical camera model. Model building starts from pinhole model:

$$u_p = K_x \frac{f}{Z} X + u_0 \tag{1}$$

$$v_p = K_y \frac{f}{Z} Y + v_0 \tag{2}$$

where (X, Y, Z) are world frame coordinates, (K_x, K_y) are coefficients for converting millimeters to pixels, (u_0, v_0) is the location of the optical axis, (u_p, v_p) coordinates according to the pinhole model, and f the focal length.

The pinhole camera model is valid only when lenses with long focal length are used. If more accurate measurements are wanted, the radial distortion has to be taken into account. The second and fourth order terms in the polynomial approximation that is used for modelling the radial distortion have been proposed by Heikkilä [2]:

$$\Delta u_r = K_x \frac{f}{Z} X\left(k_1 r^2 + k_2 r^4\right) \tag{3}$$

$$\Delta v_r = K_y \frac{f}{Z} Y\left(k_1 r^2 + k_2 r^4\right) \tag{4}$$

where $r = \frac{f}{Z}\sqrt{(K_x X)^2 + (K_y Y)^2}$, (k_1, k_2) are radial distortion coefficients and (Δu_r, Δv_r) are the effects of the radial distortion (differences from the linear model).

There also other kinds of radial distortion models proposed in literature. For example Correia et al. [1] use the third and the fifth order terms in their model.

Heikkilä [2] has modelled tangential distortion with the following model:

$$\Delta u_t = 2t_1 \frac{f^2}{Z^2} K_x K_y XY + t_2\left(r^2 + 2\frac{f^2}{Z^2} K_x^2 X^2\right) \tag{5}$$

$$\Delta v_t = t_1\left(r^2 + 2\frac{f^2}{Z^2} K_y^2 Y^2\right) + 2t_2 \frac{f^2}{Z^2} K_x K_y XY \tag{6}$$

where (t_1, t_2) are tangential distortion coefficients and (Δu_t, Δv_t) are effects of the tangential distortion.

The following accurate camera model combines the distortion models to the pinhole camera model:

$$u = u_p + \Delta u_r + \Delta u_t \tag{7}$$

$$v = v_p + \Delta v_r + \Delta v_t \tag{8}$$

2.2 Calibration routine

Many different kinds of calibration methods are presented in literature. The main difference between the methods is that many of them calculate only radial distortion coefficients and omit the tangential distortion. Tsai's RAC method is probably the most commonly used one in camera metrology [3]. With Tsai's technique it is possible to define radial distortion very accurately, but before using it, tangential distortion has to be removed. Another well known calibration technique is Weng's two step calibration process [3]. Actually, the method developed by Heikkilä [2] has similarities to Weng's method. The main idea is to first define a "good initial guess" for the nonlinear optimization routine by using a linear camera model.

During the camera calibration process, the focal length (f), the position of the optical axis on the image plane (u_0, v_0), the radial distortion coefficients (k_1, k_2) and the tangential distortion coefficients (t_1, t_2) are defined. When those parameters are known, it is possible to define the image which corresponds to the one produced by the pinhole model camera. The external parameters of the camera, its position and orientation relative to the world frame coordinates are also calculated.

Heikkilä's [2] calibration process is based on four successive steps. The actual parameter calculation is performed in three steps and the image correction is performed in the last step. At the first step, coarse camera parameters are calculated using the linear camera model. More accurate camera parameters are then calculated during the nonlinear optimization routine. The third step removes the inaccuracies caused by perspective projection.

3 CALIBRATION OBJECT

3.1 Object

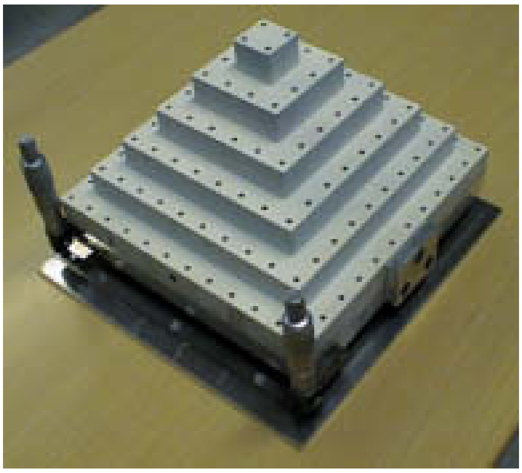


Figure 1. The calibration object which is suitable for positioning into an automatic assembly cell.

In the design of the calibration object, special attention has to be paid on the compactness (see figure 1). The idea was that it should be possible to place the object inside into the most commonly used automatic assembly cells. The area taken by the pyramidal calibration object is about 221 mm x 221 mm and it’s height is 138 mm. A three-dimensional calibration object makes it possible to calibrate the camera by using only one image. The holes in the object are used as calibration points. The material of the object is aluminum and it has been coated matt white for reducing specular reflections. The holes are drilled by an inaccuracy of 0,01 mm.

3.2 Calibration point searching

Not only the accuracy of locations of the calibration points on the object is important, but also the accuracy when searching them from the image. The positions of the points have to defined by sub-pixel accuracy. The easiest way for finding the centers of the calibration points is to calculate their central moments, i.e. weighted averages:

$$\bar{u} = \frac{\sum_v \sum_u u f(u,v)}{\sum_v \sum_u f(u,v)} \tag{9}$$

$$\bar{v} = \frac{\sum_v \sum_u v f(u,v)}{\sum_v \sum_u f(u,v)} \tag{10}$$

where *u* and *v* refer to the coordinates in an image and *f*(*u*, *v*) to the corresponding gray level. Note that the weight at the outside of the calibration point has to be 0. For good accuracy, we have to have plenty of pixels per calibration point. It has been found that they should be at least 10 pixels wide.

4 EXPERIMENTAL RESULTS

The tests of the calibration method have been divided to three parts. First we tested the ability of the method to define radial distortion parameters. Then we tested its suitability for recognizing the offset of the optical axis, and finally we checked the correctness of the external parameter calculation routine. A video camera with an 8 mm lens was used in the tests. We utilized the calibration software in Matlab code made by Heikkilä [2], and the calibration object shown in Figure 1.

A raster pattern was imaged in the radial distortion test. The distances between the raster points were equal, which means that the same number of pixels should be found between the points. Dividing the number of pixels by the distances of the points in millimeters gives us the spatial resolution. If there is radial distortion, it should cause variations to the resolution at different parts of the image area. Figure 2 shows the variation of the spatial resolution in the image area, measured in the horizontal direction. The upper line shows the resolution in the uncorrected image and the lower line in the corrected image that utilized the calibration method. In the corrected image, there are only small deviations in spatial resolution, which is mainly due to the measuring inaccuracies during the test. The effects of the aberrations have decreased clearly.

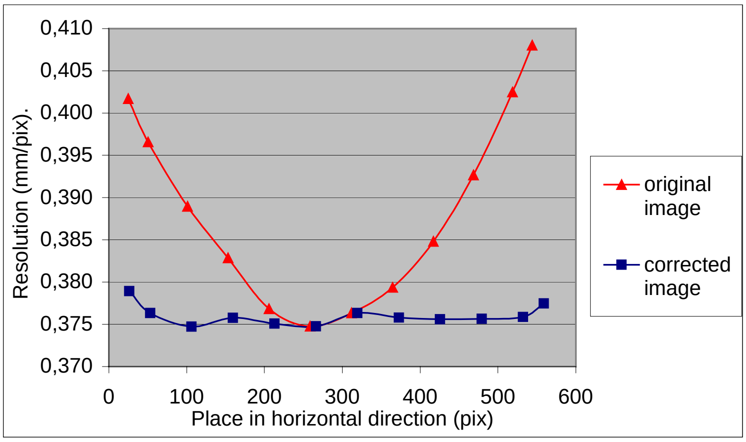


Figure 2. The measured spatial resolutions in the horizontal direction of the image.

The ability of the method to find the position of the optical axis from the image was examined by comparing the search results to an opto-mechanical method developed at VTT. The method is based on an iterative search of target points from the images of the camera while it is rotated around an axis that is close to the direction of the optical axis. The target points have to be far enough from the camera. The results of the measurements are shown in Table 1. The result of the calibration method has been calculated as an average of four measurements.

Table 1
X-Y positions of the optical axis measured with different methods.

	Nominal	Optomechanical method	Calibration method
Measure-ments	(320, 240)	(310.7, 223.9)	(313.5, 219.5) (312.3, 224.6) (309.6, 222.6) (311.0, 224.6) (309.7, 224.1)
Result	(320, 240)	(310.7±3.5,223.9±3.5)	(311.2±0.8,223.1±1.0)

The ability of the method to define the external parameters (position and orientation) of the camera was tested by measuring the distance between the calibration object and the camera. The inaccuracy of the handmade measurements can be ±2 mm. The calibration method gives the distance to the primary principal point of the lens which is difficult to determine without having the prescription data of the lens (the secondary principal point is, of course, well known). However, if we move the camera to different heights, and find the differences of the measured heights to a known reference position, then there is no dependence on the principal point separation in these values. Table 2 presents this difference of the measurements given by the calibration method and those given by manual measurements. The differences between these two methods seem to be within the error limits of the manual measurements, and the maximum difference is 0,6% of the value of the reference position, so the method seems to give reliable results for the camera height. Corresponding tests were performed for object rotation along one axis, which also gave positive results, so these tests gave confidence on the performance of the method.

Table 2
Comparisons of the manual measurements with the values given by the calibration method in position determination. The height variations of the camera were measured when the camera was lifted up from the reference level at 269,0 mm, and the corresponding differences from the calibration method were recorded.

Manual measurement (mm)	Calibration method (mm)	Difference (mm)
30	31,7	1,7
45	46,2	1,2
65	66	1
110	110,9	0,9

5 CONCLUSIONS

We presented improvements to a camera calibration method, including a compact calibration object, and possibility to perform the calibration with only one captured image. The tests of the method show that significant improvement of the accuracy was achieved for cameras that use short focal length lenses. In several tests the method was found to give reliable results when the novel calibration object was used. We also performed a few successful tests for determining the position and orientation of the camera with the same method.

REFERENCES

1. B. Correia, J. Dinis, Camera model and calibration process for high accuracy digital image metrology of inspection planes, proc, SPIE, vol 3521, Boston, Massachusetts, U.S.A. (1998).
2. J. Heikkilä, Accurate camera calibration and feature based 3-D reconstruction from monocular image sequences, Acta Universitas Ouluensis C Technica 108, Oulu, Finland (1997).
3. H. Zhuang, Z. Roth, Camera-Aided Robot Calibration, CRC Press, Florida, Florida Atlantic University. (1996).

PUBLICATION II

Calibration of the World Coordinate System with Neural Networks

In: *Proceedings of the 9th Mechatronics Forum International Conference Mechatronics 2004*. Ankara, Turkey. 30 August – 1 September 2004. Pp. 337–345.

CALIBRATION OF THE WORLD COORDINATE SYSTEM WITH NEURAL NETWORKS

Matti Kutila

VTT, P.O. Box 1302, 33101 Tampere, **FINLAND**.

email: matti.kutila@vtt.fi, tel.: +358 40 820 8334

Abstract

This paper presents a method, that partially solves calibration problems of the computer vision in robotics. The method connects image and world frame coordinates automatically. The calculation bases on neural network, which treats camera model as a black box. The world frame coordinates are mapped straight from the image frame. The coordinates are defined in relation to the distance between the camera and the target. The results show that accuracy of the proposed method is better than ±1,8 mm, when the size of the image is 250 x 150 mm. This is sufficient for many practical applications in robotics and better than achievable results with hand made measurements.

1 Introduction

Mapping the world frame coordinates from an image has been the topic of several scientific papers. The accuracy of the machine vision based measurements depends strictly on the knowledge of the camera location in relation to the world frame. Typically, it is hard to measure location of the camera with sufficient accuracy. On the other hand, the products in the assembly lines change frequently and the robots has to sophisticate to new environments without time consuming configuration changes. For this reason the coordinate mapping should be automatic and measurements with roll meters should be forgotten.

Spatial resolution can be calculated from the distance between the camera and the target. If so, the linear camera model called pinhole model is used (see. Figure 1). This means that when the distance increases, the imaging system scales the view. In reality, the optical system always contains non-linear components, which cause errors to actual spatial resolutions.

This study introduces a novel method for mapping image and world frame coordinates. The neural network creates a camera model which behaves like a black box. In other words, the idea is to find the black box which predicts world frame coordinates without knowing the complex camera models. Neural networks recognise image frame coordinates in the world frame. Fundamental benefit of this method is that the system automatically calibrates itself to world frame if the distance to the camera changes. The method presented in this paper is one step closer to automatic configuration of the vision systems in robotics.

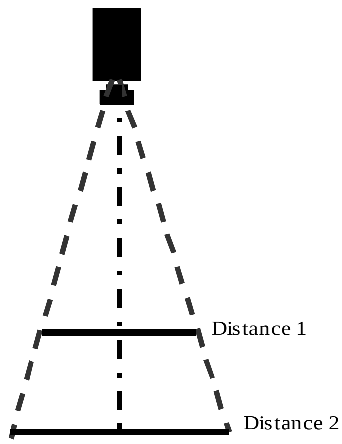


Figure 1 . Pinhole camera model

2 Network model

Numerous network models and topologies are available. Making the best choice is not a simple task. MLP (Multilayer Perceptrons) is the most commonly used model. It is quite easy to implement and can be adapted to various applications. RPF-networks could be better, especially if the output was optimised in statistical sense (Haykin 1999). However, this study bases to MLP-network. It is feasible for non-linear functions, which is fundamental requirement for the camera model.

After selecting the network model, the topology has to be optimised. In this study the topology was optimised by testing up to 3 hidden layers. Additionally, the amount of the neurons in each hidden layer varied from 2 to 15 neurons. After several test rounds it was obvious, that two hidden layers with 7 neurons on each layer was the most feasible for this application (see. Figure 2). Each layer has also its own bias term affecting all the neurons on that specific layer.

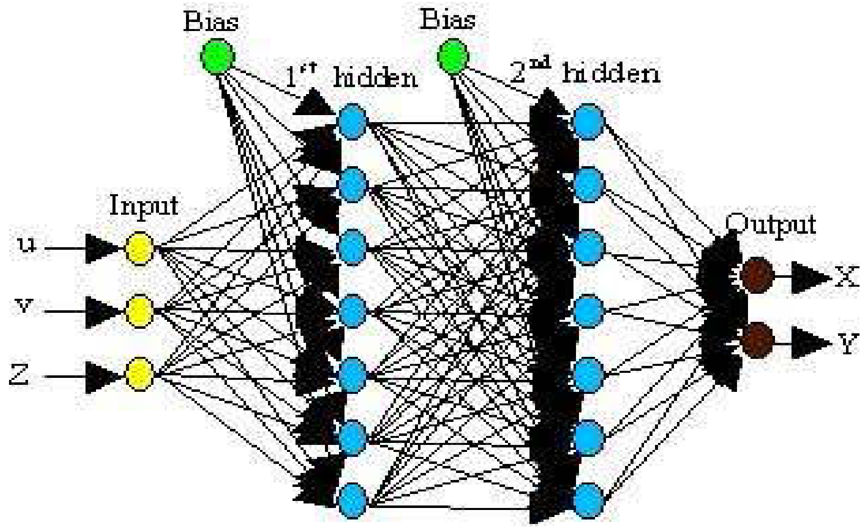


Figure 2. Topology of the proposed neural network

3 Testing arrangement

3.1Acquiring calibration data

Black and white type Sony XC-50ES camera was used in the tests (see. Figure 3). The size of the captured image was 768 x 576 pixels. The target was analysed from three distances: 441, 545 and 580 mm. The actual dimensions of the views varied from 189 x 132 mm to 260 mm x 180 mm. The neural network was trained with 99 calibration points and the result was validated with 34 points.

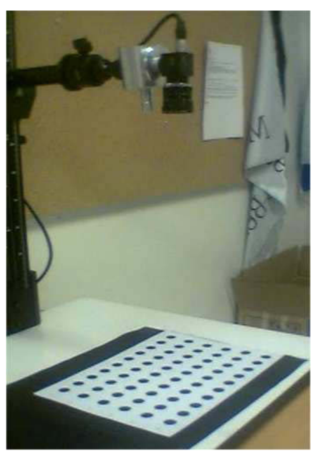


Figure 3. Testing arrangement

Sub-pixel theorem has to be utilised for creating calibration data. Location of the calibration points ($\underline{X}$, $\underline{Y}$) was defined with self written software (see. Figure 4). Calculation is based on central moments (see. Equations 1 and 2) . To achieve sufficient accuracy, the size of the calibration point has to be large enough. Diameter of the calibration points varied from 40 to 50 pixels in this case.

$$\underline{X} = \frac{\sum u_i w_j}{\sum w_j} \tag{1}$$

$$\underline{Y} = \frac{\sum v_i w_j}{\sum w_j} \tag{2}$$

u_i and v_i in the equations represent the image coordinates of a single pixel. Weight value (w_j) is 1 for the pixels inside the calibration point and 0 for the outside pixels.

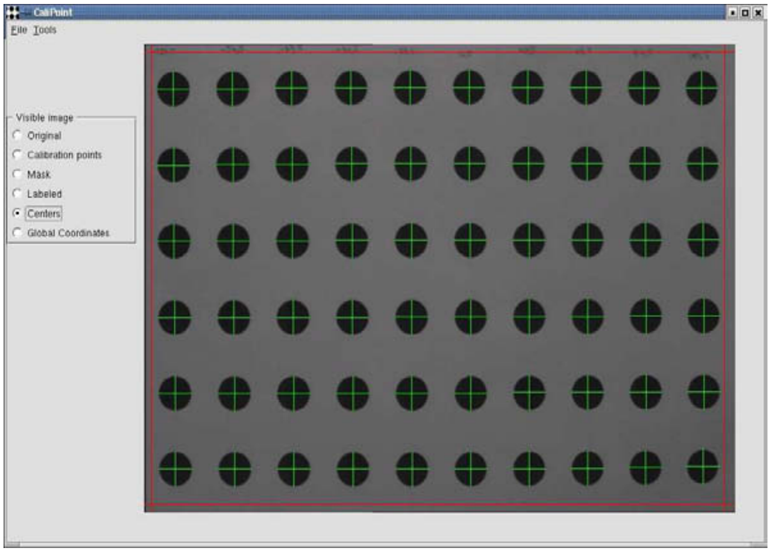


Figure 4. Application for searching locations of the calibration points

3.2Testing with neural network

Calibration data has to be shuffled before dividing it to training and validation samples. This ensures that the network does not learn the order of the calibration data.

The neural network was exited by Java application (see. Figure 5). The calibration data was loaded from the file created after shuffling. The test network was trained with following parameters:

- Learning rate: 0,02
- Number of the iterations: 8000

The same network application also executes validation stage and writes down the result file. The result file can be loaded to Microsoft Excel for analysing. The quality of the method was estimated by calculating location errors of the predicted coordinates (Equation 3).

$$e_{NN} = \sqrt{(X_p - \underline{X})^2 + (Y_p - \underline{Y})^2} \tag{3}$$

, where (X_p, Y_p) are predicted world frame coordinates.

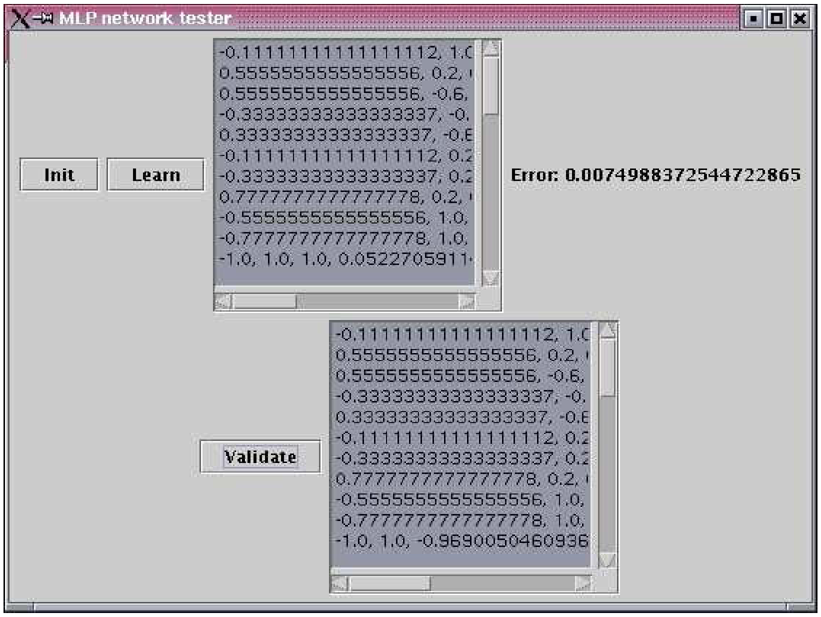


Figure 5. Test application for MLP network

3.3Reference result

The results from the neural network were compared to those defined with the pinhole camera model. The spatial resolution for the pinhole measurements was calculated as an average of the three known positions: the first in the upper, the second in the middle and the third in the lower part of the image. For linear camera model the spatial resolution is calculated as equation 4 presents. In Equation 4 *R* is spatial resolution, and *d* is distance from the target. The same calibration points were used both in test and reference measurements, because those were defined with sub-pixel accuracy.

$$R_{new} = \frac{d_{new}}{d_{old}} R_{old} \tag{4}$$

Error is defined as an absolute error between the real location (X_{real}) of the calibration point and calculated one (X_{calc}).

$$e_{HM} = |X_{real} - X_{calc}| \tag{5}$$

Comparative measurement was done only horizontally, which is sufficient to indicate that the proposed method works properly.

4 Experimental results

4.1Validation results

As already noted, the testing data was divided to training and validation samples. Figure 6 and Figure 7 visualise root mean square error of the predicted world frame coordinates horizontally and vertically. Y-axis is prediction error and X-axis location of the calibration point in world frame (see. Figure 8). The unit on both axes is millimeter.

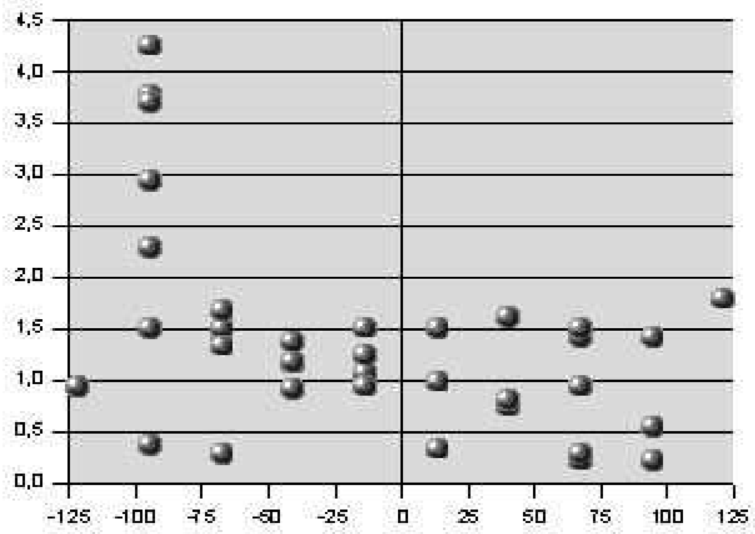


Figure 6. Mean square errors horizontally

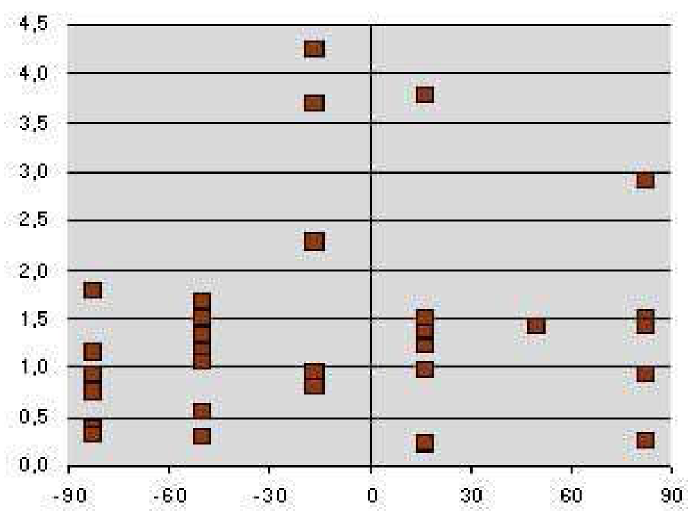


Figure 7. Mean square errors vertically

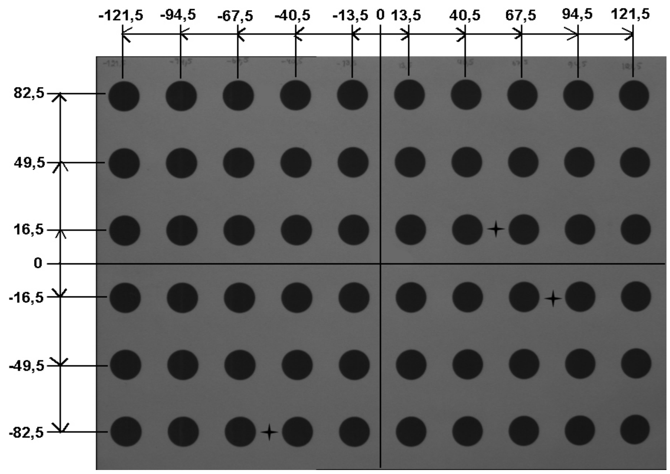


Figure 8. Coordinates of the calibration points

As the figures show, the error is below two millimeters for 86% of the test cases. Assumably the final result is even better, because one of the calibration point columns caused all the major errors (more than 1,8 mm). This can be seen from the Figure 6. The peak exists at -94,5 mm. Figure also shows that seven calibration points of that column were in validation sample. Normally only four points in each column included in the validation set. As a conclusion, learning of that column was not completed. However, the test indicates, that system is reliable and can predict the points in world frame coordinates with accuracy of ±1,8 mm.

4.2Test with interpolated coordinates

The second test predicts values of the interpolated coordinates. This indicates that the system works properly for any point in the image and not only for calibration rows and columns. The real coordinates were interpolated from neighborhoods. The results are presented on the Table 1. The utilised image points are drawn with cross to the Figure 8.

Table 1 . Errors of the interpolated points

X [mm]	Error [mm]
-54	0,51
54	0,20
81	0,61

Mean error is 0,44 mm, which corresponds with result of the validation data sample. So the system works also for the points which are out of the calibration pattern.

4.3Comparable measurement

The spatial resolution is 0,332 mm / pix when the camera distance is 545 mm from the calibration target. Moving camera to distance 441 mm gives error data presented in the Table 2. Real lengths are defined by starting measurement from the column −94,5 mm (see. Figure 8). The test distances are measured horizontally on three different locations.

Table 2. Errors caused by linear camera model

Real length [mm]	Y [mm]	Error [mm]
162,0	49,5	2,56
162,0	16,5	2,91
81	-49,5	1,77

5Conclusions

The tests prove that connection between the world frame and the image coordinate system can be established with presented method. The accuracy is better with this method than the linear camera model (pinhole model), which becomes invalid if the distance between the target and the camera changes. Non-linearity of real camera model causes unpredicted effects to behavior of the imaging system.

Possible applications for the system could be quality checking in assembly cells, where the location errors of the electronic components are not accepted. The algorithm could proceed in following way. Firstly, the image coordinates are mapped to the world frame with feeding training data to neural network. This is an off-line process, which is done in cell configuration stage. Secondly, the location marks and components on the board are extracted from the image with the machine vision software. This is the first part of the on-line process. Thirdly, the distance between two components is defined by utilising neural network, which compensates non-linearities of the imaging system. Finally, the target coordinates are compared to calculated ones and result is given if the component is in the right place.

The original aim of the this research work was to build up a system, which automatically calibrates itself as references (Barret et al. 1996) and (Choongwon & Junghee 1999) present. Calibration means not only connecting image and world frame coordinates, but also eliminating measurement inaccuracies of the optical system. Normally the effects of the aberrations are removed with constructing suitable camera model and optimising parameters of that function (Heikkilä & Silven 1997). In many cases this leads to a complex optimisation routine and camera model becomes erratic compared to the real optical system. The neural network is complex to create, if special properties are needed. However, in practical cases simple neural network gives sufficient accuracy compared to camera model functions. Furthermore, they are also faster and more reliable.

References

Haykin, S. (1999), *Neural Networks* . Prentice-Hall, Inc. Upper Saddle River. New Jersey. 842 p.

Barret, C., Berthouze, L., & Chavand, F. (1996), A camera neural model. Systems, Man, and Cybernetics, *Proceedings of the IEEE International Conference on Systems*, Volume: **3** , 14-17 Oct. 1996, p. 2124 -2126.

Choongwon, K. & Junghee, J. (1999), Robust camera calibration using neural network, TENCON 99. *Proceedings of the IEEE Region 10 Conference* , Volume: **1** , 15-17 Sept. 1999, p. 694-697.

Heikkilä, J. & Silven, O. (1997), A four step camera calibration procedure with implicit image correction. Proc. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* . San Juan, Puerto Rico, IEEE, p. 1106-1112.

PUBLICATION III

Parallel Image Compression and Analysis with Wavelets

In: *International Journal of Signal Processing* 2004.
Vol. 1, No. 1–4, pp. 65–68.
Reprinted with permission from the publisher.

Parallel Image Compression and Analysis with Wavelets

M. Kutila, J. Viitanen

***Abstract*—**This paper presents image compression with wavelet based method. The wavelet transformation divides image to low- and high pass filtered parts. The traditional JPEG compression technique requires lower computation power with feasible losses, when only compression is needed. However, there is obvious need for wavelet based methods in certain circumstances. The methods are intended to the applications in which the image analyzing is done parallel with compression. Furthermore, high frequency bands can be used to detect changes or edges. Wavelets enable hierarchical analysis for low pass filtered sub-images. The first analysis can be done for a small image, and only if any interesting is found, the whole image is processed or reconstructed.

***Keywords*—**image compression, jpeg, wavelet, vlc

I. Introduction

In recent years the interest in image compression techniques has been steadily grown. The trend is firstly to store and transmit images and video clips digitally and secondly to add multimedia properties to wireless mobile devices. The capacity of the long distance wireless networks is limited, which causes need for more efficient multimedia compression methods. The problem is to achieve sufficient compression ratio with acceptable degradation in low cost embedded consumer devices.

The basic JPEG standard for still image compression is the most commonly used intra-image compression method at the moment, but wavelet based methods have promising features for the future needs. JPEG is based on the discrete cosine transformation (DCT) of the 8x8 blocks (see. Figure 1). JPEG2000, which is a recent compression standard, utilises wavelet transformations. Wavelets normally require more computation power than DCT. The benefit is that they also retain spatial domain characteristics, which can be utilised for other purposes easier than DCT results.

This publication bases to a wavelet codec which was created at VTT for testing the feasibility of the methods to practical applications. The name of the codec is wavecodec. Wavecodec is software which is flexible for various image processing applications. The wavelet based codec is suitable when high compression ratios are needed [1]. Filtering is done in sequences of horizontal and vertical low- and high pass transformations.

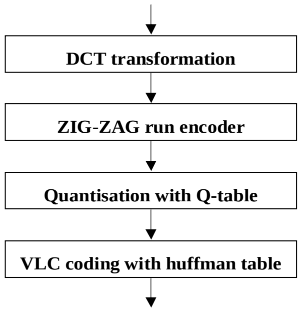


Figure 1: Simple description of the basic JPEG encoding procedure

II. Wavelet Transformation

The purpose of the transformations in image compression is to pack the significant information through sub-band coding. Then information can be carried with fewer bits. Irrelevant information on low entropy sub-bands can be totally removed.

Wavelets produce separate low and high frequency sub-bands (see. Figure 2) [3]. The interesting issue is that those bands can be investigated in spatial domain.

Figure 2: Above, the original image and below, the wavelet transformed image shown in spatial domain. The disturbances in the top left corner are a not consequence from wavelet

Manuscript received September 20, 2004.

M. Kutila is with Technical Research Centre of Finland, P.O. Box 1302, FI-33101 Tampere, Finland (corresponding author to provide phone: +358 3 316 3619; fax: +358 3 316 3494; e-mail: matti.kutila@ vtt.fi).

J. Viitanen is with Technical Research Centre of Finland, P.O. Box 1302, FI-33101 Tampere, Finland (e-mail: jouko.viitanen@vtt.fi).

transformation. They are caused by image printing technique to user interface.

The low pass filtered image is similar to original image, but edges are blurred and image size is smaller due to the narrowed sub-band width. From the high pass filtered bands, sharp edges and defects can be found.

III. Variable Length Coding

Wavelet is a transformation which compresses significant information from the original image to narrower energy sub-bands. The transformation does not reduce image size. On the contrary, after transformation more bits are needed to preserve information. For actual compression, variable length coding (vlc) is needed after the transformation [4].

The purpose of the transformation is to reduce the number of gray levels which are needed in image reconstruction. The number of the quantisation levels has a direct effect onto the image quality and compression ratio. Decreasing the number of quantisation levels causes higher quality loss, but produces smaller images.

High frequencies do not need as many quantisation levels as lower bands because their amplitude typically is much lower than the amplitudes of the lower frequency bands. Furthermore, the high frequency sub-bands contains detail information of the image. The low frequencies store most of the information and are more important for the image quality.

IV. Compression Algorithm in Wavecodec

The presented compression algorithm begins by executing a wavelet transformation (see. Figure 3). The wavelet transformation is done in four nested loops. In the first round the whole image is filtered. In the second round only the low-pass filtered portion is processed, which covers 1/4 of the image. In the third round, 1/16 and in the last round, 1/64 of the whole image size is filtered. 1-dimensional banks are convolved horizontally and vertical pixel rows sequentially.

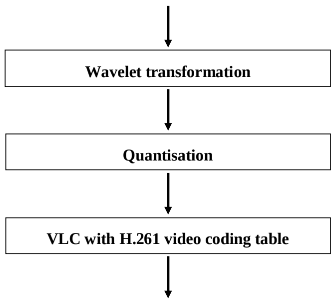


Figure 3. Description of the wavecodec application

8-tap Daubechies filters are used in wavecodec (see. Figure 4). It is not computationally efficient. Better coding efficiency would be achieved with shorter filters but the compression ratio would be worse. LeGall 5/3 and Daubechies 9/7 filters are implemented in JPEG2000. Nevertheless, the used filter is sufficient for testing purposes. In wavecodec horizontal and vertical filters are executed sequentially, starting with horizontal one.

Huffman coding is done at same time with run length encoding (RLE). This reduces computation power requirements. The used coding method corresponds to the H.261 video standard. Typical pixel values are fetched from a Huffman table. The length of the zero runs is appended in the Huffman code word. Values not found in the table are presented with 24-bit ESCAPE codes.

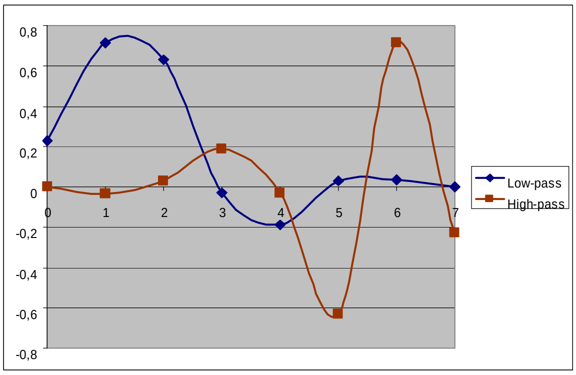


Figure 4. 8-tap Daubechies filters, which are implemented in the wavecodec algorithm.

Wavecodec is developed for efficient CPU units. The intended compression ratio was 1:10 or even better. The wavelet was chosen as the transformation method because there were interests to use small images in the pre-processing stage. Furthermore, the opportunity to execute image processing parallel with compression was an attractive feature.

V. Test Methods and Tools

The developed compression algorithm is compared with JPEG codec in the same testing environment. JPEG was implemented to the same application as wavecodec. Both compression methods were built with Borland C++ Builder 6 compiler (Figure 5). The utilised JPEG algorithm was an internal component of the compiler. The testing program runs on Microsoft Windows XP operating system. The hardware of the test platform was a laptop with Intel PIII 1,13 GHz Mobile CPU. The size of the main memory was 320 MB.

Figure 5. Program which is built for testing purpose

The Nyquist pattern (Figure 6) was used as a test image for the compression tests. The pattern contains both low and high frequencies. However, in most practical applications only low frequencies are significant. Therefore, the pattern is a kind of an extreme test for the high frequency retaining properties of a method. For this reason the fire truck images are also studied in the experimental section.

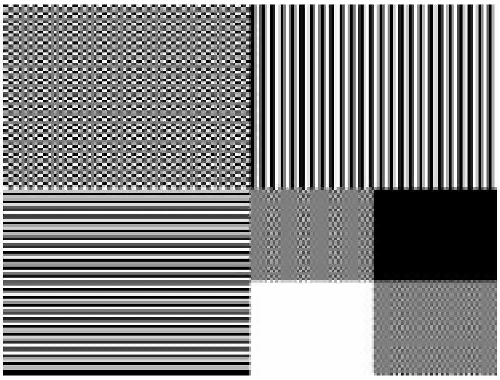


FIGURE 6. NYQUIST PATTERN

VI. EXPERIMENTAL RESULTS

As mentioned above, the most attractive feature of the wavelet-based method is the possibility to simultaneously perform other image processing functions, in addition to compression. After the wavelet transformation, horizontal and vertical filtered images are available. Figure 7 and Figure 8 show examples how to utilize high pass filters. From those images, direction related information can be extracted.

FIGURE 7. HORIZONTAL HIGH-PASS FILTERED IMAGE PARTITION

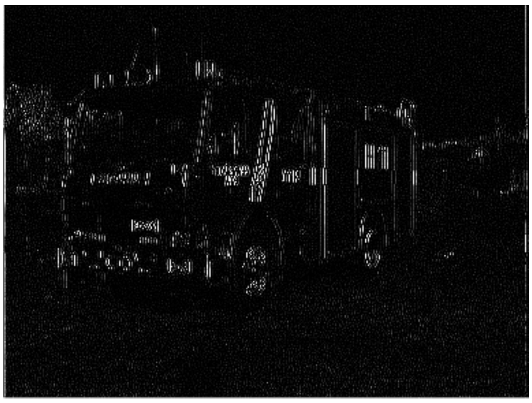


FIGURE 8. VERTICALLY HIGH-PASS FILTERED IMAGE PARTITION

After the wavelet transformation the energy of the image is mainly concentrated around the zero point, see Figure 9. The higher compression rate follows from the narrow energy band. Fewer levels are needed for maintaining image quality in the quantisation stage. The energy distribution of the fire truck image is like Gaussian and concentrated in the lower frequency bands.

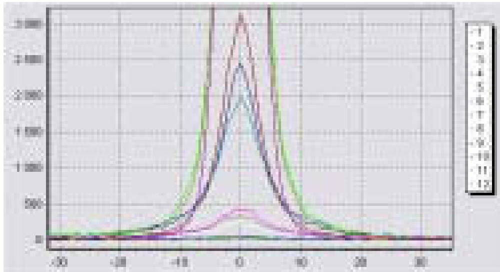


FIGURE 9. ENERGY DISTRIBUTION OF THE FIRE TRUCK AFTER THE WAVELET TRANSFORMATION

The following images show minimum and maximum values of each sub-band before and after the variable length coding. The highest frequency sub-bands are totally lost in VLC coding, because the quantisation threshold for those was high. This explains some loss of sharpness of the image in the coding stage.

	MIN	MAX
0	-337	4449
1	-1000	899
2	-1134	1009
3	-422	399
4	-992	623
5	-417	411
6	-346	212
7	-300	323
8	-358	366
9	-176	202
10	-160	180
11	-201	227
12	-105	106

	MIN	MAX
0	-57	3856
1	-1742	667
2	-1095	976
3	-417	394
4	-586	516
5	-405	389
6	-342	251
7	-261	388
8	-352	361
9	-166	196
10	-131	153
11	0	0
12	0	0

FIGURE 10. MINIMUM AND MAXIMUM VALUES AFTER WAVELET TRANSFORMATION (LEFT) AND AFTER VLC CODING (RIGHT)

Figure 11 and Figure 12 show differences between the original Nyquist pattern and the image after decompression. Practical tests indicate that wavecodec causes more losses than JPEG with a compression ratio 1:10. Peak quantisation signal-to-noise ratio for the wavelet based method is 25,9 dB and for the JPEG image 29,3 dB. This is mainly due to the coarse quantisation in the higher frequencies in wavecodec, as can be seen from the difference images.

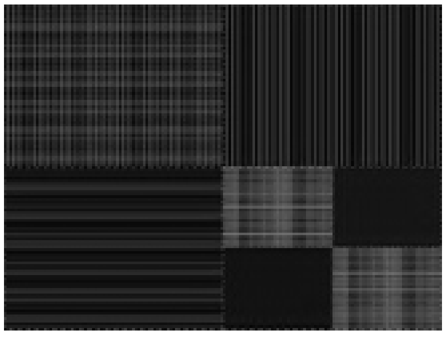


FIGURE 11. DIFFERENCE IMAGE BETWEEN WAVECODEC AND ORIGINAL NYQUIST PATTERN. PSNR RATIO IS 25,9 DB

FIGURE 12. DIFFERENCE IMAGE BETWEEN NYQUIST PATTERN AND JPEG COMPRESSED IMAGE. PSNR RATIO IS 29,3 DB

The encoding time of the wavecodec method was 311 ms for the Nyquist pattern (see. Table 1). Decoding is as fast. The wavelet transformation consumes 280 ms (see Table 2) from total coding time. The inverse transformation requires 290 ms. In other words, wavelet calculation is the bottleneck of the current implementation. However, the results are not surprising because of the iterative nature of wavelet filtering, and the rather long filter kernel used. The algorithm of the wavecodec is less complicated than that for JPEG, but because of the iterations it requires more computation power.

TABLE 1. COMPARISON OF THE JPEG AND WAVELET BASED METHODS

Method	Time	Size of the result image
JPEG	151 ms	45 KB
wavecodec	311 ms	47 KB

TABLE 2. PERFORMANCES OF THE WAVELET AND VLC ROUTINES

	Time
Wavelet transformation	280 ms
VLC encoding	31 ms

VII. CONCLUSIONS

Without any optimisation work the coding time with Borland's JPEG codec is faster than that for the implemented wavelet-based codec. The source code was not thoroughly optimised, so the performance can be increased with code analysers and processor optimised compilers. The easiest way to increase performance is to calculate the wavelet transformation with a less iterative method. Optimisation work of the algorithm is active at the moment. Thus the computation time in the final application will be lower than presented in this paper. Anyway, it is clear that the JPEG compression and decoding methods still are faster than wavelet, as reference [2] proposes. Of course, we have to keep in mind that present CPU's typically have dedicated addressing modes for DCT-type computation.

The wavelet transformation has two clear benefits compared to the conventional DCT- based compression methods. Firstly, low and high pass filtered parts are done separately, which enables frequency analysis during compression. Secondly, the small sub-images can be used in image processing (see Figure 13), which results from the nested transformations. This enables hierarchical image analysis whereas only partly reconstructing the image. If something abnormal is found, often only then the whole image needs to be decoded and analysed.

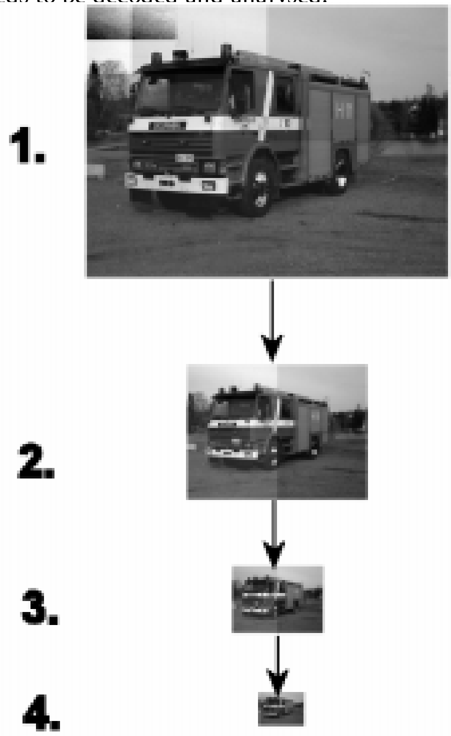


FIGURE 13. TREE OF THE LOW PASS FILTERED IMAGES

Despite the high losses, the result of wavecodec is adequate for many practical applications. One example could be detection of defects of flat surfaces and transmission of the data via a LAN link. In this case the frequency analysis in spatial domain gives the CPU efficient way to execute image processing parallel with compression.

Another example for the wavelet based method is mobile devices or wireless telecommunication. If the sub-bands are coded separately, then the quality of the image can be selected according to the transmission speed. It means that only the low-pass filtered parts of the image are transmitted which guarantees sufficient frame rate for video clips. This could solve the data overload problems of the wireless surveillance cameras.

REFERENCES

[1] M. L. Hilton, B. D. Jawerth and A. Sengupta. Compressing Still and Moving Images with Wavelets, *Multimedia Systems*, volume 2, issue 5, Springer-Verlag New York, Inc, 1994.

[2] A. Manduca, Compressing images with wavelet/subband coding, *Engineering in Medicine and Biology Magazine, IEEE*, volume: 14, issue: 5, pages 639-646, 1995.

[3] S. T. Welstead, *Fractal and Wavelet Image Compression Techniques*, SPIE Optical Engineering Press, 1999.

[4] X. Zixiang, K. Ranchandran, M. T. Orchard and Z. Ya-Qin, A comparative study of DCT and wavelet based coding, Circuits and Systems, *ISCAS '98. Proceedings of the 1998 IEEE International Symposium*, volume: 4, pages 273-276, 1998.

PUBLICATION IV

Sensor Array for Multiple Emission Gas Measurements

In: *Proceedings of IEEE International Symposium on Circuits and Systems ISCAS 2005*.
Kobe, Japan. 23–26 May 2005. Pp. 1758–1761.

Sensor Array for Multiple Emission Gas Measurements

Matti Kutila
VTT

Tampere, Finland
matti.kutila@vtt.fi

Jouko Viitanen
VTT

Tampere, Finland
jouko.viitanen@vtt.fi

***Abstract*—This paper presents a low-cost multiple gas sensing device developed for pollution measurements in a kitchen hood which is exposed to grease and dirt. The target application, which is an ozone control system, needs no accurate measurements. Low cost is achieved by using common electronics for all three sensing elements. A single processor with an user friendly interface controls the heating power of the sensing elements and reads the measurements in digital format. Two sensors can be connected to a single personal computer (PC). All the intelligence of the sensing system is in an embedded PC, which reads the raw signal from the sensor.**

The main task of the sensor system is to observe volatile organic compounds (VOC) from cooking emission and measure the ozone produced in the outlet duct. The sensor gives the total non-specific VOC amount and is not intended for special gas. Calibration with ethanol guarantees a representative enough response for most hydrocarbons in cooking emissions. The selectivity of the oxidizing type sensing element is poor. In the target application the sensing devices give an indication if more ozone is needed. The practical test showed that the response of the first VOC element is slow but stable. On the other hand, the second one reacts rapidly (< 1 s) to any changes. The combination of the sensing elements gives a very good estimate of the total VOC level. The ozone sensing element also reacts to VOCs, which is a harmful characteristic for the prospective ozone control mechanism. However, with a combination of the VOC and ozone measurements the errors in the ozone measurement can be reduced. Thus the sensor is like a simple electronic nose [1].

I. INTRODUCTION

A novel low-cost sensing device for pollution gas measurements in kitchen environment is presented in this paper. The sensing device was developed in the EU-funded Nozone Project. VOC in cooking emission are a problem in catering kitchens. VOCs cause unpleasant odour in restaurants. Use of grease implies fire hazards in ventilation ducts. Thus, the need to neutralize VOCs and eliminate smells is clear.

It is known that ozone can be used to destroy undesirable organic compounds [2]. The neutralisation process results carbon dioxide (CO_2) and vaporised water (H_20). The goal of the project was to build a system to adjust the level of ozone in the industrial hood. Ozone and VOC measurements are used for maintaining sufficient ozone level in the reaction chamber. The ozone level in the outlet must meet the EU regulations. The industrial objectives were to restrict residual ozone level to 0,1 PPM within average of 4 hours or 0,3 PPM over 10 minutes and capture VOC emission to ±0,01 mg/m^3 accuracy. Cooking produces many organic gases but only few of the resulting odours are problematic. The sensing elements were selected according to the research work carried out by other work package in the project. The work indicated that hydrocarbons are the most important emission gas.

The price was a key factor in the development process. The new sensor was developed because current commercial sensors are too expensive. It was pointed out that the metal oxidize semiconductors are the only sensing element type, which could comprise the price limitations. A typical precision gas measurements device costs more than 1500€ [3]. The electronic components of the proposed sensor with one ozone sensing element and two VOC elements cost less than €800.

One secret of the relative low price is a simple measurement principle that reduces the number of components. The same electronics were used for the three sensing elements that included two VOC and one ozone sensing heads. Digital signal processing was performed by an external computing unit. The sensor itself consists of the commercial sensor elements and units for converting analogue signals to digital format.

II. ELECTRONICS AND SOFTWARE

The electronics is protected with an EMC tested enclosure as seen in Fig 1. The computer is connected to the sensing device and electronics with serial data communication. The sensing elements were read via operation amplifiers (Fig 2), which aids to set correct gain and operating point. The electronics also includes a low pass filter for removing the peaks from the signal.

Work is done in Nozone project, which is funded by the European Commission. The project number is EU Craft EVK4-CT-2002-30009. Full title of the project is The Development of an Intelligent Responsive Pollution and Odour Abatement Technology for Cooking Emission Extraction Systems.

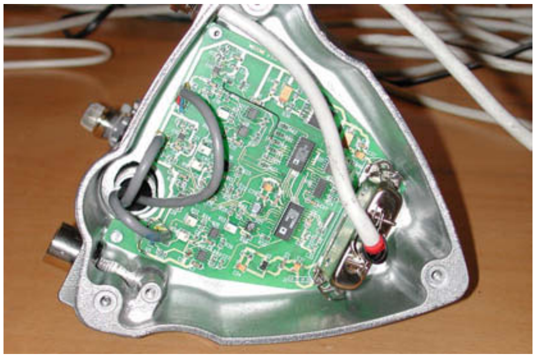


Figure 1. Electronics of the sensor device in EMC protected enclosure

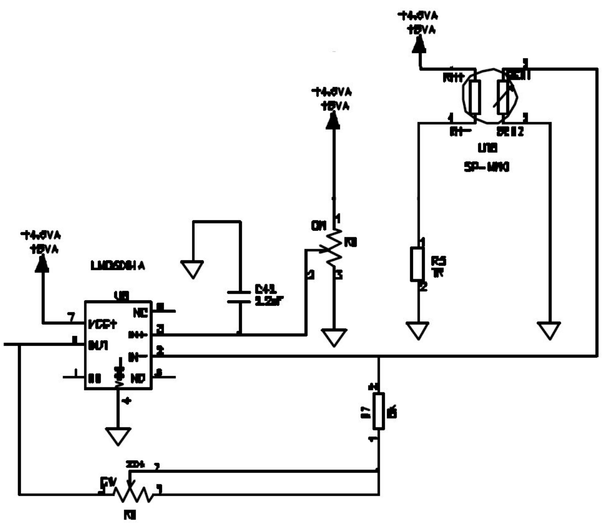


Figure 2. Schematics of the single sensing element

Both the VOC and Ozone sensing elements are based on a thin substrate layer whereby conductivity changes when the sensor absorbs gas [4]. Sensing elements, which based on oxidisation, are the cheapest ones to measure gas concentration. The drawback is the poor selectivity which is well known for this type of sensors.

The correct operating temperature is essential. The sensing elements require constant heating power. The heating period of the sensor to the correct operating temperature takes approximately 10 minutes when the sensing element is cold. The sensitivity of the sensor changes according to temperature. The control program of the sensors automatically sets the heating power to the correct value within a one second interval. The heating must be very strictly regulated, especially in the target environments where air flow cools the sensing elements.

The program reads data from two sensors simultaneously, which means that 4 VOC and 2 ozone measurements are done in parallel. The graphical user interface is compiled into the Microsoft® Windows environment. The application shows a recording for 2 hours in a line diagram. The median filter removes doubtful measurements when there is interference in communication between sensors and computer. The software also automatically resets the A/D converters whenever error bits are observed.

The rates measured during the tests are done with Intel® PIII 1,13 GHz mobile processor laptop. The sampling frequency is normally 5 Hz for the measurements. Faster sampling is possible but unnecessary in this particular application. The sampled measurements can also be stored in a file, which applies opportunity to analyze data later. The final data acquisition system is based on a low cost and compact embedded CPU board.

III. Calibration and Validation

The operating range and optimal signal level of the sensing element is adjusted with potentiometers. At first, the operation point is set so that the sensor reacts to the changes in gas concentration. Then the saturation limit is checked with maximum concentration. Finding the optimal operation point is a time consuming iterative process due to the large variation of the characteristics of individual sensor elements.

During calibration, the potentiometers are adjusted to the right operating point. The second task is to define the conversion table. The calibration table is individual for each sensing head and they cannot be duplicated directly. The look-up table (LUT) removes the errors caused by the strongly nonlinear behaviour of the sensor. Additionally, the table converts the raw electrical signal to more sensible parts per million (PPM) values. Because of the nonlinearity of the sensing head, at least seven points are needed to create the table.

Calibration of VOC elements was done with ethanol, which is a common hydrocarbon occurring in cooking emissions. The selected substance was found suitable according to practical cooking tests in the kitchen. Table 1 presents values from the calibration process of the elements. The reference was acquired with MiniRAE 2000 from RAE Systems in the laboratory. The performance of the selected VOC sensing elements is really promising.

TABLE I. Results from the validation stage of the VOC elements

	Sensor 0		Sensor 1	
Reference	VOC 1	VOC 2	VOC 1	VOC 2
90	89	97	88	97
236	232	238	228	239
322	342	304	340	311
45,6	57	44	57	43
370	386	337	384	343
182	182	165	173	165

Fig 3 shows the absolute value of the measurement error compared to ozone levels measured with the reference instrument, which was O3 41 M device from Environment s.a. The measured values are listed in Table 2. The calibration was slightly different for the two sensing elements in sensors S0 and S1. It seems that sensor S1 overshoots. Fortunately, major errors can be reduced by multiplying the values in the unit conversion table with a fixed factor. Despite the inaccuracy the sensor is still adequate for the target application.

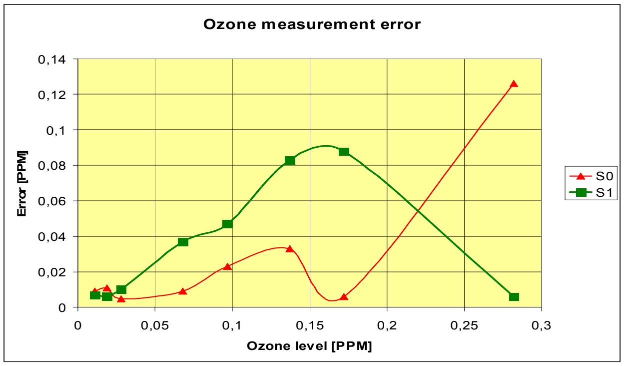


Figure 3. Errors in ozone sensing after calibration

TABLE II. MEASUREMENTS FROM VALIDATION STAGE

Ref	S0	S1
0,282	0,408	0,288
0,019	0,03	0,025
0,097	0,12	0,144
0,028	0,033	0,038
0,068	0,077	0,105
0,172	0,178	0,26
0,011	0,02	0,018
0,137	0,17	0,22

IV. SELECTIVITY OF THE SENSORS

The laboratory tests indicated lack of selectivity of the VOC sensors to the various gases. This was not surprise according to the preliminary studies [1]. The VOC sensing elements are intended to measure the total concentration of hydrocarbons. The problem is that the ozone sensor elements respond to hydrocarbons as well. The VOC gases attenuate the ozone output signal. In the kitchen hoods many gases exist, therefore the ozone value is incorrect in such an environment. The first VOC sensing element (VOC_1) responds to both hydrocarbons and ozone. The weak selectivity has also been confirmed by the manufacturer of the elements.

Fortunately, the VOC_2 sensor does not respond to ozone as can be seen in Fig 4. This means that the error of ozone measurements can be compensated using VOC_2 data at a higher priority in the ozone control system or the error in ozone sensing can be straightforwardly eliminated by using the VOC_2 readings. On the other hand, VOC_2 sensor is too sensitive to changes in emission gasses. Combination of VOC_1 and VOC_2 measurements is observed to provide a good performance for accuracy and response time.

The difference between the VOC_1 and VOC_2 sensing elements is substrate of the sensing material. Both are tin-dioxide (SnO_2) sensors, but better oxidation catalysers with suitable concentrations are added to VOC_2 sensing surface than in VOC_1.

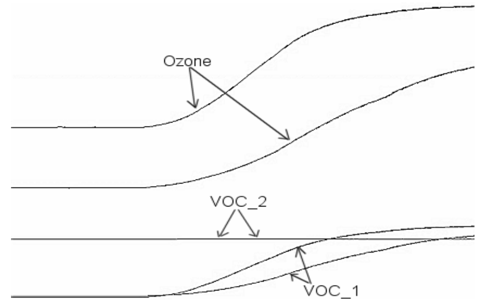


Figure 4. The VOC_2 sensor is insensitive to ozone. Pure ozone was exposed to the sensor in this experiment

V. SENSOR RESPONSE TIME

In the ozone control system a rapid response to gas level changes is essential. Fig 5 and Fig 6 show the reaction times of the sensing elements compared to the reference measurements. The problem is that changing ozone level rapidly in the chamber is nearly impossible. The ozone is mostly produced with UV tubes, which require a few seconds for ozone generation to start. However, the laboratory tests and prototype machine showed that response times are at least reasonable. Reaction to ozone and VOC level changes takes no more than 2 seconds. The sensors react fast to level changes, but reaching the final value may take up to 5-10 minutes.

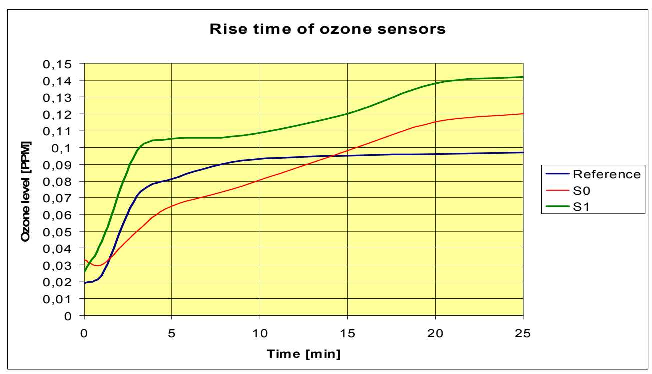


Figure 5. Response of the ozone sensing elements compared to reference measurements

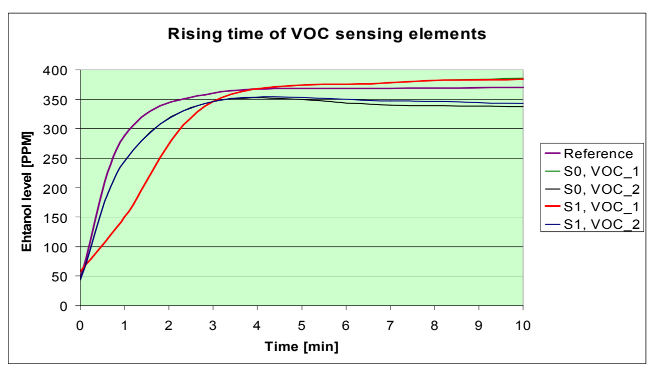


Figure 6. Response of the VOC sensing elements compared to reference measurements

VI. Experimental Results

The practical tests were done in a hood system constructed to simulate a real industrial hood system (see. Fig 7). The ventilation and air flow correspond to real kitchens (e.g. fast food restaurants). Cooking emissions in tests come from hamburgers and steaks fried under the canopy. The measurement conditions were not completely comparable for all actions in the kitchen. The goal of the test was to see if the laboratory measurements were valid in a practical environment. As already mentioned, the preliminary knowledge shows that ozone neutralizes the VOCs [2]. The test results prove this assumption.

Figure 7. Test rig which simulates industrial hood system

The sensors work well in short period tests as Table 3 indicates. The dirt did not affect the sensor readings as much as expected. In fact, the sensor needed no recalibration after three days of testing. In the tests, cooking actions were simulated and the data collected for analysis. The behaviour of the sensing elements corresponds to human observations with eyes and nose. The odours persist in the outlet duct while VOC level increased according to the sensors and the smell of the ozone corresponded with measurements as assumed.

TABLE III. The test results from prototype

Action	Ref O3	O3	VOC1	VOC2
No cooking emission		0,029	16	32
No cooking, ozone produced	0,16	0,161	0	43
More ozone		0,17	0	42
Ozone: off and cooking with oil started		0,095	33	110
VOC level stabilized		0,09	50	193
Ozone: on, VOC decreasing		0,13	0	141
No food cooked, ozone on		0,17	0	85
Steaks cooked, ozone on	0,13	0,15	3,2	170
Ozone reacts		0,16	3	110
Ozone on, cooking stopped	0,14	0,17	0	80
No cooking, ozone turned off	0,08	0,07	10	135
No ozone, sensors stabilized		0,042	21	0

The reference ozone values were measured from the same sampling point with a handheld measuring device. The results show that the ozone readings follow the reference. VOC measurements showed that VOC_1 head is slow. The VOC_2 device is actually too sensitive to the changes in the environment. The combination of the two sensing elements seems to produce really a good estimate of the VOC level.

VII. Conclusions

A combination of two different VOC sensing heads is needed for a reliable result. The accuracy of the ozone measurement is impaired by other emission gases. However, the important feature is that the VOC_2 sensor element is insensitive to ozone. This means that the error in ozone measurement can be tackled in the ozone control system. From another view, the ozone sensor is more sensitive to ozone than to other gasses. In the application, the presented accuracy of the ozone sensor is probably sufficient in practical kitchen conditions.

The calibration of the sensor should be considered carefully, when the final product is designed. Calibration of the single sensor might take even 3 hours with current method. The trimmers for adjusting correct operating point have to search iteratively. By replacing the trimmers with digital potentiometers, which can be controlled with software, would reduce calibration time to few minutes.

If no accurate measurements are needed as in the target application in a commercial kitchen environment the gas indicators work well enough. The measurement resolution of the developed low-cost sensors is appropriate. The proposed ozone control principle was shown to work with the sensors.

VIII. Acknowledgements

We would like to thank Pera Innovation Ltd., for constructing the test rig. Sintrol Oy deserves special acknowledgement for the sensing technology discussions which were fruitful and one of the key factors during the development process. We also would like to express our thanks to Enodis plc and Vent Master companies for all the input to the project. Finally, we gratefully acknowledge the support from European Commission and all the participants in the Nozone Project; Industrial Control Solutions Ltd, Dr Honle AG, Industrial Equipment Company and Wiring Solutions Ltd.

References

[1] S. Marco, A. Ortega, A. Pardo, J. Samitier, "Gas identification with tin oxide sensor array and self-organizing maps: adaptive correction of sensor drifts," *IEEE Transaction on Instrumentation and Measurement*, Vol 47, Issue 1, 1998, pp. 316-321.

[2] *Advanced Photochemical Oxidation Processes*, EPA Handbook EPA/625/R-98/004, 1998, 98 p.

[3] K. Clifford, M. Itamuram M. Kelley and R. Hughes, *Review of Chemical Sensors for In-Situ Monitoring of Volatile Contaminants*, Sandia Report SAND2001-0643, Sandia National Laboratories, 2000, 34 p

[4] S. Morrison, "Semiconducting-oxide chemical sensors," *IEEE Circuits and Devices Magazine*, Vol 7, Issue 2, 1991, pp. 32-35.

PUBLICATION V

Scrap Metal Sorting with Colour Vision and Inductive Sensor Array

In: *Proceedings of International Conference on Computational Intelligence for Modelling Control and Automation CIMCA 2005*. Vienna, Austria.
28–30 November 2005. Vol. 2, pp. 725–729.

Scrap Metal Sorting with Colour Vision and Inductive Sensor Array

Matti Kutila
VTT
Tampere, FINLAND
matti.kutila@vtt.fi

Jouko Viitanen
VTT
Tampere, FINLAND
jouko.viitanen@vtt.fi

Antero Vattulainen
Kuusakoski Oy
Heinola, FINLAND
antero.vattulainen@kuusakoski.com

Abstract

This paper presents a novel automatic scrap metal sorting system which employs a colour vision based optical sensing system and an inductive sensor array. The operation of the system is verified in a real metal recycling plant. The long period test results indicate that 80 % purity can be achieved when the feeding conveyor speed is limited below 1,5 m/s. The described system is not designed for any particular metal. However, the above separation result can only be achieved when reddish (brass, copper) and bright metals (stainless steel) are separated. The properties of aluminium, zinc, and magnesium are too similar for the current sensing principle. The results do not only depend on the sensing system, but also optimal work flow, lighting, dust and vibrations have to be considered in a practical sorting machine. The achieved purity and capacity is sufficient for industrial use. Efficient use of sensor fusion provides good performance despite the diversity of the scrap metals.

1. Introduction

This paper presents a combined machine vision and inductive sensing system intended for scrap metal sorting. The sensing principle is realised in a sorting machine called Kombi, Figure 1, in an industrial environment. The computer vision setup is based on the colour difference, or chrominance components in a CCTV signal [1]. The differences are calculated using the red channel as the common component to which amount of green and blue are compared. Such metals as copper and brass produce signals with a strong red component [2] and low response in the blue band. On the other hand, the blue component is more significant for stainless steel and aluminium.

In a practical industrial application, positions of dirt and variations in ambient lighting will complicate the separation task. Stable lighting can only be arranged in laboratory conditions. Furthermore, specular reflections from clear surfaces may cause saturated regions in the image, or direct the light rays mostly out of the view of the camera. The sorting result depends heavily on the used image pre-processing techniques that alleviate such artefacts. Proper digital filtering methods are described in this paper for achieving a reliable recognition result.

The inductive system complements the sorting process by measuring the electrical properties of the metals (i.e. it can be said that the inductive system measures invisible properties while machine vision system is for visible ones). In some cases, it is sufficient to use either the inductive sensor system or the machine vision system alone, while in some cases fusion of the decisions of the two setups is needed. An example for the latter case is coloured metals (e.g. copper and brass) for which there is no reason to use colour vision, because dirt would cause unnecessary degrade to the classification performance of the inductive system.

Figure 1. The sorting machine Kombi with inductive and colour camera sensing systems

2. System description

The hardware of the system consists of a Sony XC-003P CCD matrix camera with individual sensor elements for the three colour components. The camera allows separate adjustment of the gains of the colour channels for optimising the sensitivity according to the sorted metal. The matrix camera was selected for Kombi; however, a line scan camera can also be implemented, especially if higher resolution across the conveyer belt is needed. For noise attenuation and less demand for computational power, neighbouring pixel values are averaged. In the lighting setup, the crucial factors are sufficient intensity on the conveyer plane and suitable spectral density for colour rendering. High intensity is needed because some stained metal objects are very dark. Also, colour cameras are obviously less sensitive than black-and-white cameras. Therefore fluorescent bulbs with good efficiency were used, as shown in Figure 2. Secondly, the spectra provided by the lighting equipment should be adapted to the kinds of metals to be separated, so that both reddish metals (copper and brass) and bright metals (aluminium, steel and magnesium) receive sufficient spectral density over the bands that are important for the classifier.

Figure 2. The lighting frame in the Kombi machines

The inductive system consists of 52 sensors that measure the electric properties of metals. The output of an inductive sensor depends both on the induced eddy currents which are a function of the conductivity of the target, and the magnetic properties of the metals. Commercial inductive sensors, whose output voltages are read via an A/D converter to the computer unit, are used in Kombi. Actually, the sensors are intended to use as on/off type triggers to detect when conductive material is approaching. However, observations support an idea to utilise the sensors, which price level is low, for distinguishing metals by determining unique voltage limits for each material if calibration is done precisely.

3. Image processing

At the pre-processing stage the goal is to remove noise and irrelevant data from the image and segment the metal pieces from the background. Multiple metals may exist in a single image and they all have to be separately classified. The locations of the metal pieces are first labelled by creating a mask over the bright pixels. Dark pixels belong either to the background or they are undesired dirty spots of the objects. Dirty and saturated pixels are eliminated based on acceptable intensity levels, so only a limited number of the connected pixels around the extracted target are used. Further reduction and filtering is done by averaging the measurement points and selecting only about 30 % for the classification stage. This process eliminates most dirty pixels and other disturbances, such as specular reflections and secondary reflections from the conveyor, but still retains a large number of good pixels to be used by the classifier.

The core of the classifier is based on the use of colour difference signals i.e. chrominance values [3]. In colour space defined by the red, green and blue components, the differences relative to the red channel are calculated. The boundaries for the clusters that form the different classes are created for each metal in the two-dimensional space defined by the blue-red and green-red colour differences, as seen in Figure 3.

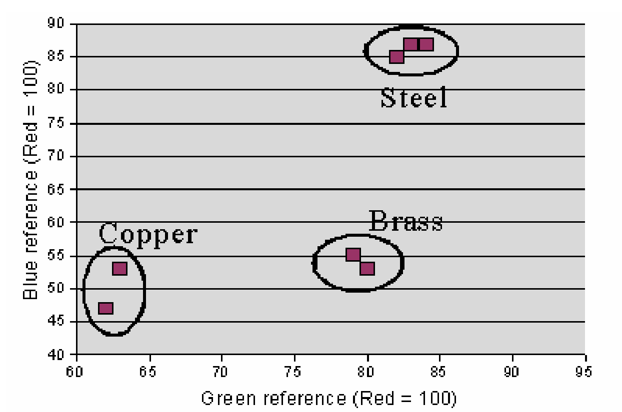


Figure 3. The two-dimensional classification space

The classification of the pixels to be used for the measurement is done in two phases. First, all the points in the metal region are used for calculating a rough estimate of the optical properties of the metal. Then the final classification is done by using only points near to the weighted centre. This method produces reliable classification because the pixels not representing good data points are ignored.

4. Inductive measurement

An inductive sensor measures the electrical properties of the metal fragments as already annotated. An electro-magnetic field is created by the transmitter in the sensor, and the induced currents are measured with a receiver. Both the conductance (through the induced eddy currents in the objects) and the magnetic properties of the objects will have an effect to the induction at the receiver. Both properties are characteristic to the specific metals or metal combinations to be sorted.

The biggest problem in an inductive sensor system is that the sensors are sensitive to the distance changes between the target and the sensor. The shape of the shredder scrap is normally complex, which causes variation in the measurements. The practical tests indicated that the response differs for planar objects and arbitrary formed scraps. When the sensors were calibrated with brass metal, the following voltages were observed in practice for the specific commercial sensor: brass 8.5 V, aluminium 6.0 V and steel 10.0 V. The range of the sensor output was 0 - 10 V.

5. Sensor fusion and software architecture

Some similarities with the proposed sensing principle can be identified from the patent [4]. Inductive and colour vision are independent but parallel systems. Practically, the reddish metals (copper and brass) are recognised with a colour vision system and remaining fragments are sorted with an inductive system. The both sensing systems provide the recognition decision before a final sorting result. If the both systems end up the same result according to sensor fusion, then a metal is accepted to a class.

Two methods are implemented to the software for connecting results of the sensing systems. When purity of the sorting is the most critical, *AND* type connection is used. It simply means that the both sensors have to accept the criteria for an entire metal region. However, typically, the camera captures metal boundaries better than the inductive system. In this case, the software includes the option that the machine vision is used to calculate metal area, and the inductive system only accepts or rejects the sorting result from the vision system.

6. Experimental Results

Two sorting machines (Kombi 1 and 2) were installed along the separation conveyor. The first one is used for separating stainless steel from the material flow. The second sorting machine recognises the reddish metals (brass and copper, Figure 4). A combination of an inductive system and a system based on computer vision was utilised in Kombi 1. In Kombi 2, colour classification was found to work even better than the combination mode of the two methods. As noted earlier, the inductive system suffers from distance variations between the sensor and the metal. For maximising the copper-brass production, the inductive system was rejected. The inductive system was dedicated to stainless steel but before Kombi 2 steel has already been separated from the material flow.

Figure 4. Copper, brass and stainless steel in camera view

Tables 1 and 2 show results from a real scrap metal sorting process in a recycling factory. Totally, 4646 kg of metal scrap (e.g. aluminium, zinc, steel, magnesium, copper, brass etc.) were sorted with the proposed technique. The speed of the conveyors was 1,5 m/s.

The practical results indicate that nearly 80 % was separated correctly in both Kombi 1 and 2. The biggest problem is caused by aluminium, which is not a surprise. The colour is similar to steel. Even with the human eye, it is hard to distinguish those two metals.

The quality of the sorted fractions is reasonable. The 80% pure fraction meets the first level industrial requirements. The purity can be increased, but as a consequence, the cost of the operation will be higher because many useful metal pieces are classified as waste like wood, plastics, etc. In fact, the current system suffers from the number of unrecognised metals, which are then separated from the residual metals. Table 3 shows that 18.9 % of waste is incorrectly separated brass and copper.

Table 1. The sorting results from the first Kombi separation unit. The purpose of this machine was to separate a stainless steel from the material flow.

Metal	Weight [kg]	Success percentage of separation [%]
Stainless steel	350.0	79.1
Al (grey)	30.3	6.8
Al (red/yellow)	3.1	0.7
Pb	1.8	0.4
Cu / Brass	22.1	5.0
Cans	2.8	0.6
Mixed	32.6	7.4
Total	442.7	100.0

Table 2. The second machine was used to find the reddish metals (copper and brass).

Metal	Weight [kg]	Success percentage of separation [%]
Cu	134.0	37.4
Brass	132.0	36.8
Cu/Brass	33.2	9.3
Al (grey)	12.0	3.3
Al (red/yellow)	10.2	2.8
Cans	0.7	0.2
Others	36.3	10.1
Total	358.4	100.0

Table 3. Material flow after separation machines

Metal	Weight [kg]	Success percentage of separation [%]
Al	48.3	50.3
Cu	5.6	5.8
Brass	12.6	13.1
Stainless steel	17.3	18.1
Cans	0.1	0.1
Others	12.0	12.5
Total	95.9	100.0

7. Conclusions

The practical implementation of the separation system has shown issues which were not considered while prototyping. In the laboratory 90 % separation accuracy was achieved, but vibrations, ambient lighting, reflections from wet belts etc. reduces the recognition result in industrial conditions. The realistic sorting result for the developed separation machine is about 80 % for stainless steel and coloured metals (brass and copper), when the speed of the belt is limited below 1,5 m/s.

An important practical aspect is that the workers in the plant normally are not engineers. Many of them have worked already before the computer era and their knowledge is restricted to normal office programs. The user interface of the machine should be simple and prevent adjustment of any configurations which could affect the recognition result. It is preferable to use fixed parameters for avoiding accidental set-up changes. On the other hand, optimal imaging conditions cannot be arranged by implementing a large number of adjustable parameters in the software. It is more desirable to use few parameters and in particular to avoid any cross references between adjustments. The camera setups and colour of the lighting are arranged for achieving maximum amount of relevant information.

The optimal configuration is not only a matter of the best adjustments in the program, but also the optimal work order must be considered. In this case, it was observed that stainless steel should be separated before the coloured metals. When the stainless steel is no longer in the material flow the coloured metals can be sorted without the inductive system which is sensitive to distance changes. In that way the effectiveness of the sensing system is enhanced. Aluminium is the biggest open issue for the Kombi machines. When the purity of the sorting result is to be improved, sorting of aluminium should be considered more carefully.

Despite the number of incorrect recognitions, the real test results indicate that a combination of the inductive measurement and colour vision works well and a sufficient separation result is achieved for practical use of the proposed sensor fusion method. The brass and copper fractions can be separated from the material flow with two-dimensional colour classification. The proposed scrap metal separation system is based on a rule-based sensor fusion method; reliable metal recognition cannot be achieved with a single sensing system. Colour vision could be seen as the heart of the system, while the inductive system acts like an assistant, allowing also measurement of the inner properties of the metals, when and the vision system only captures the externally visible features.

8. Acknowledgements

We want to acknowledge Kuusakoski Oy, which has given fruitful feedback on the development work. In particular, we would like to thank Ms. Soile Reponen, who performed an excellent job with the tests. The tests were brilliantly prepared and documented.

9. References

[1] R. C. Gonzales and R. E. Woods, *Digital Image Processing*, Addison-Wesley Publishing Company, USA, ISBN: 0-201-60078-1, 716 p., 1993.

[2] A. W. Maynard and H. S. Caldwell, Identification and Sorting of Nonferrous Scrap Metals, *Proceedings of the third mineral waste utilization symposium*, Chicago, Illinois on March 14-16, 1972.

[3] M. S. Stachowicz and D. Lemke, Image segmentation and classification using color features, *VIProm-2002, 4th EURASIP – IEEE Region 8 International Symposium on Video / Image Processing and Multimedia Communications*, Croatia, Zadar, 16-17, Jun 2002.

[4] Pat. EP0614848. 1994. *A system for identifying or validating top crowns to be used as a means for activating a circuit*. Azkoyen Ind Sa. (I. J. Echapare and I. J. L. Pina) Appl. EP19940500042, 3 Mar 1994. Pup1. 14 Sep 1994.

PUBLICATION VI

Online Detection of Driver Distraction – Preliminary Results from the AIDE Project

In: *Proceedings of the 2005 International Truck and Bus Safety and Security Symposium*.
Washington, Alexandria, Virginia, U.S.A.
14–16 November 2005. Pp. 86–96.

ONLINE DETECTION OF DRIVER DISTRACTION – PRELIMINARY RESULTS FROM THE AIDE PROJECT

Gustav Markkula[1]
Matti Kutila[2]
Johan Engström[1]
Trent W. Victor[1]
Petter Larsson[1]

[1]Volvo Technology Corporation
Gothenburg, Sweden
E-mail: gustav.markkula@volvo.com; johan.a.engstrom@volvo.com; trent.victor@volvo.com; petter.larsson@volvo.com

[2]VTT Technical Research Centre of Finland
Tampere, Finland
E-mail: matti.kutila@vtt.fi

Summary/Abstract: A preliminary version of the Cockpit Activity Assessment (CAA) module, developed as a part of the AIDE EU-funded project, is described and evaluated The CAA module is a real-time software implementing algorithms for online detection of visual and cognitive driver distraction. Algorithms for analyzing head/eye tracker output are presented, and are shown to be useful for visual distraction detection purposes although further developments are needed. The problem of cognitive distraction detection is addressed by suggesting three cognitive distraction indicators, defined so as to be robust to variations in sensor data quality, and shown to be individually sensitive to cognitive load in the driver. Finally, a support vector machine classifier, using the cognitive distraction indicators as input, is presented. On motorway, rural and suburban roads, the classifier currently reaches a 40 – 80 % accuracy at detecting a cognitive task, while maintaining an almost 80 % correct classification of non-distracted driving.

INTRODUCTION

Driver distraction is known as one of the primary causes of accidents (Neale et al., 2005). Distraction can be due to “eyes-off-road”, e.g. glances towards in-vehicle or outside targets (children, information systems, traffic signs etc.), often involving a visual time sharing between the road ahead and the target, but also to “mind-off-road” effects caused e.g. by phone conversation, voice-controlled interfaces or daydreaming. The first type of distractions will be referred to here using the general term “visual distraction”, whereas the second will be referred to as “cognitive distraction”.

Real-time monitoring of driver distraction is useful for many types of applications, e.g. distraction mitigation or warning functions and real-time adaptation of human-machine interface

(HMI) functionality (Arensberg, 2004; Almén, 2003; Claesson, 2003; Larsson and Victor, 2005; Victor, 2000, 2003). This paper describes work performed within the AIDE (Adaptive Integrated Driver-vehicle InterfacE) EU-funded project, and focuses on distraction detection mainly for the second type of application. The general goal of the AIDE project is the development of an adaptive HMI capable of integrating large numbers of Advanced Driver Assistance Systems (ADAS) and In-Vehicle Information Systems (IVIS) into a functioning whole. In this paper we report preliminary results from the development of the AIDE sub-module known as the Cockpit Activity Assessment (CAA) module, of which the main purpose is to detect driver distraction.

The specification of the AIDE system is based on a top-down approach where required functionality on the HMI level implies requirements on lower level components such as the CAA module. The required real-time distraction detection functionality of the CAA is “momentary eyes-off-road detection”, “visual time sharing detection”, and “cognitive distraction detection”. Having this information permits the AIDE system to adapt timing, modality and/or intensity of high priority messages (forward collision warning, critical vehicle diagnostics etc.) if the driver has a reduced focus on the driving, and to delay low priority information in situations where the driver is occupied with a secondary task.

Although the algorithms described in this paper have been developed for a truck platform, they are well-suited for use in a bus or car context, albeit with some minor adaptations of e.g. parameter settings.

The text will be structured as follows: First, we describe the data collection performed in order to acquire input to the algorithm development. Next, we give a description of the developed real-time algorithms. Then, results from tests of the algorithms are given. Finally, we discuss the results, suggest future work, and make conclusions.

DATA COLLECTION

To acquire data to use while developing and testing the distraction detection algorithms, twelve professional truck drivers were recruited. The drivers drove a Volvo FH12 truck, equipped with standard vehicle sensors (speed etc.), a stereo-camera based head/eye tracker and a camera based lane position sensor. The truck also contained a data logging system capable of saving all sensor and video data to a common log.

While driving, the drivers were instructed to perform a number of distracting secondary tasks, thoroughly explained before the experiment, as well as a baseline non-distraction reference task. The distracting tasks included reading sequences of numbers from stickers situated at different locations in and outside the cockpit (mirrors, speedometer etc.), using a handheld phone, operating the radio, and a cognitive task where the driver was told to repeatedly subtract an integer value, between four and seven, from an initial large integer value.

A test route was selected to contain a diversity of different traffic environments. To the extent possible the route to take was explained to the drivers before beginning the experiment. The total length of the route was about one hour.

During the experiments, two test leaders inserted annotations in real-time into the data. The annotations contained information on beginnings and ends of tasks, as well as traffic environment information. Traffic environments were defined as: "City" (50 km/h speed limit or lower in a built-up area), "Motorway" (90 km/h speed limit or higher on Swedish road class "motorväg"), and "Intermediate complexity" (not city or motorway, in practice covering rural and suburban roads).

REAL-TIME ALGORITHMS

An overview of the CAA module's distraction algorithms is given in Figure 1. In the following sections, the sub-modules depicted in this figure will be described separately.

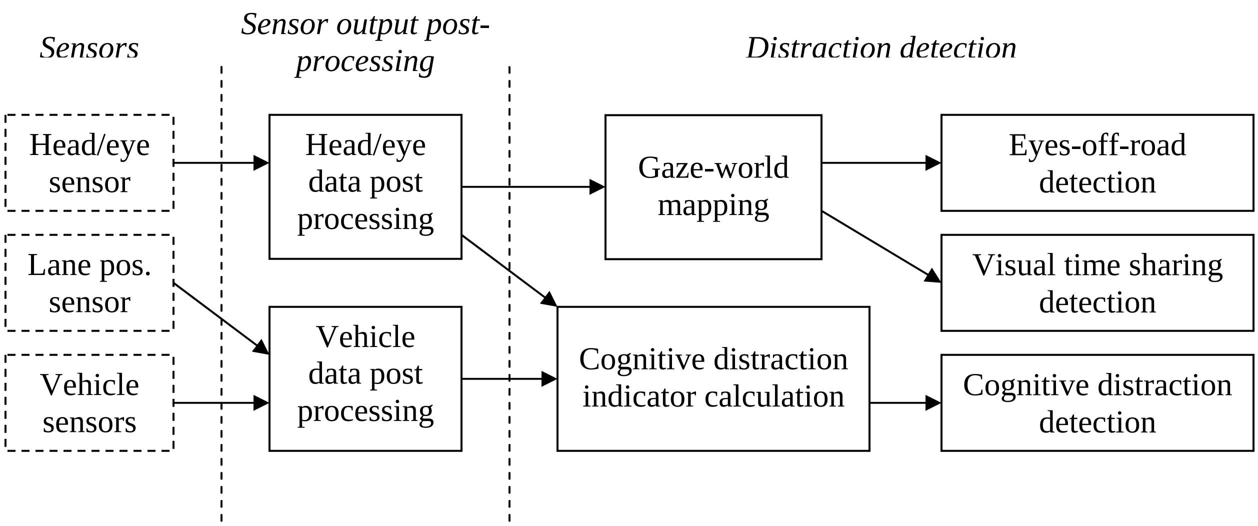


Figure 1. Overview of the distraction detection algorithms of the CAA module.

Head/Eye Data Post Processing

The output from the head/eye tracking system includes head position, head orientation, gaze orientation, saccade and blink identification, as well as confidence values for most of these quantities. The post processing of the head/eye output proceeds in a number of steps, as outlined below. These processing steps are further outlined in detail in Larsson and Victor (2005).

- *Noise filtering.* A 13 sample (≈0.2 s at 60 Hz sample rate) median filter is applied to reduce noise in head position, head orientation and gaze orientation signals. New confidence values for the filtered signals are calculated as average values of the raw signal confidence values in the 13 sample time window.
- *Saccade and blink removal.* Gaze orientation values measured at samples where a saccade (a rapid eye movement between two fixations) or an eye blink is identified, are disregarded from further calculation.
- *Head position dependence removal.* The head/eye tracker reports gaze and head orientation angles relative to a coordinate frame that moves with the driver's head. To remove the head position dependence, projections of the gaze direction vector and the head normal direction

vector are made onto the inside of an imagined sphere, of such size and position that the sphere's perimeter follows the surface of the instrument panel as closely as possible. The location of the projected point is expressed as a yaw/pitch angle pair measured along the surface of the sphere.

- *Calibration of gaze signal using road-ahead peak.* A constant error offset, different between drivers, is removed from the head/eye tracker output by using the fact that drivers' gaze angle distributions generally exhibit a sharp "road-ahead peak" that can be located in real-time (see Larsson and Victor, 2005; and Victor, Harbluk, and Engström, 2005). After finding this peak, all head and gaze angle data is translated so that the road-ahead peak lies at the origin. In this way, gaze and head angle values are comparable between drivers.

Vehicle Data Post Processing

The vehicle data post processing algorithms consist of a simple averaging filter, used to reduce noise in the vehicle speed signal, and an algorithm for calculating a robust single lane position value from the left and right lane marking distances reported from the lane position sensor.

Gaze-World Mapping And Eyes-Off-Road Detection

The purpose of the gaze-world mapping step is to map gaze and head angles onto actual real world targets of visual attention. In the distraction detection context only one such target is used, the "road-ahead" target. (The CAA has other purposes in which mapping to other objects, such as mirrors, are relevant, but this is not described here.) In the current implementation, the size and location of the road-ahead target is static, determined offline by inspecting a distribution of gaze angles for road-ahead data coming from a number of different drivers, and manually enclosing this distribution in a rectangle. A similar rectangle is fit to road-ahead head angle data. In online mode the mapping consists of determining whether or not a measured gaze angle falls within the rectangle that was fit to the gaze data, or, in situations where the sensor only manages to track the head and not the eyes, whether a measured head angle falls within the head data rectangle. The eyes-off-road detection step of the CAA algorithms merely consists of outputting "on road" or "off road", depending on the output from the gaze-world mapping.

Visual Time Sharing Detection

This algorithm has yet to be implemented in the AIDE project. However, several visual time sharing detectors are implemented and described in Larsson and Victor (2005), for example a 3-10 second moving time window over gaze classified as on/off road. The preliminary idea here is to use a simple rule reminiscent of the eyes-off-road rule, measuring the driver's division of visual attention between the road ahead of the vehicle and other visual targets. However, while the eyes-off-road detection measures short momentary distractions from the road scene ahead, the visual time sharing detection is used to measure longer term visual distractions.

Cognitive Distraction Indicator Calculation

In recent publications, a number of behavioral effects observed when cognitively loading the driver have been reported, different from the effects typically observed when using visual loads. In the cognitive distraction indicator calculation step we calculate indicators with potential of being sensitive to two of these behavioral effects. The first is the decrease in visual scanning during cognitive tasks reported in e.g. Recarte and Nunes (2003), and Victor, Harbluk, and Engström (2005) The second is the decrease in lane position variance reported in Engström, Johansson, and Östlund (2005).

The indicators implemented in the current work are: *standard deviation of gaze angle* (Victor, Harbluk, & Engström, 2005), *standard deviation of head angle* (hypothesized to also decrease with decreased visual scanning), and *standard deviation of lane position*. To calculate the head and gaze indicators we need scalar values instead of yaw/pitch angle pairs. For this purpose we calculate the Euclidean distances from the yaw/pitch angle points to the origin, as measured in the spherical reference frame. The indicators are calculated in a sliding time window of fixed length. Parts of the time windows where sensor data confidence is lower than fixed minimum values are excluded, and standard deviations are calculated on the remaining data. To control the effect of this shortening of the effective size of time windows, a "quality factor" of each indicator value is calculated as the average of sensor data confidence in the time window. In this average, confidence values count as zero if lower than the minimum required value.

Cognitive Distraction Detection

To detect cognitive distraction in the driver, a support vector machine (SVM) classifier is used. This is one of the approaches for cognitive distraction detection suggested in Lee et al. (2004) and is an alternative to the algorithm implemented in Larsson and Victor (2005). The output of an SVM is a scalar value, and a threshold value is then applied as a cut-off to get a classification into one of two classes. See e.g. Burges (1998) for a complete description of SVM classifiers. The input to our classifier consists of the cognitive distraction indicators described above. A radial basis function SVM kernel is used, as defined in Burges (1998). Training data was extracted from the experimental data by calculating indicator values for non-overlapping time windows containing data from baseline driving and driving with the cognitive task, respectively. Only time windows with "quality factor" values higher than a minimum level for all indicator values were used. Different SVM output thresholds were tested as a means of managing the trade-off between false positives and false negatives, see further under Analysis and Results.

ANALYSIS AND RESULTS

In this section we present analyses performed to test the quality of the algorithms at their current, preliminary level of development. The vehicle data post processing is not analyzed specifically, but is implicitly tested in the analysis of cognitive distraction indicator calculation.

Head/Eye Data Post Processing

As a means of qualitatively benchmarking the head/eye data post processing algorithms, measured gaze angle distributions for a number of visual attention targets (road ahead, mirrors, speedometer, tachometer, radio) were manually (and thus approximately) described for each driver separately, by enclosing the peaks in the distributions in circles. This peak identification was performed both for raw gaze angle data, and for the gaze angle data output from the post processing algorithms. The results of this is shown in Figure 2. If excluding the head position dependence removal step from the post processing algorithm, the results are very similar.

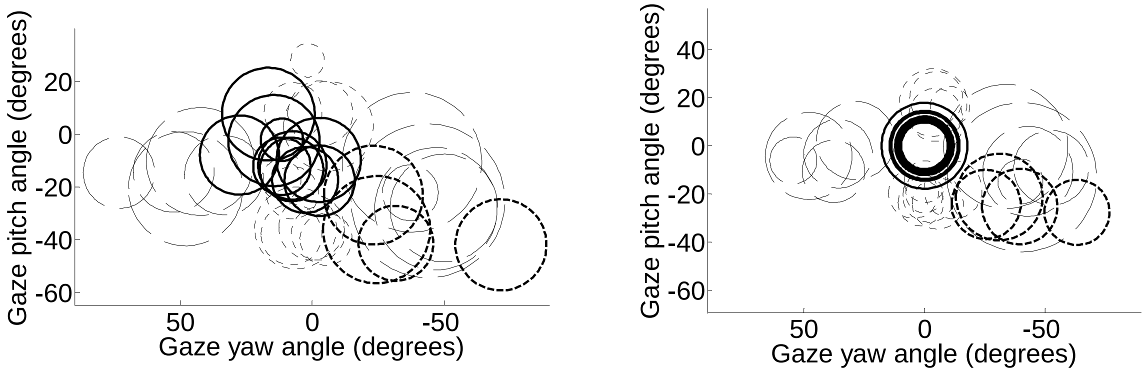


Figure 2. Approximate location and spread of gaze angle distributions for different drivers and targets, before and after applying the head/eye data post processing algorithms (left and right plot, respectively). The targets are "road-ahead" (bold line), left and right mirrors (dashed lines), radio (bold dotted line), speedometer and tachometer (dotted lines).

Gaze-World Mapping And Eyes-Off-Road Detection

The gaze-world mapping algorithm was tested by applying it to the real driving data. By comparing how the measured distribution of visual attention between on and off road targets changes from baseline driving to driving with tasks, we get a rough idea of the quality of the mapping. Figure 3 shows how the gaze-world mapping algorithm interprets the head/eye data during tasks involving different targets, for one of the drivers.

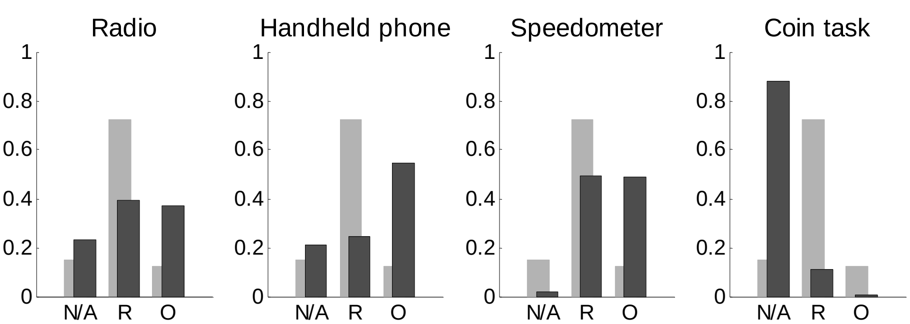


Figure 3. Distributions between targets of mapped attention for one driver during different tasks (dark bars) as compared to baseline driving (light bars). Targets are: "sensor not tracking/driver not fixating" (N/A), "road ahead" (R), and "other target" (O).

Cognitive Distraction Indicators

The sensitivity of the used cognitive distraction indicators was measured by statistically analyzing the effect of the cognitive task on extracted SVM input values from different traffic environment and task conditions. Figure 4 shows mean values and 95 % confidence intervals for all three indicators, as measured during baseline and cognitive task driving in all three traffic environments defined in Data Collection above.

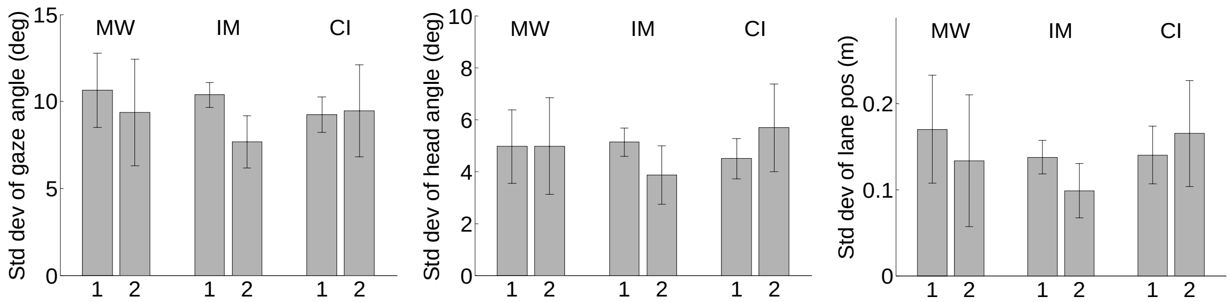


Figure 4. Mean values and 95% confidence intervals for the three different cognitive distraction indicators (15 second time window) for different tasks in different traffic environments. Tasks are baseline (1), and cognitive task (2). Environments are motorway (MW), intermediate complexity environment (IM), and city (CI).

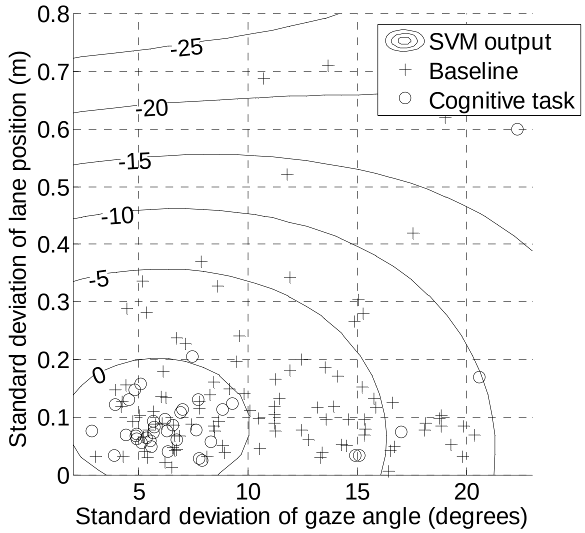
In the intermediate complexity traffic environment the cognitive task had effects significant at the 95 % confidence level on all three indicators. $t(54) = -3.29$, $p = 0.002$ for standard deviation of gaze angle; $t(54) = -2.05$, $p = 0.045$ for standard deviation of head angle; $t(67) = -2.12$, $p = 0.038$ for standard deviation of lane position. These results were obtained using a 15 second time window for the indicators. When using larger time windows (30, 60 and 120 seconds) effects are qualitatively similar, though less statistically significant since the requirement of non-overlapping time windows lowers the number of acquired indicator values as the size of the time window increases.

Cognitive Distraction Detection

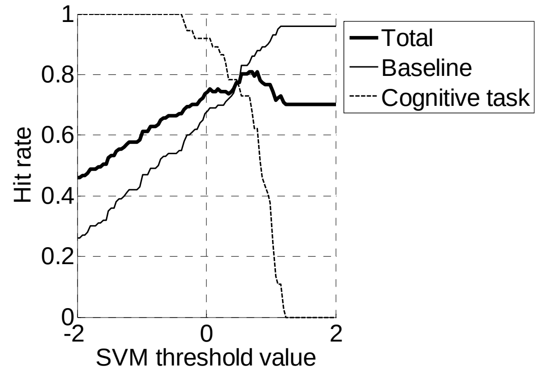
A number of different SVM classifiers were trained, using different inputs, different training data sets, etc. Here we report results for the most promising setup.

An SVM, taking all three cognitive distraction indicators as input, was trained on baseline and cognitive task data from driving in intermediate complexity environments. Using a 15 second time window, 154 baseline data points and 37 cognitive task driving data points could be extracted from the data. Ten training data sets with 20 baseline and 20 cognitive points in each, randomly selected from the complete set were created. A validation data set was also created, containing all 37 cognitive task points and 100 randomly selected baseline points. A separate SVM was then trained on each training set, and the classifier with the highest total hit rate on the validation set was selected. This classifier's output is visualized in Figure 5a. Its validation hit rate on cognitive task points is 92%, and on baseline data the hit rate is 67%. A motorway validation data set was also created, with 22 baseline points and 12 cognitive task points, and on this set the same SVM classifier has a 50% hit rate on cognitive task data, and a 68% hit rate on baseline data.

By modifying the SVM output threshold value (default is zero), we can to a certain extent manage the trade-off between false positive and false negative classifications, Figure 5b illustrates this. Judging by this figure, using a threshold value of for example 0.5 seems reasonable, yielding hit rates of 76 % and 78 % for cognitive task and baseline, respectively. Using the same threshold when classifying the motorway validation set we get hit rates of 42 % and 77 %.



a.



b.

Figure 5. SVM classifier results. a) A two dimensional representation of the output of the three dimensional SVM classifier, with the "standard deviation of head angle" input value held constant at its median value 4 degrees. The intermediate complexity environment validation data set is also shown. b) Effects of SVM output threshold value on hit rates for the intermediate complexity environment validation set.

DISCUSSION

Visual Distraction Detection Algorithms

Figure 2 suggests that the head/eye post processing algorithms are successful at translating gaze data for different visual attention targets into different, fairly separated clusters. This is a necessary prerequisite if one wants to create a working gaze-world mapping that uses the gaze data. However, our analyses have not been able to establish that the (theoretically sound) head position dependence really improves the algorithm. Its effects are minor in comparison with the effects of the "calibration" step. Insufficient sensor output quality seems a reasonable explanation to this phenomenon, but further investigations are needed.

The gaze-world mapping works fairly well, at least on an average for the specific driver for which results are shown in Figure 3. Actual attention towards the radio, a handheld phone or the speedometer causes the measured attention to decrease for the "road-ahead" target, and an increase is seen for the "other target" target. For the coin task, the reduction in "road-ahead" time is accompanied by an increase of sensor non-tracking. This could be expected, since the location of the coins was such that the drivers had to lean forward to pick them up, often out of view of the head/eye tracker cameras. It could be investigated to what extent non-tracking can be used as an indicator of visual distraction. Overall, a more thorough analysis of the gaze-world mapping would be needed to completely assess its usefulness in its present form. However, modifications to this algorithm are foreseen, why analysis efforts have been kept at a less ambitious level.

Cognitive Distraction Detection Algorithms

The effect of cognitive distraction on the suggested cognitive distraction indicators follow well the hypothesis of decreased visual scanning and lane position variance, but only for the intermediate complexity environment. That there are no clear effects for the highly complex city environment was expected, but the motorway was included in the experiment as an assumed lowest-complexity environment, where the effects of cognitive distraction could be expected to be most clearly visible. However, it may be that the assumption of the motorway environment having a low complexity does not really hold, since the motorway segments used in the data collection were close to a city, with frequent entrances and exits, and always at least some surrounding traffic. It is the impression of the test leaders that this can have caused the drivers to refuse to allocate any substantial cognitive resources to the secondary task, instead putting effort into maintaining safe levels of situational awareness.

The accuracy of the SVM classifier is fair but not excellent. It is actually obvious from Figure 5 that it will be difficult, if not impossible, for any classifier to completely separate baseline and cognitive task data points from each other, since there is a considerable overlap between the classes (at least in the two input dimensions shown in the figure). One cause of this overlap could be variations between drivers in their reactions to a cognitively loading task, why some kind of online adaptation of the algorithm to the current driver could be a viable approach.

Finally, the SVM threshold proves to be a useful parameter for tuning the algorithm behavior. It can be used to optimize hit rates for the different classes given 1) actual frequencies of the two classes in real data (not well known here), and 2) strategies on whether it is more important to avoid false positives or false negatives. For our classifier, one option could be to move the threshold value from zero to about 0.5, which then seems to give us a classification performance of just under 80 % on non-distracted data, and 40 - 80 % accurate cognitive distraction detection in intermediate complexity and motorway environments. This may or may not be an acceptable level of performance, depending on how the cognitive distraction output is to be used.

FUTURE WORK

There are many opportunities for future work with the presented algorithms, some of which will be pursued within the AIDE project. Possible examples include real-time adaptation of algorithms to individual drivers, and addition of more input dimensions to the SVM classifier. An important future work item is a forthcoming verification in online mode.

CONCLUSION

We have described a set of real-time algorithms for detection of driver distraction, of both visual and cognitive type, and we have evaluated the performance of our algorithms on real driving data collected in a distracted driving experiment using professional truck drivers. The visual distraction algorithms show promising results, but some further development and testing is needed. As for cognitive distraction detection, we have been able to reproduce results reported elsewhere in the literature on some of the effects of cognitively loading the driver, and we have described a support vector machine classifier that uses indicators sensitive to these effects as input. The performance of the classifier is fair, but tests in online mode are needed.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the contribution to the support vector machine implementation made by Maria Jokela of VTT Technical Research Centre of Finland.

REFERENCES

Almén, L. (2003).Reducing the effects of driver distraction: A comparison of distraction alerts on driver attention. Licentiate thesis No. 1017, Linköping University, Linköping, Sweden.

Arensberg, U. (2004). Tiredness and distraction are major sources of danger in traffic – the truck monitors how you are feeling. Press release 18 May, 2004. Volvo Truck Corporation.

Burges C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery 2. pp 121-167.

Claesson, C. (2003). Volvo trucks researches driver distraction. Press release 5 May, 2003. Volvo Truck Corporation. Available at: http://www.volvo.com/trucks/global/en-gb/newsmedia/press_releases/

Engström, J. A., Johansson, E. K., & Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Psychology and Behaviour*, 8(2), 97-120.

Larsson, P., & Victor, T. (2005). Method and arrangement for interpreting a subject's head and eye activity. United States Patent Application, US 2005/0073136 A1.

Lee, J., Reyes, M., Smyser, T., Liang, Y. & Thornburg, K. (2004). SAfety VEhicles using adaptive Interface Technology (Task 5) Final Report: Phase 1 of the SAVE-IT project. Available at: http://www.volpe.dot.gov/opsad/saveit.

Neale, V. L., Dingus, T. A., Klauer, S. G., Sudweeks, J., and Goodman, M. (2005). An overview of the 100-car naturalistic study and findings. *Paper presented at the 19th International Technical Conference on Enhanced Safety of Vehicles (ESV)*, Washington, DC, June 6-9, 2005.

Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: Effects on visual search, discrimination and decision making. Journal of Experimental Psychology: Applied, 9(2).

Victor, T. (2000). On the need for attention support systems. Journal of Traffic Medicine. Vol. 28, No. 2S.

Victor, T. (2003). System and method for monitoring and managing driver attention loads. United States Patent Application, US 2003/0181822 A1.

Victor, T. W., Harbluk, J. L., & Engström, J. A. (2005) Sensitivity of eye-movement measures to in-vehicle task difficulty. Transportation Research Part F: Psychology and Behaviour. 8(2), 167-190. DOI: 10.1016/j.trf.2005.04.014

PUBLICATION VII

Driver Cognitive Distraction Detection: Feature Estimation and Implementation

Submitted to *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering* on 11 April 2006.

Driver Cognitive Distraction Detection: Feature Estimation and Implementation

M Kutila[1], M Jokela[1], T Mäkinen[1], J Viitanen[1], G Markkula[2,] & T. W. Victor[2]
[1] VTT Technical Research Centre of Finland, P.O. Box 1302, FI-33101 Tampere, FINLAND
[2] Volvo Technology Corporation, Dept. 6400, M1.6, Götaverksgatan 10, 405 08 Gothenburg, SWEDEN

Abstract: This article focuses on monitoring a driver's cognitive impairment, due to talking to passengers or on a mobile phone, daydreaming or just thinking about something else than driving-related matters. This work is part of the AIDE project, which aims at scheduling the increasing information flow of navigation systems, IVIS, nomadic devices, etc. in future vehicles and thus at providing the driver with a possibility to understand and assess incoming messages and, more importantly, protect him/her against distractions when the traffic situation requires increased attention. The whole system is based on estimating the driver's momentary awareness of the surrounding traffic environment and the driver's physiological state. Cognitive workload has a major role when the driver's alertness is estimated. This paper describes an investigation of cognitive distraction, firstly giving an overall idea of its effects on the driver, and secondly discussing the practical implementation of an algorithm for detection of cognitive distraction using the Support Vector Machine classifier. The tests performed show that cognitive workload can be detected with approximately 65 - 80 % confidence despite the fact that the test material represented medium difficulty cognitive tasks (i.e. the induced workload was not very high). The assumption is that a more challenging cognitive task would yield better detection results.

The AIDE integrated project (IST-1-507674-IP) is supported by the European Commission as a part of eSafety iniative and the EUCAR Information Society Technologies program in the Sixth Framework Programme.

Keywords: Vehicle, HMI, Traffic Safety, Cognitive distraction, Support Vector Machine, Classification, Driver monitoring, Identification

1. INTRODUCTION

Driver information overload as a source of distraction has been a topic for traffic safety research since the late 1920's when the first radio receivers were introduced as an option for cars. With the landing of mobile phones, the issue of distraction has again received widespread attention. Furthermore, this problem may become even more critical in future vehicles, when an increasing number of ADAS (Advanced Driver Assistance System), IVIS (In Vehicle Information System) and nomadic devices such as PDAs, mobile phones, mp3 players, etc. are introduced to assist and entertain drivers. The information introduced from various sources to the cockpit needs intelligent management systems. To keep the driver’s levels of distraction within reasonable levels, some technical applications for monitoring a driver is needed to manage and prioritise the information presented to the driver. Therefore, the European Commission has released the AIDE (Adaptive Integrated Driver-vehicle interfacE) project to develop the necessary technologies.

It is a common interest of the automotive industry and the authorities responsible for traffic safety to develop a technology for managing the drivers' increasing information load, so that they can handle both visual and acoustic information timely and focus their main attention on the driving task. A large field study of a hundred vehicles indicated that secondary tasks such as using wireless devices and vehicle or passenger related workload have a major role in the crash or near-crash incidents [1]. The test period covered one year including 100 vehicles and 241 drivers. Some 43 000 hours of video on vehicle state and kinematics data were recorded. One possible way of addressing the inattention problem highlighted by the study is to develop an intelligent Human-Machine Interface (HMI) capable of adapting the information from various sources to the driver's state and the driving situation. This means in practice that the information to be presented needs to be prioritised according to urgency and traffic situation e.g. by selecting appropriate modalities in order to ensure that the driver can optimally concentrate on the driving.

The AIDE architecture will consist of five independently running Driver-Vehicle-Environment (DVE) monitoring modules [2], which are interacting with the Interaction and Communication Assistant (ICA) module. The purpose of the ICA is to conclude whether different driver and environmental parameters suggest that an HMI adaptation is necessary. The modules are listed in Table 1 with a short description of their function and relevance to the AIDE concept.

The development of DVE modules was initiated by addressing the use cases and defining the requirements for understanding, which driving-related variables would be useful for effectively scheduling the HMI information. The Cockpit Activity Assessment (CAA) module is dedicated to the detection of secondary task activities such as inattention to the road ahead or mental workload, which affect the controlling of a vehicle. The AIDE architecture includes also the module DAE (Driver Availability Estimator) the objective of which is to estimate the demands set for the driver by the primary driving task, i.e. how much attention is required from the driver by the actual driving context. A few exemplary scenarios are listed in Table 2 to clarify the idea behind adapting the vehicle's HMI in accordance with the CAA outputs.

This article focuses particularly on explaining the methodology, the principles and the evaluation results of examining cognitive distraction. The CAA module also provides the level of visual distraction, which is the measure of how much the driver pays attention to the road ahead and, on the contrary, scans the surrounding environment. However, the cognitive impairment is of special interest here and more challenging to estimate, since we are dealing with human mental activity. Cognitive workload may occur due to daydreaming, thinking, talking on a phone etc. The research hypothesis of this article originates from the results of the HASTE project [3], which suggest that a cognitive task creates gaze concentration to the road ahead, which in turn actually improves lane keeping. Our attempt is to prove that a multi-dimensional Support Vector Machine (SVM) -type classifier can detect cognitive workload by combining a set of driver and driving related parameters, which are discussed in greater detail in section 3.

2. PRIOR STUDIES ON COGNITIVE DISTRACTION RELATED ACTIVITIES

The HASTE project [3], which proposed methodologies and guidelines for the development of novel in-vehicle HMI technologies, identified a need of handling cognitive and visual workload separately. One of the project's major conclusions was that the lateral driving indicators, namely steering control and lane position maintenance improve, when the driver is in a cognitively impaired state [3, 4].

The idea of utilising the SVM algorithm is not only adopted by AIDE but was also raised by the SAVE-IT project [5]. Nevertheless, in SAVE-IT it was decided to implement the Hidden Markov Model (HMM) in the first phase of the project with a limited success. The benefit of the HMM is that historical events are "automatically" adopted when the workload is estimated. The input parameters of the SAVE-IT classifier were eye movements and advanced eye based measurements like saccades, fixation duration and blinking. HMM performance was evaluated by investigating the method's capability to predict [5] interaction of a driver with an IVIS (In Vehicle Information System), which induced cognitive workload. In the second phase the traditional linear regression method and the support vector machines were considered to be worth evaluating.

Engström et. al. [4] and Victor et al. [6] describe tests with fixed and moving base simulators and compared the results with those of field tests. They applied visually and cognitively loading tasks for a driver and analysed gaze movement, physiological, lateral and longitudinal data by comparing the measurements to the baseline driving. They observed speed reduction and larger steering wheel movements for maintaining the lateral position during visually loading tasks. Cognitive workload, induced by using an auditory task, was associated with improved lane-keeping and increased micro-movements of the steering wheel. Another major result was that the auditory task, which is a more cognitive than visual distraction, also caused increased gaze concentration towards the road.

In response to increased complexity of cognitive or auditory tasks, drivers increase their road viewing time and spatially focus their gaze on the road centre region at the expense of peripheral glances. This gaze concentration effect has reliably been observed during cognitive and auditory tasks [6, 7, 8, 9, 10], but also in connection with demanding driving conditions, alcohol and fatigue (see. [7]). Significant reductions in horizontal and vertical variability (i.e. standard deviation) of gaze direction, longer on-road fixations (more intense staring), and reduced glance frequency at mirrors and speedometer are typically observed.

The gaze concentration associated with cognitive tasks is strongly related to a loss of event detection capability across the entire visual field (e.g. Recarte et al. [10]), but there is little interference with path control. In a recent meta-analysis of the impact of talking on a mobile phone, Horrey et al. [11] conclude that the reduction in driving performance caused by talking is primarily on reaction time, not tracking performance. Engström et al. [4] suggest that the increased spatial gaze concentration in the auditory tasks is associated with improved lane keeping performance. This finding was consistent across six experimental sites in HASTE, and is in line with other research results.

Taken together, these findings suggest that the visual guidance of path control is more immune to interference from cognitive tasks than the identification and planning tasks that require glances at the visual periphery (e.g. speedometer, signs, vehicles, pedestrians). A by-product of increasing visual concentration on the road centre area is better lane keeping performance.

These results are congruent with biased competition approaches of attention [12] and functional differences in vision between vision-for-action and vision-for-identification [7, 13]. In short, visual guidance of path control appears to be more immune to interference because vision-for-action (the dorsal stream) utilizes non-conscious vision-action links, and because gaze spends more time on the distant path region, whereas vision-for-identification (the ventral stream) is more affected by conscious cognitive tasks. Increased attention to a subset of cognitive or auditory stimuli amplifies that stimulus but it also simultaneously inhibits other stimuli. The inhibition of unattended stimuli means poorer stimulation from stimulus-driven attention, leading to less stimulus-driven eye-movements. Likewise, the engagement of goal-directed attention to in-vehicle tasks reduces the capability to react to other goal-directed attention stimuli.

The prior studies raise the research hypothesis that the driver's cognitive distraction can be detected automatically by utilising the driver's gaze and lane keeping information since they were discovered to decrease due to cognitive workload. Victor [7] describes an implementation of a real-time PRC algorithm that identified cognitive distraction and gave a specific cognitive distraction alert (two flashing lights in the side portions of the windshield). Both off-line and real-time recognition of cognitive state have been implemented using both the Percent Road

Centre (PRC) and Standard deviation of gaze (SDG) measures [6, 7, 14]. A cognitive distraction alert was conceived, developed in several versions, and tested in a driving simulator and on-road. The real-time algorithm was based on calculation of Percent Road Centre wherein a cognitive distraction alert was issued when PRC reached 92 %. However, Victor et al. [6] show that SDG is more sensitive than PRC.

Here, the SVM method is chosen because it has capability to recognise patterns (i.e. the feature vectors) in a multi-dimensional feature space. Thus, the head movements and the lane keeping information were added to descriptors of cognitive workload for increasing reliability of the previous gaze based detection algorithms, which then is hypothesised to work also when the cognitive indicative signals are weak.

3. INDICATIVE PARAMETERS OF COGNITIVE DISTRACTION

As a part of the present work within AIDE, test data of distracted driving experiments was gathered involving twelve professional drivers driving an instrumented Volvo FH12 truck. Each of them drove the truck for about an hour, in different traffic environments described by the experimenters as ”motorway” (low complexity environment), ”city” (high complexity) and ”intermediate complexity environment”. One of the distracting secondary tasks performed by the drivers while driving was a purely cognitively loading task consisting of repeated subtractions from a large integer number. In addition to head and eye data and vehicle data, video recordings of the road ahead were also made during the data gathering.

During the data gathering procedure the subjects were asked to undertake different kinds of tasks, which were assumed to cause mental workload and thus simulate real cognitively distracting events. Examples of such tasks include using a hand-held device, picking up coins, summing up a sequence of numbers, which were stuck on the cockpit, or discussing with a test leader. The test data was manually labelled with markers, which were added during the data gathering procedure when the cognitive workload for the driver was induced.

Three parameters were chosen as the main candidates for being indicators of cognitive distraction in this test data:

- *Standard deviation of gaze angle*. This is calculated as the standard deviation, in a time window, of the quantity $\sqrt{\varphi^2 + \theta^2}$, where φ and θ are the yaw and pitch angles of the driver's gaze, respectively.
- *Standard deviation of head angle*. The same as standard deviation of gaze angle above, but calculated from the driver's head movements.
- *Standard deviation of lane position*.

All three indicators were measured in a time window of the same length. Figure 1 illustrates their sensitivity to the cognitive task sequences in the data set, for a 15 second time window. Two-sided *t*-tests, without assuming equal variances and making the Smith-Satterthwaite approximation for effective degrees of freedom [15], were performed on these results. Statistically significant ($p < 0.05$) decreases in deviations from baseline values, in accordance with the previous results discussed above, were found for all three indicators in the intermediate complexity environment, whereas the effects were less pronounced in the motorway environment and inconsistent in the city environment. The inconsistent effect in the urban area was expected, due to the high complexity of the environment. The weaker effect obtained in the motorway environment can possibly be explained by the choice of the motorway segment, featuring many exits and entrances where the traffic situation may have caused the driver to refuse to allocate any substantial cognitive resources to the secondary task. A combined set of motorway and intermediate complexity environment data was identified as a suitable data set for SVM training.

Since further analysis of the acquired data also indicated that there might be a dependency between cognitive load and observed head/gaze tracking quality output from the head/eye sensor, three additional potential cognitive distraction indicators that could be suitable SVM inputs were defined:

- *Gaze angle quality factor*, calculated as the mean gaze signal quality value given by the stereo vision system during the time window when the gaze tracking is valid and multiplied by the fraction of the time window when the data is available (i.e. this is like a combined measure of quality and completeness of the data in the captured time window).
- *Head angle quality factor*, similar to the previous parameter, but for head angle.
- *Face model quality factor*, basically a longer time window version of *head angle quality factor*. This parameter was added since it was hypothesized that the deviation of the observed *head angle quality factor* from a longer term average could be a more efficient cognitive distraction indicator than the *head angle quality factor* itself, and since it was noticed that there were considerable variations in average head tracking quality levels between individuals.

4. SUPPORT VECTOR MACHINES FOR COGNITIVE DISTRACTION DETECTION

Support vector machine (SVM) is a classification method, which optimises the locations of hyperplanes in such a way that the margin between the negative and positive feature examples is maximised. Our work utilises the well-known SVMlight algorithm [16]. It was in our interest to focus more on testing the feasibility of the SVM to the highly non-linear input data, in which the cognitive indicators are quite poorly interpretable. Thus, the previously tested SVM gave us an opportunity to avoid too heavy resource allocation for writing the classifier and allowed us to put more effort on optimising the performance. Therefore, the SVM principles are not explained here in depth but rather the goal is to identify the constraints in order to detect in-vehicle cognitive distraction.

Typically, SVM has a very good generalisation property and consequently, the classification performance of the method is not restricted to any specific application or type of data. Despite the fact that this discussion focuses on a binary type classifier since only a true or a false detection with confidence estimation is necessary, the algorithm can be adapted to multidimensional environments by splitting the classification algorithm to sub-problems.

After studying the typical applications of the different kernels and doing statistical analysis of the gathered test data, it was quite obvious that the linear kernel function cannot optimise the separation plane. Thus, the Radial Basis Function (RBF) type non-linear kernel [17], which can be applied also when the boundary is complex, was chosen, although the drawback was that the RBF kernel is sensitive to incorrect training samples and may consequently over-fit easily.

5. PROTOTYPE IMPLEMENTATION

The actual CAA module implementation uses the FaceLab stereovision system of SeeingMachines for capturing the driver-related variables (see. Figure 2). The system tracks head and gaze positions and orientations, eye blinking frequency, saccades, thus providing a basic data acquisition platform for the high-level software development. The system uses two CCD-type cameras, which have sufficient sensitivity also in low ambient illumination. The images of the stereo pair are pre-processed in real time (60 Hz) in a separate Intel Pentium-based PC. The cognitive distraction detection algorithm combines the camera vision system's output with the vehicle's internal data, which is read from the CAN bus, as well as lane tracking data.

Figure 3 shows the internal software architecture of the CAA module, in which the feature extraction is divided into two parts [18]. The left side is for detecting the visual distraction and the right side is the routine for estimating the cognitively distracted driver, which is the subject of this paper. The real-time algorithms have been written in the C programming language, and they run in a Matlab/Simulink environment. In Matlab/Simulink applies a development platform in which pre-recorded data files can be replayed. Using this platform the performance of the algorithms can be evaluated by simulating real driving (i.e. kind of hardware-in-the-loop simulation) before they are compiled to real-time PC compatible binaries.

Tuning the parameters is the most important part of optimising the SVM performance. For improving the SVM adaptation capability, a special Microsoft Windows tool was programmed (see Figure 4). The application makes the iterative parameter optimisation process easier and faster by printing the correct and incorrect hit rates. The tool is written with exactly the same source code as the core of the previously mentioned Simulink blocks. The application also assists in adaptation of the complex multi-dimensional classification routine to different types of cockpits (i.e. truck, passenger car, new vehicle models, etc.).

6. RESULTS OF SVM PARAMETER TUNING

The cognitive distraction indicative parameters (gaze, head, lane and face model quality) and the quality factors (gaze and head uniformities in a 15 s time window) were added to investigate their influence on the classification results. The following reported tests were executed in office premises, running the gathered data of real driving.

The results in Table 3 suggest that all the features improve the classification performance, despite that the test is not comprehensive. Actually, a closer analysis pointed out that the additional features compensate for the errors caused by ambiguities of the training data items, as can be seen from the graphs in Figure 5. The quality of the performance depends of course on the nature of the applied training data. Our aim was to retrieve a feasible generalisation, which works also in varying conditions, i.e. the goal was a good robustness of the algorithm. All the tests were done by dividing the data to separate training and validation samples. There were approximately 220 feature vectors in both cases. Most of the tests were verified twice by changing the training and validation samples mutually.

The two crucial parameters when fine-tuning a Support Vector Machine engine with a Radial Basis Function kernel are gamma and criterion (C). Gamma determines the spreading of a single node in the network, i.e. bigger gamma provides a better coverage, but the consequence might be an over-fit. C is the parameter for controlling the error margin for the positive detection (i.e. cognitive distraction). That is, bigger C produces a greater number of false detected non-cognitive results as Table 4 indicates.

Since there was more negative (non-cognitive) examples than positive (cognitive) in the training data, an equal balance between the criterion parameters was not desirable. Rather, different criteria for positive and negative examples (more formally C_{-}= C and C_{+} = c x C, where c = cost-factor) are preferred. Cost-factor emphasizes the total error caused by positive training examples compared to the negative ones. A large cost-factor allows bigger aggregated errors for positive training examples than for the negative ones. There was approximately 1.5 times more baseline data compared to cognitive examples in the training samples, which was taken into account when the error balance was estimated. The effects of the cost-factor test are listed in Table 5.

A threshold adjusts the location of the classification boundary. Typically, the result of the SVM routine is in the range of -1 to + 1, which, however, does not determine the absolute limits. When the threshold is positive, the size of the positive class decreases, resulting in fewer cognitive outputs (see Table 6). Figures 6 and 7 show an alternative way to retrieve the optimal value for the threshold. These graphs show percentages of correct hits for baseline and cognitive tasks, for different values of the threshold parameter, in motorway and intermediate complexity environments. It should be noted that the data used to generate these figures is different from the samples used in the prior tests (Tables 3-7), since the main objective was to investigate whether the intermediate and motorway driving complexities provide a significantly different solution. The data contains also samples from the training set and give therefore better results than the rates in the prior tests. By studying the graphs of Figures 6 and 7, a threshold value can be chosen that gives a trade-off between false positives and false negatives that is suitable for the application at hand. If e.g. designing a cognitive distraction warning system false positives can be accepted to a very small extent, so the threshold may need to be set high. In applications envisioned for the CAA module output, false positives are not as critical and we can thus set the threshold lower than for a warning application, to allow a better distraction hit rate (i.e. a lower false negative rate).

When comparing Figures 6 and 7, an assumption can be made that intermediate and motorway driving complexities provide similar results, which is not a surprise. However, this is still under consideration since it seems that the average gap between baseline and cognitive distraction levels is bigger in the motorway than in the intermediate driving complexity.

The fine-tuning of parameters has improved the classification performance compared to the prior work done during the initial module development phase. Due to over-fitting risk, the classification borders are visualised by showing two-dimensional cognitive and non-cognitive feature spaces, whereas the other features were approximated using their averages in the training sample. The idea worked well in giving a significantly better understanding of how to adjust the gamma, criterion and cost-factor parameters for achieving the desired performance level. The better visualisation properties of the laboratory tool have helped to understand the meaning and behavioural effects of the SVM kernel and to avoid over-fitting, which caused a big problem during the first test period. In Table 7 the initial before the visualisation facilities results are compared to the actual performance rates.

Since similar prior implementations are not easily available it is hard to make comparative evaluations with other works. If the results are compared with the SAVE-IT project [5], it may hardly be said that the proposed method works better, but at least the performance is equal. Additionally, the results support the HASTE achievements [3] very well, i.e. in some cases the signals induced by a cognitive workload are arbitrarily more obvious than in others (e.g. the amplitude may depend on the number of exits in the motorway, the driver's driving experience, etc.), which of course cause many false detections. Fortunately, in most of the cases the indicative signals remain within the expected limits, which indicates that the SVM implementation covers well the expectations of the research hypothesis, which specified that the cognitive workload can be detected using the vehicle's lateral position and the driver's attention variations measures.

7. DISCUSSION

The data used in the training phase were gathered during real driving, but the cognitive tasks were artificially induced. We assume the data to be quite realistic, but we have not evaluated how closely the tasks correspond to real cognitive workload. Presumably, the drivers were not highly cognitively loaded because the tasks were not very difficult in comparison to the previous work reported in [6, 7, 10], or because the driver prioritised the driving situation in the city, paying less attention on the cognitive task, thereby lowering the induced cognitive load. However, the aim of this study was to develop a practically usable module, and therefore it was of major interest to detect cognitive workload also in extreme cases (i.e. to process also very low distraction levels). Thus the module would presumably provide more accurate results when the driver is under a higher cognitive load. It should be pointed out that in the dissertation work [7], the cognitive workload was induced by asking a driver to do backward counting by subtracting 7, but in this case subtractions were in the range 3-7, which is on an average easier and will not cause an equally consistent high workload.

The doctoral thesis [7] proposed that standard deviation of the gaze signals can be utilised for detecting the cognitive distraction. Moreover, the thesis also suggested adopting a PRC (Percent Road Centre) measurement, which was investigated to reduce by cognitive workload. The idea of using SVM type classifier is actually not very far from the proposition of the thesis but the major difference is capability to increase dimensions of the feature space (i.e. adopting the vehicles lateral position among to indicate parameters) by tuning a new model and additionally "draw" non-linear classification borders. The current implementation does not use PRC measurement but as the thesis [7] shows that could be worth of a future comparison.

Another open issue is an opportunity to use the module in city driving. However, the current scenario is that detecting the cognitive distraction in those circumstances is probably not even necessary, since attention demand presumably does not allow severe cognitive impairment. Rather, visual distraction has a greater unwanted influence in heavy traffic when the secondary task types are considered. Fletcher et al. [19] have implemented an interesting scene monitoring application, which is intended for increasing robustness of the fatigue detection with gaze analysis. The method recognises the monotony driving context from a video. The same technique would potentially be beneficial for increasing performance of the cognitive distraction detection since the workloads like daydreaming are more relevant in a monotony environment (i.e. motor and intermediate ways).

An SVM type classifier was selected for this application, which turned out to be a good choice. Perhaps the neural networks would allow a more advanced processing methodology, like better robustness to "abnormal behaviour" during a cognitive workload. On the other hand, the drawback of SVM is its sensitivity to outliers in data, especially in the training or validation stages. It is also possible that the use of a simple syntactic classifier (i.e. using rules) would have provided a feasible first step approach. However, the earlier good experiences and the results of the SAVE-IT project [5] encourage adopting the SVM classifier when dealing with multi-dimensional non-linear feature space.

8. CONCLUSIONS

Literature review, prior work and applied analysis verify that gaze direction variation as well as lane deviation (i.e. deviation of the vehicle's lateral lane position) decreases when the driver is cognitively distracted. The reference [4] proposes that increased gaze concentration is actually related to the lane keeping performance. Due to these findings, measurement of the gaze, head and lane related parameters was adopted for recognising the cognitive load, which means that in a sense cognitive impairment is actually discovered as a result of improved driving measures. However, as already noted in section 3, cognitive distraction increases the driver's reaction time and is thus an important factor in decreasing driving performance.

The SVM classifier was envisioned to be an effective method for the CAA solution, since the main intention is only to detect whether the driver's attention is reduced by the cognitive load or not. The experimental results made it clear that the utilised input features created a highly non-linear boundary surface, which implicated the need of a RBF- type kernel. The selected algorithm improves the detection capability when an additional number of driving related parameters is implemented to the classifier. Moreover, the algorithm enables pretty easy adaptation to different sensor arrays, which will be the case in a passenger car demonstration when no lane position sensor exists.

As a result of iterative tests, the optimal SVM variables were retrieved for a truck cockpit. The optimisation was quite time-consuming because of the multiple fine-tuned parameters. The main issue is that there is a dependency between the classifier tuning parameters (e.g. cost-factor, gamma, etc.), i.e. changing one parameter requires subsequently retuning the others, which makes it impossible to adjust only one parameter at a time. However, robustness of the algorithm is held as the key factor and the achieved solution minimises the false detections. The performance tests in Figures 6 and 7 show that even an 80% classification capability was achieved. However, in those cases the test sample included also the training points and thus a reliable estimation would be that the rate of correct hits is more than 65%. One aspect is also the dynamic behaviour of the drivers. Therefore, the detection performance depends also on the driver's reactions to the cognitive tasks. For example, one of the drivers started continuous visual scanning to the right side of the cockpit despite there was nothing to that side of the road but trees. The visualisation facilities help to avoid the over-fitting problem by showing if there are "islands" located in the classification space.

The created software pieces worked well despite the fact that the training data mostly represented low cognitive workloads. Consequently, it be may hypothesise that in further tests the performance will not degrade. The AIDE project was formed around existing driver-monitoring techniques and improving the earlier developed algorithms. The objective was to avoid fundamental basic research work and rather focus on adopting the existing knowledge to a practical prototype implementation, which will be ultimately realised in a Volvo truck and a SEAT passenger car during 2006.

9. FURTHER WORK

The final module is intended to be run in an embedded xPC module. The installation will launch the practical test period, which will show how well the performed work fulfils the expectations of the laboratory tests. The developed software modules have an adequate performance. One important future aspect is minimising the size of the hardware, so that the driver does not see the equipment on the dashboard. At the moment, the camera modules in typical stereovision systems are too large-dimensioned, especially for passenger car applications. Nevertheless, the recently introduced miniature CMOS cameras with thin glass optics are close to the size of a stamp, so they can easily be hidden in the cockpit.

At present the module is mainly designed for truck environment, but the near-future scenario is to adapt the system to a passenger car, which does not include lane position measurement. So, the CAA module will run relying only on the gaze and head based features. The performance quality will be lower than in the case of a truck, but it is of great interest to observe the performance of the system in a totally different type of vehicle environment, and especially to see, how much work for the adaptation is required. Presumably, the SVM engine is rather easy to reconfigure.

10. ACKNOWLEDGEMENTS

We wish to express our gratitude to the European Commission and to the whole AIDE consortium for the supporting activities and fruitful discussions during the CAA module development phase.

REFERENCES

1 **Dingus, T.A., Klauer, S.G., Neale, V.L., Petersen, A., Lee, S.E., Sudweeks, J., Perez, M.A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z.R., and Jermeland, J.** 2005. The 100-car naturalistic driving study. National Highway Traffic Safety Administration. Contract No. DTNH22-00-C-07007. In press, Washington, DC.

2 **Tattegrain Veste, H., Bellet, T., Boverie, S., Kutila, M., Viitanen, J., Bekiaris, E., Panou, M., Engström, J. & Angvall, A.** 2005. Development of A Driver Situation Assessment Module In The Aide Project. *Proceedings of 16th IFAC WORLD CONGRESS. Czech Republic*. Praha. 4-8 Jul 2005.

3 **Östlund J., Nilsson L. et. al.** 2004. HASTE Deliverable 2: HMI and Safety-Related Driver Performance. EU funded project GRD1/2000/25361 S12.319626.

4 **Lee, J., Reyers, M., et al.** 2004. SAfety VEhicles using adaptive Interface Technology (Task Final Report: Phase 1 of the SAVE-IT project). [Referred 2 Feb 2006]. Available at: www.volpe.dot.gov/opsad/saveit.

5 **Engström, J., Johansson, E. & Östlund, J.** 2005. Effects of visual and cognitive load in real and simulated motorway driving. *Journal of Transportation Research Part F 8. ELSEVIER*. pages 97-120. ISSN: 1369-8478.

6 **Victor, T. W., Harbluk, J. L., & Engström, J. A.** 2005. Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Psychology and Behaviour*. 8(2), 167-190. DOI: 10.1016/j.trf.2005.04.014

7 **Victor, T.** 2005. Keeping Eye and Mind on the Road. Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 9, Uppsala: Acta Universitatis Upsaliensis.

8 **Hammel, K. R., Fisher, D. L., & Pradhan, A. K.** 2002. Verbal and spatial loading effects on eye movements in driving simulators: A comparison to real world driving. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*. pp. 2174 - 2178.

9 **Harbluk, J.L., Noy, Y.I., & Eizenman, M.** 2002. The Impact of Cognitive Distraction on Driver Visual Behaviour and Vehicle Control. *Transport Canada Technical Report*. [Referred: 2 Feb 2006]. Available at: http://www.tc.gc.ca/roadsafety/tp/tp13889/en/menu.htm

10 **Recarte, M. A., & Nunes, L. M.** 2003. Mental Workload While Driving: Effects on Visual Search, Discrimination and Decision Making. *Journal of Experimental Psychology: Applied, 9 (2)*. p 119-137.

11 **Horrey, W. J. & Wickens, C. D.** 2004. The impact of cell phone conversations on driving: A meta-analytic approach. (Technical Report AHFD-04-2/GM-04-1). Warren, MI: General Motors Corporation.

12 **Corbetta, M. & Shulman, G. L.** 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*. 3 (3). p. 201-215.

13 **Milner, A. D., & Goodale, M. A.** 1995. *The Visual Brain in Action*. Oxford: Oxford University Press.

14 **Larsson, P. & Victor, T. W.** 2005. Method and Arrangement for Interpreting Subject Head and Eye Activity. Patent Application. United States. Pub. No.: US 2005/0073136 A1

15 **Milton, J. & Arnold, J.** 2002. *Introduction to Probability and Statistics Principles and Applications for Engineering and the Computing Sciences*. 832 p. ISBN: 0071198598.

16 **Joachims, T.** 1999. Making large-Scale SVM Learning Practical. Advances in Kernel Methods. *Published in Support Vector Learning, Schölkopf, B., Burges, C. & Smola, A. (ed.).* MIT-Press. [referred: 2 Feb 2006]. Available at: http://www-ai.cs.uni-dortmund.de/DOKUMENTE/joachims_99a.pdf.
17 **Haykin, J.** 1999. Neural networks: A Comprehensive Foundation 2nd edition. U.S.A. Prentice-Hall, Inc. ISBN 0-13-273350-1.
18 **Markkula, G., Kutila, M., Engström, J., Victor, T. W. & Larsson, P.** 2005. Online Detection of Driver Distraction - Preliminary Results from the AIDE Project. *Published in Proceedings of the 2005 International Truck and Bus Safety and Security Symposium*. 14-16 Nov 2005, U.S.A., Washington, Alexandria, Virginia.
19 **Fletcher, L., Petersson, L. & Zelinsky, A.** 2005. Road scene monotony detection in a fatigue management driver assistance system. Proceedings of IEEE Intelligent Vehicles Symposium. 6-8 June 2005. pp. 484 – 489.

Table 1. Purpose of the modules in the AIDE driver/environment monitoring system

Module	Description
Cockpit Activity Assessment (CAA)	Detection of the visual and cognitive workload, which impairs the driver's attention on receiving/understanding messages via an in-vehicle HMI
Driver State Degradation (DSD)	Recognition of the driver's reduced spryness due to drowsiness (i.e. driver's physiological state degradation) using eyelid, lane tracking and steering position sensors.
Driver Availability Estimator (DAE)	Assessment of the driver's availability for primary task (=driving). The analysis fuses the cartographic context and the basic sensor measurements like steering wheel, pedals, etc.
Traffic and Environment Risk Assessment (TERA)	Measurement of the environmental risk factors i.e. risks caused by surrounding vehicles, speed in curves, road profile, etc. and additionally the driver's manoeuvre intention
Driver Characteristics (DC)	Estimation of the driver's performance ability via the driver's profile including age, driving experience, gender, type of trips, average headway , etc.

Table 2. The CAA module related adaptation examples

Conflict scenario example	CAA output	HMI adaptation example
The driver is picking up a dropped wallet from the vehicle's floor. While the driver is looking for his wallet the vehicle ahead brakes and forward collision warning initiates.	The driver is not looking at the road ahead	The forward collision warning is given earlier and possible with a greater intensity and/or with additional output modalities.
The driver is talking to a passenger when a brake failure is detected	Cognitive distraction is detected, which may reduce the driver's concentration on the vehicle status displays.	The brake failure message is enhanced with a warning sound or a voice output in addition to the standard instrument panel telltale

Table 3. Comparison of the different feature selections for classification performance. The best performance is highlighted in green colour.

Gaze rotation	Gaze rot quality	Head rotation	Head rot quality	Lane position	Face model quality	Non-cogn	Cogn
x				x		54,9	54,8
		x		x		57,7	54,8
x		x		x		64,8	53,4
x	x			x	x	76,1	54,8
		x	X	x	x	65,5	60,3
x	x	x	X	x		78,2	50,7
x	x	x		x	x	81,0	41,1
x	x	x	X	x	x	80,3	54,8

Table 4. Results of the tests for retrieving the optimal gamma - criterion adjustments

		TEST SET 1		TEST SET 2	
gamma	C	Non-cogn	Cogn	Non-cogn	Cogn
1,7	5,0	71,8	56,2	78,2	35,4
2,0	5,0	73,9	54,8	75,0	38,1
2,2	3,0	72,5	54,8	79,8	33,6
2,2	8,0	76,1	54,8	67,7	46,0
3,0	4,0	76,8	52,1	71,0	45,1
3,0	6,5	79,6	53,4	68,5	45,1
3,0	9,0	80,3	52,1	68,5	43,4
5,0	3,0	80,3	54,8	71,0	45,1
5,3	4,0	81,7	52,1	68,5	44,2
5,7	4,0	81,7	47,9	68,5	45,1
6,0	10,0	83,8	45,2	70,2	47,8
6,3	6,0	83,8	46,6	70,2	48,7
6,5	5	83,1	46,6	70,2	48,7
6,5	6	84,5	46,6	70,2	48,7
6,5	6,2	84,5	46,6	70,2	49,6
6,5	6,5	84,5	46,6	70,2	47,8
6,6	6,3	84,5	46,6	70,2	47,8
6,7	6	84,5	46,6	70,2	48,7
6,8	5	83,8	46,6	70,2	48,7
8,0	15,0	83,8	49,3	69,4	46,9
10,0	10,0	85,9	47,9	72,6	46,9
15,0	10,0	89,4	42,5	77,4	42,5
100,0	1,0	97,9	8,2	98,4	0,9

Table 5. Tests with different cost-factors

c	Gamma	C	Non-cogn	Cogn	Non-cogn	Cogn
0,5	6,5	6,2	86,6	43,8	72,6	36,3
1,0	6,5	6,2	84,5	46,6	70,2	49,6
1,5	6,5	6,2	82,4	49,3	70,2	47,8
2,0	5,0	8,0	78,2	52,1	69,4	47,8
2,0	6,5	6,2	80,3	50,7	70,2	48,7
3,0	6,5	6,2	80,3	53,4	69,4	49,6
5,0	6,5	6,2	80,3	53,4	69,4	49,6
10,0	6,5	6,2	80,3	53,4	69,4	49,6

Table 6. The following threshold adaptation tests were executed with gamma: 6.5, C: 6.2 and cost-factor: 3.0

Threshold	Non-cogn	Cogn	Non-cogn	Cogn
-0,30	37,3	71,2	62,1	54,9
-0,25	71,8	57,5	65,3	54,9
-0,23	73,2	56,2	65,3	54,0
-0,20	73,2	54,8	66,1	53,1
0,00	80,3	53,4	69,4	49,6
0,20	84,4	46,6	75,8	39,8

Table 7. Comparison to the earlier work in the CAA module development steps. The best trials are the results of the over-fitted SVM model. The non-over fitted rates are given using the SVM parameters, which are close to the currently used.

	Earlier results		**Current**
test data	**the best trials**	**non-over fitted**	
1	72,0	45,3	68,4
2	55,5	43,8	58,6
3	59,5	61,9	65,0

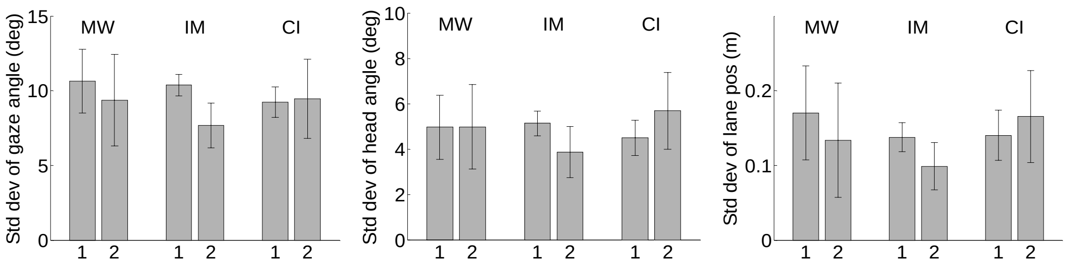


Figure 1. Mean values and 95% confidence intervals for the three main cognitive distraction indicators (15 second time window) for different tasks in different traffic environments. The tasks are baseline (1) and cognitive tasks (2). The environments are motorway (MW), intermediate complexity environment (IM), and city (CI).

Figure 2. The FaceLab system setup in the test truck

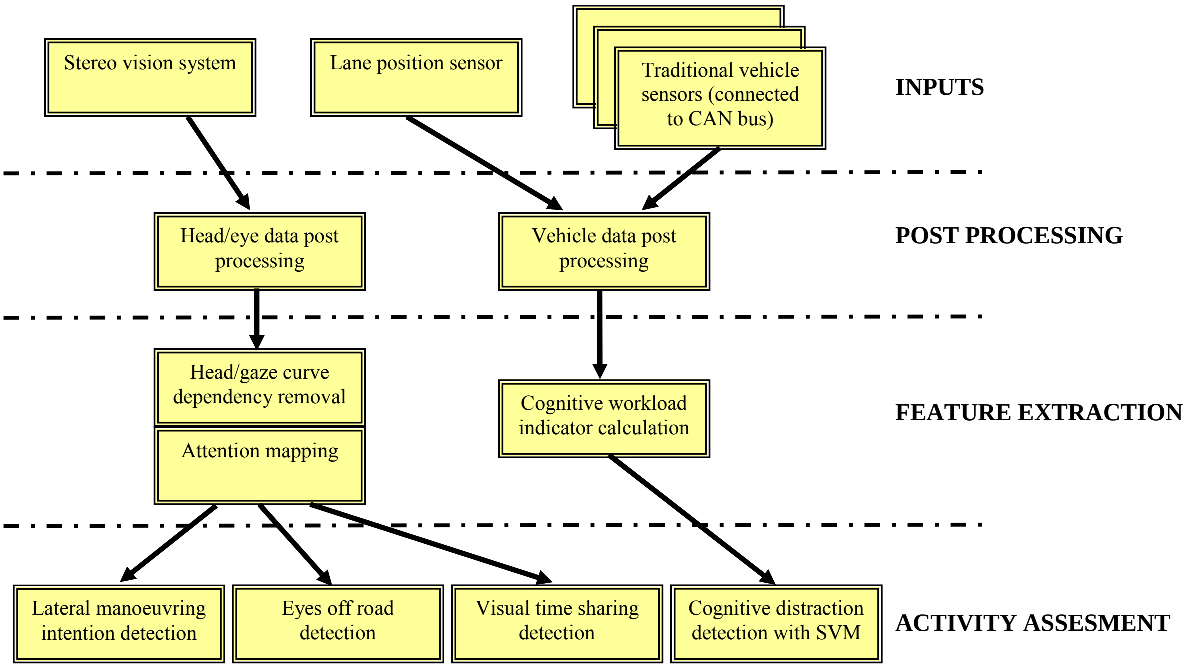
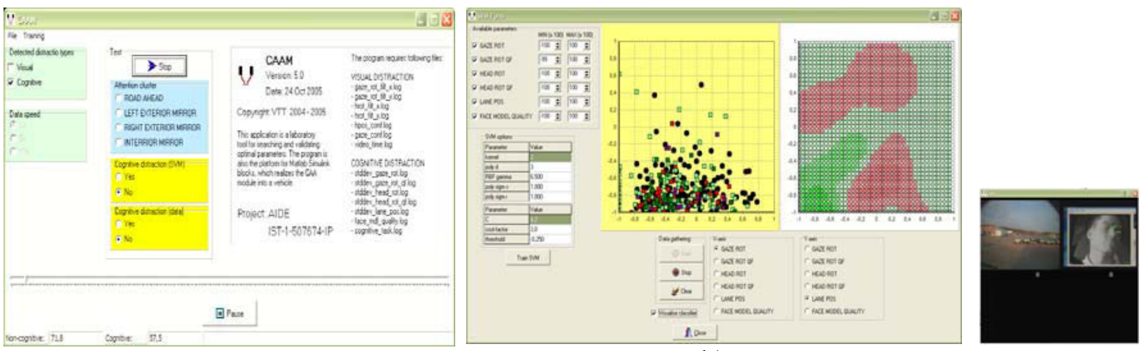
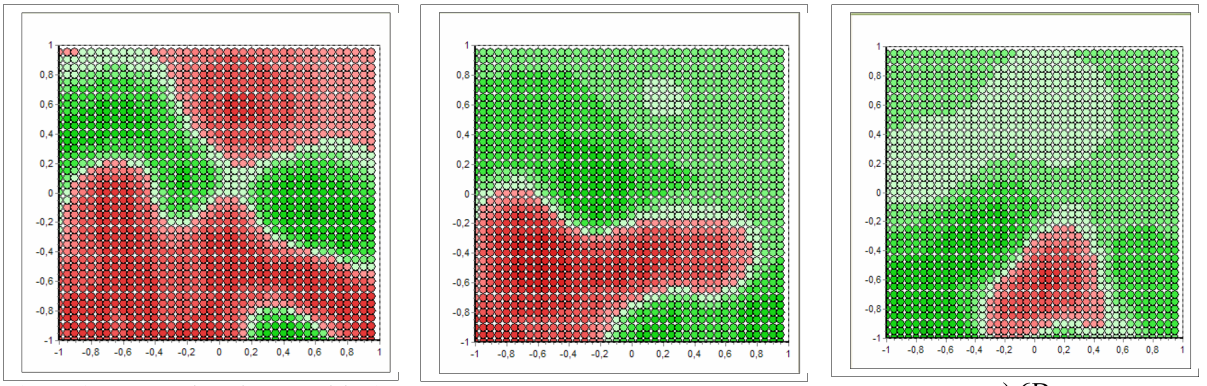


Figure 3. The internal software architecture of the CAA module



a) b) c)

Figure 4. The user interfaces of the developed laboratory tool, which are used for tuning the SVM parameters as well as estimating the performance rates. The window a) is the main form, which replays the captured data, b) is for training the SVM and visualising the reason of false detections and c) shows the related video on what is happening in the surrounding traffic and the driver's behaviour.



a) 2D (gaze rotation, lane position)

b) 3D (gaze rotation, head rotation, lane position)

c) 6D

Figure 5. Difference in gaze rotation - lane position spaces (on average level of the other features) when the number of the inputs is varied. Red colour relates to a cognitive and green colour to a non-cognitive region. When the result is close to the threshold between cognitive and non-cognitive limit the colour becomes lighter. The SVM parameters in the graphs are gamma: 6.5, criterion: 6.2, cost-factor: 3.0 and threshold: 0.

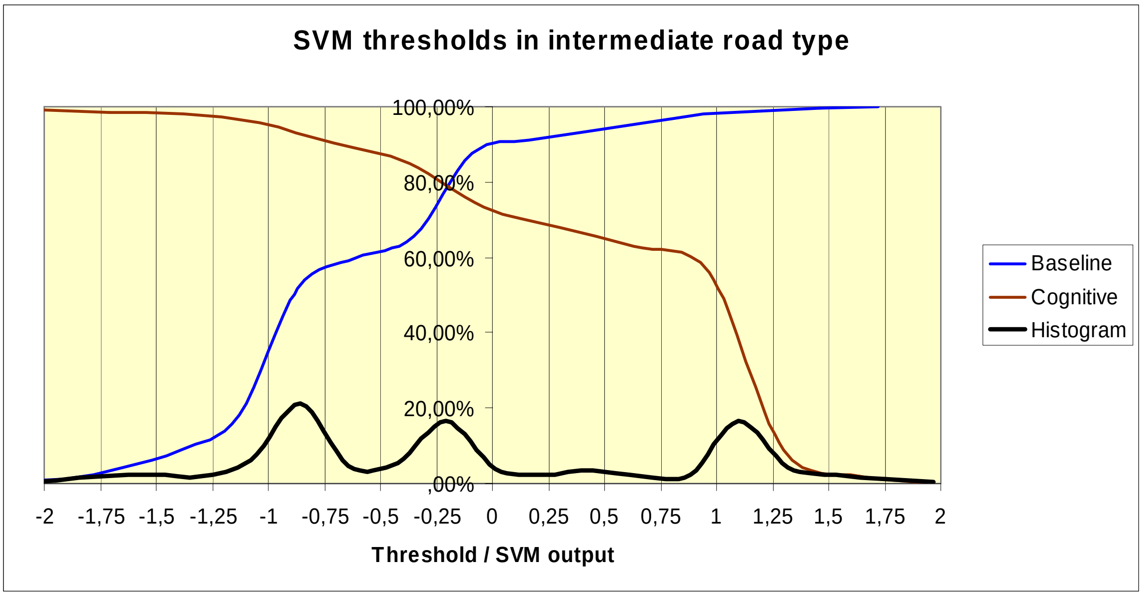


Figure 6. The hits rate aggregation as a function of the SVM threshold in the intermediate driving complexity. The horizontal axis is a SVM threshold or output and the vertical axis represents hit rates. The histogram line describes amount of data resulting the SVM output regarded to the horizontal axis. Note that the data used here include not only data points from the validation set, but also training data, which is why the hit rates here are generally higher than in Tables 3-7.

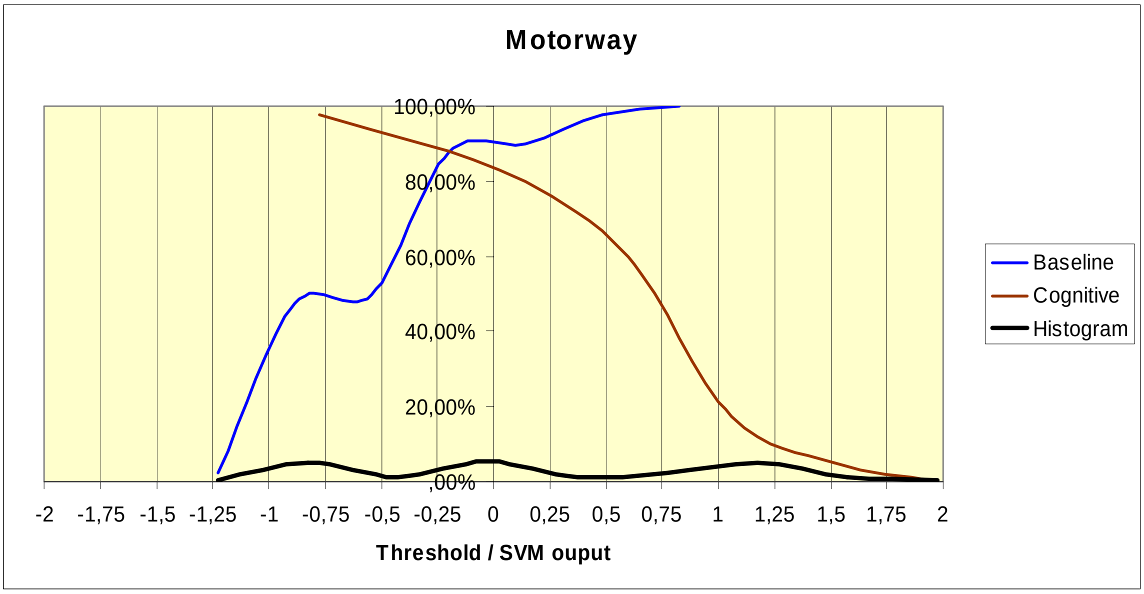


Figure 7. The hits rate aggregation as a function of the SVM threshold in the motorway. As in Figure 6 the used data set also includes training data points.

Published by

Series title, number and report code of publication

VTT Publications 621
VTT–PUBS–621

Author(s)
Kutila, Matti

Title

Methods for Machine Vision Based Driver Monitoring Applications

Abstract

An increasing number of information and driver-assistive facilities—such as PDAs, mobile phones, and navigators—are a feature of today's road vehicles. Unfortunately, they occupy a vital part of the driver's attention and may overload him or her in critical moments when the driving situation requires full concentration. The automotive industry has shown a growing interest in capturing the driver's behaviour due to the necessity of adapting the vehicle's Human–Machine Interface (HMI), for example, by scheduling the information flow or providing warning messages when the driver's level of alertness degrades. The ultimate aim is to improve traffic safety and the comfort of the driving experience.

The scope of this thesis is to investigate the feasibility of techniques and methods, previously examined within the industry, for monitoring the driver's momentary distraction state and level of vigilance during a driving task. The study does not penetrate deeply into the fundamentals of the proposed methods but rather provides a multidisciplinary review by adopting new aspects and innovative approaches to state-of-art monitoring applications for adapting them to an in-vehicle environment. The hypotheses of this thesis states that detecting the level of distraction and/or fatigue of a driver can be performed by means of a set of image processing methods, enabling eye-based measurements to be fused with other safety-monitoring indicators such as lane-keeping performance or steering activity. The thesis includes five original publications that have proposed or examined image processing methods in industrial applications, as well as two experiment-based studies related to distraction detection in a heavy goods vehicle (HGV), complemented with some initial results from implementation in a passenger car.

Keywords
driver monitoring, machine vision, distraction, fatigue, wavelets, SVM, neural networks, classification, cameras, traffic safety, vehicles, sensors, colour vision, alertness, gaze, eyes, head, workload, traffic safety and vigilance

ISBN
951–38–6875–3 (soft back ed.)
951–38–6876–1 (URL: http://www.vtt.fi/publications/index.jsp)

Series title and ISSN			Project number
VTT Publications 1235–0621 (soft back ed.) 1455–0849 (URL: http://www.vtt.fi/publications/index.jsp)			612
Date	Language	Pages	Price
November 2006	English, Finnish abstr.	82 p. + app. 79 p.	D
Name of project		Commissioned by	
AIDE (IST-1-507674-IP, FP6)		The European Commission	
Contact		Sold by	
VTT Technical Research Centre of Finland Tekniikankatu 1, P.O. Box 1300 FI-33101 TAMPERE, Finland Phone internat. +358 20 722 111 Fax +358 20 722 8334		VTT Technical Research Centre of Finland P.O.Box 1000 FI-02044 VTT, Finland Phone internat. +358 20 722 4404 Fax +358 20 722 4374	

Julkaisija	Julkaisun sarja, numero ja raporttikoodi
VTT	VTT Publications 621 VTT–PUBS–621

Tekijä(t)

Kutila, Matti

Nimeke

Menetelmiä konenäköpohjaisiin kuljettajan monitorointisovelluksiin

Tiivistelmä

Kuljettajan tukijärjestelmien määrä kasvaa tulevaisuudessa. Tämä helpottaa ajamista ja lisää ajomukavuutta, mutta tuo toisaalta mukanaan lieveilmiöitä. Muiden muassa matkapuhelimet, navigaattorit ja musiikkisoittimet kilpailevat yhä enenevässä määrin kuljettajan huomiokyvystä. Mikä pahinta, nämä laitteet saattavat haitata kuljettajan keskittymistä ja aiheuttaa onnettomuuden vaaran. Ajoneuvoteollisuus on tästä syystä osoittanut kasvavaa kiinnostusta kuljettajan tilaa monitoroivia järjestelmiä kohtaan. Nämä järjestelmät mahdollistaisivat kuljettajan ja ajotilanteen mukaan säätyvän älykkään käyttöliittymän kehittämisen. Tällainen käyttöympäristö voisi esimerkiksi viivyttää ei-kiireellisten ajoneuvon tilatietojen välittämistä, kuten tuulilasin pesunesteen loppumisesta varoittavaa viestiä, kunnes kuljettaja todetaan "valmiiksi" vastaanottamaan informaatio. Tavoitteena on siis tehdä ajamisesta entistä mukavampaa ja mikä tärkeintä myös turvallisempaa, jottei kuljettajaa häirittäisi kriittisillä hetkillä.

Tämän väitöstyön tarkoituksena on tutkia teollisessa ympäristössä kokeellisesti hyväksi havaittujen menetelmien soveltuvuutta kuljettajan havaintokyvyn ja väsymystilan arviointiin. Työn tarkoitus ei ole tuottaa syvällistä analyysia ehdotetuista menetelmistä, vaan tarkastella asiaa poikkitieteellisesti. Tämä avartaa uusia näkökulmia ja innovatiivisia lähestymistapoja olemassa oleviin monitorointijärjestelmiin ja auttaa niiden sovittamisessa ajoneuvoympäristöön. Työssä testattava hypoteesi esittää, että kuljettajan häiriytyminen ja/tai väsymystila voidaan havaita kuvankäsittelymenetelmillä. Niiden avulla on mahdollista mitata häiriytymisaste kuljettajan silmistä ja yhdistää tätä tietoa muihin indikaattoreihin kuten kaistalla vaelteluun tai epätasaisiin ohjausliikkeisiin. Tämä työ koostuu viidestä alkuperäisjulkaisusta, joissa käsitellään ja testataan kuvankäsittelymenetelmiä teollisissa sovelluksissa, sekä kahdesta julkaisusta, joissa tutkitaan kuljettajan häiriytymisen mittaamista kuorma-autossa. Näitä tuloksia on täydennetty alustavilla mittauksilla henkilöautoissa.

Avainsanat

driver monitoring, machine vision, distraction, fatigue, wavelets, SVM, neural networks, classification, cameras, traffic safety, vehicles, sensors, colour vision, alertness, gaze, eyes, head, workload, traffic safety and vigilance

ISBN

951–38–6875–3 (nid.)
951–38–6876–1 (URL: http://www.vtt.fi/publications/index.jsp)

Avainnimeke ja ISSN			Projektinumero
VTT Publications 1235–0621 (nid.) 1455–0849 (URL: http://www.vtt.fi/publications/index.jsp)			612
Julkaisuaika	**Kieli**	**Sivuja**	**Hinta**
Marraskuu 2006	Englanti, suom. tiiv.	82 s. + liitt. 79 s.	D

Projektin nimi	Toimeksiantaja(t)
AIDE (IST-1-507674-IP, FP6)	Euroopan komissio

Yhteystiedot	Myynti
VTT Tekniikankatu 1, PL 1300, 33101 TAMPERE Puh. vaihde 020 722 111 Faksi 020 722 8334	VTT PL 1000, 02044 VTT Puh. 020 722 4404 Faksi 020 722 4374

An increasing number of information and driver-assistive facilities – such as PDAs, mobile phones, and navigators – are a feature of today's road vehicles. Unfortunately, they occupy a vital part of the driver's attention and may overload him or her in critical moments when the driving situation requires full concentration. The scope of this thesis is to investigate the feasibility of techniques and methods, previously examined within the industry, for monitoring the driver's momentary distraction state and level of vigilance during a driving task. The study does not penetrate deeply into the fundamentals of the proposed methods but rather provides a multidisciplinary review by adopting new aspects and innovative approaches to state-of-art monitoring applications for adapting them to an in-vehicle environment. The thesis includes five original publications that have proposed or examined image processing methods in industrial applications, as well as two experiment-based studies related to distraction detection in a heavy goods vehicle (HGV), complemented with some initial results from implementation in a passenger car.

Tätä julkaisua myy	Denna publikation säljs av	This publication is available from
VTT PL 1000 02044 VTT Puh. 020 722 4404 Faksi 020 722 4374	VTT PB 1000 02044 VTT Tel. 020 722 4404 Fax 020 722 4374	VTT P.O. Box 1000 FI-02044 VTT, Finland Phone internat. +358 20 722 4404 Fax +358 20 722 4374

ISBN 951–38–6875–3 (soft back ed.)
ISSN 1235–0621 (soft back ed.)

ISBN 951–38–6876–1 (URL: http://www.vtt.fi/inf/pdf/)
ISSN 1455–0849 (URL: http://www.vtt.fi/inf/pdf/)

Matti Kutila