Real-time Occupant Detection in High Dynamic Range Environments

Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
at
City University
London

By
Carsten Koch

October 20, 2003

Contents

List of Figures

List of Tables

Glossary of Acronyms and Abbreviations

2-D	Two dimensions/two-dimensional
3-D	Three dimensions/three-dimensional
ACC	Adaptive Cruise Control
ADC	Analog-to-Digital Converter
AGC	Auto Gain Control
APS	Active Pixel Sensor
ASIC	Application Specific Integrated Circuit
BMW	Bayerische Motorenwerke
CAN	Controller Area Network
CCD	Charged Coupled Device
CCTV	Closed Circuit Television
CDS	Correlated Double Sampling
CMOS	Complementary Metal-Oxide Semiconductor
CWL	Center Wave Length
DSP	Digital Signal Processor
EMS	Emergency Medical Services
FFCS	Forward facing child seats
FIR	Far Infrared
FPGA	Field Programmable Gate Array
FPN	Fixed Pattern Noise
fps	Frames per Second
GMM	Gaussian Mixture Model
HBW	Half Power Bandwidth

HDR	High Dynamic Range
HMI	Halogen Metal vapor lamps
I^2C	Inter-IC Bus
JTAG	Joint Test Action Group
LED	Light Emitting Diode
LUT	Look up Table
LVDS	Low Voltage Differential Signaling
MMS	Multi Media Message
MOS	Metal-Oxide Semiconductor
MPE	Maximum Permissible Exposure limits
NHTSA	National Highway Traffic Safety Administration (USA)
NIR	Near Infrared
NOPS	Nothing present on this seat
ODFC	An object which does not fit to the other classes
OOP	out-of-position
PCA	Principal Component Analysis
PCSP	Person in correct seating position
POOP	Person out-of-position
PPS	Passive Pixel Sensor
RAM	Random Access Memory
RFCS	Rear facing child seats
ROI	Region of Interest
SNR	Signal-to-Noise Ratio
TFA	Thin Film on ASIC
TOF	Time of Flight
VIB	VisionBox, Strampe Systemelectronic GmbH
VLIW	Very Large Instruction Word

Acknowledgment

The author would like to thank Dr. T.J. Ellis and Prof. A. Georgiadis for their supervision, encouragement and valuable discussions. The author is also very grateful to Dipl.-Ing. J. Mahalek who triggered this very interesting project and to Dipl.-Ing. L. Eisenmann for fruitful discussions and contributions in the development of systems for detecting occupants and for his support at BMW.

I wish to thank Dr. S-B. Park for his introduction to CMOS cameras and smart illumination and his tireless explanations. I will never forget our in-circuit software development during high-speed test drives. I also have to thank J.J. Yoon for his support in multiple mathematical problems and value discussions about illumination strategies.

Special thanks to my fellows at BMW who helped me bring our HDR-camera from illusions and raw ideas to reality: S. Akisogulu, W. Solka, S. Weidhaas and A. Augst. I also express my gratitude to the fellows who looked carefully through this manuscript and helped to improve it with a number of corrections, suggestions and good questions.

Finally I want to thank my wonderful wife Melanie for her constant faith, love and encouragement.

The majority of this research was sponsored by BMW and performed at the BMW AG Research and Innovation Center in Munich, Germany. The opinions expressed herein are those of the author and do not necessarily represent those of BMW.

Declaration

Parts of this work have been already published in collaboration with research fellows and colleagues in the following papers and patents. I wish to express my sincere thanks to Dr. S-B. Park for his support in analyzing our illumination measurements in Section 2.4.1 and to J.J. Yoon for his contribution to the active illumination strategies in Section 4.3.

C. Koch, *'Sitzbelegungserkennung im Kfz durch digitale Bildverarbeitung*[1] *'*, Diploma thesis, FH Nordostniedersachsen FB Automatisierungstechnik, February 1999

C. Koch, S-B. Park, T.J. Ellis and A. Georgiadis, *'Illumination technique for optical dynamic range compression and offset reduction'*, Proc. of British Machine Vision Conference 2001 (BMVC2001), Manchester, UK, September 2001, pp.293-302

C.Koch, T.J. Ellis and A. Georgiadis, *'Real-time occupant classification in high dynamic range environments'*, Proc. of IEEE Intelligent Vehicle Symposium, Versailles, France, June 2002

J.J. Yoon, C.Koch and T.J. Ellis, *'ShadowFlash: an approach for shadow removal in an active illumination environment'*, Proc. of British Machine Vision Conference 2002 (BMVC2002), Cardiff, UK, September 2002, pp.636-645

C. Koch, S-B. Park and S. Sauer, *'Method and apparatus for monitoring the interior space of a motor vehicle'*, International Patent EP 1.215.619, December 2000

C. Koch and S. Akisoglu, *'Line flash'*, National Patent DE 102.45.912, patent pending, October 2002

C. Koch, J.J. Yoon and L. Eisenmann, *'ShadowFlash'*, National Patent DE 102.50.705, patent pending, October 2002

C. Koch, A. Augst and M. Fuchs *'Stereo-vision with mono chip'*, National Patent Application, patent pending, January 2003

[1] engl. transl.: *Occupant detection via digital image processing*

Abstract

The aim of this thesis is to explore strategies for real-time image segmentation of non-rigid objects in a spatio-temporal domain with a stationary camera within an optical high dynamic range environment. Camera, illumination and segmentation techniques are discussed for image processing in environments which are characterized by large intensity fluctuations and hence a high optical dynamic range (HDR), in particular for vehicle interior surveillance.

Since the introduction of the airbag in 1981 numberless lives were saved and bad injuries were avoided. But in recent years the airbag has frequently been in the headlines due to the increasing number of injuries caused by it. To avoid these injuries a new generation of 'smart airbags' has been designed which shows the ability to inflate in multiple steps and with different volumes. In order to determine the optimal inflation mode for a crash it is necessary to consider information about the interior situation and the occupants of the vehicle. This thesis presents a real-time visual occupant detection and classification system for advanced airbag deployment, utilizing a custom CMOS camera and motion based image segmentation algorithms for embedded systems under adverse illumination conditions.

A novel illumination method is presented which combines a set of images flashed with different radiant intensities, which significantly simplifies image segmentation in HDR environments. With a constant exposure time for the imager a single image can be produced with a compressed dynamic range and a simultaneously reduced offset. This makes it possible to capture a vehicle interior under adverse light conditions without using high dynamic range cameras and without losing image detail. The expansion of this active illumination experiment leads to a novel shadow detection and removal technique that produces a shadow-free scene by simulating an artificial infinite illuminant plane over the field of view. Finally a shadowless image without loss of texture details is obtained without any region extraction phase.

Furthermore, a texture based segmentation approach for stationary cameras is presented which is neither effected by sudden illumination changes nor by shadow effects.

Chapter 1

Introduction

1.1 Motivation

An increasing number of machine vision systems started in recent years to leave their smooth and gentle laboratories with restrained lighting in order to enter the rough outside world with widely varying illumination conditions. This opens new fields of applications, such as outdoor surveillance or automatic vehicle guidance. However, a couple of new problems have arisen, such as the fact that most commercial camera systems are not able to cover the full range of brightness differences which may occur in an outdoor scene, *i.e.* the optical dynamic range of the scene exceeds the dynamic range of the image sensor.

Mainstream CCD based and most of the emerging CMOS based image sensors provide an optical dynamic range of $48 \ldots 60dB$. This dynamic range is sufficient for scenes with homogeneous illumination and without extreme contrast. Extreme contrast may occur when operating an imager in direct sunlight or in scenes with areas of high brightness and deep shade, *e.g.* a building entrance. Other classic examples are looking from a dimly lit room through a window towards a bright outdoor scene, the direct view of a burning light bulb or a vehicle interior. Examples of scenes with high contrast are shown in Fig. 2.10 on Page 28, Fig. 4.6 on Page 78, Fig. 3.11 on Page 59 and Fig. 3.12 on Page 61.

1.1.1 Aims of this thesis

Usually it is possible to adjust common imagers either to the very bright areas or to the dark areas in the scene by adapting several of the imager's parameters, *e.g.* exposure time, lens aperture or by the use of optical filters. Nevertheless in extreme dynamic environments there will be areas in the scene which are over- or under-exposed, resulting in lost image detail. In particular CCD based imagers tend to suffer from this problem because only a small number of overexposed pixels can yield large saturated areas (due to blooming and smearing). This well-known limitation is shown in Fig. 4.5 on page 78. This work explores the problem of image processing in optical high dynamic range environments (HDR) by the example of real-time image processing within motor vehicles.

The term 'real-time' is a question of the number of operations and decisions to be made within a determined period of time and the necessary processing power to perform all calculations within this time. Hence, 'real-time' depends on the application. In case of the presented visual detection, classification and tracking of occupants within a motor vehicle (see Chapter 6), 'real-time' means a response within a couple of milliseconds.

Especially applications for embedded systems[1] based on image process-

[1] Processing host with μC, ASIC, FPGA or DSP core, see Section 6.5.

ing suffer from too many input data and/or too little available processing power. The benefits of the systematic and optimized use of hardware devices such as an optimized illumination, additional optical filters or a special kind of camera are often considerably underrated. The aim of this thesis is to show how to reduce processing costs by employing an optimized cooperation of hard- and software (hard- and software fusion) for real-time image segmentation on embedded systems.

Recent camera developments, in particular the revival of CMOS imagers, provide the possibility of optimizing the image acquisition and of implementing 'intelligence' into the imager (see Chapter 3). But what can be done inside the imager to support the subsequent image processing? Where are the benefits and what are the limitations? These are analyzed in this thesis using the example of a video-based car occupant detection system.

Optical sensors offer a wide range of information which cannot be provided by other sensor systems. By analyzing picture sequences from automotive adapted camera systems (imagers) automotive manufacturers and suppliers try to integrate new features into vehicles which have so far been either impossible or possible with major limitations. There are multiple imaginable applications for such camera systems: Lane tracking, traffic sign recognition, automatic vehicle guidance, electrical rearview mirrors, Adaptive Cruise Control (ACC), etc. Most of the current research and development work focuses on analyzing the outside of the car.

However, there are interesting applications where the vehicle interior will be analyzed by image data (see Section 6.1), for example an anti-theft system, air conditioning control, driver identification, drowsiness sensor and support of the airbag control box. Where are passengers sitting? Is an infant seat mounted on the passenger seat? Where are the heads of the passengers? How about the head rest position? Such information can be gained by analyzing image sequences from a vehicle interior. Even humans control the majority of the listed functions based on visual impressions.

To summarize, the aim of this thesis is to explore strategies for real-time image segmentation of non-rigid objects in a spatio-temporal domain with a stationary camera within an optical high dynamic range environment. The experiments within this work focus on image processing for vehicle interior analysis, for reasons of safety and convenience. Our investigations are not limited to cars alone, they are also valid for other transportation systems operating in open world scenes such as buses, trains, trucks, ships, etc.

1.1.2 Thesis structure

This work is divided into two parts in accordance with the general approach to tackling the problems of optical high dynamic range environments: Chapter 3 and 4 focus on hardware approaches, such as specialized camera designs and active illumination. Solutions for improved robustness of object

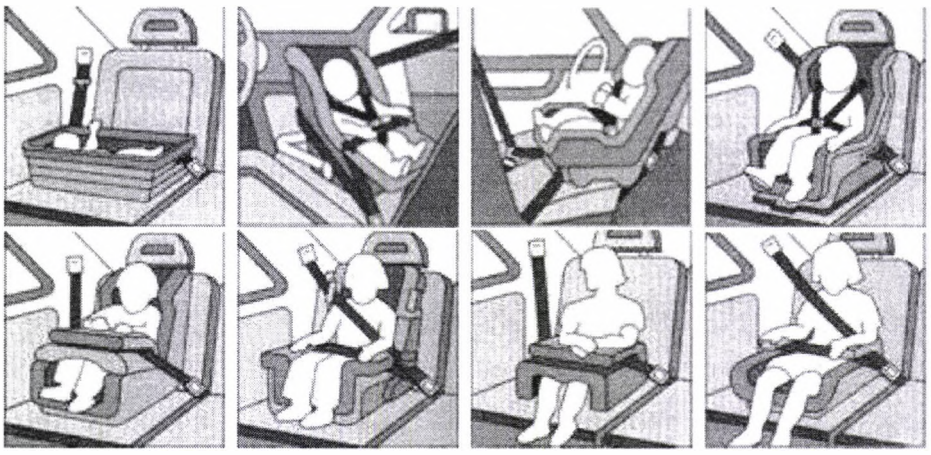


Figure 1.1: Different children and infant restraint systems.

segmentation and classification in real-time based on software alone will be presented in Chapter 5 and 6.

Chapter 2 provides an introduction to light measurement in open world scenes and defines the term optical high dynamic range. Chapter 3 starts with basics of state-of-the-art image sensors and analyzes their characteristics and suitability for image processing in HDR environments and ends by presenting a custom CMOS image sensor which was designed for fulfilling the requirements as defined in Chapter 2. Chapter 4 discusses strategies and algorithms for handling image processing in HDR environments by employing active illumination for dynamic range compression and shadow reduction.

A robust real-time segmentation algorithm using motion detection which profits by active illumination is presented in Chapter 5. The presented theoretical approaches have been tested for real applicability in Chapter 6. It illustrates the combination of the proposed active illumination techniques coupled with a texture based segmentation approach on the basis of our experimental results. An overview of the final system for real-time image segmentation in motor vehicles is given in Fig. 6.1 on Page 130.

Most of the investigations and experiments in this thesis are not limited to motor vehicles but also applicable for other machine vision tasks in high dynamic range environments, for example building surveillance, etc. However, we start with an introduction of the video-based car interior analysis because all following chapters in this thesis are related more or less to this application. The video-based car interior analysis is used as a framework in order to illustrate the derivation of optical requirements of HDR environments for a real implementation in Chapter 6.

1.2 Video-based car interior analysis

Safety equipment of modern motor vehicles combines active and passive safety. In contrast to 'active car safety' which aims to avoid accidents by controlling the car's behavior in unsafe driving situations the aim of 'passive safety' systems is to protect the vehicle occupants during and after a crash [64]. The history of passive car safety started in the fifties by examining real crashes and performing the first crash tests. The resulting knowledge led to the introduction of the safety belt which is still the number one life-saver.

The following sections give a brief overview of automotive passive safety systems in general and focus particularly on the state-of-the-art of current occupant detection systems.

1.2.1 Passive safety systems

The basis of modern passive safety systems is a body and chassis concept with a stable passenger compartment to maintain sufficient survival space and a crumple zone to absorb the kinetic energy in the event of a collision. In addition the safety belt holds the passenger to the seat to avoid serious injuries. The headrest, which is nowadays also often standard equipment for the rear seats saves the nape of the neck in case of a rear impact. Examples of restraint systems for children and infants depending on the child size, weight and age are shown in Fig. 1.1.

Another important device, which was introduced in 1981 by Mercedes-Benz to reduce the number of injuries and deaths that occur in automobile collisions, is the airbag. They quickly deploy a cushioned bag of air when a vehicle is involved in an accident. An inflated airbag for protecting passengers in case of a side impact is shown in Fig. 1.2. In addition to this modern cars have safety devices such as pyrotechnical battery cable separators designed to prevent fire after a collision and pyrotechnically tensioned seat belts. Thus, contrary to their name, modern passive safety systems are very 'active'.

1.2.2 Occupant detection systems

In 1997 automotive accidents represented about 46% of all fatal accidents worldwide [85]. This leads to the present trend that consumers prefer to buy safe cars twice as often as sporty cars. Thus the number of cars equipped with airbag systems has increased significantly in recent years. Airbags have been installed in 50 million vehicles over the past nine years and have been deployed 1.8 million times according to the estimations of the Am. NHTSA[2] [14]. But the number of airbags in modern cars also increases

[2]National Highway Traffic Safety Administration

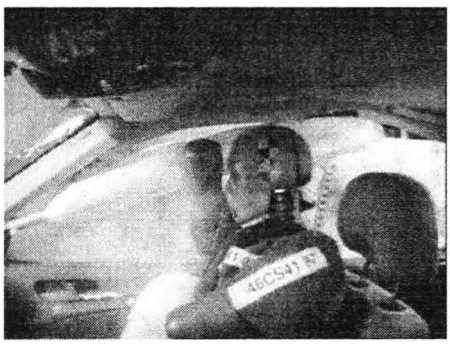


Figure 1.2: Crash with crash dummy. Demonstration of airbags for side impacts.

steadily. High end vehicles such as a BMW 7 series include up to 12 airbags for the driver, front and rear passengers. Some car manufacturers even study the feasibility of putting an airbag on the front hood of a car to protect any pedestrian who might be struck by the vehicle.

Airbags have saved several thousand lives worldwide so far [65] and protected numberless persons from serious injuries. But the number of people injured by airbags has also increased steadily. They deploy at 200-300 mph in less than 1/25th of a second, the maximum pressure being 3000 pounds per square inch. Side airbags deploy at 3 times the speed of front airbags. The size and deployment characteristics for today's airbags are calculated for belted adults sitting in an upright position. Any deviation from this position increases the likelihood of injuries.

The major shortcoming of today's airbag control units is that an inflation is triggered only by an abnormal shock (negative acceleration) inside the car, measured by a central airbag control unit. However if an occupant is too close to the airbag housing prior to deployment a large adult may be injured, a small adult is in danger but a child has no chance to escape injury. This often results in a fractured neck. Since 1990 the number of airbag fatalities had risen to 170 by the end of 2000 in the U.S., whereof about 2/3 were infants and children. Figure 1.3 shows a simulation of an unbelted child during an airbag inflation.

The federal response was to propose new regulations in order to avoid passenger injuries due to restraint systems. This demands result in so called 'smart' or 'advanced airbags' which are able to adjust deployment based on crash type (front, rear or side impact), crash severity, occupant size and position, or seat belt use, using different physical sensors:

- **Acceleration** sensors for different axis at different locations within the vehicle.
- **Torsion** of the car
- **Air pressure** sensors which measure the pressure inside a car door.

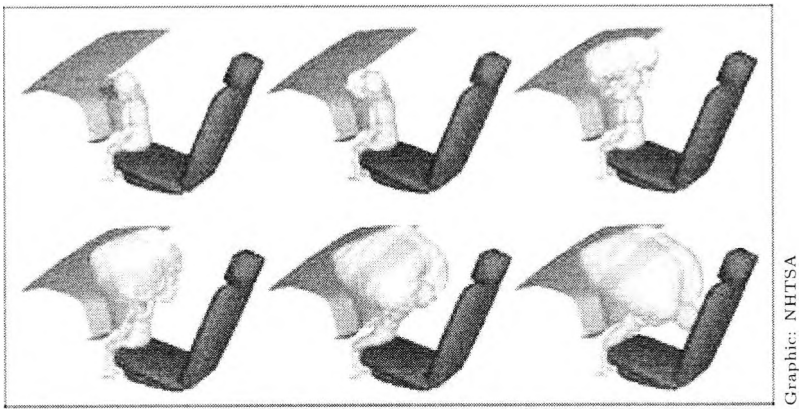


Figure 1.3: Simulation of an unbelted child during an airbag inflation.

A sudden pressure increase is a reliable hint of a side impact.

- **Gear rate** sensors for detecting roll-over crash situations

An occupation sensor is one element of this sensor tuple to control each installed airbag individually. To protect a correctly seated person ($PCSP$) the optimum protective airbag modus is to inflate to full size within a minimum of time. A front faced infant or child seat ($FFCS$) is only allowed to blow up with a reduced volume and a rear faced child seat ($RFCS$) prohibit every kind of passenger airbag. This can be realized by employing recently introduced 'dual stage' airbags, also called 'multi-stage' airbags. They contain two ignitors. One ignitor is designed to deploy (with less power) if the vehicle occupant is a small person, while the other ignitor deploys (with more power) if the occupant is a larger person. Hence, the final goal of an occupation sensor for advanced airbag control is to continuously categorize the front passenger area of the vehicle into six occupant classes as enumerated in Table 1.1. Example images of each class are shown in Fig. 6.10 on Page 140.

In addition to the kind of occupant the current occupant position is also necessary to optimize the airbag deployment. Adults may be endangered by airbags if they take up an adverse seating position or attitude, called 'out-of-position' (OOP, occupant class $POOP$).

Due to the safety relevance of this system the tolerable error ratings must ideally be zero if a human occupies the seat, in particular a child. On the contrary, the incorrect classification of a big object ($ODFC$) or an empty seat ($NOPS$) is not critical for passenger safety because an airbag deployment in this case would only increase the vehicle repair costs.

Pos.	Description	Abbr.
1	Forward facing child seats	$(FFCS)$
2	Rear facing child seats	$(RFCS)$
3	Person in correct seating position	$(PCSP)$
4	Person out-of-position	$(POOP)$
5	Nothing present on this seat	$(NOPS)$
6	An object which does not fit to the other classes	$(ODFC)$

Table 1.1: Occupant classes overview

State-of-the-art

For minimizing risks of injuries due to airbag inflation it is recommended to carry children and infants on the back seats. This is not possible in two seat cars such as sports cars and several convertibles. For these vehicles a case sensitive airbag deployment for the front seats is essential. Almost all established car manufacturers and most automotive component suppliers work on projects to solve this problem. Based on established sensor technology there are different approaches for sensing occupants in motor vehicles. Due to the high safety relevance of this application such a system must be designed redundant and error tolerant. Hence a combination of different physical measurement methods is used by most published systems for increasing the classification accuracy. The subsequent list gives a brief overview of the most important modern sensor principles for occupant detection:

- x **Weight** measurement of absolute value and/or print matrix of objects located on the passenger seat, see Fig. 1.4.
- x **Manual switch**, controlled by the driver, see Fig.1.5.
- • **Optical** analysis of multiple light beams (triangulation, see Fig. 1.6) within the near infrared (NIR) for distance measurement and shape estimation.
- • **Thermal** detection of human bodies by scanning the far infrared (FIR) spectrum
- x **Capacitive** measurements for analyzing the capacity of objects residing on the seat.
- • **Ultrasonic** distance measurements at significant spots, *e.g.* within the head, body and leg area at the passenger seat.
- x **Transponder**, which is implemented within the child seat, see Fig.1.5.
- • **Video** analysis (machine vision) of an image stream, captured by a camera, mounted within the vehicle interior.

Items with (x) mark are already employed or can be expected in the near future in high volume production, usually offered for premium cars. Ap-

proaches with (•) mark are currently in the stage of development and research.

The cheapest method to disable the passenger airbag is of course just to use a simple switch to be adjusted by the driver or the passenger itself. It is not an automatic occupant detection system, but is helps to disable the airbag in case of rear-faced child seats mounted on the front seats. The major shortcoming is that the decision depends only on the human driver. If he or she forgets to toggle the switch from 'enabled' to 'disabled' due to stress or inattention, the child or infant will be in great danger. Another disadvantage is that a static switch can not handle dynamic out-of-position situations whereby a passenger enters the danger zone in front of the airbag within milliseconds by leaning forward his upper part of his body.

An advanced version is offered by some car companies employing a transponder system for automatic child seat detection. Specially designed child seats are equipped with a transponder which consists of a transmitter and receiver coil in combination with an electronic control unit. The transmitter power supports the transponder with energy and stimulates it to return a programmable answer, which is analyzed by the passenger seat receiver. An advantage of this approach is that it is possible to detect the orientation of the child seat if two transponders are integrated. However, the major drawback of this approach is that only a limited number of special and thereby expensive child seats will be detected. In addition to this diagnosing a transponder breakdown is very tricky, increasing the possibility of erroneous airbag deployment characteristics.

High performance occupant detection and classification systems of premium car makers to fulfill the federal requirements are based on the combination of weight measurement and/or distribution by a thin matrix of measurement cells (see Fig. 1.4) and a capacity determination within the passenger seat. This system can distinguish between child seats, adults and small persons. However, movements caused by an Active Seat[3] interfere with weight measurements, the seat heating and aging influence the resilience factor of the seat foam. Furthermore, each new seat construction or modification (*e.g.* new seam design) implies an expensive re-design and calibrating of the weight mat and capacity sensor.

Research activities on video analysis within motor vehicles have increased in recent years because an occupant detection and classification approach based on machine vision promises an increased flexibility regarding system integration and reduced system costs compared to expensive weight and capacity measurements. Furthermore image data of the vehicle interior provides information for multiple safety and convenience features as discussed in Section 6.1.

[3]Active seat means an integrated passenger massage by alternating fluid pads within the seat.

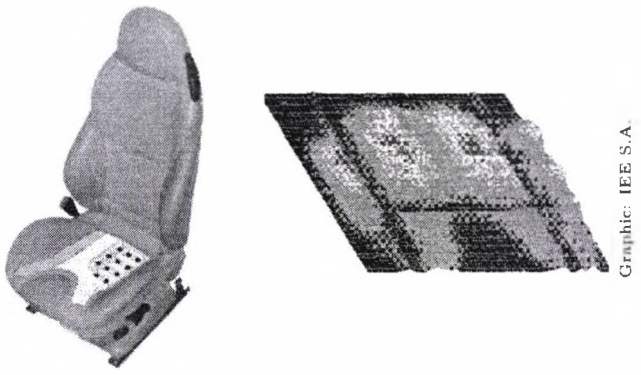


Figure 1.4: Left: seat with integrated weight measurement mat; Right: weight print of a passenger seat, occupied by a person

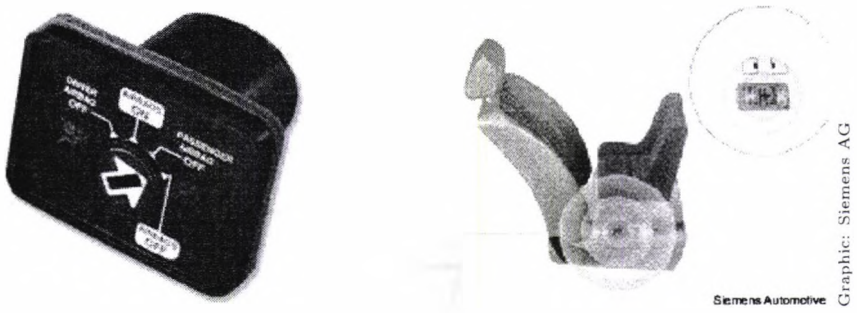


Figure 1.5: Left: manual airbag switch; Right: Siemens transponder system

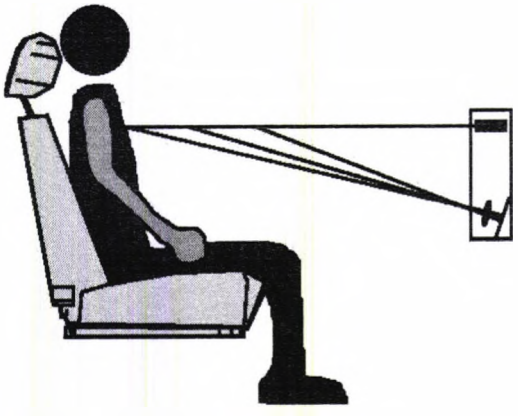


Figure 1.6: Analysis of multiple light beams (triangulation) within the near infrared (NIR) for distance measurement and shape estimation.

1.3 Previous work

Due to the fact that the recent progress of technology enables machine vision systems to leave their laboratories and comfortable industrial environments the number of research projects for outdoor applications has increased significantly. The three major pillars in this field are military applications, security applications for building surveillance and automotive applications. The following section gives a brief overview of already published approaches for monitoring the interior of intelligent vehicles for reasons of safety and convenience.

In [45] J. Krumm and G. Kirk introduced a video occupant detection system for airbag deployment. They used monochrome images taken from a single camera mounted inside a vehicle and principal components (eigenimages) nearest neighbor classifier. The classifier was trained with a set of test images of empty seats and seats with rear faced child seats ($RFCS$) which were created under varying illumination conditions to cover light fluctuations inside the vehicle. The detection system by Krumm was only able to detect a limited set of $RFCS$, that nothing is present or that an object is present which does not fit into the $RFCS$ class, but the final accuracy of the system was declared 99.5 percentage. However, the disadvantage of the principal components analysis (PCA) in this case is that the appearance of an occupied seat is so variable and not limited to trained $RFCS$s.

In [61] S.-B. Park proposed a video based system for optical occupant detection based on the idea of searching for a human face in the scene. If an adult face can be detected properly in case of a crash the airbag should be inflated, otherwise it should not. This approach was also based on principal component analysis, in this application in particular on eigenfaces (see Fig. 1.7). Thus it is limited in its real application due to the high illumination fluctuations which are present in a vehicle and the fact that a straight look into the occupant's face is not guaranteed all the time. Also great size variations of the passenger's face when they move along the optical axis of the camera yielded misclassifications. Furthermore, a more accurate distinction between $FFCS$, $RFCS$, and $POOP$ is necessary for optimized control of multi-stage airbags.

Due to the fact that PCA based approaches are not able to provide information about dynamic of out-of-position situations Owechko presented a multiple feature set system in [59]. The four types of features utilized in the architecture are range (obtained using stereo-vision), edges, motion and template matching based on the Hausdorf metric. The response rate for the system running on a 400 MHz Pentium II was specified greater than 20 updates per second, except for the template matching module which requires approximately 2 seconds to update its decision.

Other approaches for vehicle interior analysis based on range maps by

Figure 1.7: Test set of eigenfaces, used by Park *et al.* in [61].

stereo imaging have been presented by Faber in [17] and Klomark in [35]. However, three dimensional techniques are not the subject of this work due to restraints discussed in Appendix A.

To summarize, most recent research within this field has focused on software solutions for commonly available hardware. It has ignored the demanding image acquisition task of HDR scenes due to the assumption that the image quality is high enough for testing the proposed algorithms as a pre-condition. In contrast to recently published approaches this work emphasizes the benefits of a smart fusion of hard- and software (sensor and algorithm) for results optimized in terms of cost, speed and reliability. A more detailed overview of the state-of-the-art for machine vision aspects is given in the first section ('Motivation') of each chapter.

Chapter 2

Optics and Radiometry

2.1 Motivation

The optical system in conjunction with a suitable image sensor and illumination is the first step to obtain suitable pictures for the subsequent image analysis. The camera location influences the lens characteristics and *vice versa*. Hence different imager positions for analyzing a motor vehicle interior are discussed in Section 2.2 as an example for machine vision applications which suffer from HDR fluctuations.

The major subject of this chapter is to investigate the optical requirements for image processing in environments which are characterized by large intensity fluctuations and hence a high optical dynamic range (HDR). This conditions are common for open world scenes which have to be managed by outdoor surveillance systems or motor vehicles. Therefore Section 2.3 and Section 2.4 measure light intensities in open world scenes (radiometry) and define the term optical high dynamic range. State-of-the-art camera techniques will be compared for their HDR suitability in Chapter 3 based on the requirements as defined in this chapter. The measurements within this chapter are also the basis for the active illumination experiments done in Chapter 4.

The use of active illumination as detailed in Chapter 4 is limited by several constraints such as eye safety aspects (see Appendix B), camera features and kind of illumination source. Therefore Section 2.4 analysis suitable illumination sources for image processing in motor vehicles.

2.2 Optical system

Images are the two-dimensional projection of a three-dimensional world. Basically the optical information that is gathered by a camera and mapped by a lens to the image plane corresponds to brightness points, therefore it is called a 2-D intensity image. This makes it necessary not to regard the imaging sensor alone but the sensor in conjunction with the whole optical system.

The following sections provide the mathematical and physical definitions for reconstructing the experiments and results described within this work concerning digitization of 2-D intensity images. The focus is on optical terms which are particulary important for further sections and chapters dealing with optical systems and in particular for light measurement (Section 2.4.1) and active illumination (Section 4.2).

A more precise introduction to geometrical optics and image digitization can be found in various textbooks such as those by Pedrotti [66] or Sonka [81].

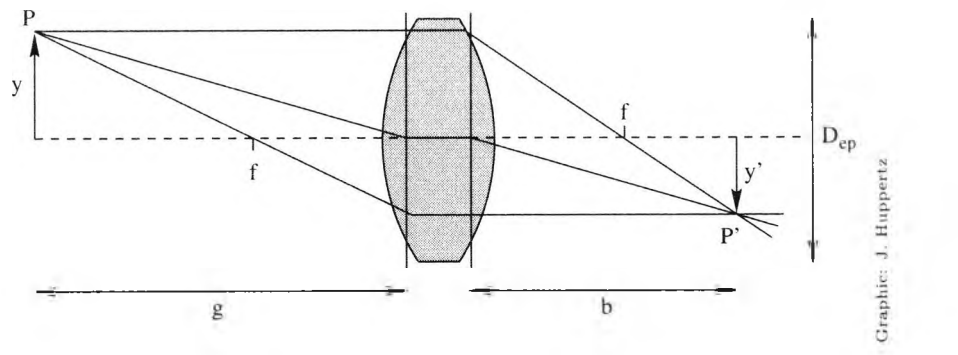


Figure 2.1: Descartes

2.2.1 Geometrical optics

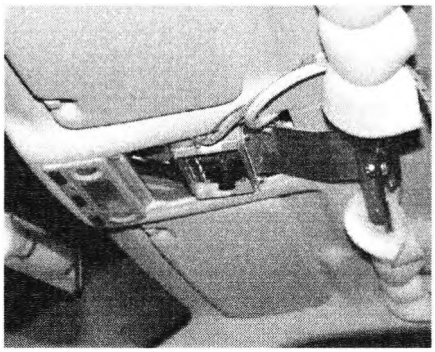
To determine the elementary lens features such as focal length f and field of view which form an intensity image the dimensions of the observed space are required. Light emitted from an object at point p is refracted by optical lenses and bundled in a point p' of the picture plane. Eqn. 2.1 shows the fundamental lens formula defined by Descartes for what is called a thin lens. The picture scale is equal to the ratio of the length g and b:

$$\frac{1}{b} + \frac{1}{g} = \frac{1}{f} \tag{2.1}$$

The field of view 2ω is defined as the maximum cone or fan of rays passing an aperture and measured at a given vertex. That means the field of view (apex angle) is the span of objects that can be captured through a lens. It is determined by the greatest width d_{roi} of the region of interest (ROI) and the distance s_{cam} between imager and scene surface. If the imager is centered above the ROI, the apex angle will be calculated by

$$\omega = \tan\left(\frac{d_{roi}}{2s_{cam}}\right) \tag{2.2}$$

For monitoring a motor vehicle interior a relative short distance between objects (passengers) and imager s_{cam} in conjunction with a large observed area d_{roi} yields a large field of view which again causes increasing distortions inside the images. The effect of large apex angles can be observed in Fig. 2.4(cp2) on Page 19. These images are made with $2\omega \simeq 120^o$ showing a slight 'fish eye' effect. The necessity of calibrating and re-computing the distortions caused by the optical system within the images depends on the subsequent image processing approaches, in particular on the final classification algorithms which will be discussed in Section 6.4.



Camera at position (cp2)

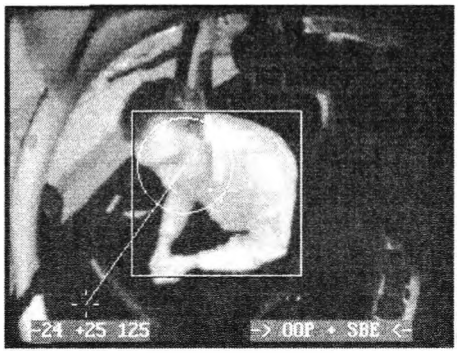


Image from (cp2)

Figure 2.2: (a) HDR camera (Fraunhofer IMS) located within the interior roof; (b) example image for OOP detection (SollyCam, see Section 3.6).

2.2.2 Camera location

This work investigates methods and techniques for an interior surveillance of motor vehicles via a monocular (see Appendix A) imager system. The following sections focus on the front seat surveillance due to safety problems caused by erroneous airbag deployment as described in Chapter 1. Infant seats located on the centre back seat are not seriously affected by airbags. However, one reason for also observing the back seat passengers is to track their heads for optimizing the inflation of airbags which are mounted next to the car windows ('airbag curtain', see Fig. 1.2 on page 6). The requirements and terms for back seat surveillance are similar to the front surveillance, but they are not part of this work.

There are three fundamental aspects which have to be taken into account for determining a suitable monocular imager location: First of all the machine vision aspect. A position to maximize the available reliable information of the sampled images by minimizing the computing costs. For instance a direct view of the sideways passenger outline.

Secondly the optical aspect: The dimensions of a vehicle interior are nearly fixed and vary only in a small range according to the vehicle model and type, *i.e.* van, roadster, bus or truck. These interior dimensions in combination with the imager position determine the lens features to ensure a sharp map of the monitored region on the sensor. An optical interior surveillance requires the placing of the imaging sensors in a position where a free view of safety relevant areas is ensured. By focusing on the front seats the back seats are mainly concealed by the front seats and head rests. This requires at least a second camera location if the surveillance of the whole car interior is desired, see Table 2.1 and Appendix A.3. Due to this geometric constraint the apex angle 2ω (field of view) is one of the most important

characteristics of the lens.

The third important aspect for determining a suitable imager location involves the mechanical integration: Not every position which provides a suitable view of the interior is possible due to design, safety and mechanical integration aspects. For instance the system should be inconspicuous for occupants and safety relevant parts of the chassis must not be weakened or impaired. A spot with already existing electronic components is to be preferred to minimize the integration effort. The illumination has to be integrated close to the camera for minimizing shadow effects (see Section 4.3).

Another point is the location temperature. A major disadvantage of plastic optics is its sensitivity to temperature fluctuation and low resistance to scratching. If operating or storage temperatures exceed $70 \sim 90^{o}C$ there will be optical deformation of the lens, which will significantly influence the image quality. Furthermore the dark noise of the imager (see Section 3.3) will increase with rising temperatures. Hence a location for the imager is preferably remote from hot areas. Hot areas within a motor vehicle are for example next to the heating or certain spots within the vehicle roof. Hence not all integration aspects can be ideally satisfied at the same time and it is necessary to find the optimum compromise between performance and constraints.

Suitable camera positions, their advantages, shortcomings and resulting apex angles are listed in Table 2.1. Every item was rated with positive (+) or negative (−) symbols. The number of symbols indicates its importance regarding feasibility and integration costs. The apex angle is an approximation for monitoring the passenger seat according to Eqn. 2.2 based on a vehicle with interior dimensions similar to a middle class sedan.

These values can also be used in general as reference for larger vehicles such as vans and off-road vehicles because the higher roof implies an increased distance between the imager and the monitored area. Hence the necessary apex angle and the resulting distortions of the image decrease. Smaller vehicles with reduced distance s_{cam} between image plane and object mean greater demands and limitations to the optical system.

Example images from each investigated position are shown in Fig. 2.4, and Fig. 2.3 shows $cp1 \ldots cp4$ a bird's view of a sketch of a car. The arrow size indicates the area of the interior (ROI) which can be monitored without interference by seats, head rests, steering wheel based on $2\omega \leq 120^{o}$.

Advantages compared with the shortcomings of particular camera locations lead to the conclusion that the interior light module is the most favorable location ($cp2$) inside a motor vehicle to mount an imager for interior surveillance. Due to this finding most experiments and measurements in the following sections have been made with the image sensor mounted close to $cp2$.

(cp1)	**Dashboard**
	$2\omega \approx 160^o$
+	Lower temperatures compared to a location within the vehicle roof
-	Limited field of view; monitoring of driver or passenger only
- -	High possibility of affected field of view, e.g by newspapers or maps etc.
(cp2)	**Interior light module**
	$2\omega \approx 120^o$
+	Perpendicular view to the keep-out zone
+ +	Less affected by occlusion than (cp1) or (cp3)
+++	Existing electronic unit means cost saving link to electronic vehicle infrastructure
-	Near to the roof and windscreen means high temperatures up to 100^oC
(cp3)	**A-pillar**
	$2\omega \approx 90^o$
++	Perpendicular view to the passenger seat, yielding simplified motion estimation
-	High possibility of adversely affected field of view due to occlusion by the driver
- -	Expensive implementation due to packaging limitations: The A-pillar already holds the airbags for side impacts
(cp4)	**Roof center**
	$2\omega \approx 150^o$
++	Passenger and driver seat monitoring is possible simultaneously
- -	The area between person and dashboard/steering wheel is concealed by the driver/passenger
-	High temperatures

Table 2.1: Feasible camera locations. The first column indicates the weighted advantage or disadvantage of each position (see also Fig. 2.3 and 2.4).

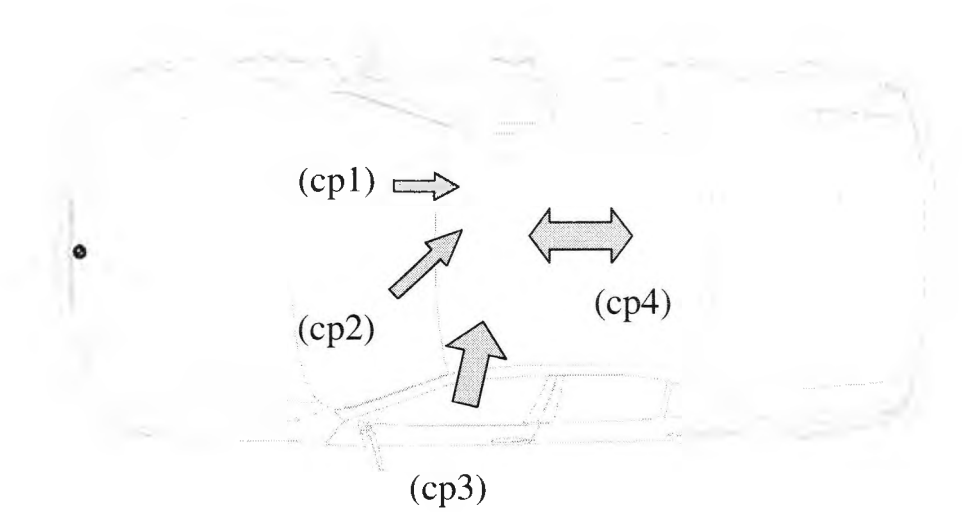
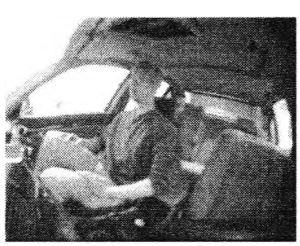
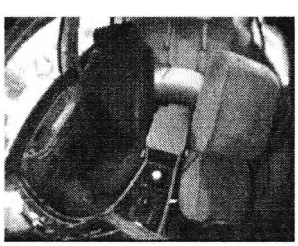
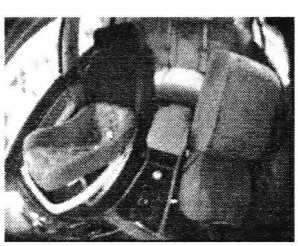
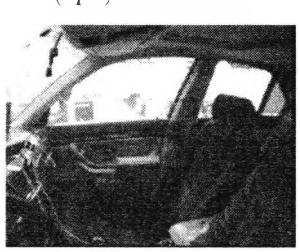


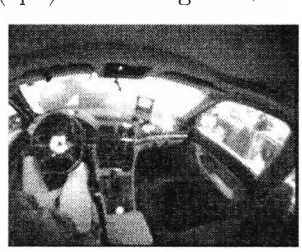
Figure 2.3: Possible camera locations. The size of the arrows indicates the area of the interior (ROI) which can be observed without restriction, assuming $2\omega \leq 120^o$.


(cp1) Dashboard


(cp2) Interior light module


(cp2) Interior light module


(cp3) A-pillar


(cp4) Roof center

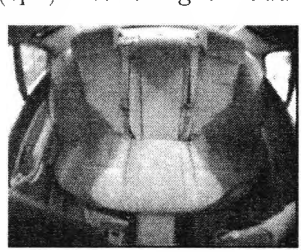

(cp4) Roof center

Figure 2.4: Positions within an interior of a motor vehicle for video-based car interior analysis according to Table 2.1 and Fig. 2.3

Symbol	Indication	Definition	Unit
Q_e	radiant energy	Product of radiant flux and time (s+r)	J, Ws
P_e, Φ_e	radiant power	From a source emitted or from an area transmitted radiant energy per time (s+r)	W
I_e	radiant intensity	From a source emitted radiant power per solid angle (s)	$\frac{W}{sr}$
L_e	radiance	From a source emitted radiant power per solid angle and emitting area (s)	$\frac{W}{cm^2 \cdot sr}$
E_e	irradiance	Incident radiant power per area (r)	$\frac{W}{m^2}$
H_e		Radiant energy per area (r)	$\frac{Ws}{cm^2}$

Table 2.2: Overview of radiometric measurement units. (s)= source (r)= receiver

2.3 Light measurement and filtering

An imager generates a two-dimensional intensity map of a three-dimensional scene. In this process the reflected amount of light from object surfaces is mapped onto a light sensitive sensor (imager) defined by the optical system as described in Section 2.2.1. This amount of light determines the power of light which is incident on the sensor.

2.3.1 Power of light

Radiant flux is a measure of radiometric power. Flux is a measure of the rate of energy flow and is expressed in watts and joules per second, respectively. Luminous flux is a measure of the power of visible light. Photopic flux, expressed in lumens, is weighted to match the responsivity of the human eye which is most sensitive to yellow-green, according to Fig. 2.9. Optical power within the infrared is therefore weighted less. Hence measurements of photopic and luminous flux are not suitable for machine vision within the near infrared spectrum, which applies for the techniques described within this work. Therefore the following deals with radiometric units only. Radiometric units are usually labelled by index e as shown in Table 2.2.

Watt (W), the fundamental unit of optical power, is defined as a rate of energy of one joule (J) per second. Optical power is a function of both

the number of photons and wavelength. Each photon carries an amount of energy that is described by Planck's equation by

$$Q = \frac{h \cdot c}{\lambda} \tag{2.3}$$

where Q is the photon energy (joules), h is Planck's constant ($6.623 \cdot 10^{-34} Js$), c is the speed of light ($2.998 \cdot 10^{8} ms^{-1}$), and λ is the wavelength of radiation (meters). All light measurement units are spectral, spatial, or temporal distributions of optical energy. Since Q is inversely proportional to wavelength, ultraviolet photons carry more energy than photons within the visible or infrared range.

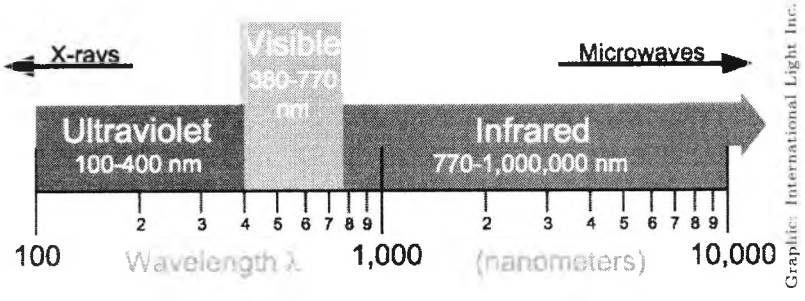


Figure 2.5: Optical portion of the electromagnetic spectrum

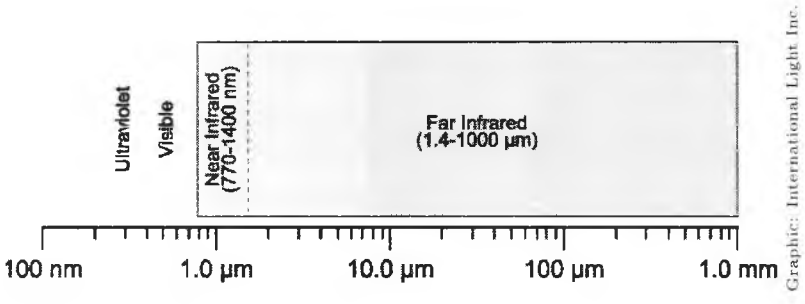


Figure 2.6: Infrared sections

One of the key concepts to understanding the relationships between measurement geometries is that of the solid angle, or steradian. A sphere contains 4π steradians. A steradian is defined as the solid angle Ω which has its vertex at the center of the sphere and cuts off a spherical surface area A equal to the square of the radius r of the sphere. For example, a one steradian section of a one meter radius sphere subtends a spherical surface area of one square meter.

$$\Omega = \frac{A}{r^2} \tag{2.4}$$

Most radiometric measurements do not require an accurate calculation of the spherical surface area to convert between units. Flat area estimates can be substituted for spherical area when the solid angle is less than 0.03 steradians resulting in an error of less than one per cent. Therefore the following calculations estimate A as a planar surface area [73].

2.3.2 Transmittance and filter

The amount of light that passes an optical lens system represents the transmittance. The transmittance factor τ is a function of the wavelength λ and, depending on the filter type, of the wave angle α_{in} of incident light. Light which is not transmitted was reflected or absorbed, for example by an optical filter (see Fig. 2.7).

$$\tau_{filter} = \frac{P_{transmit}}{P_{reflect}} = f(\lambda, [\alpha_{in}]) \tag{2.5}$$

$$P_{reflect} = (1 - P_{transmit}) \tag{2.6}$$

A longpass filter has low transmission in the shortwave region and high transmission in the longwave region and *vice versa*. A typical example is a NIR cutoff filter: Most image sensors are sensitive to emissions within the near infrared (NIR), contrary to the human eye (see Fig. 2.9 on page 27). That is why in multimedia applications an additional shortpass filter for blocking NIR is included in the lens system to assimilate the sensor response to the human eye characteristics.

The overall spectral sensitivity of a detector is equal to the product of the responsivity of the sensor and the transmission τ of the filter. Therefore multiple added filter layers yield a transmission equal to the product of the individual transmissions. Given a desired overall sensitivity and a known detector responsivity the optimum filter transmittance can be determined. Thus the colors of images showing the same scene and identical illumination situation may differ due to slight deviations of the sensitivity of imagers (see Chapter 3) and transmittance characteristics of the overall optical system.

Optical filters operate by absorption or reflection (*i.e.* interference). Most lens materials transmit both the visible and infrared spectrum.

Color filters usually consist of glass substrates or plastic which are endowed with materials that selectively absorb certain wavelengths (mass filter). The peak transmission relates to the additives while the bandwidth depends on the layer thickness. Hence varying the filter thickness makes it possible to selectively modify the spectral responsivity of a sensor to match a particular function. Glass lenses can also be coated to obtain certain transmission characteristics similar to substrate pigmentation. The boundaries between the passing band and the blocked band are not that sharp but the

performance does not depend on the input angle. Furthermore mass filters are far less expensive and thinner than interference filters and therefore easier to integrate into the imager optics.

Interference filters are made of multiple thin dielectric layers on a substrate. Each layer is separated by an optical distance d_o of a multiple of half the desired wavelength λ_{pass} that shall pass.

$$\begin{aligned} \lambda_{pass} &= f(d_l, \epsilon_{in}) \\ d_l &= n \cdot \left(\frac{\lambda_{pass}}{2} \right) \qquad n \in \mathbb{N} \qquad (2.7) \\ \Delta d &= 2 \cdot d_l \cos \epsilon_{in} \qquad (2.8) \end{aligned}$$

They use interference to selectively transmit or reflect a certain range of wavelengths. If wave fronts overlap in phase with each other, the amplitude of the wave increases. Wave fronts which are out of phase due to their optical retardation (optical path) within the layer erase each other. Hence interference between wavefronts causes rejection of wavelengths outside the pass band by reflection. A typical example is a bandpass interference filter that transmits a narrow range of wavelengths and can isolate a single emission line from a discharge lamp. The major disadvantage of filtering by interference is that the center wavelength shifts with the input angle ϵ_{in} of the incident light because the optical path Δd within each layer increases with the cosine of ϵ_{in} as shown in Eqn. 2.7.

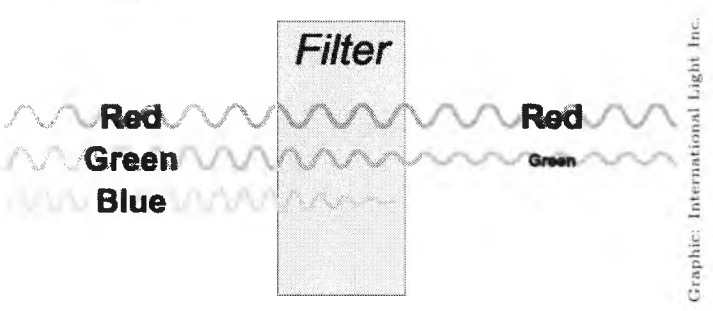


Figure 2.7: Absorption filter principle

If a supplementary illumination with a special bandwidth is used for illuminating a scene for machine vision it is obviously preferable to employ an optical bandpass (usually an interference filter) to increase the signal-to-noise ratio for the sensor (see Section 2.4). But the combination of the imager characteristics (usually consisting of silicon) and a common lowpass filter also results in a bandpass see Section 2.4.1.

The light sensitive semiconductors of an imaging sensor are covered by several thin layers of silicon oxide (see Chapter 3). Their thickness is in the order of magnitude of the wavelength of visible and near infrared radiation.

This means these thin layers build an interference filter, which is relevant to the dispersion of the absorbed light. Due to reflection at the boundary of two layers waves can eliminate themselves if the double layer thickness is equal to a multiple of the wavelength. An ideal interference filter has very uniform thickness and the surfaces are silver/metal coated. However the oxides covering the circuits of an imaging sensor have neither an uniform thickness nor a silver-coated surface.

Therefore the interference effect is not very strong, minima are not very clear compared to an ideal interference filter and the leakage is much less. But the transmission curve resulting from the layer sequence above the photo detectors is overlaid by the natural detector curve and must be mentioned if a sensor should be tuned for special wavelength (see Section 4.2.1).

2.4 Passive illumination

Image understanding consists of extracting features of the image and analyzing them, *i.e.* edges, histograms, etc. Variations in the illumination have a direct influence on the features that can be extracted from the image. A key problem is therefore to find ways of handling feature variations due to illumination changes. For instance, edges which a machine vision algorithm is looking for should be identified independently of the illumination under which the scene was captured.

Each grey value GV detected by the imager after passing lenses and optical filters consists of the product of ambient irradiance E_{amb} in the scene and the reflection factor ρ_s of the object surfaces (motor vehicle interior). A mathematical formulation of the imaging process is depicted in Eqn. (2.9). For a more precise definition of E please refer to Section 2.3.

$$GV = E_{amb} \cdot \rho_s \tag{2.9}$$

Any change in scene illumination E_{illu} yields variations of the intensity image. In order to cope with this problem, there are two options in general. Either the image processing algorithms are designed to handle this problem by taking feature variations into account for the feature extraction by an image processing filter or the illumination of the scene can be controlled and modified by hardware in such a way that $E_{illu} \approx const.$ The latter case most frequently occurs in industrial environments, indoor photography, etc. Approaches for suppressing disturbing illumination influences and variations will be presented in Section 4.2.1.

Illuminating a scene with controllable light sources is called active illumination or supplementary illumination in this work. Illumination influencing the scene which can not be manipulated is called passive or ambient illumination.

The main questions regarding illumination are: Which general illumination situations appear in open world scenes which influence motor vehicles such as cars? What is the most powerful ambient illumination source (optical noise)? How much optical power has this radiation source inside a motor vehicle? What can be done to eliminate or minimize ambient noise influence? If supplementary lighting is used: how much radiant power is necessary to achieve a sufficient signal-to-noise ratio (SNR)? What is the optimal illumination type (wavelength) for such a supplementary illumination?

The most powerful passive light source in open world scenes outside industrial or scientific domains is naturally the sun. But external lamps *e.g.* traffic lights, street lighting and headlights of other cars may also significantly influence imaging in open world scenes. Furthermore not only external light sources should be considered but also lights which are already implemented within motor vehicle interiors. There are in present premium cars up to 400 light sources including displays, LED's, incandescent lamps, etc., with diverse wavelength ranges which contribute to interior illumination fluctuations and therefore influence imaging in motor vehicles.

On its surface the sun has a power output of approximated $60MW/m^2$. Due to the distance of 150 million km to earth, only fractions of this power reach the earth atmosphere. The earth's orbit around the sun is eccentric so that the remaining power varies between its maximum of $1405W/m^2$ in January and $1308W/m^2$ in July [22]. Typical irradiance levels for outdoor situations in middle Europe are listed in Table 2.3.

Condition	**E[W/m^2]**	**photons/m^2**
summer, noon	10^3	$3 \cdot 10^{21}$
sunrise/sunset	1	$3 \cdot 10^{18}$
full moon	10^{-4}	$3 \cdot 10^{14}$
night without moon	10^{-6}	$3 \cdot 10^{12}$

Table 2.3: Typical outdoor illumination situations.

The spectral distribution changes when light travels through the atmosphere due to absorption and dispersion. The atmosphere represents an optical filter which attenuates the transmitted radiation. In particular water vapor, oxygen and carbon dioxide cause the main absorption which yields band vacancies (see Fig. 2.8). The mean for the remaining irradiance has its maximum of about $1000W/m^2$ for vertical sunlight hitting the ground [11].

The attenuation in the atmosphere depends on the climatic zone, its special composition, amount of scatter objects (*i.e.* clouds) and the length of the way through the atmosphere, which depends on the sun angle. Even

the dimension and the center wavelength of these so called 'water gaps' at about $\lambda = 770nm, 820nm, 940nm, 1120nm$ due to the H_2O content of the atmosphere changes slightly depending on the position on earth. These gaps within the sun power distribution are usually used for optical applications within the NIR, *i.e.* IR remote controls, light gates and optical networks.

Hence if the influence of the sun has to be reduced for machine vision applications it is advantageous to use these gaps for a controlled illumination environment. The largest gap is located within the NIR at $\lambda \approx 1400nm$ but the optical sensitivity and efficiency factor of mainstream camera systems based on silicon decreases at increasing wavelength and is reduced to one tenth beyond $\lambda \approx 900nm \pm 60nm$ depending on the image sensor type, see Fig. 2.12.

Infrared light (see Fig. 2.5 and 2.6) contains the least amount of energy per photon from all optical bands, see Eqn. 2.3. Therefore an infrared photon often lacks the energy required to pass the detection threshold of a quantum detector. Hence typical silicon photodiodes which are the basic part of mainstream imagers are not sensitive beyond $\lambda > 1100nm$. The aim is therefore to determine the optimum illumination wavelength where the signal-to-noise ratio of sun and ambient illumination is a maximum in relation to the efficiency factor of the sensor.

The sensitivity of mainstream image sensors is high in the visible range but a supplementary illumination for motor vehicles must be outside the visible range of the human eye. It is very important that the passengers of the vehicle must not be affected by supplementary illumination. The sensitivity of the human eye adapted to and to darkness is shown in Fig. 2.9.

Due to these findings the focus on illumination measurements in Section 2.4.1 was on wavelengths around the $770nm$ gap.

2.4.1 Optical dynamic range

The acquisition of images in a motor vehicle must be adapted to the physical influences of the environment, *i.e.* to the light conditions in open world scenes. Light conditions are determined by the maximum and minimum of occurring light (optical dynamic) and their function over time.

A distinction can be made between multimedia and machine vision applications. Both applications need an imager which provides appropriate images of the scene. But the term 'appropriate' may differ. For multimedia applications humans are able to handle blurred, smeared and saturated image areas, but a high resolution with mega pixels is usually demanded.

On the other hand, conventional machine vision systems are less robust concerning poor image quality and image areas with ambiguous information. In particular images with over- or under-exposed areas cause serious problems for image processing. This is not acceptable for safety applications such as lane departure warning, adaptive cruise control or occupant detection for

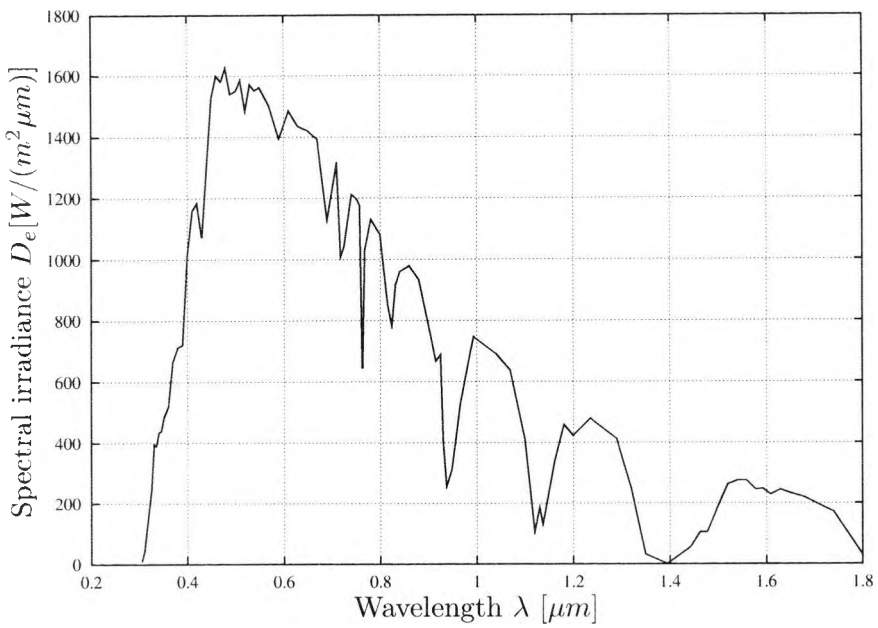


Figure 2.8: Spectral irradiance of the sunlight after earth atmosphere according to the international standard IEC 904-3 with 'water gaps' at about $\lambda \approx 770nm$, $\lambda \approx 940nm$ and $\lambda \approx 1120nm$.

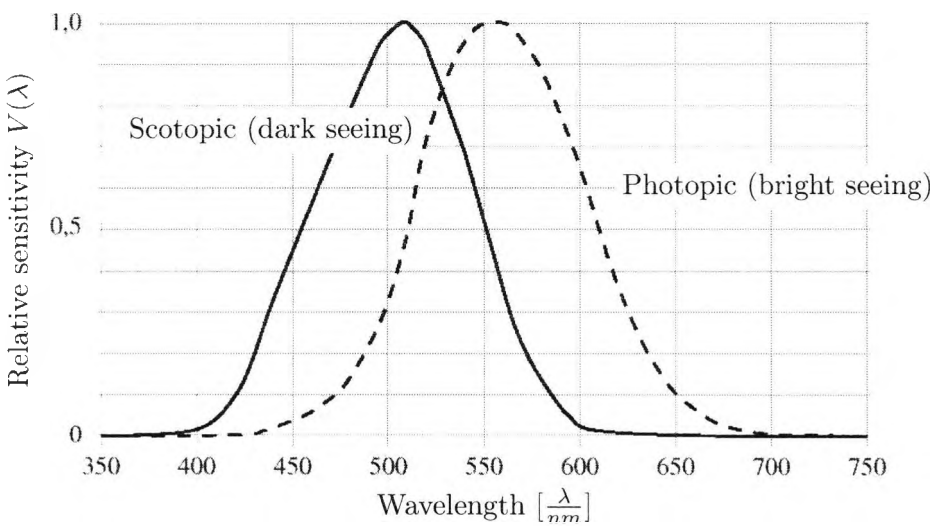
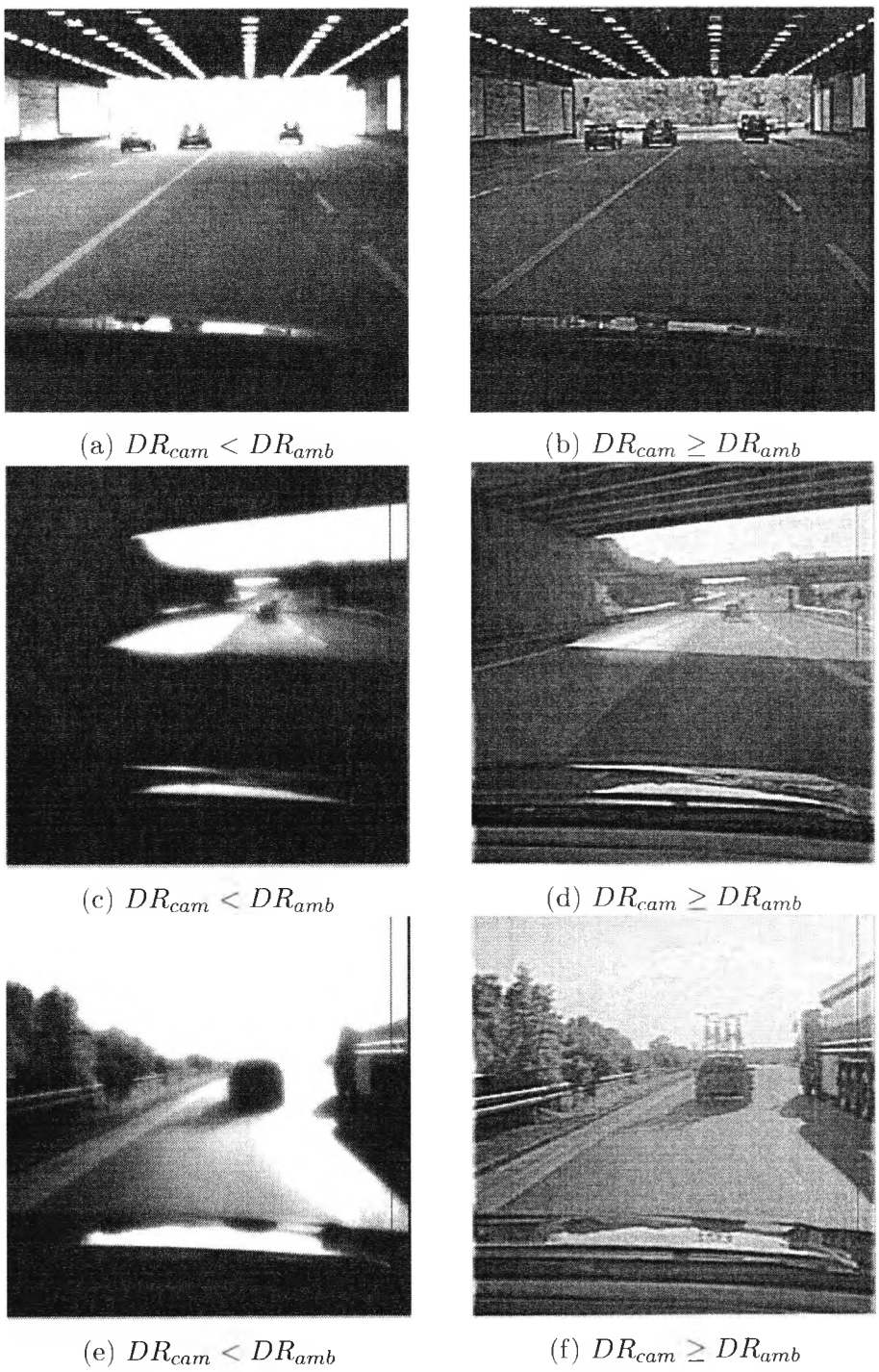


Figure 2.9: Human eye sensitivity for dark and bright environments



(a) $DR_{cam} < DR_{amb}$ (b) $DR_{cam} \geq DR_{amb}$

(c) $DR_{cam} < DR_{amb}$ (d) $DR_{cam} \geq DR_{amb}$

(e) $DR_{cam} < DR_{amb}$ (f) $DR_{cam} \geq DR_{amb}$

Graphic: Fraunhofer Gesellschaft IMS

Figure 2.10: Typical example of lost image information due to an insufficient optical dynamic of an imager (DR_{cam}) in open world scenes (DR_{amb}).

airbag control. Furthermore the maximum resolution is not necessarily used for image processing because the processing costs rise significantly according to the number of pixels.

Yamada calculated in [92] that conventional CCD cameras failed 10 % of the operating time at daylight for lane mark detection because the share of saturated pixels is larger than 20 %. The same is true for a vehicle interior application. For example, the time when the sensor is blinded (*e.g.* due to reflections within the interior) may be quite short. However, within this time some of the occupants may get too close to the airbag housing and are immediately in danger in case of an airbag deployment.

The importance of the optical dynamic range for image processing has come into focus in recent years due to the fact that machine vision tries to leave the ideal conditions of the laboratory or industrial environments with controlled illumination situations.

It is important to distinguish between local (*spatial domain*) and global dynamic range (*time domain*). Global dynamic range means the light fluctuations of the whole scene *over time* such as from sunrise till noon. Local dynamic range means brightness differences *within* the scene. There is the possibility of adapting the imager according to the global dynamic within a limited range by varying sensor parameters such as exposure time, aperture, gain, etc. But if the imager does not fit to the local dynamic range of the scene then image detail is usually lost.

To determine the radiation level inside a motor vehicle (*e.g.* car) and thus the optical dynamic range ordinary and worst-case illumination situations have been measured. The measurement setup consists of an optical power meter for sampling the global illumination in front of the sensor (time domain). The worst-case behavior for the spatial domain can also be derived from this data. The results of these measurements yield the data for determining the necessary camera features and the requirements for a supplementary illumination (see Section 4.2).

Global dynamic

The global dynamic within a motor vehicle was measured with an optical power meter [57] which consists of silicon (Si) based photodiodes operating in photovoltaic mode and converts the incident photons which hit the sensor area (A_{sensor}) into an output current (I_{out}). This current represents the integral of the environment irradiance $E_{env}(\lambda)$ multiplied by the spectral sensitivity of the sensor R_{sensor} over a determined wavelength range.

$$I_{out} = A_{sensor} \cdot \int_{\lambda_{min}}^{\lambda_{max}} E_{env}(\lambda) \cdot R_{sensor}(\lambda)\, d\lambda \tag{2.10}$$

This Si detector is sensitive for wavelengths between 400nm (VIS) and

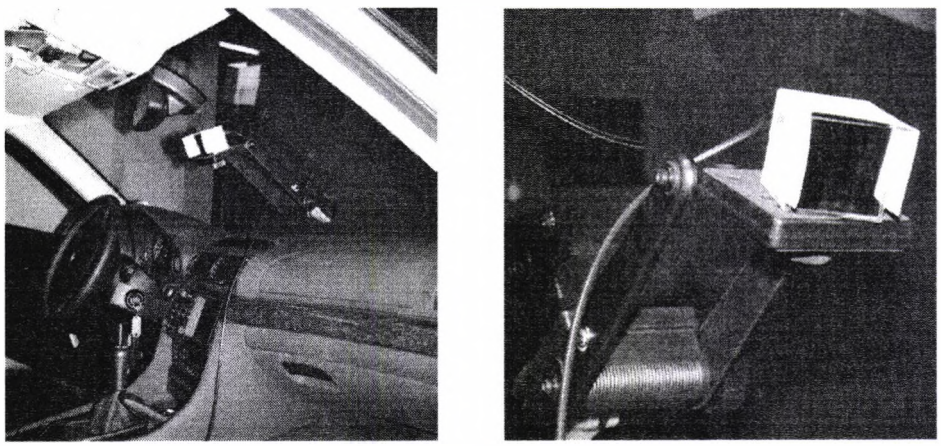


Figure 2.11: Setup for the dynamic range measurement.

1000nm (NIR). The setup for the measurement is shown in Fig. 2.11 and the sensor characteristics $R_{sensor}(\lambda)$ in Fig. 2.12.

One possibility of determining the irradiance E_{env} is first to measure the characteristic spectral gradient of the source (spectral power density) by using a spectrometer and then to calculate the present irradiance.

Assuming that the sun is the most powerful illumination source for the vehicle outside but also for the interior according to Section 2.4, $E_{env}(\lambda)$ can be approximated as the spectral power density of the sun light attenuated by the transmittance T_{optics} of the optics:

$$E_{env}(\lambda) \approx E_{sun}(\lambda) \cdot T_{optic}(\lambda) \tag{2.11}$$

$$E_{sun}(\lambda) = E_{sun,norm}(\lambda) \cdot k \tag{2.12}$$

The standardized spectrum $E_{sun,norm}$ is stretched by using a factor k which defines the total high. The spectral power density of the sun according to the international standard IEC 904-3 [11] is shown in Fig. 2.8. The power distribution differs slightly depending on the position of the measurement on earth and the characteristics of the atmosphere. If Eqn. 2.11 and 2.12 are inserted into Eqn. 2.10 and solved for k, then

$$I_{out} = A_{sensor} \cdot \int_{\lambda_{min}}^{\lambda_{max}} E_{sun,norm}(\lambda) \cdot k \cdot T_{optic}(\lambda) \cdot R_{sensor}(\lambda)\, d\lambda$$

$$k = \frac{I_{out}}{A_{sensor} \cdot \int_{\lambda_{min}}^{\lambda_{max}} E_{sun,norm}(\lambda) \cdot T_{optic}(\lambda) \cdot R_{sensor}(\lambda)\, d\lambda} \tag{2.13}$$

The whole irradiance of the source is according to Eqn. 2.12:

$$E_{env} = \int_{\lambda_{min}}^{\lambda_{max}} E_{env}(\lambda)\, d\lambda$$

$$= \int_{\lambda_{min}}^{\lambda_{max}} E_{sun,norm}(\lambda) \cdot k \cdot T_{optic}(\lambda)\ d\lambda \tag{2.14}$$

If Eqn. 2.13 is inserted into Eqn. 2.14 this shows the relationship between sensor output current and irradiance of the source:

$$E_{env} = I_{out} \cdot \underbrace{\frac{\int_{\lambda_{min}}^{\lambda_{max}} E_{sun,norm}(\lambda) \cdot T_{optic}(\lambda)\ d\lambda}{A_{sensor} \cdot \int_{\lambda_{min}}^{\lambda_{max}} E_{sun,norm}(\lambda) \cdot T_{optic}(\lambda) \cdot R_{sensor}(\lambda)\ d\lambda}}_{gain} \tag{2.15}$$

The fraction *gain* is a constant and can be calculated for each measurement setup. The Multiplication of this constant with the output current I_{out} of the sensor yields the final measured irradiance.

Two scenarios were measured: first by applying an optical bandpass filter ($T_{optics}(\lambda) = T_{bp}(\lambda)$) which cuts out a wavelength range centered at about $\lambda_{cwl} = 780nm$ within the NIR according to the results of Section 2.4. The transfer function for such an optical interference filter with a half-power bandwidth (HBW) of 10.7nm is shown in Fig. 2.13.

Secondly without any optical filter. This means $T_{optic}(\lambda)$ is equal to 1 and the wavelength integral is only determined by the range where the sensor is sensitive. These measurements serve only as reference because an illumination based on NIR provides several advantages which can not be achieved by visible light (see Section 2.4). All irradiance levels mentioned in the following sections assume that an optical filter such as a bandpass was used for blocking the visible light range unless explicitly indicated otherwise.

Finally the global dynamic range DR_{global} (in dB) for a scene is determined by the maximum $E_{env,max}$ and minimum $E_{env,min}$ irradiance for each scenario and can be calculated by

$$DR_{global} = 20 \cdot log\left(\frac{E_{env,max}}{E_{env,min}}\right) \tag{2.16}$$

Local dynamic

An imager provides a spatial resolution which means that not only the global irradiance, but also the local irradiance is an important factor for determining the dynamic range of a scene. Assume a scene where a bright light source illuminates the object surfaces. Furthermore assume that parts of the scene are shadowed by objects which intersect the direct path of the major light source. Hence the image contains areas which are very dark and very bright. This determines the local dynamic of a scene. An example is a snapshot of a motor vehicle interior while entering or leaving a tunnel (see Fig. 2.10).

The assumption is that the bright parts of the image could not be brighter than a scene where the whole scene is illuminated to the maximum

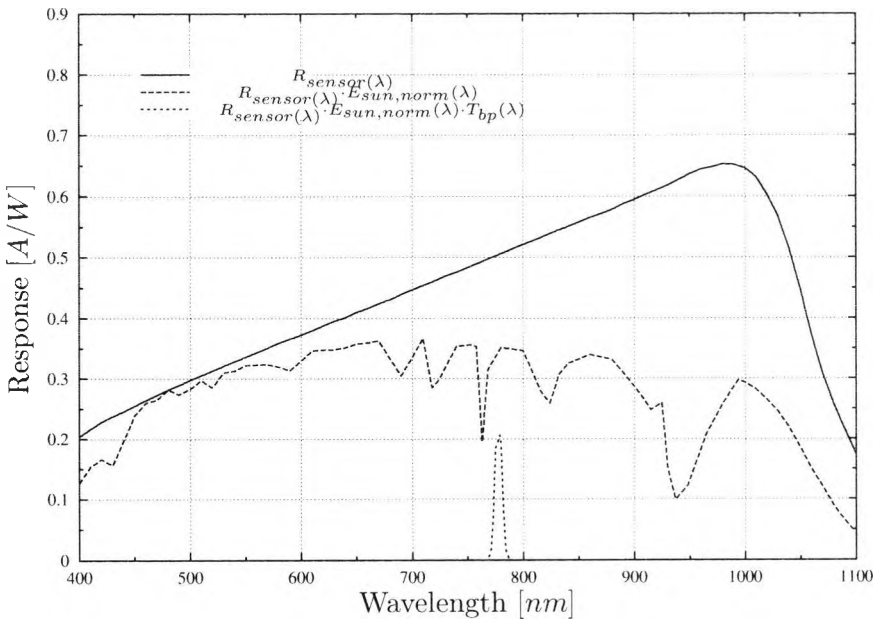


Figure 2.12: Spectral response $R_{sensor}(\lambda)$ of the power meter.

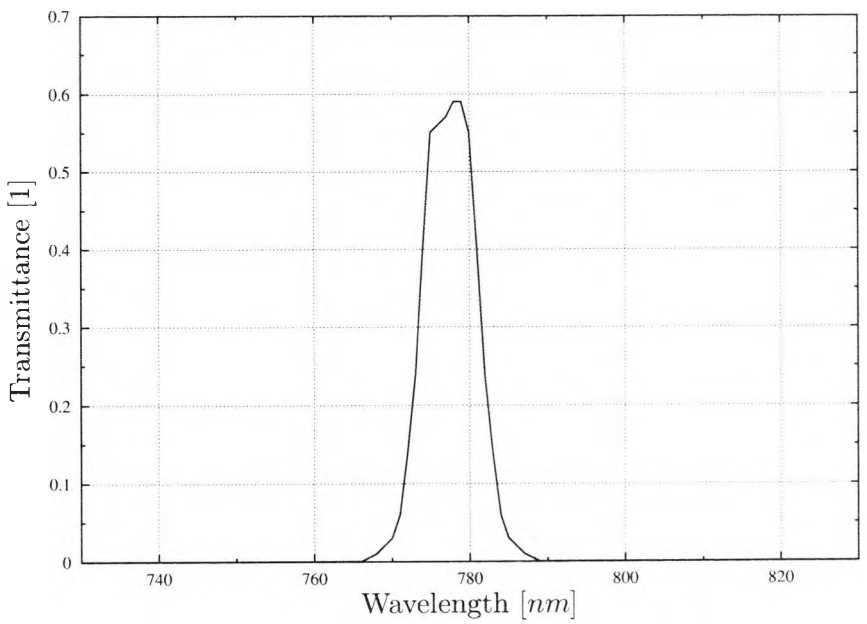


Figure 2.13: Transfer function $T_{bp}(\lambda)$ of a bandpass with CWL = 778.4nm and HBW = 10.7nm

extent by the major light source and *vice versa*. Therefore the maximum local dynamic should be comparable to the maximum global dynamic. This theory was proven by prior experiments of Park in [61]. The difference is that it is possible to adapt to global dynamic variations by changing the sensor aperture, gain, etc., but not to local dynamic variations.

2.4.2 Dynamic range in motor vehicles

After the derivation of sensor current I_{out}, irradiance E_{env} and dynamic range DR_{global} the measurements can be interpreted. The test drives were made in southern Germany during a sunny June and July in 1999. The measurements logged the irradiance within a sedan while driving and stationary in sunny and cloudy weather, with opened and closed sun roof, with different tilt angles of the sun, in the city and also on the autobahn to get data about typical illumination conditions:

- **Sunrise and sunset**: The irradiance increases and decreases slightly within the whole car. In addition the input angle of the sun into the vehicle interior changes (see Fig. 2.14).
- Sun from different sides of the vehicle: **Left and right** (see Fig. 2.18).
- **Tunnels** which yield fast but short global illumination changes (see Fig. 2.18).
- **Night drive** with short spots from head lamps of other cars implying short local illumination changes (see Fig. 2.16).
- **Crash test** for comparing this special illumination with the sun power. A crash test illumination measurement was necessary because an airbag safety system for cars will be tested by crash tests with special strong illumination for analyzing the crash results. If an airbag control system is confused by this special kind of illumination it is neither possible to test its ability nor to sell the car.

There are typically two kinds of illumination which are employed for vehicle crash tests: Halogen metal vapor lamps (HMI)[1] or pure halogen based lighting. 80% of all crash test facilities over the world use these HMI lamps emitting a spectrum which is emulating the characteristics of the sun. This means that the relative power of the source decreases significantly beyond the visible light range ($\lambda_{vis} \approx 380 \ldots 770nm$). HMI lamps were developed for the film industry to guarantee powerful but also 'natural' lighting for movies. The measurements were made within the BMW crash test facility in Aschheim near Munich. This facility employs HMI lamps.

The second popular illumination for crash tests consists only of a set of normal halogen flood lights. Common halogen light sources are characterized by increasing power within the infrared, hence crash facilities which

[1] **H**(quicksilver, chem. *Hg*)**M**(metal)**I**(halogen compound, chem. *Iodide*)

employ the above mentioned pure halogen illumination could produce interior irradiance within the NIR which is greater than the levels usually measured. However, the total radiant power of both light sources is significant but due to the high position of the lamps within the crash hall usually less irradiance shines directly into the car interior.

Experimental results

For interpreting the measurements the data gathered from various illumination conditions were divided into four separate illumination classes:

Night drive and stand between 11 p.m. and 1 a.m. If the car is driving or just activated it implies that the interior illumination (*i.e.* cockpit, indicator lamps) is enabled. The stand measurement was conducted without the interior illumination and represents the minimum irradiance within the vehicle interior on a cloudless night without moonlight.

Daylight includes all irradiance which was measured between noon and 7 p.m.

Crash test The illumination of a crash hall was measured for comparing this special light source with normal sunlight.

Maximum is defined as the total extremes of day and night scenarios, with and without interior illumination.

Fig. 2.14 ... 2.18 show some examples plotting the interior illumination as a function of time ($E_{env}(\lambda, t)$). The sample rate was 1 second in each case. During the measurements several maxima occurred caused by reflections within the interior. These reflections were usually caused by glittering parts of the interior and occurred for very brief moments. The mean measured irradiance depends on the sun, its position relative to the vehicle and is significant less than the measured maxima.

The route of Fig. 2.18 includes two tunnels. After the tunnels the vehicle was turned around and driven back. This yields at least four tunnel passages. This test series shows two interesting effects: The global irradiance hitting the sensor was approximated halved only by changing the position of the car with regard to the sun and the artificial illumination of the tunnels were not emitting radiation within the NIR.

The results for each scenario are shown in Table 2.4. The reference measurement without optical filter at daylight and crash test yields in general a maximum irradiance which was greater by a factor of 200. As mentioned before the illumination measurements were made in summer in Germany. Several places on earth obviously have an illumination situation which differs from the climate and sun power in middle Europe. Assuming a worst-case climate such as the Passat climax of the central Sahara, the maximum irradiance should be estimated as 2.4 times greater than in central Europe.

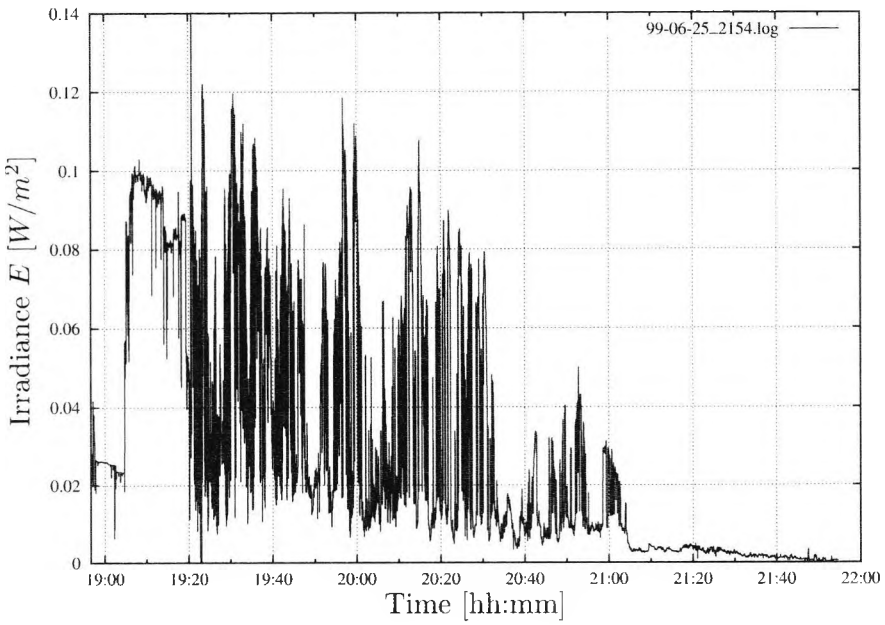


Figure 2.14: Autobahn drive ($T_{optic} = T_{bp}$). The car left the car park at 19:05h and entered the autobahn, which it left again at 21:05h.

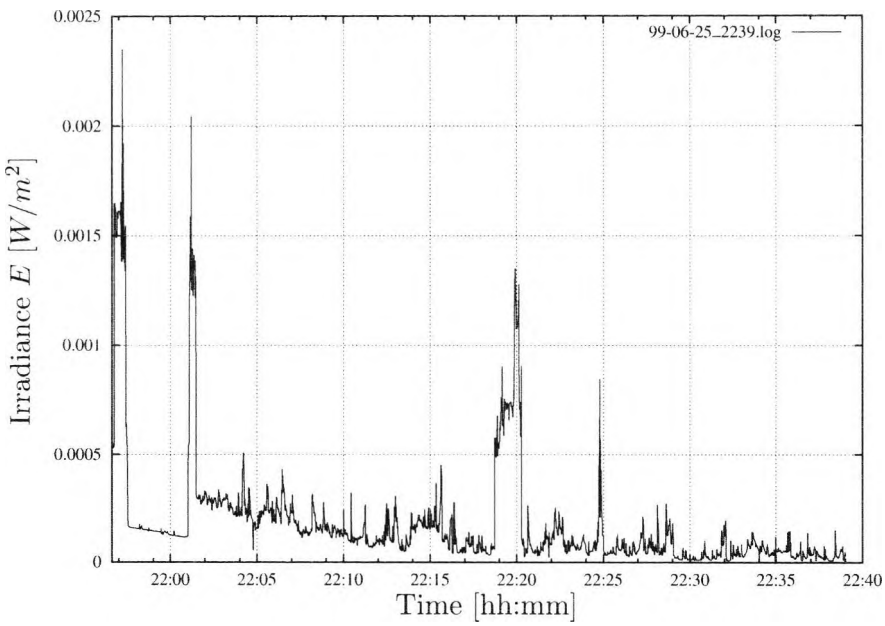


Figure 2.15: Night parking ($T_{optic} = T_{bp}$). The sequence starts with a door which was opened and active interior lighting. The peak at 22:18h was caused by the headlights of a closely following car.

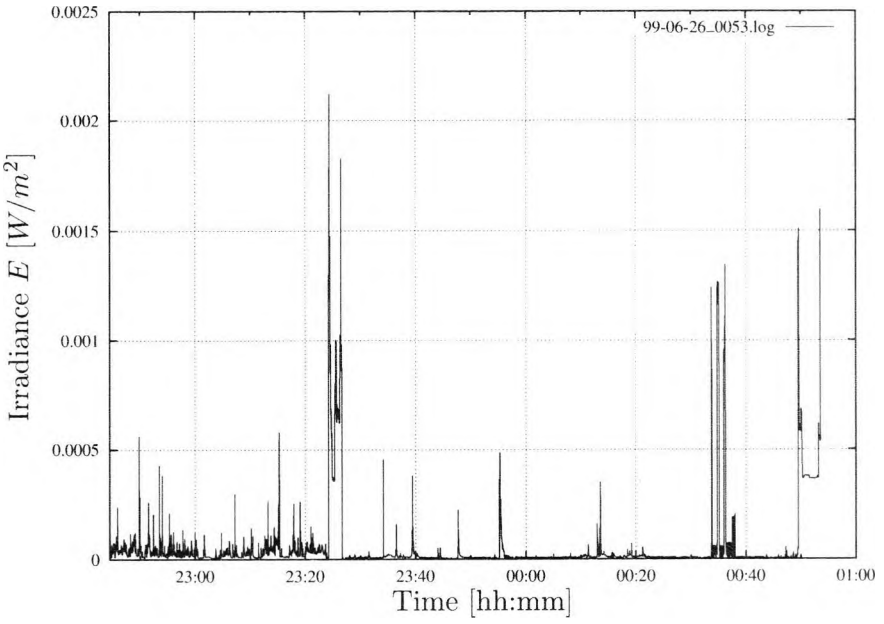


Figure 2.16: Night drive with instrumentation illumination ($T_{optic} = T_{bp}$).

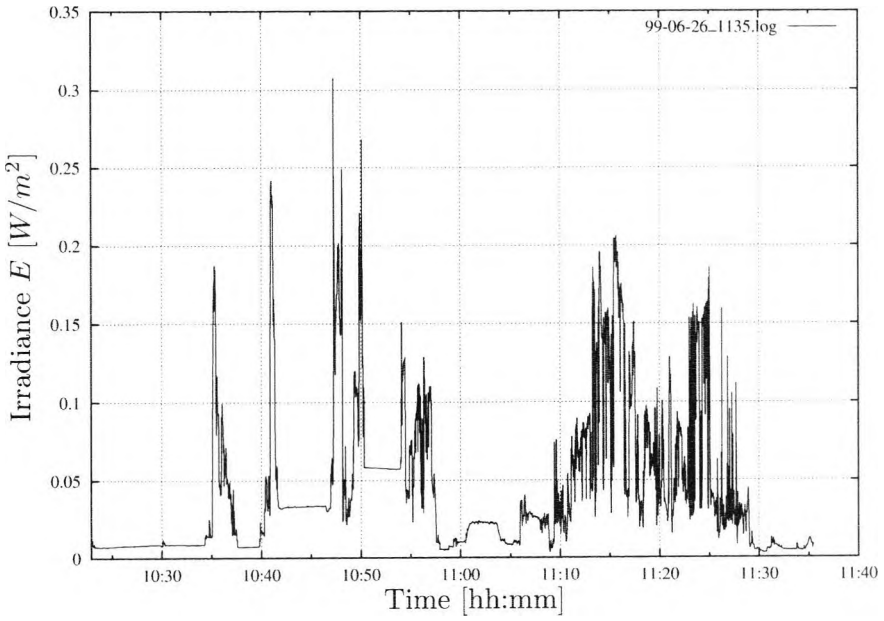


Figure 2.17: Different parking places ($T_{optic} = T_{bp}$).

Corresponding to our measurements this results in a maximum irradiance ($T_{optic} = T_{bp}$) of about $7.46 \cdot 10^{-1} W/m^2$ and thus a maximum dynamic of approximately $137dB$ assuming that the minimum irradiance is determined by the interior instrumentation illumination.

If a bandpass is applied to calculate the maximum outdoor irradiance within a given wavelength range, determined by the bandpass (see Fig. 2.13) according to Eqn. 2.17, then the remaining maximum irradiance can be estimated as $5.75 W/m^2$.

$$E_{sun,outdoor} = \int_{\lambda_{min}}^{\lambda_{max}} E_{sun}(\lambda) \cdot T_{bp}(\lambda)\, d\lambda \tag{2.17}$$

The remaining sun power inside a car is lower than outside because it does not hit the sensor directly due to the shielding of the vehicle chassis. Thus, the measured irradiance E_{env} represents according to Eqn. 2.9 the reflected light from the interior surface which is characterized by its reflectivity $\rho_{surface}$.

The highest irradiance $E_{env,max}$ was measured in the afternoon due to the lower sun position and its straight view into the interior. If a bandpass for the NIR is used ($T_{optic} = T_{bp}$), it can be estimated according to Table 2.4 at $311 mW/m^2$. The ratio between measured E_{env} and calculated $E_{sun,outdoor}$ is equal to an attenuation of 18.6 or $-25.4dB$. The maximum optical dynamic range at daylight can be estimated as $76dB$. A convertible can obviously achieve significantly greater values but not greater than $E_{sun,outdoor}$. However this work deals only with closed environments such as a car with roof but with large windows.

The smallest irradiance level $E_{env,min}$ which was detected at night without instrumentation illumination from the dash board was $802 nW/m^2$. The minimum irradiance employing the interior illumination was $10.5 \mu W/m^2$. This yields a maximum global dynamic for the front and rear interior of the vehicle of $129dB$, and $191dB$ respectively. In these illumination situations it is not possible to get images with suitable contrast for a reliable image processing without special techniques for night vision. The lower bound of illumination is in principle only limited by zero. Mainly the dimension of $E_{env,min}$ influences the optical dynamic DR_{global} and leads to results which are beyond present imager abilities.

Hence a supplementary illumination is necessary when the illumination level within the interior falls below a threshold irradiance if image processing should also handle situations with less external illumination such as night or tunnel drives, etc., without special camera techniques. Thus the upper limits are more interesting for calculating the features for a suitable sensor and a supplementary illumination. The maximum irradiance is estimated as 2.4 times greater than the measurements presented for regions with stronger sun light such as the central Sahara. Furthermore a crash illumination does

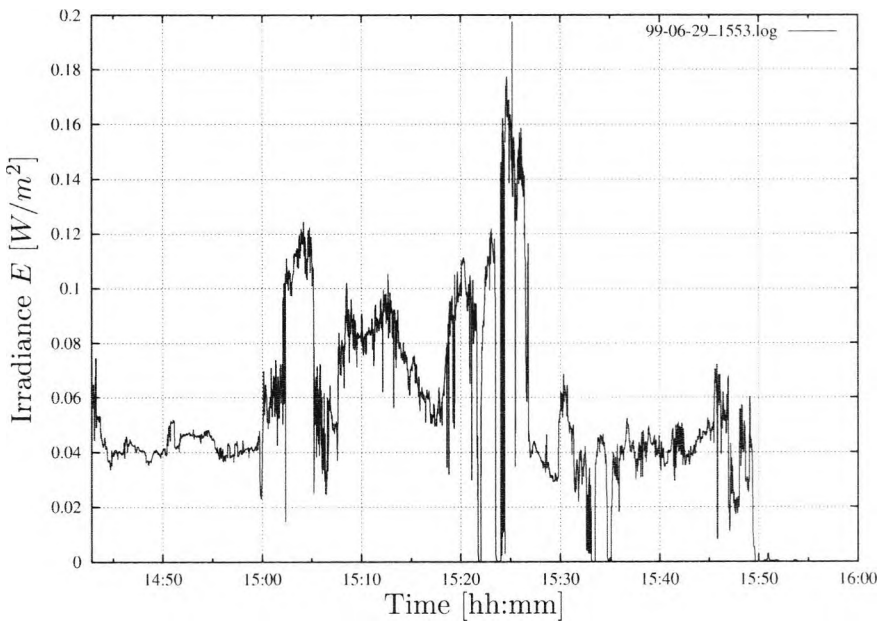


Figure 2.18: Drive through the city of Munich ($T_{optic} = T_{bp}$) with tunnels.

Scenario	$\mathbf{E_{env.min}[W/m^2]}$	$\mathbf{E_{env.max}[W/m^2]}$	$\mathbf{DR_{global}[dB]}$
$\boldsymbol{T_{optic} = T_{bp}}$			
Night, stand*	$8.02 \cdot 10^{-11}$	-	-
Night, drive	$1.05 \cdot 10^{-7}$	$2.12 \cdot 10^{-3}$	86
Daylight**	$4.92 \cdot 10^{-5}$	$3.11 \cdot 10^{-1}$	76
Maximum	$1.05 \cdot 10^{-7}$	$3.11 \cdot 10^{-1}$	129
Maximum*	$8.02 \cdot 10^{-11}$	$3.11 \cdot 10^{-1}$	192
Crash test	-	$9.50 \cdot 10^{-2}$	-
$\boldsymbol{T_{optic} = 1}$			
Daylight**	$1.45 \cdot 10^{-2}$	$7.23 \cdot 10^{1}$	74
Crash test	-	$1.82 \cdot 10^{1}$	-

Table 2.4: Measurements result for global dynamic (* minimum irradiance was measured without instrumentation illumination; ** time range 12:00-19:00h).

not represent an extreme situation with abnormal irradiance compared to powerful sunlight.

2.5 Precis

To determine the radiation level inside a motor vehicle (*e.g.* car) and thus the optical dynamic range ordinary and worst-case illumination situations have been measured. The radiometric measurements done in this chapter result in a maximum dynamic within a car interior of approximately $129dB$ which is outside the possibility of mainstream camera systems, see Chapter 3.

Advantages compared with the shortcomings of particular camera locations lead to the conclusion that the interior light module is the most favorable location ($cp2$) inside a motor vehicle to mount an imager for interior surveillance. Due to this finding most experiments and measurements in the following chapters have been made with the image sensor mounted close to $cp2$.

Another aim of this chapter was to determine the optimum illumination wavelength where the signal-to-noise ratio of sun and ambient illumination is a maximum in relation to the efficiency factor of image sensors. Due to the findings of Section 2.4 the focus on illumination measurements in this chapter was on wavelengths around the $770nm$ water gap which will also be used for active illumination experiments in Chapter 4.

Chapter 3

Image Acquisition

3.1 Motivation

Mainstream CCD based and most of the emerging CMOS based image sensors provide an optical dynamic range of $48 \ldots 60dB$. This dynamic range is sufficient for scenes with homogeneous illumination and without extreme contrasts, for example for multimedia applications, see Section 1.1.1. The investigations in Chapter 2 have shown that an optical high dynamic range environment (HDR) means special requirements for the image sensor. Established camera systems do not match these requirements, for example for monitoring a motor vehicle interior or open world scenes. Therefore a suitable imager must be adapted for such challenging illumination situations.

This chapter discusses the imager technologies available today and in the near future in Section 3.2 to 3.4 with regard to their usability for HDR scenes. After this general overview of state-of-the-art image sensor techniques a more detailed introduction to the field of imagers which are designed to handle HDR scenes follows in Section 3.5. Finally a custom imager for our active illumination experiments is presented in Section 3.6, designed to fulfill the requirements as defined in Chapter 2.

The majority of electronic cameras are nowadays employed in the field of multimedia, such as video, digital still photography and videophones (*e.g.* mobile phones with MMS[1] feature). The price of image sensors has decreased rapidly in line with the growing number of electronic cameras sold, thus opening new fields of application. In conjunction with the decreasing costs of processing power the field of potential digital image processing applications increases, too.

Today most electronic cameras use Charge Coupled Devices (CCD) based image sensors. However, the first imaging sensors created in the late 1960ies were MOS[2] based. These sensor rows included only few on-board electronic elements and a small number of pixels. Due to the former semiconductor technology the size of the CMOS[3] pixel was very large. The development of CCD based image sensors followed several years later. The advantage of CCD technology compared to MOS based imagers was the possibility of creating much smaller pixels with the same structure size.

The picture quality of CMOS cameras was inferior compared to the new CCDs, so they were displaced by the CCDs. However, due to the advancement of the CMOS technology for image sensors in the last decade new attention is paid to CMOS sensors. The pixel size decreased to dimensions which can match that of their CCD counterparts. In addition the picture quality of high end CMOS imagers was improved to a quality as good as in

[1] Multi Media Message

[2] Metal-Oxide Semiconductor

[3] Complementary Metal-Oxide Semiconductor, advanced version of the MOS technology

mainstream CCDs.

CMOS and CCD technology based imagers are both made using silicon. This creates similar sensitivity properties over the visible and near infrared (NIR) spectrum. Both technologies convert incident light (photons) into an electronic charge (electrons) by the same photoconversion process. In both technologies color sensors are realized in the same way, usually by coating each individual pixel with an optical band-pass filter for different wavelengths (*e.g.* blue $\approx 480nm$, green $\approx 550nm$, red $\approx 650nm$), organized in a mosaic pattern.

The most popular arrangement of color filters for imagers is the Bayer pattern [5], where half of the total number of the sensor pixels are covered with a green filter while a quarter of the total number is assigned to each red and blue. The final spatial image resolution of the separated color layers keeps unchanged by adaptively interpolating each color for each pixel from a pair of nearest neighbor pixels [67]. Due to the color filters covering each pixel the overall sensitivity of a color image sensor is lower than in a monochrome sensor and in fact is typically 3 times less sensitive. This is also true for the human eye, which switches to black and white mode in dark environments, too. Hence monochrome imagers are more suitable for low-light applications such as surveillance cameras.

3.2 CCD Sensors

Most of today's established imaging devices for multimedia applications such as camcorders are based on CCD (Charge Coupled Device) technology. This CCD technology is now about 30 years old. Its great advantages are the decades of design experience and a stable production process, yielding excellent image quality and low noise. Another advantage compared to CMOS based imagers is the high optical sensitivity due to the amorphous silicon employed for realizing the photoconversion process.

The surface of a CCD image sensor represents a compact mesh of electrodes which are separated into isolated rows. The gathered brightness information from a pixel (and therefore the image intensity) is determined by a read-out process after a determined global integration (exposure) time. The read-out process consists of physically routing the individual accumulated packets of electrons from the position where incident photons have been detected across the electrode mesh towards a read-out register. To get the final digital brightness information this read-out register will be shifted to an analog-to-digital converter (ADC). The block diagram of this shift operation is shown in Fig. 3.1.

These serial shift operations synchronized by a read-out clock mean that

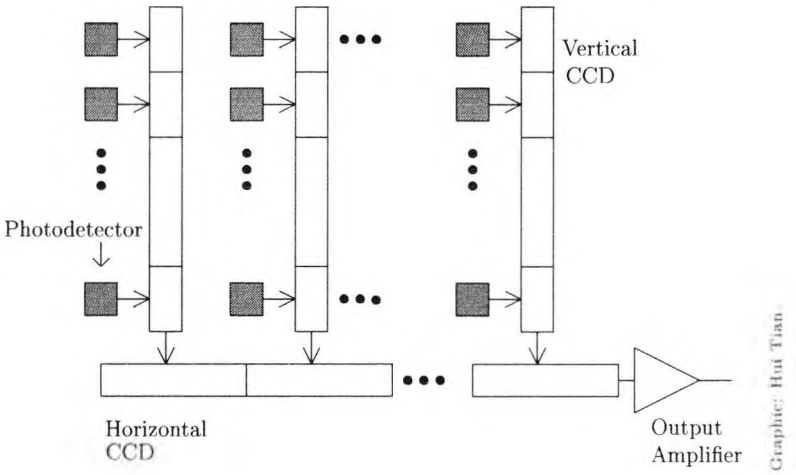


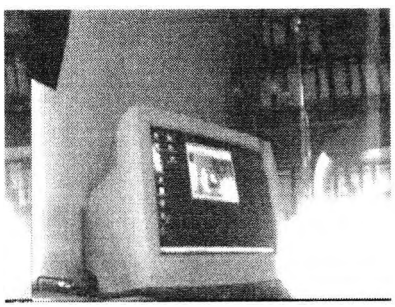
Figure 3.1: Block diagram of CCD principle for routing electron charges towards the read-out register across the sensor surface.

only the whole picture can be read out. The quality of this read-out clock (amplitude and shape) is critical for proper routing operations and significantly influences the final image quality. Therefore it is generated by a specialized clock driver chip (off-chip clock).

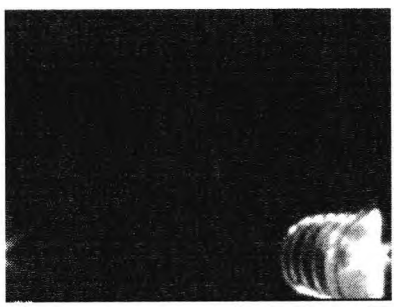
It is possible but not economically sensible to implement supplementary camera functions on the CCD sensor chip which consists of amorphous silicon, such as the read-out clock, timing logic and subsequent signal processing (see Section 3.3.3). These functions are therefore usually implemented in a couple of secondary chips. They are usually realized in standard CMOS technology and consist of crystalline silicon. This incompatibility of the CCD sensor with standard electronics leads to relatively high production costs for the final imager.

CCD based imagers suffer from three main technology-dependent shortcomings: Blooming, smearing and high power consumption. Blooming and smearing effects occur especially while capturing very bright scenes with high contrast [6]. The high power consumption is caused by the need of multiple non-standard supply voltages for the sensor, special clock device and supplementary electronics.

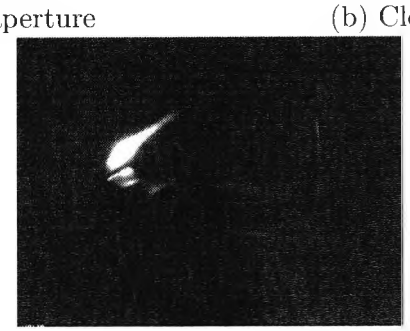
A standard CCD imager requires up to 5 or 6 different supply voltages inside. CCDs need large clock swings up to $15V$ to achieve acceptable charge transfer efficiencies. Mainstream CCD based image sensors require up to $2-5W$ while power optimized CMOS based imagers have a power consumption less than $50mW$ at the sensor level for the same pixel clock [19]. This opens up applications for mobile devices and autonomous vehicles which



(a) Opened aperture



(b) Closed aperture



(c) Fast movement of a bright object (lighter)

Figure 3.2: Blooming ((a),(b)) and smearing effects ((c)) due to satisfied pixel and uncomplete electron charge routing during read-out process.

were impossible or very difficult to realize with the high power demand of CCDs.

An example: A car anti-theft system has to work at an extremely low power consumption because the battery is not charged while the engine is turned off and would discharge quickly if too much current were drawn during stand-by. The anti-theft system has to protect the car, but it must also allow it to start the engine after several weeks of standing.

The smearing effect occurs due to delayed and therefore incomplete routing of electron charges from one pixel to another during the read-out process. Electrons which have not been transferred properly will be read-out at the next frame. This effect causes a bright tail in the image, for example if a fast bright lighting source runs through a relatively dark picture as shown in Fig. 3.2(c).

Blooming occurs due to saturated pixels (over-exposed) when the accumulated electrons from one pixel flow over to its neighbors without proper control. The worst-case is a totally bloomed picture caused by only a few saturated pixels as shown in Fig. 3.2(a) and 3.12(a). Operating temperatures above 70^oC without cooling result in bloomed images, too. Special techniques such as anti-blooming-gates have been developed to minimize smearing and blooming effects, but they can not eliminate them completely.

Smearing and the high power consumption, especially the low dynamic range and the low operating temperature range determined by the dark current sensitivity, are further factors besides blooming that had made it uneconomic in the past to implement a CCD based imaging sensor for automotive purposes.

3.3 CMOS Sensors

The principle for sensing light is almost the same for CMOS or CCD based image sensors, however the read-out philosophy is completely different. Each gathered charge packet is not transferred from one pixel to another, but instead they are processed as early as possible by charge sensing amplifiers consisting of CMOS transistors.

The first CMOS image sensors implemented just amplifiers at the top of each pixel column with the pixels themselves containing just one transistor used as a charge gate for switching the contents of the pixels to the charge amplifiers. The behavior of this CMOS pixel design is similar to analog DRAMS called *passive pixel* sensor (PPS). If both the photodetector and read-out amplifiers are implemented in each pixel they are called *active* pixel CMOS sensors (APS).

Active pixel CMOS sensors usually contain at least 3 transistors per pixel which allows the integrated charge to be converted into a voltage within the pixel, see Fig. 3.4. This voltage can then be read-out over a column bus instead of using a charge shift register as employed by CCDs [28]. Hence CCD effects such as blooming and smearing due to leak charge during charge shifting or register overflow are non-existent for CMOS based image sensors.

The ability to address separate pixels allows window/region of interest read-out (windowing, ROI). This windowing provides more flexibility for adapting the sensor to the application, for example by realizing an on chip pan/tilt control or electronic zoom. Furthermore an APS design can clock the column bus at much greater rates than passive pixel sensors and CCDs, yielding higher pixel clocks and therefore increased frame rates per second (fps).

An active pixel design yields usually less noise but on the other hand lower packing density (fill factor) than passive pixel designs (see Section 3.3.4). Both types of CMOS image sensors consist of crystalline silicon (contrary to amorphous CCDs) and can therefore be manufactured on standard CMOS foundries (fabs) all over the world by several semiconductor fabricators.

The major shortcoming of CMOS based image sensors is the problem of matching the multiple different amplifiers within each pixel which results in increasing noise and reduced image quality. This lower image quality caused by amplifier noise is the reason why CMOS based image sensors have not reached the same production volume as CCDs yet. Several approaches have

been recently presented to reduce the residual level of fixed-pattern noise to insignificant proportions, such as correlated double sampling (CDS, see Section 3.3.2), but CMOS imagers have not yet reached the image quality of CCDs after decades of optimization.

The more complex pixel structure of APS sensors yields increased noise, but it is also the major advantage of CMOS based imagers due to the ability to create 'smart' pixels. That is important for designing cameras with extended dynamic range which will be discussed in more detail in Section 3.5. Furthermore due to the fast prototyping possibilities of CMOS technology CMOS based imagers are more flexible by designing various pixel geometries [9].

Another major benefit of CMOS based imagers compared to CCDs is the high level of possible product integration that can be realized by implementing several supplementary electronic camera functions or even signal processing on the same chip [62, 21]. Due to the compatibility with standard CMOS technology it is possible to integrate

- Timing logic
- Auto exposure control
- Auto gain control (AGC)
- A/D conversion
- Anti-jitter (image stabilization) logic
- Motion detection
- Image compression algorithms
- Color encoding
- Interface circuits

on the same chip with the sensor. The results are complete one-chip cameras, which are not feasible in CCD technology. This is a striking example of progressing fusion of hard- and software for improved image processing because more and more results from image analyzing take control of the next image acquisition step.

An example is presented in Fig. 3.3 showing the LM9618 from National Semiconductor [56], which we used for most of our experiments, see Section 3.6 for details. The pixel array and supplementary logic blocks, such as timing unit, ADC and interface drivers, are completely integrated on the same chip. Only a power supply and an external clock are necessary to operate the imager.

Other advantages are the significantly reduced power consumption compared to CCDs due to the need for one single power supply voltage for the whole imager. Typically CMOS based cameras consume one fourth of the power of equivalent CCD cameras at the sensor level for an equivalent pixel clock [19] and provide a larger operating temperature range [76]. Both parameters are important in case of an implementation for mobile devices or automotive applications. Finally the price of CMOS imagers is considerably

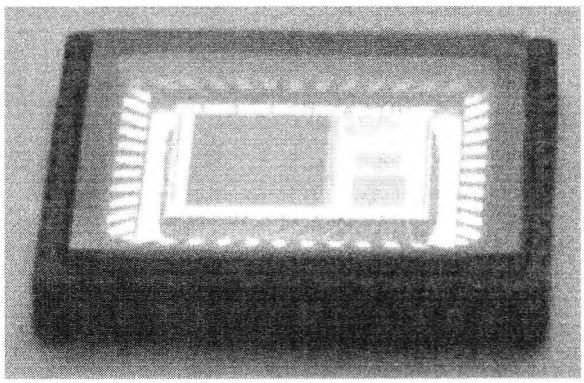


Figure 3.3: Example of an one-chip-camera: LM9618 from National Semiconductor. The large pixel array (1/3") and smaller supplementary logic blocks such as timing unit, ADC and interface drivers are visible which are completely integrated on the same chip.

lower than their CCD counterparts because of the availability of standard CMOS fabs all over the world and the ability to integrate necessary peripherals for a complete camera (sensor, timing logic, ADC, interface logic) which has substantial cost advantages over CCDs.

The revival of CMOS based imagers in recent years in combination with the decreasing prices of processing units brings machine vision applications for intelligent vehicles closer. An example for an application of an APS sensor for automotive purposes is discussed in Chapter 6.

Inside the photodiode of each pixel, the light incident on the image sensor generates a photo current proportional to the occurring irradiance. This current is converted into a voltage signal in the imaging sensor as discussed in Section 3.2. This can be performed by integrating the current on a capacitance or measuring the voltage drop of a photo current flowing through a resistor. Figure 3.4 shows an overview of the different read-out procedures, which can be divided into time continuous and time discrete read-out techniques.

3.3.1 Time continuous read-out

With time continuous techniques the voltage drop over a load resistance is measured. If a linear resistor is used the output voltage is proportional to the occurring irradiance. As an alternative it is possible to employ a non-linear resistor, usually with logarithmic response. Linear time continuous imagers are not popular due to the large size of linear resistors in CMOS technique, which lead to inacceptable pixel dimensions and less fill factor as discussed in more detail in Section 3.3.4.

The logarithmic imagers provide response characteristics which differ significantly from mainstream cameras. An example image is shown in Fig. 3.4(b). As the name implies there is a logarithmic relationship between the input signal (incident photons) and the output of the sensor. The incident light produces a photo current that is proportional to the irradiance. The photo current is converted to an output voltage using a resistor with an exponential voltage-to-current characteristic. This non-linear resistor can be realized using a MOS transistor. Here the output voltage is equal to the logarithm of the incident irradiance. This compression enables a much greater input dynamic compared to conventional imagers with linear response. Dynamic ranges above $120dB$ can be achieved by this technique. However, the photocurrent is very small $(10 \ldots 100fA)$ for continuous read-out, which means that the read-out circuitry must be extremely precise to make use of the dynamic range created by the compressed output voltage.

Since the photo current is not integrated it is theoretically possible to instantaneously read-out the intensity at any moment in time after the swing time. When the light intensity suddenly changes the pixel voltage changes towards the new equilibrium with a time constant RC, where R is the small signal impedance of the series resistor and C is the pixel capacitance. Hence the swing time of a logarithmic pixel is inversely proportional to the irradiance which means that the pixel changes faster towards the new equilibrium at increasing irradiance [28]. In reality about $1ms$ is required for room light level scenes for the swing time which means that virtually permanent scanning is not practical.

Fluctuations of the threshold voltages inside the transistors and amplifiers from pixel to pixel cause signal independent picture errors. These errors produce a fixed pattern which is overlaid to the output picture, hence this kind of noise is called Fixed Pattern Noise (FPN). By illuminating the sensor with a homogeneous light source the FPN can be measured and saved. For eliminating the disturbing FPN offset a memory with the size of an image is necessary and a simple calculation unit which subtracts the reference frame from subsequent images. Furthermore it is necessary for high quality images to compensate for multiplication errors. Hereby additional snaps will be required to interpolate the characteristic of the pixel.

The logarithmic read-out effect can be employed for advanced motion detection (see Section 5.2) within a scene with low illumination levels: Two images captured one after another which show a moving object result in different object reflection factors ρ_{s1} and ρ_{s2}. Fluctuations of the ambient illumination E_{amb} between both images are considered negligible. Therefore, according to Eqn. 2.9, the grey level difference ΔGV_{lin} between both images for an image sensor with linear response is determined as

$$\Delta GV_{lin} = \rho_{s1} E_{amb} - \rho_{s2} E_{amb} \tag{3.1}$$

$$\begin{aligned} &= E_{amb}(\rho_{s1} - \rho_{s2}) \\ &\propto E_{amb} \end{aligned} \tag{3.2}$$

and is therefore proportional to the ambient illumination. This means small differences at low light conditions and therefore it is more difficult to detect changes. The frame difference ΔGV_{log} from a sensor with logarithmic response is determined as

$$\Delta GV_{log} = \log(\rho_{s1} E_{amb}) - \log(\rho_{s2} E_{amb}) \tag{3.3}$$

$$= \log\left(\frac{\rho_{s1}}{\rho_{s2}}\right) \tag{3.4}$$

and depends therefore *not* on the amount of ambient illumination, which is useful at low light levels. However, this advantage for motion detection is reduced by the worse picture quality of logarithmic imagers which can not reach the grade of sensors with linear response. Due to this fact the use of logarithmic imagers is limited to few areas outside multimedia applications, such as visual welding control.

3.3.2 Time discrete read-out

The time continuous read-out technique causes problems with less contrast and the high Fixed Pattern Noise (FPN). To achieve images with more contrast from a CMOS camera the time discrete read-out technique is used.

The basic architectures of a time discrete CMOS imager and a time continuous one are very similar. Both use a digital logic to control the row and column addressing and amplifiers in each row. The difference is that the time discrete read-out techniques integrate the photo current Q on a capacitance controlled by an exposure logic. The straightforward case is the employment of the diode capacitance C_{pix} of the pixel. The capacitance has to be discharged at the beginning of the time discrete read-out phase. After the determined exposure time the signal amplitude is read as a voltage over the pixel capacitor:

$$V = \frac{Q}{C_{pix}} \tag{3.5}$$

Similar to time continuous imagers, CMOS imagers with time discrete read-out also tend to feature high noise. An established technique to improve the S/N ratio of CMOS image sensors with time-discrete read-out is to employ Correlated Double Sampling (CDS). For realizing CDS the pixel is read-out directly after reset of the capacitor for determining the 'dark' or 'reset' voltage level $V_r = Q_r/C_{pix}$. This sample is compared with the voltage signal after exposure $V_s = Q_s/C_{pix}$.

$$V = V_s - V_r = \frac{Q_s - Q_r}{C_{pix}} \tag{3.6}$$

Spatial and temporal noise that is common to V_r and V_s disappears from the result [68], such as

- fixed pattern noise
- spatial noise
- noise on the photo diode capacitance on the condition that this capacitance is not reset in between the two samples

but certain kinds of noise will not:

- effects due to gain non-uniformity or non-linearity
- uncorrelated temporary white noise that originated before the differencing operation, *e.g.* broadband amplifier noise is multiplied by a factor x by the differencing operation
- all downstream noise sources (downstream = after the differencing operation) such as system noise, discretisation noise
- low frequency MOSFET noise (1/f noise, flicker noise) is only partially reduced by a factor that is the logarithm of the associated reduction in bandwidth, typically a factor not more than 1...3
- signal noise, as optical shot noise is in principle not affected by CDS
- fixed pattern noise due to dark current

CDS and FPN correction (both off-chip) might be combined for improved image quality. The image read-out from the sensor is typically post-processed anyway for bad pixel correction, RGB processing [5] or compression.

The processing power and memory for CDS and FPN correction can be shared with these operations. In order to subtract the signal and its reset level the pixel must have at least the possibility to buffer its previous value, or off-chip correction will be employed by using an ADC and digital memory.

The different image characteristics depending on discrete or time continuous read-out are depicted in Fig. 3.5(a) and (b). The sensor with logarithmic response in Fig. 3.5(b) shows a low contrast picture, but it covers a very high dynamic range. In this picture both details from outside the building and the laboratory are visible. Contrary to a logarithmic response the image captured by the sensor with linear response in Fig. 3.5(a) shows great contrast but without the whole dynamic of the scene. Bright regions of the scene are overexposed. Therefore the sensor is saturated and provides no reliable intensity information for about 40 % of the image. That means that the optical dynamic of the scene exceeds the dynamic of the image sensor. Techniques for solving this problem are discussed in Section 3.5.

3.3.3 Exposure modes

CCD and CMOS image sensors with time discrete read-out as well as the traditional optochemical cameras depend on a property that is called 'shut-

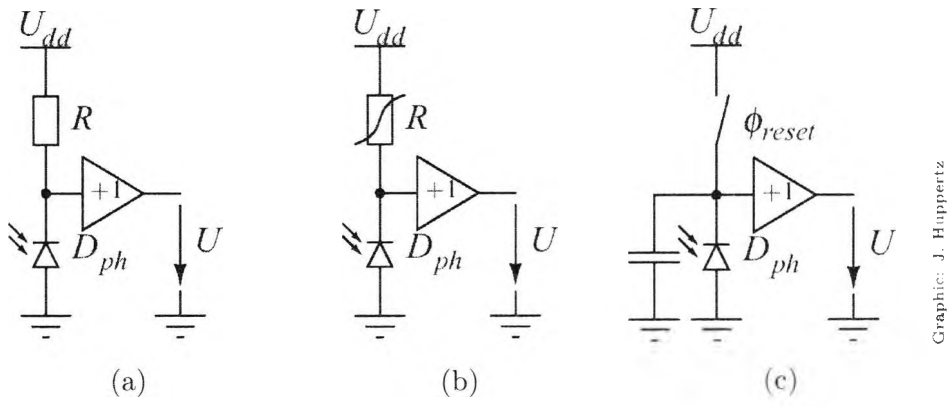


Figure 3.4: CMOS read-out techniques: (a) linear and time continuous; (b) non-linear and time continuous; (c) linear and time discrete

ter time' which is equal to the terms 'exposure' or 'integration time'.

There are two subdivisions to control the exposure process for time discrete read-out sensors: global shutters and the local shutters [87, 61]. As image processing filters a global shutter determines a particular exposure (integration) time for all pixels within the read-out zone, whereas a local shutter uses several integration times, which differ from pixel to pixel (autoadaptive pixel). These two approaches are themselves subdivided again into rolling and synchronous shutter. A synchronous shutter starts and stops exposure for every pixel at the same time. After that every pixel is read-out subsequently or simultaneously depending on the number of read-out units and ADCs. To save time during exposure and to enable a faster frame rate the rolling shutter approach is used. This means the shutter launches at a line of the picture and 'rolls' continuously from sensor line to line, starting the exposure process. Every pixel line starts exposure at a different time. Depending on the 'roll' speed the shutter returns to the first exposed pixel, reads it out and starts exposure again. This technique is shown in Fig. 4.24 on page 103.

Regarding the production volume nearly all produced CMOS based camera systems employ a global rolling shutter to enable higher frame rates. This is sufficient for most multimedia applications. However, if fast moving objects have to be observed or measured the rolling shutter technique tends to fail if the frame rate is small compared to the speed of the moving object within the scene. This causes stretched and distorted object shapes in the captured images. This effect is shown in Fig. 3.6 where a fan running at approx. 1600 rpm is captured by an imager with synchronous read-out (a) and rolling read-out (b).

A rolling shutter is also unsuitable if a flash lighting system is used to

(a) Linear read-out (b) Logarithmic read-out

Figure 3.5: Example images for different read-out modes.

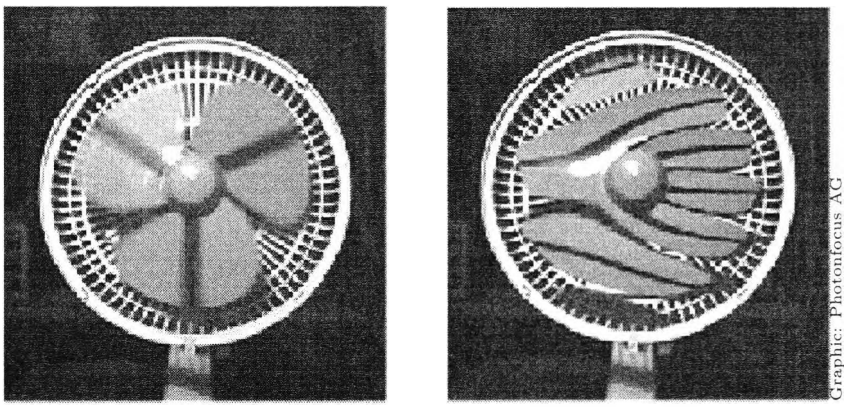
illuminate the scene. Since one frame changes to another seamlessly an accurate distinction of each frame for flash lighting synchronization is not possible. This means that for flashing one complete frame it is necessary to enable the flash light for at least two frames or to disrupt the rolling shutter frequently. This is called snapshot mode at most cameras for performing still images with video cameras in conjunction with a flash light [56].

However, for applications which require accurate distinction between subsequent frames, for example for active illumination approaches as discussed in Section 4.2, a local and synchronous shutter is desired.

Many applications, *e.g.* in automotive or machine vision, need only a small portion of the whole image to be read out. CMOS cameras usually have the ability of multiple subregion read-out. In full frame mode the regions of interest can be set. After this the smaller sub-region can be captured at a higher frame rate that is only limited by the maximum pixel rate. A common pixel rate of a state-of-the-art imager is $10MHz$. Hence 1,000 frames could be captured theoretically per second at a frame size of 100^2 pixels. This theoretical frame rate is not feasible due to time offsets caused by the timing logic and interface circuits. Modern commercial high speed CMOS cameras achieve up to 500 frames per second (fps). Research prototypes reach up to 10,000 fps [34]. This ability is very interesting for fast object tracking.

3.3.4 Fill factor

A major shortcoming of APS, and to a lesser extent of PPS, is that a significant part of the pixel surface is occupied by read-out circuits which are



(a) Synchronous shutter (b) Rolling Shutter

Figure 3.6: Example images for synchronous and rolling shutter read-out techniques. The fan is rotating at approx. 1600 rpm.

not part of the photoactive region (photo diode). Light incident on these not photoactive regions is collected by the junctions of neighboring read-out circuits, resulting in increased noise. This essentially is the reason for the low fill factor of active pixel sensors [15, 53] with its increased number of in-pixel logic. Therefore the fill factor is an important item to characterize different CMOS based imagers. The photoactive region is determined by the effective optical fill factor FF of the device that describes the portion of the pixel area A_{pix} which contributes to photosensitivity [49, 51]:

$$FF = \frac{A_{eff}}{A_{pix}} \tag{3.7}$$

Modern commercial CMOS based image sensors provide a wide spread range of FF from 30 to 60 %.

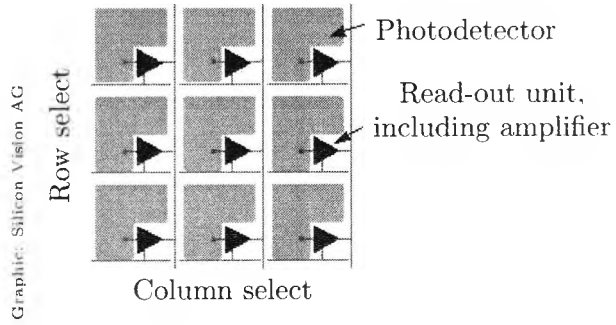


Figure 3.7: Array of active pixels of a CMOS imager

An established approach to minimize the fill factor problem is to cover the sensor array with micro lenses for each pixel. Micro lenses increase the sensor sensitivity by guiding the light that would otherwise hit the peripheral pixel circuits to the photoactive area. To minimize the influence of fluctuating incident angles of incoming light rays each micro lens has to be adapted to its spatial position on the imager surface.

3.4 TFA Sensors

As discussed in the subsequent section the photosensitive area of CMOS based image sensor pixel is limited because the pixel circuit and the detector share the same chip area (see Section 3.3.4), which leads to a reduced fill factor. A novel approach to overcome the fill factor problem of APS and to merge the advantages of the CMOS and CCD principle while avoiding their disadvantages is the TFA sensor (Thin Film on ASIC) introduced by Fischer in [18].

This hybrid imager consists of an amorphous silicon (a-Si:H) based optical detector on top of a common CMOS ASIC[4]. The ASIC performs the signal read-out or signal processing for each individual pixel and is covered by a crystalline silicon (c-Si:H) optical detector layer in a chemical vapor coating process.

In contrast to common CMOS imagers a TFA sensor is vertically integrated due to the location of the read-out units below instead of besides the optically active area of every pixel. This provides a fill factor up to 100% for the detector and the read-out circuit below. Therefore the number of employed transistors can be infinitely increased without affecting the overall imager sensitivity. Fig. 3.9 illustrates the layer sequence of a TFA sensor. The crystalline ASIC typically consists of identical pixel circuitry underneath each pixel detector and a peripheral circuit at the edge of the light sensitive area. An insulation layer separates the optical multi-layer thin film system from the ASIC chip. The thin film system is embedded between a metal rear electrode, which is usually the third metal layer of the ASIC, and a transparent front electrode.

The TFA technique makes it possible to design smart pixels consisting of more that 16 transistors for realizing more complex pixel designs. This advantage is used by Silicon Vision AG, Germany, for designing an autoadaptive pixel imager which provides a very high global dynamic range by adapting the integration time for each individual pixel according to the local illumination intensity.

Unlike mainstream linear CMOS based image sensors the integration time control takes place in the pixel itself in real time. Hence off-chip circuitry and additional computation time for the illumination adaptation are

[4] Application Specific Integrated Circuit

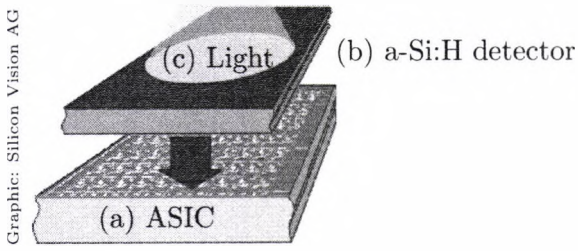


Figure 3.8: TFA principle

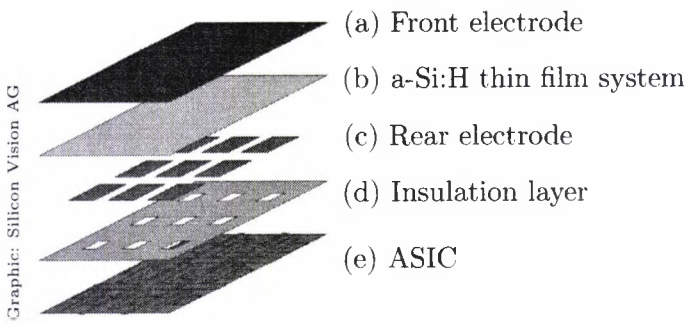


Figure 3.9: TFA layer system. Layer (a) to (c) build the optical detector

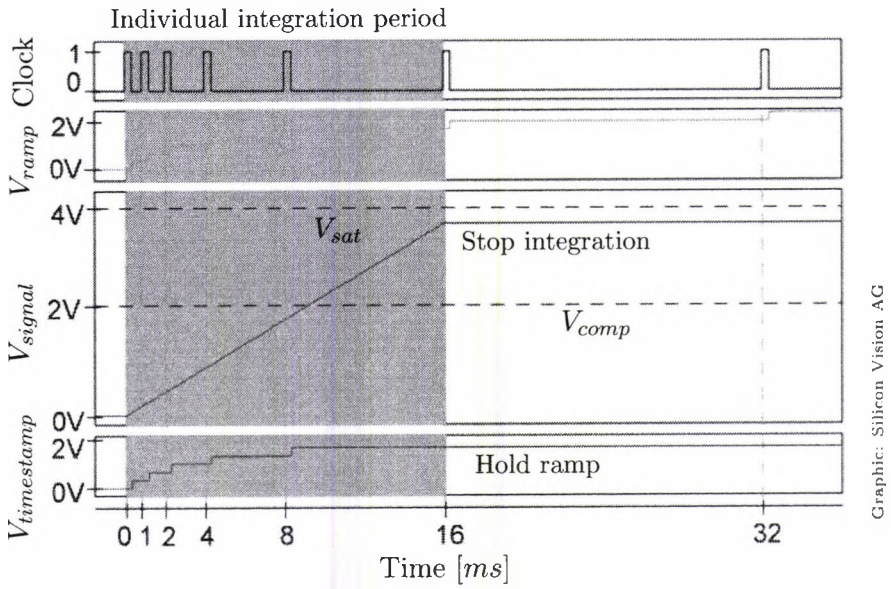


Figure 3.10: Autoadaptive timing

not required. A sudden change to a high illumination intensity is detected immediately within each pixel separately, so the integration of the photocurrent can be stopped before the integration capacitor is saturated. This yields an optimized local contrast.

Fig. 3.10 shows the timing diagram for such a locally *autoadaptive* pixel. The current of each photodiode is integrated according to Eqn. 3.5 on the integration capacitance into a signal voltage V_{signal}, similar to an APS with time discrete read-out. On every rising edge of the imager clock this voltage is compared to a reference voltage V_{comp} which is slightly below half the saturation value of V_{signal}. If the integrated signal is still below the threshold the integration continues whereas the comparator terminates the integration if the signal exceeds the reference level. With every clock a time stamp signal V_{ramp} is incremented by one step and is sampled and held in the time stamp capacitance at the moment the integration is terminated. At the end of the integration phase the information of each pixel consists of two voltages which are read out: the integrated signal and the time stamp which defines the integration duration over which the pixel signal has been integrated.

The binary increase of the integration time steps corresponds to $V_{comp} < 0.5V_{sat}$. Therefore it is ensured that the integration capacitance is not saturated in the following step and the range for the signal voltage at the end of the integration time is $0.5V_{sat} < V_{signal} < V_{sat}$. Finally the original photocurrent is reconstructed from the two signals according. In principle the global dynamic of the imager is only limited by the photodiode characteristics and the maximum number of time stamps.

The major benefit of the TFA technique compared to APS is that the amount of pixel logics is not limited and does not affect the fill factor. This allows the design of smart *pixels*, for example for extending the dynamic range or basic signal processing. Furthermore the photosensitive layer consists of amorphous silicon which features high photo sensitivity similar to CCD based imagers. The major shortcoming compared to basic CMOS image sensors is the elaborate production process for linking the amorphous and crystalline silicon layers as shown in Fig. 3.9, which is not standard. Today there is no commercial high volume product in the market which employs a TFA design due to little manufacturing experience and high production cost.

3.5 High dynamic range cameras

Numerous imaging applications require a dynamic range of more than $60 \ldots 70dB$ due to non-uniform illumination or reflection on the objects. If the light conditions are not controllable a very high dynamic range of the camera is necessary to take all details into account that could appear within the scene. This situation occurs, *e.g.*, on sunny days in open world scenes, a standard problem for machine vision in motor vehicles and outdoor surveillance. Example images of high dynamic scenes are shown in Fig. 3.11 and 3.12.

In these scenes extreme differences in irradiance occur between regions with bright (sun light) and shadowed regions. Usually it is possible to adjust common imagers either to the very bright areas or to the dark areas in the scene by adapting several imager parameters, *e.g.* exposure time or lens aperture or by the use of optical filters. However, if this entire range is covered throughout a single frame global sensitivity control is ineffective since saturation as well as signals below the noise level may occur simultaneously. Hence in extreme dynamic environments with extreme contrast CCD or even CMOS based imager systems which feature only a single exposure time are not able to cover the whole brightness range of the scene because parts of the image may be under- or overexposed. This results in loss of image detail.

The classic scene for demonstrating a camera dynamic is the direct view into a light bulb as shown in Fig. 3.11. It shows an operating 100W light bulb, whereas both the light bulb and the filament are recognizable. Even the characters on top of the bulb and the bust of Goethe in the background are visible.

There are various ways to realize a camera for HDR environments by extending the limited dynamic range of an imager. Present available HDR cameras usually consist of a CMOS based imager. But the high dynamic optical behavior of available CMOS cameras is not based on the CMOS technology itself. The range for a linear pixel signal in an CMOS based image sensor is limited to less than 80dB due to the ASIC noise floor. On the other hand pixels consisting of crystalline silicon (CMOS) are much less sensitive than comparable amorphous pixel (CCD), but CMOS technology provides the pixel designer with the possibility of reading out single pixel charges and of processing them directly besides or beneath the optical surface, see Section 3.3 and 3.4 for details. Therefore better control of the read-out process of each pixel is provided, creating the possibility of increasing the local contrast and hence the dynamic range of the sensor.

For realizing an extended dynamic range for image sensors there are two basic techniques. Either the sensor employs non-linear imaging elements

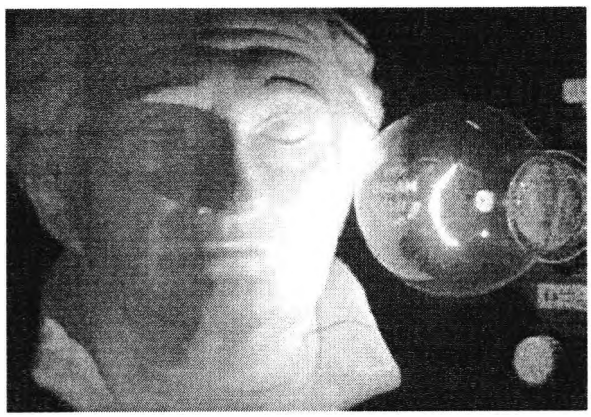


Figure 3.11: Example of a HDR scene showing a direct view of an operating light bulb.

(see Section 3.3.1) or it assembles a HDR image from a set of frames of a linear imager with multiple integration times [80]. So to achieve an extended dynamic range for image sensors with a given linear signal range below 80dB the photoresponse has to be compressed (non-linear read-out) or to be split (linear read-out with different exposure times).

3.5.1 Non-linear response

The first generation of commercial PPS based HDR cameras used intensity compression by implementing a logarithmic voltage-current response of MOS transistors as described in Section 3.3.1. Due to this compression a PPS based camera with logarithmic response can cover up to 120dB of intensities into a voltage range of a few hundred millivolts.

However, images taken from cameras with logarithmic response suffer from reduced intensity contrast at low light levels as shown in Fig. 3.5(b). This means a challenge for machine vision systems resulting mostly in reduced robustness of the applied signal processing filters and inaccurate classification results. The possibility of decompressing (*i.e.* de-logarithmize) the intensity information to a simulated linear response is not economic due to increased sensor noise at low intensity levels, see Fig. 3.5(b) and due to fluctuations from an ideal logarithmic response depending on the operating point of the non-linear MOS transistors.

3.5.2 Linear response

In contrast to sensors with time continuous read-out and logarithmic response the operating point for linear imagers is determined by the individual frame or even pixel integration time control. Therefore the most popular

approach for realizing a HDR imager uses a set of images of the same scene, assuming a stationary camera with fixed focal length and linear response. The images are successively captured with different exposure times which are then assembled into one high dynamic image by an external logic [50, 29, 71]. This yields a resulting dynamic range which is greater than a single snap.

An example is shown in Fig. 3.12(a), which shows an indoor scene illuminated by a bright spot light that was captured with a CCD based image sensor. Even with a minimum iris opening for capturing scene details, some parts are overexposed and tend to bloom. Image (b)...(e) show the same scene, captured with a CMOS based HDR camera. The final intensities are digitized to a 20 bit depth, with Fig. 3.12(b)...(d) displaying different 8 bit grey level ranges. The final HDR composed image converted to 8 bit intensity range is shown in Fig. 3.12(e), which saves the local contrast information. This feature makes it feasible to take images with a dynamic range of more than $100dB$.

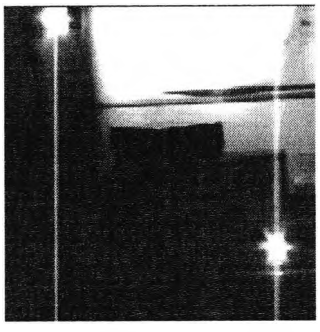
This approach implies that scene differences during image acquisition are negligible. Otherwise the computation of the final HDR output image yields incorrect scene details caused by motion or ambient illumination fluctuations. Furthermore the frame rate is affected by the need to capture a set of input images (image tuple) for calculating an accurate HDR output image. This problem was discussed in detail in Section 4.4 and 4.4.

An alternative solution for capturing HDR scenes with fast moving objects is to combine two cameras operating with different exposure times which could be linked by splitting the incoming light by a half mirror [95]. The result of assembling the long and short exposed image again yields an image with enlarged dynamic range.

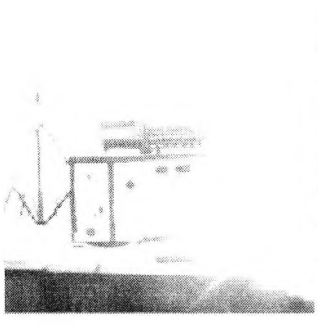
The basic principle of combining images with different brightness information due to different radiant sensitivities was already used for common photography by Wyckoff [88, 89] in 1961. He designed a multiple layer photographic emulsion with the layers differing only in their light sensitivity. That means in terms of photography to employ a very *slow* layer (ISO 2) at the bottom and a *fast* layer (ISO 600) on top. The final composite image was printed on color paper, resulting in pseudo-colored high dynamic range images. This technique was used by Wyckoff for capturing nuclear explosions and NASA's space exploration programs.

Recent developments such as those published by Street in [84] try to tackle the problem of HDR cameras without intensity compression by designing a sensor array consisting of photosensitive elements with different fixed light sensitivities. The extended linear overall dynamic range of the array is again achieved by interpolating measurements from sensor elements with high and low sensitivity. However, this pixel architecture does affect the spatial resolution of the sensor and the brightness interpolation means increased computing costs.

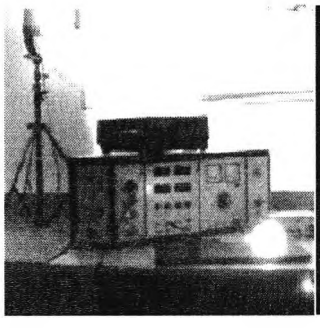
The ideal approach to guarantee that every pixel operates at its optimal



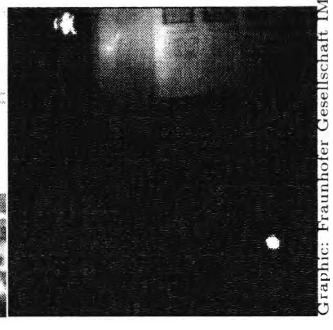
(a) CCD imager with smallest iris opening



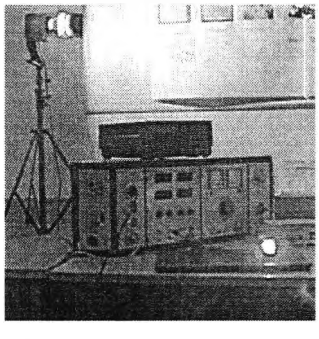
(b) CMOS bit 0...7



(c) CMOS bit 3...10



(d) CMOS bit7...14



(e) composed HDR image

Figure 3.12: (a) shows a HDR scene, captured with a CCD based image sensor. Images (b)...(d) show different 8 bit grey level ranges from a 20 bit HDR picture. The final HDR composed image, converted to 8 bit intensity range is shown in (e), saving the local contrast information.

operating range without saturation and without affecting the spatial resolution or frame rate is to use an imager with *intelligent* pixel which choose the optimal exposure time individually on chip [77]. This can be realized by using the TFA technology with autoadaptive pixels which was discussed in Section 3.4. Each smart pixel determines itself when to stop integrating. The complete illumination information is included in two signals with moderate dynamic ranges, the integration value and the time value which are both read-out from the sensor. This concept enables dynamic ranges of 120dB or more for the photosignal. An example of a HDR scene captured by a TFA imager is shown in Fig. 3.11.

3.5.3 Piecewise linear response

All discussed techniques for extended dynamic range imaging with linear CMOS sensors employ an external processing logic which calculates the final output image. In the beginning of HDR imaging such camera systems consist of specialized designs for research or industrial applications such as visual welding inspection. This limited range of applications results in small production volumes and therefore high system costs for a HDR camera.

But the consumer industry also shows increasing interest in extending the dynamic range of mainstream cameras for multimedia applications. An example is the birthday scene where a child blows out the candles from its birthday cake in a dimly lit room. The exposure control of a common digital camera focuses on the bright candles, illuminating the cake. Hence, birthday visitors in the background are not visible due to under-exposure for this image areas.

A solution to reduce the complexity of composing a HDR image captured with a linear sensor while achieving sufficient image contrast at low light conditions is to employ a *piecewise linear* response [30, 55].

A sensor with piecewise linear response is something in between linear and non-linear response, also called *dual slope* response. This technique determines the sensor's transfer curve in each individual pixel and without external control. This approach is similar to a pixel in TFA technology but with a reduced number of necessary transistors and hence an acceptable fill factor when produced in common APS technology.

At the beginning of the integration time photons are accumulated as in normal linear operation mode up to a predetermined voltage level. The slope of this integration process is determined by the integration time for low light levels t_{int_1}.

If the level is not reached (*i.e.* that part of the scene is dark) this may continue over the full integration time. In the dark parts the full integration time is available, in other words, we are in the highest sensitivity part, in the steepest slope. If the signal level reaches a determined threshold the effective

integration time is changed (*i.e.* reduced) to t_{int_2} for brighter light levels, resulting in a flatter response slope. Therefore the light power-to-voltage relation is not linear but piecewise linear as shown in Fig. 3.13.

The breakpoint U_{bp} for the resulting piecewise linear response can be positioned by programming in the analogue domain. Hence to capture a HDR scene the signal range of the sensor is subdivided in n_{slp} piecewise linear slopes with one single exposing operation, *i.e.* without combining a set of different exposed images. However, for an increased number of n_{slp} the final response curve converges to a virtual logarithmic response. To avoid this a reasonable number of breakpoints is 2...4.

The advantage of dual slope operation is improved robustness compared to logarithmic pixel compression while achieving higher contrast and less noise at low light levels. The number of available dual slope sensors has increased rapidly within recent years due to their relative simple pixel design, reducing system costs for HDR cameras. Hence the first commercial camera systems providing a dynamic range above $110dB$ are offered for applications with higher volumes such as for surveillance or automotive purposes.

The different characteristics and the resulting dynamic range for non-linear, linear and piecewise linear read-out are shown in Fig. 3.13. The example images of this chapter show monochrome scenes only because the majority of available HDR image sensors and cameras are monochrome. There is no reason why not to cover a monochrome HDR imager with a Bayer mask [5] to gain color information, but due to the rather new field of high dynamic range imaging *color* HDR imagers [23, 70] are still rare.

3.6 SollyCam

This work emphasizes the benefits of a smart fusion of hard- and software (sensor and algorithm) for cost, speed and reliability optimized results for real-time image processing in high dynamic range environments. Image processing algorithms can in principle be developed and evaluated by images from a defined environment, leading to ideal image features. However, a key feature of this work is to tackle the problems of real and unconstrained illumination environments. Hence there was a need for a suitable camera system for high dynamic range applications, such as the discussed motor vehicle interior analysis (see Chapter 1) instead of a standard multimedia version. It should provide image quality close to a real embedded system as detailed in Chapter 6.

Concerning the limitations of CCD based or CMOS based imagers with logarithmic response the TFA or linear CMOS image sensors with HDR ability are the first choice for capturing outdoor scenes. However, no commercial TFA or CMOS based camera with linear response was available at

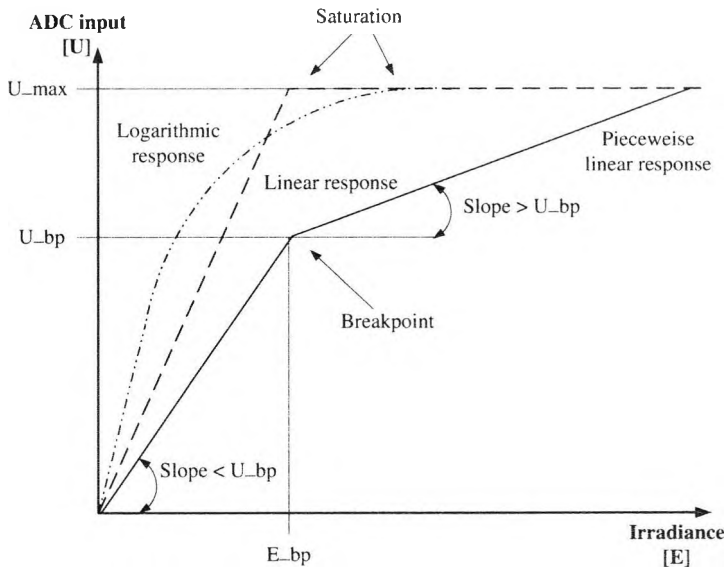
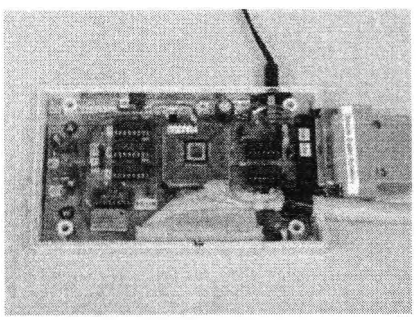


Figure 3.13: Overview of different response characteristics for CMOS based image sensors.

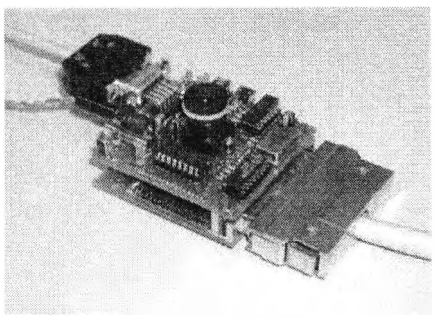
the start of this research project in 1999 which fulfilled the requirements as defined in Chapter 2. Only a couple of experimental designs of high dynamic range imagers from universities or other research laboratories (see Fig. 2.2(a)) have been available, usually coupled with unsatisfactory evaluation environments. This led to the DoubleFlash investigation for capturing an optical high dynamic scene without a HDR camera. Due to the fact that no camera was available which fulfilled our requirements, we decided to develop a custom camera optimized for our active illumination experiments (DoubleFlash, ShadowFlash, LineFlash, see Section 4.2) and automotive applications.

The following list enumerates key features of our latest imager design (SollyCam 3.0), based on the APS sensor LM9618 from National Semiconductor [56], which was employed for active illumination experiments and to capture the data for algorithm evaluation as described in Section 5.2.5 and 6.4.

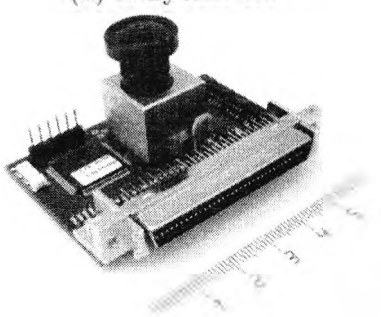
- VGA resolution (648 x 488 pixels)
- High dynamic range ($> 100dB$) due to piecewise linear response
- 12 bit digital output (RS422), programmable via PLD configuration
- Low power consumption (sensor/camera = $160mW/2W$), programmable sleep mode
- Variable timing and snapshot mode including programmable line, row and frame delays via I^2C interface
- Pixel clock up to 12 MHz



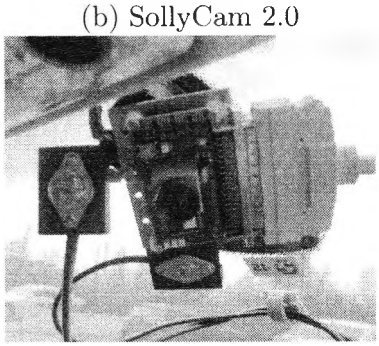
(a) Sollycam 1.0



(b) SollyCam 2.0



(c) SollyCam 3.0



(d) Car implementation

Figure 3.14: SollyCam history and automotive implementation with NIR cluster illumination.

- Still 70% photo sensitivity within the NIR at $\lambda \approx 800nm$
- Inexpensive due to high volume production
- Operating temperature -40 to $85°C$
- Special trigger logic for our active illumination experiments (DoubleFlash, ShadowFlash, LineFlash, see Section 4.2)

The development was performed in several design steps. Figure 3.14(a)...(c) shows the history from our first design for testing the sensor chip behavior concerning implementation abilities and parameterizing basic sensor scenes. This first design (SollyCam 1.0) based on the LM9617 could already be configured via I^2C interface. However, the digital sensor output line drivers have not been powerful enough to drive a standard frame grabber via TTL[5]. The next evolution step with LM9617, which was called SollyCam 2.0, was supported with an interface for digital frame grabbers using the RS 422 standard. However, a non-optimal power supply design and insufficient discrimination of digital and analog parts resulted in significantly increased noise levels and therefore an unacceptable imager dynamic. An example image is shown in Fig.3.15.

[5]Transistor-Transistor Logic

With our current design, SollyCam 3.0, the temporary noise due to improper power supply for the analog signal path was eliminated (see Fig. 3.16) and we performed the update of a HDR sensor (LM9618). The reasons were increased flexibility and the wish to also use the new design for outdoor applications such as lane tracking and rear view mirror. The dynamic range extension was realized by employing a piecewise linear operation mode as discussed in Section 3.5.3.

The camera was connected with the image processing unit (PC equipped with digital frame grabber or embedded DSP system, see Section 6.5 for details) via a 16 bit parallel RS 422/LVDS[6] link plus synchronization signals. Due to the differential data transmission the number of necessary wires exceeds three dozen, which results in large cables and plugs that are difficult to handle. Our next generation with an automotive adapted serial bus for transmitting image data and for host communication by a couple of wires is currently under development and will be available in summer 2003.

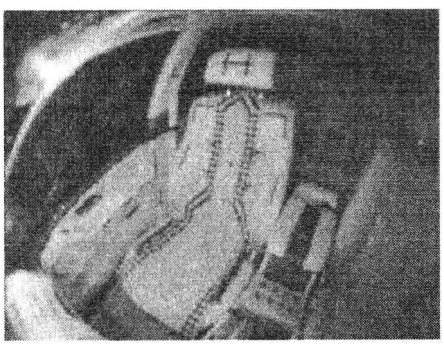
3.7 Precis

According to the measurements and experiments of Chapter 2 it is necessary for managing worst-case illumination scenarios inside a vehicle interior to employ a high dynamic range (HDR) camera or advanced active illumination. This chapter discussed the imager technologies available today and in the near future with regard to their usability for HDR scenes. A comparison of CCD, CMOS and TFA based imagers is shown in Table 3.1.

In summary CCD based image sensors are not suited for image processing in open world scenes because they tend to suffer from limited dynamic range and blooming effects. CMOS image sensors with logarithmic response provide the ability to extend the dynamic range compared to mainstream CCD cameras, but they suffer from low image contrast.

Due to the fact that no commercial TFA or CMOS based camera with linear response was available at the start of this research to fulfill the requirements as defined in Chapter 2 we decided to build our own camera system using a CMOS sensor with piecewise linear response and special logic functions for our active illumination experiments in Chapter 4. This custom imager was designed with respect to automotive requirements such as temperature range, power consumption, data transmission and was implemented in a test car to capture the data for algorithm evaluation as described in Chapter 5 and 6.

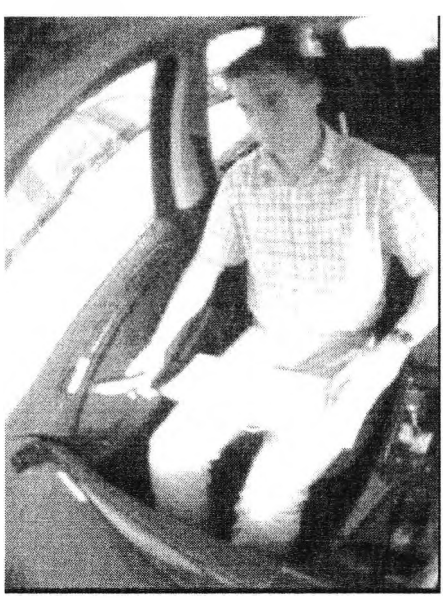
[6]Low Voltage Differential Signaling, a low noise, low power, low amplitude method for high-speed (gigabits per second) data transmission over copper wire



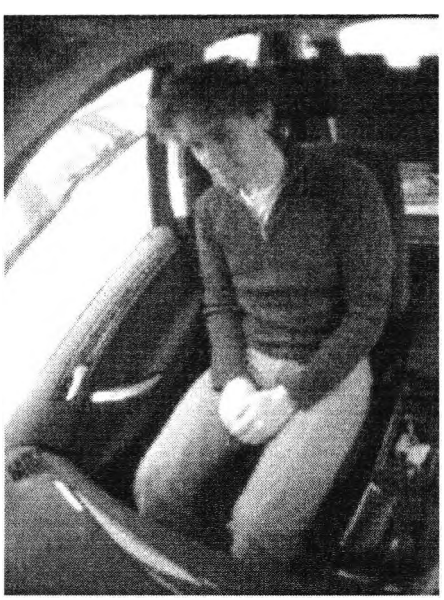
(a) Empty seat

(b) Occupied with belted person

Figure 3.15: SollyCam 2.0 example images, captured with horizontal and vertical 2:1 subsampling (324 x 244 pixel), showing significant noise and therefore a reduced sensor dynamic.



(a) $DR_{cam} < DR_{amb}$



(b) $DR_{cam} \geq DR_{amb}$

Figure 3.16: SollyCam 3.0 example images of a vehicle interior in different operation modes, showing a region of interest (240 x 320 pixel) without subsampling.

Parameter	CCD	CMOS	TFA	SollyCam
Picture quality	High due to decades of optimization	Suffer from misc. noise sources (*e.g.* FPN)	similar to CCD	see CMOS
Power consumption	High	Low, approx. ten times less than CCD	see CMOS	$132mW$
Local contrast	Low due to smearing and blooming effects	High, no blooming	see CMOS	see CMOS
Temperature range	Less than 50^o due to dominating dark current and blooming	Up to 100^o without significant loss of image quality	see CMOS	$0 \ldots 80^oC$
Dynamic range	$< 60dB$	$\geq 120dB$	$\geq 120dB$	$110dB$
Production costs	High due to external peripheral devices	Low due to single-chip capture and processing	High due to non-standard semiconductor process	Low, see CMOS, already in mass production
Optical sensitivity	High due to optically optimized amorphous silicon	Low due to crystalline silicon, compatible with standard CMOS process and due to low fill factor	see CCD	see CMOS
Random pixel access	No	Yes	Yes	Yes
Read-out	time-discrete, linear response	time-continuous and discrete, non-linear, linear and piecewise-linear response	time-discrete, linear response	time-discrete, piecewise-linear response

Table 3.1: CCD characteristics compared to CMOS and TFA based image sensors.

Chapter 4

Illumination

4.1 Motivation

For capturing images without loosing image detail in high dynamic range environments (HDR), it is necessary according to the findings of Chapter 2 and 3 to use a special camera design. A custom imager design for HDR environments was presented in Section 3.6.

This chapter discusses alternatives to special HDR cameras: the possibility of employing active illumination within a scene with significant light fluctuations to avoid the need for special (and therefore more expensive) imager designs. A novel double flash approach is proposed in Section 4.2.3 which makes it possible to capture a scene under adverse lighting conditions without using a high dynamic range camera and without losing image detail by employing active illumination.

Further investigations in smart scene illumination describe a novel shadow removal technique that produces a shadow-free scene. This proposed algorithm simulates an artificial infinite illumination plane over the field of view by using spot lights.

4.2 Active illumination

Machine vision systems in unstable illumination environments prefer to use supplementary illumination to reduce the influence of ambient illumination and thus the range of brightness variation DR_{global} within a scene as discussed in Section 2.4.

Using a bandpass filter with a center wavelength (CWL) at the wavelength of the supplementary illumination source blocks all irradiation from the ambient illumination outside the transmission wavelength band (see Section 2.3.2). A suitable wavelength range ($\Delta\lambda$) for such an illumination is the near infrared (NIR) because it is not visible to the hunan eye which is important for many surveillance tasks (see Section 2.4). In addition most cameras based on silicon are still sensitive at these wavelengths and the spectral power density of the main disturbance *noise* source (the sun) decreases significantly beyond $\lambda = 800\,\text{nm}$. This effect can be improved by utilizing a strong flashed light source which increases the signal-to-noise ratio (SNR), suppresses the noise induced by the environment and minimizes its influence on the image.

However problems with strong flashing can occur if large areas have to be illuminated because this can require an illumination source whose optical intensity can reach the threshold values for eye safety (see Appendix B). This section discusses active illumination techniques and describes a solution to this problem without the need for high dynamic range (HDR) cameras and without violating eye safety limitations whilst retaining the important scene details.

Section 4.2.1 provides an introduction to an offset reduction for ambient illumination and Section 4.2.3 introduces a double flash algorithm for dynamic range compression. Section 4.2.4 illustrates the proposed dynamic range compression with an example calculation and describes some experimental results.

4.2.1 Offset reduction

In order to reduce the necessary pre-processing steps for image analysis it is obviously preferable to capture a sequence consisting of N frames ($n = 1, \ldots, N$) without global or local illumination fluctuations, *i.e.* without varying illumination offset and noise [42, 41, 63]. This means that the irradiance E which is integrated over the exposure time (t_e) of the imager should be constant for every frame and hence the grey level of an arbitrary pixel at position $\mathbf{p}$ within a digital image $I(\mathbf{p}, n)$ only varies if the scene has changed, *i.e.* only if the reflectivity of the surface $\rho_s(\mathbf{p})$, and not the ambient illumination, has changed.

$$\begin{aligned} I(\mathbf{p}, n) &= \rho_s(\mathbf{p}) \int_{t=0}^{t_e} E(t)\, dt \\ \mathbf{p} &= (row, column) \\ 0 < \rho_s(\mathbf{p}) &< 1 \end{aligned} \tag{4.1}$$

This would simplify most image processing steps. The absence of illumination fluctuations within the scene *e.g.* supersedes the need for an adaptive background updating for motion detection (see Section 5.2).

This again reduces the necessary computing time and the cost of the processing hardware and is therefore a further step towards real-time image processing. One way to achieve this offset and noise reduction is to capture one image I_{amb} with only ambient illumination (E_{amb}) at frame n and a second image I_{flash} with additional illumination ($E_{amb} + E_{nir}$) from a supplementary light source (*e.g.* NIR) at frame (n+1).

$$I_{amb}(\mathbf{p}, n) = \rho_s(\mathbf{p}) \int_{t=0}^{t_e} E_{amb}(t)\, dt \tag{4.2}$$

$$\begin{aligned} I_{flash}(\mathbf{p}, n{+}1) &= \rho_s(\mathbf{p}) \int_{t=0}^{t_e} (E_{amb}(t) + E_{nir}(t))\, dt \\ I_{nir}(\mathbf{p}, n{+}1) &= |I_{amb}(\mathbf{p}, n) - I_{flash}(\mathbf{p}, n{+}1)| \end{aligned} \tag{4.3}$$

Assuming that the scene is static $\rho_s(\mathbf{p})$ is constant over all frames. This means that the captured grey levels within the images are only a function of E which illuminates the scene. Thus the difference between the images I_{amb} and I_{flash} yields only the received radiant power of the local illumination which is constant (see Eqn. 4.4). The variable offset of the ambient

illumination source E_{amb} is eliminated and the output sequence $I_{nir}(\mathbf{p}, n)$ is thus exempt from light fluctuations.

$$\begin{aligned} &|E_{amb} - (E_{amb} + E_{nir})| = E_{nir} = \text{const.} \\ &\rightarrow I_{nir} \propto E_{nir} \qquad if \qquad \rho_s(\mathbf{p}) = \text{const.} \end{aligned} \tag{4.4}$$

The time gap between I_{amb} and consecutive I_{flash} should be kept as small as possible to minimize the effect of changing ambient irradiation E_{amb} or surface reflectivity $\rho_s(\mathbf{p})$ caused by moving objects, which results in an inaccurate output image. This method is illustrated in Fig. 4.1. It shows the intensity difference between frame I_{amb} and I_{flash} of the same pixel over a sequence of 60 frames. The sequence begins with small illumination fluctuations and ends with a double-peak induced by a powerful external light source directed at the scene for a short time. Referring to Eqn. 4.4 the calculated difference between I_{amb} and I_{flash} should ideally be constant. The plot of I_{nir} differs marginally from an (ideal) straight line due to sensor noise, small illumination changes of E_{amb} between both input images and E_{flash} jitter. The sequence was produced with a linear HDR camera [29] (optical range of up to $120dB$, see Section 4.2.4). For applying an offset reduction via supplementary illumination the following conditions must hold:

- **Dynamic:** The camera has to cover the whole optical dynamic range of the scene and must not saturate or under-expose any pixel, otherwise the subtraction of consecutive images yields an unrepresentative result. This means that for an environment with large irradiance differences in the spatial and time domain a HDR camera is required.
- **Speed:** The time gap between I_{amb} and consecutive I_{flash} should be kept as small as possible to minimize the effect of changing ambient irradiation E_{amb} or surface reflectivity $\rho_s(\mathbf{p})$ caused by moving objects, which results in an inaccurate output image.
- **Reach:** The whole scene which has to be analyzed must be within the scope of the local illumination E_{nir}. This condition is met in closed environments, *e.g.* within a vehicle interior or building.
- **Imager response:** The response of the image sensor should be linear to minimize the effort of calculating the real difference between I_{amb} and I_{flash}. The subtraction of two images from a camera with a logarithmic response is equal to the division of one value by the other and therefore not the same. Nevertheless it is possible to use a logarithmic imager and to linearize its output. This can be done easily via a look-up table (LUT).

Another advantage of employing active illumination is that it will not only compress light fluctuations, but also suppress cast shadows within the scene due to ambient illumination. Using active illumination for shadow detection and suppression will be discussed in detail in Section 4.3.

4.2.2 Dynamic range compression

The utilization of supplementary illumination in Section 4.2.1 has an additional effect on the flashed image I_{flash}. The optical dynamic range within the image is compressed compared to the image without supplementary illumination. This is explained in the following paragraphs.

The image intensity $I(\mathbf{p}, n)$ generated by a digital imager represents the photogenerated current i_{ph} of each pixel. This current is a function of incident irradiance, sensor offset, gain and sensor noise. The dynamic range for an image sensor (DR) is commonly defined as the ratio of its largest non-saturating signal ($i_{ph,max}$) to the standard deviation of the noise under dark conditions ($i_{ph,min}$) [93].

$$DR = 20 \log \left(\frac{i_{ph,max}}{i_{ph,min}} \right) \tag{4.5}$$

Assume a scene which exhibits a wide DR in both the time and spatial domains due to fluctuating ambient illumination and different interior surface materials or in an environment with large windows (*e.g.* a vehicle interior). For simplification assume that i_{ph} is proportional to the incident irradiance E and that the imager is located in such a way that there is no irradiation directly incident on the sensor, *e.g.* within the vehicle roof. Hence the DR within the image (spatial domain) or between two frames at time n and n+1 (time domain) will be determined by the product of the reflectivity of the interior surface $\rho_s(\mathbf{p})$ and the occurring maximum and minimum irradiance E_{max}, E_{min} respectively.

$$\begin{aligned} & i_{ph} \propto E \\ \rightarrow \; & DR = 20 \log \left(\frac{\rho_s(\mathbf{p}) E_{max}}{\rho_s(\mathbf{p}) E_{min}} \right) \end{aligned} \tag{4.6}$$

Within a scene without supplementary illumination, the irradiance E_{min} and E_{max} are determined within a vehicle only by ambient illumination sources E_{amb}. This is primarily the sun.

$$\begin{aligned} E_{max} &= E_{amb,max} \\ E_{min} &= E_{amb,min} \end{aligned} \tag{4.7}$$

In the case that the vehicle interior is illuminated with a supplementary illumination E_{flash}, E_{nir} is added to the ambient E_{amb}.

$$\begin{aligned} E_{flash,max} &= E_{amb,max} + E_{nir} \\ E_{flash,min} &= E_{amb,min} + E_{nir} \end{aligned} \tag{4.8}$$

If Eqn. 4.7 and 4.8 are inserted into Eqn. 4.6, Eqn. 4.9 results which shows that the optical dynamic range of a scene with supplementary illumination

(DR_{flash}) is lower than the scene without (DR_{amb}). This effect is independent of the reflectance $\rho_s(\mathbf{p})$ which is shown in Eqn. 4.11.

$$DR_{amb} > DR_{flash} \tag{4.9}$$

$$20\log\left(\frac{\rho_s(\mathbf{p})E_{amb,max}}{\rho_s(\mathbf{p})E_{amb,min}}\right) > 20\log\left(\frac{\rho_s(\mathbf{p})(E_{amb,max}+E_{nir})}{\rho_s(\mathbf{p})(E_{amb,min}+E_{nir})}\right)$$

$$\frac{E_{amb,max}}{E_{amb,min}} > \frac{E_{amb,max}+E_{nir}}{E_{amb,min}+E_{nir}} \tag{4.10}$$

$$E_{amb,max}(E_{amb,min}+E_{nir}) > E_{amb,min}(E_{amb,max}+E_{nir})$$

$$E_{amb,max}E_{nir} > E_{amb,min}E_{nir}$$

$$E_{amb,max} > E_{amb,min} \tag{4.11}$$

$$qed.$$

This means that by using supplementary illumination which brightens the whole scene, the optical dynamic range of the scene is compressed. The power of E_{nir} determines if areas that were formerly too dark or too bright (and thus out of the dynamic range of the imager) can now be accurately captured (see Fig. 4.5). Figure 4.2 illustrates this effect. It shows the same offset reduction plot with identical data as Fig. 4.1, but with logarithmic scaling of the grey values to highlight the changes to the dynamic range of DR_{amb} and DR_{flash}.

However, if the offset reduction is applied, an input image without local illumination for determining E_{amb} is still necessary for calculating the offset free output. This image might still feature a high dynamic range. Hence, for analyzing HDR environments as described in the introduction, a HDR camera is still essential, if the offset reduction should be used. A solution to this problem is our proposed DoubleFlash technique which is described in the following section.

4.2.3 DoubleFlash

This section details an approach which combines the advantages of both offset reduction and dynamic range compression by illuminating two input images with *different radiant intensities*. Both input images $I_{flash,hi}$ and $I_{flash,low}$ are compressed in their optical dynamic range due to the supplementary illumination, and the output image is also free of illumination fluctuations. The output image I_{df} of the DoubleFlash is calculated as follows:

$$I_{flash,hi}(\mathbf{p},n) = \rho_s(\mathbf{p})\int_{t=0}^{t_e}(E_{amb}(t)+E_{nir,hi}(t))\,dt$$

$$I_{flash,low}(\mathbf{p},n+1) = \rho_s(\mathbf{p})\int_{t=0}^{t_e}(E_{amb}(t)+E_{nir,low}(t))\,dt$$

$$I_{df}(\mathbf{p},n+1) = |I_{flash,hi}(\mathbf{p},n) - I_{flash,low}(\mathbf{p},n+1)|$$

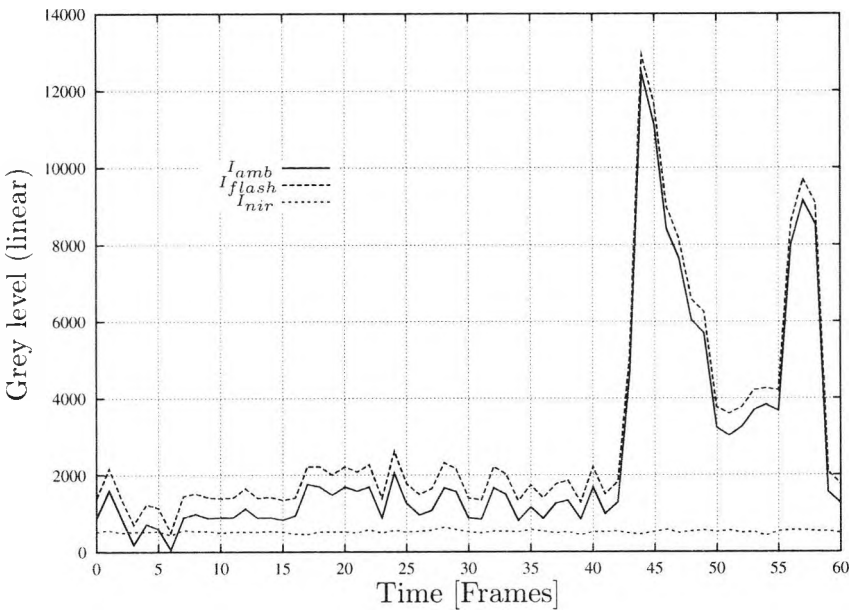


Figure 4.1: Principle of offset reduction

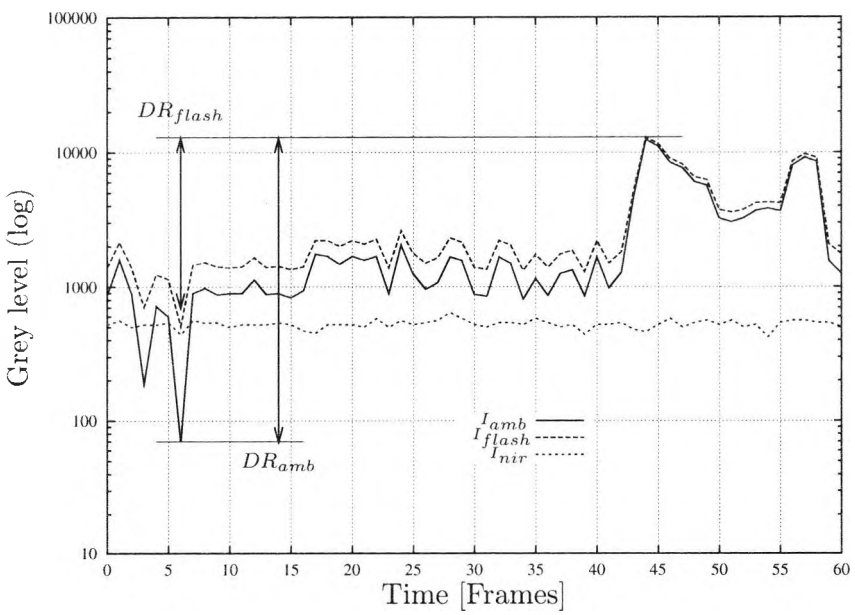


Figure 4.2: Compressed dynamic range due to supplemental illumination E_{nir}. The double arrows indicate the difference between DR_{amb} and DR_{flash}.

The influence of the ambient illumination E_{amb} has according to Eqn. 4.12 been eliminated.

$$\begin{aligned} |(E_{amb} + E_{nir,hi}) - (E_{amb} + E_{nir,low})| = \\ |E_{nir,hi} - E_{nir,low}| = \text{const.} \\ \rightarrow I_{df} \propto E_{nir,hi}, E_{nir,low} \qquad if \qquad \rho_s(\mathbf{p}) = \text{const.} \end{aligned} \tag{4.12}$$

The difference to the basic offset reduction in Eqn. 4.4 is that $E_{nir,low}$ operates as the gain for the minimum radiation E_{min} which was previously at I_{amb} and only defined by the ambient illumination E_{amb} (see Eqn. 4.4 and 4.7). The original dynamic range of *both* images are thus compressed by E_{nir} according to Eqn. 4.9.

$$\begin{aligned} E_{flash,hi,max} &= E_{amb,max} + E_{nir,hi} \\ E_{flash,hi,min} &= E_{amb,min} + E_{nir,hi} \end{aligned} \tag{4.13}$$

$$\begin{aligned} E_{flash,low,max} &= E_{amb,max} + E_{nir,low} \\ E_{flash,low,min} &= E_{amb,min} + E_{nir,low} \end{aligned} \tag{4.14}$$

This leads to the effect that the maximum dynamic range of the DoubleFlash output image (DR_{df}) is now defined by the ratio between the adjustable local illumination intensity $E_{nir,hi}$ and $E_{nir,low}$. This is shown in Eqn. 4.15 and 4.17, where Eqn. 4.13 and 4.14 are inserted into the basic dynamic range definition of Eqn. 4.6.

$$20 \log(a) = DR_{df} \tag{4.15}$$

$$a = \left| \frac{E_{flash,hi,max}}{E_{flash,hi,min}} - \frac{E_{flash,low,max}}{E_{flash,low,min}} \right| \tag{4.16}$$

$$= \left| \frac{E_{amb,max}+E_{nir,hi}}{E_{amb,min}+E_{nir,hi}} - \frac{E_{amb,max}+E_{nir,low}}{E_{amb,min}+E_{nir,low}} \right| \tag{4.17}$$

The timing diagram of the trigger logic for this illumination technique is shown in Fig. 4.3. The two main sources of IR devices are Light Emitting Diodes (LED) and laser diodes. The LEDs are slower and less efficient but also less expensive than the laser diodes. However, laser diodes require more complex drivers and lasers are a perceived hazard. Therefore the experiments have been carried out with NIR high power LED clusters as illumination sources. The diodes operate with alternating high and low currents and emit $E_{nir,hi}$ and $E_{nir,low}$, respectively.

In the case of a camera with a global and synchronous shutter it is far more efficient to activate the LED illumination only when the imager pixels are integrating light. The off-time when the image is read-out can be used to allow the LED's to cool off and also to reduce the total power consumption. Hence it is possible to run the LED's with higher pulse currents resulting in the use of smaller diodes with no reduction in radiant power.

The absolute function which is applied in Eqn. 4.4 and 4.12 implies that the illumination order of $E_{nir,low}$ and $E_{nir,hi}$, *i.e.* the capturing to $I_{flash,low}$ and $I_{flash,hi}$ respectively, is arbitrary.

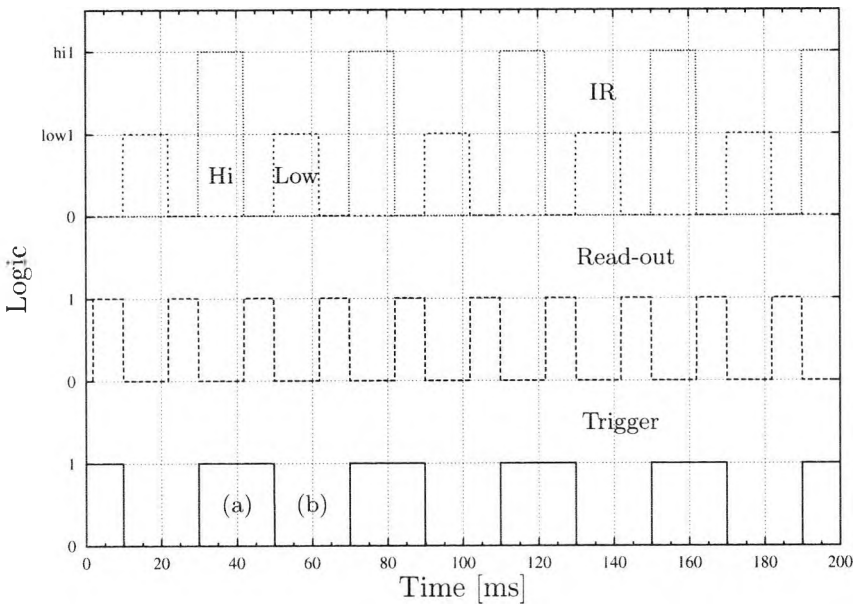


Figure 4.3: Camera timing for the DoubleFlash: One frame is equal to 20 ms, including 8 ms for read-out and transfer to the PC.

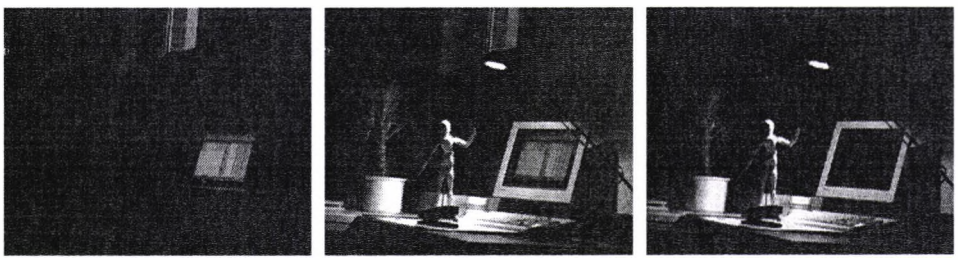
Thus, although the offset reduction and DoubleFlash need two input images, they produce one output image for every new input after the second captured image, because only an *irradiance change* between subsequent frames is required. The last captured image at time $(n-1)$ can be compared to the current image n for computing the next I_{df} output image and hence suffer only a 1 frame delay.

4.2.4 Active illumination experiments

An example of offset reduction is presented in Fig. 4.4 which shows an indoor scene captured with a mainstream CCD camera. Figure 4.4(a) shows the scene with ambient illumination E_{amb} only, Fig. 4.4(b) with supplementary illumination (E_{amb}+E_{flash}) and Fig. 4.4(c) the difference between both (I_{nir}). Notice that all scene details which are not brightened by the supplementary illumination (table lamp) are eliminated, such as the windows but also the monitor display.

Fig. 4.5 shows another indoor scene in a laboratory which includes high irradiance differences generated by a bright spot light from the right. This spot light in combination with the ambient illumination simulates E_{amb}. A mainstream CCD-based camera with a single integration time (and thus limited dynamic range) is employed to capture the scene.

The integration time in Fig. 4.5(a) was optimized for reading the labels



(a) I_{amb} (b) I_{flash} (c) I_{nir}

Figure 4.4: Offset reduction example: all scene details which are not brightened by the supplementary illumination (table lamp) are eliminated in image (c).

(a) Short exposure (b) Long exposure (c) With E_{nir} (proposed)

Figure 4.5: Example of an optical high dynamic range scene. Image (c) indicates the result of the proposed dynamic range compression by supplementary illumination. The bright light bulb and the dark background are visible simultaneously.

(a) Short exposure (b) Long exposure (c) DoubleFlash (proposed)

Figure 4.6: Car occupant detection via CMOS camera. Image (c) shows the result of the proposed DoubleFlash: Reduced dynamic, without ambient offset and FPN.

of the power supply. By extending the integration time in Fig. 4.5(b) the silhouette of a dark ring on top of the floppy disk becomes visible but most of the image is over-exposed. The imager was not able to capture the scene without losing detail with a single exposure time. A supplementary illumination E_{nir} as used in Section 4.2.2 and 4.2.3 in combination with a shortened integration time (see Fig. 4.5(c)) makes it possible to capture the labels as well as the dark ring on top of the floppy with just one exposure time. The primary direction and intensity of the disturbing light source from the right remained untouched for every image. The radiance emmitted by the supplementary illumination brightens the entire scene (E_{flash}), *i.e.* in both the dark areas and bright areas. Hence in order to capture the bright areas an integration time reduction was necessary though the additional radiance was sufficient to shift the image detail of the dark areas into the dynamic range of the camera.

The following calculation illustrates our experimental results for dynamic range reduction. Assuming a scene with a dynamic range of $E_{amb,min} = 0.44\,mW/m^2$ and $E_{amb,max} = 50\,W/m^2$ which was measured within a motor vehicle interior on a sunny summer day (see Section 2.4). This data record of the interior yields a raw optical dynamic range of up to $102dB$, which is outside the capabilities of mainstream imagers.

To capture the interior across a wide range of illumination situations without losing image detail it was necessary to employ a linear HDR camera [29] with a dynamic range of $120dB$. Figure 4.1 and 4.2 were created with the data from such a camera. Our aim was to find a way of employing a mainstream imager with limited dynamic range to capture the vehicle interior without losing the interior details. We solved this problem by using the DoubleFlash approach as proposed in Section 4.2.3 (see Fig. 4.6). A supplementary offset of $E_{nir} = 50\,mW/m^2$ produced by a NIR illumination source compresses the image dynamic to $60dB$ (see Eqn. 4.18). This is equivalent to a maximum contrast of $1000:1$ and thus corresponds to a resolution of 10 bits which is within the range of mainstream CMOS imagers.

$$DR_{amb} = 20\log\left(\frac{50\,\mathrm{W/m^2}}{0.44\mathrm{mW/m^2}}\right) = 102\mathrm{dB} \tag{4.18}$$

$$DR_{flash} = 20\log\left(\frac{50\,\mathrm{W/m^2}+E_{nir}}{0.44\,\mathrm{mW/m^2}+\mathrm{E_{nir}}}\right) = 60\mathrm{dB}$$

$$\rightarrow E_{nir} \approx 50\,\mathrm{mW/m^2} \tag{4.19}$$

To illustrate the DoubleFlash effect we extend the calculation of Eqn. 4.18. The final output image of the DoubleFlash should show a maximum dynamic of $DR_{df} = 48dB$. This is equal to a maximum contrast of $255:1$, which can be displayed and stored with 8 bits. The maximum dynamic range of mainstream CMOS imagers was already specified for the first example with $60dB$ and thus leads to a necessary supplementary illumination of $50mW/m^2$. This is the minimum illumination necessary to reduce the dy-

namic range of the scene to the requirements of the given imager. Hence it is labelled $E_{nir,low}$ because a more powerful local illumination $E_{nir,hi}$ merely yields a stronger dynamic compression and is therefore within the detection range of the imager.

$$20\log(a) = DR_{df} = 48\,\mathrm{dB} \tag{4.20}$$

$$a = \left| \frac{50\,\mathrm{W/m^2} + E_{nir,hi}}{0.44\,\mathrm{mW/m^2} + E_{nir,hi}} - \frac{50\,\mathrm{W/m^2} + E_{nir,low}}{0.44\,\mathrm{mW/m^2} + E_{nir,low}} \right|$$

$$\text{with } E_{nir,low} = 50\,\mathrm{mW/m^2}$$

$$\rightarrow E_{nir,hi} \approx 68\,\mathrm{mW/m^2}$$

A further advantage ot the DoubleFlash approach arises if the CMOS imager suffers from fixed pattern noise (FPN) [77]. FPN is an individual offset of each pixel caused by slight unwanted variations of active pixels, *i.e.* transistor characteristics (see Section 3.3). To minimize the FPN of the camera the captured data can be compared internally or externally to an offset-map for correcting the final output. This FPN correction can be performed by firmware and is thus relatively fast but still takes time.

Due to the fact that the individual FPN's represent an offset which is constant over time it will be eliminated if two subsequent images are subtracted from each other. Hence the output image sequence of a DoubleFlash system improves the final picture quality for imagers which suffer from FPN and the internal or external FPN correction can be disabled. The results of the DoubleFlash are shown in Fig. 4.6 which shows the passenger and driver seat of a car. The CMOS camera was equipped with an optical NIR bandpass as described in Section 2.3.2 which cuts off wavelengths in the visible range. A strong halogen lamp (illuminating through the sun roof) simulates E_{sun} and causes a bright region on the infant seat. The CMOS camera was not able to capture details of the infant and the driver seat within the same image just by varying the exposure time or aperture. Furthermore strong FPN effects are visible which might distort a texture analysis of the scene. Figure 4.6(c) was acquired using the DoubleFlash approach: Texture of both infant and driver seat are visible due to the compressed dynamic range. The influence of ambient illumination (the halogen lamp) was eliminated and the image quality was considerably improved due to the reduced FPN.

Summary

The irradiance $E_{amb,min}$ and $E_{amb,max}$ in a given high dynamic range scene is fixed and usually requires a HDR camera for capturing all the scene details. By using two supplementary flashes $E_{nir,low}$ and $E_{nir,hi}$ it is possible to influence the optical dynamic range of the scene so that a scene with high dynamic range (DR_{amb}) can be captured by commonly available cameras with limited optical dynamic range (DR_{df}) without sacrificing image detail

and with synchronous offset reduction. The grey-level of a pixel within the output image sequence $I_{df}(\mathbf{p}, n)$ varies only if the scene changes and not as a result of fluctuations in the ambient illumination levels. This significantly simplifies any subsequent image processing. An additional advantage referring to the otherwise necessary HDR cameras is that the supplementary illumination increases the brightness of the scenes. This enables the camera system in principle to run with a higher framerate because the integration time can be shortened due to the higher minimum irradiance level resulting in a greater signal-to-noise ratio (SNR).

The major shortcoming of the DoubleFlash is that its application is limited to environments where the whole scene to be analyzed must be within the scope of the local illumination E_{nir}. The image pre-processing described in Eqn. 4.12 is relatively simple and could be performed by firmware within the camera or integrated onto the sensor if CMOS technology is used. This yields an increase in the processing speed of the entire image processing system. Finally the costs of a mass produced camera with limited dynamic range in combination with simple trigger logic for the illumination are lower than the costs of a smart HDR camera.

4.3 Shadows and reflections

Shadows and reflections occur in a wide variety of scenes. If shadows are properly detected in images they can provide a great deal of information about the shape, relative position, and surface characteristics of objects in a scene. It is not difficult for humans to distinguish shadows from objects, but shadow effects represent a considerable challenge for machine vision tasks.

For instance, color variations caused by shadows and reflections within an image may result in an improper object shape segmentation with serious artifacts or detection of an imaginary object. This might result in shadows misclassified as objects or parts of objects due to the over/underestimation in a subsequent matching phase. Therefore many existing machine vision algorithms assume that the results of the processing are not influenced by shadows or the shadows in an image have been removed [24].

An example of a scene which is dominated by shadow effects is shown in Fig. 4.7. It shows a motor vehicle interior while a passenger is entering/leaving the car. The scene is illuminated by two bright spot lights in the near infrared (NIR) which generate large shadows within the vehicle that claim about one third of the total image area. This in turn creates problems for some established background updating mechanism and segmentations approaches, see Chapter 5.

A possibility of minimizing shadow effects is to illuminate the scene directly from the camera. However, this is not always possible for economic, ergonomic and/or design reasons. In case of monitoring occupants of a

Figure 4.7: Shadow example of a motor vehicle interior caused by bright spot light from supplementary illumination within the near infrared (NIR). Shadows claim about one third of the total image area.

motor vehicle the interior is already equipped with conventional illumination devices in the visible wavelength range. Hence it is advantageous to use these locations for additional lighting in the interior for machine vision tasks, *e.g.* within the near infrared (see Section 2.4). This reduces system costs for wiring, optics and housing. Furthermore separating the camera and the illumination provides more flexibility concerning design and integration aspects.

To prevent misclassifications of shadowed areas the shadows must be explicitly detected or efficiently removed. Several factors are required to conclude the presence of shadows in a scene: the knowledge of geometric information, the presence of obstructions, and the characteristics of both materials and light sources. Since the knowledge of these factors cannot be readily obtained under real world conditions it is still a difficult task to identify or eliminate shadows from the scene. Moreover detecting shadows also involves solving many problems such as region extraction and knowledge representation/integration. The following sections focus on shadows but also address the reflection problem because the basic effect to image processing is comparable.

A number of approaches have been studied to overcome the problem of detecting shadow regions [79, 72, 82, 54]. Existing shadow detection algorithms can be classified in terms of whether the algorithm actively uses knowledge of the environmental conditions or not. The geometric information of a scene and the known directions of light sources have been required

in identifying shadows in [44]. It also has been shown that shadows can be detected without knowledge of the geometry in an image with several assumptions [82]: a stationary camera [79], a light source that is strong enough to generate visible shadows [90], a background with a sufficient amount of texture and the dominance of a smooth-shaped background [82]. However, most of the recent shadow-related algorithms only provide the location of the shadows and cannot provide a complete solution for applications that must suppress shadow effects.

A well-established way to detect shadows or reflections is to employ the effect that shadowed areas or areas with reflections take a new color or grey level, but keep their underlying surface characteristics such as texture [1, 72]. Hence shadows are a spatial effect and therefore shadows and reflections can not be distinguished unambiguously by a pixel segmentation technique.

This fact can be used by analyzing image sequences whereby region oriented algorithms compute texture features such as mean and variance of a sliding convolution mask over time, see Section 5.2.5 for further details. For example, if the current input image is divided by an established reference background a robust hint for a shadowed area is given if the mean of the pixel's neighborhood decreases but the variance remains constant. Conversely an increased mean indicates an area which was influenced by reflections. This method works well when the surface on which the shadows are cast has texture since two equal texture patterns with different means do not explicitly imply a shadow. If, *e.g.*, a homogeneous surface is occluded by another darker homogeneous surface the criterion for a shadow is fulfilled but no shadow is present. Furthermore this was performed on the premise that a reliable reference frame had already been established.

The computing costs of determining and analyzing area textures and adjusting their results by higher image processing steps are considerable.

Another essential problem is how to deal with detected shadows. The aim of established shadow detection is to discriminate between shadows and objects. This is necessary for accurate classification of objects if the object shape is a key feature. Many segmentation algorithms for image sequences employ some kind of background updating, *e.g.* a Kalman filter [69]. But should the detected shadows be adapted to the background or not? How about areas which are shadowed for a longer period? If these areas are exempt from the background updating they will yield erroneous segmentation results after disocclusion, see Chapter 5.

This section addresses the problem of shadows in an actively illuminated environment. Section 4.3.1 gives an introduction to the nature of shadows and includes the basic definitions for our experiments. Our experiments are divided into two different approaches to handling shadow effects by active illumination. Section 4.3.2 describes an algorithm for detecting and localizing shadow regions within an image and Section 4.3.4 describes a shadow

suppression algorithm which simulates a light source with infinite dimensions. The output image or image stream is free from shadow effects and can be used for improved object segmentation and background updating. First results of this approach have been published by Yoon in [94] and Koch in [43].

4.3.1 Fundamentals

Shadows occur when objects totally or partially occlude direct light from a light source. They are visible when a viewer is not coincident with a light source. A shadow has two parts: the *self shadow* and the *cast shadow* [32, 82]. The *self shadow* is the shaded part of the object which is not illuminated by direct light. The *cast shadow* is the area projected by the object in the direction of direct light. The intensities of cast shadows and self shadows are not absolutely zero because of reflected light and ambient light. For a scene with multiple objects self shadows and cast shadows may receive reflected light from other sources. Fig.4.8 illustrates the formation of a cast shadow.

The part of a cast shadow where direct light is completely blocked by its object is called an *umbra*. The part of a cast shadow where direct light is only partially blocked is called a *penumbra*. A point light source only generates umbra shadows. An area light source generates both penumbra and umbra shadows. The penumbra will be neglected in this work for simplification due to the assumption that the bright spot light sources are infinitely small and due to the fact that the proposed algorithm suppresses umbra and penumbra as well. Furthermore a narrow penumbra may not appear in an image due to digitization effects.

4.3.2 Shadow detection

Fig. 4.9 illustrates the formation of shadows when two bright light sources exist. In this case the shadows are classified into four regions (see Fig.4.9): the cast shadows influenced by only one light source (a, b), the shadowless region perfectly irradiated by both light sources $(a \cap b)$, and the overlapped shadow region only affected by ambient illumination $(a \cup b)^c$.

The formation is interpreted as a Venn diagram as shown in Fig. 4.10. Assuming that the universal set represents an image each set stands for an area with constant illumination energy from one light source. The intersection represents the set where the surface is illuminated by two light sources at the same time. Let us define the *supplementary* irradiance E_{spl1} and E_{spl2} that are emitted by the right and left light sources in Fig. 4.9(1) and (2), respectively. Assuming that E_x is the irradiance of the region x the irradiance map $E(\mathbf{p})$ of an image can be expressed as:

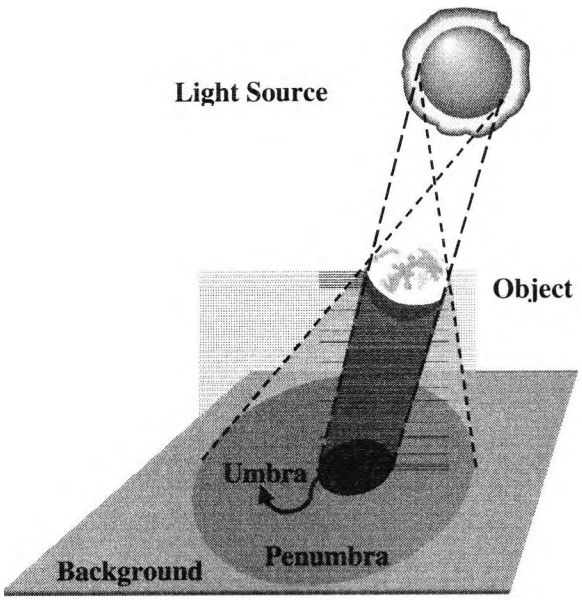


Figure 4.8: Illustration of penumbra and umbra within a shadow scene caused by a spot light.

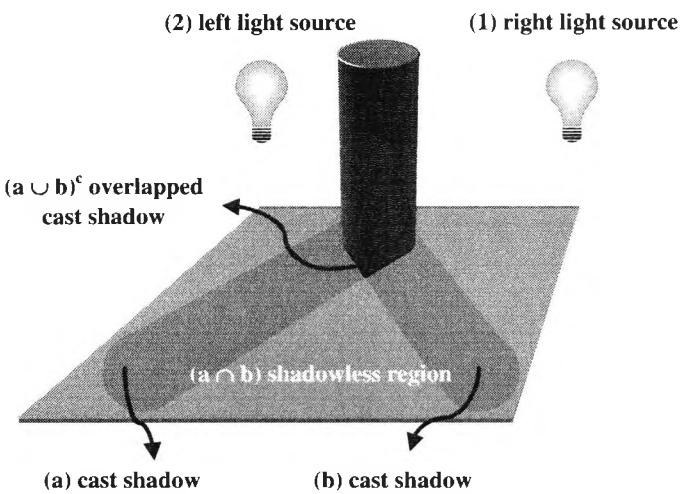


Figure 4.9: Illustration of a shadow scene illuminated by two spot light sources.

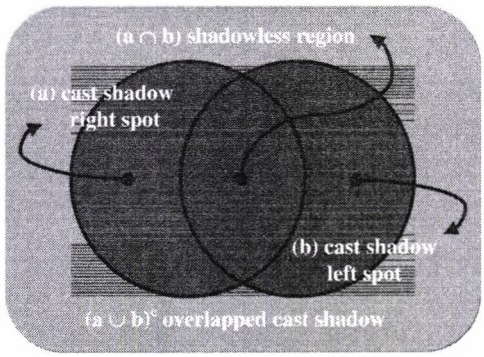
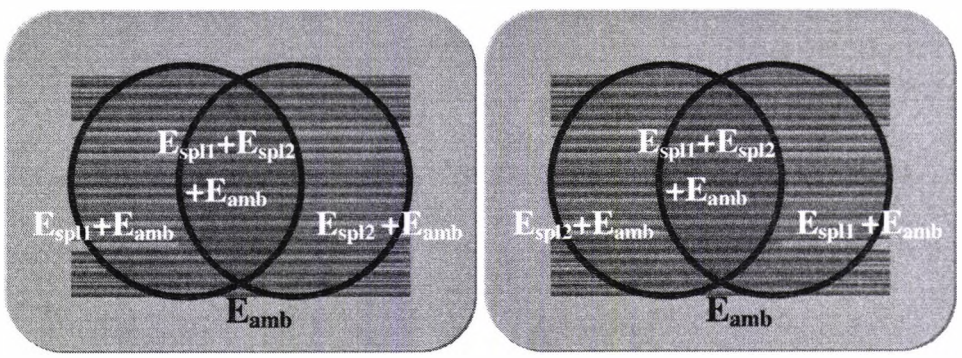


Figure 4.10: A Venn diagram based on the amount of the irradiance power.



(a) I_a (b) I_b

Figure 4.11: Venn diagram of a shadow scene illuminated by two spot lights.

$$E(\mathbf{p}) = \begin{cases} E_a = E_{spl1} + E_{amb} & \text{if } \mathbf{p} \in a \\ E_b = E_{spl2} + E_{amb} & \text{if } \mathbf{p} \in b \\ E_{a \cap b} = (E_{spl1} + E_{spl2}) + E_{amb} & \text{if } \mathbf{p} \in a \cap b \\ E_{(a \cup b)^c} = E_{amb} & \text{if } \mathbf{p} \in (a \cup b)^c \end{cases} \tag{4.21}$$

Here $\mathbf{p}$ is a position vector and E_{amb} represents ambient irradiance. Furthermore let us denote the radiant intensity I_x^e as the power emitted from the point light source x into the unit solid angle, and the irradiance E_x is the power received at the unit surface element. The radiant intensity in this work appears with the superscript e *(electromagnetic)* to distinguish it from an image I, see Table 2.2 on Page 20.

Assume that the difference of distances between an object and each light source is a small constant ϵ_x. By neglecting ϵ_x the relation between the irradiance and the radiant intensity is simplified into

$$E_x = \lim_{\epsilon_x \to 0} \frac{I_x^e}{(d+\epsilon_x)^2} \simeq c \cdot I_x^e$$

where d is the average distance and c is an adequate constant. Thus the irradiance powers caused by two light sources are equivalent if the radiant intensity of the light source I_1^e is identical with I_2^e (see Eqn. 4.1 and Fig. 4.9).

Note that all of these areas are still brightened by the ambient light so that the average intensity of an image might be disturbed considerably by an illumination change within the environment. Therefore the concept of an offset reduction technique as discussed in Section 4.2.1 is introduced to our algorithm, taking the generation of a video sequence into account.

The basic idea of shadow detection by active illumination rests upon on the following theorems:

Position: If the geometric position of a supplementary spot light changes the existing shadow borders in the scene also change. Borders of real objects do not.

Intensity: If the amount of total irradiance of a scene with a spot light increases the intensity of unshadowed areas increases in the same manner. The intensities of cast shadows do not.

This theory should be proved by a simple experiment for shadow detection without using spatial convolution filters.

For the first image I_a the left light source has the irradiance E_{spl1} while the right light source has E_{spl2}. And the second image I_b is illuminated with the opposite irradiance to I_a. The positions of both light sources are arbitrary but they must not be coincident. The distribution of the irradiance for the input images is shown in Fig. 4.11(a) and (b). Assuming that there is no illumination interference caused by the self-reflection the supplementary irradiance powers, E_{spl1} and E_{spl2} are added to the ambient irradiance E_{amb} while influencing the corresponding parts of the Venn diagram.

$$I_a(\mathbf{p}) \propto \begin{cases} E_a & = E_{spl1} + E_{amb} & \text{if } \mathbf{p} \in a \\ E_b & = E_{spl2} + E_{amb} & \text{if } \mathbf{p} \in b \\ E_{a \cap b} & = (E_{spl1} + E_{spl2}) + E_{amb} & \text{if } \mathbf{p} \in a \cap b \\ E_{(a \cup b)^c} & = E_{amb} & \text{if } \mathbf{p} \in (a \cup b)^c \end{cases} \tag{4.22}$$

$$I_b(\mathbf{p}) \propto \begin{cases} E_a & = E_{spl2} + E_{amb} & \text{if } \mathbf{p} \in a \\ E_b & = E_{spl1} + E_{amb} & \text{if } \mathbf{p} \in b \\ E_{a \cap b} & = (E_{spl1} + E_{spl2}) + E_{amb} & \text{if } \mathbf{p} \in a \cap b \\ E_{(a \cup b)^c} & = E_{amb} & \text{if } \mathbf{p} \in (a \cup b)^c \end{cases} \tag{4.23}$$

Then, assuming that $E_{spl1} > E_{spl2}$ and neglecting any camera noise, the frame ratio I_{div} between I_a and I_b can be simplified to

$$I_{div}(\mathbf{p}) = \frac{I_a(\mathbf{p})}{I_b(\mathbf{p})} \propto \begin{cases} \left\lceil \frac{E_{spl1}+E_{amb}}{E_{spl2}+E_{amb}} \right\rceil & > 1 \quad \text{if } \mathbf{p} \in a \\ \left\lceil \frac{E_{spl2}+E_{amb}}{E_{spl1}+E_{amb}} \right\rceil & < 1 \quad \text{if } \mathbf{p} \in b \\ \left\lceil \frac{(E_{spl1}+E_{spl2})+E_{amb}}{(E_{spl1}+E_{spl2})+E_{amb}} \right\rceil & = 1 \quad \text{if } \mathbf{p} \in a \cap b \\ \left\lceil \frac{E_{amb}}{E_{amb}} \right\rceil & = 1 \quad \text{if } \mathbf{p} \in (a \cup b)^c \end{cases} \tag{4.24}$$

According to Eqn. 4.24 the ratio between I_a and I_b should ideally be a constant ≈ 1 within I_{div} if every surface can be illuminated without obstruction by both illumination sources at $\mathbf{p} \in a \cap b$ or $\mathbf{p} \in (a \cup b)^c$. Deviations from this constant (or plane) indicate shadowed areas and reflections.

This is illustrated in Fig. 4.12(a) and (b) which show a surface occupied with a single object (marker). This object causes two cast shadows due to two non-coincident light sources as defined in Eqn. 4.22 and Eqn. 4.23. But self shadows on the object surface are also visible due to the convex object shape.

A three dimensional plot of Fig. 4.12(c) is shown in Fig. 4.13. $I_{div}(a \cap b)$ and $I_{div}(a \cup b)^c$ show a plane rather than a constant because $E_{spl1,2}$ at I_a was not identical to $E'_{spl1,2}$ at I_b. Hence the calculation of I_{div} according to Eqn. 4.24 yields the illumination distribution of the scene.

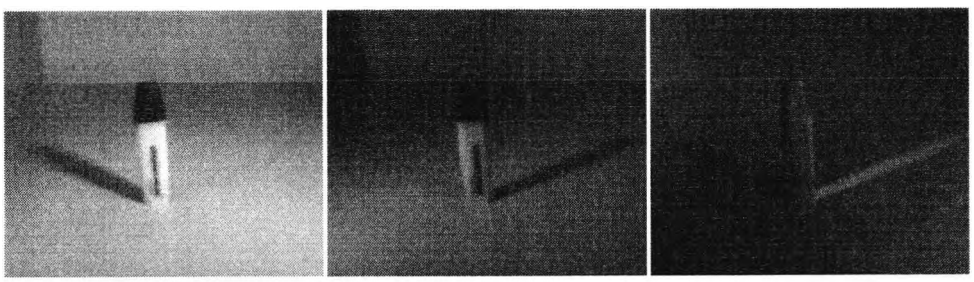
Notice that cast and even self shadows have been emphasized but the texture (characters) of the marker is gone. This is similar to the *photometric stereo* effect where at least two light sources and a *monocular* imager will be used to generate a three dimensional depth map of the scene.

4.3.3 Shadow suppression

Shadows and reflections occur due to a bright spot light as discussed in the previous sections. This effect can be reduced by using several illumination sources to guarantee an uniform lighting, or by using a diffuse light source.

Clouds present a typical example for a diffuse light source in open world environments. A cloud consists of countless water particles. A light ray passing through the cloud is evenly scattered due to the reflections against the particles, see Section 2.4. This physical phenomenon enables the photons to spread over the entire cloud and generate a spatially extended virtual light source. Consequently on an overcast day, the white sky makes an infinite size of light source and no shadows occur on the ground as shown in Fig. 4.14.

Using Gauss's law, one can easily prove that the strength of the electric field is independent from the distance from an infinite charged plane. Similarly the *irradiance*, the amount of light power per surface area, is not influenced by the distance from the light source with infinite extent (see Table 2.2 on Page 20). We cannot build an infinite plane in real life. How-



(a) I_a (b) I_b (c) I_{div}

Figure 4.12: Simple scene for explaining the basic idea of shadow detection: I_a, I_b and I_{div}. Notice, that the result of the division was gained to adapt to a common $0 \ldots 255$ grey level range.

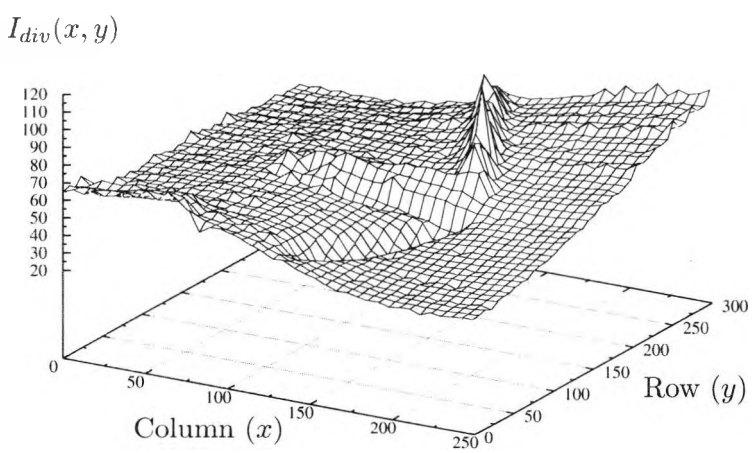


Figure 4.13: Three dimensional plot of I_{div} as shown in Fig. 4.12(c).

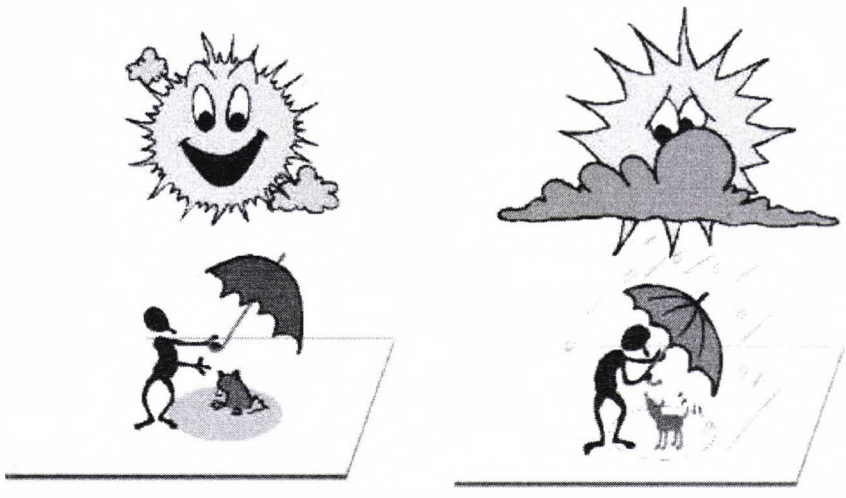


Figure 4.14: Shadows in open world scenes. Left: bright summer day with cast shadow and sharp borders. Right: Less or absent shadow effects on rainy days with occluded sun light.

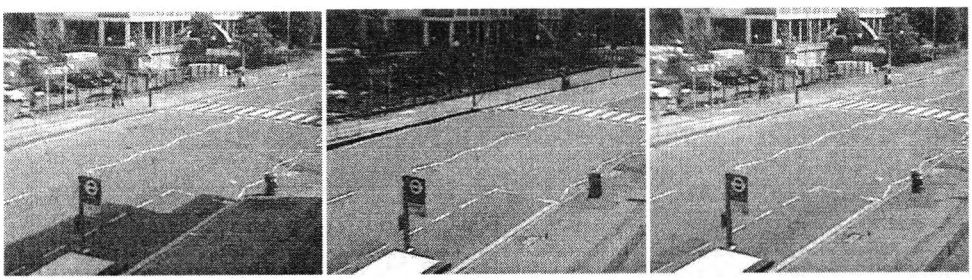
ever, the simulation of an artificial infinite illuminant plane is undoubtedly possible in the modern computing environment.

The frame ratio as presented in Section 4.3.2 is a fast and reliable tool for detecting shadowed areas. But our aim is not only to detect cast and self shadows but to suppress them to get a shadow free output image. Therefore if we can equalize the irradiance of all areas ($E_a = E_b = E_{a \cap b} = E_{(a \cup b)^c}$) the simulation of an infinite illuminant plane could be possible to deliver the constant illumination power to the entire area. Accordingly our aim in this section is to equalize the irradiance levels of the area illuminated by our active illumination to minimize shadow effects.

One simple possibility to suppress shadow effects by varying illumination is to employ two separated non-coincident light sources with equal intensity which alternatingly illuminate a scene from two different positions (I_a, I_b). If the non-linear $max()$ operation is applied, the shadowed areas will be filtered resulting in an output image I_{out} which seems to be illuminated by an infinite illuminant plane.

$$I_{out} = max(I_a, I_b) \tag{4.25}$$

An example of this approach employing outdoor images is shown in Fig. 4.15. The considerable time interval between Fig. 4.15(a) and (b) results in different illumination positions of the moving sun. Most of the shadows are successfully eliminated though some artifacts emerged in the result because of the scene difference (*e.g.* pedestrians).



(a) I_a (b) I_b (c) I_{out}

Figure 4.15: Example for outdoor images: (a) and (b) are the input for the $max()$ operator which results in (c).

However, the $max()$ operator which uses two input images is not usable if an offset reduction or dynamic range compression is required, see Section 4.2.1 and 4.2.3. Hence, our goal is to define a general description which considers more illumination situations and incorporates our findings of Section 4.2.

4.3.4 ShadowFlash

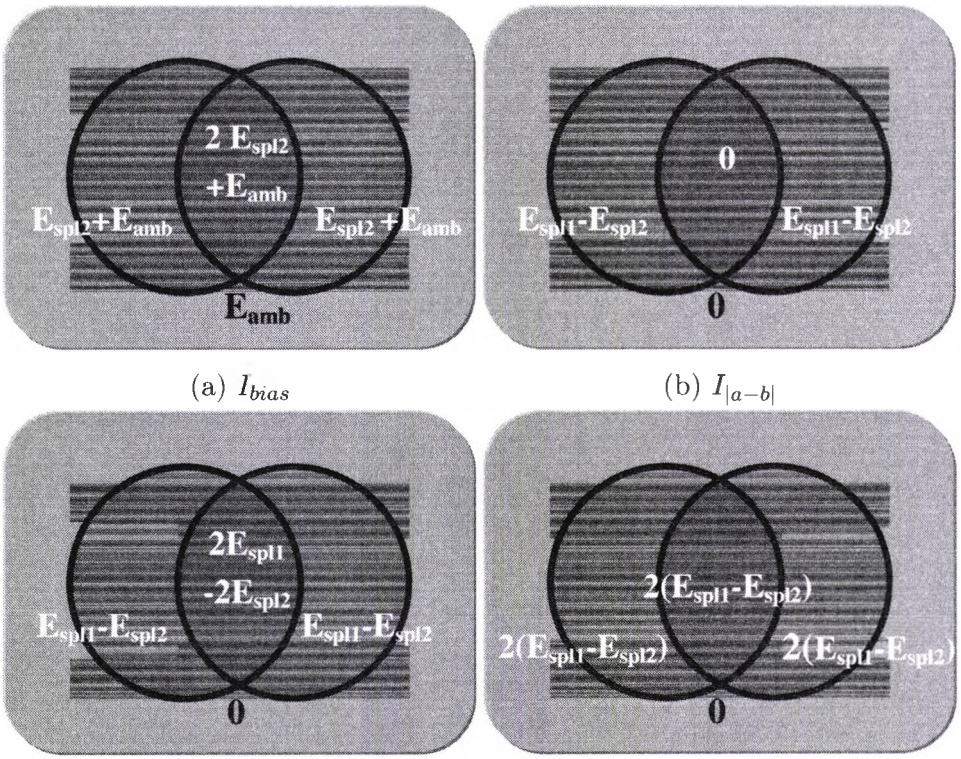
Contrary to the limited $max()$ operation as defined in Eqn. 4.25, our proposed general approach for simulating an infinite illumination source ('ShadowFlash') uses three differently illuminated images: Two separate supplementary light sources alternatingly illuminate the scene with different illumination levels.

We assume again that the acquisition time for each image is short enough to neglect scene differences between the input images (see Section 4.4). For the first image I_a the left light source has the irradiance E_{spl1} while the right light source has E_{spl2}. And the second image I_b is illuminated with the opposite irradiance to I_a. E_{spl2} is supported to both sides of the third image I_{bias}. We further assume that E_{spl1} is greater than the second supplementary irradiance E_{spl2}. The positions of both light sources are arbitrary, but they must not coincide. The distribution of the irradiance for the input images is shown in Fig. 4.11 and Fig. 4.16.

With the combination of the input images, I_a, I_b, and I_{bias} one can finally compose an irradiance-equalized image I_{out}. It is given by:

$$I_{out} = |I_a - I_b| + (I_a + I_b) - 2 \cdot I_{bias} \tag{4.26}$$

Assume that the two supplementary illumination sources can illuminate the scene with two different irradiance levels, E_{spl1} and E_{spl2} and that E_{spl1} is always larger than E_{spl2}. If we suppose that $min\,(E_{spl1}, E_{spl2}) > E_{amb}$ then the irradiance based on each region is adjusted as:



(a) I_{bias} (b) $I_{|a-b|}$

(c) I_{a+b} (d) I_{out}

Figure 4.16: An illustration of the shadow removal procedure: $I_{|a-b|} = |I_a - I_b|$, $I_{a+b} = I_a + I_b - 2 \cdot I_{bias}$, and $I_{out} = I_{|a-b|} + I_{a+b} - 2 \cdot I_{bias}$

$$
\begin{aligned}
E_{a,out} &= |(E_{spl1} + E_{amb}) - (E_{spl2} + E_{amb})| \\
&\quad + \{(E_{spl1} + E_{amb}) + (E_{spl2} + E_{amb})\} - 2 \cdot (E_{spl2} + E_{amb}) \\
&= \mathbf{2 \cdot (E_{spl1} - E_{spl2})} \qquad (4.27) \\
E_{b,out} &= |(E_{spl2} + E_{amb}) - (E_{spl1} + E_{amb})| \\
&\quad + \{(E_{spl2} + E_{amb}) + (E_{spl1} + E_{amb})\} - 2 \cdot (E_{spl2} + E_{amb}) \\
&= \mathbf{2 \cdot (E_{spl1} - E_{spl2})} \qquad (4.28) \\
E_{a \cap b,out} &= |(E_{spl1} + E_{spl2} + E_{amb}) - (E_{spl2} + E_{spl1} + E_{amb})| \\
&\quad + \{(E_{spl1} + E_{spl2} + E_{amb}) + (E_{spl2} + E_{spl1} + E_{amb})\} \\
&\quad - 2 \cdot (2 \cdot E_{spl2} + E_{amb}) \\
&= \mathbf{2 \cdot (E_{spl1} - E_{spl2})} \qquad (4.29)
\end{aligned}
$$

$$
\Rightarrow E_{a,out} = E_{b,out} = E_{a \cap b,out} \qquad (4.30)
$$

qed.

Consequently, the entire region of interest has the same irradiance power as if the image is illuminated by an infinite illuminant plane. These regions reconstruct a modified irradiance map $E'(\mathbf{p})$ with the same irradiance equal to $2 \cdot (E_{spl1} - E_{spl2})$. Given the relationship between the irradiance and image as shown in Eqn. 4.1 an output image without shadows is obtained by the modified irradiance map $E'(\mathbf{p})$. Figure 4.16(b) and (c) show the procedure of the algorithm, and the resultant irradiance map is illustrated in Fig. 4.16(d).

Since each area in each input image is illuminated at least with the minimum illumination level E_{spl2}, the difference between E_{spl1} and E_{spl2} can be used to adjust the optical dynamic of the scene $(E_{nir,hi}, E_{nir,low})$ as detailed in Eqn. 4.17 on Page 76. In case one of the supplementary illumination sources E_{spl2} is zero the algorithm can be mathematically replaced by the $max()$ operation and two input images $(max\,(I_a, I_b))$.

The third input image I_{bias} is used for suppressing influences of ambient illumination (offset reduction). If the ambient illumination is negligible the number of input images can be reduced to two by ignoring the third term in Eqn. 4.26. However, the robustness to the illumination change is lost.

Experiments

We have conducted experiments to demonstrate the basic idea of the proposed ShadowFlash algorithm. In our experiments the supplementary illumination is implemented with two identical halogen bulbs, and another halogen lamp is installed for simulating the ambient illumination. The irradiance power E_{spl2} is minimized in order to maximize the dynamic range of the output image. A CCD camera is used for the image acquisition using 640×480 pixel resolution with 8-bit intensity levels, and the positions of both the bulbs and camera are chosen to minimize the overlapped shadow regions.

Figure 4.17 shows one result of our experiments performed with a metallic object on the complex-textured background. Figure 4.17(a) and (b) are the input images with both the supplementary and ambient illuminations from different directions. Some parts of the texture on the background are obscured due to the shadows although the texture is still visible within them. Figure 4.17(c) shows the image illuminated only by ambient light.

The results of the interim stage of the procedure are shown in Fig. 4.17(d) and (e). Figure 4.17(d) shows the composite image of the two input images with the supplementary illuminations. In this step the intensity resolution of the image is temporarily doubled to 9 bits due to the addition process $(I_a + I_b)$. In principle all of the textures in the input images are identical. Thus it is clear that the pixels with an intensity greater than zero after the subtraction process $|I_a - I_b|$ have been illuminated with different irradiance powers in Fig. 4.17(e).

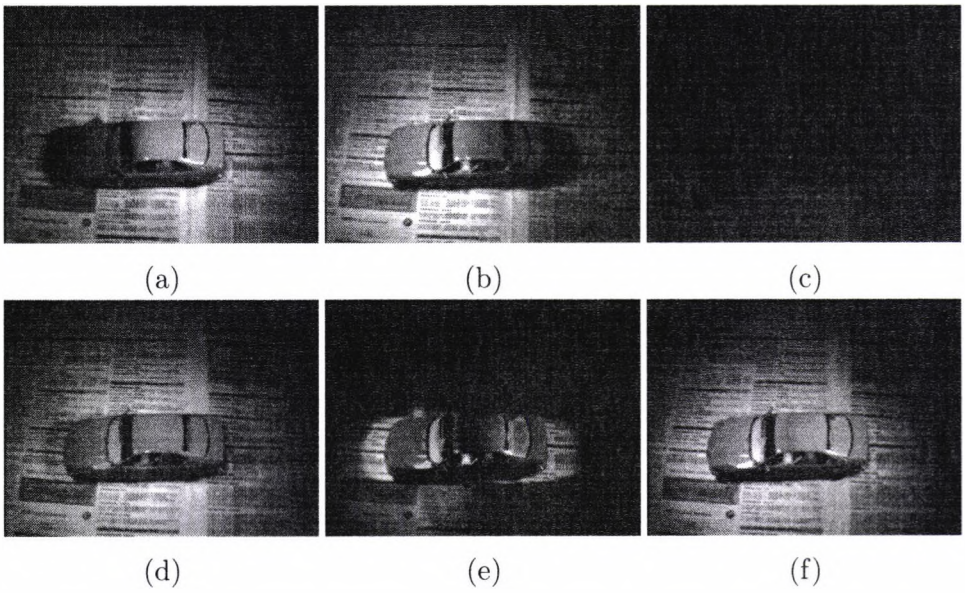


Figure 4.17: Example of ShadowFlash with ambient illumination: (a) input image with the right light source I_a, (b) input image with the left light source I_b, (c) input image only with ambient illumination I_{amb}, (d) $I_a + I_b$, (e) $|I_a - I_b|$, and (f) the result of ShadowFlash algorithm I_{out}

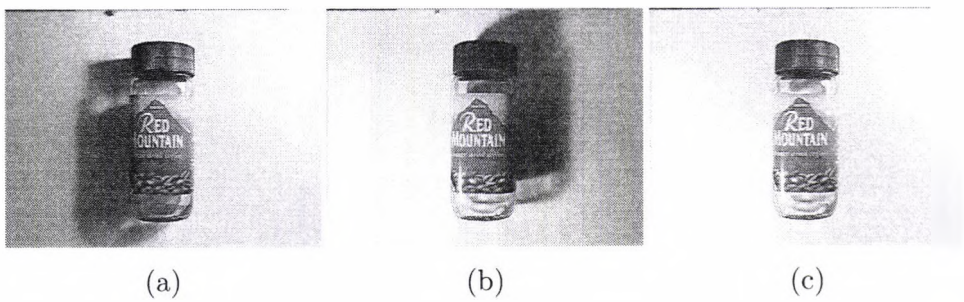
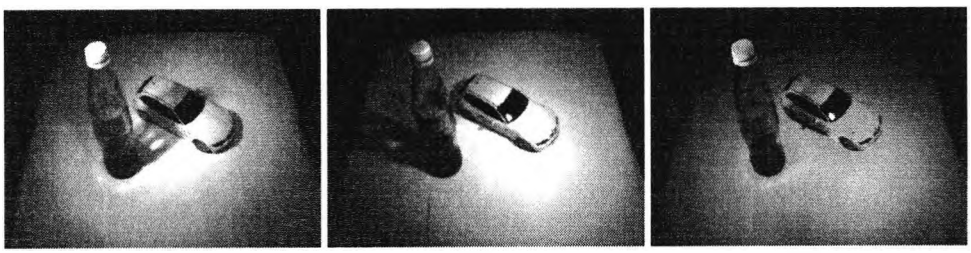


Figure 4.18: Example of ShadowFlash for color images. The images are taken with a CCD with AGC. The ambient illumination in I_{out} is not actually suppressed since I_{bias} is assumed to be negligible: (a) I_a, (b) I_b, and (c) I_{out}



(a) (b) (c)

Figure 4.19: Error case due to the uneven distributed illumination: (a) I_a, (b) I_b, and (c) I_{out}. Shadows remain in I_{out} near the right sideview mirror and at the car roof.

Fig.4.17(f) shows the result image of the ShadowFlash algorithm. The shadows which have covered the background are successfully removed, and the patterns of the background are completely restored by simulating the illumination from an infinite illuminant plane. The intensity resolution of the image has doubled in the result of the addition phase. However, the frequency per every two intensity levels has the value zero because another addition process of the algorithm makes all the intensity values turn into even numbers in the final step. Therefore the intensity resolution can be compressed to 8 bits again by eliminating the lowest bit, *i.e.* $9 \rightarrow 8$ bits.

Note that the ShadowFlash algorithm suppresses cast and self shadows only. The dark image border in Fig.4.17(f) is caused by areas which are not within the scope of the supplementary illumination ($E_{a \cup b,out}$) and not due to cast or self shadows. Since no information is obtainable to restore the original texture due to the offset reduction scheme the equalization task for these areas ($E_a = E_b = E_{a \cap b} \neq E_{(a \cup b)^c}$) is not considered in our approach and removed in the final output image:

$$\begin{aligned} E_{a \cup b,out} &= |(0 + E_{amb}) - (0 + E_{amb})| \\ &\quad + (0 + E_{amb}) + (0 + E_{amb}) \\ &\quad - 2 \cdot (0 + E_{amb}) \\ &= \mathbf{0 \neq E_{a \cap b,out}} \end{aligned} \tag{4.31}$$

Another example is shown in Fig. 4.18. These color images are obtained by a color CCD camera with both the Auto Gain Control (AGC) and Gamma correction functions. Since the shadows cast by the ambient illumination are not visible (or very weak) in the input images due to the effect of the non-linear intensity compression the algorithm could be modified to: $I_{out} = |I_a - I_b| + (I_a + I_b) \approx max(I_a, I_b)$. Consequently this results in the ambient illumination is not being suppressed in I_{out}.

The radiance emitted by each illumination source as defined in Eqn. 4.26 is considered evenly distributed, which is obviously not true for most of illumination sources. Hence employed illumination sources must emit a nearly even radiance distribution (*e.g.* a sharp focused spot or sunlight), or the quality of shadow suppression decreases towards the surface areas which are illuminated by the light source boundary. This is shown in Fig. 4.19. The shadows are not completely removed because the irradiance power of the supplementary illuminations is not evenly distributed over the field of view. The self-reflection caused by a large constant reflection surface of the objects is another reason. The overlapped shadow between those objects (the shadow of the right sideview mirror and the shadow at the car roof) still remains due to the limitation of the proposed algorithm.

Limitations and must conditions

If the optical dynamic of the scene does not exceed the dynamic of the imager and an offset reduction is not required, two input images with alternating light pulses from at least two light sources are sufficient to suppress shadow effects, see Eqn. 4.25 and Fig. 4.15. However, if the dynamic of the scene is greater than the imager dynamic (HDR scenes) and if an offset reduction is required as introduced in Section 4.2.1, the ShadowFlash approach with at least three input images with varying illumination levels is necessary, see Eqn. 4.26.

Several requirements had to be met in order to obtain the satisfactory results of the ShadowFlash experiments so far:

- For simulating the infinite plane the positions of both light sources are arbitrary but they must not coincide *and* the irradiance on the scene surface must be equal from each light source. To realize equal irradiance on the surface the radiant intensities I^e of the light sources should be approximately equivalent and the distance d from each light source to one point on the surface in the field of view should be also the same. As these two parameters are complementary to each other the adjustment for arbitrarily positioned light sources may be accomplished by modification of the emitted radiant intensities using a voltage control.
- The self-reflection on the surface of an object caused by the supplementary light sources could result in deletion effects on the recovered background. To avoid the undesirable effects the relative positions of supplementary illuminations to both the camera and object should be carefully determined.
- Finally the overlapped shadow region must be minimized.

Discussion and future work

Our work can be extended to the following researches:

- The ShadowFlash algorithm could be extended to consider the existence of more than two illumination sources to minimize overlapped shadow regions.
- With the positions of the supplementary illumination and camera being known our task will be directly applicable to the *photometric stereo method* in order to solve the problem of reconstructing the 3-D shape of an object.
- Further experiments will be performed on video sequences. Especially we can reduce the number of input images for each shadow removed image with a video sequence by selecting any 3 successive frames within the sequentially obtained frame patterns in a time axis (*e.g.* $\cdots, I_b, \underline{I_{amb}, I_a, I_b}, I_{amb}, \cdots$).
- The proposed ShadowFlash approach was designed to fulfill the requirements for image capturing in closed environments, *e.g.* a motor vehicle interior. Another application field for our ShadowFlash approach is image capturing for biometric purposes, i.e. face recognition in open world conditions. A ShadowFlash illumination unit synchronized with a camera of a cash dispenser or other access authorization systems would significantly reduce the image processing cost and would increase the recognition robustness, if the image acquisition is influenced by ambient light and a separate and closed compartment is not available due to space, cost or design constraints.
- It could also be challenging to substitute the illumination sources for other signal types such as a sound or electric wave. Application in medical ultrasonic screening, a radar system, or an industrial quality control system could be good examples.

4.4 LineFlash

By comparing two subsequent frames within an image stream it is usually assumed that the capture time for each image is short enough to neglect scene differences between the input images. Scene differences between two subsequent digitized frames may be caused by sudden light changes or fast movements within the scene. Significant scene differences yield erroneous computing results in many cases, for example if an offset reduction or shadow compression algorithm is applied as discussed in previous sections. Therefore the need of employing successively illuminated frames with different intensities is the weakest point of active illumination algorithms for real-time applications. This section describes a novel approach for illuminating a scene by a flash light, contrary to established active lighting triggered by the horizontal line trigger of a digital camera, to solve this problem.

In case of motor vehicle applications frequent and/or sudden light changes inside the car occur due to varying ambient illumination but also due to

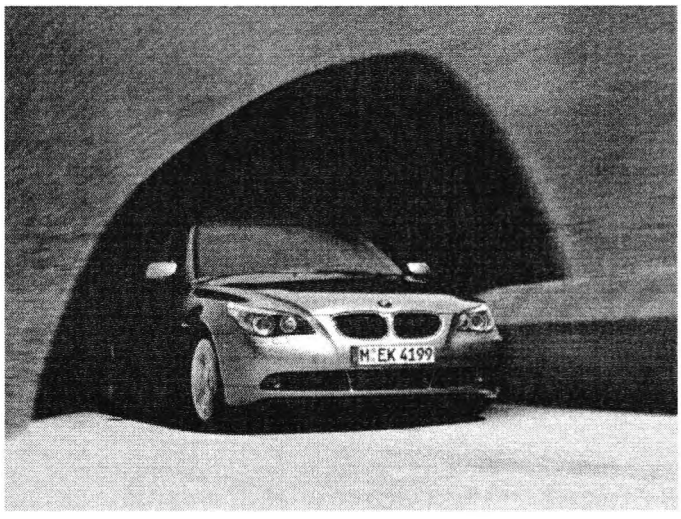


Figure 4.20: Example of sudden light changes due to movements of the vehicle.

movements of the vehicle. Sudden light changes may be caused, *e.g.*, by entering a tunnel, see Fig. 4.20. Accordingly, the vehicle is not completely shadowed during a snap, yielding high contrast inside the image. Assuming a car which drives into a tunnel at a city speed of $50km/h$ it takes $144ms$ to shadow the whole interior length of approx. $2m$, which means at least several frames.

Frequent light changes may be caused by driving through a tree alley. The relationship between vehicle speed v_{car}, distance s_{obj} between shapes which cause cast shadows and the resulting interference frequency f_{shadow} is defined as

$$f_{shadow} = \frac{v_{car}}{s_{obj}} \tag{4.32}$$

According to Fig. 4.21 driving at a vehicle speed of $v_{car} = 120km/h$ in conjunction with an object distance of $s_{obj} = 1.5m$ causes an interference frequency of approx. $22Hz$, which is within the frame rate range of mainstream cameras. To avoid constant interaction with external illumination frequencies, such as those occurring during a drive through a tree alley, it makes sense to use a slightly varying frame rate of the imager rather than a fixed one.

4.4.1 Description

To solve the problem of scene differences between subsequent input images we propose a technique where a tuple of n lines of the image ('LineFlash')

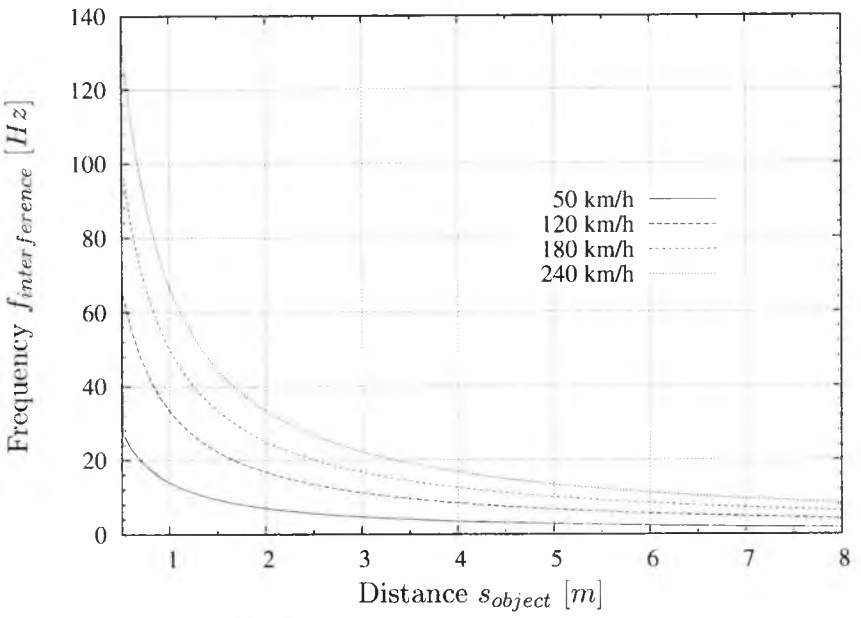


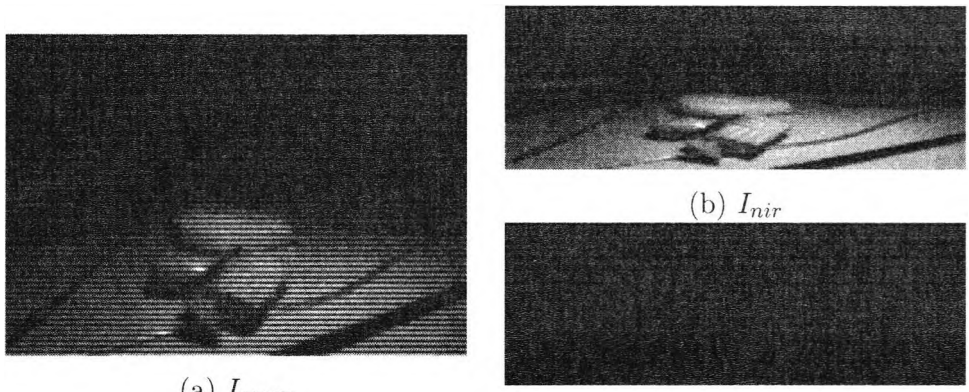
Figure 4.21: Relationship between object distance and interference frequency

is illuminated contrary to established flashing of complete frames. This approach was published in [37].

About 98% of the cameras produced world-wide, CMOS or CCD based, employ a rolling shutter approach for realizing the sensor exposure and read-out process, see Section 3.3. The amount of integrated photons and therefore the image intensity depends on the defined time between the continuous incremental reset and read-out of sensor lines. This integration time is a multiple of k line clocks (horizontal sync) provided by the sensor control unit with frequency f_{hsync}.

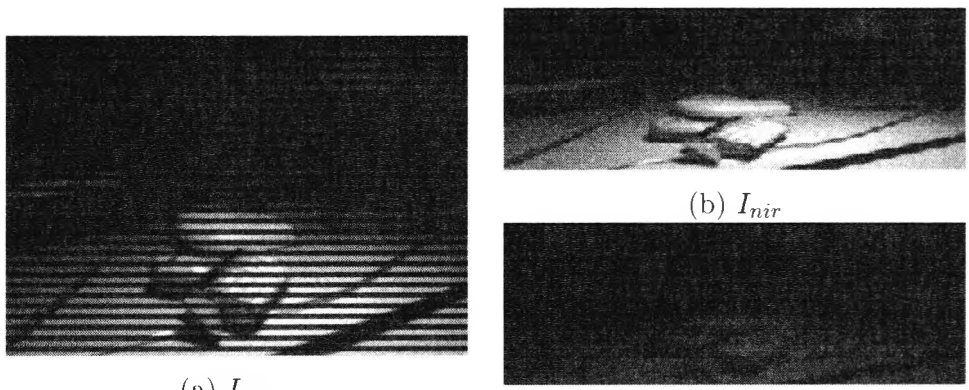
There are three options for controlling the sensor sensitivity: Varying the number of k line clocks before read-out (integration time), varying the line clock frequency f_{hsync} or varying the electronic gain g_{cam} of the read-out circuit. A sensor with rolling shutter provides a maximum sensitivity to incident radiation when k is equal to the number of available lines of the sensor.

By using flashed lines instead of full frames the vertical resolution of the image sensor is divided at least by two because the composite input frame I_{comp} includes the flashed I_{nir} and not flashed I_{amb} image of the digitized scene. The horizontal resolution of the output images stays unchanged. The basic idea of this approach is shown in Fig. 4.24.



(a) I_{comp} (b) I_{nir} (c) I_{amb}

Figure 4.22: Line flash example for $k = 2$.



(a) I_{comp} (b) I_{nir} (c) I_{amb}

Figure 4.23: Line flash example for $k = 4$.

4.4.2 Experiments

Two examples of the line flash approach are shown in Fig. 4.22 and 4.23. Figure 4.22(a) shows the different illuminated composite image I_{comp} for a tuple of two illuminated lines by a NIR flash light and two lines with ambient illumination only ($k = 2$). The CMOS sensor (see Section 3.6) employed here provided a spatial resolution of 648 by 488 pixels (VGA) operating in continuous rolling shutter mode. The line clock (horizontal sync) has been divided by k by a simple cascade of toggle flip flops (T-FF) to control the supplementary NIR flash lights.

The electronic gain g_{cam} of the sensor was increased to $\approx 3dB$ due to the short integration time resulting in significantly increased sensor noise. Figure 4.22(b) and (c) represent the separated output images of the flashed

(I_{nir}) and not flashed scene (I_{amb}).

Fig. 4.22 shows the same scene with identical supplementary illumination but with an illumination tuple of $k = 4$ flashed and not flashed sensor lines. This results in an increased brightness within the output images I_{nir} and I_{amb}. However, stronger digitization effects are visible due to the reduced spatial vertical resolution. This effect determines the upper limit of k depending on the available number of sensor lines.

4.4.3 Summary

By flashing a tuple of k lines of an image sensor rather than illuminating a tuple of frames it is possible to minimize scene changes between two input tuples for active illumination approaches.

Due to the decreased spatial resolution of the output image this approach is more effective for high resolution sensors than image sensors where a high vertical resolution is demanded for detecting scene details. Furthermore bright scenes are preferred due to the reduced sensor sensitivity caused by the limited exposure time (number of line clocks between reset and readout).

4.5 Precis

This chapter discussed different illumination techniques for environments where active illumination is feasible. After determining illumination conditions which are typical for high dynamic range environments in Chapter 2, this chapter introduced solutions for handling such environments:

DoubleFlash is an illumination technique where a scene with high dynamic range (*e.g.* vehicle interior) can be captured by commonly available cameras with limited optical dynamic range without sacrificing image detail and with synchronous offset reduction. The grey-level of a pixel within the DoubleFlash output image sequence only varies if the scene changes, but not as a result of fluctuations in the ambient illumination levels. This significantly simplifies any subsequent image processing. Finally the costs of a mass produced camera with limited dynamic range in combination with simple trigger logic for the illumination are less than for a smart HDR camera.

ShadowFlash is a novel shadow removal technique that could be employed in most image processing systems operating in an active illumination environment. With a reasonable number of controllable supplementary illuminations the proposed shadow removal algorithm simulates an infinite illumination plane over the field of view. The achievement of the proposed

method is to successfully remove shadows from a complex-textured scene without distorting the recovered background and without the support of any region extraction task. Other benefits compared to the conventional shadow detection algorithms are the lower computing costs and the improved reliability.

LineFlash describes an approach for minimizing the problem of scene differences between multi required input images. The proposed technique illuminates a tuple of lines of an image triggered by the horizontal line trigger of a digital camera, contrary to established flashing of complete frames.

Figure 4.24: Trigger sources for active illumination. Above: Frame flash controlled by v-sync; Below: Line flash controlled by h-sync

Chapter 5

Segmentation

5.1 Motivation

Image segmentation can be seen as a classification task as defined by Bartneck in [3]: Given a certain feature vector for the entity (here the pixel) it has to be decided to which of several classes this pixel belongs, *e.g.* a scene background or objects which represent the image foreground. In this context the image foreground means all kinds of objects which we want to detect, track or classify, contrary to the background class, which is the image region not providing information for our detection, tracking or classification task. The foreground/background distinction is also a popular approach to reduce the transmission data for video streams.

The detection and classification of objects within a scene could also be achieved by using template matching. Krumm and Kirk introduced in [45] a video occupant detection system for airbag deployment. They used monochrome images taken by a single camera mounted inside a vehicle and principal components (eigenimages) nearest neighbor classifier. The disadvantage of the principal components analysis (PCA) in this case is that the appearance of an occupied seat is so variable. Thus the detection system was only able to detect a limited set of *RFCS* (see Section 1.3 for more details), that nothing is present or that an object is present which does not fit into the *RFCS* class. Furthermore a drawback of basic template matching is that the computing costs are significantly high and that it depends on the illumination environment. A comparison of these basic methods is shown in Table 5.1.

Computation barriers have in the past limited the complexity of real-time image processing applications. The consequence was that many systems were either too slow to be practical or have been limited to very controlled situations. Recently faster processing units (see Section 6.5) have enabled the industry and researchers to consider more complex and therefore (usually) more robust models for real-time analysis of image streams. These new technologies and processing possibilities allow us to expand the formerly very controlled environments towards real-world situations with varying conditions. But even today it is not possible to perform every image processing operation in real-time, such as template matching for large areas.

An example is the vision based occupation sensor as introduced by Owechko in [59], where the template matching module requires approximately 2 seconds to update its decision on a PC based system running at $400MHz$. Thus extended template matching is not suitable for an embedded real-time system with limited hardware capacity today and in the near future. Due to this finding this work focuses on motion based segmentation to fulfill the real-time aspect as claimed in the title.

A robust image processing system has to be able to deal with movement through cluttered areas, objects overlapping in the scene, shadow effects,

	Motion based segmentation
+	Fast implementation possible
+	Flexible concerning unknown objects (*i.e.* templates)
-	A couple of image may lead to misclassifications (*e.g.* shadows, reflections (see Section 4.3.4))
	Template matching
+	Low error rates
-	High computing costs
-	Not flexible concerning unknown templates
-	Inflexible to illumination variations

Table 5.1: Feature overview of template matching and motion detection.

illumination changes, slow-moving or stopping objects, and objects being introduced or removed from the scene.

Until now methods mainly intended for intensity (grey-scale) image segmentation have been presented. Unfortunately the coherence between intensity changes and objects which may move is often not unambiguous, in particular if the object-background contrast is low. Therefore the master question is whether the intensity change is caused by a moving object (foreground), overlapping the background, or by a spatial illumination change caused by shadow and reflection effects, or due to global illumination changes.

The focus of this chapter is to discuss the implementation of established image sequence segmentation algorithms and their behavior in scenes with significant illumination fluctuations. Furthermore a novel texture based segmentation approach is presented which represents a reasonable trade-off between segmentation accuracy and processing costs for real-time applications in high dynamic range (HDR) environments.

A more detailed comparison of different image segmentation techniques can be found in many publications, such as by Toyama in [86] or Pal in [60] or in several established text books [4, 13, 26, 31, 81] and is not the subject of this work.

5.2 Motion based segmentation

Every three-dimensional motion in a scene is represented in a two-dimensional array by a motion field [81]. If an object in the scene moves relative to this projection surface the two-dimensional projection of that object moves within the image. The human visual system is very sensitive to motion and we tend to focus our attention on moving or altering objects. That is why a flashing traffic light causes more attention than a static one.

Motionless objects are not as easily detectable in a scene and several camouflage strategies of the animal world rely on that fact, *e.g.* the detection system of a fly. Fortunately in many object recognition applications which have to be solved by machine vision the objects of interest (foreground) are moving, whereas the background is static or can be stabilized, just as the interior of a motor vehicle. Motion segmentation can enormously simplify subsequent object recognition steps, therefore detection and segmenting moving objects in a static scene are an important computer vision task. In the following sections we will provide a brief introduction to the problem of extracting, representing and analyzing objects within image sequences using their motion. A more detailed overview of motion-based recognition is given by Cédras in [10].

The motion segmentation approaches can generally be divided into *region oriented* and *pixel oriented* processing. The pixel oriented techniques reach a decision between background and foreground only by the temporal grey-level progression of a pixel and the deduced features [33]. Region orientated motion detection employs more information from a spatial filter applied to a higher level of resolution, such as the pixel neighborhood or statistics deducted from the entire frame. Spatial motion segmentation tends to be more robust to distinguish between intensity changes due to object motion or global illumination fluctuations.

However, the major benefit of pixel based motion detection is its speed. That is why pixel based motion segmentation is still the state-of-the-art for industrial applications. Every pixel value can be processed directly after capturing. CMOS cameras provide this ability to read-out single pixels, and it is not necessary to wait for capturing a row or the whole frame (see Section 3.3). This feature in combination with pixel based filters means that the segmentation can be finished shortly after read-out of the last pixel. The time to capture the frame is not wasted by waiting but can be used for simple preprocessing of the picture, which is a good example of hard- and software fusion for image processing.

5.2.1 Requirements

Image segmentation in environments characterized by large light fluctuations is a challenging task compared to image segmentation with stable

illumination conditions. The goal of applications for such environments is in many cases any kind of surveillance, *e.g.* building surveillance or motor vehicle interior analysis. A surveillance task is defined as reliable detecting, classifying and tracking of objects in a defined area.

Object classification uses a vector of object features to calculate its affiliation to a member from a number of trained groups, see Section 6.3. Therefore a reproducible object segmentation which is not affected by object behavior or light fluctuations is a key requirement for the subsequent classification task. In many cases motion within image sequences yields sufficiently helpful hints for foreground and background discrimination.

The following section gives an overview of established motion based segmentation approaches that are suitable for real-time applications. The discussed approaches were compared in a number of test sequences showing an indoor scene which suffers from strong light fluctuations. A number of key frames (**kf1** ... **4**) have been extracted from the sequences to document their behavior under the following conditions:

Sleeping passenger [**kf1**]**:** Segmentation which employs only motion detection is not able to distinguish a motionless foreground object from the background. However, it is necessary for the majority of surveillance applications that objects or parts of objects should not be considered as part of the image background if they keep their position for an arbitrary time.

Foreground aperture [**kf2**]**:** Many classification approaches use the total area of a blob as a reliable object feature. Hence it is preferable that the detected object is represented by a full blob (correspondence problem) and not only by its motion borders. This problem occurs, *e.g.*, if the foreground object has a homogeneous texture.

Bootstrapping Every surveillance system has to be started before it runs continuously for a longer period. Each system start (boot) needs an initialization time for adapting the algorithm to the current scene. Some systems may even operate only periodically for a short time, *e.g.* a car interior surveillance. There are 'time gaps' since the imager will provide data from unlocking the vehicle (system 'wake up' or system 'start' respectively) and locking it again. After several minutes of inactivity, the vehicle systems will go into stand-by mode (system 'sleep'). Therefore the question is whether it makes sense to adapt stable objects to a reference background or not. Therefore a robust approach has to deal with scene changes which may occur during two system starts, *i.e.* when the system was turned off and proceeds to stand by. Should information from the former run be used to improve the next pass? When should the initial background be captured?

Foreground ratio: Some applications are characterized by objects in the scene which cover a significant portion of the background, *e.g.* vehicle

interior surveillance. If an object enters the vehicle most of the background, *i.e.*, the interior, is hidden. A large person occupies up to 50% of the image area depending on the camera position and optics used. Hence up to 50% or more of the background might be occluded by the foreground. This presents problems for motion detection and background updating algorithms which often rely on global background statistics.

Sudden light variation [kf3]: Most surveillance tasks are characterized by environments which suffer from light fluctuations such as smooth light variations over time or sudden light changes. The segmentation algorithm should create object shapes which are not significantly influenced by lighting conditions.

Shadows [kf4]: Shadow effects within a scene may erroneously result in foreground classified as background areas. An ideal segmentation approach should not be significantly influenced by intensity differences caused by shadow effects, see Section 4.3

Speed: Many surveillance applications require a real-time response and are finally implemented in an embedded system, *e.g.* motor vehicle occupant detection. Hence the computing costs for image segmentation algorithms are not negligible.

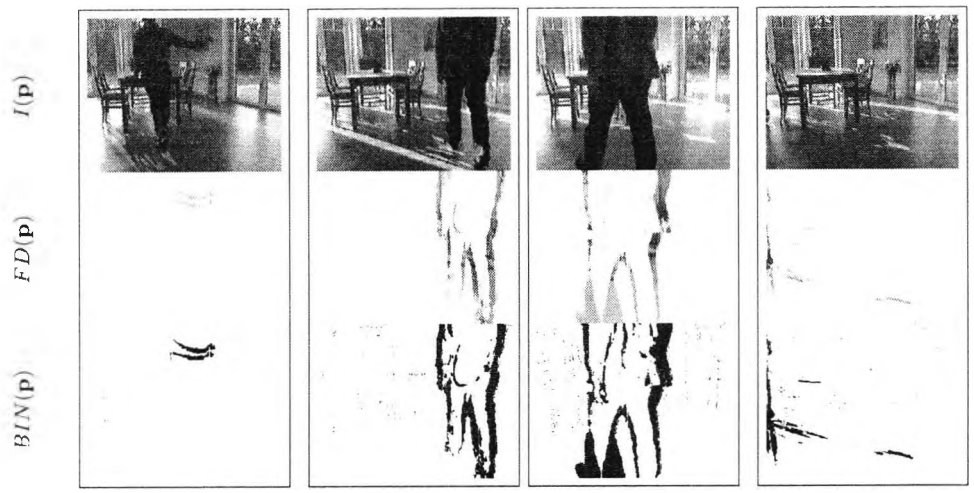
5.2.2 Difference frame technique

Adjacent frame difference

A fast and simple approach to detect motion within an image sequence $I(n)$ is to compute the arithmetic difference of two pixels representing the same given physical location $\mathbf{p}$ in two consecutive frames. It is supposed that an analysis of the difference FD will emphasize pixels which have changed their grey level due to motion [40].

$$\begin{aligned} FD(\mathbf{p}) &= |I(\mathbf{p}, n) - I(\mathbf{p}, n-1)| \\ \mathbf{p} &= (row, column) \end{aligned} \tag{5.1}$$

Assuming a stationary camera and constant illumination for the image sequence with no moving objects no differences FD will be detected. In reality it is most probable that there will never be a zero difference image due to sensor noise during the image digitization process. The major shortcoming of this approach is that objects which stop moving will obviously no longer be detected because they cause no difference between consecutive frames as shown in Fig. 5.1(a). Only the moving arm of the person is detected, the motionless body of the person within the scene is invisible. Therefore the determined threshold th which separates intensity changes due to motion and due to noise in a binary image BIN has to be well tuned and is very



(a) kf1 (b) kf2 (c) kf3 (d) kf4

Figure 5.1: Key frames 1...4 for adjacent frame difference based segmentation.

sensitive to the current scene to be observed.

$$BIN(\mathbf{p}) = \begin{cases} 1 & \text{if } FD(\mathbf{p}) > th \\ 0 & \text{otherwise} \end{cases} \tag{5.2}$$

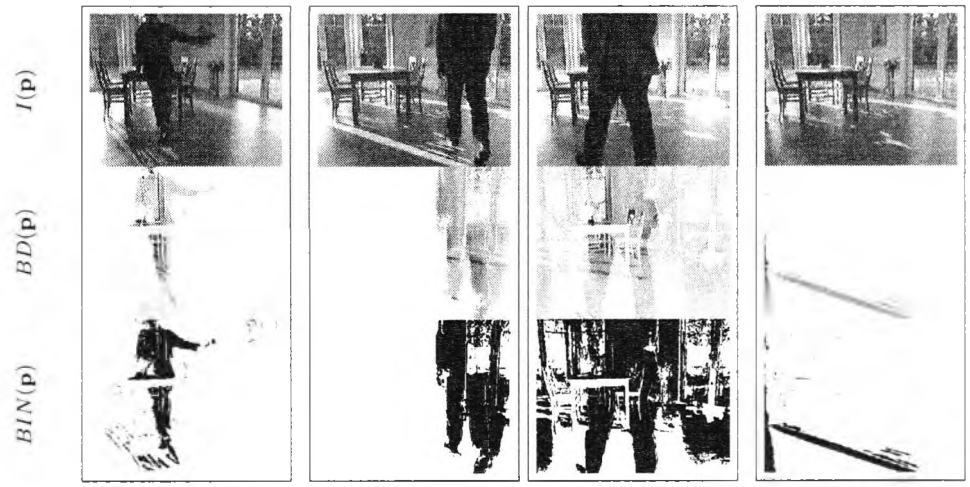
A similar effect is caused by objects without significant texture. Shifting a homogeneous plane causes no difference between two images, only the borders will be detected. This is shown in Fig. 5.1(b) and (c) showing a person with less textured clothes.

Static frame difference

To overcome the problem of detecting motionless objects within the scene a modelled background can be used to compute differences between a reference frame (background) RB and the current frame I. A standard method of generating a reference background is averaging several input frames without moving objects over time, creating a background approximation which is similar to the current static scene except where motion occurs.

$$BD(\mathbf{p}) = |RB(\mathbf{p}, n-1) - I(\mathbf{p}, n)| \tag{5.3}$$

This is shown in Fig. 5.2. A motionless object is still detected, even if only parts of the object are moving as shown in Fig. 5.2(a) and (b). However, if the illumination conditions change the static frame difference BD yields poor results due to overrated foreground areas as shown in Fig. 5.2(c) where the entering person occludes bright areas of the scene, causing the camera to adjust its automatic gain.



(a) kf1 (b) kf2 (c) kf3 (d) kf4

Figure 5.2: Key frames 1...4 for reference frame difference based segmentation.

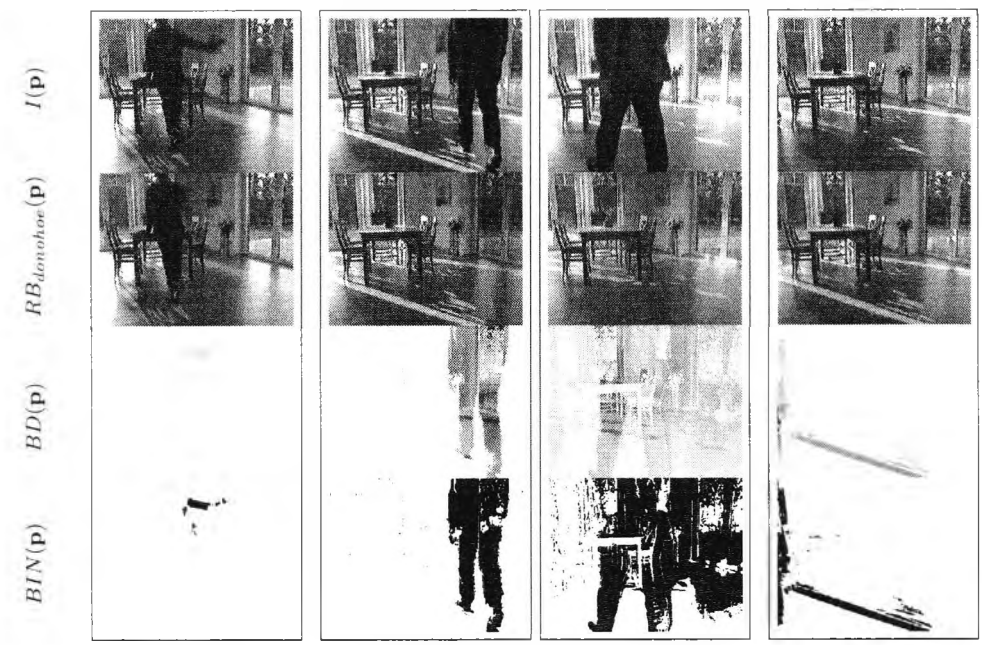
Fig. 5.2(d) shows the person walking by the table casting a long shadow which covers a formerly bright area. These shadows also yield areas in BD which are erroneously detected as foreground objects.

5.2.3 Linear prediction

A possibility of reducing the dependence on brightness changes in background subtracting is to maintain a reference background RB which adapts to illumination fluctuations of the scene by recursive filtering and linear prediction. Most researchers have ceased to implement non-adaptive methods of background subtracting because errors in the background accumulate over time, making this method useful only in highly-supervised, short-term tracking applications without significant changes in the scene. Numerous approaches to this problem have been published in recent years [83] which differ in the type of background model used and the procedure used to update the model. A more detailed overview can be found in Toyama in [86] and Brown in [8].

Donohoe

An example for time recursive updating the background model is described by Donohoe in [16, 78]. An adaption factor α defines how fast the background model RB converges towards the current input frame I. Hence smooth light variations as shown in Fig. 5.3(b) are adapted into the back-



(a) kf1 (b) kf2 (c) kf3 (d) kf4

Figure 5.3: Key frames 1...4 for segmentation based on linear prediction by Donohoe.

ground model, whereas moving objects are detected completely.

$$RB_{donohoe}(\mathbf{p}, n) = (1 - \alpha)RB(\mathbf{p}, n-1) + \alpha I(\mathbf{p}, n) \qquad (5.4)$$
$$0 \leq \alpha \leq 1$$

This approach is based on the assumption that a pixel belongs to the background if the local image luminance changes slowly in time [7]. However, this is effective in situations where objects move continuously and the background is visible a significant portion of the time. It is not robust to scenes with slow moving objects and sudden light changes. Furthermore motionless objects are still adapted to the background at times and therefore disappear (Fig. 5.3(a)). Sudden illumination changes and shadow effects are erroneously considered foreground elements (Fig. 5.3(c) and (d)). Furthermore it recovers slowly if the background is uncovered by a moving object which was adapted to the background model due to the single predetermined threshold th for the entire scene.

Park

An advanced time-recursive background model was introduced by Park who employed a second adaption coefficient β in [63, 61] to find a tradeoff between

Figure 5.4: Key frames 1...4 for linear prediction based segmentation by Park.

fast adaption to illumination changes and slower disappearance of motionless objects. The mask BD contains the moving objects and their locations in the previous image which are only very slowly adapted by β contrary to non-object areas which are considered background and will therefore be adapted faster by α.

$$RB_{park}(\mathbf{p},n) = \begin{cases} (1-\beta)RB(\mathbf{p},n-1) + \beta I(\mathbf{p},n-1) & \text{if } BD(\mathbf{p},n) > th \\ (1-\alpha)RB(\mathbf{p},n-1) + \alpha I(\mathbf{p},n-1) & \text{otherwise} \end{cases} \quad (5.5)$$
$$\beta \ll \alpha$$

Compared to an approach with single adaption coefficient this means an improved accuracy at varying illumination but still no ability to handle sudden light changes or shadow effects. The system also recovers slowly if a portion of the background is uncovered while the global illumination of the scene has changed.

Kalman prediction filter

Time recursive filter techniques as discussed in Section 5.2.3 try to model background changes by employing one or two fixed adaption coefficients for the whole image. An extension of this idea is to model each pixel as a

separate Wiener filter for dividing the image model into a background and foreground class.

This can be done by modelling a Kalman [33] filter which predicts the next step of the model by adapting the adaption coefficient regarding the previous behavior of the model, in this case the recent input images. This prediction is evaluated and controlled by its least square error of the difference from predicted to real next model value. This technique was employed by Ridder, who in [69] modelled each pixel of his reference background with a Kalman Filter, creating an increased robustness to lighting changes in the scene. Another example was presented by Koller, who has successfully integrated this method [44] in an automatic traffic monitoring application.

The major benefit of the basic Kalman Filter approach is a very robust prediction of continuous models, *e.g.* calculation of rocket trajectories. Recent published experimental results of image segmentation by employing a Kalman filter [61] yield improved foreground discrimination compared to basic linear prediction as introduced by Donohoe or Park, but with significantly increased computing costs and suffering from similar shortcomings such as erroneous foreground / background discrimination due to sudden illumination changes or shadow effects. Furthermore due to the implementation of a pixel-wise automatic threshold it still recovers slowly and does not handle multimodal backgrounds.

This work focuses on real-time approaches for embedded systems, therefore Kalman and other Wiener prediction filter will not be discussed in any further detail.

5.2.4 Gaussian mixture model

Linear prediction approaches as discussed in previous sections are based on the assumption that the state of the background model is changing slowly compared to foreground objects, resulting in a significant deviation from the background model. Unfortunately this condition is not valid in every environment, in particular not in open world scenes with less contrast. Some parts of the background may not show constant intensities, such as trees with moving leaves or a lake surface, but they are still considered background.

Stauffer introduced a segmentation approach in [83] which can overcome this disadvantage using pixel dependant linear prediction filters. He employed a multi-class statistical model for the tracked objects where each pixel of the reference background is modelled as a set of N Gaussian distributions [83, 91]. The probability of observing a grey level is defined as

$$P(\mathbf{f}_n) = \sum_{i=1}^{N} \omega_{i,n} G\left(\mathbf{f}_n, \mu_{i,n}, \Sigma_{i,n}\right) \tag{5.6}$$

where G is the Gaussian probability density function

$$G(\mathbf{f}_n, \mu_n, \Sigma_n) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{f}_n - \mu_n)^T \Sigma^{-1}(\mathbf{f}_n - \mu_n)} \tag{5.7}$$

described by its mean vector $\mu_{i,n}$ and covariance matrix $\Sigma_{i,n}$ at distribution i and frame n. Stauffer recommends a number of $3 < N < 5$ distributions to limit the necessary memory and computing costs for updating the model.

After an initialization period during which the scene is empty the statistical model of the background (the initial Gaussians) will be trained by spatial statistics over a local region. For each new frame after initialization every new observation $\mathbf{f}_n$ is checked against the N established distributions. If the current pixel value fits to one of the established N distributions the matched distribution will be updated by

$$\left.\begin{aligned} \mu_{i,n} &= (1-\varphi)\mu_{i,n-1} + \varphi \mathbf{f}_n \\ \sigma^2_{i,n} &= (1-\varphi)\sigma^2_{i,n-1} + \varphi |\mathbf{f}_n - \mu_{i,n}|^2 \end{aligned}\right\} \text{if } |\mathbf{f}_n - \mu_{i,n-1}| < c\sigma_{i,n-1} \tag{5.8}$$

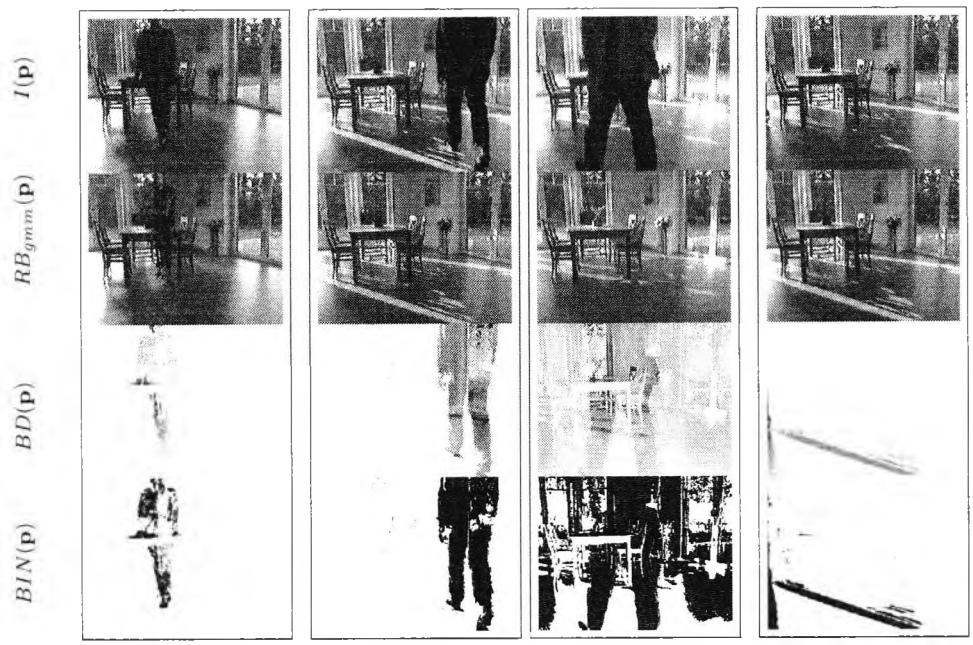
where the estimation weight $\omega_{i,n}$, which reflects the probability that the distribution belongs to the background, will be increased. The learning rate is controlled by the factor φ which determines how fast the distribution will be adapted towards a stable pixel value. This is similar to the update rate for linear prediction approaches.

If a pixel does not fit one of the established distributions a new distribution will be created if there are less than N distributions so far. Otherwise the distribution with the least background probability $\omega_{i,n}$ will be removed from the model. The distribution which comprises the highest probability value is considered as member of the current reference background RB_{gmm}.

The major benefit of a GMM is that it is able to adapt to varying illumination situations dependent on the adaptive coefficient φ by simultaneously handling multimodal backgrounds, *e.g.* trees with moving leaves. Furthermore the model of background areas which have been occluded by foreground objects for a longer period is still maintained. This means that contrary to linear prediction approaches the previous background can quickly be restored when it is revealed again. Therefore the number of pixels which are erroneously classified as foreground objects is significantly reduced.

An extended Gaussian mixture model (GMM) for monochrome and color images has been successfully implemented for outdoor surveillance by Xu and Ellis in [91], which also incorporates the ability to detect shadows by using a color chromatic representation of the images.

However, sudden light changes still cause problems if they do not occur frequently, see Fig. 5.5(c). One solution to this problem could be to model each pixel not as a mixture of intensity Gaussians but to model a mixture of region feature (texture) based Gaussians. The examination of this idea



(a) kf1 (b) kf2 (c) kf3 (d) kf4

Figure 5.5: Key frames 1...4 for segmentation based on a Gaussian Mixture Model.

is a topic for further examination and not a subject of this work due to the expected computing costs.

5.2.5 Difference texture technique

An alternative approach which tries to overcome the shortcomings of difference frame techniques and linear prediction approaches is presented in the following section and was published in [39]. It is not affected by light fluctuations nor shadow effects, even in the event of fast light changes. On the other hand it is fast to compute and needs no time to reconstruct background areas which have been adapted to the background due to longer occlusion because we decided to start with a fixed background model which does not change over time (*'master background'*).

Established approaches segment the image first by using a time recursive background that adapts to light changes. The calculated deviation from this model is used for filtering out the shadows afterwards [33, 83].

Rosin and Ellis employed in [72] an uniform local attenuation for detecting semi-transparent regions, *i.e.* shadows. Regions with homogeneous attenuation were detected by region growing and considered as shadows. These detected shadow areas were then subtracted from the foreground seg-

mentation result. This section describes an alternative approach, not looking for regions with similar texture to detect shadows (compared to an adapted background), but to segment everything which has a different texture (compared to a fixed background).

It is not the absolute difference between a master background image MB and an input image I which determines the foreground areas but its texture difference. The frame ratio FR as defined in Eqn. 5.9 is homogeneous for areas which have the same texture but heterogeneous for areas with different surface texture. The offset of 1 avoids undefined divisions, the square increases the SNR.

$$FR(\mathbf{p}, n) = \left(\frac{I(\mathbf{p},n)+1}{MB(\mathbf{p})+1}\right)^2 \tag{5.9}$$

If the texture of the master-background MB and input image I are not the same, this means that the division yields a new texture. This homogeneity of the frame ratio can be analyzed, *e.g.* by computing the local variance of each pixel as a sliding mask operation. Q and R represent the dimensions and $\mathbf{q}$ denotes the pixel location within the sliding mask:

$$FR_\sigma(\mathbf{p}) = \sqrt{\sum [FR(\mathbf{p}+\mathbf{q}) - \mu_{fr}(\mathbf{p})]^2} \tag{5.10}$$

$$\mu_{fr}(\mathbf{p}) = \frac{1}{QR} \sum FR(\mathbf{p}+\mathbf{q}) \tag{5.11}$$

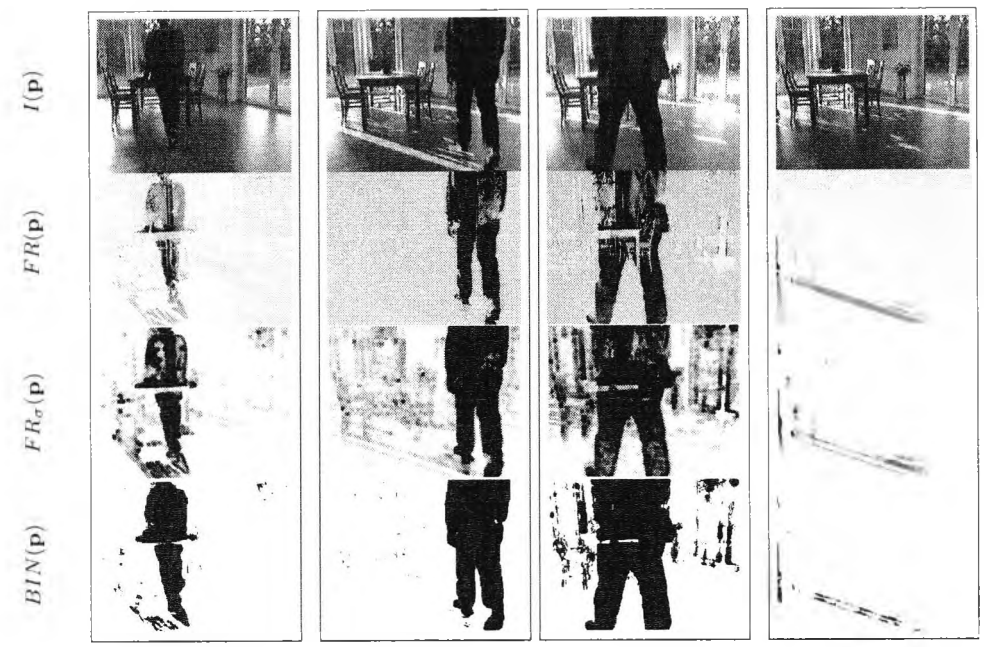
Hence FR_σ represents the intensity variance of the neighboring pixels of pixel $FR(\mathbf{p})$ within a Q by R window.

A global illumination change within the current input image yields a varying frame ratio compared to the master background. However, the local variance is still low, while areas where the texture differs from the interior are still emphasized by an increased local variance. The same is true for shadows or reflections within the image which also represent a spatial gain or attenuation. Hence the frame ratio between a shadowed region yields a constant without new texture. Only shadow borders will be emphasized by the frame ratio but are low-pass filtered due to the mask size Q by R of the local variance. Hence the influence of light variations and shadows is minimized. This is particularly advantageous in the case of sudden changes in illumination situations as shown in Fig. 5.6(a)...(d). Adaptive background approaches are usually too slow to adapt quickly to sudden light changes. In addition the computational costs for background updating, which are not negligible, can be omitted.

5.2.6 Experimental results

The results of the segmentation experiments are summarized in Table 5.2.

The tested real-time segmentation approaches are rated by the number of over- or under estimated background pixels (false positive/negative) when



(a) kf1 (b) kf2 (c) kf3 (d) kf4

Figure 5.6: Key frames 1...4 for segmentation based on texture difference.

applied to the defined key frames **kf1 ... 4** which simulate situations with strong light fluctuations as defined in Section 5.2.1. The ideal object blobs have been segmented by hand as a reference result. Values in percentage express the ratio between false segmented pixels and the total amount of pixels within the image (276x219).

Obviously the number of over- or under estimated background pixels strongly depends on the threshold levels for the binarization. However, Table 5.2 shows that the presented segmentation approach based on texture difference is more robust to sudden light changes and shadow effects (lowest overall error), which are common for HDR environments. Due to these findings the following experiments focus on texture based segmentation.

Another example for indoor scenes is illustrated in Fig. 5.7. It shows our garage with a mixture of complex (bicycles and measurement equipment) and homogenous areas (floor). We used several input images (I_1, I_2) with moving objects (persons) and two different master background images (MB_1, MB_2) as reference for computing FR.

Although MB is defined as a reference background which does not change over time, MB_2 was captured with dimmed room lighting to demonstrate that the algorithm performance does not depend on the initial light distribution of the background. Due to this simulated global illumination fluctuation, the difference of absolute intensity values between the dimmed

Key frame	*Adjacent frame difference*	*Static frame difference*	*Donohoe*	*Park*	*Gaussian mixture model*	*Difference texture*
kf1 → 'Sleeping passenger'						
False positive	368	1237	130	1480	154	2324
False negative	6591	1967	6520	3374	4083	201
Total error [%]	12	5	11	8	7	4
kf2 → 'Foreground aperture'						
False positive	3072	4523	636	739	640	2543
False negative	9497	2570	1267	1183	1323	228
Total error [%]	21	12	3	3	3	5
kf3 → 'Sudden light variation'						
False positive	5063	18794	17213	23414	20103	6693
False negative	14741	3285	4885	3158	3440	703
Total error [%]	33	37	37	44	39	12
kf4 → 'Shadows'						
False positive	1764	3491	4237	3549	3460	663
False negative	47	0	0	0	0	1
Total error [%]	3	6	7	6	6	1
Overall error [%]	17	15	14	15	14	6

Table 5.2: Segmentation results, rated by the number of over- or under estimated background pixels (false positive/negative) for situations as defined in Section 5.2.1. Values in percentage express the ratio between false segmented pixels and the total amount of pixels within the image (276x219).

reference background image and current input image stream was increased. However, the local variance distribution FR_σ was kept stable, resulting in an unaffected segmentation result as shown in Fig. 5.7(d1) and (d2).

The smoothness of the binary blob borders in the resulting images differs depending on the ratio between convolution mask and image resolution. A large convolution mask in combination with small image resolution yields blurred object shapes and *vice versa*.

Adaptive threshold

Thresholding the results of the local variance filter yields the segmented foreground. Algorithms with one fixed threshold followed by filtering to delete isolated points as defined in Eqn. 5.2 have the disadvantage of altering the profile of this object. Thresholds which heavily reduce the background noise are corrosive for the profile while the ones which obtain compact profiles do not completely delete the background noise.

From this it follows that an optimum threshold for the whole image and its texture differences does not exist. Hence for the subsequent experiments we employed an adaptive local threshold as introduced by Brofferio [7] which analyzes the neighboring environment of the current pixel to improve the discrimination between foreground and background objects. In particular the number of areas inside objects which are erroneously considered as background (called 'false gap') decreases significantly when a local threshold is used. These 'false gaps' are the results of small areas, usually within objects, which are characterized by a texture that is locally similar to the master background.

In areas where FR_σ has a small mean value μ_{th}, that is in background zones, a very large variance within the mask U by V is usually due to noise. It should be suppressed using a large threshold. In areas with predominantly large values of FR_σ, which are the ones inside the object, a small variance can reasonably be assumed to be part of a 'false gap'. Hence it should be filtered using a small threshold:

$$TH(\mathbf{p}) = \frac{k_{th}}{\mu_{th}(\mathbf{p}) + 1} \qquad k_{th} = const. \tag{5.12}$$

$$\mu_{th}(\mathbf{p}) = \frac{1}{UV} \sum FR_\mu(\mathbf{p}+\mathbf{q}) \tag{5.13}$$

Another benefit is significantly reduced noise due to the local operation in Eqn. 5.12. Every value of FR_σ which exceeds the local threshold TH is finally considered a foreground object:

$$BIN(\mathbf{p}) = \begin{cases} 1 & \text{if } FR_\sigma(\mathbf{p}) > TH(\mathbf{p}) \\ 0 & \text{otherwise} \end{cases} \tag{5.14}$$

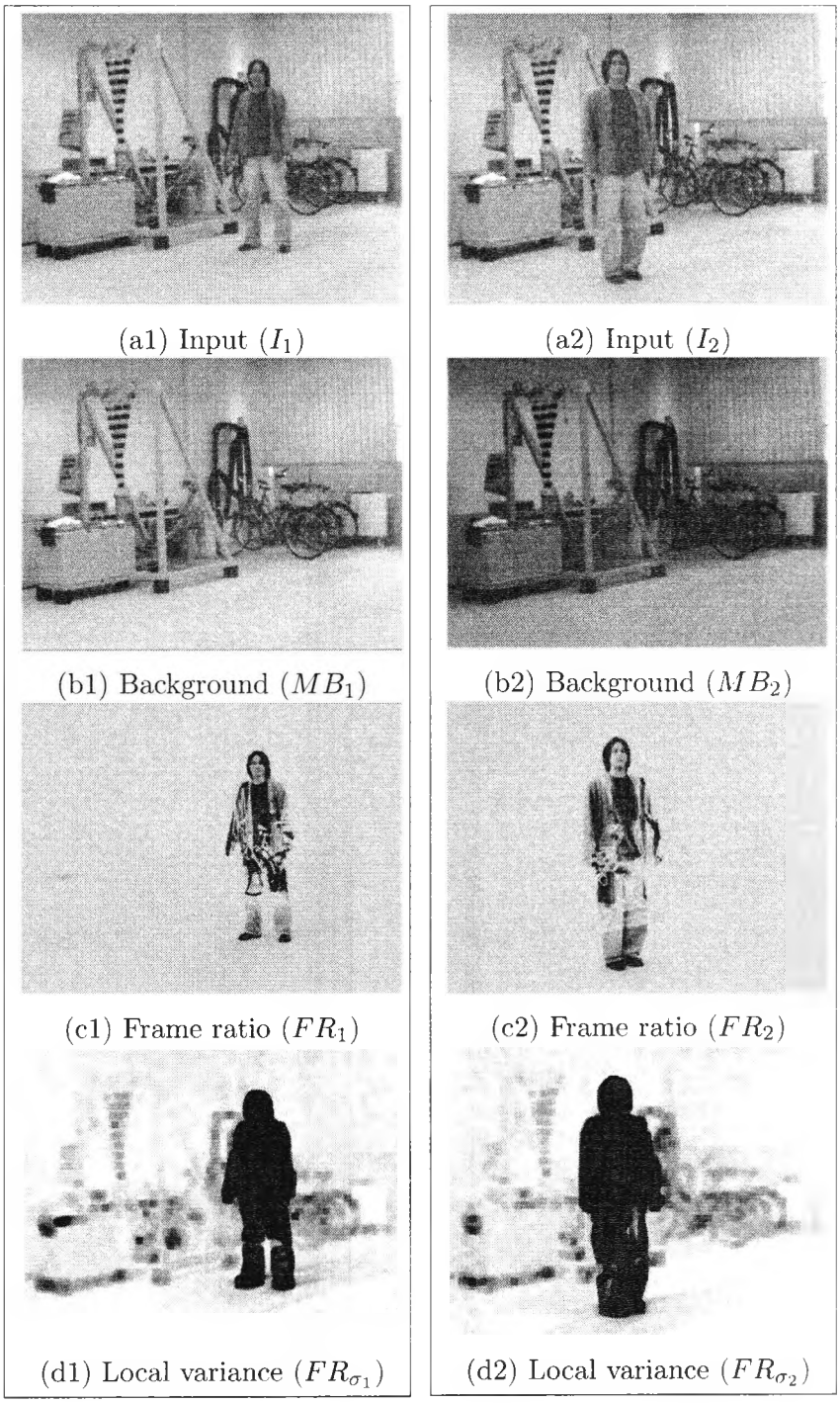
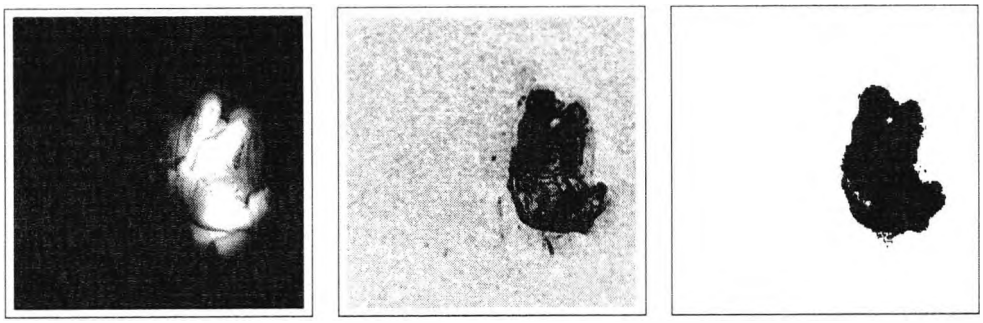


Figure 5.7: Two example image sets from texture based segmentation: Image (a) shows the input image captured by a mainstream CCD camera; (b) bright and darker background; (c) Frame ratio (FR); (d) Local variance of FR: dark pixels correspond to large variance



(a) Input (I) (b) Local variance (FR_σ) (c) Binary result (BIN)

Figure 5.8: Motor vehicle interior example showing a view of the passenger from above: Image (a) shows the DoubleFlash input image; (b) Local variance of FR: dark pixels correspond to large variance; (c) Result of the segmentation step: dark pixels represent detected foreground object(s).

In-vehicle experiments

The following section illustrates our experimental results for texture based image segmentation in conjunction with adaptive thresholding for motor vehicle interior surveillance.

It's not possible via segmentation to reconstruct image detail which is lost during image acquisition, *e.g.* due to an insufficient optical dynamic range of the imager. To handle image acquisition in HDR environments we proposed two solutions: to use a custom design imager with improved dynamic range as discussed in Section 3.6 or to employ active illumination techniques as detailed in Chapter 4.

But even if the dynamic of the scene fits to the camera characteristics, light fluctuations or shadow effects may be still present. The presented difference texture based segmentation approach was designed to especially handle such situations. Figure 5.9 shows the segmentation results of an input image captured by our custom design HDR imager within the NIR.

Figure 4.6 shows the passenger and driver seat of a car captured by using the DoubleFlash approach as discussed in Section 4.2.3. The CMOS camera was equipped with an optical NIR bandpass as described in Section 2.3 which cuts off wavelengths in the visible range. This results in a stable illumination due to the DoubleFlash approach but the intensities may vary slightly over time due to aging and contamination of the lenses of the light sources and the imager. An advantage of texture based segmentation within motor vehicles is the semi-static background which means that the master-background MB could be generated in the plant directly after vehicle production. Hence every boot up yields defined and predictable behavior which simplifies system self test and maintenance. The interior color can

differ from vehicle to vehicle since car manufacturers offer dozens of varying interior sets, while the shape and texture (*i.e.* the background) of the interior components are fixed in general.

To use a master background for image sequence segmentation appears somewhat old fashioned and outdated because the interior may differ slightly over lifetime of the car due to material aging, modification and/or damage, causing the background to be erroneously detected as foreground. However, the benefits of adapting to illumination or background variations using a time recursive background are small in this case when compared to a master background, and the risks are enormous. The final state of a time recursive background which has adapted over the lifetime of a vehicle without maintenance is unpredictable.

To employ only texture information is only reliable if there is texture. This means if a textureless area is occluded by another textureless area it will be detected as background. Hence electronic noise from the image sensor limits the range of texture based image segmentation. This is due to regions of a scene where the brightness is low and the variability of the image structure is caused by noise which becomes dominant. Therefore sufficient texture within all image regions (background *or* foreground object) is necessary and therefore a minimum image resolution is required. Fortunately a car interior is fairly heterogenous and there is always sufficient background texture available for reliable foreground discrimination, even while employing less image resolution.

Our experiments with the texture based segmentation approach have led to very encouraging results which were the basis for the occupant classification system as discussed in Section 6.2. No other image enhancement, morphological or smoothing filters have been applied to receive the blobs as shown in Fig. 5.8(c). The size of the convolution masks depends on the object resolution. Larger masks lead to better results in general, however the shape accuracy decreases and the objects tend to bloom. For the vehicle interior images we employed a 3 by 3 mask for QR and 5 by 5 for UV with an input image resolution of 256 by 256 pixels. The only parameter (k_{th}) for calculating the local threshold mask TH was chosen and tuned by hand. Tests with a varied set of image sequences have shown that k_{th} is quite robust and not as sensitive to the segmentation result as a common fixed threshold.

5.3 Precis

This chapter discussed the usability of motion based segmentation approaches for the special demands which are typical for high dynamic range environments: Sudden light fluctuations and significant shadow effects. Several established motion based segmentation approaches have been

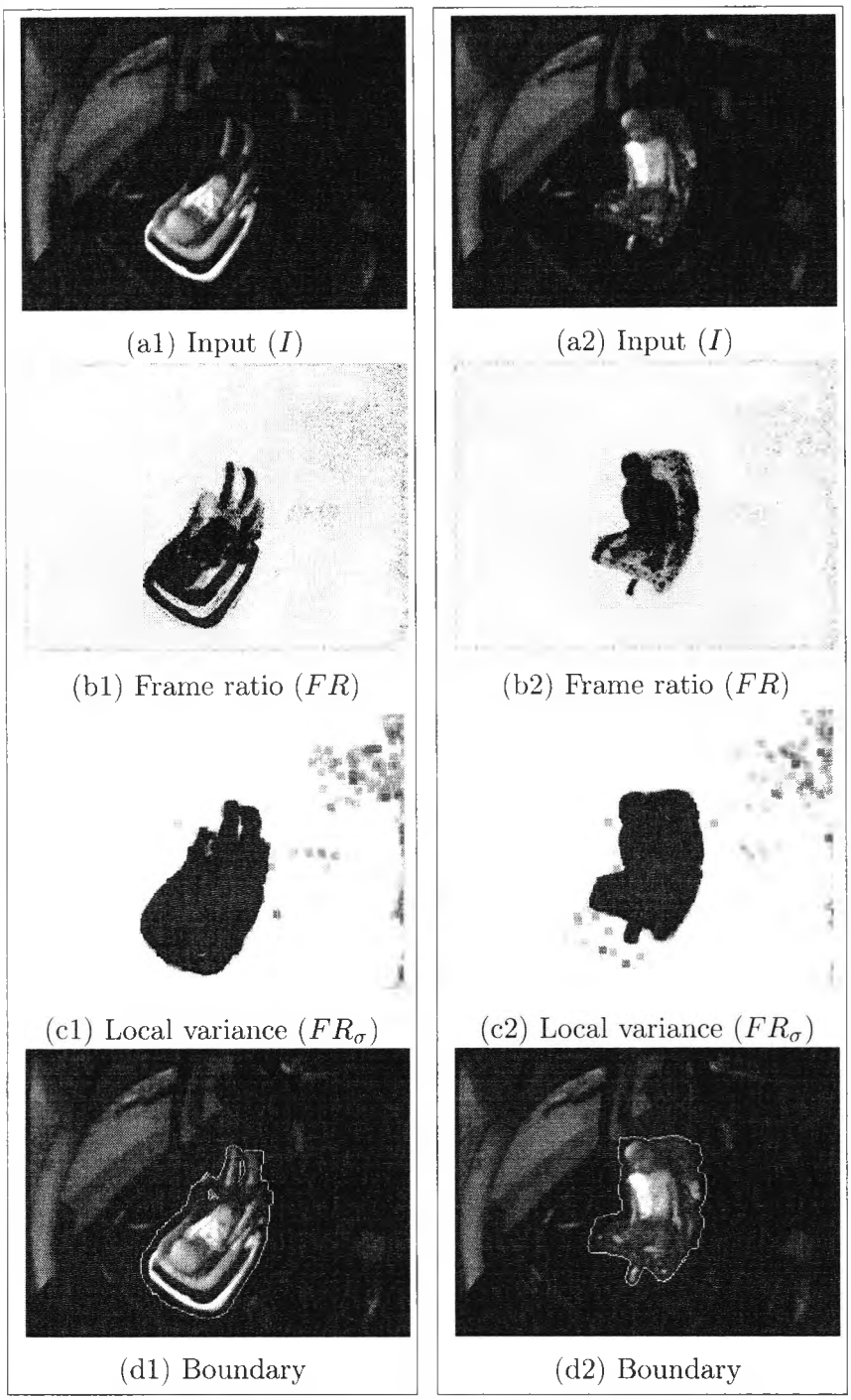


Figure 5.9: Example images from texture based segmentation for occupant detection: Image (a) shows the input image captured by our custom design HDR imager within the NIR; (b) Frame ratio FR; (c) Local variance of FR_σ: dark pixels correspond to large variance; (d) Result of the segmentation step.

tested concerning these demands.

In case of a vehicle interior analysis there are fast moving objects within our region of interest due to the small distance to the camera. Hence time recursive linear prediction filters (see Section 5.2.3) tend to be too slow to reliably estimate the uncovered background. This results in large background areas being erroneously classified as foreground, see Table 5.2. A background modelled by a mixture of Gaussians as discussed in Section 5.2.4 can do so, but the computing costs are significant and its great advantage to adapt to slightly varying illumination provides little benefit in this application.

Another major problem is the absence of available initialization time for an adaptive background and the unknown start condition (bootstrapping, see Section 5.2.1). The system has to provide initial classification results immediately after system 'wake up', but it does not know whether the seat is already occupied or empty. Furthermore after short phases of action (person enters the vehicle) there are longer periods which are characterized by less action. Only parts of the non-rigid object [2] are moving, *e.g.*, legs or arms. Hence for detecting the whole foreground object (the front seat passengers) after long periods without motion they must never be adapted into the background image.

Bright spot lights cause considerable shadow effects. Car occupant detection suffers from this problem due to the necessarily bright supplementary illumination as discussed in Section 4.2. Color variations caused by shadows and reflections within an image yield erroneous segmentation results. Shadows are a spatial problem which cannot be detected accurately at the pixel level. Hence region or frame information is necessary.

We confronted these problems directly in the foreground detection step by using texture ratio information (*i.e.* local variance) instead of raw intensity subtraction for the background and foreground object discrimination. This novel segmentation approach is based on texture difference rather than absolute intensity differences, which represents a reasonable trade-off between segmentation accuracy and real-time demands for embedded systems: Master-background division with subsequent local variance computing eliminates the influence of shadows and illumination changes. Furthermore this master-background almost completely eliminates the sleeping person and bootstrap problem, resulting in an improved reliability for motion based segmentation where the objects within the scene move slowly or stop moving for a considerable period of time, *e.g.* visual occupant detection in cars.

The major shortcoming of a static and/or slowly adapting reference background is that variations of the passenger seat position and inclination are considered negligible or that the seat varies only slowly over time. To eliminate this constraint either extended segmentation approaches have to separate the passenger seat first (*e.g.* by applying edge detection) or the

two-dimensional model database for classification has to be extended to a more flexible three dimensional model, *e.g.*, by using the findings of active illumination for an implementation of photometric stereo. This is subject to ongoing research.

Chapter 6

Implementation

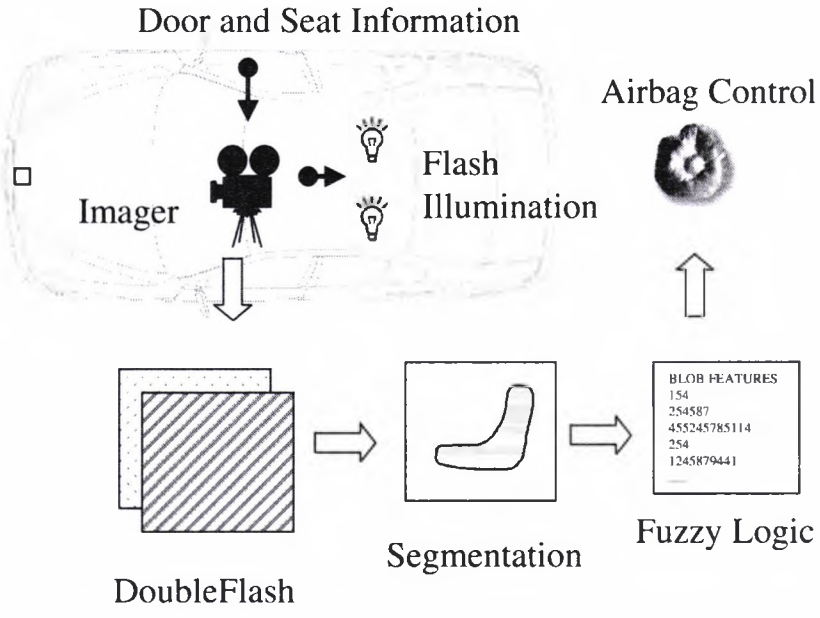


Figure 6.1: Block diagram of the system and an overview of its processing steps.

6.1 Motivation

The majority of research and development of image processing for intelligent vehicles focuses on applications which use the image sensor to analyze the vehicle environment (exterior), such as lane tracking, pre-crash sensing, night vision, pedestrian detection or blind spot detection. In contrast this work focuses on techniques and strategies for tackling the high dynamic range problem for image processing *in* motor vehicles (interior). Examples are listed in Table 6.1.

The aim of this chapter is to demonstrate the real applicability of hard- and software techniques which have been presented in prior chapters by realizing a vision based car occupant detection and classification system for enhanced airbag control. The system has to detect and distinguish car occupants such as children, adults and infant seats and to determine their present location and seating position, especially on the passenger seat. A more detailed description of the system requirements is provided in Section 1.2.2.

The final benefits and limitations of this system have been evaluated by implementing it in several test cars. Images of the latest experiments setup and a system overview are presented in Section 6.2.

Application	Description
Sound optimization	If the spatial position of the passenger seats is known the sound field distribution in the car interior for audio devices (Hi-Fi system) could be dynamically optimized [58].
OOP	Out of Position detection for advanced airbag inflation.
SBE	Sitz Belegungs Erkennung[1], for detection, classification and tracking of passengers for advanced airbag inflation [39, 45, 59], see Sec. 1.
Drowsiness detection	Detection, tracking and analyzing of the driver's eyes. Changing blinking characteristics and steadily reduced spread angle are useful hints for detection of decreasing driver attention and increasing drowsiness. 30% of accidents are caused by overtired drivers.
Climate control	The ventilation can be optimized if the current setting of the ventilation nozzles is determined by image processing. Furthermore analyzing the passenger clothing status can also be used to optimize the air distribution [61, 25].
Anti-theft system	Replacement of the state-of-the- art ultrasonic intruder sensors which feature high error rates because they can not distinguish between motion directions and object size.
Back seat surveillance	Displaying the back seats on a monitor that is visible for the driver without moving his head. This decreases the possibility of accidents due to parents which try to survey their children or infants on the back seats.
Post crash notification	The so-called 'Golden Hour' after an accident is the most effective time for Emergency Medical Services (EMS). First estimates of the number of lives that could be saved by an automatic crash notification system are 3000 lives per year [20]. A system should indicate the precise location of the traveller in trouble but also the number of passengers in need of help. Images from the car interior automatically transmitted together with an emergency call can provide helpful information of the crash severity and number of injured persons.

Table 6.1: Application examples of image processing in motor vehicles.

6.2 System overview

An overview of the test system for visual occupant detection is given in Fig. 6.1, which shows the abstract image processing chain from image acquisition to final airbag control. The system can be roughly divided into seven parts:

1. Active lighting
2. Image acquisition
3. Image pre-processing
4. Segmentation
5. Blob coloring
6. Feature extraction
7. Classification

For our experiments we used a stationary, monocular camera with a fixed focus which was mounted within the car roof at different locations with different fields of view (see Fig. 4.6 on page 78 and Fig. 5.8 on page 123).

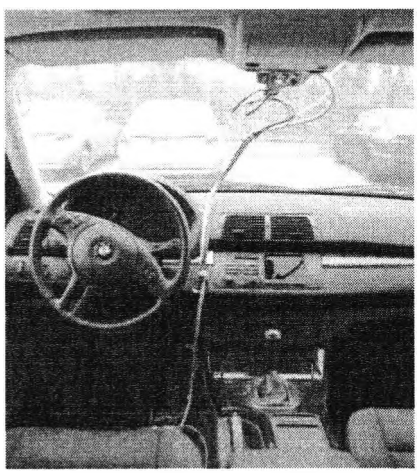
The camera consists of our custom imager (SollyCam) based on a CMOS chip from National Semiconductor with piecewise linear response. Advantages of CMOS based cameras in an automotive environment are the larger operating temperature and dynamic range coupled with lower price due to the ability to integrate all necessary peripherals and more (*e.g.* a μC) on-chip. All registers can be modified via an I^2C interface[2] from the image processing host which is described in Section 6.5. This in combination with a digital interface (RS 422) allows frame rates beyond $100Hz$, depending on the image size.

In addition the trigger output for the supplementary illumination and special input channels for door and seat information were implemented. The illumination trigger is provided to the image processing host for distinguishing and synchronizing high and low images in case of transmission errors, etc. See Section 3.3 for more details about CMOS based image sensors in general and Section 3.6 for details about our SollyCam. The car interior was illuminated using the DoubleFlash approach as discussed in Section 4.2.3 with a number of NIR clusters as active illumination devices. The NIR cluster mounted on the imager board are shown in Fig. 3.14(d).

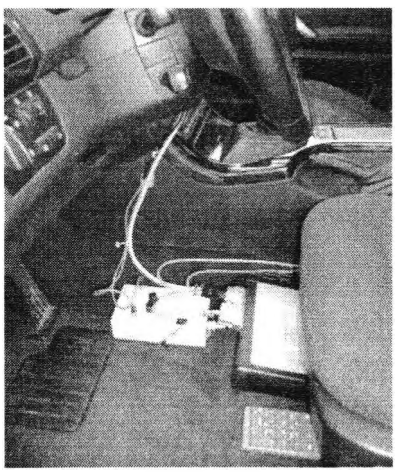
The test system employs an embedded digital signal processor (DSP) from Strampe Systemelectronic GmbH as image processing host called VisionBox (VIB). The VisionBox is discussed in more detail in Section 6.5.

Fig. 6.2(a) and 6.2(b) show the location of camera and VisionBox implemented in the test car. The VisionBox was installed below the driver's seat. The results of the image processing as well as intermediate processing steps have been displayed on the embedded car navigation system, see Fig. 6.3.

[2]Inter-IC Bus



(a) Wiring VIB ↔ Camera



(b) VIB setup

Figure 6.2: Implementation of the embedded DSP machine vision system (VisionBox) for car interior analysis.

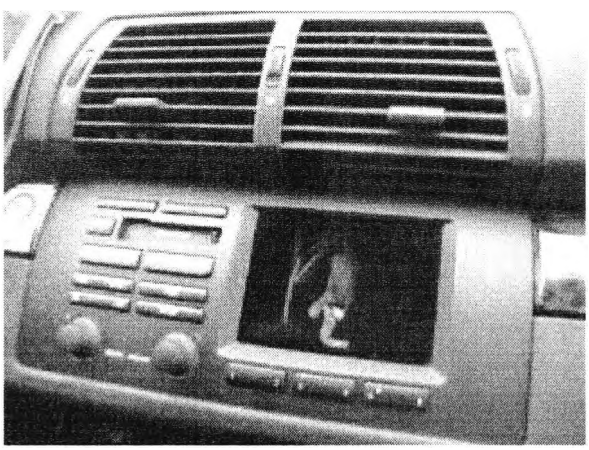


Figure 6.3: Displaying of processing results from embedded DSP via car on-board monitor.

6.3 Feature extraction

After acquisition of the input image tuple (different flashed images) and pre-processing for image enhancement the images were segmented by employing the texture based approach as introduced in Section 5.2.5. The detected foreground was segmented into connected blobs.

One assumption made to reduce the amount of input data and to speed up the classification algorithm is to determine the biggest object in the picture as the most interesting (dominant) object. This assumption is valid in a car interior. If the car interior front is divided into two regions, one for the driver and one for the passenger, a person or a child seat covers more than 40% of a region. Any other objects in the image are usually smaller than the occupants, hence they can be ignored [36]. The remained dominant object (foreground) identified by the segmentation algorithm will be measured to describe the object with numeric values and will be stored as feature vectors. Once the features are extracted and organized matching between a model and an input needs to be performed for detection or classification as described in Section 6.4.1.

The algorithm reliability and speed of the pattern matching strongly depend on the significance of the extracted object features. What features are significant, *i.e.*, what features are best suited to describe and distinguish the possible object classes? Our experiments in which we tested various shape description criteria result in the following significant object features for our 2-D model (see Fig. 6.4):

- Blob area (A_{blob})
- Blob position
  - Centre of gravity (S)
  - Extreme points ($E1, E2, E3, E4$)
  - Core distance
- Spread angle of the blob ($\alpha, \beta, \gamma, \delta$)
- Blob proportions ($a/b,\ a/d$)

The blob area A_{blob} is defined as the sum of all pixels belonging to the dominant object, excluding holes. The center of gravity $S(r, c)$ is an established method to determine the position of objects within the picture and will be calculated by

$$S_r = \frac{1}{A_{blob}} \sum_{r=0}^{R-1} \sum_{c=0}^{C-1} r f(r, c)$$
$$S_c = \frac{1}{A_{blob}} \sum_{r=0}^{R-1} \sum_{c=0}^{C-1} c f(r, c) \tag{6.1}$$

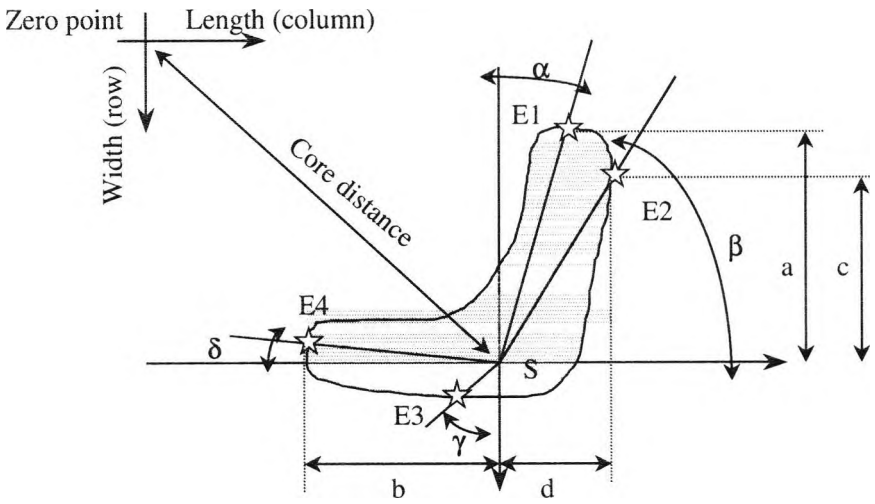


Figure 6.4: Passenger side view, showing an example blob of a forward facing child seat ($FFCS$). Blob features as used are input for the Fuzzy classifier: Spread angles $\{\alpha, \beta, \gamma, \delta\}$, blob area, core distance and dimensions $\{a, b, c, d\}$.

r and c are the picture coordinates, R and C the number of rows and columns, A_{blob} the number of pixels belonging to the object and $f(r, c)$ the picture.

Spread angles $(\alpha, \beta, \gamma, \delta)$ are defined as the angle between spread axis and a coordinate system with a zero point in the computed center of gravity S. Spread axes are the lines between the center of gravity and the extrema of the blob $(E1, E2, E3, E4)$, *i.e.* most upper, most left pixel, etc. The spread axes are in correlation with the center of gravity, so they will become invariant by translation. Hence they are suitable to describe object shapes.

The core distance represents the distance from the zero point to the center of gravity of the blob.

6.4 Classification

Many approaches for object recognition and classification such as Minimum distance classification (easy to implement), Maximum likelihood classification (fast to compute) or Neuronal Network (needs large set of training data) and hundreds of variations have been developed. To discuss each in detail and to compare them all is not the aim of this chapter, such information can be found in various books and publications [81, 13].

We choose a Fuzzy Logic based algorithm to analyze the blob data to distinguish between the occupant classes as defined in Table 1.1 on page 8.

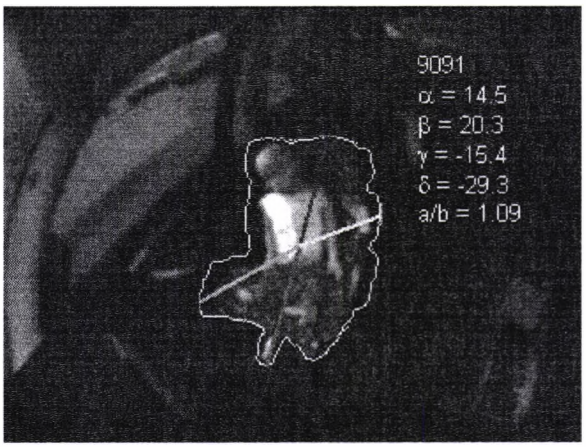


Figure 6.5: Example image of a forward facing child seat ($FFCS$) and overlay corresponding to blob features as shown in Fig. 6.4.

The advantages of a Fuzzy approach in this case are firstly that the classification produces good results even with a limited training set and secondly that it is remarkably robust to outliers. The expected number of outliers is very high due to the large number of different non-rigid 'objects'. Examples of child seats which belong to class $FFCS$ are shown in Fig. 6.6.

The variety of available child seats combined with human body shapes are enormous but the general shape can be described with simple rules for the Fuzzy classifier. Details of this formulation depend on the camera position and view. A simplified example of an idealized side view is shown in Fig. 6.4: The blob area A_{blob} of a child seat is smaller than the area occupied by an adult. The shape of a child seat generally looks like an 'L' for a rear facing infant seat and is mirrored for a forward facing infant seat. This is represented by the spread angles α and δ. The shape of a person in the

Figure 6.6: Examples of child seats which belong to class $FFCS$.

correct seating position looks similar to a forward facing child seat but the area is much greater. A person that is out of position is characterized by a negative spread angle α, etc. These unprecise master rules can be computed by Fuzzy Logic using the measured blob features.

6.4.1 Fuzzy Logic

The philosophy of Fuzzy Logic, which is very close to the human way of expressing knowledge, leads to very robust classification results, even when outliers and untrained situations occur and even if the training set is limited. We implemented a recognition system based on the standard Fuzzy Logic approach for decision making as presented by Zadeh in [96, 97] by defining a fuzzy set $\mathbf{S}$ in a fuzzy space $\mathbf{X}$,

$$\mathbf{S} = \{(x, \mu_{\mathbf{S}}(x)) | x \in \mathbf{X}\} \tag{6.2}$$

where $\mu_{\mathbf{S}}(x)$ represents the degree of membership in the fuzzy set. Usually several fuzzy sets are combined to form membership functions which describe sub-classes depending on the same input value x. Three examples of membership functions with an unit supremum as designed for the occupant detection are shown in Fig 6.7, 6.8 and 6.9.

In contrast to the basic Fuzzy Logic approach we employed membership functions which have been weighted for the final occupant classification by a weight factor η as introduced by Sasikala and Petrou in [75, 74]. This weighted factor reflects the relative importance between different membership functions that have to be combined by defining variable supremums for each membership function.

$$\sup_{x \in \mathbf{X}} \mu_{\mathbf{S}}(x) = \eta \tag{6.3}$$

The monotonic combination of each membership function with fuzzy set operators such as intersection defines the final membership of the trained occupant classes. This should be illustrated by a binary blob of a forward faced child seat mounted on the passenger seat as shown in Fig. 6.4. Assume a feature vector $\mathbf{w}$ including the segmented blob area A_{blob}, the spread angle α and the $CoreDistance$ from image origin and center of gravity,

$$\begin{aligned}
\text{if} \quad & (A_{blob} \text{ is } MEDIUM_LARGE) \cap && (6.4)\\
& (\alpha \text{ is } POSITIVE_MEDIUM) \cap && (6.5)\\
& (CoreDistance \text{ is } LARGE) && (6.6)\\
\text{then} \quad & fc \text{ is } FFCS && (6.7)
\end{aligned}$$

then fc represents the membership of input vector $\mathbf{w}$ in the class $FFCS$. Note that the logic operations in Eqn. 6.4...6.7 represent not Boolean but fuzzy operations, see [81].

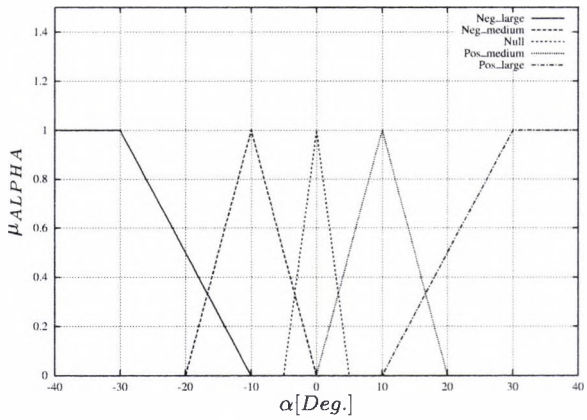


Figure 6.7: Membership function $ALPHA$

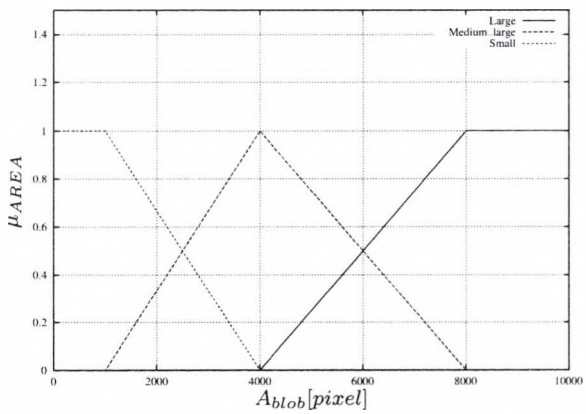


Figure 6.8: Membership function $AREA$

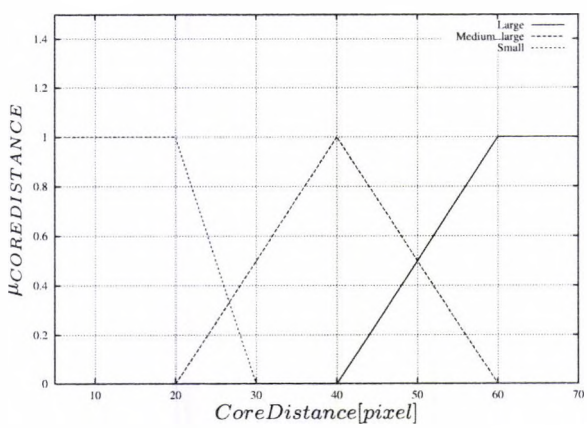


Figure 6.9: Membership function $COREDISTANCE$

The reliability of the classification can be improved by employing additional external data and an occupation history. Two examples of such additional information are: a weight mat which is implemented in the seat (see Section 1.2.2) together with the door status, which can be open or closed. A person in the correct seating position cannot change into a rear facing child seat within milliseconds without the door opening or the measured weight changing. Hence dramatic and implausible changes can be filtered by employing the occupation history and a structure of possible events for smoothing the classification output.

6.4.2 Experiments

The final results and performance estimation of our occupant classification system for single image frames without additional external data and history smoothing are shown in Table 6.2. We focused on raw occupant classification; what the airbag control unit does with these data depends on the individual safety strategy for each vehicle. This depends on the airbag module, its bag size and deployment characteristics.

Experimental data: 14 different $FFCS$, 8 different $RFCS$ and 56 different persons in correct seating position and out-of-position. Example images of the test set are shown in Fig. 6.10.

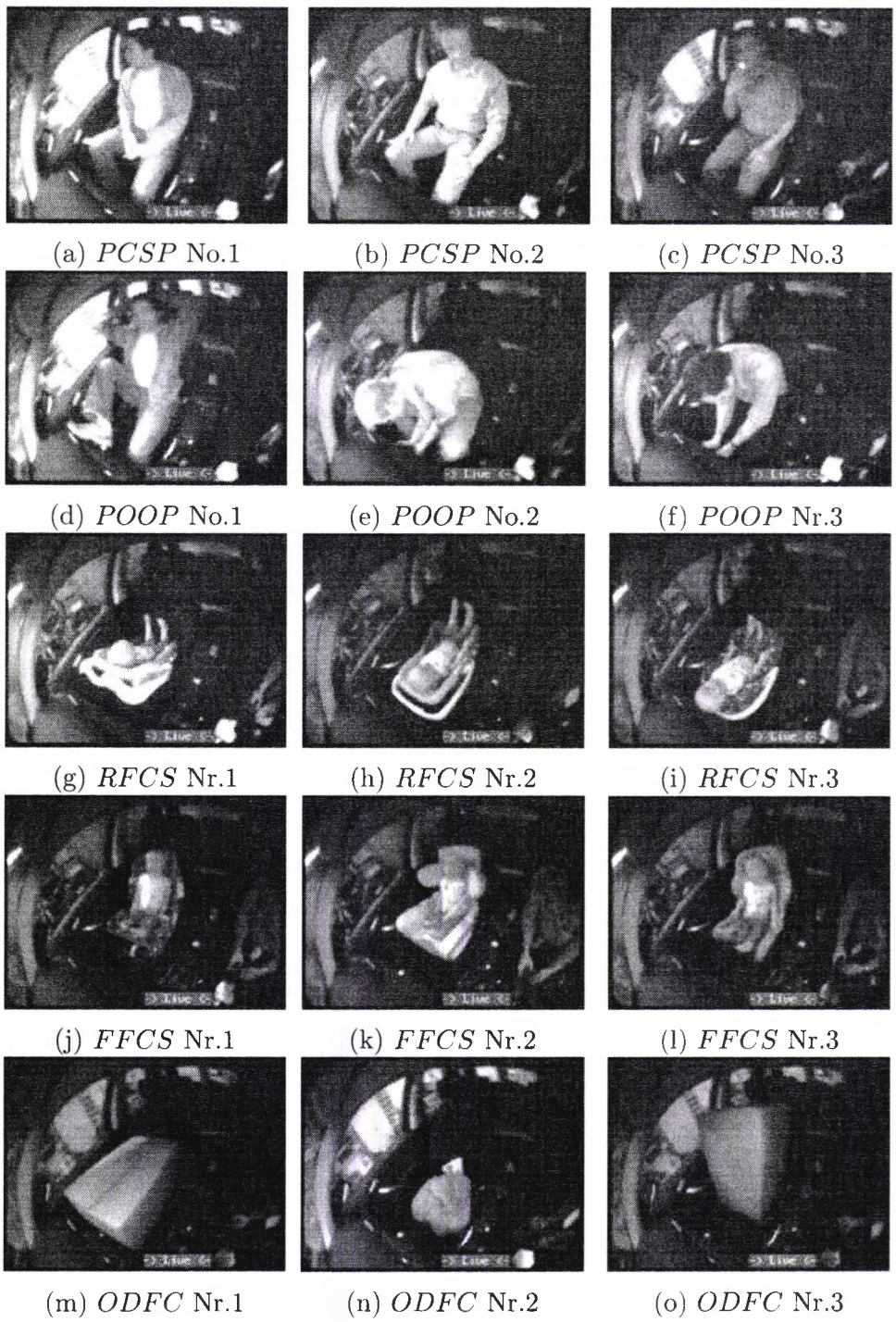
The number of test items for $FFCS$ and $RFCS$ seems to be extremely small. This is the reason why only one misclassification yields a significantly decreased percentage. But note that this small number represents about 70% of the European child seat market. The class $ODFC$ was simulated by putting several 'things' into the vehicle, *e.g.* different kind of boxes. The robustness of class $NOPS$ was tested by opening the door, sun roof, windows, etc. As mentioned in Section 6.4.1 the classification result can be improved further by employing additional external data and history smoothing. Further work will include this improvement by using knowledge gained from long term measurements.

Limitations of optical sensors

A total occlusion of the camera caused by an object which is located too close to the camera or by other optical interference limits the reliability of the visual occupant detection. This interference may be a large newspaper read by a passenger or other optical influences such as dirt and dust but also fogging of the lens. It is possible to suppress fogging by an anti-fog coating, but it is not realistic to guarantee a direct view at the passengers all times.

Therefore an automatic self-test function for maintaining and error finding is recommended for checking the system status.

A reliable system for high volume production has to detect low picture quality and to decide when the visual information provided by the image



(a) *PCSP* No.1 (b) *PCSP* No.2 (c) *PCSP* No.3

(d) *POOP* No.1 (e) *POOP* No.2 (f) *POOP* Nr.3

(g) *RFCS* Nr.1 (h) *RFCS* Nr.2 (i) *RFCS* Nr.3

(j) *FFCS* Nr.1 (k) *FFCS* Nr.2 (l) *FFCS* Nr.3

(m) *ODFC* Nr.1 (n) *ODFC* Nr.2 (o) *ODFC* Nr.3

Figure 6.10: Example data set of persons in correct seating position, out of position, various child seats and other test objects ('things').

Class	Correct	Percent	Favorite Error
FFCS	13/14	93	PCSP
RFCS	8/8	100	-
PCSP	53/56	95	FFCS
POOP	46/56	82	ODFC
NOPS	60/60	100	-
ODFC	58/60	97	FFCS
Overall	238/254	94	

Table 6.2: Classification results for single image frames without additional external data and history smoothing.

sensor is plausible, *e.g.* compared to scene features which are usually always visible. If not, the algorithm must discard the computed decision and has to release the airbag control to systems which are not based on optical data, *e.g.* a capacitive sub-system which is incorporated into the seat (see Section 1.2.2). The data from this additional sensor(s) can be imported into the existing pattern matching or be used for a second/final classification step.

6.5 Processing unit

One of the key questions of our task was how much calculation power we need to perform a stable and robust occupant classification via image processing and whether it is possible to implement such algorithms in an embedded state-of-the-art environment. The processing speed of an algorithm can be estimated by counting its use of multiplications, additions, floating-point operations, etc. However, this theoretical prediction gives just a hint of the necessary computing time to compare different algorithms.

Most image processing algorithms have been designed on PCs due to the more convenient tools for displaying, analyzing and saving processing results. However, a common PC environment is heavily influenced by operating system tasks resulting in unpredictable response times and the processing unit is not optimized for signal processing. The majority of image processing applications consist of equal filter operations which have to be performed several thousand or million times per second. This kind of signal processing is the domain of digital signal processors (DSP).

The differences between DSPs and mainstream microcontrollers (microprocessor with additional periphery for embedded designs) are listed in Table 6.3. In particular microcontroller (μC) specific polling of arbitrary I/O

Microcontroller	DSP
von Neumann architecture	Harvard architecture
One operation per clock	Several parallel operations per clock
One clock for loading each operand	Loading operands and command within one clock
Multiplication means several clocks	Special paths for multiplications within one clock

Table 6.3: Microprocessor vs DSP

ports and internal pipe structures will adulterate the theoretically calculated time values for algorithms. This is because the processing units are optimized for special tasks, *i.e.* floating or fixed point processing. Hence an algorithm using several floating operations running on such a processor must not be proportionally slower than its brother with a slower system clock using only integer operations. Hence the real processing time for algorithms varies significantly depending on the processing unit architecture, *i.e.* microcontroller or DSP, ASIC, FPGA[3], etc.

Hence for determining the necessary computing power for an application which has to be transformed finally to an embedded system, such as a system for motor vehicles, it is necessary to perform the performance evaluation directly on the target device or to implement it on a reference machine.

Therefore the proposed algorithms in Chapter 5 and Section 6.4 were implemented in a compact 32 bit off-the-shelf embedded image processing system based on the Texas Instruments C6000 DSP series called VisionBox (see Fig. 6.11). Our version provides several ports for communication with external data sources and destinations such as analog RGB output, keyboard input, etc. Key features of the employed VisionBox:

- Digital video input (LVDS/RS422)
- Digital and analog video output
- TI C3206x DSP @ $250MHz$
- $512kB$ internal memory for program, stack, fast RAM[4]
- $32MB$ SDRAM
- Interface for Compact Flash cards
- JTAG programming interface
- Misc. digital I/O interface

[3]Field Programmable Gate Array, programmable logic device
[4]Random Access Memory

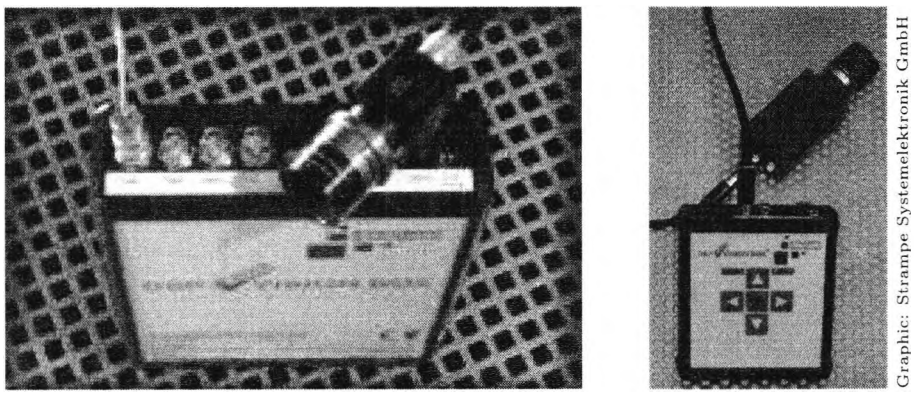


Figure 6.11: VisionBox from Strampe Systemelektronik GmbH. Left: DSPC6000 / $250MHz$ / $32MB$. Right: Compact / $150MHz$ / $32MB$

The programs are written on a PC and transferred into the DSP internal RAM after compilation via JTAG[5] interface. The final program can be copied to the Compact Flash card which is checked for a bootable program after a system reset. The system runs without any high-level operating system as a single image-processing task in a loop which enables very accurate algorithm evaluation.

According to measurements done by the manufacturer such an embedded system with a $250MHz$ clock supports more than 4 times more processing power than a $450MHz$ clocked PC depending on the implemented filter algorithms. The achieved system response for the occupant detection was up to 50 decisions (frame size = 1/4 VGA) per second including image acquisition, segmentation, classification and display control.

The next step of our system evolution would be to implement the final algorithms on a system consisting only of the pure processing unit with digital video input and a low bandwidth bus output such as CAN. Such a target hardware and its dimensions is shown in Fig. 6.11. Notice that such a system provides nearly the same computational power as its larger relatives with the dimensions of a cigarette box.

6.6 Precis

This chapter introduced an embedded real-time visual occupant detection and classification system based on hard- and software fusion. Hard- and software fusion means in this case the close link of active illumination,

[5]Joint Test Action Group or IEEE standard 1149.1. A standard specifying how to control and monitor the pins of compliant devices on a printed circuit board.

imager features and image processing to solve the problem of using machine vision in HDR environments for real-time tasks.

Characteristic blob features (*e.g.* spread angles) have been enumerated which enable a reliable distinction between persons and various types of child seat by employing a Fuzzy Logic classifier.

The next development steps include a link to one of the present automotive network (*e.g.* CAN[6]) to transmit the classification results via a network from the visual occupation sensor to the main airbag control unit.

Due to the growing market for mobile image processing applications there is a need for cost-efficient designs for demanding embedded systems with low power consumption. Another example of an experimental high performance image processing design for embedded systems was introduced by Kyo in [46]. The described parallel architecture with 128 VLIW[7] processing elements operates at a clock of $100MHz$ but performs image processing algorithms such as a 3x3 laplacian convolution up to 5 times faster than a Pentium3 clocked at $1GHz$. The power consumption is also about one tenth of conventional PC processors, which is very important for mobile designs.

Our experiments have proven that it is already possible to realize an embedded image processing based occupation detection and classification for a reasonable system price. The costs of processing power, memory and CMOS based imagers decrease each year. This means that the application of image processing in terms of artificial intelligence for motor vehicles is rather a question of years than decades.

[6]Controller Area Network
[7]Very Large Instruction Word

Chapter 7

Conclusion

7.1 Precis

This work discussed strategies for for real-time image processing in high dynamic range environments by hard- and software fusion which means employing special abilities of image sensors (such as on-chip signal processing) and active illumination to minimize necessary post-capture image enhancement steps. This leads to enhanced segmentation robustness and increased system speed, in particular in environments where the illumination of the scene is unstable, significantly influenced by ambient light fluctuations and with great contrast, such as high dynamic range (HDR) environments.

Radiometric measurements of open world scenes and detailed introduction into image acquisition techniques strengthen the *a priori* theory that mainstream CCD based image sensors are not able to cover the illumination fluctuations which are common for HDR environments.

To solve the problem of real-time image acquisition in HDR environments, three categories of solutions and improvements have been presented: a custom imager design based on CMOS technology with piecewise linear response, several novel active illumination techniques and a novel segmentation approach:

DoubleFlash is an illumination technique where a scene with high dynamic range can be captured by commonly available cameras with limited optical dynamic range without sacrificing image detail and with synchronous offset reduction. The grey-level of a pixel within the DoubleFlash output image sequence varies only if the scene changes and not as a result of fluctuations in the ambient illumination levels. This significantly simplifies any subsequent image processing.

ShadowFlash is a novel shadow detection and suppression technique, a preprocessing task that could be employed in most image processing

systems operating in an active illumination environment. With a reasonable number of controllable supplementary illuminations the proposed shadow removal algorithm simulates an infinite illuminant plane over the field of view. The achievement of the proposed method is to successfully remove shadows from a complex-textured scene without distorting the recovered background and without the support of any region extraction task.

LineFlash means to flash a tuple of lines of an image sensor rather than illuminating a tuple of frames for minimizing scene changes between a set of input images which are necessary for the majority of active illumination based image processing approaches, such as offset reduction or photometric stereo.

Difference texture based segmentation describes an alternate approach which tries to overcome the shortcomings of difference frame techniques and linear prediction approaches. This novel segmentation approach is not affected by light fluctuations nor by shadow effects, even in the event of fast light changes. On the other hand it is fast to compute and needs no time to reconstruct background areas which have been adapted to the background due to longer occlusion.

7.2 Assessment

The benefits and limitations of the presented techniques have been tested and evaluated by a typical application for HDR environments: motor vehicle interior surveillance.

This thesis introduced an approach to recognize car occupants such as children, adults, infant seats and to determine their present location and seating position, especially on the passenger seat. The final system consists of an embedded DSP as image processing host linked with a custom designed HDR CMOS camera and a couple of NIR clusters as active illumination devices. The final system was evaluated by implementing in a test car. The overall system achieved a classification accuracy of 94% in real-time, *i.e.* with up to 50 decisions per second. The accuracy of the presented system is lower than the declared accuracy by others [45, 59], but up to 100 times faster, even on an embedded system and more robust to varying illumination fluctuations due to the benefits of the presented active illumination techniques, texture based segmentation and significant classification features.

The overall classification performance of the presented car occupant detection system could be improved by

- implementing an occupation history and a structure of possible events for filtering implausible classification changes.

- an extended training set of child restraints and adults in varying sitting positions, in particular for the class *POOP*.
- incorporating a non-optical occupation sensor into the classification task, *e.g.* a capacitive sub-system which is integrated into the seat.

7.3 Outlook

Further research on this topic includes testing the applicability of the presented approaches for further outdoor applications such as building surveillance, CCTV, access authorization systems (face recognition), etc.

In particular the ShadowFlash technique can be used for applications such as a medical ultrasonic screening or a radar system by substituting the light sources for other signal types such as sonic or electric waves.

An assumption of the discussed image segmentation algorithms was that variations of the passenger seat position and slope are negligible or that the seat varies only slowly over time. To eliminate this constraint either further segmentation approaches have to separate the passenger seat first (*e.g.* by applying edge detection), or the two-dimensional model database for classification has to be extended to a more flexible three dimensional model, *e.g.* by using the findings of active illumination for an implementation of photometric stereo.

The next step of an embedded system for visual occupant detection would be to implement the final algorithms on a custom platform consisting only of the pure processing unit with digital video in and a low bandwidth bus output such as CAN for testing its behavior in a long-term vehicle implementation.

Appendix A

3D Imaging

A.1 Motivation

An alternative approach to handle illumination fluctuations within a scene is to employ depth information, *i.e.* 3-D imaging, rather than the two-dimensional projection of reflected light intensities. However, a system based on 3-D imaging, *e.g.* via stereo cameras, means increased calibrating, computing, and sensor costs. Therefore one constraint of this research project was that the motor vehicle interior surveillance system should consist of a monocular sensor to keep the hardware requirements and costs low.

Anyway it should be mentioned here because it is an upcoming approach with increased interest for a couple of intelligent motor vehicle applications. Established approaches to get a range map of a scene via image processing are:

- Stereo-vision
- Photometric stereo
- Shape from shading
- Shape from texture
- Analysis of structured light
- Several single point distance measurement in combination with a mechanical displacement unit

The 3-D imagers are very interesting for this application but there are two shortcomings: The system provides an array of measured distances, not the real volumes. To gain this topology of the object it makes sense to take a 3-D picture of the empty car and subtract this data from the current data later. But the camera cannot recognize what happens behind an object. This means the camera provides only robust information about the object shape and the distance to the detector. If, for example, a hand is held close to the camera then the camera calculates a hand shaped pillar, not the spatial volume of a real hand.

A mainstream camera and common image processing also provide only object shapes and cannot 'see' behind the shape. But the handling of vision based systems is less sensitive and more robust compared to the present 3-D imagers. Hence 3-D cameras are appropriate for applications such as out-of-position detection (OOP) but not for complex pattern matching, *e.g.* occupation detection and classification.

A better and much more detailed introduction to 3-D vision can be found in textbooks such as those by Davies [13] and Sonka [81]. The aim of the following sections is to discuss techniques for 3-D vision concerning the hard- and software *fusion* concept. In this case that means novel camera concepts which have been optimized for 3-D vision.

A.2 Time-of-flight cameras

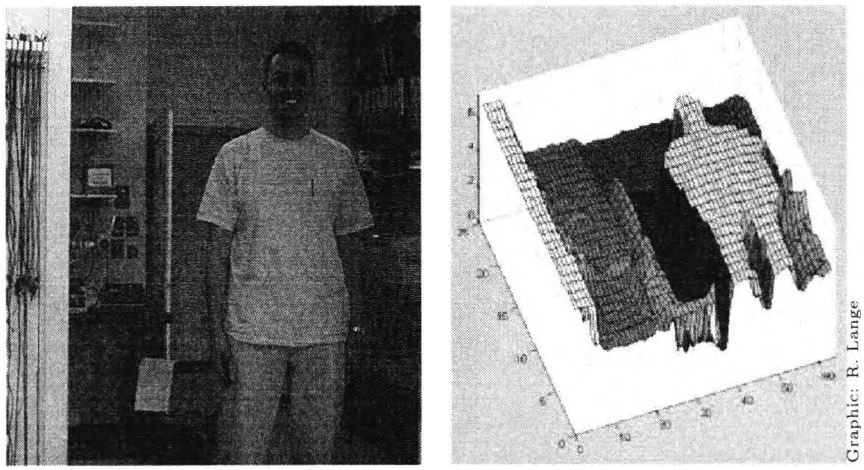
Multiple single spot distance measurement is an established method to get spatial information, which is usually employed for long distances.

Two novel range sensing cameras were introduced by Mengel in [52, 98] and Lang in [48, 47] which both consist of an array of single point distance measurement units measuring the runtime or phases of the emitted light from a supplementary light source. These time-of-flight (TOF) imagers work with a modulated light source within the visible or infrared range. The broadly emitted light pulses or waves from a bright light source (*e.g.* laser) are reflected by the objects in a scene and travel back to the imager, where their precise time of arrival or phase difference is measured locally in each pixel of a custom image sensor. This is possible because recent CMOS based imagers allow extremely short exposure times. A short exposure time is the key feature due to the speed of emitted light in combination with short distances to measure. In contrast to conventional imagers TOF range cameras determine the complete distance map of the scene on-chip. Example images of a TOF imager are shown in Fig. A.1.

The major disadvantage of this approach is that the measurement range is limited by the amount of available reflected light which depends on the maximum power of emitted light and the timing precision of the camera electronics. The latest TOF based range cameras exhibit a distance resolution of a few centimeters over a measurement range of some tens of meters. Therefore this method is not suitable for long distance measurement without violating Maximum Permissible Exposure limits (see Appendix B).

A.3 Stereo imaging

The most popular approach for getting 3-D images is stereo vision. Stereo vision means that the scene is captured by two different cameras sepa-



(a) High-resolution image (b) Measured 3-D data

Figure A.1: TOF example of a 3-D indoor scene: The distance information in (b) is coded both in the z-axis direction and in color.

rated by a distance which is called baseline.

In case of a motor vehicle interior analysis this would require more than eight image sensors as shown in Fig. A.3 if every seat has to be analyzed by a separate stereo camera system (imager cluster). However, this is not economic for space, integration and cost reasons.

An alternative approach for analyzing the whole interior by stereo-vision with a reduced number of imagers is to employ light fibers to capture images at arbitrary locations within the vehicle and to gather these image fibers at a central image sensor as shown in Fig. A.3. This technique was published in [38].

The greatest disadvantage regarding image submission via image fibers is the price of the fibers. It must consists of thousands of directed and bundled optical fibers to provide a suitable optical resolution. However, the implementation of optical fibers or prisms allows stereo vision with significantly reduced costs in mass production.

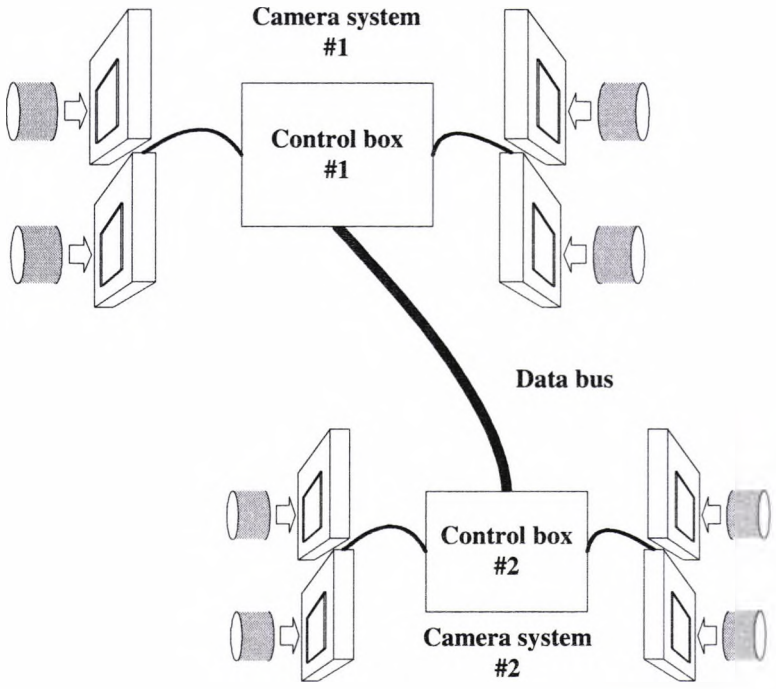


Figure A.2: Imager setup required if the front and rear seats of a car have to be observed by stereo vision.

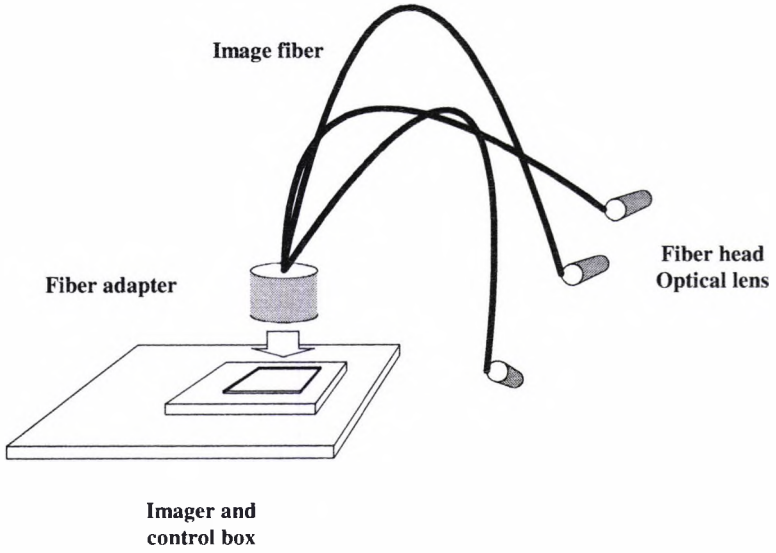


Figure A.3: Implementation of light fibers to reduce the number of image sensors.

Appendix B

Radiation safety

B.1 Motivation

Limit values for laser radiation and eye safety [12, 27] are internationally stipulated in the IEC 825-1 standard from 1993. This standard includes calculation rules to determine the Maximum Permissible Exposure limits (MPE) for laser applications and other sources of radiance which could cause hazardous levels to human observers. An MPE is that level of electromagnetic radiation, infrared, visible or ultraviolet that the eyes or skin may be exposed to over a given period of time without incurring serious effects. Based on experimental studies MPE values are intentionally set in the range of 1/10th of the known hazardous levels. However, MPE levels should be regarded as a guideline for safe exposure rather than sharp dividing lines between safe and unsafe exposure levels.

In normal use emissions from LED devices are absolutely no threat of injury to human skin, hence in this chapter only eye safety will be discussed. MPE values depend on wavelength, exposure time or pulse duration, and they are defined according to the source dimensions (which determines the image size on the retina). All sources examined in this chapter are considered sources with a relative narrow band of emission, *i.e.* $40nm$ centered around the peak wavelength. By exposure to several different wavelengths the danger of injuries may increase, and the MPE calculation becomes more difficult. To examine such sources refer to [12].

B.2 Introduction

Looking into a beam results in a minimal size image of the apparent source on the retina of an observer's eye. This minimal size image can contain sufficient energy to cause injury to the retina. The human as an observer of any light sources has a natural automatic 'blink and look-away'

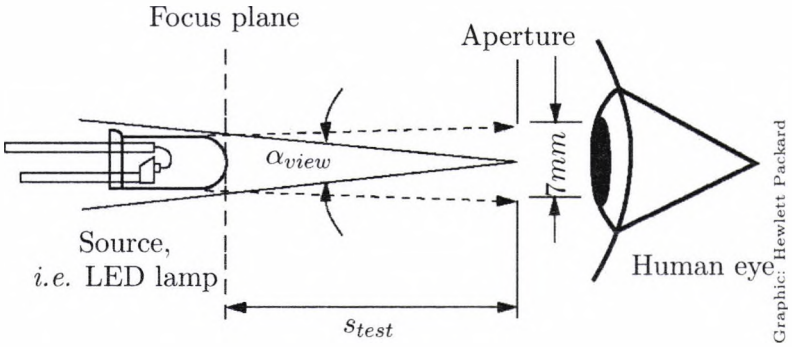


Figure B.1: Physical configuration for determining exposure values by the IEC 825-1 standard

Time	Angle
$t < 0.7s$	$\alpha_{min} = 1.5mrad$
$0.7s \leq t < 10s$	$\alpha_{min} = 2 \cdot t^{3/4} mrad$
$t \geq 10s$	$\alpha_{min} = 11mrad$

Table B.1: Minimum angle vs exposure time

protection reflex reaction that occurs in approximated 250 milli seconds and protects the eye from serious injury by light sources which appear too bright. This automatic reflex fails for radiance beyond the visible range, *i.e.* near infrared, which results in stricter MPE limits for invisible radiance. Due to the excellent suitability of near infrared (NIR) beams for creating external noise-independent illumination only the wavelengths between $700nm$ and $1050nm$ relative to MPE limits will be examined in this chapter, see Section 2.3.

The spatial MPE test conditions are depicted in Fig. B.1. A viewing aperture of $7mm$ in diameter simulates a dilated eye. The viewing angle α_{view} is determined by the apparent optical diameter of the radiance source and the measurement distance s_{test} between the aperture and the focus plane of the source. This calculated angle is limited by α_{min} and α_{max}. If α_{view} is smaller than α_{min} the MPE values are calculated independent of the apparent source dimensions. If α_{view} is greater than α_{max} α_{view} is reduced for further calculations to α_{max}. α_{min} itself is a function of the exposure time and listed in Table B.1, α_{max} is equal $0.1rad$. The viewing angle must be calculated (or measured) at the nearest distance to the apparent source (s_{test}), but not at less than $100mm$.

$$C_1 = 10^{0.002 \cdot (\lambda - 700)} \quad if \quad 700nm < \lambda < 1050nm \tag{B.1}$$

Time	Equation
$t < 10^{-9}s$	$5 \cdot 10^{6} \cdot C_1C_3 \frac{W}{m^2}$
$10^{-9}s \leq t < 1.8 \cdot 10^{-5}s$	$5 \cdot 10^{-3}C_1C_3 \frac{J}{m^2}$
$1.8 \cdot 10^{-5}s \leq t < 10^{3}s$	$18 \cdot t^{0.75}C_1C_3 \frac{J}{m^2}$
$10^{3}s \leq t < 3 \cdot 10^{4}s$	$3.2 \cdot C_1C_3 \frac{W}{m^2}$

Table B.2: Maximum Permissible Exposure limits (MPE) for wavelengths $700nm < \lambda < 1050nm$. For exposure times greater than $3 \cdot 10^4 s$ no experimental studies are available.

Angle	Factor C_3
$\alpha_{view} \leq \alpha_{min}$	1.0
$\alpha_{min} < \alpha_{view} \leq \alpha_{max}$	$\alpha_{view}/\alpha max$
$\alpha_{view} > \alpha_{max}$	$\alpha_{max}/\alpha_{min}$

Table B.3: Correction factor C_3

These limits are safe, given that no optical devices are used to focus the emitted light. This abnormal case may occur, *i.e.* if a bored child plays with its glasses and takes a direct view into the radiant source. To reduce the radiation stress usually safety glasses will be used. Of course this is not applicable for car occupants. Hence the MPE values limit the available irradiation power to the upper end for this application. The lower limit for the illumination is determined by the signal-to-noise ratio of the camera and ambient illumination which has to be suppressed. Passengers are used to drive very long distances in their car, therefore exposure times greater than eight hours must be assumed for eye safety as a worst-case . Typical radiant Intensity:

Semiconductor laser ($880nm$,$2mW$, without additional optics):
$I_e = 2...5mW/sr$

IR LED for remote control (@$100mA$):
$I_e = 10...100mW/sr$

B.2.1 Pulsed sources

MPE values for pulsed sources are limited by the subsequent three demands:

- The radiance from every single pulse in a pulse series must not exceed the MPE values for a single pulse
- The mean irradiance for a pulse series of the duration T must not exceed the MPE values for a single pulse with duration T
- The irradiance from every single pulse of the pulse series must not exceed the product of the MPE value of a single pulse and C_2

$$MPE_{series} = MPE_{single} \cdot C_2 \tag{B.2}$$

$$C_2 = N^{-1/4} \tag{B.3}$$

The number of pulses during irradiance is expressed in Eqn. B.3 with N.

B.2.2 Cluster of different sources

Suitable MPE values for a cluster of several sources are determined by the MPE constraint which results in the maximum limit considering every single source and every grouping of single sources. This complex calculation could be simplified by using a very conservative assumption that the whole energy is emitted by one point. This means that the calculated hazard is always greater than the real radiation situation. But this assumption is only useful if the results do not cause inappropriately strict safety constraints. However, the illumination to be used for this application should be used not only in safe laboratories but under rough conditions of a poorly maintained motor vehicle. Since a maximum of passenger safety must be guaranteed in this thesis we consider every source of clusters a single source.

B.3 MPE calculation example

Assuming a continuous glowing high performance GaAlAs NIR LED illuminator with data listed in Table B.4. The valid calculation rules for wavelengths between $700nm$ and $1050nm$ (NIR) and an exposure time longer than 10^3 seconds are listed in Table B.2 and are used in Eqn. B.8. The limit angle α_{min} is exposure-time dependent and stipulated by the IEC 825-1 standard as $11mrad$ for a direct view into the beam for more than $10s$.

$$\alpha_{view} = 2 \cdot \arctan\left[\frac{d_{view}/2}{s_{test}}\right] \tag{B.4}$$

$$\alpha_{view} = 2 \cdot \arctan\left[\frac{0.75mm/2}{100mm}\right] = 3.8mrad \tag{B.5}$$

Parameters	Test conditions	Typical	Units
Power output, Φ_{source}	$I_F = 300mA$	500	mW
Pulse power output, Φ_{source_p}	$I_F = 5A$	6500	mW
Peak emission wavelength, λ_p	$I_F = 50mA$	880	nm
Spectral bandwidth λ_Δ	$I_F = 50mA$	80	nm
Half intensity beam angle, α_b	$I_F = 50mA$	120	Deg.
Supply voltage, V_F	$I_F = 300mA$	13.5	Volts

Table B.4: High performance GaAlAs NIR LED illuminator

The NIR LED illuminator used for lighting the interior of a car provides a direct access to the source, so the s_{test} distance is set to the minimum value of $100mm$ (table B.1). The sample illuminator consists of nine separable diodes. Assuming the constraint of Section B.2.2 only one of them can be considered a source emitting the whole energy of the cluster. The resulting diameter of rectangular sources is determined by the mean of horizontal and vertical dimensions. A geometric view of the physical MPE test conditions leads to Eqn. B.4. Assuming an exposure time greater than $10s$ the calculated $3.8mrad$ for α_{view} is smaller than α_{min}. Hence the calculation of the MPE values for this source is independent of the source dimensions, and the correction factor C_3 is fixed for the final calculation to 1.

$$C_1 = 10^{0.002\cdot(880-700)} = 10^{0.36} = 2.29 \tag{B.6}$$

$$C_3 = 1 \quad for \quad \alpha \leq \alpha_{min} \tag{B.7}$$

$$E_{mpe} = 3.2 \cdot C_1 \cdot C_3 \tag{B.8}$$

$$\frac{W}{m^2} = 7.33\frac{W}{m^2} \tag{B.9}$$

B.3.1 Actual irradiation

The radiant intensity E_{eye} is determined by the ratio of the source radiant power Φ_{source} over the illuminated area A_{exp}. The illuminated area can be approximated as the bottom of an ideal radiation cone and is therefore a function of the beam angle α_b and the distance between source and target (s_{viewer}).

$$E_{eye} = \frac{\Phi_{source}}{A_{exp}} \tag{B.10}$$

$$A_{exp} = r_{cone}^2 \cdot \pi \tag{B.11}$$

$$r_{cone} = s_{viewer} \cdot \tan\left(\frac{\alpha_b}{2}\right) \tag{B.12}$$

$$E_{eye} = \frac{\Phi_{source}}{\left[\tan(\frac{\alpha_b}{2})\cdot s_{viewer}\right]^2 \cdot \pi} \tag{B.13}$$

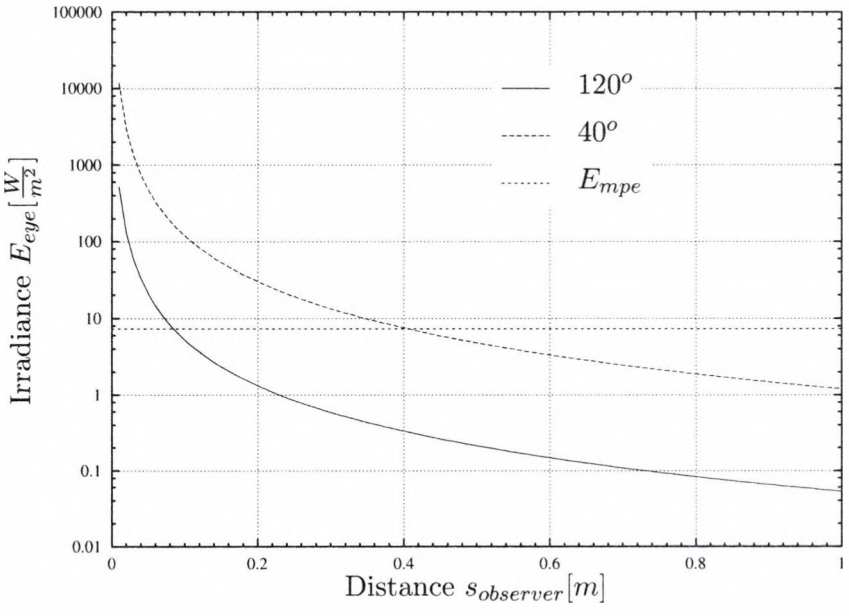


Figure B.2: Irradiation stress vs distance of the observer to the source

Substituting Eqn. B.11 and B.12 in B.10 leads to Eqn. B.13. These are approximation values due to the non-uniform radiation pattern of the diodes. The centre of the illuminated area may show peaks twice as powerful as the border regions. The resulting irradiation stress compared to the observers distance for the inspected source with a total power output approximated $500mW$ is shown in Fig. B.2 with two different half beam angles of 40^o and 120^o. The horizontal line marks the maximum permissible emission value for that source calculated in Section B.3. This means that eye safety is guaranteed for this illuminator if the observer keeps a minimum safety distance of $0.1m$ to the source, assuming a beam angle of 120^o. But if the beam is compressed to 40^o the necessary safety distance increases to approximately $0.4m$.

Bibliography

[1] M. Adjouadi. Image analysis of shadows, depressions, and upright objects in the interpretation of real-world scenes. In *Int. Conf. on Pattern Recognition, ICPR*, volume 86, pages 834–838, Paris, France, October 1986.

[2] J. Aggarwal, Q. Cai, W. Liao, and B. Sabata. Articulated and elastic non-rigid motion: A review. In *IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pages 16–22, Austin, USA, 1994.

[3] N. Bartneck and W. Ritter. Colour segmentation with polynomial classification. In *International Conference on Pattern Recognition, ICPR*, volume B, pages 635–638, 1992.

[4] H. Bässmann and J. Kreyss. *Bildverarbeitung AdOculos*. Springer Verlag, Berlin, 3 edition, 1998.

[5] B.E. Bayer. Color imaging array. National Patent US 3.971.065, 1976.

[6] M. Böhm, F. Blecher, A. Eckhardt, and B. Schneider. High dynamic range image sensors in Thin Film on ASIC technology for automotive applications. In *Advanced Microsystems for Automotive Applications*, pages 157–172, Berlin, 1998. Springer-Verlag.

[7] S. Brofferio, L. Carnimeo, D. Comunale, and G. Mastronardi. *Time-Varying Image Processing*, chapter A background updating algorithm for moving object scenes, pages 297–307. Elsevier Science Publishers, 2 edition, 1990.

[8] C.M. Brown. Tutorial on filtering, restoration, and state estimation. Technical Report TR534, University of Rochester, Computer Science Dept., June 1995.

[9] A. Bußmann. Implementierung eines Algorithmus zur Bewegungsdetektion in ein CMOS-Kamera System. Master's thesis, University Duisburg, Germany, 1998.

[10] C. Cédras and M. Shah. Motion-based recognition: A survey. *IVC*, 13(2):129–155, March 1995.

[11] International Electrotechnical Commission. Measurement principles for terrestrial photovoltaic (PV) solar devices with reference spectral irradiance data. International Standard IEC 904-3, 1989.

[12] International Electrotechnical Commission. Safety of laser products - part1: Equipment classification, requirements and user's guide. International Standard IEC 825-1, 1993.

[13] E.R. Davies. *Machine Vision: Theory, Algorithms, Practicalities.* Academic Press, San Diego, 2 edition, 1997.

[14] National Highway Traffic Safety Administration (NHTSA) Department of Transportation (USA). Occupant crash protection. 49 CFR Parts 552, 571, 585 and 595, Docket No. NHTSA 00-7013; Notice 1, RIN 2127-AG70.

[15] B. Dierickx, G. Meynants, and D. Scheffer. Near 100% fill factor CMOS active pixels. In *IEEE CCD and AIS workshop*, Brugge, Belgium, June 1997.

[16] G.W. Donohoe, D.R. Hush, and N. Ahmed. Change detection for target detection and classification in video sequences. In *IEEE Int. Conf. on Acoustic, Speech and Signal Processing*, number 2, pages 1084–1087, 1988.

[17] P. Faber. Seat occupation detection inside vehicles. In *IEEE Southwest Symposium on Image Analysis and Interpretation*, Austin, USA, April 2000.

[18] H. Fischer, J. Schulte, J. Giehl, and B. Boehm. Thin Film on ASIC - a novel concept for intelligent image sensors. In *Mat. Res. Soc. Symp.*, volume 285, pages 1139–1145, 1992.

[19] E.R. Fossum. CMOS image sensors: Electronic camera-on-a-chip. In *IEEE Transactions on Electron Devices*, volume 44, pages 1689–1698, October 1997.

[20] H.C. Gabler, R.R. Krchnavek, and J.L. Schmalzel. Development of an automated crash notification system: An undergraduate research experience. In *ASEE/IEEE Frontiers in Education Conference*, Kansas City, USA, October 2000.

[21] A. El Gamal, D. Yang, and B. Fowler. Pixel level processing - Why, What and How? In *SPIE Electronic Imaging '99 Conference*, volume 3650, January 1999.

[22] A. Gassel. *Beiträge zur Berechnung solarthermischer und exergieeffizienter Energiesysteme.* PhD thesis, University Dresden, Germany, 1996.

[23] M. Goesele, W. Heidrich, and H.-P. Seidel. Color calibrated high dynamic range imaging with ICC profiles. In *Color Imaging Conference*, number 9, 2001.

[24] G. Gordon, T. Darrell, M. Harville, and J. Woodfill. Background estimation and removal based on range and color. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR99)*, volume 2, pages 459–464, June 1999.

[25] S. Groening. Erfassung des Bekleidungszustands von Fahrzeuginsassen. BMW concept study, FORWISS Passau, Germany, 1999.

[26] P. Haberäcker. *Digitale Bildverarbeitung.* Hanser Verlag, München, 4 edition, 1991.

[27] Hewlett-Packard Co., USA. *Visible LED Device Classifications With Respect to AEL Values as Defined in the European CENELEC EN60825-1 Standard*, May 1996. Application Brief I-015.

[28] J. Huppertz. *2-D CMOS Bildsensorik mit integrierter Bildverarbeitung.* PhD thesis, Gerhard-Mercator-Universitt Duisburg, Germany, 1999.

[29] J. Huppertz, R. Hauschild, B.J. Hosticka, et al. Fast CMOS imaging with high dynamic range. In *IEEE Workshop on charge-coupled devices and advanced image Sensors*, 1997.

[30] H.Witters, T.Walschap, G. Vanstraelen, G.Chapinal, G.Meynants, and B.Dierickx. 1024 x 1280 pixel dual shutter APS for industrial vision. In *SPIE Electronic Imaging*, volume 5017, Santa Clara, USA, January 2003.

[31] B. Jähne. *Digitale Bildverarbeitung.* Springer Verlag, Berlin, 4 edition, 1997.

[32] C. Jiang and M.O. Ward. Shadow segmentation and classification in a constrained environment. *CVGIP: Image Understanding*, 59:213–225, 1994.

[33] K.P. Karmann and A. Brand. Detection and tracking of moving objects by adaptive background extraction. In *Scand. Conf. Image Anal.*, number 6, pages 1051–1058, 1989.

[34] S. Kleinfelder, S.H. Lim, X.Q. Liu, and A. El Gamal. A 10,000 frames/s CMOS digital pixel sensor. In *IEEE Journal of Solid State Circuits*, volume 36, pages 2049–2059, December 2001.

[35] M. Klomark. Occupant detection using computer vision. Master's thesis, Linköping University, Sweden, May 2000.

[36] C. Koch. Sitzbelegungserkennung im Kfz durch digitale Bildverarbeitung. Master's thesis, FH Nordostniedersachsen FB Automatisierungstechnik, February 1999.

[37] C. Koch and S. Akisoglu. Line flash. National Patent DE 102.45.912, October 2002. patent pending.

[38] C. Koch, A. Augst, and M. Fuchs. Stereo vision with mono chip. National Patent Application, January 2003. patent pending.

[39] C. Koch, T.J. Ellis, and A. Georgiadis. Real-time occupant classification in high dynamic range environments. In *IEEE Intelligent Vehicle Symposium*, Versailles, France, June 2002.

[40] C. Koch and A. Georgiadis. Quality control device for verifying glass ampoule seal integrity. National Patent DE 197.15.450, October 1997.

[41] C. Koch, S.-B. Park, T.J. Ellis, and A. Georgiadis. Illumination technique for optical dynamic range compression and offset reduction. In *British Machine Vision Conference (BMVC2001)*, pages 293–302, Manchester, UK, September 2001.

[42] C. Koch, S.-B. Park, and S. Sauer. Method and apparatus for monitoring the interior space of a motor vehicle. International Patent EP 1.215.619, December 2000.

[43] C. Koch, J.-J. Yoon, and L. Eisenmann. Shadowflash. National Patent DE 102.50.705, October 2002. patent pending.

[44] D. Koller, K. Danilidis, and H-H. Nagel. Model-based object tracking in monocular image sequences of road traffic scenes. *International Journal of Computer Vision (IJCV)*, 10(3):257–281, June 1993.

[45] J. Krumm and G. Kirk. Video occupant detection for airbag deployment. In *IEEE Workshop on Applications of Computer Vision (WACV98)*, pages 30–35, Princeton, USA, October 1998.

[46] S. Kyo, T. Koga, and S. Okazaki. IMAP-CE: A 51.2 GOPS video rate image processor with 128 VLIW processing elements. In *International Conference on Image Processing*, October 2001.

[47] R. Lange. *3D Time-of-Flight Distance Measurement with Custom Solid-State Image Sensors in CMOS/CCD-Technology.* PhD thesis, University of Siegen, Department of Electrical Engineering and Computer Science, Germany, 2000.

[48] R. Lange and P. Seitz. Solid-state Time-of-Flight range camera. *IEEE Journal of Quantum Electronics*, 37(3), March 2001.

[49] T. Lulé, S. Benthien, H. Keller, and F. Mütze. Sensitivity of CMOS based imagers and scaling perspectives. *Submitted to IEEE Transactions on Electronic Devices*, March 2000.

[50] S. Mann and R.W. Picard. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Society for Imaging Science and Technology*, number 48, pages 422–428, Washington, USA, May 1995.

[51] S.K. Mendis, S.E.Kemeny, and R.C. Gee. CMOS active pixel image sensors for highly integrated imaging systems. *IEEE Solid-State Circuits*, 32:187–197, 1997.

[52] P. Mengel, G. Doemens, and L. Listl. Fast range imaging by CMOS sensor array through multiple double short time integration (mdsi). In *ICIP01*, pages Stereoscopic and 3–dimensional Image Processing i, 2001.

[53] G. Meynants, B. Dierickx, and D. Scheffer. CMOS active pixel image sensor with CCD performance. AFPAEC Europto/SPIE, May 1998.

[54] I. Mikic, P.C. Cosman, G.T. Kogut, and M.M. Trivedi. Moving shadow and object detection in traffic scenes. In *International Conference on Pattern Recognition (ICPR00)*, volume 1, pages 321–324, September 2000.

[55] National Semiconductor Corp. *Application of the Piecewise Linear Response Feature in the LM9618/28 Image Sensors*, LM9628/18 application note 2, revision 1.1 edition, January 2002.

[56] National Semiconductor Corp. *LM9618 monochrome CMOS image sensor VGA 30 FPS*, 0.7c edition, July 2002.

[57] Oriel Instruments. *Multifunction Optical Power Meter, Model 70310*, May 1998.

[58] W. Osamu and O. Katsutake. Interior human body detecting device and vehicle equipment control device. National Patent JP 123.793, June 1994.

[59] Y. Owechko, N. Srinivasa, S. Medasani, and R. Boscolo. Vision-based fusion system for smart airbag applications. In *IEEE Intelligent Vehicle Symposium*, Versailles, France, June 2002.

[60] N. Pal and K. Pal. A review on image segmentation techniques. *Pattern Recognition*, 26(9):1277–1294, 1993.

[61] S.-B. Park. *Optische Kfz-Innenraumüberwachung.* PhD thesis, University Duisburg, Germany, December 1999.

[62] S.-B. Park, M. Schanz, B.J. Hosticka, and A. Teuner. Method and device for detecting a change between pixel signals which chronologically follow one another. International Patent EP 97.04.452, 1997.

[63] S.-B. Park, A. Teuner, B.J. Hosticka, and G. Triftshaeuser. An interior compartment protecting system based on motion detection using CMOS imagers. In *Int. Conf. in Intelligent Vehicles*, pages 297–301, October 1998.

[64] I. Paromtchik and C. Laugier. The advanced safety vehicle programme. In *Colloque sur les véhicules électriques*, pages C26–C30, Grenoble, France, February 1997.

[65] G. Paula. Sensors help make air bags safer. *Mechanical engineering magazine*, 119(8), 1997.

[66] F.L. Pedrotti and L.S. Pedrotti. *Introduction to Optics.* Pearson Education, November 1992.

[67] R. Ramanath, W. Snyder, G. Bilbro, and W. Sander. Demosaicking methods for bayer color arrays. *Journal of Electronic Imaging*, 11(3), July 2002.

[68] N. Ricquier and B. Dierickx. Active pixel CMOS image sensor with on-chip non-uniformity correction. In *IEEE Workshop on CCD and advanced image sensors*, California, USA, April 1995.

[69] C. Ridder, O. Munkelt, and H. Kirchner. Adaptive background estimation and foreground detection using kalman filtering. In *UNESCO Chair on Mechatronics (ICRAM95)*, pages 193–199, 1995.

[70] P. Rieve, M. Sommer, M. Wagner, K. Seibel, and M. Böhm. a-Si:H color imagers and colorimetry. *Journal of Non-Crystalline Solids*, 266-269:1168–1172, 2000.

[71] M.A. Robertson, S. Borman, and R.L. Stevenson. Dynamic range improvement through multiple exposures. In *International Conference on Image Processing*, pages 159–163, Kobe, Japan, October 1999.

[72] P.L. Rosin and T.J. Ellis. Image difference threshold strategies and shadow detection. In *British Machine Vision Conference (BMVC95)*, number 6, pages 347–356, Birmingham, UK, September 1995.

[73] A. Ryer. *Light Measurement Handbook.* International Light Inc., 1998.

[74] K.R. Sasikala and M. Petrou. Properties of the generalised fuzzy aggregation operators. *Pattern Recognition Letters*, 22(1):15–24, January 2001.

[75] K.R. Sasikala, M. Petrou, and J. Kittler. Fuzzy reasoning with a GIS for decision making in burned forest management. *EARSeL Advances in Remote Sensing*, 4:97–105, 1996.

[76] M. Schanz, C. Nitta, T. Eckart, B.J. Hosticka, and R. Wertheimer. A high dynamic range CMOS image sensor for automotive applications. *IEEE Journal of Solid-State Circuits*, 35(7):932–938, July 2000.

[77] B. Schneider, H. Fischer, and S. Benthien. TFA image sensors: From the one transistor cell to a locally adaptive high dynamic range sensor. In *Technical Digest of International Electron Devices Meeting*, pages 209–212, December 1997.

[78] N.L. Seed and A.D. Houghton. Background updating for real-time image processing at TV rates. In *SPIE: Image Processing, Analysis, Measurement and Quality*, volume 901, Los Angeles, USA, 1988.

[79] G.G. Sexton and X. Zhang. Suppression of shadows for improved object discrimination. IEEE Colloquium on Image Processing for Transport Applications, December 1993.

[80] C.G. Sodini and S.J. Decker. CMOS brightness adaptive imaging array with column-parallel digital output. In *IEEE Intelligent Vehicles Symposium*, pages 347–352, Stuttgart, Germany, October 1998.

[81] M. Sonka, V. Hlavacand, and R. Boyle. *Image Processing, Analysis and Machine Vision.* PWS Publishing, 2 edition, 1998.

[82] J. Stauder, R. Mech, and J. Ostermann. Detection of moving cast shadows for object segmentation. *IEEE Transactions on Multimedia*, 1(1):65–76, 1999.

[83] C. Stauffer and W.E.L Grimson. Adaptive background mixture models for real-time tracking. In *CVPR99*, volume 2, pages 246–252, Fort Colins, USA, June 1999.

[84] R. Street. High dynamic range segmented pixel sensor array. National Patent US 5.789.737, 1998.

[85] TOPTEC. Vehicle safety restraint system, August 1999.

[86] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *IEEE ICCV*, pages 255–261, 1999.

[87] M. Wäny. High dynamic CMOS image sensors. *G.I.T Imaging and Microscopy*, 3:26–28, 2001.

[88] C.W. Wyckoff. An experimental extended response film. Technical Report B-321, Edgerton, Germeshausen & Grier, Inc., Boston, Massachusetts, USA, March 1961.

[89] C.W. Wyckoff. An experimental extended response film. *SPIE Newsletter*, June 1962.

[90] F. Xiao, J.M. DiCarlob, P.B. Catrysseb, and B.A. Wandell. Image analysis using modulated light sources. In *SPIE*, 2000.

[91] M. Xu and T.J. Ellis. Illumination-invariant motion detection using color micture models. In *British Machine Vision Conf. (BMVC01)*, pages 163–172, Manchester, UK, September 2001.

[92] K. Yamada, T. Nakano, and S. Yamamoto. Effectiveness of video camera dynamic range expansion for lane mark detection. In *IEEE International Conference on Intelligent Vehicles*, 1998.

[93] D. Yang and A. El Gamal. Comparative analysis of SNR for image sensors with widened dynamic range. In *SPIE*, February 1999.

[94] J.J. Yoon, C. Koch, and T.J. Ellis. Shadowflash: an approach for shadow removal in an active illumination environment. In *British Machine Vision Conference (BMVC2002)*, pages 636–645, Cardiff, UK, September 2002.

[95] K. Yoshikawa, K. Murano, S. Sunahara, M. Mizuno, T. Tsukata, and T. Naitoand K. Yamada. Vision systems for its using wide dynamic range camera -application in license plate recognition system, parking lot monitoring system. In *World Congress on Intelligent Transport Systems*, number 5, 1998.

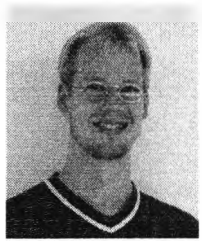
[96] L.A. Zadeh. *Fuzzy sets, Information and Control*, volume 8, pages 338–353. 1965.

[97] L.A. Zadeh. *Fuzzy sets and systems*, volume 1, chapter Fuzzy Sets as a basis for theory of possibility, pages 3–28. 1978.

[98] D. Zittlau, P. Mengel, and S. Boverie. Innovativer Kfz-Insassen- und Partnerschutz. *VDI Berichte, Gesellschaft Fahrzeug- und Verkehrstechnik*, (1471):213–227, 1999.

Index

The author



Carsten Koch was born in 1974 in Lüneburg, Germany. From 1994 to 1999 he studied at the University of Applied Sciences in Lüneburg, and he holds a degree in Automation and Electronics.

After his graduation he enrolled in City University, London, in the department of Electrical, Electronic and Information Engineering as a research student. The research was supervised by Dr. T.J. Ellis, who is leading the Machine Vision Group at City University.

In 1999 he joined the BMW research and development department in Munich, Germany, where most of this research was performed.

Today the author and his family live in northern Germany, where he founded a consulting company for image sensors in 2002. He teaches courses in digital electronics at the University of Applied Sciences in Lüneburg.