Optimized Distance Measurement with 3D-CMOS Image Sensor and Real-Time Processing of the 3D Data for Applications in Automotive and Safety Engineering

Von der Fakultät für Ingenieurwissenschaften der

Universität Duisburg-Essen

zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften

genehmigte Dissertation

von

Bernhard König

aus München

Referent: Prof. Bedrich J. Hosticka, Ph.D.
Korreferent: Prof. Dr. Josef Pauli
Tag der mündlichen Prüfung: 02.07.2008

Vorwort

Diese Arbeit entstand im Rahmen der Zusammenarbeit zwischen dem **Fraunhofer Institut für Mikroelektronische Schaltungen und Systeme (FhG IMS)** in Duisburg und der Abteilung **CT PS 9** von **Siemens Coorporate Technology** in München. Als externer Mitarbeiter von Siemens CT PS 9 durfte ich im industriellen Forschungsumfeld an interessanten Projekten mitarbeiten und die Forschungstätigkeiten meiner Promotion durchführen. Ich danke Dr. Günther Doemens und Dr. Claudio Laloni, den Abteilungsleitern von Siemens CT PS 9, die mir dies ermöglicht haben.
Genauso danke ich Prof. Bedrich J. Hosticka, Ph.D. vom FhG IMS für die Betreuung meiner Promotion. Er beobachtete stets den Fortschritt meiner Arbeiten und unterstützte mich durch sachkundige Ratschläge.
Mein besonderer Dank gilt Dr. Peter Mengel, meinem Betreuer bei Siemens CT PS 9, sowie meinem Kollegen Ludwig Listl, deren fachkundige Anleitung wesentlich zum Gelingen dieser Arbeit beitrug. In vielen Fachgesprächen und Diskussionen wurde gemeinsam das Fundament für diese Arbeit gelegt.
Weiter möchte ich meinen Kollegen vom FhG IMS und von Siemens CT PS 9, Wiebke Ulfig, Dr. Omar Elkhalili, Uwe Phillipi und Roland Bürger für die angenehme Zusammenarbeit, und meiner Familie und meinen Freunden für den Ausgleich zum Arbeitsumfeld danken.

Bernhard König

München, 31. Oktober 2007

Abstract

This thesis describes and characterizes an advanced range camera for the distance range from 2 m to 25 m and novel real-time 3D image processing algorithms for object detection, tracking and classification on the basis of the three-dimensional features of the camera output data.
The technology is based on a 64x8 pixel array CMOS image sensor which is capable of capturing three-dimensional images. This is accomplished by executing indirect time of flight measurement of NIR laser pulses emitted by the camera and reflected by the objects in the field of view of the camera.
An analytic description of the measurement signals and a derivation of the distance measuring algorithms are conducted in this thesis as well as a comparative examination of the distance measuring algorithms by calculation, simulation and experiments; in doing so, the MDSI3 algorithm showed the best results over the whole measurement range and is thus chosen as standard method of the distance measuring system.
A camera prototype was developed with a measurement accuracy in the centimeter range at an image repetition rate up to 100 Hz; a detailed evaluation of the components and of the over-all system is presented. Main aspects are the characterization of the time critical measurement signals, of the system noise, and of the distance measuring capabilities.
Furthermore this thesis introduces novel real-time image processing of the output data stream of the camera aiming at the detection of objects being located in the observed area and the derivation of reliable position, speed and acceleration estimates. The used segmentation algorithm utilizes all three spatial dimensions of the position information as well as the intensity values and thus yields significant improvement compared to segmentation in conventional 2D images. Position, velocity, and acceleration values of the segmented objects are estimated by means of Kalman filtering in 3D space. The filter is dynamically adapted to the measurement properties of the according object to take care of changes of the data properties. The good performance of the image processing algorithms is presented by means of example scenes.

Kurzfassung

Diese Arbeit beschreibt und charakterisiert eine neu entwickelte Entfernungskamera für Reichweiten von 2 m bis 25 m und spezielle 3D-Echtzeit-Bildverarbeitungsalgorithmen zum Detektieren, Tracken und Klassifizieren von Objekten auf der Grundlage der dreidimensionalen Kameradaten.
Die Technologie basiert auf einem 64x8 Pixel CMOS Bildsensor, welcher im Stande ist, dreidimensionale Szenen zu erfassen. Dies wird mittels indirekter Laufzeitmessung von NIR Laserpulsen, die von der Kamera ausgesandt und an Objekten im Blickfeld der Kamera reflektiert werden, realisiert.
Eine analytische Beschreibung der Messsignale und eine darauf aufbauende Herleitung der verschiedenartigen Entfernungsmessungsalgorithmen wird in dieser Arbeit ebenso durchgeführt, wie die vergleichende Betrachtung der Entfernungsmessungsalgorithmen durch Rechnung, Simulation und Experimente; dabei zeigt der MDSI3-Algorithmus die besten Ergebnisse über den gesamten Messbereich, und wird deshalb zum Standardalgorithmus des Entfernungsmesssystems.
Ein Kameraprototyp mit Messgenauigkeiten im cm-Bereich bei einer Bildwiederholrate bis zu 100 Hz wurde entwickelt; eine detaillierte Evaluierung der Komponenten und des Systems ist hier beschrieben. Hauptaspekte sind dabei die Charakterisierung der zeitkritischen Messsignale, des Systemrauschens und der Entfernungsmesseigenschaften.
Desweiteren wird in dieser Arbeit die neu entwickelte Echtzeit-Bildverarbeitung des Kameradatenstroms vorgestellt, die auf die Detektion von Objekten im Beobachtungsbereich und die verlässliche Ermittlung von Positions-,Geschwindigkeits- und Beschleunigungsschätzwerten abzielt. Der dabei verwendete Segmentierungsalgorithmus nutzt alle drei Dimensionen der Positionsmesswerte kombiniert mit den Intensitätswerten der Messsignale, und liefert so eine signifikante Verbesserung im Vergleich zur Segmentierung in konventionellen 2D Bildern. Position, Geschwindigkeit und Beschleunigung werden mit Hilfe eines Kalman-Filters im 3-dimensionalen Raum geschätzt. Das Filter passt sich dynamisch den Messbedingungen des jeweils gemessenen Objekts an, und berücksichtigt so Veränderungen der Dateneigenschaften. Die Leistungsfähigkeit der Bildverarbeitungsalgorithmen wird anhand von Beispielszenen demonstriert.

Contents

Chapter 1

Introduction

1.1 Motivation and Purpose

The work of this thesis was motivated by the research project **UseRCams**, which is a subproject of **PReVENT** [1]; the PReVENT project is an activity of the European automotive industry partly funded by the European Commission to contribute to road safety by developing and demonstrating preventive safety technologies and applications. Within this framework UseRCams is devoted to vulnerable road user (VRU) protection, collision mitigation and blind spot surveillance. Main objectives are thereby the development of an active camera device for 3-dimensional (3D) image acquisition of traffic scenarios and the development of adequate image processing algorithms for object segmentation, tracking and classification; these objectives are also central points of this thesis.

The following three different areas of future applications in vehicle near-to-intermediate surrounding (0.2 m - 20 m) are basic for the camera and algorithm development work within UseRCams (see also figure 1.1):

a) The **Front-View-Application** aims at the surveillance of the vehicle frontal proximity (2 m - 20 m) while driving in urban areas. The gained 3D information is being processed in real-time to provide information about potential dangers in this area. The protection of VRUs as well as the mitigation of collisions with preceding or oncoming vehicles is considered. As this scenario comprises the highest requirements to the camera device, the main focus in this thesis is set to this application.

b) The **Side-View-Application** addresses the monitoring of vehicle lateral proximity in the range from 0.2 m to 2 m; goal is the reliable detection and tracking of objects which potentially could hit the vehicle on the

side. This could improve the time-critical activation of airbags or other collision mitigation features enormously; an important feature in view of the risk potential of lateral impacts due to the poor lateral deformable zones of most vehicles.

c) The **Blind-Spot-Application** aims at the surveillance of blind areas in the surrounding of a vehicle. In UseRCams the focus is set on the monitoring of the frontal and lateral blind areas of tracks for 'start inhibit applications', i.e. if any objects are identified in the near surrounding of a truck (up to a distance of 4 m - 8 m) the driver is prevented from setting the vehicle in motion.

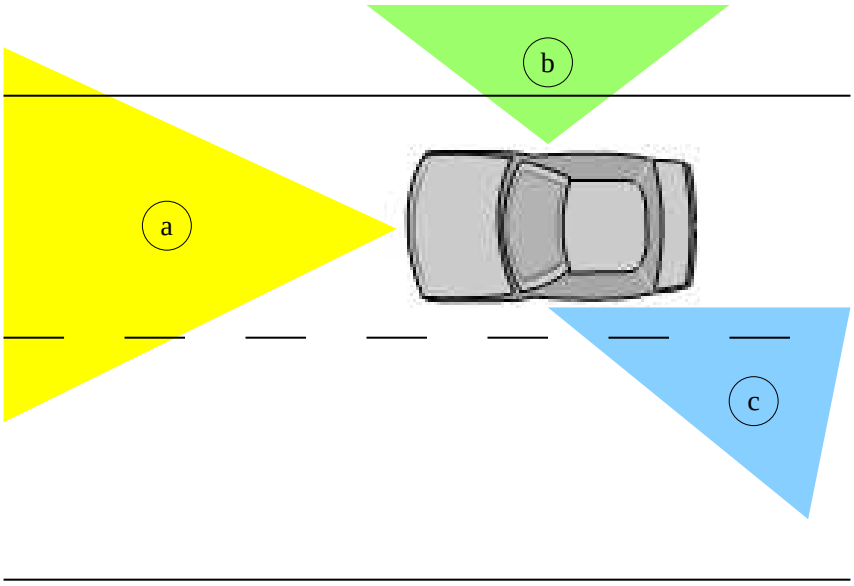


Figure 1.1: UseRCams covered areas

According to these planned future applications of the 3D imaging system following central points were derived as goals of the 3D camera system development in the specification phase of the project [2]:

- Sufficient *horizontal and vertical pixel resolution* to image persons up to distances of over 20 m without violation of the sampling theorem
- Sufficient *distance measuring accuracy* to provide accurate distance data for the subsequent image processing procedures
- Adequate *image repetition rate* to allow real-time image acquisition and image processing

- *High dynamic range* image sensor to cover the whole range of required distance (2 m - 20 m for the front view application) and object reflectivity (5% to 100% Lambertian) dynamics with a distance measurement accuracy of 3%

As targets for the development of the image processing software were defined:

- Reliable *segmentation* of objects in the 3D raw data
- *Tracing* of objects over frame sequences
- Derivation of estimates for position, velocity and acceleration data from sequences of measured object position data by *Kalman filtering*
- Object *classification* according to the application requirements

The here presented work describes the details of the UseRCams 64x8 pixel time of flight (TOF) camera development and the novel image processing approach according to the above mentioned development targets. See the following section for a detailed survey of the contents of this thesis.

1.2 Structure of this Work

This thesis describes and characterizes the advanced 3D time of flight (TOF) camera system and the novel image processing approach for the gained 3D data. Its structure is as follows:

Chapter 1 introduces the subject matter by pointing out the motivation for the work described in this thesis and gives a short overview over the contents of this thesis.
In **Chapter 2** an introduction of non-contact distance measurement techniques is given; this should make clear the preeminent advantages of the here used pulse based TOF measuring principle over other distance measurement techniques in the considered applications.
Chapter 3 introduces the theory of pulse based distance measurement. Starting from the basic concept of the TOF measurement, over the signal-theoretic description of the distance information extraction, it leads to a description of several distance derivation algorithms.
Chapter 4 is devoted to the description of the 3D-CMOS sensor system. In 4.1 all components are described, which are investigated using various measurements in 4.2. 4.3 addresses camera calibration issues.
A close examination of the different distance measuring algorithms is presented

in **Chapter 5**. Based on mathematical analysis, simulation results, and experiments, the algorithms are compared to derive the best possible system configuration. Additionally, techniques for managing the influence of adverse effects, such as ambient light, rain, and snowfall are introduced.
Chapter 6 describes the image processing algorithms that are applied to the distance image sequences. Main parts are here the object segmentation in chapter 6.3 and the object tracking based on Kalman filtering in chapter 6.4.
Chapter 7 gives an overview over the variety of possible application areas of the 3D-CMOS sensor.
Finally **Chapter 8** recapitulates the main results and innovations gained through this work and presents further approaches for promising research in this area.

Chapter 2

Overview of Non-Contact Distance Measurement

The measurement of distances is essential for many technical applications. The classical *direct distance measurement*, which means the direct comparison of the distance with a calibrated ruler, is the oldest and most obvious method, but not applicable in many cases. Therefore various *indirect distance measurement* procedures were developed throughout the centuries; here the distance is derived of any distance depending measure which is easier to access than the distance itself. The most important subgroup of the indirect distance measurement is the *non-contact distance measurement*, as many technical applications require distance measurement without any physical contact between the distance meter and the measured object. Therefore, up from the beginning of the 20^{th} century measurement procedures using sound waves and electromagnetic waves to transfer the distance information to the measuring instrument have been developed. Figure 2.1 gives an overview of todays most common non-contact distance measuring methods, which are closer described in the following sections.

2.1 Sonar and Radar

Sonar (sound navigation and ranging) systems derive the distance between the sonar device and an object via time of flight (TOF) measurement [3]. As depicted in figure 2.2, a sonic measuring impulse is emitted by the sender and propagates through the supporting medium (air, water, etc.) with the according propagation speed v. If the sound waves meet objects on their course of propagation, the waves are partly reflected back. This reflections can be detected by the sender/receiver device, what enables a measurement of the time τ_{TOF} elapsed between the sending and the receiving of the sound impulse. In-

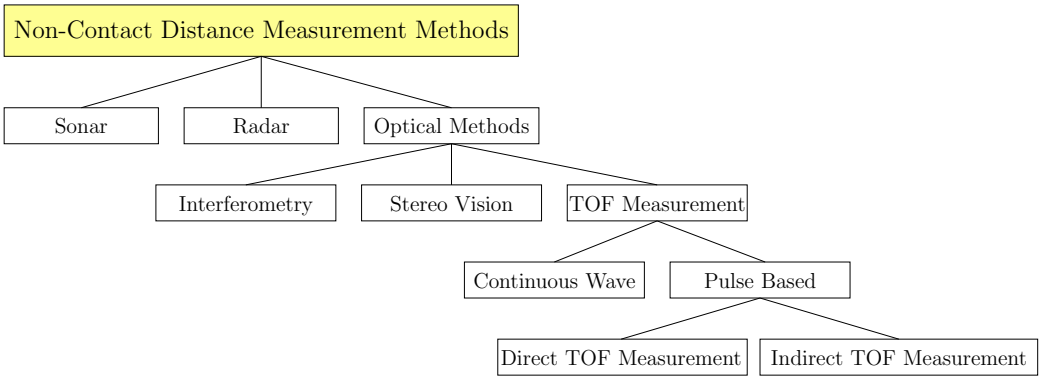


Figure 2.1: Non-contact distance measurement methods

serting τ_{TOF} and v in the TOF-*equation* (2.1) results in finding the searched distance of the object.

TOF-equation:

$$d = v \cdot \frac{\tau_{TOF}}{2}, \tag{2.1}$$

with d being the sought-after object distance from the measurement device, v the velocity of propagation of the used wave and τ_{TOF} the time measured between sending and receiving of the measurement signal.
The use of several sonar sensors in parallel or scanning of a single sensor in horizontal and/or vertical direction enables the acquisition of 3-dimensional distance images.
Many different frequency bands are used in sonar systems depending on the application, from 20 Hz for long distance sonars in submarines, over 40 kHz at ultrasonic distance sensors in the automotive and industrial fields (parking aid, etc.), up to 1 MHz - 40 MHz for diagnostic ultrasonic imaging in the medical area.
Main disadvantages of this technology are low repetition rates of the measurement resulting from the comparatively slow propagation speed of the sound waves (compared to electromagnetic waves), as well as the poor horizontal and vertical resolution due to the propagation behavior of sound waves.

Radar (radio detection and ranging) technology [4] again is based on TOF measurements. But here microwaves (100 MHz - 300 GHz) are used as measuring signal, whose time of flight between sender, reflecting object, and back is determined. As the propagation velocity of electromagnetic waves is the speed of light ($c \approx 3 \cdot 10^8$ m/s), the measured time delays of the emitted impulses are

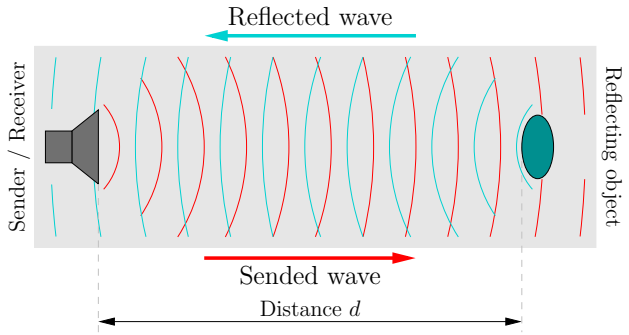


Figure 2.2: Principle of time of flight measurements

very short (6.67 ns/m). This necessitates very accurate evaluation electronics, but on the other hand results in very fast measurement repetition rates. Hence scanning procedures for 2D or 3D radar imaging are made feasible. Besides the distance information, a radar system can also determine the velocity of a target by evaluating the Doppler shift of the signal; this is e.g. used for traffic speed control.
Radar distance measurement is used for a variety of applications over a huge range of different distances. Examples are distance measurement for automatic cruise control in automobiles (up to 200 m), air traffic surveillance (distance range up to 500 km), or the mapping of planets in our solar system.
But also radar systems suffer from their poor angular resolution properties. Complex antennas can lower this problem, but this again leads to bulky systems which are not practicable for many applications.

2.2 Optical Distance Measurement Techniques

Non-contact distance measurement methods based on light waves are common techniques for many different distance measuring applications. In the following the most important ones are introduced:

2.2.1 Interferometry

Interferometry is a very precise technique to measure distance [5] indirectly from a phase shift of light waves. Figure 2.3 shows the Michelson interferometer to make clear the measuring principle.

A laser source emits a coherent laser beam of a fixed frequency f onto a semipermeable mirror that divides the beam into two parts. The first part is deflected towards a reference target with calibrated distance, and is then partly

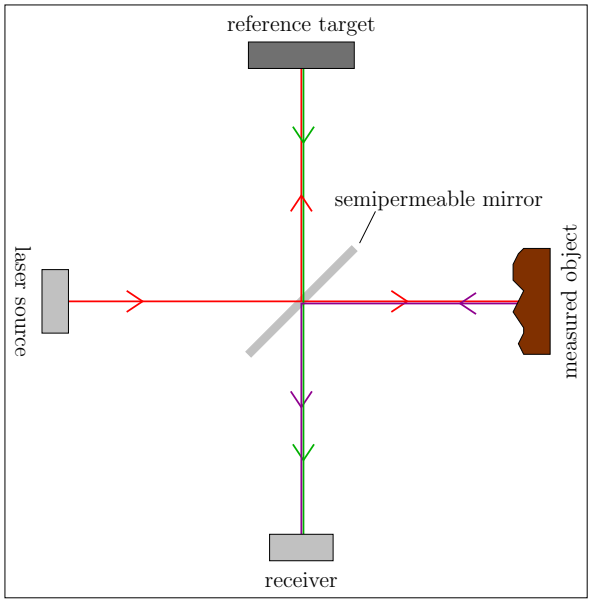


Figure 2.3: Michelson interferometer

reflected into the detector from there. The second part passes the semipermeable mirror and illuminates a point on the measurement target. Parts of the incident light are reflected from there towards the detector over the semipermeable mirror. In the detector the portions from the reference target and the measurement target interfere with each other to a resulting signal. By moving the beam splitter mirror a detector amplitude characteristics is derived which contains the relative distance information for the measured target point in its phase. By evaluating the phase shift $\Delta\phi$ of this signal, the relative distance can be determined as

$$\Delta d = c \cdot \frac{\Delta\tau_{TOF}}{2} = c \cdot \frac{\Delta\phi}{4 \cdot \pi} \cdot \frac{1}{f} \quad . \tag{2.2}$$

With up to $\lambda/100$ the achievable distance resolution is in the range of a few nm, but unambiguous measurements are only possible within a distance range of a half λ. This is no problem at relative distance measurements for i.e. surface scanning, but makes absolute distance measurements impossible. To solve this problem techniques like using several wave lengths are known.

2.2.2 Triangulation

As 2D imaging does not provide depth information of a captured scene, *triangulation* is used to generate 3D image data of a scene using standard 2D

imaging systems. The idea is thereby the observation of a scene from two different camera positions, i.e. under different observation angles. On this basis different techniques are known to get 3D data; here *stereo vision* (section 2.2.2.1) and *active triangulation* (section 2.2.2.2), two very common triangulation variants, are introduced.

2.2.2.1 Stereo Vision

Stereo vision [5] uses two cameras that observe a scene from two different positions, as depicted in figure 2.4a (intersection through a 3D scene). Thus the optical centers of the two cameras and the observed object point P form a triangle, whose basis length b and the two angles α and β are known. Consequently all other quantities, including the height d – the searched object distance – can be calculated. This is done by means of the so-called 'disparity', which contains the distance information of the object point P. Equations 2.3 - 2.6 show the derivation of the disparity p.

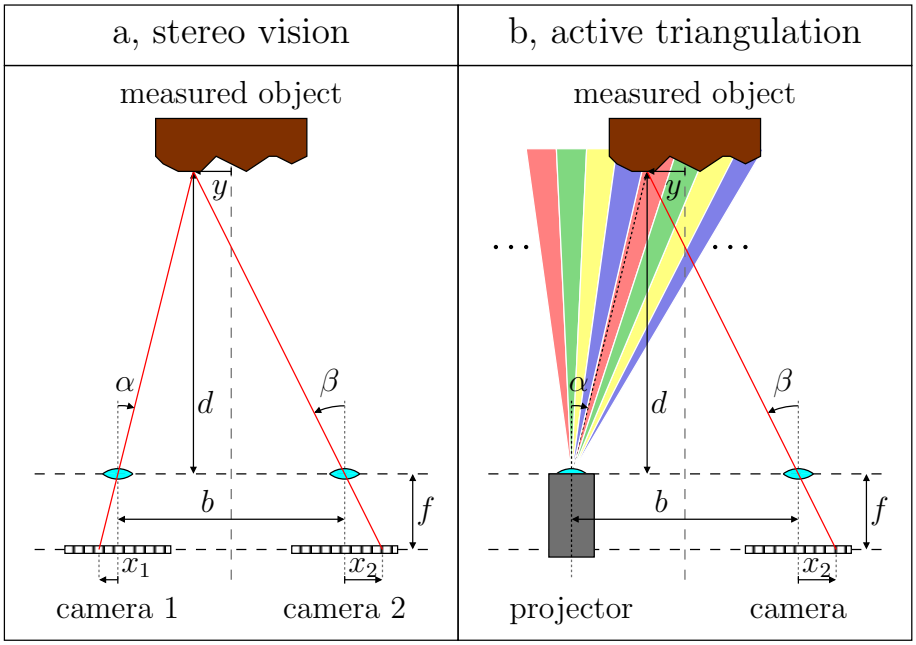


Figure 2.4: Principles of stereo vision and active triangulation

$$
\begin{aligned}
p &= x_1 - x_2 && (2.3)\\
&= f \cdot \tan(\alpha) - f \cdot \tan(\beta) && (2.4)\\
&= f \cdot \frac{y + b/2}{d} - f \cdot \frac{y - b/2}{d} && (2.5)\\
&= b \cdot \frac{f}{d} && (2.6)
\end{aligned}
$$

Resolving equation 2.6 for d and inserting equation 2.3 leads to following simple formula for the calculation of the distance d of object point P:

$$d = \frac{b \cdot f}{p} = \frac{b \cdot f}{x_1 - x_2} \ , \tag{2.7}$$

with x_1 and x_2 being the position of the image of object point P on the imaging sensor of camera 1 and camera 2 respectively.

A disadvantage of the stereo vision method is the need of separating the two cameras. The larger the distance range is, the longer the required basis width b becomes, in order to keep the measurement error in a reasonable range (cf. equation 2.8).

$$\Delta d = \frac{b \cdot f}{p^2} \cdot \Delta p = \frac{d^2}{b \cdot f} \cdot \Delta p \tag{2.8}$$

This is applicable for static measurements, but not for mobile systems. Also the necessity of having high contrast in the scene surfaces for the identification of identical points in the two pictures is often not tolerable. Further shading problems can occur; for scene points that are visible for only one camera, no distance value can be calculated.
Applications for stereo vision are for example 3D vision for robot systems and workpiece scanning in automated production lines.

2.2.2.2 Active Triangulation / CCT

Contrary to stereo vision, *active triangulation* [5] uses only one camera and a light source instead of the second camera. This light source projects a light pattern onto the observed scene which is imaged on the image sensor of the camera (see figure 2.4b). Thus again a triangle is formed, here between the light source, the object point and the camera. For calculation of the distance of a scene point, the knowledge, from which direction the point was illuminated (angle α in figure 2.4b), is required. Using a repeating pattern of different

intensities for different illumination angles, the sensor can decode this information and re-calculate the angles. But the repetition of the pattern leads to an ambiguity problem, which can severely affect the resulting 3D image, especially for scenes containing complex depth information. This can be avoided by combining several measurement steps using light patterns with different spatial frequencies; this on the other side is a time consuming procedure. A more advanced and effective method is the so-called *color coded triangulation* (CCT) [6], which uses an unambiguously coded pattern of light stripes of different colors for scene illumination.

For active triangulation the formula to calculate the distance value of point P is

$$d = \frac{b \cdot f}{p} = \frac{b \cdot f}{f \cdot \tan(\alpha) - x_2} \ , \tag{2.9}$$

with α being the angle of the light beam illuminating object point P and x_2 being the position of the image of object point P on the imaging sensor of the camera.

Disadvantage of the active triangulation method is, besides the basis width and shading problems described in 2.2.2.1, the presence of the scene illumination pattern, which is disturbing in many applications, specially if humans are part of the observed scene. Newest systems solve this problem using illumination in the infrared spectrum, invisible to the human eye [7].
Another problem of this approach is the sensitivity towards ambient light. This means that measurements can severely be disturbed, if light from outside the camera system outshines the projectors light pattern. Bandpass filtering of the incident light according to the illumination wavelength is possible to lower that influence.
3D face recognition for security applications as well as 3D measuring tasks for industrial application are only two of many possible applications for CCT systems.

2.2.3 Time of Flight Distance Measurement

According to the TOF distance measurement by usage of sound waves or microwaves (cf. chapter 2.1), several TOF distance measurement procedures using light waves are known. They all actively emit light in order to derive the distance of a reflecting object by measuring the time which the light needs to propagate to the object and return to the combined emitter/sensor device. The different approaches are described in the following paragraphs:

2.2.3.1 Continuous Wave Methods

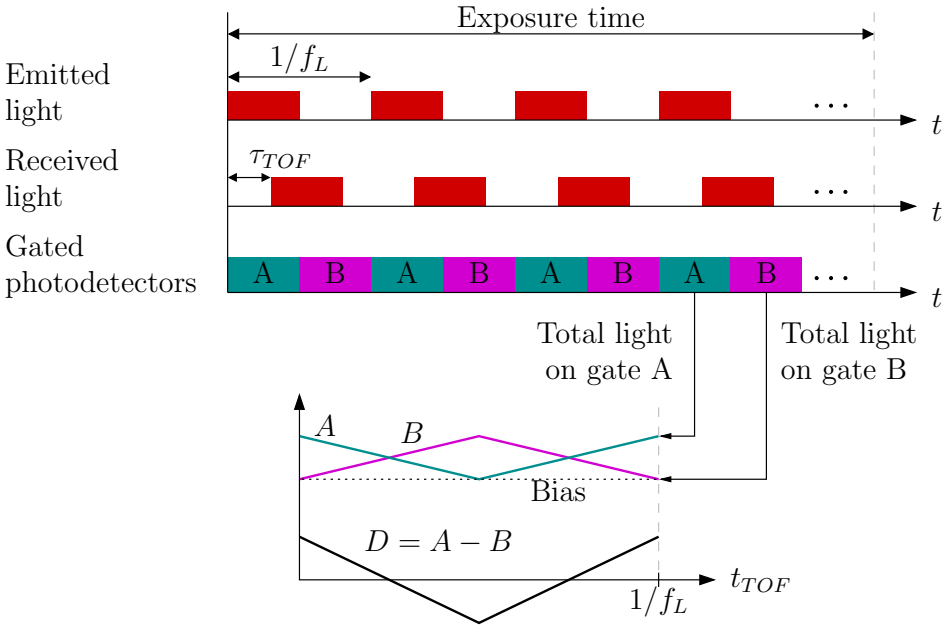


Figure 2.5: Principle of continuous wave time of flight distance measurement according to [8]

Continuous wave (CW) distance measuring systems [8,9,10] derive the desired distance information from the phase shift of an emitted light pulse sequence. For this purpose usually IR LEDs are modulated with a modulation frequency f_L (typically in the range of 5 MHz to 100 MHz) with a duty cycle of 50%. Synchronized with the emitted light, photodetectors A and B are alternately gated during the exposure time that is usually 1 ms to 50 ms (see figure 2.5). Thus the phase shift of the received signal due to the time of flight τ_{TOF} directly influences the amount of photocharges at the outputs of A and B (lower part of figure 2.5). A difference signal $D = A - B$ is calculated for cancelling of biases. For the derivation of the phase ϕ a second observation of D with the timing of the gated photodetectors A and B shifted by 90° is conducted. Phase ϕ can then be calculated according to equation 2.10 as

$$\phi = \tan^{-1}\left(\frac{D_{90}}{D_0}\right) \quad . \tag{2.10}$$

Following the object distance can be found as

$$d = c \cdot \frac{\tau_{TOF}}{2} = c \cdot T_{PW} \cdot \frac{\phi}{4 \cdot \pi} \quad . \tag{2.11}$$

A weak point of the continuous wave methods are ambiguity problems, i.e. the range of uniqueness is limited to phase shifts of $0 \leq \phi \leq \pi$. The distance of objects causing a higher phase shift is hence misinterpreted. As solution algorithms combining several different modulation frequencies are used; this for sure is a more time-consuming procedure, resulting in lower image repetition rates. Also the sensibility towards ambient light due to the comparatively long integration duration is a problem for CW distance measurement systems; it causes pixel saturation that makes a distance measurement impossible. Much effort with respect to background light suppression – like filtering the incoming light, resetting photodector bias for A and B several times during exposure, etc. – were taken to make these systems competitive in view of outdoor applications [8, 11].

2.2.3.2 Pulse Based Methods

In contrast to the CW methods, the *pulse based distance measurement* procedures utilize the time of flight of single light pulses to determine the distance of objects. Thereby one distinguishes between two different approaches, the *'direct'* and the *'indirect'* pulse based TOF measurement.

2.2.3.2.1 Direct TOF Measurement

The direct pulse based TOF measurement physically measures the time of flight of an emitted laser pulse. I.e., at the emission of a short light pulse a high-speed counter is started, which is stopped when the return of a reflection of the pulse was detected in the detector device [12]. The distance of the reflecting object is then given as

$$d = c \cdot \frac{\tau_{TOF}}{2} \quad . \tag{2.12}$$

Due to the complexity of the detection and evaluation electronics, a parallelization in a high-resolution sensor array is not feasible at the present time. Hence the available systems either provide only a few measurement points like the lidar-systems (light detection and ranging) [13] or apply scanning techniques by means of rotating mirrors, to measure one scene point after the other (laser scanners [14]).

2.2.3.2.2 Indirect TOF Measurement (3D-CMOS)

Here the time of flight of the emitted light pulse is not measured directly but via accumulation of charges on a sensor chip from which the distance information

can be calculated. Thus the procedure is equal to the CW method from chapter 2.2.3.1 using only a very short single light pulse (typically 20 ns - 200 ns) instead of a pulse sequence. For the derivation of τ_{TOF} several procedures are known [15, 16]. Concept is always based on the use of imaging pixels with extremely short exposure times in the range of the laser pulse length, which are capable to produce TOF-dependent signals. The object distance d is then calculated with the TOF-equation (2.1). In this thesis the so-called 3D-CMOS sensor, which performs indirect pulse based TOF distance measurement, and its distance derivation algorithms are addressed. See chapters 3 and 4 for a close description of these topics.

Chapter 3

Theory of Distance Measurement with 3D-CMOS Sensor

3.1 Measuring Principle

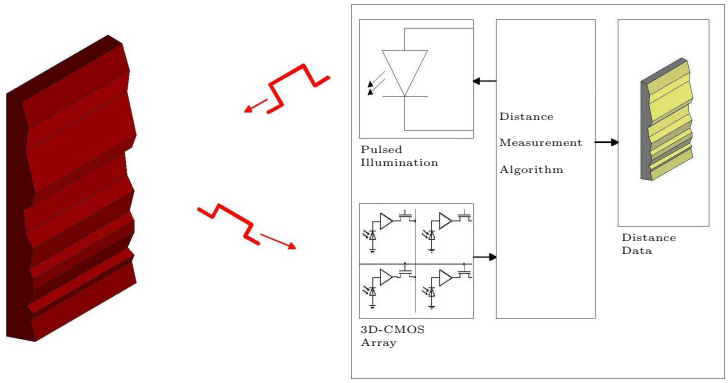


Figure 3.1: 3D-CMOS measuring principle

The *3D-CMOS sensor system* performs a pulse based indirect TOF measurement using near infrared (NIR) laser pulses in the range of 20 ns - 200 ns. As figure 3.1 shows, these pulses are sent by the internal illumination module of the camera; after diffuse reflection by the observed scene objects they return temporally delayed and with reduced amplitude to the camera device. The camera lens passes a part of the backscattered light and thus images the observed scene onto the surface of the 3D-CMOS sensor chip. The chip (cf.

chapter 4.1.1) contains light sensitive pixels which are operated as electronic short-time windowed integrators synchronized to the emitted laser pulse for TOF dependent collection of charges generated by the incident light. Different evaluation algorithms have been developed (chapter 3.4) to derive the pulse delay τ_{TOF} from a specific number of single measurements. This leads to the object distance

$$d = c \cdot \frac{\tau_{TOF}}{2} \ . \tag{3.1}$$

3.2 Camera Optics

A *lens camera model* is used for all considerations in this thesis, as shown in figure 3.2. The camera optics is thereby defined by an imaging lens of focal length f, a sensor plane parallel to the lens plane and an aperture limiting the effective diameter of the lens to D. Its optical axis is defined as the straight line through the optical center O that is perpendicular to the sensor plane.

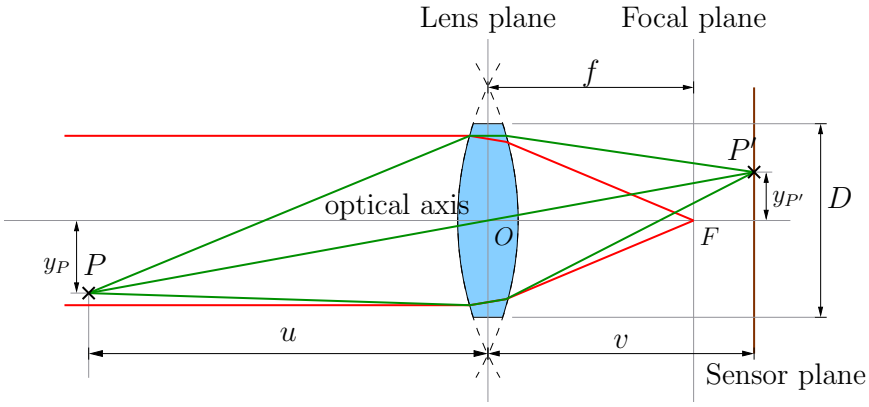


Figure 3.2: Imaging geometry

The mapping of a point P in the object space to the image space is described by the *lens law* [17]:

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \tag{3.2}$$

with u and v being the distances of P and P' normal to the lens plane, respectively.

Actually, for each object distance u an unique image distance v exists, at which the sensor has to be positioned at to obtain a sharp image P' of P. But here

we assume large object distances compared to the focal length ($u \gg f$), so that v can be approximated as

$$v = \frac{1}{\left(\frac{1}{f} - \frac{1}{u}\right)} \approx f \text{ with } u \gg f \ . \tag{3.3}$$

With equation 3.3 the magnification M of the image in relation to the object is defined as

$$M = \frac{v}{u} = \frac{f}{u} \ , \tag{3.4}$$

what leads to equation 3.5, describing the perspective projection of objects on the sensor plane:

$$y_{P'} = -y_P \cdot M = -y_P \cdot \frac{f}{u} \tag{3.5}$$

A further key parameter of the camera lens is the f-number $f_\#$, which is the ratio of the focal length f to the aperture diameter D:

$$f_\# = \frac{f}{D} \tag{3.6}$$

It is a measure for the photo-sensitivity of the camera optics, but also influences the imaging properties [17].

In the following considerations effects like chromatic or spherical aberrations or blurring due to the the approximation taken in equation 3.3 are not included. Their influence on the imaging properties of the camera is assumed as negligible in this thesis. Justification for this assumption will be given in chapter 4.2.1.4.

3.3 Signal Generation

This section deals theoretically with the generation of signals with inherent distance information using the 3D-CMOS camera. I.e., ideal signal shapes (e.g. a rectangular impulse is assumed to model a laser light impulse; in reality a laser pulse certainly has finite leading and trailing edges) and ideal signal processing operations are applied to model the processes during the distance acquisition procedure. To what extent these idealizations are justified, and which adaptations have to be done in practice will be discussed in chapter 4.

3.3.1 Pulsed Illumination

The scene illumination module of the 3D-CMOS sensor generates pulses with a pulse width in the range of some tens to hundreds of nanoseconds. Here the

pulses are assumed to be ideal rectangular pulses according to equation A.2 in appendix A.1 with amplitude $\hat{P}_L$ and pulse width T_{Pw}. See equation 3.7 as well as figure 3.3 for the laser modules radiant power characteristics for a single laser pulse.

$$P_{Laser}(t) = \hat{P}_L \cdot rect\left(\frac{t - \frac{1}{2} \cdot T_{Pw} - T_{DL}}{T_{Pw}}\right). \tag{3.7}$$

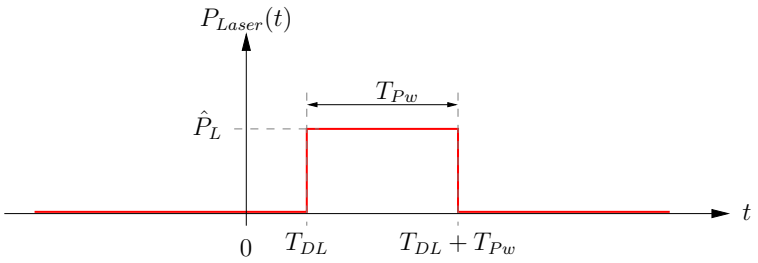


Figure 3.3: Ideal laser-pulse function

The shift by $\frac{1}{2} \cdot T_{Pw}$ towards positive t-direction is introduced to make the pulse start at $t = 0$ s, the additional shift by T_{DL} denotes an purposely inserted delay of the laser pulse by means of a delay component on the camera hardware (see chapter 4.1.4). This feature is used by some of the later examined distance derivation algorithms.

3.3.2 Reflected Signal

According to figure 3.1 the laser pulse is sent towards the scene, the distance of which is to be determined. Following boundary conditions are postulated regarding the illumination:

- The whole emitted radiant power illuminates an area A_{Object} in the object space, image of which in the image space corresponds to the active sensor area A_{Sensor}; i.e. no light power is lost by illuminating those parts of the scene that are not imaged onto the sensor
- The radiant power is homogeneously distributed over the area A_{Object}
- The laser source and the sensor chip are positioned close together; thus they can be assumed as being located at the same point

The delay of the laser light due to the distance d between camera and object (TOF) amounts to

$$\tau_{TOF} = \frac{2 \cdot d}{c} \ , \tag{3.8}$$

so that the irradiance on the sensor surface is given by

$$E_{Sensor}(t, \tau_{TOF}, \alpha) = \frac{P_{Laser}(t - \tau_{TOF})}{A_{Object}} \cdot \rho \cdot \kappa \cdot \cos^4(\alpha) \cdot \frac{1}{4 \cdot f_{\#}^2} \ , \tag{3.9}$$

with

$P_{Laser}(t - \tau_{TOF})$	radiant power of the laser module shifted by τ_{TOF}
α	angle between optical axes and camera ray
A_{Object}	irradiated object area
ρ	object surface reflectivity
κ	optical loss factor of camera and laser module optics
$f_{\#}$	f-number according to equation 3.6.

The $\cos^4$ dependency on the viewing direction α is closely investigated in [18]. For the following theoretical considerations α is assumed as 0, i.e. the object area close the the optical axis is considered.

Substituting

$$A_{Object} = A_{Sensor} \cdot \left(\frac{d}{f}\right)^2 \tag{3.10}$$

in equation 3.9, $E_{Sensor}(t, \tau_{TOF}, \alpha)$ can be calculated as

$$E_{Sensor}(t, \tau_{TOF}, \alpha = 0) = E_{Sensor}(t, \tau_{TOF}) = \frac{P_{Laser}(t - \tau_{TOF})}{A_{Sensor}} \cdot \rho \cdot \kappa \cdot \frac{f^2}{4 \cdot f_{\#}^2 \cdot d^2} \ . \tag{3.11}$$

For the following parts of this chapter the amplitude of the received laser light irradiance $E_{Sensor}(t)$ on the sensor surface is assumed as being independent from the object distance d and the object reflectivity ρ for reasons of transparency of the line of argument. Though this assumption does not correspond to the reality situation, it does not falsify the considerations.

3.3.3 Short Time Integration

The 3D-CMOS sensor chip implements short-time windowed integration of the photocharge generated by the illumination of the sensor surface. The duration

of these so-called *shutter windows* is in the range of the laser pulse duration, i.e. in the nanosecond range (cf. section 4.2.1.2). The shutter windows can be described as rectangular windows (see annex A.1) of the sensitivity value of the sensor (figure 3.4), which relates the pixel diode current to the incident radiant power. It is thus measured in A/W. Whilst the sensitivity is zero in the normal case, it becomes $\hat{S}_{\lambda S}$ during the integration time T_{Int}. For a given wavelength λ the sensitivity thus results as

$$S_{\lambda Sensor}(t) = \hat{S}_{\lambda S} \cdot rect\left(\frac{t - \frac{1}{2} \cdot T_{Int} - T_{DS}}{T_{Int}}\right). \tag{3.12}$$

where T_{DS} is an additional shift of the shutter window introduced for later use during distance measurement.

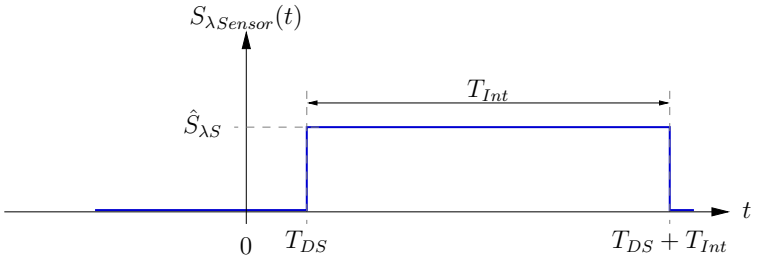


Figure 3.4: Ideal sensor sensitivity function

3.3.4 Integrator Output Signal

The incident light reflected by an observed scene generates a diode current $I_D(t, \tau_{TOF})$ in the light sensitive sensor pixels as described by following equation:

$$I_D(t, \tau_{TOF}) = S_{\lambda Sensor}(t) \cdot E_{Sensor}(t, \tau_{TOF}) \cdot A_{Pixel} \tag{3.13}$$

The integration of $I_D(t, \tau_{TOF})$ at the integration capacitance C_{Int} (which consists of the diode inherent capacitance C_D plus the sense capacitance C_{sense} as one can see in figure 4.3) during the integration time T_{Int} generates an integrator voltage U_{Int} in the considered pixel:

$$U_{Int}(\tau_{TOF} = \tau_1) = \frac{Q_{photo}(\tau_{TOF} = \tau_1)}{C_{Int}} = \frac{\int_{-\infty}^{\infty} I_D(t, \tau_{TOF} = \tau_1)dt}{C_{Int}} \tag{3.14}$$

Doing this calculation for all possible values of τ_{TOF} results in a function that describes the integrator output voltage in dependence on the time of flight of the laser pulse, i.e. in dependence on the object distance. The qualitative characteristics of $U_{Int}(\tau_{TOF})$ for given times T_{Pw} and T_{Int} can be seen in figure 3.5; note that irradiance variations due to object distance and reflectivity are – as already mentioned in section 3.3.2 – not considered here.

$$U_{Int}(\tau_{TOF}) = \frac{\int_{-\infty}^{\infty} I_D(t, \tau_{TOF})dt}{C_{Int}} \tag{3.15}$$

$$= \frac{\int_{-\infty}^{\infty} S_{\lambda Sensor}(t) \cdot E_{Sensor}(t, \tau_{TOF}) \cdot A_{Pixel}dt}{C_{Int}} \tag{3.16}$$

$$= \frac{Q_{photo}(\tau_{TOF})}{C_{Int}} \tag{3.17}$$

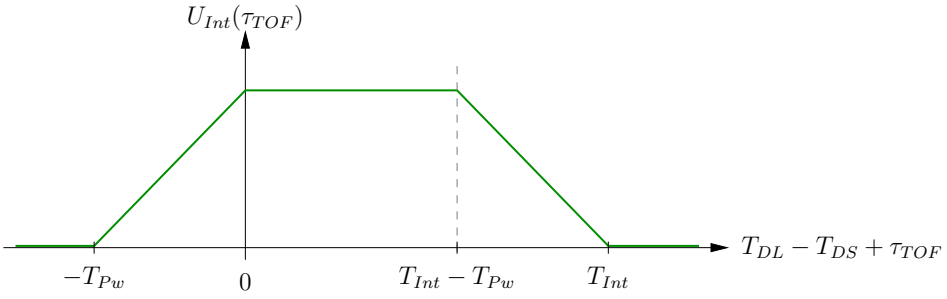


Figure 3.5: Ideal integrator output function for variation of the laser pulse delay

Over readout circuitry (cf. chapter 4.1.1.1) the integrator output voltage U_{Int} is transferred to the output buffer of the image sensor and is then available for further processing as sensor output signal.

3.4 Distance Derivation Algorithms

Using the sensor output signals derived in the previous section, the derivation of an object distance to be measured from them is the first goal of the further processing steps. For this several distance derivation algorithms exist; basic idea of all of them is the recovery of the distance information from reflected pulses by using short-time shutter windows that overlap the incident light pulse

depending on the delay τ_{TOF}. Four elementary relative positions between the incident pulse and the shutter window can be distinguished, as shown in figure 3.6:

a) The overlap decreases with increasing τ_{TOF}; the pulse 'moves out' of the shutter window for increasing distances.

b) The overlap increases with increasing τ_{TOF}; the pulse 'moves into' the shutter window for increasing distances.

c) Due to the long duration of the shutter window τ_{TOF} has no influence on the overlap and thus no influence on the sensor signal. The sensor signal is a measure for the pulse overall radiant energy. The use will become clear in the next chapter.

d) This kind of shutter window to pulse arrangement gives an signal partly independent of τ_{TOF} (as long as it is fully 'filled' by the pulse) and partly dependent on τ_{TOF}. In practice it is not used. It will thus not be taken into account in the later details.

The lower part of figure 3.6 shows the sensor output signals $U_a(\tau_{TOF})$ to $U_d(\tau_{TOF})$ for the above shutter window and pulse alignments with respect to the laser impulse delay τ_{TOF}. These signals are the basis for the calculation of the object distance one is searching for.
The following sections introduce all feasible distance acquisition algorithms without further addressing of performance behavior or other features. An investigation toward those issues will be given in chapter 5.

3.4.1 MDSI

The MDSI (multiple double short-time integration) algorithms yield the distance information from two single sensor signals (from the set of a), b) or c) in figure 3.6), as the word 'double' in the name implies. The meaning of 'multiple' will be explained later in section 4.1.1.3. Four reasonable combinations of two of these sensor signals are possible in order to obtain the distance of the measured object, consequently four MDSI variants exist:

MDSI1
MDSI1 extracts the necessary information from one shutter of category a), and one of category c). Although the first shutter operation generates distance information in the output signal $U_a(\tau_{TOF})$, the searched distance d can not be retrieved directly, as the sensor output signal is also influenced by the

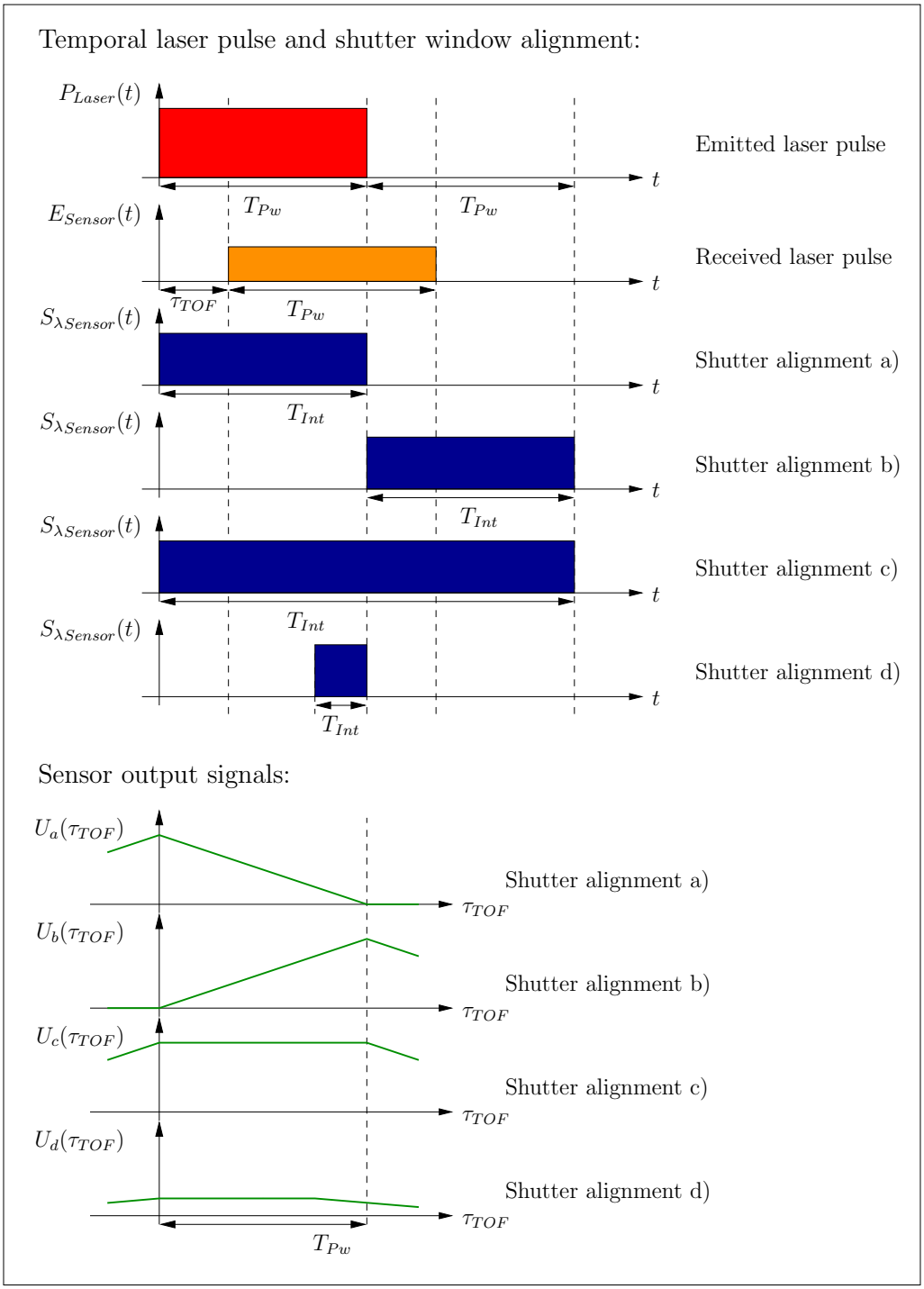


Figure 3.6: Possible alignments of a shutter window relatively to the incident light pulse

irrandiance value of the reflected light pulse. By normalizing the signal with the pulse overall irradiance value, which is measured by a shutter window operation of the type c), the distance can be calculated using the MDSI1 equation stated below:

$$d = \frac{c}{2} \cdot \left(T_{Pw} - T_{Pw} \cdot \frac{U_a}{U_c} - (T_{DL} - T_{DS}) \right) \tag{3.18}$$

where T_{Pw}, T_{DL} and T_{DS} denote the width and the relative temporal position of the according laser pulses and shutter windows as described in sections 3.3.1 and 3.3.3.

MDSI2
The MDSI2 procedure is quite similar to MDSI1. Solely a type b) instead of the type a) shutter window is applied. Thus the distance calculation equation changes to

$$d = \frac{c}{2} \cdot \left(T_{Pw} \cdot \frac{U_b}{U_c} - (T_{DL} - T_{DS}) \right) \ . \tag{3.19}$$

MDSI3
Here the long shutter window of type c) is substituted by the sum of two short shutter windows of type a) and b) for the distance calculation. This is possible, because the sum signal of a) and b) (cf. figure 3.6) equals the sensor output of c).

$$d = \frac{c}{2} \cdot \left(T_{Pw} - T_{Pw} \cdot \frac{U_a}{U_a + U_b} - (T_{DL} - T_{DS}) \right) \tag{3.20}$$

MDSI4
MDSI4 corresponds to MDSI3 unless a class b) shutter window instead of class a) is used in the enumerator of the fraction term in the d-equation:

$$d = \frac{c}{2} \cdot \left(T_{Pw} \cdot \frac{U_b}{U_a + U_b} - (T_{DL} - T_{DS}) \right) \tag{3.21}$$

All these algorithms deliver the identical results of the measured object distance. The measurement range is for all limited to

$$0 \leq d \leq c \cdot \frac{T_{Pw}}{2} \ . \tag{3.22}$$

3.4.2 Gradient Method

Contrary to the MDSI algorithms, this method does not process a TOF dependent signal, which is normalized by a second ,TOF independent, signal, but two TOF dependent shutter windows are used. Their temporal alignment is depicted in figure 3.7. Basic principle is the use of the temporal difference ΔT_{Int} between the two shutter windows. The distance of the pulse reflecting object can be calculated by employing equation 3.23 below.

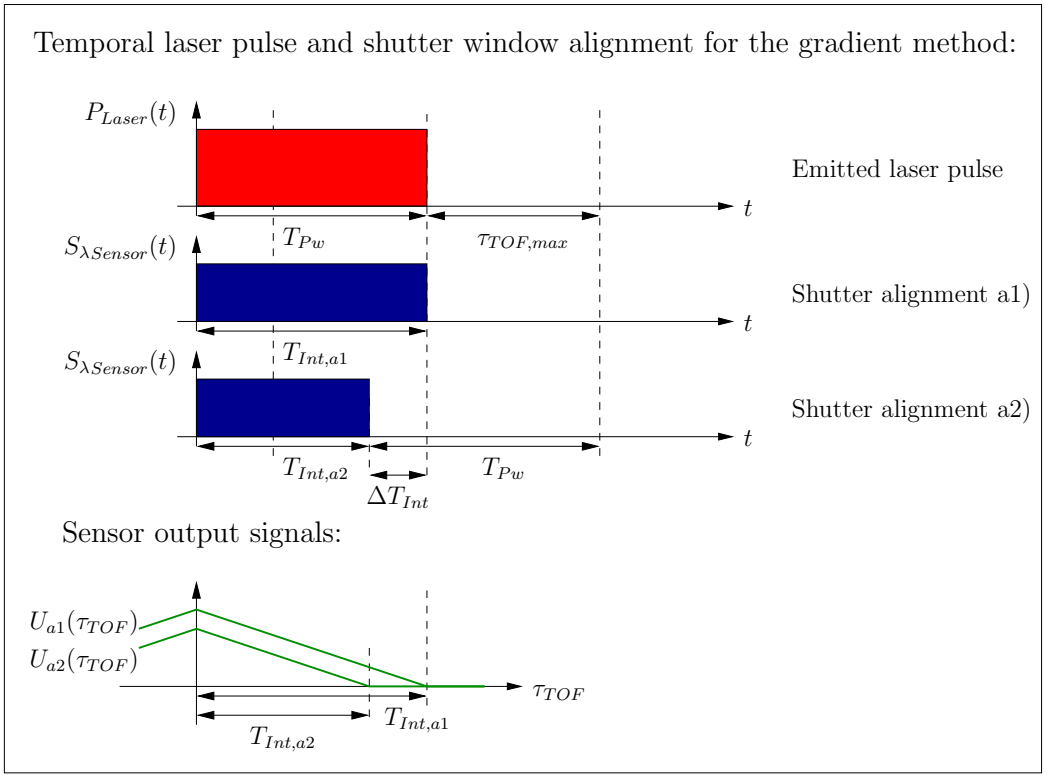


Figure 3.7: Alignment of a shutter window relatively to the incident light pulse at the gradient method

$$\begin{aligned} d &= \frac{c}{2} \cdot \left(T_{Int,a1} - (T_{Int,a1} - T_{Int,a2}) \cdot \frac{U_{a1}}{U_{a1} - U_{a2}} \right) \\ &= \frac{c}{2} \cdot \left(T_{Int,a1} - \Delta T_{Int} \cdot \frac{U_{a1}}{U_{a1} - U_{a2}} \right) \end{aligned} \tag{3.23}$$

Here also a variation of the method is possible with respect to the use of shutter windows of type b), i.e. the overlap of shutter window and reflected pulse is rising for increasing values of τ_{TOF}. The resulting equation slightly changes to

$$d = \frac{c}{2} \cdot \left(T_{Int,b1} - \Delta T_{Int} \cdot \frac{U_{b1}}{U_{b1} - U_{b2}} \right) . \tag{3.24}$$

3.4.3 CSI and DCSI

The *correlated short-time integration* (CSI) and *difference correlated short-time integration* (DCSI) procedures differ from the previous distance calculation methods as they utilize a sequence of shutter measurements instead of two single measurements one range acquisition.

CSI

In this method a sequence of shutter windows is applied to a sequence of equal shaped light pulses, whose trigger position compared to the shutter window is shifted by known delays $T_{DL,1}$, $T_{DL,2}$, etc. (cf. the schematic illustration figure 3.8). Reasonable is the usage of a shutter window width in the range of the light pulse width. Depending on the light pulse shift T_{DL} the intensity characteristics $U(T_{DL})$ can be derived (see lower part of fig. 3.8). Determination of the T_{DL}-value $T_{DL}(U(T_{DL})_{max})$, where the curve reaches its maximum, leads to the object distance by inserting into equation 3.25:

$$d = \frac{c}{2} \cdot (T_{DS} - T_{DL}(U(T_{DL})_{max})) \tag{3.25}$$

Though this procedure seems to be suitable for TOF distance measurement, its performance turned out to be inferior when compared to the other procedures introduced here [19]. Main disadvantage is the susceptibility to systematic errors at the determination of $T_{DL}(U(T_{DL})_{max})$ that makes the procedure unfeasible. Thus it is only mentioned here for the sake of completeness and will not be considered in further examinations.

DCSI

The DCSI distance derivation method is an advancement of the CSI method presented above. Two shutter signal sequences according to CSI are applied here with two shutter windows 1 and 2, that are temporally arranged as one can see in figure 3.9. The resulting output signal characteristics $U_2(T_{DL})$ of shutter 2 is then subtracted from $U_1(T_{DL})$ of shutter 1; this leads to the difference signal $U(T_{DL}) = U_1(T_{DL}) - U_2(T_{DL})$ shown in green color in figure 3.9. For distance calculation the T_{DL}-value of the zero-crossing $T_{DL}(U(T_{DL}) = 0) = T_{DL_0}$ of the medium part of the difference curve has to evaluated. Inserting this value into formula 3.26 yields the searched distance value.

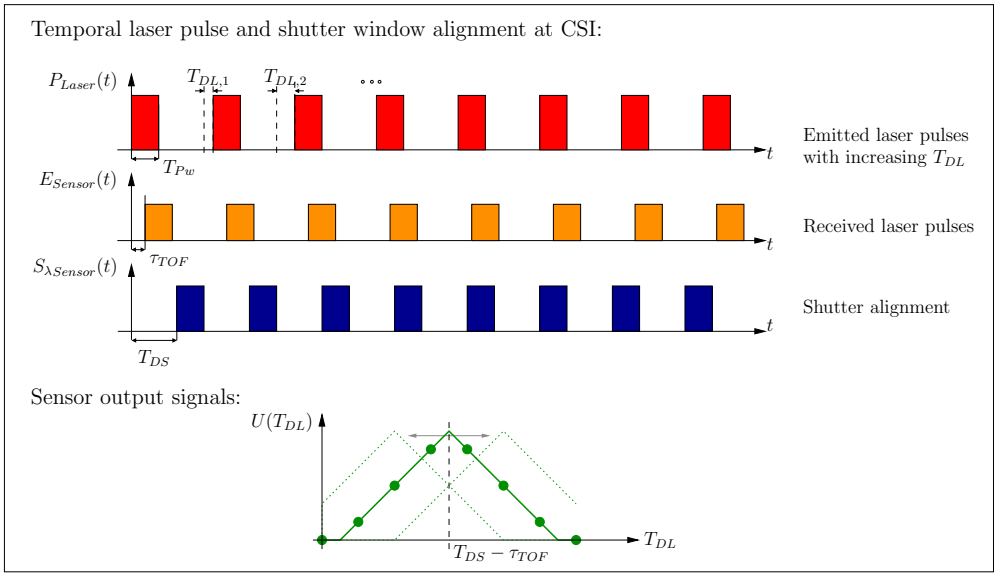


Figure 3.8: Example for the temporal alignment of laser pulses and shutter windows at the CSI method

$$\begin{aligned} d &= \frac{c}{2} \cdot \left(\frac{T_{DS,1} - T_{DS,2}}{2} - T_{DL}(U(T_{DL}) = 0) \right) \\ &= \frac{c}{2} \cdot \left(\frac{T_{DS,1} - T_{DS,2}}{2} - T_{DL_0} \right) \end{aligned} \tag{3.26}$$

The subtraction of the shutter signal curves brings many advantages compared to CSI, as elimination of potential systematic influences like offsets or nonlinearities of the shutter windows and the laser pulse; additionally the determination of the the time of flight becomes a line fitting through the measured T_{DL}-curve points for derivation of the zero crossing. See annex C for a mathematical description of this procedure.

3.4.4 Conclusion

All of the above distance measuring methods are suited to derive the distance of an scene object to be measured from an according set of sensor outputs. It is important that in all equations 3.18-3.26 any dependence on parameters P_{Laser}, A_{Object}, ρ, κ, α, and $f_{\#}$ has been eliminated. This is valid, however, only if none of these parameters has changed during the shutter window operations belonging to the same distance calculation equation.

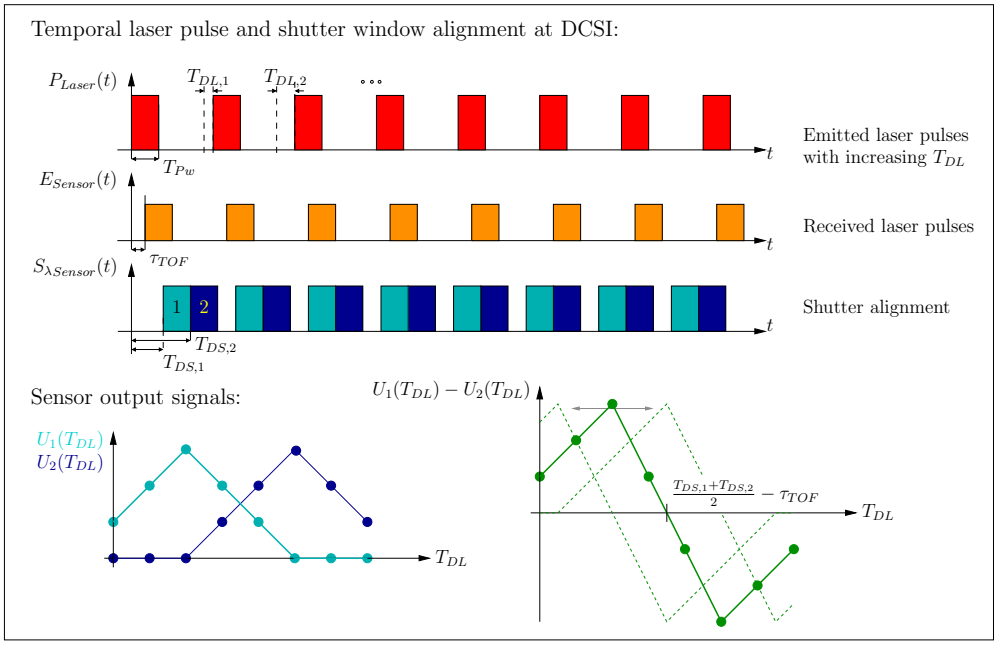


Figure 3.9: Distance measurement principle of the DCSI method

Differences between the methods are the reachable measurement accuracy as well as the applicability for the proposed sensor applications (cf. chapter 1). Chapter 5 introduces investigations regarding those issues.

Chapter 4

3D-CMOS Array Sensor System

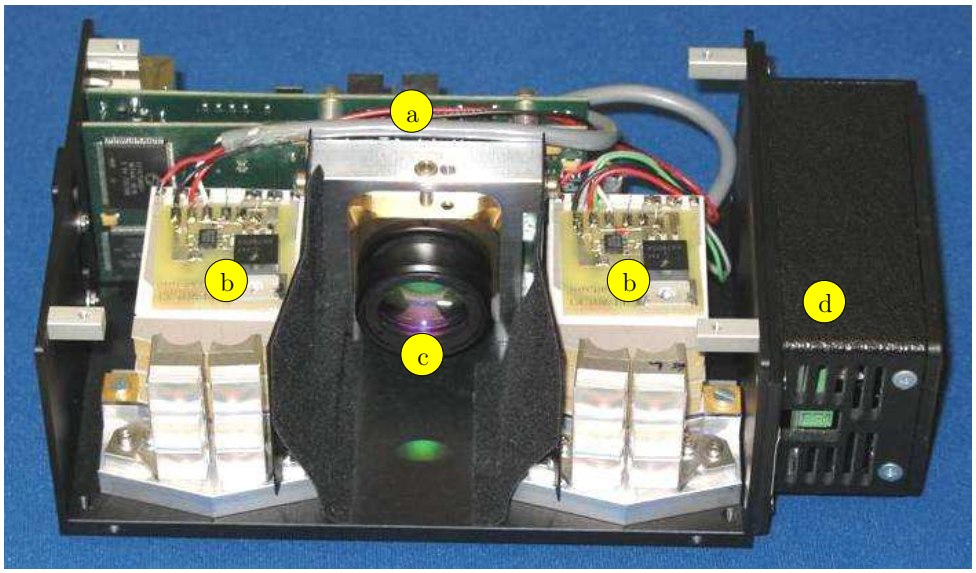


Figure 4.1: 3D-CMOS array sensor system

This chapter describes the 3D-CMOS array sensor system examined in this thesis and its performance characterization [20, 21]. As depicted in figure 4.1, the main components of the system are:

a) Electronics boards including the array sensor chip

b) Laser illumination modules

c) Imaging optics

d) Power supply

The following sections give a close description of the components and the conducted experiments.

4.1 System Components

4.1.1 Novel Array Sensor

The novel 64x8 pixel 3D-CMOS array sensor is the main component of the sensor system. It was developed by the Fraunhofer IMS within the UseRCams project on the basis of an existing 64x4 pixel 3D-CMOS sensor chip [22, 23]. Its functionality allows the indirect measurement of the time of flight of laser pulses in the ns-range.
Figure 4.2 shows the photograph of the sensor chip realized with a 0.5 μm n-well CMOS process. It includes 64·8 = 512 photo diode pixels, a suitable readout circuit per half-column including a CDS stage with analog accumulation capability (see sections 4.1.1.2 and 4.1.1.3), multiplexers for selecting the rows, as well as an analogue buffer circuit for each output channel (total 2). The light sensitive pixel area can be seen in the center of the chip. The pixel pitch is 130 μm in horizontal and 300 μm in vertical direction; this results in a sensitive area of 64·130 μm · 8·300 μm = 19.97 mm^2. As the whole peripheral pixel electronics is placed in the upper and lower areas of the chip, the active pixel areas join each other directly, i.e. the fill factor (FF) of the active pixel area is nearly 100 %.

4.1.1.1 Pixel Structure

The circuit diagram of the pixel circuit used in the 64x8 pixel 3D-CMOS array sensor is shown in figure 4.3 [23]. Each pixel circuit contains a reverse biased n-well/p-substrate photo-diode PD with a inherent capacitance C_D (5.1 pF), a reset switch ($\Phi 1$), a shutter switch ($\Phi 7x$), a sense capacitor C_{sense} (1 pF), a buffer SF (amplification factor g_{SF} = 0.85), a hold capacitor C_H, and a select switch ($\Phi 2x$). Each four pixel circuits located in one half-column always share a single readout switched-capacitor amplifier which also performs the correlated double sampling (CDS) operation (see section 4.1.1.2). This quad-to-single multiplex reduces considerably the hardware complexity.
Note that in the following the pixel index (0-3), as it is sometimes used in figure 4.3, is not considered, because the pixels of one row are absolutely identical. $\Phi 70$ for example is denoted as $\Phi 7$ and so on.
The circuit operation relies on a periodical reset of the photo-diode capacitance C_D and the sense capacitance C_{sense} to a fixed reference voltage (U_{ddpix}) and subsequent discharge due to the photo-current. The shutter switch ($\Phi 7$)

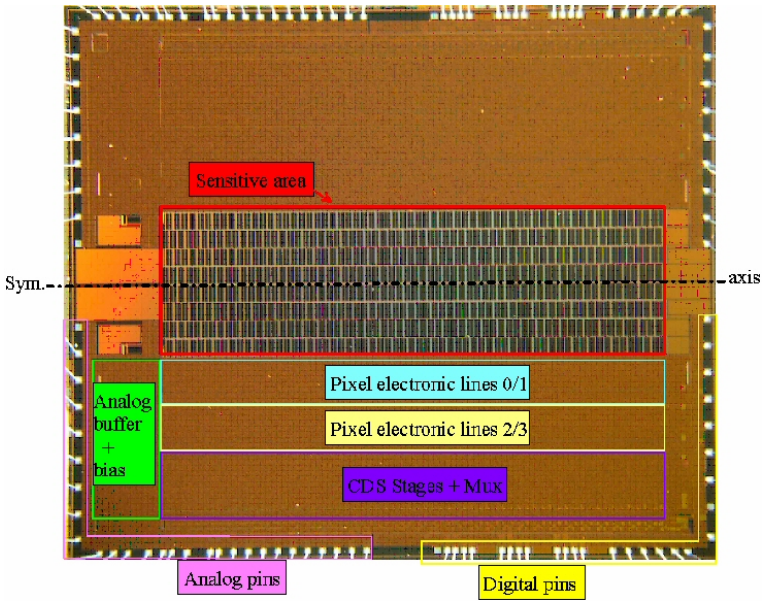


Figure 4.2: 64x8 pixel 3D CMOS sensor

controls the integration time of the discharge process, as depicted in figure 4.4. Then the remaining voltage stored at C_{sense} is read out into a second capacitor bank C_H acting as a hold capacitor. The charges held at C_H are read out using the correlated double sampling (CDS) stage by activating the select switch $\Phi 2$. Meanwhile, the acquisition of the next value on C_{sense} is performed, thus this architecture yields quasi-continuous light acquisition with minimum dead time.
After a signal acquisition cycle the pixel output signal after the CDS stage is present as $U_{Sensor,raw}$ on the sensor output. Given a sensor irradiance profile $E_{Sensor}(t)$ and a shutter window function $S_{\lambda Sensor}(t)$, this voltage is given as

$$U_{Sensor,raw} = U_{ref4} - \frac{Q_{photo}}{C_D + C_{sense}} \cdot g_{SF} \cdot \frac{C_{Cl}}{C_F} \quad , \tag{4.1}$$

where

$$Q_{photo} \overset{\text{Eq. } 3.16}{=} A_{Pixel} \cdot \int_0^{+\infty} S_{\lambda Sensor}(t) \cdot E_{Sensor}(t) dt \quad . \tag{4.2}$$

Note that this equation differs from equation 3.14 due to presence of reset voltage U_{ref4}, the sign inversion, and the signal amplification factor $g_{SF} \cdot \frac{C_{Cl}}{C_F}$ ($g_{SF} = 0.85$, $C_{Cl} = 10$ pF, $C_F = 2.5$ pF), because equation 4.1 reflects the real hardware implementation including the pixel readout circuitry.

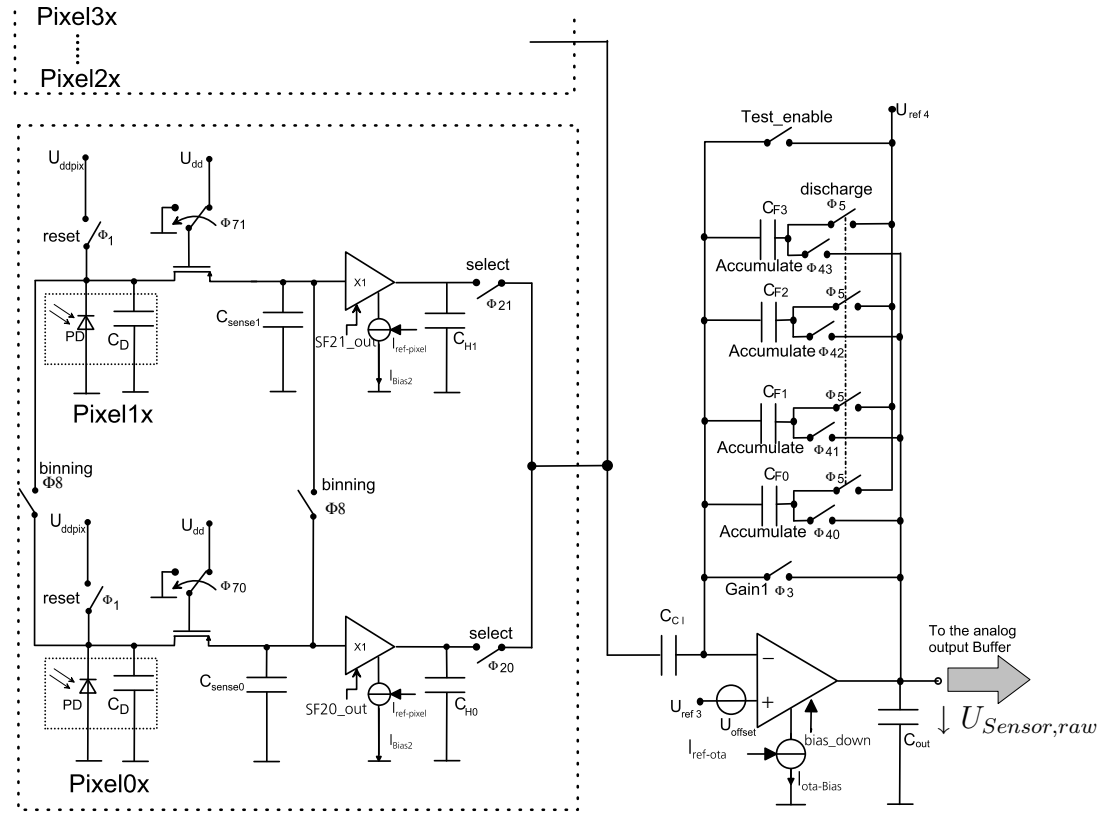


Figure 4.3: Pixel circuit with CDS readout stage [23]

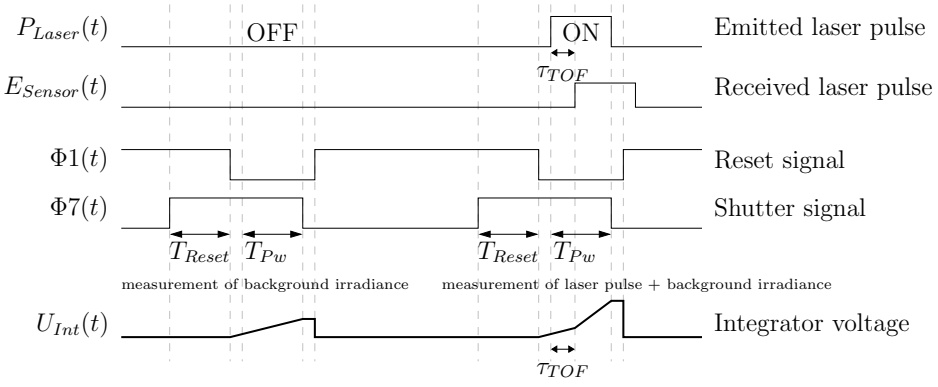


Figure 4.4: Simplified sensor timing [22]

Preprocessing of the sensor data in the system internal FPGA removes U_{ref4} of $U_{Sensor,raw}$ and inverts its sign. The result is the simplified sensor output voltage U_{Sensor}

$$U_{Sensor} = \frac{Q_{photo}}{C_D + C_{sense}} \cdot g_{SF} \cdot \frac{C_{Cl}}{C_F} \ , \tag{4.3}$$

which is used as sensor output signal in all further considerations.

4.1.1.2 Correlated Double Sampling

The amount of light received from an object not only depends on the emitted laser power $P_{Laser}(t)$, the reflectance ρ of the object and its distance, but also on the amount of $E_{Back}(t)$ due to the incident light of other light sources.
The influence of $E_{Back}(t)$ is eliminated by measuring solely the background irradiance without laser pulse, i.e. $P_{Laser} = 0$, and subtraction of this value from the measurement with laser pulse. Thus each measurement must be performed with laser pulse ON and OFF (cf. figure 4.4) and the difference being stored on the corresponding capacitance C_F in the analog memory. The whole procedure

$$1: \ U_{sense}|_{ON} \propto E_{Sensor}(t) + E_{Back}(t) \tag{4.4}$$

$$2: \ U_{sense}|_{OFF} \propto E_{Back}(t) \tag{4.5}$$

$$1 - 2: \ \stackrel{CDS}{\Longrightarrow} U_{Sensor} \propto E_{Sensor}(t) \tag{4.6}$$

results in a pixel output signal that is independent of $E_{Back}(t)$ if the two measured values of E_{Back} are correlated. This assumption is valid, if E_{Back} does not change within the CDS cycle period, what is always granted in normal environments. The corresponding control signals of the sensor are shown in figure 4.4. Note that we do not have to carry out an additional reset sampling in this case, we simply measure with the laser ON and OFF and subtract the measurements.

4.1.1.3 Analog Accumulation and Adaptive Accumulation

A unique feature of the 3D-CMOS image chip is the analogue real-time on-chip accumulation process at each individual pixel using multiple pulses for each shutter window [15]. By this means the signal-to-noise ratio increases and leads to an improved distance accuracy. Repetitive accumulation of n_{acc} laser pulses (cf. figure 4.5) is performed adaptively up to the saturation level at each pixel element. Intelligent procedures can be employed to cover a large

range of target distances and target surface reflectivities with this adaptive pulse illumination method (see section 4.2.3.3).

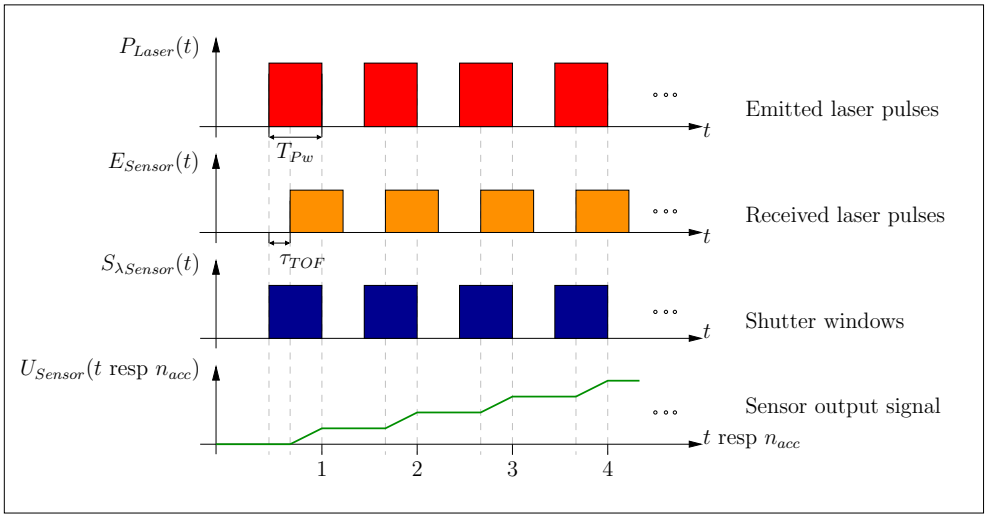


Figure 4.5: Analog accumulation principle

The resulting pixel output signal is then ideally given by

$$U_{Sensor}(n_{acc}) = \sum_{i=1}^{n_{acc}} U_{Sensor}(1) = n_{acc} \cdot U_{Sensor}(1) \ , \tag{4.7}$$

while the standard deviation of the accumulated signal increases only with the square root of n_{acc}; Gaussian distributed and uncorrelated sensor signal noise is assumed.

$$\sigma_{U_{Sensor}}(n_{acc}) = \sqrt{\sum_{i=1}^{n_{acc}} \sigma_{U_{Sensor}}(1)^2} = \sqrt{n_{acc}} \cdot \sigma_{U_{Sensor}}(1) \ , \tag{4.8}$$

Result is an improved signal-to-noise ratio (SNR) of the pixel circuit output:

$$SNR(n_{acc}) = \frac{U_{Sensor}(n_{acc})}{\sigma_{U_{Sensor}}(n_{acc})} \tag{4.9}$$

$$= \frac{n_{acc} \cdot U_{Sensor}(1)}{\sqrt{n_{acc}} \cdot \sigma_{U_{Sensor}}(1)} \tag{4.10}$$

$$= \sqrt{n_{acc}} \cdot SNR(1) \tag{4.11}$$

The SNR gain factor is thus $\sqrt{n_{acc}}$ for n_{acc} accumulated pulses. Certainly analog accumulation can only be performed as long as the resulting pixel signal does not saturate. Therefore, intelligent sensor control is implemented to adapt n_{acc} to the corresponding pixel signal. This procedure is called *adaptive accumulation*.

4.1.1.4 Nondestructive Readout

For the adaptive accumulation various laser pulses are accumulated for each pixel of the sensor. In non-destructive mode of the sensor pixel reset is carried out at the start of the bit pattern (see section 4.1.1.5 below) and the charge is accumulated with every laser pulse. Data are read out intermediate without pixel reset depending on defined integration steps. In the here examined system they were set to 1,4,16,64 and 100. By this means only data with selected pulse numbers are transferred into the memory, avoiding time consuming data transfers.

4.1.1.5 Sensor Control / Bit Pattern

For sensor control a set of reference voltages and currents (cf. U_{ddpix}, U_{ref3}, $I_{ref-pixel}$, etc. in figure 4.3) as well as several control signals (Φ1-Φ7) are provided from peripheral components on the circuit boards and the FPGA controller (see sections 4.1.4 and 4.1.5).
The signals Φ1-Φ7 entirely operate the image acquisition process, the CDS and the readout procedure. For it a signal description table, the so-called *bit pattern*, is executed in the FPGA to output the according 'high' or 'low' signals at each time step (time basis is a clock cycle of 30 ns).

4.1.2 Laser Illumination

A decisive role in the distance measuring procedure with the 3D-CMOS camera plays the scene illumination unit, since it provides the light signals, which are the carrier of the distance information to be derived. A modular concept for illumination using laser diodes as light emitters was developed, which allows assembly of modules for various applications. The basic element consists of a laser module with a variable number of laser diodes bonded directly on a ceramic substrate. Driver electronics is also integrated. Combining several basic modules with appropriate optical lenses for beam forming a variety of illumination requirements can be fulfilled.

4.1.2.1 Laser Light Source

In order to be suited for distance measuring, a laser light source was developed accomplishing following features:

- Generation of approximately rectangular pulse shape with variable pulse length from 50 ns to 200 ns
- Steep rise and fall times in the range of 10 ns
- Application dependent assembly of laser pulse peak power up to 2 kW
- Pulse repetition frequency up to 20 kHz
- Invisible to the human eye (wavelength of 905 nm in the NIR range)
- Fulfillment of eye safety regulations for laser class 1

4.1.2.2 Beam Forming Optics

Another important issue regarding the illumination module is the beam forming optics. Its task is the precise, application specific laser beam forming with a high degree of homogeneity by having low losses for laser wavelength of 905 mm and a compact design. For this reason especially designed cylindrical lens arrangements are used, which perform focusing of the light emitted by the laser diodes to illuminate exactly the area that is imaged onto the sensor. Additionally, homogenization of the resulting beam is accomplished to obtain constant light intensity over the whole illuminated area.
For the front-view application special demands are set to the illumination profile, since higher viewing ranges are desired for the center pixels (high distance viewing range along the road, less distance viewing range off the road) [2]. For this reason the single laser beams – each itself providing a homogeneous beam profile – are superposed, in order to get an overlapping region in the middle of the viewing angle.

See section 4.2.2 for a deeper investigation of the properties of the laser modules.

4.1.3 Imaging Optics / VOV

According to the application requirements [2] two different imaging lenses with focal lengths of 15.55 mm and 7.30 mm, yielding a horizontal opening angle of 30° and 60° respectively and a vertical opening angle of 9° and 18° respectively (resulting from the dimensions of the sensitive area of the sensor chip in section

4.1.1) were designed and fabricated in molding technology. In order to achieve an optimal irradiation on the sensor following constraints had to be fulfilled:

- High optical aperture
- Resolution better than the pixel size of the 64x8 3D-CMOS array chip
- No distortion for the 64x8 3D-CMOS array chip
- Image size: 8.32 mm x 2.40 mm (active sensor area)
- Anti-reflective coating for the laser wavelength at 905 nm.

Under conformance with the other constraints the resulting customized imaging lenses feature excellent light efficiency, as one can see in table 4.1.

Table 4.1: Properties of the customized imaging lenses

Parameter	30° lens	60° lens
Focal length	15.55 mm	7.30 mm
Aperture diameter	19.40 mm	7.60 mm
Resulting $f_{\#}$	0.8	0.96

Different from conventional 2D camera systems, where the *field of view* (FOV) is a well known expression for the solid angle spanned by horizontal and vertical opening angle of the camera, here the *volume of view* (VOV) of the 3D-CMOS camera system is introduced which is defined by horizontal and vertical opening angle and the minimum and maximum measurement distance of the camera (i.e. for the UseRCams front view application camera the VOV is a 30° times 9° wide region in space ranging from a distance of 2 m to 20 m). Hence the VOV of a certain 3D-CMOS camera gives the region in space where reasonable distance measurement can be performed with this camera.

4.1.4 Electronics

The electronic design is based on a multi-board architecture. Label 'a' in figure 4.1 shows the printed circuit boards of the sensor electronics. The so-called *CMOS Board*, which is placed towards the cameras objective, contains the

CMOS image sensor, current and voltage references for the sensor, an analog-to-digital converter (ADC) for the conversion of the analog sensor signal, a delay component for temporal adjustment of the laser trigger, and a FPGA for system control and data processing. For noise optimization, special emphasis has been put on the placement of analogue and digital components and to the selection and design of voltage and current regulators.
The second board acts as interface board and includes the power supply sockets for the electronics (12 V) and the laser modules (40 V - 48 V), and development and application interfaces (100 BASE-T, RS-232, JTAG, ...).

4.1.5 Firmware / Software

For operation of the camera, firmware was designed that runs on the FPGA included on the CMOS board (see above). It performs

- configuration of the system
- control of the CMOS sensor, i.e. sequence control of the sensors bit pattern and provision of the according control signals
- control of the ADC
- synchronization of shutter window and the laser source (laser trigger)
- data processing:
  - decision if a sensor output value is high enough to be a valid measurement signal by comparison with the so-called noise threshold
  - sophisticated evaluation of the adaptive accumulation according to section 4.1.1.3
  - distance calculation from the raw data and the calibration parameters (see section 4.3) appropriate to the chosen distance derivation algorithm
  - optional averaging over n_{av} images to reduce the noise; this on the other hand reduces the camera frame-rate (cf. section 4.2.3.5)

Further, a software package for camera control and image processing by a PC (connection to the 3D-CMOS camera device via LAN (100 BASE-T)) was developed. It contains following parts:

- Camera *interface library* providing functions for camera control and image acquisition

- *Image processing library* for image processing of the camera output data (cf. chapter 6)
- *User interface application* integrating the interface library and the image processing library

With these firmware/software components full operation of the camera including image processing can be carried out.

4.2 System Characterization

Within the UseRCams project, the described 3D camera was successfully developed and implemented. In order to characterize the camera performance various measurements were conducted, which are described below. Single component examination of the image sensor chip and the illumination modules was accomplished as well as a characterization of the over-all camera system performance.

4.2.1 Image Sensor Characterization

Since the properties of the image sensor chip decisively influence the performance of the over-all camera system, special emphasis is placed on the characterization of the image sensor. The examination of the basic functionalities (responsivity, shutter function, analog integration, ...) as well as the analysis of distortions is described. A closer noise examination is accomplished together with the system noise measurements in chapter 4.2.3.1.

4.2.1.1 Responsivity

A measure for the light sensitivity of the image sensor chip is the *spectral responsivity* in $\frac{\mathrm{V}}{\mathrm{J/m^2}}$. It relates the sensors output signal U_{Sensor} to the incident radiant exposure H with

$$H = \int_0^{T_{Int}} E_{Sensor}(t)dt \ . \tag{4.12}$$

Using equations 4.3, 3.13 (considering only the exposure phase of the sensor yields $S_{\lambda Sensor}(t) = \hat{S}_{\lambda S}$), 3.14, and 4.12 yields for our sensor a theoretical responsivity

$$R_\lambda = \frac{U_{Sensor}}{H} = \frac{\hat{S}_{\lambda S} \cdot A_{Pixel}}{C_D + C_{sense}} \cdot g_{SF} \cdot \frac{C_{Cl}}{C_F} \tag{4.13}$$

for a given wavelength λ.

For the measurement of H a plain target at a distance of 1 m from the 3D-CMOS camera was illuminated with homogeneously distributed laser pulse light with a wavelength of 905 nm. The reflected light was acquired with the camera and also measured with a reference photo diode [24]. The diode was placed next to the camera and furnished with the identical objective as the camera. Thus it was assured as good as possible that the irradiance on the sensor plane and on the reference diode are the same. From the temporal characteristics of the reference diode output signal, the according irradiance characteristics

$$E(t) = k \cdot U_{Diode}(t) \tag{4.14}$$

is derived, where k is a constant diode internal scaling factor. Integration over $E(t)$ according to equation 4.12 leads to the radiant exposure H. This was done for several different illumination values. Simultaneously the sensor response on the incident light was acquired. Figure 4.6 shows the measurement result; the measured points for different illumination strengths are marked with crosses. A straight line was fitted through the points, which is depicted as dashed line. The gain value of this line is the sought responsivity value $R_{905\,\mathrm{nm}}$, which was derived as

$$R_{905\,\mathrm{nm}} = 2617 \frac{\mathrm{V}}{\mathrm{J/m^2}}. \tag{4.15}$$

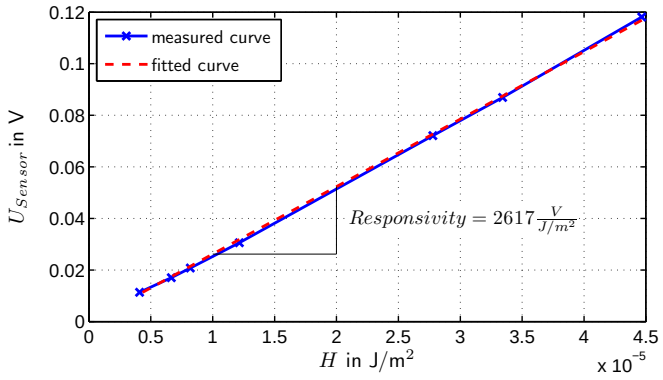


Figure 4.6: Sensor responsivity measurement at $\lambda = 905$ nm

4.2.1.2 Shutter Function

In order to obtain the shape of the shutter windows of the sensor pixels, the sensor is illuminated with very short laser pulses of about 100 ps pulse duration. As these short pulses can be regarded as Dirac impulses (see annex A.2) compared to the length of the shutter windows, shifting these pulses over the shutter window in small steps provides a sampled image of the shutter window, i.e. a sampling of the responsivity characteristic of the according sensor pixel. The result of this measurement for different shutter lengths can be seen in figure 4.7.

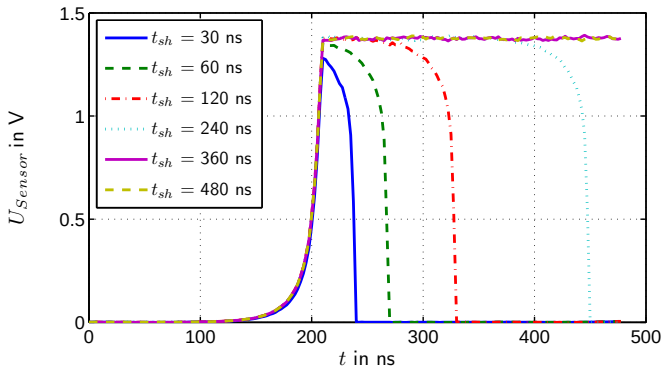


Figure 4.7: Sampled shutter window function for different shutter lengths t_{sh}

Different from the ideal rectangular shutter function the leading and trailing edges of the shutter functions are slightly distorted here. The trailing edge begins to rise about 60 ns - 70 ns before the shutter switch is closed to start the charge collection. This is due to diffusion effects in the photo diode, that cause charges generated by the laser pulse prior to the shutter closure to contribute to the photo current a few tens of ns later.
The same effect causes the bending of the shutter window characteristics 60 ns - 70 ns prior to the opening of the shutter switch; i.e., charges that are generated already by the laser pulse can not contribute to the photo current before the charge integration ends, as they did not diffuse to the diode region fast enough. Besides that, the shutter functions exhibits good characteristics; a stable responsivity level as it is required to get good linearity of the correlation function of shutter-window and laser pulse is available. The influence of the non-ideal edges will be discussed in section 4.2.3.2.

4.2.1.3 Analog Accumulation

The adaptive accumulation technique introduced in 4.1.1.3 utilizes the capability of the sensor to accumulate charges on the output capacitance of the CDS stage, called the *analog accumulation*. The accumulation characteristics were examined with respect to the linearity of the increase of the sensor signal for different numbers of analog accumulation (cf. figure 4.8).

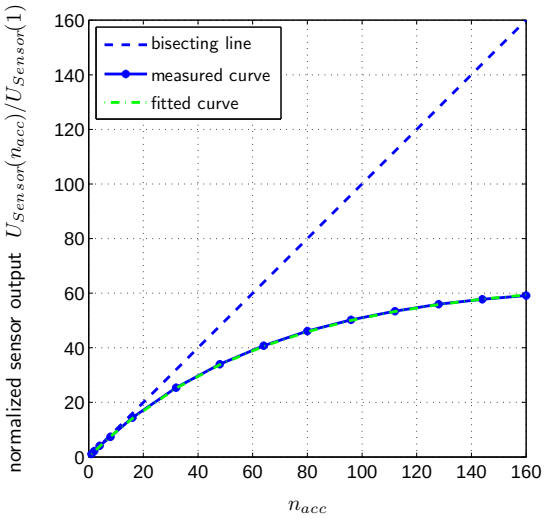


Figure 4.8: Damping of the output signal increase during analog accumulation

As one can see, the accumulation (see section 4.1.1.3) does not behave ideally according to

$$U_{Sensor}(n_{acc}) = n_{acc} \cdot U_{Sensor}(1) \qquad \text{for} \quad n_{acc} \in \mathcal{N} \tag{4.16}$$

but the increase of the sensor output is subject of a damping which lowers the signal increase for increasing analog accumulation count n_{acc}. For analysis of this damping effect, a damping function $d(n_{acc})$ is fitted into the measured sensor output values of figure 4.8:

$$d(n_{acc}) = \frac{1 - e^{-0.0153 \cdot n_{acc}}}{(1 - e^{-0.0153}) \cdot n_{acc}} \tag{4.17}$$

The real sensor output values are thus given by

$$U_{Sensor}(n_{acc}) = d(n_{acc}) \cdot n_{acc} \cdot U_{Sensor}(1) \qquad \text{for} \quad n_{acc} \in \mathcal{N} \ . \tag{4.18}$$

Though this effect does reduce the benefit of the analog accumulation, it does not falsify the distance measurement result, as it influences all shutter signals the same way. Thus the influence is eliminated by the quotient generation at the distance calculation (cf. chapter 3.4).

4.2.1.4 Optical / Electrical Crosstalk

Measurements of scenes comprising strong reflectivity gradients, which induce strong irradiance gradients between adjacent sensor pixels, revealed an intense crosstalk influence of the highly illuminated sensor parts on the signals of the lowly illuminated sensor areas. I.e., the output signals were distorted towards the high sensor signals; the magnitude of this effect strongly depends on the pitch between the observed 'low intensity' pixel and the influencing 'high intensity' region.
This *crosstalk* effect can be ascribed to optical and electrical influences:

- Optical:
  - *Blurring*: In good approximation the camera focus can be assumed to be set to infinity; thus the image of a point near to the camera is blurred, i.e. is a disc on the sensor surface instead of a sharp point. The radius of this disc is

$$r_{blur}(u) = \frac{D}{2} \cdot \frac{\left(\left(\frac{1}{f} - \frac{1}{u}\right)^{-1} - f\right)}{f} \tag{4.19}$$

    The worst case is the minimum distance to be measured, at the here considered front-view application 2 m. The resulting blurring disc radius is 76 μm (30° objective), more than half of the pixel width of 130 μm This means that blurring can strongly influence neighbor pixels, but no pixels that are farther away. The influences of chromatic or spherical aberrations can be neglected here due to the large pixel dimensions and the monochromatic illumination.
  - *Diffraction*: Diffraction happens when the incident light passes the objective of the camera. The diffraction image follows the Bessel function of first kind [25]. For the front view setup, the calculation shows, that at a distance of 3 μm from the imaged point the intensity of the diffraction image declined below 0.1 % of the peak intensity. Thus this effect is neglectible compared to the blurring and is not further examinated.

  - *Diffusion*: Small amounts of the light entering a camera objective are unpredictably diffused inside the objective and can thus fall on arbitrary positions of the sensor. In literature the amount of the diffused light is quoted between 0.5% and 4%, depending on the quality of the lenses, the clearness of the ambient air and other factors. Nevertheless even 0.5% of very bright light can be enough distortion for sensor regions with low signals to considerably influence the measured signal.
    Another potential source of diffusion is the cover glass of the sensor chip. Multiple reflections between the glass surface and the sensor surface could lead to interference over distances of multiples of the pixel width.

- Electrical:

  - Global changes of reference voltages or currents can occur due to high currents in highly illuminated pixels. Such effects can directly influence the sensor output signals of all pixels via resistive or capacitive coupling. Naturally pixels with low signal amplitudes would suffer much more than those exhibiting high signals.

Unfortunately an isolation of the effects and thus a measurement of the influence of the single effects could not be achieved. Thus it has to be assumed that the *crosstalk* is a mixture of the different effects described above.

4.2.2 Laser Illumination Characterization

Besides the image sensor, the laser illumination unit is the second important system component that directly affects the quality of the distance measurements. See the following sections for basic characterization of the laser illumination module.

4.2.2.1 Pulse Shape

Pulse shape measurement yields the temporal characteristics of the laser module output signal for different length of the laser pulses defined by trigger lengths t_{Tr} between 30 ns and 240 ns. A calibrated photo diode [24] was used to measure the curves which are depicted in figure 4.9.
Differing from the ideal rectangular pulse model, the module shows rise times of about 60 ns and fall times of about 30 ns. Also the plateau level which should be constant over a major part of the pulse duration exhibits a round

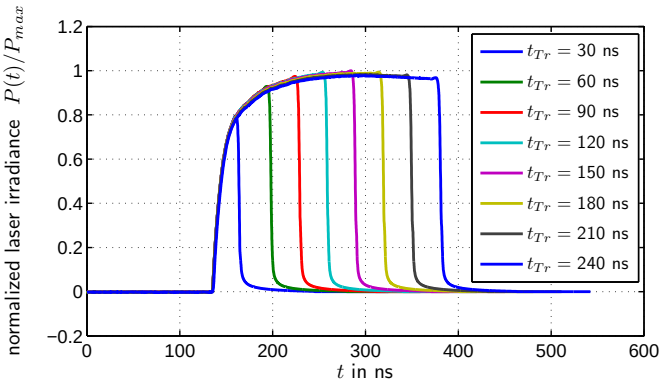


Figure 4.9: Laser module output pulse shape

shape. See chapter 4.2.3.2 for an examination of the influences of these laser pulse properties on the measurement performance of the system.

4.2.2.2 Output Power

To derive the optical output power of the laser modules the radiant power P_{Diode} on the surface of a calibrated photo diode was measured at a distance of 100 mm. From this value the overall power of the module can be calculated by scaling the measured value with the ratio of the illuminated area A_{Ill} to diode area A_{Diode}:

$$P_{Module} = P_{Diode} \cdot \frac{A_{Ill}}{A_{Diode}} = 0.245 \text{ W} \cdot \frac{3.85 \cdot 10^{-3} \text{ m}^2}{5.03 \cdot 10^{-7} \text{ m}^2} = 1.87 \cdot 10^3 \text{ W} \quad (4.20)$$

As this method is very imprecise and suffers from irradiance inhomogeneities over the illuminated area, the calculated value can only be taken as approximate value of the real output power. Additionally, the output power of the laser modules varies with changing temperature (see section 4.2.2.3 and [26]). From here a value of $P_{Module} = 2000\ W$ for the output power of a single laser module will be taken in later considerations and calculations.

4.2.2.3 Pulse Stability

Here two stability aspects are addressed, the short-term stability and the long-term stability of the laser pulses.
The short-term stability describes random fluctuations of the trigger time and the pulse amplitude; such fluctuations would directly affect the measured distance. These trigger time and amplitude jitters were tried to be analyzed

by means of measuring the distribution of the according values for a set of several thousands of pulses using the mentioned photo diode and a digital oscilloscope. But both, the trigger time jitter and the amplitude jitter turned out to be below the noise level of the oscilloscope and could thus not be measured. Since this noise level was in the range of a few tens of picoseconds (corresponding to an induced distance measurement error of a few mm), the short time stability of the laser modules can be regarded as sufficiently high for all kind of measurements to be performed with the examined sensor system.

The long-term stability of a laser module describes the characteristics of the output power over a longer range of time. Typically a laser diode starts emitting with maximal output power before the output power decreases until it converges to a stable final value after a few minutes of operation. This final power value is reached, if the laser diode reaches its thermal equilibrium; it mainly depends on the adjusted pulse length, pulse peak power and pulse repetition frequency (PRF). See figure 4.10 for three example curves ($P_{Module} \approx 2000$ W, $T_{Pw} = 180$ ns).

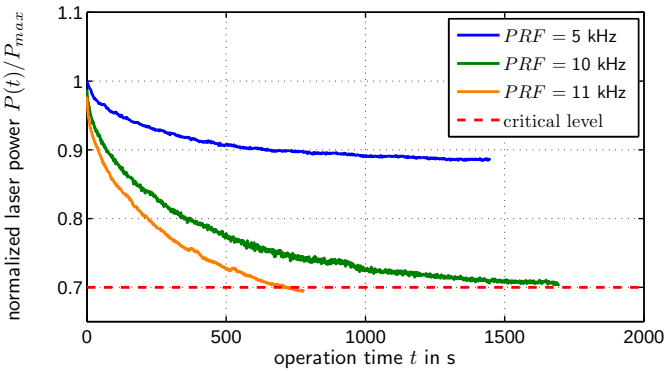


Figure 4.10: Evaluation of the PRF dependency of the laser module output power long-term characteristics

Attention has to be paid to the fact that the power level must not drop below 70% of the initial output power. This would harm the laser by thermal overstraining [27]. Thus this 70% boundary limits the maximum pulse length, pulse amplitude, and PRF. A tradeoff has to be found that fulfills this criterium as well as the requirements of the according application as good as possible. In the here considered case the modules were adjusted to

Pulse amplitude:	P_{Module}	$\approx$	2000 W
Pulse length:	T_{Pw}	$=$	180 ns ($\Leftrightarrow$ 27 m theoretical measurement range)
Pulse repetition frequency:	PRF	$=$	10 kHz

For these settings the equilibrium laser output power levels off at a value short above the 70% threshold. Further increase of the PRF causes the laser output power to drop below the 70% threshold and thus leads to thermal damage during longer operation (see output power curve for $PRF = 11$ kHz).

4.2.2.4 Eye Safety

The laser produces an intense, high directional beam of light, which can especially harm the eyes of human beings. Therefore, laser sources are classified into laser safety classes from 1 to 4 [28]. Only class 1 guarantees the operation of a laser source without any danger for people (especially their eyes) in the surrounding. Thus one major goal of the laser module development was the observance of the regulations for laser class 1. The theoretical limit between safe and potentially harmful is called *maximum permissible exposure* (MPE_1) for laser safety class 1. The MPE_i – the maximum permissible exposure for laser safety class i – levels are set by the 'International Commission on Non-Ionizing Radiation Protection' and are internationally accepted and adopted by the standardization committees such as IEC TC 76 and ANSI for the respective laser safety standards IEC 60825-1 and ANSI Z136.1. MPE_i levels are determined as a function of

- pulse duration,
- pulse repetition frequency (PRF),
- light emitting area, and
- laser wavelength.

To derive the laser safety class of a laser light source, a measured exposure value at a distance of 0.1 m from the light emitter has to be compared to the according calculated MPE_i value; in the here considered case MPE_1.
Figure 4.11 shows the measurement result for the UseRCams laser illumination module. Basis for the calculation of the MPE_1 are a pulse duration of 180 ns at a PRF of 20 kHz; for the measurement of the radiant exposure coming from the laser module, the output power was set to maximum. It homogeneously distributes on an illumination field of 30°x9°. All values are normalized to the

MPE_1 (e.g. $MPE_1(0.1\text{m}) = 0.00205\text{J/m}^2$), i.e. a normalized exposure value of 1 or below fulfills the safety class 1 requirements. This condition is not fulfilled for the specified measuring distance of 0.1 m. For measuring distances of 0.3 m and above it is. To classify the device as class 1 device, it has to be assured, that no human can get nearer to the laser source than 0.3 m, what is guaranteed under normal operation conditions when the camera is mounted behind a vehicles windshield. For additionally safety an electronic guard could be developed that switches off the laser modules if the vehicle is not moving.

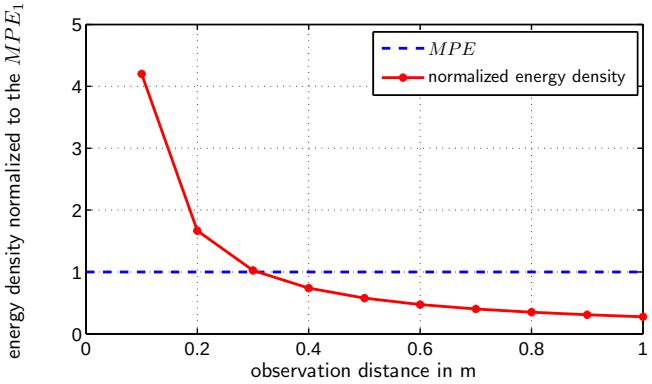


Figure 4.11: Fulfillment of laser class 1 regulations for different observation distances

4.2.3 Over-all System Characterization

This chapter characterizes the camera system as a whole after integration of all components. Goal is an overall analysis of the distance measuring capabilities of the camera by means of evaluation of different characterizing measurements, which are closely described in the following sections.

4.2.3.1 System Noise / System NEE

Noise investigation

The system noise σ_U stands for the noise which is inherent to the system output voltage U_{Sensor} (cf. section 4.1.1.1). As U_{Sensor} is the measure accessible for further processing like distance calculation, this noise value significantly influences the global system performance.
Figure 4.12 shows a model for the composition of the noise. Basic components

are the so-called sensor noise σ_S and the readout noise σ_R. The sensor noise voltage is here the standard deviation of the output voltage of the sensor (i.e. sensor noise power is the variance of the sensor output voltage) for a single accumulation. This noise contains the photodiode noise ($\sigma_{PD-noise}$), reset switch noise ($\sigma_{RS-noise}$), shutter switch noise ($\sigma_{Sh-noise}$), source follower noise ($\sigma_{SF-noise}$), noise of the operational transconductance amplifier (OTA) ($\sigma_{OP-noise}$), select switch noise ($\sigma_{Sl-noise}$), discharge switch noise ($\sigma_{DS-noise}$), and accumulate switch noise ($\sigma_{AS-noise}$). Reference voltages (U_{ddpix}, U_{ref3}, and U_{ref4}) also add some noise ($\sigma_{U_{ddpix}}$, $\sigma_{U_{ref3}}$, and $\sigma_{U_{ref4}}$). The photodiode noise consists of dark photodiode noise and photon shot noise. It must be noted that all the noise generated in front of the OTA is amplified by the capacitor ratio $\frac{C_{Cl}}{C_F}$, while the OTA noise is amplified by $\left(1 + \frac{C_{Cl}}{C_F}\right)$.
The readout noise describes the noise which is added to the sensor signal during the readout process. It includes noise contributions from the sensor readout buffer ($\sigma_{ReadoutBuffer}$), noise injected due to perturbations on the circuit board (σ_{Board}) and quantization noise ($\sigma_{Quantization}$) due to signal digitization by the ADC.

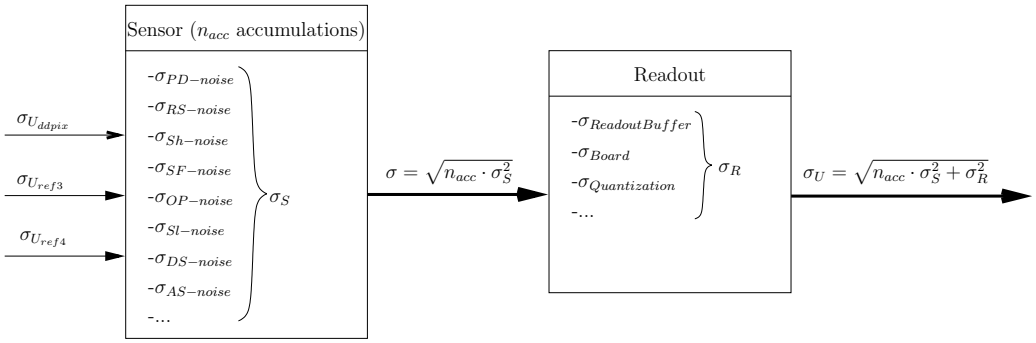


Figure 4.12: Sources of the system noise

Assuming Gaussian distribution (cf. appendix B.2) of the single noise components, the resulting over-all noise σ_U is calculated as the square root of the sum of the accumulated sensor noise and the readout noise [29]:

$$\sigma_U = \sqrt{n_{acc} \cdot \sigma_S^2 + \sigma_R^2} \tag{4.21}$$

with

$$\begin{aligned}\sigma_S &= f(\sigma_{U_{ddpix}}, \sigma_{U_{ref3}}, \sigma_{U_{ref4}}, \sigma_{PD-noise}, \sigma_{RS-noise}, \sigma_{Sh-noise}, \sigma_{SF-noise}, \\ &\quad \sigma_{OP-noise}, \sigma_{Sl-noise}, \sigma_{DS-noise}, \sigma_{AS-noise}, ...)\end{aligned}$$

$$\sigma_R = f(\sigma_{ReadoutBuffer}, \sigma_{Board}, \sigma_{Quantization}, ...)$$

A more thorough noise analysis is a quite complex task. Thus a further refinement of the noise analysis is not performed here. [30] gives insight into the image sensors noise behavior.

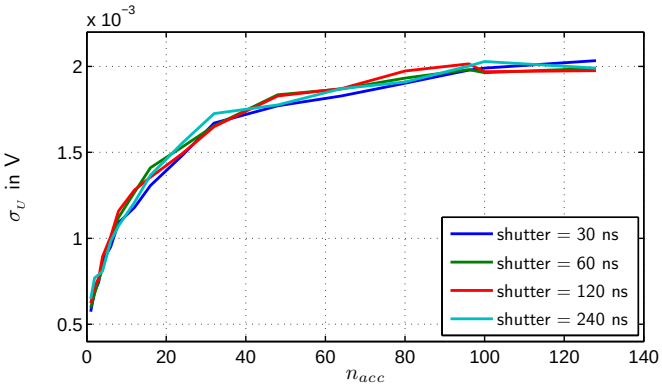
Noise measurements

Noise measurements were conducted in order to derive the noise characteristics of the system. Figure 4.13(a) shows a measurement of σ_U without laser illumination for different numbers of analog accumulations and shutter length. While the shutter length has no influence on the amplitude of the noise level, the noise increases as expected with increasing numbers of analog accumulations. The attempt to fit the function given by equation 4.21 through the measured curves to derive σ_S and σ_R yielded no good results. The reason is that the noise curves are subject of the same damping as it happens to the processed photocharges (cf. chapter 4.2.1.3). Thus in advance of the curve fitting the noise curves have to be weighted with the inverse damping function $d(n_{acc})^{-1}$ of chapter 4.2.1.3. Figure 4.13(b) shows the resulting inversely damped noise curve for a shutter length of 240 ns and the fitted curve according to equation 4.21. The parameters of the function that describe σ_S and σ_R are:

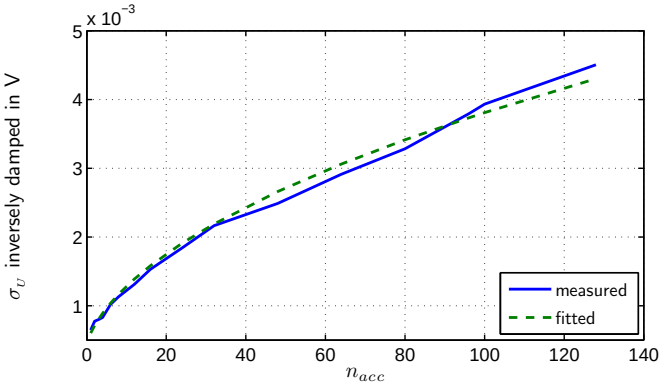
$$\sigma_S = 0.379 \text{ mV} \tag{4.22}$$

$$\sigma_R = 0.442 \text{ mV} \tag{4.23}$$

Figure 4.14 shows the probability density function (PDF, see annex B.1) of measured sensor signal noise values. The histogram is generated from 16000 single shutter values at 100 analog accumulations. Result is a Gaussian distribution (compare the histogram to the fitted Gaussian), what justifies the



(a) Dark noise of the system



(b) Dark noise weighted with inverse damping function $d(n_{acc})^{-1}$

Figure 4.13: System dark noise mean values original and inversely damped

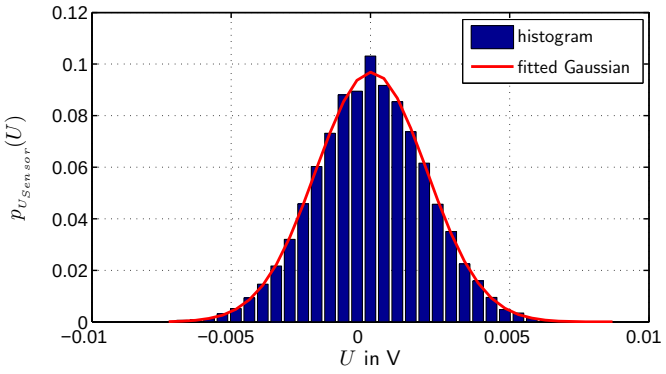


Figure 4.14: PDF of the dark noise values

assumptions given above. The Gaussian property of the signal error distribution will also be of importance for the image processing approach in chapter 6.

Another important measurement is the noise measurement in the presence of laser illumination. Figure 4.15 shows the result of noise measurements of the sensor output for different illumination values resp. sensor output signals for 16 analog accumulations; the illumination strength was thereby changed by tuning the input voltage of the laser illumination module what directly tunes the module output power.

The additional photon shot noise σ_{PS} is derived as follows:

$$\sigma_{PS}(U_{Sensor}) = \frac{\sqrt{Q_{photo}(U_{Sensor}) \cdot 6.242 \cdot 10^{18} \frac{1}{\mathrm{C}}}}{6.242 \cdot 10^{18} \frac{1}{\mathrm{C}} \cdot (C_D + C_{sense})} \cdot g_{SF} \cdot \frac{C_{Cl}}{C_F} \quad (4.24)$$

$$\overset{\text{Eq. 4.3}}{=} \frac{\sqrt{U_{Sensor} \cdot (C_D + C_{sense}) \cdot \frac{1}{g_{SF}} \cdot \frac{C_F}{C_{Cl}} \cdot 6.242 \cdot 10^{18} \frac{1}{\mathrm{C}}}}{6.242 \cdot 10^{18} \frac{1}{\mathrm{C}} \cdot (C_D + C_{sense})} \cdot g_{SF} \cdot \frac{C_{Cl}}{C_F} \quad (4.25)$$

$$= \sqrt{\frac{U_{Sensor}}{6.242 \cdot 10^{18} \frac{1}{\mathrm{C}} \cdot (C_D + C_{sense})}} \cdot \sqrt{g_{SF} \cdot \frac{C_{Cl}}{C_F}} \quad (4.26)$$

The resulting overall bright noise $\sigma^*_U(U_{Sensor}, n_{acc})$ can then be computed as equation 4.28 shows:

$$\sigma^*_U(U_{Sensor}, n_{acc}) = \sqrt{n_{acc} \cdot \left(\sigma_S^2 + \left(\frac{\sigma_{PS}(U_{Sensor})}{\sqrt{n_{acc}}} \right)^2 \right) + \sigma_R^2} \quad (4.27)$$

$$= \sqrt{n_{acc} \cdot \sigma_S^2 + \sigma_R^2 + \sigma_{PS}(U_{Sensor})^2} \quad (4.28)$$

Comparing this with the noise as defined in equation 4.21 gives the error in the computation of $\sigma_U(n_{acc})$ without considering the photon shot noise σ_{PS}:

$$\frac{\sigma^*_U(U_{Sensor}, n_{acc})}{\sigma_{U_D}(n_{acc})} = \sqrt{\frac{n_{acc} \cdot \sigma_S^2 + \sigma_R^2 + \sigma_{PS}(U_{Sensor})^2}{n_{acc} \cdot \sigma_S^2 + \sigma_R^2}} \quad (4.29)$$

$$= \sqrt{1 + \frac{\sigma_{PS}(U_{Sensor})^2}{n_{acc} \cdot \sigma_S^2 + \sigma_R^2}} \quad (4.30)$$

But inserting the values of the capacitors and of the noise contributions σ_S and σ_R known from sections 4.1.1.1 and 4.2.3.1 respectively, results in comparatively small errors when omitting the photon shot noise. Considering a maximum sensor output voltage of 1.5 V results in a error of 18% for $n_{acc} = 1$, which is the maximum error value. For increasing values of n_{acc} the error shrinks rapidely to e.g. 2.7% at $n_{acc} = 16$ and 0.46% at $n_{acc} = 100$. This shows that especially for high numbers of analog accumulations (i.e. measurements where the measurement signals and thus the SNR is low) the photon shot noise can be disregarded compared to the other noise components. For simplicity reasons the photon shot noise is not regarded at all in the further considerations of this thesis; the dark noise value is taken for all computations without making much error. Basic equation for noise calculations is thus equation 4.21. This result is important in view of the later image processing, as one can always suppose the same noise level independently from any illumination characteristics.

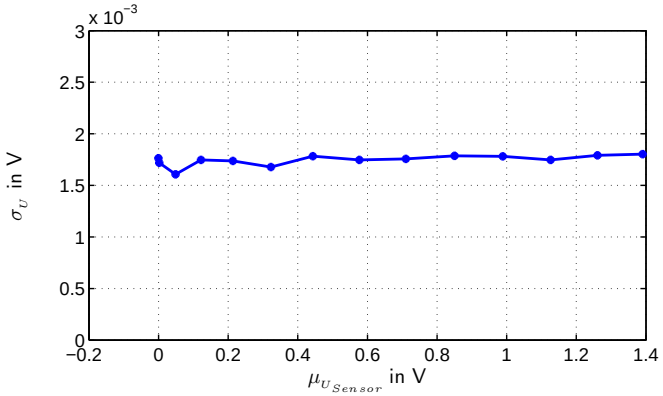


Figure 4.15: σ_U value of the sensor outputs for different sensor signal amplitudes

NEE derivation

The sensor *noise equivalent exposure* NEE_{Sensor} is defined as the x-coordinate value of the intersection point between the interpolated responsivity characteristics R_λ and the dark noise of the sensor, i.e. the value of radiant exposure that generates a sensor output signal which is equal to the sensor noise σ_S . Analogously a system NEE_{System} defines the necessary radiant exposure to obtain a sensor output similar to the system noise level σ_U. As stated earlier

in this chapter, the system noise depends on the fixed readout noise and the n_{acc}-dependent sensor noise; the system noise level is defined as the noise at one analog accumulation as this case is the worst case in view of the SNR, as here the readout noise has the largest influence.

$$NEE_{System}|_{n_{acc}=1} = 225 \text{ nJ/m}^2 \tag{4.31}$$

$$NEE_{Sensor} = 145 \text{ nJ/m}^2 \tag{4.32}$$

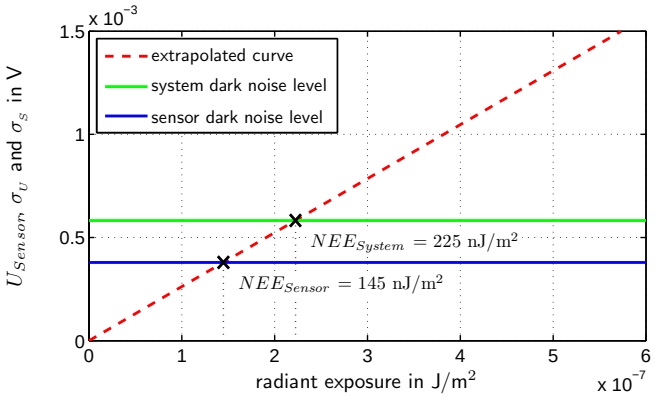


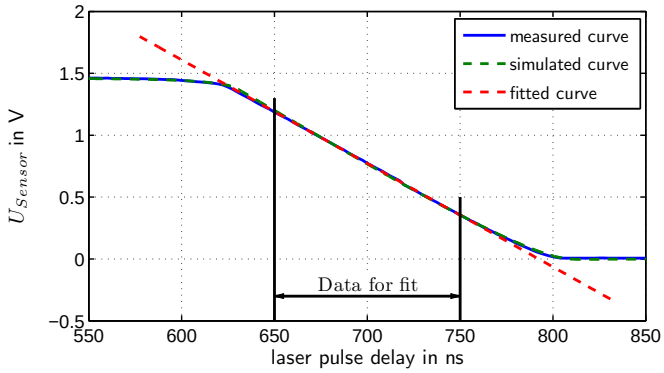
Figure 4.16: Determination of the NEE by extrapolating the fitted sensor responsivity curve of figure 4.6

Note that these values are valid for the typical laser light wavelength of $\lambda = 905$ nm. For other wavelengths the NEE-values can change due to the λ-dependency of R_λ.

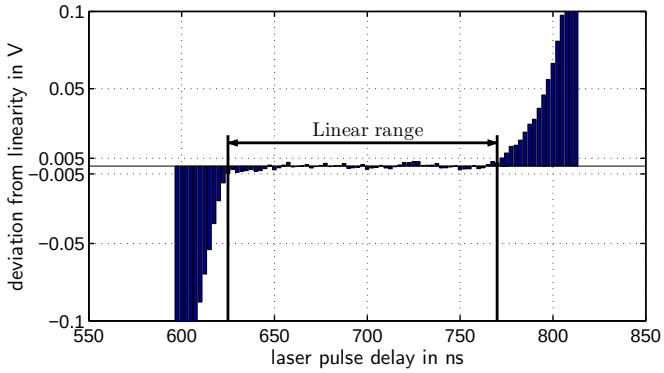
4.2.3.2 Shutter-Window-Pulse Correlation

The *shutter-window-pulse correlation* of the 3D-CMOS system was gained by measurement and by calculation. For the measurement, the laser pulses trigger was shifted in time by using a delay element located on the circuit board. The resulting sensor signal plotted against the laser delay forms the measured correlation curve. Additionally, the shutter-window-pulse correlation functions were calculated as correlation integral of the measured shutter window functions (cf. section 4.2.1.2) and the measured laser pulse shapes (cf. section 4.2.2.1). The result of both procedures for a shutter length of 390 ns and a laser pulse length of 180 ns is depicted in figure 4.17(a). Only the trailing edge of the function

is shown for reasons of the recognizability of details; the results are the same for the rest of the curve. In addition to the two curves a straight line is fitted through the middle range of the measured curve for linearity investigation, which is shown in figures 4.17(b).



(a) Trailing edge of the measured and the calculated correlation curve with linear fit



(b) Deviation of the measured values from the linear fit

Figure 4.17: Shutter-window-pulse correlation investigations

Following results can be gained from the figures:

- The measured correlation curve and the calculated correlation curve agree very well indeed. This means that the sensor model of equation 3.16 fits the real behavior of the sensor chip quite well. Thus the calculated sensor output signal can be taken for simulations or the like without the risk of getting large deviations from real sensor signals.

- For laser delays between 625 ns and 770 ns the correlation curve exhibits a range of 145 ns – corresponding to a measurement range of 21.75 m, i.e. for example 2 m - 23.75 m – in which the curve coincides with the fitted straight line quite good; in this range distance measurement is applicable without any further adaption.
- Figure 4.17(b) shows that within the measurement range the deviation of the measured signal from linearity is always below 0.005 V.
- The discrepancies of the shutter functions and the laser pulse shapes from the ideal rectangular shape, noted in chapters 4.2.1.2 and 4.2.2.1, do only affect the begin and the end of the edges of the correlation function. In between these corrupted regions good distance measurement is possible.

4.2.3.3 Dynamic Range

Dynamic range is defined as the ratio of the maximum level of a parameter to a minimum detectable value of that parameter.
Thus the dynamic range of the photodiode in terms of radiant exposure is theoretically defined as

$$DR_{PD,theor} = 20 \cdot \log \left(\frac{E_{Sensor,max} \cdot T_{Int,max}}{NEE_{Sensor}} \right) \text{dB} \tag{4.33}$$

for a single accumulation (n_{acc} =1).

Note that this yields (using equation 4.13)

$$DR_{PD,theor} = 20 \cdot \log \left(\frac{U_{Sensor,max}}{R_\lambda \cdot NEE_{Sensor}} \right) \text{dB} \tag{4.34}$$

$$= 20 \cdot \log \left(\frac{U_{Sensor,max}}{\sigma_S} \right) \text{dB} \; . \tag{4.35}$$

Taking the analog accumulation and digital averaging capabilities of the camera system into account the system dynamic range in terms of radiant exposure can be derived from this as follows:

$$DR_{System,theor} = 20 \cdot \log \left(\frac{E_{Sensor,max} \cdot T_{Int,max}}{\frac{NEE_{System}}{n_{acc,max}} \cdot \frac{1}{\sqrt{n_{av,max}}}} \right) \text{dB} \tag{4.36}$$

$$= 20 \cdot \log \left(\frac{U_{Sensor,max}}{\frac{\sqrt{n_{acc,max} \cdot \sigma_S^2 + \sigma_R^2}}{n_{acc,max}} \cdot \frac{1}{\sqrt{n_{av,max}}}} \right) \text{dB} \tag{4.37}$$

Estimating $E_{Sensor,max} = 1.5$ V (the power supply voltage is only 3.3 V) and taking into account the measured sensor and system noise values $\sigma_S = 0.379$ mV and $\sigma_R = 0.442$ mV and the maximum applied number of analog accumulations $n_{acc} = 100$ we obtain a dynamic range of the sensor system of

$$DR_{System,theor} = 91.9 \text{ dB} + 20 \cdot \log\left(\sqrt{n_{av,max}}\right) \text{dB} \ . \tag{4.38}$$

Now consider the required input range of the camera from the point of view of the application specific requirements. In the UseRCams front view application the specified target values are a range of the distance d to be measured from 2 m - 20 m and a reflectivity range from 5% - 100% (Lambertian reflectance) at a distance accuracy σ_d/d of 3%.
This leads to following required input range of the camera (see equation 3.11):

$$20{\cdot}\log\left(\frac{E_{Sensor,max} \cdot T_{Int,max}}{NEE_{System,min}}\right) = 20{\cdot}\log\left(\underbrace{\left(\frac{d_{max}}{d_{min}}\right)^2}_{\substack{\text{distance}\\ \text{dynamics}}} \cdot \underbrace{\frac{\rho_{max}}{\rho_{min}}}_{\substack{\text{reflectivity}\\ \text{dynamics}}} \cdot \underbrace{\frac{U_{Sensor}}{\sigma_U}}_{\text{SNR}_{min}}\right) \text{dB} \tag{4.39}$$

where $\frac{U_{Sensor}}{\sigma_U}$ defines the signal-to-noise ratio (SNR_{min}) which even the minimum measured signal has to satisfy to fulfill the 3% distance accuracy requirement. This SNR_{min} is derived in the following:

With equations 3.20 and 5.6 one can express the normalized distance accuracy σ_d/d (use of MDSI3 algorithm assumed, $T_{DL} = T_{DS} = 0$) as

$$\frac{\sigma_d}{d} = \frac{\frac{c}{2} \cdot T_{Pw} \cdot \frac{1}{(U_a+U_b)^2} \cdot \sqrt{U_a^2 + U_b^2} \cdot \sigma_U}{\frac{c}{2} \cdot \left(T_{Pw} - T_{Pw} \cdot \frac{U_a}{U_a+U_b}\right)} \tag{4.40}$$

$$= \frac{\frac{1}{(U_a+U_b)^2} \cdot \sqrt{U_a^2 + U_b^2} \cdot \sigma_U}{\frac{U_b}{U_a+U_b}} \tag{4.41}$$

$$= \frac{1}{(U_a+U_b) \cdot U_b} \cdot \sqrt{U_a^2 + U_b^2} \cdot \sigma_U \quad . \tag{4.42}$$

As the most critical measurements regarding the signal-to-noise ratio are measurements of object which are far away of the camera, equation 4.42 is evaluated for high target distances, i.e. $U_b \to U_c$ (see figure 3.6) and $U_a \to U_0$. This yields

$$\left.\frac{\sigma_d}{d}\right|_{\substack{U_b \to U_c \\ U_a \to U_0}} = \frac{\sigma_U}{U_{b/c}} \ , \tag{4.43}$$

and thus

$$\mathrm{SNR}_{min} = \frac{U_{Sensor}}{\sigma_U} = \frac{d}{\sigma_d} = \frac{1}{3\%} = 33.3 \ . \tag{4.44}$$

Hence the required input range of the camera results as

$$20 \cdot \log\left(\frac{E_{Sensor,max} \cdot T_{Int,max}}{NEE_{System,min}}\right) = 20 \cdot \log\left(\left(\frac{20\ \mathrm{m}}{2\ \mathrm{m}}\right)^2 \cdot \frac{100\ \%}{5\ \%} \cdot 33.3\right) \mathrm{dB} \tag{4.45}$$

$$= 20 \cdot \log(66.6 \cdot 10^3)\ \mathrm{dB} \tag{4.46}$$

$$= 96.5\ \mathrm{dB} \tag{4.47}$$

Comparison with the theoretical value shows that the system dynamic range without digital averaging ($DR_{System,theor}$ = 91.9 dB + 0 dB = 91.9 dB) is not sufficient to cope with the specified input range of the planned applications. Introducing averaging over 3 camera frames ($n_{av} = 3$) increases the theoretical dynamic range to 96.7 dB. Hence the camera system fulfills the above mentioned application requirements if digital averaging over 3 frames or more is applied. As the $DR_{System,theor}$-value of 96.7 dB exceed the required 96.5 dB (see eq. 4.47) only marginal, the better choice is an averaging value of $n_{av} = 4$ – leading to $DR_{System,theor}$ = 97.9 dB – to keep some reserve. This on the other hand lowers the camera frame rate (cf. section 4.2.3.5) by a factor of 4.

For practical examination of the dynamic range of the camera system measurements were conducted for different target distances d_r and different target reflectivities ρ with adaptive number of accumulations n_{acc} (see section 4.1.1.3); from these the standard deviations σ_{d_m} of the measured distance values d_m were calculated. Figure 4.18 shows the measurement errors normalized to the appropriate target distances d_r, i.e. the normalized distance accuracy σ_{d_m}/d_r. Distance derivation algorithm MDSI3 is used.

The curves can now be evaluated according to the above mentioned UseRCams specifications. As no target with 100% Lambertian reflectivity could be found (this value is only theoretical; a small fraction of the incident light is always absorbed by the target), a white paper target with 80% reflectivity was chosen instead. Although for calculation of the dynamic range 100% is set as ρ_{max}, as experiments showed that also for a reflectivity of 100% distance measurement over the whole specified measurement range is possible. From figure 4.18

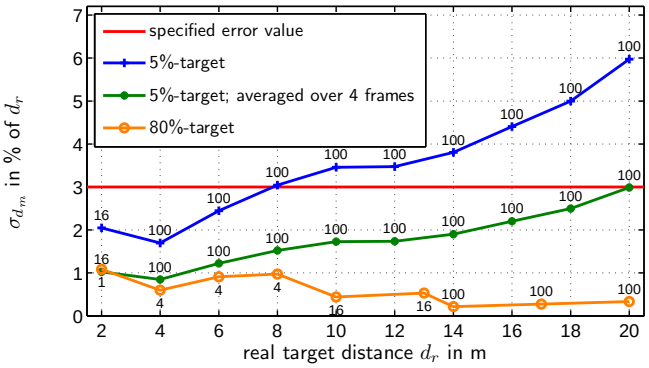


Figure 4.18: Accuracy of the measured distance d_m for different targets and camera parametrizations over a measurement range from 2 m to 20 m; the according n_{acc} is added to each data point

one can see, that the 5%-target exceeds the specified error of 3% already for distances larger then 8 m. Thus the dynamic range is not sufficient to fulfill the specified requirements without digital averaging, as stated already above in the theoretical derivation of the dynamic range. But by applying averaging over 4 subsequent frames, the accuracy can be increased what leads to a fulfillment of the requested values (see figure 4.18). This is a good match with the theoretical computations.

4.2.3.4 Distance Measurement Performance

Figure 4.19 shows the distance measuring performance of the camera system. Basis is a distance measurement of a plywood target with a reflectivity of ∼50% for distances between 3 m and 21 m; pixel 32 of sensor line 6 is examined, since this pixel coincides with the center of the target here. Adaptive accumulation is applied with $n_{acc} \in \{1, 4, 16, 64\}$. This means that at each distance step n_{acc} is chosen in such way as to get the maximum non-saturated signal.
The blue curve shows the mean values of the measured distances for the different distance steps. Good agreement of real with measured distance values (compare measured curve to dashed bisecting line) and good linearity of the distance measuring characteristics (deviation from the bisecting line, see also section 4.2.3.2) can be observed. The green line indicates the standard deviation σ_d of the measured distance values at each distance step. The sawtooth-shaped curve is characteristic for the adaptive accumulation. Each time n_{acc} is increased to the next level, the distance error drops due to the increased SNR (cf. section 4.1.1.3). The distance error is directly proportional to the target

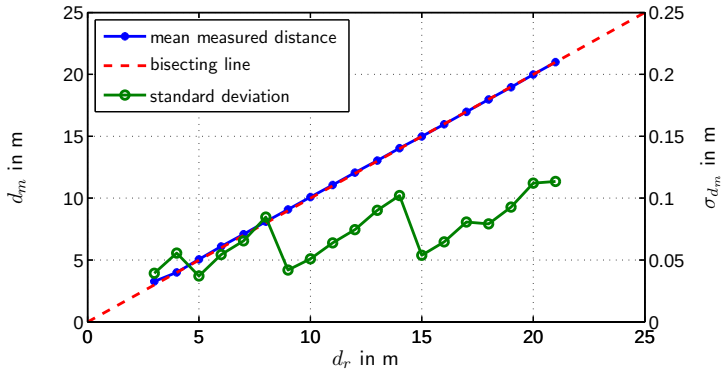


Figure 4.19: Distance measurement performance using MDSI3 algorithm and adaptive accumulation

reflectivity. Hence to derive the expected distance error for targets with other reflectivity, the error has to be weighted accordingly.

Besides the good measurement performance described in the previous sections also some shortcomings of the 3D-CMOS sensor system have to be mentioned here:

- Due to the large pixel dimensions of 130 μm times 300 μm, which are necessary to collect enough reflected photons in the extremely short integration windows, the resolution of the camera system is comparatively low. This can cause blurred distance measurement values when one pixel 'sees' two or more different objects. Especially on the edges of objects these blurring effects occur.

- The sequential acquisition of the shutters is a drawback when having fast moving objects in the observed scene, as the two shutters 'see' actually different images, as the objects are moving on during the acquisition times of the shutters. Especially on object edges this leads to corrupted distance measurement results. An improved sensor design implementing alternating acquisition of the single accumulations of the two shutters (what would require additional storage capacitors in the pixel readout stage) could reduce this effect significantly.

- Scenes with large irradiance gradients on the sensor surface reveal crosstalk between the sensor pixels. See chapter 4.2.1.4 for a deeper examination of this topic.

4.2.3.5 Camera Frame Rate

The *camera frame rate* FR defines the number of image frames that are acquired by the camera per second. The main FR-determining factors are the PRF, the maximum number of analog integrations n_{acc}^{max} per image, and the number of frames n_{av} to be averaged for noise reduction. With these FR is calculated as follows (assuming distance measurement by MDSI (see chapter 3.4.1), where 2 shutter windows are used):

$$FR = \frac{PRF}{2 \cdot n_{acc}^{max} \cdot n_{av}} \tag{4.48}$$

For the usual case, where $n_{acc}^{max} = 100$ and $n_{av} = 1$, and a laser PRF of 10 kHz (see section 4.2.2.3) one obtains a FR of 50 fps (frames per second).
In reality the cameras FR does not reach this calculated value, as sensor readout between the image acquisitions and during adaptive integration consumes additional time that slightly lowers the FR of the camera.

4.3 Camera Calibration

Before the camera can be used for distance measurements, different calibration steps are necessary, namely an *offset calibration* and a *distance calibration*.

4.3.1 Offset Calibration

Though the sensor internally applies CDS (4.1.1.2), *fixed pattern noise* (FPN) is inherent to the sensor output signals as one can see in figure 4.20. This 'spatial noise' badly affects the distance calculation of the distance derivation algorithms. In order to avoid this effect, an *offset calibration* has to be conducted. For this reason an image has to be taken without any illumination of the sensor surface for all used steps of analog accumulation (see section 4.1.1.3); these signals must be permanently stored for all pixels and subtracted later from the sensor signals each time when a distance image is taken. This removes the FPN.

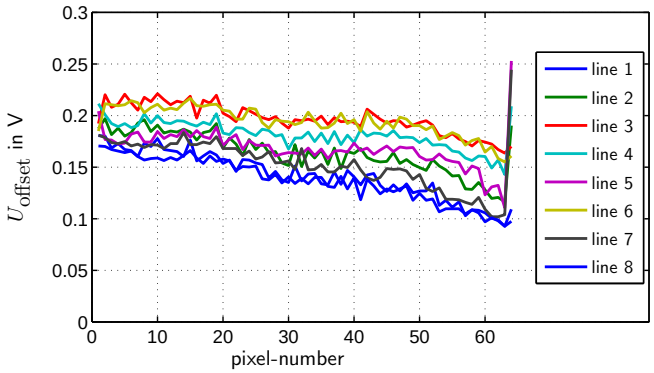


Figure 4.20: Fixed pattern noise of the sensor chip for 16 analog accumulations and a shutter length of 240 ns

4.3.2 Distance Calibration

In order to obtain accurate distance values from the raw data of the 3D-CMOS camera, also a *distance calibration* procedure has to be performed before starting measurements. According to the functional principle of the 3D-CMOS camera and the used distance derivation algorithm (cf. chapter 3.4), in the process of distance calculation an algorithm specific function $f(U_i)$, $i \in \{a, b, c\}$, of shutter signals (MDSI 1: $f(U_i) = \frac{U_a}{U_c}$, MDSI 3: $f(U_i) = \frac{U_a}{U_a+U_b}$, etc.) of each pixel is interpreted as a certain distance value. The right assignment of quotient to distance value is thereby managed by two so-called calibration parameters per pixel, which have to be derived using the camera calibration procedure.

In a first step images of a plain target at several known distances d_r are recorded, and the appropriate sensor signal function $f(U_i)$ is calculated. The second step is the fitting of a straight line through the measured quotient values, which are plotted over the known distances. Thirdly, the slope and intercept values (calibration parameters) of the fitted curve are extracted, what allows an one-to-one mapping of each possible quotient value to a distance value. See figure 4.21 to get an example of the above described procedure.

As the calibration parameters slightly differ for different pixels this procedure is conducted for all 64x8 pixels of the imaging sensor at the moment. This and the fact that the camera (or the target) has to be moved during calibration makes this procedure fairly complex. Further investigations will show, if a simplified calibration procedure, which simulates the different target distances by defined

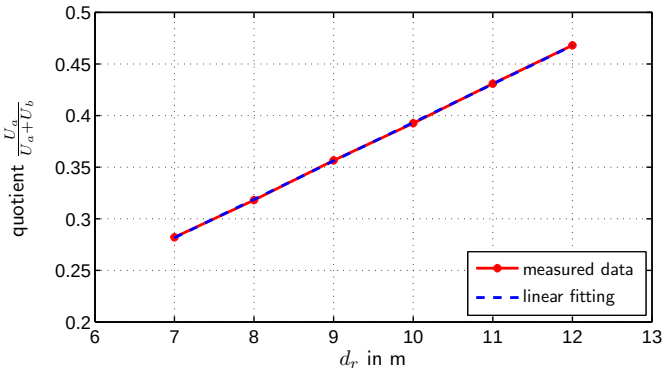


Figure 4.21: Example of a distance calibration measurement curve for 16 analog accumulations

delays of the laser pulse, is also feasible with respect to the measurement accuracy of the camera system.

Chapter 5

Optimization of the Distance Measurement

5.1 Comparison of the Distance Measurement Algorithms

In order achieve the best possible measurement results with respect to absolute and relative measurement accuracy and measuring time, an intensive comparison of the different distance measurement algorithms (cf. section 3.4) has been conducted and will be described in this chapter. It includes mathematical investigations, computer simulations, as well as laboratory experiments with the 3D-CMOS camera introduced in chapter 4.

5.1.1 Mathematical Investigation

As first investigation method a mathematical derivation of the expected distance measurement error σ_d is carried out. It enables a fast estimation of performance of the employed algorithms, but does not consider some effects that occur during the measurement in reality, like non-ideal shutter window functions and laser pulse shapes, signal quantization noise, and the use of value-discrete sensor values for distance calculation. In the following – based on the well-known law of propagation of uncertainties – the expected distance measurement error using the different algorithms is derived; additionally an evaluation example is shown in figure 5.1.

5.1.1.1 Propagation of Uncertainties

Let $f = f(x_1, x_2, ..., x_n)$ be a function depending on n variables $x_1, x_2, ..., x_n$. The uncertainty of each variable is given by Δx_i. Assuming uncorrelated variables, the random uncertainty Δf of f that results from the random uncertainties of the variables can be calculated as:

$$\Delta f = \Delta f(x_1, x_2, ..., x_n, \Delta x_1, \Delta x_2, ..., \Delta x_n) = \sqrt{\left(\sum_1^n \left(\frac{\partial f}{\partial x_i} \cdot \Delta x_i\right)^2\right)} \quad (5.1)$$

where $\frac{\partial f}{\partial x_i}$ designates the partial derivative of f for the i-th variable.

This equation is the basis for the mathematical investigation and comparison of the distance derivation algorithms. Inputs are the sensor output values $U_1, U_2, ..., U_n$, n depending on the algorithm and the according Gaussian distributed signal uncertainties. As the value of ΔU_i is not known in advance, σ_{U_i} is used as measure of the signal uncertainty, which is known from the system noise measurements in chapter 4.2.3.1. Subsequently one obtains the expected standard deviation of the measured distances σ_d as result. Thus equation 5.1 changes to

$$\sigma_d = \sigma_d(U_1, U_2, ..., U_n, \sigma_{U_1}, \sigma_{U_2}, ..., \sigma_{U_n)} = \sqrt{\left(\sum_1^n \left(\frac{\partial f}{\partial U_i} \cdot \sigma_{U_i}\right)^2\right)} \; . \quad (5.2)$$

Due to the fact that the signal amplitude U_i has no influence on the signal noise level σ_{U_i} (cf. chapter 4.2.3.1), the noise level for a certain camera setup is measured once and is then valid for any measurement with the according camera settings, yielding

$$\sigma_{U_1} \approx \sigma_{U_2} \approx ... \approx \sigma_{U_n} \quad \approxeq \quad \sigma_U \; . \quad (5.3)$$

In the following, formula 5.2 is evaluated for the different distance calculation methods described in chapter 3.4:

MDSI1

$$\begin{aligned} \sigma_d &= \sqrt{\left(\frac{c}{2} \cdot T_{Pw} \cdot \frac{1}{U_c} \cdot \sigma_{U_a}\right)^2 + \left(\frac{c}{2} \cdot T_{Pw} \cdot \frac{-U_a}{U_c^2} \cdot \sigma_{U_c}\right)^2} \\ &\approx \frac{c}{2} \cdot T_{Pw} \cdot \frac{1}{U_c} \cdot \sqrt{1 + \frac{U_a^2}{U_c^2}} \cdot \sigma_U \end{aligned} \quad (5.4)$$

MDSI2

$$\begin{aligned}\sigma_d &= \sqrt{\left(\frac{c}{2}\cdot T_{Pw}\cdot\frac{1}{U_c}\cdot\sigma_{U_b}\right)^2+\left(\frac{c}{2}\cdot T_{Pw}\cdot\frac{-U_b}{U_c^2}\cdot\sigma_{U_c}\right)^2}\\ &\approx \frac{c}{2}\cdot T_{Pw}\cdot\frac{1}{U_c}\cdot\sqrt{1+\frac{U_b^2}{U_c^2}}\cdot\sigma_U \end{aligned} \tag{5.5}$$

MDSI3

$$\begin{aligned}\sigma_d &= \sqrt{\left(\frac{c}{2}\cdot T_{Pw}\cdot\frac{U_b}{(U_a+U_b)^2}\cdot\sigma_{U_a}\right)^2+\left(\frac{c}{2}\cdot T_{Pw}\cdot\frac{-U_a}{(U_a+U_b)^2}\cdot\sigma_{U_b}\right)^2}\\ &\approx \frac{c}{2}\cdot T_{Pw}\cdot\frac{1}{(U_a+U_b)^2}\cdot\sqrt{U_a^2+U_b^2}\cdot\sigma_U \end{aligned} \tag{5.6}$$

MDSI4

$$\begin{aligned}\sigma_d &= \sqrt{\left(\frac{c}{2}\cdot T_{Pw}\cdot\frac{-U_b}{(U_a+U_b)^2}\cdot\sigma_{U_a}\right)^2+\left(\frac{c}{2}\cdot T_{Pw}\cdot\frac{U_a}{(U_a+U_b)^2}\cdot\sigma_{U_b}\right)^2}\\ &\approx \frac{c}{2}\cdot T_{Pw}\cdot\frac{1}{(U_a+U_b)^2}\cdot\sqrt{U_a^2+U_b^2}\cdot\sigma_U \end{aligned} \tag{5.7}$$

Gradient Method 1

$$\begin{aligned}\sigma_d &= \sqrt{\left(\frac{c}{2}\cdot(T_{Int,a1}-T_{Int,a2})\cdot\frac{-U_{a1}}{(U_{a1}+U_{a2})^2}\cdot\sigma_{U_a1}\right)^2+}\\ &\qquad\overline{+\left(\frac{c}{2}\cdot(T_{Int,a1}-T_{Int,a2})\cdot\frac{U_{a2}}{(U_{a1}+U_{a2})^2}\cdot\sigma_{U_a2}\right)^2}\\ &\approx \frac{c}{2}\cdot(T_{Int,a1}-T_{Int,a2})\cdot\frac{1}{(U_{a1}+U_{a2})^2}\cdot\sqrt{U_{a1}^2+U_{a2}^2}\cdot\sigma_U \end{aligned} \tag{5.8}$$

Gradient Method 2

$$\begin{aligned}\sigma_d &= \sqrt{\left(\frac{c}{2}\cdot(T_{Int,b1}-T_{Int,b2})\cdot\frac{-U_{b1}}{(U_{b1}+U_{b2})^2}\cdot\sigma_{U_b1}\right)^2+} \\ &\qquad \overline{+\left(\frac{c}{2}\cdot(T_{Int,b1}-T_{Int,b2})\cdot\frac{U_{b2}}{(U_{b1}+U_{b2})^2}\cdot\sigma_{U_b2}\right)^2} \\ &\approx \frac{c}{2}\cdot(T_{Int,b1}-T_{Int,b2})\cdot\frac{1}{(U_{b1}+U_{b2})^2}\cdot\sqrt{U_{b1}^2+U_{b2}^2}\cdot\sigma_U \qquad (5.9)\end{aligned}$$

DCSI

$$\sigma_d = \frac{c}{2}\cdot\sqrt{\left(\frac{\partial T_{DL_0}}{\partial U_1}\cdot\sigma_{U_1}\right)^2+\left(\frac{\partial T_{DL_0}}{\partial U_2}\cdot\sigma_{U_2}\right)^2+\cdots+\left(\frac{\partial T_{DL_0}}{\partial U_n}\cdot\sigma_{U_n}\right)^2} \qquad (5.10)$$

with $\frac{\partial T_{DL_0}}{\partial U_i}$ being calculated according to equation C.8 in the annex.

Figure 5.1 shows the result of the evaluation of the above σ_d-equations for a virtual measurement of distances from 5 m to 20 m. The according shutter signal amplitudes are calculated by means of equations 3.11 and 3.16; the noise level is taken from the system noise measurement in chapter 4.2.3.1. For reasons of clarity adaptive accumulation is not considered in the calculation. Though this would lower σ_d for large distances for all of the examined algorithms, it would not change the ratios of the σ_d. Thus the use of adaptive accumulation has no influence on the result of the comparison of the distance derivation methods examined here.

As a result it can be seen that especially the gradient methods denoted GRAD1 and GRAD2 exhibit bad performance in terms of σ_d compared to the other methods. Also MDSI2 shows bad measurement noise behavior; the inaccuracy is here more than twice the inaccuracy of MDSI3 and MDSI4 over long regions of the measurement range. Thus – according to this mathematical investigation – only MDSI1, MDSI3, MDSI4 and DCSI seem to be well suited for distance measurements.

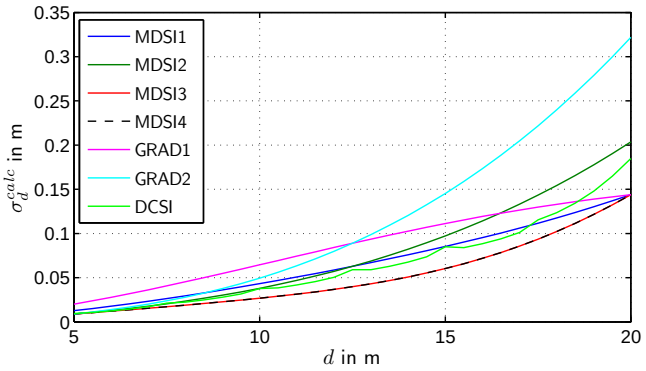


Figure 5.1: Calculation of the expected distance measurement accuracy according to equations 5.4-5.10

5.1.2 Investigation using Simulation

The next step of the comparison of the distance derivation algorithms is a computer simulation of the distance measuring process of the 3D-CMOS camera system. For good comparability to the mathematical investigation the range of the simulation is again 5 m - 20 m; adaptive accumulation is not considered for the above mentioned reasons. Basis for the simulation are calculated sensor output signal amplitudes (equations 3.11 and 3.16), to which Gaussian distributed random signals having a standard deviation according to chapter 4.2.3.1 are added. These synthetic noisy sensor output values are then processed for distance calculation using MDSI1, MDSI3, and DCSI, the algorithms that showed to be suitable in the mathematical investigation above. Iteration of this procedure for a set of 5000 values leads to the standard deviation of the received distance values as shown in figure 5.2.
Result is that the simulation leads to the same results as the calculation, though – in contrast to the calculation – quantization and value-discrete character of the sensor signals are considered here. However, these factors seem to have only marginal influence on the measured distance at the considered system setup.

5.1.3 Experimental Results

From the three methods examined by simulation, only the MDSI algorithms proved to be suitable for the current 3D-CMOS camera. The reasons for the refusal of the DCSI method are the worse accuracy characteristics and particularly the complexity of the algorithm, i.e. the use of a series of laser pulses for

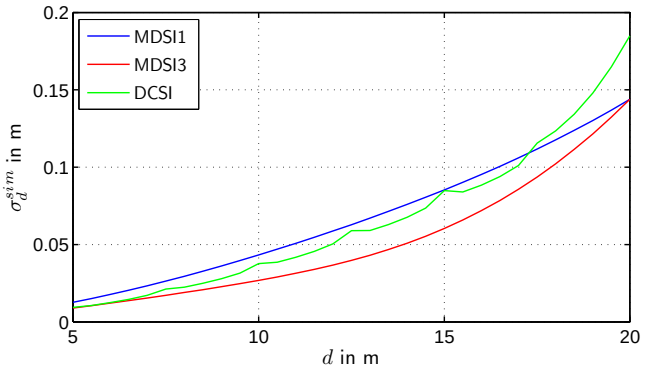


Figure 5.2: Simulation of the expected distance measurement accuracy for MDSI1, MDSI3 and DCSI

a single measurement that reduces the camera frame rate below a level which would be acceptable for the image processing described in chapter 6.
Thus the experimental exploration is directed at the comparison between measurement procedures MDSI1 or MDSI3.
Until the development of the UseRCams camera prototype, MDSI1 was the deployed distance derivation method, though the better theoretical noise performance of MDSI3 was known. But MDSI3 was not applicable with the predecessors of the UseRCams 3D image sensor due to insufficient steepness of the trailing edge of the shutter window; thus all procedures that make use of a pulse to shutter window alignment of type b) (according to figure 3.6: the pulse is emitted before the shutter window 'opens', i.e., the pulse 'moves into' the shutter window for increasing distances resp. τ_{TOF}) were not applicable with this sensor generation. The trailing edge of these sensors showed a start of increase up to 300 ns prior to reaching the final value, what made a reasonable measurement of distance dependent sensor signals impossible. The decreased trailing edge width of the 3D-CMOS array chip used here of about 60 ns (cp. section 4.2.1.2) makes shutter alignment b) and thus MDSI3 applicable.
A measurement was conducted, which determined the standard deviation of the measured distance for MDSI1 and MDSI3 at target distances between 3.0 m and 21.0 m in steps of 1.0 m and for different numbers of analog accumulations. Using a plywood target with a reflectivity ρ of $\sim$50% we got the results shown in figures 5.3 and 5.4.
It can clearly be seen that the distance noise using MDSI3 is below the distance noise using MDSI1 at any distance or integration step. This fact confirms the results of the accuracy calculations and simulations of the algorithms. To what extent the ratios between the measured distance accuracy of the two methods

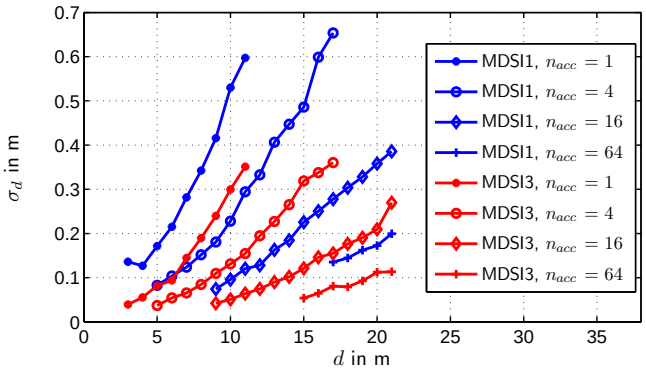


Figure 5.3: Experimental comparison of the distance measurements for MDSI1 and MDSI3 ($\rho \sim 50\%$)

match with the calculated values is closer examined in figure 5.4 for a measurement setup with 16 analog accumulations. The values are normalized with respect to the measurement accuracy using the MDSI3 method, i.e. the distance noise using MDSI3 is 1.0 for any distance step. The dashed line gives the calculated error values for the MDSI1 method, the crosses forming the blue line the measured distance errors for MDSI1, which even exceeds the value predicted by the calculation and the simulation.

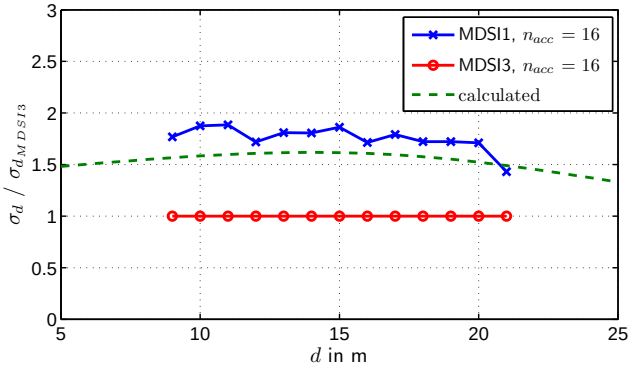


Figure 5.4: Comparison of the measurement result and the calculation result for the ratio of the errors of MDSI1 and MDSI3

Thus the calculations, simulations, and measurements correspondingly showed that the use of MDSI3 is superior due to its good distance error performance

compared to all other distance derivation procedures. Serious disadvantages of MDSI3 did not become evident during a considerable number of test measurements under various conditions. Thus MDSI3 was chosen as distance derivation method for the UseRCams camera system; also the camera performance measurements in section 4.2.3 and the considerations in chapter 6 are based on distance measurements using MDSI3. See also section 4.2.3.4 for evaluation of the system distance measurement performance using MDSI3.

5.2 Further Enhancement Approaches

Besides the internal camera noise, there are also effects from outside the 3D-CMOS camera system, that can negatively influence the measurement result. For the road safety applications, as considered in this thesis, these are environmental influences as well as sensor saturation caused by highly reflective targets. In the following the handling of these perturbations is discussed.

5.2.1 Treatment of Environmental Influences

Environmental influences account for the main group of the exterior perturbations of the 3D-CMOS distance image acquisition. They include all kind of precipitation, temperature fluctuations, and ambient light.

5.2.1.1 Temperature Fluctuations

At the present state of the system development it was operated at temperatures between 0°C and 50°C. The fluctuation of the ambient temperature turned out to influence the offset of the sensor output signals; for this reason offset calibration as described in section 4.3.1 has to be executed when large changes of the ambient temperature (10°C or more) occur. It has also to be considered that the heat dissipation of the laser diodes to the surrounding becomes less efficient for higher ambient temperatures. The diodes heat up and thus lose output power and can even suffer damage (cf. section 4.2.2.3).
Further temperature effects are not yet known.

5.2.1.2 Ambient Light

Ambient light, i.e. light from sources different from the laser modules, contributes to the photo current of the sensor pixels, and affects the measurement signal. For this reason correlated double sampling (described in chapter 4.1.1.2) has been introduced into the image acquisition procedure. This means that in addition to the normal shutter measurement a second measurement with the

same shutter parametrization, but without sending a laser pulse is performed. Subtracting the sensor output signal of this 'dark' measurement from the signal of the measurement with laser illumination eliminates the influence of any ambient light sources. Prerequisite is that the light emitted by the external sources is correlated in time over the duration of the two measurements. As the measurements are carried out within less than 10 μs, this is obvious for all natural and artificial light sources to be considered here (sunlight, street lighting, automobile headlights, ...). To make this technique effective, it has to be ensured that the amount of charges caused by the background light is only a small fraction of the overall photo charge. Otherwise the sensor dynamic range could be seriously reduced. The pulse based measurement technique with integration times in the range of 30 ns to 1 μs and illumination pulses, power of which exceeds the ambient light by a multiple, ensure this at the UseRCams sensor system. Bandpass filters for cutting the laser light wavelength from the light spectrum – indispensable at the continuous wave approaches [8, 9, 11] – could additionally lower the effect of the ambient light, but they are not really necessary.

5.2.1.3 Rain / Snowfall

As it is known, precipitation like rain or snowfall badly influences optical imaging systems, especially, if active illumination is used. The reason for this can be explained by considering the distance ratio of the interfering object (raindrop or snowflake) and the object to be measured. Thus small objects close to the camera can reflect more light into the camera objective than a large object farther away from the camera (water drop radius typically varies between 10^2 μm and 10^4 μm). For clarification figure 5.5 shows the light power on a single sensor pixel caused by a raindrop with a diameter of 6 mm and by an object with a Lambertian reflectivity of 0.1, which coveres the whole pixel. The raindrop is modeled as a sphere of water, partially reflecting the incident light from its surface according to the well-known optical reflection laws. A part of this reflected light is collected by the camera objective and imaged on the sensor surface. It can be seen that drops at near distances (0 m - 1 m) can easily induce higher light power than objects at the usual observation distances (over 2 m).
The effect of such interferences on the distance measurement is shown in figure 5.6. Here the measurement result of the distance of an object with a surface reflectivity of 0.1 at a distance of 8 m is considered under the presence of a raindrop. The red curve shows the measured object distance in dependence of the distance of the raindrop from the camera. The conclusion is that the corruption of the measured value is very significant up to a distance of 1 m;

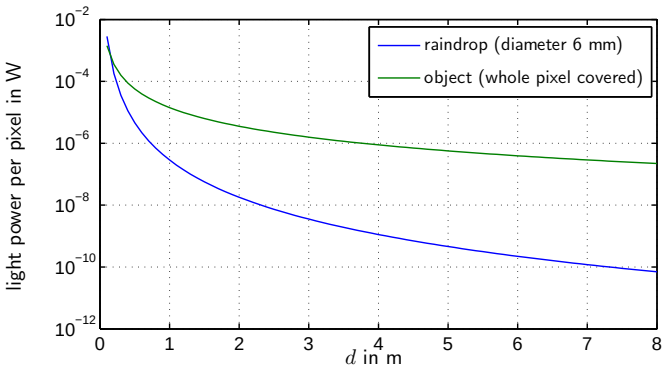


Figure 5.5: Radiant power on the sensor surface resulting from a raindrop resp. a Lambertian reflecting object

between 1 m and 2 m a small effect can be observed, for raindrop distances over 2 m the influence is not significant anymore. Reason for this fast decrease of the raindrop influence is the constant size of the raindrop in the object space for increasing distances. As the raindrop is smaller than a pixel in the object space, the size of its image on the sensor surface decreases according to square law for increasing distances. So does the irradiance induced by the raindrop. The result is a dependence of the light power on the sensor surface generated by the drop, which is proportional to d^4. In contrast to the drop the proportionality factor is only d^2 for observed objects which totally cover the pixel area in the object space. This is guaranteed for the major part of objects considered in the road safety applications.

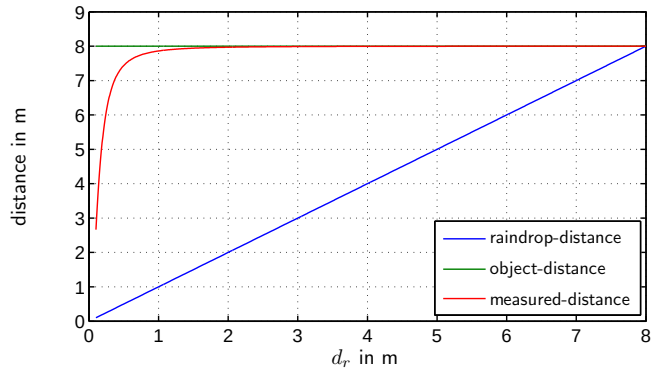


Figure 5.6: Measured distance for a object with reflectivity $\rho = 0.1$ with a single raindrop in the VOV at different distances

Thus the idea is to suppress the influences by means of shifting the shutter window in such a manner that the light reflected from objects in near distance range does not contribute to the shutter signal. Figure 5.7 illustrates the procedure. The proposed shutter window shift is only possible for a distance calculation method using two type b) shutter alignments (cf. figure 3.6). Hence the gradient method has to be applied. In order to gate a distance range from 0 m to d_G, the gating time T_G has to be adjusted to

$$T_G = 2 \cdot \frac{d_G}{c} \ . \tag{5.11}$$

Caused by the shutter alignment of the gradient method, a transition region follows the gated region, where no measurement is possible, as only one of the shutters gets a signal here. It reaches from d_G to d_T with

$$d_T = \frac{c}{2} \cdot (T_G + \Delta T_{Int}) \ , \tag{5.12}$$

where ΔT_{Int}, the delay between the two shutter windows, typically lies between 5 ns and 50 ns depending on the cameras distance range. After this transition region the normal measurement range begins. Measurements are possible up to a distance of

$$d_{max} = \frac{c}{2} \cdot (T_G + T_{Int,b1}) \ . \tag{5.13}$$

Figure 5.8 shows the results of laboratory measurements of a scene consisting of a large target at a distance of 5.0 m and a glass bead in front of the target at a distance of 0.5 m. The 4 mm diameter glass bead acted as rain drop dummy and was hung on a flimsy wire in the VOV of a certain pixel. Several distance measurements were conducted for different values of the gating time T_G. According to the theory (green curve) no influence on the distance measurement should have been present for gating times above $3.33{\cdot}10^{-9}$ s (corresponding to the bead distance of 0.5 m). But the real measurement yields an influence up to a T_G of $10{\cdot}10^{-9}$ s. This is caused by the non-ideal shape of the shutter window function and the laser pulse, introduced in chapters 4.2.1.2 and 4.2.2.1. The missing rectangular shape causes blurring of the gating distance. Nevertheless, the measurement error could be lowered by a factor of 10.5 from 1.05 m to 0.10 m by introducing a gating time T_G of $10{\cdot}10^{-9}$ s.

5.2.2 Saturation Treatment

Independent of the used distance derivation algorithm, sensor saturation can adversely influence distance measurement. This happens if the sense capacitor C_{sense} (see chapter 4.1.1.1) is completely discharged during a single integration

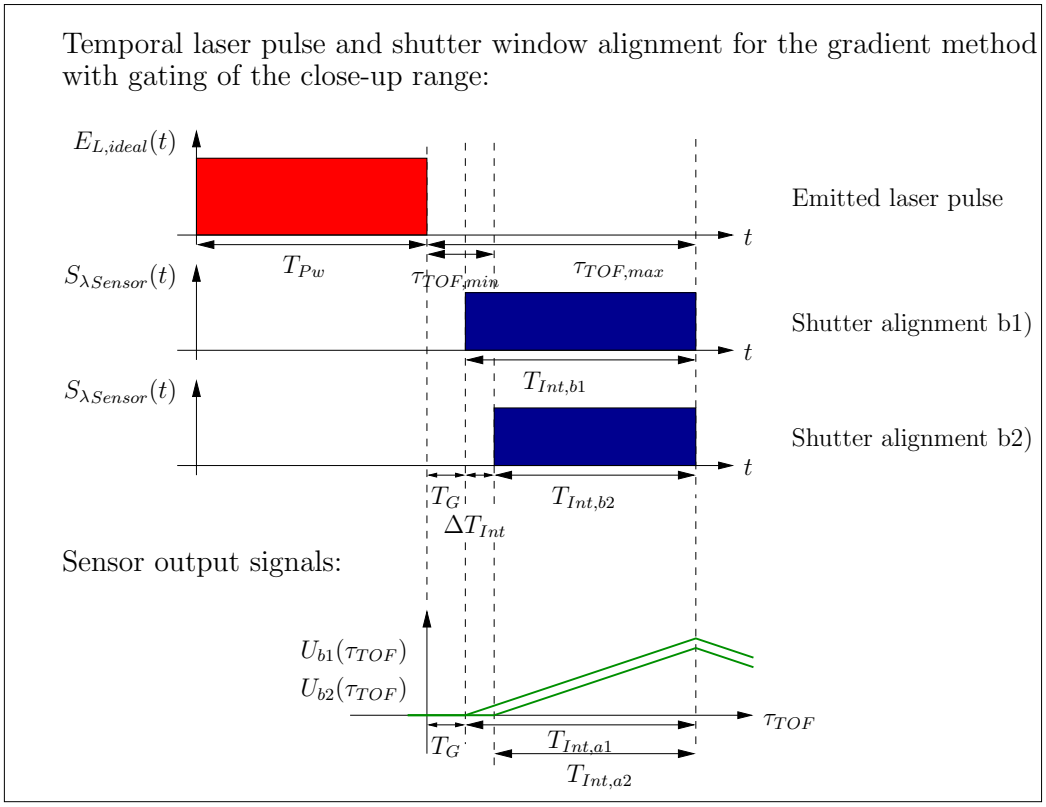


Figure 5.7: Alignment of a shutter windows relatively to the incident light pulse at the gradient method with gating of the close-up range

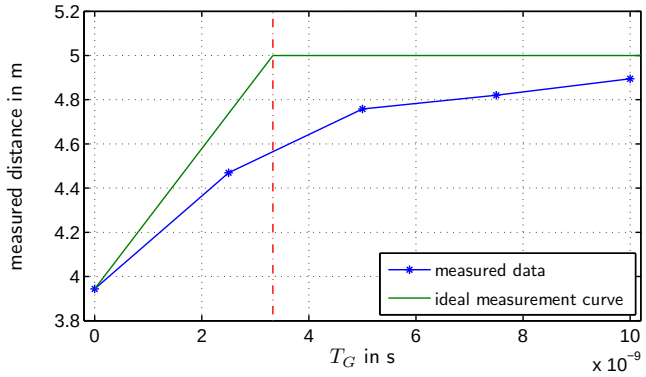


Figure 5.8: Distance measurement results for different gating times T_G for a target at 5 m and a glass bead at 0.5 m distance

step, and thus the amount of photo charges generated by the incident light cannot be evaluated. This means that with the actual sensor parametrization a distance measurement is not possible for this pixel. Reason for the saturation is a unusually high irradiance on the sensor surface, which is not covered by the sensors dynamic range (cf. section 4.2.3.3). It is ether caused by targets being very close to the camera system or by targets with retro-reflective target surfaces, because they redirect the incident light to the light source, i.e., the laser light emitted by the camera is directly reflected to the camera. Compared to Lambertian reflecting targets, which reflect the incident light into the whole half-sphere over the target surface, the resulting irradiance on the sensor chip can reach values up to a few 100 times higher at equal target distances.
A technique was developed in order to manage distance measurement for saturated pixels. The approach is similar to the gating approach above; also here the gradient method is applied. In order to decrease the generated photo charge, the overlap between laser window and shutter pulse is decreased by shifting the laser pulse by means of a electronic delay component on the sensor board. This procedure is done according to a certain shift pattern, until unsaturated sensor signals are received that allow distance calculation. Disadvantage of this algorithm is the increased time consumption; several measurement steps can be necessary to receive unsaturated sensor signals. Also the measurement accuracy drops due to the use of the gradient measurement method (cf. section 5.1.1)
The procedure was successfully tested in the laboratory; yet it was not implemented in the standard sensor firmware. The reason is that the above mentioned time consumption of the saturation mitigation algorithm is not feasible with the presently available budget of laser pulses, i.e. the technique lowers the frame rate of the system too much.

Chapter 6

Novel 3D Real-Time Image Processing Approach

Classical *3D real-time image processing* is usually based on stereo images [5], and thus suffers from the problems mentioned in 2.2.2.1. Very few applications in 3D image processing rely on direct distance image acquisition: a rare example are geometry acquisition systems which employ scanning methods [14], e.g. for machine vision. They, however, are cumbersome and slow. Since the 3D camera presented in this work provides 3D real-time images, a new approach to real-time 3D image processing had to be developed in this work [31]. Goal is thereby the extraction of the needed information from the raw data and providing it in an appropriate data format. These procedures are subject of the following chapters. Background is the use of the camera in road safety applications as described in chapter 1; though the described procedures could also be applied to many other tasks of todays computer vision. The basic flow of the proposed image processing is shown in figure 6.1.

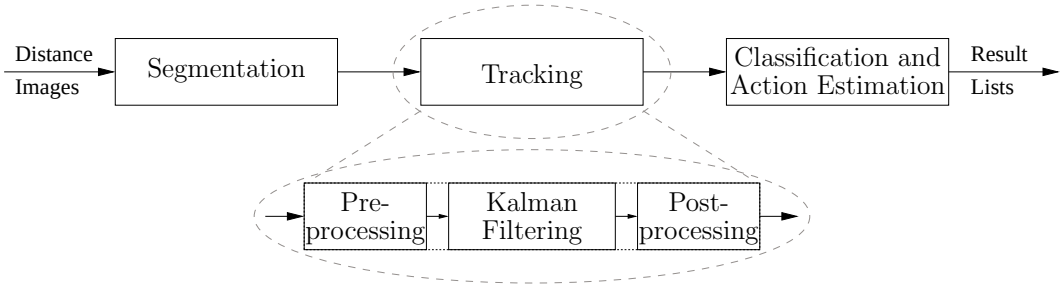


Figure 6.1: Image processing chain

The image processing chain is divided into three main tasks, which are repeated for each image frame in real time, i.e. parallel to the image acquisition of the

3D-CMOS camera. First part is the *segmentation* of the images, that mainly aims at the isolation and position determination of objects from single 3D images. Following the *tracking* module performs Kalman filtering to derive further object properties like velocity or even acceleration extracted from the temporal movement of an object over a sequence of frames. The last step is the processing of the entire information for *classification* of the detected objects and *action prediction* that should give a reliable prediction of the movements of the object in the next moments.

6.1 Raw Data Properties

Basis for the image processing are the raw data delivered by the 3D-CMOS camera. What differs from conventional 2D image processing, where only amplitude values – representing gray scale values or colors of the observed scene – are available for each pixel, is that additional distance information is present here and can be utilized in the image processing algorithms. Further, the raw data contain the number of analog accumulations (cf. section 4.1.1.3) that were used for the acquisition of the corresponding pixel signal. Together with the knowledge of the systems noise behavior described in chapter 4.2.3.1, the expected standard deviation of the sensor signals and the resulting distance uncertainty value can be derived (chapter 5.1.1). This capability will be utilized in section 6.4.3.3.
Thus the pixels raw data contain the following information as input to the image processing chain:

- Pixel position in pixel coordinates
- Distance value
- Sensor signal amplitudes
- Number of analog accumulations

6.2 Coordinate Systems and Coordinate Transformations

For description of the image acquisition with the 3D-CMOS camera and the image processing, the introduction of several different coordinate systems is necessary. Even though each coordinate system (CS) itself enables the description of point coordinates in the 3D-space sufficiently, switching between

different CSs during the image acquisition and processing processes is unavoidable for reasons of observational clearness, computational efforts or feasibility in potential applications. If the relative position between two CSs is known, point coordinates can be transformed arbitrarily between them. In the following the here used coordinate systems as well as the transformation rules are described.

6.2.1 Pixel and Sensor Coordinate System

The fundamental CS of the 3D-CMOS camera is the *pixel coordinate system*, since the camera raw data on sensor level are represented in pixel coordinates. A point position P_P is here described by three coordinates (see figure 6.2): The first two are the horizontal (defining the sensor pixel column) and the vertical (defining the sensor pixel row) pixel index u_P and v_P respectively, which describe the sensor plane position of the projection of a real-world object point. Third is the distance d_{P_P} measured for the according pixel. Thus no further knowledge of sensor or camera geometry is necessary for a unique localization of a point in the 3D-space. Up to the end of the segmentation, all image processing is performed on the level of the pixel CS.

Quite similar to the pixel CS is the *sensor coordinate system*, which is also depicted in figure 6.2. A point position in sensor coordinates is described by means of two coordinates x_{S_P} and y_{S_P}, assigning the position of the points image on the sensor surface, and the measured distance $d_{S_P} = d_{P_P}$. x_{S_P} and y_{S_P} are measured from an arbitrary origin on the sensor plane; normally the intersection point of the cameras optical axis and the sensor surface is assigned as origin.

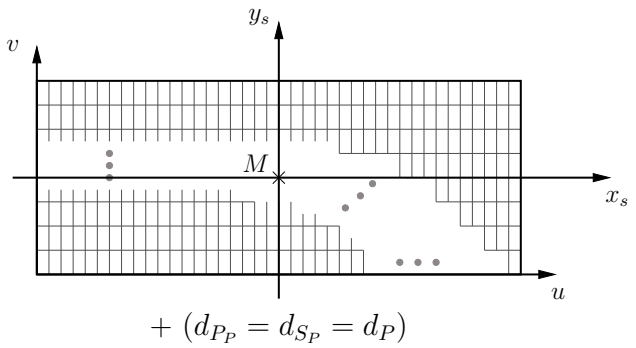


Figure 6.2: Pixel and sensor coordinate systems

In order to convert the representation of a point P from pixel to sensor coor-

dinates, following transformation has to be conducted:

$$P_S = \begin{bmatrix} x_{S_P} \\ y_{S_P} \\ d_P \end{bmatrix} = \begin{bmatrix} (u_P - u_M) \cdot \Delta x \\ (v_P - v_M) \cdot \Delta y \\ d_p \end{bmatrix} \tag{6.1}$$

where Δx and Δy define the horizontal and the vertical pixel pitch of the image sensor. Thus the knowledge of the sensors geometrical parameters is necessary for this transformation.

6.2.2 Camera and Reference / World Coordinate Systems

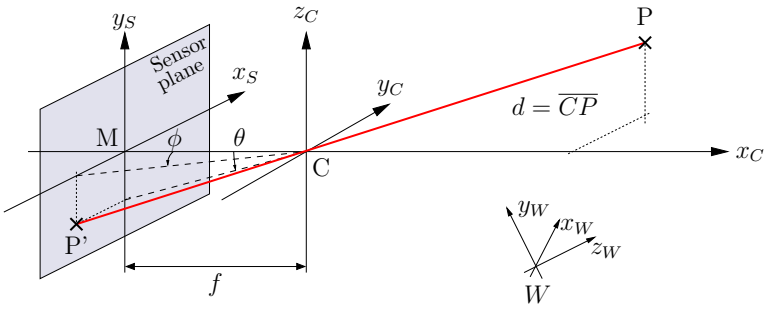


Figure 6.3: Sensor, camera, and world coordinate systems

The *camera coordinate system* represents points in common Cartesian coordinates x_C, y_C and z_C. As figure 6.3 shows, its origin C is located in the optical center of the camera, i.e. in the center of the imaging lens. The x-axis is oriented along the optical axis of the camera, the y-axis to the left, looking in positive x-direction, and the z-axis upwards, each forming an angle of 90° between each other. This results in a clockwise oriented Cartesian coordinate system. Point representation in camera coordinates is used in the tracking module of the image processing procedure. The transformation from sensor coordinates is described below.
With f being the focal length of the camera imaging optics one can derive the angles ϕ and θ describing the angle between the view ray to a point P and the xz-plane resp. xy-plane. Thus one obtains an polar point description $P_{S'}$ of P

based on the sensor coordinate description.

$$\phi = \arctan\left(\frac{x_{S_P}}{f}\right) \tag{6.2}$$

$$\theta = \arctan\left(\frac{y_{S_P}}{f}\right) \tag{6.3}$$

$$P_{S'} = \begin{bmatrix} \phi \\ \theta \\ d_P \end{bmatrix} \tag{6.4}$$

Starting with this description, the transformation from sensor to camera coordinates of a scene point P is executed by following equation:

$$P_C = \begin{bmatrix} x_{C_P} \\ y_{C_P} \\ z_{C_P} \end{bmatrix} = \begin{bmatrix} d \cdot \cos(\phi) \cdot \cos(\theta) \\ d \cdot \sin(\phi) \cdot \cos(\theta) \\ d \cdot \sin(\theta) \end{bmatrix} \tag{6.5}$$

As in most applications the camera coordinate system is not the reference, which position values are related to, the point coordinates in camera coordinate representation have to be related to a *reference coordinate system*. This CS – often called *world coordinate system* – is also clockwise oriented and Cartesian and is defined in the application using the 3D data. In the UseRCams front-view application (see chapter 1) for instance, the reference system is a vehicle inherent coordinate system with its origin being located on floor level on the intersection point of the axes in longitudinal and transverse directions of the vehicle; the x-axis is directed parallel to the longitudinal axis. Its relative position to the camera CS is described by a *translation vector* $\boldsymbol{T}$ and a *rotation matrix* $\boldsymbol{R}$ [32].

$$P_C = \begin{bmatrix} x_{C_P} \\ y_{C_P} \\ z_{C_P} \end{bmatrix} = \boldsymbol{R} \cdot \begin{bmatrix} x_{W_P} \\ y_{W_P} \\ z_{W_P} \end{bmatrix} + \boldsymbol{T} = \boldsymbol{R} \cdot P_W + \boldsymbol{T} \tag{6.6}$$

with $\boldsymbol{T}$ being the translation vector between the coordinate system origins W and C represented in world coordinates

$$\boldsymbol{T} = \begin{bmatrix} x_{W_C} - x_{W_W} \\ y_{W_C} - y_{W_W} \\ z_{W_C} - z_{W_W} \end{bmatrix} = \begin{bmatrix} x_{W_C} \\ y_{W_C} \\ z_{W_C} \end{bmatrix}, \tag{6.7}$$

and $\boldsymbol{R}$ the rotation matrix describing the orientation of the camera and thus also of the camera image with respect to the world CS. It can be assembled from three elementary rotation around the three coordinate system axes [33].

$$\boldsymbol{R}(\psi, \theta, \phi) = \boldsymbol{R}_x(\phi) \cdot \boldsymbol{R}_y(\theta) \cdot \boldsymbol{R}_z(\psi) = \tag{6.8}$$

$$= \begin{bmatrix} \cos(\theta)\cos(\psi) & \cos(\theta)\sin(\psi) & -\sin(\phi) \\ \sin(\phi)\sin(\theta)\cos(\psi) - \cos(\phi)\sin(\psi) & \sin(\phi)\sin(\theta)\sin(\psi) - \cos(\phi)\cos(\psi) & \sin(\phi)\cos(\theta) \\ \cos(\phi)\sin(\theta)\cos(\psi) - \sin(\phi)\sin(\psi) & \cos(\phi)\sin(\theta)\sin(\psi) - \sin(\phi)\cos(\psi) & \cos(\phi)\cos(\theta) \end{bmatrix} \tag{6.9}$$

where ψ, θ and ϕ denote the rotation angles around the coordinate axes. The compliance of the order of the three basic rotations is mandatory to obtain the right rotation matrix $\boldsymbol{R}$.
Analogous to equation 6.6, the transformation from a point in camera coordinates to world coordinates is carried out by applying equation 6.10:

$$P_W = \begin{bmatrix} x_{W_P} \\ y_{W_P} \\ z_{W_P} \end{bmatrix} = \boldsymbol{R}^{-1} \cdot \left(\begin{bmatrix} x_{C_P} \\ y_{C_P} \\ z_{C_P} \end{bmatrix} - \boldsymbol{T} \right) = \boldsymbol{R}^{-1} \cdot (P_C - \boldsymbol{T}) \tag{6.10}$$

6.3 Object Segmentation

The first stage of the image processing chain is the spatial segmentation of the 3D raw images on single frame level. Goal is a reliable partitioning of the images into a set of sub-regions, each of them representing a single object in the VOV of the camera. In contrary to 2D image segmentation [34,35], where segmentation is based on color, signal amplitude or texture, the (x, y, z)-information of each pixel is used here as basic decision criterion for segmentation algorithms. Thus effects of lighting conditions or object surface properties on the segmentation results are not considered.

The proposed algorithm performs following processing steps:

- Data preparation:
  Here the raw data are preprocessed for the segmentation. First 3x1 or 3x3 median filtering [36] is performed, what proved to be useful to reduce noise and outliers that are due to edge effects, adverse object surface properties (reflective/retro-reflective), and other effects. For the UseRCams front view application a subtraction of the floor level is also accomplished, where all pixels, that are supposed to image a part of the floor are removed from the image frame (only objects different from the floor level are of interest in this application). The decision if a pixel is imaging a part of the floor is done by a comparison of the pixels distance value with a stored floor plane, which is either taken from an earlier acquired image of the floor or calculatively derived from the knowledge of the cameras position and rotation in the reference coordinate system. Compare the two lower pixel lines of figures 6.5(e) and 6.5(f) to see the effect of the floor level subtraction on example scene in figure 6.5(d).

- Segmentation:
  The actual segmentation is done by a region growing algorithm [36], as regions exhibiting similar distance values in the raw image must belong to the same object in the cameras VOV with high probability. Starting at an arbitrary image point (in practice one of the image corners), the algorithm goes through the image pixel by pixel. For each pixel with a valid distance value, the adjacent pixels are checked for region affiliation. The criterion if the pixel belongs to the region or not is the difference between the distance value of the pixel and the present mean distance of the object, which has to be lower than an adjustable threshold. If a pixel (exhibiting the measured point coordinates P_{P_i} , $i \in \{u, v, d\}$) is found to be the n^{th} member of the current object, an update of the object position coordinates P_{O_i} is performed:

$$P'_{O_i} = \frac{P_{O_i} \cdot (n-1) + P_{P_i}}{n} \quad , \; i \in \{u, v, d\} \tag{6.11}$$

  Analogously, also other object description parameters (see table 6.1) are updated.
  Two segmentation examples for a synthetic model scene and a real scene are given in figure 6.5. Each starts with a 2D image of the scene (fig. 6.12(a) and fig. 6.5(d)), giving the reflectivity and shape properties of the scenes objects. Secondly, the raw 3D images are given for both scenes (fig. 6.12(b) and fig. 6.5(e)); the distance value measured by a pixel is coded in the pixel color. The distance measurement noise is clearly visible in the changing pixel color within the objects. White areas do not contain

pixels with valid distance values, which means in the most cases that no signal was received as no object is reflecting the emitted laser light. Last figures 6.5(c) and 6.5(f) show the segmented images. The objects contained in the above figures are accurately isolated as image regions, the position in pixel coordinates (u, v, d) and the number of pixel (N_{Pix}) the object consists of are given by the associated text.

- Output:
  As information source for the following image processing steps (tracking and classification), the segmentation algorithm outputs a list containing object description records for all objects found in the current frame, called object list (OL) of the k-th frame, $OL(k)$. Additionally some global parameters like the acquisition time t_{acqu} of the corresponding image frame and the number of objects N_O contained in the OL are stored. Figure 6.4 shows the basic structure of an OL and the essential contents of a object description record; table 6.1 explains these parameters.

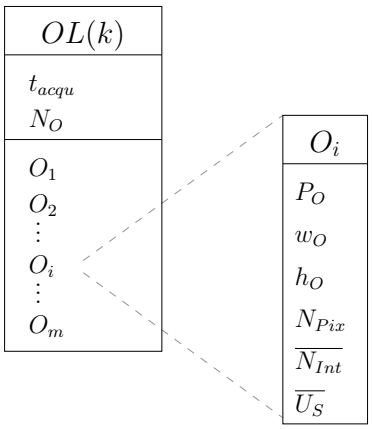
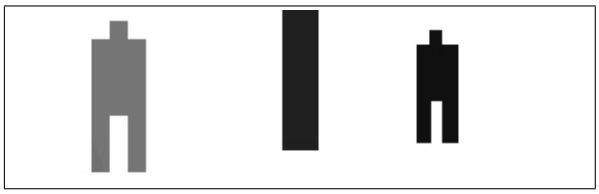


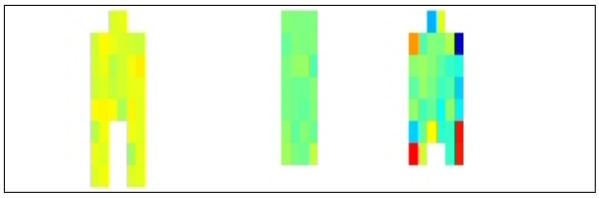
Figure 6.4: Structure of the object list of an image frame

Table 6.1: Main parameter of the description format for segmented objects

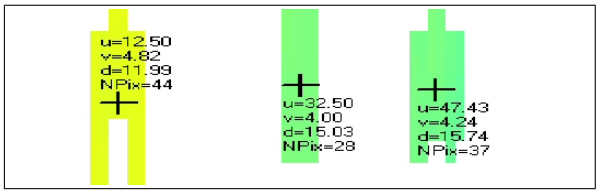
Parameter	Description
P_O	Position of the object centroid in pixel coordinates
w_O	Object width in pixels
h_O	Object height in pixels
N_{Pix}	Number of pixels the object consists of
$\overline{N_{Int}}$	Average of the analog integrations used for image acquisition over all object pixels
$\overline{U_s}$	Average of the sensor signal amplitude over all object pixels



(a) Synthetic scene

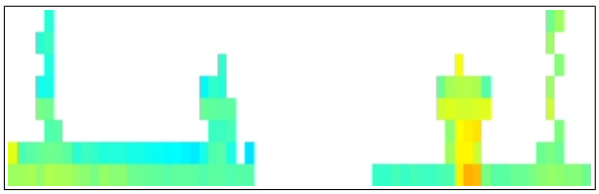


(b) Synthetic 3D-CMOS depth image of the scene above

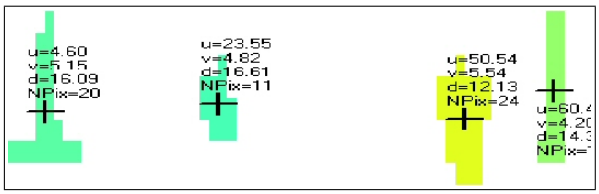


(c) Result of the segmentation of the synthetic depth image

(d) Real scene



(e) 3D-CMOS depth image of the real scene



(f) Result of the segmentation of the real depth image

Figure 6.5: Segmentation examples of synthetic and real depth images

6.4 Object Tracking

Object tracking includes all processing steps that account for the tracing and the extraction of state parameters of moving objects based on a sequence of observations over time. Several techniques – applied on 2D camera data, radar data, or the like – are known in literature [37, 38, 34].
Accordingly the object tracking module developed in this work was designed for the tracing and the extraction of kinematic parameters – including object position, velocity and acceleration – of segmented objects in the image frame sequences of the 3D-CMOS camera. The tracking procedure is divided into three parts, namely *preprocessing*, *Kalman filtering*, and *postprocessing* (cf. figure 6.1), which are addressed in the following sections.

6.4.1 Preprocessing

To provide the information necessary for the Kalman filtering, several preprocessing steps have to be carried out. They include:

- Coordinate transformation from pixel to reference coordinates according to section 6.2; in the case considered here the reference coordinate system is a vehicle inherent coordinate system (c.f. section 6.2.2). From here any position or distance measures are given in Cartesian coordinates.
- Evaluation of the time T elapsed between the last and the current image frame:
$$T = t_{acqu}(k) - t_{acqu}(k-1) \ , \qquad k \in \mathbb{N} \tag{6.12}$$
with $t_{acqu}(k)$ being the point of time of the acquisition of the k-th frame.
- Frame to frame object recovery

While the coordinate transformation and the evaluation of T are trivial mathematical operations performed on the output data of the segmentation procedure, the recovery of the objects of the last frame in the current frame, requires elaborate image processing, which is described below.

6.4.2 Frame to Frame Object Recovery

The *frame to frame object recovery* procedure (also known as tracing) links the single frames of an 3D image sequence to each other. Basis are the object lists generated during segmentation. Goal of this procedure is the reliable assignment of objects contained in the object list $OL(k-1)$ of the last frame to the objects contained in the object list $OL(k)$ of the current frame (fig.

6.6). This tracing of objects over a sequence of 3D image frames is an essential part in the proposed image processing approach. Wrong assignments entirely distort the tracking results and thus influence the scene recognition capability of the camera system in a very severe way. Sections 6.4.2.1 and 6.4.2.2 describe the algorithms implemented to solve this task.

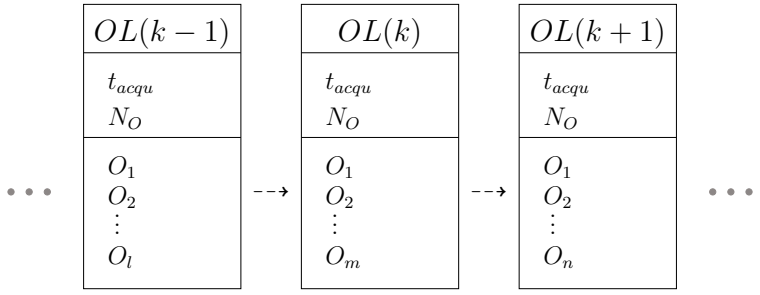


Figure 6.6: Series of consecutive object lists for the $(k-1)$-th, the k-th, and the $(k+1)$-th time step

6.4.2.1 Inter-Object Distance Computation

The object recovery algorithm developed in this work is based on the evaluation of the distance between the centroids of the objects of the $(k-1)$-th and the k-th frame. As distance between two centroids P and S at position $\vec{\boldsymbol{s}}$ and $\vec{\boldsymbol{p}}$ respectively one could choose the Euclidean distance

$$a_E = \sqrt{\sum_{i=1}^{n}(p_i - s_i)^2}, \tag{6.13}$$

with n being 3 in the here considered case for the three dimensions x, y, and z of the Cartesian space.

But using this distance computation method, it is not possible to weight the measurements acquired for different spatial directions. This is necessary, however, since the quality of distance measurement for the different spatial dimensions x, y, and z varies. A large measurement variance is an indication of noisy, and ,therefore, inaccurate determination of that particular spatial dimension. For this reason it should be handled in a different way than a measurement of another spatial dimension exhibiting small variance, which is obviously less noisy and thus more reliable. Using the Mahalanobis distance instead solves

this problem as weighting of the distance components is performed here. The Mahalanobis distance between P and S is computed by equations 6.14-6.15; thereby the component distances $(p_i - s_i)$ are weighted with the reciprocal of the according distance measurement noise σ_x, σ_y, or σ_z. This accounts for the variations in distance measurement due to the measurement noise.

$$a_M = \sqrt{(\vec{\boldsymbol{p}} - \vec{\boldsymbol{s}})^T \boldsymbol{S}^{-1} (\vec{\boldsymbol{p}} - \vec{\boldsymbol{s}})} \tag{6.14}$$

with

$$\boldsymbol{S}^{-1} = \begin{bmatrix} \frac{1}{\sigma_x} & 0 & 0 \\ 0 & \frac{1}{\sigma_y} & 0 \\ 0 & 0 & \frac{1}{\sigma_z} \end{bmatrix} \tag{6.15}$$

As shown in figure 6.7, all points with equal Mahalanobis distance a_M lie on the surface of a 3-dimensional ellipsoid with the three radii r_x, r_y, and r_z which correspond to the diagonal elements of matrix $\boldsymbol{S}$.

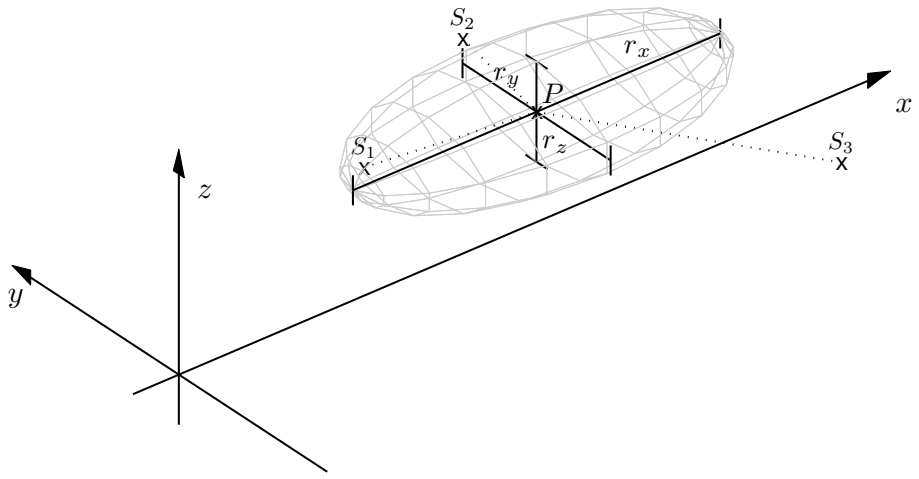


Figure 6.7: Example ellipsoid of constant Mahalanobis distance

6.4.2.2 Object to Object Assignment

For assigning an object i of $OL(k-1)$ to an object j of $OL(k)$ the Mahalanobis distance (see equation 6.14) between $O(i)$ of $OL(k-1)$ and $O(j)$ of $OL(k)$ is computed for all j. If the lowest of these distances is lower than an adjustable distance threshold, $O(j)$ of $OL(k)$ is assumed to be the same object as $O(i)$ of $OL(k-1)$.
Once an object $O(j)$ of $OL(k)$ is assigned, it is not available for further assignments anymore. As measure for the 'age' of an object the number of assignments, i.e. the number since how many frames the object exist is stored in an object associated counter.
In order to prevent wrong assignments that 'snatch away' the potential succeeding objects, prioritization rules for the order of object assignments are implemented. Basis for the prioritization is a priority value that is computed from

- the object age (counted in frames),
- the number of pixels N_{Pix} the object consists of, and
- the reliability of the measurement derived from $\overline{U_S}$ and $\overline{N_{Int}}$ (cf. table 6.1).

As a result of the object recovery procedure each object contained in $OL(k)$ is either assigned to an object of $OL(k-1)$ or identified as a new object, that appeared the first time in the current frame. This information is then used by the Kalman filter introduced below.

6.4.3 Kalman Filtering

Kalman filtering [39, 40] basically yields the best possible estimation of the present state of a time-varying system. Sources of information are thereby knowledge about the temporal behavior of the system and state measurements, which contain information about the system states plus inherent measurement noise, which is unavoidable for observations of any physical process. Retrieving the available information during minimizing the influence of the noise is thus the goal of the Kalman filtering operation. The Kalman filtering approach of this work is closely described below, following [41] and [42].

6.4.3.1 Process Description

The process to be estimated has to be discrete-time controlled and describable by a linear differential equation as shown in equation 6.16. Here the considered

process is the movement process of objects in the VOV of the camera. By adequate choice of the equation parameter, the process is being modeled as good as possible.

$$\boldsymbol{x}(k) = \boldsymbol{A}(k) \cdot \boldsymbol{x}(k-1) + \boldsymbol{B}(k) \cdot \boldsymbol{u}(k-1) + \boldsymbol{w}(k-1) \ , \tag{6.16}$$

with

$\boldsymbol{x}(k)$	:	n-dimensional system state vector, $\boldsymbol{x} \in \mathbb{R}^n$ at the k-th time step
$\boldsymbol{u}(k)$	:	l-dimensional control input vector, $\boldsymbol{u} \in \mathbb{R}^l$ at the k-th time step
$\boldsymbol{w}(k)$	:	n-dimensional process noise vector, $\boldsymbol{w} \in \mathbb{R}^n$ at the k-th time step
$\boldsymbol{A}(k)$	:	$n \times n$-dimensional state transition matrix at the k-th time step
$\boldsymbol{B}(k)$	:	$n \times l$-dimensional matrix relating $\boldsymbol{u}$ to $\boldsymbol{x}$ at the k-th time step

The state vector $\boldsymbol{x}(k)$ of the system is the smallest dimension vector that contains all the parameters of the system which are of interest in the according case of use. In the here considered application, the object *position* (x, y, z), *velocity* $(\dot{x}, \dot{y}, \dot{z})$, and *acceleration* $(\ddot{x}, \ddot{y}, \ddot{z})$ in all three space dimensions are the subjects of interest, i.e. a 3-dimensional accelerated object movement model is used. This leads to the 9-dimensional state vector:

$$\boldsymbol{x}(k) = \begin{bmatrix} \boldsymbol{x}_x(k) \\ \boldsymbol{x}_y(k) \\ \boldsymbol{x}_z(k) \end{bmatrix}_{9\times 1} \text{ where } \boldsymbol{x}_x = \begin{bmatrix} x(k) \\ \dot{x}(k) \\ \ddot{x}(k) \end{bmatrix}, \boldsymbol{x}_y = \begin{bmatrix} y(k) \\ \dot{y}(k) \\ \ddot{y}(k) \end{bmatrix} \text{ and } \boldsymbol{x}_z = \begin{bmatrix} z(k) \\ \dot{z}(k) \\ \ddot{z}(k) \end{bmatrix} \tag{6.17}$$

The state transition matrix $\boldsymbol{A}(k)$ relates the system state at the $(k-1)$-th time step to the system state at the k-th time step according to the physical behavior of the system. Thus for the accelerated movement model in the 3-dimensional space one obtains by applying the well-known *Newtonian motion equations*:

$$\boldsymbol{A}(k) = \begin{bmatrix} \boldsymbol{A}_x(k) & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_y(k) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_z(k) \end{bmatrix}_{9\times 9}, \boldsymbol{A}_i(k) = \begin{bmatrix} 1 & T & \frac{1}{2}T^2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix} \tag{6.18}$$

Analogously to $\boldsymbol{A}$, $\boldsymbol{B}$ relates the optional control input $\boldsymbol{u}$ to the system state. In the considerations of this work no control input is present at any time, i.e.

$\boldsymbol{u}(k) = \mathbf{0}_{9\times1} \forall k \in \mathbb{N}$. This means that the expression $\boldsymbol{B}(k) \cdot \boldsymbol{u}(k-1)$ does not influence the equation and is thus omitted in the following.

The noise vector $\boldsymbol{w}(k)$ describes the noise of the process at the k-th time step. It is a very important part of the process model, as it models the uncertainty of the object movements, i.e. the deviation from the course described by matrix $\boldsymbol{A}$. For the description of process noise the *piecewise constant acceleration increment model* [42] was chosen. This models the acceleration as a Wiener process, i.e. process with independent increments, which are modeled as zero-mean white noise. Thus the single noise contributions are assumed as uncorrelated. The process noise vector is:

$$\boldsymbol{w}(k) = \begin{bmatrix} \boldsymbol{w}_x(k) \\ \boldsymbol{w}_y(k) \\ \boldsymbol{w}_z(k) \end{bmatrix}_{9\times1}, \ \boldsymbol{w}_i(k) = \Gamma(k) \cdot w_i(k) = \begin{bmatrix} \frac{1}{2}T^2 \\ T \\ 1 \end{bmatrix} \cdot w_i(k) \tag{6.19}$$

$$p(\boldsymbol{w}) \sim N(\mathbf{0}, \boldsymbol{Q}) \tag{6.20}$$

The conducted measurements result in a measurement vector $\boldsymbol{z} \in \mathbb{R}^m$, which is here, like in most systems, different to $\boldsymbol{x}$, as not all system state variables are accessible to direct measurement. The matrix $\boldsymbol{H}$ relates the state vector to the measurement vector as described by equation 6.21.

$$\boldsymbol{z}(k) = \boldsymbol{H} \cdot \boldsymbol{x}(k) + \boldsymbol{v}(k) \tag{6.21}$$

with $\boldsymbol{v}$ being additive measurement noise, assumed to be white, uncorrelated and normally distributed:

$$p(\boldsymbol{v}) \sim N(\mathbf{0}, \boldsymbol{R}) \tag{6.22}$$

In the here considered case $\boldsymbol{H}$ is

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{6.23}$$

6.4.3.2 Basics of Kalman Filtering

State estimation based on Kalman filtering is a recursive process according to figure 6.8, that consists of a *prediction* and a *correction* step. Basic principle is the sophisticated combination of the knowledge about the object history

and its expected behavior (*time update*) and the new information introduced by the measurement result (*measurement update*). Each time a new measurement result $\boldsymbol{z}(k)$ is available, a *prediction/correction-cycle* is performed for the derivation of the corresponding estimate $\hat{\boldsymbol{x}}(k)$.

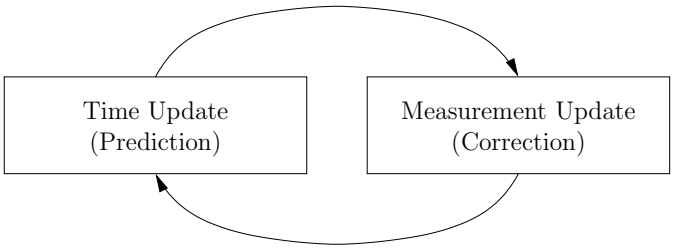


Figure 6.8: Kalman filter principle

Derivation of the Kalman Filter Equations

Let $\hat{\boldsymbol{x}}^-$ be the *a priori* and $\hat{\boldsymbol{x}}$ the *a posteriori* state estimate of a system state vector $\boldsymbol{x}$, where $\hat{\boldsymbol{x}}^-$ indicates the state determined in the *time update* and $\hat{\boldsymbol{x}}$ the result of the *measurement update*. The errors of the a priori and the a posteriori state estimates, i.e. their deviations from the state $\boldsymbol{x}(k)$, are defined by

$$\boldsymbol{e}^-(k) = \boldsymbol{x}(k) - \hat{\boldsymbol{x}}^-(k) \tag{6.24}$$

$$\boldsymbol{e}(k) = \boldsymbol{x}(k) - \hat{\boldsymbol{x}}(k) \quad . \tag{6.25}$$

From this one obtains the a priori respectively a posteriori error covariance matrix $\boldsymbol{P}^-(k)$ and $\boldsymbol{P}(k)$ as autocovariance matrix of $\boldsymbol{e}^-(k)$ and $\boldsymbol{e}(k)$ ($E[\cdot]$: expectation):

$$\boldsymbol{P}^-(k) = E[\boldsymbol{e}^-(k)\boldsymbol{e}^-(k)^T] \tag{6.26}$$

$$\boldsymbol{P}(k) = E[\boldsymbol{e}(k)\boldsymbol{e}(k)^T] \tag{6.27}$$

The Kalman filter cycle starts with the *time update*. It contains two steps, the prediction of the new a priori state estimate $\hat{\boldsymbol{x}}^-(k)$ state by inserting $\boldsymbol{x}(k-1)$ into equation 6.16 and the update of the of its error covariance matrix $\boldsymbol{P}^-(k)$. The according formulas are shown in figure 6.9 in the *time update* box. Initialization of $\hat{\boldsymbol{x}}(k-1)$ and $\boldsymbol{P}(k-1)$ has to be done at the first run of the cycle; different procedures for that are proposed in literature [42, 43].

Following the *measurement update* is executed. It contains the computation of the so-called *Kalman gain* $\boldsymbol{K}$, the calculation of the a posteriori state estimate $\hat{\boldsymbol{x}}$ – the actual Kalman filter output – as a linear combination of the a priori estimate $\hat{\boldsymbol{x}}^-(k)$ and a weighted difference between an actual measurement $\boldsymbol{z}$ and a measurement prediction ($\boldsymbol{H}\hat{\boldsymbol{x}}^-$), as well as the update of the a posteriori error covariance matrix $\boldsymbol{P}(k)$. The *measurement update* box in figure 6.9 shows the three equations, [39,40,43] give a close description of the derivation of them, especially of the weighting factor $\boldsymbol{K}$ (*Kalman gain*).

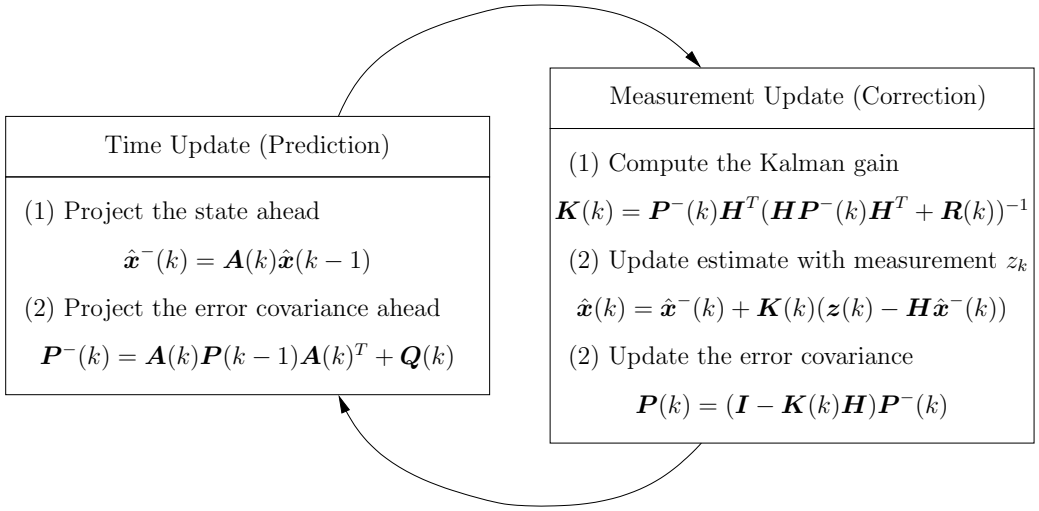


Figure 6.9: Kalman filter formulas

6.4.3.3 Derivation of the Noise Covariance Matrices $\boldsymbol{Q}(k)$ and $\boldsymbol{R}(k)$

As defined in equations 6.20 and 6.22, in Kalman filtering the uncertainties of the observed process and the uncertainty of the Kalman filter inputs, i.e. the measurement results $\boldsymbol{z}(k)$, are described by the noise covariance matrices $\boldsymbol{Q}(k)$ and $\boldsymbol{R}(k)$. While the process description matrix $\boldsymbol{A}(k)$ is governed by physical laws describing the system (cf. equation 6.18) and $\boldsymbol{H}$ only contains assignment rules between the measurement vector components and the state vector components, $\boldsymbol{Q}(k)$ and $\boldsymbol{R}(k)$ are the two parameters that directly influence the dynamical behavior of the Kalman filter. Thus it is important to reproduce the reality as good as possible when determining these matrices in order to obtain good filtering results. On the other hand specific changes of $\boldsymbol{Q}(k)$ and $\boldsymbol{R}(k)$ allow tuning of the estimation behavior of the filter (see filter tuning in chapter 6.4.3.4). The following paragraphs show the derivation of

$\boldsymbol{Q}(k)$ and $\boldsymbol{R}(k)$ in view of the traffic object tracking task considered in this thesis.

The Process Noise Covariance Matrix $\boldsymbol{Q}(k)$

$\boldsymbol{Q}(k)$ describes the random input of an observed process, i.e. the deviations of the process from the behavior described in the process description matrix $\boldsymbol{A}(k)$, as they are expected in almost every real system. In the here considered case of the estimation of position, velocity and acceleration of an object in all three dimensions of space, $\boldsymbol{Q}(k)$ is a 9×9 matrix, that consists of three 3×3 sub-matrices, which are located on the matrices diagonal, as shown in equation 6.28.

$$\boldsymbol{Q}(k) = \begin{bmatrix} \boldsymbol{Q}_x(k) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Q}_y(k) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{Q}_z(k) \end{bmatrix} \tag{6.28}$$

Using the *piecewise constant acceleration increment* model as introduced in 6.4.3.1, the process noise covariance sub-matrix $\boldsymbol{Q}_i$ results as

$$\boldsymbol{Q}_i \overset{[29]}{=} E[(\boldsymbol{w}_i(k) - \underbrace{E[\boldsymbol{w}_i(k)]}_{=0}) \cdot (\boldsymbol{w}_i(k) - \underbrace{E[\boldsymbol{w}_i(k)]}_{=0})'] \tag{6.29}$$

$$= E[\boldsymbol{w}_i(k) \cdot \boldsymbol{w}_i(k)'] \tag{6.30}$$

$$\overset{Eq.6.19}{=} E[(\Gamma(k) \cdot w_i(k)) \cdot (\Gamma(k) \cdot w_i(k))'] \tag{6.31}$$

$$= \Gamma(k) \cdot \sigma_{w_i}^2 \cdot \Gamma(k)' \tag{6.32}$$

$$= \begin{bmatrix} \frac{1}{4}T^4 & \frac{1}{2}T^3 & \frac{1}{2}T^2 \\ \frac{1}{2}T^3 & T^2 & T \\ \frac{1}{2}T^2 & T & 1 \end{bmatrix} \cdot \sigma_{w_i}^2 \quad , \tag{6.33}$$

with T being the time elapsed between the $(k-1)$-th and the k-th time step. The values of the σ_{w_i} have to be closely investigated in order to determine the parameters of the Kalman filter correctly. See chapter 6.4.3.4 for a deeper investigation of this of this topic.

The Measurement Noise Covariance Matrix $\boldsymbol{R}$

To derive the *measurement noise covariance matrix* $\boldsymbol{R}$, we start with a noise analysis on sensor level, as is obvious, that the three basic measurement noise components are related to the three dimensions of a point in sensor coordinates. Thus the basic noise components are noise of the determination of the coordinates of a point on the sensor surface (angular noises σ_ϕ and σ_θ according to figure 6.3) and noise of the distance measurement (distance noise σ_d). As a good approximation these values are assumed as zero mean and uncorrelated between each other. Hence one obtains following covariance matrix of the measurement noise in polar sensor coordinates:

$$\boldsymbol{R}_S = \begin{bmatrix} \sigma_\phi^2 & 0 & 0 \\ 0 & \sigma_\theta^2 & 0 \\ 0 & 0 & \sigma_d^2 \end{bmatrix} \tag{6.34}$$

As this representation of the measurement noise is not of use for the considered Kalman filtering approach on the basis of object movements in a Cartesian space (world coordinates), $\boldsymbol{R}_S$ has to be translated to the world coordinate system [44]. The result is the measurement covariance matrix $\boldsymbol{R}$ described by equations 6.35-6.37.

$$\boldsymbol{R} = \begin{bmatrix} \sigma_{\boldsymbol{v}_x}^2 & Cov(\boldsymbol{v}_x, \boldsymbol{v}_y) & Cov(\boldsymbol{v}_x, \boldsymbol{v}_z) \\ Cov(\boldsymbol{v}_x, \boldsymbol{v}_y) & \sigma_{\boldsymbol{v}_y}^2 & Cov(\boldsymbol{v}_y, \boldsymbol{v}_z) \\ Cov(\boldsymbol{v}_x, \boldsymbol{v}_z) & Cov(\boldsymbol{v}_y, \boldsymbol{v}_z) & \sigma_{\boldsymbol{v}_z}^2 \end{bmatrix} \tag{6.35}$$

where

$$\sigma_{\boldsymbol{v}_i}^2 = \sigma_{\phi_i}^2 + \sigma_{\theta_i}^2 + \sigma_{d_i}^2 \quad , \; i \in \{x, y, z\} \tag{6.36}$$

$$Cov(\boldsymbol{v}_i, \boldsymbol{v}_j) = \sigma_{\phi_i} \cdot \sigma_{\phi_j} + \sigma_{\theta_i} \cdot \sigma_{\theta_j} + \sigma_{d_i} \cdot \sigma_{d_j} \; , i, j \in \{x, y, z\} \; , i \neq j \tag{6.37}$$

A noise component σ_{p_c} in equations 6.36 and 6.37 represents the noise of the measurement of coordinate p, $p \in \{\phi, \theta, d\}$ in polar representation that reappears in the noise of coordinate c, $c \in \{x, y, z\}$, in Cartesian representation. The 9 different σ_{p_c} for all possible combinations of polar and Cartesian coordinates are displayed in table 6.2.

Table 6.2: Component noise contributions σ_{p_c} for $\sigma_\phi \ll 1$ and $\sigma_\theta \ll 1$

	x	y	z
ϕ	$\sigma_{\phi_x} = d \cdot \sigma_\phi \cdot \sin(\phi)$	$\sigma_{\phi_y} = d \cdot \sigma_\phi \cdot \cos(\phi)$	$\sigma_{\phi_z} = 0$
θ	$\sigma_{\theta_x} = d \cdot \sigma_\theta \cdot \cos(\phi) \cdot \sin(\theta)$	$\sigma_{\theta_y} = d \cdot \sigma_\theta \cdot \sin(\phi) \cdot \sin(\phi)$	$\sigma_{\theta_z} = d \cdot \sigma_\theta \cdot \cos(\theta)$
d	$\sigma_{d_x} = \sigma_d \cdot \cos(\phi) \cdot \cos(\theta)$	$\sigma_{d_y} = \sigma_d \cdot \sin(\phi) \cdot \cos(\theta)$	$\sigma_{d_z} = \sigma_d \cdot \sin(\theta)$

Thus the calculation of the basic noises σ_ϕ, σ_θ, and σ_d is crucial for a proper functionality of the Kalman filter. They depend on several system parameters, the segmentation results of the observed object, as well as on the utilized distance calculation procedure. 6.38-6.40 show the set of equations; use of distance derivation method MDSI3 (cf. chapter 3.4) is assumed. Using other distance derivation methods the deviation of the distance measurement σ_d has to be adapted according to section 5.1.1.1.

$$\sigma_\phi = \left| \arctan\left(\frac{x_s}{f}\right) - \arctan\left(\frac{x_s + \sigma_{x_s}}{f}\right) \right| \tag{6.38}$$

$$\text{with} \quad \sigma_{x_s} \overset{\text{Eq. 6.1}}{=} \sigma_u \cdot \Delta x$$

$$\sigma_\theta = \left| \arctan\left(\frac{y_s}{f}\right) - \arctan\left(\frac{y_s + \sigma_{y_s}}{f}\right) \right| \tag{6.39}$$

$$\text{with} \quad \sigma_{y_s} \overset{\text{Eq. 6.1}}{=} \sigma_v \cdot \Delta y$$

$$\sigma_d = \frac{c}{2} \cdot T_{Pw} \cdot \frac{1}{\overline{U_c}} \cdot \sqrt{1 + \frac{\overline{U_b}^2}{\overline{U_c}^2}} \cdot \sigma_U \tag{6.40}$$

$$\text{with} \quad \sigma_U \overset{\text{Eq. 4.21}}{\approx} \sqrt{\sigma_R^2 + \overline{N_{Int}} \cdot \sigma_S^2}$$

with

$\Delta x, \Delta y$	:	horizontal and vertical pixel pitch of the image sensor
f	:	focal length of the camera lens
T_{Pw}	:	laser pulse width used for image acquisition
x_s, y_s, d	:	coordinates of the centroid of the observed object in sensor coordinates
$\overline{U_b}, \overline{U_c}$	:	amplitudes of the two shutter signals averaged over all object pixels
$\overline{N_{Int}}$	:	analog integrations averaged over all object pixels
$\sigma_u, \sigma_v, \sigma_R, \sigma_S$	:	component uncertainties on pixel/system level

Basic task for the computation of $\boldsymbol{R}(k)$ is thus the determination of the standard deviations σ_u, σ_v, σ_R, and σ_S. This has to be done experimentally.
The standard deviations σ_R, σ_S, which describe the noise of the sensor signal amplitude, can be simply measured as described in 4.2.3.1.
The horizontal and vertical pixel uncertainty σ_u and σ_v are governed by spatial quantization noise due to the finite pixel dimensions and by sensor signal noise, which can induce flickering pixels, if the pixels sensor signal values are close to the noise threshold (cf. section 4.1.5). Both effects strongly depend on the measured objects properties (reflectivity, shape, pixel number, etc.), and thus cannot be determined definitely in advance like σ_R and σ_S. Furthermore, the horizontal and vertical errors of a object centroid determination do not fulfill the Kalman filter assumption of being normally distributed measurement noises, due to their quantized nature.
Nevertheless it is possible to find appropriate σ_u- and σ_v-values to guarantee a good filter functionality by means of filter tuning, which is described below.

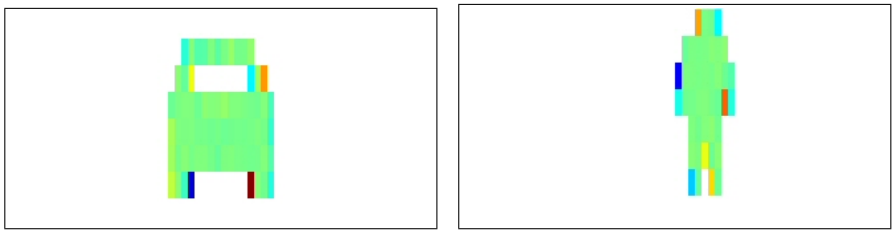
6.4.3.4 Filter Tuning

As described in 6.4.3.1, the basic process model of the Kalman filtering approach assumes uncorrelated and normally distributed noise sources that describe the uncertainty of the objects movement process and the uncertainty of the position measurement. For a proper functionality of the filter, knowledge of the description parameter is necessary. The easiest way to do this was to measure a set of test data from which the noise values can be calculated. As mentioned in 6.4.3.3 this is possible for the distance measurement noise σ_d as performed in section 4.2.3.1, but not for the horizontal and vertical measurement noises σ_u and σ_v or σ_ϕ and σ_θ, respectively. Further it is not possible to adjust the process noise descriptors σ_{w_x}, σ_{w_y}, and σ_{w_z} in advance without

any reference to the objects and scenarios to be acquired later. Another incompatibility to the Kalman filtering postulate is the temporal correlation of the process noise which is typical for movement process of real objects due to their maneuvering behavior and inertia properties [42].
Thus these parameters can only be adjusted empirically by measuring scenarios that are similar to the scenarios one expects in later measurement. By filtering these image sequences using different sets of filter parameters, and analyzing the Kalman tracking results with respect to the deviation from the real values that were measured, the sought-after parameter are obtained. This procedure is called *filter tuning*, and is subject of the following chapter.
Note that in the following all filter estimates are written without $\hat{}$, i.e. $\hat{x}(k) \rightarrow x(k)$. To prevent confusion with the real values, they are provided with the index r standing for 'real'.

Synthetic Camera Data

A principal problem in filter tuning is that for comparison of measured values with real values the exact knowledge of the real values is necessary. In the here considered case, one thus has to know the exact position, velocity, and acceleration data of the centroid of an observed object at each time step k. For traffic scenarios like walking persons or driving cars, this is actually not possible. Even with major technical efforts like differential GPS for position determination [45] and additional sensors for velocity and acceleration derivation, the required accuracy in measuring the real kinematic data can hardly be reached.
For this reason *synthetic camera sequences* are used here for filter tuning, i.e. a MATLAB© program was developed, which generates data files similar to the 3D-CMOS cameras files, containing predefined objects which are moving on programmable paths. After definition of objects and paths (i.e. all position, velocity, and acceleration parameters of the centroids of the objects at any time step are exactly known), the software calculates for each time step k the related virtual scene and the image of the scene on an imaginary sensor surface. With knowledge of all parameters describing the scene (illumination, object reflectivities, object distances, etc.) and the camera (sensor responsivity, sensor geometry, etc.) the according sensor signal amplitudes for all sensor pixels are computed; also adaptive accumulation is considered here. Adding noise to the sensor signals appropriate to the noise behavior of the system introduced in chapter 4.2.3.1, leads to sensor signals and thus distance data, which are supposed to have quite similar properties as real 3D-CMOS camera data. Figures 6.10(a) and 6.10(b) show two example frames of synthetic camera sequences.



(a) Example image frame of the synthetic 'breaking car' scenario

(b) Example image frame of the synthetic 'person crossing' scenario

Figure 6.10: Synthetically produced camera frames

To verify the usability of the synthetic camera sequences a performance comparison was conducted (see figure 6.11). A plywood target of 2 m × 0.7 m size was measured at different distances; after the segmentation of the images, the standard deviation of the calculated x-positions of theses measurements (red +-signs in the figure) were determined. The 'same' target was emulated as synthetic camera file and fed into the segmentation module. Again the standard deviation values of the x-position values after segmentation were calculculated, which are displayed as blue line. As one can see, the results are quite similar; the synthetic data can thus be accepted as modeling a real scene sufficiently realistic.

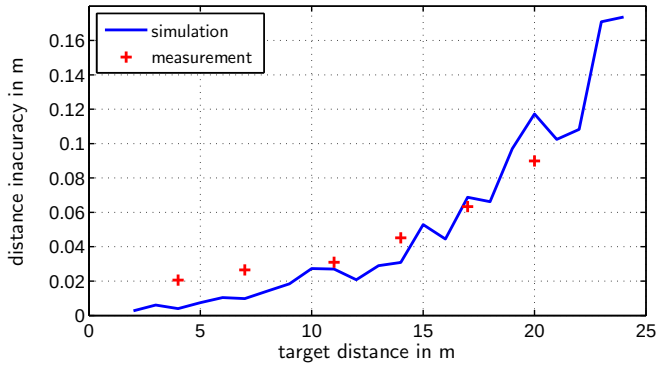


Figure 6.11: Comparison of the distance measurement inaccuracy of real and synthetic camera data

Filter Tuning Scenarios

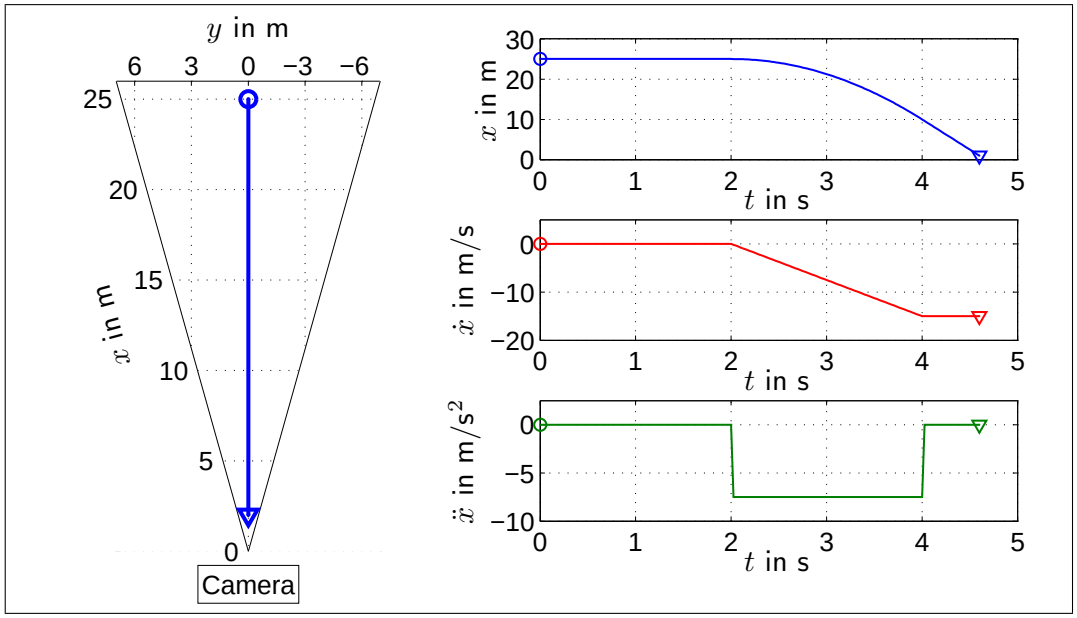
Figure 6.12 describes the two model scenarios that were used for Kalman filter tuning in the here described work. Two synthetic courses of objects moving in the VOV of the camera were used, firstly a car (as in fig. 6.10(a)) accelerating towards the camera, secondly a person (see 6.10(b)) crossing in front of the camera with a stop in the middle of the VOV. The sub-figures show the courses of the objects in bird's eye view on the left side, and the characteristic of the examined state components over the observed time steps on the right side.

Filter Tuning in x-Direction

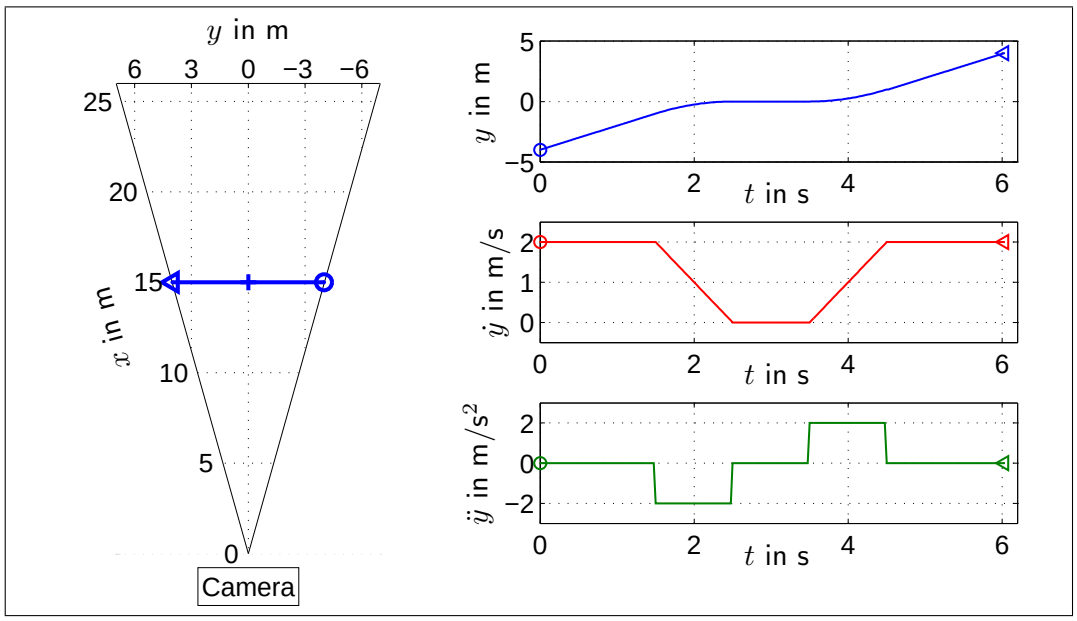
Starting point of the filter tuning procedure is the tuning by means of the observation of a object, centroid of which moves along the x-axis of the camera coordinate system, as it is defined in the 'approaching car' scenario in figure 6.12(a). This scenario was chosen, because mitigation of collisions with preceding cars is one of the main scenarios addressed in the UseRCams project (cf. chapter 1). Goal of this filter tuning step is the derivation of the optimal σ_{w_x} that minimizes the measurement error in x-direction. Therefore, the image sequence was fed into the image processing chain under variation of σ_{w_x} several times. See the resulting values for $x(k)$, $\dot{x}(k)$ and $\ddot{x}(k)$ for three selected values of σ_{w_x} and the according real values $x_r(k)$, $\dot{x}_r(k)$ and $\ddot{x}_r(k)$ (denoted with 'real' in the legends) in figure 6.13. The whole procedure was repeated 50 times for 50 realizations of the synthetic sequence (i.e. different noises for the different realizations lead to slightly different image data, while the 'real' scene remains the same), in order to be able to make statistical calculations of the behavior of the filter.
By this means, the residual errors $r_x(k)$, $r_{\dot{x}}(k)$, and $r_{\ddot{x}}(k)$ at the k-th time instance can be calculated as

$$r_x(k) = x_r(k) - E[x(k)] \tag{6.41}$$
$$r_{\dot{x}}(k) = \dot{x}_r(k) - E[\dot{x}(k)] \tag{6.42}$$
$$r_{\ddot{x}}(k) = \ddot{x}_r(k) - E[\ddot{x}(k)] \tag{6.43}$$

The standard deviations $\sigma_x(k)$, $\sigma_{\dot{x}}(k)$, and $\sigma_{\ddot{x}}(k)$ are calculated from the 50 iterations of the sequence as well; figure 6.14 shows the results for the example scenario.
From the above values the second order moments – also well known as *mean square error (MSE)* – of the Kalman filter outputs at each time step (k) can be calculated according to



(a) Scenario for the synthetic data of a car accelerating towards the camera (analogous to a preceding car breaking in a traffic scenario)



(b) Scenario for the synthetic data of a person crossing at 15 m in front of the camera from the right, stopping in the middle, and then leaving the cameras VOV on the left side

Figure 6.12: Synthetically produced example scenarios

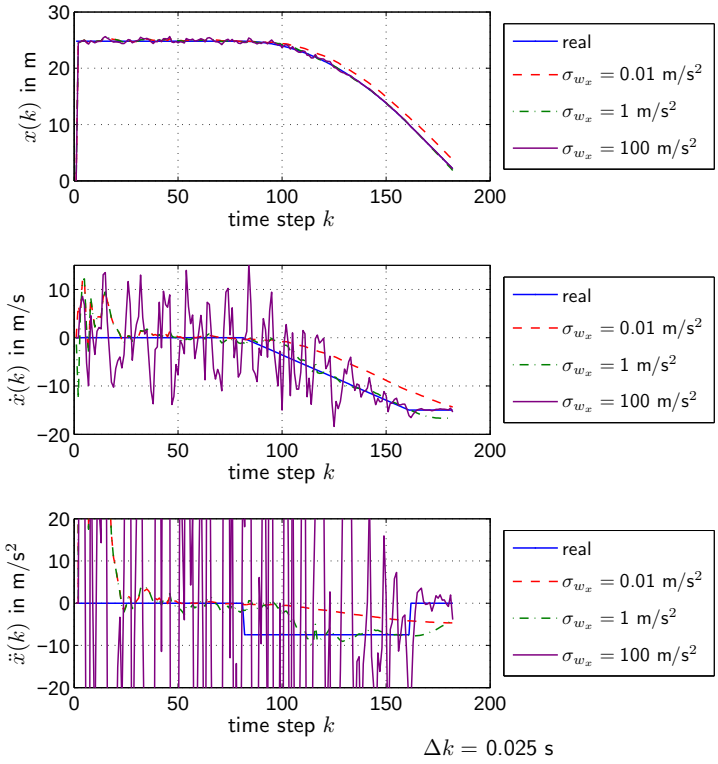


Figure 6.13: Tracking results in x-direction for the synthetic car breaking scenario for different values of σ_{w_x}

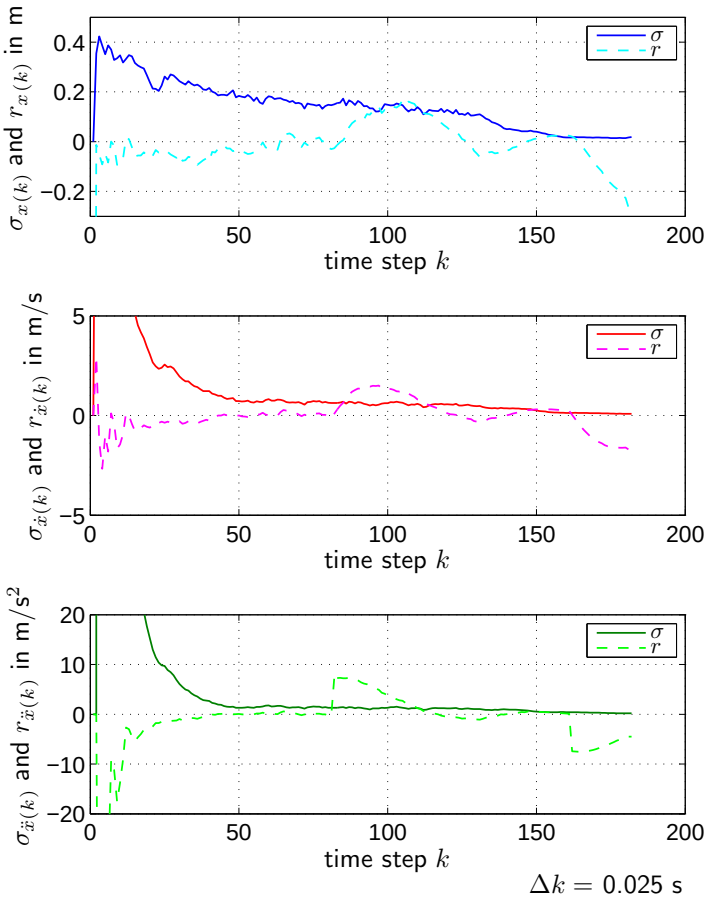


Figure 6.14: Residual error and standard deviation of the tracking results of figure 6.13

$$MSE_x(k) = \sigma_x(k)^2 + r_x(k)^2 \text{ , } MSE_{\dot{x}}(k) \text{ and } MSE_{\ddot{x}}(k) \text{ accordingly} \quad (6.44)$$

Next step is the computation of an *average MSE (AMSE)*, of the MSEs of $x(k)$, $\dot{x}(k)$ and $\ddot{x}(k)$ over a observation period $k_{min} \leq k \leq k_{max}$. In the example scene the period $40 \leq k \leq 160$ was chosen to cut off filter transient effects at the beginning of the sequence.

$$AMSE_x(\sigma_{w_x})\Big|_{\substack{k\leq k_{max}\\ k\geq k_{min}}} = E[MSE_x(k)]\Big|_{\substack{k\leq k_{max}\\ k\geq k_{min}}} \quad (6.45)$$

$$= \frac{1}{(k_{max} - k_{min}) + 1} \cdot \sum_{k=k_{min}}^{k_{max}} MSE_x(k) \quad (6.46)$$

$$AMSE_{\dot{x}}(\sigma_{w_x}) \text{ and } AMSE_{\ddot{x}}(\sigma_{w_x}) \text{ accordingly}$$

Figure 6.15 depicts the $AMSE$s of $x(k)$, $\dot{x}(k)$ and $\ddot{x}(k)$ in dependence of the adjusted σ_{w_x}. Five values – namely 0.01 m/s^2, 0.1 m/s^2, 1 m/s^2, 10 m/s^2, and 100 m/s^2 – were taken. In order to get the optimal filter parameter $\check{\sigma}_{w_x}$ minimization over $AMSE_x(\sigma_{w_x})$ is now performed:

$$\check{\sigma}_{w_x} = \min(AMSE_x(\sigma_{w_x})) \quad (6.47)$$

Analogously $\check{\sigma}_{w_{\dot{x}}}$ and $\check{\sigma}_{w_{\ddot{x}}}$ are determined by minimizing over all $AMSE_{\dot{x}}(\sigma_{w_x})$ and $AMSE_{\ddot{x}}(\sigma_{w_x})$. The +-signs in figure 6.15 indicate the according minima. In case that the results for $x(k)$, $\dot{x}(k)$, and $\ddot{x}(k)$ do not coincide, a trade-off has to be found for the setting of the process noise parameter.
Here the results coincide for all three measures, a σ_{w_x} of 1 m/s^2 minimizes the $AMSE$s of x, $\dot{x}$, and $\ddot{x}$.
As the movement behavior of objects in world coordinates is normally the same in x- and y-direction, σ_{w_y} is also set to 1 m/s^2 for the following considerations. The movement in z-direction is actually of no interest in the here addressed applications. For reasons of completeness z-movement is also filtered; therefore, σ_{w_z} is also set to 1 m/s^2.

Filter Tuning in y-Direction

Similar to filter tuning in x-direction, the filter must be tuned in y-direction. But the process noise and the measurement noise have changed their roles here. When tuning in x-direction the process noise σ_{w_x} is tuned while the measurement noise is fixed (measurement noise is known due to chapter 6.4.3.3), here

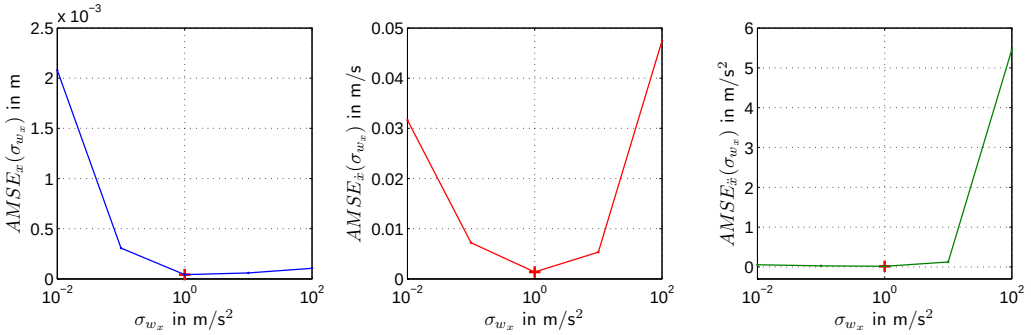


Figure 6.15: Determination of the optimal process noise value by minimizing the average normalized error of x, $\dot{x}$, and $\ddot{x}$

the measurement noise adjustment is tuned for fixed process noise σ_{w_y}, which was already set after the determination of σ_{w_x} above. This approach is necessary as the measurement noise in y-direction cannot be determined adequately by offline measurements. Also its properties to be non-Gaussian-distributed and correlated (cf. chapter 6.4.3.3) make a determination of the optimal value by filter tuning necessary. Again the procedure ends in three curves (figure 6.16) whose minima assign the optimal adjustment of the horizontal measurement noise σ_v with respect to the errors in the estimation of y, $\dot{y}$, and $\ddot{y}$.
Here the minima do not coincide; y is minimal for $\sigma_u = 5 \cdot 10^{-2}$ (pixel), $\dot{y}$ and $\ddot{y}$ for $\sigma_u = 1.5 \cdot 10^{-2}$. Though σ_u was set to $1.5 \cdot 10^{-2}$ as the error of y is only slightly higher for $\sigma_u = 1.5 \cdot 10^{-2}$ compared to $\sigma_u = 5 \cdot 10^{-2}$.

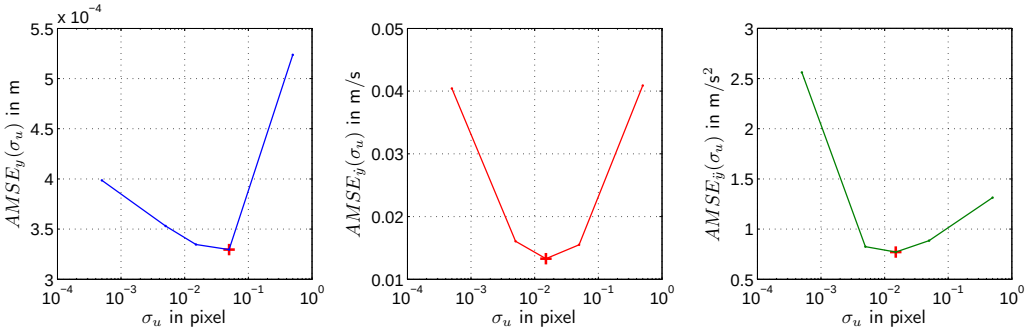
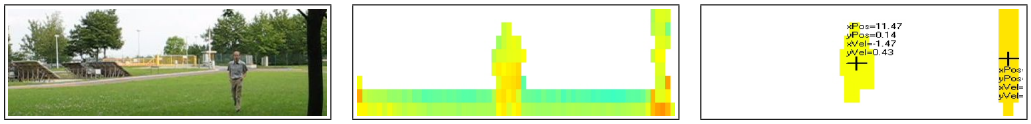


Figure 6.16: Determination of the optimal measurement noise value in y-direction by minimizing the average error of y, $\dot{y}$ and $\ddot{y}$

As it was presented here by means of the 'approaching car' and the 'crossing pedestrian' scenarios, the filter tuning procedure can be similarly executed with any other test scenario data, what would perhaps lead to similar or even better tuning results. It is only important to use tuning data, that resemble the later application area of the system best possible.

6.4.3.5 Measurement Results

After the filter tuning procedure test measurements with real scenarios were carried out in order to test the performance of the overall system including the camera and the image processing software. An example is given in the following. The scene is shown in figure 6.17(a); a person is walking in front of the camera, a tree is located on the right border of the cameras VOV. Figures 6.17(b) 6.17(c) display the camera distance image and the segmented distance image of an exemplary frame of the image sequence in color coded representation.



(a) Photograph of the example scenario (b) Single camera frame of the example scenario (c) 6.17(b) segmented

Figure 6.17: Real example scenario

The course of the walking person and the tracker outputs are given in figure 6.18. The left graph displays the projection of the x-y-positions of the segmented objects over the whole scene. Starting point of the object path is the end of the path marked with the path number (here 3 in the case of the pedestrian). First the person approaches the camera from the right, where it stands still for about 1 s at a distance of 11 m; then it moves away from the camera with another stop and a change of direction at a distance of 18 m until it is not visible anymore to the camera at distances over 23 m.
The three diagrams on the right side of figure 6.18 show the Kalman filter estimates for the pedestrian's position, velocity, and acceleration in x- and y-direction. The estimates of the z-direction are not plotted, as they are only of low importance in the here considered scenario. In general the quality of the results in z-direction is comparable to the quality of the results in y-direction.

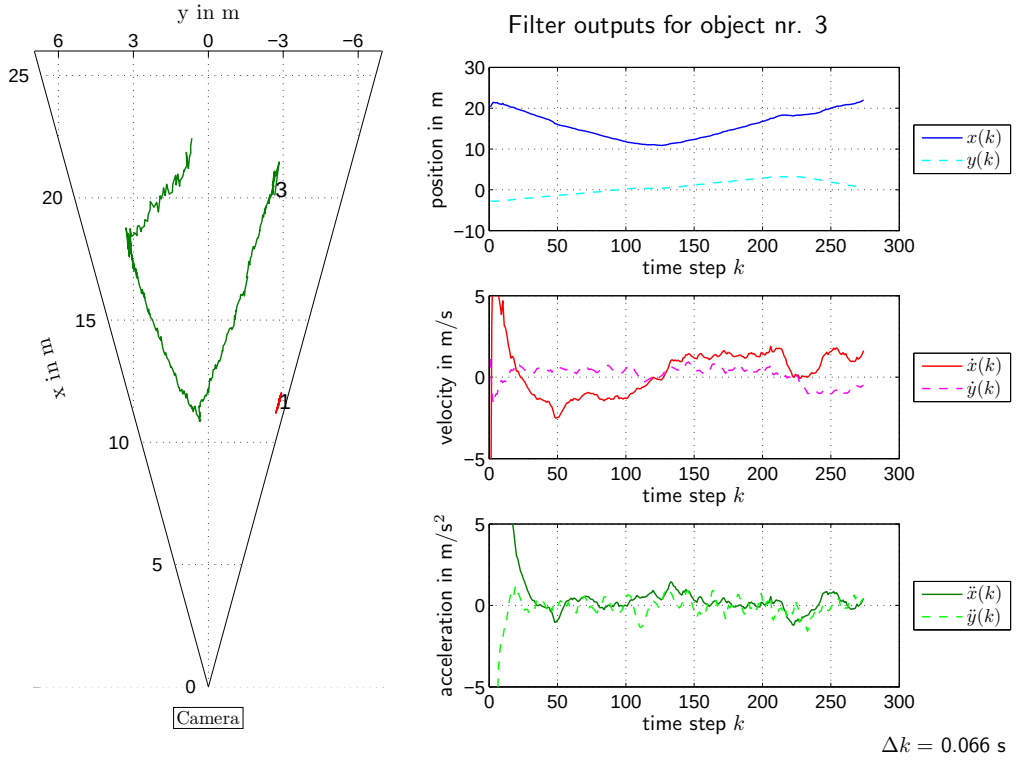


Figure 6.18: Results of the tracking of a moving person

As already remarked in section 6.4.3.4, quantitative evaluation of the measurement results of real scenes could not be performed in this work. Reason is the fact the the real values of the state vector $\boldsymbol{x}$ cannot be determined exactly for all time steps of the examined image sequence. Only a qualitative evaluation by inspection of the filter outputs is possible:
It can be stated that the position estimation works very good. Also the velocity estimation gives reasonable values; noise is superposed to the estimates, but the course of the real velocity is estimated well for both, the x- and the y-direction. The acceleration estimates yield even more noise, but at least for the x-direction the course of the real acceleration can be reconstructed quite reliably from the filter estimates. The second stop of the walking person at time step 220 and its moving on at time step 240 become clearly visible in a deflection of the $\ddot{x}(k)$-curve in negative resp. positive direction at the corresponding time steps.

One reason for the noise of the $\dot{x}(k)$- and $\ddot{x}(k)$-estimates, is the rather low frame repetition rate of 15 fps at the current prototype system. An improved system yielding image sequences with 50 fps will significantly improve the results.

6.5 Object Classification and Action Recognition

Based on the results of object segmentation and object tracking, object classification and action prediction is performed.

Goal of the classification is in the case of the road safety application the differentiation into following object classes:

- Vulnerable road users (VRUs): pedestrian, cyclist, baby buggy, ...
- Vehicles: car, truck, ...
- Obstacles: tree, wall, pole, ...

At the present state of the image processing simple evaluation algorithms were implemented that classify the objects according to their size, shape, and kinematical properties. A moving object with a width of 0.4 m for example can hardly be an obstacle, which is assumed to have a fixed position, and also no vehicle due to its narrow horizontal dimension. Thus it is correctly classified as VRU.
Of course this classification procedure yields only basic results and is quite susceptible to errors (a person standing without movement is not distinguishable from a tree trunk or a concrete column), but more complex classification algorithms with more object classes seem inapplicable at the present state of the system development. Increasing the camera resolution or combined image processing including high resolution 2D-images of the observed scene could considerably improve the object classification capabilities.

Further prediction of future incidents is tried by means of extrapolation of the image processing results. Main objectives are a reliable decision if there is the risk of a collision with one of the objects observed by the camera and – if such a risk situation emerges – the evaluation of the time to collision, which is a very important measure in view of the actions to be conducted for collision avoidance or mitigation. Starting with the x-distance $d_{OC}(k,t)$ between an

object O and the camera C at time t, where t starts with 0 at the k-th time step,

$$d_{OC}(k,t) = (x_O(k) - x_C(k)) - t \cdot (\dot{x}_O(k) - \dot{x}_C(k)) - \frac{t^2}{2} \cdot (\ddot{x}_O(k) - \ddot{x}_C(k)) \ , \quad (6.48)$$

the time when the distance in x-direction becomes zero is

$$t_0 = \frac{(\dot{x}_O(k) - \dot{x}_C(k)) - \sqrt{(\dot{x}_O(k) - \dot{x}_C(k))^2 - 2 \cdot (\ddot{x}_O(k) - \ddot{x}_C(k)) \cdot (x_0(k) - x_C(k))}}{(\ddot{x}_O(k) - \ddot{x}_C(k))} \ . \quad (6.49)$$

With this t_0 one can check, if at this or close to this time the y-distance $(y_O - y_C)$ is also zero or close to zero. If this is the case, a collision between object O and the camera C – or probably with the vehicle the camera is mounted in – is predicted. The time remaining until to the collision is then

$$t_{TTC} = t_0 \ ,$$

which is basis for the choice of mitigation actions to be performed in future applications.

6.6 Conclusion

As the measurement results of chapters 4 and 6 show, the 3D-CMOS array camera system in combination with the image processing software fulfills the application specific requirements of the UseRCams project (see chapter 1). Thus the camera development part of the project, which was our task at Siemens CT PS 9, was successfully accomplished with promising results in view of future applications in the automotive and the safety sector (see chapter 7 for some application examples).

A deeper analysis of the performance of the camera system is presently executed at the UseRCams project partners of the automotive industry, including BMW, FIAT, RENAULT, and VOLVO. They integrated the camera in test vehicles to test the system in real traffic scenarios. Results of these evaluations will soon be published on [1].

Further research for camera enhancement, especially by combination of the 3D image with a high-resolution 2D camera image, is presently done in the research project MIDIAS [46].

Chapter 7

Some Examples of Applications

The combination of direct distance measurement capability with considerable horizontal and vertical resolution represents a great advantage of the 3D-CMOS sensor over most of the established imaging or ranging techniques. Radar sensor data, for instance, provide good distance information with respect to individual or very few well-separated objects, but poor or no lateral resolution. Video data, on the other hand, feature excellent lateral resolution, but no direct depth information.
This property offers a variety of areas of applications for the 3D-CMOS technology. The following sections give an overview over the most promising approaches.

7.1 Road Safety (UseRCams)

A main application area for the 3D-CMOS camera system in connection with the 3D real-time image processing proposed in chapter 6 are the road safety applications introduced in chapter 1. For the sake of completeness they are shortly mentioned here again. The range of application as sensor for advanced driver assistance systems for vehicles comprises (see scenarios depicted in figure 7.1):

- Front looking applications for pedestrian protection and mitigation of collisions with preceeding cars
- Utilisation of a camera for recognition and mitigation of side crash situations with static objects or other vehicles
- Surveillance of the blind areas of trucks during stops for reliable recognition of objects located in this areas combined with a start inhibit function

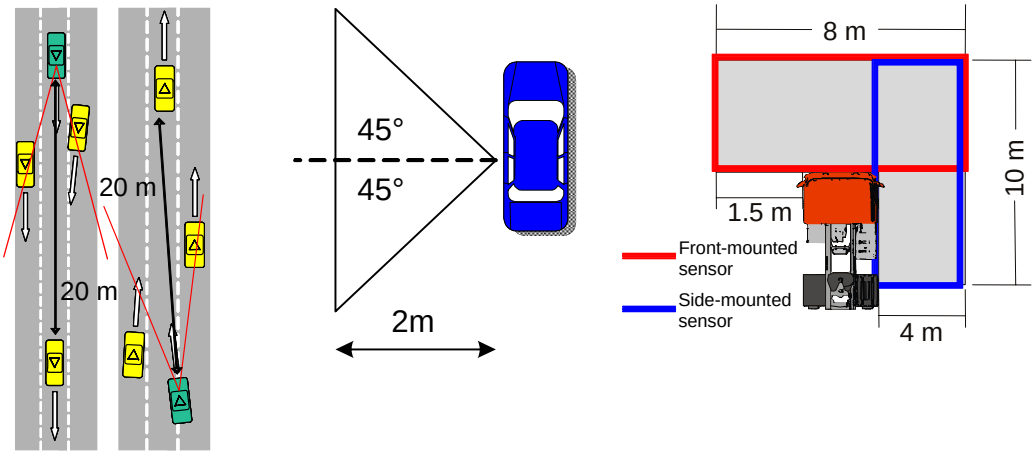


Figure 7.1: Sketches of potential collision mitigation, pre-crash, and blind spot surveillance scenarios

At the present state of development, yielding a measurement range up to 20 m, urban traffic scenarios are addressed up to a vehicle velocity of 15 m/s.
A vital component in these applications is the 3D real-time image processing software (chapter 6), which excerpts the information required by the applications from the raw data sequences of the 3D-CMOS camera.

7.2 People Counting

Another interesting application of the 3D-CMOS array camera is the surveillance of entrance areas, doors, etc. with additional counting of the people passing by or through (figure 7.2), which cannot be realized with conventional light barriers.
The people counting procedure is based on reliable object segmentation as well as object tracing over image sequences, which was successfully implemented as described earlier in this work (see chapter 6). Once the trace of an object is tracked for the whole movement from one side (e.g. line 1) to the other side of the sensor (line 8) it can be decided by means of a simple algorithm, if the according object was entering or leaving the building, room or whatever.
In addition to the counting functionality further 'intelligence' can be implemented based on the 3D data. A height discrimination of the detected objects can be introduced, to prevent the system from counting small animals or the like; similarly the distance range close the camera can be gated, i.e. disturbing influences of insects or birds can be avoided. Such sophisticated data evalua-

tion possibilities are not available when using conventional 2D-camera or light barrier systems.

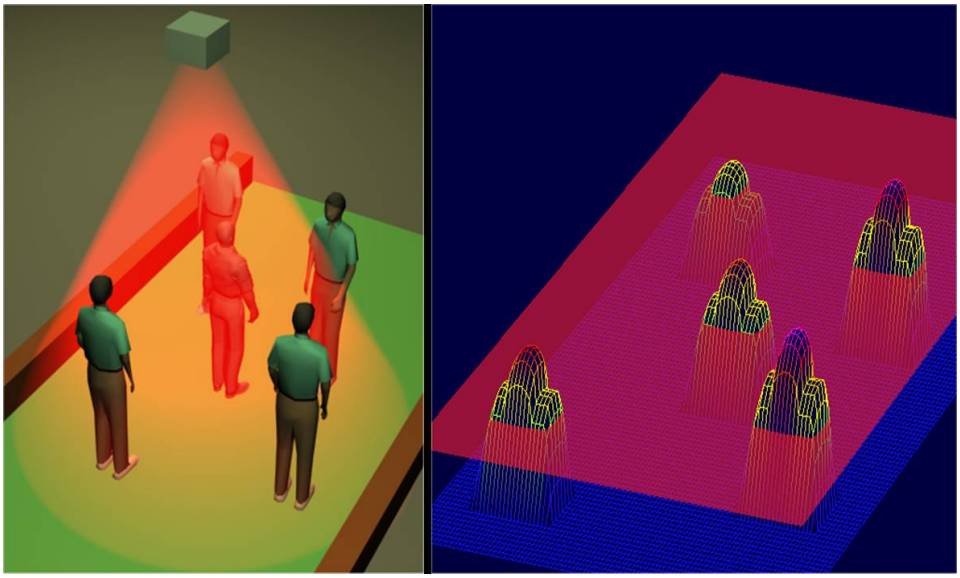


Figure 7.2: People counting application

7.3 Orientation Aid for Blind People

In the CASBliP project [47], which aims at the development of a system capable of interpreting and managing real world information from different sources to assist blind or visually impaired users, a portable 3D-CMOS sensor system in the shape of eyeglasses is being presently developed. In combination with a space/audio processing unit which maps the distance information to special sounds [48] the sensor should act as orientation aid for blind people (see figure 7.3).
A 64x1 pixel 3D-CMOS line sensor (one pixel line of a 64x4 pixel sensor [22] is used) is mounted together with the laser illumination unit and a circuit board (including power supply, sensor control and raw data processing) in a eyeglasses-like housing and serves as device for the perception of the surrounding. Sensor output of a single aquisition phase is a distance profile along the observed line. Distance range of the system is from 0.5 m to 5 m with a repetition rate of 25 fps. Due to the position of the sensor system on the users head, the environment is intuitively scanned by the head movements. The gained distance data are then transferred to a main processing unit, where

advanced data processing is accomplished, including transformation of the 3D-information into acoustic maps for a real world environment representation to the blind person through earplugs. Stereo effects for left/right encoding as well as amplitude and frequency variations of the sound signals are combined according to a complex algorithm; care is taken that the sounds do not restrict the hearing of the 'normal' soundscape, which is the main information source for blind people.
A prototype system was successfully assembled. First application tests conducted by a blind test person yielded promising results.

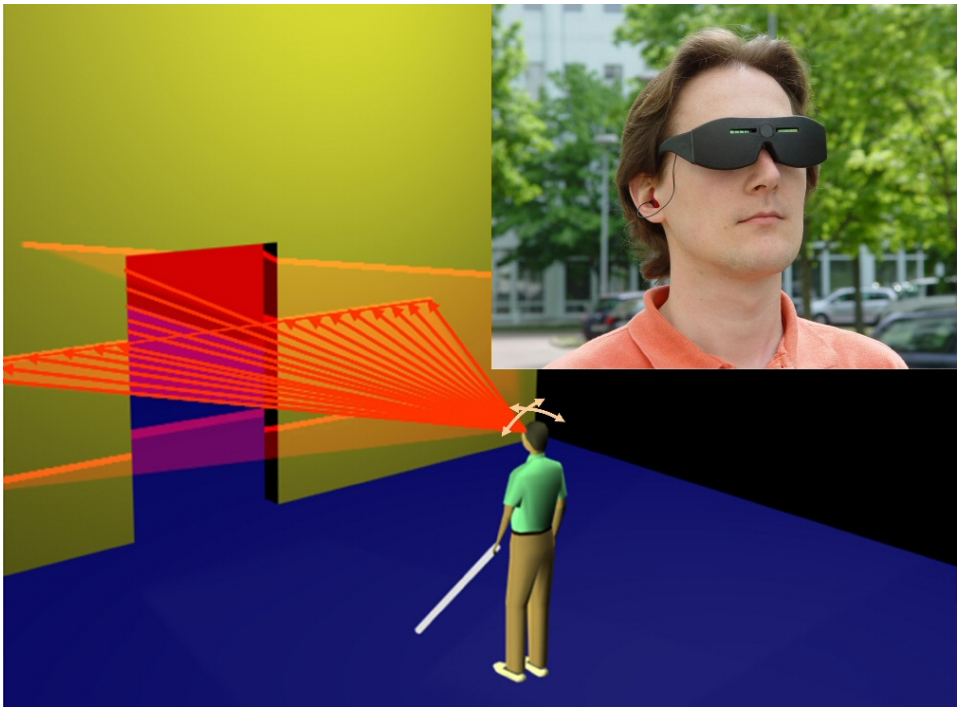


Figure 7.3: Cognitive aid for blind people by 3D-CMOS sensor

7.4 Intelligent Light Curtain

Due to its special properties the 3D-CMOS sensor system is well suited for safety and security applications; especially the distance measuring capabilities paired with wide independence of ambient illumination and surface conditions of the objects are the advantages compared to present systems as light barriers or 2D cameras.

Prototypes were assembled to demonstrate the capabilities of the systems as *reflector-less intelligent light curtain*. Special feature is here the structured illumination that allows to cover a defined area with a light curtain (see figure 7.4). Together with special compression optics, which images the illuminated shape on a 64x1 pixel line sensor, curved light curtains can be realized. In this way arbitrary surveillance fields can be realized.
The distance information can here be used for definition of active and not active regions of the light curtain.

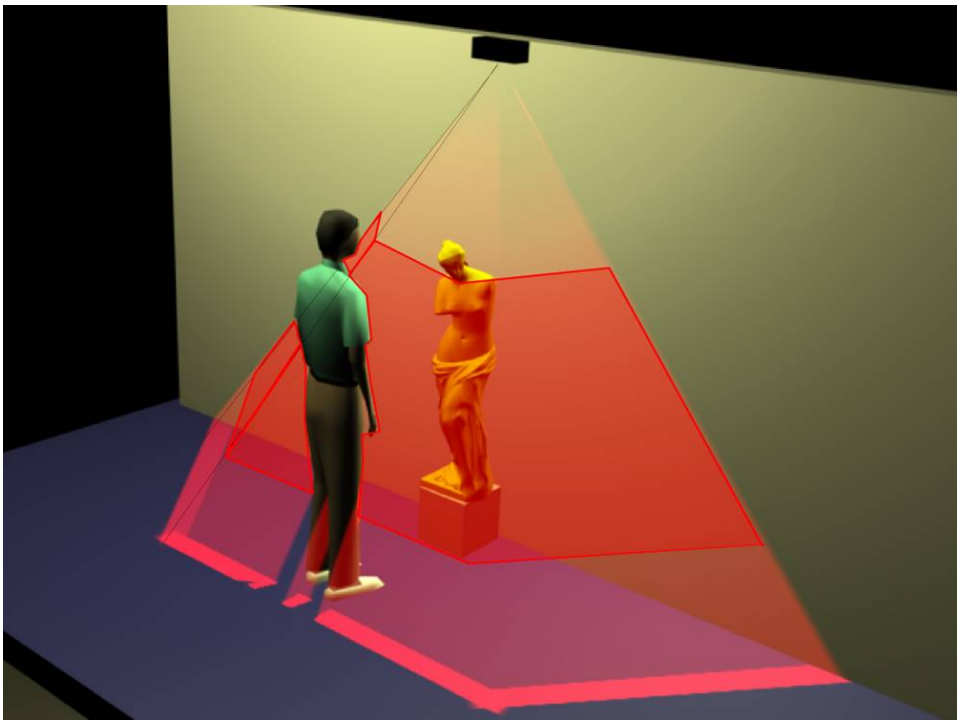


Figure 7.4: Intelligent light curtain

Many different applications can be envisioned for the intelligent light curtain. Some of them are:

- Object protection in museums (glassless show case)
- Area surveillance with possibility of defining active and non-active areas; e.g. a wall with doors and windows
- Intelligent light barrier for distances up to 100 m for small opening angles (invisible light fence), with exact localization of intrusion events by means of the distance information

Chapter 8

Conclusions and Further Work

This thesis is divided into two major parts. The first deals with the description and evaluation of the newly developed 3D-CMOS camera system, the second addresses the image processing of the data acquired with the aforesaid camera. The following paragraphs give a conclusion of the results of the research done in these areas:

A novel 64x8 pixel range measurement system was successfully developed and assembled. The system evaluation yielded good results for the basic system parameters. A noise equivalent exposure (NEE) of the sensor of 145 nJ/m^2 is measured, the NEE of the over-all system results as 225 nJ/m^2. Significant improvement of the 64x8 pixel image sensor compared to the previous image sensor generation is discovered in view of the rectangular shape of the shutter window function, which is crucial for good distance measurement results.
Further we conclude that the perfomance of the developed camera system agrees with the specified requirements (see chapter 1). Measurements show good distance measurement linearity of the 3D-CMOS system and a distance accuracy lower than 3% of the measured distance over a distance range from 2 m to over 20 m, for objects with surface reflectivities varying between 5% and 100% (Lambertian reflection). This means a dynamic range of the camera system in terms of radiant exposure of the sensor of over 96.5 dB.

The optimization of the distance measuring procedure showed that MDSI3, which was not applicable with the previous image sensor due to afore mentioned insufficient shutter window properties, performs best with regard to measurement linearity, measurement accuracy, and measurement time. This result was verified by calculations, simulations and experiments. Thus MDSI3 has been taken as standard distance derivation algorithm for the camera system.

A further result of this work is a measurement technique that makes a gating of the adverse signals, which come from raindrops or snowflakes in the close range, possible; such a feature is indispensible for outdoor applications. Drawback is increased measurement noise due to the modified measurement method. Measurement results show the enhancements compared to a measurement without gating, but also reveal the shortcomings arising from the non-ideal shutter window functions and laser-pulse shapes.

A main outcome of this thesis is the novel real-time image processing software, specifically developed for the image sequences of the 3D-CMOS camera system. It includes the segmentation, the tracking, and the classification of objects in the cameras volume of view (VOV).
Highly robust object segmentation results as compared to the segmentation of 2D images could be achieved. The procedure is based on a suitable region growing algorithm utilizing the distance information provided by the range camera.
The newly developed tracking module showed preeminent performance as well. The object tracking is here accomplished by adaptive Kalman filtering, customized to the output data properties of the 3D CMOS sensor system with real-time adaption to the scene properties. The filter tuning approach and the good results of estimation of object positions, velocities and even accelerations are demonstrated by means of example scenes.
Based on the results of object segmentation and object tracking, object classification and action prediction is performed. The basic concepts are described in this work.

Finally we can conclude – as the measurement results in chapters 4 and 6 show – that the 3D-CMOS array camera system in combination with the image processing software fulfills the application specific requirements. Thus the system development was successfully accomplished with promising results in view of future applications in the automotive and the safety sector.
A deeper analysis of the performance of the camera system is presently executed at the UseRCams [1] project partners of the automotive industry, including BMW, FIAT, RENAULT, and VOLVO. They integrated the camera in test vehicles to evaluate the system in real traffic scenarios.

The work of this thesis could be extended by following further investigations:

Further investigations should address the crosstalk problem described in chapter 4.2.1.4. This could improve the 3D imaging capabilities of the system, especially for scenes with large reflectivity and/or distance dynamics.

Efforts should also be made to further decrease the noise and/or increase the responsivity of the sensor. This could help to save laser diodes and reduce their power, which presently are the most costly components of the camera system, or to increase the systems VOV.

The sensor could be improved with respect to alternating shutters during analog accumulation instead of the sequential shutters to reduce corruption of measurement values on the edges of moving objects.

Efforts should be made to increase the pulse repetition rate of the laser light source. This would increase the frame repetition rate and decrease the aforesaid edge effects on moving objects, and thus significantly improve the image acquisition and image processing results.

The object recovery algorithm could be enhanced by shape or size comparison of the objects in addition to the presently considered Mahalanobis distance criterion. Also the number of previous frames used for object recovery should be increased from a single frame to make backtracking of object merging and splitting possible.

The tracking procedure could be improved by implementing

- autocorrelated process noise to consider the high-grade temporal correlation of the process noise in reality
- adjustable level process noise that could adapt the process noise level to the behavior of the observed object
- multiple model tracking with model switching to adapt the process model and the state vector dimensions to the properties of the observed object.

These improvement approaches are closely described in [42].

For improvement of the object classification capabilities of the 3D imaging system an increase of the sensor resolution and/or a combined image processing including high resolution 2D-images of the observed scene should be addressed. Efforts are already undertaken in the research project MIDIAS [46].

Appendix A

Basic Signals

A.1 The Rectangular Impulse

The standardized rectangular impulse is defined as

$$rect(t) = \begin{cases} 1 & ,\mid t \mid \leq \frac{1}{2} \\ & \\ 0 & ,\mid t \mid > \frac{1}{2} \end{cases} \qquad \text{for} \quad t \in \mathbb{R}. \tag{A.1}$$

Scaling by T and shifting by τ in direction of the t-axis leads to the general description of a rectangular signal:

$$rect(t, T, \tau) = rect\left(\frac{t-\tau}{T}\right) = \begin{cases} 1 & ,\mid \frac{t-\tau}{T} \mid \leq \frac{1}{2} \\ & \\ 0 & ,\mid \frac{t-\tau}{T} \mid > \frac{1}{2} \end{cases} \tag{A.2}$$

A.2 The Dirac Impulse (Dirac δ-function)

The Dirac δ-function is defined by

$$\delta(t) = \begin{cases} \infty & ,t = 0 \\ & \\ 0 & ,t \neq 0 \end{cases} \qquad \text{mit} \quad t \in \mathbb{R} \tag{A.3}$$

and

$$\int_{-\infty}^{\infty} \delta(t)dt = 1 \quad . \tag{A.4}$$

It is appropriate to describe the sampling of signals, as the integral over the product of a function $f(t)$ with $\delta(t-\tau)$ yields the sampled value $f(\tau)$ of $f(t)$:

$$f(\tau) = \int_{-\infty}^{\infty} f(t) \cdot \delta(t-\tau)dt \tag{A.5}$$

Appendix B

Random Variables

B.1 Probability Density Function (PDF)

For derivation of the pdf of a discrete random variable x, which can take the values in the set $(\xi_i, i = 1, ..., n)$, firstly the *probability mass function (pmf)* is introduced:

$$\mu_x(\xi_i) = P(x = \xi_i) \qquad i = 1, ..., n \tag{B.1}$$

with μ_i beeing the *point masses*.
By means of the Dirac impulse function (cf. chapter A.2) the discrete pdf results as

$$p_x(\xi_i) = \sum_{i=1}^{N} \mu_i \cdot \delta(x - \xi_i) \ . \tag{B.2}$$

B.2 Gaussian Distribution

A continuously distributed random variable x is called Gaussian distributed [25], if its values are distributed according to equation B.3:

$$f_x(\xi) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(\xi-\mu)^2}{2\sigma^2}} \tag{B.3}$$

Its probability density function $pdf(x)$ looks like shown in figure B.1. The two parameter μ and σ define the shape of the resulting distribution. Thereby is μ the offset of the function in ξ-direction; σ is a spread factor of the function with μ beeing the center of the spread.
A short notation equivalent to B.3 is

$$pdf(x) \sim (\mu, \sigma^2)$$

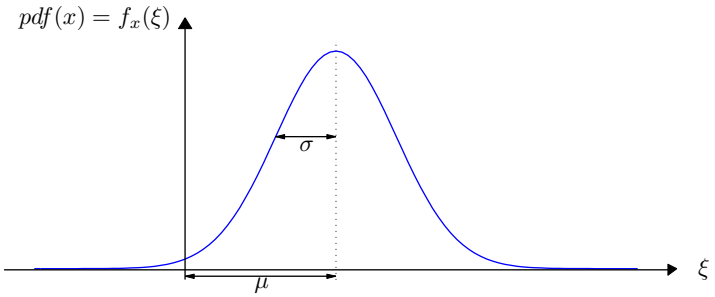


Figure B.1: Gaussian distributed probability density function

For a Gaussian distribution $f_x(\xi)$ of a random variable x one can keep in mind table B.1 which displays the portions of all drawn values x lying within an interval $\mu - n \cdot \sigma \leq x \leq \mu + n \cdot \sigma$ for different values of n.

Table B.1: Portion of x-values contained in the $\mu \pm n \cdot \sigma$-interval

Interval	Portion
$\mu - 1 \cdot \sigma \leq x \leq \mu + 1 \cdot \sigma$	0.683
$\mu - 2 \cdot \sigma \leq x \leq \mu + 2 \cdot \sigma$	0.954
$\mu - 3 \cdot \sigma \leq x \leq \mu + 3 \cdot \sigma$	0.997

These numbers count for large numbers of draws, and must be seen as approximate values for small numbers.

Appendix C

Line Fitting

For a given set of N observations consisting of the x_i- and y_i-value vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, the line

$$y_f = m \cdot x + b \tag{C.1}$$

with

$$m = \frac{\overline{\boldsymbol{x} \cdot \boldsymbol{y}} - \overline{\boldsymbol{x}} \cdot \overline{\boldsymbol{y}}}{\overline{\boldsymbol{x}^2} - (\overline{\boldsymbol{x}})^2} \tag{C.2}$$

and

$$b = \frac{\overline{\boldsymbol{x}^2} \cdot \overline{\boldsymbol{y}} - \overline{\boldsymbol{x}} \cdot \overline{\boldsymbol{x} \cdot \boldsymbol{y}}}{\overline{\boldsymbol{x}^2} - (\overline{\boldsymbol{x}})^2} \tag{C.3}$$

minimizes the quadratic error sum

$$\sum_{i=1}^{N} (y_i - (m \cdot x_i + b))^2 \quad . \tag{C.4}$$

The intersection of the fitted line with the x-axis

$$x|_{y_f=0} = x_0 = \frac{-b}{m} \tag{C.5}$$

results from the above derived parameters as

$$x_0 = \frac{\overline{\boldsymbol{x}} \cdot \overline{\boldsymbol{x} \cdot \boldsymbol{y}} - \overline{\boldsymbol{x}^2} \cdot \overline{\boldsymbol{y}}}{\overline{\boldsymbol{x} \cdot \boldsymbol{y}} - \overline{\boldsymbol{x}} \cdot \overline{\boldsymbol{y}}} \quad . \tag{C.6}$$

This can also be expressed as

$$x_0 = \frac{-\sum_{i=1}^{N-1} \left[\sum_{j=i+1}^{N} \left(\left(y_i - \frac{(y_i - y_j)}{(x_i - x_j)} \cdot x_j \right) \cdot (i-j)^2 \right) \right]}{\sum_{i=1}^{N-1} \left[\sum_{j=i+1}^{N} \left(\frac{(y_i - y_j)}{(x_i - x_j)} \cdot (i-j)^2 \right) \right]} \tag{C.7}$$

which is actually a weighted mean of all possible combinations of two data points, and thus a weighted mean of all possible MDSI3 (see chapter 3.4) distance calculations.

For accuracy calculation of the DCSI distance measuring method in chapter 5.1.1 the partial derivative $\frac{\partial x_0}{\partial y_i}$ of x_0 is required. It is computed according to following equation:

$$\frac{\partial x_0}{\partial y_i} = \frac{(\overline{\boldsymbol{x}\cdot\boldsymbol{y}} - \overline{\boldsymbol{x}}\cdot\overline{\boldsymbol{y}})\cdot(\overline{\boldsymbol{x}}\cdot\frac{x_i}{N} - \overline{\boldsymbol{x}^2}\cdot\frac{1}{N}) - (\overline{\boldsymbol{x}}\cdot\overline{\boldsymbol{x}\cdot\boldsymbol{y}} - \overline{\boldsymbol{x}^2\cdot\boldsymbol{y}})\cdot(\overline{\boldsymbol{x}}\cdot\frac{x_i}{N} - \overline{\boldsymbol{x}}\cdot\frac{1}{N})}{(\overline{\boldsymbol{x}\cdot\boldsymbol{y}} - \overline{\boldsymbol{x}}\cdot\overline{\boldsymbol{y}})^2} \tag{C.8}$$

Bibliography

[1] Ertico, "PReVENT project homepage," http://www.prevent-ip.org/, status: 06/01/2007.

[2] P. Mengel and R. Wertheimer et al., "D52.300.1 camera requirements and specifications," UseRCams project deliverable, available from http://www.prevent-ip.org/en/public_documents/deliverables/d523001_usercams_general_specifications.htm, status: 06/01/2007.

[3] Rudolf Millner, *Ultraschalltechnik: Grundlagen und Anwendungen*, Physik-Verlag, Weinheim, Germany, 1987.

[4] Erich Pehl, *Mikrowellen in der Anwendung*, Hüthig, Heidelberg, Germany, 1993.

[5] Bernd Jähne, *Digital Image Processing*, Springer, Berlin, Germany, 2005.

[6] F. Forster, *Real-Time Range Imaging for Human-Machine Interface*, Ph.D. thesis, Technische Universität München, 2005.

[7] D. Modrow et al., "3D face scanning systems based on invisible infrared coded light," in *Proceedings ISVC07*, Lake Tahoe, USA, Nov. 2007.

[8] S. Hsu et al., "Performance of a time-of-flight range camera for intelligent vehicle safety applications," in *Advanced Microsystems for Automotive Applicatioons 2006*, J. Valldorf and W. Gessner, Eds., pp. 205–219. Springer, Berlin, Germany, 2006.

[9] R. Lange, *3D Time-of-Flight Distance Measurement with Custom Solid-State Image Sensors in CMOS/CCD-Technology*, Ph.D. thesis, Universität Siegen, 2000.

[10] T. Oggier et al., "3D-imaging in real-time with miniturized optical range camera," in *Proceedings Opto-Conference*, Nürnberg, Germany, 2004.

[11] T. Oggier et al., "Novel pixel architecture with inherent background suppression for 3D time-of-flight imaging," in *Proceedings SPIE-IS&T*, San Jose, USA, 2005, vol. 5665.

[12] C. Niclass et al., "A single photon detector array with 64x64 resolution and millimetric depth accuracy for 3D imaging," in *Proceedings ISSCC*, 2005, pp. 364–365.

[13] M. Spies and H. Spies, "Automobile Lidar Sensorik: Stand, Trends und zukünftige Herausforderungen," *Advances in Radio Science*, vol. 4, pp. 99–104, 2006.

[14] K. Fürstenberg and R. Schulz, "Laserscanner for driver assistance," in *Proceedings WIT*, Hamburg, Germany, 2006.

[15] G. Doemens P. Mengel and L. Listl, "Fast range imaging by CMOS sensor array through multiple double short time integration (MDSI)," in *Proceedings ICIP*, Thessaloniki, Greece, Oct. 2001, vol. 2, pp. 333–336.

[16] D. Stoppa et al., "A 16x16-pixel range-finding CMOS image sensor," in *Proceedings ESSCIRC*, Leuven, Belgium, Sept. 2004, pp. 419–422.

[17] Graham Saxby, *The Science of Imaging*, Institute of Physics Publishing, Bristol, England, 2002.

[18] R. Kuhla, *Hardware-Modellierung und Algorithmus-Optimierung für die dreidimensionale Objekterfassung durch CMOS-Bildsensoren mit Kurzzeitintegration und Laserpulsbeleuchtung*, Ph.D. thesis, Universität Duisburg-Essen, 2007.

[19] Peter Gulden and Marc Fischer, "Verschiedene halbleiterbasierte Verfahren zur Bestimmung der Objektentfernung mit Hilfe der Lichtlaufzeit," Sept. 2003, Internal research report Siemens AG.

[20] P. Mengel et al., "Three-dimensional CMOS image sensor for pedestrian protection and collision mitigation," in *Advanced Microsystems for Automotive Applications 2006*, J. Valldorf and W. Gessner, Eds., pp. 23–39. Springer, Berlin, Germany, 2006.

[21] P. Mengel et al., "Time-of-flight camera for pedestrian protection and collision mitigation," in *Proceedings ITS*, Aalborg, Denmark, June 2007.

[22] O. M. Schrey et al., "A 4x64 pixel CMOS image sensor for 3D measurement applications," in *Proceedings ESSCIRC*, Estoril, Portugal, Sept. 2003, pp. 169–172.

[23] O. Elkhalili et al., "A 64x8 pixel 3-D CMOS time of flight image sensor for car safety applications," in *Proceedings ESSCIRC*, Montreux, Switzerland, Sept. 2006.

[24] FEMTO Messtechnik GmbH, "Photoreceiver HSA-Y-S," information available from http://www.femto.de, status: 06/01/2007.

[25] Lennart Råde and Bertil Westergren, *Springers Mathematische Formeln*, Springer, Berlin, Germany, 2^{nd} edition, 1997.

[26] Eli Kapon, *Semiconductor Lasers I*, Academic Press, San Diego, USA, 1998.

[27] "Communication with LEWICKI microelectronic GmbH," information available from http://www.lewicki.de, status: 06/01/2007.

[28] Normenstelle Informationstechnik im DIN Deutsches Institut für Normung e.V., "DIN EN 60825-1:2003-10 - Sicherheit von Lasereinrichtungen - Teil 1: Klassifizierung von Anlagen, Anforderungen und Benutzer-Richtlinien," 2003.

[29] G. Hauske, "Statistische Signaltheorie," Skriptum zur Vorlesung, Lehrstuhl für Nachrichtentechnik, Technische Universität München, 2003.

[30] O. Elkhalili, *Entwicklung von optischen 3D CMOS-Bildsensoren auf der Basis der Pulslaufzeitmessung*, Ph.D. thesis, Universität Duisburg-Essen, 2004.

[31] B. Koenig, B. Hosticka, P. Mengel, and L. Listl, "Novel long distance camera system with real-time 3D Image processing," in *Proceedings SPIE Optics & Photonics*, San Diego, USA, Aug. 2007.

[32] E. Steinbach, "Medientechnik," Skriptum zur Vorlesung, Lehrstuhl für Kommunikationsnetze, Technische Universität München, 2002.

[33] Normenstelle Luftfahrt im DIN Deutsches Institut für Normung e.V., "DIN 9300 - Begriffe Größen und Formelzeichen der Flugmechanik," 1990.

[34] David A. Forsyth and Jean Ponce, *Computer Vision: A Modern Approach*, Prentice Hall, Upper Saddle River, USA, 2003.

[35] Dana H. Ballard and Christopher M. Brown, *Computer Vision*, Prentice Hall, Englewood Cliffs, USA, 1986.

[36] Kenneth R. Castleman, *Digital Image Processing*, Prentice Hall, Upper Saddle River, USA, 1996.

[37] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, Dec. 2006, Article 13.

[38] Samual Blackman and Robert Popoli, *Modern Tracking Systems*, Artech House, Boston, USA, 1999.

[39] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transaction of the ASME-Journal of Basic Engineering*, pp. 35–45, Mar. 1960.

[40] Peter Maybeck, *Stochastic Models, Estimation, and Control*, Academic Press, San Diego, USA, 1979.

[41] G. Welch and G. Bishop, "An introduction to the Kalman filter," http://www.cs.unc.edu/~welch/kalman/, status: 06/01/2007.

[42] Yaakov Bar-Shalom and Xiao-Rong Li, *Estimation and Tracking: Principles, Techniques, and Software*, Artech House, Boston, USA, 1993.

[43] Robert Grover Brown and Patrick Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, John Wiley & Sons, New York, USA, 3^{rd} edition, 1997.

[44] Brian D.O. Anderson and John B. Moore, *Optimal Filtering*, Dover Publications, Mineola, USA, 2005.

[45] Elliott Kaplan, *Understanding GPS: Principles and Applications*, Artech House, Boston, USA, 1996.

[46] FORWISS Passau, "MIDIAS project description," http://www.forwiss.uni-passau.de/de/projectsingle/61/main.html, status: 07/01/2007.

[47] GITG Universidad Politécnica de Valencia, "CASBliP project homepage," http://www.casblip.upv.es/, status: 07/01/2007.

[48] A. F. Rodríguez Hernández et al., "Figure perception from real and virtual sounding surfaces," in *Proceedings Technology and Persons with Disabilities Conference*, Los Angeles, USA, 2003.

[49] EMVA European Machine Vision Association, "EMVA Standard 1288 - standard for characterization and presentation of specification data for image sensors and cameras," Aug. 2005, Release A1.00.

List of Abbreviations

2D	**2**-**D**imensional
3D	**3**-**D**imensional
3D-CMOS	Acronym for the sensor chip technology
ADC	**A**nalog to **D**igital **C**onverter
AMSE	**A**verage **M**ean **S**quare **E**rror
CCT	**C**olor **C**oded **T**riangulation
CDS	**C**orrelated **D**ouble **S**ampling
CMOS	**C**omplementary **M**etal **O**xide **S**emiconductor
CS	**C**oordinate **S**ystem
CSI	**C**orrelated **S**hort-time **I**ntegration
CW	**C**ontinuous **w**ave
DCSI	**D**ifference **C**orrelated **S**hort-time **I**ntegration
DR	**D**ynamic **R**ange
FF	**F**ill **F**actor
FOV	**F**ield **O**f **V**iew
FPGA	**F**ield **P**rogrammable **G**ate **A**rray
FPN	**F**ixed **P**attern **N**oise
FR	**F**rame **R**ate
IR	**I**nfra-**R**ed
LASER	**L**ight **A**mplification by **S**timulated **E**mission of **R**adiation
LED	**L**ight **E**mitting **D**iode
LIDAR	**LI**ght **D**etection **A**nd **R**anging
MDSI	**M**ultiple **D**ouble **S**hort-time **T**ntegration
MPE	**M**aximum **P**ermissible **E**xposure
MSE	**M**ean **S**quare **E**rror
NEE	**N**oise **E**quivalent **E**xposure
NIR	**N**ear **I**nfra-**R**ed
OL	**O**bject **L**ist
OTA	**O**perational **T**ransconductance **A**mplifier
PD	**P**hoto **D**iode
PDF	**P**robability **D**ensity **F**unction
PRF	**P**ulse **R**epetition **F**requency

RADAR	**RA**dio **D**etection **A**nd **R**anging
SF	**S**ource **F**ollower
SNR	**S**ignal-to-**N**oise **R**atio
SONAR	**SO**und **N**avigation **R**nd **R**anging
TOF	**T**ime **O**f **F**light
VOV	**V**olume **O**f **V**iew
VRU	**V**ulnerable **R**oad **U**ser

List of Important Symbols

$\boldsymbol{A}$	state transition matrix
A_{Ill}	illuminated area
A_{Object}	object area
A_{Pixel}	active pixel area
A_{Sensor}	active sensor area
$AMSE$	average mean square error
C	capacitance
C_{acc}	readout capacitance
C_D	diode capacitance
C_H	hold capacitance
C_{Int}	integrator capacitance
C_{sense}	sense capacitance
D	aperture diameter
DR	dynamic range
E	irradiance
$E_{Sensor}(t)$	irradiance characteristic on the sensor surface (from active illumination)
$E_{Back}(t)$	background light irradiance characteristic on the sensor surface
F	camera focus
FR	camera frame rate
$\boldsymbol{H}$	measurement matrix
H	radiant exposure
$\boldsymbol{I}$	identity matrix
I_D	photo diode current
$\boldsymbol{K}$	Kalman gain matrix
M	magnification factor, shutter switch
MSE	mean square error
N	number variable
NEE	noise equivalent exposure
$\boldsymbol{P}$	estimate error covariance matrix
P_{Diode}	incident power on the reference diode
P_{Module}	output power of a laser module
$P_{Laser}(t)$	ideal laser pulse function

$\hat{P}_L$	laser pulse amplitude
PRF	pulse repetition frequency
$\boldsymbol{Q}$	process noise covariance matrix
Q_{photo}	photo charge
$\boldsymbol{R}$	measurement noise covariance matrix
R_λ	sensor responsivity for a given wavelength λ
$S_{\lambda Sensor}$	sensor sensitivity for a given wavelength λ
$S_{\lambda Sensor}(t)$	ideal shutter window functio for a given wavelength λ
$\hat{S}_{\lambda S}$	sensor sensitivity amplitude for a given wavelength λ
T	discrete time period
T_{DL}	delay time of the laser pulse
T_{DS}	delay time of the shutter window
T_{Int}	integration time
T_{Pw}	pulse width
$U_{C_{out}}$	ideal pixel output voltage
U_i	ideal sensor output signal for shutter pulse alignment i, $i \in a, b, c, d$
U_{Int}	integrator output voltage
U_{Diode}	reference diode output voltage
U_{ddpix}	reference voltage
U_{Sensor}	ideal sensor output voltage
a_E	Euclidean distance
a_M	Mahalanobis distance
b	basis length
c	speed of light
d	distance
d_m	measured target distance
d_r	real target distance
$d(n_{acc})$	damping function
f	focal length, frequency
g_{SF}	amplification factor of the source follower
$f_\#$	f-number
k	discrete time step instance
n_{acc}	number of analog accumulations
n_{av}	number of averaged frames
p	disparity
r_i	residual of varaible i
r_{blur}	radius of a blur disc
t	time
t_{acqu}	acquisition time
t_{Tr}	laser trigger length

u	object distance, horizontal pixel index
u_{Sensor}	sensor output voltage
$\boldsymbol{v}$	measurement noise vector
v	image distance, vertical pixel index
$\boldsymbol{w}$	process noise vector
x	distance value in x-coordinate direction
$\boldsymbol{x}$	system state vector
y	distance value in y-coordinate direction
z	distance value in z-coordinate direction
$\boldsymbol{z}$	measurement vector
Δ	difference operator
Φi	digital control signal i
α	angle variable
β	angle variable
θ	angle variable
κ	optical loss factor
λ	wavelength
μ_i	mean value of signal i
ρ	surface reflectivity
σ_i	standard deviation of signal i
σ_R	readout noise
σ_S	sensor noise
σ_U	over-all system noise
σ_U^*	bright noise
τ_{TOF}	time of flight
ϕ	angle variable
ψ	angle variable

List of Figures

List of Tables