

**Kamerabasiertes System zur kontaktlosen  
Messung der momentanen Herzfrequenz für  
den Einsatz unter realen  
Umgebungsbedingungen**

Zur Erlangung des akademischen Grades eines  
**DOKTORS DER INGENIEURWISSENSCHAFTEN**  
**(Dr.-Ing.)**

von der KIT-Fakultät  
für Elektrotechnik und Informationstechnik des  
Karlsruher Instituts für Technologie (KIT)

angenommene  
**DISSERTATION**

von

M.Sc. Timon Blöcher

Tag der mündlichen Prüfung: 27.04.2020

Referent: Prof. Dr. rer. nat. Wilhelm Stork

Korreferent: Prof. Dr. rer. nat. Cornelius Neumann



*Für Gaby.*  
*Für Emilia und Maëlle.*  
*Für Esther und Eberhard.*



# Kurzfassung

In der vorliegenden Arbeit wird das Messprinzip Photoplethysmography Imaging (PPGI) detailliert analysiert und auf dieser Basis ein System zur kontinuierlichen, kamerabasierten Remote-Messung der Herzfrequenz konzipiert. Der Fokus liegt dabei auf dem Entwurf algorithmischer Ansätze, die eine robuste Erfassung auch in alltäglichen und hinsichtlich Artefakteinflüssen herausfordernden Szenarien ermöglichen. In diesem Zusammenhang werden zunächst das grundlegende Systemkonzept, eine als Entwicklungs- und Evaluationsbasis dienende Datenbank, sowie spezielle, auf die Artefaktkompensation ausgerichtete Algorithmen vorgestellt. Als Beispiel kann der Entwurf eines Multi-Patch-Algorithmus zur adaptiven Wahl geeigneter Subregionen innerhalb der gewählten Messregion Gesicht genannt werden. Zur Untersuchung der zentralen Fragestellung der Einsetzbarkeit und erzielbaren Genauigkeit in Alltagsszenarien wird das entworfene PPGI-System hinsichtlich der Anwendungsfälle „Belastungsmonitoring am Büroarbeitsplatz“ und „Fahrerbeobachtung im Automobil“ evaluiert. Während der Fokus im stationären Büro-Setting auf der Untersuchung der Machbarkeit einer PPGI-basierten Analyse der Herz- bzw. Pulsratenvariabilität (HRV/PRV) liegt, steht im mobilen Setting der Fahrerbeobachtung die Genauigkeit der Herzfrequenzschätzung unter starken Artefakteinflüssen im Vordergrund.



# Abstract

In this work the measuring principle called Photoplethysmography Imaging (PPGI) is analyzed in detail and on this basis a system for continuous, camera-based remote heart rate measurement is designed. The focus is on the design of algorithmic approaches that allow a robust detection even in everyday scenarios and under challenging conditions. In this context, the basic system concept, a database serving as a basis for development and evaluation, as well as special algorithms for artifact compensation, are presented. As an example, the design of a multi-patch algorithm for the adaptive selection of suitable subregions within the chosen measuring region of the human face can be mentioned. In order to examine the central question of applicability and achievable accuracy in everyday scenarios, the designed PPGI system is evaluated with respect to the use cases „load monitoring at the office workplace“ and „driver monitoring in automobiles“ . While the focus in the stationary office setting is on the investigation of the feasibility of a PPGI-based analysis of heart/pulse rate variability (HRV/PRV), the focus in the mobile setting of driver monitoring is on the accuracy of heart rate estimation under strong artifact influences.



# Danksagung

Die vorliegende Dissertation entstand im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter am FZI Forschungszentrum Informatik im Forschungsbereich Embedded Systems and Sensors Engineering (ESS).

Für die Möglichkeit der Promotion und die Betreuung dieser Arbeit möchte ich mich besonders bei Herrn Prof. Dr. rer. nat. Wilhelm Stork bedanken. Seine unkonventionelle Ideen, die zahlreichen Gespräche und Diskussionen, sowie der Freiraum zur eigenverantwortlichen Gestaltung meiner wissenschaftlichen Arbeit haben den Grundstein für diese Dissertation gelegt. Prof. Dr. rer. nat. Cornelius Neumann danke ich für die Übernahme des Korreferats und das damit verbundene Engagement.

Für die zahlreichen fachlichen Diskussionen zur Thematik und der Unterstützung bei der Entwicklung geeigneter Lösungsstrategien danke ich insbesondere Kai Zhou und Simon Krause. Ein großer Dank gilt auch meinen Kollegen und Freunden Johannes Schneider, Markus Schinle, Christoph Zimmermann, Johannes Bach und Friedrich Gauger. Ich danke außerdem allen von mir betreuten Studenten, die durch ihre Arbeiten wichtige Beiträge zur Entwicklung der „Pulscam“ geleistet haben.

Ganz herzlich danke ich meinen Eltern Esther und Eberhard für ihre stetige Ermutigung und ihren Rückhalt auch in schwierigen Zeiten. Von ganzem Herzen danke ich meiner Frau Gaby. Ohne ihre Zuneigung, ihr Verständnis und die zahlreichen Tage, in denen sie sich voller Liebe alleine um unsere Kinder gekümmert hat, wäre diese Arbeit nicht möglich gewesen.

Karlsruhe, im April 2019

*Timon Blöcher*



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Motivation	3
1.2	Zielsetzung und eigener Beitrag	5
1.3	Gliederung	7
<b>2</b>	<b>Grundlagen</b>	<b>9</b>
2.1	Physiologische Grundlagen	10
2.1.1	Herz-Kreislauf-System	10
2.1.2	Vitalparameter	16
2.1.3	Haut	23
2.2	Messtechnische Erfassung der Herzaktivität	26
2.2.1	Elektrokardiographie	26
2.2.2	Photoplethysmographie und Pulsoximetrie	27
2.2.3	Photoplethysmography Imaging	29
2.2.4	Kamera als PPGI-Sensor	31
2.3	Optische Grundlagen der oberflächennahen Lichtreflexion	32
2.3.1	Dichromatisches Reflexionsmodell	32
2.3.2	Erweiterung des Reflexionsmodells hinsichtlich PPGI	34
2.3.3	Einfluss der spektralen Zusammensetzung des Lichts	37
2.4	Bild- und Signalverarbeitung	38

2.4.1	Principal Component Analysis . . . . .	38
2.4.2	Active Appearance Modell . . . . .	41
2.4.3	Signale und Systeme . . . . .	49
<b>3</b>	<b>Stand der Wissenschaft und Technik . . . . .</b>	<b>55</b>
3.1	Kontaktlose Messung von Vitalparametern . . . . .	56
3.2	Fitting Methoden für das 2D Face Alignment . . . . .	59
3.2.1	Fitting von AAMs . . . . .	59
3.2.2	Constrained Local Model . . . . .	62
3.2.3	Supervised Descent Method . . . . .	62
3.3	Stand der Wissenschaft bei PPGI . . . . .	63
3.3.1	Messtechnischer Aufbau . . . . .	64
3.3.2	Datenbasis . . . . .	65
3.3.3	Detektion und Tracking der Messregion . . . . .	66
3.3.4	Signalverarbeitungsalgorithmen . . . . .	68
3.3.5	Ableitung weiterer Vitalparameter . . . . .	70
3.3.6	Einsatz von PPGI im Automobil . . . . .	72
3.4	Ausgangslage und Abgrenzung . . . . .	73
<b>4</b>	<b>Grundlegende Datenerhebung und Analyse . . . . .</b>	<b>75</b>
4.1	Anforderungen an die Datenbasis . . . . .	76
4.2	Entwicklung eines Beleuchtungsmoduls für aktive Mes- sungen . . . . .	78
4.2.1	Ansteuerung der LEDs . . . . .	79
4.2.2	Ansteuerung der Beleuchtung . . . . .	83
4.3	Versuchsaufbau . . . . .	84
4.3.1	Bildaufnahme . . . . .	85
4.3.2	Beleuchtung . . . . .	86
4.3.3	Referenz Vitalparameter . . . . .	87
4.3.4	Referenz Artefaktsignale . . . . .	88
4.3.5	Synchronisation . . . . .	89

4.4	Szenarien . . . . .	90
4.4.1	Szenarien mit unterschiedlicher Art der Beleuchtung	90
4.4.2	Szenarien mit unterschiedlicher Art der Bewegung	92
4.5	Datenaufnahme und -aufbereitung . . . . .	94
4.6	SNR als Gütekriterium der Signalqualität . . . . .	95
4.7	Analyse der Signalqualität . . . . .	96
4.7.1	Analyse der Szenarien . . . . .	96
4.7.2	Analyse unterschiedlicher Messregionen im Gesicht	100
4.8	Diskussion und Fazit . . . . .	101
<b>5</b>	<b>Systemkonzept . . . . .</b>	<b>103</b>
5.1	Systemanforderungen . . . . .	104
5.2	Systemspezifikation . . . . .	105
5.2.1	Systemaufbau . . . . .	105
5.2.2	Softwarearchitektur . . . . .	106
5.3	Abgeleitetes Vorgehen . . . . .	109
<b>6</b>	<b>Modell und Algorithmus zur Detektion der Messregion . . . . .</b>	<b>113</b>
6.1	Face-Tracking-Modul . . . . .	115
6.1.1	Modellerstellung . . . . .	117
6.1.2	Trackingprozess . . . . .	127
6.2	Fixed-Region-Ansatz . . . . .	129
6.3	Adaptiver Multi-Patch-Ansatz . . . . .	130
6.3.1	Definition der Subregionen . . . . .	131
6.3.2	Selektion der Subregionen . . . . .	133
6.4	Vergleich der Ansätze . . . . .	135
6.5	Diskussion und Fazit . . . . .	139
<b>7</b>	<b>Algorithmus zur Extraktion des Messsignals und Bestimmung der Herzfrequenz . . . . .</b>	<b>141</b>
7.1	Extraktion des BVP-Signals . . . . .	143
7.1.1	Projektion unter Einfluss von Artefakten . . . . .	147

7.1.2	Projektion bei geringem Artefakteinfluss . . . . .	154
7.2	Bestimmung der Herzfrequenz aus dem BVP-Signal . . . . .	156
7.2.1	Adaptiver Bandpassfilter . . . . .	158
7.2.2	Peak Detektion im BVP-Signal . . . . .	162
7.3	Messungen und Ergebnisse . . . . .	162
7.3.1	Optimierung des SNR . . . . .	162
7.3.2	Verbesserung der Herzfrequenzschätzung . . . . .	167
7.4	Diskussion und Fazit . . . . .	169
<b>8</b>	<b>Onlinefähiges System zum Einsatz unter Realbedin-</b>	
	<b>gungen . . . . .</b>	<b>173</b>
8.1	Umsetzung als Live-Prototypensystem . . . . .	175
8.2	Messung der Pulsratenvariabilität im Setting Büro . . . . .	178
8.2.1	Motivation . . . . .	178
8.2.2	Messsetup und Probandenkollektiv . . . . .	181
8.2.3	Auswertung der Herzfrequenzmessung . . . . .	184
8.2.4	Vergleich von Puls -und Herzratenvariabilität . . . . .	184
8.2.5	Diskussion und Fazit . . . . .	189
8.3	Einsatz von PPGI im Automobil . . . . .	190
8.3.1	Motivation . . . . .	190
8.3.2	Abgrenzung und Zielsetzung der Studie . . . . .	192
8.3.3	Systemintegration . . . . .	193
8.3.4	Positionierung der Kamera . . . . .	194
8.3.5	Studie mit initialem Prototypensystem . . . . .	196
8.3.6	Studie mit optimierten Prototypensystem . . . . .	201
8.3.7	Diskussion und Fazit . . . . .	214
<b>9</b>	<b>Zusammenfassung und Ausblick . . . . .</b>	<b>215</b>
	<b>Literaturverzeichnis . . . . .</b>	<b>220</b>
	<b>Eigene Veröffentlichungen . . . . .</b>	<b>241</b>

<b>Abbildungsverzeichnis</b> . . . . .	<b>243</b>
<b>Tabellenverzeichnis</b> . . . . .	<b>247</b>
<b>Einwilligungserklärung Datenaufzeichnung</b> . . . . .	<b>249</b>



# Abkürzungsverzeichnis

<b>AAM</b> Active Appearance Model	<b>GUI</b> Graphical User Interface
<b>ACOR</b> AC-Anteil der Oberflächenreflexion	<b>HKS</b> Herz-Kreislauf-System
<b>ADC</b> Analog-Digital-Converter	<b>HRV</b> Herzratenvariabilität
<b>bpm</b> beats per minute	<b>HZV</b> Herzzeitvolumen
<b>BSS</b> Blind Source Separation	<b>IBI</b> Inter-Beat-Interval
<b>BVP</b> Blutvolumenpuls	<b>ICA</b> Independent Component Analysis
<b>EKG</b> Elektrokardiographie	<b>LMS</b> Least-Mean-Square
<b>EEG</b> Elektroenzephalographie	<b>NIR</b> Nahinfrarot
<b>FDA</b> Food and Drug Administration	<b>PCA</b> Principal Component Analysis
<b>FFT</b> Fast Fourier Transformation	<b>POE</b> Plane orthogonal to Energy
<b>fps</b> frames per second	<b>POS</b> Plane orthogonal to Skin
<b>HF</b> Herzfrequenz	<b>PPG</b> Photoplethysmographie

<b>PPGI</b> Photoplethysmography Imaging	<b>ROI</b> Region of Interest
<b>PRV</b> Pulsratenvariabilität	<b>RSA</b> Respiratorische Sinusarrhythmie
<b>PTT</b> Pulse-Transit-Time	<b>SDM</b> Supervised Descent Method
<b>PWG</b> Pulswellengeschwindigkeit	<b>SNR</b> Signal-Rausch-Verhältnis
<b>RMSE</b> Root Mean Square Error	<b>SPO2</b> Sauerstoffsättigung
<b>RMSSD</b> Root Mean Sum of Squared Distance	

# 1

## Einleitung

**D**IE MESSUNG MENSCHLICHER VITALPARAMETER gibt Aufschluss über den Gesundheitszustand einer Person und ist daher fester Bestandteil bei Diagnose, Monitoring und Therapie von Patienten im medizinisch-klinischen Umfeld. In den letzten Jahren hält die Erfassung physiologischer Signale zudem vermehrt Einzug in den Alltag. Zum einen gibt es Bestrebungen den menschlichen Gesundheitszustand durch kontinuierliches Monitoring im nicht-klinischen Umfeld mit medizinisch zertifizierten Geräten fortwährend zu erfassen, um auf diese Weise Diagnose und Therapie zu optimieren und Krankenhausaufenthalte oder Arztbesuche zu reduzieren. Durch dieses „Personal Health Monitoring“ [99] sollen Krankenhäuser und Arztpraxen entlastet und Behandlungskosten gesenkt werden. Zum anderen ist im Zuge der Quantify-Self-Bewegung [57] eine deut-

liche Zunahme an nicht-medizinischen Consumer-Geräten zu erkennen, die es ermöglichen bestimmte Vitaldaten wie die Herzfrequenz (HF) quasi-kontinuierlich zu tracken. Dabei werden die Geräte als Fitnessstracker zur Optimierung der körperlichen Verfassung und des Trainingsverlaufs eingesetzt oder geben Information über psychophysiologische Parameter wie Stressbelastung und Schlafqualität aus.

Während im medizinischen Bereich aus Gründen der Genauigkeit in erster Linie auf an der Brust getragene Elektrokardiographie (EKG)-Systeme zurückgegriffen wird, werden im Consumer-Bereich meist am Handgelenk getragene, Photoplethysmographie (PPG)-basierte Wearables genutzt. Einen aktuellen Versuch die beiden Bereiche zu verbinden, liefert die Apple Watch 4 mit integrierter und von der Food and Drug Administration (FDA) zertifizierter EKG-Funktion zur frühzeitigen Detektion von Herzrhythmusstörungen.

Neben der tragbaren Sensorik werden vermehrt kontaktlose Messmethoden erforscht und entwickelt. Beim Ansatz der kapazitiven EKG-Messung werden beispielsweise die Vitalparameter durch die Kleidung hindurch ohne direkten Hautkontakt erfasst. In der Literatur wird in diesem Zusammenhang oft der englische Begriff der „Unobtrusive Vital Sign Measurements“ (unaufdringliche Messmethoden) verwendet [21]. Bei entsprechender Integration in Alltagsgegenstände wird zudem von ambienter Sensorik gesprochen. Das Verfahren des Photoplethysmography Imaging (PPGI) geht dabei noch einen Schritt weiter und bietet die Möglichkeit der kamerabasierten Erfassung der Herzfrequenz über Entfernungen einiger Meter hinweg („Remote-Messung“). Neben der Erfassung der HF sind auch Untersuchungen zur Messung der Herzratenvariabilität (HRV), der Sauerstoffsättigung (SPO2) und des Blutdrucks Gegenstand aktueller Forschung.

## 1.1 Motivation

Ein großer Vorteil von PPGI ist die Möglichkeit der Remote-Messung. Das Anlegen von Sensoren vor einer Messung entfällt, was insbesondere bei eigenständiger Messung ohne geschultes Fachpersonal eine häufige Fehlerquelle darstellt. Gleichzeitig lässt sich die Sensorik ambient integrieren, ohne dass ein geringer Maximalabstand zum Messobjekt im Millimeterbereich gefordert wird, wie beispielsweise beim kapazitiven EKG. Auf diese Weise wird eine kontinuierliche Langzeitmessung ermöglicht, ohne dass die Bewegungsfreiheit der untersuchten Person spürbar eingeschränkt ist. Dies eröffnet zahlreiche potentielle Einsatzmöglichkeiten eines PPGI-Systems in verschiedensten Domänen. Zu nennen ist zunächst der medizinisch-pflegerische Bereich mit der Vitaldatenüberwachung von Frühgeborenen oder älteren Menschen im Kontext des Ambient Assisted Living, auf Intensiv- oder Sonderisolierstationen oder in der Schlafanalyse. Zusätzliche Anwendungsdomänen bieten der Security-Bereich (Beispiel Flughafenüberwachung), der Lifestyle-Bereich (Quantify-Self über Smartphone), die Nutzer-Authentifizierung (Liveness Detection) oder die Arbeitsmedizin/-wissenschaft (Belastungsmonitoring). Darüber hinaus ist der Einsatz als Monitoring-System in Cockpits mobiler Systeme, z.B. in Zügen oder Flugzeugen, besonders aber in Automobilen denkbar.

Während der Fokus vieler wissenschaftlicher Arbeiten und Publikationen im PPGI-Kontext zunächst, auch begründet durch die noch junge Geschichte des Verfahrens, auf der Entwicklung grundlegender Systemansätze und Algorithmen lag, steht in dieser Arbeit die Überführung in konkrete Anwendungsszenarien im Vordergrund. Um zukünftige Anwendungen im Alltag zu ermöglichen, gilt es daher unter anderem die systemischen und algorithmischen Voraussetzungen für den Einsatz der Messmethode unter dem Einfluss von starken Artefakten zu untersuchen. Auch wurden bisher nur wenige Studien mit realitätsnahen Bedingungen veröffentlicht [163].

Das in dieser Arbeit fokussierte Anwendungsgebiet ist die Vitalparametermessung im Automobil als Beitrag zur Fahrerzustandserkennung. Die Erfassung physiologischer Signale im automobilen Kontext hat in den letzten Jahren stark an Bedeutung gewonnen. Dies trifft sowohl für die Forschung und Vorausbildung, als auch für den Bereich der Serienentwicklung von Automobilen zu. Dabei lag der Fokus bisher auf kontaktbasierten Ansätzen mit ins Lenkrad integrierten Sensoren oder auf quasi-kontaktbasierten Ansätzen mittels kapazitivem EKG über im Fahrersitz integrierte Textilelektroden. Bisher sind diese Ansätze in Automobilen nicht in Serie, da die zusätzliche Sensorik und der sich ergebende Integrationsaufwand mit zu hohen Kosten verbunden ist. Vor dem Hintergrund der klaren Tendenz zum generellen Einbau von Kameras im Fahrzeuginnenraum zur Fahrer- und Insassenbeobachtung, wäre im Idealfall kein zusätzlicher Sensor notwendig.

Bei der Fahrerbeobachtung liegt der Fokus bisher auf der Analyse der Augenpartien zur Erkennung kurz- oder langfristiger Unaufmerksamkeit, welche durch Ablenkung oder Müdigkeit hervorgerufen wird. Nach Untersuchungen des Allianz-Zentrums für Technik sind ca. 30 Prozent aller Unfälle auf Unaufmerksamkeit zurückzuführen [43]. Die Erweiterung um kardiovaskuläre Parameter könnte präzisere Aussagen über die Leistungsfähigkeit oder das Müdigkeitslevel als bisherige Systeme ermöglichen und zusätzlich Informationen über die Emotion oder das Stress- und Fitnesslevel des Fahrers liefern. Des Weiteren ist auch ein Beitrag zur Erkennung medizinischer Notfallsituationen denkbar, beispielsweise bei Bewusstseinsverlusten von Bus- oder Berufskraftfahrern, welche ein hohes Gefahrenpotential darstellen. Für hochautomatisierte und vollautonome Fahrzeuge (SAE Level 4 und 5<sup>1</sup>) verfolgen verschiedene Automobilhersteller zudem das Ziel das Wohlbefinden der Insassen zu erfassen, um u.a. Komfortsysteme oder das Fahrzeugverhalten automatisiert anzupassen. Ein Bei-

---

<sup>1</sup> <https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic>

spielszenario wäre in diesem Zusammenhang die vitalparameterbasierte Früherkennung der Reisekrankheit (Kinetose), um plötzliche Übelkeit der Insassen zu vermeiden. Neben diesen perspektivischen Anwendungsfällen kann ein Remote-Vitaldatenmonitoring-System bereits heute einen wichtigen Beitrag im Bereich der Forschung- und Vorausbildung leisten. Beispielsweise können die erfassten Vitalparameter im Bereich der Fahrerprobung oder Nutzerstudien als objektives Maß und Referenz für die subjektive Empfindung des Fahrerlebnisses dienen.

Außer der Erfassung der Herzfrequenz, kann PPGI prinzipiell auch zur Messung der HRV eingesetzt werden. Dabei werden über die Analyse der Schlag-zu-Schlag-Distanzen des Herzens wichtige Information über die Aktivität des vegetativen Nervensystems und das Zusammenspiel der Teilsysteme Sympathikus und Parasympathikus gesammelt. Auf dieser Basis lassen sich Rückschlüsse über sekundäre Parameter wie Stress oder die kognitive und psychische Belastung ziehen. Für den möglichen Einsatz eines PPGI-Systems, beispielsweise zum kontaktlosen Belastungsmonitoring im Arbeitsumfeld, gilt es jedoch zunächst zu untersuchen, in wie weit die mittels PPGI erfassten Distanzen zwischen Pulswellen und die daraus abgeleitete Pulsratenvariabilität (PRV) mit der über das EKG und die elektrische Herzaktivität abgeleitete HRV vergleichbar ist.

## 1.2 Zielsetzung und eigener Beitrag

Diese Arbeit soll einen Beitrag zur Entwicklung von PPGI-Systemen leisten, um ein belastungsfreies Remote-Vitaldatenmonitoring mittels Kamera zu ermöglichen. Als Messregion wird dabei, unter anderem aus Gründen der Zugänglichkeit und Durchblutungsstärke, das menschliche Gesicht genutzt. Im Rahmen der Arbeit werden das Messverfahren selbst, Haupteinflussfaktoren auf die Signalqualität und die sich daraus ergebenden Randbedingungen und Systemgrenzen detailliert untersucht. Darauf aufbauend werden systemische und algorithmische Konzepte, sowie eine

Realisierung als onlinefähiges System zur Schlag-zu-Schlag-Analyse der Herzfrequenz abgeleitet. Der Fokus liegt dabei auf den Anwendungsszenarien „Fahrerbeobachtung im Automobil“ und „Belastungsmonitoring am (Büro-)Arbeitsplatz“ und den darin bestehenden Herausforderungen durch den Einfluss von Bewegungs- und Lichtartefakten. Während beim Einsatz im Automobil die grundsätzliche Machbarkeit der Erfassung der HF unter starken Artefakteinflüssen im Vordergrund steht, soll bei der Anwendung am Arbeitsplatz die Möglichkeit der Ableitung von Parametern der HRV, als Voraussetzung für ein mögliches Belastungsmonitoring, analysiert werden. Daraus ergeben sich folgende zentrale Fragestellungen für die Arbeit:

- Welche systemischen und algorithmischen Konzepte sind notwendig, um ein im Alltag einsatzfähiges PPGI-System zu realisieren? Welcher Ansätze bedarf es, um eine onlinefähige Schlag-zu-Schlag-Analyse der HF zu ermöglichen?
- Welche Signalqualität der Herzfrequenzmessung kann im Vergleich zu kontaktbasierten Systemen erreicht werden und was sind die entscheidenden begrenzenden Einflussfaktoren?

Darauf aufbauend sollen im Rahmen der Evaluation in den beiden Anwendungsszenarien folgende Fragen untersucht werden:

- Mit welcher Qualität lässt sich mithilfe des entwickelten Systems die PRV in einem alltäglichen Szenario wie der Arbeit am PC ableiten? In wie weit kann diese als Ersatz der EKG-basierten HRV dienen?
- Mit welcher Genauigkeit kann die HF unter starken Artefakteinflüssen wie bei der Fahrerbeobachtung im Automobil erfasst werden?

## 1.3 Gliederung

Die vorliegende Arbeit ist in 9 Kapitel untergliedert. In Kapitel 2 werden zunächst die für das Verständnis der Arbeit notwendigen physiologischen und technischen Grundlagen dargestellt. Der Stand der Wissenschaft und Technik wird in Kapitel 3 beschrieben. Dabei wird auf die Vor- und Nachteile unterschiedlicher kontaktloser Verfahren zur Erfassung der Herzfrequenz eingegangen und anschließend die verschiedenen Ansätze der PPGI-Messung vorgestellt. Nach der Abgrenzung zu bestehenden Ansätzen wird im darauffolgenden Kapitel 4 die Erstellung einer Datenbank als Grundlage der Systementwicklung und Validierung beschrieben. Kapitel 5 beinhaltet das Systemkonzept mit der Beschreibung der Systemanforderungen und dem daraus abgeleiteten Aufbau des Systems. In Kapitel 6 werden unterschiedliche algorithmische Ansätze zur Detektion und Verfolgung des Gesichts als Messregion dargestellt, bevor in Kapitel 7 die entwickelten Methoden zur Extraktion des Messsignals und Bestimmung der Herzfrequenz präsentiert werden. Kapitel 8 beinhaltet die Umsetzung als onlinefähiges System und die Validierung in den zuvor genannten Anwendungsszenarien. Die Arbeit schließt mit der Zusammenfassung und dem Ausblick in Kapitel 9.



# 2

## Grundlagen

**D**IESES KAPITEL beschreibt die für das Verständnis der Arbeit notwendigen Grundlagen. Dabei werden zunächst die physiologischen Grundlagen des Herz-Kreislauf-Systems und die optischen Eigenschaften der Haut als Messregion beschrieben. Anschließend werden verschiedene Messtechnik-Arten zur Erfassung der menschlichen Herzaktivität und die optischen Grundlagen der oberflächennahen Hautreflexion thematisiert. Abgeschlossen wird das Kapitel mit den notwendigen Grundlagen der Bild- und Signalverarbeitung.

## 2.1 Physiologische Grundlagen

### 2.1.1 Herz-Kreislauf-System

Im Herz-Kreislauf-System (HKS) bildet das Herz das zentrale Pumporgan, die Blutgefäße dienen als Transportwege. Die Aufgabe besteht in der Versorgung sämtlicher Organe mit Blut, welches den für die Organtätigkeit notwendigen Sauerstoff und Nährstoffe enthält. Das Herz wird durch die Herzscheidewand (Septum) in die linke und rechte Herzhälfte unterteilt, welche sich wiederum aus Vorhof (Atrium) und Herzkammer (Ventrikel) zusammensetzen. Die Aufgabe des Vorhofs ist die Regulation der Befüllung der Kammer, deren anschließende Pumpaktivität das restliche Gefäßsystem versorgt [17]. Die Pumpaktivität besteht aus zwei Phasen: der Systole (Anspannung) und der Diastole (Erschlaffung). Zusätzlich kontrollieren die Herzklappen den Blutfluss im Herzen, um einen möglichen Rückfluss in die falsche Richtung zu verhindern.

Die Regulation der Herzaktivität erfolgt über die Anpassung der Herzfrequenz und des Schlagvolumens. Auf diese Weise wird das Herzzeitvolumen (HZV) <sup>1</sup> kontinuierlich an den Gewebebedarf angepasst. Die Adaption der Herzfrequenz erfolgt durch das Zusammenspiel des stimulierenden Sympathikus und des innervierenden Parasympathikus des vegetativen Nervensystems. Ein weiteres autonomes Regelsystem, der Frank-Starling-Mechanismus, ermöglicht die Anpassung der Kammeraktivität bzw. des Auswurfsvolumens an ein erhöhtes Füllungsvolumen am Ende der Diastole (Vorlast) oder einen erhöhten Aortendruck (Nachlast) [71].

Grundsätzlich nutzt der Blutkreislauf die Hintereinanderschaltung zweier Kreisläufe:

- *Körperkreislauf* (gr. Kreislauf) : Versorgung der Organe und Gewebe
- *Lungenkreislauf* (kl. Kreislauf): Sauerstoffanreicherung des Blutes

---

<sup>1</sup> Blutvolumen, welches pro Zeit in den Körperkreislauf gepumpt wird

Über die linke Kammer wird zunächst Blut in den Körperkreislauf gepumpt, bis die immer kleiner werdenden arteriellen Gefäße (Arteriolen und Kapillare) der Körperperipherie erreicht werden. In den kleinsten Gefäßen, den dünnwandigen Kapillaren, erfolgt der Gas- und Sauerstoffaustausch mit den Organen. Von den Kapillaren kehrt das sauerstoffarme Blut über die Venolen, die kleinen Venen und die obere Hohlvene in den rechten Vorhof zurück, bevor es über den rechten Ventrikel in den Lungenkreislauf befördert wird. Dort findet der Gasaustausch mit der Lunge statt, bevor es wieder der linken Herzhälfte zugeleitet wird [133].

### **Hoch- und Niederdrucksystem**

Funktionell wird der Blutkreislauf in Hoch- und Niederdrucksystem unterteilt. Das Hochdrucksystem besteht aus den Arterien, der linken Herzhälfte und den Gefäßen des Körperkreislaufs [80]. Aufgrund der rhythmischen Kontraktion des Herzens ergibt sich hier ein wellenförmiger Verlauf des Blutflusses. Bei einem gesunden Erwachsenen liegt die Herzfrequenz, also die Zahl der Kontraktionen pro Minute, in Ruhe zwischen 60 und 80, wobei jeder Kontraktion eine Pulswelle folgt. Die Laufzeit einer Pulswelle von einem Ort des arteriellen Systems zu einem anderen bezeichnet man als Pulstransitzeit (*engl.*: Pulse-Transit-Time (PTT)). Das Herz besitzt dabei verschiedene Funktionen um Dauer und Stärke einer solchen Welle zu beeinflussen. So bewirkt die Dehnbarkeit der elastischen Arterien (Compliance) eine Abnahme des pulsatilen Charakters des Blutflusses zur Peripherie hin, was auch als Windkessel-Funktion bezeichnet wird.

Zum Niederdrucksystem zählen die Venen, die rechte Herzhälfte und die Gefäße des Lungenkreislaufs. Auf Grund der hohen Dehnbarkeit und Kapazität erfüllt das Niederdrucksystem die Funktion eines Blutspeichers. So befinden sich etwa 85% der Gesamtmenge des Blutes von 5-6 Litern im Niederdrucksystem. Der Rückstrom des venösen Blutes zum Herzen beruht auf der Kontraktionen der Skelettmuskulatur (Muskelpumpe), der at-

mungsbedingten Exkursion des Thorax und der Sogwirkung des rechten Ventrikels im herznahen Bereich [143]. Auf diese Weise besitzt der Verlauf des Blutflusses hier einen unregelmäßigeren und weniger pulsatilen Charakter. Eine Übersicht über den Verlauf des Drucks, der mittleren linearen Strömungsgeschwindigkeiten und des Gesamtquerschnitts der Gefäße in den unterschiedlichen Bereichen des Herzkreislaufsystems wird in Abbildung 2.1 gegeben.

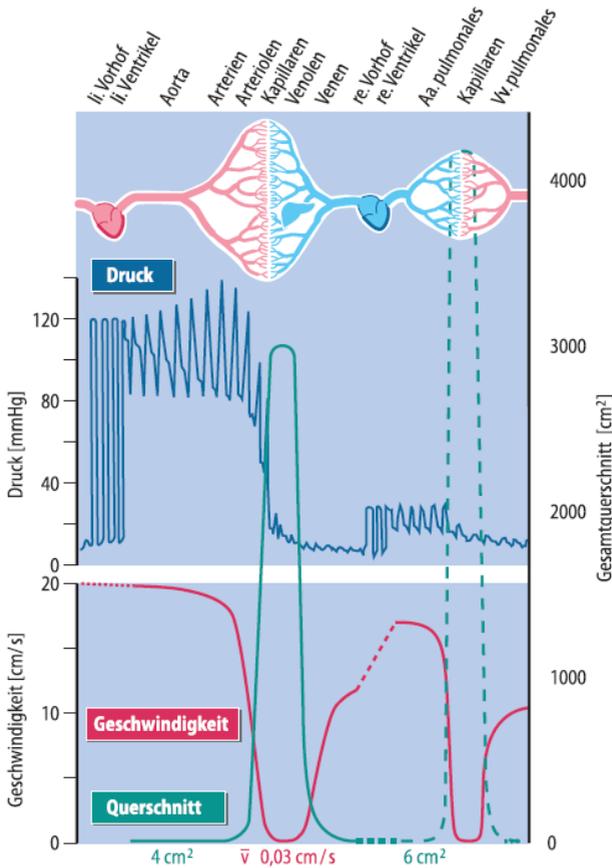


Bild 2.1: Verlauf von Druck, Strömungsgeschwindigkeit und Gefäß-Gesamtquerschnitt im HKS, *Quelle: [142]*

## Erregungsphysiologie des Herzens

Die Kontraktion des Herzmuskels erfolgt durch die elektrische Erregung der Herzmuskelfasern. Die Fasern bestehen aus einem Verbund von Herzmuskelzellen, die von einer gemeinsamen Hülle, der Perimembran, umgeben sind. Ein großer Teil der Herzmuskelzellen gehört zum Arbeitsmyokard, welches die mechanische Pumparbeit bei der Herzmuskelkontraktion verrichtet. Der kleinere Teil der Zellen zählt zum Erregungsbildungs- und Erregungsleitungssystem, welches für die Weiterleitung des Erregungsimpulses im Herzmuskel zuständig ist. Zu diesem Zweck sind die Herzmuskelzellen über sogenannte Gap Junctions miteinander verbunden und somit elektrisch gekoppelt. Wird eine Herzmuskelzelle erregt, breitet sich die Erregung auf sämtliche Zellen aus, was als funktionelles Synzytium bezeichnet wird. Das Herz agiert dabei nach dem Alles-oder-nichts-Gesetz. Es besagt, dass das Herz auf einen elektrischen Reiz mit der Erregung und Kontraktion aller Fasern antwortet oder gar nicht reagiert. Die Auslösung der rhythmischen Kontraktion entsteht im Herzen selbst. Diese Eigenschaft wird Autorhythmie des Herzens genannt.

Die Erregungsbildung erfolgt in der Regel im Sinusknoten (siehe Abb. 2.2). Von dort erreicht die Erregung zunächst die Vorhöfe und breitet sich anschließend über den AV-Knoten auf das His-Bündel aus, welches in die Tawara-Schenkel der rechten und linken Kammer unterteilt ist. Über die Purkinje-Fasern erfolgt schließlich die Erregung des Kammermuskels. Fällt die Erregungsbildung im Sinusknoten aus oder wird die Weiterleitung in den Vorhöfen unterbrochen, übernimmt der AV-Knoten die Funktion als Schrittmacher [5].

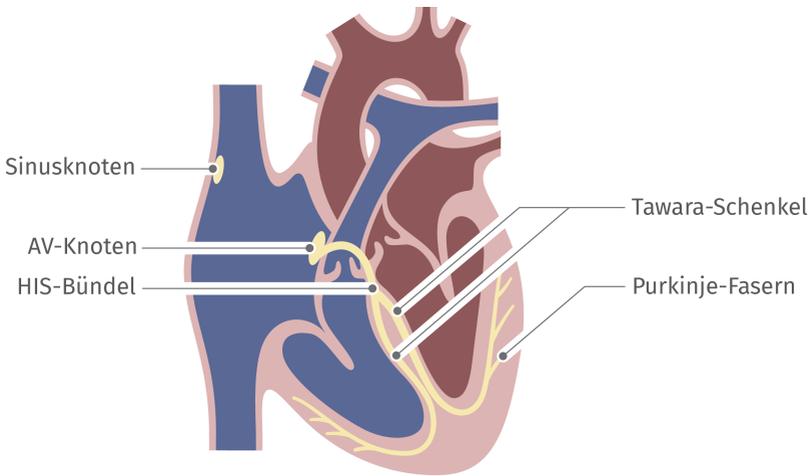


Bild 2.2: Erregungsbildung/-leitung im Herzen

### Form und Ausbreitung der Pulswelle

Bei jeder Kontraktion des Herzens entsteht am Ausgang des linken Ventrikels eine Pulswelle, welche in den Körperkreislauf propagiert. Eine Pulswelle besitzt dabei die gleichen Eigenschaften wie eine akustische Welle und kann durch ihre Frequenz(-anteile) und Amplitude beschrieben werden. Innerhalb des kardiovaskulären Systems können Pulswellen reflektiert (Echo) und überlagert (Rückkoppelung), verstärkt (Trichter) oder gedämpft (Drossel) werden. Die Basisschwingung einer Welle entspricht der Herzfrequenz. Rücklaufende Wellen, wie sie bei der Reflexion an Gefäßwänden entstehen, treten als Vielfache dieser Schwingung auf (ab ca. 4Hz bei einer Grundschiwingung von 1Hz). Die Pulswelle entspricht daher der Summe der Schwingungen unterschiedlicher Frequenzen. Hämodynamischen Einfluss haben Wellen mit Frequenzen unter 15Hz, wobei Wellen mit Frequenzen größer 3 Hz, auf Grund der Massenträgheit des Blutes, den Blutfluss kaum verbessern, sondern lediglich den Blutdruck in den Gefäßen erhöhen [160]. Grundsätzlich lassen sich bei der Ausbreitung der Pulswelle

drei Arten des Pulses unterscheiden: der Strompuls durch Flüssigkeitsströmung, der Druckpuls durch Druckänderungen und der Volumenpuls durch Querschnitts- bzw. Volumenänderungen. Das messtechnische Prinzip des in dieser Arbeit betrachteten Ansatzes beruht auf der Erfassung und Analyse des Volumenpulses, welcher unter anderem über plethysmographische Verfahren gemessen werden kann (vgl. Abschnitt 2.2.2).

Form und Geschwindigkeit einer Pulswelle sind, wie in Abschnitt 2.1.1 beschrieben, stark von Faktoren wie Durchmesser und Elastizität des jeweiligen Gefäßabschnitts abhängig. Beispielsweise dämpfen elastische Gefäße hochfrequente Wellenanteile, wodurch sich die Wellenform der herznahen Aorta gegenüber peripheren Arterien unterscheidet. Im gesunden Körper kann die Elastizität der Gefäße gesteuert werden. Ist dieser Steuermechanismus der Gefäße gestört, steigt der arterielle Blutdruck und die Gefäße versteifen sich. Dadurch erhöht sich die Wellengeschwindigkeit, sowie reflexionsbedingt die Anzahl hochfrequenter Wellen und es kommt zu einem sich verstärkenden, negativen Kreislauf [160]. Gefäßeigenschaften wie die Elastizität hängen von Parametern wie Alter, Geschlecht und Größe, aber auch von Nahrung, Krankheit oder Medikamenten ab und beeinflussen so auch die Pulswellenform. Durch eine Pulswellenanalyse in Bezug auf Morphologie und Geschwindigkeit können daher Rückschlüsse auf gesundheitliche Risikoindikatoren wie arterieller Gefäßsteifigkeit oder Bluthochdruck gezogen werden [82].

In Abbildung 2.3 ist beispielhaft die Form der Pulswelle in peripheren Arterien skizziert. Der Beginn der Pulswelle wird durch den systolischen Auswurf des Blutes in die Aorta in Folge der Ventrikelkontraktion bestimmt, bis hin zu einem ersten Maximum, der sogenannten Percussion-Welle. Die nachfolgende Tidal-Welle entsteht durch Reflexionen an der Aortenwand, wobei Form und Amplitude insbesondere von der Elastizität der Aortenwand abhängen. Das Ende der Tidal-Welle wird durch einen Druckabfall (Inzisur) markiert und entspricht gleichzeitig dem Ende der Systole. In der Diastole folgt die Dikrote Welle, die durch Reflexionen der Pulswellen an

Orten mit stark erhöhtem Wellenwiderstand (z.B. durch Gefäßverzweigungen) entsteht.

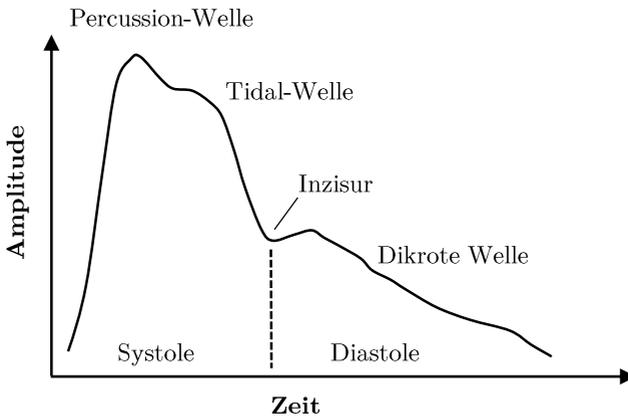


Bild 2.3: Exemplarische Form der Pulsform in den peripheren Arterien

### 2.1.2 Vitalparameter

Körperfunktionen, deren unterbrechungsfreie Funktion unabdingbar ist, um die Lebensvorgänge des Organismus zu gewährleisten, werden als Vitalfunktionen bezeichnet. Dazu gehören in erster Linie die Atmung und die Herz-Kreislauffunktion, im weiteren Sinne auch die höheren Funktionen (Bewusstsein) [112]. Zu den wichtigsten Vitalparametern zählen daher die Atemfrequenz, die Sauerstoffsättigung, die Herzfrequenz, der arterielle Blutdruck oder die Körpertemperatur. In den folgenden Abschnitten wird lediglich auf die im weiteren Verlauf relevanten Parameter eingegangen. Vitalparameter außerhalb der Normwerte können Anzeichen für akute oder generelle Störungen des Organismus darstellen. Bei der Messung von Vitalparametern wird zwischen zwei Messarten unterschieden. In der Notfallmedizin findet die Messung meist zu einem diskreten Zeitpunkt

statt, bei der beispielsweise die Herzfrequenz oder die Körpertemperatur einmalig bestimmt wird. Es existieren aber auch kontinuierliche Messmethoden wie die Elektrokardiographie (EKG), die Elektroenzephalographie (EEG) oder die Pulsoxymetrie, bei der Werte über einen längeren Zeitraum aufgezeichnet und ausgewertet werden (Monitoring).

## Herzfrequenz

Die Herzfrequenz beschreibt die Zahl der Herzaktionen pro Minute. Meist stimmt sie mit der Pulsfrequenz überein. Bei speziellen Formen der Arrhythmie kann die Herzkontraktion jedoch ausbleiben und eine Differenz zwischen Herz- und Pulsfrequenz entstehen <sup>2</sup>. Die Herzfrequenz in Ruhe hängt zunächst von Alter, Geschlecht und körperlicher Verfassung ab. Sie beträgt bei Neugeborenen im Schnitt 120 Schläge pro Minute (*engl.*: „beats per minute (bpm)“) und sinkt im Alter von 65 Jahren auf ungefähr 70 bpm. Unabhängig von diesen Faktoren steigt die Herzfrequenz bei körperlicher, mentaler oder kognitiver Anstrengung. Die Regulation der Herzfrequenz erfolgt hauptsächlich durch Sympathikus und Parasympathikus des autonomen Nervensystems oder über Hormonfreisetzung (vgl. 2.1.1). Des Weiteren steigt die Herzfrequenz beim tiefen Einatmen und sinkt beim tiefen Ausatmen, was als Respiratorische Sinusarrhythmie (RSA) bezeichnet wird. Bezogen auf die altersüblichen Normwerte teilt man die Herzfrequenz in drei Bereiche ein:

- Bradykardie: erniedrigte Herzfrequenz (z.B. < 60 bpm)
- Normofrequenz: normale Herzfrequenz (z.B. 60 bpm – 100 bpm)
- Tachykardie: erhöhte Herzfrequenz (z.B. > 100 bpm)

---

<sup>2</sup> Bei der Bestimmung der Frequenz der Herzkontraktionen wird im Rahmen dieser Arbeit der Einfachheit halber stets von Herzfrequenz gesprochen, auch wenn die Pulsfrequenz erfasst wird. Eine Differenzierung erfolgt lediglich für die Erfassung der Herzratenvariabilität (siehe Abschnitt 2.1.2)

Die maximale Herzfrequenz, die bei größter körperlicher Anstrengung erzielt werden kann, wird als Maximalpuls bezeichnet und näherungsweise über die Formel  $HF_{max} = 220 - \text{Lebensalter}$  bestimmt.

### Herzratenvariabilität

Wie im vorigen Abschnitt beschrieben besitzt das Herz die Fähigkeit die Herzfrequenz an physische oder psychische Beanspruchung zu adaptieren. Neben der reinen Frequenzanpassung wird auch die Regelmäßigkeit mit der das Herz schlägt an verschiedene Belastungen angepasst. Bei körperlicher Anstrengung oder Stress steigt die Herzfrequenz und das Herz schlägt in gleichen zeitlichen Abständen. In Ruhe und bei Entspannung sinkt die Herzfrequenz und der Herzschlag erfolgt in unregelmäßigeren Abständen. Diese ständige Anpassung der zeitlichen Intervalle zwischen Herzschlägen an die Bedürfnisse des gesamten Organismus wird als Herzfrequenzvariabilität (HRV) bezeichnet. Ein variabler Herzschlag deutet auf einen guten Gesundheitszustand oder Wohlbefinden hin, ein monotoner Herzschlag auf das Gegenteil. Daher dient die HRV in der Kardiologie als wichtiger Prognoseparameter für Herz- und Immunerkrankungen und liefert ein gutes Maß für die Regulationsfähigkeit des Organismus. Aber auch bei der Diagnose von psychischen Erkrankungen wie Depressionen oder Burnout liefert die HRV-Analyse wichtige Erkenntnisse [10].

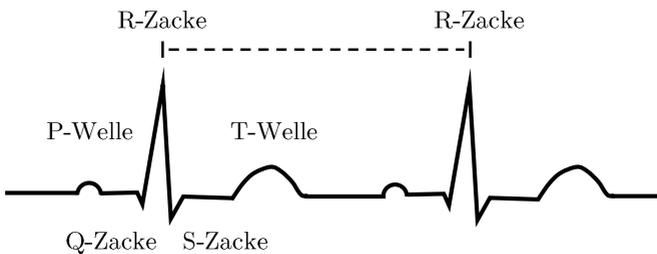


Bild 2.4: Definition des RR-Intervalls im EKG

Die Analyse der HRV erfolgt auf Basis des zeitlichen Abstands zweier Ventrikelkontraktionen. Im Elektrokardiogramm (siehe Abb. 2.4) entspricht dies der R-Zacke, weshalb der Abstand als RR-Intervall bezeichnet wird (in der Literatur auch NN-Intervall („normal-to-normal“)). Die Schwankung dieser Intervalle kann im Zeitbereich, im Frequenzbereich und nichtlinear quantisiert werden. Hinsichtlich der Messdauer wird zwischen Langzeit(24h)-, Kurzzeit(5min)- und Ultrakurzzeit(< 5min)-Messungen unterschieden. Im Folgenden werden zunächst die wichtigsten Parameter des Zeitbereichs aufgeführt:

$$SDNN = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (NN_i - \overline{NN})^2} \quad (2.1)$$

mit  $\overline{NN}$  = Mittelwert und  $N$  = Anzahl NN-Intervalle. Die Standardabweichung der NN-Intervalle wird als SDNN bezeichnet und ist ein Maß für die generelle Anpassungsfähigkeit des Körpers an Belastungen. Bei der Analyse ist zu beachten, dass der SDNN-Wert mit der Länge der Messzeit steigt, da mehr unterschiedliche Aktivitätspegel in die Berechnung einfließen. Es sollten folglich nur Werte mit der gleichen Messdauer bei ähnlicher Aktivität der Probanden und gleichen Messbedingungen verglichen werden. Besser geeignet für die Analyse der Kurzzeit-Variabilität sind die Parameter Root Mean Sum of Squared Distance (RMSSD) oder pNN50.

$$RMSSD = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (NN_{i+1} - NN_i)^2} \quad (2.2)$$

Wie Formel 2.2 zeigt, wird der HRV-Parameter RMSSD über die Quadratwurzel des Mittelwerts der Summe aller quadrierten Differenzen zwischen benachbarten RR-Intervallen berechnet [23]. Der RMSSD beschreibt hauptsächlich die parasymphatische Aktivierung und gilt als wichtiges Maß für

die Erholungsfähigkeit. Ein niedriger RMSSD-Wert kann als Hinweis für mentale oder körperliche Belastung gedeutet werden.

$$pNN50 = \frac{NN50}{N-1} \cdot 100\% \quad (2.3)$$

Der HRV-Parameter pNN50 entspricht dem prozentualen Anteil aufeinanderfolgender NN-Intervalle mit einer Differenz größer 50 Millisekunden. Während der RMSSD bessere statistische Eigenschaften besitzt, ist der pNN50 sensitiver gegenüber ersten Verschlechterungen in Bezug auf die Erholungsfähigkeit [23].

Die Auswertung im Frequenzbereich erfolgt auf Basis des Leistungsdichtespektrums (LDS), welches beispielsweise mittels Fast Fourier Transformation (FFT) bestimmt werden kann. Auf Grund der nicht-äquidistanten Abtastung der NN-Intervalle in Folge der variierenden NN-Abstände ist zuvor eine Interpolation der Zeitreihe notwendig [127]. Für die Analyse wurden durch die Task Force der European Society of Cardiology und der North American Society of Pacing and Electrophysiology vier Spektralbereiche definiert: den Ultra Low Frequency (ULF)-, Very Low Frequency (VLF)-, Low Frequency (LF)- und High Frequency (HF)-Bereich (siehe Tabelle 2.1 nach [23]). Die Angabe der Leistungswerte erfolgt entweder absolut mit der Einheit  $ms^2$  oder relativ in normierten Einheiten (n.u.) [130]. Für die Normierung wird die absolute Leistung eines Frequenzbandes auf die Gesamtleistung des LF- und HF-Bandes bezogen. Auf diese Weise kann ein Vergleich zwischen Personen trotz individueller Unterschiede in der spezifischen Bandleistung oder der Gesamtleistung erfolgen [76]. Das ULF-Band enthält Informationen über Langzeitvariationen mit einer Periodendauer größer 5 min und gibt Hinweise zu zirkadianen und metabolischen Effekten. Dieser Parameter wird nur bei Langzeitmessungen (24 h) bestimmt. Die Periodendauern der Variationen im VLF-Band liegen zwischen 25 s und 300 s. Es umfasst Effekte der Thermoregulation und des sympathischen Nervensystems. Das LF-Band enthält Variationen mit Perioden im

Tabelle 2.1: HRV-Parameter im Frequenzbereich nach [23]

Variable	Beschreibung	Frequenzbereich	Einheit
Gesamtleistung	Varianz aller NN Intervalle	nährungsweise $\leq 0,4$ Hz	$\text{ms}^2$
ULF	Leistung im ULF-Band	$\leq 0,003$ Hz	$\text{ms}^2$
VLF	Leistung im VLF-Band	0,003 Hz-0,04 Hz	$\text{ms}^2$
LF	Leistung im LF-Band	0,04 Hz-0,15 Hz	$\text{ms}^2$
LF norm	Normierte Leistung im LF-Band	0,04 Hz-0,15 Hz	n.u.
	$\frac{LF}{Gesamtleistung-VLF} \cdot 100$		
HF	Leistung im HF-Band	0,15 Hz-0,4 Hz	$\text{ms}^2$
HF norm	Normierte Leistung im HF-Band	0,15 Hz-0,4 Hz	n.u.
	$\frac{HF}{Gesamtleistung-VLF} \cdot 100$		
LF/HF	Verhältnis LF zu HF		%

Bereich 7 s bis 25 s. Es enthält Informationen über die Aktivität des Baroreflexes und wird durch sympathische und parasympathische Effekte beeinflusst. Es kann zur Beurteilung physischer und psychischer Belastung genutzt werden. Die Periodendauer der HF-Band-Variationen liegt zwischen 2,5 s und 7 s. Es unterliegt parasympathischen und respiratorischen Effekten. Die parasympathische Aktivität gilt als Indikator für die Anpassungs- und Regulationsfähigkeit und gibt Hinweise hinsichtlich der psychischen Gesundheit. Generell kann Aktivierung und Stress eine sympathische Aktivierung auslösen, die in einer Reduzierung der Leistung im HF- und einer Erhöhung im LF-Band resultiert. Einen zusätzlichen Kennwert bei der HRV-Analyse im Frequenzbereich bildet der Quotient aus LF und HF, welcher meist zur Beurteilung des Zusammenspiels von Sympathikus und Parasympathikus genutzt wird [121].

Auf Grund der komplexen physiologischen Mechanismen zur Regulierung der HRV werden auch nichtlineare Messungen für die HRV-Analyse

genutzt. Die erfassten Indizes korrelieren mit spezifischen Frequenz- und Zeitbereichsmessungen, solange sie von den gleichen Prozessen erzeugt werden. Während Stressoren und Störungen wie Diabetes einige nicht-lineare Messungen beeinträchtigen können, signalisieren erhöhte Werte nicht immer einen gesunden Organismus. So ist beispielsweise bei Patienten nach einem Herzinfarkt eine erhöhte nichtlineare HRV ein unabhängiger Risikofaktor für die Mortalität [130]. Die Visualisierung nichtlinearer Parameter erfolgt meist mit Hilfe eines Poincaré-Diagramms [64]. Dabei werden aufeinander folgende NN-Intervalle als Punkte in einem 2D-Diagramm eingetragen. Die Koordinaten ergeben sich dabei zu  $P_n = (NN_n, NN_{n+1})$ . Bei gesunden Probanden entsteht dabei meist eine Punktwolke in Form einer Ellipse, deren Analyse Informationen über die HRV-Dynamik enthält. Im Gegensatz zu Frequenzbereichsmessungen ist die Poincaré-Plotanalyse unempfindlich gegenüber Trendänderungen in den RR-Intervallen [130]. Die am häufigsten genutzten Parameter zur Poincaré-Plotanalyse sind die Standardabweichungen SD1 und SD2 zur Beschreibung der Streuung in Richtung der ersten und zweiten Winkelhalbierenden. SD1 entspricht der Breite und SD2 der Länge der Ellipse. Während der Effekt der kurzzeitigen Variabilität durch SD1 beschrieben wird, enthält SD2 in erster Linie Informationen über die Langzeitvariabilität. Bei der Auswertung ist zu beachten, dass RMSSD und SD1 ähnliche Effekte beschreiben [28].

### **Puls Transit Zeit**

Die Pulswellenlaufzeit (*engl.*: Pulse Transit Time PTT) ist die Zeitspanne, die eine Pulswelle für das Zurücklegen einer bestimmten Strecke im Gefäßsystem innerhalb eines Herzzyklus benötigt. Sie wird über den Quotient aus Pulswellengeschwindigkeit (PWG) und der Distanz ( $d$ ) bestimmt.

$$PTT = \frac{d}{PWG} \quad (2.4)$$

Die PWG ergibt sich als Funktion von Elastizität, Gefäßdurchmesser, Gefäßwanddicke und Dichte des Blutes. Sie entspricht nicht der Strömungsgeschwindigkeit des Blutes, welche auf Grund der Dehnungsfähigkeit geringer ist. Durch die Abhängigkeit von Gefäßparametern dient die Messung der PWG, respektive der PTT, zur Beurteilung kardiovaskulärer Risikofaktoren wie der Gefäßversteifung. In zahlreichen Studien wurde gezeigt, dass mittels der PTT der absolute Blutdruck bestimmt werden kann. Dazu bedarf es jedoch einer Kalibrierung, um eine individuelle Angleichung der Funktion zwischen Blutdruck und PTT zu erreichen. Zudem konnte bisher lediglich eine Kurzzeitstabilität der Messmethode nachgewiesen werden [107].

### 2.1.3 Haut

#### **Aufbau und Durchblutung der Haut**

Die menschliche Haut erreicht eine Fläche von bis zu zwei Quadratmetern und ist das größte Organ des Menschen. Sie schützt den Körper nach außen hin und erfüllt auf Grund ihrer Sinnesrezeptoren ihre Aufgabe als wichtiges Sinnesorgan. Sie ist zudem ein wesentliches Stoffwechselorgan und reguliert über die Abgabe von Schweiß den Wasserhaushalt und die Körpertemperatur. Die menschliche Haut ist grundsätzlich in drei Schichten aufgebaut, der äußeren Epidermis (Oberhaut), der Dermis (Lederhaut, *lat. Corium*) und der Hypodermis (Unterhaut, *lat. Subcutis*). Die Dermis kann nochmals in zwei Schichten unterteilt werden. Sie umfasst die Papillenschicht, die die Versorgung der Epidermis gewährleistet, und die Retikularschicht, die der Haut ihre Festigkeit und Elastizität verleiht [29].

Die Durchblutung der Haut wird durch ein Netzwerk aus Arterien und Venen in der Hypodermis gewährleistet. Dieses Netzwerk wird als subkutaner Plexus bezeichnet. Abzweigungen der Arterien des subkutanen Plexus versorgen die darüber liegenden Hautschichten mit Blut. Zwischen der Retikular- und Papillarschicht der Dermis existiert ein weiteres Netz-

werk aus Adern, der dermale Plexus, aus welchem die Kapillaren für die Versorgung der Epidermis aufsteigen. Die Epidermis besteht dagegen aus undurchblutetem Gewebe [29]. Der Aufbau wird in Abbildung 2.6 in Abschnitt 2.3.1 visualisiert. Durch die Lage an der Körperoberfläche und der hohen Dichte an Blutgefäßen bietet die Haut die Möglichkeit zur nichtinvasiven Durchblutungsmessung mit dem Ziel der Ableitung wichtiger Vitalparameter (z.B. Herzfrequenz, Sauerstoffsättigung, etc.), sowie der Untersuchung der Hämodynamik des Blutes oder des Zustands des Gefäßsystems [50]. Im Vergleich zu anderen Körperregionen ist die Hautdurchblutung im Gesicht besonders hoch [149], ein Grund für die Wahl als Messregion im Rahmen dieser Arbeit.

### **Durchblutung der Gesichtshaut**

Nach [103] wird das Gesicht von zahlreichen Ästen aus den äußeren und inneren Halsschlagadern versorgt. Mit Ausnahme der Hauptadern haben diese Gefäße einen kleinen Durchmesser und nur wenige von ihnen gehen als direkte Arterien zur Gesichtshaut über. In der Haut bilden die Hautarterien ein Gefäßmuster, das durch feine dermale Verbindungen gekennzeichnet ist. Die Anzahl der Kapillarschlingen unter der Epidermis des Gesichtes variiert von Region zu Region deutlich. Auch der Durchmesser der Kapillaren ist in verschiedenen Hautbereichen unterschiedlich. Die Blutgefäße der oberen Dermis sind in annähernd konische Gefäßeinheiten organisiert, deren Basen sich unter der Epidermis überlappen. Die Analysen der oberflächennahen Durchblutungsstärke in [103] zeigen, dass die Stirn, die periorale Region, die Nase, das Philtrum, die Lippe, das Kinn und die Augenregion reichlich kapillar durchblutet sind. Der Schläfenbereich, die seitlichen Wangen und der Kiefer sind weniger und das Ohr mäßig gefäßreich.

## Regulation der Hautdurchblutung

Die Hautdurchblutung wird autoregulatorisch gesteuert und durch thermische, mechanische, chemische und psychische Reize beeinflusst. Dabei spielen sowohl lokale Einflüsse (Kältereiz) als auch Fernwirkungen (nerval oder humoral) eine Rolle. Gleichzeitig bestehen große lokale Unterschiede in der Auswirkung solcher Reize auf die Durchblutung. Die Durchblutung der Haut wird durch verschiedenartige Mechanismen geregelt. Diese Regelung erlaubt, dass der Blutstrom in manchen Arealen bis auf das Tausendfache der minimalen Durchblutung ansteigen kann [83]. Auf diese Weise trägt die Hautdurchblutung maßgeblich zur Thermoregulation des menschlichen Körpers bei, da sich die Wärmeleitfähigkeit der Haut mit ihrer Durchblutungsgröße ändert. Besonders interessant ist auch der psychische Einflussfaktor auf die Hautdurchblutung. So führen Angst oder Ekel zu einer verringerten Durchblutung und daraus resultierender Blässe, wohingegen Scham und Zorn die Durchblutung fördern und es zum Erröten der Haut (*engl.*: „Blushing“) kommt. Bei Stress kann es zu partiell gesteigerter Durchblutung und daraus entstehenden fleckig erröteten Flächen kommen [105]. Für die Messung der Hautdurchblutung und Beurteilung möglicher Gefäßkrankheiten beim Menschen gibt es bisher drei medizinisch anerkannte Verfahren: die Venenverschluss-Plethysmographie, die Wärmeleitmessung (Wärme-Clearance) und die Radioisotopen-Clearance. Für nähere Information zu den Verfahren wird auf [63] verwiesen.

## Hautfarbe

Die Farbe der Haut hängt neben dem Perfusionsgrad, von weiteren Faktoren wie dem Sauerstoffpartialdruck ab. So lässt sich an dünnen Hautschichten wie den Lippen die Sauerstoffsättigung des Blutes erkennen. Wenn diese unter 70% liegt sind die Lippen bläulich gefärbt, was als Zyanose bezeichnet wird. Gründe hierfür können zu geringe Durchblutung in Folge von Kälte, mangelnder Gasaustausch in der Lunge oder Herzerkrankun-

gen sein. Einen eher statischen Einfluss auf die Hautfarbe hat die Pigmentierung durch Melanin. Abhängig von der genetisch und umweltbedingten Melaninbeladung durch UV-Bestrahlung können so verschiedene Hauttypen, von Hauttyp I (sehr hell) bis Hauttyp VI (dunkelbraun bis schwarz), definiert werden. Es gibt zwei Arten von Melanin. Eumelanin ist ein braunes bis schwarzes Pigment und bestimmt den Hauttyp beziehungsweise die Hautfarbe. Phäomelanin ist ein rotes bis gelbes Pigment und erzeugt vor allem bei hellen Hauttypen einen rötlichen oder gelblichen Unterton. Die perfusionsbedingte Farbänderung ist beim hellen Hauttyp stärker ausgeprägt.

## 2.2 Messtechnische Erfassung der Herzaktivität

### 2.2.1 Elektrokardiographie

In Abschnitt 2.1.1 wurde die elektrische Erregung des Herzens beschrieben. Die elektrischen Potentialänderungen am Herzen können mit Hilfe der Elektrokardiographie (EKG) messtechnisch erfasst werden. Die Aufzeichnungen über die Zeit werden als Elektrokardiogramm (ebenfalls EKG) bezeichnet. Es stellt die Projektion eines Spannungsvektors dar, der durch die Summation der Potentialänderungen im Herzen entsteht. Zur Messung werden Elektroden an der Körperoberfläche angebracht und die darüber gemessenen Potentiale, welche auch Ableitungen genannt werden, nach geeigneter EKG-Vorverstärkung im EKG-Gerät aufgezeichnet. Das Standard-EKG umfasst zwölf Ableitungen mit zehn Elektroden.

In Abbildung 2.4 wurde bereits der typische Verlauf eines EKGs gezeigt. Die elektrische Erregung der Vorhöfe ist als P-Welle zu erkennen. Die drei darauf folgenden Zacken Q, R und S werden als QRS-Komplex bezeichnet und beschreiben die Erregung der Kammern. Die anschließende T-Welle wird durch die Repolarisation der Kammern verursacht. Je nach Position der Elektroden ergibt sich eine veränderte Form des EKGs.

Aus diagnostischer Sicht gibt das EKG Auskunft über den Herzrhythmus, Störungen der Erregungsbildung oder Erregungsleitung, den Zustand der Herzmuskulatur und den Lagetyp des Herzens im Brustkorb. Außerdem können Durchblutungsstörungen der Herzkranzgefäße und daraus resultierende Herzinfarkte erkannt werden. Bei der Anwendung wird zwischen Ruhe-, Belastungs- und Langzeit-EKG unterschieden, welche jeweils unterschiedliche Indikatoren besitzen. Beispielsweise können mittels Belastungs-EKG die koronare Herzkrankheit (KHK) und Herzrhythmusstörungen besser diagnostiziert werden. Da das EKG in dieser Arbeit in erster Linie als Referenzsystem genutzt wird, wird das Messprinzip an dieser Stelle nicht ausführlicher behandelt. Für eine detaillierte Beschreibung wird auf die einschlägige Literatur verwiesen.

### **2.2.2 Photoplethysmographie und Pulsoximetrie**

Neben der elektrischen Erregung des Herzens kann die mechanische Ausbreitung der Pulswelle im Körper messtechnisch erfasst werden. Anders formuliert werden zur Untersuchung der Herzaktivität Blutvolumenschwankungen über bestimmte Zeitintervalle sensorisch aufgezeichnet und ausgewertet. Generell werden Verfahren die eine graphische Darstellung der Volumenschwankungen eines Körperteils oder Organs ermöglichen [117], als Plethysmographie bezeichnet.

Eine bekannte hämodynamische Untersuchungsmethode ist die Photoplethysmographie PPG zur Analyse des venösen Gefäßsystems. Mit Hilfe des optischen Verfahrens werden Blutvolumenschwankungen im subkutanen Venenplexus erfasst. Der zu Grunde liegende Effekt ist die Lichtabsorption durch den Blutfarbstoff Hämoglobin. Zur Messung wird infrarotes Licht einer bestimmten Wellenlänge auf ein Hautareal eingestrahlt und der reflektierte Anteil während der Durchführung standardisierter Bewegungsprogramme gemessen. Der reflektierte Lichtanteil nimmt dabei mit steigender subkutaner Blutfüllung absorptionsbedingt ab. Relative Än-

derungen zwischen untersuchten Messorten können schließlich Auskunft über lokale Dysfunktionen des Venensystems geben. Auch kann auf diese Weise die Funktion der Muskel-Venen-Pumpe oder der Venenklappen untersucht werden [68].

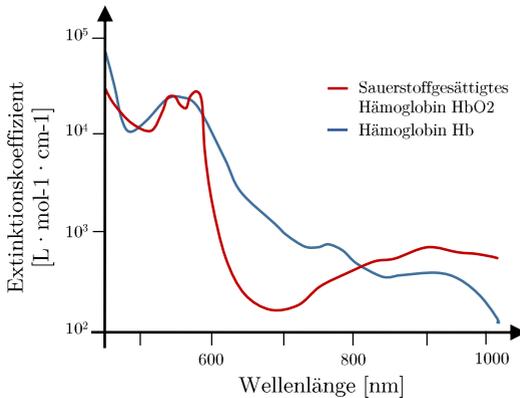


Bild 2.5: Lichtabsorption von menschlichem Blut

Die sogenannte Pulsoximetrie basiert auf dem gleichen Prinzip, dient jedoch der Analyse des arteriellen Systems. Der optische Sensor ist dabei meist in Form eines Fingerclips realisiert. Zwei LEDs mit verschiedenen Wellenlängen senden rotes und nahinfrarotes Licht (660 nm und 905 nm – 920 nm) aus, welches durch Hämoglobin zu einem großen Teil absorbiert wird. Je nach Sensortyp wird der Transmissions- oder Reflexionsanteil des Lichts mittels Photodiode oder Phototransistor gemessen. Dabei absorbieren oxygeniertes und desoxygeniertes Hämoglobin die spektralen Anteile des Lichts unterschiedlich stark (siehe Abbildung 2.5):

- desoxygeniertes Hämoglobin: Absorptionsmaximum bei 555 nm
- oxygeniertes Hämoglobin: Absorptionsmaxima bei 541 nm und 577 nm

Durch die Verwendung zweier unterschiedlicher Beleuchtungswellenlängen und geeigneter Verhältnissbildung der Messwerte, lässt sich auf diese

Weise die Sauerstoffsättigung des Blutes bestimmen, was der eigentlichen Indikation der Pulsoximetrie entspricht. Zusätzlich erhält man das Photoplethysmogramm, welches die durch Herzkontraktion induzierten Blutvolumenschwankungen des arteriellen Systems abbildet und so die Bestimmung der momentanen Herzfrequenz ermöglicht.

### 2.2.3 Photoplethysmography Imaging

Als Photoplethysmography Imaging (PPGI) wird die kontaktlose Messung der Perfusion in oberflächennahen Hautschichten und daraus ableitbarer Parameter wie der Herzfrequenz mittels Kamera als optischem Sensor bezeichnet. In der Literatur finden sich auch die Bezeichnungen Remote Photoplethysmography (rPPG) oder Imaging Photoplethysmography (iPPG). PPGI nutzt mit der blutvolumenabhängigen Lichtabsorption von Blutgefäßen dasselbe Grundprinzip wie die namensgebende PPG. Anstelle eines einzelnen Sensors in Form einer Photodiode, wird jedoch ein Array von Einzelsensoren, den Bildpixeln des Kamerasensors, für die messtechnische Erfassung verwendet. Auf diese Weise kann die Perfusionsrhythmik orts aufgelöst und über Entfernung hinweg erfasst werden. Zur Erzeugung des Messeffekts wird ein bestimmtes Hautareal entweder vom vorhandenen Umgebungslicht oder von speziell auf die Anwendung ausgerichteten Lichtquellen beleuchtet. Folglich wird zwischen passiven und aktiven PPGI-Systemen unterschieden.

Um das Signal-Rausch-Verhältnis (SNR) zu erhöhen, erfolgt die Signalextraktion über Mittelwertbildung der Pixelwerte in einer definierten Messregion (Region of Interest (ROI)) auf der Haut. Als ROI werden aus Gründen der Zugänglichkeit zum Hautareal entweder der Unterarm [148] [55], die Handinnenfläche [131] [77], am häufigsten aber das Gesicht [52] [109] genutzt. Das extrahierte Signal setzt sich aus verschiedenen Komponenten zusammen. Es enthält statische Anteile aus der direkten Oberflächenreflexion der Haut und je nach Eindringtiefe des Lichtes, von Gewebe und

Knochen. Gleichzeitig besitzt das Signal eine zeitveränderliche Komponente, welche durch Blutvolumenschwankungen hervorgerufen wird. Das Ziel eines PPGI-Algorithmus ist es, dieses zeitveränderliche Blutvolumenpuls (BVP)-Signal aus dem Mischsignal zu extrahieren. Eine besondere Herausforderung entsteht, wenn durch Bewegung oder Beleuchtungsänderung die statischen Anteile der Oberflächenreflexion einen dynamischen und ebenfalls zeitveränderlichen Charakter bekommen. Im Speziellen ist dies der Fall, wenn sich die Frequenzanteile dieser Störkomponenten im Bereich der Herzfrequenz befinden.

Die Anwendungsgebiete eines solchen Systems zur Remote-Messung von Vitalparametern sind äußerst vielfältig (vgl. Kapitel 1.1). So kann es zu Monitoringzwecken im medizinisch-pflegerischen Bereich eingesetzt werden, beispielsweise zur Überwachung von Frühgeborenen oder in der Schlafanalyse. Das Ziel bei diesen Anwendungen ist es, kabelgebundene und kontaktbasierte Lösungen zu ersetzen. Auch Szenarien in denen das Anlegen von Sensoren erschwert sein kann, wie beispielsweise auf Isolierstationen, sind für den Einsatz von PPGI-Systemen geeignet. Voraussetzung für den Einsatz in dieser Domäne ist jedoch eine sehr hohe Messgenauigkeit, welche medizinischen Standards genügt. Weitere Anwendungsdomänen umfassen den Lifestyle-Bereich mit der Anwendung Indoor-Cardio-Fitness [156], sowie den Security-Sektor mit den Beispielapplikationen Überwachung von öffentlichen Plätzen und Anti-Spoofing- bzw. Liveness-Detection-Systemen [87]. Durch die Korrelation der Herzfrequenz oder der Herzratenvariabilität mit Stress und Müdigkeit [152] [104] kann ein PPGI-System auch wichtige Informationen zur Zustandsanalyse einer Person beitragen. Neben der Erfassung und Analyse der Herzfrequenz werden auch PPGI-basierte Ansätze zur Messung der Sauerstoffsättigung [55], der PTT [100] und des Blutdrucks [60] erforscht.

### 2.2.4 Kamera als PPGI-Sensor

Auf eine Beschreibung der technischen und optischen Grundlagen von Kameras wird an dieser Stelle verzichtet und auf [47] und [45] verwiesen. Für PPGI ergeben sich zunächst Einflüsse durch den messtechnischen Aufbau des Gesamtsystems, beispielsweise durch Typ und Qualität des Objektivs, welches unter anderem eine hohe Lichthelligkeit gewährleisten sollte, den Bildöffnungswinkel, den ISO-Wert oder die Verschlusszeit des Bildsensors, aber auch die spektralen Eigenschaften der Lichtquelle. Gleichzeitig gibt es Parameter wie die Haut-Pixeldichte in der Messregion, die direkt vom Abstand der Apertur zum Messobjekt abhängen.

Hinsichtlich der Auswahl des Bildsensors wurde in mehreren Arbeiten nachgewiesen, dass auch einfache Sensoren wie sie in Webcams verbaut werden, den Messeffekt im Vergleich zu höherwertigen Sensoren ausreichend gut erfassen [86] [140]. Dies ist jedoch auch vor dem Hintergrund anwendungsspezifischer Artefakteinflüsse zu sehen, welche durch bestimmte Eigenschaften höherwertiger Bildsensoren teils hardwareseitig gemindert werden können. In der folgenden Tabelle ist eine Einschätzung über den Einfluss verschiedener Bildsensoreigenschaften auf die Messqualität eines PPGI-Systems gegeben, welche auf Basis publizierter Studien wie [139] getroffen wurde. Darüber hinaus können Kameras über spezielle, vorteilhafte Funktionen verfügen, wie beispielsweise die Möglichkeit zur automatischen Kameraregelung auf die ROI. Erfolgt die Aufnahme nicht im RAW-Format sind zudem negative Einflüsse durch Kompressionseffekte zu beachten [96] (siehe auch 4.3.1).

Tabelle 2.2: Bewertung des Einflusses verschiedener Kameraeigenschaften auf die PPGI-Messqualität

<b>Kameraeigenschaft</b>	<b>Einfluss</b>
Sensortyp (CMOS,CCD)	mittel
Framerate (>20 Hz) [Hz]	mittel
SensorgroÙe [Zoll]	hoch
Auflösung [Pixel]	niedrig
Dynamikbereich (Bittiefe) [dB]	mittel
Sensorrauschen [dB]	hoch
spektrale Effizienz [% bei $\lambda$ ]	hoch

## 2.3 Optische Grundlagen der oberflächennahen Lichtreflexion

### 2.3.1 Dichromatisches Reflexionsmodell

Bei der Modellierung der Lichtreflexion durch menschliche Haut sind mehrere Effekte zu beachten. Trifft Licht auf die Haut wird ein Teil an der Oberfläche reflektiert, der andere Teil dringt in die Haut ein und wird dort von den unterliegenden Schichten absorbiert, gestreut oder reflektiert. Diese Zusammenhänge werden unter anderem durch das dichromatische Reflexionsmodell nach Shafer [129] beschrieben. Es unterteilt die von einem beleuchteten Körper reflektierte Strahlung in zwei Komponenten, welche sich in ihrer spektralen Zusammensetzung unterscheiden (siehe Abbildung 2.6). Eine Komponente ist die durch Spiegelung an der Oberfläche des Körpers verursachte Oberflächenreflexion. Da diese Strahlung nicht in den Körper eindringt und einer einfachen Spiegelung entspricht, ist das Spektrum dieser Strahlung identisch mit dem der Lichtquelle. Die zweite Komponente ist die Körperreflexion. Sie beschreibt die Interaktion des Lichtes mit den tiefer gelegenen Schichten des beleuchteten Körpers. Dabei dringt das Licht ab-

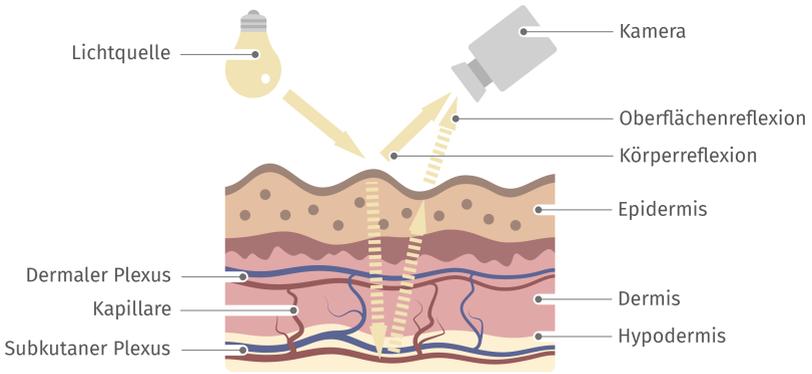


Bild 2.6: Modell der oberflächennahen Lichtreflexion der Haut

hängig von der Wellenlänge unterschiedlich tief in die Haut ein, sodass es aus den unterschiedlichen Hautschichten mit veränderter spektraler Zusammensetzung reflektiert wird. Diese Strahlung hat keine Vorzugsrichtung und ist daher diffus.

Die folgende mathematische Beschreibung orientiert sich an [70]:

$$L = L_o(\lambda, n, s, v) + L_k(\lambda, n, s, v) \quad (2.5)$$

$$= g_o(n, s, v) \cdot c_o(\lambda) + g_k(n, s, v) \cdot c_k(\lambda) \quad (2.6)$$

$$= g_o(n, s, v) \cdot \begin{pmatrix} c_{o,R} \\ c_{o,G} \\ c_{o,B} \end{pmatrix} + g_k(n, s, v) \cdot \begin{pmatrix} c_{o,R} \\ c_{o,G} \\ c_{o,B} \end{pmatrix} \quad (2.7)$$

Die Gesamtstrahlungsdichte  $L$  ergibt sich als Summe der Oberflächen- und Körperreflexionskomponente. Diese Komponenten sind jeweils abhängig von der Wellenlänge der Strahlung  $\lambda$ , der Oberflächennormalen  $n$ , der Beleuchtungsrichtung  $s$  und der Blickrichtung der Kamera  $v$ . Es wird nun angenommen, dass sich die Komponenten  $L_o$  und  $L_k$  in geometrische Komponenten  $g(n, s, v)$  und Farbkomponenten  $c(\lambda)$  zerlegen lassen. Der Vorteil be-

steht nun darin, dass die geometrische Komponente der Oberflächenreflexion  $g_o(n, s, v)$  nicht modelliert werden muss. Die geometrische Körperreflexionskomponente kann durch das Lambertsche Gesetz beschrieben werden. Die Farbvektoren lassen sich auf drei Wellenlängenbereiche des roten, grünen und blauen Spektralbereich eingrenzen. Somit kann  $L$  über einen dreidimensionalen RGB-Farbvektor beschrieben werden.  $g_o(n, s, v)$  und  $g_k(n, s, v)$  können als Skalierungsfaktoren angesehen werden. Die Vektoren  $c_o$  und  $c_k$  spannen im RGB-Farbraum eine Ebene auf, die Dichromatische Ebene oder Farbsignalebene. Durch Analyse der Farbvektoren in der Ebene können Reflexionskomponenten separiert werden, was beispielsweise für die Glanzlichteliminierung in Bildern angewendet wird. Für PPGI kann diese Eigenschaft genutzt werden, um die pulsinduzierten Komponenten von anderen Reflexionskomponenten zu trennen.

### 2.3.2 Erweiterung des Reflexionsmodells hinsichtlich PPGI

Aufbauend auf den Gleichungen des vorigen Abschnitts, ist es möglich eine mathematische Beschreibung der PPGI-Problestellung zu formulieren. Teil des mathematischen Modells ist neben der Haut eine natürliche oder künstliche Lichtquelle, sowie eine Kamera als optischer Sensor. Der erfasste Hautpixelwert ist zunächst von der Lichtquelle, den Hauteigenschaften und der verwendeten Kamera abhängig. Zudem wird er im zeitlichen Verlauf von den Distanzen zwischen Lichtquelle und Haut, sowie Haut und Kamera beeinflusst. Der Farbwert variiert somit mit der Bewegung, Rotation und Durchblutung der Haut.

Wang et al. [157] beschreiben das Modell mit der folgenden Basisformel. Dabei wird als Kamera eine RGB-Farbkamera vorausgesetzt. Die Lichtquelle besitzt eine variable Intensität, ihre spektrale Zusammensetzung wird als konstant angenommen.

$$C_k(t) = I(t) \cdot (\mathbf{v}_s(t) + \mathbf{v}_d(t)) + \mathbf{v}_n(t) \quad (2.8)$$

$C_k(t)$  beschreibt den von der Kamera registrierten zeitlichen Verlauf des Farbwertes für das  $k$ -te Hautpixel. Der Farbwert wird mit Hilfe des RGB-Farbraums dargestellt und besitzt daher drei Komponenten: den Rotkanal, den Grünkanal und den Blaukanal.  $I(t)$  bezeichnet die Helligkeit der Lichtquelle, die einerseits durch die Helligkeit der Lichtquelle selbst, andererseits durch die Veränderung des Abstandes zwischen Lichtquelle, Haut und Kamera verursacht wird.  $\mathbf{v}_s(t)$  und  $\mathbf{v}_d(t)$  beschreiben die spiegelnde Oberflächenreflexion und den diffusen Anteil der Körperreflexion der in der Kamera detektierten Strahlung.  $\mathbf{v}_n(t)$  entspricht dem durch Quantisierung entstehenden Fehler in der Kamera. Die Oberflächenreflexion  $\mathbf{v}_s(t)$  wird in zwei weitere Komponenten zerlegt: eine stationäre Komponente  $s_0$  und eine sich mit der Bewegung der Haut verändernde, zeitabhängige Komponente  $s(t)$ . Da das Spektrum der Oberflächenreflexion identisch mit dem Spektrum der Lichtquelle ist, ist auch deren normierter Farbwert  $\mathbf{u}_s$  identisch. Insgesamt ergibt sich so für die Oberflächenreflexion  $\mathbf{v}_s(t)$  der folgende Ausdruck:

$$\mathbf{v}_s(t) = \mathbf{u}_s \cdot (s_0 + s(t)) \quad (2.9)$$

Auch der diffuse Anteil wird in zwei Komponenten zerlegt. So ergibt sich erneut eine stationäre Komponente, die abhängig vom normierten Farbwert der Haut  $\mathbf{u}_d$  und der Stärke der Reflexion  $d_0$  ist. Die zweite, zeitabhängige Komponente setzt sich aus dem Blutvolumenpulssignal  $p(t)$  und dessen relativer Stärke  $\mathbf{u}_p$  in den unterschiedlichen Farbkanälen zusammen:

$$\mathbf{v}_d(t) = \mathbf{u}_d \cdot d_0 + \mathbf{u}_p \cdot p(t) \quad (2.10)$$

Einsetzen der Gleichungen 2.9 und 2.10 in Gleichung 2.8 ergibt:

$$C_k(t) = I(t) \cdot (\mathbf{u}_s \cdot (s_0 + s(t)) + \mathbf{u}_d \cdot d_0 + \mathbf{u}_p \cdot p(t)) + \mathbf{v}_n(t) \quad (2.11)$$

Die stationären Anteile der beiden Reflexionskomponenten lassen sich zu einer einzelnen Komponente zusammenfassen:

$$\mathbf{u}_c \cdot c_0 = \mathbf{u}_s \cdot s_0 + \mathbf{u}_d \cdot d_0 \quad (2.12)$$

Eingesetzt in Gleichung 2.11 ergibt sich:

$$C_k(t) = I(t) \cdot (\mathbf{u}_c \cdot c_0 + \mathbf{u}_s \cdot s(t) + \mathbf{u}_p \cdot p(t)) + \mathbf{v}_n(t) \quad (2.13)$$

Die Intensität der Lichtquelle  $I(t)$  wird ebenfalls in zwei Anteile zerlegt: Die erste Komponente ist ein fester Anteil  $I_0$ , die zweite Komponente ist die zeitliche Veränderung  $i(t)$ , welche von der Abstandsänderung zwischen Lichtquelle, Haut und Kamera abhängig ist. Zudem wird die zeitlich nicht fixierte Intensität der Lichtquelle berücksichtigt:

$$I(t) = I_0 \cdot (1 + i(t)) \quad (2.14)$$

Durch Einsetzen von Gl. 2.14 in Gl. 2.13 ergibt sich folgende Gleichung für das von der Kamera gemessene Signal  $C_k(t)$ :

$$C_k(t) = I_0 \cdot (1 + i(t)) \cdot (\mathbf{u}_c \cdot c_0 + \mathbf{u}_s \cdot s(t) + \mathbf{u}_p \cdot p(t)) + \mathbf{v}_n(t) \quad (2.15)$$

Durch Mittelung über eine große Zahl an Hautpixeln wird das Quantisierungsrauschen  $v_n(t)$  minimiert und kann als vernachlässigbar angesehen werden. Das Ziel von PPGI-Algorithmen ist es, aus dem gemittelten Messsignal  $C(t)$  die blutvolumenpulsinduzierte Komponente  $p(t)$  zu extrahieren.

### 2.3.3 Einfluss der spektralen Zusammensetzung des Lichts

In dem beschriebenen Modell wird die spektrale Zusammensetzung der Lichtquelle als konstant angenommen. Trotzdem gilt es den Einfluss verschiedener Beleuchtungswellenlängen genauer zu betrachten. Wie bereits erwähnt dringt das Licht abhängig von der Wellenlänge unterschiedlich tief in die Haut ein, sodass sich die beschriebene Pulskomponente aus Reflexionen unterschiedlicher Schichten bildet, was auch Einfluss auf die Signalmorphologie hat. Kürzere Wellenlängen im gelben bis grünen Wellenlängenbereich erreichen nur die kleineren und weniger pulsierenden Gefäße der oberflächennahen Retikular- und Papillarschicht der Dermis. Jedoch ist die Empfindlichkeit für Blutvolumenschwankungen auf Grund der Absorptionskurve von Hämoglobin (Abbildung 2.5) in diesem Wellenlängenbereich stärker als im (infra-)roten Bereich. Bei Verwendung einer roten oder infraroten Beleuchtung gelangt das Licht bis in die stärker pulsierenden Gefäße des subkutanen Plexus, bei gleichzeitig geringerer Absorption durch Hämoglobin. Der nichtlineare Zusammenhang zwischen der PPGI-Signalstärke und der Beleuchtungswellenlänge wurde von Coral et al. gezeigt [1]. Die Messungen zeigen, dass die PPGI-Signalstärke im Bereich des Absorptionsmaximums bei  $550\text{nm}$  am stärksten ist und danach schnell in beide Richtungen abfällt. Lediglich im Wellenlängenbereich zwischen  $800\text{nm}$  und  $925\text{nm}$  ergibt sich nochmals ein leichter Anstieg. In Bezug auf die PPGI-Signalstärke besitzen die Absorptionseigenschaften von Hämoglobin somit stärkeren Einfluss als das Erreichen der tieferen, stärker durchbluteten Hautschichten des subkutanen Plexus. Mit Blick auf die Signalmorphologie weisen kürzere Wellenlängen, auf Grund der geringeren Pulsatilität der kleinen Gefäße, einen weniger ausgeprägten Hochfrequenzanteil des Pulssignals auf und erscheinen gleichzeitig leicht verzögert [101].

## 2.4 Bild- und Signalverarbeitung

Ein großer Teil der Arbeit befasst sich mit der Entwicklung von Algorithmen zur robusten Detektion des Gesichts als Messregion (Kapitel 6) und der nachfolgenden Extraktion des Messsignals (Kapitel 7) innerhalb der ROI. Im Folgenden werden die zum Verständnis notwendigen Methoden der Bild- und Signalverarbeitung einführend beschrieben.

### 2.4.1 Principal Component Analysis

Normalerweise besitzen Merkmale, die der Beschreibung der Gesichtstexturinformation dienen, eine relative hohe Dimensionalität, was sich negativ auf die benötigte Rechenleistung auswirkt. Die Principal Component Analysis (PCA) ist eine mathematische Methode, welche sich unter anderem zur Dimensionalitätsreduktion eines Merkmalsraums eignet, wobei die größtmögliche Menge an Merkmalsinformationen erhalten bleibt. Dabei wird versucht die hohe Zahl an Merkmalsvektoren durch eine geringere Zahl möglichst aussagekräftiger Linearkombinationen, den Hauptkomponenten, zu nähern. Das Verfahren ist ein wichtiger Bestandteil des im nächsten Abschnitt beschriebenen Active Appearance Model (AAM)-Ansatzes für das Tracking von Gesichtern. Der folgende Abschnitt liefert die zugehörige mathematische Beschreibung.

Ein gegebener Merkmalspunkt  $\mathbf{x}_i$  kann im Merkmalsraum geringerer Dimension wie folgt ausgedrückt werden:

$$\widehat{\mathbf{x}}_i = \bar{\mathbf{x}} + \sum_{j=1}^M \mathbf{a}_{ij} \mathbf{e}_j, \quad i = 1, \dots, n, \quad M < N, \quad (2.16)$$

mit  $n$  als Anzahl von Merkmalsvektoren und  $N$  als Dimension des originalen Merkmalsraumes. Der Satz an Vektoren  $\mathbf{e}_j$  ( $j = 1, \dots, M$ ) entspricht dem Koordinatensystem des reduzierten Merkmalsraumes. Der Merkmalsvektor  $\mathbf{x}_i$  wird durch einen  $M$ -dimensionalen Parametervektor  $\mathbf{a}_i = [\mathbf{a}_{i1}, \dots, \mathbf{a}_{iM}]^T$

repräsentiert. Der Informationsverlust der Merkmale bei der Transformation wird dabei wie folgt definiert:

$$J = \sum_{i=1}^n \left( \mathbf{x}_i - \bar{\mathbf{x}} - \sum_{j=1}^M \mathbf{a}_{ij} \mathbf{e}_j \right)^2 \quad (2.17)$$

Dieser kann somit als Energieverlust in den Vektoren ausgedrückt werden, orthogonal zum Unterraum der durch  $\mathbf{e}_j$  aufgespannt wird. Der minimale Informationsverlust für ein gegebenes  $\mathbf{x}_i$  wird nur erreicht, wenn  $\bar{\mathbf{x}}$  dem geometrischen Zentrum der Merkmalspunkte entspricht und  $\mathbf{e}_j$  ( $j=1, \dots, M$ ) als Eigenvektoren der Kovarianzmatrix durch die  $M$  größten Eigenwerte beschrieben werden. Zur Berechnung der Koordinatenvektoren wird zunächst eine Kovarianzmatrix der Merkmalsdaten gebildet:

$$C = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T, \quad (2.18)$$

wobei  $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i)$  das Zentrum der Merkmalsvektoren repräsentiert. Im Anschluss werden die Eigenvektoren  $\mathbf{e}_j$  und die Eigenwerte  $\lambda_j$  der Kovarianzmatrix berechnet und in absteigender Reihenfolge entsprechend den Eigenwerten sortiert. Die ersten  $M$  Eigenvektoren werden dann als Koordinaten des  $M$ -dimensionalen Merkmalsraums gewählt. Da  $\mathbf{e}_j$  ( $j=1, \dots, M$ ) einem Satz orthogonaler Basisvektoren entspricht, können die Parameter  $\mathbf{a}_{ij}$  als Projektion des Merkmals  $\mathbf{x}_i$  auf  $\mathbf{e}_j$  gesehen werden:

$$\mathbf{a}_{ij} = \mathbf{e}_j^T (\mathbf{x}_i - \bar{\mathbf{x}}) \quad (2.19)$$

Somit kann die Varianz der Merkmalsprojektion auf  $\mathbf{e}_j$  wie folgt ausgedrückt werden:

$$C_j = \frac{1}{n-1} \sum_{i=1}^n a_{ij}^2 = \frac{1}{n-1} \sum_{i=1}^n \mathbf{e}_j^T (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{e}_j = \mathbf{e}_j^T \mathbf{C} \mathbf{e}_j = \lambda_j \quad (2.20)$$

Dies entspricht den  $j$ -größten Eigenwerten der Kovarianzmatrix. Dadurch repräsentieren die Koordinaten  $\mathbf{e}_j$  ( $j=1, \dots, M$ ) ebenfalls die Richtungen, in welchen die Merkmalsdaten die größte Variation besitzen.

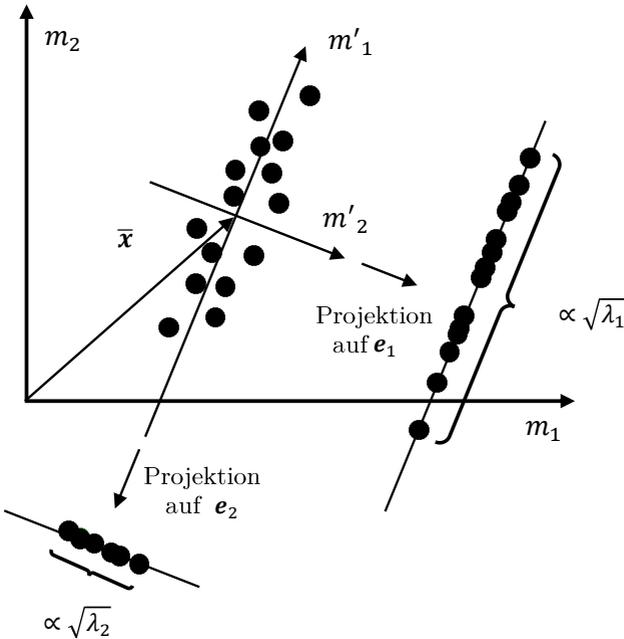


Bild 2.7: PCA im zweidimensionalen Merkmalsraum

Die PCA kann als eine orthogonale lineare Transformation betrachtet werden, die die Daten in neue Koordinaten projiziert, sodass diese dabei die größte Varianz aufweisen. Abbildung 2.7 zeigt die PCA im zweidimen-

sionalen Merkmalsraum. Die Vektoren  $e_1$ ,  $e_2$  entsprechen den Eigenvektoren der Kovarianzmatrix. Sollen die Daten über lediglich einen Parameter beschrieben werden, wird  $e_1$  genutzt, da dieser die größere Varianz beinhaltet, und somit eine größere Informationsdichte der Daten beibehält als Vektor  $e_2$ . Die PCA ermöglicht somit die Transformation der Daten aus dem 2D-Koordinatensystem  $m_1, m_2$  in ein eindimensionales Koordinatensystem mit Achse  $m'_1$ , bei der ein Großteil der Information erhalten bleibt.

### 2.4.2 Active Appearance Modell

Das Active Appearance Modell (AAM) von Cootes, Edwards und Taylor [30] liefert einen robusten Ansatz zur Erkennung und Verfolgung deformierbarer Objekte in Bildern, wie zum Beispiel Gesichtern. Der Ansatz entspricht der Erweiterung des Active Shape Modells. Dabei werden zur Detektion und Verfolgung eines Objektes dessen Kanten betrachtet und nach dem Prinzip der aktiven Konturen eine zuvor berechnete Durchschnittsform an das Objekt angepasst. Das AAM betrachtet zusätzlich zur Form die Textur des Objekts. Somit besteht die Basis in einem Modell aus Formen und Strukturen, wodurch die Güte der Schätzung insbesondere vom Training des Modells abhängt. Der Einsatz des AAM für die Gesichtsanalyse wurde erstmals in [37] vorgestellt. Da das Einsatzgebiet des AAM in dieser Arbeit ebenfalls die Gesichtsanalyse ist, orientiert sich die folgende Beschreibung an der genannten Publikation.

#### Active Shape Modell

In einem ersten Schritt wird das Formmodell erstellt. Die Basis dafür bildet ein Trainingsdatensatz mit  $N$  Bildern mit je  $v$  annotierten Gesichtsländern. Die Landmarken werden üblicherweise händisch annotiert und befinden sich in auffälligen Gesichtspartien wie Mund, Nase, Augen, Augenbrauen und Gesichtskontur. Die Form  $s$  ist dann über die Positionen

der Gesichtslandmarken im Bild definiert und kann mathematisch wie folgt ausgedrückt werden:

$$\mathbf{s} = (x_i, y_i), i = 1, \dots, v \quad (2.21)$$

Die Form  $\mathbf{s}$  besitzt somit die Dimension  $\mathbf{n} = 2v$ .

Zur Darstellung eines AAM wird ein trianguliertes Netz (*engl.*: Mesh) genutzt (siehe Abbildung 2.8). Die Einteilung des Gesichts in Dreiecksbereiche ermöglicht das Mappen von Gesichtstexturinformation (z.B. Grauwerte) innerhalb der Bereiche von einer Gesichtsform auf eine andere. Dies erfolgt mittels „Image Warping“ [98], was mathematisch einer affinen Transformation entspricht (vgl. Abschnitt 2.4.2).

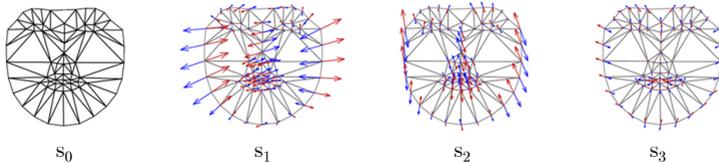


Bild 2.8: Darstellung der Basisform als Mesh, sowie lineare Formvariationen (Quelle: [37])

Per Definition kann eine Form  $\mathbf{s}$  durch eine Basisform  $\mathbf{s}_0$  und eine Linearkombination von  $n$  Formvektoren  $\mathbf{s}_j$  ausgedrückt werden:

$$\mathbf{s} = \mathbf{s}_0 + \sum_{j=1}^n p_j \mathbf{s}_j \quad (2.22)$$

mit  $p_j$  als Formparameter (Koeffizienten) der Formvektoren  $\mathbf{s}_j$ . Die Bestimmung der Basisform  $\mathbf{s}_0$  für einen Trainingsdatensatz erfolgt mittels Procrustes-Analyse. Im ersten Schritt wird der Schwerpunkt einer jeden Forminstanz subtrahiert, das neue Zentrum ist der Ursprung  $(0,0)$ . Im nächsten Schritt wird eine beliebige Forminstanz so skaliert, dass gilt  $|\mathbf{s}| = 1$ . Die anderen Forminstanzen werden zu dieser ersten Schätzung ausgerichtet (skaliert und rotiert). Durch Mittelung entsteht die erste Schätzung

der Basisform (*engl.* Base Shape) als normalisierte Durchschnittsform aller Trainingsdaten. Auf Basis der erhaltenen Schätzung wird erneut ausgerichtet und gemittelt, bis der Abstand der Basisform von den Trainingsformen möglichst gering ist. Es ergibt sich ein Minimierungsproblem der Art:

$$\min_{a,b} \sum \left\| \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} - \begin{bmatrix} x_{c,i} \\ y_{c,i} \end{bmatrix} \right\|^2, \quad (2.23)$$

mit  $x_{c,i}, y_{c,i}$  als  $i$ -te Landmarkskoordinaten der „Canonical Shape“. Die Transformationsparameter  $a, b$  werden durch Least-Squares-Schätzung bestimmt.

Die Bestimmung der Formvektoren  $s_i$  erfolgt mit Hilfe der PCA. Ziel ist die Beschreibung der Streuung der Landmarken entlang der Hauptachsen, was eine deutlich kompaktere Beschreibung der Formvarianzen ermöglicht (vgl. Abschnitt 2.4.1). Dazu wird die Kovarianzmatrix für die ausgerichteten Formen aufgestellt. Durch Lösen des Eigenwertproblems ergeben sich die Eigenwerte und zugehörigen Eigenvektoren. Die ersten  $n$  Eigenvektoren entsprechen dem Satz an Formvektoren  $s_i$  in Gleichung 6.1. Meist wird die PCA für auf die Basisform ausgerichtete Formen nach der Procrustes Analyse durchgeführt. Somit sind die Formvariationen, welche durch globale, rigide Formtransformationen wie Translation, Rotation oder Skalierung entstehen, bereits entfernt. Die Eigenvektoren der PCA beschreiben somit lokale, nicht rigide Transformationen, welche im Folgenden mit  $s_{non-rigid}$  bezeichnet werden. Nicht-rigide Transformationen entstehen beispielsweise durch mimische Ausdrücke oder Sprechbewegungen. Rigide Transformationen entstehen hingegen durch starre Bewegungen des Gesichts wie Kopfbewegungen. Sie lassen sich über Translation, Rotation oder Skalie-

zung abbilden (siehe Abbildung 2.9), was einer Ähnlichkeitstransformation entspricht. Dies kann in kompakter Form ausgedrückt werden:

$$\begin{bmatrix} a+1 & -b \\ b & a+1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}, \quad (2.24)$$

mit  $a, b$  als Skalierungs- und Rotationsparameter und  $t_x, t_y$  als Translationsparameter. Für sämtliche Landmarken  $[x_i, y_i]$  der Form  $\mathbf{s}$  gilt somit:

$$\begin{bmatrix} \begin{bmatrix} a+1 & -b \\ b & a+1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \\ \vdots \\ \begin{bmatrix} a+1 & -b \\ b & a+1 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \end{bmatrix} = \begin{bmatrix} x_1 & -y_1 & 1 & 0 \\ y_1 & x_1 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & -y_n & 1 & 0 \\ y_n & x_n & 0 & 1 \end{bmatrix} \begin{bmatrix} a+1 \\ b \\ t_x \\ t_y \end{bmatrix} \quad (2.25)$$

Wie zu sehen ist, lässt sich die Ähnlichkeitstransformation für eine Form  $\mathbf{s}$  als Linearkombination von 4 Basisvektoren darstellen. Dabei beschreiben die ersten beiden Vektoren die Skalierungs- und Rotationsparameter  $a, b$ . Sie setzen sich aus den Koordinaten der Landmarken zusammen. Die letzten beiden Vektoren entsprechen der Translation und bestehen aus binären Elementen. Die Formvektoren werden als  $\mathbf{s}_{rigid}$  bezeichnet und spannen den Ähnlichkeitsraum für  $\mathbf{s}$  auf. Gleichzeitig bilden die Eigenvektoren der PCA einen Deformationsraum. Für ausgerichtete Daten steht dieser senkrecht zum Ähnlichkeitsraum, wodurch  $\mathbf{s}_{rigid}$  und  $\mathbf{s}_{non-rigid}$  eine Orthonormalbasis bilden. Die Formparameter  $\mathbf{s}_i$  in Gleichung 6.1 setzen sich dann aus 4 Vektoren  $\mathbf{s}_{rigid}$  und  $n$  Eigenvektoren  $\mathbf{s}_{non-rigid}$  zusammen. So

kann jede Gesichtsform mit Hilfe der Formparameter beschrieben werden. Ist die Gesichtsform gegeben können die Formparameter über

$$p_j = \mathbf{s}_j^T (\mathbf{s} - \mathbf{s}_0), j = 1, \dots, n \quad (2.26)$$

bestimmt werden.

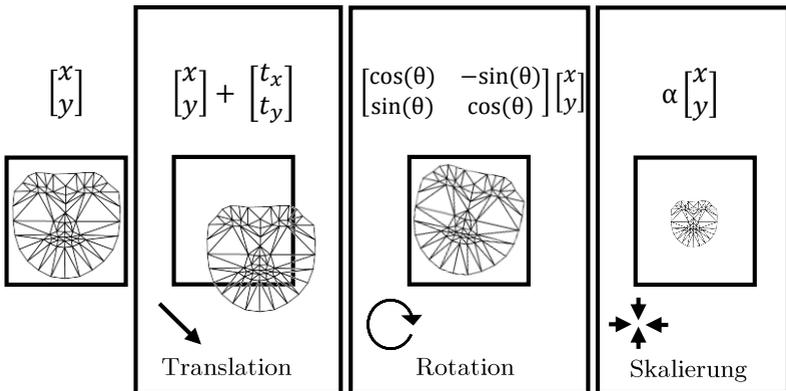


Bild 2.9: Ähnlichkeitstransformation der Face Shape

## Erweiterung um Texturinformation

Die Erweiterung des Modells um Texturen erfolgt, wie bereits im letzten Abschnitt erwähnt, mittels Image Warping. Dabei wird die Texturinformation der Trainingsbilder (z.B. Grauwerte), durch die entsprechende Verformung des Gesichts im Trainingsbild, auf das triangulierte Netz der Basisform (*engl.*: Base Mesh)  $\mathbf{s}_0$  gemappt. Zunächst wird davon ausgegangen, dass alle Pixel in einem Dreiecksbereich die gleichen Warping-Parameter besitzen. Die Zuordnung des Pixels  $(x^0, y^0)$  von der Basisform zu einer For-

minstanz (Gesichtsform in einem Beispielbild) kann dementsprechend wie folgt formuliert werden:

$$\begin{bmatrix} x^0 & y^0 & 1 \end{bmatrix} \begin{bmatrix} p_1 & p_2 \\ p_3 & p_4 \\ p_5 & p_6 \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix}, \quad (2.27)$$

mit  $p_i, i = 1, \dots, 6$  als Warping-Parameter und  $(x, y)$  als Koordinaten der Pixel der Forminstanz (siehe Abbildung 2.10). Für die Ecken der Dreiecke

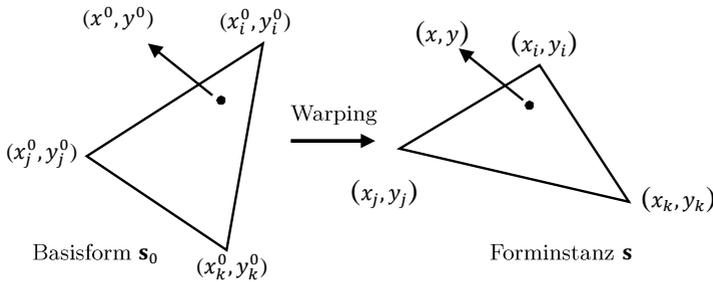


Bild 2.10: Warping von der Basisform in die Forminstanz

ergibt sich:

$$\begin{bmatrix} x_i^0 & y_i^0 & 1 \\ x_j^0 & y_j^0 & 1 \\ x_k^0 & y_k^0 & 1 \end{bmatrix} \begin{bmatrix} p_1 & p_2 \\ p_3 & p_4 \\ p_5 & p_6 \end{bmatrix} = \begin{bmatrix} x_i & y_i \\ x_j & y_j \\ x_k & y_k \end{bmatrix} \quad (2.28)$$

mit  $(x_i^0, y_i^0)$ ,  $(x_j^0, y_j^0)$ ,  $(x_k^0, y_k^0)$  als Koordinaten der Ecken des Basisform-Dreiecks  $s_0$  und  $(x_i, y_i)$ ,  $(x_j, y_j)$ ,  $(x_k, y_k)$  für die Ecken der Instanz  $s$  im Trai-

ningsbild. Die Eck-Koordinaten sind durch die Label gegeben, sodass die Warping-Parameter über folgende Gleichung bestimmt werden können:

$$\begin{bmatrix} p_1 & p_2 \\ p_3 & p_4 \\ p_5 & p_6 \end{bmatrix} = \begin{bmatrix} x_i^0 & y_i^0 & 1 \\ x_j^0 & y_j^0 & 1 \\ x_k^0 & y_k^0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x_i & y_i \\ x_j & y_j \\ x_k & y_k \end{bmatrix} \quad (2.29)$$

Einsetzen von 2.29 in 2.27 ergibt:

$$\begin{bmatrix} x & y \end{bmatrix} = [\alpha \ \beta \ \gamma] \begin{bmatrix} x_i & y_i \\ x_j & y_j \\ x_k & y_k \end{bmatrix}, \quad (2.30)$$

mit

$$[\alpha \ \beta \ \gamma] = \begin{bmatrix} x^0 & y^0 & 1 \end{bmatrix} \begin{bmatrix} x_i^0 & y_i^0 & 1 \\ x_j^0 & y_j^0 & 1 \\ x_k^0 & y_k^0 & 1 \end{bmatrix}^{-1} \quad (2.31)$$

Es ist zu erkennen, dass die Koeffizienten  $\alpha$ ,  $\beta$ ,  $\gamma$  lediglich von den Koordinaten der Ecken in der Basisform abhängen, wodurch die Koeffizienten in der Trainingsphase berechnet werden können. Für ein Pixel der Basisform  $(x_0, y_0)$  können die Koordinaten  $(x, y)$  im Beispielbild  $I$  über Gleichung 2.30 bestimmt werden. Der Grauwert eines Pixels im Base-Mesh wird dann über  $\mathbf{A}(x_0, y_0) = \mathbf{I}(x, y)$  bestimmt.

Nach der Erweiterung der des Modells zum Active Appearance Modell wird erneut eine PCA angewendet:

$$\mathbf{A} = \mathbf{A}_0 + \sum_{i=1}^m \lambda_i \mathbf{A}_i, \quad (2.32)$$

mit  $\lambda_i, i = 1, \dots, m$  als Appearance-Parameter. Kongruent zum Formmodell, kann die Appearance  $A(x)$  als Appearance der Basisform  $\mathbf{A}_0$  und einer Linearkombination von Appearance-Vektoren  $\mathbf{A}_i$  ausgedrückt werden. Abbildung 2.11 zeigt die Darstellung der Basis-Appearance und dessen Kombination mit den ersten drei Appearance-Komponenten.

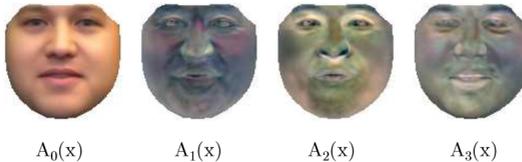


Bild 2.11: Darstellung der Basis-Appearance und Variationen (Quelle: [37])

## Modell Instantiierung

Abbildung 2.12 veranschaulicht die Rekonstruktion eines Gesichtsbildes auf Basis der Modellparameter. Im ersten Schritt wird die Appearance des Gesichts in der Basisform  $\mathbf{s}_0$  mit Hilfe von Gleichung 2.32 bestimmt. Dann wird die Form  $\mathbf{s}$  des zu rekonstruierenden Gesichts über die Formparameter  $p_j$  und Gleichung 2.22 berechnet. Um im nächsten Schritt die Intensität bzw. den Grauwert eines Pixels  $x$  in  $\mathbf{s}$  zu bestimmen, wird dessen korrespondierende Position  $\mathbf{W}(x; \mathbf{p})$  in  $\mathbf{s}_0$  mittels der Warping-Gleichung 2.30 berechnet, wobei  $\mathbf{p}$  den mit Gleichung 2.31 bestimmten Warping-Parametern entspricht.  $M$  entspricht dem rekonstruierten Gesicht in der Basisform.

Die Idee des Fittings des Modells an ein unbekanntes Bild beruht darauf, den besten Satz an Form- und Appearance-Parametern zu finden, sodass

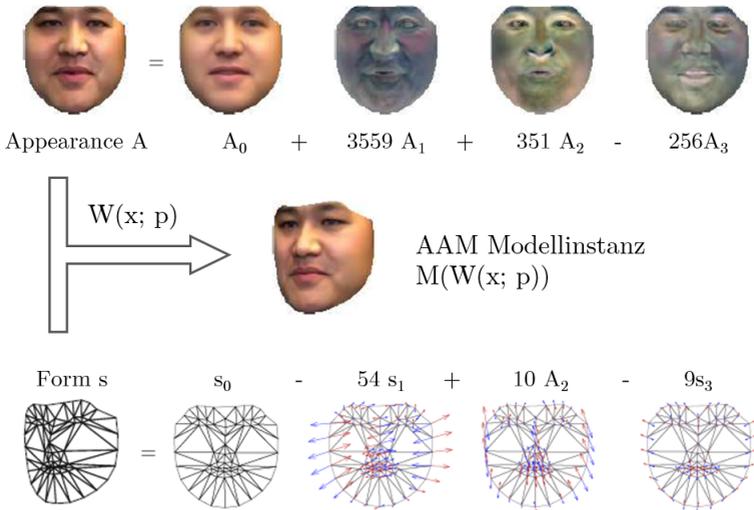


Bild 2.12: Modellinstanz im AAM (adaptiert von [37])

der Unterschied zwischen der Rekonstruktion des Gesichts und dem gegebenen Bild minimal wird. Aktuell verwendete Methoden werden in Kapitel 3.2 diskutiert.

### 2.4.3 Signale und Systeme

Für eine detaillierte Einführung in die Signalverarbeitungstheorie wird auf [69] verwiesen. Auf dieser Basis wird eine kurze Einführung in zentrale Operationen und Transformationen gegeben.

## Fast-Fourier-Transformation

Die Fourier-Transformation ist ein auf der Fourier-Reihe basierendes mathematisches Werkzeug, welches besagt, dass eine stetige Funktion durch eine Reihe von Sinus- und Kosinusfunktionen ausgedrückt werden kann. Sie sagt auch aus, dass ein Zeitsignal als Superposition von Sinus- und Kosinusschwingungen unterschiedlicher Frequenz und Phase dargestellt werden kann.

Die kontinuierliche Fourier-Transformation für ein Zeitsignal  $x(t)$  ist gegeben durch

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{j2\pi ft} dt. \quad (2.33)$$

Die diskrete Fourier-Transformation für abgetastete Werte im zeitlichen Abstand  $T$  ist wie folgt definiert:

$$X_*(f) = \sum_{n=-\infty}^{\infty} x(nT)e^{j2\pi fnT}. \quad (2.34)$$

Um das Frequenzspektrum zu erhalten, also die Amplituden der Schwingungen aus denen sich das Signal zusammensetzt, wird der Betrag der komplexen Fourier-Transformierten  $|X_*(f)|$  gebildet.

## Faltung

Ein in der Signalverarbeitung ebenfalls häufig eingesetzter Operator ist die Faltung. Sie ist für zwei Signale  $x(t)$  und  $y(t)$  definiert als [69]

$$(x * y)(t) = \int x(\tau)y(t - \tau)d\tau \quad (2.35)$$

und im diskreten Fall als

$$(x * y)(n) = \sum_k x(k)y(n - k) \quad (2.36)$$

Die Faltung kann mittels der Fourier-Transformation  $F(\cdot)$  ebenfalls im Frequenzbereich durchgeführt werden :

$$(x * y)(t) = (F^{-1}(F(x) \cdot F(y)))(t). \quad (2.37)$$

Die Faltung zweier Zeitsignale entspricht der Multiplikation im Frequenzbereich.

### Analog-Digital-Converter

Um eine analoge Messgröße in eine digitale zu überführen, benötigt man zunächst einen Sensor, der die analoge Größe elektrisch erfasst. Diese elektrische Größe (typischerweise Spannung) bewegt sich in einem zuvor festgelegten Arbeitsbereich von  $[U_l, U_h]$ , respektive der niedrigsten und höchsten erlaubten Spannung. Der Wert innerhalb dieses Intervalls soll als Information kodiert werden und wird zu diesem Zweck diskretisiert. Ein Analog-Digital-Converter (ADC) überführt die elektrische Spannung dabei in eine festgelegte Anzahl von Bits. Typischerweise tastet ein ADC das Eingangssignal in regelmäßigen Zeitabständen ab, was als Abtastrate oder Abtastfrequenz bezeichnet wird.

### Abtasttheorem

Ein abgetastetes Signal kann zeitdiskret dargestellt werden, indem man es mit einem Impulszug multipliziert [69]:

$$s_a(t) = s(t) \cdot \sum_{n=-\infty}^{\infty} \delta(t - nT) \quad (2.38)$$

Im Frequenzbereich entspricht dies:

$$S_a(f) = S(f) * \left( \frac{1}{T} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T}\right) \right) \quad (2.39)$$

Das Signal ist dann vollständig rekonstruierbar, wenn gilt:

$$f_A > 2f_{max} \tag{2.40}$$

Die Abtastfrequenz  $f_A = \frac{1}{T}$  muss demnach doppelt so groß sein, wie die höchste im Signal vorkommende Frequenz. Andernfalls tritt sogenanntes Aliasing auf. Außerhalb des Basisbands (0Hz bis  $\frac{f_A}{2}$ ) liegende Frequenzen werden auf das Basisband projiziert und stören das Nutzsignal. Um dies zu vermeiden, wird das abzutastende Signal zuvor mittels eines analogen Tiefpassfilters vorverarbeitet, sodass Frequenzen oberhalb von  $\frac{f_A}{2}$  stark gedämpft werden.  $\frac{f_A}{2}$  ist nach H. Nyquist's Abtasttheorem auch als Nyquist-Frequenz bekannt.

### Digitale Filter

Wie in Gl. 2.37 beschrieben, entspricht eine Faltung im Zeitbereich einer Multiplikation im Frequenzbereich. Dieser Sachverhalt kann für die Umsetzung eines Frequenzfilters genutzt werden. Für ein gegebenes Signal  $g[n]$ , dessen Frequenzspektrum für alle gewünschten Frequenzanteile 1 und für alle unerwünschten 0 sein soll, kann ein Signal  $y[n]$  mittels Faltung von  $g[n]$  mit einem Signal  $x[n]$  gewonnen werden. Ein digitales Filter kann demnach mittels Faltung implementiert werden [69]:

$$y[n] = \sum x[k] \cdot g[n-k] \tag{2.41}$$

Ein Filter dieser Form wird auch FIR-Filter (Finite Impulse Response) genannt. Der Name bezieht sich auf den Umstand, dass die Impulsantwort des Filters endlich ist. Es existiert auch eine Form von digitalem Filter, des-

sen Impulsantwort unter Umständen nicht abklingt: das IIR-Filter (Infinite Impulse Response):

$$\sum a[k] \cdot y[n-k] = \sum b[k] \cdot x[n-k]$$

$$y[n] = \frac{1}{a[0]} \left( \sum_{k=0} b[k]x[n-k] - \sum_{k=1} a[k]y[n-k] \right) \quad (2.42)$$

Der Unterschied zum FIR-Filter besteht darin, dass die Ausgabe  $y[n]$  von früheren Ausgaben  $y[n-k]$  abhängt. Ein weiterer Unterschied ist, dass FIR-Filter mit linearer Phase entworfen werden können, während dies im Allgemeinen für IIR-Filter nicht möglich ist. Das ist wichtig, da linearphasige Filter die Wellenform des Signals beibehalten, während es im nichtlinearen Fall zu Verzerrungen des zu filternden Signals kommen kann.

### Segmentierte Faltung

Die segmentierte Faltung (*engl.*: Overlap Add (OLA)) ist ein Verfahren zur schnellen Faltung und wird in der digitalen Signalverarbeitung eingesetzt. Dabei wird die diskrete Eingangsfolge eines beliebig langen Signals  $x[n]$  durch Faltung mit der Impulsantwort  $h[n]$  in einander nicht überlappende Teilfolgen zerlegt:

$$y[n] = x[n] * h[n] \stackrel{\text{def}}{=} \sum_{m=-\infty}^{\infty} h[m] \cdot x[n-m] = \sum_{m=1}^M h[m] \cdot x[n-m], \quad (2.43)$$

mit  $h[m] = 0$ , wenn  $m$  außerhalb  $[1, M]$  liegt [65].

Die Idee besteht darin, die Problemstellung in mehrere Faltungen von  $h[n]$  mit kleinen Segmenten von  $x[n]$  zu zerlegen:

$$x_k[n] \stackrel{\text{def}}{=} \begin{cases} x[n+kL] & n = 1, 2, \dots, L \\ 0 & \text{sonst,} \end{cases} \quad (2.44)$$

mit  $L$  als zu wählende Segmentlänge. Dann können  $x[n] = \sum_k x_k[n - kL]$ , und  $y[n]$  als Summe kurzer Faltungen beschrieben werden:

$$y[n] = \left( \sum_k x_k[n - kL] \right) * h[n] = \sum_k (x_k[n - kL] * h[n]) \quad (2.45)$$

$$= \sum_k y_k[n - kL], \quad (2.46)$$

mit  $y_k[n] \stackrel{\text{def}}{=} x_k[n] * h[n] = 0$  außerhalb  $[1, L + M - 1]$ .

Der Vorteil der segmentierten Faltung ist, dass die zirkulare Faltungsoperation zur Bildung der Teilfolgen  $y_k$  effizient über die FFT oder die inverse FFT (IFFT) ausgedrückt werden kann:

$$y_k[n] = \text{IFFT}(\text{FFT}(x_k[n]) \cdot \text{FFT}(h[n])) \quad (2.47)$$

In den Anwendungen wird die segmentierte Faltung u.a. zur effizienten Implementierung von FIR-Filtern höherer Ordnung verwendet, da sie in digitalen Signalprozessoren (DSP) ressourcensparend umgesetzt werden kann.

## Kreuzkorrelation

Die Kreuzkorrelation ist ein Maß für den Zusammenhang zweier unterschiedlicher Variablen [69]. Auf die Analyse von zwei diskreten Energiesignalen  $x[k]$  und  $y[k]$  bezogen, ist sie definiert als:

$$R_{xy}[n] = \sum_{k=-\infty}^{\infty} x^*[k] y[k + n] \quad (2.48)$$

Ein Anwendungsfall für die Kreuzkorrelation ist die Ermittlung der zeitlichen Verschiebung zweier stark korrelierender Signale.

# 3

## Stand der Wissenschaft und Technik

**Z**IEL DIESER ARBEIT ist die Entwicklung eines PPGI-Systems zur robusten, berührungslosen Remote-Messung der Herzfrequenz in realen Anwendungsszenarien. Durch die ambiente Integration der Kamera in der Messumgebung ist dabei kein am Körper zu tragender Sensor mehr notwendig. Dadurch entfällt bei Messungen die Notwendigkeit des An- und Ablegens von Sensoren, sowie negative Begleiterscheinung beim Tragen von Sensoren (z.B. allergische Hautreaktionen). In der Folge ergeben sich vielfältige Einsatzmöglichkeiten eines derartigen Systems, insbesondere für kontinuierliche Langzeitmessungen. In der Literatur findet man im Zusammenhang mit ambienter Vitalsensorik auch den Begriff der „Un-

obtrusive Measurements“, also unaufdringlicher Messmethoden. Das folgende Unterkapitel gibt zunächst einen Überblick über berührungslose Methoden zur Messung kardiovaskulärer Parameter, bevor in Abschnitt 3.3 der Stand der Forschung und Technik für PPGI beschrieben wird.

### 3.1 Kontaktlose Messung von Vitalparametern

Eine Möglichkeit zur berührungslosen Messung der bioelektrischen Herzaktion liefert das kapazitive EKG. Im Vergleich zum herkömmlichen EKG (vgl. Abschnitt 2.2.1) besitzen kapazitive EKG-Elektroden eine dielektrische Oberfläche. Dadurch wird zwischen Sensor und Körper des Probanden ein Isolator eingebracht, wodurch eine kapazitive Kopplung entsteht. Damit wird der zuvor nötige ohmsche Kontakt überflüssig, was gleichzeitig den Einsatz von Trockenelektroden ohne leitfähige Elektrolyte ermöglicht. Dadurch ist das Verfahren insbesondere für das Langzeit-Monitoring oder die Messung durch Kleidung von Interesse [48]. Die erste Umsetzung eines ambienten EKG-Systems wurde von Ishijima vorgestellt [59]. Der Nachteil ambienter EKG-Systeme besteht darin, dass die maximale Distanz zwischen Körper und Sensor für eine Ableitung mit ausreichender Qualität i.d.R. im Millimeterbereich liegt. Somit handelt es sich streng genommen um einen quasi-kontaktbasierten Ansatz, wobei bereits die Art der Kleidung einen hohen Einfluss auf die Signalqualität haben kann. Für detailliertere Information zur Messtechnik und Umsetzung ambienter EKG-Systeme wird auf [48] verwiesen.

Neben der elektrischen Erregung der Herzkontraktion lassen sich auch die mechanischen Effekte, die beim Pumpen des Blutes durch das Gefäßsystem entstehen, messen. Diese Effekte äußern sich in einer messbaren Bewegung der Körperoberfläche. Als berührungslose Sensortechnologien deren Messprinzip darauf beruht, sind die Ballistokardiographie [56], Radar- und Laser-basierte Verfahren [136][115][102][91], sowie die videobasierte Analyse der Bewegungen (*engl.*: Video Motion Analysis (VMA)) [9][161]

zu nennen. Im Gegensatz zu PPGI werden bei der VMA nicht die absorptionsbedingten Helligkeitsänderungen der Haut, sondern kleinste Positionsänderungen der Hautoberfläche in Bildsequenzen bestimmt. Ein Vorteil dieser Messmethoden besteht darin, dass sie neben der Herzaktivität auch respiratorische Informationen extrahieren können. Wie bereits in den Abschnitten 2.2.2 und 2.2.3 des Grundlagenkapitels beschrieben, lässt sich die Pulsation des Blutes dank der Lichtabsorptionseigenschaften des Blutfarbstoffes Hämoglobin mittels PPG und PPGI auch optisch erfassen.

Ein weiterer messbarer Effekt liegt in der sich ändernden elektrischen Impedanz durch den Blutfluss und die kardiorespiratorische Aktivität in Form von Organbewegungen. Dies lässt sich mittels Magnetischer Induktionsmessung (MI) auch berührungslos ohne angelegte Elektroden, wie bei der Impedanzplethysmographie (IPG), erfassen [31]. Für einen detaillierten Überblick und eine genaue Beschreibung der einzelnen Verfahren wird auf [20] und [85] verwiesen. Die folgende Tabelle zur Bewertung der einzelnen Verfahren wurde ebenfalls [20] und [85] entnommen und in Teilen adaptiert.

Hinsichtlich der Messinformation ist zunächst hervorzuheben, dass das kapazitive EKG ein medizinisch sehr gut erforschtes Signal liefert. Dem entgegen steht die bereits erwähnte erforderliche maximale Distanz zum Messobjekt im Millimeterbereich. Bei der Messgenauigkeit der Ansätze, welche die Blutpulsation erfassen, ergeben sich bei bewegungsbasierten Verfahren messprinzipbedingte Vorteile bei sich ändernden Umgebungsbedingungen (z.B. Helligkeitsänderungen), wohingegen die optischen Verfahren robuster gegenüber Bewegungsartefakten sind und somit als geeigneter für Langzeitmonitoringzwecke eingeordnet werden können. Eine generelle, vergleichende Bewertung erscheint jedoch schwierig, da die Signalqualität stets stark von der Anwendung und der Qualität der jeweils eingesetzten Sensor-Hardware abhängt.

Tabelle 3.1: Übersicht über kontaktlose Verfahren zur Messung der Herzfrequenz

Verfahren	Kontakt	Messgröße	eingebrachte Energie	Distanz	Positionsensitiv	Kosten	Anzahl Messobj.	TRL
kap. EKG	kapazitiv	el. Biopotential	nein	mm	+	0	1	+
BCG	mechanisch	Verschiebung	nein	0	?	+	1	+
Radar	elektromagn.	Verschiebung	ja	m	?	0	>1	-
Laser	optisch	Verschiebung	ja	m	-	-	1	-
VMA	optisch	Verschiebung	nein	m	?	+	>1	0
PPG	optisch	Lichtabsorption	ja	mm	0	+	1	+
PPGI	optisch	Lichtabsorption	ja	m	?	+	>1	0
MI	elektromagn.	el. Bioimpedanz	ja	cm	0	0	1	?

Ein Vorteil von PPGI ist, dass über die Kamera prinzipiell Vitaldaten mehrerer Personen gemessen werden können. Für die Anwendung im Automobil ergibt sich zudem der Vorteil, dass unter Verwendung zukünftig verbauter Fahrerbeobachtungs- oder Innenraumkameras gegebenenfalls kein zusätzlicher Sensor für das Vitaldatenmonitoring verbaut werden müsste.

## 3.2 Fitting Methoden für das 2D Face Alignment

Bevor in Abschnitt 3.3 auf den State of the Art von PPGI eingegangen wird, folgt an dieser Stelle ein Überblick über mögliche Fitting-Methoden für das in Kapitel 2.4.2 eingeführte AAM und daraus abgeleitete Ansätze, da die Entwicklung eines Face Alignment Moduls ein wichtiger Bestandteil dieser Arbeit ist.

### 3.2.1 Fitting von AAMs

Active Appearance Modelle haben eine über 20-jährige Forschungshistorie bei der Erkennung und Verfolgung nicht-rigider Objekte in Bildern im Computer Vision Bereich. Nach Kapitel 2.4.2 beschreiben sie eine Objektklasse anhand eines kleinen Satzes an Shape- und Appearance-Parametern. Das Prinzip des Fittings eines AAM an ein neues Bild besteht darin, den besten Satz von Shape- und Appearance-Parametern zu finden, sodass der Unterschied zwischen der rekonstruierten Appearance und dem gegebenen Bild minimal wird:

$$\min_{\mathbf{p}, \lambda} \left\| \mathbf{d}(\mathbf{W}(\mathbf{x}; \mathbf{p})) - \mathbf{A}_0 + \sum_{i=1}^m \lambda_i \mathbf{A}_i \right\|^2 \quad (3.1)$$

Die Appearance-Differenz ( $\mathbf{d}(\mathbf{W}(\mathbf{x}; \mathbf{p})) - \mathbf{A}_0 + \sum_{i=1}^m \lambda_i \mathbf{A}_i$ ) wird als Fehlerbild (*engl.*: Error Image) bezeichnet. Das Auflösen nach Shape-Parameter  $\mathbf{p}$  ergibt zunächst ein nicht-lineares Optimierungsproblem. Um dieses zu

Linearisieren wird eine initiale Schätzung der Landmarks-Positionen benötigt, was in der Regel mit Hilfe eines Face Detectors erfolgt [153]. Wie in Abbildung 3.1 zu sehen ist, stellt der Face Detector die Gesichtsp position in Form einer Bounding Box (blaues Rechteck) bereit. Um die initiale Schätzung der Landmarks-Positionen zu generieren, wird die Basisform in die Mitte der Box gelegt. In jeder Iteration des AAM Fittingprozesses werden anschließend die aktuellen Modellparameter geschätzt. Typischerweise entspricht das Parameterupdate einer Funktion des Fehlerbilds.

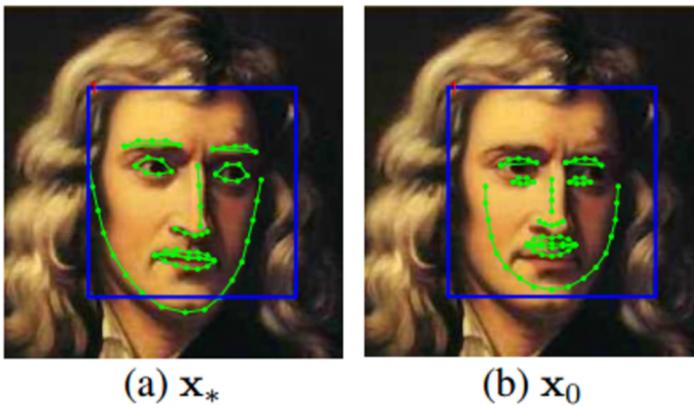


Bild 3.1: Initiale Prädiktion der Landmarken mittels Face Detection: (a) Ground Truth Landmarken (b) Initiale Prädiktion, *Quelle: [162]*

Die erste Anwendung von Gradientenabstiegsverfahren für das Alignment von Bildern wurde in [30] beschrieben. Dabei ist das AAM-Fitting als Lukas-Kanade-Problem [88] formuliert und kann iterativ mittels Gauss-Newton-Optimierung gelöst werden. Da Hesse- und Jacobi-Matrix im Testbild und somit bei jeder Iteration neu berechnet werden, gilt der Ansatz als äußerst rechenintensiv. In [93] wird daher vorgeschlagen die Rollen des rekonstruierten Appearance-Bildes und des Testbildes im Algorithmus zu vertauschen. Das Shape-Parameter-Inkrement wird somit in gegensätzlicher bzw. „inverser“ Richtung geschätzt. Zudem wird die Appearance-

Variation aus der Beschreibung des Differenzbildes ausgelagert, wodurch die Jacobi-Matrix einmalig in einem Vorverarbeitungsschritt bestimmt werden kann. Im Vergleich zu den Lukas-Kanade Standardansätzen sind die sogenannten „Project-Out Inverse Compositional (POIC)“ Algorithmen somit deutlich effizienter, können jedoch weniger gut mit „ungesehenen“ Objekten umgehen, die nicht Teil der Trainingsdaten sind. In [8] wird mit der Simultaneously Inverse Compositional (SIC)-Methode ein ähnlicher Ansatz präsentiert. Da dabei die Appearance-Parameter jedoch in jeder Iteration bestimmt werden, können Hesse- und Jacobi-Matrix nicht wie beim POIC-Ansatz in einem Vorverarbeitungsschritt berechnet werden. Um die Effizienz der Berechnung zu steigern, wird angenommen, dass die Appearance des Gesichts über die Iterationen unverändert bleibt, wodurch diese nur in erster Iteration berechnet werden muss.

Die zweite Klasse an AAM-basierten Face Alignment Verfahren basiert auf dem Lernen einer Mapping-Funktion zwischen Fehlerbild und Modellparametern. In der originalen AAM-Publikation [30] wird die lineare Regression zwischen dem Modellparameterinkrement und dem Fehlerbild erlernt. Die Regressionsbeziehung entspricht einer Näherung an den steilsten Abstieg („steepest descent“) in [93] und wird als linear und unabhängig von den aktuellen Modellparametern angenommen. Saragih geht in [122] auf die Nachteile eines festen, linearen Modells ein, in dem er dessen Einsatz in realen AAM-Fitting-Szenarien simuliert. In [123] führt Saragih daraufhin ein nichtlineares Update-Modell ein, welches die POIC-Methode [93] und die lineare Regression [30] hinsichtlich Genauigkeit und Konvergenz übertrifft. Tresadern [145] setzt Boosting als Regressionsverfahren für lineare und nichtlineare Modelle ein. Es wird gezeigt, dass sich nichtlineare Modelle besser für Vorhersagen in größeren Entfernungen eignen, wohingegen lineare Regressoren Vorteile bei kleineren Distanzen besitzen. Folglich wird ein hybrides Modell aus linearen und nichtlinearen Regressoren zur Lösung des Fitting-Problems vorgeschlagen.

### 3.2.2 Constrained Local Model

Eine weitere Möglichkeit zur Lösung des Face Alignment Problems bieten „Constrained Local Models (CLMs)“ [32][124][33]. Im Gegensatz zu AAMs, bei denen die ganzheitliche Appearance von Gesichtern zur Beschreibung der Texturinformation genutzt wird, werden bei CLMs einzelne Templates für die jeweiligen Bereiche um die Gesichtslanmarken (*engl.*: „Patches“) erzeugt. Während des Fittingprozesses werden die Templates gesampelt und in das Modell projiziert. Auf diese Weise wird für die Landmarken eine Response Map berechnet, welche die Kostenfunktion für die Pixel innerhalb der Patches darstellt. Typischerweise wird die Response Map durch normierte Korrelation des Sample-Patch’s mit dem Modell-Template erzeugt. Das Formmodell wird anschließend verwendet, um die optimale Kombination zu finden.

### 3.2.3 Supervised Descent Method

Xiong et al. präsentierten in [162] mit der „Supervised Descent Method (SDM)“ einen weiteren, modernen Face Alignment Ansatz. Wie in Abbildung 3.2 dargestellt, wird die Landmarkenverschiebung direkt durch Multiplikation des Regressionsmodells  $R_k$  mit den extrahierten Merkmalen bestimmt, ohne dass Hesse- und Jacobi-Matrix wie bei den Ansätzen auf Basis des Newton-Verfahrens [93] berücksichtigt werden müssen. Die SDM ist vergleichbar mit dem in [30] vorgeschlagenen Verfahren, bei dem nur eine lineare Regression zwischen dem Modellparameterinkrement und dem Error Image erlernt wird. Der zweite Unterschied zwischen der SDM und der Methode nach [30] besteht darin, dass die SDM SIFT-Features verwendet, um eine robustere Beschreibung der Textur gegenüber Beleuchtungsvariationen zu erhalten. Das in dieser Arbeit entwickelte Face Alignment Modell basiert auf der Idee der SDM und wird in Kapitel 6.1 beschrieben.

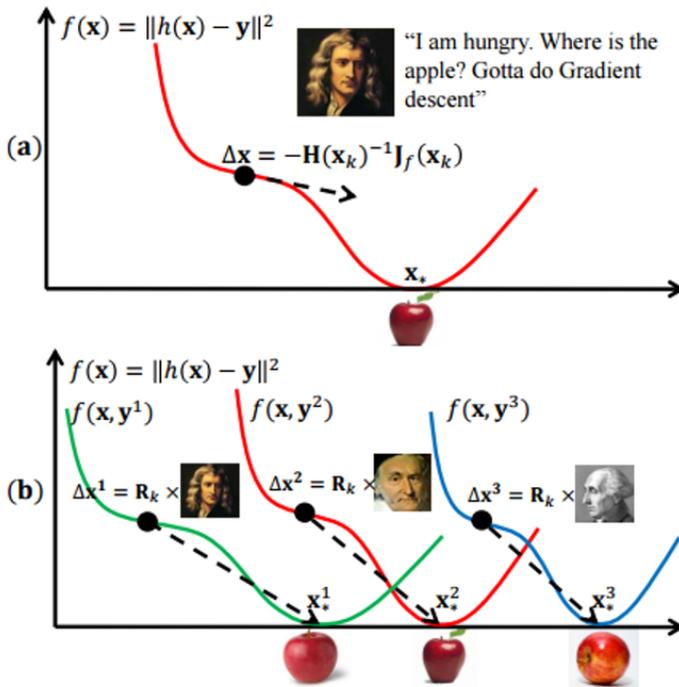


Bild 3.2: Vergleich zwischen (a) Gauss-Newton-Methode and (b) SDM , *Quelle:* [162]

### 3.3 Stand der Wissenschaft bei PPGI

Nachfolgend wird der Stand der Wissenschaft von PPGI unterteilt nach Aspekten wie dem messtechnischen Aufbau oder den eingesetzten Signalverarbeitungsalgorithmen beschrieben. Zaunseder et al. liefern in [163] einen sehr detaillierten Überblick über den State of the Art, an dem sich die folgende Beschreibung in weiten Teilen orientiert. Am Ende eines Abschnitts wird jeweils kurz skizziert, in welcher Form die Arbeit einen wissenschaftlichen Beitrag zu dem jeweiligen Teilaspekt zu liefern versucht.

### 3.3.1 Messtechnischer Aufbau

In der Literatur am häufigsten zu finden ist der passive Systemansatz, bei dem das Umgebungslicht als Lichtquelle und eine RGB-Kamera als Sensor genutzt wird. Der Ansatz bietet zunächst den Vorteil des einfachen Messaufbaus [109][86][95]. Zudem wird der Messeffekt, welcher im grünen Wellenlängenbereich am stärksten ist (siehe Absorptionsmaxima von Hämoglobin in 2.2.2), ideal ausgenutzt. In 2008 haben Verkruyssen et al. [151] gezeigt, dass es bei Aufnahmen mit herkömmlichen Digitalkameras bei Tageslicht möglich ist, sowohl die Herz- als auch die Atmungsrate im Frequenzspektrum zu identifizieren. Neben der Grundfrequenz seien die Oberschwingungen bis hin zur vierten Harmonischen erkennbar. Lewandowska et al. [86] konnten in 2011 nachweisen, dass der Messeffekt durch die geschickte Kombination der Farbkanäle schon mit einfachen Webcams ausreichend gut erfasst werden kann. Auf Grund der hohen Verfügbarkeit, sowie der geringen Kosten werden sie heute in zahlreichen Publikationen genutzt. Nichtsdestotrotz wurde der positive Effekt einer höheren Farbtiefe als die für Webcams typischen 8-Bit in mehreren Publikationen nachgewiesen [22][84]. Da der Einfluss der Auflösung auf die Signalqualität, auf Grund der Mittelung der Pixel innerhalb der ROI, als weniger stark einzustufen ist, variieren die verwendeten Auflösungen stark (z.B. 320x240 in [73], 1920x1200 in [108]). Die Abtastfrequenzen liegen typischerweise bei 30 frames per second (fps) oder darunter. Für bestimmte Zwecke, insbesondere zur besseren Auflösung zeitlicher Merkmale, wurden jedoch auch höhere Abtastfrequenzen bis zu 420 fps genutzt [3][140][148].

Um die Signalqualität zu steigern, kann die Messregion mit Licht grüner Wellenlänge aktiv beleuchtet werden. Hülsbusch [50] hat dazu ein System bestehend aus 4 LED-Modulen vorgestellt und die Wichtigkeit der möglichst homogenen Ausleuchtung herausgestellt. Für viele Anwendungen wie beispielsweise Driver Monitoring, ist eine aktive Beleuchtung im sichtbaren Spektralbereich nicht zweckmäßig, weshalb im nahinfraroten Spek-

trum beleuchtet wird [44]. Dies bietet den Vorteil des möglichen System-einsatzes bei Tag und Nacht und der Reduktion des Einflusses des Sonnenlichts bei entsprechender optischer Filterung. Demgegenüber steht jedoch der deutlich geringer ausgeprägte Messeffekt im Nahinfrarot (NIR)-Bereich (ca. Faktor 50). Hinsichtlich spezieller Optiken wurde beispielsweise der Einsatz von Polarisationsfiltern untersucht. Während Hülsbusch [50] keinen positiven Effekt nachweisen konnte, zeigten Trumpp et al. [146], dass orthogonale Polarisationsfilter zu einem höheren SNR führen können.

*Zur vergleichenden Analyse des passiven und aktiven Ansatzes, sowie dem Einfluss unterschiedlicher Beleuchtungs-Wellenlängen, wird in Kapitel 4.2 ein im Rahmen der Arbeit entwickeltes Beleuchtungsmodul zur selektiven Auswahl verschiedener Wellenlängen für die aktive Beleuchtung vorgestellt. Jedoch wird in dieser Arbeit das Ziel eines einfachen und kostengünstigen Hardwareaufbaus angestrebt, um die Einsatzchancen des algorithmischen Systems für eine Vielzahl von Applikation zu maximieren, indem lediglich eine zusätzliche Softwarekomponente integriert werden müsste. Daher steht der passive Ansatz mit RGB-Kameras ohne aktive Beleuchtung im Fokus der Untersuchungen.*

### 3.3.2 Datenbasis

Ein Großteil der in Forschungsarbeiten verwendeten Datensätze wurde zum Ziel der Algorithmenentwicklung aufgezeichnet. Die Aufnahme erfolgt daher meist unter Laborbedingungen. Dabei befinden sich die Probanden häufig in Ruhe (z.B. [86]), teilweise werden bestimmte Arten von Bewegungen betrachtet [109][58][97]. Oft fehlt den Daten die Diversität hinsichtlich des Alters und der Hautfarbe der Probanden. Insbesondere aber fehlt es an Untersuchungen, die nicht-stationäre Bedingungen z.B. durch sich ändernde Lichtverhältnisse betrachten oder unter realen Umweltbedingungen aufgezeichnet wurden. Die Zahl der öffentlich zugänglichen Datenbanken für PPGI beschränkt sich bisher auf einige wenige.

Zu nennen sind die MAHNOB-HCI- [134], die UBFC-RPPG- [16] und die DEAP-Datenbank [72]. Weitere öffentliche Datenbanken sind unbedingt erforderlich, um einen objektiven Vergleich unterschiedlicher algorithmischer Ansätze zu ermöglichen.

*Mit dem Ziel der Algorithmenentwicklung für den herausfordernden Einsatz unter dem Einfluss häufiger und teils starker Licht- und Bewegungsartefakte, wird zunächst eine Datenbank aufgezeichnet, welche in 10 verschiedenen Szenarien unterschiedliche Arten von Beleuchtung, Bewegung und Lichtintensitätsänderungen untersucht (Kapitel 4). Bei der Aufzeichnung soll durch die Betrachtung technischer (z.B. geeignete Synchronisation mit Referenzdaten) und datenschutzrechtlicher Rahmenbedingungen (z.B. Einholen DSGVO-konformer Einwilligungserklärungen), die Voraussetzungen für eine Veröffentlichung der Datenbank geschaffen werden.*

*Für die Untersuchungen unter Realbedingungen werden im Rahmen dieser Arbeit die Anwendungsszenarien „Belastungsmonitoring am (Büro-) Arbeitsplatz“ und „Fahrerbeobachtung im Automobil“ untersucht. Ziel der Aufnahmen ist es, neben einer entsprechenden Probandenzahl, durch eine adäquate Länge der Videos (>30 Minuten) Aussagen über die Stabilität des Systems bei größeren Messdauern treffen zu können.*

### 3.3.3 Detektion und Tracking der Messregion

Mit Blick auf die für Messungen verwendete ROI gibt es lediglich vereinzelte Untersuchungen an Armen [148][55] oder Handinnenflächen [131][77] [66]. Die am häufigsten verwendete Messregion ist das Gesicht, da es in der Regel frei zugänglich ist. Meistens werden die gut durchbluteten Bereiche der Stirn und Wange [74][84][79] oder das komplette Gesicht als Messregion [52] [109] verwendet.

Zur Detektion des Gesichts wird in der Regel die bekannte Viola-Jones-Methode [153] genutzt. Die Bereiche innerhalb des Gesichts werden an-

schließlich manuell oder automatisch, z.B. auf Basis von geometrischen Verhältnissen innerhalb der Bounding Box oder in Bezug auf die Augenpositionen [86], bestimmt. McDuff et al. [95] nutzen zum Ausschluss der Augenregion, mit dem Ziel Störeinflüsse durch Blinzeln zu vermeiden, den in [92] vorgestellten Facial-Landmark-Detector. Alternativ oder zusätzlich können Segmentierungsverfahren eingesetzt werden. Um das Gesicht im Bild von Hintergrund und Oberkörper zu trennen, setzen Tarassenko et al. [141] beispielsweise eine nichtparametrische Bayes'sche Bildsegmentierung ein. Sahindrakar et al. [120] nutzen den HSV-Farbraum für eine Hautpixel-Segmentierung, bei der mittels empirisch ermitteltem Schwellwert eine Wahrscheinlichkeitsmatrix für Hautpixel im Bild bestimmt wird.

Bei Bewegung der Probanden ist neben der Detektion des Gesichts ein Tracking-Algorithmus für das Verfolgen von Bild zu Bild erforderlich. Häufig wird dazu die im vorigen Unterkapitel eingeführte Methode des Feature-Tracking nach Kanade-Lucas-Tomasi [88][144] genutzt, etwa in [74][58][41]. Der Einsatz komplexerer Modelle für die Aufgabe des ROI Tracking ist eher selten. Beispiele liefern Stricker et al. [137], die den Fitting-Ansatz für deformierbare Modelle nach Saragih et al. [124] nutzen. Bousefsaf et al. [18] nutzen die in [6] eingeführte Methode. Gleichzeitig wurde gezeigt, dass es durch den Einsatz von Landmark-Tracking zu Jittering kommen kann, was sich negativ auf das SNR auswirkt [22]. In diesem Fall sind wiederum spezielle Methoden zur Reduktion solcher Einflüsse erforderlich [147].

*Aus den bereits genannten Gründen wird in dieser Arbeit ebenfalls das Gesicht als Messregion genutzt. In Kapitel 6 wird in diesem Zusammenhang zunächst eine Weiterentwicklung der in Abschnitt 3.2.3 eingeführten SDM, als alternative Methode für das Face Alignment bei PPGI eingeführt. Auf Basis des Modells und der extrahierten Tracking-Landmarken wird in der Folge ein Ansatz zur Online-Adaption definierter Pixelberei-*

*che innerhalb des Gesichts, mit Hilfe eines Index für die Haut-Konfidenz und die Hochfrequenz-Rauschenergie, vorgestellt. Beide Verfahren sind mit dem Ziel verbunden, die Robustheit gegenüber häufigen Bewegungsartefakten, wie sie beispielsweise beim Fahrer während der Fahrt im Auto vorkommen, zu steigern.*

### 3.3.4 Signalverarbeitungsalgorithmen

In Abschnitt 3.3.1 wurde bereits beschrieben, dass auf Grund des stark ausgeprägten Messeffekts im grünen Wellenlängenbereich meist RGB-Kameras als Sensoren für PPGI eingesetzt werden. Während die ersten Systemansätze Algorithmen nutzten, um die menschliche Herzfrequenz aus Bildsequenzen einzelner Farbkanäle zu extrahieren [151], wurden in der Folge insbesondere Ansätze mit dem Ziel entwickelt, durch geschickte Kombination der Farbkanäle eine bessere Signalqualität zu erreichen. In 2011 stellten Lewandowska et al. [86] und Poh et al. [109] Verfahren zur Störkompensation durch Source Separation auf Basis der Principal Component Analysis (PCA) bzw. der Independent Component Analysis (ICA) vor. De Haan et al. [34] präsentierten 2013 einen alternativen Ansatz, bei welchem die Farbinformationen in den Chrominanz-Farbraum transformiert wird, um so die Robustheit gegenüber Bewegungen zu verbessern. In 2014 beschrieben de Haan et al. [35] ein weiteres Verfahren, dass das Pulssignal durch Projektion in den RGB-Raum mit Hilfe eines empirisch ermittelten Blutvolumenpulsvektors (BVP-Vektor) bestimmt. Im selben Jahr stellten Tarassenko et al. [141] einen mathematischen Ansatz für die Beschreibung der blutvolumenpulsinduzierten Helligkeitsschwankungen mit Hilfe eines autoregressiven Modells vor. Die dominante Frequenz kann dabei als Pol der Übertragungsfunktion dargestellt werden. Wang et al. präsentierten 2015 den datengetriebenen 2SR-Ansatz [159] und führten 2017 den Plane orthogonal to Skin (POS)-Algorithmus ein [157], welcher den von de Haan eingeführten BVP-Vektor zur RGB-Projektion mit Hilfe von

getroffenen Annahmen über die optischen Eigenschaften der Hautreflexionen bestimmt. Qi et al. [113] stellten 2017 eine Weiterentwicklung des ICA-basierten Ansatz von Poh vor, wobei Signale verschiedener ROIs über einen sogenannten Joint-Blind-Source-Separation-Ansatz (JBSS) fusioniert werden. Im Jahr 2018 führten Chen und McDuff mit „DeepPhys“ den ersten Ansatz auf Basis neuronaler Netze ein [25].

Neben den wissensbasierten Ansätzen zur Farbkanalkombination und den eher datengetriebenen Verfahren zur Source Separation, werden weitere Techniken der Signalverarbeitung eingesetzt, mit dem Ziel das SNR des extrahierten Signals zu erhöhen. In erster Linie handelt es sich dabei um den Einsatz von Detrending-Methoden oder Bandpassfiltern. Die Cut-Off-Frequenzen der Bandpassfilter orientieren sich dabei am physiologischen Bereich der Herzfrequenz (z.B. 40 – 240 Schläge pro Minute in [34]). Während in zahlreichen Arbeiten Filter mit festen Grenzen eingesetzt werden, wurden erste adaptive Filter mit variablen Cut-Off-Frequenzen eingeführt. Huang und Dung [53] verwenden beispielsweise einen klassischen Ansatz adaptiver Filter auf Basis eines Least-Mean-Square (LMS)-Schätzers, wobei die Gesichtsposition zur Bildung des Differenzsignals genutzt wird, um auf diese Weise die Robustheit gegenüber Bewegungen zu steigern. Weitere adaptive Filtermethoden wurden in [159] und [40] beschrieben. Bei Einkanalanalysen (z.B. NIR-Ansatz) kann die Empirical Mode Decomposition (EMD) [26] oder ein Kalman Filter [61][111] zur Störkompensation eingesetzt werden.

Die Bestimmung der Herzfrequenz auf Basis des extrahierten Signals erfolgt entweder im Zeitbereich (z.B. [54][140][4]) oder im Frequenzbereich nach entsprechender Fouriertransformation (z.B. [109][137][35]). Der Vorteil der Analyse des Signals im Zeitbereich liegt in der Möglichkeit der Schlag-zu-Schlag-Detektion der Herzaktivität, was beispielsweise Voraussetzung für die Bestimmung der HRV ist [14]. Um den Verlauf der Herzfrequenz zu analysieren werden gleitende Zeitfenster eingesetzt. Die Grö-

ße der Zeitfenster variieren mit dem verwendeten algorithmischen Ansatz und der Applikation (z.B. 60s in [74] oder 12s in [34]).

*Die Entwicklung innovativer Signalverarbeitungsalgorithmen für eine erfolgreiche Signalrekonstruktion unter dem Einfluss mittlerer bis starker Artefakte ist ein zentraler Bestandteil dieser Arbeit. In Kapitel 7 wird daher zunächst die Weiterentwicklung der durch de Haan [34] und Wang [157] eingeführten Signalprojektion im RGB-Farbraum vorgestellt. Dabei liegt der Fokus auf der Anpassung der Projektionsrichtung für bestimmte Arten und Intensitäten von Kopfbewegungen.*

*Zudem wird eine in [15] eingeführte Methode zur adaptiven Bandpassfilterung auf Basis einer Spektrogramm-Analyse als Alternative zu [53] und [40] vorgestellt. Dabei wird das extrahierte BVP-Signal mit engen und variablen Filtergrenzen gefiltert, mit dem Ziel durch das artefaktbereinigte Signal die nachfolgende Peak-Detektion für die Schlag-zu-Schlag-Analyse im Zeitbereich zu optimieren.*

### 3.3.5 Ableitung weiterer Vitalparameter

Wie bereits in Kapitel 2.2.2 beschrieben, bietet der optische Messeffekt verbunden mit der unterschiedlichen Absorptionsstärke von sauerstoffgesättigtem und ungesättigtem Hämoglobin theoretisch die Möglichkeit die Sauerstoffsättigung mittels PPGI zu erfassen. Dafür bedarf es einer Methode die ROI mit selektiven Lichtwellenlängen aktiv zu beleuchten. Humphreys et al. [55] beleuchten den Unterarm mit Licht der Wellenlängen  $760\text{nm}$  und  $880\text{nm}$  und setzten eine Monochromkamera als Sensor ein. In Kapitel 4.2 wird ein im Rahmen der Arbeit entwickeltes Beleuchtungsmodul für die PC-gesteuerte, selektive Auswahl von fünf Wellenlängen im sichtbaren und Nahinfrarotbereich vorgestellt.

Die Alternative zu alternierender, selektiver Beleuchtung besteht in dem Einsatz von schmalbandigen optischen Filtern. Fan und Li [39] nutzen zu diesem Zweck ein motorisiertes Filterrad mit Bandpassbereichen bei  $520\text{nm}$

und  $660\text{nm}$  und gleichzeitiger, aktiver Beleuchtung mit weißem Licht. Verkruysse et al. [150] verwendet zwei Kameras, eine mit optischem Filter bei  $675\text{nm}$  und eine bei  $842\text{nm}$ , was jedoch eine exakte Synchronisation der Kamerasysteme voraussetzt. Verkruysse et al. gehen dabei auch auf die Herausforderung der Kallibration solcher Systeme ein. Insbesondere die schwer zu reproduzierenden Messbedingungen bei einer Remote-Messung auf Grund sich ändernder Abstände und Winkel zwischen Proband, Beleuchtung und Sensor oder veränderter Umgebungsbeleuchtung stellen hier eine Herausforderung dar.

Neben der Sauerstoffsättigung als weiteren Vitalparameter, gibt es mehrere Publikationen die das Ziel verfolgen, auf Basis der Erfassung der PTT eine Korrelation mit dem Blutdruck nachzuweisen. Ein häufig verwendeter Messansatz liegt in der Nutzung der Handinnenfläche und des Gesichts als ROI bei Aufzeichnung von Videos mit ein oder zwei Kameras [131] [60]. Teilweise wird auch ein kombinierter Ansatz aus PPGI und EKG [67] oder PPGI und PPG [164] untersucht. Für den rein PPGI-basierten Ansatz konnte bisher keine signifikante Korrelation zwischen PTT-basiertem Blutdruck und dessen Referenz nachgewiesen werden [138]. Auf die speziellen Herausforderungen der PTT-Messung nur über das Gesicht wird in [100] eingegangen.

Zusätzlich zur Herzfrequenz hat sich vor allem die HRV als quantitatives Maß für die Aktivität des vegetativen Nervensystems und der daraus abgeleiteten Beurteilung des Stresslevels und der Erholungsfähigkeit des Körpers etabliert. Daher ist die HRV auch Gegenstand zahlreicher Publikationen im PPGI-Kontext [97][95][54][2]. Die Aufnahmen erfolgen dabei meist in Laborumgebung und unter Ruhebedingungen bzw. bei Ausführung zuvor definierter Bewegungen. Während die Publikationen die prinzipielle Eignung des PPGI-Verfahrens zur Bestimmung der HRV bestätigen, ergeben sich insbesondere durch Bewegungen [58] oder sich ändernde Umgebungsbedingungen [42] Unterschiede zur Referenz. McDuff hat daher in [97] für die Messung der PRV einen komplexen Hardware-Aufbau be-

stehend aus 9 Kameras verwendet, um eine möglichst große Anzahl nutzbarer Frames bei rotatorischer Bewegung zu generieren.

*Um zu untersuchen, inwiefern sich die mittels PPGI bestimmte PRV als mögliche Alternative für die EKG-basierte HRV eignet (u.a. gefordert in [126]), wird in Kapitel 8.2 eine Studie mit einer Aufnahmedauer von 40 – 60 Minuten je Messung vorgestellt. Die Aufnahmen erfolgen dabei im Setting „Büroarbeitsplatz“ mit zufälligen Bewegungen der Probanden ohne spezielle Vorgaben.*

### 3.3.6 Einsatz von PPGI im Automobil

Während in Kapitel 1.1 bereits die zahlreichen Anwendungsgebiete von PPGI erwähnt wurden, wird hinsichtlich des State of the Arts der Einsatz zur Fahrerbeobachtung im Automobil näher beschrieben, da diese Anwendung im Fokus der Arbeit steht. In 2015 veröffentlichten Kuo et al. [75] eine Studie bei der zehn Probanden bestimmte Fahraufgaben durchführten. Bei vier der zehn Probanden konnte eine aus Sicht der Autoren ausreichende Qualität der Herzfrequenzschätzung (innerhalb +/- 5 Schläge zur Referenz in 48 – 75% der Fahrt) nachgewiesen werden. Zhang et al. [165] veröffentlichten in 2017 Ergebnisse eines Webcam-basierten Ansatzes zur Messung der Herzfrequenz, wobei die Pilotmessung mit einem einzigen Probanden durchgeführt wurden. Ebenfalls in 2017 veröffentlichten Blöcher et al. [14] einen RGB-basierten Ansatz auf Basis der ICA. Bei dem zeitbasierten Ansatz wurde eine Sensitivität der Peak-Detektion > 90% erreicht, allerdings mindert die geringe Anzahl an Probanden (2) die Aussagekraft der Ergebnisse.

*Mit den in Kapitel 6 und 7 beschriebenen Algorithmen wird u.a. das Ziel eines robusten Ansatzes zum Einsatz von PPGI im Automobil verfolgt. Bei den Auswertungen wurde speziell auf eine signifikante Messdauer*

*der Fahrten und Anzahl der Probanden geachtet, um fundierte Aussagen über die Eignung während der Fahrt treffen zu können.*

### 3.4 Ausgangslage und Abgrenzung

Durch das Remote-Messprinzip von PPGI eröffnen sich zahlreiche Anwendungsgebiete. Diese Vielfalt verbunden mit immer kostengünstigeren Kamerasensoren, sowie der Möglichkeit der on-Chip-Verarbeitung, haben das Interesse am Messansatz in den letzten Jahren stark gesteigert, was sich an der deutlich gestiegenen Zahl an Publikationen quantifizieren lässt [118]. Während zahlreiche Arbeiten darauf fokussiert waren, die grundlegende Messtechnik [50][94] oder Algorithmik [109] [34][141][157] zu untersuchen, steht in dieser Arbeit die Weiterentwicklung und Ergänzung bestehender Prinzipien, für den Einsatz in hinsichtlich des Einflusses von Artefakten herausfordernden Realweltszenarien wie der Fahrerbeobachtung im Automobil, im Vordergrund. Der wissenschaftliche Beitrag der Arbeit wurde themenbezogen am Ende der Teilabschnitte des vorigen Unterkapitels dargestellt.

Mit Blick auf die Ausgangslage am Institut verfolgt die Arbeit die Weiterentwicklung bestehender Ideen und Konzepte zur unaufdringlichen Vitalparametermessung im Automobil, welche beispielsweise im Rahmen der Dissertation von Heuer [48] auf Basis der kapazitiven EKG-Messung umgesetzt und untersucht wurden. Im Bereich von PPGI bestanden keine Vorerfahrungen.



# 4

## Grundlegende Datenerhebung und Analyse

**I**M FOLGENDEN KAPITEL wird eine Datenbank für PPGI beschrieben, welche auf der Analyse des Standes der Wissenschaft und Technik aufbaut und die Basis für den Vergleich algorithmischer Ansätze und einer datengetriebenen Systementwicklung darstellt. Mit dem Ziel des Systemeinsatzes in herausfordernden Umgebungen werden spezielle Szenarien zur Simulation des Einflusses von unterschiedlichen Beleuchtungsarten, sowie Licht- und Bewegungsartefakten entworfen und aufgezeichnet. Neben der Beschreibung des Versuchsaufbaus und den konzipierten Szenarien werden auch die Entwicklung des eingesetzten Beleuchtungsmoduls, sowie Methoden zur exakten Synchronisation von Videoaufnahmen und Refe-

renzsystem beschrieben. Für die aufgezeichneten Daten erfolgt zudem eine Analyse hinsichtlich der Signalqualität in Bezug auf die unterschiedlichen Szenarien und verschiedenen Unterregionen innerhalb der Messregion Gesicht.

### 4.1 Anforderungen an die Datenbasis

Das Ziel der Erstellung der Datenbank ist die Schaffung einer umfangreichen Datenbasis, auf deren Grundlage die Entwicklung der Bild- und Signalverarbeitungsalgorithmen des PPGI-Systems (Kapitel 5-7) erfolgt. Dies ist auch durch den fehlenden Zugang zu öffentlichen Datenbanken motiviert, welche neben Videodaten des Gesichts parallel aufgezeichnete Referenzvitaldaten aus EKG- oder PPG-Systemen enthalten. Vor diesem Hintergrund wird die Veröffentlichung der Datenbank angestrebt, um anderen Wissenschaftlern einen objektiven Vergleich entwickelter, algorithmischer Ansätze zu ermöglichen. Aus technischer Sicht liegt der Fokus in der Analyse des Einflusses von Licht- und Bewegungsartefakten auf die Signalqualität anhand von Aufnahmen in verschiedenen Szenarien. Durch den Einsatz unterschiedlicher Beleuchtungen und Kameratypen soll zudem eine Datenbasis für Entwurf und Analyse von passiven und aktiven Systemansätzen geschaffen werden. Um die Einflüsse von Kompressionsverfahren zu vermeiden, werden die Daten im Rohformat aufgezeichnet. Nachfolgend werden die definierten Anforderungen an die Datenbank aufgeführt:

- [D-AF-01] Die Datenbank soll verschiedene Szenarien hinsichtlich des Einflusses von Licht- und Bewegungsartefakten enthalten.
- [D-AF-02] Für jedes Szenario soll nur ein Einflussfaktor geändert werden.
- [D-AF-03] Die Mindestlänge einer Datenaufzeichnung beträgt zwei Minuten pro Szenario und Proband.

- [D-AF-04] Es bedarf einer signifikanten Anzahl an Probanden (Anzahl > 25) mit ausreichender Diversität hinsichtlich Alter, Geschlecht und Ethnie.
- [D-AF-05] Es bedarf eines Beleuchtungsmoduls für Aufnahmen im Sinne eines aktiven Systems in unterschiedlichen Lichtwellenlängenbereichen.
- [D-AF-06] Die Aufzeichnung eines Videos sollte parallel mit einer RGB-Farbkamera und einer NIR-Monochromkamera erfolgen
- [D-AF-07] Die Referenzaufnahme der Herzfrequenz sollte mit einem zertifizierten Medizinprodukt durchgeführt werden.
- [D-AF-08] Es bedarf einer Referenzaufnahme der Einflussfaktoren Licht und Kopfbewegung.
- [D-AF-09] Es bedarf einer möglichst exakten Synchronisation der Video- und Referenzdaten.

Darüber hinaus gilt es die datenschutzrechtlichen Voraussetzung für die Veröffentlichung der Datenbank zu schaffen. Der auf Basis der Anforderungen entwickelte Versuchsaufbau wird in Abschnitt 4.3, der Entwurf der Szenarien in Abschnitt 4.4 beschrieben. Im nächsten Unterkapitel wird zunächst auf die Entwicklung eines Moduls zur aktiven Beleuchtung eingegangen.

## 4.2 Entwicklung eines Beleuchtungsmoduls für aktive Messungen

Das Ziel der aktiven Messungen im Rahmen der Arbeit besteht darin, den Einfluss verschiedener Wellenlängen im sichtbaren und nahinfraroten Bereich hinsichtlich der Signalqualität quantifizieren zu können. Für die Durchführung aktiver Messungen [D-AF-04] bedarf es zunächst der Entwicklung eines geeigneten Beleuchtungsmoduls für Aufnahmen bei Illumination mit unterschiedlichen Wellenlängen. Ähnlich wie in [50] und [55] nutzt das entwickelte Beleuchtungsmodul ein LED-Array mit LEDs verschiedener Wellenlänge. Leichte Unterschiede bestehen in der Wahl der Wellenlängen, wobei in diesem Fall LEDs der Wellenlänge grün (525 nm), rot (655 nm) und infrarot (780 nm, 850 nm, 940 nm) eingesetzt werden. Die grüne Wellenlänge liegt im Bereich des Absorptionsmaximums von Hämoglobin und entspricht der idealen Beleuchtung hinsichtlich des Messeffekts. Die Beleuchtung im nicht sichtbaren Nahinfrarotbereich ist interessant, da sie die Beleuchtung bei Dunkelheit ohne Blendwirkung ermöglicht. Die Kombination mit der roten Wellenlänge wird ebenfalls für die kontaktbasierte Messung der Sauerstoffsättigung bei der Pulsoximetrie genutzt.

Die LEDs einer Wellenlänge sind in Form einer 6x6-Matrix auf dem LED-Modul mit einem jeweiligen Abstand von  $d = 25 \text{ mm}$  angebracht. Insgesamt besteht das LED-Modul somit aus 36 LED-5er-Gruppen. Beim gewählten LED-Öffnungswinkel von je  $\varphi = 20^\circ$  ergibt sich bei einem Abstand zum Messobjekt von  $r=0,5 \text{ m}$  nach Gleichung 4.1 eine beleuchtete Kreisfläche von  $900 \text{ cm}^2$ . Es gilt:

$$D_A = (n - 1) \cdot d + 2r \tan \frac{\varphi}{2} = 30 \text{ cm} \quad (4.1)$$

Die Bündelung der Lichtstrahlen in der Mitte des Lichtkegels führt zu einem leichten Abfall der Beleuchtungsintensität nach außen hin. Auf Grund der Größe der beleuchteten Fläche im Vergleich zur Gesichtsgröße ist die

Ausleuchtung des Gesichts ausreichend homogen. Bei Bedarf wird eine Diffusorscheibe zur weiteren Homogenisierung des Lichts eingesetzt. Abbildung 4.1 zeigt ein bestücktes LED-Modul.



Bild 4.1: Bestücktes LED-Beleuchtungsmodul mit 5 Wellenlängen in 6x6-Matrix-Anordnung

### 4.2.1 Ansteuerung der LEDs

Für die Beleuchtung mit Leuchtdioden bedarf es einer geeigneten Ansteuerung. LEDs müssen aufgrund ihrer physikalischen Eigenschaften in ihrem Vorwärtsstrom beschränkt werden, um sie vor Schäden zu schützen. Weitere Anforderungen liegen in der Langzeitkonstanz der Helligkeit der Dioden, sowie der Möglichkeit diese für eine Wellenlänge separat anpassen zu können. Nach der Analyse verschiedener, schaltungstechnischer Realisierungskonzepte und vorgefertigter LED-Treiberbausteine kommt beim finalen Entwurf der LED-Treiber *AL3065* von *Diodes Incorporated* zum Einsatz. Dieser integrierte Schaltkreis (*engl.*: „integrated circuit (IC)“) verbindet eine einstellbare Stromquelle mit einem DC/DC-Aufwärtswandler. Dadurch lässt sich mit einer 10 V-Versorgungsspannung die für den jeweili-

gen LED-Typ passende Spannung erzeugen, wodurch eine energieeffiziente Ansteuerung realisiert wird. Die Schaltfrequenz des Steuerbausteins lässt sich mittels Widerstand für den Bereich 100 kHz bis 1 MHz linear einstellen. Durch die Wahl einer größeren Frequenz kann die Schaltung, insbesondere die Induktivität, verkleinert werden, gleichzeitig steigen die Umschaltverluste. Als Kompromiss wird eine Frequenz von 400 kHz gewählt. Der Treiberbaustein besitzt vier Stromsenken (Kanäle), welche jeweils einen Maximalstrom von 400 mA führen können. Pro Kanal werden jeweils neun LEDs in Reihe angeschlossen. Insgesamt werden die 36 LEDs einer Wellenlänge mit einem LED-Treiber angesteuert. Zur Einstellung der Helligkeit wird mittels Potentiometer eine Referenzspannung zwischen 0V und 1,7V angelegt. So lässt sich der Strom linear von 0 mA bis  $I_{max}$  einstellen. Abbildung 4.4 zeigt das Anschlussdiagramm des LED-Treibers.

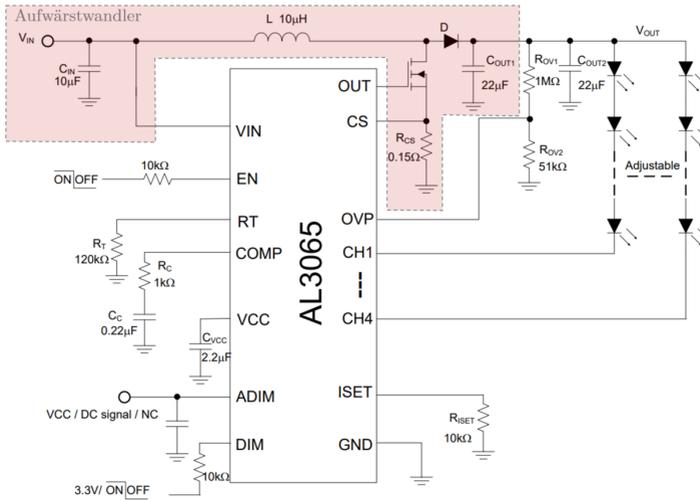


Bild 4.2: Anschlussdiagramm des LED-Treibers aus dem Datenblatt

Die Grundschaltung eines Aufwärtswandlers bestehend aus einer Induktivität, einem elektrischen Schalter (in der Regel ein MOSFET oder Transistor), einer Diode und einem Kondensator, ist in Abbildung 4.4 rot mar-

kiert. Im eingeschalteten Zustand des Schalters beginnt durch die Spule ein Strom zu fließen. Dieser ist nach

$$u_L(t) = L \cdot \frac{d}{dt} i_L(t) \longleftrightarrow i_L(t) = \frac{1}{L} \cdot \int u_L(t) dt \quad (4.2)$$

abhängig von der Größe der Induktivität  $L$ , der Eingangsspannung  $V_{IN}$  und der Zeitdauer  $T_{EIN}$ . Wird der Schalter geöffnet, versucht die Spule aufgrund ihrer physikalischen Eigenschaften den Strom weiterzutreiben. Da dieser nicht mehr über den Schalter nach Ground (GND) fließen kann, steigt die Spannung an der Induktivität, bis sie ein höheres Potential als die Ladespannung des Ausgangskondensators erreicht hat. Nun kommutiert der Strom auf die Diode und lädt den Kondensator auf. Da aus Sicht des Ausgangskondensators die Diode in Sperrrichtung betrieben wird, kann dieser sich nicht entladen und die Spannung steigt bei jedem Schaltzyklus weiter an. Theoretisch kann mit idealen Bauelementen auf diese Weise eine unendlich hohe Spannung am Ausgang erzeugt werden. Die Ausgangsspannung ergibt sich aus der Eingangsspannung und dem Tastverhältnis  $d$ . Die Spulenstromänderung während der Einschaltzeit (4.3) und während der Ausschaltzeit (4.4) ist gegeben durch:

$$\Delta I_L = U_E \cdot \frac{T_{EIN}}{L} \quad (4.3)$$

$$-\Delta I_L = (U_E - U_A) \cdot \frac{T - T_{EIN}}{L} \quad (4.4)$$

Da sich diese Werte bei nicht-lückendem Betrieb betraglich entsprechen müssen, ergibt sich die Ausgangsspannung zu:

$$U_A = \frac{U_E}{1-d} \longleftrightarrow d = 1 - \frac{U_E}{U_A} \quad (4.5)$$

Im lückenden Betrieb erreicht der Spulenstrom seine Anfangsbedingung bevor der Schaltzyklus abgeschlossen ist und kann in der verbleibenden

Zeit keine weitere Energie mehr übertragen. Diese Betriebsart sollte durch entsprechende Regelung vermieden werden.

Für die Dimensionierung der benötigten Bauteile muss zunächst der maximale und der minimale Tastgrad festgelegt werden. Die benötigte Ausgangsspannung für den Treiber der roten LEDs beträgt laut Datenblatt bei einer Flussspannung von 1,5 V und einem Strom von  $I = 30 \text{ mA}$  etwa  $U_{Fluss} = 15 \text{ V}$ . Auf Grund etwaiger Schwankungen wird mit maximal  $U_A = 15 \text{ V}$  gerechnet. Durch Gl. (4.5) ergibt sich ein Tastgrad von  $d_{min} = 0,26$  und  $d_{max} = 0,33$ . Der maximale Stromrippel folgt aus:

$$\Delta I_L = U_E \cdot \frac{T_{EIN}}{L} = U_E \cdot \frac{d}{f \cdot L} \quad (4.6)$$

Für die Induktivität gilt somit:

$$L \geq U_E \cdot \frac{d_{max}}{f \cdot \Delta I} \quad (4.7)$$

Der maximale Rippel wird in der Regel auf 30 % bis 40 % festgelegt [46], womit sich eine Induktivität von  $L \approx 200 \mu\text{H}$  ergibt. Zusätzlich sollte der ohmsche Widerstand der Induktivität möglichst niedrig gehalten werden. Bei der Dimensionierung der Diode sollte die Durchbruchspannung einen niedrigen Wert haben, wofür sich Schottkydioden gut eignen. Die Sperrspannung sollte über der maximalen Ausgangsspannung des Aufwärtswandlers liegen, da es andernfalls zum Kurzschluss des Kondensators im Einschaltmoment kommen könnte. Der Ausgangskondensator muss mindestens für die maximale Ausgangsspannung ausgelegt werden. Der Spannungsrippel der Ausgangsspannung ergibt sich dabei aus der Kapazität des Kondensators und der Schaltfrequenz. Bei einer Schaltfrequenz von 400 kHz und der verwendeten Kapazität von  $44 \mu\text{F}$  ist der Rippel mit wenigen Millivolt vernachlässigbar klein. Abbildung 4.3 zeigt die fertig bestückte Platine zur Ansteuerung der LED-Matrix, welche über ein Flachbandkabel mit dem LED-Modul verbunden wird.

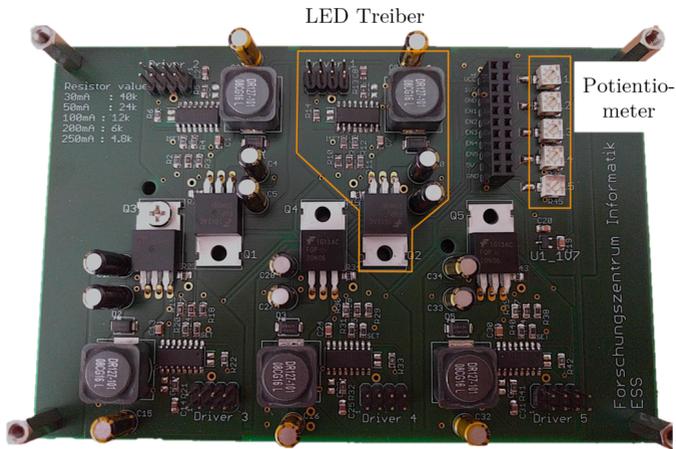


Bild 4.3: Platine zur Ansteuerung der Beleuchtungsmatrix mit insgesamt fünf Treiber-Modulen zur Ansteuerung von jeweils einem LED-Typ

### 4.2.2 Ansteuerung der Beleuchtung

Der in Abb. 4.4 gezeigte finale Aufbau des Beleuchtungsmoduls besteht aus dem LED-Modul, der unterhalb montierten Steuerplatine und einem Microcontroller. Die Stromversorgung von Mikrocontroller und Beleuchtung erfolgt über ein 10V-Gleichstromanschluss. Die Kommunikation mit dem Steuer- und Auswerterechner erfolgt über die USB-Schnittstelle. Die am Rechner über eine Benutzeroberfläche eingegebenen Befehle werden vom Microcontroller übersetzt und über die serielle UART-Schnittstelle an die Steuerplatine übertragen. Darüber hinaus kann das Modul über eine RS232-Schnittstelle mit dem externen Trigger der Kamera verbunden werden, um einen bildweisen Wechsel der Beleuchtungswellenlänge zu erwirken. Durch die getaktete Beleuchtung könnten entsprechende Multiwellenlängen-Aufnahmen im nahinfraroten Bereich realisiert werden.

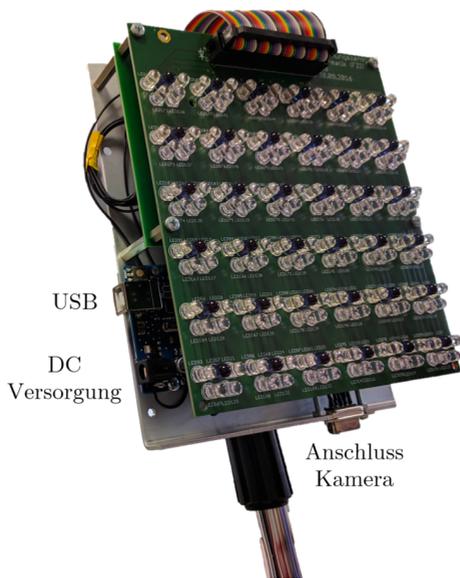


Bild 4.4: Finaler Aufbau des Beleuchtungsmoduls

### 4.3 Versuchsaufbau

Der Versuchsaufbau für die Datenaufnahme besteht aus Bildaufnahmesystem, dem bereits eingeführten Beleuchtungsmodul und Referenzsystemen zur Aufzeichnung der Ground Truth der Vitalparameter und der Artefakt-signale. Auf die einzelnen Bestandteile wird im Folgenden näher eingegan-gen.

### 4.3.1 Bildaufnahme

Wie in Kapitel 3 beschrieben wird für PPGI in erster Linie zwischen dem passiven Ansatz mit Umgebungslicht und dem aktiven Ansatz mit meist monochromer Beleuchtung unterschieden. Die Bildaufnahme erfolgte daher parallel mit der Logitech Pro Webcam HD C920 als Farbkamera und der IDS UI3370CP-Rev. 2 als Monochromkamera [D-AF-06], welche im nahinfraroten Wellenlängenbereich sensitiv ist. Die Bildwiederholfrequenz wurde zu 30 fps gewählt. Die Logitech Webcam wurde mit einer Auflösung von  $640 \times 480$  Pixeln bei einer Bittiefe von 8 Bit betrieben. Für die Aufnahmen mit der IDS Kamera wurden eine Auflösung von  $1024 \times 1024$  Pixeln und eine Bittiefe von 12 Bit genutzt. Durch die Beschränkung auf Einkanal-Analysen bei monochromen Daten, verbunden mit dem verringerten Absorptionskoeffizienten im NIR-Bereich, wurde im Vergleich zur Farbvariante ein höherwertiger Sensor bei größerer Auflösung gewählt, um ein für die Analysen ausreichendes Mindest-SNR zu gewährleisten. Die Kameras wurden in geringem Abstand zueinander auf einem Stativ angebracht und in einer Distanz von 70 – 80 cm zentral vor den Probanden positioniert (s. Abbildung 4.6).

Die Datenaufzeichnung wurde zentral aus Matlab heraus gesteuert. Vor dem Hintergrund des geringen Nutzsignalanteils wurden die verschiedenen von Matlab zur Verfügung gestellten Videoformate hinsichtlich der Signalqualität verglichen. Tabelle 4.1 zeigt eine Aufstellung der verschiedenen Formate. Abbildung 4.5 zeigt Beispiele für einen BVP-Signalausschnitt (vgl. Abschnitt 2.2.3) bei Verwendung der verschiedenen Videoformate. Der negative Einfluss der Kompression auf die BVP-Signalqualität ist deutlich zu erkennen, sodass die Videos der Logitech Webcam im unkomprimierten AVI-Format und die Aufnahmen der IDS-Kamera im verlustfreien Motion JPEG 2000 („Archival“) gespeichert wurden.

Tabelle 4.1: Übersicht der verschiedenen Videoformate in Matlab

<b>Format</b>	<b>Beschreibung</b>
'Archival'	Verlustfrei komprimierte Motion JPEG 2000 Datei
'Motion JPEG AVI'	AVI Datei mit Motion JPEG Enkodierung (verlustbehaftet)
'MPEG-4'	MPEG-4 Datei mit H.264 Enkodierung (verlustbehaftet)
'Uncompressed AVI'	unkomprimierte AVI Datei

### 4.3.2 Beleuchtung

Die Aufnahmen erfolgten im SmartHome des FZI House of Living Labs <sup>1</sup>, welches unter anderem mit dem Ziel der Bereitstellung einer Umgebung für Forschungsstudien, inklusive entsprechender technischer Schnittstellen für die Anpassung der Umgebungsbedingungen, eingerichtet wurde. Das natürlich einfallende und das künstliche Licht konnte so während der Datenaufzeichnung über die per Rest-API an das Aufnahme-Framework angebundene Deckenleuchten und Jalousien gesteuert werden. Für die aktive, monochrome Beleuchtung wurde das in Abschnitt 4.3.2 beschriebene LED-Beleuchtungsmodul bei den Wellenlängen 525 nm (grün) und 940 nm (NIR) genutzt. Das Beleuchtungsmodul wurde zentral hinter und 20 cm über den Kameras platziert und dabei so ausgerichtet, dass eine möglichst homogene Ausleuchtung der Gesichter der Probanden gewährleistet wird.

---

<sup>1</sup> <https://www.fzi.de/forschung/fzi-house-of-living-labs/>

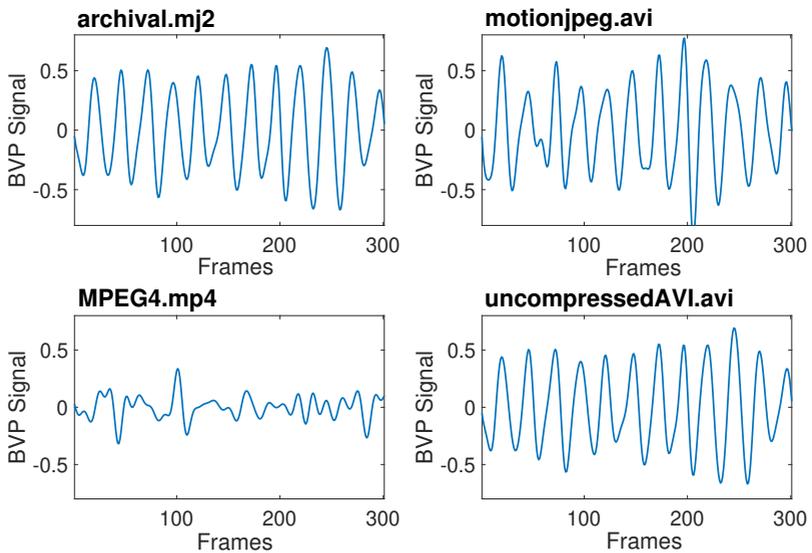


Bild 4.5: Vergleich eines BVP-Signalausschnittes für verschiedene Videoformate

### 4.3.3 Referenz Vitalparameter

Als Referenzsystem zur Aufzeichnung der Vitalparameter wurde das Polysomnographiesystem SOMNOscreen Plus<sup>2</sup> der Firma SomnoMedics verwendet [D-AF-07]. Es bietet die Möglichkeit zum Anschluss einer Vielzahl an Sensoren, deren Daten synchron auf einer integrierten Speicherkarte aufgezeichnet werden. Die für die Datenbank als Referenz aufgenommenen Parameter sind in Tabelle 4.2 mit zugehöriger Abtastrate aufgelistet. Die parallele Aufzeichnung von EKG und PPG sorgt dabei für eine Redundanz bezüglich der Groundtruth-Herzfrequenz für den Fall, dass einer der beiden Sensoren artefaktbedingt keine ausreichende Signalqualität liefert. Neben den Daten zur Herzrhythmicität wurde auch das Atmungs-Referenzsignal mittels Bauch- und Brustgurt (Effort Thorax + Abdomen) erfasst. Der in-

<sup>2</sup> <https://somnomedics.de/produkte/schlafdiagnostik/polysomnographie/somnoscreen-plus/>

tegrierte Lichtsensor diente zur Synchronisation mit dem Videosignal (vgl. Abschnitt 4.3.5).

Tabelle 4.2: Aufgezeichnete Referenzdaten des *SOMNOscreen*-Systems mit verwendeter Abtastrate.

<b>Vitalparameter</b>	<b>Abtastrate</b>
EKG	512 Hz
Pleth	256 Hz
Puls	16 Hz
SpO <sub>2</sub>	16 Hz
Effort Abdomen	64 Hz
Effort Thorax	64 Hz
Effort Thorax + Abdomen	64 Hz
Licht	64 Hz

#### 4.3.4 Referenz Artefaktsignale

Für die Aufnahme der Artefaktsignale Helligkeitsänderung und Kopfbewegung [D-AF-08] wurde das Lux-Meter Extech HD450 der Firma Extech, sowie die IMU Brick 2.0 von TinkerForge verwendet. Die IMU liefert die Orientierungswinkel Roll, Pitch, Yaw, sowie die lineare Beschleunigung entlang der X-, Y- und Z-Achse und wurde per Tragereif am Kopf der Probanden angebracht. Das Lux-Meter wurde hinter den Probanden und oberhalb des Kopfes montiert. Der finale Aufbau mit Bildaufnahmesystem, Beleuchtungsmodul und Referenzsystemen ist in Abbildung 4.6 zu sehen.

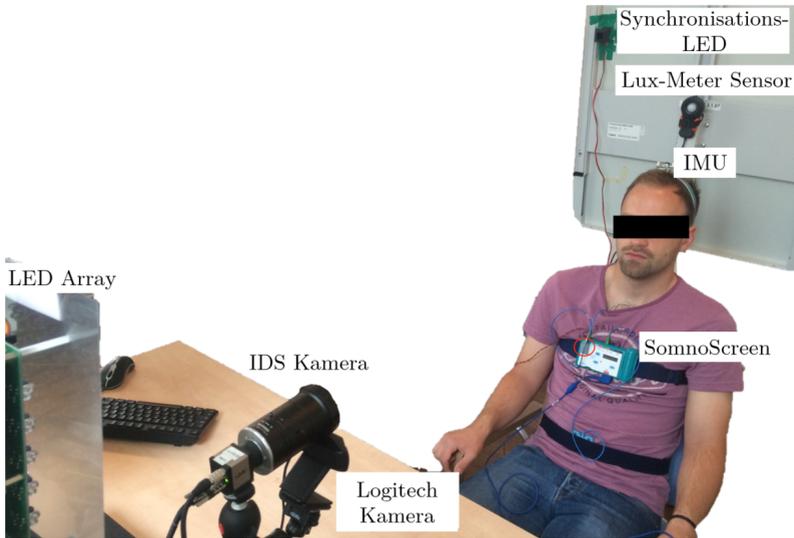


Bild 4.6: Gesamtüberblick der Versuchsumgebung zur Aufnahme der verschiedenen Szenarien

### 4.3.5 Synchronisation

Ein essentieller Teil beim Aufbau der Datenbank war die Synchronisation der Video- und Referenzdaten, da nur mit synchronisierten Daten belastbare Aussagen über die Güte der Algorithmen getroffen werden können. Die Synchronisation erfolgte auf Basis eines binären Lichtsignals, welches durch ein Microcontroller-Board und daran angebundene LEDs erzeugt wurde. Das Lichtsignal wurde zum einen auf den integrierten Lichtsensor des *SOMNOscreen* eingebracht, zum anderen wurde eine LED an der Stellwand hinter den Probanden platziert, sodass diese im Videobild zu sehen ist. Mit Hilfe der durch Schalten der LEDs zu Beginn und am Ende der Aufnahmen generierten Rechtecksignale, konnten die Signale beider Systeme synchronisiert und ein möglicher, zeitlicher Offset korrigiert werden. Die IMU-Daten wurden mittels des Aufnahme-Frameworks parallel zu den Videodaten in Matlab eingelesen. Für die Daten des Lux-Meters erfolgte

die Synchronisation offline über die enthaltenen Zeitstempel, was bei einer Abtastrate von 1 Hz als ausreichend angesehen werden kann.

## 4.4 Szenarien

Das SNR für PPGI hängt von verschiedenen Faktoren wie der Art der Beleuchtung, der Wahl der Kamerasensorik oder der Störung durch Artefakte ab (vgl. Abschnitt 4.6). Die Szenarien der Datenbank sollen die verschiedenen Einflussfaktoren möglichst gut abbilden, um eine entsprechende Analyse zu ermöglichen (vgl. [D-AF-01], [D-AF-02]). Mit diesem Ziel wurden die in Tabelle 4.3 aufgeführten Szenarien definiert. Die Szenarien 101, 102, 106 und 107 nutzen unterschiedliche Arten der Beleuchtung. Die Szenarien 103 und 104 untersuchen den Einfluss langsamer und schneller Lichtänderungen. Der Einfluss von Kopfbewegungen wird in den Szenarien 201-204 analysiert. Die einzelnen Szenarien werden in den folgenden Abschnitten näher beschrieben.

### 4.4.1 Szenarien mit unterschiedlicher Art der Beleuchtung

In den Szenarien zur Analyse des Einflusses unterschiedlicher Beleuchtungsarten befinden sich die Probanden in Ruhe [D-AF-02]. Für Szenario 101 wurde das einfallende Tageslicht als natürliche Beleuchtungsquelle genutzt. Dies stellt für die Analyse mit RGB-Kameras das Best-Case-Szenario dar. Für die Aufnahmen in Szenario 102 wurde der Raum zuvor abgedunkelt (vgl. 4.3.2). Die Ausleuchtung des Gesichtes erfolgte über künstliches Licht der Deckenleuchte. Ziel dieses Szenarios ist die Untersuchung, wie sich die Beleuchtung durch rein künstliches Licht auf die Signalqualität auswirkt. Die Szenarien 106 und 107 beschreiben Aufnahmen, bei denen im Vergleich zu den Szenarien 101 und 102 mit einer selektiven Wellenlänge anstelle eines breiten Lichtspektrums beleuchtet wird. In Szenario 106 wurde grünes Licht der Wellenlänge 525 nm, in Szenario 107 infrarotes Licht der Wellenlänge 940 nm genutzt (siehe Abbildung 4.7). Die NIR-Auf-

Tabelle 4.3: Übersicht der Szenarien

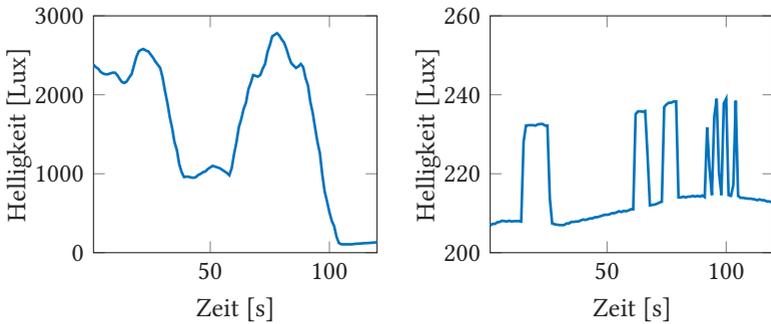
<b>Szenarioname</b>	<b>kurze Beschreibung</b>
101_natural_lighting	Szenario mit Tageslicht ohne Bewegung
102_artificial_lighting	Szenario mit rein künstlicher Beleuchtung ohne Bewegung
103_abrupt_changing_lighting	Szenario mit schnellen, geringen Lichtänderungen
104_dim_lighting_auto_exposure	Szenario mit langsamen, starken Lichtänderungen
106_green_lighting	Szenario mit Beleuchtung im grünen Wellenlängenbereich
107_infrared_lighting	Szenario mit Beleuchtung im Nahinfrarotbereich
201_shouldercheck	Bewegungsszenario mit rotatorischer Bewegung
202_scale_movement	Bewegungsszenario mit skalierender Bewegung
203_translation_movement	Bewegungsszenario mit translatorischer Bewegung
204_writing	Textschreibeszenario

nahmen sind insbesondere für Night Vision Anwendungen von Interesse, auf Grund der entfallenden Blendwirkung bei Beleuchtung im nicht sichtbaren Spektralbereich.

Szenario 103 und 104 beinhalten Aufnahmen unter dem Einfluss von langsamen und schnellen Lichtänderungen, welche über die Steuerung der Jalousien und der Deckenleuchte herbeigeführt wurden. Die zeitliche Abfolge der Lichtänderung wird beispielhaft durch das mittels Lux-Meter gemessene Lichtsignal in Abbildung 4.8 visualisiert.



Bild 4.7: Aufnahmen bei monochromer Beleuchtung mit grünem Licht und im Nahinfrarotbereich



a) Langsame Lichtänderung

b) Schnelle Lichtänderung

Bild 4.8: Zeitliche Verläufe der Helligkeit, die sich aus den Zeitmustern ergeben

#### 4.4.2 Szenarien mit unterschiedlicher Art der Bewegung

Der negative Einfluss der Bewegung auf das zu extrahierende BVP-Signal entsteht durch die Winkeländerungen zwischen Lichtquelle und Haut, sowie Haut und Kamera, da diese sich entsprechend den Ausführungen in Ka-

pitel 2.3.2 direkt auf das Farbsignal der Kamera auswirken. Die Stärke des Einflusses hängt dabei von der Art und Geschwindigkeit der Bewegung ab. In den ersten drei Bewegungsszenarien werden unterschiedliche Arten der Bewegungen anhand vordefinierter rotatorischer (201), skalierender (202) und translatorischer (203) Bewegungen untersucht. Die Abbildungen 4.9, 4.10 und 4.11 zeigen entsprechende Beispiele.

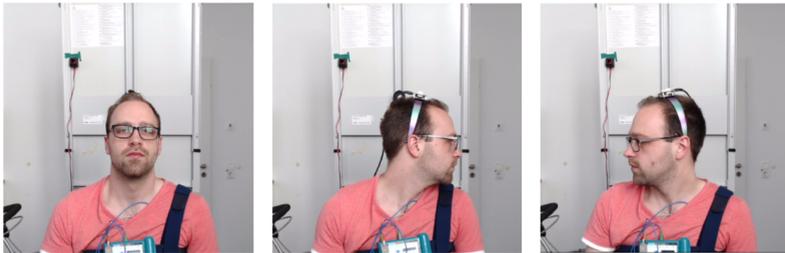


Bild 4.9: Ablauf der rotatorischen Bewegung

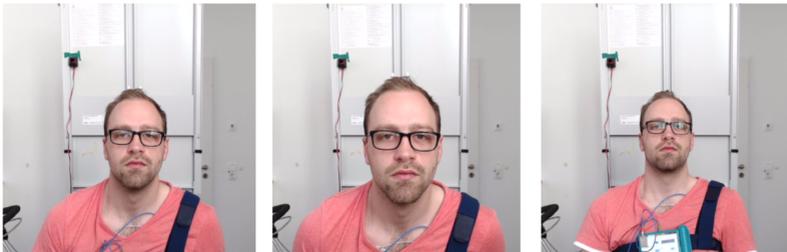


Bild 4.10: Ablauf der skalierenden Bewegung

Die Bewegungen wurden vom Studienleiter entsprechend eines definierten Zeitmusters angewiesen. Das Muster war so angelegt, dass zunächst drei Bewegungen mit einem größeren zeitlichen Abstand verlangsamt ausgeführt wurden und im Anschluss fünf kurz aufeinander folgende, schnellere Kopfbewegungen des selben Typs erfolgten. Auf diese Weise sollten möglichst viele unterschiedliche Bewegungsmodi des Kopfes abgebildet

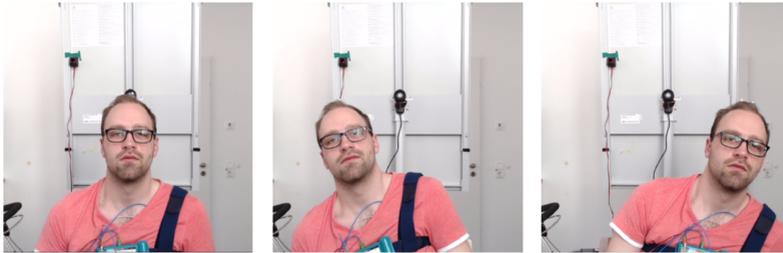


Bild 4.11: Ablauf der translatorischen Bewegung

werden. In Szenario 204 wurden die Probanden angewiesen einen Text abzuschreiben. Dabei war es das Ziel, neben den vorgegebenen Bewegungen in den Szenarien 201-203, auch nicht definierte Kopfbewegungen zu untersuchen. Der Kopf der Probanden war während der Aufnahme leicht geneigt, was zusätzlich den Zugang zur Messregion einschränkt. Das Szenario soll unter anderem die Bedienung der Mittelkonsole während einer Autofahrt simulieren.

### 4.5 Datenaufnahme und -aufbereitung

Insgesamt enthält der Datensatz Videos von 26 Probanden in 10 verschiedenen Szenarien. Das Probandenkollektiv umfasst 20 männliche und 6 weibliche Probanden im Alter von 23 bis 33 Jahren [D-AF-04]. Bei einer Videolänge von jeweils zwei Minuten und der parallelen Aufzeichnung mit zwei Kameras entspricht dies insgesamt 1040 Minuten Videomaterial [D-AF-03]. Vor Beginn einer Messung wurden datenschutzkonform ein Informationsgespräch auf Basis der ausgehändigten Probandeninformation geführt und gegebenenfalls die freiwillige Zustimmung zur Messaufnahme in Form einer signierten Einwilligungserklärung dokumentiert (siehe Anhang). Zur Qualitätssicherung der Datenaufnahme wurde ein spezielles SW-Framework zur automatisierten Aufnahme und Nachbereitung der Daten entwickelt. Während der Aufnahmen wurde der korrekte Ablauf

anhand eines Durchführungsprotokolls kontrolliert. Die synchronisierten Daten liegen innerhalb einer vordefinierten Ordnerstruktur im Unisens-Datenformat<sup>3</sup> vor. Über das SW-Tool UnisensViewer wird eine schnelle und einfache Visualisierung der Daten, über die Unisens Matlab Toolbox ein schneller Zugriff für die Datenverarbeitung ermöglicht.

## 4.6 SNR als Gütekriterium der Signalqualität

Eine gängige Größe zur Beurteilung der Signalqualität eines von Störkomponenten überlagerten Signals ist das Signal-zu-Rausch-Verhältnis (*engl.* „Signal-to-Noise-ratio“(SNR)). Es beschreibt das Verhältnis der Leistung des Nutzsignals zur Rauschleistung der Störkomponenten. Für PPGI kann es zum einen auf das Rohsignal angewendet werden, um die Aufnahmequalität oder die Intensität von Artefakten zu beurteilen, zum anderen auf das BVP-Signal um die Qualität des eingesetzten Algorithmus zu evaluieren. Die Berechnung des SNR's eines BVP-Signals folgt im Rahmen dieser Arbeit der in [34] eingeführten Metrik. Dabei wird das SNR zunächst als Energiequotient der über den Messzeitraum aufgezeichneten Signal- und Rauschanteile definiert:

$$\text{SNR} = 10 \cdot \log_{10} \left( \frac{E_{\text{signal}}}{E_{\text{noise}}} \right) \quad (4.8)$$

Das Nutzsignal wird im Frequenzspektrum als der Signalbereich um die vom Referenzsystem gelieferte Referenz-Herzfrequenz festgelegt, ergänzt um den Bereich der ersten harmonischen Oberschwingung. Da der Nutzsignalanteil im Bereich der ersten Harmonischen größerer Streuung unterliegt, wird dieser mit doppelter Breite gewählt. Für die Wahl der Fensterbreite finden sich in der Literatur unterschiedliche Angaben. In dieser Arbeit werden Fensterbreiten von  $\frac{1}{6} \text{ Hz}$  für den Bereich der Referenz-Herzfrequenz und  $\frac{1}{3} \text{ Hz}$  für den Bereich der ersten Harmonischen definiert.

---

<sup>3</sup> <http://www.unisens.org/>

Der Störanteil entspricht der Energie des übrigen Spektrums. Abbildung 4.12 zeigt beispielhaft das Frequenzspektrum eines BVP-Signals mit Fensterfunktion (grün) und Referenz-Herzfrequenz (rot).

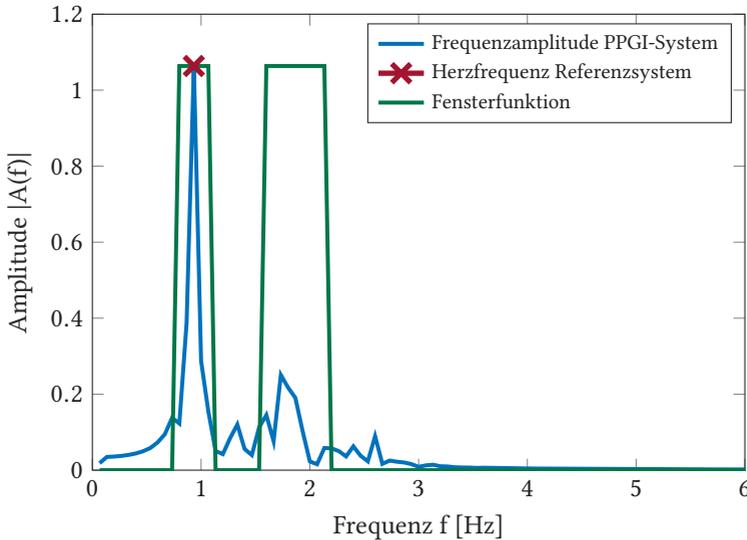


Bild 4.12: Beispiel Frequenzspektrum eines BVP-Signals mit Fensterfunktion (grün) und Referenz-Herzfrequenz (rot)

## 4.7 Analyse der Signalqualität

### 4.7.1 Analyse der Szenarien

Die Analyse der Datenbank erfolgt anhand der im vorigen Abschnitt eingeführten SNR-Metrik. Das übergeordnete Ziel ist es, den Einfluss von Bewegungs- und Lichtartefakten, sowie unterschiedlichen Beleuchtungsarten auf die Messaufnahmen der RGB-Farbkamera und NIR-Monochromkamera zu untersuchen. Zu diesem Zweck wird das SNR für die Farbkanäle der Farbkamera und dem monochromen Kanal der NIR-Kamera für jedes

Szenario einzeln bestimmt, wobei stets der Mittelwert über alle 26 Probanden gebildet wird. Die Detektion der Messregion erfolgt auf Basis einer frühen Version des in Kapitel 6.2 eingeführten Face Alignment Systems mit fester Regionenwahl im Gesicht. Auf Basis der extrahierten Landmarken werden dabei die Stirn- und Wangenbereiche als ROI definiert, wie in Abbildung 4.13 beispielhaft dargestellt.

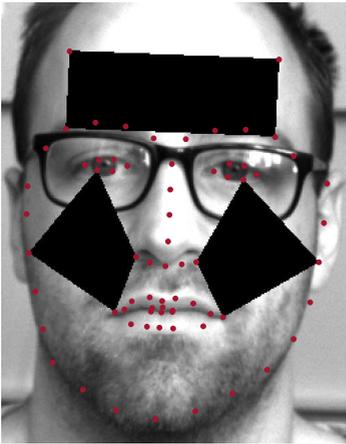


Bild 4.13: Definition der ROI für die Szenario-Analyse

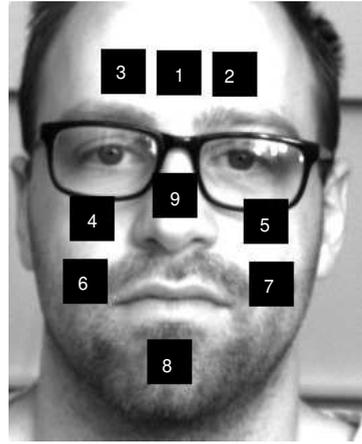


Bild 4.14: Definition der Subregionen für die ROI-Analyse

Tabelle 4.4 enthält die SNR-Werte der Einzelkanäle in Dezibel (dB) <sup>4</sup>. Für die RGB-Farbkamera erzielt der grüne Farbkanal erwartungsgemäß die höchsten SNR-Werte. Dabei wird die beste Signalqualität bei natürlicher Beleuchtung erreicht (Szenario 101). Für die konstante Beleuchtung mit künstlichem weißen, grünen und infraroten Licht (Szenarien 102, 106 und 107) zeigt sich im Vergleich eine Verringerung des SNR. In diesen Szenarien wurde die natürliche Beleuchtung über die Jalousien reduziert. Die Luxmeter-Messungen für Szenario 102 liegen etwa Faktor 4 unter den Mes-

<sup>4</sup> Im Vergleich zur Auswertung in [13] wurde der amplitudenselektive Filter nach [155] nicht eingesetzt, wodurch sich im Vergleich leicht abweichende Werte ergeben können.

sungen für Szenario 101. Die intensitätsärmere, künstliche Beleuchtung führt zu einem erhöhten Bildrauschen, was sich insbesondere auf preiswerte Sensoren auswirkt. Bei selektiver Beleuchtung mit grünem Licht verhindert der automatische Weißabgleich der Farbkamera eine bessere Signalqualität. Unter Einfluss wechselnder Beleuchtungsintensität nimmt das SNR auf Werte unter  $-5\text{ dB}$  ab.

Tabelle 4.4: Gemittelte SNR-Werte der Kameras für die verschiedenen Szenarien in dB

<b>Szenario</b>	<b>Roter Kanal Farbkamera</b>	<b>Grüner Kanal Farbkamera</b>	<b>Blauer Kanal Farbkamera</b>	<b>Mono- chrom- Kamera</b>
101	-5,51	-0,65	-2,96	-4,27
102	-5,90	-3,07	-5,30	-5,58
103	-7,52	-5,16	-6,93	-7,80
104	-8,28	-5,93	-7,14	-6,25
106	-4,69	-2,32	-3,50	-1,72
107	-6,41	-3,61	-5,28	-3,38
201	-8,64	-8,09	-8,46	-9,34
202	-7,93	-5,68	-7,08	-8,47
203	-8,00	-5,90	-7,28	-8,27
204	-8,59	-8,12	-8,58	-11,39

In den Bewegungsszenarien (Szenarien 201-204) ergibt sich eine weitere Verringerung der Signalqualität aufgrund der sich ändernden Position der Messregion. Während translatorische und skalierende Bewegungen der Probanden (Szenarien 202-203) einen ähnlich negativen Einfluss auf die Signalqualität haben, wie die sich ändernde Beleuchtung in den Szenarien 103 und 104, ergibt sich für die rotatorische Bewegung (Szenario 203) und

das Textschreibszenario 204 nochmals ein deutlicher Abfall des SNR. In diesen Szenarien ist die ROI meist nur teilweise und in bestimmten Sequenzen gar nicht zugänglich. Gerade diese Szenarien kommen beispielsweise im Fahrzeug in Form des Schulterblicks, der Interaktion mit dem Beifahrer oder der Bedienung der Mittelkonsole vermehrt vor. Hierfür gilt es im Rahmen der Arbeit geeignete Methoden zur Artefaktkompensation zu entwickeln. Die Analyse des SNR lässt darauf schließen, dass Bewegungsartefakte einen größeren negativen Einfluss auf die Signalqualität haben, als eine sich ändernde Beleuchtung. Allerdings ist hierbei zu beachten, dass die Intensitätsänderung, insbesondere in Szenario 103, im Vergleich zu einem Fahrszenario mit Licht-Schatten-Wechseln bei starker Sonneneinstrahlung, geringer ausfallen. Zur Untersuchung solcher Einflüsse ist es notwendig, Realdaten im Feld zu erheben.

Die Ergebnisse für die NIR-Kamera weisen eine ähnliche Charakteristik auf. In Szenarien mit weißer Beleuchtung ergeben sich im Vergleich zu den Farbkanälen niedrigere Werte für das SNR, da auf Grund der Empfindlichkeit im Nahinfrarotbereich bei geringerer Absorptionsstärke von Hämoglobin ein vergleichsweise erhöhter Rauschanteil in das Frequenzspektrum eingeht. Vorteile ergeben sich hingegen bei der monochromatischen Beleuchtung in den Szenarien 106 und 107. Dennoch liegen die Werte in diesen Szenarien unter dem Wert des grünen Kanals der Farbkamera bei natürlicher Beleuchtung in Szenario 101. Vor dem Hintergrund der vorteilhaften Eigenschaften der NIR-Kamera hinsichtlich Bildqualität, verdeutlicht dies die Notwendigkeit einer hohen Beleuchtungsintensität beim aktiven Ansatz mit künstlicher, monochromer Beleuchtung. Im Vergleich der Szenarien 106 und 107 zeigt sich für die NIR-Kamera zudem der deutliche Unterschied im SNR zwischen der Beleuchtung im grünen und infraroten Wellenlängenbereich.

### 4.7.2 Analyse unterschiedlicher Messregionen im Gesicht

Neben der Analyse der einzelnen Szenarien mit fester und großflächiger ROI wird auch die Signalqualität unterschiedlicher Messregionen im Gesicht untersucht. Dazu wird das Gesicht eines Probanden, wie in Abbildung 4.14 dargestellt, in neun Subregionen unterteilt. Um die Vergleichbarkeit zu gewährleisten, werden Subregionen gleicher Flächengröße definiert.

Die Ergebnisse der neun Messregionen sind in Tabelle 4.5 beispielhaft für Szenario 101 aufgeführt. Durch die geringere Größe der Subregionen und der damit verbundenen verringerten räumlichen Mittelung, kommt es im Vergleich zur Szenario-Analyse zu einer leichten Verringerung der Messgenauigkeit und des mittleren SNR. Bei der monochromen NIR-Kamera ist der Effekt weniger stark ausgeprägt, da sie eine höhere Bildauflösung aufweist. Darüber hinaus reduziert die Bittiefe von 12 Bit den Quantisierungsfehler. In Szenario 101 nimmt das SNR für den grünen Kanal der RGB-Kamera von  $-0,65$  dB auf  $-1,99$  dB ab, für die monochrome Kamera ist die Differenz mit  $-4,27$  dB und  $-4,77$  dB geringer.

Vergleicht man die Werte der einzelnen Subregionen, so zeigt Subregion 1 der Stirn die beste Signalqualität, gefolgt von den Wangenregionen 4 und 5, sowie den Stirnregionen 2 und 3. Dies lässt sich unter anderem mit der stärkeren Hautdurchblutung dieser Regionen und der im Vergleich zu anderen Regionen geringeren Zahl an Nicht-Haupixeln (Bart oder Brille) begründen. Vor allem aber liegen diese Teile der Messregion in etwa auf einer räumlichen Ebene, die orthogonal zur Blickrichtung der Kamera liegt. Dies resultiert in einer erhöhten Zahl an von den Gefäßen reflektierten Photonen, welche zum gleichen Zeitpunkt auf den Bildsensor treffen. Der Effekt der Helligkeitsänderung mit der Frequenz der Blutvolumenschwankung ist in diesen Subregionen somit am stärksten ausgeprägt. Auffallend ist zudem, dass die Ergebnisse der Subregionen 5 und 2 im Vergleich zu den Subregionen 4 und 3 leicht erhöht sind. Dies deutet auf einen einseitig

Tabelle 4.5: SNR-Werte für verschiedene ROIs in Szenario 101

<b>Region</b>	<b>Roter Kanal Farbkamera</b>	<b>Grüner Kanal Farbkamera</b>	<b>Blauer Kanal Farbkamera</b>	<b>Mono- chrom- Kamera</b>
1	-4,07	0,72	-2,66	-3,66
2	-5,98	-2,56	-5,19	-4,96
3	-6,16	-2,76	-5,12	-5,34
4	-4,54	-0,56	-3,55	-4,09
5	-4,39	-0,44	-3,38	-3,78
6	-6,55	-3,35	-5,44	-5,76
7	-6,26	-3,19	-5,35	-4,81
8	-6,13	-3,67	-5,85	-5,55
9	-5,37	-2,08	-4,92	-4,94
Mean	-5,49	-1,99	-4,61	-4,77

verstärkten Lichteinfall oder eine leichte Seitendifferenz in der Gesichtsdurchblutung hin.

## 4.8 Diskussion und Fazit

In diesem Kapitel wurde die Erstellung einer Datenbank für PPGI beschrieben, welche durch die Definition unterschiedlicher Szenarien die Analyse des Einflusses von verschiedenen Beleuchtungsarten, sowie Licht- und Bewegungsartefakten auf die Signalqualität ermöglicht. Mit Aufnahmen von 26 Probanden und insgesamt 1040 Minuten Videomaterial dient sie als Grundlage des weiteren Systementwurfs. Die Analyse der Daten zeigt, dass rotatorische Kopfbewegungen die Signalqualität stärker negativ beeinflussen, als translatorische und skalierende Bewegungen. Auch die dau-

erhaft geneigte Kopfhaltung im Textschreibeszenario führt zu vergleichsweise niedrigen SNR-Werten, da die ROI meist nur teilweise zugänglich ist. Da diese Bewegungen im angestrebten Anwendungsszenario im Fahrzeug in Form des Schulterblicks, der Interaktion mit dem Beifahrer oder der Bedienung der Mittelkonsole vermehrt vorkommen, werden im Rahmen der Arbeit geeignete Methoden zur Artefaktkompensation entwickelt. Die Auswertung der Beleuchtungsszenarien lässt auf einen leicht geringeren Einfluss von Lichtänderungen gegenüber Bewegungsänderungen schließen. Dies gilt es anhand von Realdaten mit stärkeren Lichtintensitätsänderungen, wie sie im Fahrzeug insbesondere bei hoher Sonneneinstrahlung häufig vorkommen, detaillierter zu untersuchen. Gleichzeitig ergeben sich bessere Resultate unter Verwendung von natürlichem Tageslicht im Vergleich zu künstlicher Beleuchtung. Insbesondere für den NIR-Bereich zeigt sich die Notwendigkeit einer hohen monochromatischen Beleuchtungsintensität. Generell führt eine absinkende Lichtintensität, abhängig von den Sensoreigenschaften der Kamera, zu einer verringerten Signalqualität. Mit Blick auf unterschiedliche Messregionen im Gesicht liefern die Stirn- und Wangenbereiche das stärkste Signal. Auf Basis dieser Erkenntnisse wird im folgenden Kapitel das Systemkonzept mit Fokus auf die Entwicklung geeigneter algorithmischer Ansätze zur Kompensation von Artefakten vorgestellt.

# 5

## Systemkonzept

**I**M FOLGENDEN KAPITEL wird das Systemkonzept als Grundlage der Entwicklung des PPGI-Systems im Rahmen dieser Arbeit vorgestellt. Zunächst werden aufbauend auf den Erkenntnissen aus der Analyse des Standes der Technik (Kap. 3), sowie der initialen Auswertung der Datenbankmessungen im vorherigen Kapitel, die Anforderungen an das System definiert. Im Anschluss wird der grundsätzliche Hardware-Systemaufbau, die Softwarearchitektur des Systems auf Komponentenebene und die zugehörigen Schnittstellen beschrieben. Da der Fokus der Arbeit auf der Konzeption und Entwicklung von PPGI-Algorithmen liegt, wird der Analyseteil des Systems ausführlich betrachtet. Abschließend wird das prinzipielle Vorgehen für die Entwicklung der einzelnen Teilkomponenten, welche Inhalt der Folgekapitel sind, kurz erläutert.

## 5.1 Systemanforderungen

Mit dem Ziel der Entwicklung eines onlinefähigen PPGI-Systems zur Remote-Erfassung der Herzaktivität über das menschliche Gesicht in realen und herausfordernden Einsatzszenarien, werden zunächst folgende Systemanforderungen definiert:

- [S-AF-01] Das System soll eine robuste Erfassung der Herzfrequenz in herausfordernden Umgebungen ermöglichen, d.h. auch bei starker Bewegung der Person oder bei Integration in mobile Systemumgebungen.
- [S-AF-02] Das System soll grundsätzlich die Bestimmung der Pulsratenvariabilität ermöglichen.
- [S-AF-03] Das System soll als onlinefähiges System eine kontinuierliche Messung inklusive Live-Feedback ermöglichen.
- [S-AF-04] Das System soll die Güte der Schätzergebnisse rückmelden können.
- [S-AF-05] Die Funktionalität soll unabhängig von Alter, Geschlecht oder Ethnie gegeben sein.
- [S-AF-06] Das System soll als mobiles System mit möglichst geringem Aufwand in unterschiedliche Umgebungen integrierbar sein, mit dem Ziel der Nutzung in unterschiedlichen Anwendungsszenarien.
- [S-AF-07] Das System soll eine unaufdringliche Messmethode zur Remote-Messung bei Entfernungen im Ziel-Distanzbereich von 0,5 – 1,5 m bereitstellen.
- [S-AF-08] Mit dem Ziel einer anwendungsorientierten Umsetzung sollen kostengünstige Bildsensoren und herkömmliche Verarbeitungsprozessoren genutzt werden.

[S-AF-09] Das System soll die Möglichkeit zur Anbindung verschiedener Kameras bieten.

## 5.2 Systemspezifikation

### 5.2.1 Systemaufbau

Für die Spezifikation des Systems werden neben den funktionalen Anforderungen an das Gesamtsystem auch die Erkenntnisse aus der Basisanalyse der Datenbank des vorigen Kapitels berücksichtigt. Als erste Konsequenz wird für die Entwicklung der passive Systemansatz ohne aktive Beleuchtung und unter Nutzung von RGB-Farbkameras gewählt. Dies liegt insbesondere im vergleichsweise niedrigen SNR bei Messungen mit aktiver Beleuchtung im nicht sichtbaren Infrarotbereich begründet, was im Widerspruch zu [S-AF-01] steht. Eine Beleuchtung im grünen Wellenlängenbereich wird auf Grund der Blendwirkung ausgeschlossen, da diese Anforderung der unaufdringlichen Messung [S-AF-07] widerspricht und somit keine kontinuierliche Messung ermöglicht [S-AF-03]. Im Gegensatz dazu kann der blutvolumenabhängige Absorptionseffekt des Lichts mittels RGB-Kameras (CCD- oder CMOS-Siliziumsensoren) bei entsprechender Umgebungsbeleuchtung optimal erfasst werden, da die Absorptionsmaxima von oxygeniertem Hämoglobin bei 541 nm und 577 nm und somit im grünen Wellenlängenbereich liegen (vgl. Kapitel 2.2.2). Darüber hinaus wurde im Stand der Technik bereits aufgezeigt, dass unter Verwendung einer RGB-Kamera als Sensor, die Signale der drei Farbkanäle eine Möglichkeit zur Störkompensation mittels Multikanalanalyse bieten.

Die Nutzung von RGB-Sensoren ermöglicht zudem die Auswahl aus einer großen Vielfalt an kostengünstigen Off-the-Shelve-Kameras [S-AF-08], welche nach Lewandowska et al. [86] eine ausreichende Erfassung des Messeffekts ermöglichen. Für die Aufzeichnung der Datenbank (4) wurde daher die Logitech Webcam c920 als Sensor eingesetzt. Die dabei gewählten Parameter (4.3.1) stellen somit die Minimalanforderung an die Bild-

erfassung dar. Für einen Prototypen-Aufbau soll das Bildaufnahmesystem durch gleich- oder höherwertige Sensoren ersetzt werden können [S-AF-09]. Eine genauere Analyse von Kameraeigenschaften erfolgt im Rahmen der Kameraauswahl für das InCar-System in Kapitel 8.3.6.

Ein weiterer Vorteil liegt beim passiven Ansatz im einfachen Messaufbau bestehend aus Kamera und Recheneinheit, ohne spezielles Beleuchtungssystem [S-AF-06]. Ergänzt um eine Anzeige-Komponente zur Live-Visualisierung der Ergebnisse [S-AF-03] ergibt sich der in Abbildung 5.1 visualisierte Hardware-Systemaufbau. Die HW-Anbindung zwischen Ka-

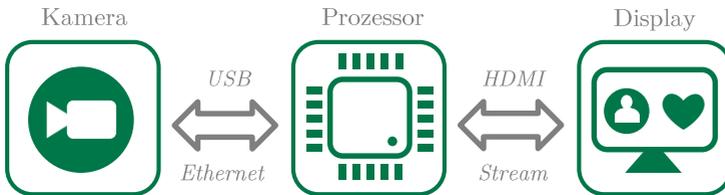


Bild 5.1: Hardware-Aufbau des Gesamtsystems

mera und Verarbeitungseinheit erfolgt wahlweise über USB oder Ethernet. Für die Umsetzung als mobiles System [S-AF-06] sollte die Kamera-Hardware die direkte Stromversorgung über die Schnittstelle ermöglichen. Dies erleichtert die prototypische Integration in neue Messumgebungen.

### 5.2.2 Softwarearchitektur

Das SW-System kann grundsätzlich in vier Komponenten unterteilt werden:

- *Steuerkomponente*: Steuern und Überwachen der Systemprozesse, z.B. Anzahl der Bilder im Image Buffer

- *Einlesekomponente*: Komponente zum Einlesen und Zwischenspeichern der Bilddaten und optional der Referenzdaten (z.B. eines EKG-Systems)
- *Analysekomponente*: Verarbeitungseinheit zur Analyse der Bilddaten und Bestimmung der Herzfrequenz
- *Bedienkomponente*: Graphische Benutzeroberfläche (*engl.*: Graphical User Interface (GUI)) zur Interaktion mit dem System und Live-Visualisierung der berechneten Herzfrequenz, sowie optionaler Meta- und Kontextinformationen

Die Konzeption und Implementierung der verschiedenen Komponenten erfolgt dabei stets im Hinblick auf [S-AF-03], dem Ziel der Umsetzung als Live-Prototypensystem mit Online-Lauffähigkeit auf Standard-Prozessoren. Die Einlesekomponente übernimmt das Einlesen und Zwischenspeichern der Kamerabilder und Referenzdaten. Dabei sollen neben der Online-Verarbeitung auch aufgezeichnete Videos zur nachgelagerten Offline-Analyse eingelesen werden können. Das Steuern und Überwachen von Prozessen ist Aufgabe der Steuerkomponente. Steigt beispielsweise die Anzahl der Frames im Image Buffer in Folge einer zu hohen CPU-Auslastung, wird unter anderem die Zahl der gleichzeitig verarbeiteten Bilder durch die Steuerkomponente an die zur Verfügung stehenden Ressourcen angepasst und über die Bedienkomponente eine entsprechende Meldung an den Nutzer ausgegeben. Die Benutzeroberfläche soll neben der Live-Visualisierung auch die Möglichkeit zum Aktivieren und Deaktivieren des Datenloggings bereitstellen.

Im Fokus der Arbeit steht die Entwicklung des algorithmischen Konzepts, sowie die prototypische Umsetzung für die Machbarkeitsanalyse beim Einsatz in Realszenarien. Die Portierung auf Zielsysteme, wie etwa einer Steuergeräteeinheit im Fahrzeug, ist nicht Bestandteil dieser Arbeit. Daher werden im nächsten Schritt des konzeptionellen Systementwurfs -

der Anforderungsanalyse der Komponenten - lediglich die aus den Systemanforderungen [S-AF-xx] in Abschnitt 5.1 abgeleiteten Anforderungen an die Analysekomponente explizit aufgeführt und beschrieben. Der Algorithmus der Analysekomponente lässt sich dazu erneut in drei Teilkomponenten zerlegen:

BV *Bildverarbeitungskomponente*: Detektion der Messregion zur Extraktion des Rohsignals

SV *Signalverarbeitungskomponente*: Signalverarbeitung zur Extraktion des BVP-Signals

SC *Schätzkomponente*: Bestimmung der Herzfrequenz auf Basis des BVP-Signals

Die definierten Anforderungen an die Analysekomponente [A-AF-xx] werden der jeweiligen Teilkomponente zugeordnet:

[A-AF-01] (BV) Das Gesicht muss robust detektiert und getrackt werden.

[A-AF-02] (BV) Die Beschreibung der Gesichtspose muss eine geeignete Definition der ROI ermöglichen.

[A-AF-03] (BV) Auf Grund der Bewegungs- und Lichtartefakte sollte die ROI adaptiv sein.

[A-AF-04] (SV) Das BVP-Signal muss möglichst rauscharm extrahiert werden.

[A-AF-05] (SV) Es bedarf geeigneter Methoden zur Kompensation von Bewegungs- und Lichtartefakten.

[A-AF-06] (SC) Das System soll eine Schlag-zu-Schlag-Detektion für die PRV-Analyse ermöglichen.

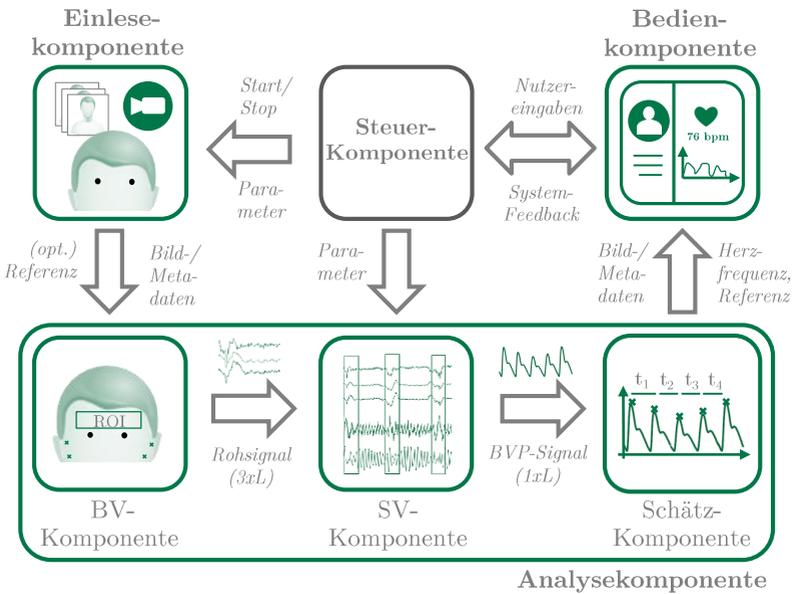


Bild 5.2: SW-Systemarchitektur des Gesamtsystems

Insgesamt ergibt sich dadurch die in Abb. 5.2 gezeigte, grundlegende SW-Systemarchitektur. Die definierten Schnittstellen sind in Form der Ein- und Ausgangsdaten der Software-Komponenten in Tabelle 5.1 nochmals genauer beschrieben.

### 5.3 Abgeleitetes Vorgehen

Die Systembeschreibung anhand der Systemanforderungen [S-AF-xx], der Systemkomponenten, sowie zugehöriger Schnittstellen dient als Basis der weiteren Systementwicklung und soll gleichzeitig ein grundsätzliches Systemverständnis für den Leser schaffen. Im Zuge dieses Kapitels wurde zudem die Entwicklung der algorithmischen Teilkomponenten der Analysekomponente, im Hinblick auf den Einsatz des PPGI-Systems in herausfor-

Tabelle 5.1: Definierte Schnittstellen mit Ein- und Ausgangsdaten

<b>Komponente</b>	<b>Eingabe</b>	<b>Ausgabe</b>
<i>Steuer-Komponente</i>	Benutzereingaben	Start-/Stopbefehle, Parameter
<i>Einlese-Komponente</i>	Image Frames von der Kamera mit $f_k$	Frames in Image Buffer, Metainformation
<i>BV-Komponente</i>	Image Frames aus Memory mit $f_s$	RGB Rohsignal der Länge L in Raw Signal Buffer ( $3 \times L$ )
<i>SV-Komponente</i>	RGB Rohsignal der Länge L aus Raw Signal Buffer BVP	Signal der Länge L in BVP Signal Buffer ( $1 \times L$ )
<i>Schätz-Komponente</i>	BVP Signal der Länge L in BVP Signal Buffer	Berechnete Herzfrequenz in bpm
<i>Bedien-Komponente</i>	Herzfrequenz System und Referenz, Meta-/ Kontextinformationen	Start/Stop-Flag, Parameter

dernden Realszenarien, als ein wesentlicher Bestandteil der Arbeit herausgestellt. Die Konzeption und Umsetzung dieser Teilkomponenten erfolgt unter Beachtung der definierten Anforderungen [AF-xx-BV/SV/SC] und ist Inhalt der folgenden Kapitel.

Kapitel 6 beschreibt die Entwicklung der Ansätze zur Detektion und Tracking der ROI und zur Extraktion des RGB-Rohsignals der BV-Komponente. Im Folgekapitel werden die Signalverarbeitungsmethoden zur Extraktion des BVP-Signals aus dem RGB-Rohsignal (Kap. 7.1), sowie zur Bestimmung der Herzfrequenz mittels Schlag-zu-Schlag-Detektion im BVP-Signal (Kap. 7.2) beschrieben (SV- und SC-Komponente). Die algorithmischen Teilkomponenten werden jeweils auf Basis der Daten der Datenbank aus Kapitel 4 hinsichtlich üblicher Kenngrößen für die Bewertung

von PPGI-Ansätzen evaluiert. Durch Kombination der entwickelten Teilkomponenten erfolgt anschließend die Umsetzung als onlinefähiges Prototypensystem (8.1). Abschließend wird das Gesamtsystem hinsichtlich der Erfassung der PRV im Setting eines Büroarbeitsplatzes (8.2), sowie der robusten Erfassung der Herzfrequenz im Szenario der Fahrerbeobachtung im Automobil (8.3) im Feld evaluiert.



# 6

## Modell und Algorithmus zur Detektion der Messregion

**D**IESES KAPITEL beinhaltet die Beschreibung der entwickelten Ansätze zur Detektion und zum Tracking der Messregion (ROI) im Gesicht in unbekanntem Bildsequenzen, sowie der darauf aufbauenden Extraktion des RGB-Rohsignals. In der Übersicht der Softwarearchitektur entspricht dies der Bildverarbeitungskomponente (siehe Abb. 6.1). Die automatisierte Identifikation geeigneter Hautareale stellt die Basis für die kontinuierliche Signalerfassung und somit der PPGI-basierten Onlinemessung der Herzfrequenz dar. Besonders unter Einfluss von Bewegung sind geeignete Verfahren notwendig, die ROI über Bildsequenzen hinweg zu verfolgen. Die Signalqualität eines PPGI-Systems hängt dabei direkt von der Güte des

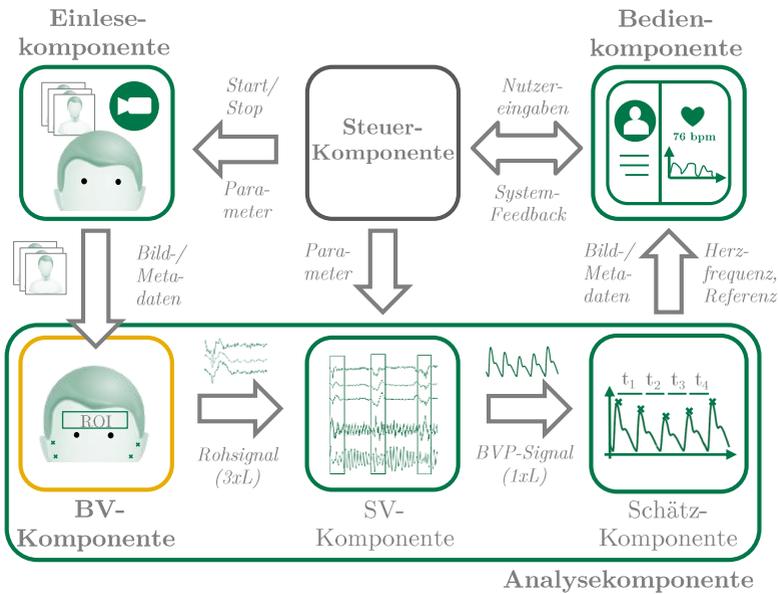


Bild 6.1: Softwarearchitektur des Gesamtsystems - Bildverarbeitung

ROI-Tracking-Algorithmus ab. Daher wurde für die Detektion des Gesichts und das Verfolgen von Bild zu Bild, ein Face-Tracking-Modul entwickelt, welches Position und Form des Gesichts anhand von 68 Merkmalspunkten (Landmarken) beschreibt. Das Modul bezieht sich auf Anforderung [A-AF-01](BV) des Systemkonzepts und wird im folgenden Abschnitt ausführlich vorgestellt. Auf Basis der Landmarken werden im nächsten Schritt Pixelareale innerhalb des Gesichts als ROI definiert, welche Hautbereiche mit möglichst hohem SNR enthalten [A-AF-02](BV). Hierzu werden zwei Verfahren eingeführt.

- Im Fixed-Region-Ansatz werden auf Grundlage der Position bestimmter Landmarken zuvor festgelegte Teilbereiche des Gesichts als ROI definiert. Die Auswahl der Teilbereiche basiert auf den Datenbank-

Untersuchungen aus Kapitel 4.7.2. Der Durchschnitt der Pixelwerte innerhalb der gesamten ROI entspricht dem Rohsignalwert und ist somit Eingang der nachfolgenden Signalverarbeitung.

- Der Multi-Patch-Ansatz bietet eine adaptive Alternative, bei welchem kleine Pixelgruppen innerhalb der Gesichtsregion als einzelne Sensoren dienen. Dabei fließen nur Signale von Sensoren mit ausreichender, momentaner Signalqualität in die weitere Verarbeitung mit ein [A-AF-03](BV) .

Die beiden Verfahren werden in den Abschnitten (6.2) und (6.3) beschrieben.

## 6.1 Face-Tracking-Modul

Das Face-Tracking-Modul vereint die Methoden der Face Detection und des Face Alignment. Die Aufgabe der Face Detection, ist die initiale Detektion des Gesichts einer Person im Bild und die Bereitstellung der Position über eine Bounding Box. Die Face Detection erfolgt mit Hilfe der Viola-Jones-Methode [153]. Das Face Alignment sorgt nachfolgend für die möglichst exakte Positionsschätzung der Gesichtslanmarken und definiert die geometrische Struktur. In Kapitel 2.4 wurden das Active Appearance Model (AAM) und die darauf aufbauende Supervised Descent Methode (SDM) zur Lösung des Face Alignment Problems vorgestellt. In der Literatur wird für das Tracking in PPGI-Systemen bisher meist der Kanade-Lucas-Tomasi-Algorithmus eingesetzt [74][58][41]. Dieser basiert wie das AAM auf einem nichtlinearen Gradientenabstieg, mit dem Unterschied, dass beim AAM die Registrierung zwischen dem Eingangsbild und einem parametrischen Erscheinungsmodell, anstelle eines statischen Templates erfolgt. Die AAM gilt daher als besonders geeignet für das Tracking von Gesichtern. Als ein Vorteil der SDM wurde die vergleichsweise geringe Rechenintensität beschrieben, eine wichtige Anforderung für die angestrebte Umsetzung als

Live-Analyse-System [S-AF-03]. Darüber hinaus besitzt die Methode den Anspruch, durch die Nutzung von SIFT-Merkmalen anstelle der Grauwertverteilung zur Beschreibung der Gesichtstextur, robuster gegenüber Beleuchtungsänderungen zu sein. Je nach Anwendungsszenario, ist dies eine wichtige Eigenschaft für ein PPGI-System. In dieser Arbeit wird daher ein auf der SDM basierender Ansatz konzipiert und hinsichtlich der PPGI-Signalverarbeitung erweitert. So werden SIFT-Features zur Beschreibung der Textur und ein Regressionsmodell mit kaskadierten linearen Regressoren zur Lösung des Minimierungsproblems des Fitting-Prozesses genutzt. Auf die Unterschiede zur originalen SDM wird in Abschnitt 6.1.1 eingegangen.

Insgesamt nutzt das Face-Tracking-Modul drei Modelle für das Face Alignment:

- Das Formmodell, bestehend aus einer Basisform und Formvektoren zur Beschreibung der geometrischen Gesichtsstruktur mit 68 Landmarken
- Das initiale Vorhersagemodell, zur groben Schätzung der Landmarkspositionen basierend auf der Face Detection
- Das Regressionsmodell, zur möglichst exakten Schätzung der Landmarkspositionen durch Minimierung der Fehlerfunktion

Das Ziel des Vorgehens besteht darin, in einem unbekanntem Gesichtsbild die optimale Kombination aus Basisform und Formvektoren des Formmodells zu finden. Die Aufgabe des initialen Schätzmodells ist es, auf Basis des Ausgangs der Face Detection eine erste Modellinstanz zu liefern, die die Form des Gesichts näherungsweise beschreibt. Diese erste Schätzung wird anschließend mit Hilfe des Regressionsmodells optimiert. Das Ergebnis entspricht der gesuchten Modellinstanz. Im Folgenden wird zunächst der Trainingsprozess zur Erstellung der verschiedenen Modelle beschrieben. Anschließend wird auf den Tracking-Prozess für unbekanntes Bildsequenzen eingegangen.

### 6.1.1 Modellerstellung

Der Trainingsprozess gliedert sich in vier Schritte. Abbildung 6.2 zeigt den Prozess. Im ersten Schritt wird das Formmodell (siehe 2.4.2) auf Basis der

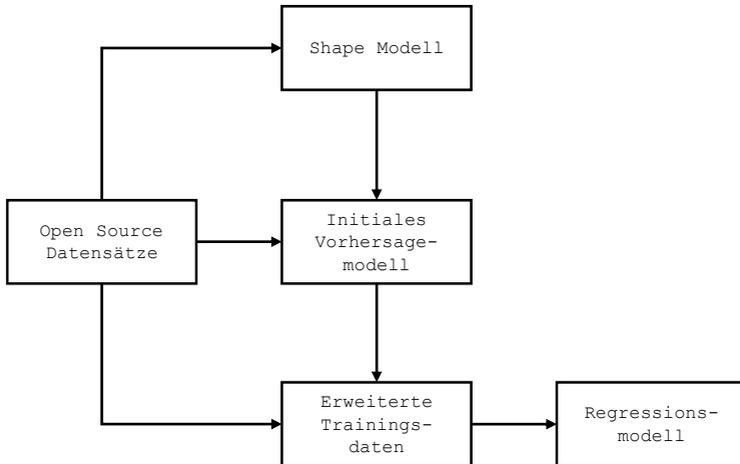


Bild 6.2: Trainingsprozess für die Modellerstellung

annotierten Landmarken der verwendeten Datensätze trainiert. Als Datensätze wurden unter anderem fünf bekannte Open Source Datensätze genutzt, welche jeweils 68 Landmarken zur Beschreibung der Gesichtsform nutzen (vgl. Abbildung 6.3).

- Labeled Face Parts in the Wild (LFPW) [12]
- Helen dataset [81]
- Annotated Faces in the Wild (AFW) [166]
- iBug - 300 Faces in the Wild [119]
- iBug - 300 Videos in the Wild [27]

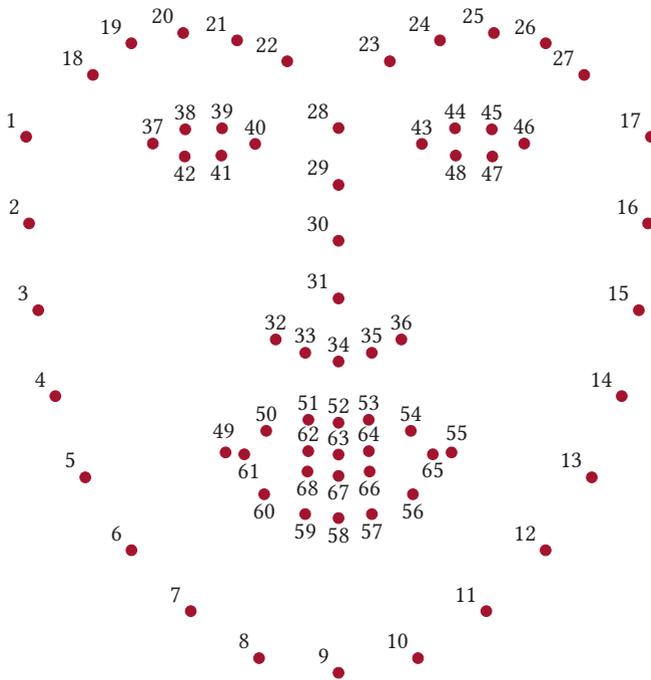


Bild 6.3: Annotation der Landmarken

Im zweiten Schritt wird das initiale Vorhersagemodell angelehrt. Das Modell lernt dabei die relativen Positionen und Skalen der Landmarken innerhalb der Bounding Boxes der Face Detection für die initiale Schätzung der Landmarkspositionen in einem unbekanntem Bild. Auf Basis der bestehenden Modelle wird anschließend mittels Monte-Carlo-Sampling der Trainingsdatensatz, für das Trainieren des Regressionsmodells im vierten und letzten Schritt, künstlich erweitert.

## Shape Modell

Die Erstellung des Formmodells folgt der Beschreibung in Kapitel 2.4.2. Grundlage ist der Trainingsdatensatz mit  $N$  Bildern, in denen die Gesichtsformen  $s$  über die annotierten Landmarken  $x_v, y_v$  nach  $s = (x_1, y_1, \dots, x_v, y_v)$  beschrieben sind. Für das nach Gleichung 6.1 definierte Modell erfolgt zunächst die Bestimmung der Basisform  $s_0$  mittels Procrustes-Analyse. Dabei werden die unterschiedlich skalierten und orientierten Trainingsdaten normiert, beziehungsweise einheitlich ausgerichtet. In Abbildung 6.4 a) - c) sind die verschiedenen Schritte der Procrustes-Analyse für den LFPW-Datensatz visualisiert. In Unterabbildung a) ist zunächst die initiale Verteilung der Landmarken im Trainingsdatensatz zu sehen. Unterabbildung b) zeigt die Verteilung nach der Schwerpunktverschiebung zum Koordinatenursprung und dem Entfernen der translatorischen Variationen. Nach dem Ausrichten der Formen ist die Procrustes Analyse abgeschlossen. Das Ergebnis ist in Abb. 6.4 c) dargestellt. Die Basisform  $s_0$  entspricht der sich daraus ergebenden Durchschnittsform. Gleichzeitig sind die rigid Formvektoren  $s_{\text{rigid}}$  des Formmodells bestimmt.

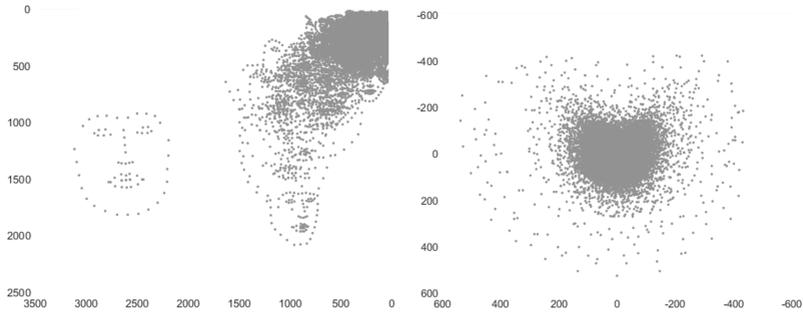
Im nächsten Schritt wird für die normierten Landmarken die Hauptkomponentenanalyse (PCA) durchgeführt. Ziel ist die Bestimmung der Eigenvektoren zur möglichst kompakten Beschreibung der nicht rigid Formtransformationen. Diese werden über die Aufstellung der Kovarianzmatrix und Lösen des Eigenwertproblems gewonnen. Die ersten 132 der nach der Größe der Eigenwerte sortierten Eigenvektoren entsprechen den nicht rigid Formvektoren  $s_{\text{nonrigid}}$ .

Es ergibt sich das trainierte Modell:

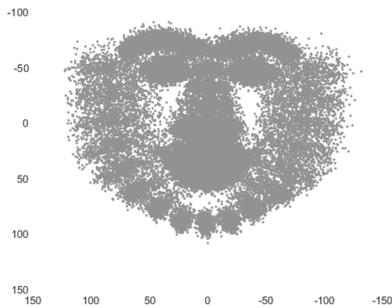
$$\mathbf{s} = \mathbf{s}_0 + \sum_{j=1}^n p_j \mathbf{s}_j \quad (6.1)$$

mit bekannter Basisform  $s_0$ , den Formvektoren  $s_i$ , sowie den Parametern  $p_i$  zur Verformung des Modells und Beschreibung der Form  $s$ . Dabei ent-

sprechen die Formvektoren  $s_{1-4}$  den rigiden Formvektoren  $s_{\text{rigid}}$  und die Formvektoren  $s_{5-136}$  den nicht rigiden Formvektoren  $s_{\text{nonrigid}}$ . Abbildung 6.4 d) veranschaulicht die Basisform (Mitte) und Beispiele für sich durch Kombination mit den Formvektoren ergebende Formvarianten.

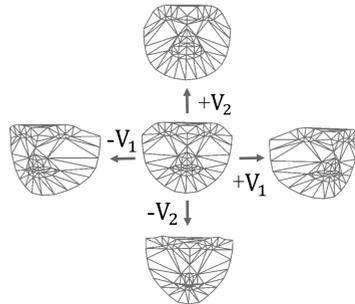


a) Initiale Verteilung der Landmarken



c) nach Procrustes Analyse

b) nach Schwerpunkt-Verschiebung



d) Veranschaulichung des Formmodells

Bild 6.4: Ausrichten der Trainingsdaten des Formmodells

## Initiales Schätzmodell

Soll in in einem unbekanntem Bild ein Gesicht mit Hilfe des Formmodells beschrieben werden, bedarf es zunächst einer ersten groben Schätzung der Landmark-Positionen, welche in den nachfolgenden Schritten optimiert werden kann. Diesem Zweck dient das „initiale Schätzmodell“. Hierfür nutzt das Modell zunächst die Face Detection, welche die Position des Gesichts in Form einer Bounding Box bereitstellt. Das Modell lernt während des Trainings die Basisform des Formmodells möglichst passend in die Bounding Box zu legen. Der Mittelpunkt der Bounding Box entspricht dem geschätzten Zentrum des Gesichts, die Größe impliziert die Skalierung. Die Analyse der Trainingsdaten zeigt jedoch eine häufige Abweichung zwischen dem geschätzten Zentrum des Face Detectors ( $X_{c,b}$ ,  $Y_{c,b}$  (rot markiert in Abbildung 6.5 a)) und dem Schwerpunkt der annotierten Landmarken ( $X_{c,s}$ ,  $Y_{c,s}$ ) (gelb markiert in Abbildung 6.5 a)). Für eine gute initiale Schätzung muss das Modell diese Verschiebung kompensieren. Insgesamt werden während des Trainings drei Offsets aus den Trainingsdaten gelernt:

- X-Koordinatenverschiebung des Schwerpunkts in Bezug auf das Zentrum der Bounding Box
- Y-Koordinatenverschiebung des Schwerpunkts in Bezug auf das Zentrum der Bounding Box
- Skalierung der Face Shape in Bezug auf die Größe der Bounding Box

Die Skalierung ist dabei als die euklidische Norm des Formvektors definiert. Die Offsets werden auf die Seitenlänge der Bounding Box normiert und können wie folgt ausgedrückt werden:

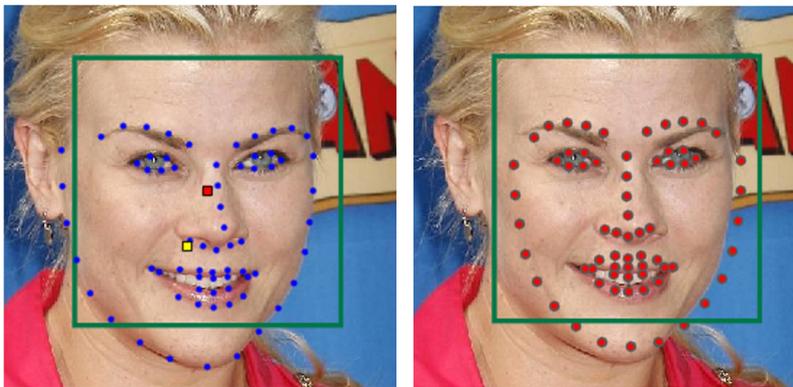
$$Offset_x = \frac{X_{c,s} - X_{c,b}}{L_b} \quad (6.2)$$

$$Offset_y = \frac{Y_{c,s} - X_{c,b}}{L_b} \quad (6.3)$$

$$Offset_{scale} = \frac{S_s}{L_b} \quad (6.4)$$

wobei  $S_s$  der Skalierung der Face Shape und  $S_b$  der Seitenlänge der Bounding Box entsprechen. Für jedes Trainingsbild ergibt sich so ein (1x3)-Offset-Vektor. Das initiale Schätzmodell basiert auf dem Mittelwert der Vektoren.

Für ein unbekanntes Gesicht in einem Bild ergibt sich folgender Prozess: Zunächst wird der Schwerpunkt der Face Shape über das Zentrum der Bounding Box des Face Detectors und die Verschiebungsparameter des initialen Schätzmodells berechnet. Anschließend wird die entsprechend skalierte Basisform um den Schwerpunkt gelegt, die initiale Position der Landmarken ist bestimmt (siehe Abbildung 6.5 b)). Die Offsetparameter des Schätzmodells berechnen sich jeweils relativ in Abhängigkeit der Seitenlänge  $L_b$  der Bounding Box.



a) Abweichung zw. Schwerpunkt der Landmarken u. Bounding Box Zentrum

b) Initiale Schätzung der Landmarken

Bild 6.5: Veranschaulichung zum initialen Schätzmodell

*\* Ursprungsbild entstammt LFPW-Datensatz*

## Regressionsmodell

Ziel des Regressionsmodells ist es, für ein unbekanntes Bild das Ergebnis des initialen Schätzmodells hinsichtlich einer möglichst exakten Schätzung der Landmarkspositionen zu optimieren. Der Trainingsdatensatz ergibt sich folglich zu  $D^1, \dots, D^N$ , mit  $D^t = \{\mathbf{d}^t, \mathbf{x}_*^t, \mathbf{x}_0^t\}$ . Dabei entspricht  $\mathbf{d}^t$  dem  $t$ -ten Bild,  $\mathbf{x}_*^t$  den zugehörigen Ground Truth Landmarken und  $\mathbf{x}_0^t$  den geschätzten Landmarkspositionen des initialen Schätzmodells. Um die Robustheit des zu erstellenden Regressionsmodells zu erhöhen, wird zunächst der bestehende Datensatz mittels Monte-Carlo-Sampling künstlich erweitert. Mit Hilfe des erstellten initialen Schätzmodells kann eine erste Schätzung der Landmarken vorgenommen werden. Es wird angenommen, dass die relativen Offsets zwischen Bounding Box und Gesichtslandmarken über die Trainingsdaten normalverteilt sind. Die Verteilungsparameter Mittelwert und Varianz können mittels Face Detection auf den Trainingsdaten erlernt werden. Auf diese Weise kann für jedes Bild des Trainingsdatensatzes mehr als ein Offset-Vektor bestimmt werden. Abbildung 6.6 veranschaulicht wie so zusätzliche Trainingsdaten für das Regressionsmodell erzeugt werden können.

Die erste Iteration des Trainingsprozesses für das Regressionsmodell ist in Abbildung 6.7 dargestellt. Da  $\mathbf{x}_*^t, \mathbf{x}_0^t$  für ein Bild des Trainingsdatensatzes gegeben sind, können die Formparameter der Ground Truth  $\mathbf{p}_*$  und des Schätzmodells  $\mathbf{p}_0$  über Gleichung 2.26 des Grundlagenkapitels bestimmt werden. Das Parameterinkrement  $\Delta\mathbf{p}_*$  ergibt sich zu  $\Delta\mathbf{p}_* = \mathbf{p}_* - \mathbf{p}_0$ . Gleichzeitig wird in einer definierten Region um die Gesichtslandmarke (Patch) das SIFT-Feature  $\boldsymbol{\varphi}_0$  extrahiert. So ergibt sich für jedes Trainingsbild  $D^t$  eine Gleichung  $\Delta\mathbf{p}_*^t = \mathbf{R}_0\boldsymbol{\varphi}_0^t + \mathbf{b}_0$  mit den zu bestimmenden Variablen  $\mathbf{R}_0, \mathbf{b}_0$ . Eine Möglichkeit  $\mathbf{R}_0, \mathbf{b}_0$  zu berechnen, besteht in der in der Definition als lineares Least-Square-Problem:

$$\min \sum_{D^t} \|\Delta\mathbf{p}_*^t - \mathbf{R}_0\boldsymbol{\varphi}_0^t - \mathbf{b}_0\|^2 \quad (6.5)$$

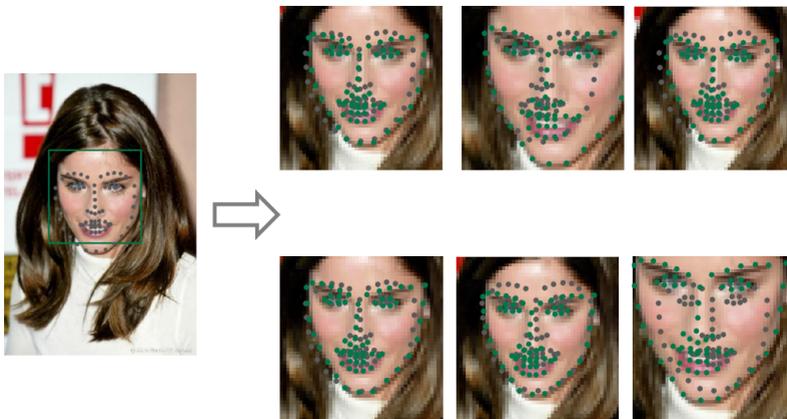


Bild 6.6: Monte-Carlo-Sampling  
 \* Ursprungsbild entstammt LFPW-Datensatz

Zur Lösung werden die Parameterinkremente und extrahierten SIFT-Features in Form zweier Matrizen neu angeordnet:

$$\Delta P_* = [\Delta p_*^1, \dots, \Delta p_*^N] \quad (6.6)$$

$$\Phi_0 = [\varphi_0^1, \dots, \varphi_0^N]; \quad (6.7)$$

Das Minimierungsproblem in 6.5 kann nun über folgende Gleichung gelöst werden:

$$[R_0, b_0] = \Delta P_* M^T (M M^T + \lambda I)^{-1} \quad (6.8)$$

mit  $M = [\Phi_0; \mathbf{1}]$  und  $I$  als Einheitsmatrix.  $\lambda$  ist ein Regularisierungsparameter für den Fall, dass  $M M^T$  schlecht konditioniert ist. Alternativ können die Parameterinkremente als lineare Funktionen der SIFT-Features betrachtet und das Regressionsmodell mit Hilfe der Support Vector Regression trainiert werden [36].

Nach Bestimmung von  $R_0$ ,  $b_0$  ist das Regressionsmodell in erster Iteration angelernt. Das Formparameterinkrement kann für jedes Trainingsbild

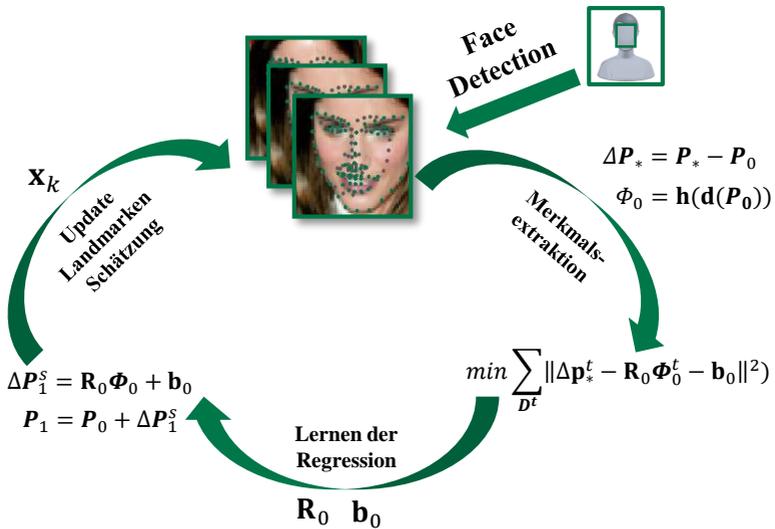


Bild 6.7: Anlernen des Regressionsmodells

über  $\Delta p_1^s = R_0 \Phi_0 + b_0$  berechnet werden. Mit Hilfe des Formparameterinkrements wird anschließend der Wert des aktuellen Formparameters aktualisiert  $p_1 = p_0 + \Delta p_1^s$ . Mit dem aktualisierten Parameterinkrement  $\Delta p_* = p_* - p_1$  und den SIFT-Features um die aktualisierten Landmarken  $\Phi_1 = h(d(p_1))$  kann ein neuer Trainingsdatensatz für die nächste Iteration generiert werden. Auf diese Weise wird iterativ eine Folge von Regressionsmodellen  $R_k, b_k$  mit  $k = 1, \dots, 4$  zwischen Parameterinkrement und SIFT-Features gelernt.

### Unterschied zur SDM

Nach Kapitel 3.2 wird das Fitting eines Modells für ein unbekanntes Bild (Sample-Bild) durch Minimierung der Fehlerfunktion erreicht, welche als Differenz der Appearance Parameter im Sample und in der Modellinstanz definiert ist. Der Hauptunterschied zur originalen SDM liegt in der Wahl

des Bezugssystems, in welchem die Fehlerfunktion definiert ist. In der originalen SDM wird das Bezugssystem des Sample-Bildes gewählt, in welches die Modellinstanz gewarpt wird. Die SIFT-Features werden im normalisierten Sample-Bild berechnet. Die entwickelte Methode berechnet die SIFT-Features hingegen nachdem das normalisierte Gesichtsbild auf die Modellinstanz gewarpt wurde. Es wird das Bezugssystem des Modells genutzt. In der originalen Methode erfolgt die Regression somit zwischen den Positionsinkrementen  $\Delta \mathbf{x}_* = \mathbf{x}_* - \mathbf{x}_0$  und den SIFT-Features, in diesem Ansatz hingegen zwischen den Formparameterinkrementen  $\Delta \mathbf{p}_* = \mathbf{p}_* - \mathbf{p}_0$  und den SIFT-Features. Die sich ergebenden unterschiedlichen Merkmalstypen zur Beschreibung der Appearance werden in der Literatur als Similarity Normalized Shape Features (SPTS) bei der SDM und Canonical Normalized Appearance (CAPP) bei der entwickelten Methode bezeichnet [89]. Folglich wird der Ansatz im weiteren Verlauf als „CAPP-SDM“ bezeichnet.

Die Abbildungen 6.8 und 6.9 zeigen das Vorgehen der SIFT-Feature-Extraktion mit verschiedenen Auflösungen und verdeutlichen den Unterschied. Wie erwähnt erfolgt die Extraktion der SIFT-Features  $\mathbf{h}(\mathbf{d}(\mathbf{p}))$  in Patches um die prädizierten Landmarken. Die roten Markierungen entsprechen der Prädiktion der Landmarken in jeder Iteration. In der ersten Iteration werden die SIFT-Merkmale aus einem niedriger aufgelösten Gesichtsbild extrahiert. In der zweiten und dritten Iteration erfolgt die Merkmalsextraktion in höher aufgelösten Bildern. Wie zu sehen ist, werden bei der CAPP-SDM neben den rigiden geometrischen Variationen, auch die lokalen, nicht-rigiden Variationen entfernt. Daher hat der Ansatz den Anspruch robuster gegenüber Einflüssen von Kopfbewegungen, Mimik und Sprechbewegungen zu sein.



Bild 6.8: Merkmalsextraktion bei der originalen SDM (SPTS)

*\*Ursprungsbild entstammt LFPW-Datensatz*

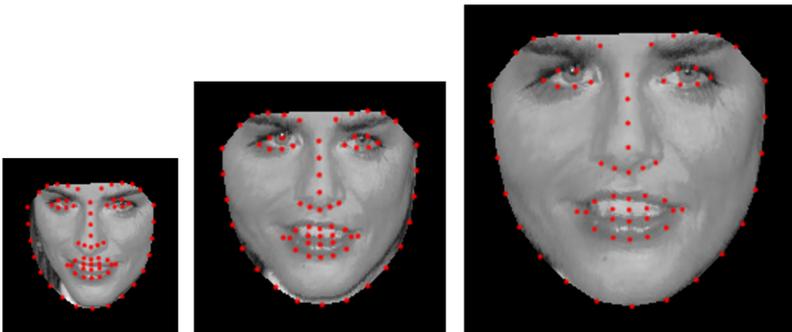


Bild 6.9: Merkmalsextraktion bei der CAPP-SDM

*\*Ursprungsbild entstammt LFPW-Datensatz*

## 6.1.2 Trackingprozess

Die Detektion der Landmarken in einem unbekanntem Bild folgt dem in Abbildung 6.10 dargestellten Prozess. Die Face Detection liefert die Position des Gesichts, das initiale Schätzmodell stellt die initialen Positionen der Landmarken bereit. In Patches um diese Landmarken werden entsprechend dem CAPP-SDM-Ansatz die SIFT-Features extrahiert. Mit Hilfe des trainierten Regressionsmodells werden anschließend die Formparameter über

$\mathbf{P}_k = \mathbf{P}_{k-1} + \mathbf{R}_{k-1} \boldsymbol{\Phi}_{k-1} + \mathbf{b}_{k-1}$  aktualisiert. Auf Basis der aktualisierten Formparameter wird das Gesicht entsprechend rekonstruiert und die Position der Landmarken angepasst. Die aktualisierten Landmarken dienen dann als Eingang der nächsten Iteration der Regression. Aus Gründen der Performanz wird die Anzahl der Iterationen auf vier festgelegt. Das Face Alignment für ein Bild ist abgeschlossen.

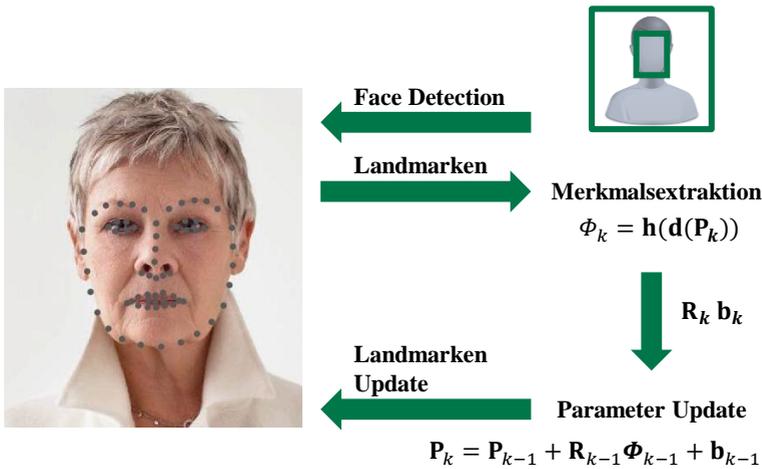


Bild 6.10: Prozess zum Update der Landmarken

\* Ursprungsbild entstammt LFPW-Datensatz

Für ein Live-PPGI-System gilt es Gesichter in Bildsequenzen kontinuierlich zu tracken. Die initiale Detektion des Gesichts mittels Face Detection mit Hilfe des Viola-Jones-Algorithmus ist dabei vergleichsweise rechenintensiv. Die Anwendung auf jedes Frame wäre somit ineffizient. Aus diesem Grund wird alternativ die letzte Schätzung der Landmarken als Eingang für den Alignment-Prozess genutzt. Erst wenn die Konfidenz der letzten Schätzung einen vordefinierten Schwellwert unterschreitet, wird die Face Detection und das initiale Schätzmodell zur Re-Detection genutzt. Um PPGI-

Signale auch bei starker Drehung des Kopfes extrahieren zu können, wurde ein Frontal- und Profil-Modell umgesetzt. Der Wechsel zwischen den Modellen erfolgt auf Basis der Schätzung der Kopfpose (*engl.*: „Head Pose“) nach der initialen Positionsschätzung. Abbildung 6.11 fasst den algorithmischen Ablauf des Face-Tracking-Moduls nochmals zusammen.

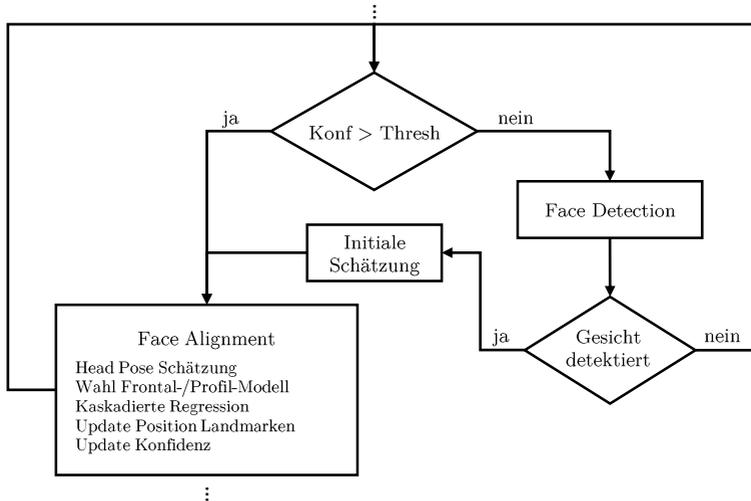


Bild 6.11: Ablaufdiagramm des Face-Tracking-Moduls

## 6.2 Fixed-Region-Ansatz

Mit Hilfe der Position der 68 vom Face-Tracking-Modul gelieferten Landmarken, können im Gesicht einer Person geeignete Messregionen für die Herzfrequenzmessung beschrieben werden. Die Verbindungen der Landmarken markieren ein trianguliertes Mesh, welches zahlreiche Subregionen im Gesicht definiert. Eine Möglichkeit zur Definition der ROI besteht in der Vorauswahl und Kombination geeigneter Subregionen. Die Ergebnisse der Analyse unterschiedlicher Messregionen im Gesicht aus Abschnitt

4.7.2 zeigen, dass sich die beste Signalqualität für die zentralen Stirn- und Wangenbereiche ergibt. Diese Regionen werden somit fest als ROI definiert (siehe Abbildung 6.12). Das Rohsignal ergibt sich durch Mittelung der Pixel der ausgewählten Subregionen und ist Eingang des nachfolgenden Signalverarbeitungsblocks. In der Live-Anwendung mit GUI lässt sich die Vorauswahl der Region per Klick auf den entsprechenden ROI-Bereich anpassen. Dies kann beispielsweise bei Verdeckung der Stirnregion durch Haare notwendig sein.

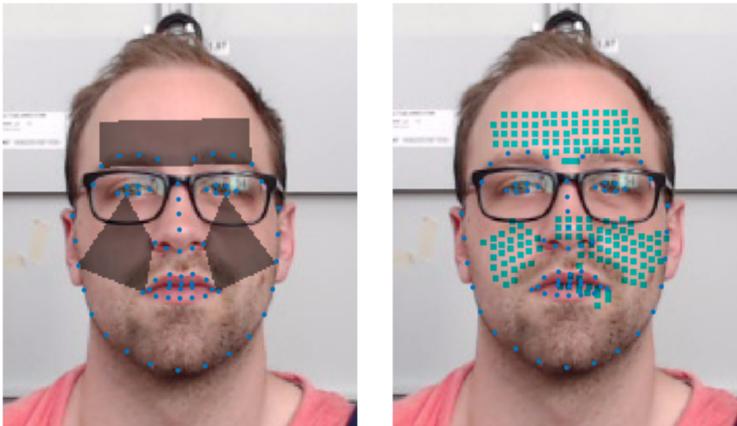


Bild 6.12: Vergleich Fixed-Region-Ansatz (li.) und Multi-Patch-Ansatz (re.)

### 6.3 Adaptiver Multi-Patch-Ansatz

Der im vorigen Abschnitt beschriebene Fixed-Region-Ansatz besitzt den Nachteil, nicht automatisiert auf eine mögliche Überdeckung der Messregion reagieren zu können. Zudem kann es innerhalb einer Messregion Bereiche geben die unterschiedlich starke PPGI-Signalanteile besitzen, beispielsweise auf Grund von Sprechbewegungen, Mimiken oder Blinzeln. Als

Lösung wurde der Multi-Patch-Ansatz entwickelt, bei dem kleine Bereiche im Gesicht, sogenannte Patches, als einzelne Sensoren definiert werden. Auf Basis verschiedener Kriterien für die vom Sensor gelieferten Signalqualität, fließen die jeweiligen Signale in die weitere Berechnung ein oder werden ignoriert. Ein ähnliches Prinzip wird auch in [158] beschrieben. Im Unterschied zu diesem Ansatz werden Anzahl und Position der Sensoren nicht über den optischen Fluss, sondern mit Hilfe der Basisform des Formmodells aus Abschnitt 6.1.1 definiert und rauschabhängig selektiert.

### 6.3.1 Definition der Subregionen

Die Pixelzahl im Gesichtsbereich ist aufgrund von Kopfbewegungen und mimischen Veränderungen nicht konstant. Um den Einfluss einer variierenden Zahl an Gesichtspixeln zu reduzieren, werden die Pixelsensoren innerhalb der Basisform des Formmodells definiert. Die Basisform ist grundsätzlich auf 1 normiert. Für den beschriebenen Ansatz wird die Basisform entsprechend hochskaliert, bis diese 490 Pixel umschließt (siehe Abbildung 6.13). Dabei definiert jedes Pixel innerhalb der skalierten Basisform des Gesichtsmodells das Zentrum eines Einzelsensors.

Die Position eines Sensors für ein unbekanntes Gesichtsbild kann durch affines Warping bestimmt werden. Dazu wird erneut die Mesh-Struktur des Formmodells, in Abbildung 6.13 grün dargestellt, genutzt. Nach dem Alignment der Landmarken können die affinen Transformationsparameter über die Eckpunkte einer Subregion des Mesh sowohl in der Basisform, als auch im Gesichtsbild berechnet werden. Die Position der Pixelsensoren wird dann durch Anwendung der affinen Transformation der jeweiligen Subregion mit zugehörigen Parametern bestimmt. Auf diese Weise kann die Zahl der Pixelsensoren konstant gehalten werden, auch wenn einige Pixel durch Kopfdrehungen verdeckt sind. Gleichzeitig ist sie unabhängig von der Auflösung.

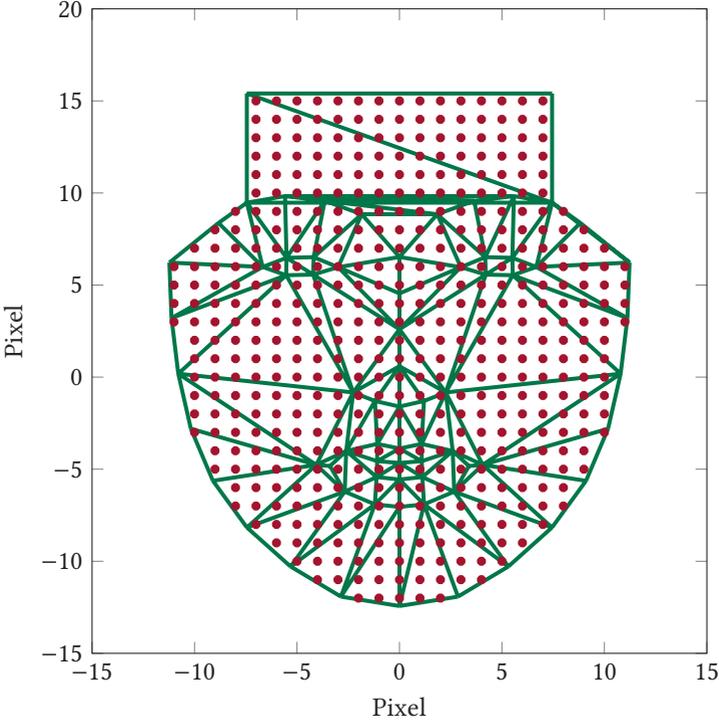


Bild 6.13: Definition der Pixelsensoren in der Basisform

Aufgrund des geringen SNR bei PPGI ist eine Mindestgröße der Einzelsensoren erforderlich. Entsprechend werden Patches um die Zentren definiert. Eine ausreichend große Fläche der Patches verringert das Sensorrauschen, gleichzeitig kann es zu räumlichen Überschneidungen zwischen benachbarten Pixelsensoren kommen und so die Trennung zwischen artefaktbehafteten Bereichen und solcher mit guter Signalqualität verhindert werden. Als empirisch ermittelten Trade-Off wird die Größe der Patches zu  $13 \times 13$  Pixel definiert (bezogen auf eine Auflösung von  $640 \times 480$ ).

### 6.3.2 Selektion der Subregionen

Die Selektion der Einzelsensoren erfolgt anhand zweier Kriterien, dem Haut-Konfidenzindex und der Hochfrequenz-Rauschenergie. Grundlage für die Schätzung des Haut-Konfidenzindex ist die Segmentierung der Haut (*engl.*: „Skin Segmentation“) im detektierten Gesicht. Hierfür wird nach dem Start des Systems ein One-Class-SVM-Klassifikator (OC-SVM) trainiert. Um die Onlinefähigkeit des Systems zu gewährleisten erfolgt der Trainingsprozess einmalig. Als Trainingsdatensatz dienen die ersten Frames nach initialer Gesichtsdetektion. Die normalisierten RGB- und YCbCr-Intensitäten werden als Merkmale genutzt. Auf diese Weise ergibt sich zusätzlich zu jedem extrahierten Gesicht eine Haut-Segmentierungsmatrix. Für die Selektion wird das Zeitfenster der letzten 1,2 Sekunden der Einzelsensoren analysiert. Der Haut-Konfidenzindex wird bestimmt als der zeitliche Anteil des Signalauschnitts, in dem sich der Sensor in einem segmentierten Hautbereich befunden hat. Wenn ein Pixelsensor mehr als 70% Hautpixel in über 80% der Frames eines Signalauschnitts enthält, wird der Pixelsensor für die BVP-Signalextraktion genutzt, andernfalls verworfen.

Als zweites Selektions-Kriterium wird das Hochfrequenzrauschen der Einzelsensorsignale im selben Zeitfenster verwendet. Aufgrund der niedrigen spektralen Frequenzauflösung bedingt durch das kurze Beobachtungszeitfenster kann keine ausreichende Abgrenzung von niederfrequenter Störung und Nutzsignal erfolgen. Gleichzeitig zeigen Artefakte durch Kopfbewegungen oder Lichtänderungen einen klaren Anstieg hochfrequenter Störanteile. Aus diesem Grund wird die Hochfrequenz-Rauschenergie der Spektralkomponenten mit  $f_n > 4Hz$  außerhalb des Nutzsignalbereichs im Zeitraum der letzten 1,2 Sekunden als Kriterium berechnet. Abbildung 6.14 zeigt beispielhaft für einen Signalauschnitt einer Datenbankmessung die Verteilung der normierten Rauschenergie der 490 Einzelsensoren ohne und mit Einfluss von Kopfbewegungen. Ohne Kopfbewegungen konzentriert sich die Verteilung stark im unteren Rauschenergiebereich. Unter

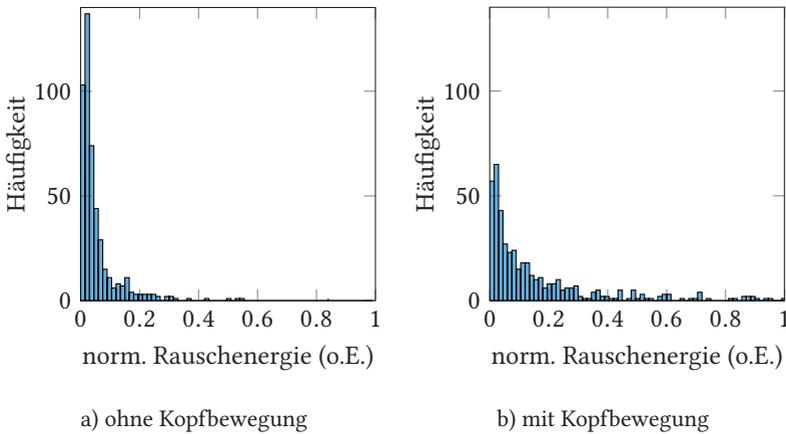


Bild 6.14: Häufigkeitsverteilung des Hochfrequenzrauschens

Einfluss von Kopfbewegungen treten vermehrt Rauschenergien im mittleren und hohen Bereich auf. Der Schwellwert für die Selektion eines Einzelsensors wird nach Analyse der Datenbank zunächst bei 0,25 der normierten Rauschenergie festgelegt. Da sich die Rauschenergie abhängig von Kamera und Umgebungsbedingungen ändert, gilt es den Schwellwert für das jeweilige Setting neu zu bestimmen.

Am Ende des zweischrittigen Prozesses ergibt sich eine Auswahl von Pixelsensoren für jedes Frame, deren gemittelter Pixelwert als Eingang für die weitere BVP-Signalverarbeitung dient. Abbildung 6.15 zeigt beispielhaft die Selektion von Einzelsensoren. Sensoren im Bereich der Augenbrauen, der Brille und der Gesichtskontur werden größtenteils ausgeschlossen.

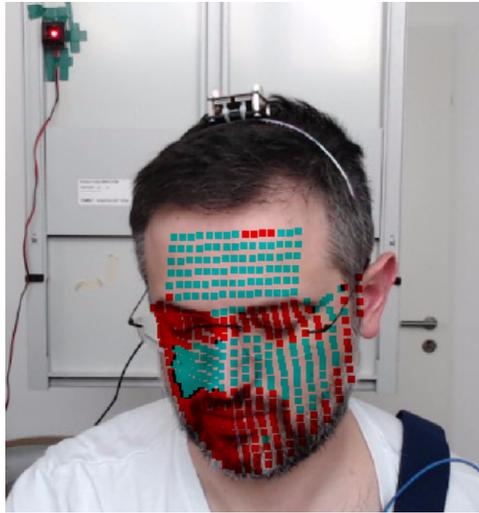


Bild 6.15: Pixelsensorpatches im Gesichtsbild.

## 6.4 Vergleich der Ansätze

Für den Vergleich der beiden Ansätze zur Extraktion des Rohsignals auf Basis des Outputs des Face-Tracking-Moduls wird, entsprechend der Auswertung in Kapitel 4.7, das SNR für die verschiedenen Szenarien bestimmt. Die Bewertung anhand des SNR erfolgt dabei für das extrahierte BVP-Signal anstelle der RGB-Rohsignale, um die Auswirkungen auf das zu rekonstruierende Signal besser analysieren zu können. Als Algorithmus zur Extraktion des BVP-Signals wird zunächst der in [157] vorgestellte POS-Algorithmus als State-of-the-Art-Ansatz genutzt. Auf die Algorithmen zur Signalextraktion wird im Folgekapitel eingegangen.

Es zeigt sich, dass bei Nutzung des adaptiven Multi-Patch-Ansatzes zur Wahl der ROI eine Verbesserung für sämtliche Szenarien erzielt werden kann. Schon für Szenario 101 unter Ruhebedingungen und bei konstanten Lichtverhältnissen verbessert sich das SNR, beispielsweise auf Grund des

Tabelle 6.1: Vergleich der SNR-Werte für die verschiedenen Szenarien in dB

<b>Szenario</b>	<b>Fixed-Region</b>	<b>Multi-Patch</b>
101	2,92	3,62
102	-1,75	-0,79
103	-0,27	-0,19
104	2,27	2,82
201	-2,83	-1,97
202	0,96	1,82
203	-0,55	0,15
204	-0,99	0,23

Ausschlusses möglicher durch Bart oder Brille verdeckter Bereiche innerhalb der ROI. Eine deutliche Verbesserung der Absolutwerte ergibt sich für die Szenarien 102 und 104. Dabei werden insbesondere Patches mit starkem Hochfrequenzrauschen, welches sich durch die künstliche Beleuchtung (vgl. Abschnitt 4.4) oder die abrupten Helligkeitsänderungen im jeweiligen Szenario ergibt, durch den Selektionsalgorithmus eliminiert. In den Bewegungsszenarien 201-204 führt der Ausschluss von Non-Skin-Pixeln, deren Zahl sich in Folge der Kopfbewegungen und deren Einfluss auf die Genauigkeit des Face-Tracking-Moduls deutlich erhöhen kann, ebenfalls zu einem Anstieg der SNR-Werte.

Abbildung 6.16 veranschaulicht die Auswirkung anhand der Translationsbewegung eines Probanden in Szenario 203. Beim Fixed-Region-Ansatz enthält ein kleiner Bereich der ROI (linke Wange des Probanden) auf Grund von Ungenauigkeiten des Face-Trackers in Folge der Bewegung keine Hautpixel. Zudem werden Pixel, in denen das SNR durch die Bewegung und den Einfluss auf das reflektierte PPGI-Signal stark abnimmt, nicht eliminiert.

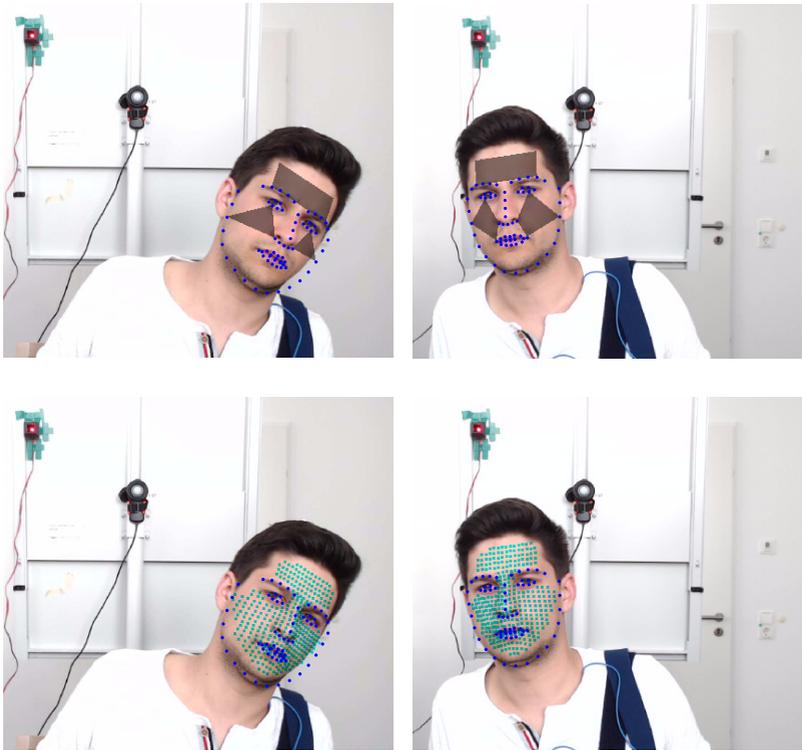


Bild 6.16: Vergleich der ROI des Fixed-Region- (o.) und Multi-Patch-Ansatzes (u.) bei translatorischer Kopfbewegung

Abbildung 6.17 zeigt die zugehörigen rekonstruierten BVP-Signale für diesen Ausschnitt. Zusätzlich ist die Kopfbewegung durch die Änderung der Landmark-Koordinaten dargestellt. Für den Frame-Bereich 90-200 ist eine klare Störung des BVP-Signals für den Fixed-Region-Ansatz zu erkennen, wohingegen sich für den Multi-Patch-Ansatz ein rekonstruiertes BVP-Signal ohne sichtbaren Einfluss der Störung ergibt. Grundsätzlich ist für den Multi-Patch-Ansatz durch die höhere Anzahl an Pixeln in der ROI

und den Ausschluss von Patches mit hohem HF-Rauschen eine Tiefpasswirkung mit niedrigerem Rauscheinfluss zu erkennen.

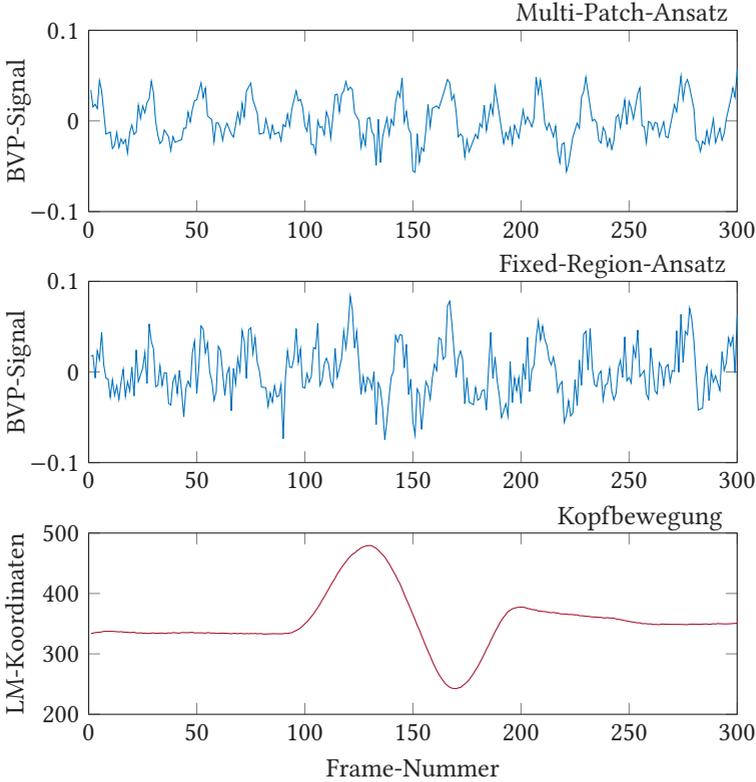


Bild 6.17: Vergleich des BVP-Signals bei fester und adaptiver Selektion der ROI

## 6.5 Diskussion und Fazit

Inhalt dieses Kapitels war die Konzeption und Umsetzung eines geeigneten Verfahrens zur automatisierten Detektion einer geeigneten Messregion (ROI) innerhalb des Gesichts einer Person und zur Extraktion des RGB-Rohsignals. In diesem Zusammenhang wurde ein Face-Tracking-Modul zum Landmarken-basierten Verfolgen des Gesichts vorgestellt. Auf dieser Basis wurden zwei Verfahren zur Regionenauswahl präsentiert, mit fester Definition der ROI (Fixed-Region-Ansatz) und adaptiver Regionenauswahl (Multi-Patch-Ansatz). Dabei konnten die Vorteile der adaptiven Wahl mittels eines Hautindex- und Hochfrequenz-Rauschkriteriums anhand der Auswertung des SNR für die verschiedenen Szenarien der Labordatenbank gezeigt werden. Insbesondere für den Einsatz des PPGI-Systems in bewegten Szenarien mit wechselnden Umgebungsbedingungen ist die Kombination aus präzisiertem landmarkenbasierten Tracking und adaptiver Regionenauswahl eine wichtige Voraussetzung für eine erfolgreiche Extraktion des BVP-Signals.



# 7

## Algorithmus zur Extraktion des Messsignals und Bestimmung der Herzfrequenz

IM RAHMEN DIESES KAPITELS werden die verschiedenen Schritte der Signalverarbeitung beschrieben, um aus den gemittelten RGB-Rohdaten der ROI als Ergebnis des Bildverarbeitungsblocks, das BVP-Signal möglichst genau zu rekonstruieren und daraus die Herzfrequenz zu berechnen. Bezüglich der algorithmischen Verarbeitungskette entspricht dies dem zweiten und dritten Prozessschritt der Analysekomponente (vgl. Abb. 7.1). Unter Berücksichtigung von Anforderung [A-AF-05](SV) des Systemkonzepts umfasst dies für die Rekonstruktion des BVP-Signals zunächst einen

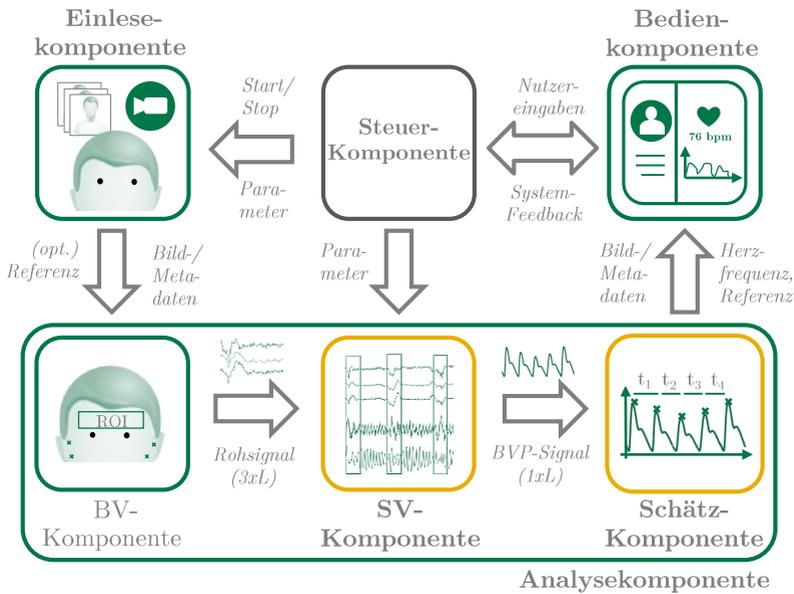


Bild 7.1: Softwarearchitektur des Gesamtsystems - Signalverarbeitung

Algorithmus zur Trennung der Signal- und Störkomponenten, speziell unter Einfluss von Kopfbewegungen (Kap. 7.1). Inhalt des darauffolgenden Unterkapitels 7.2 ist die Bestimmung der Herzfrequenz auf Basis des extrahierten BVP-Signals, wobei unter anderem auf die Vorzüge einer Schlag-zu-Schlag-Detektion der Pulswelle im Zeitbereich eingegangen wird [A-AF-06](SC). Eine weitere Möglichkeit, die Signalqualität hinsichtlich der Peak-Detektion zu optimieren, besteht in der gezielten Filterung von Störfrequenzkomponenten, welche in den Nutzfrequenzbereich von  $[0.6 - 4\text{ Hz}]$  fallen. Im Rahmen der Arbeit wurde zu diesem Zweck ein adaptiver Bandpassfilter entwickelt, welcher in Abschnitt 7.2.1 vorgestellt wird. Die Methode dient der Erfüllung von Anforderung [A-AF-04](SV) aus Kapitel 5.2.2. Abschließend werden die entwickelten Ansätze anhand der aufgezeichneten Datenbasis aus Kapitel 4 evaluiert.

## 7.1 Extraktion des BVP-Signals

In Kapitel 3.3.4 wurden bereits mehrere Ansätze zur Trennung der verschiedenen Komponenten des innerhalb der ROI erfassten RGB-Signals eingeführt. Die ICA- und PCA-basierten Ansätze nutzen den Umstand aus, dass der Beitrag der Blutvolumenpuls-induzierten Komponente und der Störkomponenten in den einzelnen Farbkanälen, auf Grund des variierenden Absorptionsgrades von Hämoglobin mit der Wellenlänge, unterschiedlich stark ist. Spezielle Annahmen über die Verteilung der Komponenten in den RGB-Signalen werden bei diesen Blind Source Separation (BSS)-Ansätzen nicht getroffen. Chrom [34] und POS [157] hingegen sind vorwiegend modellbasierte Ansätze, die Vorwissen über die Verteilung der Signalkomponenten in den RGB-Farbkanälen nutzen, um den Prozess der Komponententrennung zu kontrollieren. Der im Rahmen der Arbeit entwickelte Algorithmus baut auf der Grundidee dieser Algorithmen - der gezielten Projektion der Signale in eine Ebene im Farbraum (vgl. Kapitel 2.3.1) - auf, erweitert diese jedoch hinsichtlich der Robustheit gegenüber Bewegungsartefakten mit dem Ziel des Einsatzes in stark artefaktbehafteten Realeinsatzszenarien, wie der Fahrerbeobachtung im Automobil.

Die Beschreibungen in [34] und [157] basieren unter anderem auf dem in Kapitel 2.3.1 vorgestellten dichromatischen Reflexionsmodell. In diesem Zusammenhang führen Wang et al. in [157] die Begriffe der spiegelnden Reflexion („specular“) für den an der Oberfläche reflektierten Teil des Lichts und der diffusen Reflexion („diffuse“) für den in den Körper eindringenden und vom Körperinneren reflektierten Teil ein. Streng genommen ist die Bezeichnung nicht ganz korrekt, da diese Begriffe aus dem Bereich der Optik rein geometrisch definiert sind. Dabei entspricht die Spiegelreflexion der gerichteten Reflexion an glatten Oberflächen, die diffuse Reflexion der Streuung an rauen Oberflächen. Gemeint ist jedoch der unterschiedliche Einfluss auf die spektrale Zusammensetzung des reflektierten Lichts für an der Oberfläche und im Körper reflektiertes Licht, der entscheidende Ef-

fekt für die entwickelten Ansätze. Im Zusammenhang mit dem dichromatischen Modell haben sich die Begriffe der Oberflächen- und Körperreflexion etabliert, welche im weiteren Verlauf der Arbeit entsprechend verwendet werden.

Nach Wang et. al [157] lässt sich das Signal, welches sich bei Mittelung über einen Hautbereich ergibt, mathematisch näherungsweise mit Gleichung 7.1 beschreiben (Herleitung siehe Kap. 2.3.2). Danach gilt:

$$C(t) = \underbrace{I_0 \cdot (1 + i(t))}_{\text{Beleuchtung}} \cdot \underbrace{(\mathbf{u}_c \cdot c_0 + \mathbf{u}_s \cdot s(t) + \mathbf{u}_p \cdot p(t))}_{\text{Hautreflexion}} \quad (7.1)$$

mit  $I_0$  als mittlere Beleuchtungsstärke,  $c_0$  als Reflexionsstärke,  $\mathbf{u}_c$ ,  $\mathbf{u}_s$  als Farbvektoren von Haut und Lichtspektrum der Quelle, sowie  $\mathbf{u}_p$  als relative, blutvolumenpulsbezogene Stärke in den Farbkanälen. Dabei modelliert  $I_0 \cdot (1 + i(t))$  den stationären und zeitveränderlichen Teil der Beleuchtung. Der Term  $\mathbf{u}_c \cdot c_0$  beschreibt den stationären Anteil der Oberflächen- und Körperreflexion.  $\mathbf{u}_s \cdot s(t)$  entspricht dem zeitveränderlichen Teil der Oberflächenreflexion, welcher hauptsächlich durch Bewegung hervorgerufen wird. Der zeitveränderliche Anteil der durch die Änderung des Blutvolumenpulses hervorgerufen wird, wird durch  $\mathbf{u}_p \cdot p(t)$  beschrieben. Das Quantisierungsrauschen der Kamera wird in der Beschreibung auf Grund der angenommenen Mittelung des Signals innerhalb der ROI vernachlässigt. Nach Umformulierung der Gleichung 7.1 und Streichen der hinsichtlich ihres Beitrags zur Signalamplitude vernachlässigbar kleinen Produkte der AC-Komponenten ( $i(t) \cdot p(t)$  und  $i(t) \cdot s(t)$ ) ergibt sich:

$$C(t) = \underbrace{\mathbf{u}_c \cdot I_0 \cdot c_0}_{\text{DC-Anteil}} + \underbrace{\mathbf{u}_c \cdot I_0 \cdot c_0 \cdot i(t)}_{\text{Intensitätsvariationen}} + \underbrace{\mathbf{u}_s \cdot I_0 \cdot s(t)}_{\text{Oberflächenreflexion}} + \underbrace{\mathbf{u}_p \cdot I_0 \cdot p(t)}_{\text{BVP-induziert}} \quad (7.2)$$

Der AC-Anteil des Signals entspricht somit näherungsweise der Summe folgender drei Quellsignalkomponenten:

- einer *lichtintensitätsabhängigen Komponente*, welche sich abhängig von der Beleuchtungsstärke der Lichtquelle oder dem Abstand Lichtquelle - Objekt - Kamera ändern kann.
- dem *AC-Anteil der Oberflächenreflexion (ACOR)*, welcher sich durch die Änderung der Winkel zwischen Lichtquelle - Objekt - Kamera ändern kann.
- dem gesuchten *blutvolumenpulsinduzierten Anteil (BVP)*, dem Teil der Körperreflexion, der entsprechend der Herzfrequenz variiert.

Dies lässt die Vermutung zu, dass sich das Mischsignal durch geeignete, lineare Projektion in die drei Quellsignale zerlegen lässt. Für eine geeignete Projektion gilt es nach [34] und [157] das RGB-Rohsignal zunächst zeitlich zu normieren, was über die Division durch das zeitliche Mittel im Analysezeitintervall erreicht wird. Anschließend wird der DC-Anteil subtrahiert, es ergibt sich das zeitlich normierte und mittelwertfreie RGB-Rohsignal  $\tilde{C}_n(t)$ :

$$\tilde{C}_n(t) = \mathbf{1} \cdot i(t) + \mathbf{N} \cdot \mathbf{u}_s \cdot I_0 \cdot s(t) + \mathbf{N} \cdot \mathbf{u}_p \cdot I_0 \cdot p(t) \quad (7.3)$$

mit Normierungsmatrix  $N = \overline{C}(t)^{-1}$  und  $\mathbf{1}$  als Vektor  $(1,1,1)^T$ .

Im CHROM-Algorithmus werden für  $\tilde{C}_n(t)$  zwei orthogonale Chrominanz-Signale definiert, mit dem Ziel auf diese Weise die ACOR-Komponente zu unterdrücken und die Dimensionalität des Demixing-Prozesses zu reduzieren:

$$X(t) = \tilde{C}_n^R(t) - \tilde{C}_n^G(t) \quad (7.4)$$

$$Y(t) = 0.5 \cdot \tilde{C}_n^R(t) + 0.5 \cdot \tilde{C}_n^G(t) - \tilde{C}_n^B(t) \quad (7.5)$$

Das Bilden der Chrominanzsignale kann dabei als Projektion von  $\tilde{C}_n(t)$  mit einer  $(2 \times 3)$ -Projektionsmatrix  $\mathbf{P} = \begin{pmatrix} 1 & -1 & 0.5 \\ 0.5 & 0.5 & -1 \end{pmatrix}$  angesehen werden. Nach Korrektur der Projektionsmatrix mit einem empirisch ermittelten Standard-Haut-Vektor  $\mathbf{u}_{skin} = [0.77, 0.51, 0.38]^T$ , entspricht deren Anwendung ei-

ner Projektion auf eine Ebene, die orthogonal zur Richtung der ACOR-Variationen liegt. Dies ist mit dem Ziel verbunden, diese Variationen in den projizierten Signalen  $S_1(t)$  und  $S_2(t)$  größtenteils zu eliminieren. Dabei wird die Richtung der Projektionsvektoren auf der Ebene so gewählt, dass sich die bewegungsinduzierten Intensitätsvariationen auf beide Signale gleich auswirken, die BVP-Variationen unterschiedlich. Durch Subtraktion der Chrominanz-Signale ergibt sich eine Schätzung von  $p(t)$  zu  $\hat{p}(t) = S_1(t) - \alpha \cdot S_2(t)$ , mit  $\alpha = \frac{\sigma(S_1)}{\sigma(S_2)}$  als Quotient der Standardabweichungen. Durch das als „Alpha-Tuning“ bezeichnete Verfahren, addieren sich die BVP-Komponenten auf, die bewegungsinduzierten Komponenten werden entfernt.

Im POS-Algorithmus wird eine Projektion mit veränderten Vorgaben für die Projektionsmatrix angewendet. Hier erfolgt die Projektion auf die Ebene orthogonal zur Richtung der Intensitätsschwankungen, da diesen der größte Störeinfluss zugeordnet wird. Als Richtung wird der Vektor  $\mathbf{1} = [1,1,1]^T$  gewählt, basierend auf der Annahme, dass sich die Intensitätsvariationen, etwa in Folge der Änderung des Abstands zwischen Lichtquelle, Objekt und Kamera, auf alle Farbkanäle gleich auswirken. Die Intensitätsschwankungen verteilen sich somit hauptsächlich entlang der Ebenen-Normalen  $\mathbf{1}$  und sollen durch die Ebenen-Projektion, ähnlich wie die ACOR-Variationen bei CHROM, eliminiert werden. Die AC-Signalamplituden bei Projektion auf die Ebene orthogonal zum Vektor  $\mathbf{1}$  sind somit in erster Linie der ACOR- und BVP-Komponente zuzuordnen. Die Analysen in [157] zeigen, dass die AC-Signalamplituden der beiden Komponenten unterschiedlich in der Ebene verteilt sind. Aus dieser Erkenntnis ergibt sich die Vorgabe, zwei orthogonale Projektionsvektoren zu finden, sodass die zeitlich normalisierten Rohsignale  $\tilde{C}_n(t)$  in Richtungen auf der Ebene projiziert werden, für die die BVP-Komponenten in der Analyse ein positives Vorzeichen aufweisen und „in Phase“ liegen, die ACOR-Komponenten unterschiedliche Vorzeichen und somit gegensätzliche Phasen besitzen. Auf diese Weise sollen die BVP-Anteile verstärkt

und die ACOR-Anteile unterdrückt werden. Mit dieser Vorgabe wird die Projektionsmatrix wie folgt gewählt:

$$P_{pos} = \begin{bmatrix} 0 & 1 & -1 \\ -2 & 1 & -1 \end{bmatrix} \quad (7.6)$$

Das finale Projektionssignal ergibt sich erneut durch Alpha-Tuning, wobei auf Grund der „in Phase“ liegenden BVP-Komponenten, eine Addition der projizierten Signale  $\hat{p}(t) = S_1(t) + \alpha \cdot S_2(t)$  erfolgt.

### 7.1.1 Projektion unter Einfluss von Artefakten

Der im Rahmen dieser Arbeit entwickelte Algorithmus zum Trennen der Signalkomponenten nutzt ebenfalls das Prinzip der Projektion. Jedoch zeigen Analysen auf Basis der Datenbankaufnahmen, dass die getroffenen Annahmen in [157] zur Trennung der Komponenten, je nach Bewegungsart und -intensität, nicht immer gegeben sind. Dies soll nachfolgend anhand eines Beispiels gezeigt werden. Für die Verteilung der jeweiligen Komponente auf die Farbkanäle wurde in [157] folgendes angenommen:

$$\begin{aligned} G_p &> B_p > R_p \\ B_s &> G_s > R_s, \end{aligned} \quad (7.7)$$

wobei  $G_p, B_p, R_p$  dem jeweiligen Anteil an der BVP-Komponente und  $B_s, G_s, R_s$  an der ACOR-Komponente entspricht. Mit diesen Annahmen besitzt die BVP-Komponente bei Projektion mit Matrix  $P_{pos}$  folgende Eigenschaften:

$$\begin{aligned} G_p - B_p &> 0 \\ -2R_p + G_p + B_p &> 0 \end{aligned} \quad (7.8)$$

Für die ACOR-Komponente gilt hingegen:

$$\begin{aligned} G_s - B_s &< 0 \\ -2R_s + G_s + B_s &> 0 \end{aligned} \quad (7.9)$$

Gelten diese Annahmen wird nach der Projektion und Anwendung des Alpha-Tunings die BVP-Komponente verstärkt und die ACOR-Komponente abgeschwächt.

Ein Beispiel, für welches die Annahmen in 7.7 gegeben sind, wird in Abbildung 7.2 gezeigt. Abgebildet ist das Frequenzspektrum von  $\tilde{C}_n(t)$  bei langsam ausgeführter Kopfbewegung von links nach rechts. Die Frequenzamplitude wird dabei hauptsächlich durch die ACOR-Komponente bestimmt. Es gilt:  $B_s > G_s > R_s$ . Jedoch wird diese Annahme nicht immer er-

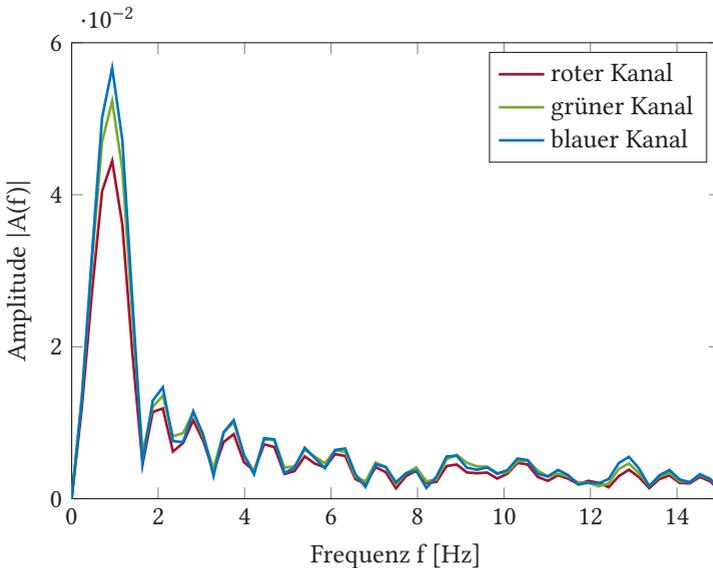


Bild 7.2: Frequenzspektrum von  $\tilde{C}_n(t)$  bei seitlicher Kopfbewegung

füllt. Abbildung 7.3 zeigt ein Gegenbeispiel bei Auf- und Abbewegung des Kopfes. Dabei ist zu sehen, dass die ACOR-Komponente des blauen Kanals

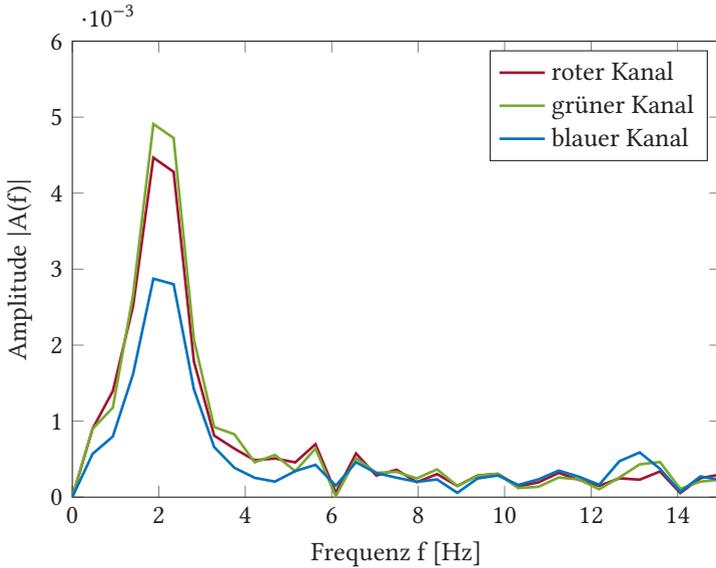


Bild 7.3: Rohsignal-Spektrum bei Auf- und Abbewegung des Kopfes

die geringste Energie aufweist. Für das gegebene Beispiel gilt  $B_s < R_s < G_s$  und somit  $(G_s - B_s > 0)$ . Die Komponenten sind folglich „in Phase“. Nach Anwendung des Alpha-Tuning würden die ACOR-Komponenten im POS-Algorithmus verstärkt anstatt gedämpft werden.

Speziell bei stark artefaktbehafteten Signalen  $\tilde{C}_n(t)$ , sei es durch starke Kopfbewegungen oder sich ändernde Lichtverhältnisse, wie sie beim Einsatz in mobilen Umgebungen wie dem Automobil häufig vorkommen, kann ein primär modellbasierter Ansatz eine erfolgreiche Rekonstruktion des BVP-Signals nicht immer ermöglichen. Als Lösungsansatz wurde daher eine Methode entwickelt, welche die Wahl der Ebene nicht rein auf Basis von Vorwissen trifft, sondern abhängig von der Richtung der minimalen Rauschenergie des Projektionssignals. Die Lage der Ebene im Raum ist so-

mit nicht fest vorgegeben, sondern wird signalabschnittsweise bestimmt. Abbildung 7.4 veranschaulicht die Unterschiede in der Wahl der Ebene im RGB-Farraum. Dies basiert auf der Erkenntnis, dass gerade unter starken

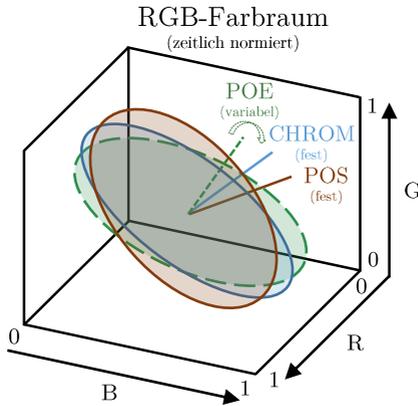


Bild 7.4: Vergleich der Wahl der Projektionsebenen im RGB-Farraum für POS, CHROM und POE

Störeinflüssen die Störkomponenten  $i(t)$  und  $s(t)$  nicht als klar trennbar und auch das Produkt  $i(t) \cdot s(t)$  nicht als vernachlässigbar klein angesehen werden können. In diesem Fall liegen die AC-Amplituden der Störkomponenten im RGB-Raum sehr verteilt vor. Es gilt in erster Linie, für das betrachtete Zeitintervall eine Projektionsrichtung mit möglichst geringem Störanteil zu finden, um die Wahrscheinlichkeit für die Trennung der BVP- und Störkomponenten und somit der erfolgreichen Rekonstruktion des Signals  $p(t)$  zu erhöhen. Dazu wird zunächst die Ebene nicht fest als orthogonal zu Vektor  $[1,1,1]$  angenommen, sondern, im Falle einer erhöhten Signalenergie in Folge von Artefakten, orthogonal zur Richtung der Signalenergie gewählt. Die Energie des normalisierten RGB-Signals ist dabei gegeben durch:

$$e = \begin{bmatrix} e_r & e_g & e_b \end{bmatrix} \quad (7.10)$$

Für den gesuchten Projektionsvektor  $\boldsymbol{v}$  soll nun gelten:

$$\boldsymbol{e} \cdot \boldsymbol{v} = 0 \quad (7.11)$$

Demnach entspricht Vektor  $\boldsymbol{e}$  der Normalen der definierten Ebene,  $\boldsymbol{v}$  entspricht jedem Einheitsvektor auf die Ebene. Die Normale der Ebene wird beim Überschreiten eines Schwellwertes für die Signalenergie wie folgt berechnet:

$$\boldsymbol{e} = \sqrt{\sum_{i=0}^N \tilde{C}_{n,i}^2}, \quad (7.12)$$

mit  $i$  als Anzahl der Abtastpunkte des Signalausschnitts des zeitlich normierten Signals  $\tilde{C}_n(t)$ . Da sich die Energie des Rohsignals bei Kopfbewegungen hauptsächlich in der ACOR-Komponente befindet, wird unter der Bedingung 7.11 eine Projektionsrichtung  $\boldsymbol{v}$  auf der Ebene gesucht, auf der die Rauschenergie  $E_{noise}$  des Projektionssignals  $S_v(t)$  möglichst gering ist, also:

$$E_{noise}(S_v(t)) \rightarrow \min, \quad (7.13)$$

mit

$$S_v(t) = \boldsymbol{v} \cdot \tilde{C}_n(t). \quad (7.14)$$

Daher wird der Ansatz in Anlehnung an die Namensgebung in [157] als „Plane orthogonal to Energy (POE)“ bezeichnet.

Zur Bestimmung der Rauschenergie wird zunächst das RGB-Rohsignal mittels FFT in seine spektralen Anteile zerlegt:

$$Y_c = FFT(\tilde{C}_n(t)) \quad (7.15)$$

Der Rauschvektor  $\boldsymbol{v}_{c,noise}$  wird anschließend definiert als:

$$\boldsymbol{v}_{c,noise} = [Re(Y_{c,f_1}), Im(Y_{c,f_1}), Re(Y_{c,f_2}), Im(Y_{c,f_2}), \dots, Re(Y_{c,f_m}), Im(Y_{c,f_m})] \quad (7.16)$$

mit  $c$  als Index der Farbkanäle (r,g,b) und  $f_1, f_2, \dots, f_m$  als Indizes der Rauschfrequenz-Bins des Spektrums für die gilt  $f > 4\text{Hz}$ . Die Rauschenergie  $e_{c,noise}$  eines Farbkanals kann somit repräsentiert werden durch

$$e_{c,noise} = v_{c,noise} v_{c,noise}^T \quad (7.17)$$

Fasst man die Rauschvektoren zu einer Rauschmatrix  $V_{noise}$  zusammen,

$$V_{noise} = \begin{bmatrix} v_{r,noise} \\ v_{g,noise} \\ v_{b,noise} \end{bmatrix}^T \quad (7.18)$$

kann die Rauschenergie des in Richtung  $\mathbf{v}$  projizierten Signals über

$$\begin{aligned} E_{noise}(\mathbf{v}) &= \mathbf{v}^T V_{noise}^T V_{noise} \mathbf{v} \\ &= V^T M_{noise} \mathbf{v}, \end{aligned} \quad (7.19)$$

mit  $M_{noise} = V_{noise}^T V_{noise}$  berechnet werden. Da der gesuchte Projektionsvektor  $\mathbf{v}$  auf der Ebene orthogonal zur Energie definiert wurde, kann er als Linearkombination von zwei beliebigen, zueinander orthogonalen Vektoren  $\mathbf{v}_{e,1}, \mathbf{v}_{e,2}$  auf der Ebene angesehen werden. Es gilt:

$$\mathbf{v} = \begin{bmatrix} \mathbf{v}_{e,1}, \mathbf{v}_{e,2} \end{bmatrix} \boldsymbol{\alpha} = P_e \boldsymbol{\alpha}, \quad (7.20)$$

mit  $\boldsymbol{\alpha}$  als normierten Koeffizienten-Vektor zur Erstellung des Projektionsvektors  $\mathbf{v}$ . Da  $\mathbf{v}$  als Einheitsvektor im dreidimensionalen Raum angenommen wurde, ist der Koeffizienten-Vektor  $\boldsymbol{\alpha}$  ein  $(2 \times 1)$ -Einheitsvektor im Koordinatensystem von  $\mathbf{v}_{e,1}, \mathbf{v}_{e,2}$ . Auf diese Weise ergibt sich:

$$E_{noise}(\mathbf{v}) = \boldsymbol{\alpha}^T P_e^T V_{noise}^T V_{noise} P_e \boldsymbol{\alpha} \quad (7.21)$$

mit  $L_{noise} = P_e^T V_{noise}^T V_{noise} P_e$ .

$E_{noise}$  kann durch Lösung eines Optimierungsproblems bestimmt werden:

$$\begin{aligned} \min_{\alpha} \quad & \alpha^T L_{noise} \alpha \\ \text{s.t.} \quad & \alpha^T \alpha = 1; \end{aligned} \quad (7.22)$$

Der optimale Koeffizient  $\alpha$  ist der Eigenvektor von  $L_{noise}$  mit minimalen Eigenwert. Da  $\alpha$  nur aus zwei Elementen besteht, kann das Optimierungsproblem analytisch gelöst werden, was die Echtzeitgeschwindigkeit des Algorithmus gewährleistet.

Der Algorithmus hat den Anspruch robuster gegenüber Kopfbewegungen

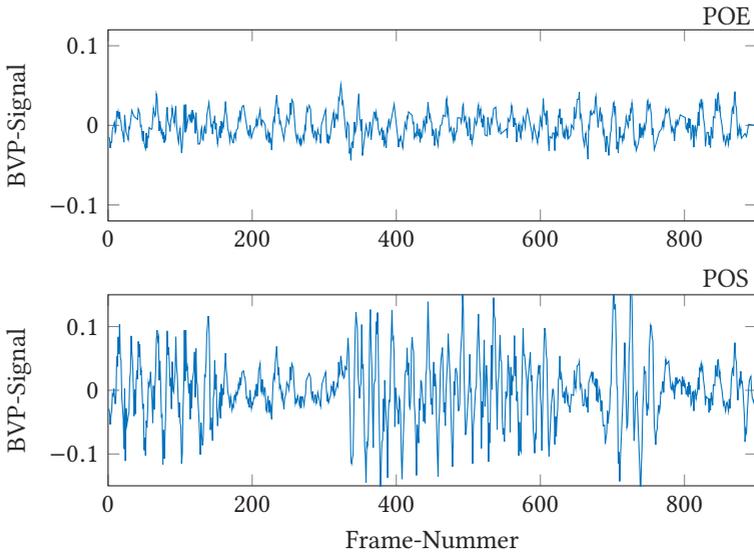


Bild 7.5: Vergleich der algorithmischen Ansätze bei vertikaler Kopfbewegung

zu sein. Vorteile bietet der Algorithmus insbesondere bei vertikalen Kopfbewegungen, wie sie unter anderem im Fahrzeug bei Fahrt auf unebener

Straße vermehrt vorkommen. Abbildung 7.5 zeigt die Auswirkungen derartiger Bewegungen auf das jeweils mit dem POE- und POS-Ansatz extrahierete BVP-Signal. Man erkennt für diesen Fall, dass im POS-Ansatz die ACOR-Komponente in Folge der Verletzung der Annahme in 7.7 verstärkt wird, die BVP-Komponente ist nur abschnittsweise erkennbar. Der POE-Ansatz ermöglicht hingegen die Unterdrückung der Störung durch die Wahl der Projektion in Richtung der minimalen Rauschenergie, wodurch das BVP-Signal für den gesamten Abschnitt erfolgreich rekonstruiert werden kann. Ein ausführlicher Vergleich auf Grundlage der Datenbasis findet sich in Abschnitt 7.3.1 am Ende des Kapitels.

### 7.1.2 Projektion bei geringem Artefakteinfluss

Für artefaktbehaftete Signalabschnitte kann durch die Wahl des Projektionsvektors  $v$  in Richtung der minimalen Rauschenergie des projizierten Signals der Einfluss von Artefakten reduziert und das BVP-Signal in vielen Fällen erfolgreich rekonstruiert werden. In Fällen ohne größeren Einfluss von Artefakten wird der Frequenzanteil des Signals mit  $f > 4\text{ Hz}$  jedoch nicht durch die Störkomponente, sondern durch die harmonischen Oberschwingungen der BVP-Komponente bestimmt. Der „Rauschvektor“ und der „BVP-Vektor“ auf der Ebene zeigen in diesem Fall in dieselbe Richtung. Der Projektionsvektor  $v$  würde folglich orthogonal zum BVP-Vektor gewählt werden und die BVP-Anteile im projizierten Signal  $S_v(t)$  minimiert. Für Signalabschnitte mit geringem Störeinfluss erfolgt daher eine alternative Berechnung des Projektionsvektors. Analog zum Rauschvektor  $v_{c,noise}$  wird im stationären Szenario ein Signalvektor  $v_{c,signal}$  definiert zu:

$$v_{c,signal} = [Re(Y_{c,f_1}), Im(Y_{c,f_1}), Re(Y_{c,f_2}), Im(Y_{c,f_2}), \dots, Re(Y_{c,f_k}), Im(Y_{c,f_k})], \quad (7.23)$$

mit  $f_1, f_2, \dots, f_k$  als Indizes der Signal-Frequenz-Bins für die gilt  $f < 4\text{ Hz}$ .

Zur Bestimmung des Projektionsvektors  $v$  wird für diesen Fall ein neues Optimierungsproblem der Art

$$\min_{\alpha} \frac{\alpha^T L_{noise} \alpha}{\alpha^T L_{signal} \alpha}, \quad (7.24)$$

mit  $L_{signal} = P_e^T V_{signal}^T V_{signal} P_e$  definiert.  $\alpha^T L_{signal} \alpha$  repräsentiert dabei die „Niederfrequenzenergie“ ( $f < 4\text{Hz}$ ) des projizierten Signals  $S_v(t)$ . Die Objektfunktion kann also vereinfacht als das inverse SNR des projizierten Signals  $S_v(t)$  angesehen werden. Frequenzkomponenten unter 4Hz werden verstärkt, Frequenzkomponenten über 4Hz abgeschwächt. Da unter dem Einfluss von Artefakten in der Regel auch die Niederfrequenzenergie stark erhöht wird, kann die Objektfunktion in diesem Fall nicht angewendet werden. Es wird entsprechend die zuvor eingeführte Objektfunktion aus Gl. 7.22 genutzt. Die Umschaltung zwischen den beiden Methoden zur Bestimmung des Projektionsvektors  $v$  erfolgt, ähnlich zur Wahl der Ebene, abhängig von der Signalenergie im grünen Kanal.

Das Optimierungsproblem in Gl. 7.24 liegt in Form eines generalisierten Rayleigh-Quotienten vor.  $L_{signal}$  kann dabei zerlegt werden zu  $L_{signal} = \lambda \Lambda \lambda^T = \lambda \Lambda^{\frac{1}{2}} \Lambda^{\frac{1}{2}} \lambda^T$ , mit  $\lambda$  und  $\Lambda$  als Matrizen, bestehend aus den Eigenvektoren und diagonalisierten Eigenwerten. Somit kann  $\alpha^T L_{signal} \alpha$  formuliert werden zu:

$$\alpha^T L_{signal} \alpha = \alpha^T \lambda \Lambda^{\frac{1}{2}} \Lambda^{\frac{1}{2}} \lambda^T \alpha = \beta^T \beta, \quad (7.25)$$

mit

$$\alpha = \lambda \Lambda^{-\frac{1}{2}} \beta. \quad (7.26)$$

Dadurch kann Gleichung 7.24 in der Form

$$\min_{\alpha} \frac{\beta^T \Lambda^{-\frac{1}{2}} \lambda^T L_{noise} \lambda \Lambda^{-\frac{1}{2}} \beta}{\beta^T \beta} \quad (7.27)$$

ausgedrückt und auf die selbe Weise wie Gl. 7.22 gelöst werden.

## 7.2 Bestimmung der Herzfrequenz aus dem BVP-Signal

Im letzten Schritt des PPGI-Algorithmus erfolgt die Bestimmung der Herzfrequenz auf Basis des extrahierten BVP-Signals  $BVP(t)$ , für das unter Verwendung des POE-Algorithmus gilt:  $BVP(t) = S_v(t)$  (siehe Gl. 7.14). Eine gängige Methode zur Schätzung der Herzfrequenz ist die Analyse des Frequenzspektrums nach Anwendung der FFT (z.B. [109][137] [35]). Über eine Detektion der maximalen Frequenz im Spektrum innerhalb des physiologischen Bereichs der Herzfrequenz (in der Literatur meist  $[0.6 - 4.0\text{Hz}]$ , z.B. [34]) ergibt sich die gesuchte Frequenz in  $\text{Hz}$  bzw. „Schläge pro Minute“ (bpm). Ein Beispiel wird in Abbildung 7.6 gezeigt. Ein Nachteil die-

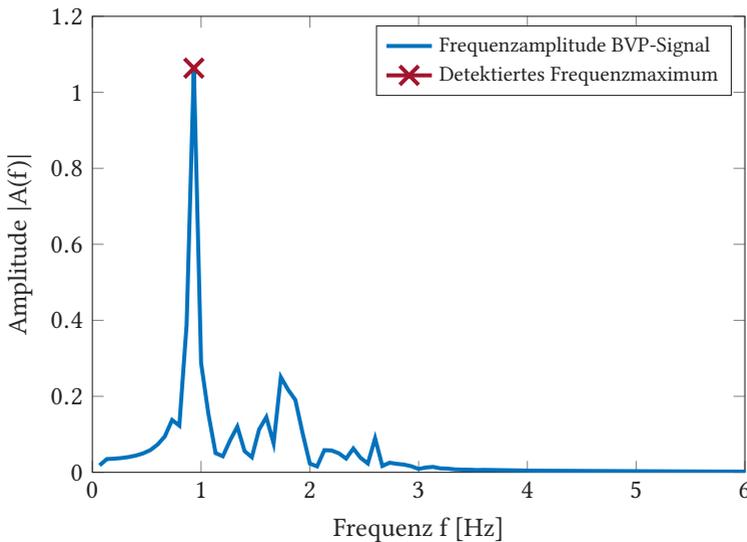


Bild 7.6: Beispiel für die Herzfrequenzschätzung im Frequenzspektrum

ser Methode liegt im Zusammenhang von Beobachtungszeitraum und Frequenzauflösung. Für kurze Beobachtungszeiträume im Bereich weniger Se-

kunden kann das Spektrum für die Bestimmung der momentanen Herzfrequenz nur unzureichend genau aufgelöst werden. Es gilt:

$$\Delta f = \frac{f_a}{N} \quad (7.28)$$

mit  $f_a$  als Abtastrate und  $N$  als Anzahl der Abtastwerte. Für eine Abtastrate von  $f_a = 30\text{Hz}$  und einen Beobachtungszeitraum von 300 Samples ergibt sich eine Frequenzauflösung von  $\Delta f = 0,1\text{Hz} = 6\text{bpm}$ . Die Frequenzauflösung im Sinne der Abtastabstände im Spektrum kann durch Methoden wie Zero-Padding zwar erhöht werden, besser geeignet ist für kurze Beobachtungszeiträume jedoch die Analyse im Zeitbereich.

Bei der Analyse im Zeitbereich wird das BVP-Signal meist durch einen Bandpassfilter mit Durchlassbereich  $[0.6 - 4.0\text{Hz}]$  vorverarbeitet. Anschließend erfolgt die Detektion der Peaks im gefilterten Signal  $BVP_p(t)$ <sup>1</sup>, wobei nach Invertierung des Signals und unter geringen Störeinflüssen ein Signal-Peak dem Maximum der Pulswelle (Percussion-Welle, vgl. Abschnitt 2.1.1) entspricht. Auf diese Weise wird eine Schlag-zu-Schlag-Analyse ermöglicht, welche neben der Berechnung der momentanen Herzfrequenz, insbesondere die Bestimmung der Pulsratenvariabilität ermöglicht. Die Herzfrequenz errechnet sich über den inversen Mittelwert der Schlag-zu-Schlag-Distanzen (*engl.*: Inter-Beat-Interval (IBI)) des gefilterten BVP-Signals  $BVP_p(t)$  in einem vordefinierten Zeitintervall.

$$\text{HF} = \frac{60}{\text{IBI}} \quad (7.29)$$

Für die möglichst exakte Bestimmung der HF bedarf es eines robusten Verfahrens zur Peak-Detektion. In dieser Arbeit wurde der Ansatz verfolgt, das BVP-Signal durch gezielte Vorverarbeitung mittels eines adaptiven Bandpassfilters so aufzubereiten, dass die eigentliche Maxima-Detektion über einen K-Nearest-Neighbor (KNN)-Algorithmus, unter Beachtung we-

---

<sup>1</sup> Index p für „processed“

niger Regeln für die Gültigkeit eines Maximums, realisiert werden kann. Die Anwendung bestehender Algorithmen zur Peak-Detektion bei kontaktbasierten PPG-Verfahren ist insofern begrenzt, als dass sich die Morphologie von PPGI- und PPG-Pulswellen zu stark unterscheidet. Beim passiven Ansatz ist beispielsweise die Tidal-Welle (Kap. 2.1.1), auf Grund der geringeren Eindringtiefe des sichtbaren Lichts und der vergleichsweise geringen Stärke der Wellenreflexion der oberflächennahen Schichten der Dermis (vgl. Kap. 2.3.3), nur minimal ausgeprägt.

### 7.2.1 Adaptiver Bandpassfilter

Aufgabe des in Abschnitt 7.1 vorgestellten POE-Algorithmus ist es, aus dem zeitlich normierten RGB-Rohsignal  $\tilde{C}_n(t)$  das (ungefilterte) BVP-Signal  $BVP(t)$  zu rekonstruieren. Im nächsten Schritt soll nun ein möglichst rauscharmes BVP-Signal  $BVP_p(t)$  für die Schlag-zu-Schlag-Analyse zur Herzfrequenzbestimmung im Zeitbereich erzeugt werden, indem durch adäquate Filterung die Komplexität für die Peak Detektion minimiert wird. Zu diesem Zweck wurde eine adaptive Filtermethode entwickelt, welche im folgenden Abschnitt vorgestellt wird.

Im Bereich der EKG-Signalverarbeitung werden adaptive Filter schon länger zur Störkompensation von Bewegungsartefakten eingesetzt [106] [132] [62]. Dabei bilden das zu filternde Messsignal und eine parallel aufgezeichnete Störkomponente (Referenzsignal) den Eingang des Filters. Das Ziel ist es, die Filterparameter kontinuierlich anzupassen, sodass das Differenzsignal aus Messsignal und Referenzsignal möglichst klein wird. Auf dieser Basis kann anschließend der Störanteil im Messsignal bestimmt werden. Eine adaptive Filtermethode für PPGI wird von Huang und Dung in [53] präsentiert. Sie nutzen einen kaskadierten, adaptiven Filter, wobei als Referenzsignal die x- und y-Koordinaten der Gesichtspose dienen. Ziel ist es, die durch Gesichtsbewegung verursachte Störkomponente zu elimi-

nieren. Zur Minimierung des Differenzsignals wird ein LMS-Schätzer eingesetzt, dessen Koeffizienten adaptiv angepasst werden.

Im Rahmen dieser Arbeit wird das Konzept eines adaptiven Bandpassfilters vorgestellt, welcher auf dem Prinzip eines adaptiven Filters aufbaut, jedoch anstelle der Differenzsignalbildung und LMS-Schätzung eine Spektrogramm-Analyse zur Anpassung der Filtergrenzen nutzt. Während das Verfahren in [53] auf die Reduktion geringer Störeinflüsse zielt, soll mit der Methode in dieser Arbeit insbesondere die Unterdrückung von stark gestörten Signalanteilen für eine verbesserte Peak-Detektion unter hohen Artefakteinflüssen realisiert werden. In zahlreichen Arbeiten in der Literatur wird ein Bandpassfilter mit festen Filtergrenzen entsprechend des physiologischen Bereichs der Herzfrequenz von  $[0.6 - 4.0\text{Hz}]$  eingesetzt (vgl. Abschnitt 7.2). Das Ziel eines solchen Filters liegt in der Eliminierung hoch- und niederfrequenter Signalanteile, welche physiologisch bedingt nicht Blutvolumenpuls-induziert sein können und somit eindeutig einer Störquelle zuzuordnen sind. Jedoch kommt es unter Einfluss von Artefakten häufig zu überlagernden Störkomponenten im physiologischen Herzfrequenzbereich, was die Wahrscheinlichkeit für Fehldetektionen bei der Peak Detektion erhöht. Die Idee des entwickelten Bandpassfilters ist es, die Filtergrenzen enger zu wählen und adaptiv hinsichtlich einer initial geschätzten Herzfrequenz  $f_p$  unter Berücksichtigung identifizierter, überlagernder Störkomponenten anzupassen, um so einen schmaleren und beweglichen Durchlassbereich  $[f_p - 0.5\Delta f, f_p + 0.5\Delta f]$  mit möglichst geringen Störanteilen zu erhalten.

Mit diesem Ziel wird eine Zeit-Frequenz-Analyse des BVP-Signals und zusätzlich extrahierter Artefaktsignale auf Basis der jeweiligen Spektrogramme durchgeführt. Die Spektrogramme werden dabei mittels Kurzzeit-Fourier-Transformation (*engl.*: „Short-Time Fourier Transform (STFT)“) unter Nutzung eines Hanning-Fensters mit Fensterbreite  $L_H = 5\text{s}$  für das Zeitintervall der letzten  $L_t = 30\text{s}$  des Signals erzeugt. Somit ergibt sich ein Analysezeitraum des Spektrogramms von  $L_f = 25\text{s}$ . Die geringe Frequenz-

auflösung des Spektrogramms bei der gewählten Fensterbreite von 5s kann für das Ziel einer groben, initialen Herzfrequenzschätzung ( $f_p$ ) als ausreichend angesehen werden. Als Artefaktsignale dienen die horizontale und vertikale Kopfbewegung, welche über das Face-Tracking-Modul extrahiert werden. Abbildung 7.7 zeigt beispielhaft die erzeugten Spektrogramme eines Signalausschnitts eines Probanden bei rotatorischer Bewegung des Kopfes in Szenario 201 der Datenbank, anhand derer die einzelnen Schritte der Zeit-Frequenz-Analyse erklärt werden.

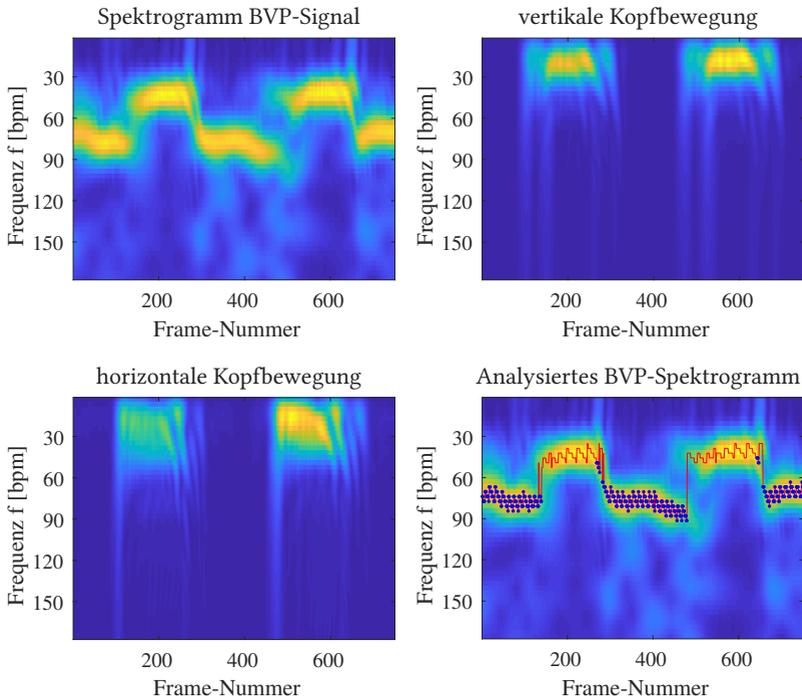


Bild 7.7: Spektrogramm Analyse.

Im ersten Schritt werden mittels einfacher Maximum-Detektion die Hauptfrequenzen im Spektrogramm des BVP-Signals markiert. Im Anschluss

werden anhand zweier Kriterien bestimmte Hauptfrequenzen als fehlerhaft markiert. Zunächst wird dazu für jede Hauptfrequenz die Peakhöhe (Amplitude) bezogen auf die Gesamtenergie im untersuchten Frequenzbereich, als relatives Maß für die Klarheit des Peaks („Peak-Clearity“), berechnet. Wird ein definierter Schwellwert unterschritten, gilt das Spektrum als durch Artefakte verschmiert („blurred“) und der Peak wird als fehlerhaft bzw. störbehaftet markiert (rote Markierung im analysierten BVP-Spektrum in Abb. 7.7). In einem zweiten Schritt wird überprüft, ob sich die übrigen Hauptfrequenzen des BVP-Spektrums auch als Hauptfrequenzen in den Spektrogrammen der Artefaktensignale wiederfinden. In diesem Fall sind diese mit hoher Wahrscheinlichkeit auf Kopfbewegungen zurückzuführen und werden im BVP-Spektrum ebenfalls als artefaktinduziert markiert. Die initiale Schätzung der Herzfrequenz  $f_p$  erfolgt anschließend durch Mittelung der übrigen, nicht als fehlerhaft markierten Hauptfrequenzen im BVP-Spektrum (blaue Markierung im analysierten BVP-Spektrum in Abb. 7.7). Die variablen Parameter des Filters sind der Analysezeitraum  $L_f$  und die Bandbreite  $\Delta f$  des Durchlassbereichs, welche applikationsabhängig zu wählen sind. Ein kleinerer Analysezeitraum erhöht die Adaptionfähigkeit bei schnellen Herzfrequenzänderungen, wie sie beispielsweise in Fitnessapplikationen vorkommen können. Jedoch wird gleichzeitig die Empfindlichkeit gegenüber Störeinflüssen erhöht. Ähnlich verhält es sich bei der Wahl der Bandbreite. Ein großes  $\Delta f$  erhöht die Wahrscheinlichkeit für verbleibende Störanteile im Signal. Dennoch sollte  $\Delta f$  nicht zu klein gewählt werden, um zu verhindern, dass wichtige, pulsinduzierte Signalinformationen für die Analyse der Schlag-zu-Schlag-Distanzen zur Schätzung der Pulsratenvariabilität entfernt werden. Als Trade-Off wird ein Analysezeitraum von  $L_f = 25\text{s}$  und eine Bandbreite von  $\Delta f = 0.8\text{Hz} \hat{=} 50\text{bpm}$  gewählt. Die Filtergrenzen des adaptiven Bandpassfilters ergeben sich entsprechend zu  $[f_p - 0.4\text{Hz}, f_p + 0.4\text{Hz}]$ .

## 7.2.2 Peak Detektion im BVP-Signal

Durch den Einsatz des adaptiven Bandpassfilters ergibt sich das Signal  $BVP_p(t)$ , welches im Vergleich zum Signal nach Bandpassfilterung mit festen Grenzen  $BVP_{p, fest}(t)$ , deutlich weniger Störeinflüssen unterliegt. Abbildung 7.8 zeigt das Beispiel eines BVP-Signals ohne Filterung, sowie nach Filterung mit festen und adaptiven Grenzen, unter dem Einfluss rotatorischer Kopfbewegungen. Insbesondere in stärker gestörten Signalabschnitten (rot markiert) scheint eine korrekte Identifikation der Peaks in  $BVP_{p, fest}(t)$  nicht mehr möglich. In  $BVP_p(t)$  kann die Peak Detektion durch die präzisere Filterung des Signals unter Verwendung eines einfachen K-Nearest-Neighbor(KNN)-Ansatzes korrekt durchgeführt werden. Die detektierten Peaks sind in Abbildung 7.8 rot markiert. Dabei werden lediglich folgende Regeln für die Gültigkeit eines Peaks definiert:

- Zwischen zwei Peaks gibt es einen Mindestabstand (abhängig von der maximalen Herzfrequenz).
- Zwischen zwei Peaks liegt ein Minimum kleiner Null.
- Es sind nur Peaks  $> 0$  zulässig, da das Signal mittelwertfrei ist.

Die Verbesserungen, welche sich unter Nutzung des adaptiven Bandpassfilters hinsichtlich der Genauigkeit der Herzfrequenzschätzung ergeben, werden in Abschnitt 7.3.2 dargestellt.

## 7.3 Messungen und Ergebnisse

### 7.3.1 Optimierung des SNR

In Abschnitt 7.1 des Kapitels wurde der POE-Algorithmus zur Trennung der BVP- und Störkomponenten auf Basis der Projektion in Richtung der minimalen Rauschenergie vorgestellt. Der Algorithmus verfolgt das Ziel einer gesteigerten Robustheit gegenüber Artefakten, um unter anderem

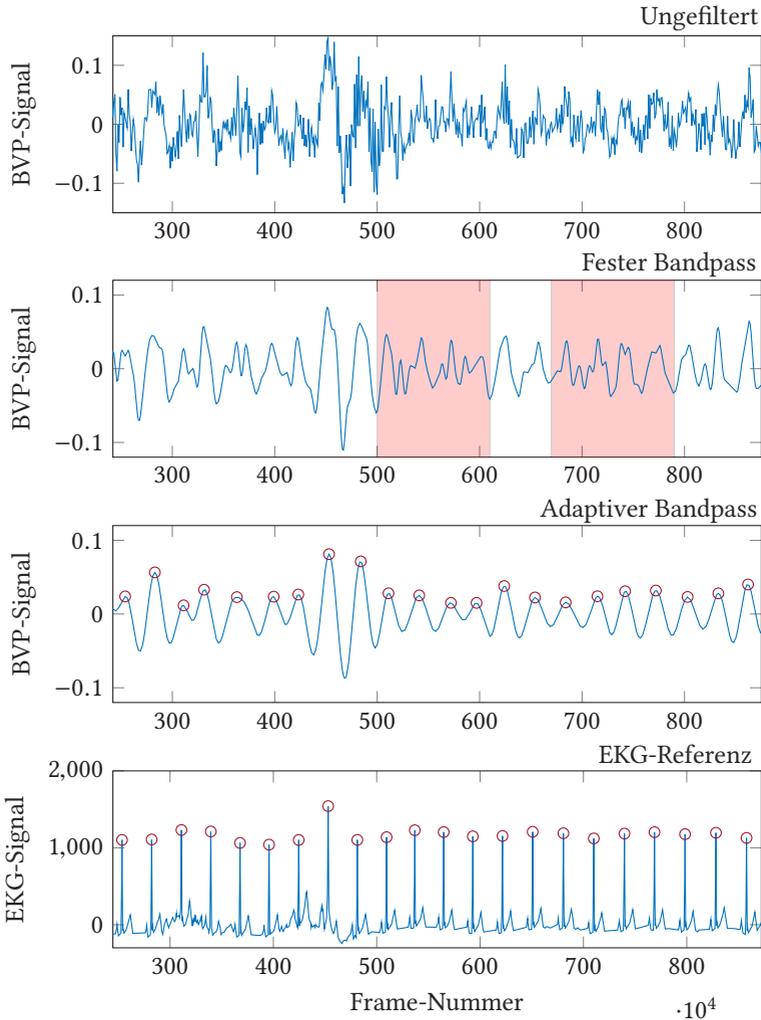


Bild 7.8: BVP-Signale mit unterschiedlicher Filterung

unter dem Einfluss starker Kopfbewegungen eine erfolgreiche BVP-Signalrekonstruktion zu ermöglichen. In den folgenden beiden Abbildungen werden rekonstruierte BVP-Signale für den POS-Algorithmus und den entwickelten POE-Algorithmus gezeigt. In Abbildung 7.9 sind die Signale eines 1-minütigen Videoausschnittes eines Probanden in Szenario 203 bei translatorischer Kopfbewegung zu sehen. Abbildung 7.10 stellt den in Abbildung 7.9 markierten Signalausschnitt vergrößert dar. Es ist zu erkennen, dass die veränderte Wahl der Projektionsrichtung im POE-Algorithmus unter dem Einfluss translatorischer Kopfbewegungen eine Rekonstruktion der BVP-Signalkomponente mit verringertem Bewegungsanteil ermöglicht. Tabelle 7.1 zeigt die SNR-Auswertung des POE-Algorithmus

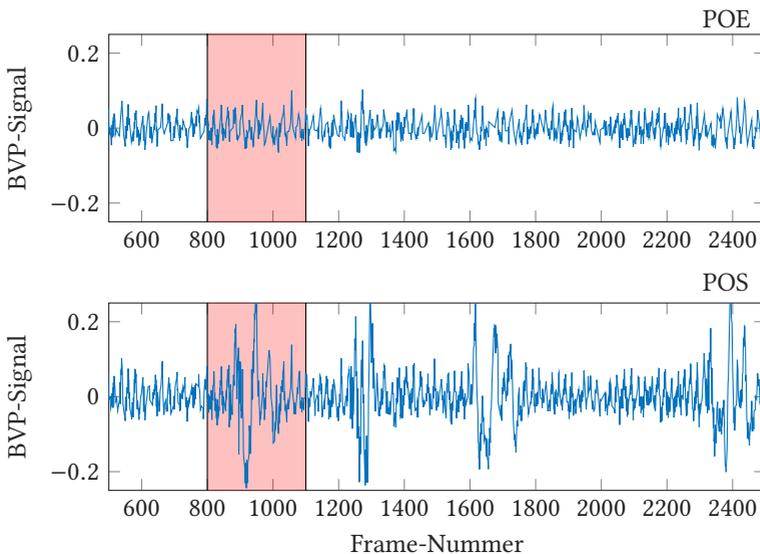


Bild 7.9: Vergleich des BVP-Signals bei POE und POS für das gesamte Video

im Vergleich zu den State-of-the-Art Algorithmen ICA, CHROM und POS für die verschiedenen Datenbankszenarien. Für den ICA-Algorithmus wurde die JADE-Implementierung nach Cardoso [24] genutzt, ergänzt durch

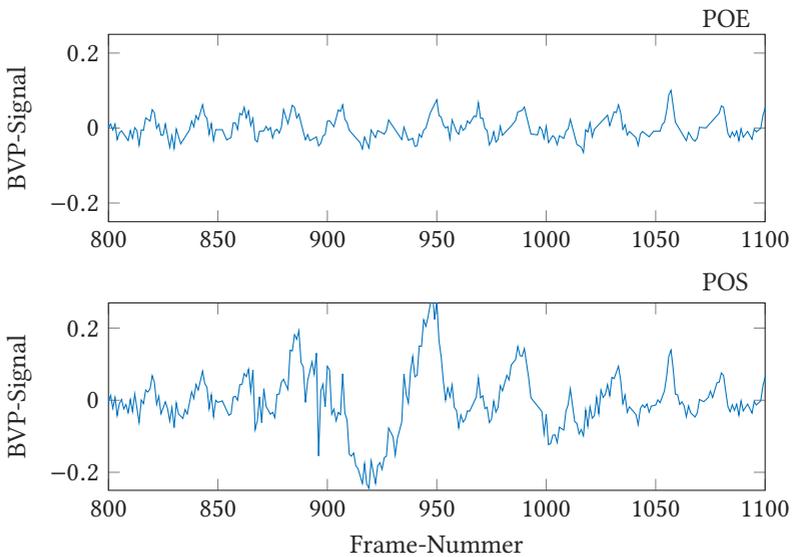


Bild 7.10: Vergleich des BVP-Signals bei POE und POS für eine Kopfbewegung

das in Abschnitt 2.4.3 beschriebene Verfahren der segmentierten Faltung. CHROM und POS wurden entsprechend [34] und [157] implementiert. Die ROI wurde für alle Algorithmen in gleicher Form über den Multi-Patch-Ansatz (6.3) bestimmt. Dabei wurde das Selektions-Kriterium des Hochfrequenzrauschens deaktiviert, um die Aussagekraft der SNR-Analyse für die Kernalgorithmen nicht zu beeinträchtigen. Insofern ergeben sich für den POS-Algorithmus, welcher als State-of-the-Art-Algorithmus für die SNR-Analyse der ROI-Detektion im letzten Kapitel eingesetzt wurde, leicht veränderte Werte.

Unter Ruhebedingungen und bei natürlicher Beleuchtung weist POE eine Steigerung des SNR im Vergleich zu ICA und CHROM auf. In Szenario 102 bei künstlicher Beleuchtung mit geringer Intensität liefert ICA das beste Ergebnis. Die gute Performance des ICA-Algorithmus, auch im Vergleich zu Analysen aus dem Stand der Technik (z.B. [78]), kann durch den Ein-

satz der segmentierten Faltung begründet werden. Für die Szenarien 103 und 141 mit wechselnder Beleuchtungsintensität wird unter Verwendung von POE der jeweils höchste Wert für das SNR erzielt. Für das Szenario mit skalierender Bewegung liefert POS die höchsten Werte für das SNR. Da sich die Ausrichtung des Gesichtes während der Bewegungen nicht ändert und die Bewegung orthogonal zur Richtung des Bildaufnahmesystems ausgeführt wird, ist die Wahl der Projektionsrichtung orthogonal zu Vektor  $\mathbf{1}$  in diesem Fall gut geeignet. Eine deutliche Verbesserung des SNR gegenüber den State-of-the-Art-Ansätzen liefert POE für die Bewegungsszenarien 201,203 und 204. Die Vorteile des POE-Algorithmus bei Auf- und Abbewegung des Kopfes wurden bereits am Ende von Abschnitt 7.1 dargestellt. Dies zeigt die besondere Eignung des Algorithmus für den Einsatz in herausfordernden Umgebungen mit wechselnden Lichtverhältnissen, insbesondere jedoch bei häufigen und stark ausgeprägten Bewegungen, wie sie beispielsweise bei der Anwendung im Fahrzeug vorkommen. In diesen Fällen scheint für die Projektions-Algorithmen die Wahl der Projektion in Richtung der geringsten Rauschenergie gut geeignet.

Tabelle 7.1: Vergleich der SNR-Werte für verschiedene Algorithmen in den Datenbank-Szenarien in dB

Szenario	ICA	CHROM	POS	POE
101	2,74	3,08	<b>3,45</b>	3,43
102	<b>-0,66</b>	-1,17	-0,90	-0,95
103	-0,28	-0,40	-0,39	<b>-0,26</b>
104	2,42	2,33	2,64	<b>3,06</b>
201	-3,17	-3,01	-2,47	<b>-1,13</b>
202	0,91	1,21	<b>1,58</b>	1,56
203	-0,06	-0,33	-0,17	<b>0,52</b>
204	-1,48	-1,39	-0,86	<b>-0,71</b>

### 7.3.2 Verbesserung der Herzfrequenzschätzung

Zur Evaluation des Einflusses des adaptiven Bandpassfilters auf die Genauigkeit der Herzfrequenzschätzung wird die „Success Rate“ (SR) als Gütemaß für verschiedene Szenarien der aufgezeichneten Datenbank ermittelt. Die Success Rate ist definiert als der Anteil an geschätzten Frequenzen, deren Abweichung zur Ground Truth des EKG-Referenzsystems unterhalb eines definierten Fehlertoleranzwertes liegt. Die Fehlertoleranz wurde für die Auswertung zu  $e_t = 3bpm$  gewählt. Die Tabellen 7.2-7.6 zeigen die über die Probanden gemittelten  $SR_3$ -Werte für die Kombination verschiedener Signalverarbeitungsalgorithmen zur Trennung der Signalcomponenten (Spalten) und verschiedener Ansätze zur Bestimmung der Herzfrequenz aus dem extrahierten BVP-Signal (Zeilen). Das Frequenzspektrum wurde über eine 2048-Punkte-FFT nach Zero-Padding ermittelt. Für den festen Bandpass wurden die Filtergrenzen  $[0.6 - 4.0Hz]$  angenommen, die Parameter des adaptiven Bandpassfilters wurden entsprechend der Beschreibung in Abschnitt 7.2.1 gewählt. Der Ansatz der Hilbert Transformation zur Schätzung der Herzfrequenz aus dem BVP-Zeitsignal ist in [14] beschrieben. Für Szenario 101 zeigen sich unabhängig von der Kombination der algorithmischen Ansätze hohe Werte für das  $SR_3$ -Kriterium. Unter Verwendung des adaptiven Bandpasses liegen die geschätzten Herzfrequenzen für die verwendeten SV-Algorithmen fast vollständig innerhalb der Fehlertoleranz. Befinden sich Proband und Aufnahmesystem in Ruhe bei gleichbleibender Beleuchtung ist die Genauigkeit der Messmethode vergleichbar mit kontaktbasierten Sensoren.

Für die Bewegungsszenarien (201-204) sinken die  $SR_3$ -Werte erwartungsgemäß leicht ab. Dabei zeigen die Werte einen Anstieg der Schätzgenauigkeit unter Verwendung des adaptiven Bandpassfilters für sämtliche Bewegungsszenarien, unabhängig von der Wahl des Signalverarbeitungsalgorithmus. Dies bestätigt den positiven Einfluss der Filterung mittels der entwickelten Methode auf die Genauigkeit der Peak-Detektion und die

Tabelle 7.2: Success Rate für Szenario 101

<b>Ansatz</b>	<b>ICA</b>	<b>CHROM</b>	<b>POS</b>	<b>POE</b>
<b>Frequenzspektrum</b>	0,84	0,84	0,85	0,84
<b>Fester Bandpass</b>	0,87	0,96	0,97	0,91
<b>Hilbert Transform.</b>	0,91	0,96	0,97	0,93
<b>Adaptiver Bandpass</b>	0,98	0,99	0,99	0,99

Schätzung der Herzfrequenz. Mit Blick auf die Bewegungsszenarien ist dies vor allem auf die Spektrogrammanalyse der Kopfbewegungssignale zur geeigneten Bestimmung der initialen Hauptfrequenz  $f_p$  zurückzuführen. Obwohl der adaptive Bandpassfilter in erster Linie zur Eliminierung von Bewegungsartefakten konzipiert wurde, zeigt sich der positive Einfluss auch bei den Beleuchtungsszenarien. So kann beispielsweise die Schätzgenauigkeit in Szenario 103 von 78% Prozent für die frequenzbasierte Schätzung (gemittelt über die SV-Algorithmen) auf 87% bei zeitbasierter Schätzung unter Verwendung des Filters gesteigert werden. Dies lässt sich dadurch begründen, dass durch das Kriterium der Peak-Clearity bei der Schätzung der Hauptfrequenz auch die Störeinflüsse durch Lichtänderungen weitestgehend eliminiert werden können.

Hinsichtlich der Kombination mit den verschiedenen Signalverarbeitungsalgorithmen bestätigen die Werte der Tabellen 7.6 die Ergebnisse aus Abschnitt 7.3.1. Während die Kombination von POS und adaptivem Bandpass für das Szenario mit skalierender Bewegung die höchste Genauigkeit liefert, gilt dies in den Szenarien 201,203 und 204 für die Kombination aus POE und adaptivem Bandpass. Dies bestätigt die Eignung von POE für PPGI-Messungen unter Einflüssen von Artefakten, insbesondere stärker ausgeprägten Kopfbewegungen.

Tabelle 7.3: Success Rate für Szenario 201

Ansatz	ICA	CHROM	POS	POE
<b>Frequenzspektrum</b>	0,51	0,56	0,58	0,64
<b>Fester Bandpass</b>	0,34	0,44	0,47	0,58
<b>Hilbert Transform.</b>	0,42	0,53	0,58	0,59
<b>Adaptiver Bandpass</b>	0,64	0,76	0,75	<b>0,78</b>

Tabelle 7.4: Success Rate für Szenario 202

Ansatz	ICA	CHROM	POS	POE
<b>Frequenzspektrum</b>	0,74	0,77	0,76	0,76
<b>Fester Bandpass</b>	0,76	0,85	0,87	0,87
<b>Hilbert Transform.</b>	0,78	0,87	0,89	0,87
<b>Adaptiver Bandpass</b>	0,95	0,96	<b>0,97</b>	0,96

Tabelle 7.5: Success Rate für Szenario 203

Ansatz	ICA	CHROM	POS	POE
<b>Frequenzspektrum</b>	0,65	0,65	0,67	0,69
<b>Fester Bandpass</b>	0,57	0,64	0,65	0,70
<b>Hilbert Transform.</b>	0,60	0,65	0,64	0,70
<b>Adaptiver Bandpass</b>	0,87	0,85	0,87	<b>0,90</b>

## 7.4 Diskussion und Fazit

In diesem Kapitel wurde ein Algorithmus zur Trennung der BVP- und Stör-signal-komponenten der RGB-Rohsignale vorgestellt. Dieser projiziert, ähnlich wie die Algorithmen CHROM und POS, das zeitlich normalisierte RGB-

Tabelle 7.6: Success Rate für Szenario 204

<b>Ansatz</b>	<b>ICA</b>	<b>CHROM</b>	<b>POS</b>	<b>POE</b>
<b>Frequenzspektrum</b>	0,64	0,66	0,69	0,71
<b>Fester Bandpass</b>	0,40	0,55	0,55	0,57
<b>Hilbert Transform.</b>	0,43	0,52	0,54	0,55
<b>Adaptiver Bandpass</b>	0,69	0,75	0,77	<b>0,79</b>

Rohsignal in den RGB-Farbraum. Während CHROM und POS eine auf Annahmen basierende (2x3)-Projektionsmatrix nutzen, sucht der entwickelte Algorithmus einen (1x3)-Projektionsvektor in Richtung der minimalen Rauschenergie, daher die entsprechende Benennung in „POE - Plane Orthogonal to (Noise-)Energy“. Dies ist mit dem Ziel verbunden, auch bei starken Artefakten, unter deren Einfluss die bei CHROM und POS getroffenen Annahmen nicht immer gegeben sind, eine erfolgreiche Rekonstruktion des BVP-Signals und somit den Einsatz in herausfordernden Umgebungen zu ermöglichen. Die Analysen auf Basis der Labordatenbank zeigen, dass der Algorithmus insbesondere bei starker Änderung der Kopfpose vorteilhafte Eigenschaften besitzt.

Des Weiteren wurde ein Verfahren zur adaptiven Bandpassfilterung des extrahierten BVP-Signals auf Basis der Spektrogrammanalyse von BVP- und Artefaktensignalen vorgestellt. Durch die Wahl eines verkleinerten und adaptiven Durchlassbereichs des Filters wird das BVP-Signal so vorverarbeitet, dass die Schlag-zu-Schlag-Analyse durch einfache Ansätze für die Peak-Detektion ermöglicht wird. Der positive Einfluss auf die Genauigkeit der Schätzung der Herzfrequenz konnte für verschiedene Szenarien in Kombination mit unterschiedlichen PPGI-Kernalgorithmen gezeigt werden. Unter Verwendung der Success Rate mit Fehlertoleranz  $e_t = 3$  bpm als Gütemaß kann durch die Kombination der Ansätze Multi-Patch, POE und der adaptiven Bandpassfilterung unter Ruhebedingungen eine 99-

prozentige Übereinstimmung mit der EKG-Referenz nachgewiesen werden. Im stationären Setting liefert das System somit eine mit kontaktbasierten Sensoren vergleichbare Messgenauigkeit. In den Bewegungsszenarien liegt der Wert im Durchschnitt bei 86%. Im nächsten Kapitel wird unter anderem untersucht, ob beim Einsatz des Systems in mobilen Fahrzeugumgebungen mit zufälligen Bewegungen der Probanden und wechselnden Beleuchtungsintensitäten eine vergleichbare Messgenauigkeit erzielt werden kann.



# 8

## Onlinefähiges System zum Einsatz unter Realbedingungen

**D**AS ÜBERGEORDNETE ZIEL der Arbeit ist die Entwicklung eines PPGI-Systems für die Live-Analyse [S-AF-03] in Alltagsszenarien unter realen Bedingungen, um einerseits die Herzfrequenz, aber auch Parameter wie die Pulsratenvariabilität [S-AF-02] kontinuierlich zu erfassen. Eine zu untersuchende Fragestellung dabei ist, ob auch in sehr herausfordernden Szenarien, beispielsweise beim Einsatz in mobilen Systemen wie dem Automobil, trotz starker Bewegungs- und Lichtartefakten eine ausreichende Genauigkeit und Verfügbarkeit des Systems erreicht werden kann [S-AF-01]. Die Umsetzung als Gesamtsystem und Evaluation in Realszenarien sind Inhalt dieses Kapitels.

Im nächsten Abschnitt wird zunächst die Umsetzung der bereits unter Laborbedingungen evaluierten algorithmischen Teilkomponenten der Bild- und Signalverarbeitung in ein onlinefähiges Analysesystem beschrieben (vgl. Abb 8.1). Im Anschluss wird die Eignung des Systems für den

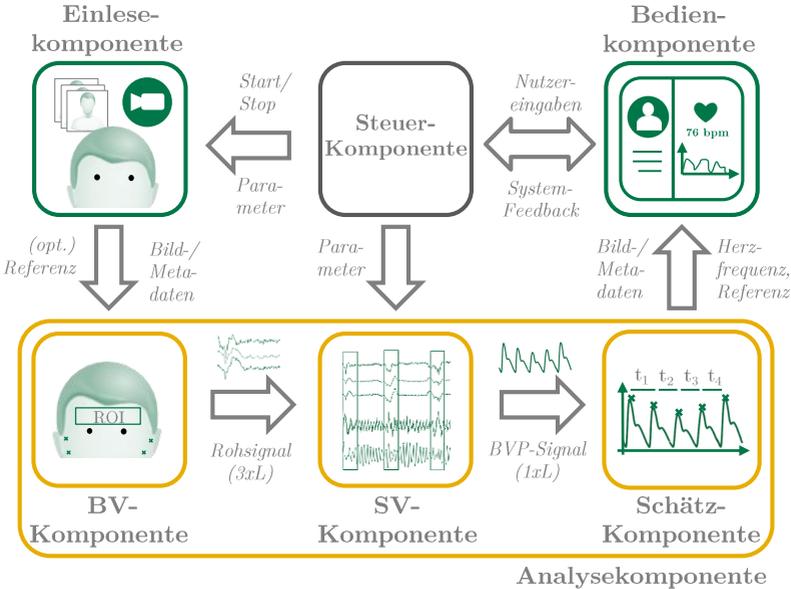


Bild 8.1: Softwarearchitektur des Gesamtsystems - Analysekomponente

Realeinsatz anhand von zwei Anwendungsfällen untersucht [S-AF-06]. Zunächst wird im Setting eines Büroarbeitsplatzes analysiert, mit welcher Qualität sich die Pulsratenvariabilität im Vergleich zur EKG-basierten Herzfrequenzvariabilität ableiten lässt, da die HRV als wichtiger Indikator für Stress und Überlastung gilt. Im Unterschied zu den Labormessungen steht das System unter dem Einfluss nicht vorgegebener, zufälliger Probandenbewegungen und Lichtbedingungen. Als weiterer wichtiger Anwendungsfall wird der Einsatz des Systems zur Vitaldatenerfassung im Kontext der Fahrerzustandserkennung im Automobil evaluiert. Dabei wird die

Genauigkeit und Verfügbarkeit des Systems bei dauerhafter Präsenz von Bewegungs- und Lichtartefakten untersucht und die entwickelten Methoden zur Artefaktkompensation analysiert. In diesem Zusammenhang werden unter anderem spezielle Randbedingungen des Systems für den Einsatz in einem solch dynamischen Setting herausgearbeitet und ein Ansatz zur Konfidenzschätzung vorgestellt [S-AF-04].

## 8.1 Umsetzung als Live-Prototypensystem

Für die Umsetzung als onlinefähiges Prototypensystem gilt es die einzelnen algorithmischen Teilkomponenten der Analysekomponente in geeigneter Form miteinander zu verbinden. Abbildung 8.2 zeigt den algorithmischen Ablauf von der Bildaufnahme bis zur Ausgabe der Herzfrequenz. Die rechte Hälfte der Grafik enthält die jeweiligen Ein- und Ausgabeinformationen der algorithmischen Teilkomponenten.

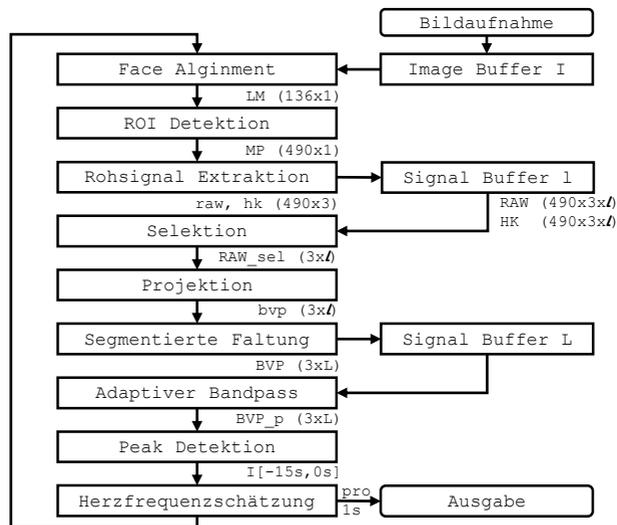


Bild 8.2: Ablauf des Algorithmus im Online Framework

Zunächst werden die mit der Framerate  $f_b$  aufgezeichneten Bilder in einem Image Buffer zwischengespeichert. Die übliche Framerate liegt bei  $f_b = 30\text{fps}$ . Die Verarbeitung des Imagestreams durch das System erfolgt bildweise. Nach dem Einlesen des Bildes „ $IM$ “ erfolgt die Extraktion der Positionen der 68 Gesichtslanmarken „ $LM$ “ über das Face Tracking Modul (vgl. Kap. 6.1.2). Auf Basis der Landmark-Koordinaten werden die Positionen der 490 Patches entsprechend der Beschreibung in Kapitel 6.3.1 bestimmt. Durch Mittelung der RGB-Werte innerhalb der Patches ergeben sich die zugehörigen Rohsignalwerte „ $raw$ “ für das Einzelbild. Gleichzeitig werden die jeweiligen Hautkonfidenz-Parameter „ $hk$ “ bestimmt. Die bildweise erfassten Daten werden anschließend in einem Signalbuffer der Länge „ $l$ “ zwischengespeichert. Die Matrizen „ $RAW$ “ und „ $HK$ “ werden anschließend zur Selektion der Patches anhand der Hautkonfidenz und des Hochfrequenzrauschens im Zeitfenster „ $l$ “ genutzt (vgl. Abschnitt 6.3.2). Im nächsten Schritt wird auf Basis der gemittelten RGB-Werte der selektierten Patches „ $RAW\text{-}sel$ “ und des in Abschnitt 7.1 beschriebenen Verfahrens der Projektionsvektor berechnet. Das Ergebnis der Projektion ist das BVP-Signal „ $bvp$ “ der Länge „ $l$ “. Durch Anwendung der segmentierten Faltung (Abschnitt 2.4.3) entsteht das BVP-Signal „ $BVP$ “ der Länge „ $L$ “, welches in einem zweiten Signalbuffer abgelegt wird. Dieser wird beim ersten Start des Systems initial gefüllt. Im Anschluss fungiert dieser, ebenso wie der Signalbuffer „ $l$ “, als gleitendes Fenster über den kontinuierlich extrahierten Signalen „ $bvp$ “ und „ $BVP$ “. Die Default-Werte der Buffer-Größen liegen für die gewählte Framerate von  $f_b = 30\text{fps}$  bei  $l = 36\text{frames}$  und  $L = 900\text{frames}$ . Durch Post-Processing mittels des adaptiven Bandpassfilters (7.2.1) entsteht das Signal  $BVP_p$ , auf dessen Basis die Peak Detektion und die Berechnung der Herzfrequenz (7.2) durchgeführt wird. Die Mittelung über die Schlag-zu-Schlag-Distanzen erfolgt dabei im Zeitfenster der letzten 15 Sekunden.

Für die Live-Visualisierung wird über die graphische Benutzeroberfläche das mit den aktuellen Landmarken und den aktiven Patches überlagerte Live-Videobild dargestellt. Gleichzeitig wird das BVP-Signal nach dem Post-Processing ausgegeben. Über eine integrierte Logging-Funktion lassen sich die Daten für nachgelagerte Analysen speichern.



Bild 8.3: GUI zur Live-Visualisierung des PPGI-Systems

## 8.2 Messung der Pulsratenvariabilität im Setting Büro

### 8.2.1 Motivation

In Kapitel 7 konnte gezeigt werden, dass das entwickelte System in der Lage ist, die Herzfrequenz unter Laborbedingungen mit einer mit kontaktbasierten, medizinischen Sensoren vergleichbaren Genauigkeit zu erfassen. Im nächsten Schritt sollen insbesondere zwei Fragestellungen untersucht werden:

- Kann in realen Anwendungsszenarien in Alltagssituationen bei zufälliger Bewegung der Probanden und unterschiedlichen Umgebungsbedingungen eine ähnlich hohe Genauigkeit für die Erfassung der HF erzielt werden?
- Können neben der HF weitere Parameter der Herzaktivität wie die PRV mit einer ausreichenden Genauigkeit aus dem extrahierten BVP-Signal abgeleitet werden?

Neben der Analyse der HF hat sich vor allem die HRV als quantitatives Maß für die Aktivität des vegetativen Nervensystems und der daraus abgeleiteten Beurteilung des Stresslevels und der Erholungsfähigkeit des Körpers etabliert. So wird sie unter anderem zur Analyse der physischen und psychischen Belastung in der Arbeitsmedizin und Arbeitswissenschaft eingesetzt (siehe [121]). Vor diesem Hintergrund wurde das System zunächst im Setting eines Büroarbeitsplatzes evaluiert. Dabei wurde für die Probanden die HF und die HRV während üblicher Tätigkeiten am PC über das entwickelte PPGI-System und ein 1-Kanal EKG-Referenzsystem abgeleitet.

Für die PPGI-basierte Messung der Schlag-zu-Schlag-Varianzen sind jedoch einige Randbedingungen zu beachten. Der Gold-Standard zur Erfassung der HRV ist das EKG, wobei die Abstände der R-Zacken zweier aufeinanderfolgenden QRS-Komplexe zur Berechnung genutzt werden. Bei optischen Messverfahren wie PPGI wird hingegen die Pulsratenvariabilität gemessen, da messtechnisch nicht die elektrische Erregung der Herzkon-

traktion (EKG), sondern die mechanische Ausbreitung der Pulswelle am Messort erfasst wird. Die sich deutlich unterscheidende Signalmorphologie der Pulswelle führt im Vergleich zur R-Zacken-Detektion bei der EKG-Messung zu weniger exakt definierten Peaks (vgl. Abschnitt 2.2.2). Zudem ergibt sich eine höhere Varianz der PRV in Folge physiologischer Einflüsse auf die Ausbreitung der Pulswelle, wie die atmungsabhängige Veränderung des arteriellen Blutdrucks [127]. Aus diesem Grund gilt die PRV schon bei Erfassung mittels kontaktbasierten Verfahren wie der Pulsoximetrie nicht als „Eins-zu-eins-Ersatz“ für die HRV.

Durch das kontaktlose Messprinzip mittels Kamera sind weitere Einflüsse auf die Erfassung der PRV zu erwarten. McDuff hat daher in [97] für die Messung der PRV einen komplexen Hardware-Aufbau bestehend aus neun Kameras verwendet, um eine möglichst große Anzahl nutzbarer Frames bei rotatorischer Bewegung zu generieren. In dieser Arbeit wird hingegen bewusst auf spezielle und komplexe Hardware-Aufbauten verzichtet, da die Vorteile des Systems, neben der Messung über Entfernung hinweg, im einfachen Messsetup liegen. Hinzu kommt, dass im Gegensatz zu den Aufnahmen der Labordatenbank die Einflussfaktoren nicht gezielt zu vorgegebenen Zeitpunkten verändert werden. Die Probandenbewegungen erfolgen zufällig hinsichtlich Zeitpunkt, Dauer und Ausprägung. Auch kommt es zu möglichen Einflüssen durch Teilverdeckungen des Gesichts, sowie nicht einheitlichen Lichtbedingungen. Zusätzlich wird die Untersuchung mit dem Ziel eines leicht handhabbaren Messsetups unter Verwendung von Standard-Kamerahardware (Webcam) durchgeführt. Auf Grund der beschriebenen Unterschiede zum Referenz-Messsystem liegt der Fokus der Analyse nicht im Vergleich der absoluten HRV-Parameter, sondern in der Untersuchung der Korrelation des Verlaufs zwischen PRV und HRV. Daher wird im Vergleich zu McDuff (zwei Minuten) eine längere Aufnahmedauer von 40-60 Minuten je Messung gewählt. Der Nachweis eines ver-

gleichbaren Verlaufs würde den Einsatz des Systems für verschiedene Anwendungszwecke motivieren.

Der Einsatz des Systems für medizinische Untersuchungen wird auf Grund der beschriebenen Unterschiede zur HRV-Messung zunächst ausgeschlossen. Mögliche Anwendungsgebiete sind der AWMF-Leitlinie zur Nutzung der HF und HRV in der Arbeitsmedizin und -wissenschaft [121] entnommen, in der unter anderem der Einsatz als objektives Maß für

- die ergänzende Untersuchung zur Identifizierung von betrieblichen Belastungsschwerpunkten
- den Verlauf der physischen und psychischen Beanspruchung während eines Arbeitstages
- die Evaluierung von allgemeinen Präventionsmaßnahmen, wie z. B. bei Stressbewältigungskursen, Ernährungsumstellungen und sportlichen Betätigungen
- Selbstmonitoring-Zwecke im Bereich des Stressmanagements, sowie Biofeedback- und Entspannungsübungen im Arbeitsalltag

genannt wird. Als konkretes Beispiel kann die Tätigkeit eines Fluglotsen angeführt werden, dessen Arbeits- und Stressbelastung trotz inzwischen zahlreicher Lotsen-Unterstützungssysteme enorm hoch ist. Gleichzeitig ist menschliches Versagen in der Flugsicherung mit einem hohen Risiko für Leib und Leben von Dritten verbunden. Um Fehler durch überhöhte Arbeits- und Stressbelastung zu vermeiden, könnte das System bei Erkennung eines starken Abfalls der PRV eine zuvor definierte Pausenphase motivieren <sup>1</sup>. Generell gewinnt die Untersuchung des Herzkreislaufsystems und das Ableiten psychophysiologischer Größen mit Blick auf dem demografischen Wandel und der sich verändernden Arbeitswelt in der Arbeitswissenschaft und -medizin zunehmend an Bedeutung [121], was die Entwicklung und den Einsatz solcher Systeme in diesem Bereich zusätzlich

---

<sup>1</sup> Das Beispiel des Fluglotsen ist nur als ein potentieller Anwendungsfall anzusehen, eine Belastungsstudie mit Fluglotsen ist nicht Teil der vorgestellten Ergebnisse.

motiviert. Der Hauptvorteil des entwickelten PPGI-Systems liegt dabei im einfachen Messsetup, ohne die Notwendigkeit des An- und Ablegens von am Körper zu tragender Sensorik.

### 8.2.2 Messsetup und Probandenkollektiv

Die Extraktion des BVP-Signals inklusive der zugehörigen Schlag-zu-Schlag-Intervalle für die PRV-Analyse erfolgt mittels des in Abschnitt 8.1 beschriebenen Systems. Als Bildaufnahmesystem wird die Webcam Logitech c920 eingesetzt und auf dem oberen Rand eines externen Bildschirms platziert. Die Bildrate der Kamera liegt bei  $f_b = 30$  fps, die Auflösung wird zu  $640 \times 480$  gewählt. Der Abstand zwischen Proband und Kamera liegt im Bereich von 80 – 110 cm. Als Referenzsystem dient der EKG-Brustgurt ecgMove2 der Firma movisens, welcher ein 1-Kanal-EKG mit einer Abtastfrequenz  $f_a = 256$  Hz liefert. Die ecgMove-Sensoren wurden bereits in zahlreichen wissenschaftlichen Studien eingesetzt (u.a. [127] [128] [125]). Die Analyse der EKG-Daten, inklusive der R-Zacken-Detektion, erfolgt auf Basis des DataAnalyzer-Tools von movisens<sup>2</sup>. Die analysierten EKG-Daten werden zusammen mit den mittels des PPGI-Systems extrahierten Daten, im Unisens-Format (vgl. Abschnitt 4.5) gespeichert und bei Bedarf nochmals händisch synchronisiert. Für die Peak-Detektion der PPGI-Daten werden die extrahierten Signale zuvor auf  $f_{b,int} = 256$  Hz interpoliert. Die Berechnung der HRV- und PRV-Parameter erfolgt offline in Matlab mit Hilfe der mhrv-Bibliothek der PhysioZoo-Plattform [11], welche die gängigen HRV-Parameter im Zeit- und Frequenzbereich aus RR-Intervall-Zeitreihen bestimmt und in der Literatur regelmäßig eingesetzt wird [90].

Die Aufnahmen werden in unterschiedlichen Bürumgebungen und zu unterschiedlichen Tageszeiten durchgeführt. Dabei sind die Probanden instruiert, sich während ihrer üblichen Tätigkeit (PC-Arbeit, Telefongespräche, etc.) möglichst natürlich zu verhalten. Abbildung 8.4 veranschaulicht

---

<sup>2</sup> <https://www.movisens.com/de/produkte/dataanalyzer-2/>

das Mess-Setting. Der Auf- und Abbau des Messsystems inklusive des Anle-

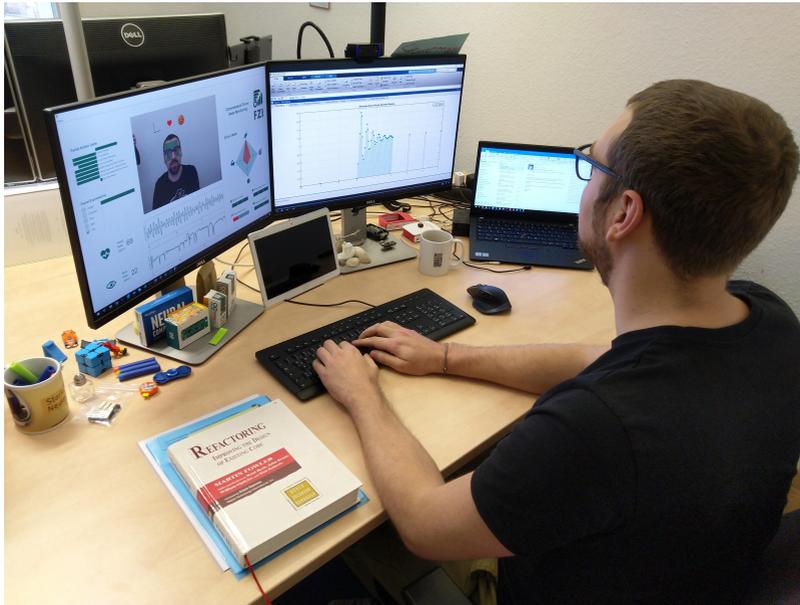


Bild 8.4: Beispiel für Aufnahmen mit dem PPGI-System im Setting Büro

gens des Referenzsystems erfolgt durch den Studienleiter. Die Information der Probanden und die Unterzeichnung der Probandenerklärung verläuft entsprechend dem in Kapitel 4.5 beschriebenen Prozess. Das Probandenkollektiv besteht aus acht männlichen und vier weiblichen Teilnehmern im Alter von 22 bis 34 Jahren. Es wurden 12 Messungen mit einer durchschnittlichen Messdauer von 44 min 20 s und einer Gesamtmessdauer von 8 h 52 min aufgenommen. Tabelle 8.1 zeigt eine Übersicht der Messungen.

Tabelle 8.1: Übersicht der Messungen

Mess-ID	Geschl.	Messdauer	Artefakt- anteil	Anzahl NN	Anzahl PP	Signal- qualität	Besonderheiten	SR <sub>5</sub> [%]	RMSE [bpm]
1	m	47 min 6 s	3,35 %	3947	3935	sehr gut	guter Abstand zum Probanden	97,08	2,83
2	m	53 min 31 s	5,08 %	4167	4151	gut	teilweise Gesicht außerhalb des Bildfeldes	97,37	2,67
3	m	48 min 0 s	5,09 %	3602	3577	gut	Proband mit Brille, häufig skal. Bewegung	97,18	2,05
4	m	37 min 25 s	8,31 %	2962	2949	sehr gut	häufig Hand vor Kinnbereich	97,14	2,04
5	m	37 min 25 s	6,14 %	3160	3126	gut	Proband mit Brille und Bart	95,81	2,60
6	m	45 min 12 s	13,19 %	4060	3941	mittel	Frisurbedingte Vollverdeckung der Stirn	95,91	2,44
7	m	41 min 45 s	12,49 %	2614	2571	gut	Tragen von Kopfhörern	95,14	2,69
8	m	39 min 50 s	10,53 %	2890	2863	mittel	Proband mit Bart	95,60	2,36
9	w	41 min 2 s	1,54 %	3129	3122	sehr gut	Tragen von Kopfhörer, starker seitlicher Lichteinfall	92,92	4,94
10	m	44 min 46 s	8,72 %	2523	2522	sehr gut	Teils Telefonate und Trinkvorgänge	92,72	5,01
11	w	43 min 50 s	1,58 %	3617	3615	sehr gut	Gesichtschnitt im Bild vergleichsweise klein	93,14	3,38
12	w	52 min 10 s	3,61 %	5073	5037	gut	Teils Hand vor Gesicht und Trinkvorgänge	93,29	3,28
<b>Mittel</b>		44 min 20 s	6,64 %	3478	3450			<b>95,28</b>	<b>3,02</b>

### 8.2.3 Auswertung der Herzfrequenzmessung

Bei der Auswertung hinsichtlich der Genauigkeit der Herzfrequenzschätzung wird im Durchschnitt eine Success Rate von  $SR_5 = 95,28\%$  ( $e_t = 5$  bpm) erzielt. Zusätzlich wird der Root Mean Square Error (RMSE) <sup>3</sup> als Maß für die durchschnittliche Abweichung bestimmt. Dieser liegt bei  $RMSE = 3,02$ . Die guten Ergebnisse sind zum einen auf das Anwendungsszenario zurückzuführen, bei dem rotatorische Kopfbewegungen und Verdeckungen des Gesichtsbereichs nur vereinzelt vorkommen. Zum anderen zeigen sie die effektive Umsetzung der entwickelten Algorithmen in ein onlinefähiges System, welches auch bei größerer Messdauer (45 min) robust funktioniert. Die Wahl des Anwendungsszenarios ist, wie bereits erwähnt, mit dem Ziel verbunden, eine Auswertung der Pulsratenvariabilität, ohne zu starken Einfluss von Artefakten, zu ermöglichen.

### 8.2.4 Vergleich von Puls -und Herzratenvariabilität

Nach der AWMF-Leitlinie [121] gilt es für Kurzzeitmessungen unter einer Stunde im Sinne der Qualitätssicherung der HRV-Messung, nichtstationäre Einschwingphasen zu Untersuchungsbeginn und stark artefaktbehaftete Anteile von der Analyse auszuschließen. Daher wird jeweils die erste Minute der Messung von der Berechnung der Parameter ausgenommen und der restliche Teil der Messung mit einem automatischen Artefaktkorrekturverfahren vorverarbeitet. Dies beinhaltet, dass stark artefaktbehaftete Bereiche bei der Berechnung der Zeitparameter ausgeschlossen werden. Der nutzbare Anteil liegt durchschnittlich bei 93,36%. Für die Berechnung der Frequenzparameter werden diese Bereiche mit Hilfe eines Moving-Median-Filters nachbearbeitet, um Unstetigkeiten im Signal zu vermeiden. Die Berechnung der verschiedenen HRV-/PRV-Parameter erfolgt anschließend auf Basis der bereinigten Ballistogramme.

---

<sup>3</sup> <https://www.spektrum.de/lexikon/geowissenschaften/rmse/13702>

Aus den in Abschnitt 8.2.1 genannten Gründen wird an dieser Stelle auf eine detaillierte Auswertung der absoluten PRV-Parameter verzichtet. Bereits für den SDNN-Parameter liegt die prozentuale, mittlere Abweichung zwischen HRV und PRV im Probandenmittel bei  $\mu_{\% \Delta p} = 11,67\%$  und Standardabweichung  $\sigma_{\% \Delta p} = 12,37\%$ , bei den weiteren Parametern ist die Abweichung nochmals größer. Grundsätzlich ergibt sich eine positive, mittlere Abweichung der Absolutwerte im Zeitbereich und eine Tendenz zu höheren Werten der PRV. Dies kann, wie in Abschnitt 8.2.1 bereits dargestellt, einerseits physiologisch durch die Abhängigkeit vom Blutdruck, andererseits durch die Messungenauigkeit des Verfahrens begründet werden. Die hohen Abweichungen für die Parameter zeigen, dass eine direkte Übertragbarkeit von auf HRV-Absolutwerten basierenden Forschungserkenntnissen auf die PRV mit dem aktuellen System nicht möglich ist.

Mit dem Ziel der Beurteilung von Belastungsverläufen am Arbeitsplatz gilt es in erster Linie HRV-Parameter, die sich für Kurzzeitmessungen eignen zu analysieren. So werden nach der beschriebenen Vorverarbeitung die Zeitparameter AVNN, SDNN, RMSSD und pNN50, die Frequenzparameter LF, HF, VLF und HFLF, sowie die nichtlinearen Parameter SD1 und SD2 berechnet (vgl. Abschnitt 2.1.2). Die Berechnung der Parameter erfolgt dabei mit Hilfe eines „Sliding Windows“ mit einer Länge von fünf Minuten (Kurzzeit-HRV [154]) und einer Schrittweite von 30 Sekunden. Zur Untersuchung der Ähnlichkeit des Verlaufs zwischen PRV und HRV wird die Pearson-Korrelation als Metrik genutzt. Der Korrelationskoeffizient nach Pearson ist ein dimensionsloses Maß für die Stärke des linearen Zusammenhangs zwischen den mit der jeweiligen Messmethode erfassten Parameter. Er kann Werte zwischen -1 und +1 annehmen. Ein Wert  $r > 0$  entspricht einem positiven linearen Zusammenhang, ein Wert  $r < 0$  folglich einem negativen. Bei  $r = 0$  liegt kein linearer Zusammenhang zwischen

den Parametern vor. Für den jeweiligen Parameter  $p$  wird der Korrelationskoeffizient mit Fenster  $i$  wie folgt bestimmt:

$$r = \frac{\sum_{i=1}^N (p_{hrv,i} - \overline{p_{hrv}}) (p_{prv,i} - \overline{p_{prv}})}{\sqrt{\sum_{i=1}^N (p_{hrv,i} - \overline{p_{hrv}})^2 \sum_{i=1}^N (p_{prv,i} - \overline{p_{prv}})^2}}, \quad (8.1)$$

Tabelle 8.2 zeigt die Werte der Korrelation der verschiedenen Parameter für die einzelnen Probanden und im Mittel <sup>4</sup>. Im Durchschnitt liegen die Werte der Parameter AVNN, SDNN, LF, VLF und SD2 im hohen bis sehr hohen Bereich. Die Werte für RMSSD, pNN50, HF, SD1 liegen hingegen im geringen bis mittleren Bereich. Für die Probanden 1 und 12 weisen sämtliche Parameter eine mittlere bis sehr hohe Übereinstimmung auf. Abbildung 8.5 zeigt exemplarisch den zugehörigen Verlauf der Parameter für Proband 1 mit hoher Verlaufsähnlichkeit zwischen HRV und PRV. Die hohe Korrelation für eine Vielzahl der untersuchten Parameter, zeigt trotz der Abweichung in den Absolutwerten die prinzipielle Eignung des Systems zum PRV-Verlaufsmonitoring im untersuchten Anwendungsszenario, jedoch beschränkt auf die Analyse relativer Änderungen im Zeitverlauf. Gleichzeitig besteht Optimierungsbedarf hinsichtlich der probandenunabhängigen Qualität der Messergebnisse.

<sup>4</sup> Für die Bewertung der Ähnlichkeit der Verläufe werden fünf Bereiche definiert:  $[0,8 - 1,0]$  → *sehr hoch*;  $[0,6-0,8]$  → *hoch*;  $[0,4-0,6]$  → *mittel*;  $[0,2-0,4]$  → *gering*;  $[0-0,2]$  → *sehr gering*

Tabelle 8.2: Korrelationskoeffizienten zw. PRV und HRV (Fensterbreite 300 s, Schrittweite 10 s)

Mess ID	1	2	3	4	5	6	7	8	9	10	11	12	Mittel	Median
<b>AVNN</b>	0,99	0,97	0,99	0,99	0,94	0,92	0,98	0,67	0,98	0,98	1	0,95	<b>0,95</b>	<b>0,98</b>
<b>SDNN</b>	0,96	0,90	0,92	0,96	0,89	0,67	0,87	0,93	0,80	0,83	0,92	0,82	<b>0,87</b>	<b>0,89</b>
<b>RMSSD</b>	0,81	0,04	-0,13	0,84	0,24	0,41	0,26	0,12	-0,30	0,24	-0,15	0,82	0,27	0,24
<b>pNN50</b>	0,83	0,08	0,26	0,78	0,27	0,14	0,21	-0,25	0,04	-0,4	0,03	0,84	0,24	0,18
<b>LF</b>	0,90	0,86	0,90	0,78	0,35	0,46	0,52	0,89	0,68	0,41	0,94	0,83	<b>0,71</b>	<b>0,81</b>
<b>HF</b>	0,79	0,05	0,26	0,51	0,26	0,28	0,32	0,39	-0,15	0,06	0,15	0,57	0,29	0,27
<b>VLF</b>	0,91	0,96	0,88	0,93	0,66	0,58	0,83	0,77	0,92	0,89	0,99	0,67	<b>0,83</b>	<b>0,88</b>
<b>LFHF</b>	0,69	0,48	0,53	-0,02	0,28	0,19	0,27	0,67	0,55	0,39	0,49	0,53	0,42	0,48
<b>SD1</b>	0,81	0,04	-0,13	0,84	0,24	0,41	0,26	0,12	-0,3	0,24	-0,15	0,82	0,27	0,24
<b>SD2</b>	0,95	0,93	0,93	0,96	0,87	0,72	0,84	0,84	0,87	0,92	0,97	0,81	<b>0,88</b>	<b>0,90</b>

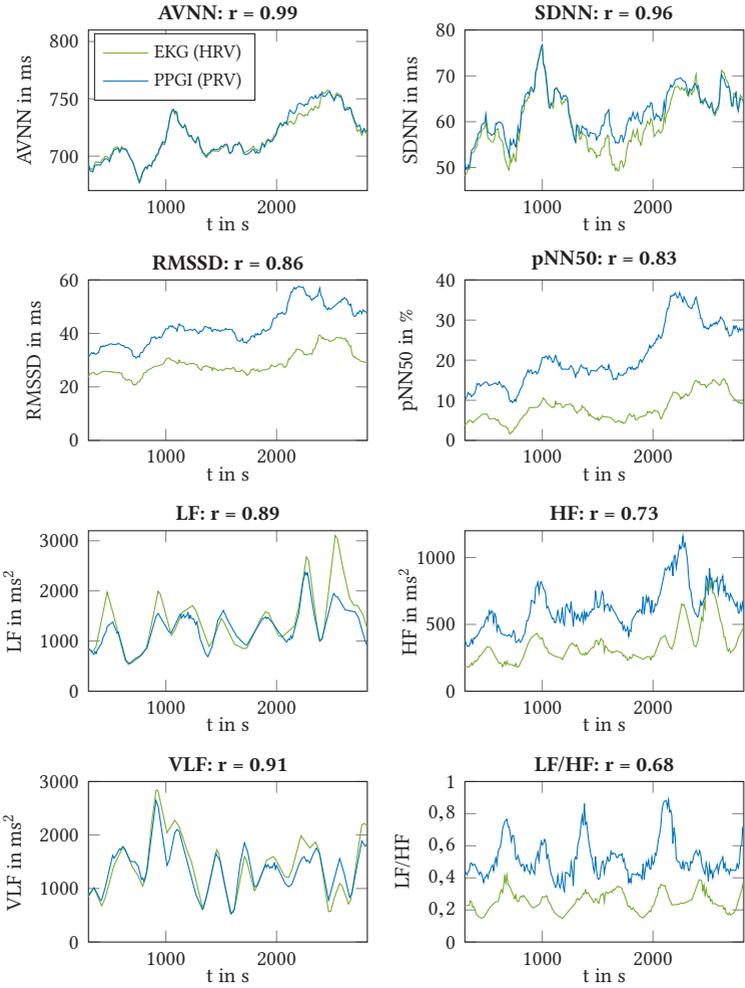


Bild 8.5: Verlauf der HRV-/PRV-Parameter für Proband 1  
(Fensterbreite: 300 s, Schrittweite: 10 s)

### 8.2.5 Diskussion und Fazit

Im Rahmen der Studie wurde der Einsatz des PPGI-Systems unter Verwendung von Off-the-Shelf-Webcams im Setting „Büroarbeitsplatz“ evaluiert. Die Ergebnisse zeigen eine hohe Übereinstimmung der erfassten Herzfrequenz im Vergleich zum EKG-Referenzsystem ( $SR_5 = 95,28\%$ ,  $RMSE = 3,02$ ). Beim Vergleich der PPGI-basierten PRV-Messung mit der EKG-basierten Erfassung der HRV ergibt sich größtenteils ein deutlicher Unterschied in den Absolutwerten einzelner Parameter. Eine Nutzung des PPGI-Systems für die Messung absoluter HRV-/PRV-Werte ist in dieser Form nicht möglich. Dagegen kann durch die Werte der Korrelation für die Parameter AVNN, SDNN, LF, VLF und SD2 (im Mittel  $> 0,7$ ) eine hohe Ähnlichkeit des zeitlichen Verlaufs nachgewiesen werden, was den Einsatz für relative Verlaufsanalysen ermöglicht. Diesbezüglich ist zu erwähnen, dass es in der HRV-Forschung generell an geeigneten alters- und geschlechtsspezifischen Norm- und Grenzwerten der HRV für fundierte gesundheitsbezogene Aussagen oder die Erfassung psychosozialer Belastungen fehlt. Daher wird eine Einschätzung auch beim Einsatz von EKG-Systemen meist nicht auf Basis von Einzelmessungen, sondern mittels standardisierter Verlaufsmessungen getroffen [121]. Die geringere Korrelation beim HF-Parameter ist durch die Bildrate von  $30\text{ fps}$  erklärbar.

Vor dem Hintergrund der nachgewiesenen Verlaufsähnlichkeit kann somit der Einsatz des PPGI-Systems zur Erkennung von Belastungsveränderungen im Tagesverlauf oder längerfristigen Trends (bei annähernd gleichbleibenden Umgebungsbedingungen oder entsprechender Kalibration) motiviert werden. Beispielsweise könnte die positive Auswirkung von Präventionsmaßnahmen durch die Veränderung des sympathisch-parasympathischen Verhältnisses des autonomen Nervensystems hin zu einer höheren parasympathischen Grundaktivität (z.B. erhöhte SDNN oder RMSSD, erniedrigte LF/HF-Ratio) messbar gemacht werden, ohne die Notwendigkeit des Anlegens von Körpersensorik. Gleichwohl gilt es den Zu-

sammenhang zwischen PRV- und Belastungs- bzw. Erholungsverläufen, ebenso wie die Rückwirkungsfreiheit kamerabasierter Systeme, in zukünftigen Studien genauer zu untersuchen.

## **8.3 Einsatz von PPGI im Automobil**

### **8.3.1 Motivation**

Die Erfassung physiologischer Signale im automobilen Kontext hat in den letzten Jahren verstärkt an Bedeutung gewonnen, da diese Informationen über den psychophysiologischen Zustand des Fahrers enthalten, welcher sich wiederum auf dessen Fahrverhalten und Fahrtauglichkeit auswirkt. Denn selbst wenn der technologische Fortschritt im Bereich der Fahrerassistenzsysteme (FAS) das Fahren immer sicherer gestaltet, stellt der Fahrer weiterhin ein Hauptrisiko für Unfälle dar. Laut statistischem Bundesamt wurden in Deutschland im Jahr 2016 1871 Übermüdungsunfälle registriert. Die Dunkelziffer liegt jedoch weit darüber. Untersuchungen des Allianz-Zentrums für Technik haben ergeben, dass bei rund einem Drittel der Unfälle Unaufmerksamkeit, dazu zählt auch Schläfrigkeit, eine Rolle spielt. Hier könnte das physiologische Monitoring einen wichtigen Beitrag zur Erfassung der Vigilanz [19], Aufmerksamkeit [135] oder Stressbelastung [7] während der Fahrt liefern.

Bei Projekten und Studien im Bereich der Forschung- und Voraentwicklung lag der Fokus bisher auf kontaktbasierten Ansätzen mit ins Lenkrad integrierten Sensoren [49] oder auf quasi-kontaktbasierten Ansätzen mittels kapazitiver EKG-Messung über im Fahrersitz integrierte Textilelektroden [38] [128]. Bisher sind diese Ansätze unter anderem deshalb nicht in Serienfahrzeugen integriert, da die zusätzliche Sensorik und der sich ergebende Integrationsaufwand mit erhöhten Kosten verbunden ist. Gleichzeitig schränkt der benötigte Kontakt oder der geringe Maximalabstand zum Messobjekt, die Verfügbarkeit derartiger Systeme ein. Eine kamerabasier-

te Lösung bietet nun den Vorteil einer berührungslosen Remote-Messung. Darüber hinaus gibt es eine klare Tendenz zum generellen Einbau von Innenraumkameras, sowohl für die Fahrerbeobachtung als auch für das Insassenmonitoring, wodurch gegebenenfalls keine zusätzliche Sensorik integriert werden müsste.

Im Bereich der Forschungs- und Vorausentwicklung von FAS kann ein Remote-Vitaldatenmonitoring-System bereits heute interessante Informationen liefern. Dabei kann es sowohl für die Fahrerprobung zur Objektivierung der Bewertung von FAS durch Applikations- und Testingenieure oder in Studien zur Analyse von Fahrkomfort und Fahrspaß eingesetzt werden. Über die erfassten psychophysiologischen Parameter könnten Rückschlüsse auf den kognitiven und emotionalen Zustand des Fahrers gezogen werden, als objektives Maß und Referenz für die subjektive Empfindung des Fahrers. Ein wesentlicher Vorteil liegt dabei in der einfachen Anwendbarkeit, da kein Anlegen von Sensoren notwendig ist.

Für den Bereich der Serienentwicklung wird das System zukünftig insbesondere dann interessant, wenn es als SW-Komponente unter Nutzung von bestehender Serienkamerahardware in Fahrzeugen integriert werden könnte. Bei der Entwicklung von Fahrerbeobachtungskameras liegt der Fokus in der Automobilindustrie derzeit auf dem Monitoring der Augenregion des Fahrers, um auf Basis von Blickrichtung, Augenöffnungsgrad und Lidschlussverhalten Warnungen bei erkannter Müdigkeit oder Ablenkung des Fahrers zu geben. Darüber hinaus gelten Fahrerbeobachtungskameras laut einer Studie von ABI Research mit einer Prognose von 17,5 Mio. verkauften Einheiten für das Jahr 2026 als Schlüsseltechnologie für das (teil-)autonome Fahren [116]. Dabei sollen die Systeme unter anderem die Übernahmeregelung von Level-3-Fahrzeugen sicherer und komfortabler gestalten (vgl. Abschnitt 1.1). Im Bereich der passiven Sicherheit werden diese 2D-Systeme durch 3D-Systeme ergänzt, welche mittels Tiefeninformation Größe, Körperhaltung, etc. bestimmen, um beispielsweise das

Auslösen des Airbags zielgerichtet auszuführen. Unabhängig von der Dimensionalität arbeiten die Systeme mit aktiver Nahinfrarot-Beleuchtung, um unter anderem die Verfügbarkeit bei Tag und Nacht zu gewährleisten, ohne den Fahrer zu blenden. Ob ein aktives PPGI-System bei NIR-Einkanalmessungen mit gepulster Beleuchtung einen Beitrag für Sicherheitssysteme leisten kann, gilt es in zukünftigen Arbeiten zu untersuchen. Für hochautomatisierte Level-4- und vollautonome Level-5-Fahrzeuge verfolgen verschiedene Automobilhersteller das Ziel das Wohlbefinden der Insassen zu erfassen, um u.a. Komfortsysteme oder das Fahrverhalten automatisiert anzupassen. Gleichzeitig soll das Infotainment-Angebot beispielsweise um InCar-Office-Anwendungen wie Videotelefonie erweitert werden. In diesem Zusammenhang gibt es Bestrebungen zukünftig neben NIR-Systemen auch RGB- bzw. RGB/IR-Systeme im Fahrzeuginnenraum einzusetzen. Dabei kann ein passives PPGI-System, wie es in dieser Arbeit entwickelt wurde, mit der Erfassung der Herzfrequenz bzw. der Herzratenvariabilität einen wichtigen Parameter für die Einschätzung des Wohlbefindens der Insassen liefern. Ein konkretes Szenario wäre beispielsweise die vitalparameterbasierte Früherkennung der Reisekrankheit (Kinetose), um plötzliche Übelkeit der Insassen, z.B. durch zu langes Lesen bei autonomer Fahrt, zu vermeiden.

### 8.3.2 Abgrenzung und Zielsetzung der Studie

Wie bereits in Kapitel 3.3.6 dargestellt, beschränkt sich die Evaluation von PPGI-Ansätzen im automobilen Kontext im Stand der Forschung und Technik auf vereinzelte Studien, mit einfachen algorithmischen Ansätzen [75] oder begrenzt differenzierter Darstellung der Ergebnisse hinsichtlich des Probandenkollektivs oder Umwelteinflüssen [114][165][14]. Das in dieser Studie eingesetzte System zeichnet sich hingegen durch speziell entwickelte Algorithmen für das Face Alignment, eine adaptive ROI-Bestimmung, spezielle Methoden zur Artefaktkompensation, sowie eine

Schlag-zu-Schlag-Analyse zur Bestimmung der Herzfrequenz aus. Darüber hinaus wurden spezielle Methoden zur Artefaktdetektion entwickelt und integriert, um stark gestörte Signalanteile zu eliminieren. Im Rahmen der Arbeit wurden zwei Studien mit unterschiedlichem Reifegrad des Prototypensystems durchgeführt, wobei das Monitoring des Fahrers im Vordergrund steht. Ziel der ersten Studie ist eine grundsätzliche Einschätzung der Machbarkeit der kamerabasierten Vitaldatenerfassung im automobilen Kontext. Aufbauend auf den Erkenntnissen wurden für die zweite Studie eine neue Kamera integriert und die Algorithmen des Prototypensystems optimiert und erweitert. Bei der Analyse werden unter anderem die Einflüsse der Randbedingungen wie Wetter oder Straßenart, aber auch der Interaktion des Fahrers mit dem Beifahrer auf die Schätzgenauigkeit untersucht. In diesem Zusammenhang wird auch auf den Trade-Off von Genauigkeit und Verfügbarkeit des Systems eingegangen und eine Methode zur Berechnung der System-Konfidenz vorgestellt.

### **8.3.3 Systemintegration**

Das mobile Prototypensystem besteht aus Kamera, USB-Verbindung für Datentransport und Stromversorgung der Kamera, sowie einem Notebook für die Datenauswertung und Live-Visualisierung über die GUI. Zur Stromversorgung ist das Notebook per DC/AC-Spannungswandler an die Autobatterie angeschlossen. Durch das einfache Setting lässt sich das System schnell in ein neues Fahrzeug integrieren. Als Kamera kommen für die erste Studie die Logitech Webcam c920, für die zweite Studie die FLIR CM3-U3-31S4C-CS zum Einsatz. Die Positionierung der Kamera im Fahrzeug zum Monitoring des Fahrers wird in Abschnitt 8.3.4 diskutiert. Als Notebook wird das Lenovo P51 mit Intel Core i7-7700HQ Prozessor (2,8GHZ) und 4 GB NVIDIA Quadro M1200M Grafikprozessor genutzt. Die Analyse-Software läuft dabei als Standalone-Anwendung und benötigt lediglich die Matlab Runtime als Vorabininstallation. Neben der Live-Visualisierung der

Vitalparameter über die GUI bietet das System die Möglichkeit, die extrahierten Parameter und Signale für die nachfolgende Auswertung zu loggen oder über verschiedene Schnittstellen (z.B. TCP/IP) online zur Verfügung zu stellen.

### 8.3.4 Positionierung der Kamera

Im Rahmen dieser Arbeit steht das Monitoring der Herzfrequenz des Fahrers im Vordergrund. Für die Integration der Kamera im Fahrzeuginnenraum kommen dabei mehrere Positionen in Frage. Die möglichen Einbaupositionen sind in Abbildung 8.6 dargestellt und umfassen:

1. Dachmodul (DM)
2. Innenspiegel (Spiegelfuß) (IS)
3. A-Säule (AS)
4. Hutze (HU)
5. Lenksäule bzw. Kombi-Instrument (LS)
6. Mittelkonsole (MK)

Tabelle 8.3 zeigt das Ergebnis der Bewertung der verschiedenen Einbaupositionen. Dabei wurden folgende Kriterien bewertet:

- I) Häufigkeit und Grad der Verdeckung des Fahrergesichts
- II) Gesicht frontal erfassbar
- III) Stärke der direkten Sonneneinstrahlung auf den Sensor
- IV) Stärke der Sicht- und Bedieneinschränkung des Fahrers
- V) Eignung für Integration des Systems

Die Kriterien i.-iii. beziehen sich auf die Eignung für die PPGI-Anwendung, die Kriterien iv.-v. auf Usability und Integrationsaspekte. Die Wahl der Kriterien deckt sich mit der Auswahl in [51], die Bewertung weicht jedoch in Teilen ab. Für Anwendungen im Bereich der Gesichtsanalyse zur Fahrerbeobachtung, ist die frontale Sicht auf das Gesicht wichtig, um ausreichend



Bild 8.6: Mögliche Kamerapositionen für die fahrergerichtete PPGI-Messung

Merkmale zu erfassen und die Komplexität durch Teilverdeckungen zu reduzieren. Um gleichzeitig eine möglichst geringe Einschränkung hinsichtlich Sicht und Bedienung zu gewährleisten liegt die aktuell bevorzugte Position im Bereich der Lenksäule und des Kombi-Instruments (Position 5). Ein aktuelles Beispiel liefert das SuperCruise System des Cadillac CT6 <sup>5</sup>.

Für die im Rahmen der Arbeit durchgeführten Studien wird Position 4 auf der Hutze gewählt, da die Analyse der Einflüsse von Licht- und Bewegungsartefakten zunächst im Vordergrund steht und eine grundsätzliche Machbarkeit gezeigt werden soll. Um eine Einschätzung hinsichtlich des Einflusses von Kamera-Sichtverdeckungen durch das Lenkrad auf die Verfügbarkeit des Systems zu bekommen, wurde in der zweiten Studie eine Fahrt mit Kameraposition 5 durchgeführt.

<sup>5</sup> <https://www.cadillac.com/world-of-cadillac/innovation/super-cruise>

Tabelle 8.3: Bewertung verschiedener Kameraposition für PPGI

Kriterium	1	2	3	4	5	6
	DM	IS	AS	HU	LS	MK
Verdeckung des Gesichts im Bild	o	+	o	+	-	+
Gesicht frontal erfassbar	-	-	-	++	+	-
Sonneneinstrahlung auf Sensor	+	+	-	o	+	o
Sicht-/ Bedieneinschränkung (für Fahrer)	+	+	+	-	o	++
Integrationsaspekte	o	o	+	-	o	+

### 8.3.5 Studie mit initialem Prototypensystem

Die Validierung des Prototypensystems im Rahmen der ersten Studie erfolgte in einem Zwischenstadium der Systementwicklung. Die Zielsetzung war die Untersuchung, ob sich das entwickelte Prototypensystem grundsätzlich für die Anwendung im automobilen Kontext eignet und in welchen Bereichen die größten Herausforderungen, beziehungsweise die größten Optimierungspotentiale liegen. Die gewonnenen Erkenntnisse waren zentraler Bestandteil der weiteren Systementwicklung (vgl. Abschnitt 8.3.6). Als Kamera kam entsprechend der Erstellung der Datenbasis in Kapitel 4 die Logitech Webcam c920 zum Einsatz. Die Aufzeichnung der Videodaten erfolgte im Raw-Format. Für die Erfassung der Messregion wurde dem Entwicklungsstand entsprechend der Fixed-Region-Ansatzes (6.2) genutzt.

## Messungen und Ergebnisse

Für die Studie wurden Fahrten mit einer weiblichen und fünf männlichen Probanden im Alter zwischen 27 und 33 Jahren durchgeführt, von denen drei einen Bart und einer eine Brille trug. Die Dauer der Testfahrten lag zwischen 30 und 40 Minuten. Als Referenzsystem diente der EKG-Gurt Polar-H7 [110], welcher die momentane Herzfrequenz per Bluetooth bereitstellt. Insgesamt wurden 213 Minuten an Daten bei unterschiedlichen Wetter- und Straßenbedingungen aufgezeichnet. Zur synchronen Speicherung der Umwelt- und Umgebungseinflüsse wurde ein Skript erstellt, welches neben dem Live-Bild und dem Status von PPGI- und Referenzsystem, die Möglichkeit zum Logging von Straßenart, Wetter, Tageszeit und Fahrerinteraktion bietet.

Als Evaluationskriterien für die Einschätzung der grundsätzlichen Eignung des Systems dienen das SNR, der RMSE und die  $SR_5$ . Tabelle 8.4 zeigt die über die Probanden gemittelten Ergebnisse. Bei der Auswertung wird die gesamte Dauer einer Messung betrachtet (100%-Systemverfügbarkeit), sodass auch starke Abweichungen im Vergleich zur Referenz, z.B. in Folge von Fehldetektionen des Face-Tracking-Moduls, mit in die Ergebnisse einfließen. Das SNR wird auf dem vorverarbeiteten BVP-Signal (vor der adaptiven Bandpass-Filterung) berechnet. Der vergleichsweise niedrige Durchschnittswert des SNR von  $-9,02$  dB (*im Vergleich:  $-1,8$  dB bei System und Proband in Ruhe*) spiegelt den deutlichen Einfluss der Licht- und Bewegungsartefakte auf die Signalqualität im bewegten, automobilen Setting wieder. Die erzielten Werte für den RMSE und die  $SR_5$  zeigen die grundsätzliche Machbarkeit der PPGI-basierten Herzfrequenzmessung im automobilen Setting, gleichzeitig aber auch den Optimierungsbedarf. Im folgenden Abschnitt werden die Haupteinflussfaktoren auf die Signalqualität beschrieben.

Tabelle 8.4: Ergebnisse der ersten InCar-Studie

<b>Proband</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>Mittel</b>
SNR [dB]	-9,54	-9,02	-9,54	-12,32	-11,19	-10,06	-9,02
RMSE [bpm]	5,64	6,50	11,60	6,42	4,76	6,32	6,87
SR <sub>5</sub> [%]	67,36	69,38	49,80	73,15	76,04	62,81	65,68

## Haupteinflussfaktoren

Die Ermittlung der Haupteinflussfaktoren auf die Genauigkeit der Herzfrequenzschätzung erfolgt auf Basis der Untersuchung stark gestörter Bereiche. Zur Identifikation der Bereiche werden Abweichungen zur Referenz von mehr als 8 bpm und einer Dauer von mehr als 20 Sekunden im Signalverlauf der Herzfrequenz markiert. Diese Bereiche werden anschließend einer visuellen Analyse im Videobild unterzogen. Abbildung 8.7 zeigt beispielhaft für die Testfahrt mit Proband 2 den mittels PPGI- und Referenzsystem ermittelten Verlauf der Herzfrequenz (oben), inklusive der Messabweichung (unten).

Insgesamt ergeben sich sechs Teilbereiche mit starker Störung, welche zu drei Signalabschnitten zusammengefasst sind und jeweils einem Einflussfaktor zugeordnet werden können. Im ersten Bereich kommt es zu mehreren Tracking-Verlusten des Face-Tracking-Moduls bei seitlicher Kopfhaltung während der Interaktion des Fahrers mit dem Beifahrer. Im zweiten Bereich sind die Hautpartien des Gesichts auf Grund direkter, frontaler Sonneneinstrahlung überbelichtet, da die Webcam lediglich eine Regelung auf den gesamten Bildbereich, inklusive dunklem Bildhintergrund, ermöglicht. Der dritte Bereich enthält Störungen durch Kopfbewegungen (Auf- und Abbewegungen) in Folge schlechter Straßenbedingungen.

Im Folgenden sind die Haupteinflussfaktoren [EF-xx], die in den Messfahrten zu einer Verschlechterung der Schätzgenauigkeit führen, zusam-

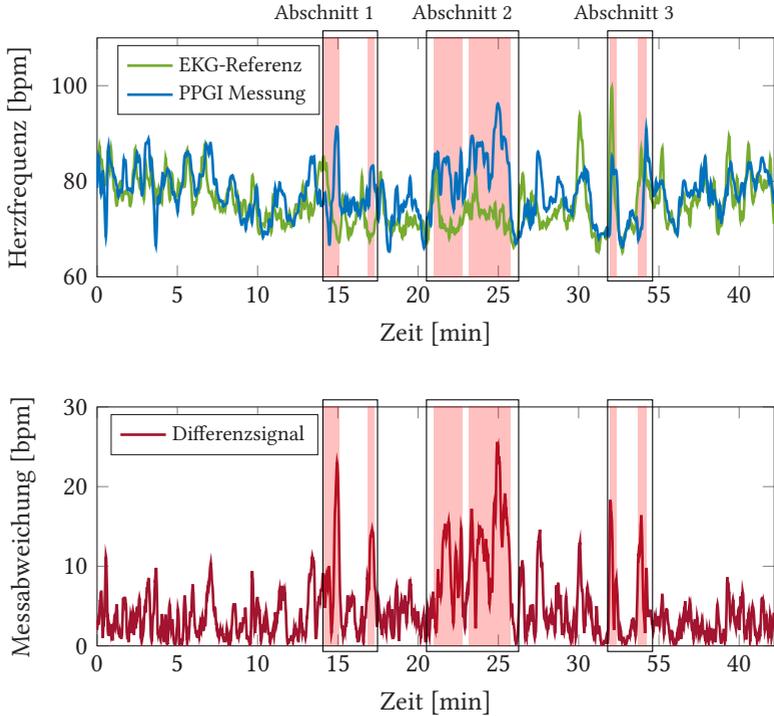


Bild 8.7: Vergleich der Herzfrequenzverläufe unter Einfluss von Artefakten

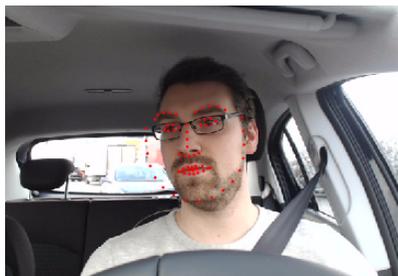
mengefasst. Abbildung 8.8 zeigt zugehörige Beispielbilder aus den Testfahrtvideos.

- [EF-01] *Face Alignment*: Fehldetektion des Face-Trackers, z.B. durch Tragen von Bart und Brille
- [EF-02] *Face Alignment*: Tracking-Verlust des Gesichtes bei starker Kopfdrehung, z.B. bei Schulterblick oder Interaktion mit dem Beifahrer
- [EF-03] *Bewegungsartefakte*: Schnelle Auf- und Abbewegungen des Kopfes bei Bodenwellen und Schlaglöchern

[EF-04] *Lichtartefakte*: Schnelle Lichtwechsel, z.B. bei der Durchfahrt unter Brücken

[EF-05] *Lichtartefakte*: Teilabschattungen des Gesichts bei seitlichem Lichteinfall, die zu über- oder unterbelichteten Regionen führen

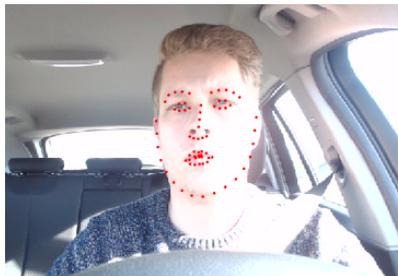
[EF-06] *Kameraregelung*: Überbelichtete Gesichtsregion durch Regelung auf den gesamten Bildbereich



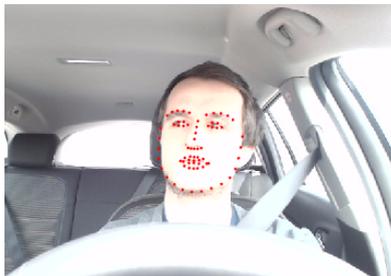
Face Alignment: Bart und Brille



Face Alignment: Interaktion Beifahrer



Lichtartefakte: Teilüberbelichtung



Kameraregelung: Gesicht überbelichtet

Bild 8.8: Beispiele für Artefakte in den Messfahrten bei Studie 1

### 8.3.6 Studie mit optimierten Prototypensystem

Aufbauend auf den Erkenntnissen der ersten Studie werden folgende Lösungsansätze [LA-EF-xx] zur Systemoptimierung definiert:

LA-EF-01 Erweiterung des Trainingsdatensatzes des Face-Trackers

LA-EF-02 Optimierung des Profil-Face-Modells des Face-Trackers

LA-EF-03 Optimierung der Signalverarbeitung hinsichtlich der Artefaktkompensation

LA-EF-04 Artefaktdetektion zur Eliminierung kurzzeitig stark gestörter Signalabschnitte

LA-EF-05 Algorithmus zur adaptiven Anpassung der ROI auf Teilbereiche im Gesicht

LA-EF-06 Integration der Kameraregelung auf den Bildbereich der ROI

Bei der Erweiterung der Trainingsdatensätze [LA-EF-01] des Face Alignment Moduls wurde darauf geachtet eine höhere Diversität der Daten zu erreichen, unter anderem durch stärkere Einbindung herausfordernder Fälle wie Teilverdeckungen (z.B. durch Bart, Brille, Frisur) oder Posen- und Helligkeitsvariationen. Zur weiteren Verbesserung des Face Alignment Modells [LA-EF-02] und der Methoden zur Artefaktkompensation [LA-EF-03] wurde eine Parameteroptimierung hinsichtlich der eingefahrenen Daten unter Realbedingungen durchgeführt. Neben den Optimierungsarbeiten wurde eine Artefaktdetektion zur Eliminierung kurzzeitig stark gestörter Signalabschnitte integriert, welche in Abschnitt 8.3.6 beschrieben ist [LA-EF-04]. Die Analyse der Haupteinflussfaktoren der ersten Studie bestätigt zudem nochmals die Notwendigkeit einer adaptiven ROI-Selektion, deren Entwicklung zum Zeitpunkt der ersten Studie nicht abgeschlossen war. Für die Evaluation des optimierten Prototypensystems wurde daher der in Abschnitt 6.3 beschriebene adaptive Multi-Patch-Ansatz integriert [LA-EF-05]. Hardwareseitig wurde nach den Ergebnissen der ersten Studie eine neue Kamera integriert und angebunden [LA-EF-06]. Die Auswahl der

Kamera inklusive zugehöriger Kriterien wird im folgenden Abschnitt beschrieben. Abschließend werden die Ergebnisse der zweiten Studie mit optimiertem Prototypensystem vorgestellt und diskutiert.

## **Auswahl der Kamera**

Für die Wahl der Kamera für die InCar-PPGI-Messungen in der zweiten Studie wurden zunächst entsprechende Auswahlkriterien erarbeitet. Beispielsweise ist mit Blick auf die BVP-Signalqualität einem niedrigen Sensorausachen eine höhere Priorität zuzuordnen, wie einer hohen Auflösung (vgl. Abschnitt 2.2.4). Tabelle 8.5 zeigt die verschiedenen Kriterien inklusive zugeordneter Priorität von 3 (niedrig) bis 1 (hoch), nach denen der Sensor für die zweite Entwicklungsstufe des Prototyps ausgewählt wurde. In den Spalten zwei und drei sind ein geeigneter Wertebereich und die tatsächlichen Werte der gewählten Kamera FLIR CM3-U3-31S4C-CS mit dem Bildsensor Sony IMX265 angegeben. Dieser besitzt als Global-Shutter CMOS-Sensor gute Eigenschaften für die PPGI-Anwendung. Speziell das geringe zeitliche Rauschen und die hohe spektrale Effizienz im grünen Wellenlängenbereich führen im Vergleich zu anderen Sensoren zu einer für die Anwendung optimierten Bildqualität. Der hohe Dynamikbereich bietet, neben der generellen Verbesserung der Signalqualität, Vorteile bei unterschiedlicher Belichtung innerhalb des Gesichts (vgl. [EF-05]). Ein weiteres wichtiges Merkmal ist die mögliche Regelung der Kamera auf einen bestimmten Bildausschnitt, welcher abhängig von der Position des Fahrgesichts im Bild gewählt werden kann. Auf diese Weise lässt sich insbesondere das Risiko einer Überbelichtung des Gesichtes bei direkter Sonneneinstrahlung reduzieren (vgl. [EF-06]). Auf Grund der geringen Abmessung und der USB3-Vision-Schnittstelle lässt sich die Kamera zudem einfach und schnell integrieren. Als Objektiv wurde ein lichtstarkes 1-Megapixel-Objektiv von Lensagon mit einer Brennweite von  $f = 7,8\text{mm}$  gewählt.

Tabelle 8.5: Kriterien der Kameraauswahl mit optimalem Wertebereich

Parameter	Prio	Wertebereich	Werte FLIR
Framerate [Hz]	2	>30	55
Auflösung [Pixel]	3	>480x600	2048x1536
Zeitliches Dunkel- rauschen [dB]	1	<2,5	2,22
spektr. Effizienz [% bei $\lambda$ ]	1	>50/60/50	53/66/56
Sensorgroße [Zoll]	1	>2/3"	1/1.8"
Pixelgröße [ $\mu\text{m}$ ]	3	>3,5	3,45
Dynamikbereich [dB]	2	>60	71,38
Kamerareglung auf ROI	2	ja	ja

### Artefakt detektion

Bei der Algorithmen-Entwicklung in Kapitel 6 und 7 liegt der Fokus auf der Kompensation von Artefakten, um auch unter Einfluss von Licht- und Bewegungsänderungen ein BVP-Signal mit möglichst guter Signalqualität zu extrahieren. Die Analyse der Messfahrten der ersten Studie hat gezeigt, dass es im automobilen Setting vermehrt zu kurzfristig auftretenden und gleichzeitig stark ausgeprägten Artefakteinflüssen kommen kann, in denen der Nutzsignalbereich so stark von Störung überlagert ist, dass die Extraktion des BVP-Signals nicht mehr möglich ist und die Systemgrenze erreicht ist. In diesem Fall ist die Detektion und Exklusion dieser Signalanteile für die nachgelagerte Herzfrequenzberechnung der Störkompensation vorzuziehen.

Beispiele für Ursachen sind starke Rotationen des Kopfes ( $>90^\circ$ ) beim Schulterblick während Überhol- oder Abbiegevorgängen oder große Heligkeitsänderungen, bei denen es in Folge schneller Licht-Schatten-Wechsel

zu starken Überbelichtungen im Gesichtsbereich kommt. Ein Input der „Artefaktdetektion“ ist daher die Konfidenz des Face-Tracking-Moduls. Dabei wird das Artefaktsignal für das entsprechende Frame zu eins gesetzt, wenn der Konfidenzschwelligwert für die Gesichtserkennung unterschritten wird. Ein weiterer Teil der Artefaktdetektion wird über den Multi-Patch-Ansatz realisiert. Im Algorithmus werden Patches mit einer Vielzahl zu heller oder dunkler Farbwerte in Folge von Über- oder Unterbelichtung einzelner Gesichtspartien für die weitere Verarbeitung ausgeschlossen. Beim Unterschreiten einer kritischen Mindestanzahl nutzbarer Patches wird dies ebenfalls im Artefaktsignal markiert. Auch Hochfrequenz-Störungen durch schnelle Licht-Schatten-Wechsel, wie sie beispielsweise beim Durchfahren von Baumalleen vorkommen, werden durch den Multi-Patch-Ansatz abgefangen. In diesem Fall steigt das Hochfrequenz-Rauschniveau stark an, was zum Ausschluss der Patches (vgl. Abschnitt 6.3.2) und gegebenenfalls zum Unterschreiten der kritischen Mindestanzahl führt.

Das kontinuierlich erfasste Artefaktsignal wird anschließend bei der Berechnung der Herzfrequenz auf Basis der IBIs genutzt. Sind mehr als 20% der zu einem IBI gehörigen Frames im Artefaktsignal markiert, wird es von der Berechnung der Herzfrequenz ausgenommen. Je nach Einstellung des Schwellwertes erhält das System entweder eine höhere Genauigkeit der Schätzung bei geringerer Verfügbarkeit oder umgekehrt eine geringere Genauigkeit bei höherer Verfügbarkeit.

### **Konfidenzzahl**

Im letzten Abschnitt wurde beschrieben, dass sich die Zahl der verfügbaren IBIs für die Herzfrequenzberechnung in einem Signalabschnitt abhängig von der Artefaktdetektion ändern kann. Dabei ändert sich auch die Konfidenz der Schätzung. Mit dem Ziel dem Nutzer die Information der aktuellen Konfidenz online bereitzustellen, wurde ein Ansatz zur Berechnung einer Konfidenzzahl auf Basis momentaner Bild- und Signalinformationen

entworfen. Als Maß zur Bewertung der Qualität dient dabei, neben dem prozentualen Anteil der nicht als Artefakt markierten Bereiche eines Signalabschnitts, die Qualitätsbewertung des nutzbaren Signalanteils. Diese erfolgt anhand der aufgeführten Kriterien:

- Bewegung der ROI (*Differenz bestimmter Landmarkspositionen*)
- Größe der ROI (*Anzahl der verfügbaren Patches*)
- Orientierung des Gesichts (*Kopfposenschätzung des Face Trackers*)
- Änderung der Lichtverhältnisse (*Differenz der Helligkeit im Bild*)
- Analyse des Signalspektrums [0.6 – 4.0Hz] (*Klarheit des Herzfrequenz-Peaks (Peakhöhe bezogen auf Signalenergie)*)
- Analyse des Rauschspektrums [0 – 0.6Hz] (*Peakhöhe der Rauschenenergie*)

Manche Größen zur Qualitätsbewertung lassen sich direkt im Bild bestimmen, wie die Größe bzw. das Vorhandensein der ROI. Für andere Größen, z.B. der Bewegung der ROI, gilt es die Differenz zwischen zwei Frames zu bilden oder wie bei der Analyse der Spektren ein Zeitfenster festzulegen. Insgesamt ist es sinnvoll, sämtliche Größen auf den selben Zeitabschnitt, der auch für die Berechnung der Herzfrequenz genutzt wird, zu beziehen. Gleichzeitig ist auf eine geeignete Normierung der einzelnen Werte zu achten. Auf diese Weise wird zu jedem geschätzten Wert der Herzfrequenz eine zugehörige Konfidenz bestimmt. Die einzelnen Kriterien fließen dabei unterschiedlich gewichtet in die Berechnung der Konfidenzzahl ein:

$$K = \frac{1}{N} \sum_{i=1}^N (a_i \cdot k_i) \quad (8.2)$$

Die jeweiligen Faktoren  $a_i$  lassen sich auf Basis eines bestehenden Datensatzes empirisch ermitteln, beispielsweise über den Zusammenhang zwischen Schätzfehler und Konfidenz.

## Messungen und Ergebnisse

Zur Evaluation des finalen Prototypensystems wurden insgesamt 12 Messfahrten mit einer Dauer von jeweils 35-55 Minuten eingefahren. Die Gesamtdauer der Messfahrten beträgt 528 Minuten. Das Probandenkollektiv setzt sich aus zwei weiblichen und neun männlichen Probanden im Alter zwischen 26 und 64 Jahren zusammen. Keiner der Probanden hat eine Erkrankung des HKS angegeben. Die Messfahrten wurden bei unterschiedlichen Randbedingungen hinsichtlich Wetter (sonnig, bewölkt, regnerisch) und Tageszeit (Vormittag, Nachmittag, Abend) durchgeführt. 10 der 12 Messfahrten erfolgten auf einer definierten Strecke mit einer Gesamtlänge von 44,3 km (Abbildung 8.9), um eine vergleichbare Analyse hinsichtlich des Streckentyps (Innerstädtisch, Landstraße, Autobahn) zu ermöglichen. Bei zwei Fahrten war die Strecke nicht vorgegeben.



Bild 8.9: Vorgegebene Teststrecke mit den Kategorien Stadt (rot), Landstraße (gelb), Autobahn (blau) - (Quelle: Karte erzeugt mit Google Maps)

Bei 11 der 12 Fahrten wurde Kameraposition 4 gewählt (vgl. Abschnitt 8.3.4), mit dem Ziel die Performance des Systems ohne integrationsbedingte Verdeckungen des Kamerasichtfeldes zu analysieren. Eine Fahrt wurde mit Kameraposition 5 durchgeführt, um den Einfluss von zufälligen Verdeckungen durch das Lenkrad initial einschätzen zu können (Abb. 8.10).



Bild 8.10: Visualisierung der Kamerapositionen 4 (li.) und 5 (re.) bei Studie 2

Die Bildrate der Kamera wurde zu  $f_b = 30$  fps, die Auflösung zu  $800 \times 600$  gewählt. Als Referenzsystem wurde erneut der EKG-Brustgurt ecgMove2 der Firma movisens eingesetzt (vgl. Abschnitt 8.2.2). Tabelle 8.6 zeigt eine Übersicht der Messungen zusammen mit den erzielten Werten für den RMSE und die  $SR_5$ .

Tabelle 8.6: Übersicht und Ergebnisse der Messfahrten

Fahrt/ Prob.	Geschl. $m/w$	Strecke	Wetter	Tages- zeit	Inter- aktion	KFZ- Typ	Kam.- pos.	Dauer	SR <sub>5</sub> [%]	RMSE [bpm]
1/1	m	beliebig	sonnig	NM	Keine	Kompakt	4	35 min 6 s	94,00	3,10
2/2	w	beliebig	sonnig	VM	Mittel	Kombi	4	34 min 13 s	88,82	3,74
3/3	m	vorgeg.	sonnig	VM	Hoch	Kombi	4	54 min 9 s	80,05	7,40
4/4	m	vorgeg.	sonnig	A	Hoch	Kombi	4	48 min 59 s	79,23	5,45
5/5	m	vorgeg.	sonnig	NM	Mittel	Kombi	4	44 min 3 s	86,49	5,71
6/6	m	vorgeg.	Regen	NM	Hoch	Kombi	4	49 min 42 s	80,14	4,38
7/7	w	vorgeg.	Regen	NM	Mittel	Kombi	4	40 min 48 s	83,00	4,62
8/8	m	vorgeg.	bewölkt	VM	Hoch	SUV	4	46 min 33 s	79,79	5,00
9/9	m	vorgeg.	bewölkt	VM	Hoch	SUV	4	41 min 43 s	82,54	4,84
10/10	m	vorgeg.	bewölkt	NM	keine	SUV	4	44 min 1 s	92,85	2,63
11/11	m	vorgeg.	bewölkt	NM	Mittel	Kombi 2	4	44 min 23 s	90,50	5,53
12/1	m	vorgeg.	Regen	A	keine	Kombi 2	5	43 min 12 s	86,95	6,04
<b>Mittel</b>								43 min 59 s	<b>85,36</b>	<b>4,87</b>

Im Mittel wird eine Success Rate von  $SR_5 = 85,36\%$  erreicht, der RMSE liegt bei 4,87bpm. Bei der gewählten Einstellung der Artefaktdetektion liegt die Systemverfügbarkeit bei  $\approx 99\%$ . Dies bedeutet, dass nur bei 1% der Messungen nicht genügend IBIs mit ausreichender Qualität für die Aktualisierung der Herzfrequenz verfügbar waren. Bei Erhöhung der Sensitivität der Artefaktdetektion verbessern sich die Werte zu  $SR_5 = 87,23\%$  und  $RMSE = 3,99\text{bpm}$ , gleichzeitig sinkt die Verfügbarkeit des Systems auf  $\approx 92\%$ . Für die Einstellung der Artefaktdetektion ist stets der Trade-Off zwischen Genauigkeit und Verfügbarkeit des Systems zu beachten. Die erzielten Werte zeigen eine deutliche Verbesserung der Genauigkeit der Herzfrequenzschätzung gegenüber dem initialen Prototypensystems ( $SR_5 = 65,68\%$  u.  $RMSE = 6,87\text{bpm}$ , vgl. Abschnitt 8.3.5) und bestätigen die Wirksamkeit der zu Beginn des Unterkapitels beschriebenen Optimierungen.

Tabelle 8.7: Auswertung hinsichtlich verschiedener Einflussfaktoren

<b>Wetter</b>	$SR_5$ [%]	RMSE [bpm]	<b>Inter- aktion</b>	$SR_5$ [%]	RMSE [bpm]	<b>Strecke</b>	$SR_5$ [%]	RMSE [bpm]
sonnig	81,92	6,19	keine	91,27	3,92	Stadt	79,23	5,90
Regen	83,25	5,22	mittel	87,20	4,90	Landstr.	84,15	4,88
Wolkig	86,31	4,16	hoch	80,35	5,41	Autobahn	83,36	4,81

Tabelle 8.7 zeigt die Auswertung der Fahrten hinsichtlich der Einflussfaktoren Wetter, Interaktion mit dem Beifahrer und Streckentyp. Die Ergebnisse weisen einen lediglich geringen negativen Einfluss der Sonne auf die Schätzgenauigkeit des Systems auf. Abb. 8.11 zeigt den Signalverlauf der Herzfrequenzschätzung von Proband 2 bei sonnigen Wetterbedingungen. Der Verlauf der EKG-basierten Schätzung kann durch das entwickelte PPGI-System trotz der herausfordernden Bedingungen (vgl. EF03-EF06 in Abschnitt 8.3.5) gut nachgebildet werden. Dies kann unter anderem durch die adaptive Wahl der ROI und der Exklusion überbelichteter Gesichtspartien begründet werden (siehe Abbildung 8.12).

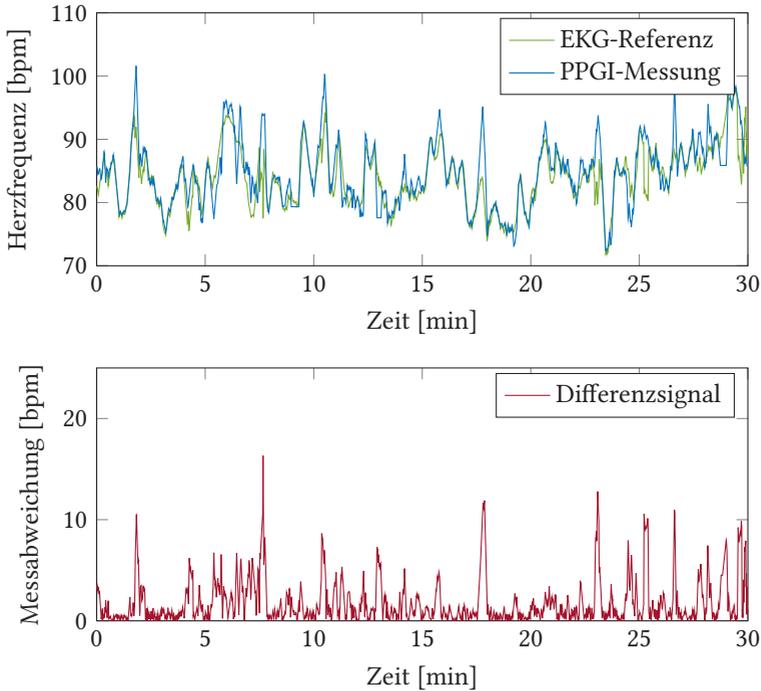


Bild 8.11: Beispielverlauf der Herzfrequenzschätzung bei sonnigen Wetterbedingungen

Zudem kann der Einfluss starker Licht-Schatten-Wechsel, welche beim Durchfahren von Baumalleen oder durch Reflexionen von anderen Verkehrsteilnehmern entstehen, durch die Artefaktdetektion reduziert werden. Die Abbildungen 8.13 und 8.14 zeigen Bildausschnitte und zugehörige Signalverläufe beim Durchfahren einer Brücke. Es ist zu erkennen, dass die automatische Regelung der Kamera dem Wechsel von Schatten und Licht in diesem Fall nicht schnell genug folgen kann, was zunächst zu einer Unterbelichtung (Abbildung 8.13 r.o.), dann zu einer Überbelichtung (l.u.) des Gesichts führt. Die adaptive Patch-Wahl in Kombination mit der Artefaktdetektion ermöglicht, dass die geschätzte Herzfrequenz dem

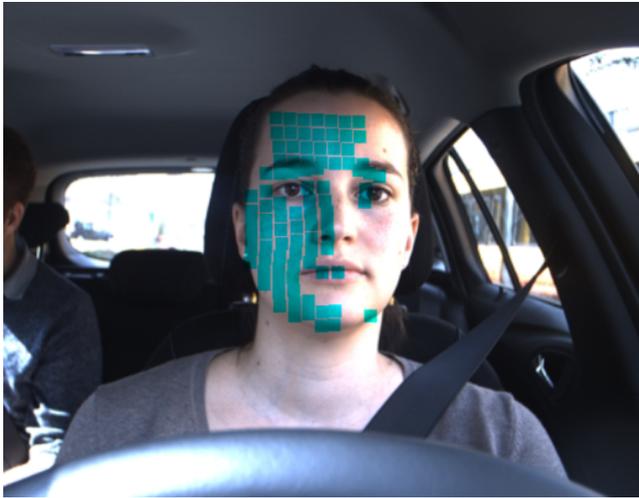


Bild 8.12: Beispiel Sonne

Verlauf der Referenz davon unbeeinträchtigt folgen kann. Im betroffenen Abschnitt  $[-5s; -2s]$  bewirkt die Artefaktdetektion das Unterdrücken des BVP-Signals (Abbildung 8.14 oben), wodurch der Einfluss dieses Bereichs auf die Herzfrequenzschätzung minimiert wird und es lediglich zu einer geringen Abweichung im Vergleich zur Referenz kommt (Abbildung 8.14 unten).

Die Auswertung nach Streckentyp zeigt zunächst, dass die größte Genauigkeit bei Fahrt auf der Autobahn oder Landstraße erzielt werden kann ( $SR_5 \approx 84\%$  und  $RMSE \approx 4,8bpm$ ). Jedoch kann die Abnahme der Genauigkeit bei Stadtfahrten als gering eingestuft werden ( $SR_5 = 79,23\%$  und  $RMSE = 5,90bpm$ ). Dies ist unter anderem auf den POE-Algorithmus und dessen Eignung zur Kompensation von Bewegungsartefakten durch Auf- und Abbewegungen des Kopfes in Folge von Schlechtweg-Strecken oder Start-Stop-Vorgängen zurückzuführen. Die Ergebnisse lassen den Schluss zu, dass das System unabhängig vom Strecken- bzw. Straßentyp eingesetzt werden kann.

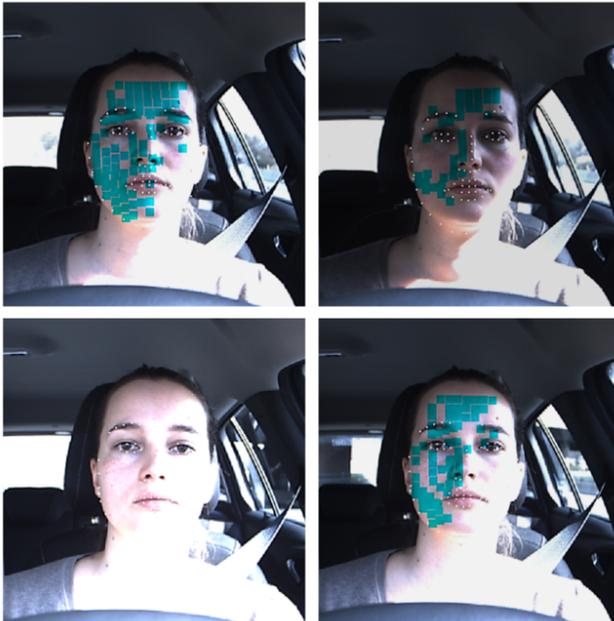


Bild 8.13: Beispiel Brücke: Bildsequenz mit Visualisierung aktiver Patches (grün)

Die größte Herausforderung für das System bleibt die Interaktion mit dem Beifahrer mit Fehldetektionen des Face-Trackers bei starker Rotation des Kopfes. Im Vergleich zur Fahrt ohne Interaktion nimmt die Success-Rate von  $SR_5 = 91,27\%$  auf  $SR_5 = 80,35\%$  bei häufiger Interaktion während der Fahrt ab. Ähnlich verhält es sich bei Nutzung von Kameraposition 5, auch wenn auf Grund von lediglich einer Messfahrt diesbezüglich nur eine erste Einschätzung getroffen werden kann. Der Wert für die Success-Rate sinkt für Messfahrt 12 im Vergleich zu Messfahrt 1 (gleicher Proband) von  $SR_5 = 94,00\%$  auf  $SR_5 = 86,95\%$ , der RMSE steigt von  $RMSE = 3,10$  bpm auf  $RMSE = 6,04$  bpm. Dennoch kann festgehalten werden, dass weder ein hoher Grad an Beifahrer-Interaktion, noch die Positionierung der Kamera an der für Fahrerbeobachtungskameras typischeren Verbauposition 5 im Be-

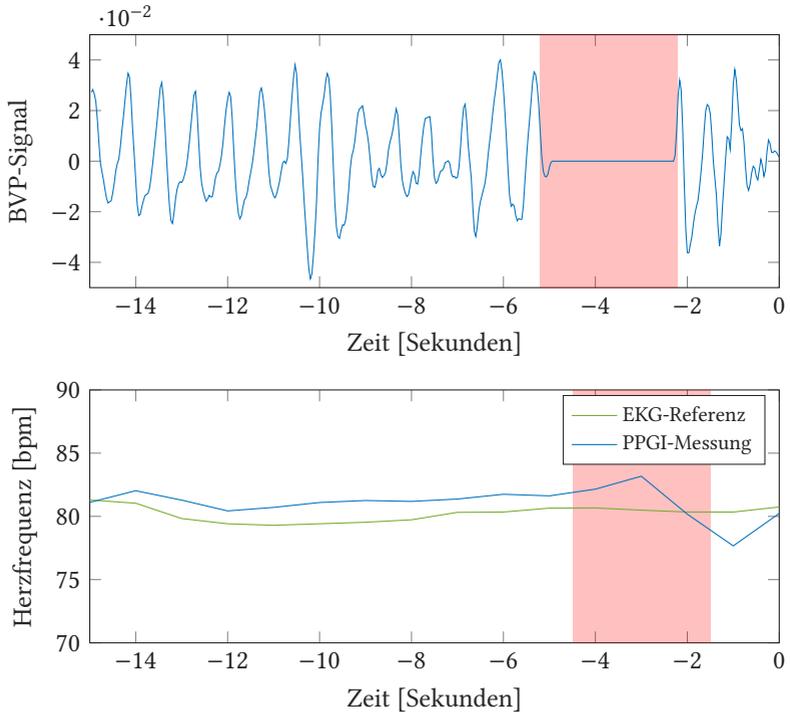


Bild 8.14: Beispielverlauf von BVP-Signal und Herzfrequenzschätzung bei Unterqueren einer Brücke

reich des Kombiinstruments, den Einsatz des PPGI-Systems für die Herzfrequenzmessung ausschließen.

Neben diesen Untersuchungen konnten im Rahmen der Studie keine signifikanten Unterschiede der Genauigkeit in Bezug auf das Geschlecht oder Alter der Probanden festgestellt werden. Da alle Probanden den Hauttypen II-IV entsprechen, kann im Rahmen dieser Studie keine Aussage hinsichtlich des Einflusses besonders heller (Hauttyp I) oder dunkler Haut (Hauttyp VI) getroffen werden. Dies gilt es in zukünftigen Studien zu untersuchen.

### 8.3.7 Diskussion und Fazit

Auf Basis der Ergebnisse der Studie des finalen PPGI-Prototypensystems kann die Machbarkeit der berührungslosen Remote-Herzfrequenzmessung mittels RGB-Farbkameras im Szenario „Fahrerbeobachtung im Automobil“ erfolgreich nachgewiesen werden ( $SR_5 = 85,23\%$  und  $RMSE = 4,87$  bpm). Dabei wurden insgesamt 12 Messfahrten mit einer durchschnittlichen Messdauer von ( $\approx 44$  min) eingefahren. Die Analyse zeigt den positiven Einfluss der Kombination der entwickelten Algorithmen (adaptiver Multi-Patch-Ansatz, POE-Algorithmus, Artefaktdetektion, etc.) auf die im Rahmen der Analyse der Datenbank aus Kapitel 4 und der initialen InCar-Studie herausgearbeiteten, negativen Haupteinflussfaktoren (Bewegungsartefakte, Licht-Schatten-Wechsel, Halbabschattungen des Gesichts, etc.), sodass eine hohe Verfügbarkeit ( $\approx 99\%$ ) und Genauigkeit des Systems unabhängig von Wetterbedingungen, Streckentyp oder Verhalten des Fahrers erzielt werden kann. Der größte Unterschied in der Schätzgenauigkeit ergibt sich für Szenarien mit und ohne Beifahrer-Interaktion. Hardwareseitig beschreibt Abschnitt 8.3.6 die Mindestanforderungen an das Bildaufnahmesystem für den erfolgreichen Einsatz des Systems unter dem Einfluss solch starker Artefakte. Die Verwendung einer Webcam ist für das automobiler Setting nicht mehr ausreichend.

# 9

## Zusammenfassung und Ausblick

Während die Messung physiologischer Parameter zur Beurteilung des Gesundheitszustands seit Jahrzehnten fester Bestandteil im medizinisch klinischen Umfeld ist, erhält sie aktuell vermehrt Einzug in den menschlichen Alltag. Zum einen sollen Krankenhäuser und Ärzte durch das sogenannte „Personal Health Monitoring“ bei Diagnose und Therapie unterstützt und so das Gesundheitssystem entlastet werden. Zum anderen ist eine deutliche Zunahme an nicht-medizinischen Consumer-Geräten zu erkennen, die es ermöglichen, Vitaldaten wie die Herzfrequenz quasi-kontinuierlich zu tracken, um sich selbst, beispielsweise beim Sport oder während des Schlafs, zu monitoren. Mit dem Ziel der Umsetzung möglichst unaufdringlicher Messmethoden („Unobtrusive Measurements“) wurden in diesem Zuge kontaktlose Ansätze für etablierte kontaktbasierte Messverfahren

entwickelt (Beispiel kapazitives EKG) oder die Sensoren in alltagstaugliche Wearables integriert.

Das Verfahren des Photoplethysmography Imaging bietet nun darüber hinaus die Möglichkeit der kamerabasierten Erfassung der Herzfrequenz über Entfernungen einiger Meter hinweg. Der Fokus der wissenschaftlichen Arbeiten lag dabei bisher verstärkt auf dem Entwurf grundlegender Systemansätze und Algorithmen. In dieser Arbeit wurde, durch Analyse, gezielte Weiterentwicklung und Ergänzung bestehender algorithmischer Ansätze, ein gegenüber Artefakten robustes PPGI-System für die kontinuierliche Live-Messung der Herzfrequenz in realen und herausfordernden Anwendungsszenarien entwickelt. Die fokussierten Anwendungsszenarien waren dabei die „Fahrerbeobachtung im Automobil“ und das „Belastungsmonitoring am (Büro-)Arbeitsplatz“. Im Zuge der Systementwicklung und Evaluation wurden unter anderem folgende zentrale Fragestellungen untersucht: Welche systemischen und algorithmischen Konzepte sind für den Einsatz in Alltagsszenarien notwendig? In wie weit lässt sich dabei mithilfe des entwickelten Systems neben der Herzfrequenz auch die Herzfrequenz- bzw. Pulsratenvariabilität ableiten? Mit welcher Genauigkeit kann die Herzfrequenz unter starken Artefakteinflüssen erfasst werden?

Als Grundlage der Entwicklungsarbeiten wurde zunächst eine Datenbank mit Aufnahmen von 26 Probanden in 10 verschiedenen Szenarien mit insgesamt 1040 Minuten Videomaterial erstellt. Das Ziel lag insbesondere in der initialen Einschätzung zum Einfluss von unterschiedlichen Beleuchtungsarten, Bewegungen und Lichtintensitätsänderungen. Bei der Analyse konnte beispielsweise gezeigt werden, dass rotatorische Kopfbewegungen die Signalqualität stärker negativ beeinflussen, als translatorische und skalierende Bewegungen.

Beim Entwurf des Software-Systems wurde der algorithmische Teil in drei Teilkomponenten untergliedert: Einer Bildverarbeitungs Komponente zur kontinuierlichen Erfassung der Messregion, einer Signalverarbei-

---

tungskomponente zur Extraktion und Aufbereitung des BVP-Signals und einer Schätzkomponente zur Bestimmung der Herzfrequenz auf Basis der Schlag-zu-Schlag-Analyse im Zeitbereich. Die zentralen und für den Einsatz in herausfordernden Alltagsszenarien als notwendig identifizierten algorithmischen Bestandteile werden im Folgenden nochmals herausgestellt:

- *Face-Tracking-Modul* zur kontinuierlichen Erfassung des Gesichts als Messregion und darauf basierender Multi-Patch-Algorithmus zur adaptiven ROI-Bestimmung innerhalb des Gesichts
- *POE-Algorithmus* zur Projektion des RGB-Rohsignals in den RGB-Farbraum in Richtung der minimalen Rauschenergie, mit dem Ziel der Trennung von BVP- und Störsignalkomponenten auch unter Einfluss starker Artefakte
- Verfahren zur *adaptiven Bandpassfilterung* des extrahierten BVP-Signals auf Basis einer Spektrogramm-Analyse der BVP- und Artefaktsignale zum Zweck einer verbesserten *Schlag-zu-Schlag-Detektion* für die Herzfrequenzschätzung im Zeitbereich

Unter Verwendung der als Gütemaß eingeführten „Success Rate“ konnte durch die Kombination der Ansätze für das Datenbank-Szenario mit ruhenden Probanden eine Übereinstimmung der Herzfrequenzschätzung mit der EKG-Referenz von 99% bei einer Fehlertoleranz von  $e_t = 3$  bpm nachgewiesen werden. Bei der Evaluation des Gesamtsystems im Setting des „Büroarbeitsplatzes“ lagen die Werte in einem ähnlichen Bereich ( $SR_5 = 95,28\%$  mit  $e_t = 5$  bpm,  $RMSE = 3,02$  bpm). Im stationären Setting liefert das System somit eine mit kontaktbasierten Sensoren vergleichbare Messgenauigkeit für die Herzfrequenz. Beim Vergleich der PPGI-basierten PRV-Messung mit der EKG-basierten Erfassung der HRV ergab sich größtenteils ein deutlicher Unterschied in den Absolutwerten der HRV-Parameter. Eine Nutzung des PPGI-Systems für die Messung absoluter HRV-/PRV-Werte ist in dieser Form nicht möglich. Dagegen konnte durch die erzielten Korre-

lationswerte für die Parameter AVNN, SDNN, LF, VLF und SD2 (im Probandenmittel  $r \geq 0.7$ ) eine hohe Ähnlichkeit des zeitlichen Verlaufs nachgewiesen werden, was den Einsatz für relative Verlaufsanalysen, etwa zur Erkennung von Belastungsveränderungen im Tagesverlauf, motiviert. Dabei gilt es den Zusammenhang zwischen PRV- und Belastungsverläufen in zukünftigen Studien detailliert zu untersuchen.

Abschließend wurde das System im mobilen Setting der Fahrerbeobachtung im Automobil evaluiert und die Haupteinflussfaktoren auf die Signalqualität und die Systemgrenzen herausgearbeitet. Nach einer initialen Probandenstudie hat sich gezeigt, dass der Einsatz einer Webcam in diesem Szenario nicht mehr ausreichend ist ( $SR_5 = 65,68\%$ ,  $RMSE = 6,87$  bpm). Zudem wurde auf Basis der gewonnenen Erkenntnisse eine Artefaktdetektion zur Löschung stark gestörter Signalbereiche integriert, bei denen eine Rekonstruktion des BVP-Signals nicht mehr möglich ist. Die Evaluation des finalen Prototypensystems ergab für 12 Messfahrten mit unterschiedlichen Probanden, Streckentypen und Wetterbedingungen bei einer Gesamtdauer von  $528$  min und einer Systemverfügbarkeit von 99% eine durchschnittliche Success Rate von  $SR_5 = 85,23\%$  und einen Root-Mean-Square-Error von  $RMSE = 4,87$  bpm. Dies zeigt die grundsätzliche Machbarkeit der PPGI-basierten Herzfrequenzmessung auch in stark artefaktbehafteten Szenarien wie der automobilen Anwendung. Weiterhin konnte gezeigt werden, dass der Einfluss der Sonne, sowie von häufigeren Bewegungen im Stadtverkehr durch die entworfenen Algorithmen stark reduziert wird. Den größten Einfluss auf die Schätzgenauigkeit hat eine häufige oder dauerhafte Interaktion mit dem Beifahrer. Insgesamt wurden im Rahmen der Arbeit in vier Datenbanken etwa  $38$  h an Messdaten für die Entwicklung und Evaluation des PPGI-Systems aufgezeichnet.

---

## Ausblick

Um einen zukünftigen Einsatz in Seriensystemen zu motivieren, gilt es für den entwickelten Ansatz im Kontext der Fahrerbeobachtung, weitere Studien mit sehr großen Probandenzahlen (>100) unter Verwendung seriennaher Hardware-Lösungen und Sensorpositionen durchzuführen. Ein Nachteil des gewählten passiven Systemansatzes ist zudem die Notwendigkeit von ausreichendem Umgebungslicht. Dies erschwert den Einsatz des Systems bei Nacht. Derzeit wird in dieser Hinsicht ein aktiver Ansatz mit getakteter Beleuchtung mittels dreier Wellenlängen im Nahinfrarotbereich zur Vermeidung der Blendwirkung untersucht (vgl. Kap. 4.2). Dies ist mit dem Ziel verbunden, große Teile der entwickelten Algorithmen des Mehrkanal-RGB-Ansatzes auf den NIR-Bereich übertragen zu können.

Bei der PPGI-basierten HRV-Analyse sollte die Aussagekraft hinsichtlich sekundärer Parameter wie Stress oder psychischer Belastung genauer analysiert werden. In diesem Zuge gilt es auch die Rückwirkungsfreiheit kamerabasierter Ansätze zu untersuchen. Zudem bietet das System Potentiale zur Erfassung weiterer Vitalparameter wie der Pulse Transit Time (PTT) oder dem relativen Blutdruck.



# Literaturverzeichnis

- [1] *Optimal wavelength selection for noncontact reflection photoplethysmography*, Bd. 8011, 2011. <https://doi.org/10.1117/12.903190>.
- [2] Alghoul, K., S. Alharthi, H. Al Osman und A. El Saddik: *Heart rate variability extraction from videos signals: ICA vs. EVM comparison*. IEEE Access, 5:4711–4719, 2017.
- [3] Amelard, R., C. Scharfenberger, A. Wong und D.A. Clausi: *Illumination-compensated non-contact imaging photoplethysmography via dual-mode temporally coded illumination*. In: *Multimodal Biomedical Imaging X*, Bd. 9316, S. 931607. International Society for Optics and Photonics, 2015.
- [4] Antink, C. H., H. Gao, C. Brüser und S. Leonhardt: *Beat-to-beat heart rate estimation fusing multimodal video and sensor data*. Biomedical optics express, 6(8):2895–2907, 2015.
- [5] Antoni, H.: *Erregungsphysiologie des Herzens*, S. 472–497. Springer Berlin Heidelberg, Berlin, Heidelberg, 1997.
- [6] Asthana, A., S. Zafeiriou, S. Cheng und M. Pantic: *Incremental face alignment in the wild*. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, S. 1859–1866, 2014.

- [7] Baek, H. J., H. B. Lee, J. S. Kim, J. M. Choi, K. K. Kim und K. S. Park: *Nonintrusive biological signal monitoring in a car to evaluate a driver's stress and health state*. *Telemedicine and e-Health*, 15(2):182–189, 2009.
- [8] Baker, S. und I. Matthews: *Lucas-kanade 20 years on: A unifying framework*. *International journal of computer vision*, 56(3):221–255, 2004.
- [9] Balakrishnan, G., F. Durand und J. Guttag: *Detecting pulse from head motions in video*. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, S. 3430–3437, 2013.
- [10] Beauchaine, T.P. und J.F. Thayer: *Heart rate variability as a transdiagnostic biomarker of psychopathology*. *International Journal of Psychophysiology*, 98(2, Part 2):338 – 350, 2015. <http://www.sciencedirect.com/science/article/pii/S0167876015300209>, *Psychophysiological Science and the Research Domain Criteria*.
- [11] Behar, J., A. Rosenberg, I. Weiser-Bitoun, O. Shemla, A. Alexandrovich, E. Konyukhov und Y. Yaniv: *PhysioZoo: A Novel Open Access Platform for Heart Rate Variability Analysis of Mammalian Electrocardiographic Data*. *Frontiers in Physiology*, 9:1390, Okt. 2018.
- [12] Belhumeur, P.N., D.W. Jacobs, D.J. Kriegman und N. Kumar: *Localizing parts of faces using a consensus of exemplars*. *IEEE transactions on pattern analysis and machine intelligence*, 35(12):2930–2940, 2013.
- [13] Blöcher, T., S. Krause, K. Zhou, J. Zeilfelder und W. Stork: *VitalC-amSet - a dataset for Photoplethysmography Imaging*. In: *2019 IEEE Sensors Applications Symposium (SAS)*, S. 1–6, March 2019.
- [14] Blöcher, T., J. Schneider, M. Schinle und W. Stork: *An online PPGI approach for camera based heart rate monitoring using beat-to-beat*

- detection. In: *2017 IEEE Sensors Applications Symposium (SAS)*, S. 1–6, March 2017.
- [15] Blöcher, T., K. Zhou, S. Krause und W. Stork: *An Adaptive Bandpass Filter Based on Temporal Spectrogram Analysis for Photoplethysmography Imaging*. In: *2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)*, S. 1–6. IEEE, 2018.
- [16] Bobbia, S., R. Macwan, Y. Benezeth, A. Mansouri und J. Dubois: *Unsupervised skin tissue segmentation for remote photoplethysmography*. *Pattern Recognition Letters*, Okt. 2017.
- [17] Bolz, A. und W. Urbaszek: *Technik in der Kardiologie: Eine interdisziplinäre Darstellung für Ingenieure und Mediziner*. Springer Berlin Heidelberg, 2013. <https://books.google.de/books?id=jpf3BQAAQBAJ>.
- [18] Bousefsaf, F., C. Maaoui und A. Pruski: *Automatic selection of webcam photoplethysmographic pixels based on lightness criteria*. *Journal of Medical and Biological Engineering*, 37(3):374–385, 2017.
- [19] Brookhuis, K. A. und D. de Waard: *Monitoring drivers' mental workload in driving simulators using physiological measures*. *Accident Analysis & Prevention*, 42(3):898–903, 2010.
- [20] Brüser, C., C. H. Antink, T. Wartzek, M. Walter und S. Leonhardt: *Ambient and Unobtrusive Cardiorespiratory Monitoring Techniques*. *IEEE Reviews in Biomedical Engineering*, 8:30–43, 2015.
- [21] Brueser, C., C. Antink, T. Wartzek, M. Walter und S. Leonhardt: *Ambient and Unobtrusive Cardiorespiratory Monitoring Techniques*. *IEEE reviews in biomedical engineering*, 8, März 2015.
- [22] Butler, M., J. A. Crowe, B. R. Hayes-Gill und P. I. Rodmell: *Motion limitations of non-contact photoplethysmography due to the optical and topological properties of skin*. *Physiological measurement*, 37(5):N27, 2016.

- [23] Camm, A. J. M. M., M. Malik, J. T. G. B. Bigger, G. Breithardt, S. Cerutti, R. Cohen, P. Coumel, E. Fallen, H. Kennedy, R. E. Kleiger *et al.*: *Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology.* *Circulation*, 93(5):1043–1065, 1996.
- [24] Cardoso, J. . und B. H. Laheld: *Equivariant adaptive source separation.* *IEEE Transactions on Signal Processing*, 44(12):3017–3030, Dec 1996.
- [25] Chen, W. und D. McDuff: *Deepphys: Video-based physiological measurement using convolutional attention networks.* In: *Proceedings of the European Conference on Computer Vision (ECCV)*, S. 349–365, 2018.
- [26] Cheng, J., X. Chen, L. Xu und Z. J. Wang: *Illumination variation-resistant video-based heart rate measurement using joint blind source separation and ensemble empirical mode decomposition.* *IEEE journal of biomedical and health informatics*, 21(5):1422–1433, 2016.
- [27] Chrysos, G. G., E. Antonakos, S. Zafeiriou und P. Snape: *Offline deformable face tracking in arbitrary videos.* In: *Proceedings of the IEEE international conference on computer vision workshops*, S. 1–9, 2015.
- [28] Ciccone, A., J. Siedlik, J. M. Wecht, J. Deckert, N. D. Nguyen und J. Weir: *Reminder: RMSSD and SD1 are Identical Heart Rate Variability Metrics: RMSSD and SD1, Identical Metrics.* *Muscle Nerve*, 56, Jan. 2017.
- [29] Claus, W. und C. Claus: *Humanbiologie kompakt.* Spektrum Akademischer Verlag, 2009.
- [30] Cootes, T. F., G. J. Edwards und C. J. Taylor: *Active Appearance Models.* *IEEE Trans. Pattern Anal. Mach. Intell.*, 23:681–685, 2001.
- [31] Cordes, A.: *Monitoring von Atmung und Herzaktivität Mittels Magnetischer Induktionsmessung.* Dissertation, 2013.

- [32] Cristinacce, D. und T. Cootes: *Automatic feature localisation with constrained local models*. Pattern Recognition, 41(10):3054–3067, 2008.
- [33] Cristinacce, D. und T.F. Cootes: *Feature detection and tracking with constrained local models*. In: *Bmvc*, Bd. 1, S. 3. Citeseer, 2006.
- [34] de Haan, G. und V. Jeanne: *Robust Pulse Rate From Chrominance-Based rPPG*. IEEE Transactions on Biomedical Engineering, 60(10):2878–2886, Oct 2013.
- [35] De Haan, G. und A. Van Leest: *Improved motion robustness of remote-PPG by using the blood volume pulse signature*. Physiological measurement, 35(9):1913, 2014.
- [36] Drucker, H., C.J.C. Burges, L. Kaufman, A.J. Smola und V. Vapnik: *Support Vector Regression Machines*. In: Mozer, M.C., M.I. Jordan und T. Petsche (Hrsg.): *Advances in Neural Information Processing Systems 9*, S. 155–161. MIT Press, 1997. <http://papers.nips.cc/paper/1238-support-vector-regression-machines.pdf>.
- [37] Edwards, G.J., C.J. Taylor und T.F. Cootes: *Interpreting face images using active appearance models*. Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, S. 300–305, 1998.
- [38] Eilebrecht, B., T. Wartzek, J. Lem, R. Vogt und S. Leonhardt: *Capacitive electrocardiogram measurement system in the driver seat*. ATZ worldwide eMagazine, 113(3):50–55, Mar 2011. <https://doi.org/10.1365/s38311-011-0034-3>.
- [39] Fan, Q. und K. Li: *Noncontact imaging plethysmography for accurate estimation of physiological parameters*. Journal of Medical and Biological Engineering, 37(5):675–685, 2017.

- [40] Feng, L., L. M. Po, X. Xu und Y. Li: *Motion artifacts suppression for remote imaging photoplethysmography*. In: *2014 19th International Conference on Digital Signal Processing*, S. 18–23. IEEE, 2014.
- [41] Gambi, E., A. Agostinelli, A. Belli, L. Burattini, E. Cippitelli, S. Fioretti, P. Pierleoni, M. Ricciuti, A. Sbröllini und S. Spinsante: *Heart rate detection using microsoft kinect: Validation and comparison to wearable devices*. *Sensors*, 17(8):1776, 2017.
- [42] Gil, E., M. Orini, R. Bailon, J. M. Vergara, L. Mainardi und P. Laguna: *Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions*. *Physiological measurement*, 31(9):1271, 2010.
- [43] GmbH, A. A.: *Ablenkung durch moderne Informations- und Kommunikationstechniken und soziale Interaktion bei Autofahrern*. online. <https://www.allianzdeutschland.de/wp-content/uploads/2018/10/allianz-ablenkungsstudie-2016.pdf>, besucht: 2018-10-01.
- [44] Gücüyener, İ.: *A novel design of heartbeat monitoring system for the motor vehicle*. *International journal of injury control and safety promotion*, 23(4):395–399, 2016.
- [45] Haferkorn, H.: *Optik: physikalisch-technische Grundlagen und Anwendungen*. John Wiley & Sons, 2008.
- [46] Hauke, B.: *Basic Calculation of a Boost Converter's Power Stage*. Techn. Ber., Texas Instruments, 2014.
- [47] Hering, E. und R. Martin: *Optik für Ingenieure und Naturwissenschaftler: Grundlagen und Anwendungen*. Carl Hanser Verlag GmbH Co KG, 2017.
- [48] Heuer, S.: *Ambiente kapazitive EKG-Messung - Elektroden, Systeme und Konzepte*. Dissertation, 2011.

- [49] Heuer, S., B. Chamadiya, A. Gharbi, C. Kunze und M. Wagner: *Unobtrusive in-vehicle biosignal instrumentation for advanced driver assistance and active safety*. In: *2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, S. 252–256, Nov 2010.
- [50] Hülsbusch, M.: *Ein bildgestütztes, funktionelles Verfahren zur optoelektronischen Erfassung der Hautperfusion*. Dissertation, RWTH Aachen, 2008.
- [51] Hoch, S.: *Kontextmanagement und Wissensanalyse im kognitiven Automobil der Zukunft*. Dissertation, 2009.
- [52] Hsu, Y., Y.L. Lin und W. Hsu: *Learning-based heart rate detection from remote photoplethysmography features*. In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, S. 4433–4437. IEEE, 2014.
- [53] Huang, R. und L. Dung: *A motion-robust contactless photoplethysmography using chrominance and adaptive filtering*. In: *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, S. 1–4, Oct 2015.
- [54] Huang, R. Y. und L. R. Dung: *Measurement of heart rate variability using off-the-shelf smart phones*. Biomedical engineering online, 15(1):11, 2016.
- [55] Humphreys, K., T. Ward und C. Markham: *Noncontact simultaneous dual wavelength photoplethysmography: a further step toward noncontact pulse oximetry*. Review of scientific instruments, 78(4):044304, 2007.
- [56] Inan, O. T., P. F. Migeotte, K. S. Park, M. Etemadi, K. Tavakolian, R. Casanella, J. Zanetti, J. Tank, I. Funtova, G. K. Prisk *et al.*: *Ballistocardiography and seismocardiography: A review of recent advances*. IEEE journal of biomedical and health informatics, 19(4):1414–1427, 2014.

- [57] Institute, Q. S.: *What is quantified self?* online. <http://qsinstitute.com/about/what-is-quantified-self/>.
- [58] Iozzia, L., L. Cerina und L. Mainardi: *Relationships between heart-rate variability and pulse-rate variability obtained from video-PPG signal using ZCA*. *Physiological measurement*, 37(11):1934, 2016.
- [59] Ishijima, M.: *Monitoring of electrocardiograms in bed without utilizing body surface electrodes*. *IEEE Transactions on Biomedical Engineering*, 40(6):593–594, June 1993.
- [60] Jeong, I. C. und J. Finkelstein: *Introducing contactless blood pressure assessment using a high speed video camera*. *Journal of medical systems*, 40(4):77, 2016.
- [61] Jiang, W. J., S. C. Gao, P. Wittek und L. Zhao: *Real-time quantifying heart beat rate from facial video recording on a smart phone using Kalman filters*. In: *2014 IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom)*, S. 393–396. IEEE, 2014.
- [62] Joshi, V., A. Verma und Y. Singh: *De-noising of ECG Signal Using Adaptive Filter Based on MPSO*. *Procedia Computer Science*, 57:395–402, Dez. 2015.
- [63] K., G.: *Die Hautdurchblutung des Menschen – Möglichkeiten zur Objektivierung von Hautreaktionen*. *Fette, Seifen, Anstrichmittel*, 70(3):177–182. <https://onlinelibrary.wiley.com/doi/abs/10.1002/lipi.19680700309>.
- [64] Kamen, P. W., H. Krum und A. M. Tonkin: *Poincare plot of heart rate variability allows quantitative display of parasympathetic nervous activity in humans*. *Clinical science*, 91(2):201–208, 1996.
- [65] Kammeyer, K. D. und K. Kroschel: *Digitale Signalverarbeitung: Filterung und Spektralanalyse mit MATLAB-Übungen*. Springer-Verlag, 2013.

- [66] Kamshilin, A. A., S. Miridonov, V. Teplov, R. Saarenheimo und E. Nipolainen: *Photoplethysmographic imaging of high spatial resolution*. Biomedical optics express, 2(4):996–1006, 2011.
- [67] Kamshilin, A. A., I. S. Sidorov, L. Babayan, M. A. Volynsky, R. Giniatullin und O. V. Mamontov: *Accurate measurement of the pulse wave delay with imaging photoplethysmography*. Biomedical optics express, 7(12):5138–5147, 2016.
- [68] Kügler, C.: *Venenkrankheiten*, Kap. Kap. 8: Spezielles Management von Venenkrankheiten, S. 326. ABW Wissenschaftsverlag, 2010.
- [69] Kiencke, U., F. P. Leon und H. Jäkel: *Signale und Systeme*. De Gruyter Oldenbourg, 2008.
- [70] Klette, R., A. Koschan und K. Schlüns: *Computer Vision: Räumliche Information aus digitalen Bildern*. Springer-Verlag, 2013.
- [71] Klinker, R., H. Pape, A. Kurtz und S. Silbernagl: *Physiologie*. Georg Thieme Verlag, 2009. <https://books.google.de/books?id=Ew9vAPFduSMC>.
- [72] Koelstra, S., C. Muhl, M. Soleymani, J. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt und I. Patras: *DEAP: A Database for Emotion Analysis; Using Physiological Signals*. IEEE Transactions on Affective Computing, 3(1):18–31, Jan 2012.
- [73] Kong, L., Y. Zhao, L. Dong, Y. Jian, X. Jin, B. Li, Y. Feng, M. Liu, X. Liu und H. Wu: *Non-contact detection of oxygen saturation based on visible light imaging device using ambient light*. Optics express, 21(15):17464–17471, 2013.
- [74] Kumar, M., A. Veeraraghavan und A. Sabharwal: *DistancePPG: Robust non-contact vital signs monitoring using a camera*. Biomedical optics express, 6(5):1565–1588, 2015.

- [75] Kuo, J., S. Koppel, J. L. Charlton und C. M. Rudin-Brown: *Evaluation of a video-based measure of driver heart rate*. Journal of safety research, 54:55–e29, 2015.
- [76] Kuusela, T.: *Heart rate variability (HRV) signal analysis*. Methodological Aspects of Heart Rate Variability Analysis, S. 9–40, 2012.
- [77] Kviesis-Kipge, E. und U. Rubīns: *Portable remote photoplethysmography device for monitoring of blood volume changes with high temporal resolution*. In: *2016 15th Biennial Baltic Electronics Conference (BEC)*, S. 55–58. IEEE, 2016.
- [78] Kwon, S., H. Kim und K. S. Park: *Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone*. In: *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, S. 2174–2177. IEEE, 2012.
- [79] Kwon, S., J. Kim, D. Lee und K. Park: *ROI analysis for remote photoplethysmography on facial video*. In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, S. 4938–4941. IEEE, 2015.
- [80] Lang, F. und P. Lang: *Basiswissen Physiologie*. Springer-Lehrbuch. Springer Berlin Heidelberg, 2007. <https://books.google.de/books?id=jwUjBAAAQBAJ>.
- [81] Le, V., J. Brandt, Z. Lin, L. Bourdev und T. S. Huang: *Interactive facial feature localization*. In: *European conference on computer vision*, S. 679–692. Springer, 2012.
- [82] Lekakis, J. P., N. A. Zakopoulos, A. D. Protogerou, T. G. Papaioannou, V. T. Kotsis, V. C. Pitiriga, M. D. Tsitsirikos, K. S. Stamatelopoulos, C. M. Papamichael und M. E. Mavrikakis: *Arterial stiffness assessed by pulse wave analysis in essential hypertension: relation to 24-h blood pressure profile*. International Journal of Cardiology, 102(3):391 – 395,

2005. <http://www.sciencedirect.com/science/article/pii/S0167527304005467>.
- [83] Lembeck, F.: *Neurogene Mechanismen der Hautdurchblutung*. In: Schnyder, U. W., B. Wüthrich und A. Eichmann (Hrsg.): *Verhandlungen der Deutschen Dermatologischen Gesellschaft*, S. 6–11, Berlin, Heidelberg, 1986. Springer Berlin Heidelberg.
- [84] Lempe, G., S. Zauneder, T. Wirthgen, S. Zipser und H. Malberg: *ROI selection for remote photoplethysmography*. In: *Bildverarbeitung für die Medizin 2013*, S. 99–103. Springer, 2013.
- [85] Leonhardt, S., L. Leicht und D. Teichmann: *Unobtrusive Vital Sign Monitoring in Automotive Environments—A Review*. *Sensors*, 18(9):3080, 2018.
- [86] Lewandowska, M., J. Rumiński, T. Kocejko und J. Nowak: *Measuring pulse rate with a webcam—a non-contact method for evaluating cardiac activity*. In: *2011 federated conference on computer science and information systems (FedCSIS)*, S. 405–410. IEEE, 2011.
- [87] Liu, S., P. C. Yuen, S. Zhang und G. Zhao: *3D mask face anti-spoofing with remote photoplethysmography*. In: *European Conference on Computer Vision*, S. 85–100. Springer, 2016.
- [88] Lucas, B. D., T. Kanade *et al.*: *An iterative image registration technique with an application to stereo vision*. 1981.
- [89] Lucey, P., J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar und I. Matthews: *The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression*. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, S. 94–101. IEEE, 2010.
- [90] M. Moen, J., M. G. Matt, C. Ramirez, K. V. Tarasov, K. Chakir, Y. S. Tarasova, Y. Lukyanenko, K. Tsutsui, O. Monfredi, C. Morrell, S. Tagi-

- rova, Y. Yaniv, T. Huynh, K. Pacak, I. Ahmet und E. G. Lakatta: *Over-expression of a Neuronal Type Adenylyl Cyclase (Type 8) in Sinoatrial Node Markedly Impacts Heart Rate and RhythmData*. *Frontiers in Neuroscience*, 13, Juni 2019.
- [91] Martinek, R., J. Nedoma, M. Fajkus, R. Kahankova, J. Konecny, P. Jan-ku, S. Kepak, P. Bilik und H. Nazeran: *A phonocardiographic-based fiber-optic sensor and adaptive filtering system for noninvasive continuous fetal heart rate monitoring*. *Sensors*, 17(4):890, 2017.
- [92] Martinez, B., M.F. Valstar, X. Binefa und M. Pantic: *Local evidence aggregation for regression-based facial point detection*. *IEEE transactions on pattern analysis and machine intelligence*, 35(5):1149–1163, 2012.
- [93] Matthews, I. und S. Baker: *Active appearance models revisited*. *International journal of computer vision*, 60(2):135–164, 2004.
- [94] McDuff, D., S. Gontarek und R. W. Picard: *Improvements in remote cardiopulmonary measurement using a five band digital camera*. *IEEE Transactions on Biomedical Engineering*, 61(10):2593–2601, 2014.
- [95] McDuff, D., S. Gontarek und R. W. Picard: *Remote detection of photoplethysmographic systolic and diastolic peaks using a digital camera*. *IEEE Transactions on Biomedical Engineering*, 61(12):2948–2954, 2014.
- [96] McDuff, D.J., E. B. Blackford und J. R. Estep: *The impact of video compression on remote cardiac pulse measurement using imaging photoplethysmography*. In: *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, S. 63–70. IEEE, 2017.
- [97] McDuff, D.J., E. B. Blackford und J. R. Estep: *Fusing Partial Camera Signals for Noncontact Pulse Rate Variability Measurement*. *IEEE Transactions on Biomedical Engineering*, 65(8):1725–1739, Aug 2018.

- [98] McMillan, L.: *An Image-Based Approach to Three-Dimensional Computer Graphics*. 1997.
- [99] Mittelstadt, B.: *Personal Health Monitoring*. Jan. 2013.
- [100] Moço, A., S. Stuijk, M. van Gastel und G. de Haan: *Impairing factors in remote-PPG pulse transit time measurements on the face*. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, S. 1358–1366, 2018.
- [101] Moço, A. V., S. Stuijk und G. de Haan: *Skin inhomogeneity as a source of error in remote PPG-imaging*. *Biomedical optics express*, 7(11):4718–4733, 2016.
- [102] Morbiducci, U., L. Scalise, M. De Melis und M. Grigioni: *Optical vibrocardiography: A novel tool for the optical monitoring of cardiac activity*. *Annals of biomedical engineering*, 35(1):45–58, 2007.
- [103] Moretti, G., R. A. Ellis und H. Mescon: *Vascular patterns in the skin of the face*. *Journal of Investigative Dermatology*, 33(3):103–112, 1959.
- [104] Munla, N., M. Khalil, A. Shahin und A. Mourad: *Driver stress level detection using HRV analysis*. In: *2015 International Conference on Advances in Biomedical Engineering (ICABME)*, S. 61–64, Sep. 2015.
- [105] Niestroj, I.: *Praxis der Orthomolekularen Medizin: Physiologische Grundlagen. Therapie mit Mikro-Nährstoffen*. Anwenden & auswählen. Hippokrates, 2000. <https://books.google.de/books?id=CAVc4qAtliMC>.
- [106] Ottenbacher, J., M. Kirst, L. Jatoba, M. Huflejt, U. Grossmann und W. Stork: *Reliable motion artifact detection for ECG monitoring systems with dry electrodes*. In: *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, S. 1695–1698, Aug 2008.

- [107] Patzak, A., Y. Mendoza, H. Gesche und M. Konermann: *Continuous blood pressure measurement using the pulse transit time: comparison to intra-arterial measurement*. Blood pressure, 24(4):217–221, 2015.
- [108] Paul, M., B. Venema, V. Blazek, J. Mühlsteff und S. Leonhardt: *A camera-based multispectral setup for remote vital signs assessment*. In: *EMBECC & NBC 2017*, S. 968–971. Springer, 2017.
- [109] Poh, M. Z., D. J. McDuff und R. W. Picard: *Non-contact, automated cardiac pulse measurements using video imaging and blind source separation*. Optics express, 18(10):10762–10774, 2010.
- [110] Polar GmbH Deutschland: *Produkte - Polar GmbH Deutschland*. online. <https://www.polar.com/de/produkte/>, besucht: 2018-05-11.
- [111] Prakash, S. K. A. und C. S. Tucker: *Bounded Kalman filter method for motion-robust, non-contact heart rate estimation*. Biomedical optics express, 9(2):873–897, 2018.
- [112] Pschyrembel, W.: *Pschyrembel*. Walter de Gruyter, 2013.
- [113] Qi, H., Z. Guo, X. Chen, Z. Shen und Z. J. Wang: *Video-based human heart rate measurement using joint blind source separation*. Biomedical Signal Processing and Control, 31:309–320, 2017.
- [114] Qi, H., Z. J. Wang und C. Miao: *Non-contact driver cardiac physiological monitoring using video data*. In: *2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*, S. 418–422, July 2015.
- [115] Ren, L., H. Wang, K. Naishadham, O. Kilic und A. E. Fathy: *Phase-Based Methods for Heart Rate Detection Using UWB Impulse Doppler Radar*. IEEE Transactions on Microwave Theory and Techniques, 64(10):3319–3331, Oct 2016.

- [116] Research, A.: *Camera-Based Driver Monitoring Systems to be Chief Enablers of Safe, Semi-Autonomous Driving*. online. <https://www.abiresearch.com/press/camera-based-driver-monitoring-systems-be-chief-en/>.
- [117] Rosenow, G.: *Die Plethysmographie und ihre Anwendung als klinische Methode*. In: Kraus, F., E. Meyer, O. Minkowski, F. MÄ¼ller, H. Sahli, A. Schittenhelm, A. Czerny, O. Heubner und L. Langstein (Hrsg.): *Ergebnisse der Inneren Medizin und Kinderheilkunde*, S. 80–115. Springer Berlin Heidelberg, 1919. [http://dx.doi.org/10.1007/978-3-642-90623-7\\_4](http://dx.doi.org/10.1007/978-3-642-90623-7_4).
- [118] Rouast, P. V., M. T. Adam, R. Chiong, D. Cornforth und E. Lux: *Remote heart rate measurement using low-cost RGB face video: a technical literature review*. *Frontiers of Computer Science*, 12(5):858–872, 2018.
- [119] Sagonas, C., E. Antonakos, G. Tzimiropoulos, S. Zafeiriou und M. Pantic: *300 faces in-the-wild challenge: Database and results*. *Image and vision computing*, 47:3–18, 2016.
- [120] Sahindrakar, P., G. de Haan und I. Kirenko: *Improving motion robustness of contact-less monitoring of heart rate using video analysis*. Technische Universiteit Eindhoven, Department of Mathematics and Computer Science, 2011.
- [121] Sammito, S., B. Thielmann, R. Seibt, A. Klussmann, M. Weippert und I. Böckelmann: *AWMF Leitlinie: Nutzung der Herzschlagfrequenz und der Herzfrequenzvariabilität in der Arbeitsmedizin und der Arbeitswissenschaft*, Juni 2014.
- [122] Saragih, J. und R. Göcke: *Learning AAM fitting through simulation*. *Pattern Recognition*, 42(11):2628–2636, 2009.
- [123] Saragih, J. und R. Goecke: *A nonlinear discriminative approach to AAM fitting*. In: *2007 IEEE 11th International Conference on Computer Vision*, S. 1–8. IEEE, 2007.

- [124] Saragih, J. M., S. Lucey und J. F. Cohn: *Deformable model fitting by regularized landmark mean-shift*. International Journal of Computer Vision, 91(2):200–215, 2011.
- [125] Schaaff, K. und M. T. Adam: *Measuring emotional arousal for online applications: Evaluation of ultra-short term heart rate variability measures*. In: *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, S. 362–368. IEEE, 2013.
- [126] Schäfer, A. und J. Vagedes: *How accurate is pulse rate variability as an estimate of heart rate variability?: A review on studies comparing photoplethysmographic technology with an electrocardiogram*. International journal of cardiology, 166(1):15–29, 2013.
- [127] Schneider, J.: *System zur Erfassung kardialer Aktivität mittels Impedanzplethysmographie am Handgelenk*. Dissertation, Karlsruher Institut für Technologie (KIT), 2019.
- [128] Schneider, J., C. Koellner und S. Heuer: *An approach to automotive ECG measurement validation using a car-integrated test framework*. In: *2012 IEEE Intelligent Vehicles Symposium*, S. 950–955. IEEE, 2012.
- [129] Shafer, S. A.: *Using color to separate reflection components*. Color Research & Application, 10(4):210–218, 1985.
- [130] Shaffer, F. und J. Ginsberg: *An Overview of Heart Rate Variability Metrics and Norms*. Frontiers in Public Health, 5:258, Sep. 2017.
- [131] Shao, D., Y. Yang, C. Liu, F. Tsow, H. Yu und N. Tao: *Noncontact monitoring breathing pattern, exhalation flow rate and pulse transit time*. IEEE Transactions on Biomedical Engineering, 61(11):2760–2767, 2014.
- [132] Sharma, I., R. Mehra und M. Singh: *Adaptive filter design for ECG noise reduction using LMS algorithm*. In: *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)*, S. 1–6, Sep. 2015.

- [133] Silbernagl, S.: *Taschenatlas Physiologie*. DeL. Thieme, 2012. [https://books.google.de/books?id=nvff2\\_nmqn4C](https://books.google.de/books?id=nvff2_nmqn4C).
- [134] Soleymani, M., J. Lichtenauer, T. Pun und M. Pantic: *A Multimodal Database for Affect Recognition and Implicit Tagging*. IEEE Transactions on Affective Computing, 3(1):42–55, Jan 2012.
- [135] Solovey, E. T., M. Zec, E. A. Garcia Perez, B. Reimer und B. Mehler: *Classifying driver workload using physiological and driving performance data: two field studies*. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, S. 4057–4066. ACM, 2014.
- [136] Staderini, E. M.: *UWB radars in medicine*. IEEE Aerospace and Electronic Systems Magazine, 17(1):13–18, Jan 2002.
- [137] Stricker, R., S. Müller und H. M. Gross: *Non-contact video-based pulse rate measurement on a mobile service robot*. In: *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, S. 1056–1062. IEEE, 2014.
- [138] Sugita, N., K. Obara, M. Yoshizawa, M. Abe, A. Tanaka und N. Homma: *Techniques for estimating blood pressure variation using video images*. In: *2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, S. 4218–4221. IEEE, 2015.
- [139] Sun, Y., V. Azorin-Peris, R. Kalawsky, S. Hu, C. Papin und S. E. Greenwald: *Use of ambient light in remote photoplethysmographic systems: comparison between a high-performance camera and a low-cost webcam*. Journal of biomedical optics, 17(3):037005, 2012.
- [140] Sun, Y., S. Hu, V. Azorin-Peris, R. Kalawsky und S. E. Greenwald: *Noncontact imaging photoplethysmography to effectively access pulse rate variability*. Journal of biomedical optics, 18(6):061205, 2012.

- [141] Tarassenko, L., M. Villarroel, A. Guazzi, J. Jorge, D. Clifton und C. Pugh: *Non-contact video-based vital sign monitoring using ambient light and auto-regressive models*. *Physiological measurement*, 35(5):807, 2014.
- [142] Thews, G. und P. Vaupel: *Vegetative Physiologie*. Springer Medizin Verlag, 2005.
- [143] Thews, G. und P. Vaupel: *Vegetative Physiologie*. Springer-Lehrbuch. Springer Berlin Heidelberg, 2006. <https://books.google.de/books?id=i3jDUQMwfPcC>.
- [144] Tomasi, C. und T.K. Detection: *Tracking of point features*. Techn. Ber., Tech. Rep. CMU-CS-91-132, Carnegie Mellon University, 1991.
- [145] Tresadern, P. A., P. Sauer und T. F. Cootes: *Additive Update Predictors in Active Appearance Models*. In: *BMVC*, Bd. 2, S. 4. Citeseer, 2010.
- [146] Trumpp, A., P. L. Bauer, S. Rasche, H. Malberg und S. Zaunseder: *The value of polarization in camera-based photoplethysmography*. *Biomedical optics express*, 8(6):2822–2834, 2017.
- [147] Trumpp, A., S. Rasche, D. Wedekind, M. Schmidt, T. Waldow, F. Gajetjen, K. Plötze, H. Malberg, K. Matschke und S. Zaunseder: *Skin detection and tracking for camera-based photoplethysmography using a Bayesian classifier and level set segmentation*. In: *Bildverarbeitung für die Medizin 2017*, S. 43–48. Springer, 2017.
- [148] Trumpp, A., J. Schell, H. Malberg und S. Zaunseder: *Vasomotor assessment by camera-based photoplethysmography*. *Current Directions in Biomedical Engineering*, 2(1):199–202, 2016.
- [149] Tsuchida, Y.: *Regional differences in the skin blood flow at various sites of the body studied by xenon 133*. *Plastic and reconstructive surgery*, 80(5):705–710, 1987.

- [150] Verkruysse, W., M. Bartula, E. Bresch, M. Rocque, M. Meftah und I. Kirenko: *Calibration of contactless pulse oximetry*. *Anesthesia and analgesia*, 124(1):136, 2017.
- [151] Verkruysse, W., L. O. Svaasand und J. S. Nelson: *Remote plethysmographic imaging using ambient light*. *Optics Express*, 16(26):21434, 2008.
- [152] Vicente, J., P. Laguna, A. Bartra und R. Bailón: *Drowsiness detection using heart rate variability*. *Medical Biological Engineering Computing*, 54, Jan. 2016.
- [153] Viola, P. und M. Jones: *Rapid Object Detection using a Boosted Cascade of Simple Features*. Bd. 1, S. I–511, Feb. 2001.
- [154] Voss, A., R. Schroeder, A. Heitmann, A. Peters und S. Perz: *Short-Term Heart Rate Variability—Influence of Gender and Age in Healthy Subjects*. In: *PloS one*, 2015.
- [155] Wang, W., A. C. den Brinker, S. Stuijk und G. de Haan: *Amplitude-selective filtering for remote-PPG*. *Biomedical optics express*, 8(3):1965–1980, 2017.
- [156] Wang, W., A. C. den Brinker, S. Stuijk und G. de Haan: *Robust heart rate from fitness videos*. *Physiological Measurement*, 38(6):1023–1044, may 2017. <https://doi.org/10.1088%2F1361-6579%2Faa6d02>.
- [157] Wang, W., A. den Brinker, S. Stuijk und G. de Haan: *Algorithmic Principles of Remote-PPG*. *IEEE Transactions on Biomedical Engineering*, PP, Sep. 2016.
- [158] Wang, W., S. Stuijk und G. De Haan: *Exploiting spatial redundancy of image sensor for motion robust rPPG*. *IEEE transactions on Biomedical Engineering*, 62(2):415–425, 2014.

- [159] Wang, W., S. Stuijk und G. De Haan: *A novel algorithm for remote photoplethysmography: Spatial subspace rotation*. IEEE transactions on biomedical engineering, 63(9):1974–1984, 2015.
- [160] Wassertheurer, S.: *Pulswelle und Blutdruck: Kurz und bündig!* Journal für Hypertonie - Austrian Journal of Hypertension, 14(2):45–46, 2010.
- [161] Wu, H. Y., M. Rubinstein, E. Shih, J. Guttag, F. Durand und W. Freeman: *Eulerian video magnification for revealing subtle changes in the world*. 2012.
- [162] Xiong, X. und F. De la Torre: *Supervised descent method and its applications to face alignment*. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, S. 532–539, 2013.
- [163] Zaunseder, S., A. Trumpp, D. Wedekind und H. Malberg: *Cardiovascular assessment by imaging photoplethysmography—a review*. Biomedical Engineering/Biomedizinische Technik, 63(5):617–634, 2018.
- [164] Zhang, G., C. Shan, I. Kirenko, X. Long und R. Aarts: *Hybrid optical unobtrusive blood pressure measurements*. Sensors, 17(7):1541, 2017.
- [165] Zhang, Q., Q. Wu, Y. Zhou, X. Wu, Y. Ou und H. Zhou: *Webcam-based, non-contact, real-time measurement for the physiological parameters of drivers*. Measurement, 100:311–321, 2017.
- [166] Zhu, X. und D. Ramanan: *Face detection, pose estimation, and landmark localization in the wild*. In: *2012 IEEE conference on computer vision and pattern recognition*, S. 2879–2886. IEEE, 2012.

# Eigene Veröffentlichungen

- Blöcher, T., J. Schneider, M. Schinle und W. Stork: *An online PPGI approach for camera based heart rate monitoring using beat-to-beat detection*. In: *2017 IEEE Sensors Applications Symposium (SAS)*, März 2017
- Blöcher, T., L. Garralda Iriarte, J. Schneider, C. Zimmermann und W. Stork: *A Low-cost Life Sign Detection Method based on Time Series Analysis of Facial Feature Points*. In: *BIOSIGNALS 2017*, Februar 2017
- Blöcher, T., K. Zhou, S. Krause und W. Stork: *An Adaptive Bandpass Filter Based on Temporal Spectrogram Analysis for Photoplethysmography Imaging*. In: *2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)*, August 2018
- Blöcher, T., S. Krause, K. Zhou, J. Zeilfelder und W. Stork: *VitalC-amSet - a dataset for Photoplethysmography Imaging*. In: *2019 IEEE Sensors Applications Symposium (SAS)*, März 2019
- Schinle, M., J. Schneider, T. Blöcher, J. Zimmermann, S. Chiriac und W. Stork: *A Modular Approach for Smart Home System Architectures Based on Android Applications*. In: *2017 5th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud)*, April 2017.

- Schneider, J., M. Schroth, M. Holzhey, T. Blöcher und W. Stork: An approach to improve impedance plethysmography on the wrist by using adaptive feedback control. In: *2017 IEEE Sensors Applications Symposium (SAS)*, März 2017.

# Abbildungsverzeichnis

2.1	Verlauf von Druck, Strömungsgeschwindigkeit und Gefäß-Gesamtquerschnitt im HKS, <i>Quelle: [142]</i> . . . . .	12
2.2	Erregungsbildung/-leitung im Herzen . . . . .	14
2.3	Exemplarische Form der Pulswelle in den peripheren Arterien	16
2.4	Definition des RR-Intervalls im EKG . . . . .	18
2.5	Lichtabsorption von menschlichem Blut . . . . .	28
2.6	Modell der oberflächennahen Lichtreflexion der Haut . . . . .	33
2.7	PCA im zweidimensionalen Merkmalsraum . . . . .	40
2.8	Darstellung der Basisform als Mesh, sowie lineare Formvariationen ( <i>Quelle: [37]</i> ) . . . . .	42
2.9	Ähnlichkeitstransformation der Face Shape . . . . .	45
2.10	Warping von der Basisform in die Forminstanz . . . . .	46
2.11	Darstellung der Basis-Appearance und Variationen ( <i>Quelle: [37]</i> )	48
2.12	Modellinstanz im AAM (adaptiert von [37]) . . . . .	49
3.1	Initiale Prädiktion der Landmarken mittels Face Detection: (a) Ground Truth Landmarken (b) Initiale Prädiktion , <i>Quelle: [162]</i>	60
3.2	Vergleich zwischen (a) Gauss-Newton-Methode and (b) SDM , <i>Quelle: [162]</i> . . . . .	63

4.1	Bestücktes LED-Beleuchtungsmodul mit 5 Wellenlängen in 6x6-Matrix-Anordnung . . . . .	79
4.2	Anschlussdiagramm des LED-Treibers aus dem Datenblatt . . . . .	80
4.3	Platine zur Ansteuerung der Beleuchtungsmatrix mit insgesamt fünf Treiber-Modulen zur Ansteuerung von jeweils einem LED-Typ . . . . .	83
4.4	Finaler Aufbau des Beleuchtungsmoduls . . . . .	84
4.5	Vergleich eines BVP-Signalausschnittes für verschiedene Videoformate . . . . .	87
4.6	Gesamtüberblick der Versuchsumgebung zur Aufnahme der verschiedenen Szenarien . . . . .	89
4.7	Aufnahmen bei monochromer Beleuchtung mit grünem Licht und im Nahinfrarotbereich . . . . .	92
4.8	Zeitliche Verläufe der Helligkeit, die sich aus den Zeitmustern ergeben . . . . .	92
4.9	Ablauf der rotatorischen Bewegung . . . . .	93
4.10	Ablauf der skalierenden Bewegung . . . . .	93
4.11	Ablauf der translatorischen Bewegung . . . . .	94
4.12	Beispiel Frequenzspektrum eines BVP-Signals mit Fensterfunktion (grün) und Referenz-Herzfrequenz (rot) . . . . .	96
4.13	Definition der ROI für die Szenario-Analyse . . . . .	97
4.14	Definition der Subregionen für die ROI-Analyse . . . . .	97
5.1	Hardware-Aufbau des Gesamtsystems . . . . .	106
5.2	SW-Systemarchitektur des Gesamtsystems . . . . .	109
6.1	Softwarearchitektur des Gesamtsystems - Bildverarbeitung . . . . .	114
6.2	Trainingsprozess für die Modellerstellung . . . . .	117
6.3	Annotation der Landmarken . . . . .	118
6.4	Ausrichten der Trainingsdaten des Formmodells . . . . .	120
6.5	Veranschaulichung des initialen Schätzmodells . . . . .	122

---

6.6	Monte-Carlo-Sampling . . . . .	124
6.7	Anlernen des Regressionsmodells . . . . .	125
6.8	Merkmalsextraktion bei der originalen SDM - SPTS . . . . .	127
6.9	Merkmalsextraktion bei der CAPP-SDM . . . . .	127
6.10	Prozess zum Update der Landmarken . . . . .	128
6.11	Ablaufdiagramm des Face-Tracking-Moduls . . . . .	129
6.12	Vergleich Fixed-Region-Ansatz (li.) und Multi-Patch-Ansatz (re.)	130
6.13	Definition der Pixelsensoren in der Basisform . . . . .	132
6.14	Häufigkeitsverteilung des Hochfrequenzrauschens . . . . .	134
6.15	Pixelsensorpatches im Gesichtsbild. . . . .	135
6.16	Vergleich der ROI des Fixed-Region- (o.) und Multi-Patch-Ansatzes (u.) bei translatorischer Kopfbewegung . . . . .	137
6.17	Vergleich des BVP-Signals bei fester und adaptiver Selektion der ROI . . . . .	138
7.1	Softwarearchitektur des Gesamtsystems - Signalverarbeitung .	142
7.2	Frequenzspektrum von $\tilde{C}_n(t)$ bei seitlicher Kopfbewegung . .	148
7.3	Rohsignal-Spektrum bei Auf- und Abbewegung des Kopfes . .	149
7.4	Vergleich der Wahl der Projektionsebenen im RGB-Farbraum für POS, CHROM und POE . . . . .	150
7.5	Vergleich der algorithmischen Ansätze bei vertikaler Kopfbe- wegung . . . . .	153
7.6	Beispiel für die Herzfrequenzschätzung im Frequenzspektrum	156
7.7	Spektrogramm Analyse. . . . .	160
7.8	BVP-Signale mit unterschiedlicher Filterung . . . . .	163
7.9	Vergleich des BVP-Signals bei POE und POS für das gesamte Video . . . . .	164
7.10	Vergleich des BVP-Signals bei POE und POS für eine Kopfbe- wegung . . . . .	165
8.1	Softwarearchitektur des Gesamtsystems - Analysekomponente	174

8.2	Ablauf des Algorithmus im Online Framework . . . . .	175
8.3	GUI zur Live-Visualisierung des PPGI-Systems . . . . .	177
8.4	Beispiel für Aufnahmen mit dem PPGI-System im Setting Büro	182
8.5	Verlauf der HRV-/PRV-Parameter für Proband 1 ( <i>Fensterbreite:</i> <i>300 s, Schrittweite: 10 s</i> ) . . . . .	188
8.6	Mögliche Kamerapositionen für die fahrergerichtete PPGI-Messung	195
8.7	Vergleich der Herzfrequenzverläufe unter Einfluss von Arte- fakten . . . . .	199
8.8	Beispiele für Artefakte in den Messfahrten bei Studie 1 . . . . .	200
8.9	Vorgegebene Teststrecke mit den Kategorien Stadt (rot), Land- straße (gelb), Autobahn (blau) - ( <i>Quelle: Karte erzeugt mit Goog- le Maps</i> ) . . . . .	206
8.10	Visualisierung der Kamerapositionen 4 (li.) und 5 (re.) bei Stu- die 2 . . . . .	207
8.11	Beispielverlauf der Herzfrequenzschätzung bei sonnigen Wet- terbedingungen . . . . .	210
8.12	Beispiel Sonne . . . . .	211
8.13	Beispiel Brücke: Bildsequenz mit Visualisierung aktiver Patches (grün) . . . . .	212
8.14	Beispielverlauf von BVP-Signal und Herzfrequenzschätzung bei Unterqueren einer Brücke . . . . .	213

# Tabellenverzeichnis

2.1	HRV-Parameter im Frequenzbereich . . . . .	21
2.2	Bewertung des Einflusses verschiedener Kameraeigenschaften auf die PPGI-Messqualität . . . . .	32
3.1	Übersicht über kontaktlose Verfahren zur Messung der Herz- frequenz . . . . .	58
4.1	Übersicht der verschiedenen Videoformate in Matlab . . . . .	86
4.2	Aufgezeichnete Referenzdaten des <i>SOMNOscreen</i> -Systems mit verwendeter Abtastrate. . . . .	88
4.3	Übersicht der Szenarien . . . . .	91
4.4	Gemittelte SNR-Werte der Kameras für die verschiedenen Sze- narien in dB . . . . .	98
4.5	SNR-Werte für verschiedene ROIs in Szenario 101 . . . . .	101
5.1	Definierte Schnittstellen mit Ein- und Ausgangsdaten . . . . .	110
6.1	Vergleich der SNR-Werte für die verschiedenen Szenarien in dB	136
7.1	Vergleich der SNR-Werte für verschiedene Algorithmen in den Datenbank-Szenarien in dB . . . . .	166

7.2	Success Rate für Szenario 101 . . . . .	168
7.3	Success Rate für Szenario 201 . . . . .	169
7.4	Success Rate für Szenario 202 . . . . .	169
7.5	Success Rate für Szenario 203 . . . . .	169
7.6	Success Rate für Szenario 204 . . . . .	170
8.1	Übersicht der Messungen . . . . .	183
8.2	Korrelationskoeffizienten zw. PRV und HRV (Fensterbreite 300 s, Schrittweite 10 s) . . . . .	187
8.3	Bewertung verschiedener Kameraposition für PPGI . . . . .	196
8.4	Ergebnisse der ersten InCar-Studie . . . . .	198
8.5	Kriterien der Kameraauswahl mit optimalem Wertebereich . . . . .	203
8.6	Übersicht und Ergebnisse der Messfahrten . . . . .	208
8.7	Auswertung hinsichtlich verschiedener Einflussfaktoren . . . . .	209

# Anhang

Die folgende Seite zeigt die Einwilligungserklärung, welche vor Aufzeichnung von Videodaten im Rahmen der Datenbankerstellung oder Probandenstudien von den Probanden zur Zustimmung unterzeichnet wurde.

# Einwilligungserklärung

Im Rahmen der Studie

"Erstellung einer Videodatenbasis mit Vitalparameter Referenzdaten zur Evaluierung von PPGI- Algorithmen"

werden folgende personenbezogene Daten von Ihnen erfasst:

- Videoaufnahmen des Gesichts
- Aus Kameradaten ermittelte Herzfrequenz (sekundengenau aufgelöst)
- EKG-Herzfrequenz ermittelt aus Referenzgurt (sekundengenau aufgelöst)

Die für die Datenerhebung verantwortliche Stelle ist das **FZI Forschungszentrum Informatik, Haid- und-Neu-Straße 10-14, 76131 Karlsruhe (datenschutz@fzi.de)**.

Die erhobenen Daten werden ausschließlich für folgende Zwecke genutzt:

- Forschung

Ich: \_\_\_\_\_ (Name, Vorname) willige hiermit ein, dass

- Die oben genannten personenbezogenen Daten für die oben genannten Zwecke vom FZI Forschungszentrum Informatik erhoben und gespeichert werden können.
- Es wird gewährleistet, dass eine Weitergabe der Daten nur zu Forschungszwecken und nach Unterzeichnung einer entsprechenden vertraglichen Vereinbarung zum datenschutzkonformen Nutzung der Daten an Dritte weitergegeben werden. Die erhobenen Daten werden in jedem Fall ohne eine Zuordnung zu meinem Namen an Dritte übermittelt.
- Die erhobenen Daten, ohne eine Zuordnung zu meinem Namen, an andere **Dritte** übermittelt werden können, sofern diese die Daten selbst ausschließlich **nicht-kommerziell** für **Forschungszwecke** nutzen. Dies wird durch das FZI Forschungszentrum Informatik durch den Abschluss einer vertraglichen Vereinbarung sichergestellt.

Ich bestätige, dass ich darüber informiert worden bin, dass

- Mir jederzeit das Recht zusteht diese Einwilligung ohne Angabe von Gründen zu **widerrufen**. In diesem Fall werden alle von mir erhobenen, personenbezogenen Daten **unverzüglich gelöscht**. Wurden meine Daten an Dritte übermittelt, werden diese über den Widerruf informiert und müssen, gemäß vertraglicher Verpflichtung, die Daten ebenfalls **löschen**.
- Die Studienteilnahme und die Unterzeichnung dieser Erklärung vollständig freiwillig ist.
- *Nur für FZI-Angestellte*: Das FZI Forschungszentrum Informatik mich als Arbeitnehmer nicht zur Teilnahme an dieser Studie verpflichtet und für diesen Zweck explizit keinen Gebrauch von seiner Weisungsbefugnis macht. Bei einer Verweigerung der Teilnahme drohen mir keine disziplinarischen oder arbeitsrechtlichen Konsequenzen.

\_\_\_\_\_  
(Ort, Datum)

\_\_\_\_\_  
(Unterschrift)