

FACHHOCHSCHULE AACHEN
Fachbereich 6
Luft- und Raumfahrttechnik

Master Thesis
Range Imaging Applications
for Dynamic Deformation
Measurement

von
Tobias Augspurger
Matriculation Number: 357059

Referent: Prof. Dr.-Ing. H. Funke
Korreferent: Prof. Dr.-Ing. B. Burbaum

Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die im Literaturverzeichnis angegebenen Quellen benutzt habe.

Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder noch nicht veröffentlichten Quellen entnommen sind, sind als solche kenntlich gemacht.

Die Zeichnungen oder Abbildungen in dieser Arbeit sind von mir selbst erstellt worden oder mit einem entsprechenden Quellennachweis versehen.

Diese Arbeit ist in gleicher oder ähnlicher Form noch bei keiner anderen Prüfungsbehörde eingereicht worden.

Aachen, January 26, 2015

Abstract

Three-dimensional Time-of-Flight (ToF) range cameras improved significantly in accuracy, speed, price and size over the last years. By delivering the range from the scene to the image sensor in every pixel, static and dynamic scenes can be reconstructed into a three-dimensional model contactless from a single point of view. The observed area is illuminated by infrared diodes and a IR CMOS camera measures the time delay of the returning light [1]. Out of it the range is acquired parallel with no time delay between the pixels [2]. This offers the potential for various close to medium range applications. One of this is the measurement of vibrations and structural deformations under dynamic loads [3]. In contrast to other present measurement methods, complete surface movements can be acquired in parallel without a movable laser beam or a complex stereo triangulation between multiple cameras . The Kinect 2 range camera is used to show the performance of vibration measurements and to analyze how the range images can be filtered and processed. It is the first ToF camera on the consumer market and at the time of release the one with the highest number of pixels (512x424 at 30 frames per seconds). It delivers 3D informations from 0.6 m to 8.0 m at a depth resolution of 1 mm [4]. A Laser Doppler Vibrometer is used, as high precision reference measurement device, delivering the highest standard in deformation measurement on a single surface point. This thesis will prove, that vibrations of a body can be reconstructed at amplitudes of 4 mm with an accuracy below 1 mm . The performance depends on the distance to the camera, angle of view, surface material, number of investigated pixels and the measurement time. At the ideal distance of 1.3 m to 1.5 m , the error to the reference is reduced down to 0.05 mm after 8 seconds of measurement at a frequency of 3 Hz . Even the first and second harmonics are visible in the frequency domain on white paper, offering a high reflectivity for the infrared light of the ToF measurement principle. The influence on the accuracy of different angles to the surface normal is another topic. A vibration of 3 Hz can be reconstructed correctly in frequency up to an angle of 45°. The amplitude is overestimated with higher angles. Since every pixel delivers a range information, the complete video stream can be processed by a Fast-Fourier Transform to derive the dynamic behavior of the whole field of view. The vibration can be illustrated in images, representing the amplitude and frequency distribution in the complete field of view over a certain time span.

|

This development offers new applications in the static and dynamic observation and investigation of machines and structures. Positions, geometries, movements, orientations and other object properties can be analyzed and reconstructed in real time. Industrial installations and processes such as engines can be observed and the production quality can be improved.

Acknowledgment

The work described in this thesis would not be possible without the extensive support from Prof. Funke and the team of propulsion laboratory at the University of Applied Science Aachen. From the rough idea to the finished experimental results I got a lot of support and feedback that helped me to find the right way even if the tasks seemed impossible to fulfill. Prof. Kameier from the University of Applied Science Duesseldorf delivered us a high precision reference measurement device. He also contributed Matlab source code that was necessary for the evaluation. I would also like to thanks Prof. Franke who allocated the oscilloscope for the acquisition of the Laser Doppler Vibrometer. Many thanks goes also to my family and friends, who supported me over the last months, corrected my texts and gave me a lot of feedback.

Contents

1	Introduction	13
2	State of the Art	15
2.1	The Fundamentals of Light	15
2.2	Photometric and Radiometric Quantities	17
2.2.1	Radiant and Luminous flux	17
2.2.2	Radiant and Luminous Intensity	18
2.2.3	Luminous Efficacy	19
2.2.4	Radiance and Luminance	19
2.2.5	Transformation from Radiometric into Photometric	20
2.3	Diffuse, Spread and Spectral Reflection	20
2.4	Semiconductor Light Sources and Sinks	23
2.4.1	Light-Emitting Diodes	24
2.4.2	Laser Diodes	26
2.4.3	Photo Diodes	26
2.5	Image Sensors	28
2.5.1	Charged-Coupled Device Sensor	28
2.5.2	Complementary Metal-Oxide-Semiconductor Sensor	29
2.5.3	Attributes of Image Sensors	30
2.5.4	Rolling and Global Shutter	31
2.5.5	Camera Optic	33
2.5.6	Motion Blur	34
2.6	Range Imaging Techniques	35
2.6.1	Stereo Triangulation	35
2.6.2	Structured Light Sensor	36
2.6.3	Interferometry	37
2.6.4	Time-of-Flight	38
2.7	Time-of-Flight Imaging Errors	41
2.7.1	Systematic Errors	41
2.7.2	Non-systematic Errors	42

2.8	Acquisition of Digital Dynamic Data	43
2.9	Evaluation of Dynamic Data	44
2.9.1	Time Domain	44
2.9.2	Frequency Domain	45
3	Time-of-Flight Market Research	48
3.1	Kinect 2 System Overview	50
4	Data Processing of Range Images	54
4.1	Transformation into Point Cloud Data	55
4.2	Meshing of Point Cloud Data	56
4.3	Analysis and Filter Methods	58
4.3.1	Creation of Frame Arrays	58
4.3.2	Mean and Variance	59
4.3.3	Gradient	60
4.3.4	Denoising with Mean or Median of a Pixel Area	61
4.3.5	Total Variation Denoising	62
4.3.6	Processing of Dynamic Depth Data	64
5	Data Processing of Laser Doppler Vibrometer	66
6	Analysis of Depth Images	68
6.1	Variance Depending on the Distance	68
6.2	Variance Depending on Geometries and Surfaces	69
6.3	Influence of Background Lighting on Static Measurements	71
7	Evaluation of Membrane Dynamic	72
7.1	Physical Model of Membrane Oscillation and Observation	72
7.2	Scientific Problem	74
7.3	Hypothesis	74
7.4	Measurement Setup	75
7.5	Test Procedure	80
7.5.1	Variable Distance	81
7.5.2	Variable Angle	81
7.5.3	Variable Frequency	82
7.5.4	Variable Displacement	82
7.5.5	Sweep	82

7.6	Measurement Results	83
7.6.1	Variable Distance	84
7.6.2	Variable Angle	89
7.6.3	Variable Frequency	90
7.6.4	Variable Displacement	92
7.6.5	Sweep	94
7.6.6	Influence of Background Lighting on Dynamic Measurement	96
7.7	Conclusion	98
8	Outlook	100
8.1	Advancements in the Experimental Setup and Analysis	100
8.2	Static Calibration and Angle Calculation	102
8.3	Applications in Structure and Machine Observation	102
8.4	Dynamic Investigations of Rotating Parts	103
8.5	Application in Aerodynamic and Structural Test Facilities	103
A	Annex	105
B	Source code extracts	115
C	Sourcecode and Documentation on CD	117

List of Abbreviations

ToF	Time-of-Flight
IR	Infrared
CMOS	Complementary Metal–Oxide–Semiconductor
LDV	Laser Doppler Vibrometer
VIS	Visible
UV	Ultraviolet
CIE	Commission International de l'Eclairage
LED	Light-Emitting Diodes
CRI	Color Rendering Index
RGB	Red, Green, Blue
CCD	Charged-Coupled Device
SPAD	Single-Photon Avalanche Diodes
ADC	Analog-to-Digital Converter
DNR	Dynamic Range
SNR	Signal-to-Noise Ratio
FOV	Field of View
LiDAR	Light Detection and Ranging
SAR	Synthetic Aperture Radar
CW	Continuous-Wave
FFT	Fast Fourier Transform
DFT	Discrete Fourier Transform
STFT	Short-Time Fourier Transform
HD	High Definition
AMP	Amplifier
TIFF	Tagged Image File Format
PCL	Point Cloud Library
STL	Surface Tesselation Language
TVD	Total Variation Denosing
CSV	Comma-Separated Values
PIV	Particle Image Velocimetry

BNC	Bayonet Neill Concelman
PSP	Pressure Sensitive Paint

Table 1: List of Abbreviations

List of Symbols

Symbol	Units	Description
c	$m \cdot s^{-1}$	Speed of light
\vec{E}	$V \cdot m^{-1}$	Electric field
\vec{M}	$N \cdot m^{-1} \cdot A^{-1}$	Magnetic field
λ	m	Wave length of light
ν	Hz	Frequency of light
E	J	Energy
P	W	Power
h	$6,62606957 \cdot 10^{-34} J \cdot s$	Planck constant
m	kg	Mass
Φ_e	W	Radiant flux
Φ_v	lm	Luminous flux
Ω	sr^{-1}	Solid angle
I_e	$W \cdot sr^{-1}$	Luminous intensity
η_v	lmW^{-1}	Luminous efficacy
ϱ	%	Reflectance
τ	%	Transmittance
α	%	Absorption
U_{th}	V	Threshold voltage
R_λ	AW^{-1}	Resposivity
$Q.E.$	—	Quantum efficiency
fps	$1/s$	Frames-Per-Second
t_r	s	Response time
t_e	s	Exposure time
F	m	Focal length
θ	DEG	Angle of View
\vec{s}_{blur}	m	Motion blur
\vec{v}	ms^{-1}	Speed
n	Hz	Rotational speed
d	m	Distance

t_{ToF}	s	Time-of-Flight
f_{mod}	Hz	Modulation frequency
φ	rad	Phase shift
σ^2	-	Variance
d_{amb}	-	Ambiguity distance
n_{bits}	Bit	Bit depth
\dot{n}	$\frac{bit}{s}$	Data stream
$f_{sampling}$	Hz	Sampling frequency
f_{max}	Hz	Maximum frequency in signal
$f_{nyquist}$	Hz	Nyquist-Shannon frequency
A	-	Amplitude
A_{pp}	-	Peak-to-peak amplitude
l_{hv}	m	Vertical / horizontal length
∇	-	Nabla operator
λ_r	-	TVD regularization parameter
U_{AC}	-	Alternate voltage
E	$mm^{-1}s^{-1}V^{-1}$	Sensitivity of LDV
$T_{sampling}$	s	Time between two sample points
\vec{n}	-	Surface normal vector
α	DEG	Viewing angle
β	DEG	Cone angle of membrane

Table 2: List of symbols

1 Introduction

The interest in new innovative kinds of human/computer interaction lead to multiple low cost developments to detect the movement of a human body. The Kinect 2, developed by Microsoft, measures the propagation time of light by the Time-of-Flight (ToF) principle, to derive 3D data from the observed scene. The field of view is illuminated by IR diodes and a IR CMOS camera measures the time that the light take to reflect back into the camera. With the constant speed of light and the measured time of flight the distance, that the light traveled, can be calculated. Since the appearance of ToF cameras in the year 2000, the pixel resolution, dynamic range and the accuracy have significantly improved. In contrast to stereo triangulation between multiple cameras or 3D scanning with a moving laser beam, this technique allows low cost and small size applications for 3D scanning. High frame rates can be achieved and the 3D data is acquired on every pixel at the same time.

The acquisition of real-time 3D data also offers various applications in experimental setups. In crash test facility for example the geometry and orientation of the structures before and after the test can be measured. In windtunnel applications, the angle of attack and vibrations of the aerodynamic body can be observed by the camera without disturbing markers on the surfaces. The measurement of the static geometries and the dynamic behavior delivers a way to merge data between different calibrated camera systems in a common set of data. Aeroelasticity investigations are possible on the complete body in every point of the field of view.

This thesis will reveal how static and dynamic ToF data can be investigated and processed. Xiao Qi and Derek Lichti already showed the performance of dynamic beam deformation measurement with ToF [3]. A 1 Hz sinus oscillation of an concrete beam with an amplitude 4 mm was reconstructed with an accuracy of 0.061 mm. This was possible with two phase synchronized ToF cameras in a stereo arrangement for a single point on the beam. The noise and uncertainty between both sensors are canceled out between both measurements. The advantages of static investigation with a Time-of-Flight camera for structural deformation measurement have already been proven with last generation devices [5]. In previous investigations of 3D scanning techniques for static deformation, Terrestrial Laser Scanning

also showed a high performance, but the systems are slow, expensive and large [6].

Advanced ToF image sensors with a higher pixel resolution, accuracy and improved noise characteristic promise the dynamic investigations of complete surfaces with a single camera. A depth resolution up to 0.3 mm is possible, with current CMOS technology [7]. Other devices feature high frame rates of 470 fps in a 1280×1024 pixel area [8]. The dynamic performance of the Kinect 2 ToF camera, that is originally designed for human/computer interaction, will be investigated. The accuracy of the measurement will be compared to a Laser Doppler Vibrometer (LDV). A speaker membrane delivers a constant dynamic movement with a calibrated sinus function as input. Therefore, deformation measurement like Stereo Pattern Tracking can be investigated on a controlled test body [9]. Boundary conditions like surface material, geometry and angle of the camera to the surface normal will be analyzed in terms of precision and accuracy. Another major advantage of range data is the simple image processing that will be shown in Matlab for some basic examples for the static and dynamic structure investigation. The creation of surface dynamics images, showing the frequency and amplitude distribution, is one of the main subjects in this context. This delivers a simple and easy way to obtain the dynamic behavior of complete surfaces. Applications of Time-of-Flight cameras are shown in multiple small experiments. The revealed procedures are only supposed to show the performance of the measurement principle in a rough way and demonstrate how a real calibration of the ToF System in a controlled environment can look like. ToF cameras have not been investigated previously at the Department of Aerospace Technology which offers the possibility to build up knowledge from a unprejudiced basis. The source code for image processing has been developed without any examples. Improvements are still necessary in the experimental setup, source code and suggested theories.

2 State of the Art

2.1 The Fundamentals of Light

Light is a substantial part of the universe. It lets us see the world, the stars and has remarkable properties in the space-time. For example it is the fastest way of energy transportation with a speed of $c = 299792458 \text{ m/s}$ in vacuum. The definition of electromagnetic radiation gathers all kinds of light like visible (VIS), infrared (IR), ultraviolet light (UV) and other kinds of radiation like microwaves or radio waves. James Clerk Maxwell discovered that light can be described by an electric \vec{E} and a magnetic \vec{M} field that stand perpendicular to each other and to the direction of wave propagation. Figure 1 shows the propagation of one light ray. Both fields oscillated with the same wavelength of the light λ and the frequency ν moving into the wave direction of the wave vector propagation. Green light for example is defined within 520 and 565 nm while radio waves have a length of several meters. With shorter wavelength, the energy of the photon increases and interaction with material changes. Figure 2 shows the complete electromagnetic spectrum from γ -rays to long radio waves. With shorter wavelengths, the radiation causes interactions at the atomic and subatomic level.

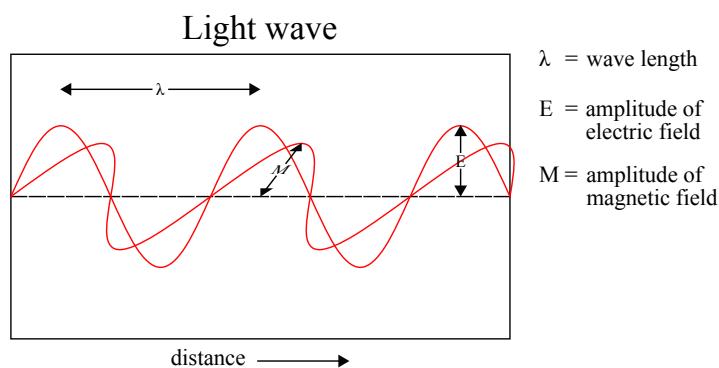


Figure 1: Light as electromagnetic wave Gpvos 2007 wikipedia.org CC0

The polarization of light characterizes how the fields \vec{E} and \vec{M} orientation in space is distributed between different light waves. An unpolarized light source emits a chaotic orientated wave form. After polarized filtering only waves with the same orientation remain. The coherence of light describes correlations in the wavelength and the phase relation

between the \vec{E} and \vec{M} oscillation. Coherent light sources cause interference between each other, resulting in a new coherent wave form. The coherence length is the distance of propagation where the wave is still of the same wavelength and phasing and can interfere with itself in the same way. Two interfering coherent waves in the same phase cause constructive interference. The amplitudes will sum up. A phase shift of 180° between two coherent waves will result in a destructive interference and a subtraction of amplitudes. Incoherent waves that are polarized can still be monochromatic, meaning that the wavelength is in a narrow band in the spectrum. However the phase and directions of the waves are randomly distributed between each other.

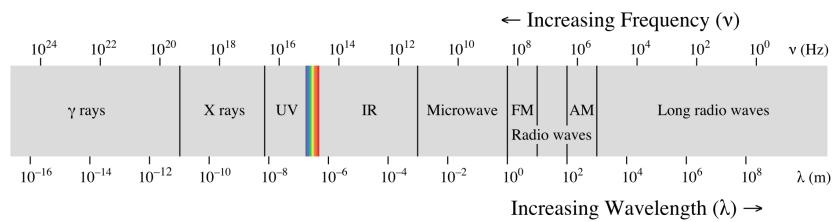


Figure 2: Electromagnetic spectrum Philip Ronan wikipedia.org ©①©

The Planck constant h specifies the relation between energy E and frequency of light ν . The radiant energy of one light beam can also be considered as a particle with the mass m , a so called photon.

$$E = h \cdot \nu = mc^2 \quad [J] \quad (1)$$

$$\lambda = \frac{c}{\nu} \quad [m] \quad (2)$$

The mass-energy equivalence proves that light waves can be transformed into and from mass. This can be observed in the sun. The nuclear fusion of two hydrogen atoms results in a helium atom. A small amount of mass is lost in this reaction and transformed into electromagnetic energy that lightens our universe. The different wavelengths of a light source are represented in the electromagnetic spectrum. The sun's irradiance provides the complete electromagnetic power per square meter that is delivered on a surface in a certain distance. This value is corrected by the wavelengths of the specific light spectrum in the spectral irradiance. Figure 3 shows the VIS and IR components of the sunlight and how they are influenced by the atmosphere of the earth in the spectral irradiance versus the wavelength.

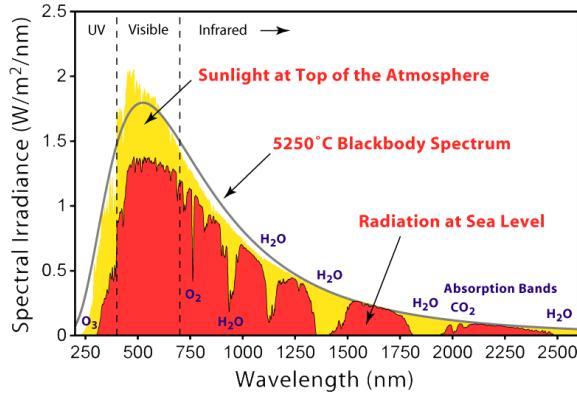


Figure 3: Solar radiation spectrum on earth Nick84 wikipedia.org 2013 ©○○

The atmosphere of the earth absorbs various wavelengths that are related to the gas composition. Water (H_2O) and carbon dioxide (CO_2) absorb the infrared light of the solar spectrum while ozone (O_3) absorbs the high energy UV radiation. Nitrogen (N_2), however, is transparent for most components of the sunlight.

2.2 Photometric and Radiometric Quantities

Photometric (index v) and radiometric (index e) quantities are used for the quantification of light sources and lighting conditions. The human spectral sensitivity for visible light is considered in the luminosity function. The eye is capable to adapt its spectral sensitivity to different light conditions. Figure 4 shows the CIE standard luminosity function $V(\lambda)$ in the photopic view under well-lit lighting and the scotopic view under low light. The peak of the photopic sensitivity can be found at about 555 nm which is the base for the normalization of the curve. The CIE luminosity function is a standard established by the Commission International de l'Eclairage and forms the central color matching function for light sources. The radiometric quantities consider the complete electromagnetic spectrum and are independent of the sensitivity of the human eye [10].

2.2.1 Radiant and Luminous flux

The radiant flux Φ_e is a scale for the radiant energy Q emitted per time unit on a surface. It is calculated by a given radiant intensity I_e and the solid angle Ω . In other contexts it is also called electromagnetic power:

$$\Phi_e = \frac{dQ}{dt} = I_e \cdot \Omega \quad [W] \quad (3)$$

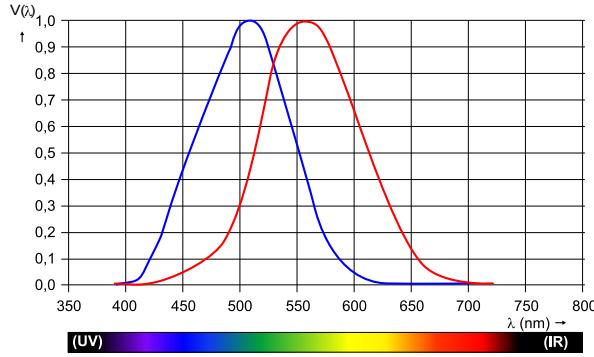


Figure 4: CIE Luminosity function $V(\lambda)$ of the human eye under low light (blue curve) and high light (red curve) conditions HHahn wikipedia.org ©①②

The complement is called luminous flux Φ_v in the visible spectrum, with the unit lumen [lm]. A light source with 1 cd, that emits in every direction in a solid angle of 4π , will produce a luminous flux of 12.6 lm. Every material above 0 kelvin radiates light, depending on the surface temperature. The radiated power can be calculated for a ideal black body that absorbs all radiation with the Stefan-Boltzmann-Constant σ , the material temperature T in kelvin, the emitting area A and the emissivity ϵ of the surface.

$$\Phi_e = \epsilon(T) \cdot \sigma \cdot A \cdot T^4 \quad [W] \quad (4)$$

2.2.2 Radiant and Luminous Intensity

The radiant intensity I_e is used for quantization of the total electromagnetic power emitted by a light source. The power is distributed on sphere areas while moving through space. One steradian (sr) defines an reference area $A = r^2$ on a sphere with the radius r . Figure 5 shows the propagation of light under the solid angle Ω .

$$\Omega = \frac{A}{r^2} \quad [sr^{-1}] \quad (5)$$

$$I_e = \frac{\Phi_e}{\Omega} \quad [Wsr^{-1}] \quad (6)$$

The radiant power of visible light can be judged by the luminous intensity I_v . The power of the spectrum is adapted to the sensitivity of the human eye for different colors. The result

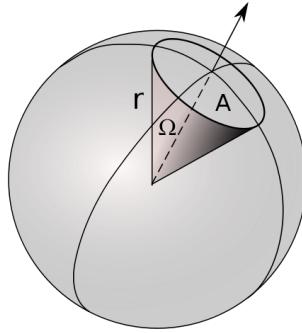


Figure 5: Propagation of light in one steradian Haade wikipedia.org CC0

is the wavelength-weighted power emitted by the light source per unit solid angle. The unit of luminous intensity is the candela (cd), an SI base unit:

One candela is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of a wavelength of 555 nm and that has a radiant intensity in that direction of 1/683 watt per steradian[11].

A 100 W light bulb offers a luminous intensity of about 1.1 cd. Modern LEDs in contrast can have up to 5 cd, due to the concentration of the emitted light in one direction and a higher luminous efficacy η of up to 208 lmW^{-1} .

2.2.3 Luminous Efficacy

The amount of power P that is transformed in a light source into visible light is represented by the luminous efficacy η_v . The remaining energy will be radiated in other invisible wavelengths or transferred into a heat flow that vanishes into ambient material and air.

$$\eta_v = \frac{\Phi_v}{P} \quad (7)$$

2.2.4 Radiance and Luminance

The radiance defined in equation 8 is a measure for the quantity of radiation that passes through an area. It can also be used to quantify the amount of radiation, which is reflected from a surface in a specific direction. It changes with the cosine of the incident angle φ

normal to the emitting surface area. This equation is also called Lambert's cosine law and applies to diffuse reflecting surfaces or Lambertian surfaces.

$$L_e = \frac{I_e}{A_e} = \frac{d^2\Phi_v}{dA d\Omega \cos\varphi} \quad [Wm^2sr^{-1}] \quad (8)$$

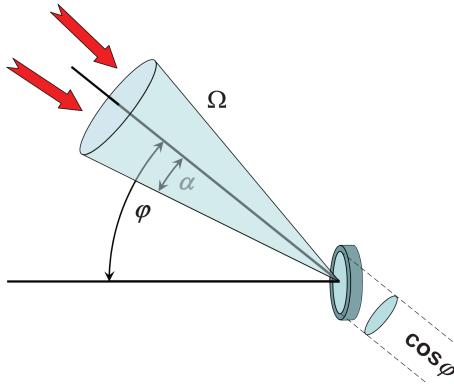


Figure 6: Radiance and Luminance on a surface

The luminance L_v is the analog quantity with respect to the eye. It is a surface's grade of brightness to the human eye.

2.2.5 Transformation from Radiometric into Photometric

The radiometric quantities can be transformed into photometric quantities with the help of the Luminosity function $V(\lambda)$ and the efficacy of radiation η_e . With a radiant flux given in watt, this value can be compared with photometric quantities in lumen. Even if a light source emits low luminous flux, the destructive impact on the human eye can still be significant from the IR or UV spectrum. That is why for the hazard consideration, the radiometric quantities with specific wave lengths have to be considered.

$$\Phi_v = \eta_e \cdot V(\lambda) \cdot \Phi_e \quad [W] \quad (9)$$

2.3 Diffuse, Spread and Spectral Reflection

Depending on the wavelength, light interacts differently with material. The surface structure and material properties play a further role in the reflection and absorption characteristics. An ideal spectral reflecting area will radiate all light backwards with the same angle mirrored on the normal vector of the surface. Most materials can give specular reflection, provided that the surface is polished to eliminate irregularities within the magnitude of the

wavelength. Transmission is the passage of the electromagnetic waves through a medium. The exit angle of the beam depends on the entry angle and the properties of the medium and light. Diffuse reflection is a combination between transmission and spectral reflection. The waves enter the material and are reflected on the surface multiple times before leaving it again. The direction of reflection is independent from the angle of the incident light. Ideal matt surfaces like white paper have this property and follow the Lambert's cosine law from equation 8. The mixture of diffuse and spectral reflection is called spread reflection and can be observed on most common material. In this case, a diffuse propagation will take place around the direction of ideal spectral reflection, depending on the incident angle [12]. Figure 7 shows the ideal diffuse, spread and spectral reflection of an incident light ray. The reflectivity is the fraction of the reflective ϕ_r electromagnetic power to the incident electromagnetic power ϕ_i .

$$\varrho = \frac{\phi_r}{\phi_i} \quad (10)$$

The transformation of radiation in other forms of energy in a medium is called absorption. Heat and other atomic reactions are created by the radiation interacting with the atoms. Some fluorescent materials, like phosphorous, emit the energy over a longer timespan in shifted wave lengths. The ratio of reflectance ϱ , transmittance τ and the absorption α must correspond to the conservation of energy. As a conclusion, equation 11 must be fulfilled.

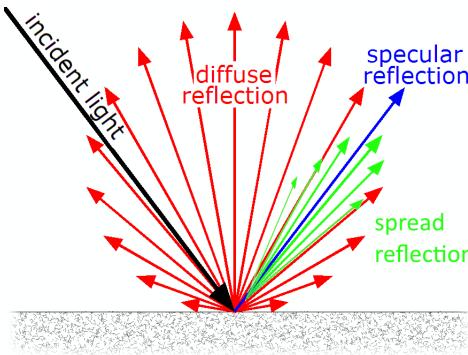


Figure 7: Reflection on an ideal mirroring surface and a diffuse surface GianniG46 wikipedia.org ©○○

$$\varrho + \tau + \alpha = 1 \quad (11)$$

The reflectivity depends on the wavelength, the surface structure and the material. Infrared light is reflected in another way than visible light. White paper, for example, is a highly

reflecting material for IR light with a reflectance of about 80 %. Black paper, in contrast, will absorb most of the incident light. In the reflectance spectrum, the reflectivity is plotted as a function of the wavelength. Aluminum reflects most of the incident light in the IR and VIS spectrum. Gold absorbs most of the radiation in the UV region and becomes reflective for red and IR wavelengths. Figure 8 shows the reflection versus the wavelength for different metals.

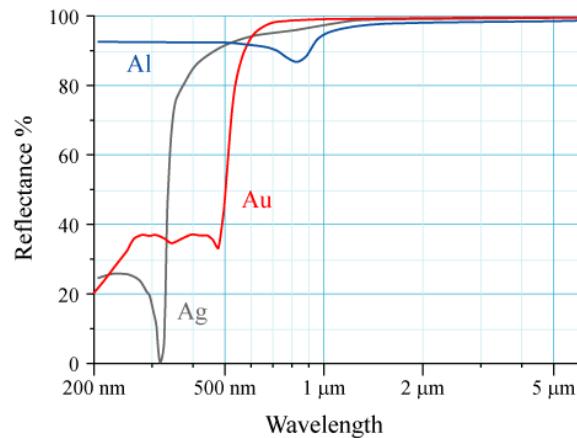


Figure 8: Reflectance spectrum of aluminium (Al), silver (Ag) and gold (Au) Bob Mellish wikipedia.org 2005

2.4 Semiconductor Light Sources and Sinks

Nowadays, light generation in semiconductor material plays a crucial role in all regions of illumination. Diodes consist of a chip of two layers of semi-conducting material, which are doped with impurities. As in most diodes, current flows from the anode to the cathode when a threshold voltage U_{th} is reached between both layers. The anode consists of negatively charged holes (p) and the cathode delivers positive charges (n) in form of electrons. The contact area between both materials is called space charge region or depletion area. It gets conductive when the threshold voltage is reached. A current flow in the reverse direction is not possible under normal conditions. Figure 9 shows the distribution of impurities in a common diode.

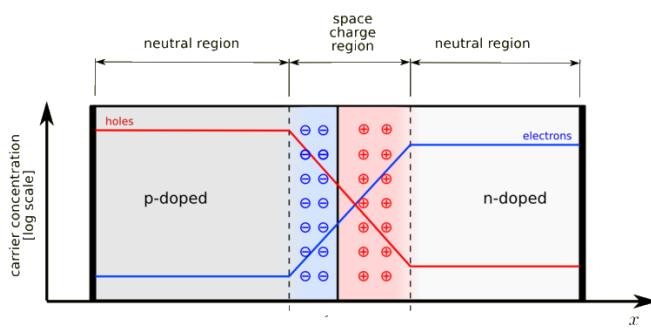


Figure 9: p-n junction of a diode TheNoise wikipedia.org

The recombination of the charges releases energy in the form of radiation. The energy difference between the top of the valence band (holes) and the bottom of the conduction band (electrons) is described by the band gap. The emitted wavelength depends on the size of the gap and the corresponding voltage that drops in this region. For infrared LEDs gallium arsenide (GaAs) or aluminium gallium arsenide (AlGaAs) are used as semi-conducting materials. The emitted wavelength depends on the voltage drop and on the semiconductor material.

A current will flow through a GaAs IR diode at a voltage higher than 1.63 V. Blue light is emitted from indium gallium nitride diodes and a U_{th} higher than 2.48 V. This process does also work in the opposite direction. Incident radiation with a wavelength according to the band gap of the diode induces a voltage. In general, this process is possible in every diode, but photo diodes are very sensitive to incident light. That is why this components are used to measure radiation. Figure 10 shows the current curve versus voltage of an silicon diode. U_{th} is defined as the point with a current flow of 1 mA. Higher radiant intensity can be achieved through an increase in voltage, which causes a higher current flow

[13]. With a decreasing temperature, the threshold voltage U_{th} will be shifted to higher values.

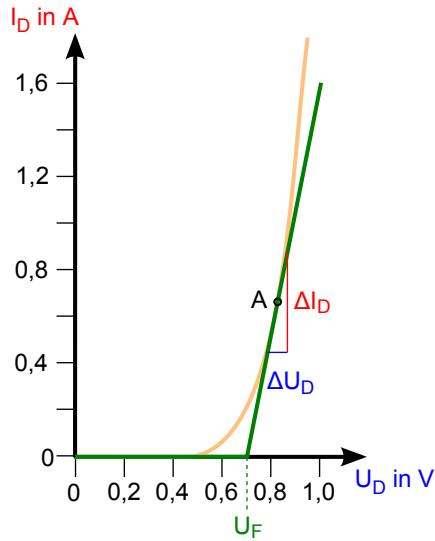


Figure 10: Diode current versus voltage Stuendle 2011 wikipedia.org

2.4.1 Light-Emitting Diodes

With the development of blue LEDs in 1994 new ways of lightning were possible. Other colors, such as red, were discovered in 1968, but only with the discovery of the color blue, the total visible spectrum was covered. The high efficiency and the long lifetime of LEDs offer huge steps in illumination in comparison to light bulbs or halogen bulbs. For the human sensation of white light, another problem had to be solved with LEDs. The color rendering index (CRI) is a measure how the LED reveals the color of illuminated objects in reference to the sunlight. The little fraction of red light in the white spectrum modifies the color of illuminated objects. The light seems to be "cold". More red light on the other hand reduces the luminous efficacy, but increases the CRI value.

The electromagnetic spectrum of a LED can be approximated by the Gaussian distribution. A blue LED will emit light with wavelengths between 450 and 500 nm. This range is called spectral width. The speed of respond is often given by the rise and fall time; the time required to go from 10 % to 90 % power. The shortest possible pulse length with modern fast LEDs can be shorter than 1 ns. As a conclusion results the maximum switching frequency can be up to multiple megahertz for LEDs.

White light can be generated by the combination of red, green and blue. Another method

is the coating of a blue LED with phosphor: A fraction of the blue light undergoes a wavelength shift, being transformed from shorter to longer wavelengths.

There are two basic types of LED structures: Edge emitters and surface emitters. Surface emitters have a comparatively simple structure. The light will be radiated in all directions from the surface of the semiconductor. The total optical output power is higher than the one from edge emitters. In contrast, edge emitters have a smaller spectral width, a very high power density and a internal optical gain like a laser. The light leaves the chip between the layers on the edge of the semiconductor [14].

The normalized radiant intensity as function of the angle of radiation can be determined experimentally and characterize from different angles. These diagrams are often published by manufacturers in the data sheets of the light source as shown in figure 11 for a common IR LED.

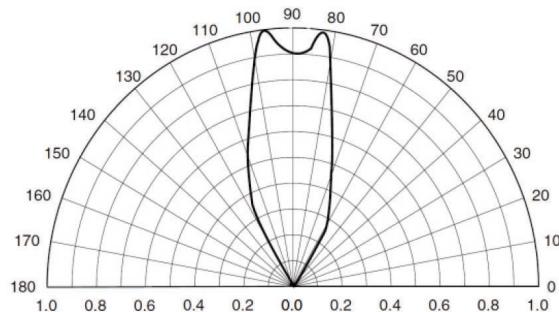


Figure 11: Radiation diagram of a IR LED QED222 www.fairchildsemi.com

2.4.2 Laser Diodes

Laser diodes in contrast to LED create monochromatic light in a narrow bandwidth. An additional undoped intrinsic semiconductor is implemented between the p-n junction. It emits light waves with chronologically steady polarization and phases to each other. Therefore, ideally only a fixed wavelength is created. As a conclusion the oscillation of the emitted light is therefore phase-synchronous. The beam divergence is low. This property is called collimated in optics [15]. An important advantage of laser diodes is the fact, that it can be modulated in intensity with high frequencies up to several gigahertz and extreme short pulses in the region down to femtoseconds. In most technical implementations, the laser light intensity will be additionally measured with a photo diode near the laser diode to get a feedback for precise control. Unlike a regular diode used in electronics, the goal for a laser diode is that all carriers recombine in the I region, and produce light. Thus, laser diodes are fabricated using direct bandgap semiconductors. Figure 12 illustrates the setup of a laser diode on its heat sink, the monitoring photo diode and the protecting window.

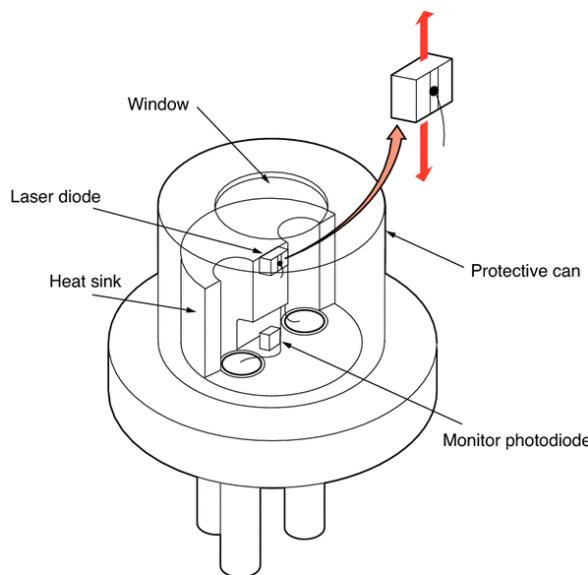


Figure 12: Setup of a common laser diode [16]

2.4.3 Photo Diodes

Photo diodes are used for the conversion from radiant into electrical energy. The physical principle behind this is the inner photoelectric effect. If absorption of light occurs in the junction, holes move to the anode and electrons to the cathode. A current is induced and therefore a voltage can be measured. The responsivity R_λ , also known as responds, is the ratio between the current flow relative to the radiant power in different wavelengths. The

quantum efficiency (Q.E.) is defined as the fraction of the incident photons, that contribute to photo current.

$$R_\lambda = \frac{I_P}{P} \quad [AW^{-1}] \quad (12)$$

$$Q.E. = R_\lambda \cdot \frac{hc}{\lambda q} \quad [1] \quad (13)$$

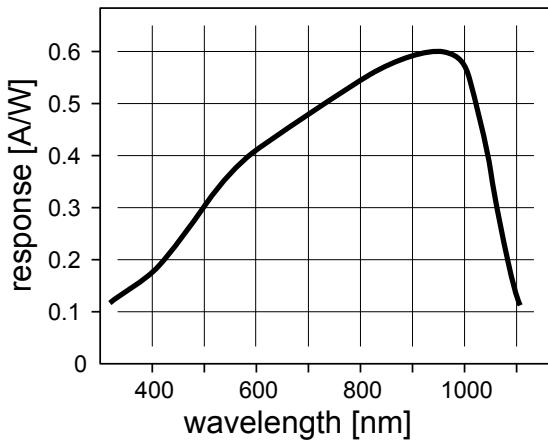


Figure 13: Response over wavelength of a silicon photodiode Kai Martin 2007 wikipedia.de

Every signal of the photo diode will be overlayed by dark current and thermal current. This noise results from electromagnetic background radiation and thermal movement of the charge carriers. To achieve the raw noise characteristics of a photo diode, the output signal must be measured in an absolute dark room with no background radiation. The rise and fall time of a photo diode can be expressed as the frequency response at which the output amplitude decreases by 3 dB. Three factors influence the response time:

1. The charge collecting time of the carriers in the depleted region.
2. The charge collecting time in the undepleted region.
3. The resistance and capacity of the diode circuit combination.

$$t_r \simeq \frac{0.35}{f_{3dB}} \quad [s] \quad (14)$$

Single-photon avalanche diodes (SPAD) are very sensitive to incident light and enable the measurement of very low intensities. The electric field between the junction is so high, that the arrival of a single photon can be determined down to ten picoseconds. A photo-generated carrier triggers an avalanche current due to the impact of ionization mechanism. The circuit behind the diode must fulfill multiple tasks [17]:

1. Sense the leading edge of the avalanche current.
2. Generate a standard output pulse synchronous with the avalanche build-up.
3. Quench the avalanche by lowering the bias down to the breakdown voltage.
4. Restore the photo diode to the operative level.

2.5 Image Sensors

A radiant flux can be converted into a 2D data matrix by an image sensor, an array of photo diodes. Every single element of the array, a so-called pixel, consists of one or multiple photo diodes, that are arranged in columns and rows. The pixel pitch is the dimension of a single pixel. The area that is actually sensitive to the light is determined by the fill factor times the pixel pitch. Without a micro lense, this area corresponds to the photodiode surface. The physical extent of the complete flat sensor is called optical size. An ordinary picture of visible light consists of a mixtures of red, green and blue (RGB) colors, but single photo diodes can only be optimized for a narrow band in the light spectrum. The band gap must correspond to the wave length of the light to detect it. Therefore three diodes are needed to deliver the RGB information on one pixel. They can be arranged side by side or stacked. The classical Bayer filter separates the light with red, green and blue filters. Foveon X3 sensors, in contrast, use three stacked photo diodes with various band gaps. IR image sensors are made of special semiconductor material like Indium Gallium Arsenide.

2.5.1 Charged-Coupled Device Sensor

The first digital image sensor was based on the charged-coupled device (CCD) architecture. In the first place Bell Labs invented this type of integrated circuit for data storage in shift register. Charges could be shifted between the different capacitors that hold the energy of each diode. Later, it was discovered that this circuit can also be used as an image sensor. After the array is exposed to light, every pixel contains information about the incident light in form of charges. This data is transferred into a storage pixels array. Gathered charges are transformed into a corresponding voltage by a single amplifier in series order. The signal can be converted afterwards into a digital representation or into a continuous analog

signal. The bit depth or color depth of the analog-to-digital converter (ADC) is given in bits and specifies the number of possible steps for the discretization. The color precision depends on this value and on the amplifier characteristics.

2.5.2 Complementary Metal-Oxide-Semiconductor Sensor

Modern semiconductor technology lead to the development of image sensors that are produced based on complementary metal–oxide–semiconductors (CMOS), also called active-pixel sensor. In contrast to CCD, every pixel is implemented with his own amplifier. Other components, like the ADC or the oscillator are integrated onto the sensor chip. This allows production under a much lower cost. Thus most consumer market cameras nowadays are based on this technique. The simplest realization of an CMOS pixel can be done by a photo diode and three CMOS transistors as shown in figure 14. The charges will be gained inside every pixel after the exposure time. The circuit can be discharged via the reset connection, for every new image.

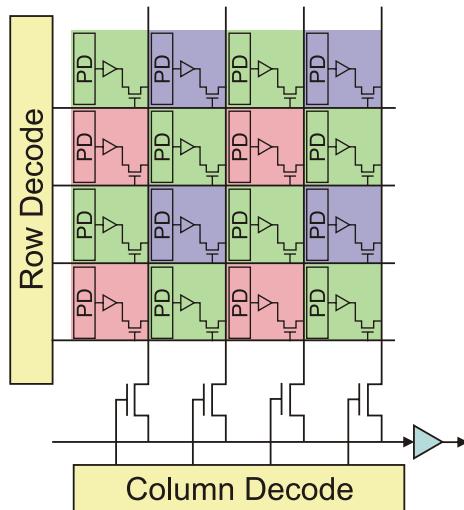


Figure 14: Frame transfer in a CMOS image sensor Shape 2008 wikipedia.de

2.5.3 Attributes of Image Sensors

For the foreseeable future, CCD and CMOS sensors will play both a significant role in imaging procedures. The long experience with CCD sensor production and the associated reliability is one of the major advantages in comparison to CMOS for scientific applications. Nevertheless, CMOS technology gains growing interests for scientific applications with developments driven by the consumer market. The following attributes are used to differ between the characteristics of both images sensors [18] [19].

Dynamic Range

The ratio between the saturation of a pixel to its threshold to detect photons is called the dynamic range (DNR). Saturation is reached when additional light will not cause more charges. The DNR is often represented in decibel. A low threshold is only possible with a low Signal-to-Noise Ratio (SNR). It describes the power ratio between the signal that needs to be measured and the unwanted background noise [20].

$$DNR = 20 \cdot \log \frac{P_{saturation}}{P_{threshold}} \quad (15)$$

$$SNR = \frac{P_{signal}}{P_{noise}} \quad (16)$$

The human eye has its highest dynamic range in the green spectrum. It enables us to see very small differences in color and brightness in this range. The contrast radio describes the same property for display systems. A high contrast radio is desired for most displays. Small shades between colors can be distinguished better.

In the past CCDs enjoy significant noise advantages over CMOS because of less circuits on the chip that interfere the measurement. ADCs and other components are outsourced from the image sensor. This issue has been solved by the development of modern CMOS and the possibility to integrate the signal processing at the sides of image sensor, which has substantially damped many noise sources and improved CMOS performance.

Uniformity

Ideally, every pixel should behave in the same way under identical illumination conditions. In practice, inaccuracies in the semiconductor production cause small variance in the structure of the photo diodes. This problem occurs with CCD and CMOS. CCDs benefit from the

single amplifier, that every charge goes through. As a conclusion all values are increased with the same gain. In CMOS variations between the amplifier causes a higher uniformity. This is a significant issue in high-speed applications with a limited SNR. The short exposure time gathers only small amounts of charges. Consequently uniformity between the pixels will emerge under the short exposure time.

Speed

The highly integrated CMOS technology makes this type of sensor much faster than CCD. Capacitance and propagation delays are shorter, because the circuitry is integrated into the chip. The charge transfer time is much shorter and the ADC processing can be done directly in the image sensor. This results in much higher speeds with CMOS. Images can be made in a very short distance to each other. Therefore, CMOS allows higher frames-per-second (fps).

Windowing

One limitation for the maximum speed of a camera is the ability to transfer and save all captured images. CMOS sensors allow the independent operation of parts on the image sensor. The number of working pixels is decreased to increase the maximum possible speed between every capture. This function is called region-of-interest (ROI) or windowing in modern industrial cameras.

Antiblooming

Blooming is the unwanted movement from charges between pixel in saturation. CMOS is immune for this effect, because every pixel is separated from each other. Modern CCD reduce this effect with the help of Anti-Blooming Gates. Unfortunately, this limits the size of the pixel and therefore the dynamic range.

2.5.4 Rolling and Global Shutter

In traditional analog photo cameras, the shutter speed is the time in which the film is exposed to the light. The shutter opens and closes mechanically. The time in between is also called exposure time t_e . Today this is done without any mechanic in a digital camera. The sensor starts, stops and resets based on a digital control signal. Two different types of electronic shutters can be distinguished. The rolling shutter is used in most CMOS consumer cameras. The rows of the image sensor start their measurement shifted to each other. This causes the rolling shutter effect when fast moving scenes are measured in the order of the exposure time or smaller. The time delay between the lines can produce a blur

in the picture. This kind of shutter is not desired for scientific and industrial investigations of fast moving objects. Figure 15 shows the timing of a rolling shutter image sensor.

The global shutter is used in most scientific and industrial cameras. Every row of the image is exposed at the same time without any phase shifting. In the past, this was one of the main benefits of CCD cameras. New developments allow the combination between global shutter and CMOS. Figure 16 shows the timing of a CMOS global shutter camera. Images can be captured without geometric distortions. A longer readout time is necessary for the processing of charges in parallel. The camera is "blind" in this time slot for incident light. A flashlight synchronized to the camera trigger is a indicator that is commonly used to investigate the shutter of the camera. A trigger input or output at the camera allows the synchronization of the image exposure to the flash or other devices. The camera and the lighting can be shifted to each other to evaluate the timing of the reset, exposure, readout and charge transfer. For the synchronization between multiple cameras, the trigger channel enables the fixed phase relation between images from different perspectives. This is especially necessary when dynamic processes are investigated. The exposure time t_e must always be smaller than the frame periode. Consequently, the exposure time must be reduced with higher frame rates (fps):

$$t_e < \frac{1}{fps} \quad (17)$$

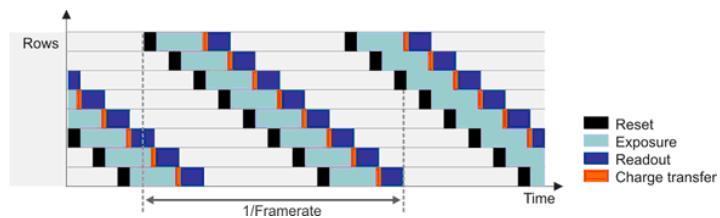


Figure 15: Timing of rolling shutter CMOS Sensor <http://www.cihansari.com/>

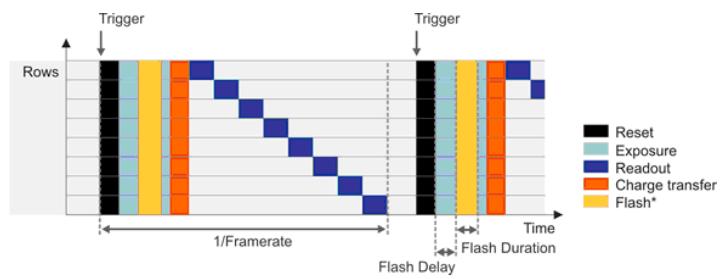


Figure 16: Timing of triggered global shutter CMOS sensor <http://www.cihansari.com/>

2.5.5 Camera Optic

The image sensor without a lens in front of it is not able to capture a sharp image. The ambient light from different angles will overlap on multiple pixels. That is why a lens is needed to direct the path of the light rays as parallel as possible on the sensor. Every point of the scene should ideally be represented on the image sensor only on one position. For different distances this is not possible with a fixed lens position. A focus is needed to change the plane of the lens and therefore the focal length F . As a consequence the plane in which the scene is captured sharply is moved. Figure 17 shows some basic dimensions of the camera optic. With a higher maximum F , sharp pictures can be made from a long distance d . The angle of view θ is reduced in contrast. The focal length is a scale for the optical power of a lens and to measure how strong the rays are bended. The optical axis stands normal to the middle of the image plane and point into the middle of the investigated scene.

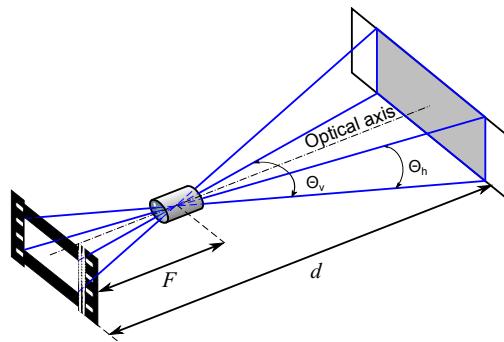


Figure 17: Focal length in the camera optic Christophe Dang Ngoc Chan 2007 wikipedia.org ©○○

The field of view (FOV) is the image section of the scene that is observed. In an optical instrument it describes the solid angle through which a detector is sensitive to incident electromagnetic radiation. The aspect ratio of the resulting images corresponds to the aspect ratio of the field of view. It can be calculated by the vertical and horizontal angle of view $\theta_{v,h}$, as described by equation 18.

$$\text{Aspect Ratio} = \frac{\tan(\theta_v)}{\tan(\theta_h)} \quad (18)$$

The lens speed N is the ratio between the maximum possible aperture and the focal length of the objective. The illuminance, that can be captured by the lens, increases with the

square of the aperture. A lens with a larger maximum aperture is called a fast lens because it delivers more light in the same exposure time.

$$N = \frac{F}{D} \quad (19)$$

2.5.6 Motion Blur

Motion blur occurs when the sensor or an object in the scene moves within the exposure time of an image. One observed point will be projected to multiple positions on the image sensor. This can be avoided with a very short exposure time t_e or with a stroboscope in a dark room that controls the amount of light in the scene. The movement \vec{s}_{blur} of every point inside a picture can easily be calculated when the velocity \vec{v}_p parallel to the picture plane is known and constant.

$$\vec{s}_{blur} = \vec{v}_p \cdot t_e \quad [m] \quad (20)$$

Figure 18 shows the motion blur on a propeller. The speed on every point on the blades \vec{v} can be calculated with angular velocity $\vec{\omega}$ and the radius \vec{r} to the axis of rotation. The velocity \vec{v}_p is the projection of \vec{v} on the image plane. The blur increases linear with the radius to the axis.

$$\vec{v} = \frac{d\vec{r}(t)}{dt} = \vec{\omega} \times \vec{r} \quad [ms^{-1}] \quad (21)$$

The scalar angular velocity can be calculated by the rotational speed n . With the axis of rotation as z the vectorial angular velocity can be derived. This can also be used to calculate the rotational speed from a blurred picture with known exposure time and geometry of the rotor.



Figure 18: Motion blur of a fast moving rotor Tech. Sgt. Ben Bloker wikipedia.org

2.6 Range Imaging Techniques

Multiple 3D scanning approaches are used, to derive three-dimensional data from the real world. In this section, a comparison of different range imaging techniques and their characteristics is done. They all produce 2D images, that represents the distance to every point in the scene, a so called range or depth image. This data can be transformed into a 3D space, when the system is calibrated adequately to the field of view, which depends on the focal length or freedom of movement of the laser. Various methods have been developed, to determine the distance to the image sensor.

2.6.1 Stereo Triangulation

The position of a point in space can be reconstructed from the perspective of two or more cameras. This requires a calibrated position and field of view of the cameras to each other. With two images of the same scene, taken from different points of view, the correspondence problem has to be solved. It describes the challenge of finding a set of points in one image, which can be identified as the same points in the other image from a different point of view. Two basic methods are distinguished:

Correlation-based – Check if identical location can be found in both images.

Feature-based – Find multiple features in the images and see if a layout of subsets is similar.

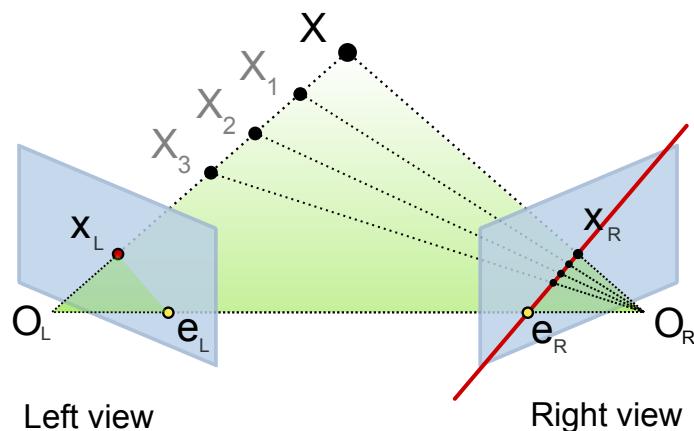


Figure 19: Basic principles of stereo triangulation Arne Nordmann 2007 ©©©

Markers are easy additives to identify single points from multiple perspectives. Figure 19 shows one scene that is observed from two positions. The point X is projected as X_R

and X_L on both image sensors. The epipolar line between O_L and O_R is the base for the triangulation [21]. The position of the point X cannot be distinguished in the depth with just the left image. With the second image, the range of every point can be calculated. The nearer the objects are to the image sensor, the higher the displacement of the object between the two images is. Therefore, the triangulation is simplified and more accurate. The human eyes work by the same principle to determine the depth in the human field of view. The basic principle is also called stereo vision and allows the measurement of distances down to 10^{-7} m [1].

2.6.2 Structured Light Sensor

The projection of a structured light pattern is an established method to derive the depth of a scene into the room and to analyze it by a camera. The orientation of the camera and the lighting projection must be known to reconstruct the 3D surface from the 2D picture. The light pattern is created by interfering of two coherent laser beams or by a display that is projected. From the perspective of the camera, the pattern is deformed by the 3D surface. The new distribution is the base for the depth reconstruction. In most technical implementations, the wavelength of the light can be found in the IR spectrum. The human eye cannot see the structure. The measurement will be disturbed by surfaces that absorb or spectral reflect. Ideally, the light reflects in a diffuse way from the surface. Figure 20 is an image of the point structure that is projected on a model aircraft with a Kinect 1 camera. From that, the raw uncalibrated depth image is calculated where every pixel corresponds to a value of one depth measurement. The vertical stabilizer at the rear end cannot be acquired, because it lies under the minimal distance of about 0.8 m . A hole seems to appear in this region. One major disadvantage is the low range of structured light sensors. Thus, most implementations are in the close-up range for applications, like gesture recognition. An enhancement of this method is the Programmable Structured Light (DLP). The light pattern can be adapted onto the target object to capture physical measurements, analyze locations or inspect a surface.

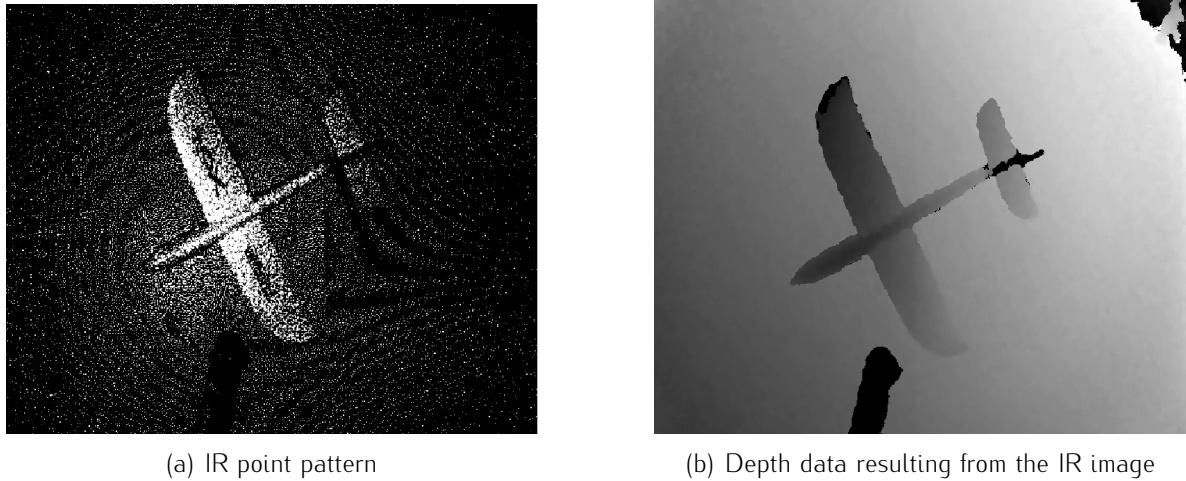


Figure 20: Kinect 1 structured light

2.6.3 Interferometry

The Doppler effect is the change in frequency that occurs when the observer moves relative to the wave direction. In the interferometry this effect is used to measure the absolute velocity v of an objects relative to stationary coherent light sources with a constant frequency f_E . The interference between the reflected and the emitted light results in a new waveform. The interfering light has the new frequency f_D that is proportional to the speed v of the reflecting body. A photo diode measures the sum of the spectral components f and can therefore calculate the frequency shift [22].

$$f = f_D \pm \frac{1}{\lambda} \quad (22)$$

$$f_D = \frac{2 \cdot |v|}{\lambda} \quad (23)$$

Laser Doppler vibrometer is an industrial standard measurement approach that uses a single laser beam to derive the velocity of vibrating point on a surface. Small fluctuations in the wavelength of the laser light would decrease the accuracy. Therefore high stable light sources are necessary. The measurement returns the velocity in a single pixel on one point in space. The absolute distance to the objects can not be derived. The integration of the speed over time results in displacement. Thus, only static scenes can be investigated. For the measurement of complete surfaces, the laser beam is adjusted by a mirror. The displacement of a surface cannot be measured in parallel. Every point is acquired with a

time delay. The interferometry also allows to determine depth, directly using a technique called phase-unwrapping. The earth surface is scanned with this technique, known as terrestrial SAR interferometry. The resolution of such a system is the range of 10^{-9} mm under ideal conditions.

2.6.4 Time-of-Flight

The Time-of-Flight (ToF) principle describes a variety of methods that measure the time that a wave or a particle needs to travel an unknown distance. With the constant speed of light c and the measured time of flight t_{ToF} , the way that a photon has traveled can be calculated. The established Light Detection and Ranging (LiDAR) 3D scanner uses this technique to measure the distance d to an object with a single laser beam. A laser diode emits a movable beam, which is reflected backwards and measured within a photo diode. The light travels two times the distance d back and forward:

$$d = \frac{c \cdot t_{ToF}}{2} \quad (24)$$

The measurement of an absolute distance that can be derived to velocity and acceleration without further information is one advantage of the Time-of-Flight method in comparison to the interferometry. The resulting 3D images are merged together by the sum of pixels that are captured in series over the field of view. The laser beam is readjusted by a movable mirror for every captured pixel. This allows the scan of a complete static scene. Since the image cannot be scanned in parallel but in series, movement between individual measurements results in errors that can be compared to the rolling shutter effects of image sensors.

New IR global shutter CMOS sensors, in contrast, are able to acquire the total reflecting light in parallel on multiple pixels. Most ToF Cameras use near-infrared ($\approx 850 \text{ nm}$) LEDs or laser diodes that illuminate the scene. The incident power density falls off with the square of distance according to equation 5. The phase shift between the illumination and the reflection is measured and translated into a distance. Every measurement delivers up to three images: The depth, the intensity and in some cases the amplitude of the reflection. Two different methods can be found to detect the time between the emitted and the reflected light. The pulsed method uses a single amount of light that is reflected in the room. The emitter illuminates for a brief moment Δt and the reflecting light is sampled in parallel in

every pixel in two out-of-phase windows over the same Δt . Both electrical charges Q_1 and Q_2 are used to compute the distance:

$$d = \frac{1}{2} \cdot c \cdot \Delta t \cdot \frac{Q_1}{Q_1 + Q_2} \quad (25)$$

The continuous-wave (CW) approach modulates a square or sinus function on the light intensity. Very sensitive single-photon avalanche diodes are able to measure the phase shift φ for every pixel with high speed counters. In contrast to the pulsed approach, the waves are measured over multiple periods to derive the phase shift.

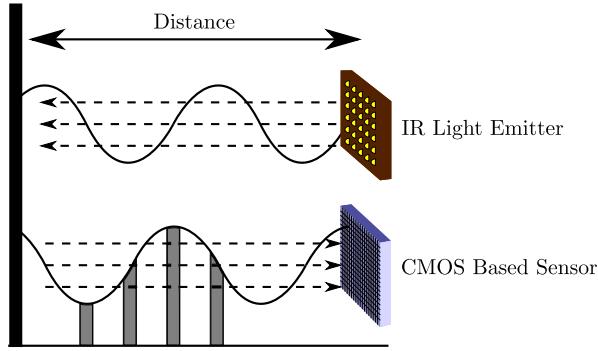


Figure 21: Principle of Continuous-wave Time-of-Flight [23]

The time window for the CW measurement is the integration time. It is comparable to the exposure time of an ordinary image sensor. Figure 21 illustrates a radiated and reflected square wave IR signal and four phase shifted time windows. Q_1 to Q_4 represent the amount of charges that are collected in every window. In contrast to the pulsed control, the ambient light can be canceled out by measuring it out of the phase to the IR Signal [2].

$$\varphi = \arctan \frac{Q_3 - Q_4}{Q_1 - Q_2} \quad (26)$$

$$d = \frac{c}{4 \cdot \pi \cdot f} \cdot \varphi \quad (27)$$

The reflected amplitude A and the ambient light offset B are proportional to the depth measurement variance σ^2 :

$$\sigma^2 = \frac{c}{4 \cdot \sqrt{2} \cdot \pi \cdot f} \cdot \frac{\sqrt{A + B}}{c_d \cdot A} \quad (28)$$

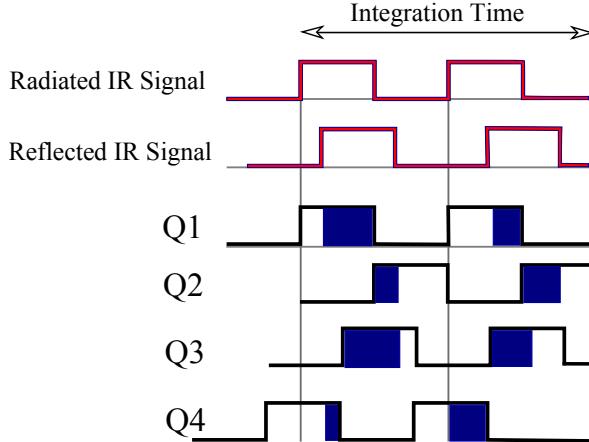


Figure 22: Measuring windows of continuous-wave ToF

Equation 28 shows that the variance can be reduced with a higher modulation frequency f_{mod} and amplitudes A. The constant c_d is a indicator, how well the ToF sensor separates and collects the photoelectrons. The amplitude can be increased with shorter distances to the scene or with a higher radiant intensity. One disadvantage of the CW measurement is that the phase warps around every 2π . The ambiguity results in multiple possible distances. The intensity of the reflecting light helps to identify the true distance. This limits the range that can be observed. A reduction in the modulation frequency would increase the ambiguity distance but reduce the accuracy at the same time. With the maximum phase shift angle $\varphi = 2 \cdot \pi$ and equation 27, the ambiguity distance can be calculated by the following equation:

$$d_{amb} = \frac{c}{2 \cdot f_{mod}} \quad (29)$$

Ambient light in the same IR spectral region disturbs and results in noise. A monochromatic filter in front of the lens can reduce the incident light to the wavelength of the emitting diodes. Unfortunately, the sunlight emits in a broad bandwidth that cannot be filtered out completely.

In comparison to the other range imaging techniques, ToF has the advantage of a small size, no moving parts, a high range and a very fast scanning speed. The only disadvantage is the low depth accuracy of 1 mm. New senor developments could increase the resolution significantly down to 0.3 mm [1].

2.7 Time-of-Flight Imaging Errors

Depth measurement with ToF cameras faces the appearance of both, systematic and non-systematic errors. Generally, systematic errors can be reduced by a calibration [23]. This section gives a rough overview of different errors sources and their reasons.

2.7.1 Systematic Errors

Range Distortion

The IR light can not be generated in an ideal sinusodial or square waveform, as a consequence range distortion appears. Usually, the error plotted against the distance follows a sinusoidal shape that can be compensated by a calibration.

Integration-Time-Related Error

Scientific ToF cameras allow to adjust the integration time for the depth measurement. For the same scene, different values cause changes in the results. The precision versus different depths depends on the integration time.

Build in Pixel-Related Errors

Inaccuracies in the production of the image sensor causes error in the depth measurement. The diodes or capacitors differ in the responsibility and latency-related offset error. This type of errors are related to the position of the pixel on the sensor array. A Fixed Pattern Noise (FPN) table can be created to obtain the computed depth with a reference distance.

Amplitude-Related Errors

Depth accuracy is highly related to the amount of incident light. High reflected amplitudes result in a high accuracy. Low amplitudes appear more often at the border of the image as the emitted light power is lower. When the object is too close to the camera or the integration time has been chosen to high, saturation appears and the depth measurement will be invalid.

Temperature-Related Errors

Depth values suffer from the operation temperature of the sensor. After the activation of the camera it is heating up, until it reaches a stable temperature. In this phase, the depth error changes versus time. The camera should be turned on several minutes with activated illumination before the usage to achieve an precise measurement.

2.7.2 Non-systematic Errors

Low Illuminated Regions

Scenes that are not illuminated uniformly will suffer from regions with low illumination. This causes a high amount of noise and a reduction in the signal-to-noise ratio. The amplitude image can be used to filter out these kinds of pixels.

Multiple Light Reflection Error

The interference of different light reflection in a single pixel occurs depending on the shape of the scene. Surface edges and concavity reflect the wave in multiple ways. The light travels in successive paths before it reaches the sensor. Therefore, a depth measurement is corrupted. Flying pixels on sharp edges are one artifact that is observed in all ToF 3D scans. Light from the background interferes with the reflection on the edge in the foreground. This result is a noisy pixel that seems to jump in space and cannot be allocated to a connected surface.

Light Scattering

Reflections with high amplitudes from nearby objects can cause light scattering between the image sensor and its lens. This influences the measurement of other nearby pixels and results in noise. New sensor material with low reflectivity could make scattering negligible in the future.

Motion Blur

As with traditional cameras, the capture of dynamic scenes results in motion blur, depending on the integration time and the speed of the observed object. The error can be classified in two different types of artifacts: Lateral and axial blur.

Flying Pixels

False distance information can also occur due to the pixels relative large solid angle. Different distances inside a solid angle lead to superimposed reflected light reducing its amplitude and introducing a false phase shift. These pixels are commonly known as flying pixels, because they lie in-between the fore-and background.

2.8 Acquisition of Digital Dynamic Data

The discretization of an analog signal, like from a pixel, is done in the analog to digital converter (ADC). The signal value will be divided into single samples depending on the bit depth n_{Bits} as shown in figure 23 with 3 Bits $\equiv 2^3$ steps. The sampling is controlled by a digital signal with the frequency $f_{sampling}$ that determines the conversion from the analog to the digital value. The Nyquist-Shannon theorem in equation 30 proves, that the minimal possible sampling frequency $f_{sampling}$ must be double the maximum frequencies f_{max} of the signal to preserve the spectrum. For the estimation of amplitude and slope a factor of two is not enough. A variety of samples per period is required, depending on the shape of the signal, to ensure a sample as near as possible to the peak of the deflection. High resolutions and sampling frequency mean less quantization errors. Therefore, differences between the digital and analog signal gets smaller. The averaging of multiple samples is a common way, used to reduce the noise and the data stream of the acquisition.

$$f_{sampling} \geq f_{nyquist} = f_{max} \cdot 2 \quad (30)$$

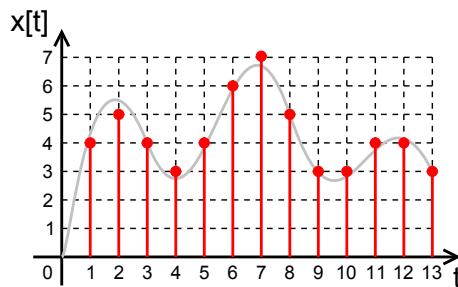


Figure 23: Illustration of the analog to digital conversion wdwd wikipedia.org

Higher frequencies above f_{max} , resulting from noise or other sources, still have an undesirable impact on lower spectral components. These effects are called aliasing artifacts and lead to artificial frequency components. A low-pass filter applied on the analog signal with a cutoff frequency f_{max} can reduce this phenomenon. Higher frequencies than $f_{max} = f_{sampling} / 2$ will not arrive at the ADC input.

An aperture error results from the fact that the sample is obtained as a time average within a sampling region and not at an instant moment. The effects are corresponding to the motion blur of image sensors. Oversampling is the process of sampling with frequencies much higher than $f_{nyquist}$, leading to an improved resolution and less quantization noise. For an accurate measurement of the amplitude, a minimum sampling frequency of $f_{sampling} = f_{nyquist} \cdot 5$ is

used in practice, to reduce influences like aperture errors or aliasing. The factor between the $f_{Nyquist}$ and $f_{Sampling}$ dimension is called Oversampling Factor N. The data stream of a ADC converter can be calculated by the sampling frequency $f_{Sampling}$ and the digital resolution n_{Bits} by the following equation:

$$\dot{n} = n_{Bits} \cdot f_{Sampling} \left[\frac{bit}{s} \right] \quad (31)$$

2.9 Evaluation of Dynamic Data

2.9.1 Time Domain

The analysis of mathematical function with respect to time is called time domain. The time is represented in form of real numbers, for the case of continuous time or in discrete number resulting in a discrete time. An oscilloscope is a common tool for the investigation of real-world signals in the time domain. The graph describes how a value changes with time, whereas a frequency-domain graph describes how much of the signal amplitude lies within a given frequency band versus a range of frequencies. Certain terms are used for the description of the time domain. Figure 24 shows different properties of a sinus oscillation.

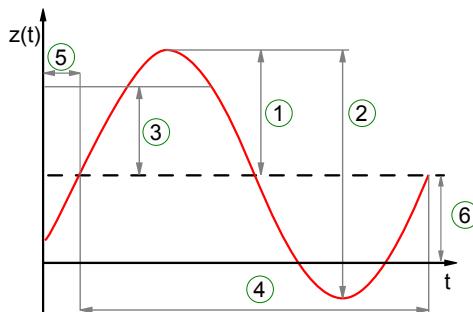


Figure 24: Properties of sinus oscillation Saure wikipedia.org

1. Peak amplitude A_p – Maximum absolute value from zero to highest values
2. Peak-to-peak amplitude $A_{pp} = A_p \cdot 2$ – Change between highest value and lowest amplitude value
3. Root mean square amplitude $A_{rms} = A_p / \sqrt{2}$ – The square root of the mean versus time of the square in the vertical distance of the graph
4. Wave Period $T = \frac{1}{f_p} = \frac{2\pi}{\omega}$ – Time difference between one periodic oscillation
5. Phase Shift φ – Time difference between the $t = 0$ and the zero crossing of the oscillation or another reference time

6. Offset Z_{DC} - Constant displacement from the origin to the equilibrium

An undamped sinus oscillation can be described in two mathematical forms: One is the common analytic sinus form as a function of time t . The value Z_{DC} is a static offset in the coordinate system that superimposes the dynamic by a constant static value. In this case, the oscillation is not around the point of origin but around the equilibrium.

$$z(t) = A_p \cdot \sin(2 \cdot \pi \cdot f \cdot t + \varphi) + Z_{DC} \quad (32)$$

The oscillation can also be represented with the help of the Euler equations and imaginary numbers. The imaginary part has no physical meaning, but it simplifies the processing of a dynamic function.

$$\bar{z}(t) = A \cdot e^{j(\omega \cdot t + \varphi)} + Z_{DC} \quad (33)$$

The energy keeps constant at a free undamped sinus oscillation:

$$E = \frac{1}{2} m \omega^2 A^2 \quad (34)$$

2.9.2 Frequency Domain

Every steady periodic function can be transformed into a sum of sinus oscillations, by a so called Fourier transformation. This allows the investigation and the comparison of dynamic signals. To do so, the data is transformed from the time domain $z(t)$ into the frequency domain $|Y(f)|$. The signals can now be investigated without a common time base. The frequency domain is represented by a spectrum that shows the individual amplitudes components versus the corresponding frequencies. This can be especially useful in the analysis of vibrations of structures. Undesired eigenmotions can be compared and analyzed. In figure 25, the waveform of equation 35 is illustrated, which is superimposed by some random noise.

$$z(t) = 0.7 \cdot \sin(2 \cdot \pi \cdot 50 \text{ Hz} \cdot t) + 1.4 \cdot \sin(2 \cdot \pi \cdot 120 \text{ Hz} \cdot t) + y_{noise} \quad (35)$$

The two sinus wave forms and the noise are difficult to separate in the time domain. In the frequency domain, in contrast, noise and the original function can be distinguished. The two peaks indicate that the original function consists of two sinus functions with 50 Hz and 120 Hz. The noise floor lies underneath with multiple random peaks in the whole spectrum. The amplitudes from the original signal converge with the peaks in the frequency domain over a longer timespan and a higher $f_{sampling}$. More detailed information on this example can be found in the official Matlab documentation of the `fft()` function [24].

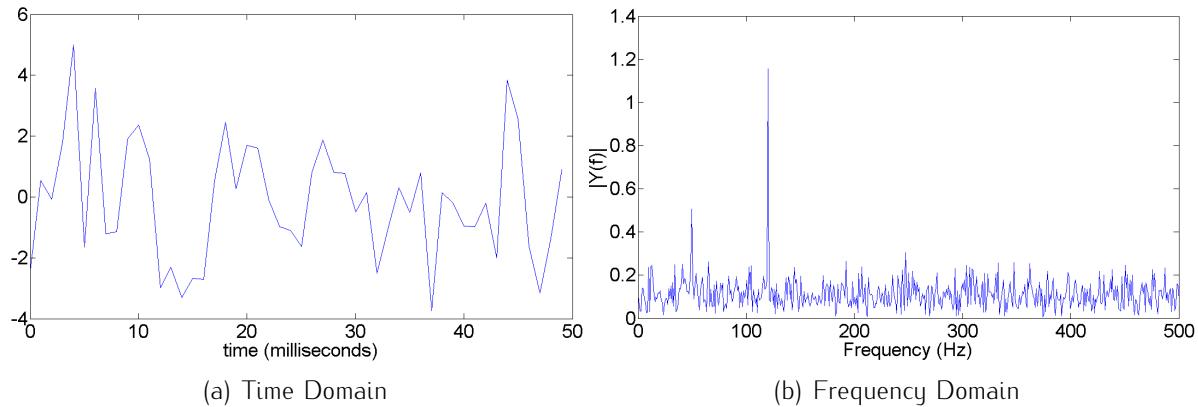


Figure 25: Fast Fourier Transformation in Matlab

The Fast Fourier transform (FFT) is a widely used algorithm to convert a digital signal from one domain into another. The input data must be in a discrete digital form with 2^n samples. In this way, the processing power is much lower compared to the original Discrete Fourier Transform (DFT). The signal is transformed into its complex components. The real parts of the frequencies represents half of the sinus function amplitude. $f_{sampling}$ of the computed signal must be known to calculate the right frequencies out of the raw data. The following code is taken from the `fft()` example and plots the frequency domain out of the raw data array `y` with the number of elements `L`:

```
NFFT = 2^nextpow2(L); % Next power of 2 from length of y
Y = fft(y,NFFT)/L;
f = Fs/2*linspace(0,1,NFFT/2+1);
plot(f,2*abs(Y(1:NFFT/2+1)))
```

`NFFT` contains the number of samples that are processed. `Y` is half of the amplitudes while `f` provides the corresponding frequency array that goes from zero to $f_{nyquist}$. These data are plotted against each other. The single-sided amplitude is two times the real part of the complex representation of the amplitude components from `Y`.

The original digital signal can be reproduced from the frequency domain into the time domain without any loss through the inverse fast Fourier transform. This is useful for the filtering of data. Undesired frequency components can be identified and removed from the data. When the FFT of a non-periodic signal is computed, the resulting spectrum suffers from leakage. This can occur when only parts of a period are measured at the start or the end of the acquisition. The variance of the spectrum increases. Different FFT windows are used to reduce the influence of signal at the beginning and ending, like the hamming window.

window [25]. When measuring a self-windowing signal, like an impulse, a shock response, a sine burst or a noise burst a window should not be used. Such signals are used in modal analysis. Applying a window function would just deteriorate the signal-to-noise ratio.

The Short-time Fourier Transform (STFT) is used to determine the sinusoidal frequency and phase content of local sections of a signal as it changes versus time. Therefore variation versus time can be analyzed better than with the FFT of the complete signal. The Matlab function spectrogram() delivers a way to create a STFT out of a signal. The trend of amplitude and frequency is plotted versus time in intervals that are represented by NFFT as a number of samples for a separated Fourier Transform. The overlapping factor describes the range between two windows that is analyzed by both. The following source code represents an example for the creation of a spectrogram in Matlab:

```

Fs = 30; %Samplingrate
amp=2;%peak-Wert
zeitmax=5;
zeit=linspace(0,zeitmax,zeitmax*Fs);
daten=amp*sin(2*pi*zeit*frequenz)
nfft = 32;%Blockgroesse 2er-Potenz
window = hanning(nfft);
% z.B. 80 fuer 80% overlap
Prozent_overlap = 5;
overlap = round(nfft/100)*Prozent_overlap;
[B,F,T] = spectrogram(daten,window,overlap,nfft,Fs)
figure
imagesc(T,F,abs(B));
colormap(jet);
colorbar
axis xy
ylim([0 30])

```

3 Time-of-Flight Market Research

Since the semiconductors technology became fast enough for the development of ToF cameras in the year 2000, multiple companies developed image sensors for civil applications. The following section shall give an overview of the ToF cameras that are currently on the market or will soon be released. Detailed specifications like resolution, size, accuracy, hardware/software trigger and integration times can be found in the official associated datasheets. In the following table specifications of sample image sensors and complete camera systems from the industrial and consumer market are gathered:

	Pixel Array	Speed [fps]	FOV [°]	Range [m]	Size [mm]
Mesa SR4000	176x144	30	69x55	8	65x65x76
pmdtec CamBoardpico	160x120	45	82x66	1	89x17x6
Fotonic E70 16W	160x120	52	70x53	30	80x80x86
Argos 3D - P100	160x120	160	90x90	3	27x75x57
odos real.iZ-1K	1280x1024	100	not fixed	10	182x80x92
Basler ToF-6c	640x480	30	48x38	5	-
Microsoft Kinect 2	512x424	30	70x60	8	249x66x67
Polytec epc660 Sensor	320x240	264	-	240	-

Table 3: Time-of-Flight camera systems

The Argos 3D - P100 camera offers a high frame rate of 160 fps at a low depth resolution and number of pixels. It also has a hardware trigger which makes it useful for synchronization with other devices. Multiple sensors can be synchronized phase shifted to each other. The interaction of the IR light between different sensors is therefore prevented. The Real.iZ-1K brings the highest fps with the largest number of pixels. In contrast to the other system the pulsed wave ToF approach is used. Additional illuminators can be added to the camera to increase the accuracy at high distances. The standard C mount allows the exchange of the lens. Therefore, the field of view can be adjusted. The high flexibility of this system requires a calibration before it can be used. The output data must be preprocessed by a calibration matrix to obtain depth data. A trigger input allows the synchronization with other systems. The Fotonic E70 shows good motion blur characteristics and is the only

sensor based on CCD technology. It is designed for rough environments and shows good outdoor characteristics. The company pmdtec offers ToF cameras, such as the highly compact Camboard Pico and the Camboard Nano, where the lenses and illuminators are fixed. Since there is no hardware trigger, a synchronization frame by frame is not possible. Mesa SR4000 is one of the pioneer in ToF technology. Lots of scientific work has been published based on this system. Basler is a new player in this sector, offering the first prototype with a high number of pixels at 30 fps and an external trigger. The Development of image sensors with 1280x1024 is already announced. Microsoft brings ToF with the Kinect 2 into the consumer market. Compared to the other industrial system, the price is very low with 200 \$ for a complete development kit. Nevertheless the resolution, range and accuracy can still keep up. A higher frame rate and range has been implied [26] [4]. A disadvantage is the absence of raw data access and low level configurations. Polytec makes a huge advantage with the image sensor epc660. 264 fps at 320x240 pixels should allow the investigation of fast dynamic scenes. A region of interest enables 4000 fps at 160x30 pixels and therefore various applications in the investigation of fast unsteady deformations. Currently there is no operational camera on the market with this sensor. Overall, the ToF market is rapidly growing. The cancelling of ambient light will allow outdoor applications soon, since companies like pmdtec are interested in outdoor application in the automotive industry. The increase in range, frame rate, depth resolution and number of pixels over the last few years enables many of industrial, civil and scientific applications.



Figure 26: odos real.iZ-1K 1280x1024 with changeable lens and illuminator

3.1 Kinect 2 System Overview

The Kinect 2 is the first Time-of-Flight consumer camera and brings the highest total amount of pixels compared to other ToF devices at the time of the release. It is originally designed for gesture recognition and enables the control of games and other applications by hand or body movements. This thesis concentrates on the ToF camera even if other sensors are included, like a microphone array and a HD CMOS rolling shutter camera for visible light. In contrast to the Kinect 1, the developer decided to change the technique of the range imaging from a structured light sensor to Time-of-Flight. This increases the accuracy, field of view, resolution, number of pixels and the maximum distance. Thus movements of multiple persons can now be captured in parallel and applications can be controlled by the users even by figure gestures. No official technical specifications of the sensor have been released yet. In Payne's publication, an image sensor is presented that probably is the one, which is integrated in the camera [26]. The following table represents the basic specifications from the publication:

Pixel Pitch	$10 \mu m * 10 \mu m$
Chip Size	$8.2 mm * 14.2 mm$
Pixel Array	512x424
Dynamic Range	$> 2500 = 68 db$
Modulation Frequency	$10 - 130 MHz$
Average Modulation Frequency	$80 MHz$
Field of View	$70 (H) \times 60 (V)$ degrees
Depth Uncertainty	<0.5 % of range
Operating Wavelength	$860 nm$
Frame Rate	max. $60 fps$ (typical $30 fps$)
ADC Data Stream	$2 GS/s$
Effective Fill Factor	60%
Reflectivity	$15\% - 95\%$
Responsivity @ $860 nm$	$0.144 A/W$
Readout Noise	$320 \mu V$ differential
ADC Resolution	$10 Bits$

Table 4: Kinect 2 system specifications

The illumination is based on three 860 nm IR laser diodes. A rough measurement of the emitted light with a single IR photo diode and an oscilloscope showed that the diodes flash with a rate of 30 Hz. The frame rate of the depth video stream also corresponds to 30 fps, visible in figure 79 and 80. Every IR flash period comes with a burst of three different intensities and frequencies of 120 MHz, 80 MHz and 16 MHz for a time of approximately 5.4 ms resulting in one depth frame [4]. After about 16 ms of IR light, the diodes are turned off for 16 ms and a new image is created. The captures are taken in rapid temporal succession to avoid motion blur. Every image results out of three different modulation frequencies to achieve the high dynamic range over large distance from 0.5 m to 8 m. Microsoft does not allow low level hardware access yet to change the modulation frequency or integration time. The sensor diodes feature a responsivity of 0.14 A/W at the laser wavelength. To increase the fill-factor to 60%, micro lenses are integrated into every pixel for a larger detecting area. With the given lens speed of 1.07, the focal length is approximately 80 mm with a aperture measured by hand of about 70 mm. Figure 27 shows a Kinect 2 disassembled. The three laser diodes are heavily shielded. The beam is scattered and filtered by the optic in front of the diodes. The IR camera lens and the case are equipped with monochromatic filters. Disturbing IR light sources such as the sun are reduced in this way.

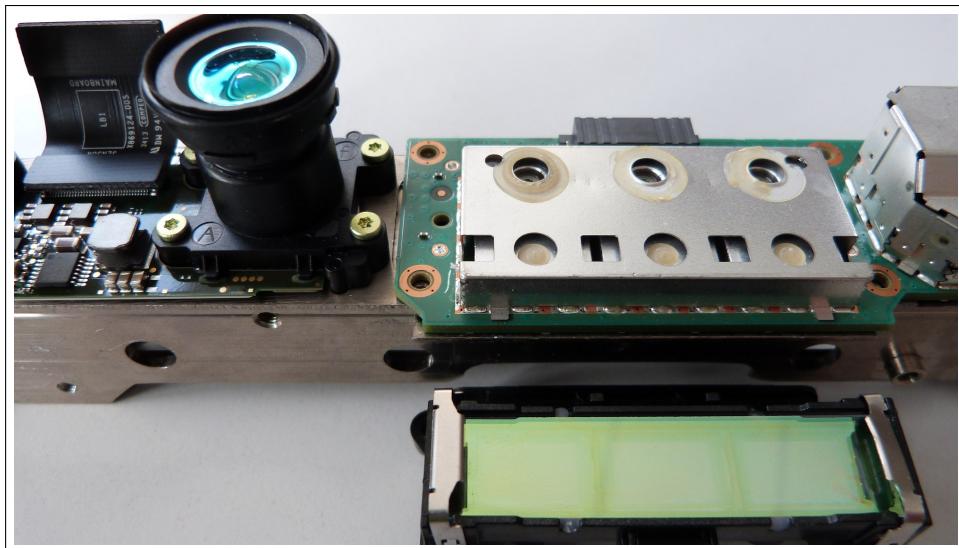


Figure 27: Kinect 2 IR camera and IR illumination with scattering optic

A high dynamic range is necessary to simultaneously render high-reflectivity objects near the camera and low-reflectivity objects far from the camera to provide the depth resolution for all pixels. Every pixel selects the best shutter time and amplifier (AMPs) gain for the ADCs by

itself. The digitized output from the ADCs is combined in the shutter engine. The MIPI interface provides the communication to other embedded systems. Figure 28 shows a microscope image of the Kinect 2 Image Sensor. The electronic for the processing of the data is located on the sides to avoid interferences with the image sensor, common on modern CMOS sensors.

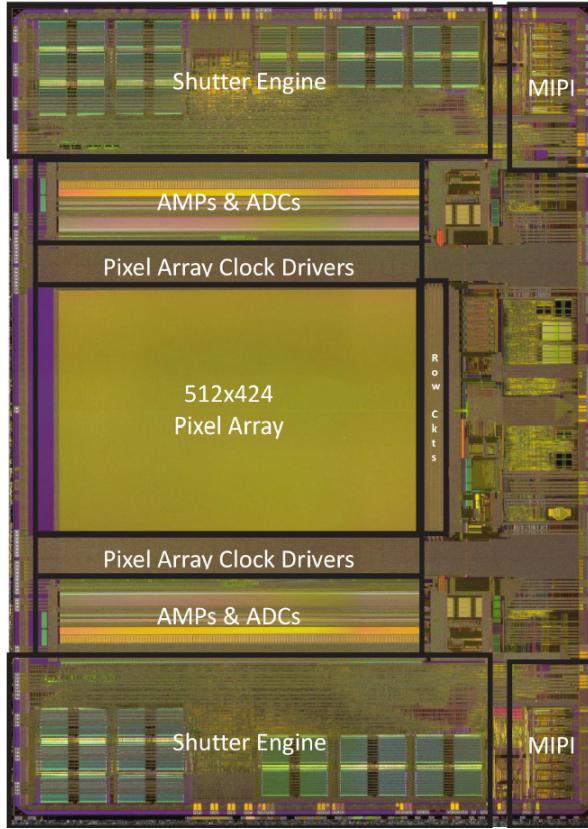


Figure 28: Kinect 2 image sensor [26]

A USB 3.0 connection allows the readout of the images with a latency less than 20 ms. A 12 V power supply delivers additional power to the camera when it is plugged to a computer. The data and the power supply are combined in a proprietary cable. Thus, the Kinect Hub is required to use a standard USB 3.0 plug for the PC connection. With the Microsoft Software Kinect 2 Studio, the data steam can be observed in a 3D point cloud but direct access to raw data is not possible in this application. Microsoft offers the Kinect for Windows Development Kit 2.0 to create customized programs for the processing of the Kinect 2 data stream. Several demonstration programs with source code in C++ and C# are included but raw data acquisition of a complete video stream is not provided yet. Wai Ho Li offers the first open source tool dumpK4W that saves the "raw" data from the video stream to a mass storage [27]. These images are given in an integer resolution

in millimeters and therefore the data stream must already be processed by the camera itself. Wai Ho Li uses the OpenCV library for real-time computer vision to save every picture from the data stream. The images cannot be processed in real-time for further investigations. A driver for the Matlab environment would be useful for a real-time processing.

The field of view of the IR camera is fixed to one lens position. Adjustment of the focal length is not possible. The advantage for the user is, that the system can be calibrated before it is shipped to the customer. A calibration before the operation of the camera is therefore not necessary. The range images are converted into millimeters inside the camera before they are transferred via USB. The area A that is observed by the sensor in a certain plain results out of the distance and the angle of view θ in the horizontal and vertical direction:

$$l_{h,v} = \tan \frac{\theta_{h,v}}{2} \cdot d \cdot 2 [m] \quad (36)$$

$$A = l_v \cdot l_h = \tan \frac{70^\circ}{2} \cdot \tan \frac{60^\circ}{2} \cdot d^2 \cdot 4 [m] \quad (37)$$

The area projected on a single quadratic pixel A_i is the fraction of one pixel to the complete area A . Figure 29 illustrates the field of view, the distance to the observed scene d , the coordinate system of the observed plane (index p) and the camera (index c).

$$A_i = \frac{A}{512 \cdot 424} [m^2] \quad (38)$$

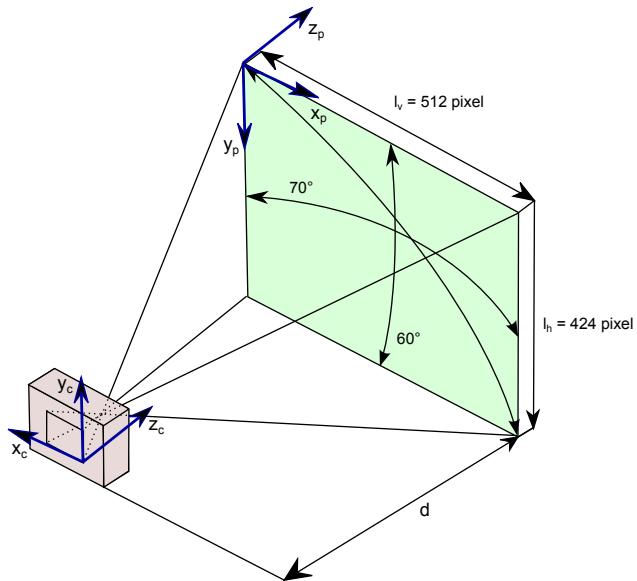


Figure 29: Kinect 2 field of view

4 Data Processing of Range Images

The video stream, saved with dumpK4W, consists of a bunch of individual pictures. A frame rate of 30 *fps* every second of acquisition will result in 60 images from the IR camera. One capture is represented in a depth and intensity image. The data is given in the Tagged Image File Format (TIFF) in 512x424 pixel with a non-standard aspect ratio. TIFF is a common format for the lossless storage of image data. The pixels consist of a 16 *Bit* integer number that represents the distance in millimeters to the image plane. About 13 bits are used corresponding to the 8000 *mm* of maximum distance. Values under 500 *mm* and beyond 8000 *mm* are set to zero. When opened in a standard image viewer, the files seem to be only black. The brightness and the contrast must be changed to foreground the shades of grey to distinguish different pixels. Matlab provides a set of tools for the processing of the raw data. The Matlab Image Processing Toolbox delivers a variety of ways for the processing of images and video streams. It is widely used in scientific environment and features offline and live processing of image data. Figure 30 represents a raw image that is plotted with the `imagesc()` function in Matlab.

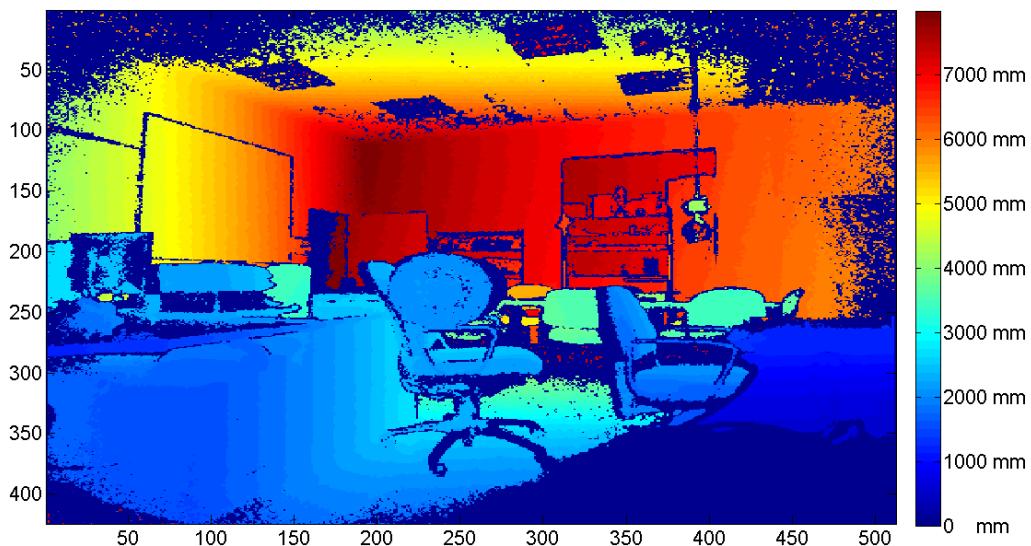


Figure 30: Raw image plotted with `imagesc()` function

The distance is represented in a red, green and blue colorspace that is fitted to the depth resolution. Red indicates far distances and blue the near fields. Isolines in the color space show equal regions of distance. Figure 30 shows, that the depth values are given perpendicular to the image plane. The isolines run parallel to the image plane, possible due to the fixed calibrated focal length and simplifies the processing of the data considerably. Noise is one of the major challenges in the interpretation of ToF Data. It can be considered as an unwanted fluctuation of the range measurement error. The noise increases at the corners of the FOV. This is caused by the optic of the camera and the light cone of the illumination. The lens reduces the light intensity at the outer boarders of the image. The threshold in this area can barely be archived and therefore the Signal-to-Noise ratio is reduced. The live display of the point cloud in Kinect Studio can be used for a rough estimation of the noise. The pixels seem to jump when observed live on a static surface. A detailed analysis of the noise characteristics can be found in section 6. The orientation of the raw data is flipped in the vertical axis, since it is given in the camera coordinate system. The Matlab command `fliplr()` transforms the image into the observed plane coordinate system.

4.1 Transformation into Point Cloud Data

The data must be transformed into a so called point cloud to gain 3D information out of the 2D data. The images are already given in a calibrated form happens inside the camera. Other camera devices with adjustable FOV, low level configuration and flexible illumination need a calibration of the raw images into the corresponding distance. Systematic errors are reduced in this way [28] [29]. Every pixel in the 512x424 array corresponds to a single point in 3D space, depending on the field of view and distance. Kinect Studio enables to rotating of live point cloud data. The regions which are not reached by the IR light, are highlighted in the form of shadows. Figure 31 shows one frame of the video stream that is shifted to the perspective of the camera. The person in the foreground shows an ideal diffuse reflection of the IR light and therefore a high signal-to-noise ratio. The transformation into a point cloud from raw data is also possible with the open source program MATLAB to Point Cloud Library (PCL) created by Peter Corke. It delivers a bundle of tools for visualization, surface reconstruction, filtering and segmentation. It is a standard environment for 3D image processing, that is developed by companies from different industries.

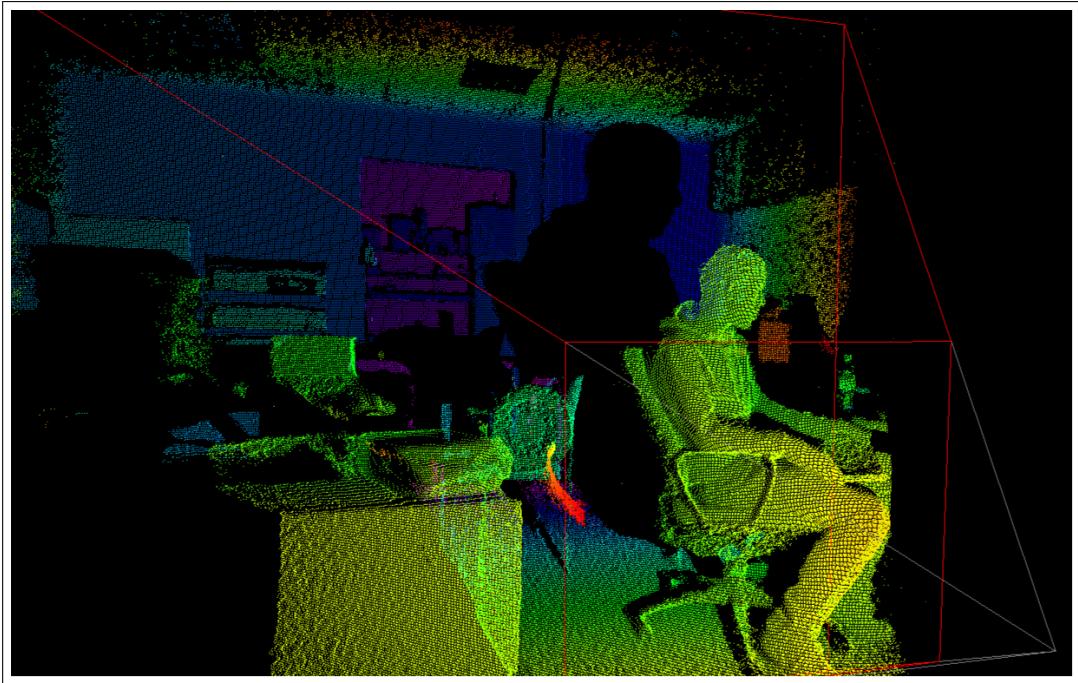


Figure 31: Point cloud in kinect studio

4.2 Meshing of Point Cloud Data

For the reconstitution of a scanned object, 3D points are not enough. The point cloud needs to be meshed. The vertices in the form of points represent knots for this connection. The lines in between are called edges. A face is a set of edges that can be combined into a closed surface. Figure 32 shows the transformation of a cube from vertices to triangle faces.

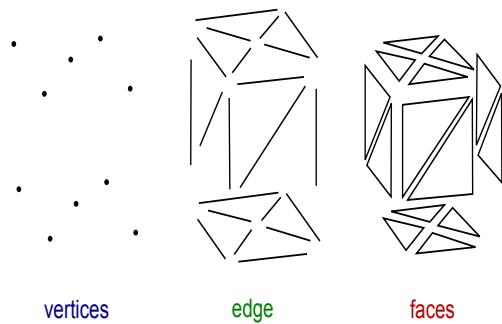


Figure 32: Transformation from vertices into faces Rchoetzein wikipedia.org CC BY-SA

Kinect Fusion is a Simultaneous Localization and Mapping (SLAM) algorithm which is capable to mesh the ToF video stream in real time. It is designed for accurate real-time mapping of complex and arbitrary indoor scenes. The algorithm is capable to recalculate the camera's orientation and position when it is moved. In the Kinect Fusion Explorer Demo,

Microsoft allows the export of this data into a mesh format like the Surface Tessellation Language (STL). It is a standard interface that is supported by most 3D programs for further processing. Stlwite is another tool for Matlab to create STL files from the raw depth data. Figure 33(a) shows a single frame of an aircraft model that is taken with the Kinect 2 camera, transformed into STL with Matlab and imported into the Open Source program Meshlab. Data from the far background and foreground was removed from the point cloud in Meshlab. The flying pixels at the edges lead to "flying surfaces" after the meshing of the raw image.

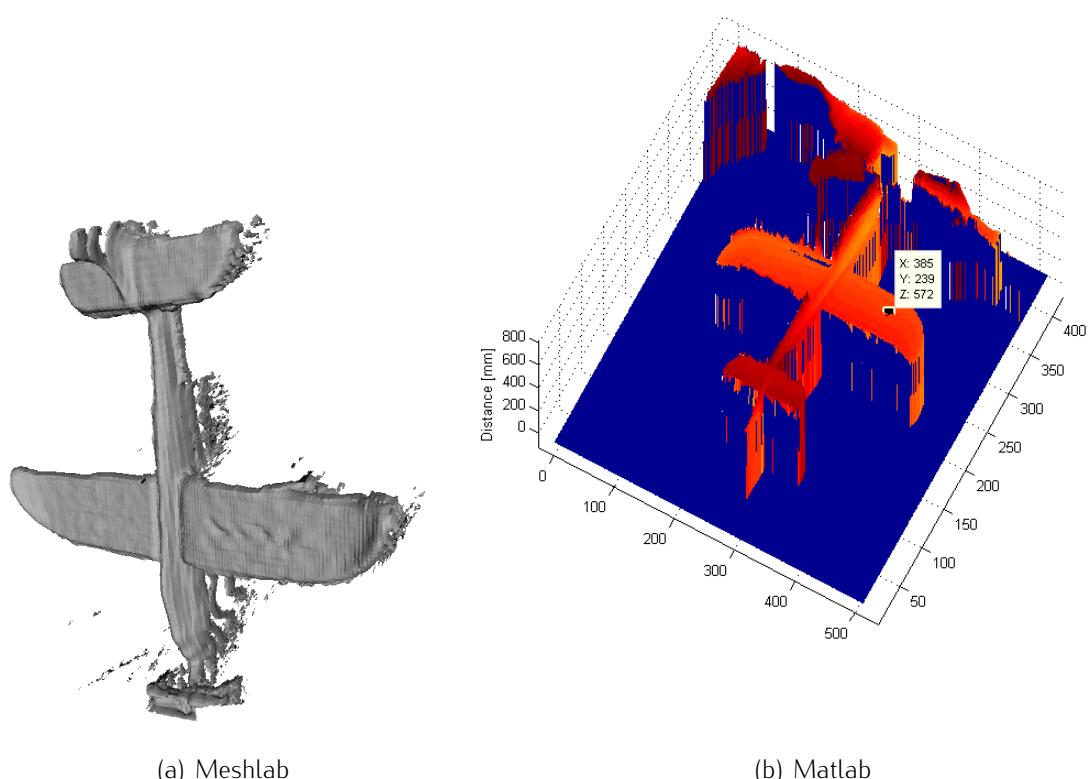


Figure 33: Meshing of point cloud data

The camera must be moved to provide a complete 3D scan of an object from all sides. To fuse multiple images into a single 3D space, the orientation of the object to the camera must be known. Reconstruct Me and other software toolkits offer the scan of a room or objects from different points of view. The position of the moving camera is calculated out of the video stream. By merging the range data from all sides, the 3D space is reconstructed. A simple construction of mesh data from range images can also be done in Matlab. The function `mesh(Z)` draws a wireframe with a surface color corresponding to the Z axis. It is a simple and fast way to determine the quality of the raw data. Before the meshing

of the image, it needs to be transformed in double values by the function `double()`. For a reduction in processing time, the data can be reduced to the distance of the investigated object. Measurements in other regions are cut out. The data cursor in an already plotted mesh helps to identify this value by the user. Figure 33(b) shows the resulting plot and the cursor to identify the region of interests with a cut off distance of 800 mm.

```
Imagedouble = double(ImageInteger)
Imagedouble(Imagedouble > CutoffDistance) = 0
mesh(Imagedouble)
```

4.3 Analysis and Filter Methodes

Multiple filtering approaches are made, to improve the quality of the image data. They all have in common that they aim to remove the measurement noise which results from different types of error specified in section 2.7. Fluctuations between multiple depth images are useful to separate noise from the so called ground truth image.

4.3.1 Creation of Frame Arrays

For the processing of a video stream the single pictures have to be aligned into a single matrix that consists out of the rows times columns times number of images L. 90 frames for example recorded in 3 s create an 512x428x90 3D matrix called Framearray. The data is gathered from a folder that contains all depth images. For keeping them in the right order, they have to be numbered consecutively in the file name. The time delta between two frames correspond to $T_{sampling} = \frac{1}{fps}$. The following code shows how to create an array of frames in Matlab out of an image folder. The images are transferred into double data type from unsigned integer.

```
ImageFolder=uigetdir('C:\ImageFolder','Folder RAW Data Images');
L=length(dir(strcat(ImageFolder,'*.tiff')));
for g = 1:L
    Frame=double(imread(strcat(ImageFolder,'\',tiffFilesS(g).name)));
    ;
    Framearray(:,:,g) = Frame;
end
```

4.3.2 Mean and Variance

The ground truth image is simply the mean of all data from the stream. Noise in the depth information is canceled out by multiple measurements. This is a standard way to remove non-systematic noise for a higher precision. The additional parameter in the function declares the dimension over which the matrix is averaged. The number of images determines the quality of the result. In some publications up to 500 images are used. Movements in the recording time corrupt the resulting image.

```
GroundTruthImage = mean(Framearray , 3)
```

The variance of a single pixel depth measurement versus time determines how much the value fluctuates around the mean value. With the ground truth image as the expected value μ the variance σ^2 of a single pixel can be derived with the Var() function:

```
PixelVariance = var(Framearray (Column , Row , :)) ;
```

$$\sigma^2(X) = \text{Var}(X) = \sum_{n=1}^N (x_i - \mu)^2 \quad (39)$$

The variance for a complete depth image plotted into a new image provides a simple way to estimate noise characteristics of different material. Figure 34 represents the variance of a scan on wooden plane area in a distance of about 700 mm. Values with a higher variance than 7 mm are set to zero. The reason for that is the high amount of noise in the corners of the image. The variance reaches values up to 470 mm, corresponding to the low radiant flux and darkening of the lens in the corners. The distribution is separated in elliptic uniform regions. The middle of the field of view shows the lowest variance. The variance increases in the image's corner direction. Apart from the corners, the variance alternates between 2 mm to 0.8 mm. In a box between 100 to 450 pixel horizontal and 50 to 350 pixel vertical the mean variance is 1.205 mm. Source code 1 demonstrates the principle how the variance image can be created. The mean variance in a pixel area delivers a single value that is better comparable.

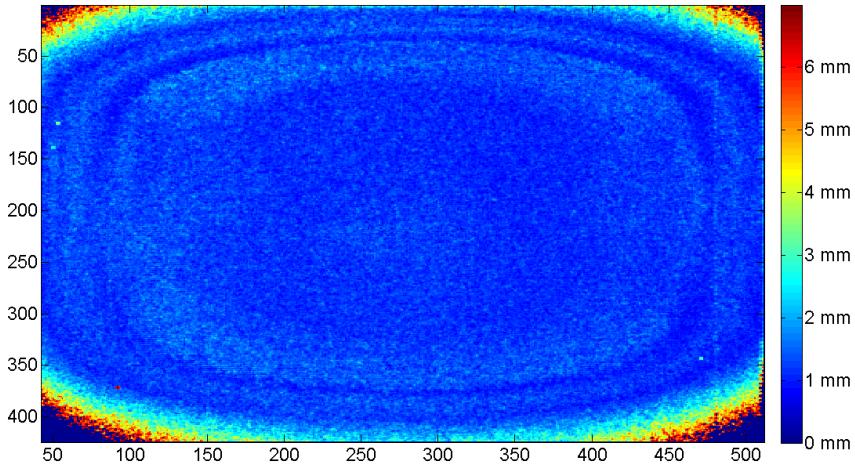


Figure 34: Raw variance of a wooden plate in 700 mm distance

4.3.3 Gradient

The gradient of an image specifies the alteration in the 2D data space. Mathematically the gradient of a two-variable function at each image pixel is a 2D vector with the components given by the derivatives in the horizontal and vertical direction. At each image point the gradient vector points in the direction of the largest possible intensity increase. The length of the vector is a measure for the rate of change. The gradient of an image is given by the formula with x pointing in the horizontal and y in the vertical direction:

$$\nabla f = \frac{\delta f}{\delta x} \hat{x} + \frac{\delta f}{\delta y} \hat{y} \quad (40)$$

`[Gmag, Gdir] = imgradient (Frame)`

The corresponding Matlab function `imgradient()` contains the magnitude of the vector `Gmag` and the direction `Gdir` in degree to the vertical axis. The following Figure 35 shows the image gradient of the office room. Values above 200 mm have been cut off to fit the colorspace and to highlight small differences.

The cupboard in front of the image stands parallel to the image plane. That is why the gradient is small. The floor features small changes in the foreground and with higher distance the magnitude increases. At the end of the room with a distance of 6812 mm the noise increases the gradient. Ideally, the flat wall should show only one constant value like the cupboard in the foreground. In general, high angles to the image plane will result

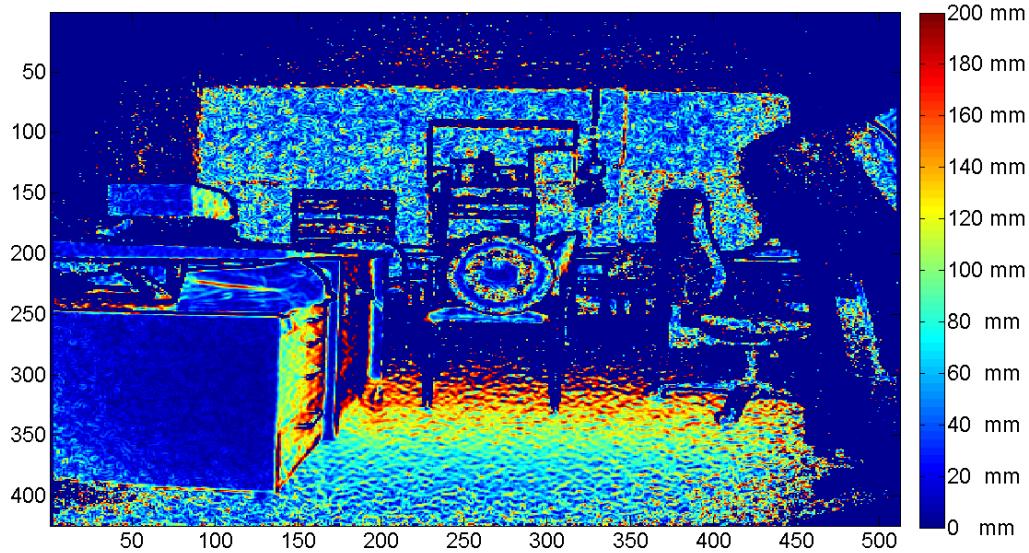


Figure 35: Gradient magnitude of the office room

in high gradients and low angles in small ones. This can be useful in different subjects. Edges around surfaces and planes are recognized. Flying pixels can be removed from the image, since they feature a high gradient to nearby pixels. Plane regions with various gradients conclude to areas with low reflectance or saturation of the image. The flat area in the middle shows a very low gradient and is therefore perpendicular to the sensor plane.

4.3.4 Denosing with Mean or Median of a Pixel Area

One basic approach to reduce the noise in the depth data is the mean or median between nearby pixels. The depth is smeared to a single value. Such kind of filtering benefits from the Kinect's high number of pixels. The median sorts every element by its value and takes the middle element. Therefore, the impact of outlier can be reduced. The mean sums up every value and divides it by the number of elements. The following source code shows the calculation of a median and a mean value from a pixel area:

```

for g = 1:NumberOfPictures
    k=0;
    for a=1:v_windowlength
        for b=1:h_windowlength
            k=k+1;
            Pixelsector(k) = FramearrayDyn_Raw(v_edge+a, h_edge+b, g);
        end
    end

```

```

DynPixelMean(g) = mean( Pixelarea );
DynPixelMedian(g) = median( Pixelarea );
end
    
```

A pixel area with the vertical dimension v_windowlength and horizontal dimension h_windowlength is collected in the array Pixelsector. The position is determined by v_edge and h_edge. DynPixelMedian collects the median value of this area for every pixel.

4.3.5 Total Variation Denoising

Rudin et al. pioneered the Total Variation Denoising (TVD) filter method that suppress noise profiles like Gaussian noise. In contrast to other approaches, like linear smoothing or median filtering, TVD simultaneously preserves edges whilst smoothing away noise in flat regions, even at low signal-to-noise ratios. It is based on the principle that signals with excessive and possibly disturbing details have high absolute gradients [30]. With a given signal, the absolute gradient can be calculated by the following formula:

$$V(X) = \sum_n |(x_{i+1} - y_n)| \quad (41)$$

One measure of closeness is the sum of square errors:

$$E(x, y) = \frac{1}{2} \sum_n (x_n - y_n)^2 \quad (42)$$

The aim of TVD is to minimize the following discrete functional versus the function y_n :

$$E(x, y_n) + \lambda V(y_n) \quad (43)$$

The regularization parameter λ_r controls the impact of the total variation on the smoothing. A high value will minimize the total variation while small values will preserve the original signal. Figure 36 plots a noisy signal and the filtered curve after TVD.

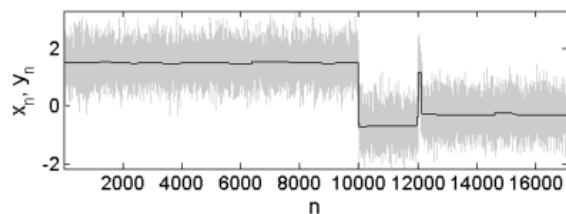


Figure 36: Total Variation Denoising of a 1D signal MAL 2010 wikipedia.org ©

Noise from 2D data like range images helps significantly in the reduction of different noise qualities. An adaptive TVD approach is proposed based on an extended structure tensor using the amplitude and phase of the ToF signal as reference [31]. In contrast to other filter methods, depth noise is removed from the 3D shape, while edges are still preserved. This algorithm has been implemented by Guy Gilboa into Matlab [32]. λ is set to one by default. The number of iterations is adjustable. Even if we cannot access the phase and amplitude of ToF data, the filter can still be used on the raw images with fixed parameters. Figure 37 illustrates the variance of the same data which figure 34 is based on. The TVD filter significantly reduced the mean variance of the complete plate from 1.205 mm to 0.1540 mm based on 20 iterations and $\lambda_r = 1$. The circular isolines of variance are highly reduced. Except for the corners, the variance could be reduced.

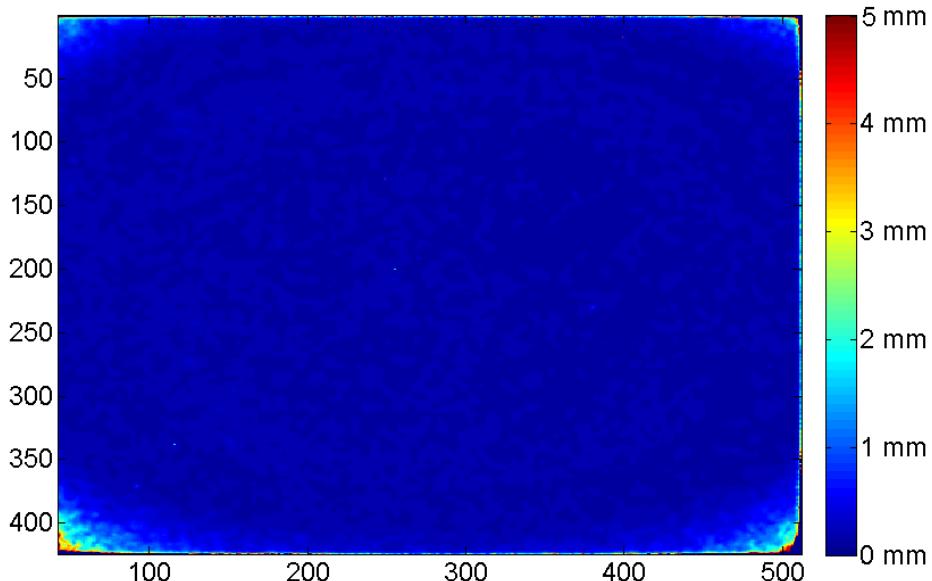


Figure 37: TVD filtered variance of a wooden plate in 700 mm distance

4.3.6 Processing of Dynamic Depth Data

The range information that is represented in the whole picture allows an dynamic investigation on every region of the field of view. For this purpose, the amplitudes of the noise must be smaller than the dynamic that is investigated. The number of images should correspond to 2^n for the FFT algorithm. If this is not the case, the remaining data is set to zero, fitting to 2^n samples. A new 2D array with the depth of every frame is created to investigate a single pixel over the complete video stream. Afterwards, the static offset of the data to the image sensor plane is removed. Every value of the array is reduced by the mean of the whole data set. Following this, the dynamic is investigated without the static offset to the surface. This process can be done for every pixel in the field of view. It remains a video stream, which contains just the dynamic. Out of this, new images are generated that represent the amplitude or frequency of the maximum peak in the spectrum. Source code 1 shows how to create frequency and amplitude images out of the FFT. The variable f_p represents the frequencies between zero and the maximum possible value of 15 Hz. AmpsDyn contains the amplitudes A_p divided by four and is given in the same dimension as the frequency array. The abs() function removes the complex values which result from the FFT and only the real part remains. The max() function returns the maximum values from the spectrum that we need to create the images for the f_p frequency and A_p amplitude distribution. Figure 38 illustrates the maximum amplitudes of a canvas in a frame that is excited with about 5 Hz by a student over 8 seconds. Higher values than 40 mm have been cut out to distinguish small changes in the color space.

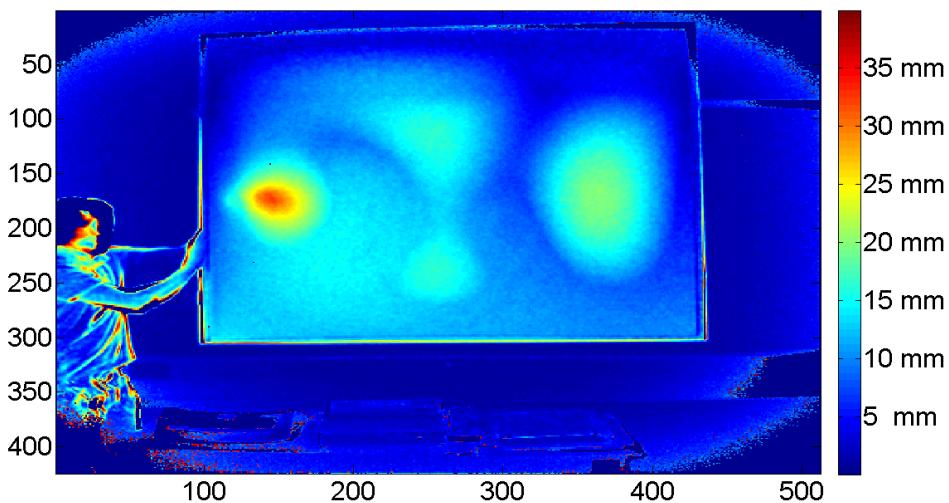


Figure 38: Peak amplitudes of every pixel of a canvas excited by a person

For a better investigation of the image, a cursor can be used to investigate the dynamic of a single pixel or a pixel area. Figure 39 shows the time domain and figure 40 the frequency domain of a pixel in the region of the excitation. The highest peak indicates the dominating frequency $f_p \approx 4.8 \text{ Hz}$ in the spectrum at an dominating amplitude $A_p \approx 15 \text{ mm}$. The second peak at about 2.5 Hz results out of the offset of the signal. The first harmonic emerges at about 9.6 Hz. Every periodic signal consists of the dominating frequency f_p and its harmonics, which are integral multiples of the dominating frequency f_p . The strongest harmonic peak is indicated as f_{ph} and A_{ph} .

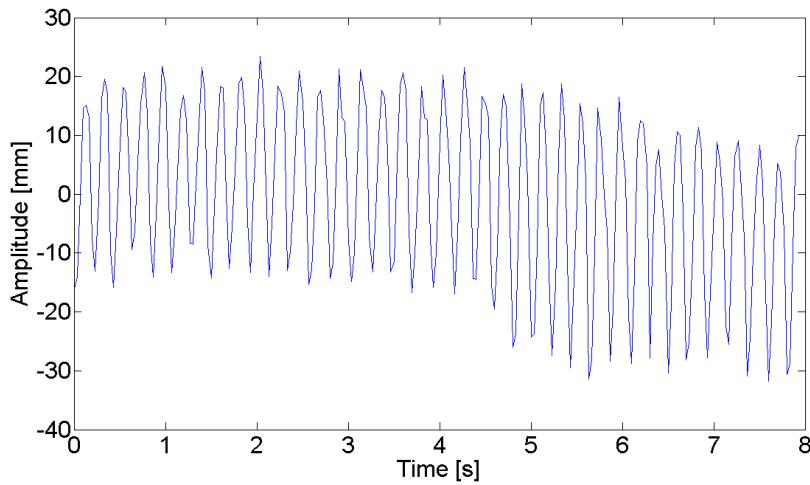


Figure 39: Time domain of excitation area

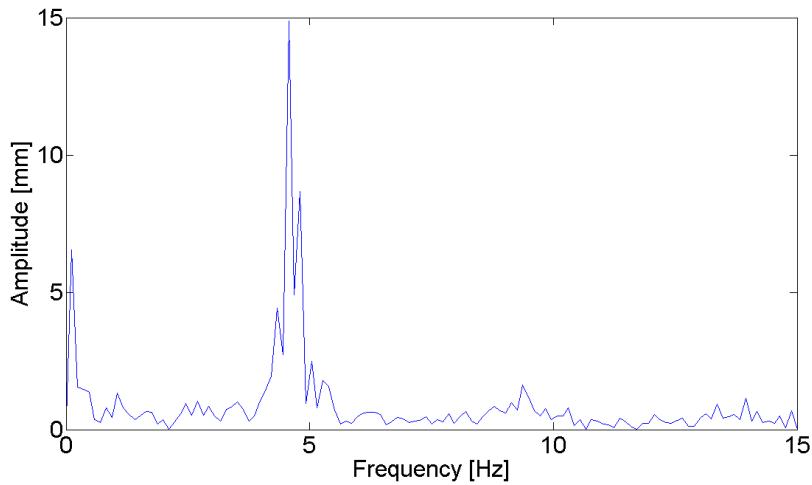


Figure 40: Frequency domain of excitation area

5 Data Processing of Laser Doppler Vibrometer

As an accurate and precise contactless measurement reference for the ToF camera, a Laser Doppler Vibrometer (LDV) runs parallel to the Kinect 2. The signal is transformed from fiber optic into a voltage signal, it is filtered and amplified before it is measured inside the Agilent DSO6054A oscilloscope. A digital interface delivers the access to the scope data. High sampling rates of common devices require a fast connection to gather the raw data over a long time. The digital resolution and sampling rate can also be reduced before it is transferred. Therefore, the data can be stored on a common USB stick in time domain CSV format. Columns are separated by commas and rows by new lines. This returns the time versus the voltage of every data point in two arrays of the same dimensions; Volt and Second. Equation 44 presents the way, how the reduced sampling rate is calculated depending on the TimePerDIV configuration on the screen, the number of samples $n_{samples}$ that are taken and the 10 rasters on the screen. TimePerDIV is adjusted by the knobs on the scope screen and $n_{samples}$ in the configurations.

$$f_{sample} = \frac{n_{samples} - 1}{TimePerDIV \cdot 10} \quad (44)$$

The time between every datapoint is the period T_{sample} and is used for the scaling of the time axis in the time domain and for the FFT analysis.

$$T_{sample} = \frac{1}{f_{sample}} \quad (45)$$

For the transformation of the alternate voltage $U(t)$ in a corresponding speed v , the sensitivity E adjusted at the LDV at the time of the measurement needs to be known. The CSV file format of the raw data can simply be imported in Matlab and transformed by equation 46.

$$v(t) = U(t) \cdot E \quad (46)$$

The direct current noise voltage needs to be removed from the data. This is done by a reduction of every value by the mean of all data. A better way would be to use a real high pass filter to remove the DC parts. The data needs to be integrated to get the displacement $z_{dyn}(t)$ from the speed $v(t)$. The library offers a function `cumtrapz()` for this purpose. It computes an approximation of the cumulative integral via the trapezoidal method. The following source code extract shows the processing of the LDV data:

```
% Scope and Laser Doppler Vibrometer Data
E = 125 %%(mm/s)/V
n_samples = 1001; % zero based
TimePerDIV= 0.5 %s
f_samples = (n_samples -1) / (TimePerDIV*10);
speed =(Volt-mean(Volt))*E;
displacement=cumtrapz(speed) / f_samples;
second=second+max(second);
```

In the last step the time vector is corrected to start at zero. Afterwards, the displacement array can be processed in the same way as described in section Dynamic Analysis of Pixels 4.3.6. The integration acts like a low pass filter. Sharp edges are removed and a smooth sinus remains. Equation 47 shows the analytic mathematical integration of the speed to obtain the displacement. The constant offset c must be calculated from other boundary conditions. This is helpful for a rough estimation of the displacement $z(t)$, directly at the scope with just the frequency f , sensitivity E and the speed $v(t)$. The complete source code of the LDV data processing, which is done later in the experimental setup, can be found in the source code extract 2.

$$z(t) = \int v(t)dt = \int A\sin(2\pi ft)dt = -\frac{A}{2\pi f}\cos(2\pi ft) + c \quad (47)$$

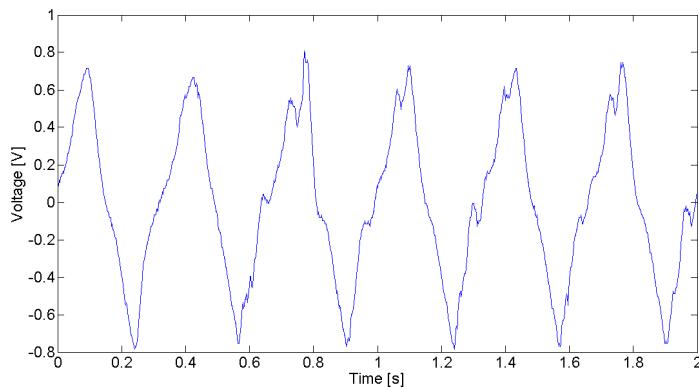


Figure 41: 3 Hz Oscillating row data sampled with a scope

6 Analysis of Depth Images

For a better understanding of the dynamic, the static behavior is analyzed in this section. The variance is the main dimension that will be investigated for that purpose. The static analysis is a measure for the noise floor. The variance should be as small as possible for a precise and accurate investigation of the dynamic investigation. It behaves as a disturbing dynamic that needs to be distinguished from the real dynamic. The analysis of absolute depth errors is not possible, since this requires a very precise experiment set-up. The positions must be adjusted over the complete range of 8 m in the resolution of the camera of 1 mm. For a smaller range, this could be done on a lathe in the future. The following approaches shall give some rough estimations on the static behaviors.

6.1 Variance Depending on the Distance

The function of the variance σ^2 over different distances gives the fluctuation of a measurement in itself. As a consequence, the precision can be judged over the distance. Microsoft delivers its own measurement data of the standard derivation depending on the distance [26]. A direct comparison is not possible. The area that is kept fixed for the analysis is unknown. A translation between the variance σ^2 and the standard deviation σ is possible by the following equation:

$$\sigma = \sqrt{\sigma^2} \quad (48)$$

A white screen for overhead projectors with a very high diffuse reflectance of about 80% has been captured over 3 seconds for the estimation of the variance. The investigated area is a 4x4 pixel window in the middle of the complete image. The area that is used for the dynamic investigations corresponds to this window. Therefore, the regions of the highest variance in the corners do not lie in the area. At the longest distance, the investigated area must still be completely on the screen. Afterwards, the mean variance of all pixels 16 is calculated in a time span of 3 seconds. The analysis has been done with the raw and TVD filtered data. As parameters for the filtering 20 iterations and $\lambda = 1$ are used. The variance versus the distance is plotted additionally even if the measurement setup like the window size or the reflectivity is unknown [26].

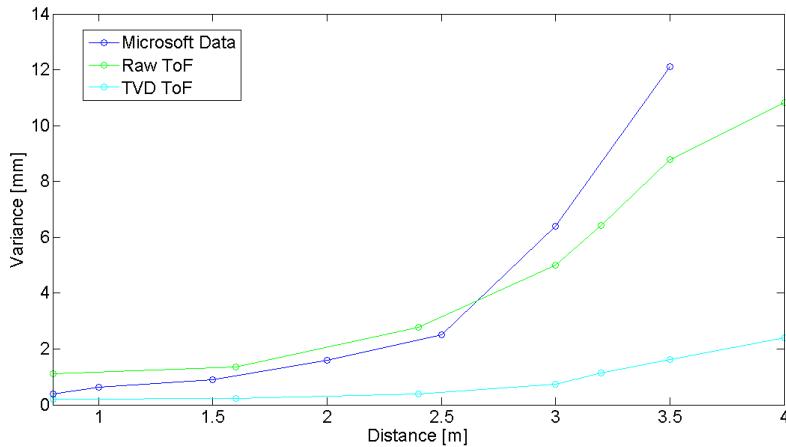


Figure 42: Variance vs. distance Microsoft data, raw and TVD after 3 Seconds

The comparison between the raw and filtered data in figure 42 shows a significant reduction in the variance over the complete distance with TVD. The raw data has a quadratic trend. This results out of the quadratic decrease of the light intensity versus the distance to the camera. The light propagates on sphere area into the room as shown in equation 5. The noise increases and the signal decreases with higher distances that correspond to a lower SNR.

6.2 Variance Depending on Geometries and Surfaces

A simple way to investigate the variance difference between high reflecting white and low reflecting black surfaces is possible on a checkerboard. Squares of the same size on a flat surface are easy to distinguish with different reflectivity properties even if they lie on the same surface. This is also a common approach to calibrate the distortion of a standard camera objective. Figure 43 illustrates the change of the variance between the black and white areas in a one meter distance after a recording time of 3 s. The black fields feature a higher variance than the white ones. The sticker material of the checkerboard has spread reflection characteristics. Therefore in the middle of the image more light is reflected backwards into to the sensor than on the outer areas. The variance in the middle is reduced. More light is reflected backwards. A calibration of the Kinect 2 is not necessary since the focal lens is fixed and the correction of the distortion is done in the image processing of the camera. The reflection characteristics on various material does not always behave in the same way as for visible light. This is shown on some random objects measured in there variance for a rough estimation between different surfaces. In figure 44, a collection of geometries and surfaces are investigated. Except for the black flat plastic with a thick layer

of clear paint and the polished aluminum pipe, all material probes show a low variance. The black plastic absorbs and reflects most of the IR light and the pipe scatters the light in a spectral way to all sides. Only the middle part of the pipe parallel to the image plane is of low variance. More light is reflected backwards. Even if the heck spoiler and the copper sheet are spectral reflecting for visible light, this is not the case in the IR spectrum.

Flying pixels cause a high variance at the border of all objects. To distinguish small changes in the color space, larger variance values than 140 mm and distances behind the objects are set to zero. The influence of the angle to a plane surface has not been investigated. The change of variance on different geometries and surfaces should be part of further investigations.

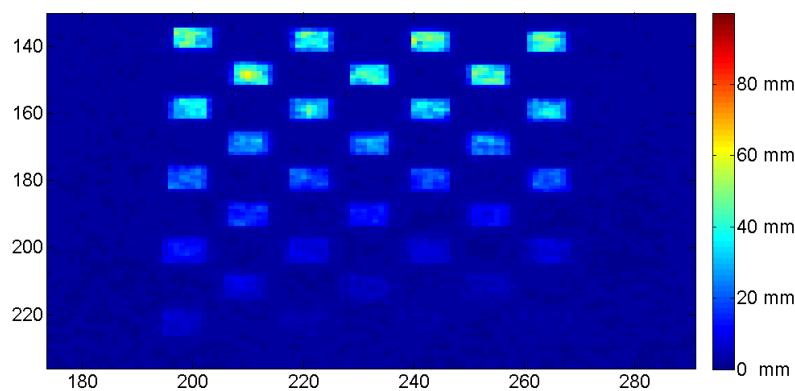


Figure 43: Variance on a checkerboard in 1 m distance

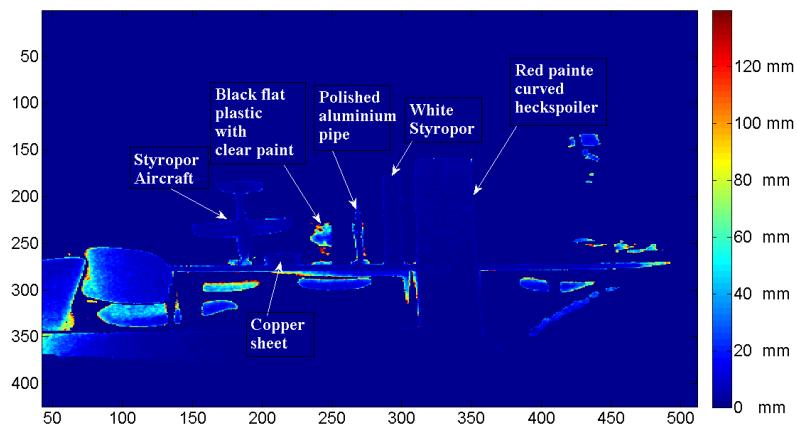
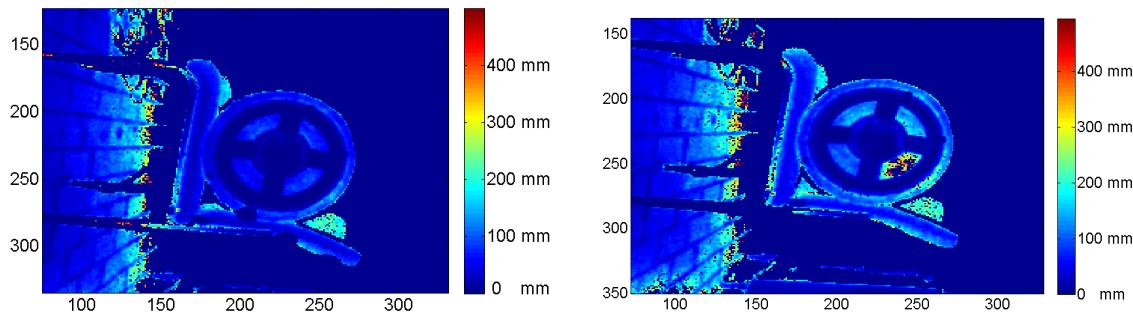


Figure 44: Variance on various geometries and surfaces

6.3 Influence of Background Lighting on Static Measurements

The variance of the static bass membrane plays an important role for further dynamic analysis. A reduction in the noise characteristic is done with 5 white matt paper stickers attached to the membrane. This gives a clear sharp border to the original black surface. 2000 W of halogen light are pointed on the membrane to disturb the measurement from the position of the camera by a second IR light source. This data is compared with a recording in absolute darkness. The position and 1.4 m distance of the camera to the membrane are unchanged. The distribution of variances is illustrated in figure 45(a) for the dark and in figure 45(b) under halogen lighting. The depth data of the background behind the membrane is set to zero. The light impact is remarkable on the black cone surface. The white paper is influenced much less by the IR lighting. The reason for that is the low SNR on the black cone. The IR light of the ToF measurement is strongly absorbed and scattered on this surfaces. With the IR impact of the halogen lighting, the noise is additionally increased to values higher than 500 mm. This is caused by a spread spectral reflection, which directs a high amount of IR halogen light directly into the camera. The RGB image of the Kinect in figure 78 shows the same saturation caused by the spread reflection.



(a) Variance on the static membrane in a dark room (b) Variance on the static membrane and 2 kW halogen

Figure 45: Influence of halogen lighting on the variance

7 Evaluation of Membrane Dynamic

7.1 Physical Model of Membrane Oscillation and Observation

This section delivers a physical model for the vibration of a subwoofer membrane in the investigated low frequency area $f_{ref} \leq 10 \text{ Hz}$. Due to the low frequency excitation, every point moves with the same speed $\vec{v}(t)$ and the same phase into the normal direction \vec{n} of the middle flat surface. The measurement of the vibration under a bias angle is characterized by the angle between the optical axis and the surface normal, called viewing angle α . Since LDV measures only the speed into the direction of the optical axis, the amplitude A_{LDV} is reduced under higher angles to the surface normal. Figure 46 illustrates the basic geometries of the membrane observation. The viewing angle α is defined between the surface normal \vec{n} and the optical axis turning into the negative mathematical direction. In contrast to that, the ToF amplitude A_{ToF} is increased by the bias observation since the investigated pixels are moving sideways and ToF delivers the absolute distance and not a relative speed like LDV. With a cone angle β , the oblique areas deliver different normal surface angles and therefore α is changed. This physical model shows the relationship between the viewing angle α and the resulting amplitude of A_{LDV} and A_{ToF} represented by equation 49 and 50. This model assumes that the ToF camera has an infinite number of pixels. In practice this is not fulfilled. One pixel will therefore cover multiple distances under high angles. On the upper and lower surfaces the angle $\alpha_{l,u}$ depends on the oblique position of the cone surface and the angle β to the middle surface. The coordinate system of the membrane is represented by the index (m) and has the same orientation as the scene plane coordinate system that is given by the depth images.

$$A_{ToF} = A_m(\cos \alpha + 1) \quad (49)$$

$$A_{LDV} = A_m \cdot \sin \alpha \quad (50)$$

$$\alpha_{l,u} = \alpha \pm \beta \quad (51)$$

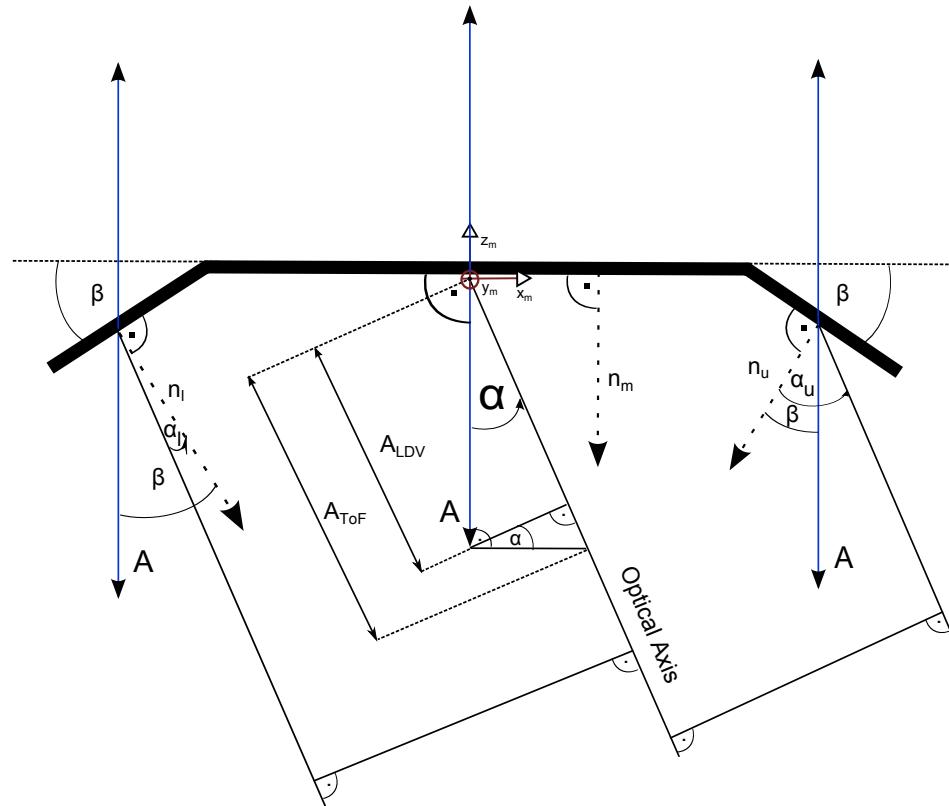


Figure 46: Geometry of bias membrane observation

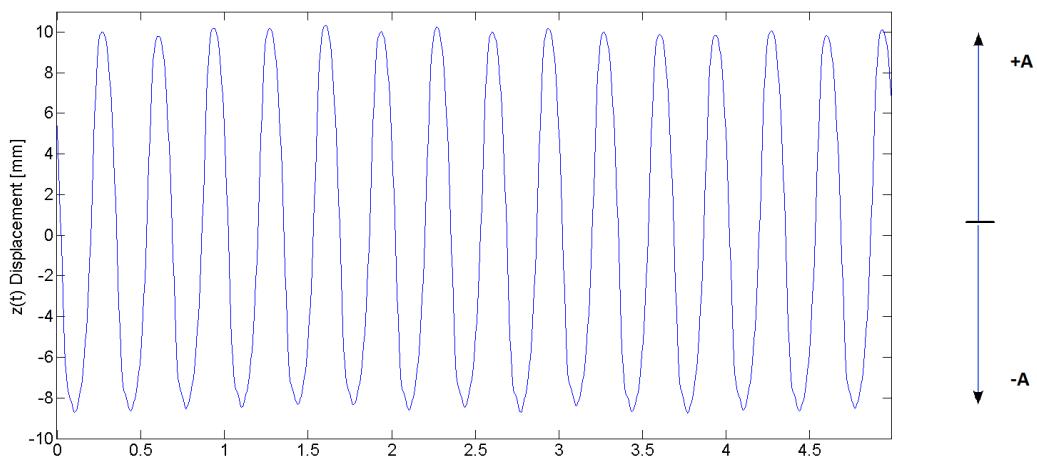


Figure 47: Displacement of the membrane measured on the middle surface with a Laser Dopper Vibrometer under $\alpha = 0^\circ$

7.2 Scientific Problem

The dynamic behavior of a membrane is the test body for a detailed analysis of the ToF principle for dynamic measurement of surface vibrations. The membrane is excited and acquired under variable conditions to investigate multiple scenarios. A bass tube is plugged to a waveform generator, creating a defined vibration that is recovered from multiple ToF 3D scans. This data is compared with a Laser Doppler Vibrometer (LDV) as a reference. LDV delivers the accurate data of a single point on the membrane, whereas ToF measures the complete surface movement. Multiple questions will arise from that: One is the precision and accuracy distribution in frequency and amplitude over the membrane under different conditions. Which part cannot be investigated and which area is overlaid by noise? The influence of angle, distance, frequency and amplitude is questioned. How is the measurement influenced by an additional IR light source that emits in the same spectrum as the ToF laser diodes? Another question is the influence of static filter method for the dynamic investigation. Is a reduction of the noise possible without an impact on the actual dynamic? Is the complete signal-to-noise-ratio influenced by the filtering? Can overtones like the first or second harmonic still be derived from the ToF data? How does ToF behave in a continuous change in frequency and amplitude? ToF allows the investigation of the complete field of view. How can the dynamic of a surface be illustrated? The following section tries to investigate all these questions in detail. The membrane movement should deliver an ideal test body, because it vibrates in only one direction at a stable frequency and amplitude. The vibration is controllable in frequency and amplitude. This approach could later be useful for the calibration of ToF for vibrations measurements with a scientific camera and under more controlled conditions. The experimental setup shall deliver first steps into this direction, showing the performance of ToF and, how the data can be processed in Matlab.

7.3 Hypothesis

With the depth resolution of 1 mm and a variance lower than 1.2 mm up to a distance of 2 m , the system should be capable to distinguish changes of amplitude in the magnitude of the 1 mm depth resolution. Except for the corner of the images, a dynamic investigation in the whole field of view should be possible for distances below 2.5 m , because the variance in this distance is below or in the region of the depth resolution. Processed by a Fast-Fourier Transform, a vibration with a amplitude higher than 1 mm should always return the strongest peak in the frequency spectrum. With a sampling rate of 30 fps and a maximum vibration frequency of 10 Hz , the limits of the Nyquist-Shannon theorem for dynamic measurements are satisfied. A precise determination in frequency should be possible on all areas of the

membrane. An oversampling factor of 5 should be enough for the reconstruction of the wave form and amplitude. Therefore, every sinus period is sensed by 10 samples. Consequently, a precise reconstruction of a 3 Hz should be possible in frequency and amplitude. The change of the viewing angle α influences the amplitudes magnitude into the direction of the ToF camera and the SNR. The ideal trend should follow the physical model that is represented in section 7.1. The displacement of ToF is increased with higher viewing angles and the low reflection backwards to the sensor will cause a high inhomogeneous in the amplitude distribution. Single points of the membrane do not correspond to single pixels in the depth image any more. The impact of the filter method depends on the amplitude and the parameter of the filtering. A high smoothing of the image pixels should not just decrease the variance but also the vibration's amplitude. The impact of black versus white surfaces should be significant in the dynamic analysis. Figure 45(a) already implies, that the black oblique parts have a higher variance, overlying the vibration. The influence of a second IR light source should decrease the SNR ratio and therefore also the accuracy of the measurement.

Amplitudes higher than 1 mm with lower frequencies than 10 Hz will emerge in a Short-time Fourier transform. The resulting spectrogram should deliver a simple and detailed way to investigate multiple frequencies with ToF and LDV in a single signal.

By extracting maximum values out of every pixel's FFT, the frequency and amplitude distribution can simply be illustrated. Differences in the surface's dynamic are recognizable. The distribution of the first and second harmonics will be visible even on high reflecting surfaces with a low SNR.

7.4 Measurement Setup

The direct comparison between ToF and LDV requires the measurement from preferably the same point and the same distance to the vibration. Since this is not possible, both devices are mounted together on a single tripod. The direction of the perspective should be as close as possible to avoid bias differences in the viewing angle α . The LDV is mounted near the IR camera of the Kinect 2 device. Figure 50 shows the assembly of both devices. The LDV should never point into eye-height to avoid hazard of the human eye. The experiment is performed in a dark room without windows and just two halogen lights at the ceiling. The Kinect 2 images are recorded by a desktop PC via an USB 3.0 connection. The open source software dumpK4W was used for the recording of the "raw"

images. It saves every image from the video stream into an uncompressed tiff image. The depth information in every pixel is already represented in millimeter dimensions. Thereby, a calibration is not necessary. The USB 3.0 controllers require a high data throughput. Some older USB 3.0 chipsets are overcharged by the high data stream from the two cameras and the three microphones. Microsoft delivers a list of chipsets which are capable to handle that. The data stream is captured for 3 or 8 seconds. Before an import into Matlab is possible, all other images like RGB have to be removed by hand. The acquisition has to be repeated, if the number of images does not fit to the capture time. Running a D2D example program with a live video stream at the same time reduces the number of lost images. The Polytec CLV700 is the LDV that is used as a reference device. An optical fiber connection delivers the high monochromatic laser beam to the sensor head. From there, the laser beam radiates on the surface and reflects backwards into the sensor head. The resulting interfering light is transported backwards into the analog input module CLV-M200 for the evaluation. The decoder module CLV-M030 amplifies the voltage to a adjustable range. $125 \frac{\text{mm}}{\text{s}}/\text{V}$ is configured in this stage by a knob. A lower value was not possible since the high amplitudes of the membrane lead to a saturation in the decoder module. Up to 1.250 mm/s can be measured in this configuration at resolution of $2 \mu\text{m/s}$ [22]. The BNC connection delivers the analog signal to the oscilloscope, that is necessary to sample the high voltages up to 10 V . With the Agilent DSO6054A Digital Oscilloscope, the analog signal is sampled depending on the configuration shown in section 5. Only the TimePerDIV parameter of the screen and the $n_{samples}$ influences the way the data is acquired. The compressed data is saved on the USB Stick in the CSV file format. This has to be configured inside the print options. Matlab provides pre-built functions for the automatic input of the CSV formated data. Since the frequencies of the input signal f_{ref} are lower than 10 Hz , the oscilloscope is not able to trigger the screen on the input signal. The refreshing of the screen occurs in free-run operation. The membrane in the subwoofer tube is driven by a 240 W Yamaha A-S500 integrated amplifier in pure direct mode without any filtering. The inner part of the membrane should be as flat a possible. Consequently, ToF and LDV have a uniform plane surface that can be investigated for deformations. The tube is located on a chair at the height of the LDV and TOF perspective. The Agilent Waveform Generator 33120A creates the audio signal that is feed into the amplifier on a mono line by a self made BNC to chinch cable. Before a direct connection is established, the output voltage is checked to match the standard line level criteria. Levels above $U_{pp} = 3.472 \text{ V}$ for professional equipment and $U_{pp} = 0.894 \text{ V}$ for consumer equipment can damage the electronic. The line level is an alternating current signal without a direct current offset. A standard audio output, that is measured with the scope, should deliver the same U_{pp} value at the maximum volume. All

measurement devices that are connected to each other are plugged together in one mains socket to avoid ground loops and therefore interference between devices like 50 Hz noise. For a higher reflectivity of the ToF IR light, white paper stickers are attached to the flat surface and the oblique outer parts. It should feature a high paper density to achieve a reflectance ρ of 80 % and higher. The middle sticker has a dimension of 76x39 mm. The LDV requires a small reflecting tape on the inner surface to increase the intensity of the reflected laser beam as seen in figure 51. Figure 48 narrows the experimental setup down to the important components. The distance and angle to the membrane is illustrated. The reconfiguration of this parameters is done in a roughly way with a common measuring tape and a radian measure.

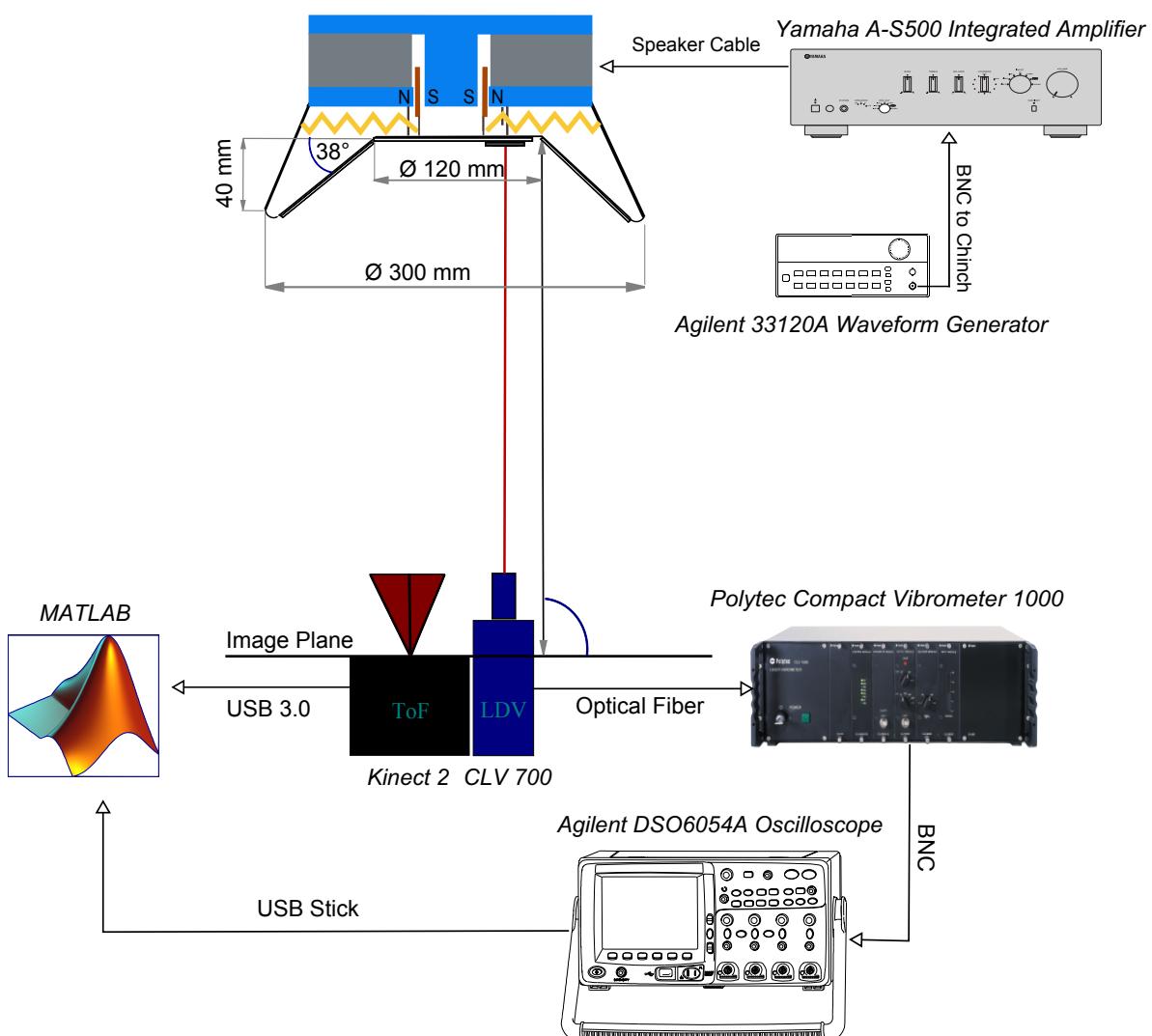


Figure 48: Sketch of the experimental setup

Device	Information	Configuration
Polytec CLV700	HeNe Laser Safety II	Variable laser focus
Input Module CLV-M200	$v_{max} = 1.250 \text{ mm/s}$	$2 \mu\text{m}$ Resolution
Decoder Module CLV-M030	Resolution = $2 \mu\text{m/s}$	125 mm/s/V
Output Module CLV-M001	-	5 kHz Low-pass filter
Kinect 2	512×424 30 fps	Depth resolution 1 mm
Agilent 33120A Waveform Generator	Uncalibrated	$0.894 V_{pp}$
Agilent DSO6054A Oscilloskop	500 MHz	1000, 500 or 250 ms per DIV
Yamaha A-S500 Amplifier	150 W Mono	pure direct / no filtering
Matlab R2013a	-	-
Subwoofer	30" JBL Car Tube	Paper Sticker / Reflector

Table 5: Devices of experimental setup



Figure 49: Experimental setup



Figure 50: LDV and Kinect 2 mounted on one tripod

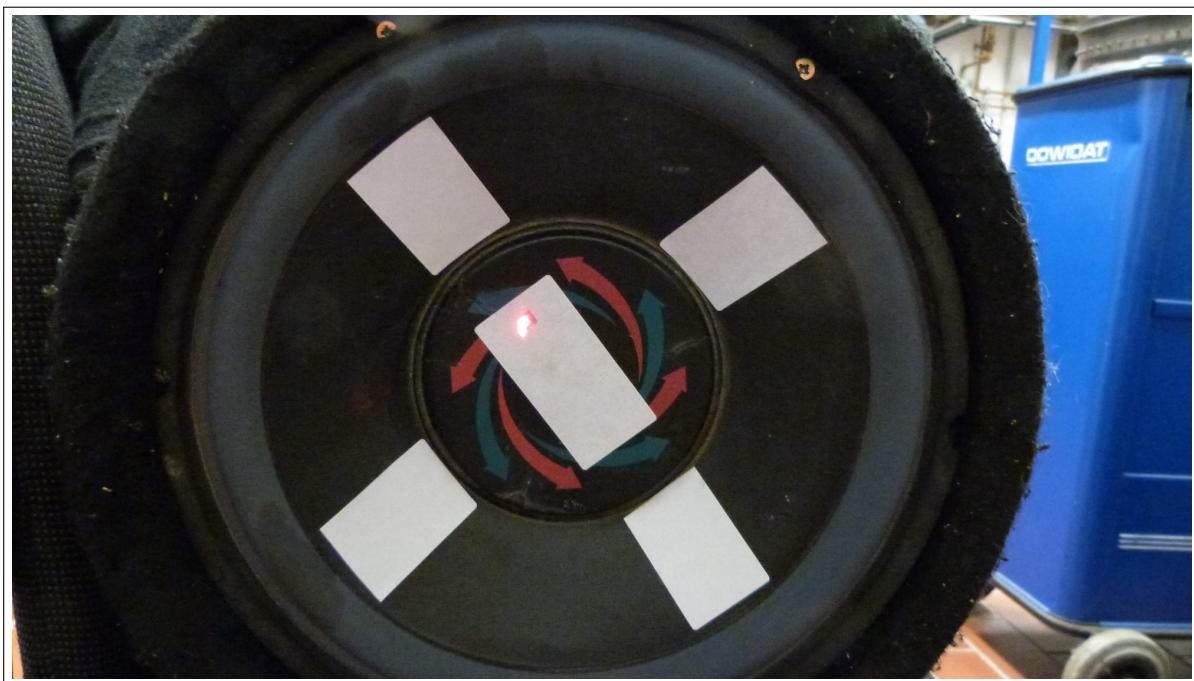


Figure 51: Membrane with white paper stickers and reflector

7.5 Test Procedure

Some fundamental boundary conditions apply for all experiments that are done. Every device should be turned on for some minutes before the execution to achieve a stable temperature. ToF cameras are in general sensitive in their accuracy for fluctuation in temperature [23]. The IR lighting and electronic heats up the device over time. The LDV is pointed together with the ToF camera on the membrane with reflector and the white paper attached to it. The membrane is in the middle of the field of view in the distance d and the viewing angle α to the normal vector \vec{n} of plane middle surface. If no angle is given, it is kept constant at $\alpha = 0^\circ$. The LDV signal indicator should be at least at a level of about 30 % to guaranty a high SNR. Afterwards, the volume of the amplifier can carefully be turned up. Ideally, no sound should be hearable with frequencies lower than 30 Hz. In practice, the membrane creates noise in higher frequencies even if it excited in the 1 to 10 Hz range. The LDV is capable to measure this sounds as spikes on the sinus oscillation. It is a sign for an overdrive of the subwoofer. The high amplitudes can cause the membrane to detach from the tube. The attachment is checked under the highest amplitude and frequency according to equation 34. The tube and the structure itself should not get excited. The peak-to-peak value is displayed on the oscilloscope and should stay in a narrow band. It is an indicator for a stable oscillation at a constant amplitude 5. Every measurement point is the mean from a 4x4 pixel window that is chosen by the user as near a possible to the LDV measurement at the middle of the white paper. Since the laser reflector is also spectral reflecting the ToF IR lighting, an investigation of the same point is not possible. The reflector leads to a hole in the 3D scan. Some safety guidance for humans, environment and devices have to be kept in mind. The laser beam should never be attached in the height of the human eye to avoid damages to the retinal after long exposure. The membrane is another source of danger. High amplitudes up to 10 mm require the volume to be turned up to a maximum. The same volume in the audible range can damage the eardrum. So a change in frequency at the waveform generator has to be done with caution. Structures in the environment of the test are also exposed by the strong vibrations and can be excited in there eigenfrequencies. The influence of background lighting is characterized in a distance of 1.4 m at 3 Hz with 2000 W halogen lighting at the position of the cameras pointing on the membrane. An RGB image from the Kinect's perspective is visible in figure 78

The data is captured in a unsynchronized way. This means, that single data points cannot be compared to each other. It is only possible to analyze single deflections to each other in the variable displacement setup. In all other configurations it is assumed, that the oscillations

is kept stable over the complete acquisition time. In this way, small time delays between the ToF and LDV measurement have a small impact on the results in the frequency domain. Nevertheless a time synchronization is beneficial. Both measurements must take place in the same time span. The acquisition of both devices is started manually. At first, by the commandline on the PC to start ToF and afterwards by pressing the "Quick Print" button on the oscilloscope to acquire the LDV data. Since low frequencies can not trigger the screen, the measurement should be started when the signal trend ends at the right side of the screen.

7.5.1 Variable Distance

The first parameter being changed, is the distance d of the tripot to the middle of the membrane. A high precision horizontal movement is not possible with the given equipment. The rough estimation is done with a measuring tape. The 0° viewing angle of the laser beam to the surface normal vector should be kept constant under all conditions. The volume and the frequency of the waveform generator $f_{ref} = 3 \text{ Hz}$ is unchanged for all measurement points. The amplitude of about 4 mm is high enough to distinguish it from noise under the maximum distance. For every measurement point the focus of the LDV has to be readjusted to a sufficient signal level. The noise is reconsidered for every distance and can be minimized by gently moving the sensor head. The first run of the experiment is done for 3 seconds of ToF and 2 seconds of LDV. The second run will be done for 10 seconds of LDV and 8 seconds of ToF. At a distance of 1.4 m the setup is added with 2000 W of halogen lighting at the position of the camera to disturb the ToF measurement on the membrane with additional IR light.

7.5.2 Variable Angle

Since there is no precise setup for the viewing angle α determination, it is estimated by a goniometer on the floor. In the distance of 1.4 m and 2.0 m the angle α is changed from 90° to 0° in 5° steps at a constant frequency of $f_{ref} = 3 \text{ Hz}$. α is the angle between the optical axis of the camera and the middle surface turning in the mathematical positive direction. The focus must be readjusted for every angle. It is ensured, that the laser points on the reflector over the complete displacement of the membrane at every angle. The area of the LDV and ToF measurement is therefore increased with higher angles. The simplest way to change the angle between the perspective and the membrane is by turning the subwoofer on the measurement point. The paper stickers are compared with each other. The one in

the middle delivers a surface under the angle α . The lower sticker has an angle of $\alpha + 38^\circ$ and the lower of $\alpha - 38^\circ$ due to the cone.

7.5.3 Variable Frequency

The change of frequency from 10 Hz to 1 Hz is done in a distance of 1.4 m at a constant volume in one 1 Hz steps at $\alpha = 0^\circ$ to the surface normal. Since the frequency change at the functions generator cannot be separated from a change in the maximum displacement, the amplitude cannot be kept constant for different configurations. The membrane should still be out the overdrive region to archive comparable data, at the highest frequency. The acquisition takes place at every parameter for 8 seconds to capture enough periods at the lower frequencies.

7.5.4 Variable Displacement

The volume is turned up from 0% slowly for a brief moment and reduced afterwards for the investigation of different displacement of the membrane. This configuration can also be considered as an amplitude sweep, because the volume is changed during the data recording. Only a small number of oscillations is analyzed and can be compared with the idle state before and after the oscillation. The acquisition takes place for 8 second to allow enough time for turning the volume up and down manually.

7.5.5 Sweep

The adjustment of the frequency versus time is called Sweep. At a constant volume, the frequency is increased manually in 0.1 Hz steps from 1 to 10 Hz . Ideally, this is done automatically by the function generator in a smaller resolution. The measurement takes place over 10 seconds with ToF and LDV. To gather all the date with the scope in this time span, $f_{sampling}$ has to be decrease to 100 Hz . No noise should occur at the highest frequency in order to avoid deviations of the ideal sinus trend.

7.6 Measurement Results

All data is processed by the *dynKinect_0_3_thesis_evaluation.m* Matlab script. The TimePerDIV parameter is the only number that has to be chosen manually depending on the acquisition time of the oscilloscope. From this $f_{sampling}$ results according to equation 44. The wrong parameter will change the amplitudes, times and frequencies in the domains. For a comparison between the devices, raw data is transformed into the frequency domain by the `fft()` function shown in chapter 4.3.6, 5 and 2.9.2. The Short-time Fourier transform method is used for the sweep analysis. In the frequency domain, noise and dynamic are easy to separate. The strongest peak in a 4x4 pixel windows is the value that is compared and plotted versus the change of parameters. In a distance of 2.6 m, the window still fits on the sticker of 76x39 mm according to equation 38 and figure 52. Every data point is written into a CSV file by hand. The file is shifted into the Matlab workspace afterwards. A small script called *xy_plotter.m* is used to plot the data into graphs automatically.

7.6.1 Variable Distance

Figure 53 shows the progress of a the raw and TVD filtered data versus the reference LDV measurement after 3 seconds. LDV delivers two distortions at 1.1 m and at 1.8 m that cannot be explained by the instability of the membrane. A detailed analysis in the time domain shows an error in the processing with Matlab. Plotting clarifies an unsteady trend. This behavior is not noticeable in the raw data of the oscilloscope. Equally to the 1.8 m position, a strong jump of about 3 mm can be observed in the frequency and time domain additionally. A second peak near 0 Hz appears in the time domain and is a measure for the offset. Such behavior cannot be observed in the ToF data. A reason for that could be noise in the LDV measurement or an incorrect integration procedure. That is why those two points are not used as a reference and not plotted in the LDV data. Apart from that, the amplitude is reconstructed with a maximum error of about 1 mm by ToF. The membrane works stable in a range of about 0.4 mm in all distances at around 3.5 mm amplitude according to LDV. An influence of the distance on the accuracy is not clearly visible. The TVD filter could increase the accuracy in the distance between 0.9 m and 1.6 m with fixed parameters. Strong distortions like in 1.3 m and 1.6 m are reduced. In other distances, the filtering reduces the accuracy in contrast. Figure 54 visualizes the influence of a longer acquisition time of LDV and ToF in a second series of tests. The amplitude is reconstructed by LDV in smaller range of about 0.2 mm with no distortions. ToF delivers a strong deviation at 0.8 m . The reason for that could be light scattering between the sensor and the lens. Too much light is reflected from the white paper surface. Apart from that point, the amplitude is reconstructed in a band of about 0.7 mm . In the distance of 1.3 m to 1.5 m , the absolute error is minimized down to a value of 0.05 mm . The increased acquisition time allows a better separation of the variance from the actual dynamic. Figure 52 delivers the projected area of the 4×4 pixel window versus the distance at a viewing angle of 0° .

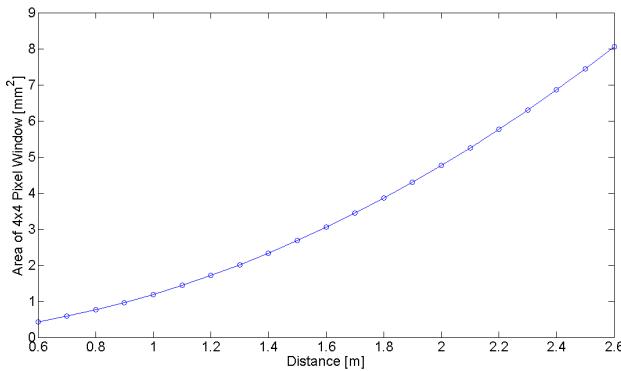


Figure 52: Theoretical area of 4×4 pixel window versus distance

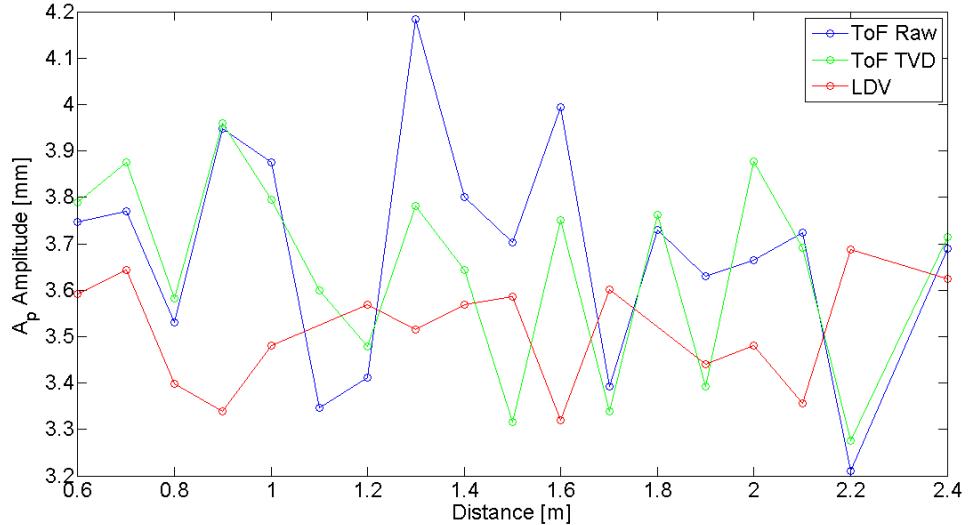


Figure 53: Amplitude A_p versus distance after 3 seconds of ToF measurement

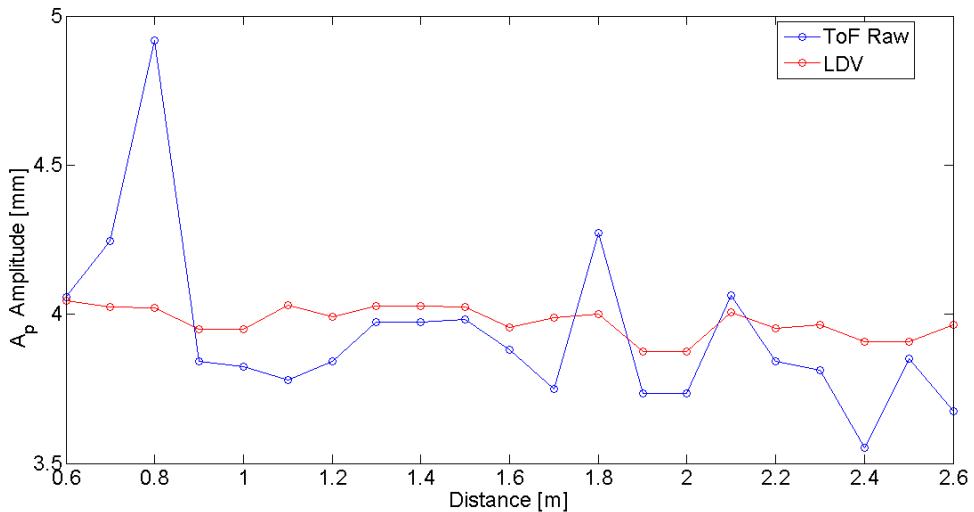


Figure 54: Amplitude A_p versus distance after 8 seconds of ToF measurement

A direct comparison of the time domain data of ToF in figure 55 and LDV in figure 56 shows the huge difference in the sampling rate and resolution. The digital resolution of $T_{sample} = 33.3 \text{ ms}$ in time and 1 mm in depth is noticeable in the ToF data. Regions of high acceleration at the peaks of the displacements have sharp digital edges in the sinus trend. This leads to additional fluctuations in the peak values. The LDV data proves the high short time stability of the membrane vibration. The short spacing in time of $T_{sample} = 5 \text{ ms}$ and in displacement of $2 \mu\text{m}$ very precisely reconstructs the trends. A better comparison between a set of ToF and LDV data can be done in the frequency domain after a FFT. Figure 57(a) and

57(b) shows the ToF and LDV in the frequency domain in 1.4 m distance. The difference in the amplitude A_p is 0.063 mm and in frequency 0.117 Hz. The obvious error in frequency can be explained by the digital resolution in time of $T_{sample} = 0.033$ s. Even the first and second harmonic component are clearly defined and match to the spectrum of LDV. On other parts of the membrane, the harmonics cannot be distinguished from noise that clearly as on the oblique black cone surfaces. Figure 58(a) and 58(b) illustrate the frequency domain in 2 m distance. The first and second harmonic can still be seen in the ToF spectrum. The error of the absolute peak is increased to +0.402 mm. In the LDV data, the third harmonic is still visible. An offset peak at 0.1953 Hz of 0.3618 mm appears in the LDV data but not in the ToF data.

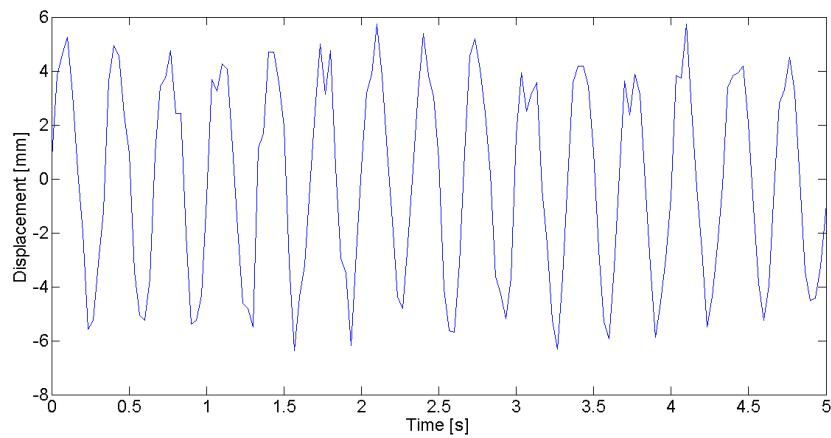


Figure 55: Time domain of ToF after 8 s in 1.4 m distance

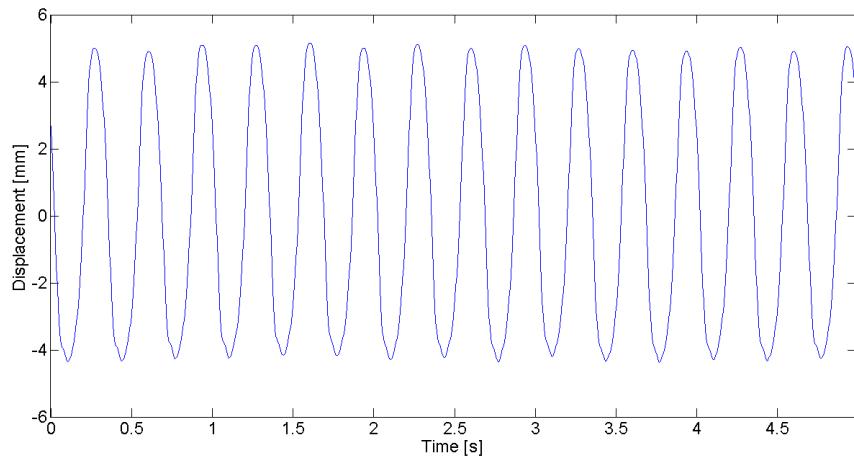
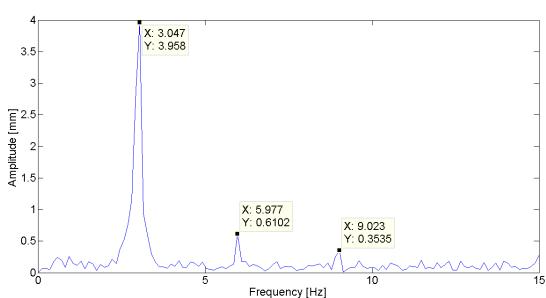
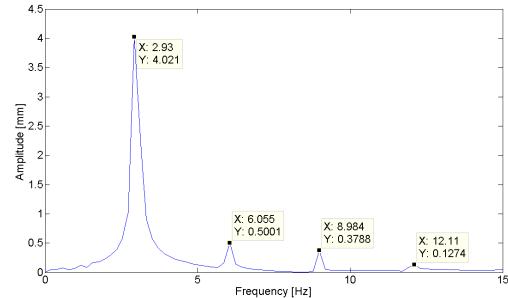


Figure 56: Time domain of LDV after 8 s in 1.4 m distance

For a detailed illustration of complete surface dynamics, two new images are created from the 1.0 m data. Every pixel of the ToF video stream is analyzed in the FFT and a new image is plotted depending on the highest peak A_p in the spectrum. Section 4.3.6 explains the creation of such images more detailed. The challenge is to integrate the frequency or the amplitude into a 2D plot in a huge range with a high contrast in the colorspace. The dynamic information from the video data is transferred into a single image. Figure 59(a) illustrates the dominating frequency distribution of f_p on the membrane. Most pixels measure the correct 3 Hz frequency. Some areas on black oblique surfaces are dominated by noise of a random spectrum. The frequencies are shifted slightly. At the edges of the stickers, the amplitudes and frequencies are highly increased. An explanation is the fluctuation of variance between the white and black field in the area that is observed by the pixel. The change in variance overestimates the actual deflection.

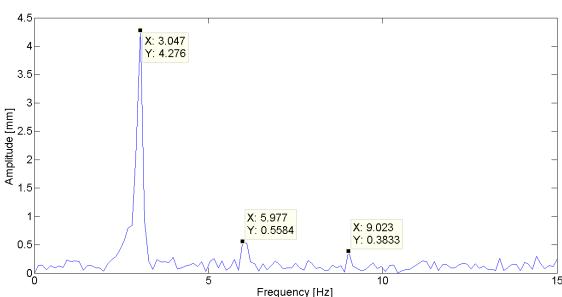


(a) ToF Frequency Domain

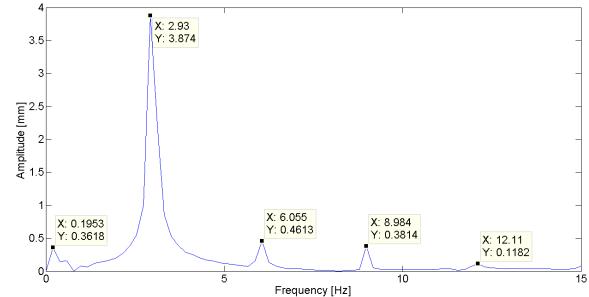


(b) LDV Frequency Domain

Figure 57: 3 Hz excitation in 1.4 m distance



(a) ToF Frequency Domain



(b) LDV Frequency Domain

Figure 58: 3 Hz excitation in 2.0 m distance

Another abnormality is given by the LDV reflector. The IR light is reflected spectral on this surface. A depth measurement is not possible here and therefore also no dynamic investigation. The static surfaces around the membrane illustrates this in chaotic colors. The dominating frequencies on the static surfaces depend on the variance that shows a very high entropy. Figure 59(b) clarifies the amplitude distribution. Higher values than 5 mm are removed to distinguish small changes in the colorspace. The white paper delivers a homogeneous variance distribution and therefore a steady amplitudes distribution. The values are underestimated or overestimated in the black oblique regions. The connection between the static and dynamic part of the membrane shows a transition of the amplitude from 4 mm to 0 mm . In the upper oblique areas, the amplitude is higher than in the lower ones. An explanation for that could be the variance depending on the incident angle φ according to equation 8. Figure 43 illustrates the same effect on the checkerboard. In tabel 8 and 9, a complete overview of the distribution in frequency f_p and amplitude A_p is given from 0.6 m to 1.3 m for the 8 seconds measurement time. In 0.6 m four paper stickers cannot be analyzed. This can be explained by a saturation of the ToF Sensor by light scattering. The amount of reflected light is too high for the depth measurement. With increasing distance, the oblique black surfaces display errors in the frequency and amplitude measurement. In total, the white paper surfaces could be measured correctly in frequency and amplitude from 0.7 m to 2.6 m . The amplitude distribution in 0.8 m shows the same error as in figure 54. Even in a distance of 1.3 m , the amplitude in the oblique black areas is not higher than 3 mm of the LDV measurement. 3D plotting of the amplitudes can help to recognize patterns in the deformations. Figure 77 shows the 3D plotting by the mesh() function in Matlab of the 0.8 m data. Regions of different amplitude distribution can be distinguished easier. In this illustration, the sinusoidal trend of the error at the edges of the stickers becomes visible. The appearing waveform is at a constant wavelength with various amplitudes. A explanation for this behavior could be fluctuations of the variance between black and white in the pixels when the membrane is moving. To illustrate the frequency and the amplitude distribution in a single image, the frequency is plotted by mesh() and the surface color is determined by the amplitude in figure 76. This delivers the complete dominating dynamic in a single image.

Table 11 shows the frequency distribution f_{ph} of the strongest harmonic peak. It is obvious that only the white paper surface delivers correct harmonic frequencies. In the near distance the second harmonic is stronger in some areas. The 1.2 m and 1.3 m distance delivers a more homogeneous distribution of the first harmonic. Therefore the first becomes stronger

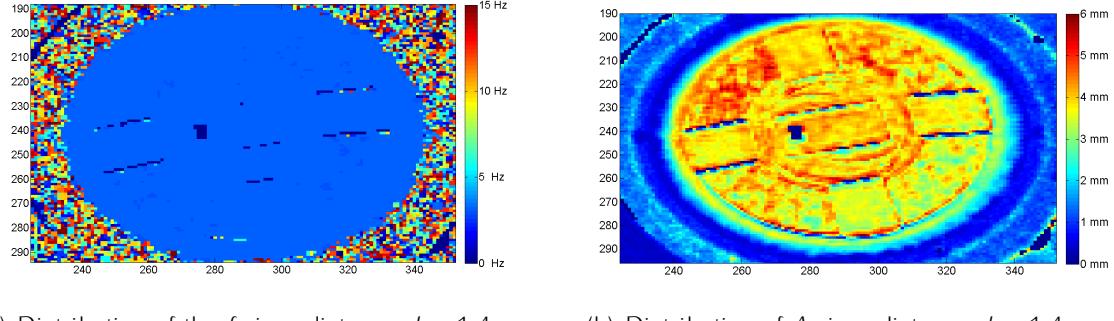


Figure 59: Distribution of the dominating frequency f_p and dominating amplitude A_p

than the second harmonic. Comparing this with figure 54, it seems that the areas where correct harmonics are indicated, is a measure for the accuracy of the complete FFT.

7.6.2 Variable Angle

For the analysis of different angles to the membrane, oscillating at a constant frequency $f_{ref} = 3 \text{ Hz}$ and amplifier volume, LDV is compared on the middle, lower and upper sticker with ToF. From figure 48 and the photograph in figure 51 it becomes evident that the lower sticker has an viewing angle of $\alpha_l = \alpha + 38^\circ$ and the upper sticker of $\alpha_u = \alpha - 38^\circ$. Therefore, multiple parameters can be investigated in a single image. The reference measurement with LDV shows a trend that fits to the expected values given by the physical model in section 7.1 for the 1.4 m and 2.0 m distance. The amplitude A_p decreases with the cousins of the angle α . The ToF results on the lower sticker can follow the trend of LDV with a small offset of about -0.2 mm up to 20° in both distances. From 35° to 45° this gets even smaller. This effect can be explained by the reduction of the normal angle to the image sensor with higher angles up to 38° . The middle and upper sticker delivers an overestimating of the amplitude with increasing angle. The sequence of images in table 7 shows regions of high amplitudes up to 10 mm on the sicker. The inhomogeneous distribution between single pixels is reduced by the 4×4 pixel windows. It is used to plot the results from figure 60 and 61. In the 2 m distance, the trend is similar up to 20° . The middle sticker loses the correct frequency measurement at 30° . Noise overlies the actual dynamic of the membrane. The middle surface however still delivers the correct frequency with amplitude beneath the LDV values up to an angle of 25° . This fits to the physical model. At viewing angles $\alpha > 25^\circ$, the size of the pixels play an additional role. Various distances are measured inside one pixel. The area observed in the window increases. Therefore, the model does not apply anymore. Another problem is that the analysis of the high angles is done on a

inhomogeneous distributions. The value strongly depends on the position where the window was placed by the user. Tabel 6 delivers a detailed overview on the regions that can be investigated in frequency and therefore in amplitude. The black mat region around the sticker in the middle of the membrane provides the right frequency in all distances from 0° to 45°. Only the upper sticker shows variation in the frequency on single pixel under high angles in the long distance.

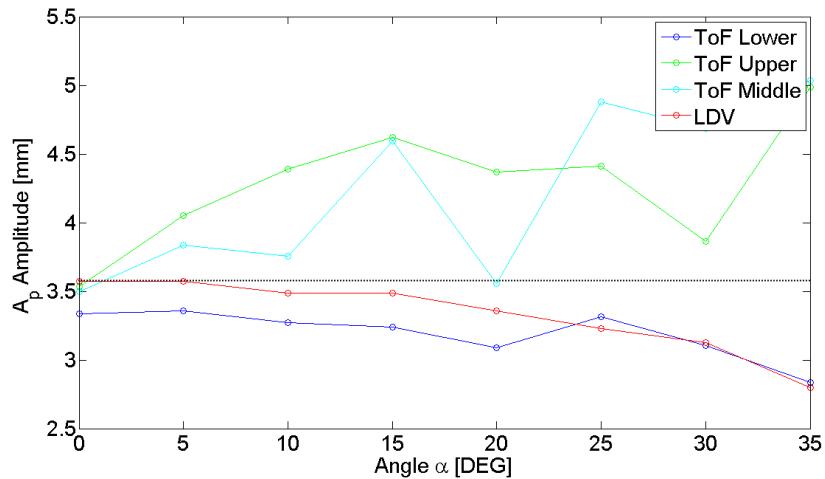


Figure 60: Dominating amplitude A_p under variable angle in 1.4 m

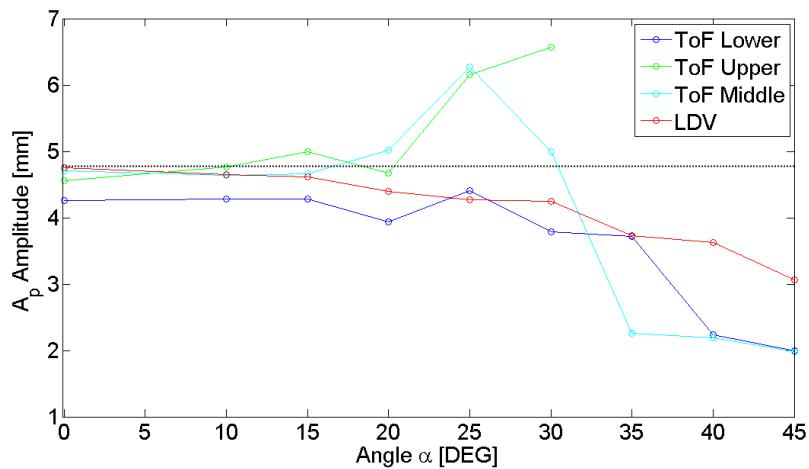


Figure 61: Dominating amplitude A_p under variable angle in 2 m

7.6.3 Variable Frequency

The measurement of different steady frequencies at a constant volume over a time span of 3 s is plotted in figure 62 and figure 63. The measurement at a frequency of 4 Hz shows

the strongest variation in the amplitude A_p of about 1.4 mm. At 2 Hz, the strongest peak in the amplitude is recognizable with 6.6 mm ToF and 5.6 mm LDV. This behavior matches the observations by sight. After that, the amplitude decreases steadily down to 3.7 mm at 10 Hz. The error in the amplitude is also reduced with higher amplitudes. This matches the results from figure 54. More periods are acquired in the 3 s at higher frequencies. Therefore, the actual amplitude can be estimated better. The frequency measurement matches between LDV, ToF and the Waveform Generator for every data point. ToF has an error in the period smaller than $T_{sample} = 0.033$ s. A connection between the error and the frequency is not visible on the white markers. The frequency distribution in table 10 contrary, shows more errors on the black oblique surface with higher oscillating frequency.

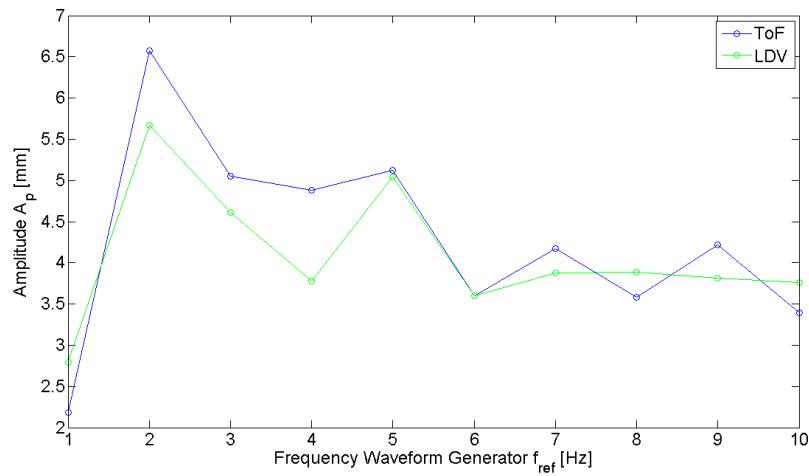


Figure 62: Amplitude A_p versus different oscillation frequencies f_{ref} of the membrane

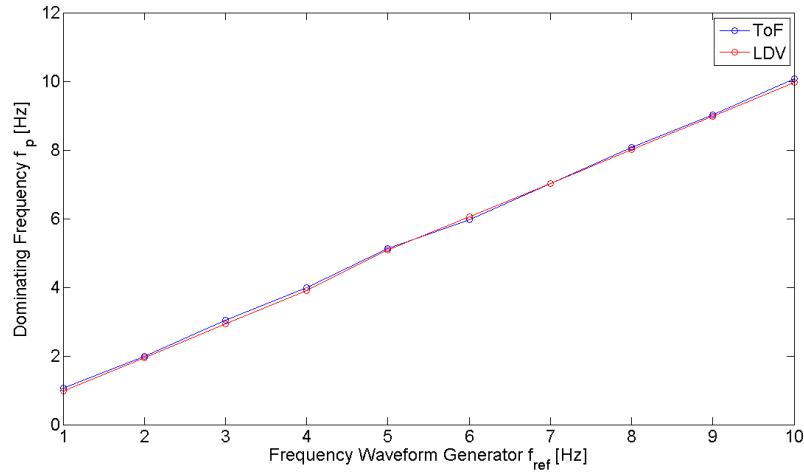


Figure 63: Frequency f_p versus different oscillation frequencies f_{ref} of the membrane

7.6.4 Variable Displacement

The analysis of the variable displacement offers the opportunity to directly compare deflections on the time axis. Ideally, the samples should be phase synchronized for a direct comparison, but since both measurement are running in free-run mode, this is not possible. Figure 66 shows the ToF data in a time window of 5 seconds. At the beginning and the end of the signal where no deflection appears, the noise is visible. In the LDV data, plotted in figure 65, the variance in the measurement is not recognizable at the start and end of the signal. The signal course shows a sinusoidal trend without any abnormality. The deflection in the positive direction seems to be minimal higher than in the negative direction. In both signals, the 12 deflections can be identified and assigned. The low sampling rate of the ToF measurement causes a mismatch in the amplitudes. To compensate the difference in the offset between both signals the peak-to-peak A_{pp} amplitudes are plotted against each other for every deflection in figure 64. ToF shows an overestimating between 1 and 2 mm until deflection 9, that can be explained by the high variance of ToF. Subsequently, the ToF data underestimates the true LDV data. This point of change is the same deflection where both A_{pp} values are decreasing. In figure 68 and figure 69, the frequency domain is plotted. ToF overestimates the amplitude by 1.428 mm. The first and second harmonic can still be found in the ToF spectrum. Also, the offset peak is visible in both data.

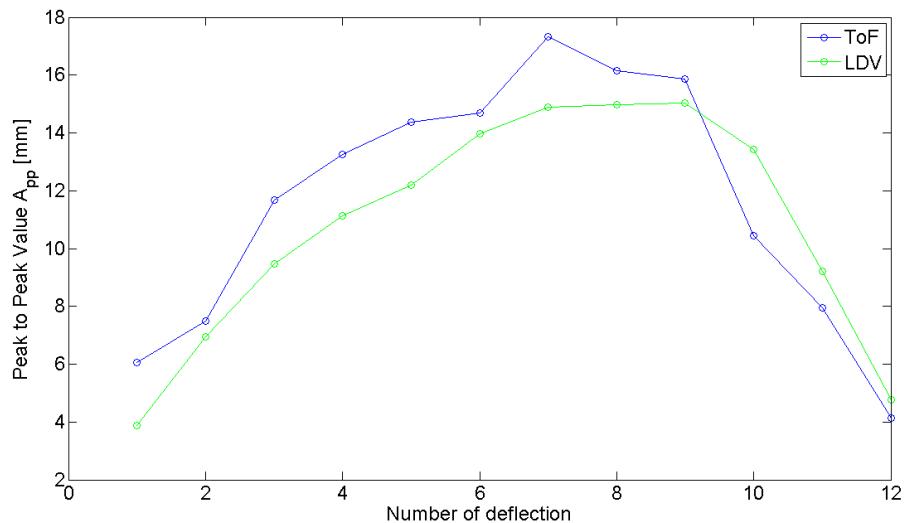


Figure 64: Direct deflection comparison in peak-to-peak A_{pp} value.

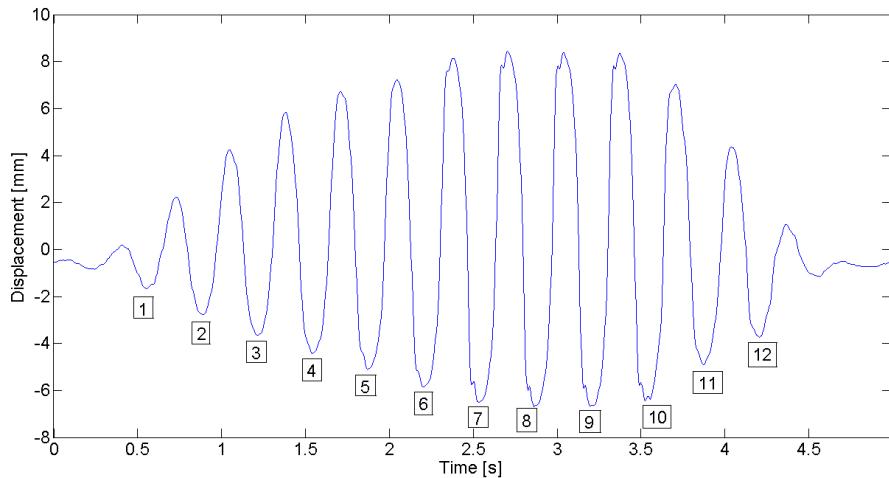


Figure 65: Variable displacement of the LDV data 1.4 m distance

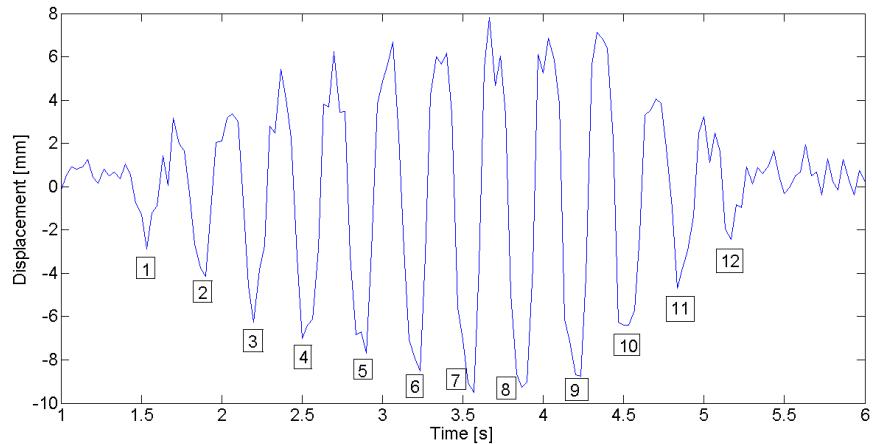


Figure 66: Variable displacement of the ToF data in 1.4 m distance

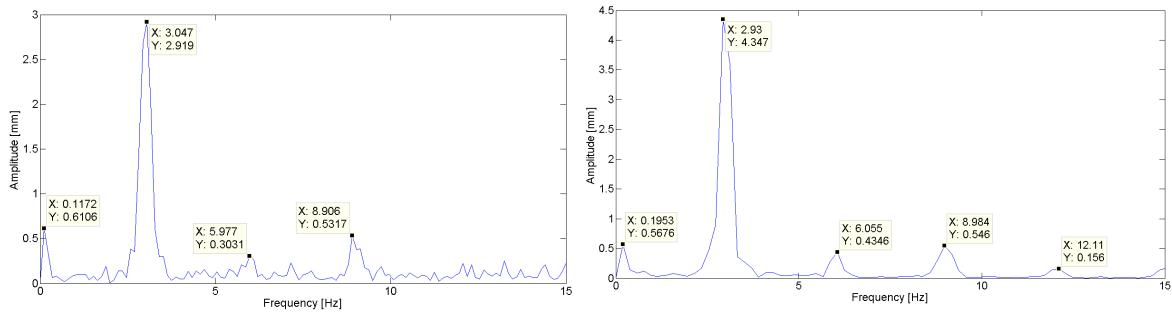


Figure 67: ToF and LDV comparision of 3 Hz oscillation in 1.4 m distance

7.6.5 Sweep

The sweep is a simple way to analyze multiple frequencies in a single measurement in order to derive the behavior of a system in a broad band. Figure 68 plots the Time domain of the sweep from 1 Hz to 10 Hz . The offset of the sinus trend is constant over the complete time. The quantization error and the variance cause a fluctuation in the amplitude. The LDV data, in contrast, has an offset that is unsteady over time. A reason could be a numerical integration error or noise that was not properly filtered by a high-pass filter before the analysis.

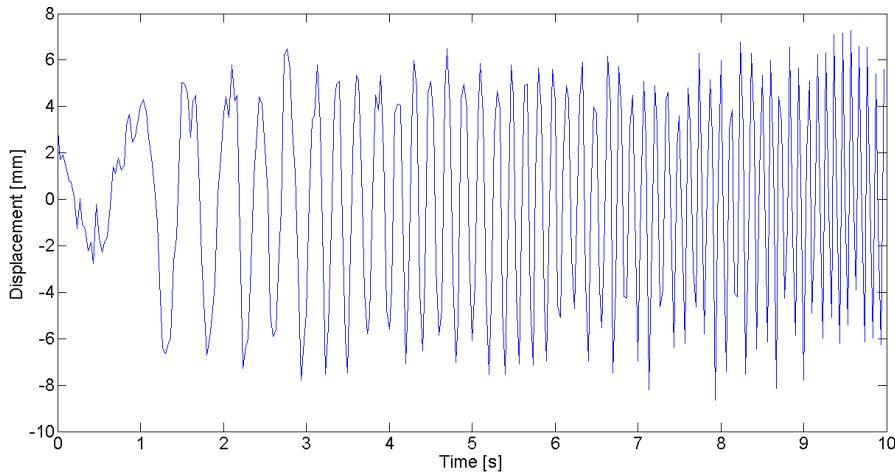


Figure 68: 1 Hz to 10 Hz sweep in 1.4 m distance captured with ToF

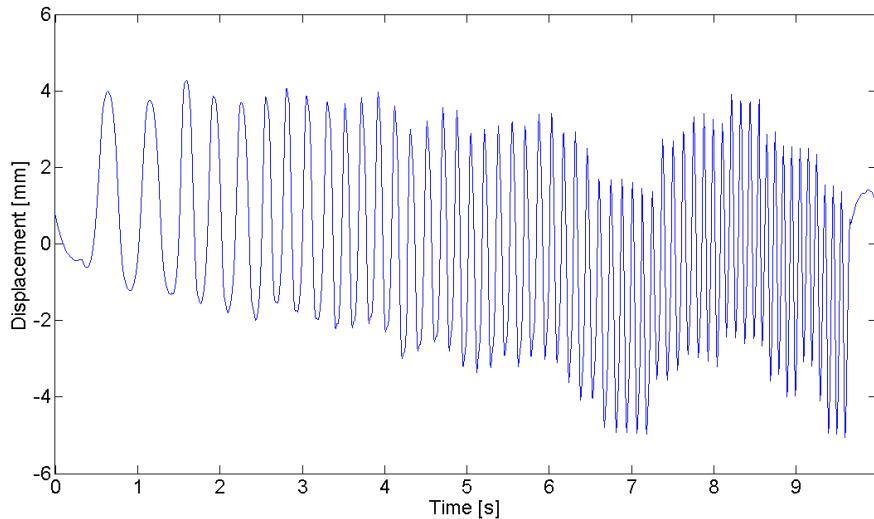


Figure 69: 1 Hz to 10 Hz sweep in 1.4 m distance captured with LDV

Since a sweep contains all frequencies in the spectrum, the standard Fourier transform of the complete signal shows superimpositions between the single peaks. An analysis for separated frequencies is not possible. For that reason, the signal is analyzed by the Short-time Fourier transform and plotted in a spectrogram. The 10 seconds of measurement time are separated into 9 slots for the frequencies of 1 Hz to 10 Hz . This is done by a fixed number of samples for the processing represented by NFFT of 128 for LDV and 32 for ToF. A hanning window is used to reduce the influence at the window edges. Figure 70 represents the spectrogram of the ToF data. Every step from 1 Hz to 10 Hz is visible without any offset. The 1 Hz range indicates a reduced amplitude that is also visible in the time domain. Afterwards the amplitudes stay stable at about 45 dB . Harmonic frequencies can be identified for some ranges with amplitudes of about 15 dB .

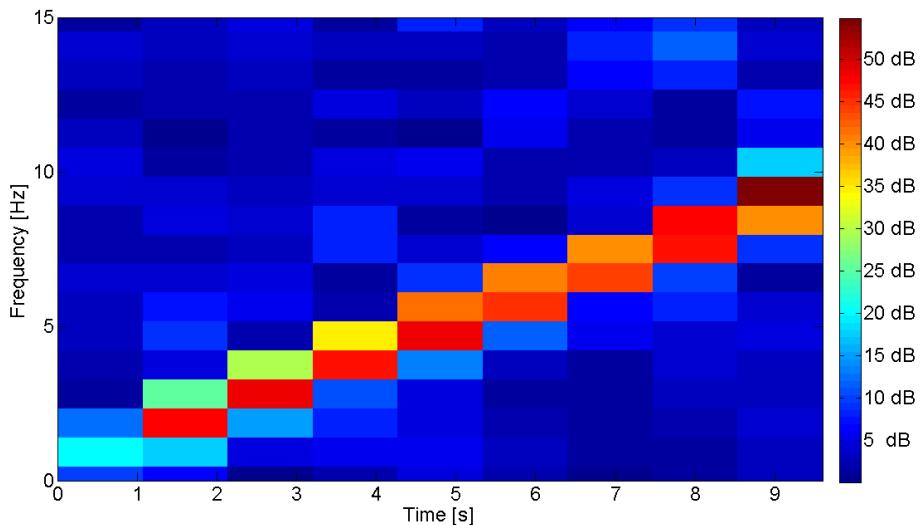


Figure 70: Spectrogram of the sweep ToF data in 1.4 m distance

Figure 71 shows the same sweep measured by LDV. In contrast to ToF, an offset of the signal appears at 1 Hz , 2 Hz , 3 Hz , 7 Hz and 9 Hz that is also visible in the time domain. Apart from that, every frequency is identified. The difference between the amplitudes is caused by a number of samples that are processed. Due to the low variance and high resolution of LDV, the first, second and third harmonic for every frequency are represented in the spectrogram.

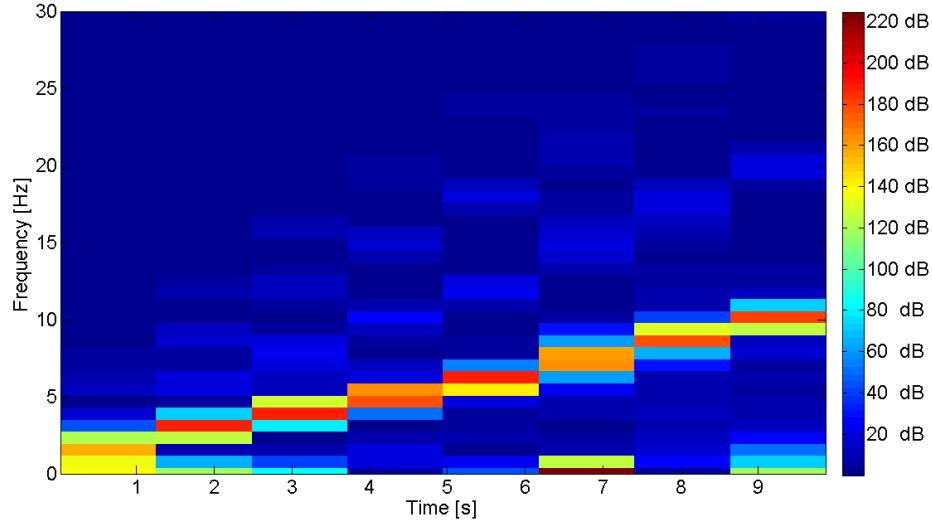


Figure 71: Spectrogram of the sweep LDV Data in 1.4 m distance

7.6.6 Influence of Background Lighting on Dynamic Measurement

For estimating the influence of the 2 kW halogen lighting, which is also done for the static in section 6.3, the distribution of the harmonics will be used as an indicator for the quality of the measurement. Figure 72(a) clarifies the distribution of the harmonics of a 3 Hz oscillation in 1.4 m distance under ideal dark conditions. The white paper surface delivers the first and the second harmonic. Figure 72(b), in contrast, shows the impact of 2000 W halogen lighting. The harmonics are overlaid by the increase of the variance.

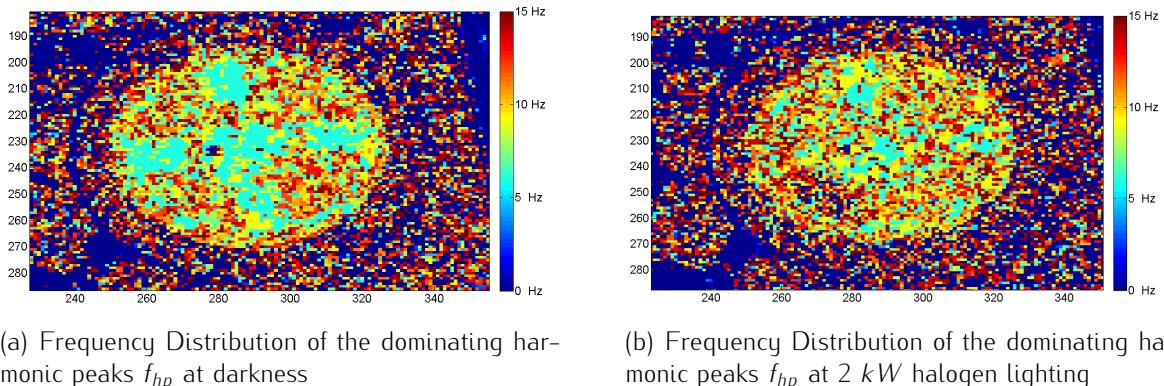


Figure 72: Harmonic frequency distribution f_{hp}

Looking at the frequency distribution in figure 73 of the dominating peak shows that especially on the black oblique the frequency is not estimated correctly. The influence is explicit in the left lower corner. A spread reflection from the halogen lighting directly points into the camera.

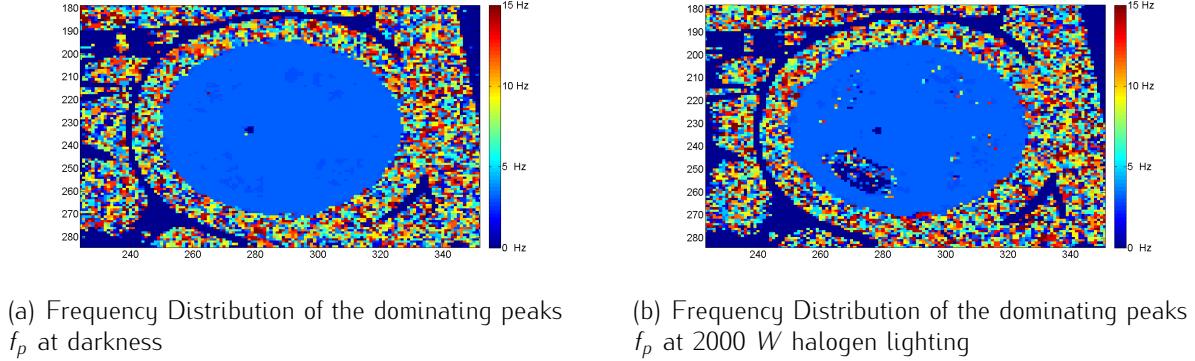


Figure 73: Dominating frequency distribution f_p

From the amplitude distribution represented in figure 74 it becomes obvious, that the markers are influenced less than the black oblique region. The dynamic of the static part around the membrane also increases in its amplitude. The results match to the static investigation that has been done. The high amount of IR lighting disturbs the dynamic measurement. The noise level is increased and superimposes the measurement of the dynamic.

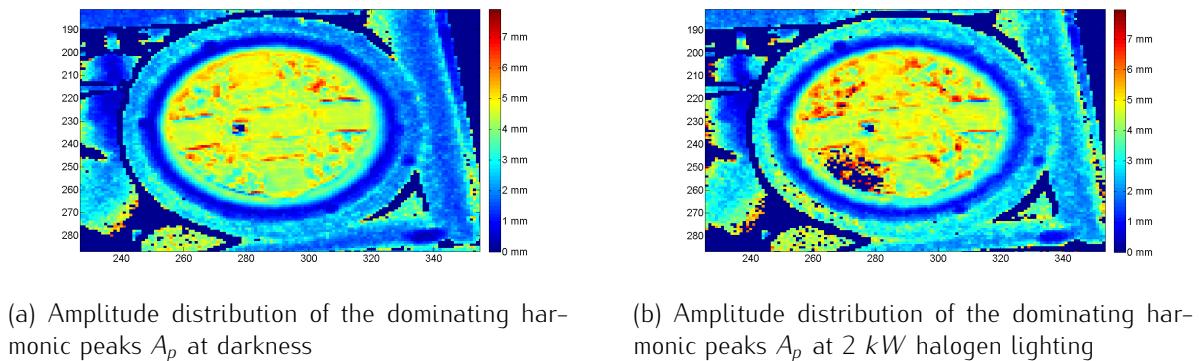


Figure 74: Amplitude distribution A_p

7.7 Conclusion

Even if the experimental setup and analysis does not correspond to all scientific standards, the overall result can be used to get a wide overview on the dynamic behavior of ToF in comparison to LDV. The variable distance setup shows errors in the LDV and ToF measurement, caused by systematic errors like range distortion or non systematic errors like light scattering. The unsteady membrane vibration additionally influences the interpretation of the data. A total fluctuation of 4 mm in the amplitude is recognizable in the reference LDV data of the variable distance. Electrical noise in the experimental setup and the quality of the membrane itself could be a reason for that behavior. The fact that the Kinect 2 functionality and configuration is not completely open is another challenge in the analysis. The detailed preprocessing of the images inside the camera is still unknown. Nevertheless, the device is able to measure the dynamic of an oscillating system starting at an amplitude of 1 mm up to a frequency of 10 Hz . The distance of about 1.4 m to the camera provides the smallest error of 0.05 mm at a measurement time of 8 s and amplitude of 4 mm . In theory, a measurement up to 15 Hz is possible by the Nyquist Shannon sampling theorem. The ToF data delivers the correct frequency under high viewing angles α up to 45° on highly diffuse reflecting surfaces like paper in the middle of the membrane. The amplitudes are overestimated with increasing viewing angles. In contrast, LDV shows an underestimating of the amplitude A_p with higher viewing angle. The speed vector in the direction of the laser decreases. The sweep analysis indicates that even amplitude and frequency are reconstructed in very short time spans of about 1 s from 1 Hz to 10 Hz . It was possible to synchronize a point in time of two measurements with the variable displacement function. Unsteady processes could be compared in both data sets without a precise synchronization of the LDV sampling rate and exposure time of the ToF camera. A direct comparison between single displacements showed a maximal difference of 2 mm in the peak-to-peak value A_{pp} . The measurement of the 3 Hz oscillation confirms this result. The amplitudes A_p are estimated from 0.6 m to 2.6 m with a maximum difference of 1 mm . Even the first and second harmonic are reconstructed on the white paper surface. The distribution of the harmonics is an indicator for the accuracy of the measurement. The influence of background IR lighting shows a strong impact on the black oblique surface. The vibration dynamics disappear in the noise floor. The dominating frequency is still identifiable, but the second harmonic disappears on the white paper surfaces. In general, most hypotheses could be proven with a small exception. The advantages of the TVD filter is not clearly visible in the variable distance measurement. A reason for that could be processing of the raw data with the wrong parameters. Intensity images and the ToF phase shift are necessary to apply the

filter in the correct manner.

Dynamic investigations with ToF deliver a high improvements compared to other methods like strain gauge, stereo triangulation or LDV. Complete surfaces can be investigated fast and in parallel without a direct contact to the test body. Only a highly diffuse reflectance surface under a low viewing angle $\alpha < 45^\circ$ must be guaranteed for a high signal-to-noise ratio. Apart from that, the Kinect 2 delivers a precise low cost device for dynamic investigations. The images can easily be processed with Matlab. A complex calibration is not necessary. The frequency and amplitude distribution delivers a high validation of a surface dynamic.

8 Outlook

8.1 Advancements in the Experimental Setup and Analysis

A scientific or industrial rated ToF camera is indispensable for future precise analysis for deformation measurement. A real raw data access is necessary for a widespread post-processing. Low level access for configurations like the IR modulation frequency f_{mod} or the integration time Δt are vital for adjusting the camera to different applications. A hardware trigger input enables a precise synchronization between LDV and ToF. Therefore, individual measurement samples can be compared between both data sets. An exchangeable lens is also an important advantage. The field of view can be adjusted to the region of interest. This also requires a modification of the illumination to the new field of view. The odos real.iZ-1K ToF camera, visible in figure 26, fulfills all these requirements with the highest number of pixels that can actually be found on the market. The high speed of 100 fps enables the investigation of an ordinary membrane up to 50 Hz . Therefore, the range of most common structural vibrations is covered.

As test body the subwoofer membrane is not the ideally device. A professional shaker for structural investigations would be a better solution, because frequency and amplitude can be adjusted precisely and held constant over time. The configured parameters deliver an additional reference to the LDV measurement. The superimposing of multiple vibrations enables the investigation of complex excitations between different sources. A high-end subwoofer in a high noise free experimental setup at audible operations frequency could also deliver a more stable behavior. With higher frame rates of the camera the membrane can be investigated in the designed operation range.

For a more precisely acquisition of the LDV data, an oscilloscope with a higher data storage capability is necessary. Consequently, the data can be captured with a higher sampling rate over a longer timespan. With frequencies above 10 Hz , the trigger mode can be used for an automatic acquisition of the data. Additional investigations in the LDV measurement is necessary. The reason why some data is lost is still unknown. A low pass filter for the

filtering of the offset should deliver a signal that can be compared better to the ToF data.

One strong error source in the experimental setup is the user defined measurement window with the mouse cursor. Since the white marker delivers a clear border to the black background, it should be possible to automatically detect the position and angle of the windows in every image. Position and number of pixels that are averaged can be fitted to the size of the marker. The existence of the harmonic frequencies in the spectrum could additionally be used to filter pixels with a low SNR out of the measurement window. The distribution of the harmonic peaks shows, that regions with a high SNR correspond to regions with a high reflectivity and therefore to higher accuracy. The distribution of the harmonics depending on the viewing angle is another investigation that delivers more information on this experimental setup.

The creation of the frequency and amplitude distribution images offers improvements: The images could be filtered by a proper TVD or a common image filter to improve the validity. Other algorithms for the filtering of the images or the resulting depth could also help to increase the quality of the data.

A direct interface between the camera and the software is desirable for dealing with Kinect 2 inside of Matlab. Therefore, a direct live processing of the data is possible. This should simplify the scientific work with the camera remarkably. With the near-field firmware update, the range is reduced from 10 cm to 1 m [33]. This option is also very interesting for the deformation measurement. Since the accuracy of ToF decreases with the distance, areas can be analyzed with a very high depth accuracy at a high number of pixels. The RGB images of the camera in combination with the ToF images deliver another way for increasing the depth accuracy. Therefore, a combination between stereo triangulation and ToF is possible [34].

8.2 Static Calibration and Angle Calculation

The static calibration of a ToF system is an important step that has to be done before further dynamic investigations. In this thesis this was only possible by some rough estimations in a rough experimental setup. The influence of different surfaces, geometries and perspective delivers the base for the dynamic investigation. When a geometry cannot be measured in the static range, a dynamic investigation is not possible. The checkerboard with different colors and surface materials could be a first step. With a complex system like odos real.iZ-1K, it is anyways necessary to calibrate the camera, since depth data depends on the field of view and other configurations. A calibration of the Kinect 2 was not necessary, because the camera is already calibrated by Microsoft to a fixed configuration.

The influence of the variance versus the angle to the surface normal vector is an important investigation which is missing in this thesis. Only with this data, the dynamic behavior can be fully interpreted. An automatic calculation of the normal surface vector to the optical axis is possible with the gradient function presented in equation 40. Therefore it enables to calculate the angle of the investigated window and the error that results out of this.

8.3 Applications in Structure and Machine Observation

In the next step, the vibration investigations with ToF should be enlarged to a real engine or a dynamic structure like a aircraft wing. With vibration amplitudes higher than 1 mm, the Rotax 912 ULS in the propulsion laboratory of the University of Applied Science Aachen should deliver an ideal test body. The vibrations under different frequencies in multiple directions on a complex surface represents a challenge in the experimental setup and analysis. The observation of machines could be a large application field for ToF. Disturbances in the operation are often indicated by a change in the dynamic behavior. This does also apply to structures like bridges or buildings. A change of the dynamic behavior could be a indicator for a weakening of the structure.

The analysis of an aircraft wing should deliver detailed data of the aerodynamic effects like flatter and other aerodynamic load under different flight conditions. The landing phase is the most interesting situation for an deformation analysis, because most structural damages emerge here. A high resolution, high depth accuracy, long range and high speed ToF measurement should deliver all deformation of the outer geometry. The determination of flight control surfaces in 3D and the resulting angles and deformations delivers detailed information on the aerodynamic behavior.

8.4 Dynamic Investigations of Rotating Parts

The trend to high speed ToF cameras with very short integration times could soon enable the recording of fast rotating parts like turbines or compressors. Polytec announced with the epc660 ToF image sensor a device, that allows up to 4000 fps at 160×30 pixel. In this way the movement of the blades is recorded multiple times per rotation in 3D. Geometries and deformations of the surfaces are acquired contact-less and parallel. For a detailed investigation of the rotation, a synchronization between the camera and the rotor is necessary to capture the 3D at defined positions. A laser barrier or another optical trigger can deliver a very precise and fast trigger for the camera setup.

One major problem that will occur is motion blur that depends on the integration time. Equation 20 delivers the connection between the exposure time, comparable to the integration time, and the motion blur that occurs. Assuming a blade tip speed of 340 m/s and an allowed motion blur of 0.1 mm the integration time needs to be $0.29 \mu\text{s}$. A very strong ToF light source will be necessary to create sufficient light for a high SNR ratio leading to a proper 3D scan.

8.5 Application in Aerodynamic and Structural Test Facilities

Test facilities for aerodynamic and structural investigations have a high need for a detailed investigation of the surface dynamic and geometry. Since ToF cameras work with no forcing need of markers, static and dynamic investigation are possible with a single device from one point of view. Without the disturbing markers, the boundary layers are not influenced anymore. The small size and construction without any mechanical parts enables various integration possibilities. From the 3D data multiple informations like angles, distances or dynamics can be captured from a single measurement.

The 3D data can additionally be used to integrate other measurements devices into a common point cloud. Therefore a ToF camera enables the integration of different optical measurement systems like Particle Image Velocimetry (PIV) or Pressure Sensitive Paint (PSP) into a common 3D mesh. With a proper synchronization between all devices the integration of all optical measurement data into one time line is possible. In such an experimental setup the deformation, aerodynamic boundary layer and vortexes can be compared to each other. The ToF data delivers in real time all geometries and dynamics, which are necessary to integrate the other measurements. The field of view, viewing angle and position of every camera must be calibrated for this setup. Figure 75 illustrates a combination between PIV

and PSP for aerodynamic investigations on a delta wing. The angular velocity ω is given in the planes of the PIV measurement. PSP delivers the pressure distribution of the complete surface. With a additional ToF measurement all deformations, angles of control surface and geometries are integrated. This allows the connection between surface dynamic and aerodynamic.

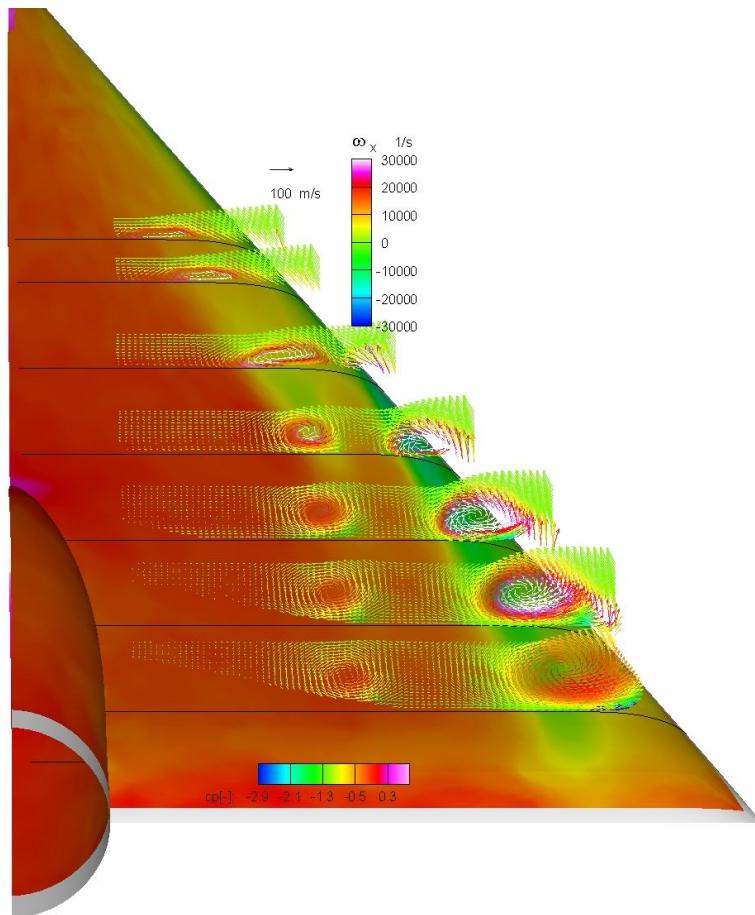


Figure 75: PIV and PSP on a delta wing DLR Institute of Aerodynamics and Flow Technology 2014

A Annex

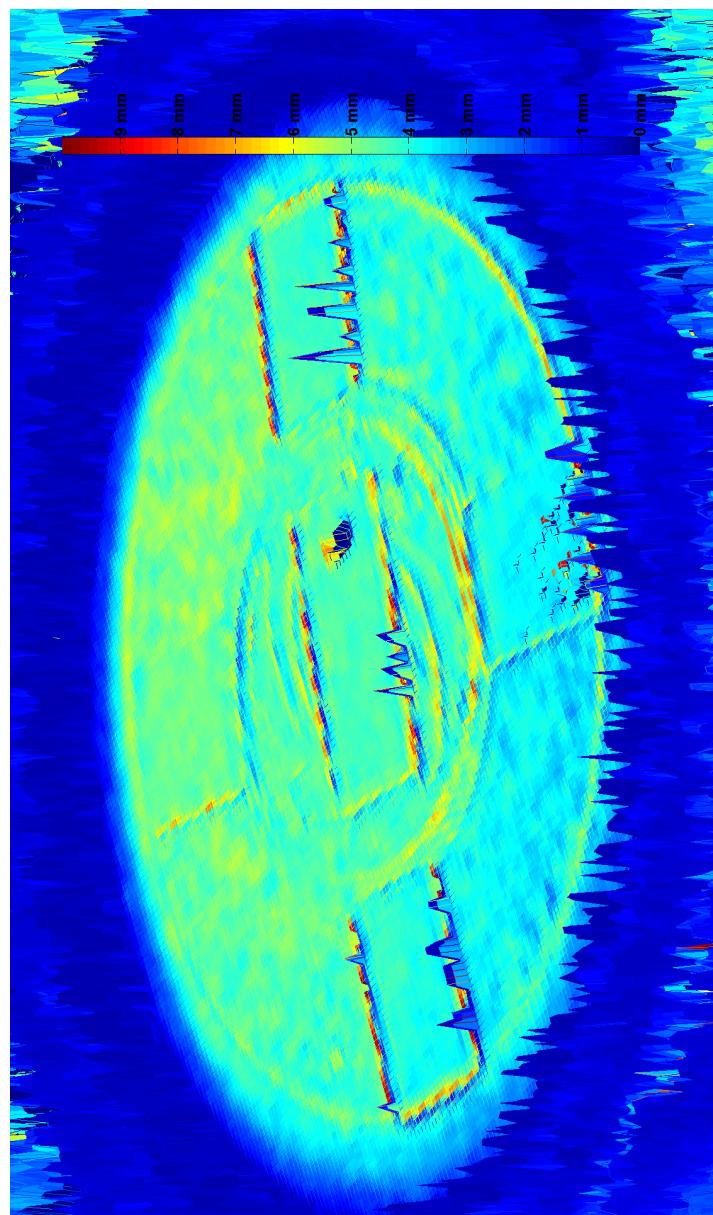


Figure 76: Distribution of dominating amplitudes (color) and frequency (z-axis) in 0.7 m distance after 8 seconds

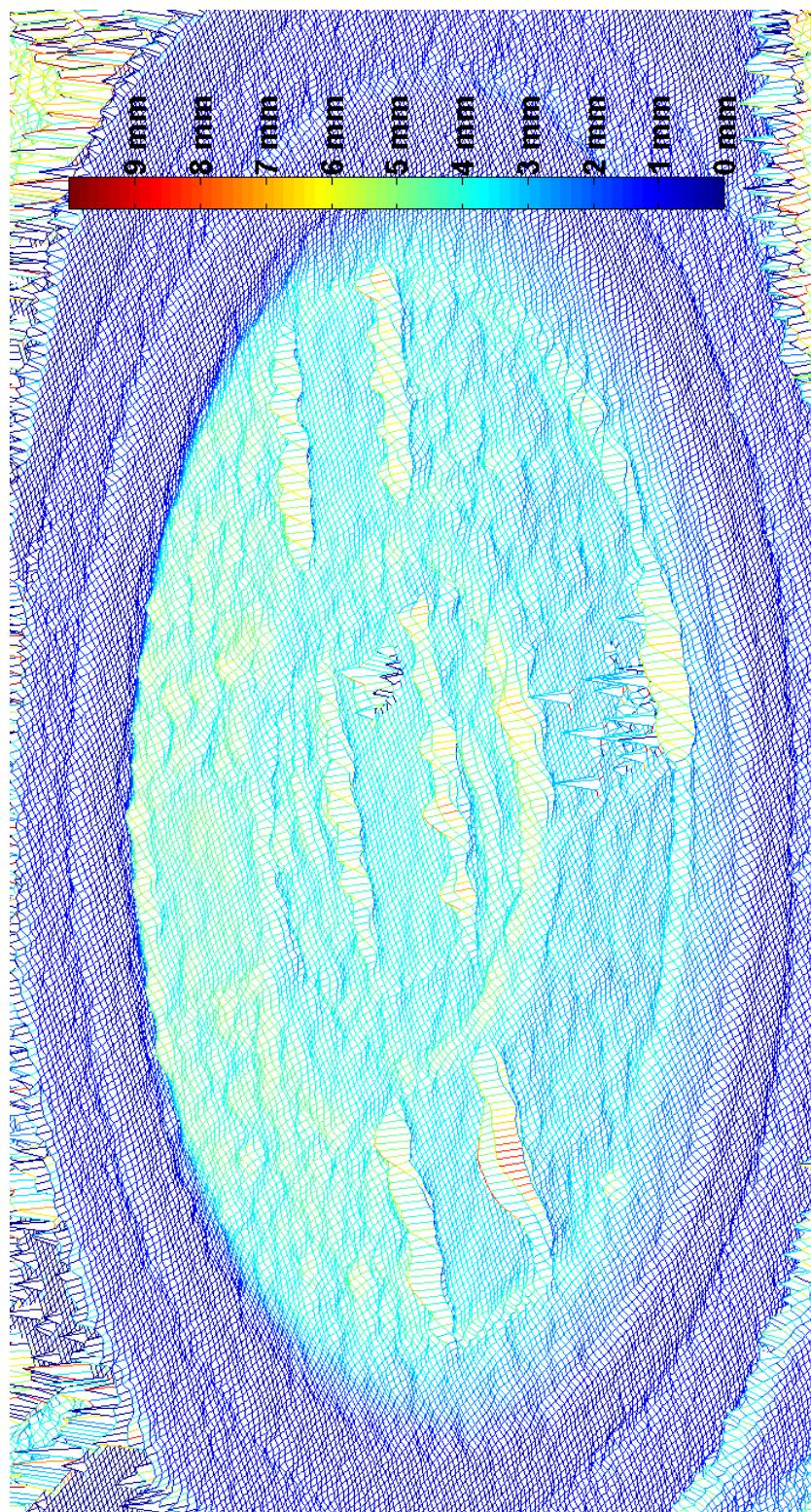


Figure 77: Distribution of dominating amplitudes (z-axis) in 0.7 m after 8 seconds

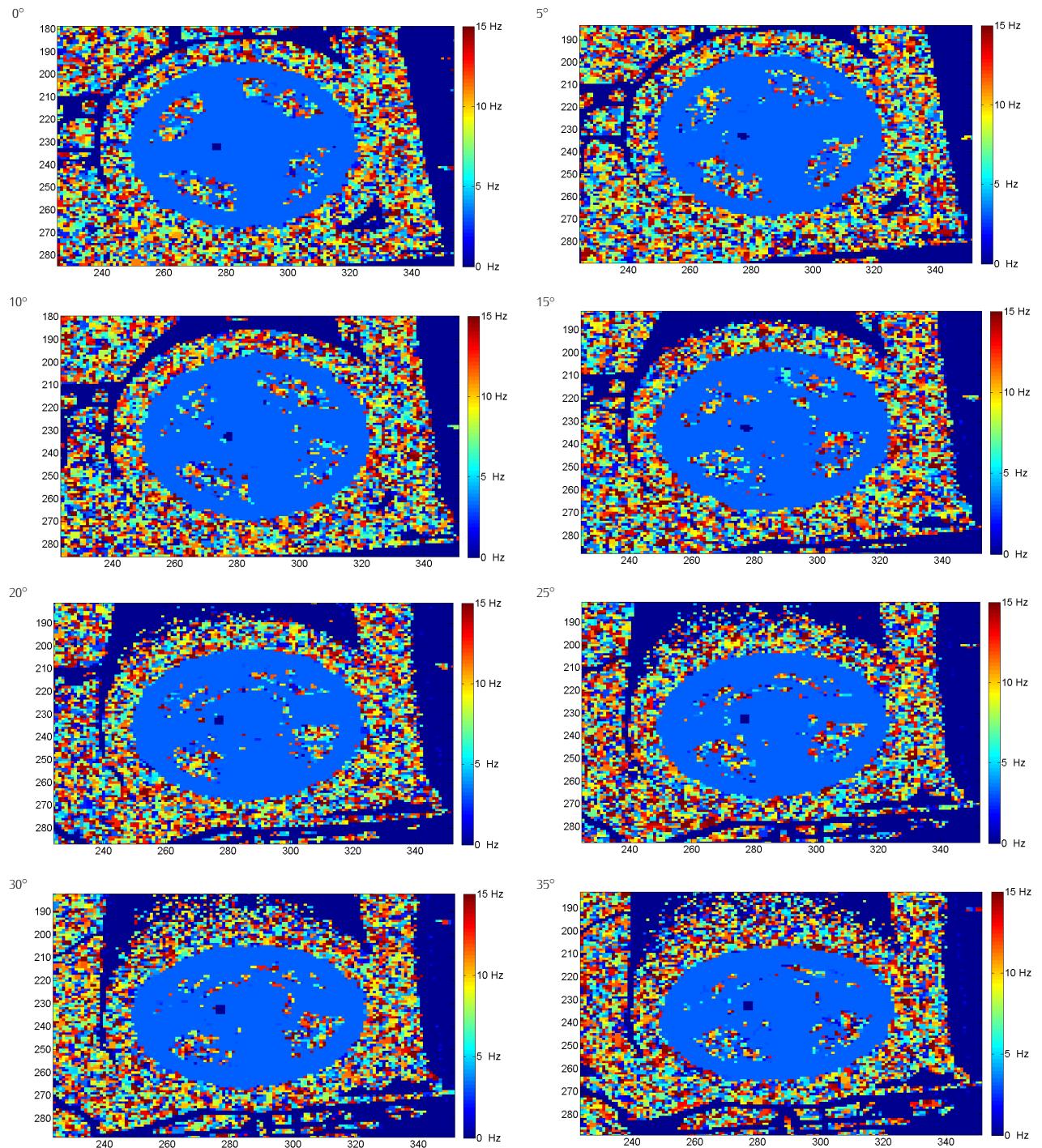


Table 6: Dominating frequency f_p for 0° to 35° at 1.4 m after 8 s

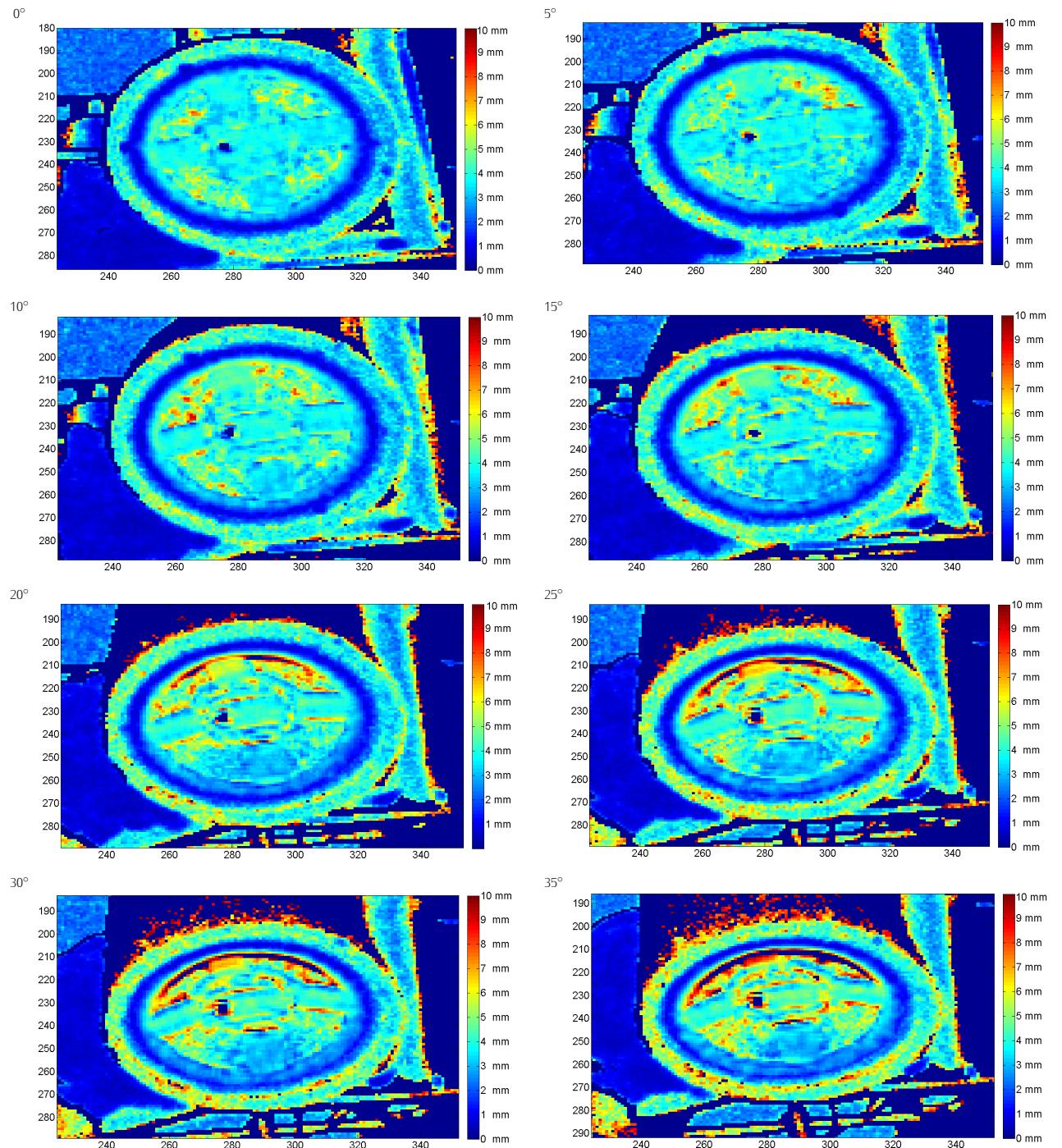


Table 7: Dominating amplitude A_p for 0° to 35° at 1.4 m after 8 s

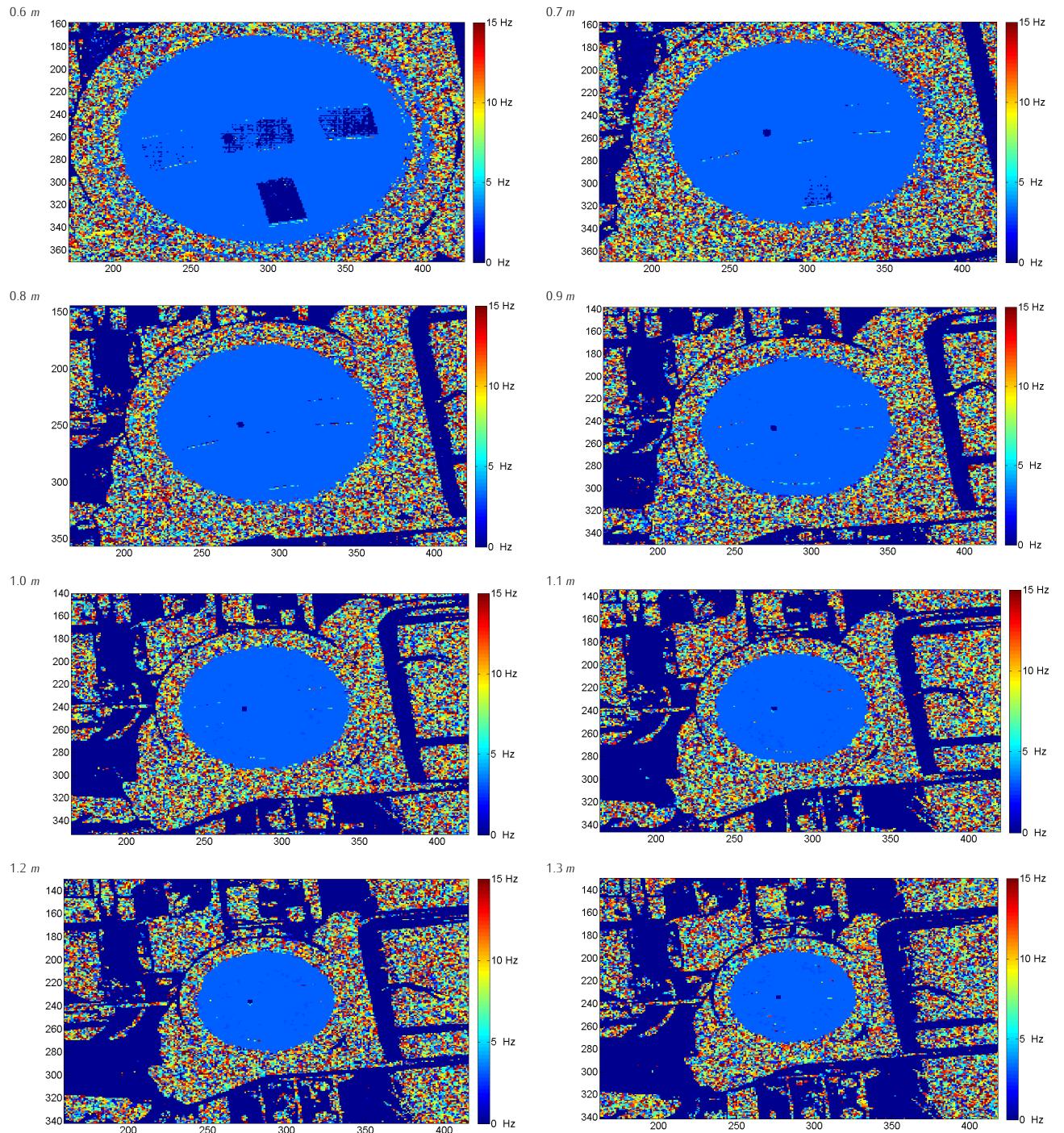


Table 8: Dominating frequency f_p in a distance of 0.6 m to 1.3 m after 8 s

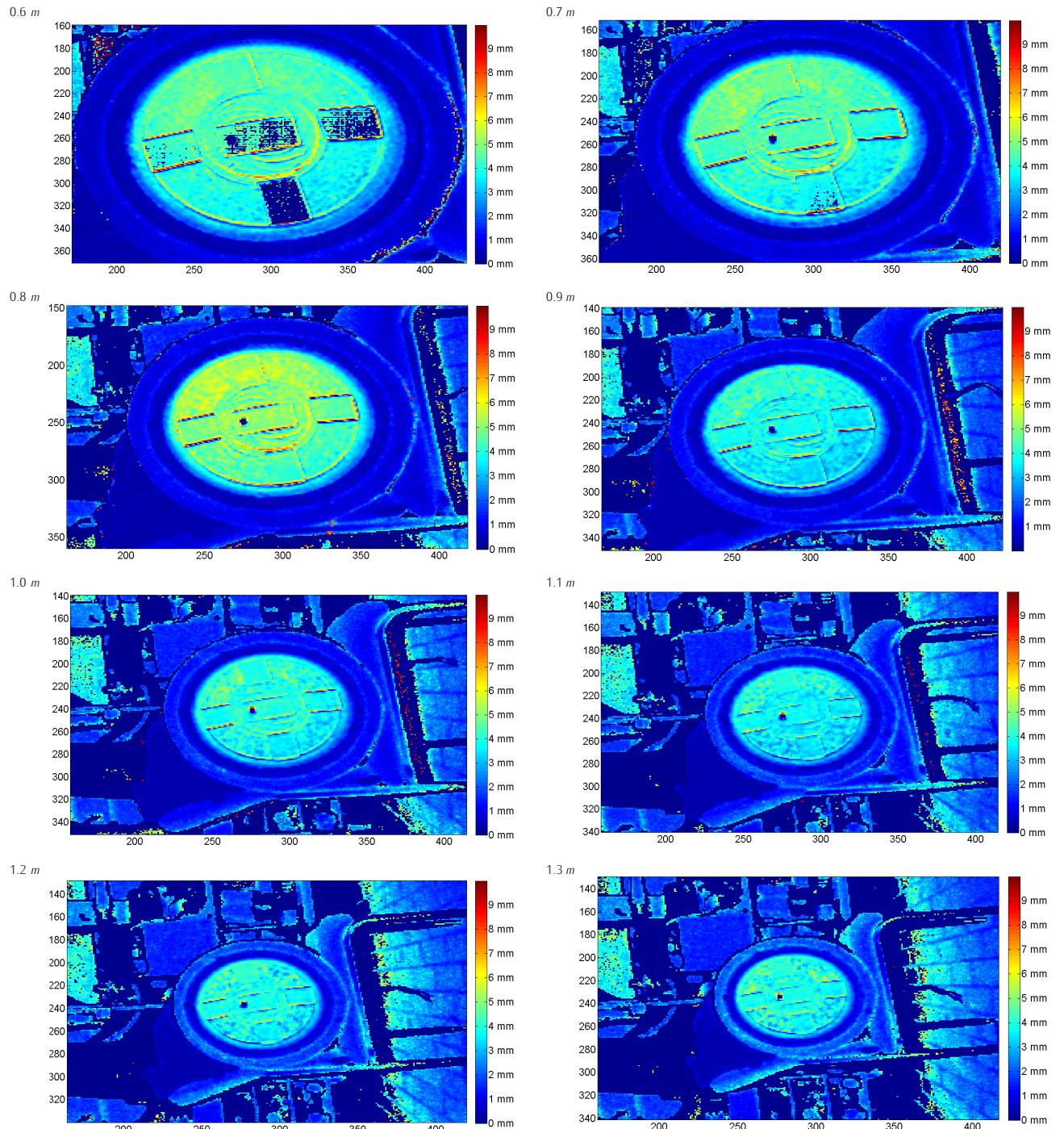


Table 9: Dominating amplitude A_p in a distance of 0.6 m to 1.3 m after 8 s

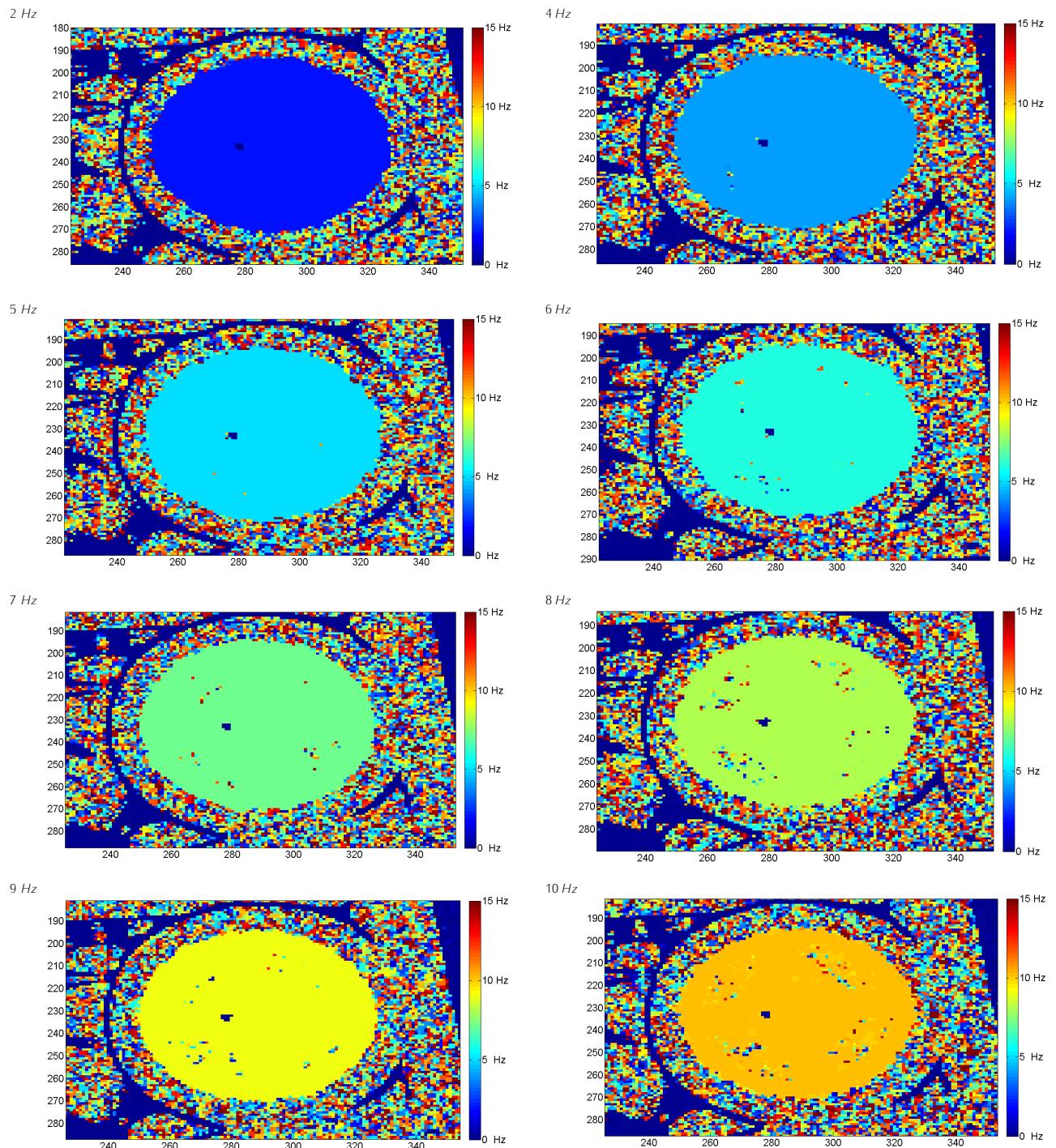


Table 10: Dominating frequency distribution f_p of vibrations from 2 Hz to 10 Hz after 3 s

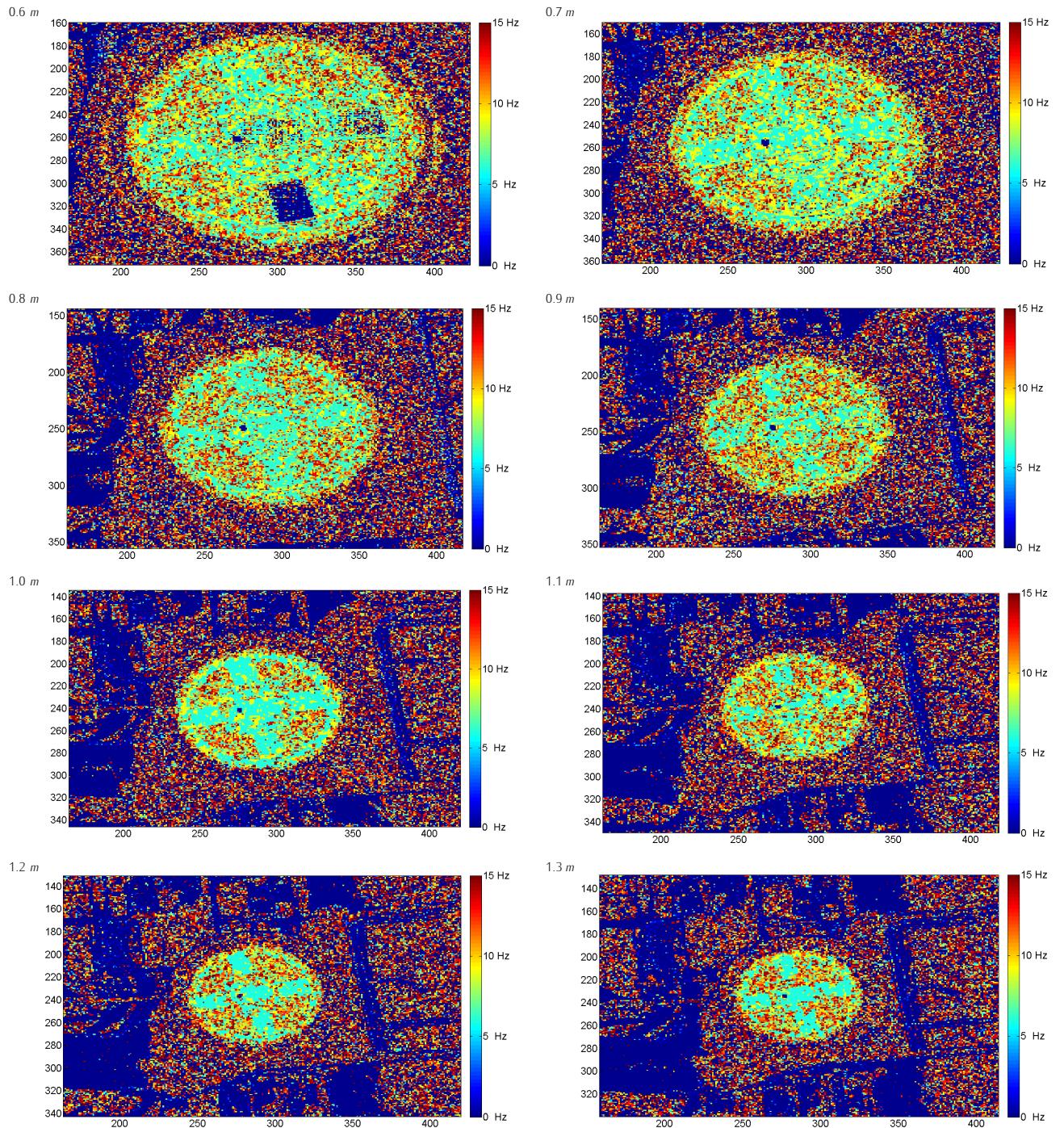


Table 11: Dominating harmonic frequency f_{hp} distribution of 3 Hz oscillation in a distance of 0.6 m to 1.3 m after 8 s



Figure 78: Influence of 2000 W halogen lighting on the RGB image

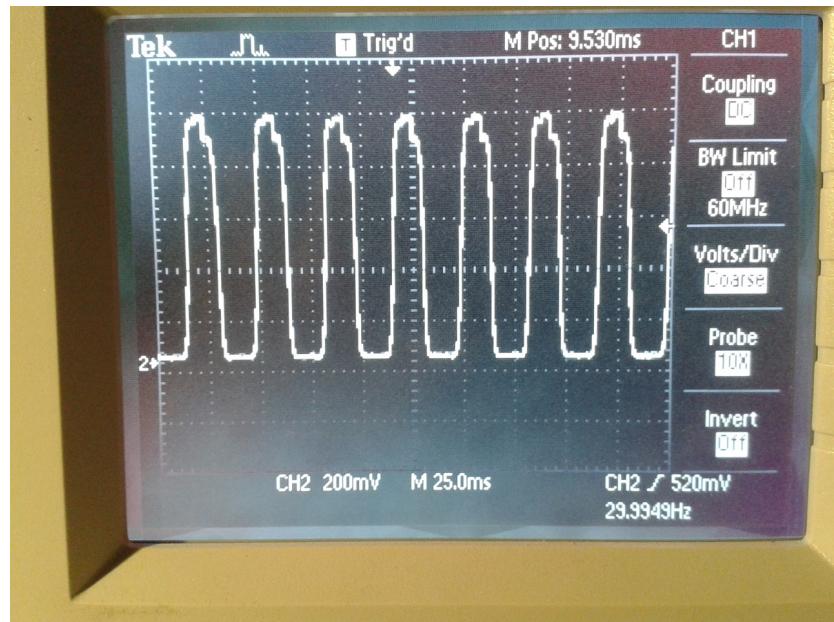


Figure 79: Modulation of Kinect 2 IR lighting measured with a photodiode

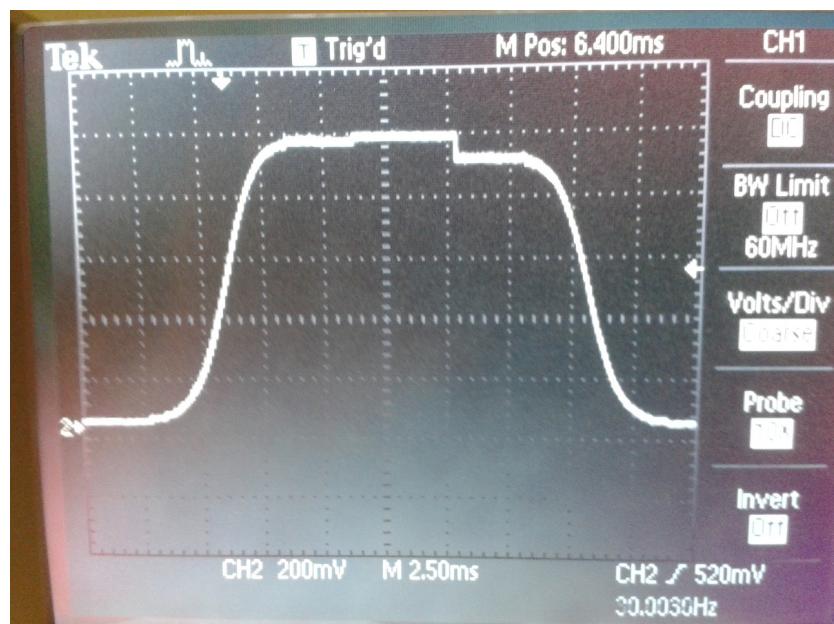


Figure 80: Three different light intensities in one image frame

B Source code extracts

```
% FFT and variance
for row = 1:424
    for column = 1:512
        for g= 1:L
            DynPixel(g)= FramearrayDyn(uint16(row),uint16(column),g);
        end
        DynPixel=DynPixel-mean(DynPixel);

        %Variance
        V = var(DynPixel)

        % FFT
        Y = fft(DynPixel,NFFT)/(L);
        Y_AmpsDyn = abs(Y(1:NFFT/2+1));

        % Find peaks in the spectrum
        [peakamplitude, peakindex]= max(4*Y_AmpsDyn);
        freq=f(peakindex);

        % Color Pixel in image
        FreqMap(row,column) = freq;
        AmpMap(row,column) = peakamplitude;
        VarMap(row,column) = V;

    end
end
```

Listing 1: Creation of variance,amplitude and frequency images

```
% Scope and Laser Doppler Vibrometer Data
sensitivity = 125 %(mm/s)/V
samples = 1001; % zero based
TimePerDIV = 0.2% $s$ 
samplingrate_scope = (samples-1)/(TimePerDIV*10)

csvFilesS=dir(strcat(ImageFolder,'*.csv'));
if ~isempty(csvFilesS)
second=csvread(strcat(ImageFolder,'\',csvFilesS.name),2,0,[2,0,samples-0]);
Volt=csvread(strcat(ImageFolder,'\',csvFilesS.name),2,1,[2,1,samples-1]);
%Remove Offset
speed =(Volt-mean(Volt))*sensitivity;

%Integrate the Speed to get the displacement
displacement=cumtrapz(speed)/samplingrate_scope;
%Remove Offset
displacement = displacement-mean(displacement);
%Add time offset
second=second+max(second);
figure;
subplot(2,2,1:2);
plot(second,displacement);
title('Time Domain of Scope Data');
xlabel('Time [s]');
ylabel('Displacement [mm]');
xlim([min(second) max(second)]);

subplot(2,2,3:4);
Lscope = length(displacement);
NFFT_scope = 2^nextpow2(Lscope); % Next power of 2 from length of y
Y = fft(displacement,NFFT_scope)/(Lscope);
f_scope = samplingrate_scope/2*linspace(0,1,NFFT_scope/2+1);
Y_Laser = abs(Y(1:NFFT_scope/2+1));
plot(f_scope,2*Y_Laser);
title('Frequency Domain of Scope Data');
xlabel('Frequency [Hz]');
ylabel('Amplitude [mm]');
xlim([0 15]);
end
```

Listing 2: Evaluation of LDV data from scope

C Sourcecode and Documentation on CD

dynKinect_0_3_thesis_evaluation.m

dynKinect_0_31_harmonics.m

FastVarMap.m

Raw and processed measurement data

Reference papers and documents

Pictures of experimental setup

Latex source code and all figures

Bibliography

- [1] Thierry Oggier. Image sensor technologies for 3d time-of-flight range imaging. page <http://www.imagesensors.org/>, 2009.
- [2] Larry Li. Time-of-flight camera—an introduction. 2014.
- [3] Xiaojuan Qi, Derek D Lichti, Mamdouh El-Badry, Ting On Chan, Sherif Ibrahim El-Halawany, Hervé Lahamy, and Jeremy Steward. Structural dynamic deflection measurement with range cameras. *The Photogrammetric Record*, 29(145):89–107, 2014.
- [4] John Sell and Patrick O'Connor. The xbox one system on a chip and kinect sensor. 2014.
- [5] Sonam Jamtsho. Geometric modelling of 3d range cameras and their application for structural deformation measurement. *Master Sciences Thesis, Department of Geomatics Engineering, Calgary, Alberta, December*, 2010.
- [6] HS Park, HM Lee, Hojjat Adeli, and I Lee. A new approach for health monitoring of structures: terrestrial laser scanning. *Computer-Aided Civil and Infrastructure Engineering*, 22(1):19–30, 2007.
- [7] Keita Yasutomi, Takahiro Usui, Sang-Man Han, Taishi Takasawa, Keiichiro Kagawa, and Shoji Kawahito. 7.5 a 0.3 mm-resolution time-of-flight cmos range imager with column-gating clock-skew calibration, 2014.
- [8] odosimaging. realiz1k. http://www.odos-imaging.com/uploads/pages/p-realiz-1K/DSOIVS1000_datasheet_2014_web.pdf, 2014.
- [9] Jennifer Mantik, H Quix, and J Quest. Enhancement of the stereo pattern tracking technique for model deformation assessment at etw, 2013.
- [10] William Ross McCluney. Introduction to radiometry and photometry, 1994.
- [11] B Taylor and Peter J Mohr. The nist reference on constants, units and uncertainty. 1999.
- [12] Eugene Hecht. Optik, 1987.

- [13] Sven Tschirley. Elektronik i–grundlagen.
- [14] D Botez and M Ettenberg. Comparison of surface-and edge-emitting led's for use in fiber-optical communications. *Electron Devices, IEEE Transactions on*, 26(8):1230–1238, 1979.
- [15] George Horace Brooke Thompson. Physics of semiconductor laser devices. *Chichester, Sussex, England and New York, Wiley-Interscience, 1980.* 572 p., 1, 1980.
- [16] newport.com. Laser diode technology.
- [17] S Cova, M Ghioni, A Lacaita, C Samori, and F Zappa. Avalanche photodiodes and quenching circuits for single-photon detection. *Applied optics*, 35(12):1956–1976, 1996.
- [18] CF Neugebauer and A Yariv. A parallel analog ccd/cmos neural network ic, 1991.
- [19] Dave Litwiller. Cmos vs. ccd: Maturing technologies, maturing markets—the factors determining which type of imager delivers better cost performance are becoming more refined. *Photonics Spectra*, 39(8):54–61, 2005.
- [20] Jerrold T Bushberg and John M Boone. The essential physics of medical imaging, 2011.
- [21] Sebastian Finsterwalder. Die geometrischen grundlagen der photogrammetrie. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 6:1–42, 1897.
- [22] Polytec. Clv compact laser vibrometer.
- [23] Sergi Foix, Guillem Alenyà, and Carme Torras. Lock-in time-of-flight (tof) cameras: a survey. *Sensors Journal, IEEE*, 11(9):1917–1926, 2011.
- [24] Mathworks. Fast fourier transform documentation. page
<http://de.mathworks.com/help/matlab/ref/fft.html>.
- [25] John G Proakis. Digital signal processing: principles algorithms and applications, 2001.
- [26] Andrew Payne, Andy Daniel, Anik Mehta, Barry Thompson, Cyrus S Bamji, Dane Snow, Hideaki Oshima, Larry Prather, Mike Fenton, Lou Kordus, et al. 7.6 a 512× 424 cmos 3d time-of-flight image sensor with multi-frequency photo-demodulation up to 130mhz and 2gs/s adc, 2014.
- [27] Wai Ho Li. dumpk4w, 2014.

- [28] Marvin Lindner, Ingo Schiller, Andreas Kolb, and Reinhard Koch. Time-of-flight sensor calibration for accurate range sensing. *Computer Vision and Image Understanding*, 114(12):1318–1328, 2010.
- [29] Damien Lefloch, Rahul Nair, Frank Lenzen, Henrik Schäfer, Lee Streeter, Michael J Cree, Reinhard Koch, and Andreas Kolb. Technical foundation and calibration methods for time-of-flight cameras. In *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*, pages 3–24. Springer, 2013.
- [30] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.
- [31] Frank Lenzen, Henrik Schäfer, and Christoph Garbe. Denoising time-of-flight data with adaptive total variation, 2011.
- [32] Guy Gilboa. Total variation denoising, 2003.
- [33] Lower Ben. Hackers put kinect for windows v2 through its paces. <http://blogs.msdn.com/b/kinectforwindows/archive/2014/07/02/hackers-put-kinect-for-windows-v2-through-its-paces.aspx>, 2014.
- [34] Carlo Dal Mutto, Pietro Zanuttigh, and Guido M Cortelazzo. A probabilistic approach to tof and stereo data fusion. *3DPVT, Paris, France*, 2, 2010.

List of figures

1 Light as electromagnetic wave	Gpvos 2007 wikipedia.org	15
2 Electromagnetic spectrum	Philip Ronan wikipedia.org	16
3 Solar radiation spectrum on earth	Nick84 wikipedia.org 2013	17
4 CIE Luminosity function $V(\lambda)$ of the human eye under low light (blue curve) and high light (red curve) conditions	HHahn wikipedia.org	18
5 Propagation of light in one steradian	Haade wikipedia.org	19
6 Radiance and Luminance on a surface		20
7 Reflection on an ideal mirroring surface and a diffuse surface	GianniG46 wikipedia.org	21
8 Reflectance spectrum of aluminium (Al), silver (Ag) and gold (Au)	Bob Mellish wikipedia.org 2005	22
9 p-n junction of a diode	TheNoise wikipedia.org	23
10 Diode current versus voltage	Stuendle 2011 wikipedia.org	24
11 Radiation diagram of a IR LED	QED222 www.fairchildsemi.com	25
12 Setup of a common laser diode [16]		26
13 Response over wavelength of a silicon photodiode	Kai Martin 2007 wikipedia.de	27
14 Frame transfer in a CMOS image sensor	Shape 2008 wikipedia.de	29
15 Timing of rolling shutter CMOS Sensor	http://www.cihansari.com/	32
16 Timing of triggered global shutter CMOS sensor	http://www.cihansari.com/	32
17 Focal length in the camera optic	Christophe Dang Ngoc Chan 2007 wikipedia.org	33
18 Motion blur of a fast moving rotor	Tech. Sgt. Ben Blöker wikipedia.org	34
19 Basic principles of stereo triangulation	Arne Nordmann 2007	35
20 Kinect 1 structured light		37
21 Principle of Continuous-wave Time-of-Flight [23]		39
22 Measuring windows of continuous-wave ToF		40
23 Illustration of the analog to digital conversion	wdwd wikipedia.org	43
24 Properties of sinus oscillation	Saure wikipedia.org	44
25 Fast Fourier Transformation in Matlab		46
26 odos real.iZ-1K 1280x1024 with changeable lens and illuminator		49
27 Kinect 2 IR camera and IR illumination with scattering optic		51
28 Kinect 2 image sensor [26]		52
29 Kinect 2 field of view		53
30 Raw image plotted with imagesc() function		54

31 Point cloud in kinect studio	56
32 Transformation from vertices into faces <small>Rhoetzelin wikipedia.org ©©©</small>	56
33 Meshing of point cloud data	57
34 Raw variance of an wooden plate in 700 mm distance	60
35 Gradient magnitude of the office room	61
36 Total Variation Denoising of a 1D signal <small>MAL 2010 wikipedia.org ©©©</small>	62
37 TVD filtered variance of an wooden plate in 700 mm distance	63
38 Peak amplitudes of every pixel of a canvas excited by a person	64
39 Time domain of excitation area	65
40 Frequency domain of excitation area	65
41 3 Hz Oscillating row data sampled with a scope	67
42 Variance vs. distance Microsoft data, raw and TVD after 3 Seconds	69
43 Variance on a checkerboard in 1 m distance	70
44 Variance on various geometries and surfaces	70
45 Influence of halogen lighting on the variance	71
46 Geometry of bias membrane observation	73
47 Displacement of the membrane measured on the middle surface with a Laser Dopper Vibrometer under $\alpha = 0^\circ$	73
48 Sketch of the experimental setup	77
49 Experimental setup	78
50 LDV and Kinect 2 mounted on one tripot	79
51 Membrane with white paper stickers and reflector	79
52 Theoretical area of 4x4 pixel window versus distance	84
53 Amplitude A_p versus distance after 3 seconds of ToF measurement	85
54 Amplitude A_p versus distance after 8 seconds of ToF measurement	85
55 Time domain of ToF after 8 s in 1.4 m distance	86
56 Time domain of LDV after 8 s in 1.4 m distance	86
57 3 Hz excitation in 1.4 m distance	87
58 3 Hz excitation in 2.0 m distance	87
59 Distribution of the dominating frequency f_p and dominating amplitude A_p	89
60 Dominating amplitude A_p under variable angle in 1.4 m	90
61 Dominating amplitude A_p under variable angle in 2 m	90
62 Amplitude A_p versus different oscillation frequencies f_{ref} of the membrane	91
63 Frequency f_p versus different oscillation frequencies f_{ref} of the membrane	91
64 Direct deflection comparison in peak-to-peak A_{pp} value.	92
65 Variable displacement of the LDV data 1.4 m distance	93
66 Variable displacement of the ToF data in 1.4 m distance	93

67 ToF and LDV comparision of 3 <i>Hz</i> oscillation in 1.4 <i>m</i> distance	93
68 1 <i>Hz</i> to 10 <i>Hz</i> sweep in 1.4 <i>m</i> distance captured with ToF	94
69 1 <i>Hz</i> to 10 <i>Hz</i> sweep in 1.4 <i>m</i> distance captured with LDV	94
70 Spectrogram of the sweep ToF data in 1.4 <i>m</i> distance	95
71 Spectrogram of the sweep LDV Data in 1.4 <i>m</i> distance	96
72 Harmonic frequency distribution f_{hp}	96
73 Dominating frequency distribution f_p	97
74 Amplitude distribution A_p	97
75 PIV and PSP on a delta wing DLR Institute of Aerodynamics and Flow Technology 2014	104
76 Distribution of dominating amplitudes (color) and frequency (z-axis) in 0.7 <i>m</i> distance after 8 seconds	105
77 Distribution of dominating amplitudes (z-axis) in 0.7 <i>m</i> after 8 seconds	106
78 Influence of 2000 <i>W</i> halogen lighting on the RGB image	113
79 Modulation of Kinect 2 IR lighting measured with a photodiode	114
80 Three different light intensities in one image frame	114