

Kamerabasierte Messung von Vitalparametern mit verbesserter Störsicherheit

Dissertation

zur Erlangung des akademischen Grades

Doktoringenieur (Dr.-Ing.)

von M.Sc. Michał Rapczyński
geb. am 23.01.1987 in Gdynia

genehmigt durch die Fakultät für Elektrotechnik und Informationstechnik
der Otto-von-Guericke-Universität Magdeburg

Gutachter:

Prof. Dr.-Ing. habil. Ayoub Al-Hamadi

Prof. Dr. rer. nat. Frank Ortmeier

Prof. Dr. rer. nat. Gunther Notni

Promotionskolloquium am 29. Juni 2023

This document is set in Palatino, compiled with pdfL^AT_EX₂ε and Biber.

The L^AT_EX template from Karl Voit is based on KOMA script and can be found online: <https://github.com/novoid/LaTeX-KOMA-template>

Zusammenfassung

Die kontaktfreie, kamerabasierte Messung von Vitalparameter des Menschen bietet diverse Vorteile gegenüber klassischen kontaktbasierten Methoden. Verfahren aus dem Stand der Technik haben jedoch verschiedene Probleme in realistischen Anwendungsszenarien. So verursachen Bewegungen oder Verdeckungen, zum Beispiel durch Haare oder Brillen, Störsignale, welche die Messgenauigkeit einschränken. In dieser Arbeit werden unterschiedliche Ansätze zur Verbesserung der Störsicherheit für die Messung verschiedener Vitalparameter vorgestellt. Dabei werden alle Teile der Signalverarbeitungskette betrachtet. Neue Methoden für die Wahl der Region of Interest, Signalverarbeitung und Korrekturverfahren werden vorgestellt. Zudem wurden grundlegende Untersuchungen des Einflusses der Videokompression auf die Genauigkeit und die Messung der Vitalparameter in einem breiten Spektrum im Nahinfrarotbereich, durchgeführt. Die kontaktlose Vitalparametermessung wurde zudem in zwei Anwendungsfällen validiert, der Lebenderkennung und der Messung der Vitalparameter in einem MRT. Die Messgenauigkeit und Störsicherheit konnte durch die vorgestellten Methoden deutlich verbessert und neue Ansätze für die zukünftige Nutzung der Technologie erschlossen werden.

Abstract

The non-contact, camera-based measurement of human vital signs offers various advantages over classical contact-based methods. However, state-of-the-art methods have various problems in realistic application scenarios. For example, movements or occlusions, e.g. by hair or glasses, cause interfering signals that limit the measurement accuracy. In this work, different approaches to improve the noise robustness for the measurement of different vital signs are presented. All parts of the signal processing chain are considered. New methods for region of interest selection, signal processing and correction techniques are presented. In addition, fundamental studies of the influence of video compression on the accuracy, and measurement of vital signs in a wide spectrum in the near-infrared range, have been carried out. The non-contact vital sign measurement was also validated in two use cases, life detection and vital sign measurement in an MRI. The measurement accuracy and noise immunity could be significantly improved by the presented methods and new approaches for the future use of the technology could be developed.

Liste der Veröffentlichungen

Publikationen

Michal Rapczynski, Philipp Werner und Ayoub Al-Hamadi. »Effects of Video Encoding on Camera Based Heart Rate Estimation«. In: *IEEE Transactions on Biomedical Engineering (Impact Factor : 4.5)* 66.12 (2019), S. 3360–3370. DOI: 10.1109/TBME.2019.2904326

Michal Rapczynski, Philipp Werner, Sebastian Handrich und Ayoub Al-Hamadi. »A Baseline for Cross-Database 3D Human Pose Estimation«. In: *Sensors (Impact Factor: 3.6)* 21.11 (2021), S. 3769. DOI: 10.3390/s21113769. URL: <https://doi.org/10.3390/s21113769>

Marc-Andre Fiedler, Michal, Rapczynski und Ayoub Al-Hamadi. »Fusion-Based Approach for Respiratory Rate Recognition From Facial Video Images«. In: *IEEE Access (Impact Factor: 3.4)* 8 (2020), S. 130036–130047. DOI: 10.1109/ACCESS.2020.3008687

Michal Rapczynski, Philipp Werner, Frerk Saxon und Ayoub Al-Hamadi. »How the Region of Interest Impacts Contact Free Heart Rate Estimation Algorithms«. In: *2018 25th IEEE International Conference on Image Processing (ICIP)*. Okt. 2018, S. 2027–2031. DOI: 10.1109/ICIP.2018.8451846

Michal Rapczynski, Chen Zhang, Ayoub Al-Hamadi und Gunter Notni. »A Multi-Spectral Database for NIR Heart Rate Estimation«. In: *2018 25th IEEE International Conference on Image Processing (ICIP)*. Okt. 2018, S. 2022–2026. DOI: 10.1109/ICIP.2018.8451104

Michal Rapczynski, Philipp Werner und Ayoub Al-Hamadi. »Continuous Low Latency Heart Rate Estimation from Painful Faces in Real Time«. In: *23th International Conference on Pattern Recognition (ICPR)*. 2016

Michal Rapczynski, Erik Lilienblum, Sebastian von Enzberg und Ayoub Al-Hamadi. »Simultaneous multi-camera calibration based on phase-shift measurements on planar surfaces«. In: *IEEE International Instrumentation and Measurement Technology Conference, I2MTC 2014, Proceedings, Montevideo, Uruguay, May 12-15, 2014*. IEEE, 2014, S. 175–180. DOI: 10.1109/I2MTC.2014.6860728. URL: <https://doi.org/10.1109/I2MTC.2014.6860728>

Michal Rapczynski, Chen Zhang, Maik Rosenberger und Ayoub Al-Hamadi. »Multispektrale Vermessung der Haut zur Verbesserung kontaktloser Herzraterschätzung«. In: 22. *Workshop Farbbildverarbeitung - Ilmenau*. 2016

Marc-Andre Fiedler, Michal, Rapczynski und Ayoub Al-Hamadi. »Facial Video-Based Respiratory Rate Recognition Interpolating Pulsatile PPG Rise And Fall Times«. In: 2021 *IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. 2021, S. 545–549. DOI: 10.1109/ISBI48211.2021.9434132

Michal Rapczynski, Frerk Saxen, Philipp Werner und Ayoub Al-Hamadi. »Der Einfluss von Hautfarbensegmentierung auf die kontaktfreie Schätzung von Vitalparametern«. In: 22. *Workshop Farbbildverarbeitung - Ilmenau*. 2016

Michal Rapczynski, Christopher Lang und Ayoub Al-Hamadi. »Verhinderung der Überwindung von Gesichtserkennung durch kamerabasierte Vitalparameterschätzung«. In: 24. *Workshop Farbbildverarbeitung*. 2019

Patente

Ayoub Al-Hamadi; Maria Nisser; Gunther Notni; Thomas Pertsch; Michal Rapczynski; Jan Sperrhake; Chen Zhang. »Verfahren und Vorrichtung zur kontaktfreien Bestimmung von zeitlichen Farb- und Intensitätsveränderungen bei Objekten«. DE. 102020108064A1. 2020

Inhaltsverzeichnis

1	Einleitung	1
2	Stand der Technik	7
2.1	Physiologische Grundlagen	8
2.1.1	Aufbau der menschlichen Haut	8
2.1.2	Lichtabsorption des Gewebes	8
2.1.3	Herzratenvariabilität und Atemrate	10
2.2	Kamerabasierte Vitalparameterschätzung	11
2.2.1	Videokompression	12
2.2.2	Region of Interest	18
2.2.3	Signalverarbeitung	22
2.2.4	Atemratenschätzung	28
2.3	Multispektrale Messungen	29
2.4	Machinelles Lernen	31
2.4.1	DeepPhys	31
2.4.2	PhysNet	32
2.4.3	Extractor-Estimator-CNN	34
2.4.4	RhythmNet	35
3	Neue Methoden für die kamerabasierte Vitalparametermessung	37
3.1	Hauterkennung	37
3.2	Adaptiver Bandpass	40
3.3	Graphenbasierte Peakselektion (IBI-Graph)	42
3.4	Pulserkennung durch LSTM	46
3.5	Bestimmung der Atemrate	48
3.5.1	Vorverarbeitung	48
3.5.2	Modulationen	50
3.5.3	Postprocessing	54
3.5.4	Artefaktreduzierung	54

3.5.5	Bestimmung der Atemrate	55
3.6	Lebenderkennung	56
3.6.1	Gesichtsverifizierung	58
3.6.2	Vitalparameterschätzung	58
4	Experimentelle Ergebnisse	61
4.1	Gesichtsdetektion	61
4.2	Fehlerberechnung	63
4.2.1	Grundwahrheiten	63
4.2.2	IEC Genauigkeit (Herzrate)	63
4.2.3	DR und FDR Genauigkeit (Atemrate)	64
4.3	Datenbanken	65
4.3.1	BioVid	66
4.3.2	BP4D+	67
4.3.3	MMSE HR	68
4.3.4	PURE	68
4.3.5	AtemDB	69
4.4	Videoeigenschaften	70
4.4.1	Allgemeiner Kompressionsfehler	74
4.4.2	Constant Rate Factor (CRF)	75
4.4.3	Farbunterabtastung	76
4.4.4	Vergleich der Region of Interest (ROI)	78
4.4.5	Vergleich der Signalextraktionsmethoden	80
4.4.6	Auflösung	81
4.4.7	Bildwiederholungsrate	86
4.5	Region of Interest (ROI)	89
4.5.1	Landmarkenbasierte ROIs	89
4.5.2	Hautsegmentierung	89
4.5.3	Daten und Signalverarbeitung	91
4.5.4	Ergebnisse	94
4.6	Adaptiver Bandpass	101
4.7	Pulserkennung durch LSTM	102
4.8	Atmung	106
4.8.1	PPG-Signal Generierung	106
4.8.2	Datenbanken	107
4.8.3	Ergebnisse	109
4.8.4	Vergleich	114

4.9	Lebenderkennung	116
4.10	Messung im MRT	119
4.11	Multispektrale Messung	123
4.11.1	Multispektrale Messung 450 - 950 nm	124
4.11.2	Multispektrale Messung 675 - 950 nm	127
5	Zusammenfassung und Diskussion	135
5.1	Videoeigenschaften	135
5.1.1	Wahl des CRF-Wertes	136
5.1.2	Einfluss von räumlicher Unterabtastung	137
5.1.3	Einfluss der zeitlichen Unterabtastung	141
5.1.4	Empfehlungen für die Kompression der Videodaten	142
5.2	Region of Interest (ROI)	143
5.3	Signalextraktion	145
5.3.1	Klassische Verfahren	145
5.3.2	LSTM Netze	147
5.4	Atmung	148
5.5	Lebenderkennung	149
5.6	Messung im MRT	150
5.7	Multispektrale Messung	150
6	Ausblick	153
6.1	Videoeigenschaften	153
6.2	Region of Interest	154
6.3	Signalverarbeitung	155
6.4	Lebenderkennung	156
6.5	Sauerstoffsättigung	156
6.6	Herzratenvariabilität	157
	Literatur	159

Abbildungsverzeichnis

1.1	Ablauf der kamerabasierten Vitalparametermessung inklusive der üblichen Teilschritte und der in dieser Arbeit präsentierten neuen Methoden zur Vitalparameterschätzung und Experimente.	3
2.1	Schematischer Aufbau der Gesichtshaut mit mittleren Schichttiefen und das Reflexions- und Reemissionsverhalten des einfallenden Lichtes (nach [Goe+17], [Cho+15] und [AP81]). . .	9
2.2	Molarer Extinktionskoeffizient in Abhängigkeit der Wellenlänge für Hämoglobin mit (Hb) und ohne gebundenen Sauerstoff (HbO ₂) in Wasser.	10
2.3	Beispielhafte Veränderungen der Herzratenvariabilität im PPG-Signal durch die Atmung mittels Amplitudenmodulation (blau), Basislinienmodulation (rot) und Frequenzmodulation (gelb). Das zugrundeliegende Signal ist schwarz dargestellt.	11
2.4	Schematischer Ablauf der kamerabasierten Vitalparametermessung	12
2.5	Beispiel aus der MMSE-HR Datenbank mit einem CRF=0. . .	14
2.6	Beispiel aus der MMSE-HR Datenbank mit einem CRF=37. . .	14
2.7	Farbhistogramm der Abbildung 2.5 (CRF=0).	14
2.8	Farbhistogramm der Abbildung 2.6 (CRF=37).	14
2.9	Darstellung (2x2 Pixel) von YUV ₄₄₄ mit voller Farbinformation.	16
2.10	Darstellung (2x2 Pixel) von YUV ₄₂₀ mit Farbunterabtastung.	16
2.11	Exemplarische Beispiele der verschiedenen Skalierungsalgorithmen für das Downsampling um 50%.	17
2.12	Beispiele für die verschiedenen landmarkenbasierten <i>Region of Interest</i>	20
2.13	Schematischer Ablauf der Inversen Fast Fourier (IFFT) PPG-Signalverarbeitung.	25

2.14	Beispiel für die Abhängigkeit des PPG-Signales (schwarz) und der Phasenänderung des durch die Hilbert Transformation abgeleiteten Signals (rot).	26
2.15	HSV-Farbe in Abhängigkeit des Hue-Wertes.	27
2.16	Schematischer vereinfachter Aufbau des DeepPhys Netzes, mit C als Farbbild zum Zeitpunkt t . (nach [CM18])	32
2.17	Schematischer vereinfachter Aufbau der zwei PhysNet Modelle.	33
2.18	Schematischer vereinfachter Aufbau des Extractor-Estimator Netzes. (nach [SFM18])	34
2.19	Schematischer vereinfachter Aufbau des RhythmNet Modells. (nach [Niu+19])	35
3.1	Beispiel der Hautwahrscheinlichkeit (BioVid Datenbank).	39
3.2	Beispiel der Hautwahrscheinlichkeit (PURE Datenbank).	39
3.3	Darstellung der als Haut klassifizierten Farben (mit $p > 0.1$) der verwendeten Look-Up-Table.	39
3.4	Beispiel des Verhaltens des adaptiven Bandpasses und den abgeleiteten Cutoff-Frequenzen, mit einer fehlenden Herzraten-schätzung bei Zeitschritt 5.	42
3.5	Beispiele für (a) die Generierung der Inter-Beat-Intervalle (IBI), (b) den Graphen und (c) der Peakauswahl zur Herzratenbestimmung. Durch Störpeaks entstandene Elemente sind rot markiert.	44
3.6	Beispiel für normierte (-1 bis 1) RGB Eingangssignale. (Das rote und blaue Signal wurden für eine bessere Übersicht um ± 2 verschoben.)	47
3.7	Beispiel für die im Training verwendeten Zielsequenzen und eine Ausgabe eines trainierten Netzes.	47
3.8	Aufbau des LSTM Netzes	47
3.9	Schematischer Ablauf der Atemratenmessung.	49
3.10	Beispielhafte Darstellung von (a) dem PPG-Signal mit erkennbarem Pulssignal, (b) den resultierenden modulierten respiratorischen Signalen für jede der drei Modulationsarten und (c) der Atmungsgrundwahrheit.	51
3.11	Beispiel für die berechneten Pulsamplituden (rote Pfeile) des PPG-Signals (blau).	53

3.12	Beispiel für die berechneten Basislinien (rot) für die BMs im PPG-Signal (blau).	53
3.13	Beispiel für die berechneten Periodendauern (rote Pfeile) für die FMs des PPG-Signals (blau).	53
3.14	Beispiel für die berechneten systolischen Anstiegszeiten (grüne Pfeile) und diastolischen Abfallzeiten (magentafarbene Pfeile) der Rise-Fall-Modulation des PPG-Signals (blau).	53
3.15	Schematische Darstellung des Verfahrensablaufes der Lebensderkennung.	57
3.16	Beispiel des Spektrums eines PPG Signales und den abgeleiteten Peakmerkmalen.	59
4.1	Beispiel der Landmarken und <i>Bounding Box</i>	62
4.2	Beispielbild der BioVid Datenbank.	66
4.3	Beispielbild der BP4D+ und MMSE-HR Datenbank.	67
4.4	Beispielbild der PURE Datenbank.	69
4.5	Beispielbild der AtemDB Datenbank.	69
4.6	Ablauf der Auswertung der Videokompression für verschiedene CRF mit den Codecs <i>x264</i> und <i>x265</i> für zwei ROI und vier PPG-Signalverarbeitungsmethoden. Dabei wird für jede Codec/CRF/PPG/ROI Kombination eine Herzrate bestimmt.	72
4.7	Mittlerer quadratischer Fehler der Pixel-RGB-Werte der ersten 10 Sekunden aller Videos im Vergleich zu den Originalbildern für verschiedene CRF-Werte.	75
4.8	Mittlere IEC-Genauigkeit für verschiedene CRF-Werte (Zahlen) und die Datensatzgröße für die <i>x264</i> und <i>x265</i> Codecs (YUV420) auf dem MMSE -Datensatz.	77
4.9	Mittlere IEC-Genauigkeit für verschiedene CRF-Werte (Zahlen) und die Datensatzgröße für die <i>x264</i> und <i>x265</i> Codecs (YUV420) auf dem PURE -Datensatz.	77
4.10	Mittlere IEC-Genauigkeit für verschiedene CRF-Werte (Zahlen) und die Datensatzgröße für <i>x264</i> und verschiedene Farbformate auf dem PURE -Datensatz.	79
4.11	Mittlere IEC-Genauigkeit für verschiedene CRF-Werte (Zahlen) und die Datensatzgröße für <i>x265</i> und verschiedene Farbformate auf dem PURE -Datensatz.	79

4.12	Mittlere IEC-Genauigkeit für verschiedene CRF-Werte für <i>Skin</i> und <i>FaceMid</i> ROIs mit den <i>x264</i> - und <i>x265</i> -Codecs auf dem MMSE -Datensatz.	80
4.13	Mittlere IEC-Genauigkeit für verschiedene CRF-Werte für <i>Skin</i> und <i>FaceMid</i> ROIs mit den <i>x264</i> - und <i>x265</i> -Codecs auf dem PURE -Datensatz.	81
4.14	IEC-Genauigkeit für verschiedene CRF-Werte unter Verwendung verschiedener Signalextraktionen und des x264 -Codecs auf dem PURE -Datensatz.	82
4.15	IEC-Genauigkeit für verschiedene CRF-Werte unter Verwendung verschiedener Signalextraktionen und des x265 -Codecs auf dem PURE -Datensatz. PURE dataset.	82
4.16	IEC-Genauigkeit für verschiedene CRF-Werte unter Verwendung verschiedener Signalextraktionen und des x264 -Codecs auf dem MMSE -Datensatz. PURE dataset.	83
4.17	IEC-Genauigkeit für verschiedene CRF-Werte unter Verwendung verschiedener Signalextraktionen und des x265 -Codecs auf dem MMSE -Datensatz. PURE dataset.	83
4.18	Mittlere IEC-Genauigkeit in Abhängigkeit der gemittelten ROI Bounding Box-Größe verschiedener Videoauflösungen mit drei Skalierungsalgorithmen auf dem MMSE -Datensatz (<i>x265</i> , $CRF=0$).	85
4.19	RMS-Fehler der PPG-Signale (<i>normG</i>) von verschiedenen Videoauflösungen in Bezug auf das Originalvideo bei Verwendung von drei Skalierungsalgorithmen auf dem MMSE -Datensatz (<i>x265</i> , $CRF=0$).	85
4.20	Mittlere IEC-Genauigkeit in Abhängigkeit der effektiven Bildwiederholrate (gleiche interp. FPS verbunden) auf dem MMSE -Datensatz (<i>x264</i> , $CRF=0$).	87
4.21	Mittlere IEC-Genauigkeit in Abhängigkeit der effektiven Bildwiederholrate (gleiche interp. FPS verbunden) auf dem PURE -Datensatz (<i>x264</i> , $CRF=0$).	87
4.22	Mittlere IEC-Genauigkeit in Abhängigkeit der effektiven Bildwiederholrate (gleiche interp. FPS verbunden) auf dem BioVid -Datensatz (<i>x264</i> , $CRF=17$).	87

4.23	Auf Gesichtsmerkmalen und Hautsegmentierung basierende ROIs: (a) Mitte des Gesichts (FaceMid), (b) Stirn (Forehead), (c) Wahrscheinlichkeitsschwelle $t = 0,3$ und (d) Flächenschwelle $A = 10\%$ (ROI-Pixel in pink).	90
4.24	IEC Mittelwert über alle 10 Probanden für jeden Algorithmus auf dem <i>Steady</i> -Teil des <i>PURE</i> Datensatzes.	94
4.25	IEC Genauigkeit der einzelnen Verfahren auf der <i>PURE</i> Datenbank für Werte des Wahrscheinlichkeitsschwellwertes t	96
4.26	IEC Genauigkeit der einzelnen Verfahren auf der <i>PURE</i> Datenbank für Werte des Flächenschwellwertes A	96
4.27	Boxplot der IEC Genauigkeit in Abhängigkeit der Neuronen n	104
4.28	Boxplot der IEC Genauigkeit in Abhängigkeit der Trainingsgewichtes w	104
4.29	Beispiel für ein normiertes Atemsignal (dimensionslos) der BP4D+, bei dem die Grundwahrheit bestimmt werden kann.	109
4.30	Beispiel für ein verworfenes normiertes Atemsignal (dimensionslos) der BP4D+, bei dem die Grundwahrheit nicht ermittelt werden konnte.	109
4.31	Vergleich der Frequenzspektren eines AM-Signales (a) ohne und (b) mit Artefaktreduktion mit einer Grundwahrheitsfrequenz von 0,28 Hz.	112
4.32	Beispiele der HKBU Datenbank. Probanden ohne und mit verschiedenen Masken.	117
4.33	Konfusionsmatrix des Tree-Klassifikators	118
4.34	Konfusionsmatrix des SVM-Klassifikators	118
4.35	Konfusionsmatrix des KNN-Klassifikators	118
4.36	Messaufbau für die Herzratenschätzung im MRT. (Blau: Triggersignal, Rot: Videodaten, Grün: Vitaldaten)	120
4.37	Ausschnitt eines MRT-Videos (Proband 2) mit Darstellung der ROI (rot) und den Augenbereichen (blau). Helligkeit und Kontrast wurden für diese Darstellung angepasst.	121
4.38	Die IEC Genauigkeit in Abhängigkeit der Größe der ROI in Pixel (gefittete Potenzreihe 2. Ordnung der Daten in rot).	123
4.39	Verwendeter Versuchsaufbau mit 8-kanalige Multispektralkamera und Halogen-Lampen.	125
4.40	Mittlerer absoluter Fehler in Abhängigkeit der Wellenlänge.	126

4.41	Experimentaler Aufbau mit Hyperspektralkamera und zwei Halogenlampen.	128
4.42	Normierte multivariate Normalverteilung zur gewichteten Mittelwertbildung in Pixel.	129
4.43	Beispiel der berechneten spektralen Leistungsdichte, der Grundwahrheit und den gefundenen Peaks mit Höhe und Breite. . .	130
4.44	Gemittelter absoluter Fehler der Messung der Herzraten für alle Probanden.	132
4.45	Periodogramm mit dominanten Störfrequenzen und fehlerhafter Messung der Grundwahrheit (Proband 24, 860nm, Grundwahrheit schwarz).	132
4.46	Gemittelter absoluter Fehler der Messung der Herzraten in Abhängigkeit der Wellenlänge.	132
4.47	Gemittelter Anteil der Leistung der Pulsfrequenz vom PPG-Signal in Abhängigkeit der Wellenlänge.	133
4.48	Gemitteltetes Verhältnis der Höhe des <i>GW-Peaks</i> zum <i>Rausch-Peak</i> in Abhängigkeit der Wellenlänge.	133
4.49	Mittlere Intensität (Helligkeit) in Abhängigkeit der Wellenlänge.	134
4.50	Anteil der Pulsfrequenz in Abhängigkeit der Intensität (Probanden farblich gruppiert).	134
5.1	Mittlere IEC-Genauigkeit für verschiedene CRF-Werte (Zahlen) und die Datensatzgröße für den x264 codecs (YUV420) auf dem MMSE -Datensatz.	137
5.2	Beispiel PPG-Signal (<i>normG</i>) aus dem MMSE-HR Datensatz (videoID: F005/T10) der ersten 300 Bilder mit einer Auflösung von 976x1306 Pixel unter Verwendung verschiedener Skalierungsalgorithmen.	139
5.3	Beispiel PPG-Signal (<i>normG</i>) aus dem MMSE-Datensatz (videoID: F005/T10) der ersten 300 Bilder mit einer Auflösung von 130x174 Pixel unter Verwendung verschiedener Skalierungsalgorithmen.	139
5.4	Beispiel des gemittelten Grün-Kanal Signales (MMSE-HR, videoID: F005/T10, x265, CRF=0).	139
5.5	Beispiel für die Skalierung mit dem <i>nearest neighbor</i> Verfahren.	140
5.6	Beispiel für die Skalierung mit dem <i>area</i> oder <i>bicubic</i> Verfahren.	140

5.7 Boxplot der IEC Genauigkeit der LSTM Modelle aus Tabelle 4.9.147

Tabellenverzeichnis

2.1	Eindringtiefe des Lichtes in das Hautgewebe in Abhängigkeit der Wellenlänge, bis 37% der ursprünglichen Energiedichte (aus [AP81]).	9
3.1	Detektionsmatrix der Täuschungsansätze der einzelnen Systemkomponenten und der möglichen Überwindung der Sicherheitskomponenten.	57
4.1	Überblick der verwendeten Datenbanken mit Angaben über Geschlecht(M/F), Gesamtlänge der Videos (im Format hh:mm), Messmethoden der Herzrate (HR) und Atemrate (AR), Auflösung in Pixel, Bildrate in FPS und verwendete Encodiermethode. . .	65
4.2	Mittelwert μ und Standardabweichung σ der absoluten Fehler der Herzfrequenzschätzungen auf den Datensätzen PURE und MMSE unter Verwendung verschiedener ROIs und coders in Bezug auf den CRF-Wert.	88
4.3	Mittelwert μ und Standardabweichung σ der absoluten Fehler der Herzfrequenzschätzungen auf dem PURE -Datensatz unter Verwendung verschiedener Pixelformate in Bezug auf den CRF-Wert.	88
4.4	Mittelwert μ und Standardabweichung σ der absoluten Fehler der Herzfrequenzschätzungen auf dem MMSE -Datensatz für verschiedene Auflösungen unter Verwendung verschiedener Skalierungsalgorithmen in Bezug auf den Mittelwert der Bounding-Box-Pixel des Gesichts.	88
4.5	IEC-Genauigkeit (in %) von ROIs (Spalten) und Algorithmen (Zeilen) für den PURE Datensatz. Beste ROI für jeden Algorithmus fett markiert.	95

4.6	Ergebnisse der Kombinationen von ROIs (Spalten) und Algorithmen (Zeilen) für den <i>BioVid</i> Datensatz. IEC-Genauigkeit in %. Original-ROI der Algorithmen fett.	99
4.7	Ergebnisse der Kombinationen von ROIs (Spalten) und Algorithmen (Zeilen) für den <i>MMSE-HR</i> Datensatz. IEC-Genauigkeit in %. Original-ROI der Algorithmen fett.	99
4.8	Einfluss des dynamischen und statischen Bandpasses auf zwei spektrale (FFT) und auf Peakanalyse basierende (IBI-Graph) Herzratenschätzung.	102
4.9	Mittelwerte und Standardabweichungen der IEC Genauigkeit und des Messfehlers, über alle trainierten LSTM Modelle mit $n \geq 16$ und $3 \leq w \leq 8$ für verschiedene Eingangssignale und Datenbanken.	105
4.10	IEC Genauigkeit und mittlere Fehler μ und Standardabweichungen σ der zehn LSTM Modelle mit den besten IEC Raten (mit Anzahl der Neuronen n und dem Trainingsgewichtungsfaktors w) und gleicher Test- und Trainingsdatenbank.	105
4.11	Ergebnisse für FuseMod (Haut(gewichtet)-ROI, 30s-Fenster) unter Verwendung verschiedener PPG- und Mittlungsmethoden auf der BP4D+	110
4.12	Ergebnisse für FuseMod (Forehead-ROI, 30s-Fenster) unter Verwendung verschiedener PPG- und Mittlungsmethoden auf der BP4D+	110
4.13	Ergebnisse für FuseMod (Haut(gewichtet)-ROI, 30s-Fenster) mit verschiedenen PPG- und Mittlungsmethoden auf der AtemDB	110
4.14	Ergebnisse für FuseMod (Forehead-ROI, 30s-Fenster) mit verschiedenen PPG- und Mittlungsmethoden auf der AtemDB	110
4.15	Ergebnisse mit und ohne Artefaktreduktion für FuseMod (30s-Fenster) auf der BP4D+	113
4.16	Ergebnisse mit und ohne Artefaktreduktion für FuseMod (30s-Fenster) auf der AtemDB	113
4.17	Ergebnisse der FuseMod -Implementierungen (30 und 60 Sekunden Fenster) und den Vergleichsalgorithmen (30 Sekunden und vorgeschlagene Fensterlänge aus der jeweiligen Originalpublikation) auf der BP4D+	114

4.18 Ergebnisse der **FuseMod**-Implementierungen (30 und 60 Sekunden Fenster) und den **Vergleichsalgorithmen** (30 Sekunden und vorgeschlagene Fensterlänge aus der jeweiligen Originalpublikation) auf der **AtemDB**. 115

4.19 IEC Genauigkeit, Mittelwert und Std. Abweichung der im MRT durchgeführten Herzratenschätzung. 122

4.20 Absolute Fehler in BPM (Fehler >10 BPM hervorgehoben) . 127

5.1 Mittlere IEC Genauigkeit und Standardabweichung (in %) von ausgewählten Algorithmen (min IEC >70%) für verschiedene Datenbanken und Regions of Interest, basierend auf den Werten aus den Tabellen 4.5 - 4.7. (Für die PURE Datenbank wurde $t = 0.2$ für die Spalte *Haut* verwendet.) 144

5.2 Zusammenfassung der IEC Ergebnisse aus der Abbildung 4.24 und den Tabellen 4.5-4.7 und 4.9 für ausgewählte Algorithmen und deren PPG-Signalverarbeitungsmethode, Bandpassfrequenzen, Bestimmung der Herzrate (HR) und verwendeten Korrekturmethode. 145

Abkürzungsverzeichnis

aGRD	adaptive Green-Red-Difference
AM	Amplitudenmodulation
BM	Basislinienmodulation
BPM	Herzschläge pro Minute (beats per minute)
BrPM	Atemzüge pro Minute (breaths per minute)
BSS	Blind-Source-Separation
CHROM	PPG Chrominanzansatz von DeHaan und Jeanne [dJ13]
CNN	Convolutional Neural Network
CRF	Constant Rate Factor
DR	Detection Rate
DRLSE	Distance Regularized Level Set Evolution
GRD	Green-Red-Difference
FDR	False Detection Rate
FFT	Fast Fourier Transformation
FIR	Filter mit endlicher Impulsantwort
FM	Frequenzmodulation
FPS	frames per second
Hb	ungebundenen Hämoglobin
HbO₂	an Sauerstoff gebundenen Hämoglobin

HRV	Herzratenvariabilität
HSV	HSV (Hue, Saturation, Value) Farbraum
IBI	Inter-Beat-Interval
ICA	Independent Component Analysis
IEC	International Electrotechnical Commission
IFFT	Inverse Fast Fourier Transformation
ITU	International Telecommunication Union
LSTM	Long Short-Term Memory
LUT	Look-up-Table
NIR	Nahinfrarotspektrum
nm	Nanometer, Wellenlänge des Lichtes
normG	normalisierte Grünkanal Ansatz
PCA	Hauptkomponentenanalyse (Principal Component Analysis)
PPG	Photoplethysmographie
RGB	Rot, Grün, Blau
RMS	Root Mean Square
RNN	rekurrente neuronale Netze
ROI	Region of Interest
RSA	respiratorische Sinusarrhythmie
SRV	Signal-Rausch-Verhältnis
x264	Implementierung des H.264 Videokompressionsstandards
x265	Implementierung des H.265 Videokompressionsstandards
YUV	Luminanz/Chrominanz Farbmodell

1 Einleitung

Motivation Die Erfassung verschiedener Vitalparameter liefert aussagekräftige Informationen über den gesundheitlichen Zustand des Menschen. Insbesondere Herzrate, Atmung und Herzratenvariabilität sind von großer Bedeutung. Diese sind nicht nur in lebensbedrohlichen Situationen relevant, sondern auch zur Diagnose und Risikobewertung von Krankheiten oder um bei sportlichen Aktivitäten optimale Trainingsbelastungen zu erreichen.

Der Hauptanwendungsbereich für kontaktlose Messungen liegt im Gesundheitssektor. Dazu zählen viele verschiedene Szenarien in der telemedizinischen Diagnostik, Home-Care, Früherkennung, Prävention und Langzeitbeobachtung. Zudem können auch in weiteren Bereichen, wie Sport und Fitness, der Fahrerzustandserfassung im Bereich der Mobilität, oder bei der Erfassung des Nutzerzustandes im Kontext der Mensch-Maschine-Interaktion potenzielle Anwendungen gefunden werden. Die Notwendigkeit für automatisierte kontaktfreie Gesundheitsüberwachungssysteme steigt insbesondere in Ländern mit einer alternden Bevölkerung, da dies mit einer großen Anzahl chronischer Erkrankungen einhergeht.

Existierende Methoden auf Basis dieses kontaktfreien Messprinzips sind entweder nicht hinreichend robust gegenüber Bewegungen, Mimik und Beleuchtungsänderungen oder benötigen zu lange Zeitfenster für eine Messung. Das Ziel dieses Promotionsvorhabens ist die Entwicklung einer bildbasierten, kontaktfreien Messmethode, die den Nutzenden maximale Bewegungsfreiheit und Komfort bietet, robust und schnell funktioniert sowie einfach zu verwenden ist. Dies wird durch die Verbesserung der Störsicherheit gegenüber den oben genannten Ursachen erreicht.

Nachteile kontaktbasierter Messsysteme Aktuell vertriebene Geräte zur Messung von Vitalparametern verwenden ausschließlich kontaktbasierte

Messmethoden. Diese weisen einige Nachteile auf. Ein Messkopf muss am Körper angelegt werden, was für die tragende Person meist unangenehm ist. Es können zudem Hautirritationen bzw. Schmerzen auftreten, wenn zur Fixierung beispielsweise Klebe-Elektroden bzw. Federklemmen eingesetzt werden. Zudem kann es zur Übertragung von Krankheitserregern kommen, was zu zusätzlichen Kosten zur Sicherstellung der Keimfreiheit der Messinstrumente führt. Die Einrichtung und Auswertung des Messgerätes sowie die Anbringung benötigter Kontaktpunkte am menschlichen Körper, können durch kontaktlose Systeme vermieden werden, was Patienten befähigt, die eigenen Vitalparameter leichter selbstständig aufzuzeichnen. Weiterhin ließe sich der Messaufwand (Elektroden, Kabel, etc.) für den Patienten minimieren und damit einhergehend dessen Mobilität verbessern. Dies versetzt die Person in die Lage, während der Messung, Alltagstätigkeiten durchzuführen und hierbei von keinerlei zusätzlichen körperlichen Einschränkungen beeinträchtigt zu werden.

Überblick kamerabasierte Vitalparametermessung Das zugrundeliegende Messverfahren wird bildbasierte Photoplethysmographie (PPG) genannt. Bei dieser Methode wird optisch die periphere Durchblutung gemessen. Dabei wird Licht, welches entweder das Gewebe durchdrungen hat, oder von der Haut reflektiert wurde, gemessen und ausgewertet. Das Hämoglobin im Blut absorbiert mehr Licht, als das umliegende Gewebe, sodass sich der Blutvolumenpuls in einer periodischen Änderung der Lichtreflexion widerspiegelt. Die Abbildung 1.1 zeigt sowohl die üblichen Teilschritte der kamerabasierten Vitalparameterschätzung, als auch die in dieser Arbeit präsentierten neuen experimentellen Methoden für die jeweiligen Teilschritte der Messung, sowie die durchgeführten Experimente.

Die Aufnahme der Videos wird im Stand der Technik in der Regel mit Rot, Grün, Blau (RGB) Kameras durchgeführt, kann aber auch im Nahinfrarotspektrum (NIR) geschehen. Die Videodaten werden üblicherweise für die Auswertung und Archivierung mittels eines Videocodecs komprimiert und abgespeichert. Diese Schritte zählen nicht zur eigentlichen kamerabasierten PPG-Messung hinzu, sind jedoch von großer Bedeutung für die Qualität der Daten und der daraus abgeleiteten Vitalparametermessungen. In dieser Arbeit werden daher sowohl die Messung im NIR Bereich (Kapitel 4.11), als



Schritte	Beschreibung	Neue Methoden und Experimente
Kamera	Aufnahme der Person im RGB oder Nahinfrarotspektrum.	Analyse des Herzratensignales in verschiedenen Bändern des Nahinfrarotspektrums (650-950nm). Messung der Herzrate im MRT mittels eines Spiegelsystems (monochromatisch).
Video	Kompression und Speicherung der Daten.	Untersuchung des Einflusses der Videokompression und Auflösung auf die Qualität der Vitalparametermessung.
Region of Interest	Segmentierung der Haut und Mittelung der RGB-Farbkanäle in den einzelnen Bildern.	Einsatz und Anpassung von wahrscheinlichkeitsbasierter Hautdetektion zur Bestimmung der ROI.
RGB Signale		
Signalverarbeitung	Verarbeitung und Analyse der RGB Signale zur Isolation der Pulsinformationen.	Entwicklung eines adaptiven Bandpassfiltersystems auf Basis einer Look-Up-Table zur kontinuierlichen Signalfilterung in Echtzeit.
PPG Signal		
Herzrate, Atemrate	Messung der Vitalparameter durch Analyse der Maxima, oder des Signalspektrums.	Entwicklung eines Algorithmus zur Peakselektion mittels der Generierung und Nutzung eines gerichteten Graphen. Modellierung und Training eines LSTM basierten neuronalen Netzes zur Peakselektion. Entwicklung eines Algorithmus zur Atemraten-erkennung durch die Fusionierung und Auswertung verschiedener Modulationen des PPG-Signales.

Abbildung 1.1: Ablauf der kamerabasierten Vitalparametermessung inklusive der üblichen Teilschritte und der in dieser Arbeit präsentierten neuen Methoden zur Vitalparameterschätzung und Experimente.

auch der Einfluss der Videokompression auf die PPG-Messung untersucht (Kapitel 2.2.1 und 4.4). Zudem wurde eine Messung der Herzrate mittels eines Spiegelsystems und einer monochromen Kamera durchgeführt (Kapitel 4.10). Dadurch wurde die Messung der Vitalparameter sowohl in einem anderen Spektralbereich, als auch die Genauigkeit der Messung auf größere Distanzen getestet.

Für ein vorliegendes Video wird für jedes Einzelbild eine Region of Interest (ROI) definiert. Dazu wird, in der Regel, zunächst das Gesicht des Probanden gesucht. In dem gefundenen Bereich werden dann bestimmte Pixel als ROI definiert. Im Anschluss werden dann die RGB-Farbmittelwerte der ROI-Pixel berechnet. So kann jedem Bild ein Rot, Grün, Blau (RGB)-Farbwert-Triple zugeordnet werden. Der Stand der Technik für die ROI ist in Kapitel 2.2.2 beschrieben, während der neue Ansatz der wahrscheinlichkeitsbasierten ROI in den Kapiteln 3.1 und 4.5.2 beschrieben und experimentell untersucht wird.

Aus den drei RGB Signalen der Videobilder wird das PPG-Signal extrahiert. Hierzu wurde eine Vielfalt an Verfahren entwickelt, welche im Kapitel 2.2.3 vorgestellt werden. Dabei ist das Pulssignal häufig von Bewegungs- und Mimikartefakten sowie Lichtänderungen oder dem Sensorrauschen überlagert. Zur Filterung der Störsignale werden häufig die spektralen Eigenschaften der unterschiedlichen Farbkanäle und verschiedene Filterverfahren genutzt, um das Pulssignal aus den RGB Daten zu isolieren. Im Rahmen dieser Arbeit wurde ein adaptiver Bandpassfilter entwickelt, welcher in Kapitel 3.2 beschrieben ist.

Um die Herzfrequenz aus dem PPG-Signal zu bestimmen, werden meist entweder ein spektralbasierter, oder ein auf den zeitlichen Pulsabständen, den Inter-Beat-Interval (IBI), basierter Ansatz verwendet. Verschiedene Methoden zur Bestimmung der Herzrate aus dem Stand der Technik sind in Kapitel 2.2.3 beschrieben. Dabei wird im spektralen Ansatz die dominante Frequenz im Band 0,5-4Hz als Herzfrequenz angenommen, während für das IBI-Verfahren die zeitlichen Abstände zwischen den Pulsmaxima, die IBI, berechnet werden. Um die Atemfrequenz zu schätzen, werden die zeitlichen Änderungen der IBI betrachtet, um daraus die Herzratenvariabilität und damit die Atemrate abzuleiten (siehe Kapitel 2.1.3). Zur Bestimmung der

Herzrate wurde im Rahmen dieser Arbeit eine graphen-basierte Peakselektion entwickelt, welche verschiedene mögliche Pulsabfolgen im PPG-Signal bewertet und so den Einfluss möglicher Störsignale minimiert. Diese wird in Kapitel 3.3 beschrieben. Weiterhin wurde eine Methode zur Pulsdetektion auf Grundlage von Long Short-Term Memory (LSTM) Netzen entwickelt (siehe Kapitel 3.4), welche in der Lage sind, die Pulsinformationen aus den RGB Daten zu extrahieren. Ein Verfahren, welches die kurzfristige Modulation der Herzrate analysiert, wurde für die Bestimmung der Atemrate entwickelt (siehe Kapitel 3.5.2). Dabei werden physiologisch verursachte Schwankungen der Herzratenvariabilität genutzt, um Rückschlüsse auf die Atemrate zu schließen.

Zusätzlich wurden Versuche für zwei proof-of-concepts durchgeführt, um die Verwendung der Vitalparameter in speziellen Settings zu untersuchen. Zum einen wurde, auf Grundlage der kamerabasierten Vitalparameterschätzung, eine Methode zur Verhinderung der Überwindung von Gesichtserkennung entwickelt (siehe Kapitel 3.6 und 4.9). Dieses beinhaltet eine Kombination aus RGB und 3D Tiefeninformationen, um Masken, Fotos sowie lebende Gesichter zu unterscheiden. Zum anderen wurde in Kapitel 4.10 die Verwendung der Vitalparameterschätzung in einem MRT Gerät untersucht. Bedingt durch die hohen magnetischen Feldstärken muss dort üblicherweise spezielle Messtechnik eingesetzt werden. Die Messung der Herzrate wurde in dem Versuch erfolgreich kontaktlos, durch den Einsatz eines Spiegelsystems, aus einem abgeschirmten Nebenraum durchgeführt.

2 Stand der Technik

Der in der Einleitung vorgestellte Überblick der kamerabasierten Vitalparametermessung wird in diesem Kapitel vertieft. Die einzelnen Schritte werden erläutert und Verfahren aus dem Stand der Technik vorgestellt. Dabei wird der Ablauf des Verfahrens, von der menschlichen Haut, über die Signalverarbeitung, bis zur Messung, als Leitfaden für den Aufbau des Kapitels genutzt. Zudem wird die Messung im, für den Menschen nicht sichtbaren, nahinfraroten Lichtspektrum beschrieben. Am Ende des Kapitels werden weiterhin verschiedene Messansätze mit Convolutional Neural Network (CNN) vorgestellt, welche die Herzrate ohne Zwischenschritte aus den Videodaten ableiten.

In Abschnitt 2.1 werden die physiologischen Grundlagen und biologischen Prozesse, wie der Aufbau der menschlichen Haut und die Lichtabsorptionsverhalten des Gewebes erläutert, auf deren Grundlage die Messung basiert. Der allgemeine Ablauf der Vitalparameterschätzung und die Teilschritte werden in Kapitel 2.2 erläutert. Dabei werden zunächst der Einfluss verschiedener Videoeigenschaften (Kapitel 2.2.1), wie die Wahl des Videoformates und der Encodierparameter erläutert. Weiterhin wird der Stand der Technik für die einzelnen Schritte der Vitalparameterschätzung, wie die Region of Interest (ROI) (Kapitel 2.2.2), die Signalverarbeitung (Kapitel 2.2.3) und die Atemratenschätzung (Kapitel 2.2.4) beschrieben. Die im Stand der Technik publizierten Methoden zur Messung der Vitalparameter im Nahinfrarotbereich werden in Kapitel 2.3 und die Verwendung von tiefen neuronalen Netzen zur Vitalparameterschätzung wird in Kapitel 2.4 beleuchtet.

2.1 Physiologische Grundlagen

2.1.1 Aufbau der menschlichen Haut

Die menschliche Haut besteht aus der Epidermis (Oberhaut), der Dermis (Lederhaut) und der Subkutis (Unterhautfettgewebe), wie in Abb. 2.1 schematisch dargestellt. Daneben sind in den verschiedenen Hautschichten diverse Adnexorgane wie Haare, Nägel oder Drüsen angesiedelt. Die Epidermis besteht aus verhornten, geschichteten Epidermalzellen, welche etwa alle 4 Wochen komplett erneuert werden. Der äußerste Teil der Haut, das Stratum Corneum, wurde in der Abb. 2.1 zusätzlich dargestellt, da diese bis zu 7% des Lichtes reflektieren kann [AP81]. Dieses Reflexionsverhalten kann dadurch großen Einfluss auf die kontaktfreie Messung der Vitalparameter nehmen. Neben verschiedenen Funktionen koordiniert die Dermis biophysikalische Vorgänge zur Erhaltung der Homöostase (Konstanthaltung des inneren Milieus) des umliegenden Gewebes. Zudem verläuft in der Dermis ein Netz aus Blutgefäßen sowohl zur Versorgung der unteren Epidermis, als auch der in der Haut befindlichen Nerven, Haare und sonstigen Organen. Diese sind durch vertikalen Blutgefäße mit der Subcutis verbunden, welche die oberen Hautschichten mit Blut versorgen.

2.1.2 Lichtabsorption des Gewebes

Die Abbildung 2.1 zeigt neben dem Aufbau auch ein Modell des Reflexionsverhaltens der Haut. Für die kamerabasierte Vitalparameterschätzung werden die Pulsinformationen in dem reemittierten Licht der Haut gemessen. Diese photoplethysmographischen (PPG) Informationen werden daher größtenteils in den Hautschichten mit vielen Blutgefäßen generiert.

Die Eindringtiefe des Lichtes ist dabei stark von der Wellenlänge des einfallenden Lichtes abhängig (siehe Tabelle 2.1). Je länger die Wellenlänge des Lichtes, desto tiefer dringt es in das menschliche Gewebe ein. Laut Anderson und Parrish erreicht [AP81] blaues und ultraviolettes Licht (<400 nm) nur die oberflächlichen Kapillare, während Wellenlängen ab 600 nm

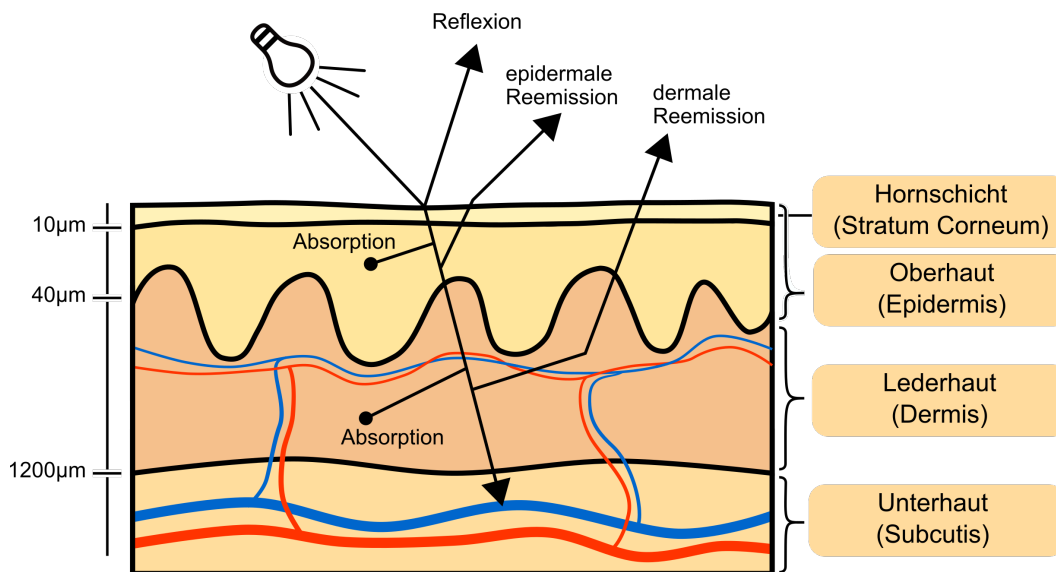


Abbildung 2.1: Schematischer Aufbau der Gesichtshaut mit mittleren Schichttiefen und das Reflexions- und Reemissionsverhalten des einfallenden Lichtes (nach [Goe+17], [Cho+15] und [AP81]).

Tabelle 2.1: Eindringtiefe des Lichtes in das Hautgewebe in Abhängigkeit der Wellenlänge, bis 37% der ursprünglichen Energiedichte (aus [AP81]).

Wellenlänge in nm	250	280	300	350	400	450	500	600	700	800	1000	1200
Tiefe in µm	2	1,5	6	60	90	150	230	550	750	1200	1600	2200

deutlich tiefer in das Gewebe eindringen können. Mit zunehmender Wellenlänge wird das Signal jedoch auch anfälliger für Bewegungsartefakte, welche durch Veränderungen im inneren Gewebe, wie z. B. Muskelbewegungen, verursacht werden [Spi+07]. Der Anteil des reemittierten Lichtes aus den tieferen Hautregionen wird zudem durch Absorption in der Haut stark verringert.

Die PPG-Informationen sind weiterhin vom Absorptionsverhalten des Blutes abhängig. Bei jedem Herzschlag wird, durch den kurzzeitig erhöhten Blutdruck, mehr Blut durch die Kapillaren der Haut gepumpt, was zu einem zeitlich variablen Blutvolumen in der Haut führt. Abbildung 2.2 zeigt die Koeffizienten der Lichtextinktion für ungebundenes Hämoglobin (Hb) und an Sauerstoff gebundenes Hämoglobin (HbO₂). Die Sauerstoffsättigung

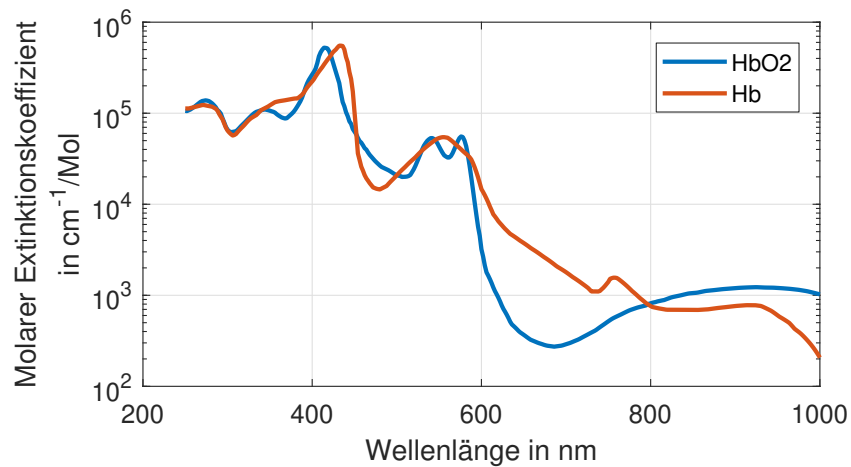


Abbildung 2.2: Molarer Extinktionskoeffizient in Abhängigkeit der Wellenlänge für Hämoglobin mit (Hb) und ohne gebundenen Sauerstoff (HbO₂) in Wasser.

des Blutes liegt bei einem gesunden Menschen über 95%. Für Licht über 600 nm Wellenlänge liegt ein starker Abfall der Extinktion vor, wodurch die Signalstärke der Pulsinformationen für diesen Spektralbereich deutlich abnimmt. So ist der Absorptionskoeffizient der roten Blutkörperchen im Bereich des blau-grünen Lichtes etwa siebenmal höher als bei rotem Licht [Vol+17]. Aus diesen Gründen hat sich grünes Licht (500-600 nm) als Farbkanal für die kamerabasierte PPG-Messungen durchgesetzt [Svi+18]. Da in diesem die stärkste photoplethysmografischen Information enthalten sind [VSN08; Fen+15; Cas+18; Rui+14].

2.1.3 Herzratenvariabilität und Atemrate

Die Herzratenvariabilität (HRV) beschreibt die kurzfristige Variation der Zeitintervalle zwischen aufeinanderfolgenden Herzschlägen und lässt direkte Rückschlüsse auf die körperliche Verfassung einer Person zu. Sie wird bei einer Vielzahl von Diagnostiken zu Rate gezogen und ist wichtig für die Bewertung der Aktivität des autonomen Nervensystems. Die Atmung ist mit der HRV durch die respiratorische Sinusarrhythmie (RSA) gekoppelt. Diese verursacht unter anderem einen Anstieg der Herzfrequenz während des Einatmens und einen Abfall beim Ausatmen [ZRF94]. Durch eine genaue

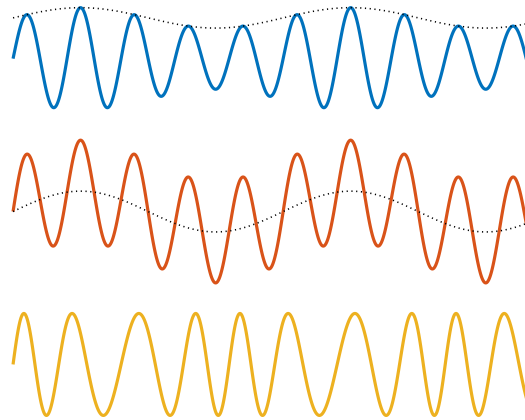


Abbildung 2.3: Beispielhafte Veränderungen der Herzratenvariabilität im PPG-Signal durch die Atmung mittels Amplitudenmodulation (blau), Basislinienmodulation (rot) und Frequenzmodulation (gelb). Das zugrundeliegende Signal ist schwarz dargestellt.

Analyse der HRV kann daher die Atemrate aus der zeitlichen Variation einzelner Herzschläge im PPG-Signal abgeleitet werden [Mad+11].

Dabei können verschiedene Modalitäten der HRV verwendet werden, welche in Abbildung 2.3 dargestellt sind. Neben der Frequenzänderung können auch die Amplitude und die Basislinie des Signales Informationen über die Atemrate enthalten. Die Änderung der PPG Amplitude kann auf ein reduziertes Schlagvolumen während des Einatmens zurückgeführt werden. Dies verursacht eine Änderung des intrathorakalen Drucks, was ein Absinken der Pulsamplitude bewirkt [Mer+11]. Die Basislinie des Signales wird durch Änderungen des Gewebeblutvolumens beeinflusst. Diese sind eine Folge von Änderungen des intrathorakalen Drucks und der Vasokonstriktion der Arterien während der Einatmung, wenn das Blut in die Venen fließt [NFFo6].

2.2 Kamerabasierte Vitalparameterschätzung

Die beschriebenen physiologischen Eigenschaften und deren Einfluss auf das Reflexionsverhalten der Haut werden genutzt, um die Herz- und Atem-

rate kontaktlos mittels Videodaten auszuwerten. Die Vorgehensweise und die Schritte der Signalverarbeitungskette wurden in der Einleitung dargestellt (siehe Abbildung 1.1). Die meisten Ansätze der aktuellen Forschung lassen sich in vier Teilschritte untergliedern (Siehe Abb. 2.4).

Es werden die Grundlagen der Videokompression (Kapitel 2.2.1), der Stand der Technik für die Region of Interest (ROI) (Kapitel 2.2.2), verschiedene in der Literatur vorgestellte Signalverarbeitungsansätze und Verfahren für die Schätzung der Herzrate vorgestellt (Kapitel 2.2.3). Abschließend werden Ansätze zur Messung der Atemrate (Kapitel 2.2.4) vorgestellt.

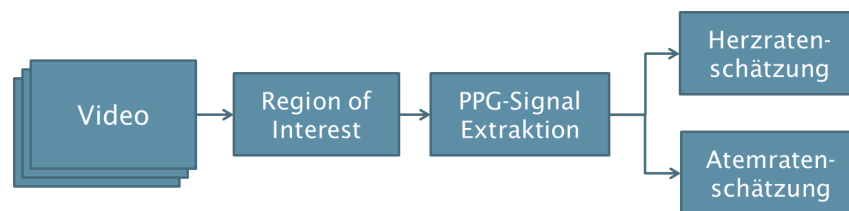


Abbildung 2.4: Schematischer Ablauf der kamerabasierten Vitalparametermessung

2.2.1 Videokompression

Moderne Standard-Videokomprimierungsalgorithmen wie H.264 und H.265 sind psycho-visuell optimiert. Sie komprimieren die Videodaten mit dem Ziel, dass die Informationsreduzierung möglichst unsichtbar für die menschliche Wahrnehmung bleibt und die Bildqualität nicht merklich beeinträchtigt wird. Dazu gehören oft Reduzierung der Farbinformationen (*Color subsampling*), reduzierte Bildqualität bei schnellen Bewegungen und das Filtern von geringen Farbänderungen. Diese Optimierungsschritte helfen dabei, die Datengrößen der Videoinformationen zu reduzieren, um Video-Streaming und Archivierung mit einer hohen Bildqualität zu ermöglichen.

Jedoch kann die in diesen Algorithmen angewandte Informationsreduktion einen starken Einfluss auf das PPG-Signal und die darin enthaltenen Pulsinformationen haben. Dieses Problem wurde, insbesondere in der Anfangszeit der kamerabasierten Herzratenschätzung, in der Literatur oft vernachlässigt. So wurden zum Beispiel häufig Datenbanken mit hohen Kompressionsgraden in der Forschung verwendet. In den meisten Veröffentlichungen

fehlen Details über die verwendeten Codecs, Encodierungsparameter und Videocontainerformate. Dies erschwert den Vergleich und die Reproduzierbarkeit der veröffentlichten Ergebnisse. Teile dieses Kapitels wurden vorab in [RWA19] publiziert.

Codecs

Die Grundlage einer Videokompression ist der Codec (**c**oder, **d**ecoder), welcher die Algorithmen und Regeln für das Kodieren und Dekodieren der Daten beschreibt. Moderne Codecs reduzieren die Datenraten dynamisch und in Abhängigkeit des Videoinhaltes. Damit erreichen sie Kompressionsraten von $> 99\%$ ohne sichtbare Qualitätsverluste. Am weitesten verbreitet ist zum einen der *H.264* Standard, auch genannt *Advanced Video Coding* (AVC), welcher im Video-Streaming (z. B. *YouTube*, *iTunes Store*), HDTV-Übertragungen oder Blu-rays verwendet wird. Zum anderen der neuere, fortschrittlichere *H.265*-Standard oder *High Efficiency Video Coding* (HEVC). Dieser Codec bietet ca. 50% Bitrateneinsparung bei gleicher Wahrnehmungsqualität im Vergleich zu früheren Standards [Sul+12, p. 1667]. Im Folgenden werden explizit die im *FFMPEG*-Projekt [Bel] veröffentlichten Implementierungen *x264* und *x265* der *H.264* und *H.265* Codecs betrachtet.

Beide Codecs versuchen, redundante Informationen in verschiedenen Bereichen des Videos zu finden. Sowohl innerhalb einzelner Frames (Intraframe), als auch in vorherigen oder nachfolgenden Frames (Interframe). Daher kann sich insbesondere die *interframe*-Kompression nachteilig auf die Qualität des PPG-Signals auswirken, wenn dieselben Farbinformationen in verschiedene Einzelbilder kopiert werden.

Constant Rate Factor (CRF)

Der Constant Rate Factor (CRF) ist der *default rate control mode* (Standardmaß zur Qualitätssteuerung) für *x264* und *x265* und wird als Parameter für die vom Menschen wahrgenommene Videoqualität verwendet. Der Wert reicht von 0 (verlustfrei) bis 51 (höchste Komprimierung).

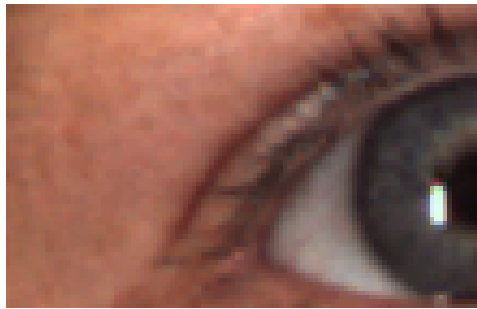


Abbildung 2.5: Beispiel aus der MMSE-HR Datenbank mit einem CRF=0.

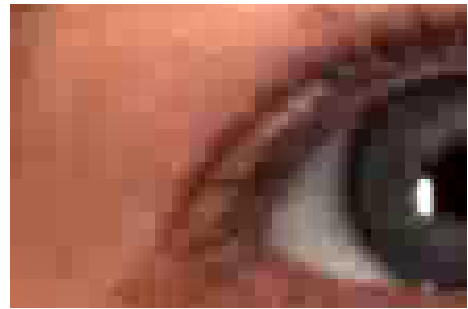


Abbildung 2.6: Beispiel aus der MMSE-HR Datenbank mit einem CRF=37.

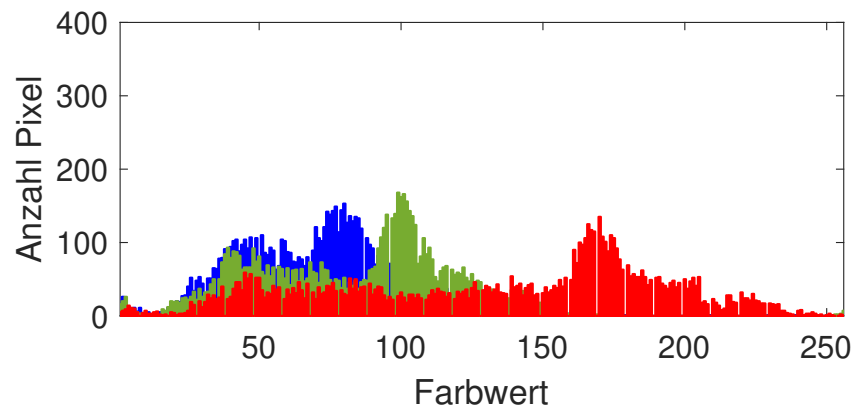


Abbildung 2.7: Farbhistogramm der Abbildung 2.5 (CRF=0).

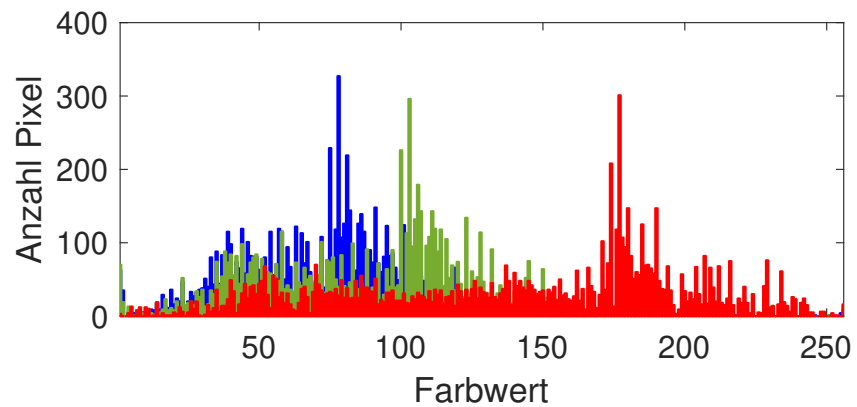


Abbildung 2.8: Farbhistogramm der Abbildung 2.6 (CRF=37).

Die effektive Komprimierungsrate kann im gesamten Video zeitlich variieren, um die Kodierung zu optimieren. Die Bitrate kann zum Beispiel bei bewegungsintensiven Frames reduziert werden, um die Dateigröße zu verringern. Dies ist möglich, da das menschliche Auge bei bewegten Objekten weniger Details wahrnimmt. Zudem wird die Größe der verglichenen Blöcke, Schwellwerte für die Datenreduktion und auch Quantisierungstabellen dynamisch in Abhängigkeit des CRF und des Videoinhaltes angepasst. Abbildungen 2.5 und 2.6 zeigen die unterschiedliche Qualität für verschiedene CRF Werte. Im Bild mit der starken Kompression (CRF=37) ist die Blockbildung zur Informationsreduktion deutlich sichtbar. Damit geht auch ein Verlust der Details, wie scharfe Kanten und schmale Objekte einher, wie zum Beispiel an den Wimpern zu sehen ist. Zudem ist eine Änderung der Farbe zu erkennen, welche bei der hohen Kompression deutlich rötlicher wirkt. In den Abbildungen 2.7 und 2.8 sind die Verteilungen der RGB-Farbwerte der beiden Bilder als Histogramme dargestellt. Die Verteilungen der Histogramme zeigen die Informationsreduktion der stärkeren Kodierung. Das Bild mit CRF=0 zeigt eine Verteilung der Farben auf einem größeren Farbbereich. Mit einem CRF=37 werden die Pixelfarbwerte durch die Kompression auf weniger Farben reduziert, wodurch mehr Pixel dieselben Werte vorweisen. Auffällig sind zudem regelmäßige Nullstellen im Farbspektrum bei CRF=0.

Farbunterabtastung

Die meisten modernen Kameras und Bildschirme verwenden RGB-Sensoren und RGB-Displays. Digitale Videos werden in der Regel in RGB aufgezeichnet und angezeigt, jedoch für die Speicherung in das YUV-Farbmodell konvertiert. Dieses Farbmodell hat seinen Ursprung im analogen Farbfernsehen und teilt die Farbinformationen in eine Luminanz- (Y) und zwei Chrominanzkomponenten (U,V). Dabei enthält der Y-Kanal die Helligkeits- und die U,V-Kanäle die Farbinformationen.

In dieser Arbeit werden zwei *FFMPEG*-Videofarbpixelformate verwendet, YUV444p und YUV420p. Es wurden nur progressive Pixelformate (p) verwendet. Während das YUV444-Format (siehe Abb 2.9) die Werte aller drei

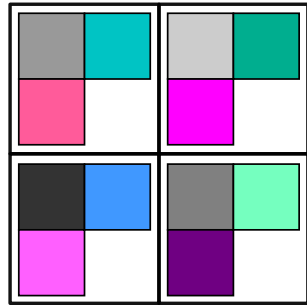


Abbildung 2.9: Darstellung (2x2 Pixel) von YUV₄₄₄ mit voller Farbinformation.

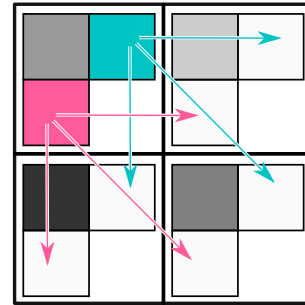


Abbildung 2.10: Darstellung (2x2 Pixel) von YUV₄₂₀ mit Farbunterabtastung.

Kanäle für jedes Pixel speichert, implementiert das YUV₄₂₀-Format eine Farbunterabtastung (chroma subsampling), die zu einer reduzierten Auflösung in den Chrominanzkanälen führt. In jedem 2x2-Pixel-Block werden vier Y-Werte und nur ein U- und ein V-Wert für den gesamten Block gespeichert (siehe Abb 2.10). Dadurch sinkt die Bandbreite eines Pixels von 24 Bit auf 12 Bit und halbiert die gespeicherten Farbinformationen.

Aufgrund der höheren Sehschärfe des menschlichen Auges für Helligkeit als für Farben, ist eine Reduzierung der Farbinformationen und der Dateigröße möglich, ohne eine Verschlechterung des Bildes für die visuelle, menschliche Wahrnehmung. Aus diesem Grund ist YUV₄₂₀ das Standard-Pixelformat für die meisten modernen Video-Streaming- und -Speicherverfahren.

Die etwa 16,7 Millionen möglichen RGB-Farben (8 Bit) werden, nach den Vorgaben (rec.601 [Int11b] und rec.709 [Int15]) der International Telecommunication Union (ITU), auf nur etwa mögliche 11 Millionen YUV-Farben (8 Bit) abgebildet. Dadurch können Farbinformationen verloren gehen, da die Farbtransformation von RGB nach YUV nicht für alle Farben umkehrbar ist. Neben diesen sogenannten *Out-of-Gamut* Farben, welche nicht korrekt in beiden Farbräumen abgebildet werden, können auch Rundungsfehler bei der Quantisierung, während der Kodierung (RGB→YUV) und der Dekodierung (YUV→RGB) der Farbtransformation auftreten.

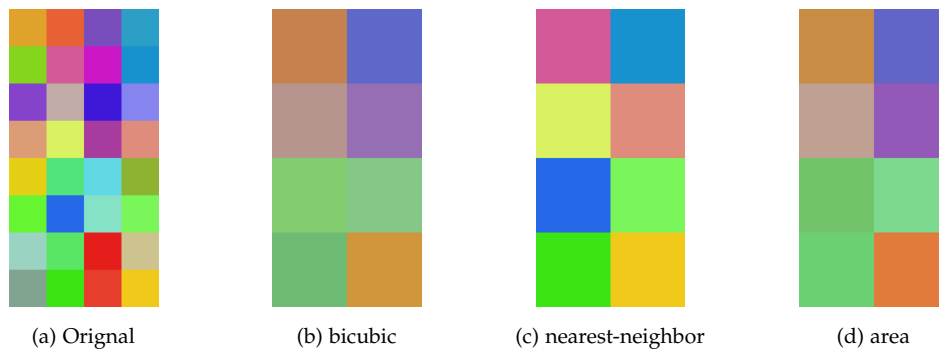


Abbildung 2.11: Exemplarische Beispiele der verschiedenen Skalierungsalgorithmen für das Downsampling um 50%.

Skalierungsalgorithmen

Nach aktuellem Stand existieren diverse Skalierungsverfahren für Bilder und Videos, welche unterschiedliche Vor- und Nachteile aufweisen. Zur Untersuchung der Auswirkungen von Videogrößen auf die Vitalparameterschätzung wurden mehrere Skalierungsalgorithmen geprüft. Aus den in *FFMPEG* implementierten Verfahren wurden drei ausgewählt. Abbildung 2.11 zeigt die unterschiedlichen Auswirkungen der verschiedenen Skalierungsalgorithmen für die Reduzierung der Auflösung um 50%.

bicubic Bei der bikubischen Skalierung werden kubische Polynome verwendet, um die Pixelinformationen zu interpolieren. Der errechnete Pixelwert wird aus der gewichteten Summe der Farbinformationen der Pixel in der nächstgelegenen 4x4-Nachbarschaft bestimmt. Dies erzielt harmonische Farbübergänge und kann aufgrund des nichtlinearen Modells auch Kanteninformationen erhalten.

nearest-neighbor Bei diesem Ansatz werden die Farbinformationen des neuen Pixels aus dem einen Pixel im Originalbild übernommen, welcher der neuen relativen Position im Bild am nächsten ist. Bei dieser Methode bleibt ein Teil der ursprünglichen Farbinformationen des Originalbildes erhalten. Bei starker Skalierung kann dadurch jedoch Blockbildung erfolgen,

da Informationen über nichtlineare Farbgradienten verloren gehen. Zudem werden Kanteninformationen, welche unterhalb der Abtastschwelle liegen, verloren.

area Bei der *area* Skalierung werden, die Farbinformationen durch die Mittlung der Farbwerte der Fläche bestimmt, welche der Pixel im kleineren Bild repräsentiert. Dies erzielt ähnliche Resultate wie das *bicubic* Verfahren, wobei durch die lineare Transformation mehr Informationen über hochfrequente Bildinhalte verloren gehen.

2.2.2 Region of Interest

Für die Messung der Pulssignale, ist das Identifizieren und Finden günstiger Hautregionen im Video von großer Bedeutung. Die verwendete Region, aus der die PPG Informationen abgeleitet werden, wird die Region of Interest (ROI) genannt. PPG-Messungen lassen sich an vielen Körperstellen durchführen [Nil+07]. Die Stirn und andere Gesichtsbereiche eignen sich besonders gut für diesen Zweck, da diese häufig unverdeckt sind. Nilsson [Nil13] konnten zudem zeigen, dass die extrahierten Signalenergien an der Stirn sechsmal höher waren als am Finger, weil diese eine hohe Dichte an Blutgefäßen aufweist und der Schädel von einer vergleichsweise dünnen Haut bedeckt ist [Cas+18].

Im Allgemeinen kann erwartet werden, dass die Verwendung von einer größeren Zahl Hautpixeln das Signal-Rausch-Verhältnis verbessert und zu einem saubereren PPG-Signal sowie einer besseren Herzfrequenzschätzung führt [VSN08]. Die Auswahl eines großen Detektionsbereiches kann mehrere Teile des Gesichts, wie den Mund oder die Augen einschließen. Dies kann zu unerwünschtem Rauschen, aufgrund von Bewegungen des Gesichtes (Reden, Blinzeln, etc.) führen. Verschiedene Störfaktoren (wie Haare, Bärte, Blinzeln, Brillen oder Kleidung), welche die Haut in einem vordefinierten ROI verdecken, können ebenfalls die Genauigkeit deutlich verringern. Eine saubere Bildsegmentation ist daher wichtig, für einen ausreichend hohen Signalrauschabstand und die Genauigkeit der Messung. Eine häufig verwendete Variante, diesen Effekten entgegenzuwirken, besteht darin, große Teile

des Gesichts (z. B. Augen oder Mundregion) aus der ROI auszuschließen und die Farbinformationen in diesen Regionen nicht zu verwenden. Die Wahl der verwendeten ROI ist demnach für die Qualität des berechneten PPG Signals von großer Bedeutung. Ein flexibler Ansatz zur Bestimmung der ROI ist einer festen geometrischen ROI vorzuziehen.

Die ROI basiert in den aktuellen Fällen auf einem Gesichtsverfolgungsalgorithmus [PMP10; PMP11; Mon+17; ICM16; Tar+17] und/oder Gesichtslan-
dmarken, wie beispielsweise die Stirn [Lew+11; Blö+17] oder die Wangen [Nis+16; GMR16b], zu definieren. Dabei wird häufig die *Bounding Box* der Gesichtslan-
dmarken verwendet, welche das Rechteck beschreibt, das alle Landmarken einschließt. Andere Ansätze wenden einen Hauterkennungs-
algorithmus innerhalb der *Bounding Box* des Gesichts an, um die mögliche Einbeziehung jedes Pixels in die ROI zu schätzen [RWA16; dJ13].

Im Folgenden werden verschiedene ROI vorgestellt, welche im Stand der Technik häufig Verwendung finden. Teile dieses Abschnittes wurden bereits in [Rap+18b], [Rap+16a], und [FRA20] publiziert.

FaceMid Die **FaceMid** ROI (siehe Abb. 2.12a) ist eine weit verbreitete ROI [PMP11; Mon+17; Wei+12; HNM15] im Bereich der Herzfrequenz-
Schätzung und wurde von Poh, McDuff und Picard [PMP10] erstmals vorgestellt. Die Region schließt die volle Höhe der *Bounding Box* ein, welche die
Gesichtsmerkmale umschließt, beschneidet aber die Seiten und nutzt nur die mittleren 60% der Region. Dies soll das Signal-Rausch-Verhältnis (SRV)
des extrahierten PPG-Signals verbessern, da viele Nichthautpixel an den Rändern entfernt werden.

Forehead Die Stirn oder **Forehead** ROI (siehe Abb. 2.12b) wird ebenfalls häufiger in der Literatur verwendet [Blö+17] [ICM16] [Tar+17] [Wer+14].
Die Verwendung der Stirn als ROI hat mehrere Vorteile. Zum einen ist in der Regel weniger Bewegung durch Mimik innerhalb der Stirnregion, welche zu
Störungen führen könnte. Weiterhin ist diese Region der Haut relativ dünn und gut durchblutet. Zudem ist sie häufig die größte zusammenhängende
Hautregion im Gesicht. Einige Herausforderungen dieser ROI sind die Abschätzung der tatsächlichen Größe, da die Stirnpartie von Mensch zu

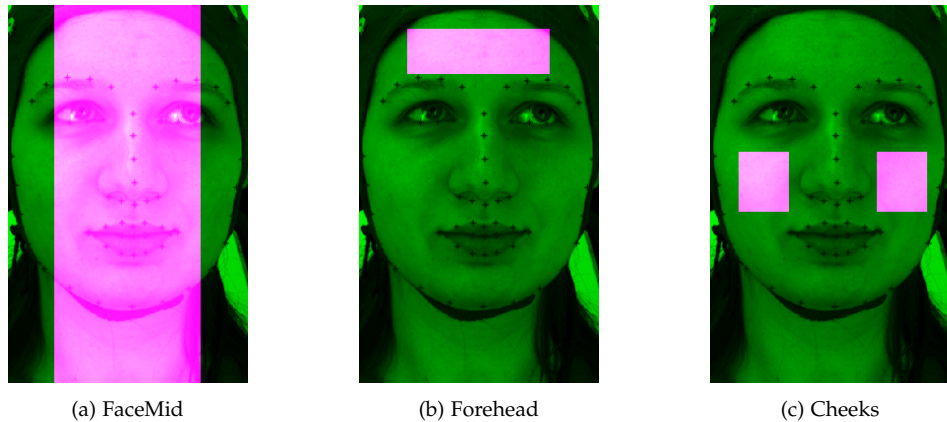


Abbildung 2.12: Beispiele für die verschiedenen landmarkenbasierten *Region of Interest*.

Mensch stark variieren kann und das Verfolgen der Region über die Zeit. Da die Landmarkendetektoren aus dem Stand der Technik keine Punkte für die Stirn enthalten, müssen Landmarken in der Nähe der Stirn (Augenbrauen, obere Nase) verwendet werden, um die ROI zu definieren. Dabei wird die ROI aus den vorhandenen Punkten nach oben extrapoliert, was leicht zu Fehlern führen kann, da die ROI von nur wenigen Punkten abhängig ist.

Cheeks Eine weitere häufiger verwendete landmarkenbasierte ROI sind die Wangen oder **Cheeks** [Nis+16] [Wer+14] [GMR16a] (siehe Abb. 2.12c). Dabei werden in der Regel zwei ROI auf den beiden Wangen im Gesicht definiert. Dies geschieht entweder über die Landmarken der Augen, Nase, Mund und dem Kinn oder dem Tracking von Texturmerkmalen im Gesicht. Das in [Fen+15] vorgeschlagene Verfahren, verwendet beispielsweise zwei Regionen auf den Wangen als aktive Bereiche. Die ROIs werden durch die Verfolgung von Speeded-Up-Robust-Feature (SURF)-Punkten [BTV06] in der Mitte des Gesichts berechnet und eine affine Transformation auf die ROIs angewendet, um Kopfbewegungen zu kompensieren und die ROIs zu stabilisieren.

Van Gastel Van Gastel, Stuijk und De Haan [VSD16] haben eine weitere ROI vorgestellt, bei der dynamisch Regionen des Gesichtes gewichtet in die

Berechnung des PPG-Signales einfließen. Dabei werden Punkte im Gesicht, welche durch eine Minimierung der Eigenwerte bestimmt werden, mittels des Kanade-Lucas-Tomasi [TD91] Tracking-Verfahrens verfolgt. Die Transformation der Punkte zwischen zwei Frames wird berechnet und auf die *Bounding Box*, welche manuell im ersten Frame definiert wird, angewendet. Im Anschluss wird die *Bounding Box* in 30 gleich große Unterregionen unterteilt.

Die Signale aus den verschiedenen Unterregionen werden dann verwendet, um mittels einer optimalen Linearkombination das PPG-Signal zu rekonstruieren. Dazu werden, nach einer Bandpassfilterung der Signale, zunächst die Gewichtung der Einzelsignale mittels einer Blind-Source-Separation (BSS) Verfahrens ermittelt. Die Gewichte mit dem höchsten Signal-Rausch-Verhältnis werden dann genutzt, um die Störsignale zu unterdrücken. Das Signal-Rausch-Verhältnis wird dabei durch eine Hauptkomponentenanalyse (Principal Component Analysis) (PCA) bestimmt. Der Eigenvektor, aus den ersten fünf der PCA, welche die höchste Korrelation mit dem getrimmten ($\alpha = 0.7$) Mittelwert aller Einzelsignale aufweist, wird dabei als Pulssignal definiert. Analog dazu wird auch das Atemsignal und die Atemfrequenz bestimmt, dazu wird das Passband für die Filterung in einen niedrigen Frequenzbereich gelegt.

Hautdetektion Im Gegensatz zu der Gesichts- und Landmarkendetektion lässt sich die Region of Interest auch durch eine farbbasierte Hautdetektion realisieren. Der Stand der Technik kann in zwei Kategorien unterteilt werden: pixelbasierte und regionenbasierte Ansätze. Während die pixelbasierten Ansätze, aufgrund der Unabhängigkeit der einzelnen Pixeldaten voneinander, schneller sind, benötigen regionenbasierte Ansätze deutlich mehr Rechenzeit [SA14b]. Der sehr schnelle ($\sim 20\text{ms}$ pro Bild) und leistungsstarke Look-up-Table (LUT)-Ansatz von Jones und Rehg [JR02] ist der aktuelle Stand der Technik bei den pixelbasierten Ansätzen. In den letzten Jahren wurden immer mehr regionenbasierte Ansätze mit besserer Hautklassifikation vorgeschlagen. Da Farbe bereits ein sehr gutes Merkmal für die Hauterkennung ist, basieren alle regionenbasierten State-of-the-Art Ansätze auf einem LUT Ansatz, gefolgt von einer Textur-[Kaw13; Kaw+14] oder Superpixel-[Hua+15; SP15; SA14b] Analyse. Diese Ansätze haben aller-

dings auch deutlich höhere Rechenkosten (~ 1000 ms pro Bild) und sind damit für die Echtzeitverarbeitung von Videodaten nicht geeignet.

Eine auf dem *BayesBGR* Ansatz [JR99] basierende Hauterkennungs-
methode wurde für die Nutzung in der Vitalparametererkennung angepasst
und ist in Kapitel 3.1 beschrieben. Dabei wird das Bayes'sche Theorem als
Grundlage verwendet, um jeder RGB-Farbe, und damit jedem Pixel eine
Hautfarbenwahrscheinlichkeit zuzuordnen.

2.2.3 Signalverarbeitung

Aus den identifizierten Regionen der Haut werden zunächst nur die rohen
Farbinformationen extrahiert. Die Pulsinformationen sind in den RGB Wer-
ten der ROI enthalten, werden jedoch aufgrund der niedrigen Signalstärke
von verschiedenen Störquellen, wie Bewegung, überlagert. Um das Photople-
thysmographie (PPG) Signal zu isolieren, werden üblicherweise verschiede-
ne Filterverfahren verwendet und das Vitalsignal aus dem Signalspektrum
oder den zeitlichen Abständen der Pulse bestimmt. Häufig wird zudem
eine Form der Fehlerkorrektur verwendet.

Normiertes Grün (normG) Der normalisierte Grünkanal Ansatz (normG)
wurde in der Literatur von Stricker et al. [SMG14] und Rapczynski et al.
[RWA16] als Signalextraktionsmethode verwendet. Dabei wird der Grünkanal
durch die Summe aller Kanäle normiert, um räumliche, unterschiedliche
oder sich über die Zeit ändernde Lichtintensitäten im Video zu kompensie-
ren. Das PPG-Signal wird für jedes Bild aus dem Mittelwert aller normierten
Grünwerte g_n der roten R , grünen G und blauen B Kanäle der i ROI-Pixel
berechnet (Gl. 2.1).

$$g_n = \frac{1}{i} \sum_i \frac{G_i}{R_i + G_i + B_i} \quad (2.1)$$

Adaptive Green-Red-Difference (aGRD) Die aGRD wurde von Feng et al. [Fen+15] vorgestellt und basiert auf der Green-Red-Difference (GRD) Methode von Hülsbusch et al. [Hueo8]. Der Ansatz entfernt diffuses und gestreutes Licht in den grünen und roten Signalen, um das PPG-Signal zu berechnen. Wie in Kapitel 2.1.2 beschrieben, sind deutlich mehr Pulsinformationen im grünen Spektrum des Lichtes, als im Roten enthalten. Diese unterschiedlichen Signal-Rausch-Abstände können ausgenutzt werden, um die Störsignale, welche im roten und grünen Kanal enthalten sind, von den Pulsinformationen zu trennen. Diese sind deutlich stärker im Grün-Kanal enthalten.

Dabei wird für jedes Bild t das aGRD Signal aus den bandpassgefilterten Signalen, der Grün- und Rotkanäle I_{Rf} und I_{Gf} , berechnet:

$$aGRD(t) = \frac{I_{Gf}(t)}{\tilde{\alpha}_G(t)\tilde{\beta}_G(t)} - \frac{I_{Rf}(t)}{\tilde{\alpha}_R(t)\tilde{\beta}_R(t)} \quad (2.2)$$

Mit dem normalisierten Beleuchtungsspektrum $\tilde{\alpha}$ und dem normalisierten diffusen Reflexionsspektrum $\tilde{\beta}$ der Grün- und Rotkanäle definiert als:

$$\tilde{\alpha}_G(t)\tilde{\beta}_G(t) = \frac{I_G(t)}{\sqrt{I_R^2(t) + I_G^2(t) + I_B^2(t)}} \quad (2.3)$$

$$\tilde{\alpha}_R(t)\tilde{\beta}_R(t) = \frac{I_R(t)}{\sqrt{I_R^2(t) + I_G^2(t) + I_B^2(t)}} \quad (2.4)$$

Chrominance (CHROM) DeHaan und Jeanne [dJ13] entwickelten einen auf Chrominanz basierenden Ansatz (CHROM), um den Effekt der durch Bewegung erzeugten Spiegelreflexionen zu eliminieren. Dieser ist eine Weiterentwicklung des GRD-Verfahrens, bei der durch die Definition zweier orthogonaler Chrominanzsignale, das bewegungsinduzierte Rauschen, von dem durch die Blutvolumenänderung induzierten Pulssignal getrennt wird.

Das von der Kamera erfasste reflektierte Licht besteht gemäß dem dichromatischen Reflexionsmodell aus zwei verschiedenen Komponenten X und Y .

Diese werden aus den normierten Farbkanälen R_n, G_n und B_n abgeleitet.

$$X = 3R_n - 2G_n \quad (2.5)$$

$$Y = 1.5R_n + G_n - 1.5B_n \quad (2.6)$$

Das Signal X modelliert das diffuse Licht, das von der Körperoberfläche zurückgeworfen wird und Farbänderungen der Haut enthält. Das Signal Y modelliert die reflektierende Komponente, welche die Farbe der Lichtquelle widerspiegelt und dadurch keine photoplethysmografischen Informationen enthält. Ein Faktor α wird aus dem Verhältnis der Standardabweichungen der bandpassgefilterten Signale X_f und Y_f berechnet, um starke Störungen im Ausgangssignal zu minimieren. Dies ermöglicht eine dynamische Standardisierung der Faktoren bei der Linearkombination der bandpassgefilterten RGB Signale R_f, G_f und B_f .

$$PPG = 3\left(1 - \frac{\alpha}{2}\right)R_f - 2\left(1 + \frac{\alpha}{2}\right)G_f + \frac{3\alpha}{2}B_f \quad \text{mit} \quad \alpha = \frac{\sigma(X_f)}{\sigma(Y_f)} \quad (2.7)$$

Inverse Fast Fourier Transformation (IFFT) Wang et al. [Wan+17] stellen einen Ansatz zur Signalextraktion vor, welcher auf der Inverse Fast Fourier Transformation (IFFT) basiert. Abbildung 2.13 zeigt den schematischen Ablauf des Verfahrens. Dabei werden die RGB Signale zunächst mittels einer Fast Fourier Transformation (FFT) in den Frequenzbereich transformiert. Die für den menschlichen Puls relevanten Frequenzen (0,5-4 Hz) der drei RGB Signale werden zurück in den Zeitbereich transformiert. Für jede Frequenz wird dabei aus den RGB Informationen ein PPG-Teilsignal berechnet. Dazu kann, zum Beispiel, der oben beschriebene *normG* oder *CHROM* Ansatz verwendet werden.

Die einzelnen PPG-Teilsignale werden im Anschluss durch eine gewichtete Summierung in das finale PPG-Signal zusammengeführt. Die Teilsignale mit größeren Intensitätsschwankungen in den Farbkanälen, relativ zu deren abgeleiteten PPG-Teilsignal, werden niedriger gewichtet. Laut den Autoren können mittels der Analyse der Intensitätsschwankungen durch Bewegungen verursachte Störsignale erkannt und gefiltert werden.

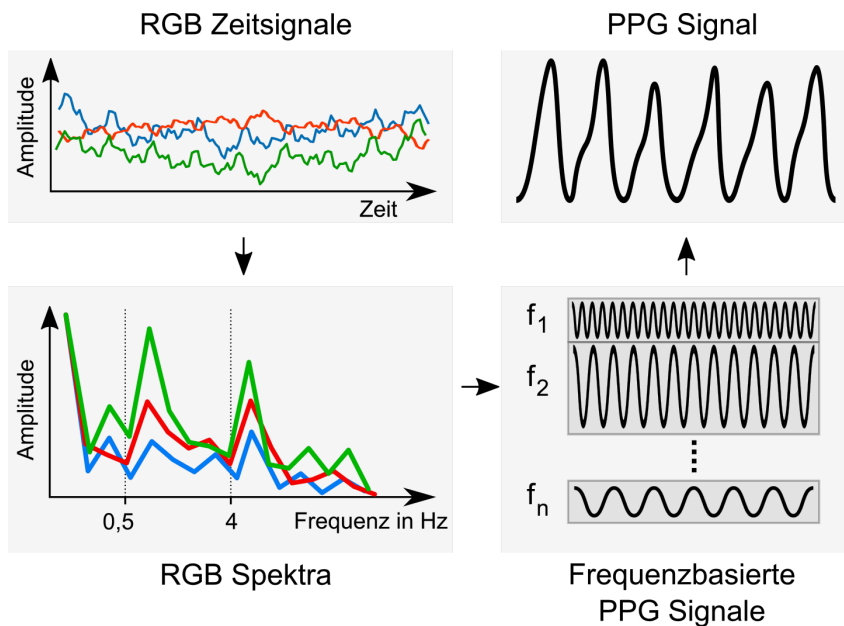


Abbildung 2.13: Schematischer Ablauf der Inversen Fast Fourier (IFFT) PPG-Signalverarbeitung.

Hilbert-Transformation Blöcher et al. (2017) [Blö+17] verwenden eine Independent Component Analysis (ICA) mit dem *Jade*-Algorithmus [RJ13] zur Berechnung des PPG-Signals aus den RGB Signalen. Das von dem Verfahren generierte Ausgangssignal wird bandpassgefiltert (0,75-4 Hz). Dieser Ansatz verwendet anschließend die Hilbert-Transformation, um den Phasenwinkel des Pulssignals zu berechnen. Dabei werden das gefilterte PPG Zeitsignal $x(t)$ und die Hilbert-Transformation $y(t) = \mathcal{H}(x(t))$ zu dem komplexen Signal $z(t)$ kombiniert.

$$z(t) = x(t) + j \cdot y(t) \quad (2.8)$$

Aus z_t können die Amplitude $A_z(t)$ und Phase $\phi_z(t)$ bestimmt werden. Die Phasenverschiebungen des komplexen Signals $z(t)$ korrelieren dabei mit dem PPG-Signal, wobei die Phasensprünge $\phi_z(t)$ den Herzschlägen in $x(t)$ entsprechen (siehe Abb. 2.14). Dies ermöglicht eine präzise Peak-Lokalisierung, anhand der Sägezahnform des Phasensignals und senkt den Einfluss der Peakhöhenschwankungen und lokalen Maxima auf die Pulsbestimmung.

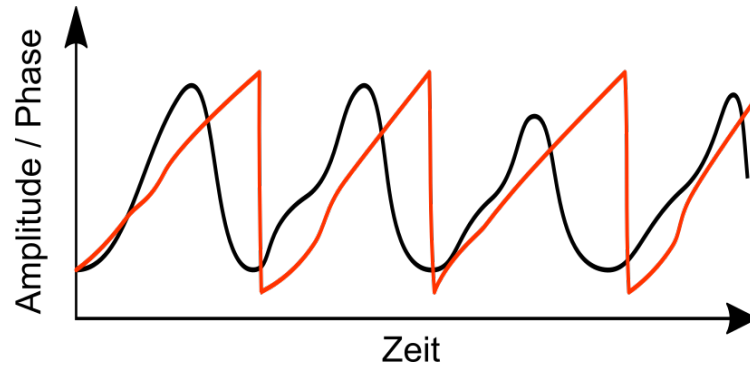


Abbildung 2.14: Beispiel für die Abhängigkeit des PPG-Signales (schwarz) und der Phasenänderung des durch die Hilbert Transformation abgeleiteten Signals (rot).

Beleuchtungsrektifizierung Li, Chen, Zhao und Pietikainen [Li+14] stellen eine Signalverarbeitungsmethode vor, welche Gesichtserkennung und adaptive normalisierte Least-Mean-Square-Filtermethoden einsetzt, um Störsignale durch Bewegungen zu minimieren. Dazu wird der durchschnittliche Grünwert g_{face} der ROI in jedem Bild berechnet. Zusätzlich wird der Hintergrund im Bild mittels der Distance Regularized Level Set Evolution (DRLSE) Methode [Li+10a] bestimmt und ebenfalls der Wert des Grünkanals g_{bg} berechnet. Ein Normalized-Least-Mean-Squares-Filter wird verwendet, um den optimierten Koeffizienten h des Modells zu bestimmen. Die Beleuchtungsschwankungen der ROI werden modelliert und das beleuchtungskorrigierte Signal g_{IR} berechnet.

$$g_{IR} = g_{face} - h \cdot g_{bg} \quad (2.9)$$

Der optimale Wert von $h(j)$ wird iterativ für jeden Zeitpunkt j bestimmt. Dabei wird eine Schrittweite μ und der Initialisierungswert $h(0)$ verwendet.

$$h(j+1) = h(j) + \mu \cdot g_{IR}(j) \cdot g_{bg}(j) \quad (2.10)$$

Zur Stabilisierung des Signales, wird das Hintergrundsignal normalisiert, indem g_{bg}^H , das hermitesch transponierte Signal von g_{bg} , mit g_{bg} multipliziert wird, um die Signalenergie von g_{bg} zu bestimmen.

$$h(j+1) = h(j) + \frac{\mu \cdot g_{IR}(j) \cdot g_{bg}(j)}{g_{bg}^H(j) \cdot g_{bg}(j)} \quad (2.11)$$

Segmente des PPG-Signales, welche durch plötzliche Bewegungen starke Varianzen aufweisen, werden verworfen. Die restlichen Fenster des PPG-Signales werden mit einem Bandpass (0,7-4 Hz) gefiltert. Zuletzt wird *Welch's* Methode zur Schätzung der Leistungsspektraldichte verwendet, um die Herzfrequenz zu schätzen.

HSV Farbraum Sanyal *et al.* [SN18] berichteten, dass eine Farbraumtransformation von RGB in den HSV (Hue, Saturation, Value) Farbraum bessere Ergebnisse, sowohl für die Herzraten- als auch für die Atemraten-Schätzung liefert. Die besseren Ergebnisse des Grün-Kanals, als des Rot-Kanals, in der Literatur, sind laut Sanyal und Nundy *ein Artefakt der Parametrisierung des RGB-Farbraumes*. Sie schlagen die Nutzung der Hue-Komponente im HSV-Farbraum vor (siehe Abbildung 2.15).

Dazu wird in der ROI jeder Pixel vom RGB- in den HSV-Farbraum konvertiert. Dabei entspricht jeder Hue-Wert einer anderen Farbe. Durch die Wahl eines Bereichs kann so effektiv die gewünschte Absorptionsfrequenz gewählt werden. Das PPG Signal wird aus den HSV-Pixeln der ROI mit den Werten von $0 < H < 0.1$, was einer Wellenlänge von 700 - 800 nm entspricht, gemittelt. Das Signal wird anschließend in den Frequenzbereich transformiert und mit einem Bandpass gefiltert. Für die Herzratenschätzung wird der Bereich 0,8 – 2,2 Hz und für die Atemrate 0,18 – 0,5 Hz das Passband gewählt. Die Maxima der gefilterten Spektren werden als Herz- beziehungsweise Atemrate angenommen.

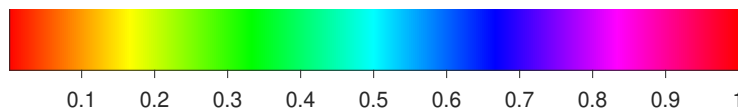


Abbildung 2.15: HSV-Farbe in Abhängigkeit des Hue-Wertes.

Statistische Verfahren In der Literatur werden eine Vielzahl an statistischen Signalverarbeitungsmethoden verwendet, um die PPG Informationen aus den RGB Signalen zu berechnen. Dazu wird in der Regel auf Methoden zur BSS [CJ10] zurückgegriffen. Diese versuchen eine Trennung von Quellsignalen aus einem Satz von gemischten Signalen zu erreichen, wenn

keine oder wenige Informationen über die Quellsignale oder den Mischprozess zur Verfügung stehen. Die häufigst verwendeten Methoden sind, die Hauptkomponentenanalyse (Principal Component Analysis) (PCA) [FRSo1], die Independent Component Analysis (ICA) [JH91] und deren Variationen, wie zum Beispiel der Single-Channel ICA (SCICA) [D]07].

Die **PCA** bestimmt eine Anzahl von Hauptkomponenten, welche zueinander orthogonale Richtungsvektoren darstellen. Die Komponenten minimieren den durchschnittlichen quadratischen Abstand der Daten zur Hauptkomponente und spannen ein neues linear unabhängiges Koordinatensystem auf. Dabei hat die erste Hauptkomponente die größte Varianz, welche für jede folgende Komponente abnimmt.

Die **ICA** findet die unabhängigen Quellsignale durch Maximierung der statistischen Unabhängigkeit der geschätzten Komponentensignale. Dies geschieht unter der Annahme, dass es sich bei den Teilkomponenten um nicht-gaußsche Signale handelt und diese statistisch unabhängig voneinander sind. Die Definition der Unabhängigkeit variiert für die unterschiedlichen ICA-Algorithmen. Generell wird versucht, eine Minimierung von überschneidenden Informationen und die Maximierung der Nicht-Gaußianität der Quellsignale zu erreichen.

2.2.4 Atemratenschätzung

Die meisten Forschungsarbeiten, die sich mit der Messung von kamera-basierten physiologischen Vitalparametern befassen, beschränken sich auf die Messung der Herzfrequenz. Wenn Atemsignale abgeleitet wurden, sind diese, bis auf wenige Ausnahmen, kein Fokus der Forschung gewesen und wurden daher keiner genaueren Analyse unterzogen. Alternativ dazu, gibt es Forschungsergebnisse aus dem Bereich der Atemratenerkennung, aus konventionellen PPG-Signalen, welche nicht aus Videobildern generiert wurden, sondern aus Kontakt-PPG-Aufzeichnungsgeräten.

In der Forschung gibt es keinen Konsens, welcher Bereich für die Bestimmung plausibler Atemfrequenzen optimal ist, da diese je nach betrachteter Personengruppe stark variieren können [Vil+14]. Die Atemrate von erwachsenen, gesunden Personen, in Ruhe und ohne körperliche Anstrengung,

liegt in der Regel zwischen 12 und 18 Atemzüge pro Minute (breaths per minute) (BrPM). Atemraten unter 12 BrPM werden als langsam und über 18 BrPM als schnell angesehen [Bec+17]. Der aktuelle Stand der Technik hinsichtlich der Atemratenerkennung, basiert auf Modulationen verschiedener Signalparameter der Herzschläge, welche durch die Atmung beeinflusst werden. Diese atmungsinduzierten Variationen sind hauptsächlich in den Amplituden, Intensitäten und Frequenzen des PPG-Signals zu beobachten [Cha+18; CVS16; Kar+13; Her+17].

Die Amplitudenmodulation (AM) kann auf ein reduziertes Schlagvolumen während der Einatmung zurückgeführt werden. Dies wird durch eine Änderung des intrathorakalen Drucks verursacht, was ein Absinken der Pulsamplitude bewirkt [Mer+11]. Die Basislinienmodulation (BM) wird durch Änderungen des Gewebeblutvolumens hervorgerufen. Diese sind eine Folge von Änderungen des intrathorakalen Drucks und der Vasokonstriktion der Arterien, während der Einatmung, wenn das Blut in die Venen fließt [NFFo6]. Die Frequenzmodulation (FM) wird durch die respiratorische Sinusarrhythmie (RSA) (siehe Abschnitt 2.1.3) verursacht, welche zu einem Anstieg der Herzrate während der Einatmung und einem Abfall während der Ausatmung [ZRF94] führt.

Um die Robustheit der Atemraten-Schätzung zu erhöhen, wird häufig eine Fusion der verschiedenen Modulationen durchgeführt und die Ergebnisse mehrerer Einzelverfahren gemeinsam ausgewertet. In der Literatur wird dies zur Verbesserung der Ergebnisse für die kontaktbasierte Erkennung der menschlichen Atemrate genutzt [Cha+18]. Dies insbesondere bei Bewegungen nützlich, da die einzelnen Methoden, in diesen Fällen eine hohe Störanfälligkeit aufweisen. Zudem sind die einzelnen atmungsinduzierten Schwankungen je nach Untersuchungsperson unterschiedlich stark. Einflussfaktoren sind z. B. individuelle Atemrate [Láz+14], Geschlecht [Li+10b] oder Alter [Cha+16].

2.3 Multispektrale Messungen

In der Literatur finden sich wenige Arbeiten, welche nicht den sonst üblichen RGB Ansatz zur Vitalparameterschätzung verfolgen. Einige Ansätze ver-

wenden ausschließlich Licht einer oder mehreren definierten Wellenlängen ohne Kamerafilter. Andere Ansätze verwenden optische Filter oder Spezialkameras für die Messung bestimmter Wellenlängen.

Wierenga et al. [WMS05] verwendeten LEDs mit drei verschiedenen Wellenlängen (660nm, 810nm und 940nm) und eine Monochromkamera, um eine räumliche Verteilung des PPG-Signals zu erfassen und schlugen eine mögliche Anwendung für die berührungslose Pulsoximetrie vor. Sun et al. [Sun+12] untersuchten den Einfluss der Kamera-Bildrate auf die Herzrate, unter Verwendung von Infrarot-Lichtquellen (880nm) sowie einer Hochgeschwindigkeitskamera mit 200 frames per second (FPS). Ihre Ergebnisse zeigten "keinen signifikanten Unterschied zwischen den verschiedenen Abtastraten" (20-200 FPS).

Spigulis et al. [Spi17] stellten verschiedene Prototypen zur Fernbestimmung von Haut- und Vitalparametern mittels multispektraler Beleuchtung vor. Darunter ein Gerät zur Überwachung der Anästhesieeffizienz während Operationen. Spigulis, Jakovels und Rubins [SJR10] verwendeten ebenfalls Laserbeleuchtung mit den Wellenlängen 532 nm und 635 nm, um eine multispektrale Bildgebung mittels einer RGB-Kamera ohne zusätzliche Kamerafilter zu erreichen.

Verkuyse et al. [Ver+17] präsentierten eine Kalibrierung für die kontaktfreie Pulsoximetrie, unter Verwendung von zwei monochromen Kameras mit Bandpassfiltern bei 675 nm und 842 nm. Amelard et al. [ACW16] entwickelten ein probabilistisches Modell zur Quantifizierung der ortsgebundenen arteriellen Pulsatilität. Sie verwendeten eine Monochromkamera mit einem optischen Bandpass (850 nm bis 1000 nm).

McDuff et al. [MGP14] verwendeten eine 5-Band-DSLR-Mosaikkamera mit einem RGBCO Sensor (RGB, Cyan, Orange). Eine Kombination der GCO-Farbkanäle übertraf, in ihren Experimenten, die Kombination aus den üblichen RGB-Kanälen.

Gupta et al. [GMR16b] kombinierten eine RGB-Kamera, eine monochrome Kamera mit einem *Magenta*-Filter und eine Wärmebildkamera und testeten verschiedene Eingabekombinationen mit dem von Poh. et al. [PMP11] vorgestellten Ansatz zur Herzratenmessung. Die Kombination aus grünen, roten und thermischen Kanälen ergab in ihren Experimenten die genauesten

Ergebnisse. Das *Philips-MedIT*-System wurde von Paul. et al. vorgestellt. [Pau+17]. Es besteht aus vier monochromen Kameras, einer Farbkamera, einer Wärmebildkamera und einer spezifischen Beleuchtung in verschiedenen Wellenlängen (weiß und NIR), die das sichtbare, nahinfrarote und langinfrarote Spektrum abdecken.

Martinez et al. [MPS11] haben ein Spektrometer verwendet, um das Signal-Rausch-Verhältnis (SRV) im sichtbaren und Nahinfrarotspektrum (NIR) (380 nm - 980 nm) zu berechnen und schlagen optimale Wellenlängenbereiche, mit hohem SRV für die Schätzung der Herz- und Atemrate, vor. Dieser Ansatz verwendet keine Kamera mit 2D Matrixsensor, kann aber als Ausgangspunkt für die kamerabasierte Schätzung der Vitalparameter verwendet werden.

2.4 Machinelles Lernen

Bisher gibt es in der aktuellen Literatur nur wenige Ansätze, welche maschinelles Lernen für die kamerabasierte PPG-Messung verwenden. Das vorherrschende Verfahren in der aktuellen Forschungslage sind tiefe neuronale Netze, auch CNN genannt. Das Fehlen von umfangreichen und öffentlich zugänglichen Datensätzen erschwert die Entwicklung sowie Validierung dieser Verfahren. Eine nicht zu unterschätzende Herausforderung ist zudem, die Frage der Echtzeitanwendbarkeit, bei steigender Komplexität der entwickelten Systeme. Allerdings zeigt die Verwendung von CNN bei der kamerabasierten PPG-Messung ein großes Potenzial gegenüber traditionellen Ansätzen der Bild- und Signalverarbeitung.

2.4.1 DeepPhys

Der von Chen und McDuff [CM18] vorgestellte **DeepPhys** Ansatz nutzt ein VGG CNN [SZ14] als Ausgangspunkt für Frame-basierte Generierung des PPG-Signales (siehe Abbildung 2.16). Im ersten Schritt werden die Videobilder, mittels bikubischer Interpolation (siehe Abschnitt 2.2.1) auf 36x36 Pixel herunterskaliert. Für die Bestimmung der PPG-Informationen in einem

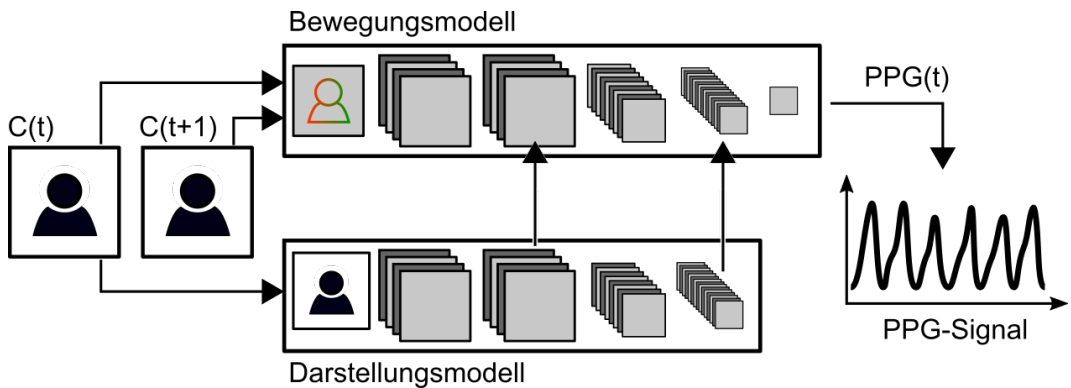


Abbildung 2.16: Schematischer vereinfachter Aufbau des **DeepPhys** Netzes, mit C als Farbbild zum Zeitpunkt t . (nach [CM18])

Videoframe, wird ein Bewegungsmodell verwendet, welches auf einer normierten Farbdifferenz von zwei aufeinanderfolgenden Frames ($C(t)$ und $C(t + 1)$) als Input basiert. Durch die Normierung und Differenzbildung werden jedoch viele Farbinformationen entfernt. Um dies auszugleichen, wurde ein separates Darstellungsmodell dem Netz hinzugefügt. Dieses Modell hat die gleiche Architektur, wie das Bewegungsmodell, mit Ausnahme der letzten drei Schichten. Für das Darstellungsmodell werden die Farbbilder (zentriert auf den Mittelwert 0 und skaliert auf eine Standardabweichung von 1) als Input genutzt. Anhand des Darstellungsmodells können zudem *Soft-Attention*-Masken bestimmt werden, um Hautbereichen mit stärkeren PPG-Signalen höhere Gewichtung zu geben, um die Messgenauigkeit zu verbessern. Diese Informationen werden an zwei Stellen in das Bewegungsmodell weitergegeben.

Der Ansatz übertraf, laut den Autoren, die Ergebnisse der verglichenen State-of-the-Art Ansätze, insbesondere bei weiten und schnellen Kopfdrehungen. Es wurden keine Aussagen über die benötigte Rechenzeit und Echtzeitfähigkeit des Systems getroffen.

2.4.2 PhysNet

Yu, Li und Zhao [YLZ19] stellten zwei räumlich-zeitliche Netzwerkmodelle zur Rekonstruktion von PPG-Signalen aus nicht vorverarbeiteten Gesichts-

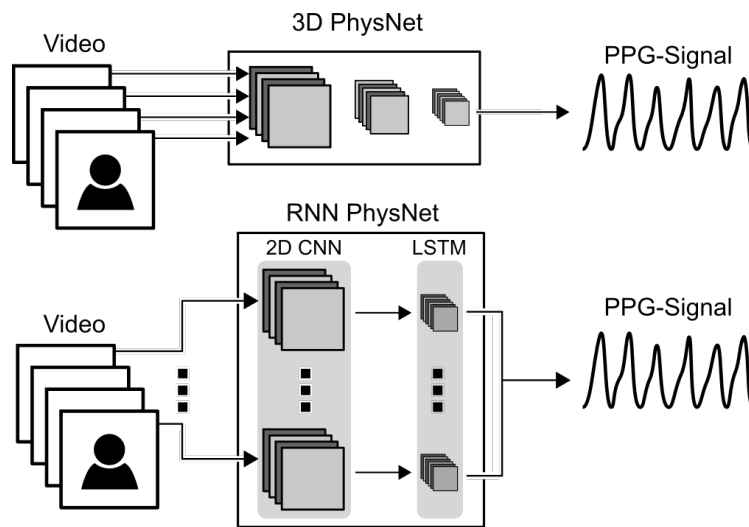


Abbildung 2.17: Schematischer vereinfachter Aufbau der zwei **PhysNet** Modelle.

videos vor. Das **3DCNN DeepPhys** und das auf rekurrenten neuronalen Netzen (RNN) aufbauende **RNN DeepPhys** Modell (siehe Abbildung 2.17). Im Gegensatz zu anderen CNN Verfahren wurde zusätzlich der Fokus auf die Messung der Herzratenvariabilität gelegt. In beiden Ansätzen werden mehrere Bilder gleichzeitig, als Input für das Modell verwendet, um die semantischen PPG-Merkmale im räumlichen und zeitlichen Bereich simultan zu extrahieren.

Der **3DCNN** Ansatz verwendet dreidimensionale Faltungen mehrerer aufeinanderfolgender Inputbilder, sowohl über die zwei Bilddimensionen (Höhe, Breite) als auch die Zeit. Dadurch sollen robustere Merkmale erlernt und die PPG-Signale mit den geringsten zeitlichen Schwankungen ermittelt werden können.

Beim **RNN** Modell werden die einzelnen Frames zunächst in ein 2D CNN Modell gegeben, um die räumlichen Merkmale zu extrahieren. Anschließend wird ein LSTM Modell, für die Transformation der räumlichen Merkmale in den Zeitbereich verwendet.

Die **DeepPhys** Ansätze erreichten gute Erkennungsraten, sowohl für die Herzrate, als auch für verschiedene Parameter der Herzvariabilität. Sie

zeigen zudem auch gute cross-data Generalisierungsfähigkeiten auf Daten mit stärkeren Kompressionsraten.

2.4.3 Extractor-Estimator-CNN

Spetlík, Franc und Matas [SFM18] haben ein CNN Modell vorgestellt, welches aus Videobildern direkt die Herzrate bestimmt. Das Modell besteht aus zwei Teilnetzen, dem Extractor (EX) und dem HR-Estimator (ES). Die beiden Komponenten werden einzeln trainiert. Die Bounding Box der gefundenen Gesichter wurden auf ein Seitenverhältnis 3:2 angepasst, ausgeschnitten und auf die Inputgröße des Netzes von 192×128 Pixel skaliert.

Der EX-Teil des Netzes konvertiert ein Videoframe in einen einzelnen skalaren Wert. Dabei wird der EX zunächst darauf trainiert, das Signal-Rausch-Verhältnis (SRV) zu maximieren. Dazu werden die Grundwahrheiten als Zielvorgaben verwendet. Für eine Abfolge von Videoframes kann so aus den Werten ein PPG-Signal generiert werden. Das PPG-Signal wird als Input für den ES-Teil des Netzes verwendet, welches daraus eine Herzrate schätzt. Das Training des ES hat als Ziel, den absoluten mittleren Fehler (MAE), zwischen der geschätzten Herzfrequenz und der Herzfrequenz der Grundwahrheit, zu minimieren.

Ein großer Nachteil der vorgestellten Methode ist, dass die Architektur des ES für jede Datenbank angepasst und neu trainiert werden muss, wodurch keine Generalisierung möglich ist. Es wurde zudem hauptsächlich auf stark komprimierten Daten validiert, wobei nur mäßige Ergebnisse erzielt wurden. Dies erschwert Aussagen über die Qualität des Ansatzes.

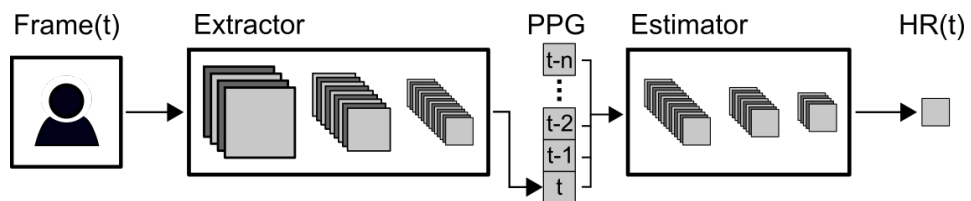


Abbildung 2.18: Schematischer vereinfachter Aufbau des **Extractor-Estimator** Netzes. (nach [SFM18])

2.4.4 RhythmNet

Niu, Shan, Han und Chen [Niu+19] haben das *RhythmNet* Modell vorgestellt, welches wie das *Extractor-Estimator* Modell aus einem Video direkt die Herzrate bestimmt. Dem CNN vorgesetzt ist eine räumlich-temporale Analyse, welche die Herzrhythmus-Signale isolieren und andere für die Herzfrequenz irrelevanten Informationen zu unterdrücken.

Ein Video wird zunächst in einzelne Videoclips (v_1, v_2, \dots, v_t) unterteilt, unter Verwendung eines Fensters mit 300 Frames Länge, das mit einer Schrittweite von 0,5 Sekunden verschoben wird. Die Videoclips werden zur Berechnung von räumlich-temporalen Karten der einzelnen Clips verwendet. Dabei werden die Gesichter der einzelnen Frames aneinander ausgerichtet und in den Luminanz/Chrominanz Farbmodell (YUV) Farbraum konvertiert. In diesem wird das Gesicht in 5×5 Unterregionen aufgeteilt. Aus den YUV Signalen aller Regionen wird dann eine $25 \times 3 \times 300$ Werte große räumlich-temporale Repräsentationen der Videoclips generiert, die als Eingabe für das folgende CNN Modell dient.

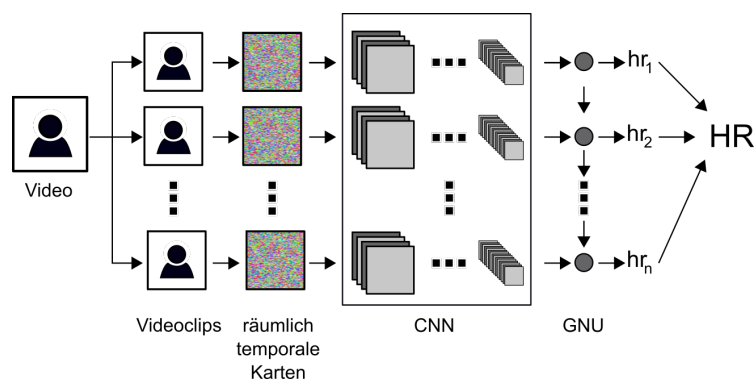


Abbildung 2.19: Schematischer vereinfachter Aufbau des **RhythmNet** Modells. (nach [Niu+19])

Die generierten Daten werden als Input in ein Modell auf Basis des *ResNet-18* [He+16] gegeben. Dieses wurde mit einer zusätzlichen Ausgabeschicht erweitert. Die aus den CNNs extrahierten Merkmale werden in eine einschichtige Gated Recurrent Unit (GRU) [Cho+14] eingespeist. Die Ausgabe jeder GRU wird verwendet, um die Herzraten-Werte der einzelnen Video-

clips zu regressieren. Für jedes Video wird der Durchschnitt aller Werte der einzelnen Videoclips als finale Herzrate berechnet.

Das **RhythmNet** übertraf bei der Validierung verschiedene Vergleichsalgorithmen, wenn dieselbe Datenbank für das Training und Testen verwendet wurde. Bei der Verwendung von unterschiedlichen Datenbanken stieg der mittlere Fehler an. Auf der stark komprimierten *Mahnob-HCI*, um mehr als den Faktor 12 und auf der MMSE-HR um den Faktor 3. Dies deutet auf eine noch schlechte Generalisierungsfähigkeit des Modells aufgrund der gegebenen beschränkten Datenlage. Es wurden keine Aussagen über die benötigte Rechenzeit und Echtzeitfähigkeit des Systems getroffen.

3 Neue Methoden für die kamerabasierte Vitalparametermessung

Aufbauend auf den Ansätzen im Stand der Technik, wurden neue experimentelle Methoden zur kamerabasierten Vitalparameterschätzung entwickelt und validiert. Da die Vitalparameterschätzung zu Beginn der Promotion ein junges Forschungsfeld war, gab es zu vielen Aspekten der Forschung keine aussagekräftigen Daten. Daher wurden während der Promotion neue Methoden für alle Teilschritte der Verarbeitungskette erarbeitet. Es wurden Ansätze für die Region of Interest (ROI) (Kapitel 3.1) und Signalverarbeitung (Kapitel 3.2, 3.3 und 3.4) entwickelt. Außerdem wurde eine kamerabasierte Methode zur Schätzung der Atemrate aus der Herzratenvariabilität (HRV) entwickelt (Kapitel 3.5).

3.1 Hauterkennung

Um eine schnelle und zuverlässige Hauterkennung zu erreichen, wurde die in [SA14a] vorgestellte Implementation einer Look-up-Table (LUT) verwendet und für die Vitalparameterschätzung angepasst, und so eine parameterfreie ROI-Generierung ermöglicht. Die neue ROI erlaubt mit geringerem Rechenaufwand die Auswertung der Vitalparameter in Echtzeit.

Diese Hauterkennungsmethode basiert auf dem *BayesBGR* $3^2 \times 3$ Ansatz [JR99]. Das Bayes'sche Theorem wird als Grundlage verwendet, um für jede RGB-Farbe eine Hautfarbenwahrscheinlichkeit p zu modellieren. Die

Hautfarbenwahrscheinlichkeit für einen Pixel mit der Farbe c_{RGB} wird nach dem *Bayesschen* Satz bestimmt:

$$P(\text{Haut}|c_{RGB}) = \frac{P(c_{RGB}|\text{Haut})P(\text{Haut})}{P(c_{RGB})} \quad (3.1)$$

Dabei ist $P(c_{RGB}|\text{Haut})$ die relative Häufigkeit der Hautfarbenpixel von der Gesamtmenge der Pixel mit der Farbe c_{RGB} , $P(\text{Haut})$ der Anteil der Hautpixel von allen Pixeln und $P(c_{RGB})$ der Anteil der Pixel mit der Farbe c_{RGB} von allen Pixeln. Daraus ergibt sich:

$$\frac{P(c_{RGB}|\text{Haut})P(\text{Haut})}{P(c_{RGB})} = \frac{\frac{n(c_{RGB}, X_{\text{Haut}})}{N(X_{\text{skin}})} \frac{N(X_{\text{skin}})}{N(X)}}{\frac{n(c_{RGB}, X)}{N(X)}} = \frac{n(c_{RGB}, X_{\text{Haut}})}{n(c_{RGB}, X)} \quad (3.2)$$

wobei X und X_{Haut} die Mengen aller Pixel und Hautpixel, $n(c_{RGB})$ die Anzahl der Farbe c_{RGB} und $N(X)$ die Gesamtanzahl aller Pixel im Datensatz darstellt. Die Hautfarbenwahrscheinlichkeit kann auf die Darstellung in der letzten Gleichung reduziert werden und resultiert dadurch, für jedes Pixel mit der Farbe c_{RGB} , in dem Verhältnis der Anzahl an Hautpixeln der Farbe $n(c_{RGB}, X_{\text{Haut}})$ zu allen Pixeln der Farben $n(c_{RGB}, X)$ im Datensatz. Abbildungen 3.1 und 3.2 zeigen zwei Beispiele der resultierenden Hautfarbenwahrscheinlichkeit. Durch das Setzen einer Schwelle τ für die Wahrscheinlichkeit können die Pixel in Haut/nicht-Haut klassifiziert werden. Diese Schwelle kann durch die empirische Minimierung des Klassifikationsfehlers auf dem Datensatz bestimmt werden.

Die Farbhäufigkeiten können für jede RGB-Kombination gespeichert werden und als Look-Up-Table während der Laufzeit verwendet werden. Zur Reduzierung des Speicheraufwandes wird die Farbtiefe der Pixel von 8 Bit (256 Stufen) auf 5 Bit (32 Stufen) reduziert. Dies hat laut [JR99] kaum Auswirkungen auf die Segmentationsgenauigkeit.

Unter Verwendung der Hautwahrscheinlichkeit p_i als Gewichtungsfaktor (siehe Abb. 3.1 und 3.2), wird jeder Pixel i bei der Berechnung des Pixelmittelwerts C_n (für je R,G,B) im Bild mit seiner Hautwahrscheinlichkeit p_i

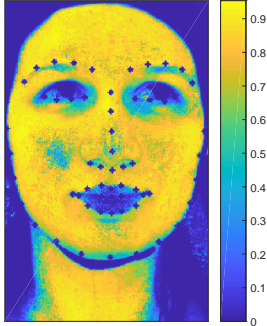


Abbildung 3.1: Beispiel der Hautwahrscheinlichkeit (BioVid Datenbank).

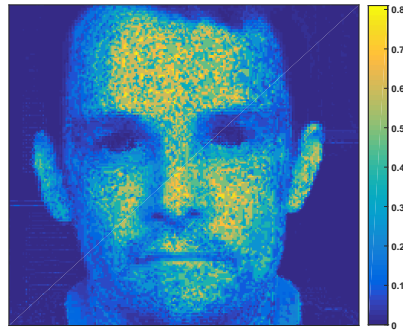


Abbildung 3.2: Beispiel der Hautwahrscheinlichkeit (PURE Datenbank).

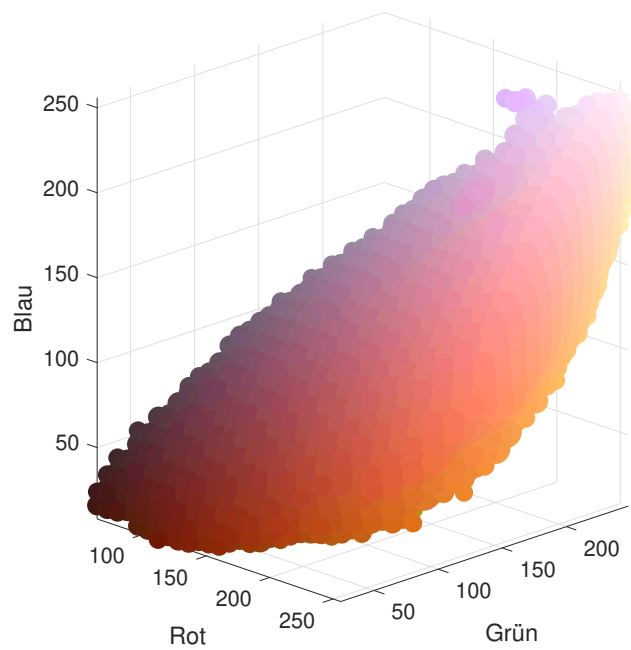


Abbildung 3.3: Darstellung der als Haut klassifizierten Farben (mit $p > 0.1$) der verwendeten Look-Up-Table.

gewichtet und aufsummiert (siehe Gleichung 3.3). Zudem wird die Normierung durch die Summe der Gewichte p_i durchgeführt:

$$C_n = \frac{1}{\sum_{i=1}^n p_i} \cdot \sum_{i=1}^n p_i c_i \quad (3.3)$$

So wird eine parameterfreie normierte haut-basierte ROI generiert, anstelle der binären Maskierung auf Haut/Nicht-Haut-Pixel. Dadurch kann dieser Ansatz flexibler mit unterschiedlichen datenbankspezifischen Unterschieden, wie Beleuchtung oder Verdeckungen verwendet werden.

Für die Generierung der LUT in dieser Arbeit wurde der *ECU face and skin detection* Datensatz [PBC05] verwendet. Dieser enthält 4000 Bilder und korrespondierende Hautmasken, welche durch manuelle Segmentierung des Gesichts und der Hautregionen erstellt wurden und enthält insgesamt 208,8 Millionen Haut- und 901,8 Millionen nicht-Hautpixel. Abbildung 3.3 zeigt eine Darstellung der in der LUT als Haut klassifizierten Farben.

3.2 Adaptiver Bandpass

Für die Stabilisierung der Herzratenschätzung wurde ein neues adaptives Filterverfahren entwickelt, um sowohl eine genaue Bestimmung der Herzrate als auch schnell auf dynamische Änderungen des Pulses zu ermöglichen. Die in Kapitel 2.2.3 vorgestellten Signalverarbeitungsmethoden verwenden Bandpässe mit festen Cutoff-Frequenzen und einem weiten Passband, in der Regel von 0,5 – 4 Hz, über die komplette Spanne der physiologisch möglichen Herzfrequenzen. Dadurch können Störfrequenzen in diesem Bereich nicht herausgefiltert werden und überlagern unter Umständen das PPG-Signal und verhindern die erfolgreiche Messung der Herzfrequenz. Insbesondere die hohen Frequenzen (> 120 BPM) werden in Ruhesituationen nur sehr selten erreicht und sind daher häufig eine Störquelle im Frequenzspektrum.

Das neu entwickelte dynamische Bandpassverfahren, das sich an die zuvor geschätzte Herzfrequenz anpasst, kann diese Störungen dynamisch filtern.

Der Ansatz verwendet Filter mit endlicher Impulsantwort (FIR) und nutzt eine Kombination aus einem statischen Bandpass mit großem Passband und den oberen und unteren Cutoff-Frequenzen f_{min} und f_{max} und einem dynamischen Bandpass mit einem Passband niedrigerer Breite und den Cutoff-Frequenzen f_{low} und f_{high} . Abbildung 3.4 zeigt einen beispielhaften Verlauf des dynamischen Bandpasses während der Herzratenschätzung.

In einem Initialisierungsschritt wird zunächst der statische BP (f_{min} bis f_{max}) für die Filterung verwendet. Für die weiteren Zeitschritte wird der dynamische Bandpass genutzt. Falls keine gültige Herzfrequenz vorliegt, kann entweder auf einen früheren Zeitschritt zurückgegriffen, oder der Filter neu initialisiert werden. Wenn eine Herzfrequenz in einem vorhergehenden Zeitschritt geschätzt wurde, werden die dynamischen Passbandfrequenzen f_{low} und f_{high} auf ± 15 BPM ($0,125\text{Hz}$), um die geschätzte Herzfrequenz definiert.

Damit eine schnelle Filterung in Echtzeitanwendungen sichergestellt werden kann, wird für den dynamischen Bandpassfilter eine Look-up-Table der Filterparameter im Voraus berechnet und bereitgestellt, um rechenintensive Neuberechnungen der Parameter während der Laufzeit zu vermeiden. Nach Angabe der Parameter für die Cutoff-Frequenzen, die dynamische Bandpassbreite, die verwendete Bildwiederholrate und die Zeitfensterlänge, werden die Filterparameter für alle ganzzahligen Herzraten innerhalb f_{min} und f_{max} in $1/60$ Hz (1 BPM) Schritten berechnet. Dabei werden für jeden Filter mit der korrespondierenden Herzrate f_{hr} die Cutoff-Frequenzen $f_{hr} \pm 0,125\text{Hz}$ verwendet. Die Ordnung der Filter O_{BP} wird anhand der Bildwiederholrate FPS und der Zeitfensterlänge L berechnet, um eine möglichst hohe Filterordnung zu erzielen (siehe Gleichung 3.4). Dabei wird die Ordnung auf einen ganzzahligen Wert gerundet.

$$O_{BP} = \lfloor (FPS * L/3) \rfloor - 1 \quad (3.4)$$

Zudem wird ein FIR Bandpassfilter für den Initialisierungsfall mit den Cutoff-Frequenzen f_{min} bis f_{max} erstellt. Für die Filterung während der Herzratenschätzung wird die zuletzt geschätzte Herzfrequenz an den Filter übergeben, welchen dann entweder, im Initialisierungsfall, den statischen

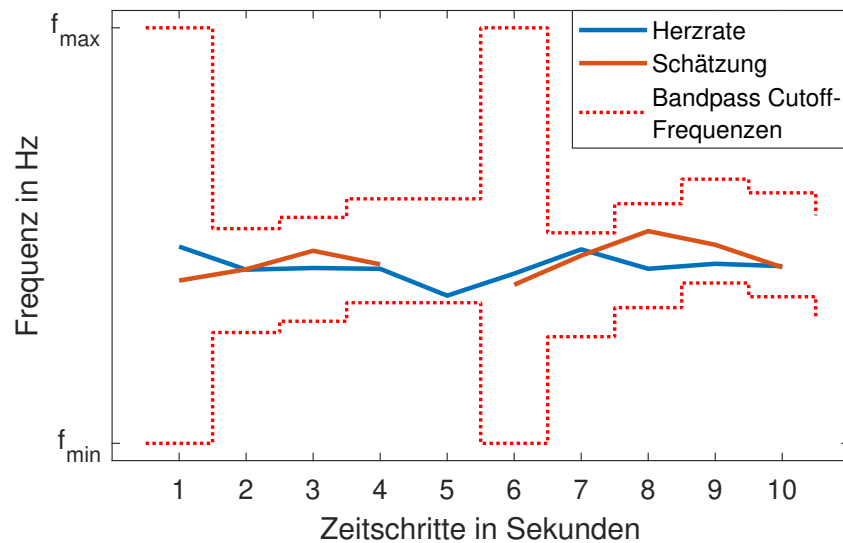


Abbildung 3.4: Beispiel des Verhaltens des adaptiven Bandpasses und den abgeleiteten Cutoff-Frequenzen, mit einer fehlenden Herzratenschätzung bei Zeitschritt 5.

Bandpass, oder im dynamischen Fall den Filter der nächstliegenden ganzzahligen Herzrate (in BPM) für die Filterung des Signales verwendet.

3.3 Graphenbasierte Peakselektion (IBI-Graph)

Nach der Signalverarbeitung und Generierung des PPG-Signales treten häufig noch Artefakte aus verschiedenen Störquellen im Signal auf. Diese führen häufig zu kleineren lokalen Maxima, welche die Berechnung der Herzrate erschweren. Bei einer einfachen Mittlung der zeitlichen Pulsabstände (Inter-Beat-Interval, IBI) können schon wenige Störpeaks großen Einfluss auf die korrekte Bestimmung der Herzrate haben, da ein zusätzlich erkannter Störpeak zu zwei falsch bestimmte IBIs führt (siehe Abbildung 3.5a). Der in [RWA16] vorgestellte Algorithmus sucht daher nach der Folge von Maxima (Peaks) welche eine möglichst gute Repräsentation der zugrundeliegenden Herzrate sind. Um dies zu erreichen, werden die zeitlichen Abstände der gefundenen Peaks (IBI) mithilfe eines Graphen dargestellt

und ein optimaler Pfad gesucht, um mögliche Störungen im Signal zu kompensieren.

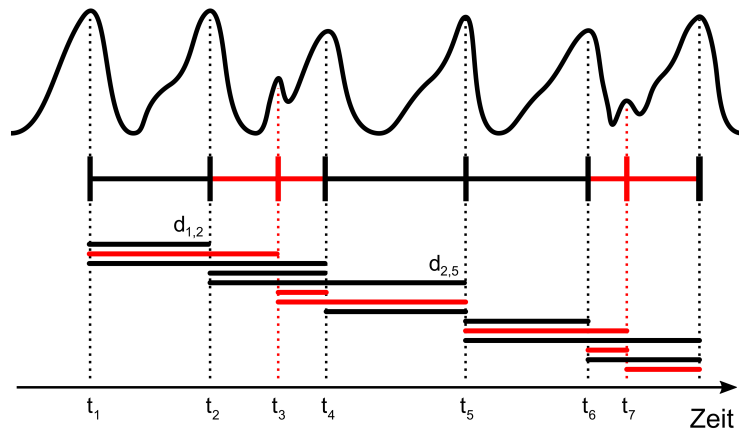
Als ersten Schritt werden zunächst die Peaks im Signal erkannt (siehe Abbildung 3.5a)). Für alle Peaks p_i wird das IBI $d_{i,j}$ zu jedem zukünftigen Peak p_j berechnet. Jedes IBI wird dabei mit der Zeit t_i des ersten Peaks und dem IBI $d_{i,j}$ beschrieben. Aufgrund der angenommenen Beschränkung der Herzrate werden nur IBIs mit d Werten von 0,3 bis 2 Sekunden betrachtet.

Jedes IBI enthält damit die Informationen von zwei Peaks. Diese werden anschließend als Knoten zu einem Graphen verbunden (siehe Abbildung 3.5b), wodurch nun jede Verbindung zwischen zwei Knoten Informationen aus drei Peaks enthält. Es werden nur die Knoten verbunden, welche einen zeitlichen Abstand von 0,3 bis 2 Sekunden haben, welches einer menschlichen Herzrate von etwa 30-200 BPM entspricht. Die Knoten und Kanten des Graphen beschreiben alle möglichen Peak-Folgen in dem betrachteten Zeitfenster, welche die Mindestvoraussetzungen der Herzratenschätzung erfüllen. Auf Grundlage dieses Graphen wird ein Pfad gesucht, welcher eine kontinuierliche Herzrate im Zeitfenster annimmt und versucht die Störpeaks zu umgehen.

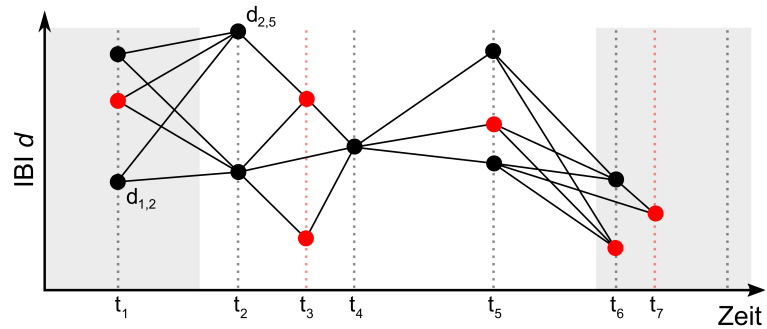
Um eine schnellere und genauere Herzratenbestimmung zu ermöglichen, werden bei der Erstellung des Graphen und der Pfadsuche die Informationen des vorherigen Zeitschrittes berücksichtigt. Dadurch wird das Verfahren in zwei Fälle unterteilt, den Initialisierungsfall und den Fortsetzungsfall. Im **Initialisierungsfall** wird der Pfad mit der minimalen mittleren quadratischen Abweichung \bar{e} der Pfadknoten vom Pfad-IBI Mittelwert \bar{d} gesucht. Im **Fortsetzungsfall** wird der von der letzten Herzratenschätzung ermittelte Pfad, in dem Graphen des neuen, zeitlich überlappenden, Fensters rekonstruiert und als Ausgangspunkt für die Erweiterung des aktuellen Pfades verwendet.

Initialisierungsfall Im Initialisierungsfall stehen keine Informationen über eine vorhergehende Herzrate zur Verfügung. Daher werden mehrere Pfade durch den Graphen gebildet und bewertet. Alle Knoten mit $t < 2$ Sekunden und ohne Eingangskanten werden als mögliche Startpunkte definiert.

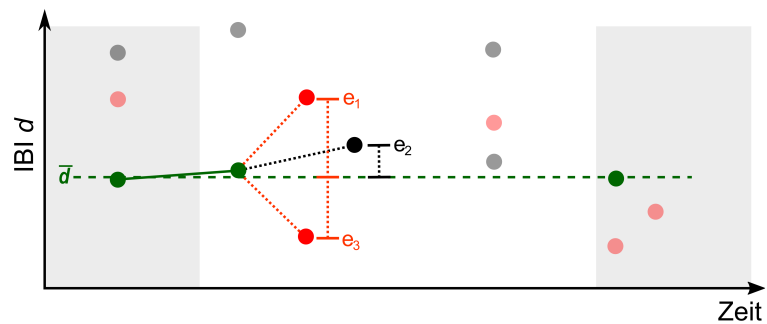
3 Neue Methoden für die kamerabasierte Vitalparametermessung



(a) PPG-Signal mit eingezeichneten Peaks und Peakabständen (IBIs) $d_{i,j}$.



(b) IBI-Graph mit verbundenen Knoten. Jeder Knoten repräsentiert einen IBI aus Abbildung 3.5a mit der Höhe $d_{i,j}$ am Zeitpunkt t_i . Grau hinterlegte Bereiche enthalten die Start und Endknoten für die Pfadsuche.



(c) Beispiel für die Bestimmung des nächsten Pfadknotens. Grüne Elemente sind bereits Teil des Suchpfades und \bar{d} der Mittelwert des Pfades. Der Knoten mit dem geringsten Abstand e zu \bar{d} wird dem Pfad hinzugefügt.

Abbildung 3.5: Beispiele für (a) die Generierung der Inter-Beat-Intervalle (IBI), (b) den Graphen und (c) der Peakauswahl zur Herzratenbestimmung. Durch Störpeaks entstandene Elemente sind rot markiert.

Analog werden alle Knoten in den letzten zwei Sekunden und ohne Ausgangskanten als mögliche Endknoten bestimmt. Diese Bereiche sind in den Abbildungen 3.5b und 3.5c grau hinterlegt. Alle Kombinationen der Start und Ende Knoten werden auf eine mögliche Verbindung durch den Graphen geprüft. Nicht durch den Graphen verbundene Paare und Paare, deren d Werte um mehr als 0,5 Sekunden auseinander liegen, werden nicht weiter betrachtet.

Für jedes Start-Ende Knotenpaar wird einen Pfad durch den Graphen gebildet. Aus den d -Werte der Start und Endknoten wird der Pfadmittelwert \bar{d} gebildet (siehe Abbildung 3.5c). Ausgehend vom Startknoten wird dem Pfad nun der nächste Knoten hinzugefügt. Dabei werden die mit dem Startknoten verbundenen folgenden Knoten betrachtet. Der Knoten mit der geringsten Abweichung e vom Pfadmittelwert \bar{d} , wird dem Pfad hinzugefügt und der Pfadmittelwert \bar{d} aktualisiert. Ausgehend vom neu hinzugefügten Knoten werden weitere Knoten dem Pfad hinzugefügt, bis ein Endknoten erreicht ist.

Falls der erreichte Endknoten nicht zu dem gewählten Start-Ende Knotenpaar gehört, wird vom ursprünglichen Endknoten aus, analog ein zweiter Pfad rückwärts durch den Graphen gebildet. Wenn der rückwärtige Pfad einen Knoten des vorwärts gebildeten Pfades erreicht, werden die beiden Pfade an diesem Knoten verbunden.

Nachdem für jedes Start-Ende Knotenpaar ein Pfad durch den Graphen bestimmt wurde, werden diese verglichen und bewertet. Für jeden Pfad wird zunächst der IBI-Mittelwert \bar{d} aus den d -Werten der Knoten berechnet. Anschließend kann die mittlere quadratische Abweichung \bar{e} des Pfades mit n Knoten wie folgt berechnet werden:

$$\bar{e} = \frac{1}{n} \sum_{i=1}^n (d_i - \bar{d})^2 \quad (3.5)$$

Der Pfad mit dem geringsten Fehler \bar{e} wird für die Berechnung der Herzrate hr ausgewählt, welche aus dem Mittelwert der IBIs \bar{d} bestimmt wird.

$$hr = \frac{1}{\bar{d}} * 60 \quad (3.6)$$

Fortsetzungsfall Falls bei der letzten Herzratenschätzung ein zulässiger Pfad bestimmt werden konnte, werden diese Informationen bei der Generierung des neuen Graphen und der Pfadsuche verwendet. Mit dem Ziel, den Pfad der letzten Schätzung im neuen Zeitfenster fortzusetzen.

Der gerichtete Graph wird zunächst analog zum *Initialisierungsfall* gebildet. Die Knoten des vorherigen Pfades im neuen Graphen identifiziert und zum aktiven Pfad hinzugefügt. Dabei wird der letzte Knoten ignoriert, um eine größere Varianz in der Pfadsuche zu ermöglichen und eine mögliche suboptimale, letzte Endknotenwahl des Algorithmus auszugleichen. Dies ermöglicht zudem eine schnellere Anpassung auf eine sich ändernde Herzrate. Dem Pfad werden, nach den im Initialisierungsfall beschriebenen Regeln, weitere Knoten hinzugefügt bis dieser einen Endknoten ohne Ausgangskanten erreicht. Im Fortsetzungsfall findet keine rückwärtige Pfadsuche statt. Nachdem der Pfad durch den Graphen bestimmt wurde, wird die Herzrate aus dem Mittelwert \bar{d} der Pfadelemente bestimmt.

3.4 Pulserkennung durch LSTM

Alternativ zu den klassischen Signalverarbeitungsmethoden, welche in den vorherigen Kapiteln vorgestellt wurden, wurde eine Methode zur Pulsdetektion auf Grundlage von Long Short-Term Memory Netzen (LSTM) entwickelt, welche eine Sonderform der CNN darstellen. LSTMs basieren auf rekurrente neuronale Netze (RNN) und sind optimiert, um Sequenzen von Daten zu verarbeiten. Grundlage für das Verfahren ist die *ecg2rr Peak Detektion* von Laitala, Jiang, Syrjälä, Naeini, Airola, Rahmani, Dutt und Liljeberg [Lai+20], welche zur Detektion der R-Peaks und QRS-Komplexe in EKG Signalen entwickelt wurde.

Das in [Lai+20] vorgestellte LSTM Modell wurde für die kamerabasierte Schätzung der Vitalparameter angepasst. Anstelle der EKG-Signale werden die gemittelten Farbwerte der ROIs als Eingangssignale des Netzes verwendet. Nach der Gesichtserkennung und der ROI Bestimmung werden für ein Video mit l Bildern die gemittelten RGB Werte extrahiert. Diese können als Eingangssignale für das LSTM Netz verwendet werden, oder durch andere Verfahren, wie dem *CHROM* Ansatz vorher weiterverarbeitet werden. Die

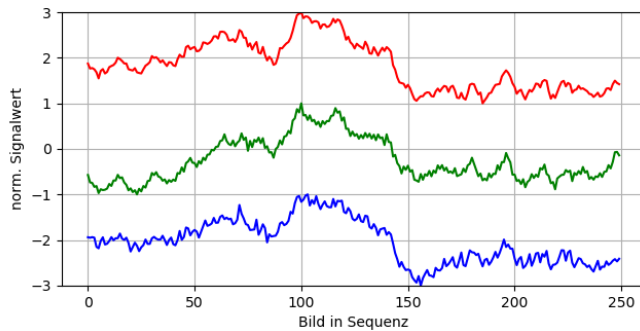


Abbildung 3.6: Beispiel für normierte (-1 bis 1) RGB Eingangssignale. (Das rote und blaue Signal wurden für eine bessere Übersicht um ± 2 verschoben.)

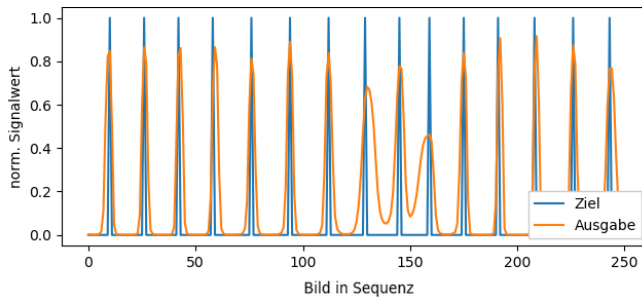


Abbildung 3.7: Beispiel für die im Training verwendeten Zielsequenzen und eine Ausgabe eines trainierten Netzes.

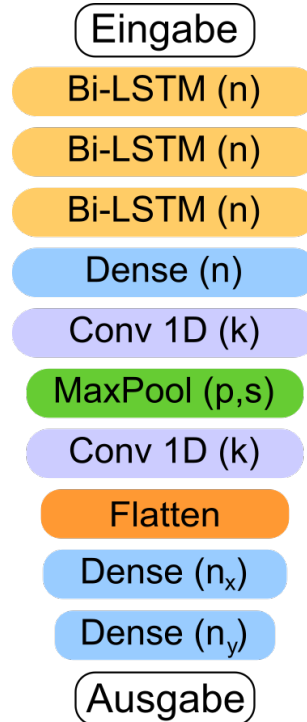


Abbildung 3.8: Aufbau des LSTM Netzes

Eingangssignale werden für jeden Kanal auf den Bereich -1 bis 1 normiert (siehe Abb. 3.6).

Für das Training werden die zu den Videosignalen korrespondierenden Grundwahrheiten verwendet. Dabei wird für jede Eingangssequenz ein Zielsignal, mit der Länge l , aus den Herzschlägen der Grundwahrheit generiert. Das Zielsignal hat an den Zeitpunkten mit Herzschlag den Wert 1 ansonsten 0 (siehe Abb. 3.7).

Das LSTM-Netz führt eine Sequenz-zu-Sequenz-Zuordnung durch, bei der sowohl die Eingangs- als auch die Ausgangssequenz die gleiche Länge haben. Abbildung 3.8 zeigt den Aufbau und die aufeinanderfolgenden

Schichten. Das Netz hat eine Eingangsschicht mit der Größe $l \times c$, mit der Länge der Sequenz l und c Kanälen. Diese werden durch drei bidirektionale LSTM Schichten mit jeweils n Ausgangsdimensionen verarbeitet. Anschließend wird durch eine Reihe von *Dense*, *Convolution* und *MaxPooling* Schichten eine Ausgangssequenz generiert. Die *Convolution* Schichten werden mit einem Kernel der Größe k nur in der Zeitdimension angewendet. Das *MaxPooling* wird mit den Parametern p für die *Pooling*-Größe und s für den *Stride* ebenfalls nur in der Zeitdimension durchgeführt. Zuletzt werden alle vorhandenen n Dimensionen mit einer *Flatten* Schicht in einen 1D Signalvektor umgewandelt. Die *Dense* Schichten am Ende des Netzes haben eine sich verringernde Zahl an Neuronen $n_x > n_y$ und enden in der Ausgangsschicht mit einer Größe von $l \times 1$. Aus der Ausgangssequenz (siehe Abb. 3.7) können die Pulse erkannt und daraus die Herzrate bestimmt werden.

3.5 Bestimmung der Atemrate

Im Rahmen dieser Arbeit wurde ein neuer Algorithmus zur Erkennung der Atemrate aus Videosequenzen von menschlichen Gesichtern entwickelt. Dieser basiert auf der Modulation von atmungsbedingten Schwankungen der Herzrate, eines aus Videosequenzen extrahierten PPG-Signals und wurde bereits in [FRA20] und [FRA21] publiziert. Dabei wurden die in Kapitel 2.2.4 beschriebenen Methoden für die kontaktbasierte Atemratenmessung für die kamerabasierte Messung angepasst. Die Schätzung der Atemrate kann in Vorverarbeitung, Modulation, Nachbearbeitung und Fusion der Ergebnisse unterteilt werden und ist in Abbildung 3.9 schematisch dargestellt.

3.5.1 Vorverarbeitung

Für die Schätzung der Atemrate wird das Signal in kleinere Abschnitte einer konstanten Länge unterteilt und diese für die kontinuierliche Berechnung um ein definiertes Zeitintervall zu verschieben.

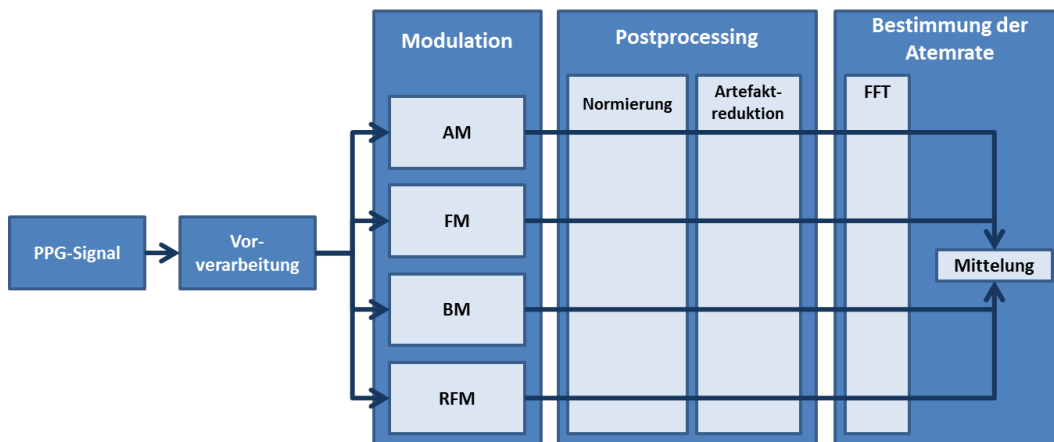


Abbildung 3.9: Schematischer Ablauf der Atemratenmessung.

Bislang gibt es in der Literatur keinen Konsens, welche Länge und Schrittweite des Signalfensters als optimal ist. Die meisten Studien verwenden Fensterlängen zwischen 30 und 90 Sekunden [Cha+18]. Je nach Anwendung und Einsatz kann eine der beiden Fensterlängen die geeignetere Lösung sein, z. B. 30 Sekunden für eine zeitlich flexiblere und 90 Sekunden für eine stabilere Langzeitüberwachung, bei der kurze Änderungszeiträume nicht stark berücksichtigt werden sollten. Kürzere und längere Fensterlängen sind unserer Ansicht nach für die Atmungserkennung nicht geeignet. Die untere Grenze muss jedoch mindestens 20 Sekunden betragen, wenn man davon ausgeht, dass die minimal mögliche Atmung eines gesunden Menschen auf sechs BrPM fallen kann. Um die Erkennung für diesen Fall zu gewährleisten, müssen mindestens zwei vollständige Atemzüge im Signal detektierbar sein [SN18].

Die einzelnen PPG-Signalfenster werden zunächst mit einem zero-phase Butterworth-Bandpass zweiter Ordnung gefiltert. Dabei wird eine untere Grenzfrequenz von 0,5 Hz und eine obere Grenzfrequenz von 4 Hz verwendet. Anschließend werden alle systolischen Maxima (max) und Minima (min) erfasst.

3.5.2 Modulationen

Die Atmung moduliert das PPG-Signal aufgrund verschiedener physiologischer Ursachen, wie der respiratorische Sinusarrhythmie (RSA) (siehe Kapitel 2.1.3) und kann anhand der Minima und Maxima des PPG-Signals abgeleitet werden. Aus verschiedenen Signalparametern kann dadurch die im PPG-Signal modulierte Atmung extrahiert werden.

Dabei werden die verschiedene Signalparameter wie die Amplitude (AM), mehrere Basislinien (BM) und Frequenz (FM) und verschiedene Herzschlagspezifische Zeitverläufe (RFM) verwendet, um die Atemrate zu schätzen. Jeder Herzschlag im PPG-Signal wird anhand der oben genannten Parameter in einen Zeit- und Signalwert umgewandelt. Insgesamt wurden zu diesem Zweck sieben Varianten implementiert: eine AM, drei BMs und drei FMs. Abb. 3.10 zeigt ein Beispiel eines PPG-Signals, die resultierenden modulierten Atmungssignale (vor der FFT-Analyse) für jede der drei Modulationsarten (AM, BM und FM) und das Grundwahrheits-Atmungssignal, um die verarbeiteten Signale zu veranschaulichen.

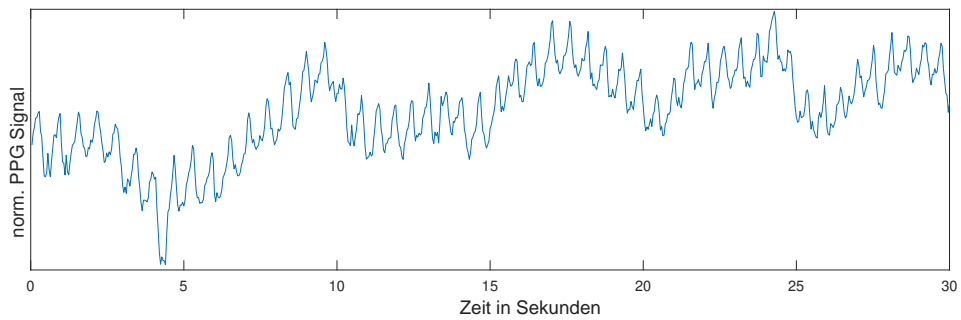
Amplitudenmodulation (AM) Die AM basiert auf Änderungen der Spitze-Spitze-Amplitude. Für jeden Herzschlag i im Zeitfenster werden zunächst zwei Punkte definiert, das lokale Maximum und das lokale Minimum zwischen dem betrachteten Herzschlagmaximum und dem Maximum des folgenden Herzschlages.

Dies ist beispielhaft in Abb. 3.11 dargestellt. Aus den Zeitpunkten $t_{i,max}$ und $t_{i,min}$ dieser Extrempunkte und den Signalwerten $PPG(t)$ an diesen Punkten werden die Werte t_i und PPG_i für das AM-Signal bestimmt. Dabei wird PPG_i aus den beiden zugehörigen PPG-Werten gemittelt (siehe Gleichung 3.7) und t_i aus dem Mittelwert ihrer zeitlichen Abstände bestimmt (siehe Gleichung 3.8).

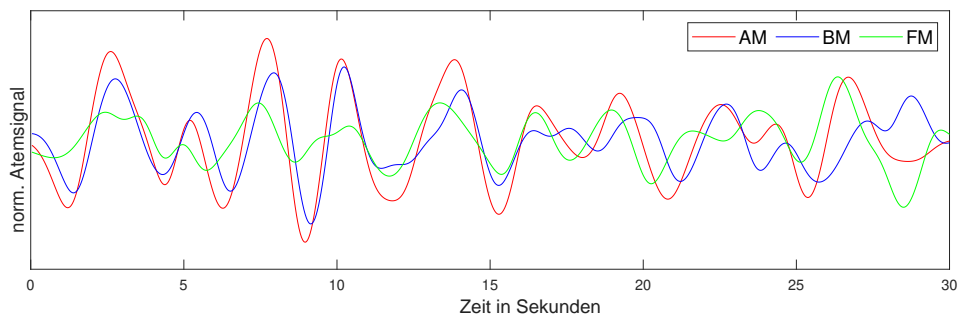
$$PPG_{i, AM} = PPG_{i, max} - PPG_{i, min} \quad (3.7)$$

$$t_{i, AM} = \frac{t_{i, max} + t_{i, min}}{2} \quad (3.8)$$

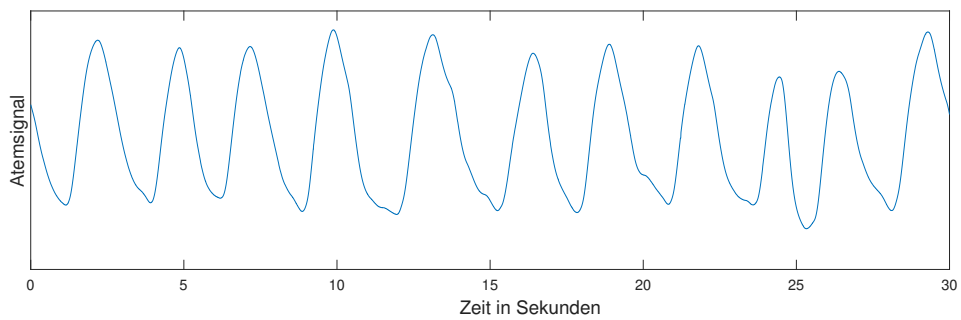
3.5 Bestimmung der Atemrate



(a) Normiertes PPG-Signal mit erkennbarem Pulssignal



(b) modulierte Atemwegssignale



(c) Atmungssignal Grundwahrheit

Abbildung 3.10: Beispielhafte Darstellung von (a) dem PPG-Signal mit erkennbarem Pulssignal, (b) den resultierenden modulierten respiratorischen Signalen für jede der drei Modulationsarten und (c) der Atmungsgrundwahrheit.

Basislinienmodulation (BM) Drei unterschiedliche Basislinien (Halbwert, Maxima und Minima) werden aus dem PPG Signal extrahiert. Es wird je eine Basislinie für alle Maxima und alle Minima gebildet. Dabei werden die Zeitpunkte $t_{i,max}$ und $t_{i,min}$ der Extrempunkte und deren Signalwert PPG_i an diesen Punkten verwendet.

Die verschiedenen BMs sind beispielhaft in Abb. 3.12 dargestellt. Für die Halbwert-BM wird für jeden Herzschlag i erst der Mittelwert der beiden Pulsextrema PPG_i berechnet:

$$PPG_{i, BM(Halb)} = \frac{PPG_{i, max} + PPG_{i, min}}{2} \quad (3.9)$$

Für t_i wird der Zeitpunkt auf der rechten Flanke des Maximums verwendet, welcher dem berechneten PPG_i Wert am nächsten liegt.

Frequenzmodulation (FM) Für die Berechnung der drei FM (Maxima, Minima und HR-Maxima) werden unterschiedliche frequenzabhängige Parametern betrachtet.

Die ersten beiden Ansätze verwenden die Zeitabstände zwischen aufeinanderfolgenden Extremwerten, einmal zwischen den Maxima und einmal zwischen den Minima.

$$PPG_{i, FMmax} = t_{i+1, max} - t_{i, max} \quad (3.10)$$

$$PPG_{i, FMmin} = t_{i+1, min} - t_{i, min} \quad (3.11)$$

Die dritte FM basiert auf der Herzfrequenz in BPM. Für die Bestimmung der Herzrate werden dabei die Inter-Beat-Intervalle der Maxima verwendet:

$$PPG_{i, frequency} = 60 \cdot \frac{1}{t_{i+1, max} - t_{i, max}} \quad (3.12)$$

Für den t_i Wert werden die t_{i+1} Werte der jeweils nächsten Extrema genutzt. Abb. 3.13 zeigt die bestimmten Intervalle der FM. Die ermittelten atmungsinduzierten Parameter werden anschließend linear interpoliert. Als Frequenz wird die ursprüngliche Bildrate verwendet.

3.5 Bestimmung der Atemrate

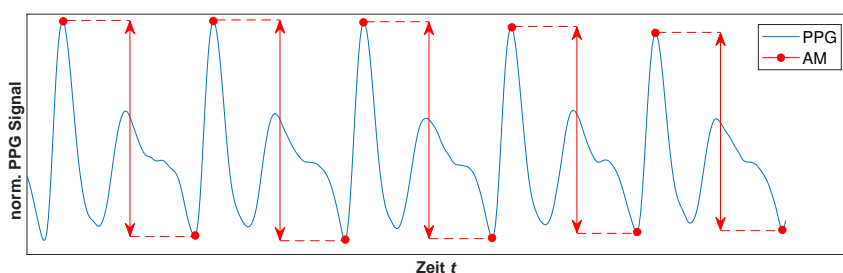


Abbildung 3.11: Beispiel für die berechneten Pulsamplituden (rote Pfeile) des PPG-Signals (blau).

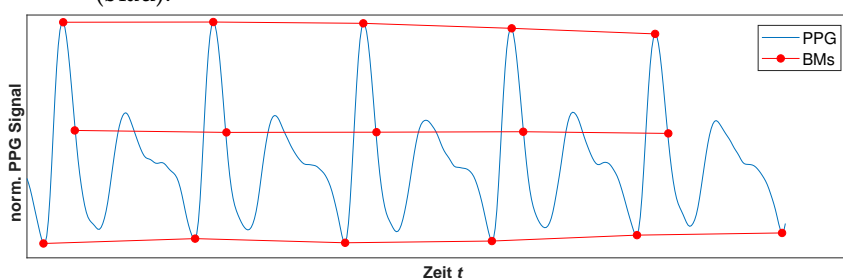


Abbildung 3.12: Beispiel für die berechneten Basislinien (rot) für die BMs im PPG-Signal (blau).

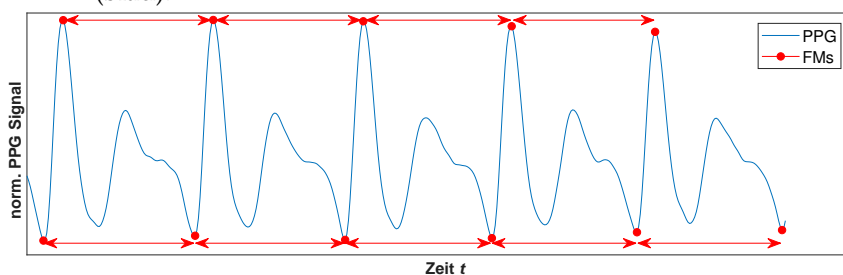


Abbildung 3.13: Beispiel für die berechneten Periodendauern (rote Pfeile) für die FMs des PPG-Signals (blau).

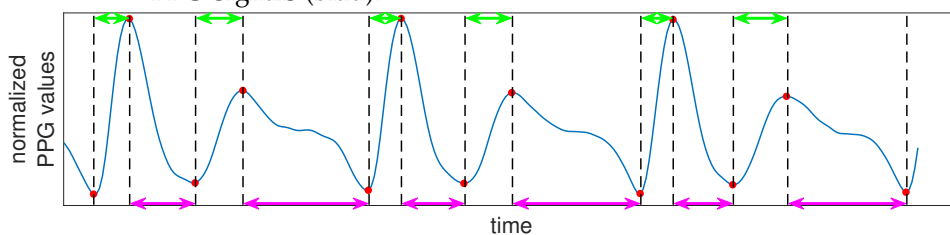


Abbildung 3.14: Beispiel für die berechneten systolischen Anstiegszeiten (grüne Pfeile) und diastolischen Abfallzeiten (magentafarbene Pfeile) der Rise-Fall-Modulation des PPG-Signals (blau).

Rise-Fall-Modulation (RFM) Aus dem PPG-Signal können zudem die Anstiegs- und Abfallzeiten von der aufeinanderfolgenden systolischen (Kontraktionsphase) und diastolischen (Erschlaffungsphase) lokalen Extrema abgeleitet werden, um daraus die Differenzen zwischen deren Anstiegs- und Abfallzeiten zu berechnen. Die berechneten Differenzen bilden dabei die neuen Signalwerte PPG_i für jeden Herzschlag i , die jeweiligen systolischen und diastolischen Maxima die $t(i)$ Werte. Ein Beispiel für die Bestimmung der RFM Werte ist in Abb. 3.14 dargestellt.

$$PPG_i = (PPG_{i, \max} - PPG_{i, \min}) - (PPG_{i, \max} - PPG_{i+1, \min}) \quad (3.13)$$

3.5.3 Postprocessing

Nach der Modulation wird eine Normalisierung der Atemsignale durchgeführt. Der Mittelwert der Signale μ wird vom Signal subtrahiert, um den restlichen Gleichanteil zu entfernen, welcher nach dem Bandpass im Frequenzspektrum verbleiben kann. Nach der Normalisierung wird ein Verfahren zur Entfernung von systematisch auftretenden Artefakten durchgeführt. Dieses ist im folgenden Abschnitt 3.5.4 beschrieben.

Das Signal wird anschließend mit einem Butterworth-Bandpass im Bereich der üblichen menschlichen Atemfrequenzen gefiltert. Dazu wurde eine untere Grenzfrequenz von 0,1 Hz und eine obere von 0,5 Hz gewählt. Der gewählte Bereich erlaubt es, einen möglichst großen Teil der Atemrate von Erwachsenen in Ruhe und bei moderater körperlicher Belastung zu messen, ohne Frequenzen im Bereich der Herzrate zu erfassen, was bei einer Erhöhung der oberen Grenzfrequenz auftreten kann.

3.5.4 Artefaktreduzierung

Über alle Modulationsarten hinweg wurde eine dominante Frequenz im unteren Bereich des bandpassgefilterten Frequenzspektrums gemessen, welche die Atemratenfrequenz überlagerte. Die dadurch falsch identifizierten Atemraten lagen meist zwischen 6 bpm und 10 bpm, in wenigen Fällen

wurde eine falsche Rate bis zu 12 bpm festgestellt. Der Frequenzbereich dieser wiederholt gemessenen Störfrequenzen liegt also im Intervall von 0,1 Hz bis 0,2 Hz.

Diese Art von Störungen ist auch in der Literatur zu finden. Einige Forscher nennen Bewegungsartefakte als Ursache für dieses Phänomen. Moraes *et al.* [Mor+18] und Lee *et al.* [Lee+07] schreiben, dass die von der Testperson ausgeführten Bewegungen oft Frequenzen um 0,1 Hz oder leicht höher aufweisen. Diese Theorie deckt sich auch mit den Forschungsergebnissen von Ram *et al.* [Ram+12], die eine umfassende Frequenzanalyse menschlicher Bewegungen durchgeführt haben. Untersucht wurden insbesondere normale und natürliche Körperbewegungen, zum Beispiel beim Sprechen oder aufgrund von Veränderungen der Gesichtszüge. Die stärkste Frequenzkomponente in ihrer Auswertung war eine Rate von etwa 0,1 Hz, aber auch minimal höhere Frequenzen waren in großem Umfang vorhanden. Als eine weitere Hypothese für den Ursprung dieser Interferenzen nannten Charlton u. a. [Cha+18] die Traube-Hering-Mayer-Wellen. Diese Wellen werden vom sympathischen Nervensystem erzeugt und ermöglichen eine bewusste Regulierung der Organaktivität [LUO92] und liegen beim Menschen konstant um einen Frequenzwert von etwa 0,1 Hz.

Die genaue Herkunft und Entstehungsweise dieser Frequenzen konnte nicht abschließend geklärt werden. Um sie bei der Atmungserkennung zu unterdrücken, wird eine Differenzierung des Signals nach der Normierung und vor der letzten Bandpassfilterung durchgeführt. Die Gradienten $PPG_{i, \text{grad}}$ des normierten Signals $PPG_{i, \text{norm.}}$ werden für jeden Signalwert i wie folgt berechnet:

$$PPG_{i, \text{grad}} = \frac{PPG_{i+1, \text{norm.}} - PPG_{i-1, \text{norm.}}}{2} \quad (3.14)$$

3.5.5 Bestimmung der Atemrate

Abschließend wird für jedes modulierte Signal eine Fast Fourier Transformation (FFT) zur Frequenzanalyse verwendet und die dominante Frequenz mit der größten Signalstärke ermittelt. Diese wird anschließend in die Atemrate in BPM umgerechnet. Um die Robustheit der Schätzung der Atemrate zu

erhöhen, wird eine Mittlung der Ergebnisse der verschiedenen Modulationssignale durchgeführt und mehrere Einzelverfahren kombiniert und zusammen ausgewertet. Da einzelne atemungsinduzierte Schwankungen je nach Untersuchungsperson unterschiedlich stark sein können, wird durch die Berechnung des Mittelwertes und/oder Medians der Einfluss von einzelnen Ausreißern reduziert.

3.6 Lebenderkennung

Gesichtserkennungstechnologie verbreitet sich zunehmend in unserer modernen Gesellschaft. Sowohl im öffentlichen Raum durch staatliche Sicherheitskräfte, als auch neuerdings als Zugangskontrolle für Mobilgeräte wie Smartphones, wie zum Beispiel Apples *Face ID* und für Online Sicherheitsabfragen durch Gesichtserkennungssysteme.

Aufgrund der einfachen "Bedienung" der Gesichtserkennungssysteme sind diese leider auch anfällig für Umgehungs- und Täuschungsversuche. Das Gesicht eines Menschen ist relativ einzigartig, wird aber im Gegensatz zu einem Passwort im öffentlichen Raum gezeigt und verbreitet. Nur geringer Aufwand und Kosten sind heutzutage notwendig, um eine Kopie des Gesichtes in Form eines Fotos zu erlangen, insbesondere seit dem Aufkommen des Internets und den sozialen Netzwerken. Daher sind Fotoausdrucke und Darstellungen von Gesichtern auf Monitoren die verbreitetsten Methoden zur Umgehung von Gesichtserkennungssystemen. Verschiedene Methoden und Anstrengungen, sowohl zur Durchführung von Angriffen auf Gesichtsdarstellungen, als auch deren Verhinderung wurden in der Literatur untersucht [RPR12; PS00; MHP11; KD14].

Mit dem Aufkommen neuer videobasierter Rekonstruktionsmethoden von dreidimensionalen Objekten aus Bildern [Rag+13] und schnellen 3D Fertigungsmethoden wie dem 3D Druck sind in den letzten Jahren neue Möglichkeiten der Umgehung von Gesichtserkennungssystemen, durch individuell angepasste und hergestellte Masken, kostengünstig verfügbar geworden. Zur Detektion dieser realistischen Masken bietet sich die kamerabasierte Vitalparameterschätzung an. Das Pulssignal kann zur Unterscheidung von durchblutetem Gewebe und einer künstlichen Maske

Systemkomponente	Präsentation			
	Bild	Bildschirm	Maske	Gesicht
Gesichtserkennung	✓	✓	✓	✓
3D Analyse	✗	✗	✓	✓
Vitalparameter	✗	✓	✗	✓

Tabelle 3.1: Detektionsmatrix der Täuschungsansätze der einzelnen Systemkomponenten und der möglichen Überwindung der Sicherheitskomponenten.

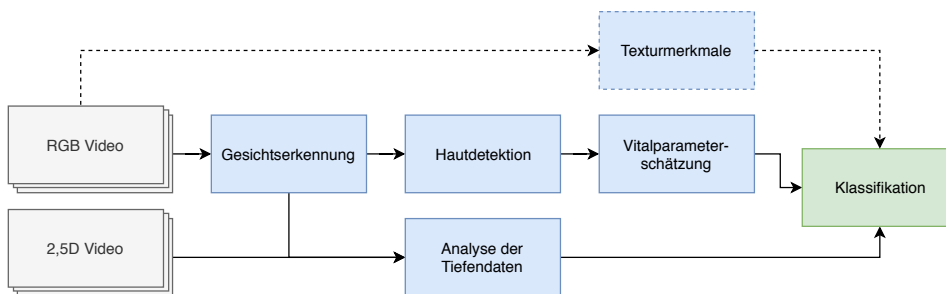


Abbildung 3.15: Schematische Darstellung des Verfahrensablaufes der Lebenderkennung.

genutzt werden, welche dieses Signal nicht generiert. Tabelle 3.1 zeigt die unterschiedlichen Täuschungsverfahren und eine Auflistung einzelnen Systemkomponenten. Durch die Kombination von Tiefendaten und Vitalparameterschätzung können menschliche Gesichter und versuchte Angriffe unterschieden und klassifiziert werden.

Die Methodik des vorgestellten Verfahrens zur Erkennung von Gesichtern ist in Abbildung 3.15 dargestellt. Ausgehend von einem RGB Video und korrespondierenden 2,5D Tiefendaten wird eine Klassifikation durchgeführt, die zwischen menschlichen Gesichtern, Masken und Bildern unterscheiden kann. Dazu werden die Tiefendaten analysiert, um eine Täuschung durch Bild oder Video zu erkennen und die RGB Daten verarbeitet, um im präsentierten Gesicht nach messbaren Vitalparametern zu suchen. In den folgenden Abschnitten werden die einzelnen Komponenten des Verfahrens näher betrachtet.

3.6.1 Gesichtsverifizierung

Zusammen mit den RGB Bildern, welche für die Gesichtserkennung verwendet werden, wurden 2,5D Tiefeninformationen aufgenommen. Diese dienen der Erkennung von gedruckten Bildern und Monitoren, wie Smartphones und Tablets, welche zur Überwindung der Gesichtserkennung des Systems genutzt werden. Dazu werden Tiefeninformationen des gefundenen Gesichtes analysiert und auf erwartbare Höhenunterschiede untersucht.

Die Tiefenanalyse wird durch den Vergleich der Informationen aus mehreren Regionen (ROI) im Gesicht realisiert. Dabei werden die Tiefendaten der Augenpartie, des Kinns und der Nase verwendet. Die Entfernungen der beiden Augenregionen, werden mit der Tiefeninformation der Nasenspitze verglichen, um Rotationen der vertikalen Achse zu erkennen. Die Kinnregion wird zusammen mit den Augenregionen genutzt, um Rotation in der horizontalen Achse zu detektieren. Wenn alle Rotationskriterien eingehalten werden, müssen für eine positive Klassifikation eines Gesichtes die Augen und Kinnregionen hinter der Nase liegen. Zur Verbesserung der Genauigkeit wird die Nase zudem relativ zu ihrem direkten Umfeld im Gesicht betrachtet und die Tiefe der Nase zu diesem bestimmt. Dabei muss dieser Abstand einen Schwellwert von mindestens 1,5 cm erreichen.

Für jeden Frame eines betrachteten Videos wird eine Klassifikation durchgeführt. Wenn mindestens 90% der Frames als Gesicht klassifiziert wurden, wird das Video als Gesicht eingeordnet.

3.6.2 Vitalparameterschätzung

Der verwendete Ansatz zur Herzratenschätzung kann in drei Teile unterteilt werden. Zunächst wird das PPG Signal aus den Bilddaten extrahiert. Anschließend wird es mittels eines Bandpasses gefiltert. Das Spektrum des so vorverarbeiteten Signales wird daraufhin analysiert und klassifiziert.

Zur Erzeugung des Signales mittels des Videomaterials wird in jedem Einzelbild das Gesicht mit dem *IntraFace Facial Feature Detection & Tracking (v1.1)* Algorithmus detektiert. Darauf folgt eine Erkennung von bestimmten Punkten im Gesicht mittels der „dlib C++ Library landmark detection“. Die

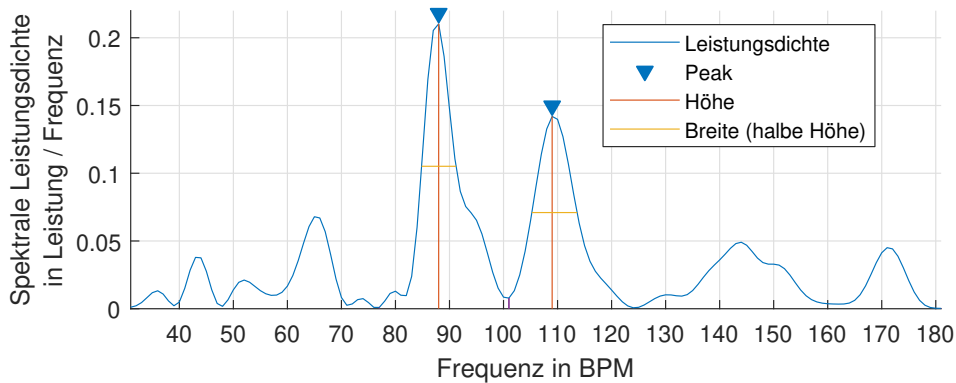


Abbildung 3.16: Beispiel des Spektrums eines PPG Signales und den abgeleiteten Peakmerkmalen.

generierten Punkte werden genutzt, um einen Bildausschnitt des Gesichtes für die Hautdetektion aus dem Bild zu extrahieren. Dieser Bildausschnitt wird um 20% der Gesamthöhe des Bildausschnittes nach oben und unten vergrößert, um die Stirn und den Hals der Probanden mitzuerfassen.

Für eine schnelle und zuverlässige Hautdetektion wird das in Abschnitt 3.1 beschriebene Hautdetektionsverfahren verwendet. Dabei wird eine Look-Up-Table (LUT) genutzt, die einem Pixel, in Abhängigkeit seiner RGB Farbwerte, eine Hautfarbenwahrscheinlichkeit zuordnet. Die Pixel in jedem Bild werden mit dieser Wahrscheinlichkeit gewichtet gemittelt, sodass für jede der drei Farben ein Vektor gebildet wird.

Aus den Farbvektoren wird durch den Einsatz des CHROM Ansatzes (siehe Abschnitt 2.2.3) das PPG-Signal generiert und für die weitere Analyse mittels einer Fast-Fourier Transformation (FFT) in seine Frequenzanteile zerlegt. Betrachtet werden dabei nur die Frequenzbereiche, welche die Spanne der menschlichen Herzraten abdecken und bei etwa 0,5-3Hz oder 30-180 Schlägen-pro-Minute (BPM) liegen. Das Signalspektrum wird daraufhin normiert und mögliche Signalspitzen isoliert.

Für die beiden höchsten Peaks im Spektrum werden verschiedene Merkmale bestimmt. Diese sind die relative Höhe, Breite und Energie der Peaks (siehe Abb. 3.16). Dabei wird die Breite als der Abstand der Flanken auf halber Höhe bestimmt. Die Energie eines Peaks ist der summierte Anteil

des Gesamtsignales, im als Breite des Peaks definierten Bereiches. Darüber hinaus werden der Mittelwert und Median aller Frequenzen im Spektrum bestimmt. Somit werden jedem Video mehrere spektrale Merkmale zugeordnet. Diese spektralen Merkmale werden für die Klassifikation des PPG Signales genutzt, um zwischen Signalen vom menschlichen Gesichtern und Signalen von Masken zu unterscheiden.

Die experimentelle Auswertung zu dem beschriebenen Verfahren ist in Abschnitt 4.9 beschrieben.

4 Experimentelle Ergebnisse

Das folgende Kapitel beschreibt die durchgeführten Experimente und Versuche für die in Kapitel 3 vorgestellten neu entwickelten Methoden, sowie umfangreiche Grundlagenuntersuchungen und mehrere Anwendungsfälle.

Die Kapitel 4.1 bis 4.3 beschreiben übergreifende Informationen, wie den Ablauf der Gesichtsdetektion, die Fehlerberechnung und die genutzten Datenbanken, welche in mehreren Kapiteln verwendet werden. In Kapitel 4.4 werden ausführlich verschiedene Formen der Videokompression und deren Einfluss auf die Genauigkeit der Vitalparameterschätzung untersucht. In den darauf folgenden Kapiteln werden der Einfluss der Region of Interest (ROI) (Kapitel 4.5), die Verwendung eines adaptiven Bandpasses und des IBI-Graphen-Korrekturverfahrens (Kapitel 4.6), die Verarbeitung der RGB Daten durch ein LSTM-Netz (Kapitel 4.7), sowie die Messung der Atemrate (Kapitel 4.8) beschrieben.

Weiterhin wurden ausführliche Grundlagenuntersuchungen zur Erkennung von lebendem Gewebe (Kapitel 4.9), der Messung der Herzrate im Nahinfrarotbereich (Kapitel 4.11) durchgeführt sowie die Verwendung der kamerabasierten Vitalparameterschätzung in einem MRT betrachtet (Kapitel 4.10).

4.1 Gesichtsdetektion

Für alle in diesem Kapitel vorgestellten Experimente wurde das gleiche Gesichts- und Landmarkendetektionsverfahren verwendet. In den verwendeten Videos wurden in jedem Bild sowohl 68 Gesichtspunkte, als auch eine *Bounding Box* um das Gesicht des Probanden definiert (siehe Abb. 4.1). Der

in *OpenCV 2.4* vorhandene *Haar-Cascade* Klassifikator wurde genutzt, um das Gesicht in dem Videobild zu finden.

Im nächsten Schritt wurde der *DLib facial landmark detector* verwendet, um die Pixelkoordinaten (u, v) von 68 Gesichtslandmarken zu detektieren. Beide Schritte wurden in C++ implementiert. Die Landmarken wurden über mehrere Frames stabilisiert. Aufgrund der in den verschiedenen Datenbanken zum Teil unterschiedlichen Bildwiederholraten wurde, anstatt einer festen Frameanzahl, eine Minimaldauer von 100 Millisekunden für die Stabilisation gewählt. Ein längerer Zeitraum würde das Rauschen der Landmarken weiter reduzieren, jedoch die Genauigkeit bei schnelleren Kopfbewegungen verringern. Dieser Wert wurde für das entsprechende Video in eine Frameanzahl umgerechnet und auf die nächste ganze ungerade Zahl aufgerundet. Für die Stabilisierung wurde der Mittelwert für die u - und v Koordinaten aus den Frames vor und nach dem jedem Bild berechnet. Konnten in einem Frame keine Landmarken detektiert werden, wurde dieses Frame nicht für die Stabilisierung verwendet. Falls für ein Bild keine stabilen Punkte bestimmt werden konnten, wurden die zuletzt berechneten stabilen Punkte des Videos verwendet.

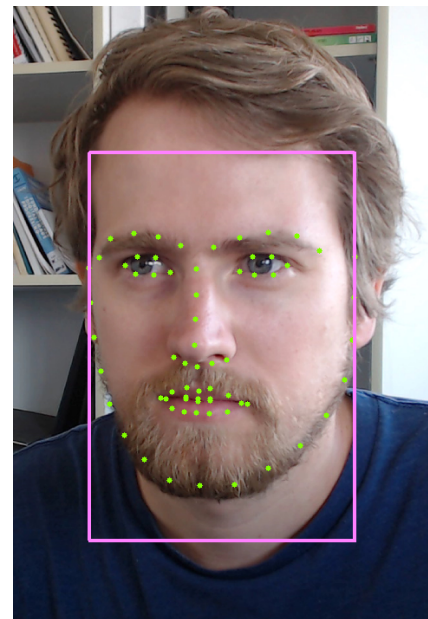


Abbildung 4.1: Beispiel der Landmarken und *Bounding Box*.

Anschließend wurde ein Bildausschnitt mithilfe der *Bounding Box* um die Gesichtslandmarken extrahiert. Dieser wurde oben und unten um die halbe Länge des Abstands zwischen den Ecken der Augen und der Nasenspitze vergrößert. Diese Referenzpunkte wurden aufgrund ihrer relativ fixen Position im Gesicht gewählt. Die Größe der *Bounding Box* kann ansonsten durch die Bewegung der Augenbrauen und des Mundes in kurzer Zeit stark variieren. Dieser Ansatz vergrößerte die ROI vertikal, um zusätzliche Haut

am Hals und an der Stirn abzudecken und ist gleichzeitig unabhängig von der Pixelauflösung des Gesichts in den Videodaten.

4.2 Fehlerberechnung

Für die Berechnung der Messgenauigkeit der verschiedenen Vitalparameter wurde eine einheitliche Methode zur Bestimmung der Grundwahrheiten und der Fehlerberechnung definiert. Diese sind in Kapitel 4.2.1 beschrieben. Neben den üblichen Fehlermaßen, wie dem arithmetischen Mittelwert des Fehlers und der Standardabweichung, wurden weitere Genauigkeitsmaße verwendet, welche aus den Zulassungsanforderungen für die Herzratenmessung von medizinischen EKG-Geräten abgeleitet wurden und in den Kapiteln 4.2.2 und 4.2.3 beschrieben sind.

4.2.1 Grundwahrheiten

Die Grundwahrheit der Herzfrequenz wurde mit Hilfe einer QRS-Komplex-Erkennungsmethode [Sch+14] aus den vorliegenden EKGs oder Kontakt-PPG Daten berechnet, gefolgt von einer manuellen Überprüfung auf nicht detektierte oder falsch positiv klassifizierte Herzschläge. Für jede Herzfrequenzschätzung wurde das gleiche Zeitfenster im Video (PPG) und der Grundwahrheit analysiert. Aus den im Zeitfenster der Grundwahrheit gefundenen Herzschläge werden die Inter-Beat-Intervalle (IBIs) gebildet. Aus deren Mittelwert wird die Grundwahrheitsherzfrequenz HR_{GT} (in BPM) berechnet. Der Fehler e für jede Herzratenschätzung wird als $e = HR_{GT} - HR_{PPG}$ berechnet.

4.2.2 IEC Genauigkeit (Herzrate)

Die von der International Electrotechnical Commission (IEC) in der Norm 60601-2-27 für medizinische EKG-Geräte [Int11a] beschriebene Fehlerdefinition wird in dieser Arbeit als Validierungsmaßstab für die Herzfrequenzschätzung genutzt. Unter Verwendung der obigen Norm gilt eine

Schätzung als gültig, wenn der absolute Fehler zwischen der geschätzten Herzfrequenz und der Grundwahrheit weniger als 10% der Grundwahrheit oder mindestens 5 BPM beträgt. Der Prozentsatz der Messungen einer Messreihe, die dieser IEC-Norm entsprechen, wird in dieser Arbeit als **IEC-Genauigkeit** (in %) bezeichnet.

4.2.3 DR und FDR Genauigkeit (Atemrate)

Um die Genauigkeit der Atemratenmessungen in Kapitel 4.8 zu bestimmen, wurden zwei weitere auf der IEC-Genauigkeit basierende Fehlermaße definiert. Zum einen die Detection Rate (DR) (in%), bei der eine Atemraten-schätzung als gültig gilt, wenn der Fehler zwischen dem gemessenen Wert und der Grundwahrheit kleiner oder gleich 2 Atemzüge pro Minute (breaths per minute) (BrPM) ist. Dieses wurde in ähnlicher Weise von Charlton u. a. [Cha+18] als statistisches Fehlermaß für die Atemratenerkennung vorgeschlagen. Die DR beschreibt den prozentualen Anteil der korrekt erkannten Messungen n_k von der Gesamtanzahl der untersuchten Zeitfenster n :

$$DR [\%] = \frac{n_k}{n} \quad (4.1)$$

Um die Auswirkung der Artefaktreduktion (siehe Kapitel 3.5.4) besser zu beurteilen, wurde ein zusätzliches Fehlermaß auf Grundlage der DR definiert, um den Prozentsatz der falsch erkannten Fenster im niederfrequenten Bereich zu bestimmen. Die False Detection Rate (FDR) gibt das Verhältnis der Messungen n_f an, welche einen Fehler von $> \pm 2$ BrPM und eine dominante niederfrequente Komponente im Bereich zwischen 0,1 Hz und 0,2 Hz aufweisen, relativ zur Gesamtzahl der Messungen n :

$$FD [\%] = \frac{n_f [0.1-0.2 \text{ Hz}]}{n} \quad (4.2)$$

4.3 Datenbanken

Im folgenden Abschnitt werden die häufiger verwendeten Datenbanken vorgestellt. Die Tabelle 4.1 gibt einen Überblick über verschiedene Eigenschaften der Datensätze. Zwei der Datenbanken, die *HKBU* (siehe Kapitel 4.9) und *MRT DB* (siehe Kapitel 4.10), werden nur einmalig in den entsprechenden Kapiteln verwendet und dort beschrieben, sind der Vollständigkeit halber in der Tabelle mit aufgeführt.

Für die kamerabasierte Vitalparameterschätzung kann nur auf wenige Datensätze zurückgegriffen werden, da eine bestimmte Kombination von Grundwahrheiten vorliegen muss. Insbesondere müssen Video- und Vitalparameter synchron aufgezeichnet werden, was in anderen Forschungsfeldern selten notwendig ist. So wurde auf Datenbanken der Schmerzerkennung (*BioVid*) oder multimodalen Datenbanken für die allgemeine Zustandserkennung (*BP4D+*, *MMSE-HR*) des Menschen zurückgegriffen.

Viele in anderen Forschungsfeldern häufig verwendeten Datenbanken, für zum Beispiel die Emotionserkennung, enthalten häufig keine Vitalparameter-Grundwahrheiten und Datenbanken für die Vitalparameter-Analyse (EKG, EEG, etc.) fehlen die Videodaten. Ein weiterer wichtiger einschränkender Faktor sind die verwendeten Kompressionsverfahren zur Reduzierung des Speicherplatzes der Datenbanken. So haben einige Datenbanken wie *AMI-GOS* [Cor+18], *COFACE* [HAM17], *MAHNOB-HCI* [Sol+12] oder *DEAP*

Tabelle 4.1: Überblick der verwendeten Datenbanken mit Angaben über Geschlecht(M/F), Gesamtlänge der Videos (im Format hh:mm), Messmethoden der Herzrate (HR) und Atemrate (AR), Auflösung in Pixel, Bildrate in FPS und verwendete Encodiermethode.

Datenbank	Prob.	M/F	Alter	Videos	Länge	HR	AR	Bildmaße	FPS	Codec
AtemDB	12	10/2	23-36	48	02:31	Finger	Gurt	1388 x 1038	25	x264, CRF 0
BioVid	87	44/43	18-65	87	36:50	EKG	-	1388 x 1038	25	x264, CRF 17
BioVidEmo (Teil D)	87	44/43	18-65	87	17:30	EKG	-	1388 x 1038	25	x264, CRF 17
BioVidEmo (Teil E)	87	44/43	18-65	87	10:30	EKG	-	1388 x 1038	25	x264, CRF 17
BP4D+	140	58/82	18-66	1400	17:17	Finger	Gurt	1040 x 1392	25	JPG
MMSE-HR	40	17/23	18-66	102	01:11	Finger	-	1040 x 1392	25	JPG
PURE	10	8/2	-	60	01:08	Finger	-	640 x 480	30	PNG
HKBU-MARs V1+	12	-	-	170	00:28	Finger	-	1280 x 720	30	MJPEG
MRT DB	8	4/4	21-31	10	00:50	Finger	-	1000 x 1000	25	x264, CRF 0

[Koe+11] starke Kompressionsartefakte, welche die Verwendung für die videobasierte Vitalparameterschätzung stark einschränken.

Diese Herausforderungen lagen, zum einen an dem fehlenden Kenntnisstand über den Einfluss der Kompression auf die Vitalparameterschätzung und zum anderen an den technischen Herausforderungen der Videoaufnahme sowie der Wahl des Codecs und deren Parameter. Es wurde, zum Beispiel, die *COHFACE* Datenbank mit Hilfe der Webcam-Software des Herstellers aufgezeichnet. Diese encodierte die Videos in der Standardinstellung im *fragmented MP4* (FMP4) Format, welches für Livestreaming optimiert ist und starke Störsignale im Frequenzbereich um 1,2Hz oder 72BPM erzeugte, was im Bereich der menschlichen Herzrate liegen.

4.3.1 BioVid

Die **BioVid Heat Pain Database** wurde von Werner, Al-Hamadi, Niese, Walter, Gruss und Traue [Wer+13] vorgestellt. Die 87 Teilnehmer der Datenbank wurden wiederholt schmerzhaften Hitzereizen ausgesetzt. Die Probanden bestanden aus drei Altersgruppen (18-35, 36-50 und 51-65 Jahre), wobei jede der Gruppen aus 15 Männern und 15 Frauen bestand.



Abbildung 4.2: Beispielbild der **BioVid** Datenbank.

Die Experimente wurden mit mehreren Videokameras und physiologischen Sensoren aufgezeichnet. Es wurden drei AVTPike F145C Farbkameras mit einer Auflösung von 1388x1038 und 25 Hz verwendet. In den in dieser Arbeit beschriebenen Experimenten wurde nur die Frontal vor dem Probanden positionierte Kamera verwendet. Zusätzlich zu den Videoaufnahmen wurden Hautleitwert (SCL), Elektrokardiogramm (EKG), Elektromyogramm (EMG) von drei schmerzrelevanten Muskeln und das Elektroenzephalogramm (EEG) aufgenommen.

Den Teilnehmern wurde ausdrücklich erlaubt, sich frei zu bewegen. Daher enthält die Datenbank eine große Menge an Kopfbewegungen und Gesichts-

ausdrücke. Für die Validierung der verschiedenen Experimente wurden die Videodaten des Teil C der Datenbank verwendet. Diese Version der Datenbank wurde genutzt, da durch die langen ungeschnittenen Videodaten (ca. 25 Min. je Proband) die Möglichkeit zur Berechnung von kontinuierlichen Herzfrequenz-Schätzungen über einen größeren Zeitraum möglich sind.

Für die Validierung der LSTM basierten Pulserkennung in Kapitel 4.7 wurden zudem die Teile D und E der Datenbank verwendet, welche in dieser Arbeit als **BioVidEmo** bezeichnet werden. Diese beinhalten den BioVid Teil D mit 450 Videos mit spontanen Emotionen und einer Gesamtlänge von 17,5 Stunden und in Teil E 630 Videos mit gespielten Emotionen mit einer Gesamtlänge von 10,5 Stunden.

4.3.2 BP4D+

Die **BP4D+** ist eine Erweiterung des BP4D Datensatzes und enthält verschiedene Modalitäten, darunter synchronisierte 3D-, 2D-, thermische, physiologische Datensequenzen (z. B. Herzfrequenz, Blutdruck, Hautleitwert (EDA) und Atemfrequenz) und weitere Metadaten wie Gesichtsmerkmale und Werte für das *Facial Action Coding System* [Ell87]. Abbildung 4.3 zeigt ein Beispielbild aus der Datenbank. Die Datenbank stand erst spät für die Forschung in dieser Arbeit zu Verfügung und konnte daher nur in einigen Experimenten verwendet werden.



Abbildung 4.3: Beispielbild der **BP4D+** und **MMSE-HR** Datenbank.

Es enthält Videos von 140 Probanden und je 10 Aufgaben (Emotionen) für jeden Probanden, darunter 58 Männer und 82 Frauen, mit einem Alter zwischen 18 und 66 Jahren. Es ist ein sehr diverser Datensatz mit Probanden vieler verschiedener ethnischer Herkünfte.

Die Videodaten wurden mit einem *Di3D Dynamic Imaging System* in Farbe und einer Auflösung von 1040x1392 bei 25 FPS aufgenommen. Zwei Lampen wurden, zusätzlich zum normalen Laborlicht verwendet, um die Szene zu beleuchten. Jedes Videobild wurde als separate *jpg*-Bilddatei mit einer *Qualitäts*-Einstellung von 100% und 2x2 color subsampling gespeichert. Die physiologischen Daten wurden mit einem *Biopac MP150*-System erfasst, welches den Blutdruck und die Herzfrequenz mit 1000 Hz aufzeichnete. Andere physiologische Signale wurden aufgezeichnet, sind aber nicht Teil der MMSE-HR. Die Probanden wurden angewiesen, in einem Abstand von ca. 1,3 Meter zu den Kameras zu sitzen.

4.3.3 MMSE HR

Der **MMSE-HR** ist ein Teil des **BP4D+**-Datensatzes und wurde von Zhang et al. [Zha+16] in 2016 vorgestellt. Der Datensatz wurde speziell für das Testen und die Validierung von Algorithmen zur videobasierten Herzfrequenzschätzung erstellt.

Er enthält 102 Videos unterschiedlicher Länge von 40 verschiedenen Probanden (17 männlich, 23 weiblich; 18-66 Jahre alt) mit unterschiedlichen ethnischen Hintergründen. Während eines Interviews wurden die Teilnehmer verschiedenen Stimuli ausgesetzt, um emotionale Reaktionen hervorzurufen. Der Messaufbau und der Ablauf der Aufnahmen sind analog zu denen des BP4D+ Datensatzes.

4.3.4 PURE

Der PURE-Datensatz wurde von Stricker, Müller und Gross [SMG14] im Jahr 2014 als Benchmark-Datenbank vorgestellt, um "die verschiedene Gesichtsegmentierungsansätze zu vergleichen und die Artefakte, die durch Kopfbewegungen entstehen, genauer zu untersuchen". Abbildung 4.4 zeigt ein Beispielbild aus der Datenbank.

Die Videos haben eine Länge von jeweils 60s und wurden mit einer Farbkamera mit einer Auflösung von 640x400 Pixel und 30 FPS aufgenommen.

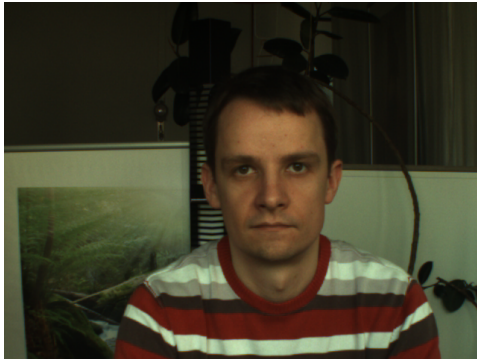


Abbildung 4.4: Beispielbild der **PURE** Datenbank.



Abbildung 4.5: Beispielbild der **AtemDB** Datenbank.

Jedes Bild wurde als separate *png*-Bilddatei gespeichert. Die physiologischen Signale wurden mit einem *Pulox CMS50E* Fingerpulsoximeter aufgezeichnet. Herzfrequenz und SpO_2 -Messwerte wurden mit einer Abtastrate von 60 Hz aufgezeichnet. Der Aufbau wurde mit Tageslicht durch ein großes Fenster frontal zum Gesicht beleuchtet. Die Beleuchtungsbedingungen änderten sich im Laufe der Zeit leicht, je nach Wolkenbedeckung .

Der PURE-Datensatz enthält 10 Probanden (8 männlich, 2 weiblich), die jeweils sechs kontrollierte Kopfbewegungen ausführen. Diese beinhalten:

- *Steady*: ohne Bewegung
- *Talking*: Reden ohne Bewegung des Kopfes
- *Slow Translation*: kontrollierte Bewegung parallel zur Kamera
- *Fast Translation*: wie *Slow Translation*, mit doppelter Geschwindigkeit
- *Small Rotation*: zufällige Rotation des Kopfes um bis zu 20deg
- *Medium Rotation*: zufällige Rotation des Kopfes um bis zu 30deg

4.3.5 AtemDB

Für die Erforschung und Validierung der videobasierten Atemratenalgorithmen standen wenige Datenbanken mit Atemgrundwahrheiten zur Verfügung. Es waren nur zwei Datenbanken in der Literatur bekannt, welche synchronisierte Atem- und Videosignale enthielten. Die **BPD4+** und die

OBF [Li+18] Datenbank, welche aus datenschutzrechtlichen Gründen zum Zeitpunkt der Experimente nicht öffentlich zur Verfügung stand.

Daher wurde eine eigene Datenbank (**AtemDB**) für die Erforschung und Validierung der Atemratenalgorithmen erstellt. Abbildung 4.5 zeigt ein Beispielbild aus der Datenbank. Es wurden robuste, unkomprimierte Daten generiert, um diese als Benchmark für die Atemratenerkennung zu verwenden. Sie enthält Aufnahmen von 12 Erwachsenen im Alter zwischen 23 und 36 Jahren, 10 Männern und 2 Frauen. Pro Person wurden vier Videos mittels einer RGB-Kamera (Modell Pike F-145) mit einer Auflösung von 1388x1038 Pixeln aus einer Entfernung von etwa 1,5 Metern aufgenommen. Jede Sequenz dauerte 3 Minuten und wurde mit einer Bildrate von 25 Bildern pro Sekunde aufgenommen.

Für jeden Probanden wurden vier verschiedene Atemszenarien aufgezeichnet. Im ersten Szenario wird die natürliche Atmung des Probanden ohne Auflagen aufgezeichnet. Bei den nächsten drei Messungen sollten die Probanden versuchen, einem vorgegebenen Atemmuster zu folgen, das auf einem Monitor vor ihnen angezeigt wurde. Die Atemfrequenzen für die einzelnen Szenarien waren 10, 15 und 20 BrPM. Über alle Messungen hinweg wurden die Teilnehmer gebeten, sich wenig zu bewegen, um dadurch induzierte Störungen zu minimieren. Die Referenzsignale wurden über einen Brustgurt (Modell NB-RSP1A) mit einer Abtastrate von 512 Hz unter Verwendung eines Triggersignals zur Synchronisation mit der Kamera aufgezeichnet.

4.4 Videoeigenschaften

Um den Einfluss verschiedener Videoeigenschaften auf das PPG-Signal zu bestimmen, wurde eine umfassende Analyse verschiedener Videokodiereinstellungen durchgeführt. Dabei wurden der Constant Rate Factor (CRF), die Farbunterabtastung, Bildwiederholrate und die Auflösung variiert und der Einfluss auf die Herzratenschätzung untersucht. Teile der Ergebnisse dieses Kapitels wurden vorab in [RWA19] publiziert.

Videocodierung Es wurden zwei Codecs ausgewählt. Erstens, der *H.264* oder *Advanced Video Coding (AVC)* Standard, der heute weit verbreitet in Video-Streaming (z.B. *YouTube*, *iTunes Store*), HDTV-Sendungen oder Blu-rays eingesetzt wird. Zweitens, den neueren, fortschrittlicheren *H.265* oder *High Efficiency Video Coding* Standard (HEVC). Dieser Codec bietet „eine Bitratensparnis von ca. 50 % bei gleichwertiger wahrgenommener visueller Qualität im Verhältnis zur Leistung früherer Standards [...]„[Sul+12, p. 1667, Übersetzung des Autors]. Weiterführende Informationen über die verwendeten Videocodierungsverfahren sind in Kapitel 2.2.1 zu finden.

Die Auswirkungen verschiedener Videokodierungsparameter auf die Herzfrequenz Schätzung wurden auf zwei öffentlich verfügbaren Datensätzen mit insgesamt 161 Videos von 50 Probanden systematisch ausgewertet. Variiert wurden der Constant Rate Factor (**CRF**), die Farbunterabtastung (**YUV420**, **YUV444**), Bildwiederholrate und die Reduzierung der Auflösung mit verschiedenen Skalierungsalgorithmen (**Bilinear**, **Area**, **Neighbour**). Ein vereinfachter Ablauf der Auswertung der Videokompression ist in Abb. 4.6 beispielhaft dargestellt. In Kombination mit den zwei genannten Videocodecs wurden insgesamt 13.084 Videos mit einer Gesamtgröße von 1,2 TB generiert und analysiert.

Die genannten Parameter sind in Kapitel 2.2.1 beschrieben. Vier verschiedene Methoden zur Extraktion von PPG-Signalen (**normGrün**, **CHROM**, **aGRD**, **IFFT**) und zwei ROIs (**faceMid**, gewichtete Hautdetektion **skin**) wurden verwendet, um den Einfluss der Kodierungsparameter auf die Herzfrequenzschätzung zu beurteilen. Die verwendeten ROIs sind in Kapiteln 2.2.2 & 4.1 und die Signalextraktionsverfahren in Kapitel 2.2.3 beschrieben.

Alle Videos für diese Arbeit wurden mit FFmpeg [Bel] kodiert. Die in dieser Arbeit verglichenen Kodierungsmethoden beschränken sich auf einige wenige wichtige Parameter, die einen großen Einfluss auf die Qualität der Videodaten und dadurch auf die Genauigkeit der Herzfrequenzschätzung haben.

Für die Generierung der Videodaten wurden die in FFmpeg enthaltenen *x264* und *x265* Implementierungen der Codecstandards unter Verwendung der Voreinstellung *ultrafast* encodiert. Die nicht genannten Kodierungsparametern wurden auf den voreingestellten Werten gelassen. Die Videos

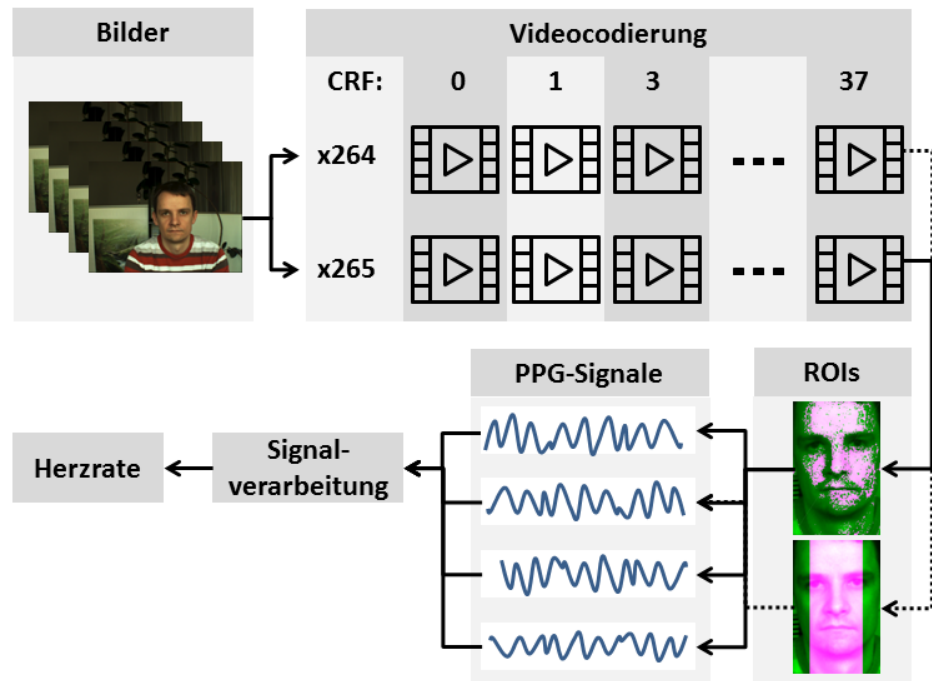


Abbildung 4.6: Ablauf der Auswertung der Videokompression für verschiedene CRF mit den Codecs *x264* und *x265* für zwei ROI und vier PPG-Signalverarbeitungsmethoden. Dabei wird für jede Codec/CRF/PPG/ROI Kombination eine Herzrate bestimmt.

wurden mit *OpenCV* (Version 2.4.2) und der auf *FFMPEG* basierenden *VideoCapture*-Klasse in C++ dekodiert und verarbeitet.

Bestimmung der Herzrate Für jedes Video wird jeweils ein PPG-Signal für alle acht Kombinationen aus den zwei ROIs und den vier Signalextraktionsverfahren bestimmt. Dieses wird zunächst mit einem adaptiven Bandpass gefiltert (siehe Kapitel 3.2). Im Initialisierungsfall wird ein weiterer (30 - 240 BPM) Bandpass auf einem 30s langen Signalfenster verwendet. In den folgenden Zeitschritten (1Hz) wird ein kürzeres Signalfenster (10s) mit einem kleineren Passband (± 15 BPM) gefiltert, welches um den zuletzt geschätzten Herzfrequenzwert zentriert ist, um das Signal-Rausch-Verhältnis zu erhöhen und eine schnelle Reaktion auf eine sich ändernde Herzfrequenz

zu ermöglichen.

Die Maxima des gefilterten Signales werden dann isoliert und stellen die mögliche Pulspeaks dar. Für die Bestimmung der Pulspeaks wurde der graphenbasierten Algorithmus zur Herzfrequenzschätzung [RWA16] (Siehe Kapitel 3.3) verwendet. Aus dem Mittelwert der Inter-Beat-Intervalle (IBI) der resultierenden Pulspeaks wurde die geschätzte Herzrate HR_{PPG} (in BPM) berechnet. Die Generierung und Auswertung der Grundwahrheiten sind in Kapitel 4.2 beschrieben. Neben dem arithmetischen Mittel des Fehlers und der Standardabweichung wird die Güte der Herzratenschätzung für die einzelnen Versuchsreihen mithilfe der IEC Genauigkeit angegeben (siehe Kapitel 4.2.2).

Datenbanken Wir verwendeten zwei verschiedene Datensätze, um den Einfluss der gewählten Kodierungstechniken zu testen, die **MMSE-HR** und die **PURE** Datenbanken. Beide Datensätze sind öffentlich zugänglich, enthalten Videodaten und dazu synchron aufgezeichnete physiologische Daten. Beide Datensätze sind als separate Bilddateien abgespeichert und haben daher keine Interframe-Kompression, sodass die zu untersuchenden Videokodierungsparameter und Verarbeitungsschritte kontrolliert werden können.

Der **MMSE-HR**-Datensatz hat eine größere Bildauflösung (>1 Megapixel), wodurch der Einfluss von verschiedener Downsampling-Algorithmen getestet werden kann. Sie ist im *JPEG* Format mit 2x2-Pixel color-subsampling und verlustbehafteter Komprimierung gespeichert. Der **PURE**-Datensatz besteht aus separaten *PNG*-Dateien, in denen alle Farbinformationen verlustfrei gespeichert wurden (YUV444), jedoch hat diese Datenbank eine niedrigere Bildauflösung (640x400).

Für die Untersuchung des Einflusses der Bildwiederholrate wurde zusätzlich die **BioVid** Datenbank verwendet. Jedoch liegt die **BioVid** nur verlustbehaftet encodiert vor und kann daher nicht für die anderen Untersuchungen des Kompressionsfehlers verwendet werden.

4.4.1 Allgemeiner Kompressionsfehler

Um die Ergebnisse des Videocodierungsprozesses zu überprüfen, wurden die Farbinformationen aus den kodierten und dekodierten Videos mit den Originalbildern verglichen. Dieser Vergleich wurde für verschiedene CRF-Werte durchgeführt. Gemäß der FFMPEG-Dokumentation [FFM] sollten Videos mit einem CRF = 0 verlustfrei Farbinformationen kodieren. Da die **MMSE-HR** in einem bereits verlustbehafteten Format vorlag, können in dieser Arbeit nur die Kompressionsunterschiede zur gespeicherten *JPEG* Bildqualität verglichen werden und nicht zu den ursprünglichen Daten.

Abb. 4.7 zeigt den mittleren quadratischen Fehler (MSE) der Pixelwert-RGB-Differenzen zwischen den enkodierten Videos und den vorliegenden Bildern aller Frames der ersten 10s für alle Videos in den PURE- und MMSE-Datensätzen. Dabei wurden für die einzelnen RGB Kanäle in jedem Pixel jeweils ein Fehlerwert bestimmt. Beide Encoder (x264 und x265) wurden mit beiden Datensätzen verwendet. Auf der PURE-Datenbank wurde zusätzlich der Einfluss der Farbunterabtastung (**YUV420**, **YUV444**) verglichen.

Die hohen Fehlerwerte für das YUV420-Format unter Verwendung des PURE-Datensatzes können auf die Farbunterabtastung zwischen den *png*-Bildern und den Videos zurückgeführt werden, welche bei der Verwendung der Vollfarbinformationen im YUV444-Format deutlich niedriger sind. Die **MMSE-HR**-Datenbank ist in ihrem ursprünglichen JPG-Format bereits *color-sampled*, sodass nur das YUV420-Format verwendet wurde.

Beide Codecs zeigen ansteigende Fehler bei höheren CRF-Werten, mit geringeren Fehlern im PURE-Datensatz als bei der MMSE, wenn für beide das in den Bildern vorliegende Pixelfarbformat verwendet wird. Der x264-Codec weist bei allen CRF-Werten einen geringeren Fehler als x265 auf. Kleine Farbänderungen liegen auch bei CRF= 0 vor, was im Widerspruch zur angegebenen Verlustlosigkeit der Videocodierung steht. Dies könnte durch Quantisierungs- und Rundungsfehler bei der Konvertierung von RGB in YUV und zurück verursacht werden.

Die Fehlerraten unter Verwendung von x264 mit YUV420 zeigen ähnliche „Stufen“ in beiden Datensätzen, deutlicher zu sehen im PURE Datensatz. Auf dem PURE-Datensatz bei einem CRF= 25 nimmt der Fehler bei einem

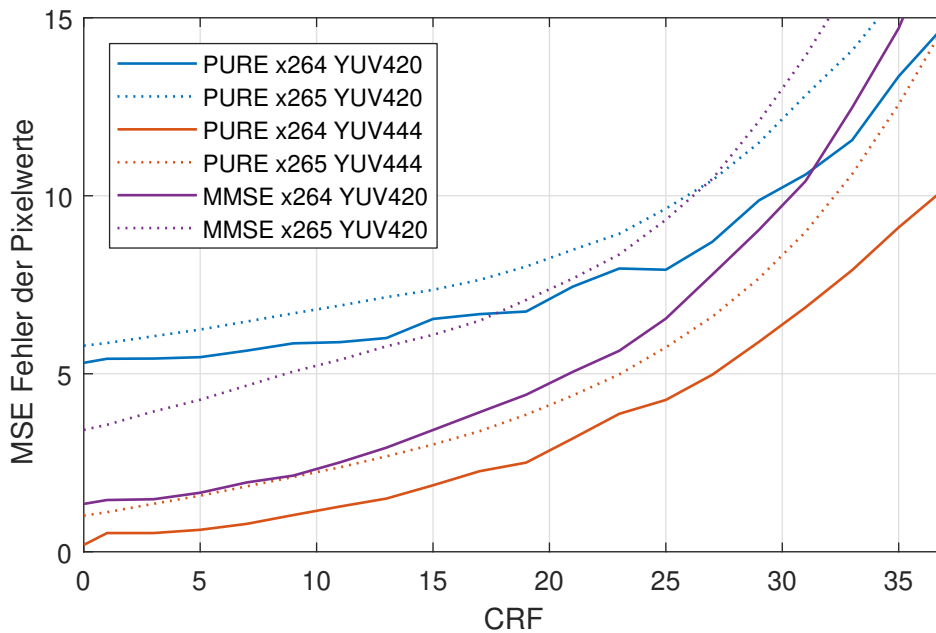


Abbildung 4.7: Mittlerer quadratischer Fehler der Pixel-RGB-Werte der ersten 10 Sekunden aller Videos im Vergleich zu den Originalbildern für verschiedene CRF-Werte.

höheren CRF-Wert einmalig ab. Die möglichen Ursachen dafür werden in Kap 5.1.1 diskutiert.

Um Kodierungs- oder Dekodierungsfehler bei der Berechnung oder beim Laden der Bilder oder Videos auszuschließen, wurde ein Video, unter Verwendung des verlustfreien *HuffYUV*-Codecs, kodiert und mit den Originalbildern verglichen, was zu einem MSE-Fehler von 0 führte.

4.4.2 Constant Rate Factor (CRF)

Es wurden verschiedene CRF-Werte zwischen 0 – 37 getestet, um ein breites Spektrum möglicher Kompressionsstufen zu berücksichtigen. Die Berechnungen wurden für beide Datensätze mit den Codecs *x264* und *x265* durchgeführt. Alle vier verschiedenen Signalextraktionsmethoden (siehe

Abschnitt 2.2.3) wurden verwendet, und die mittlere *IEC-Genauigkeit* wurde für jeden CRF-Wert berechnet.

Abb. 4.8 und Abb. 4.9 zeigen die *IEC-Genauigkeit* und Größe des Datensatzes (im Vergleich zu unkomprimiertem Video) in Abhängigkeit des CRF-Wertes für die Codecs *x264* und *x265*. Für beide Codecs werden die höchsten Genauigkeiten bei der geringsten Komprimierung mit $CRF=0$ erreicht. Der Vergleich der Ergebnisse bei gleichen CRF-Werten zeigt, wie erwartet, eine deutlich höhere Kompressionsrate für *x265*. Damit einhergehend hat der *x265*-Codec auch eine schneller abnehmende *IEC-Genauigkeit* bei Erhöhung des CRF-Wertes und eine auch geringere Maximalgenauigkeit als *x264*. Der *x264*-Codec zeigt in beiden Datensätzen eine Absenkung der *IEC Genauigkeit* bei etwa 96 – 99% der reduzierten Dateigröße und einem CRF-Wert von 15 – 21. Die Genauigkeit fällt schnell ab, um dann wieder im Anschluss auf ein höheres Niveau als vor dem Einbruch zu steigen. Der *x264*-Codec erreichte insgesamt die besseren *IEC-Genauigkeiten* auf Kosten der größeren Dateien. Die Mittelwerte und Standardabweichungen des Fehlers sind in Tab. 4.2 am Ende des Kapitels dargestellt.

4.4.3 Farbunterabtastung

Für die Untersuchung des Einflusses der Farbunterabtastung auf die Genauigkeit der Herzfrequenzschätzung wurde nur der *PURE*-Datensatz verwendet, da die *JPEG*-Bilder des *MMSE*-Datensatzes bereits farblich unterabgetastet vorlagen. Die Videos wurden in das Standardformat *YUV420* und das *YUV444* kodiert (siehe Abschnitt 2.2.1). Die Abbildungen 4.10 und 4.11 zeigen die mittlere *IEC-Genauigkeit* in Abhängigkeit von der gespeicherten Dateigröße.

Beide Codecs, *x264* und *x265*, wurden unter Verwendung des *YUV420*- und des *YUV444*-Pixelformats getestet. Bei Verwendung von *x265* übertraf das *YUV420*-Pixelformat in der Regel das *YUV444*-Format, sowohl bei gleichem CRF-Wert als auch gleicher Dateigröße. Beim *x264* Codec zeigten beide Pixelformate bis zu einem CRF-Wert von 17 ähnliche Ergebnisse, wobei die Genauigkeit bei *YUV444* stark zwischen den einzelnen CRF Varianten schwankte. Beim *x265* Codec zeigten die Verläufe eine deutliche Ähnlichkeit

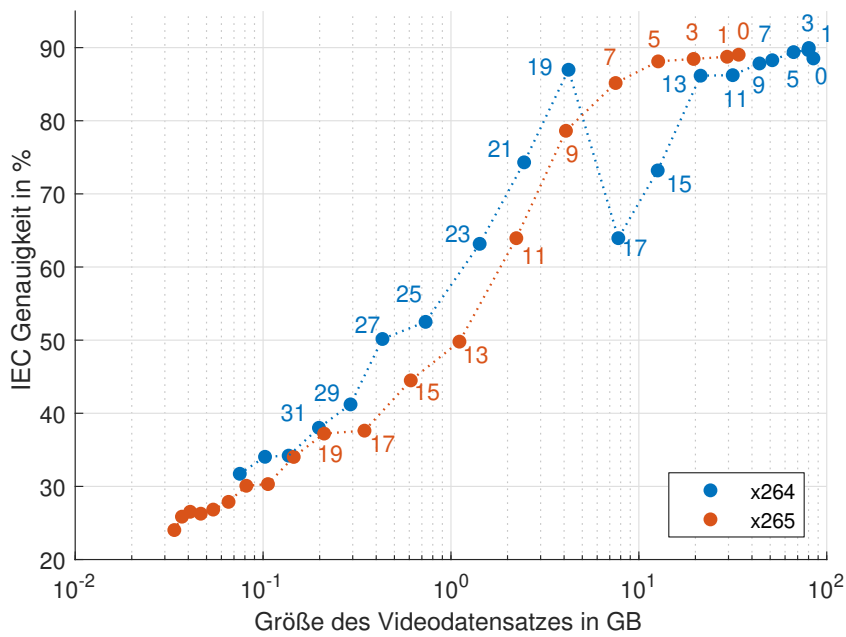


Abbildung 4.8: Mittlere IEC-Genauigkeit für verschiedene CRF-Werte (Zahlen) und die Datensatzgröße für die x264 und x265 Codecs (YUV420) auf dem MMSE-Datensatz.

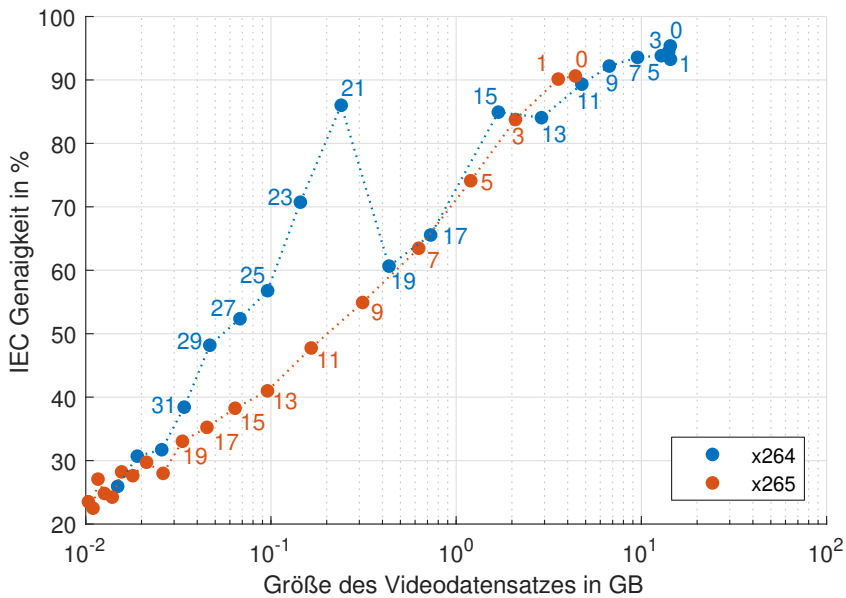


Abbildung 4.9: Mittlere IEC-Genauigkeit für verschiedene CRF-Werte (Zahlen) und die Datensatzgröße für die x264 und x265 Codecs (YUV420) auf dem PURE-Datensatz.

für beide Pixelformate. Während die Dateigrößen bei Verwendung des x265 Codec nur geringe Unterschiede zeigten, waren die mit dem x264 encodierten YUV₄₄₄-Dateien etwa doppelt so groß wie das YUV₄₂₀-Format. Dies entspricht der Verdoppelung des mittleren Pixelfarbformats von 12 auf 24 Bit und deutet auf unterschiedliche Farbformatkompressionsmethoden bei den beiden Codecs hin. Die Mittelwerte und Standardabweichungen des Fehlers sind in der Tabelle 4.3 am Ende des Kapitels dargestellt.

4.4.4 Vergleich der Region of Interest (ROI)

Zwei Regions of Interest Verfahren wurden verglichen, um die Auswirkung der Bildkompression auf die ROIs zu testen. Die hautbasierte Methode basiert auf einer Farb-Lookup-Tabelle und ist daher möglicherweise zusätzlichen Herausforderungen durch hohe Kompressionsartefakte ausgesetzt.

Die Abbildungen 4.12 und 4.13 zeigen die mittlere IEC-Genauigkeit über alle vier Extraktionsmethoden (siehe Abschnitt 2.2.3) für verschiedene CRF-Werte. Der Mittelwert und die Standardabweichungen des Fehlers werden in Tab 4.2, am Ende des Kapitels, dargestellt. Qualitativ verändern sich die Ergebnisse beider ROIs in vergleichbarer Weise mit dem CRF-Wert, jedoch zeigt die *Skin*-ROI in fast allen Fällen (außer CRF = 37) bessere Ergebnisse als die der *FaceMid*-ROI. Daher wird in der Analyse nur die *Skin*-ROI weiter betrachtet.

Weiterhin wurden die Fehlerunterschiede zwischen den ROIs *Skin* und *FaceMid* für die einzelnen Probanden untersucht, um mögliche systematische Unterschiede auszuschließen. Die *Skin*-ROI hat bei allen Probanden des PURE-Datensatzes bessere Ergebnisse. Im MMSE-Datensatz hatten nur zwei Testpersonen signifikant schlechtere Ergebnisse (> 5% mittlerer IEC) bei Verwendung der Haut-ROI. Bei der einen Person handelte es sich um eine kaukasische Frau (ID: Foo8), bei der anderen um eine afroamerikanische Frau (ID: Foo9). Im Vergleich zu den anderen Probanden in der Datenbank konnte keine physiognomische Ursache wie Hautfarbe, Brille oder andere Unterschiede zu den restlichen Probanden gefunden werden, welche die Abweichungen erklärten. Beide Frauen machten im Vergleich zu anderen Versuchs-

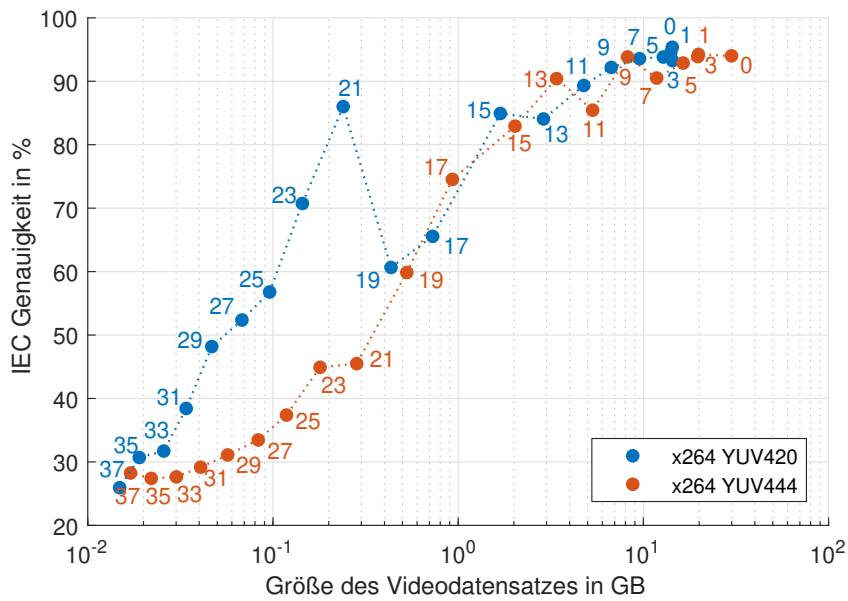


Abbildung 4.10: Mittlere IEC-Genauigkeit für verschiedene CRF-Werte (Zahlen) und die Datensatzgröße für **x264** und verschiedene **Farbformate** auf dem **PURE**-Datensatz.

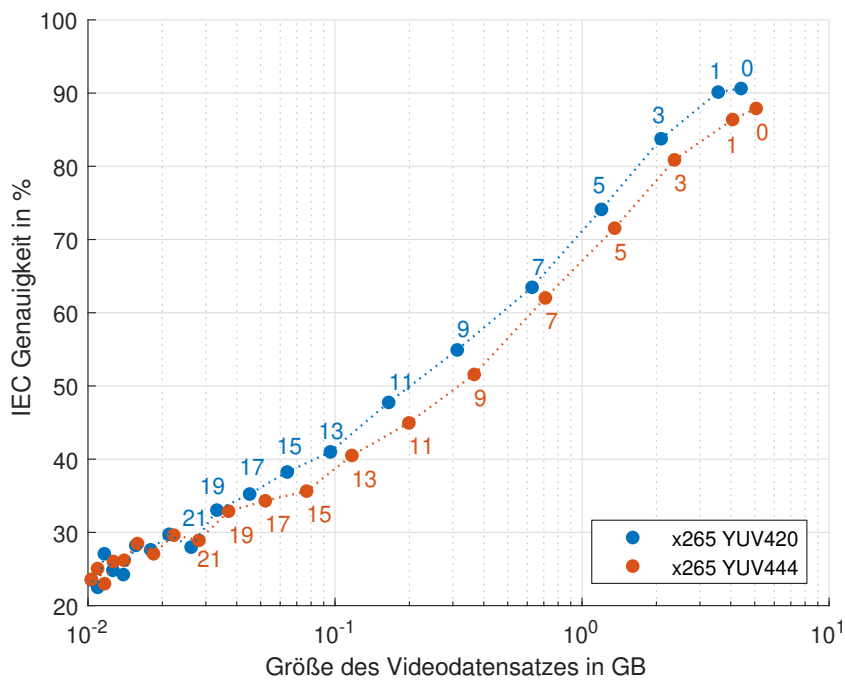


Abbildung 4.11: Mittlere IEC-Genauigkeit für verschiedene CRF-Werte (Zahlen) und die Datensatzgröße für **x265** und verschiedene **Farbformate** auf dem **PURE**-Datensatz.

4 Experimentelle Ergebnisse

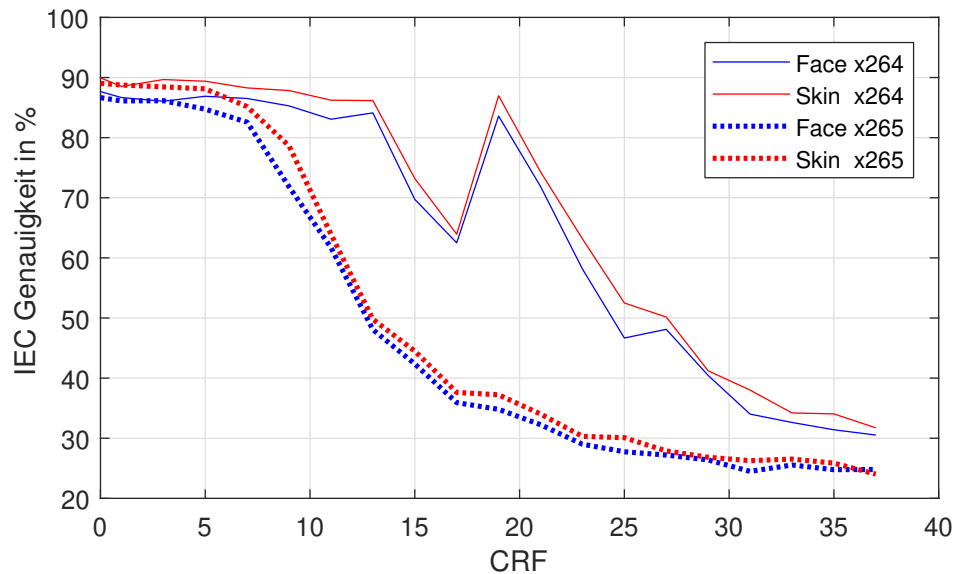


Abbildung 4.12: Mittlere IEC-Genauigkeit für verschiedene CRF-Werte für *Skin* und *Face-Mid ROIs* mit den *x264*- und *x265*-Codecs auf dem *MMSE*-Datensatz.

personen viele Mundbewegungen (Öffnen/Schließen/Lächeln/Sprechen). Dabei wurden die Zähne durch den *Skin*-ROI-Algorithmus als Haut klassifiziert. Dadurch wurden Nicht-Hautpixel bei der Berechnung des PPG-Signals berücksichtigt, was eine hohe Amplitudenänderung des Signals zur Folge hat und die Effizienz der Signalfilterung einschränkt.

Weiterhin zeigt der *x264*-Codec einen Einbruch in beiden Datensätzen, um einen CRF von 15 – 21. Die Genauigkeit fällt sehr steil ab, um anschließend wieder deutlich auf ein höheres Niveau als vor dem Einbruch zu steigen. Dieser Effekt wurde bereits in den vorherigen Abschnitten beobachtet und wird in Kapitel 5.1.1 diskutiert.

4.4.5 Vergleich der Signalextraktionsmethoden

Die Abbildungen 4.14, 4.15, 4.16 und 4.17 zeigen die IEC-Genauigkeit für die verschiedenen Signalextraktionsmethoden. Die IEC-Genauigkeit der verschiedenen Methoden war bei niedrigen CRF-Werten vergleichbar. Keine

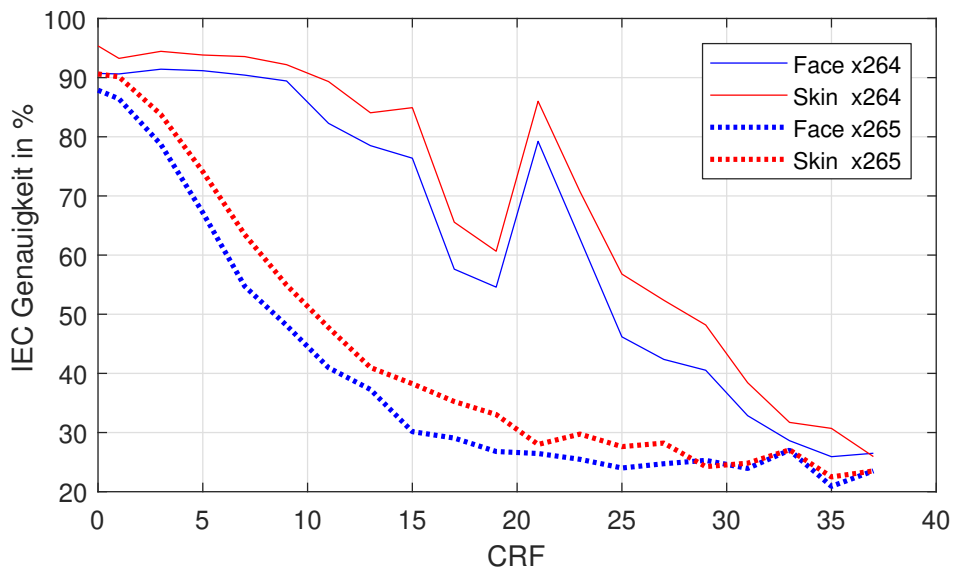


Abbildung 4.13: Mittlere IEC-Genauigkeit für verschiedene CRF-Werte für *Skin* und *Face-Mid ROIs* mit den *x264*- und *x265*-Codecs auf dem **PURE**-Datensatz.

Extraktionsmethode war in beiden Datensätzen, Codecs und CRF-Werten eindeutig dominant. *GRD* erzielte bei höheren CRF-Werten im Allgemeinen bessere Ergebnisse, als die anderen Methoden. Dabei waren die Unterschiede beim stärker komprimierten *x265*-Codec deutlicher zu erkennen. Die Ergebnisse des *IFFT*-Ansatzes verschlechterten sich auf dem **PURE**-Datensatz im Vergleich zu den anderen Methoden mit dem *x265*-Codec.

4.4.6 Auflösung

Um den Einfluss der Videoauflösung auf die Genauigkeit zu testen, wurden die Videos mit drei in **FFMPEG** implementierten Skalierungsmethoden auf fünfzehn Auflösungsstufen herunterskaliert. Aufgrund der geringen ursprünglichen Auflösung des **PURE**-Datensatzes (640x400 Pixel) wurde für diese Untersuchungen nur der **MMSE**-Datensatz verwendet.

Die Videos wurden von der ursprünglichen Auflösung 1040x1392 Pixel linear in 1/16 Schritten der Original-Pixelauflösung bis zu einem Minimum von 130x174 Pixel neu encodiert. Dabei entstandene ungerade Pixelabmessungen

4 Experimentelle Ergebnisse

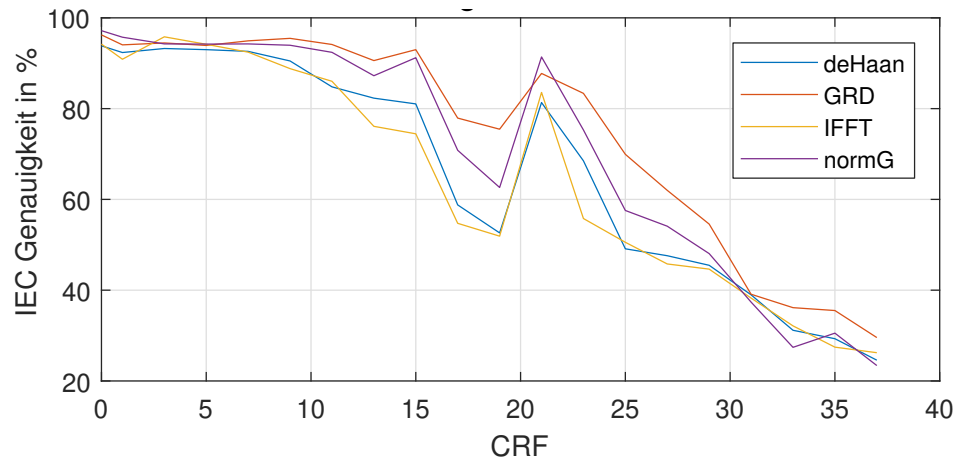


Abbildung 4.14: IEC-Genauigkeit für verschiedene CRF-Werte unter Verwendung verschiedener **Signalextraktionen** und des **x264**-Codecs auf dem **PURE**-Datensatz.

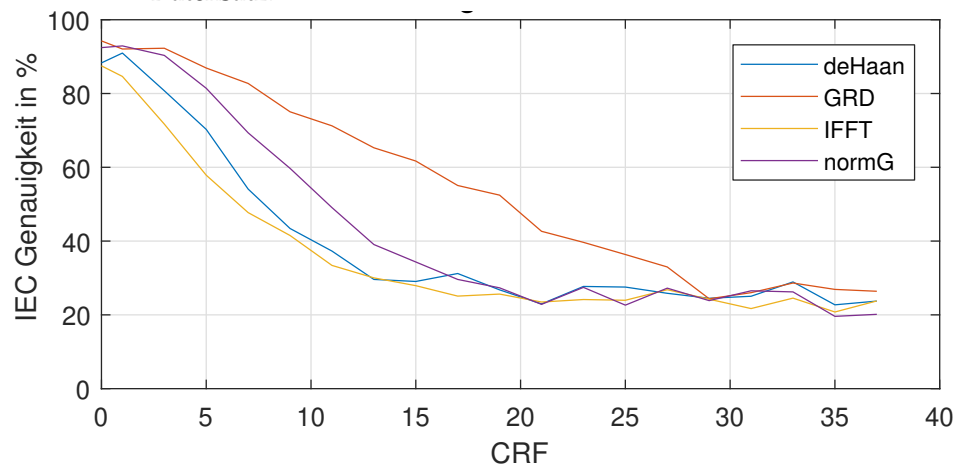


Abbildung 4.15: IEC-Genauigkeit für verschiedene CRF-Werte unter Verwendung verschiedener **Signalextraktionen** und des **x265**-Codecs auf dem **PURE**-Datensatz.

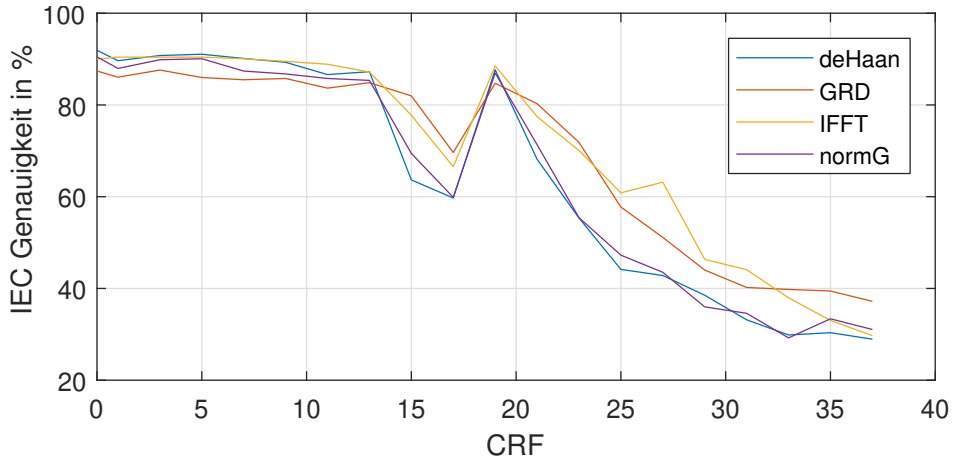


Abbildung 4.16: IEC-Genauigkeit für verschiedene CRF-Werte unter Verwendung verschiedener **Signalextraktionen** und des **x264**-Codecs auf dem **MMSE**-Datensatz. **PURE** dataset.

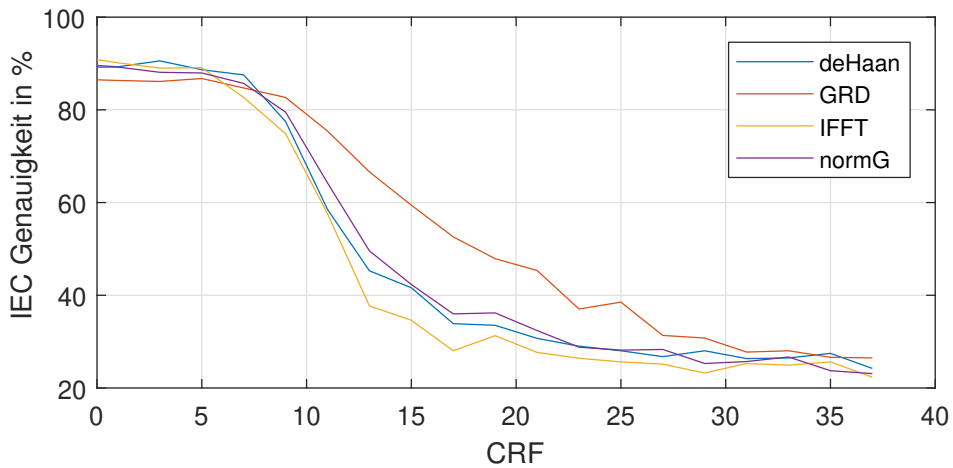


Abbildung 4.17: IEC-Genauigkeit für verschiedene CRF-Werte unter Verwendung verschiedener **Signalextraktionen** und des **x265**-Codecs auf dem **MMSE**-Datensatz. **PURE** dataset.

wurden aufgerundet, da die verwendeten Codes gerade Werte der Bildhöhe und -breite benötigen. Alle Videos zur Analyse der Auflösungs-skalierung wurden aus den ursprünglichen *jpg*-Bildern unter Verwendung des *x265*-Codecs mit $CRF = 0$ erstellt. Die verwendeten Skalierungsalgorithmen sind *nearest neighbor*, *area* und *bicubic* (FFMPEG Standard). Während die *area*- und *bicubic*-Algorithmen ihren neuen Pixelwert aus der Informationen mehrerer Pixel berechnen, setzt der *nearest neighbor*-Ansatz die Zielfarbe aus der Farbwert des räumlich nächstgelegenen Pixels im Originalbild und verwirft die restlichen Farbinformationen. Weitere Informationen bezüglich der Skalierungsmethoden sind in Kap. 2.2.1 zu finden.

Abb. 4.18 stellt die mittlere IEC-Genauigkeit, über alle vier Signalextraktionsmethoden gemittelt, für verschiedene Videoauflösungen dar. Die Mittelwerte und Standardabweichungen des Fehlers sind in Tab. 4.4 am Ende des Kapitels dargestellt. Sie zeigt einen stabilen Wert bis zu etwa 100.000 Pixel (~ 316 Pixel zum Quadrat) im Gesichts-Bounding-Box. Die *area*- und *bicubic*-Skalierungsalgorithmen weisen ab diesem Zeitpunkt einen merklichen Genauigkeitsabfall auf, während die Genauigkeit der *nearest neighbor*-Algorithmen mit Ausnahme der kleinsten getesteten Auflösung eine Genauigkeit von über 85% erreicht.

Die Unterschiede in der PPG-Signalabweichung unter Verwendung der verschiedenen Skalierungsalgorithmen sind in Abb. 4.19 zu sehen. Sie zeigt den Root Mean Square (RMS) Fehler der *normG* PPG-Signale des MMSE-Datensatzes zu den PPG-Signalen der ursprünglichen 1040×1392 -Pixel Videos. Die RMS-Fehler haben eine ähnliche Tendenz wie die IEC-Genauigkeit, die in Abb. 4.18 zu sehen ist. Der Fehler steigt anfangs langsam mit der Verringerung der Auflösung für alle drei Methoden bis zu etwa 100.000 Gesichtspixeln. Der *nearest neighbor*-Ansatz zeigt den geringsten Fehler und bleibt im Gegenzug zu den *area*- und *bicubic*-Skalierungsalgorithmen auch bei Werten von unter 100.000 ROI-Pixeln stabiler, während der Fehler bei den anderen Ansätzen deutlich ansteigt, was mit der Abnahme der IEC-Genauigkeit bei denselben Auflösungen korrespondiert. Unterhalb einer Größe von etwa 20.000 Pixeln (~ 141 Pixel zum Quadrat) steigt der Fehler bei allen Algorithmen erheblich an.

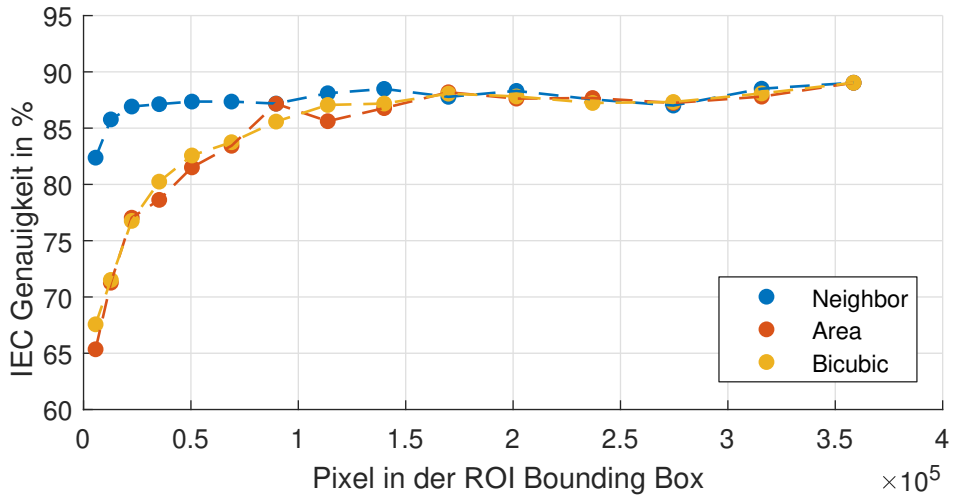


Abbildung 4.18: Mittlere IEC-Genauigkeit in Abhängigkeit der gemittelten ROI Bounding Box-Größe verschiedener Videoauflösungen mit drei Skalierungsalgorithmen auf dem MMSE-Datensatz (x265, CRF=0).

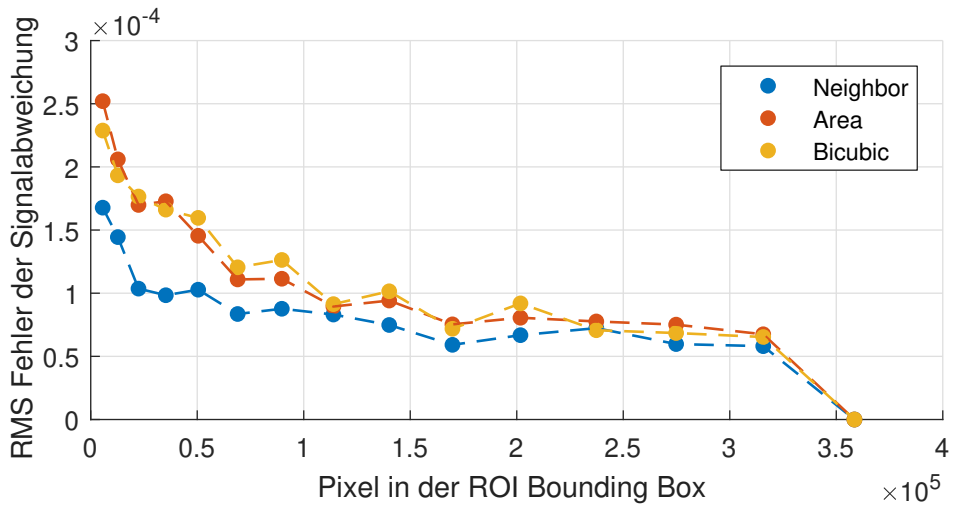


Abbildung 4.19: RMS-Fehler der PPG-Signale (*normG*) von verschiedenen Videoauflösungen in Bezug auf das Originalvideo bei Verwendung von drei Skalierungsalgorithmen auf dem MMSE-Datensatz (x265, CRF=0).

4.4.7 Bildwiederholungsrate

Um den Einfluss von verschiedenen Bildwiederholraten (FPS) zu testen wurden die berechneten PPG Signale durch Unterabtastung modifiziert. Das PPG Signal wurde mit verschiedenen Werten für die Bildwiederholraten (7, 10, 15, 20, 25 FPS) neu generiert, um eine geringere Aufnahmezeit zu simulieren. Das ursprüngliche PPG-Signal wurde mittels *nearest neighbor* Verfahren mit gleichmäßig verteilten Abtastpunkten interpoliert. Dabei wurde der interpolierte Wert an einem Punkt, als der Wert des am zeitlich nächstgelegenen Originalwertes übernommen. Dadurch werden im Signal enthaltenen Informationen reduziert. Zusätzlich wurde ein Teil der (0, 5, 10, 20 oder 30%) der interpolierten Bilder des Zeitfensters entfernt, um während einer Live-Aufnahme verlorene Bilder zu simulieren. Die interpolierten Signale wurden im Anschluss auf 25 Bilder pro Sekunde hoch skaliert, um die Effekte der nachfolgenden Schritte, wie der Filterung, zu vereinheitlichen.

Alle Kombination der beiden Bildwiederholraten-Parameter wurden für die drei Datenbanken auf einem Zeitfenster von 10 Sekunden berechnet. Beide Parameter wurden für die Auswertung zur *effektiven Bildwiederholrate* kombiniert, welche der durchschnittlichen Anzahl der Bilder pro Sekunde entspricht. So korrespondieren 10 FPS mit 20% verlorenen Bildern einer effektiven Bildwiederholrate von 8 FPS.

Abbildungen 4.20, 4.21 und 4.22 zeigen die mittlere IEC-Genauigkeit in Abhängigkeit der effektiven Bildwiederholrate. Bei allen Datenbanken ist diese bis zu einem Wert von etwa 10 FPS stabil, jedoch leicht abfallend. Unterhalb davon fällt die Genauigkeit der Herzratenschätzung stark ab.

4.4 Videoeigenschaften

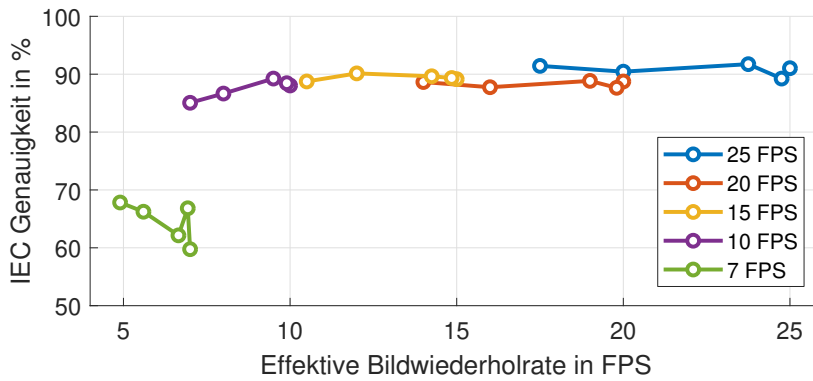


Abbildung 4.20: Mittlere IEC-Genauigkeit in Abhängigkeit der effektiven Bildwiederholrate (gleiche interp. FPS verbunden) auf dem MMSE-Datensatz (x264, CRF=0).

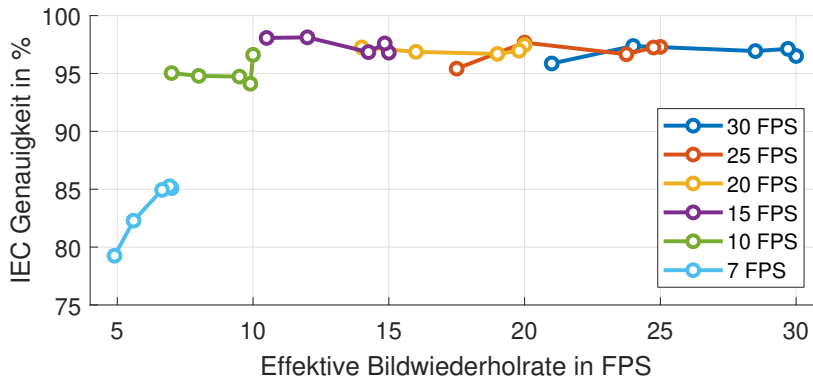


Abbildung 4.21: Mittlere IEC-Genauigkeit in Abhängigkeit der effektiven Bildwiederholrate (gleiche interp. FPS verbunden) auf dem PURE-Datensatz (x264, CRF=0).

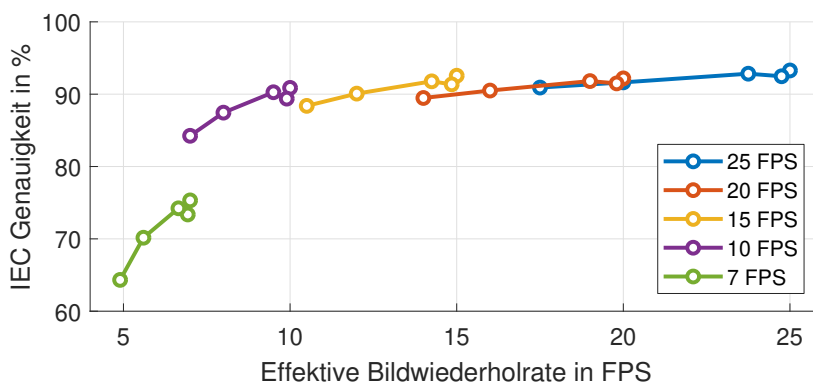


Abbildung 4.22: Mittlere IEC-Genauigkeit in Abhängigkeit der effektiven Bildwiederholrate (gleiche interp. FPS verbunden) auf dem BioVid-Datensatz (x264, CRF=17).

4 Experimentelle Ergebnisse

Tabelle 4.2: Mittelwert μ und Standardabweichung σ der absoluten Fehler der Herzfrequenzschätzungen auf den Datensätzen **PURE** und **MMSE** unter Verwendung verschiedener **ROIs** und **codexs** in Bezug auf den CRF-Wert.

	CRF	0	1	3	5	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35	37			
MMSE	FaceMid	x264	0.10	-0.15	-0.13	-0.02	0.14	0.22	-0.18	0.04	-0.84	-3.10	-0.30	-3.62	-3.75	-3.99	-3.76	-3.84	-3.58	-3.43	-2.43	-2.40		
		x265	-0.26	-0.12	-0.02	0.20	0.73	0.00	-0.64	-1.82	-0.84	-2.73	-3.35	-3.62	-3.45	-3.75	-3.18	-3.58	-3.95	-2.63	-3.73	-2.27	-3.36	
	Haut	x264	-0.12	-0.40	-0.33	-0.10	-0.38	-0.02	0.00	-0.11	-0.64	-1.78	-3.18	-3.42	-3.31	-3.42	-3.47	-3.31	-4.28	-3.32	-3.10	-2.56	-2.87	
		x265	-0.30	-0.22	0.06	0.04	0.75	0.43	-0.96	-1.97	-2.72	-3.50	-3.24	-3.06	-4.50	-3.42	-3.47	-3.69	-3.38	-2.97	-3.10	-2.94	-2.94	
		x264	2.99	3.43	2.48	2.95	3.43	2.79	3.35	5.38	6.93	3.54	5.50	6.57	6.78	6.85	5.00	6.27	6.66	6.65	7.45	8.99	8.99	
PURE	FaceMid	x264	4.33	3.35	3.56	4.18	3.72	3.14	2.13	2.36	1.55	1.89	2.98	3.10	2.17	2.97	3.26	3.13	5.67	4.71	2.71	5.51	5.51	
		x265	1.58	2.65	2.07	2.35	2.46	2.97	3.43	3.43	3.45	6.39	4.98	4.64	5.80	6.36	5.56	6.21	6.08	6.68	7.29	8.55	8.55	
	Haut	x264	3.80	3.77	4.32	5.18	4.16	4.16	3.55	2.71	2.26	3.25	3.10	4.56	2.75	2.86	3.40	2.85	4.93	3.96	3.96	2.78	4.63	4.63
		x265	9.42	9.77	10.51	10.04	10.26	10.99	10.99	11.23	11.16	15.07	15.80	14.47	14.39	17.34	19.13	18.46	20.01	21.64	21.64	21.34	22.32	22.32
		x264	9.39	9.71	10.30	11.04	11.49	14.17	16.44	16.44	19.32	20.48	21.06	14.47	14.39	17.34	19.13	18.46	20.01	21.64	21.64	21.34	22.32	22.32
MMSE	FaceMid	x264	9.11	9.77	10.30	11.04	11.49	14.17	16.44	16.44	19.32	20.48	21.06	14.47	14.39	17.34	19.13	18.46	20.01	21.64	21.64	21.34	22.32	22.32
		x265	8.98	9.25	9.81	10.15	11.14	12.83	16.03	19.25	19.84	21.21	21.33	22.06	22.69	22.21	22.72	24.22	24.05	25.74	25.74	25.14	25.14	25.14
	Haut	x264	15.54	15.10	13.03	13.63	15.22	14.91	16.57	17.16	18.68	21.10	20.95	19.18	21.98	23.23	23.57	24.22	24.05	25.74	25.74	25.14	25.14	25.14
		x265	16.63	16.65	18.03	21.44	23.00	24.22	25.17	25.47	26.31	26.21	26.37	26.11	26.79	27.06	26.56	26.05	26.56	25.18	25.43	26.35	25.88	25.88
		x264	10.12	12.66	11.74	12.18	12.66	13.57	15.75	17.37	15.74	21.94	21.14	17.43	20.02	23.26	22.61	23.13	24.77	24.77	25.01	25.17	25.17	25.17
PURE	FaceMid	x264	16.13	15.54	17.08	19.89	21.41	23.08	23.57	24.43	24.67	24.23	24.58	25.48	25.33	24.97	24.92	25.79	25.33	25.36	25.56	26.13	25.71	
	Haut	x265	16.13	15.54	17.08	19.89	21.41	23.08	23.57	24.43	24.67	24.23	24.58	25.48	25.33	24.97	24.92	25.79	25.33	25.36	25.56	26.13	25.71	

Tabelle 4.3: Mittelwert μ und Standardabweichung σ der absoluten Fehler der Herzfrequenzschätzungen auf dem **PURE**-Datensatz unter Verwendung verschiedener **Pixelformate** in Bezug auf den CRF-Wert.

	CRF	0	1	3	5	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35	37		
BPM	YUV420	x264	1.58	2.63	2.07	2.33	2.46	2.97	3.57	3.43	3.45	6.39	4.08	4.64	5.80	6.36	5.56	6.21	6.08	6.68	7.29	8.55	8.55
		x265	3.80	3.77	4.32	5.18	4.16	3.55	2.71	2.26	3.25	3.10	4.36	4.36	8.26	3.40	4.04	2.85	4.93	3.96	2.78	4.63	4.63
	YUV444	x264	2.48	2.49	2.74	2.29	3.26	2.30	4.29	3.62	3.25	5.02	7.57	8.56	8.62	8.24	7.21	7.14	6.98	6.65	7.39	7.37	7.37
		x265	3.50	3.19	5.02	5.22	5.52	4.00	3.04	2.77	3.27	3.98	3.98	4.10	4.31	3.67	5.22	5.50	5.50	4.83	5.54	6.65	6.65
		x264	10.12	12.66	11.74	12.18	12.66	13.57	15.75	17.37	15.74	21.94	21.14	17.43	20.02	23.26	22.61	23.13	24.77	24.77	25.01	25.17	25.17
BPM	YUV420	x264	16.13	15.54	17.08	19.89	21.41	23.08	23.57	24.43	24.67	24.23	24.58	25.48	25.33	24.97	24.92	25.79	25.33	25.36	25.56	26.13	25.71
		x265	12.94	13.50	13.25	15.70	11.82	16.84	15.18	17.64	17.01	21.79	23.20	23.59	24.45	24.64	25.14	25.71	25.63	25.26	24.48	24.48	24.48
	YUV444	x264	16.41	18.50	20.78	22.27	24.01	23.69	24.87	25.18	24.62	24.40	26.05	25.45	25.85	25.66	26.29	25.87	26.13	25.06	25.90	25.90	25.90
		x265	16.38	16.41	18.50	20.78	22.27	24.01	23.69	24.87	25.18	24.62	24.40	26.05	25.45	25.85	25.66	26.29	25.87	26.13	25.06	25.90	25.90
		x264	10.12	12.66	11.74	12.18	12.66	13.57	15.75	17.37	15.74	21.94	21.14	17.43	20.02	23.26	22.61	23.13	24.77	24.77	25.01	25.17	25.17

Tabelle 4.4: Mittelwert μ und Standardabweichung σ der absoluten Fehler der Herzfrequenzschätzungen auf dem **MMSE**-Datensatz für verschiedene **Auflösungen** unter Verwendung verschiedener Skalierungsalgorithmen in Bezug auf den Mittelwert der Bounding-Box-Pixel des Gesichts.

	Anzahl der Pixel	5.607	12.711	22.438	35.232	50.456	68.892	89.638	113.698	139.958	169.875	201.549	237.015	274.592	315.720	358.524
μ in BPM	Neighbor	5.05	3.84	3.35	3.41	2.83	3.41	3.19	2.78	2.55	3.00	2.57	2.95	3.23	2.76	-0.30
	Area	10.27	8.33	6.61	6.67	5.72	5.11	4.14	4.14	3.57	2.75	3.05	2.90	3.11	2.80	-0.30
	Bi-cubic	9.87	8.48	6.60	5.90	5.09	4.94	3.81	3.42	2.92	2.81	2.81	2.88	2.96	3.16	2.85
σ in BPM	Neighbor	14.37	12.46	11.52	11.66	10.18	12.43	10.57	10.64	10.01	11.50	10.13	11.14	11.67	10.99	8.98
	Area	18.25	16.39	15.33	15.84	15.65	14.80	14.80	13.14	12.68	10.48	11.15	10.74	11.13	10.63	8.98
	Bi-cubic	17.50	16.81	15.11	14.79	14.16	14.40	14.40	12.95	11.71	10.53	10.80	11.07	10.83	11.34	10.78

4.5 Region of Interest (ROI)

Für die Untersuchung des Einflusses der ROI auf die Herzfrequenzschätzung wurden verbreitete ROIs aus dem Stand der Technik ausgewählt und mit den neu entwickelten, in Kapitel 3.1 vorgestellten, Hautsegmentierungsansätzen verglichen. Teile der Ergebnisse dieses Kapitels wurden vorab in [Rap+16a] und [Rap+18b] publiziert.

4.5.1 Landmarkenbasierte ROIs

Es wurden drei ROI-Methoden implementiert, um die Ergebnisse der vorgestellten Hautsegmentierungsansätze mit gängigen ROIs aus der Literatur zu vergleichen. Diese wurden in Kapitel 2.2.2 ausführlicher beschrieben.

Die **FaceMid** ROI verwendet 60% der zentrierten Breite und die volle Höhe der Bounding Box des Gesichts als aktiven Bereich (siehe Abb. 4.23(a)). Die **Forehead** ROI verwendet 50% der zentrierten Breite und 30% der Länge des Abstands zwischen den Augenwinkeln und der Unterseite der Nase als Höhe, beginnend bei den Augenbrauen (siehe Abb. 4.23(b)). Die in [Fen+15] vorgeschlagene **Feng-ROI** verwendet zwei Regionen auf den **Wangen** als aktiven Bereich. Die ROIs werden durch die Verfolgung von Speeded-Up-Robust-Feature (SURF)-Punkten in der Mitte des Gesichts berechnet und eine affine Transformation auf die ROIs angewendet, um Kopfbewegungen zu kompensieren.

4.5.2 Hautsegmentierung

Für die Hautsegmentierung verwenden wir *BayesBGR* 32^3 Ansatz [JR99] basierende Hauterkennungsmethode welcher in Kapitel 3.1 ausführlicher beschrieben ist. Dieser Ansatz weist, durch eine Look-up-Table (LUT), jeder RGB-Farbe, und damit jedem Pixel eine Hautfarbenwahrscheinlichkeit p zu (siehe Abb. 3.1 und 3.2). Auf Grundlage der Hautfarbenwahrscheinlichkeit wurden drei verschiedene ROI abgeleitet und deren Einfluss auf die Herzratenschätzung untersucht.

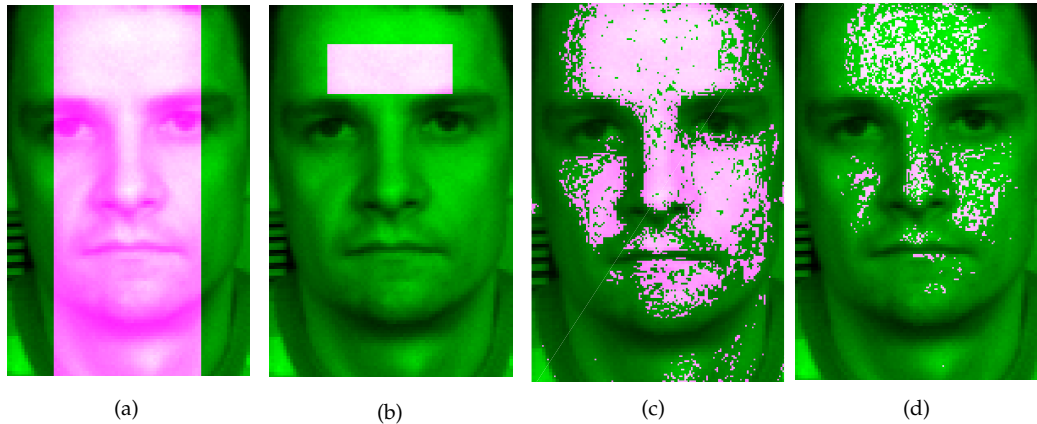


Abbildung 4.23: Auf Gesichtsmerkmalen und Hautsegmentierung basierende ROIs: (a) Mitte des Gesichts (FaceMid), (b) Stirn (Forehead), (c) Wahrscheinlichkeitsschwelle $t = 0,3$ und (d) Flächenschwelle $A = 10\%$ (ROI-Pixel in pink).

Hautsegmentierung mit einer Wahrscheinlichkeitsschwelle t Diese Schwelle segmentiert alle Pixel mit $p > t$ (siehe Abb. 4.23 (c) und 3.2) als Haut. Dabei werden bei einem Schwellwert von zum Beispiel $t = 0.30$ alle Pixel als Haut segmentiert, deren RGB Farbwerte eine Hautfarbenwahrscheinlichkeit von mindestens 30% aufweisen.

Hautsegmentierung mit einem Flächenschwelle A Mit dieser Schwelle werden die Pixel mit der höchsten Wahrscheinlichkeit als Haut segmentiert (siehe Abb. 4.23(d)). So werden für einen Schwellwert von zum Beispiel $A = 10\%$ die 10% der Pixel mit den höchsten Hautfarbenwahrscheinlichkeiten als ROI Pixel ausgewählt. Dieser Ansatz hält die Anzahl der Hautpixel konstant, auch wenn z. B. eine plötzliche Änderung der Beleuchtung einen großen Einfluss auf die Ergebnisse der Hautwahrscheinlichkeiten $p(c)$ hat.

Hautsegmentierung mit $p(c)$ als gewichtetes Mittel Bei diesem Ansatz wird die berechnete Hautwahrscheinlichkeit p_i für jedes Pixel i (siehe Abb. 3.2) als Gewichtungsfaktor bei der Berechnung der RGB Werte für das PPG-Signal verwendet (siehe Kapitel 3.1, Gleichung 3.3).

4.5.3 Daten und Signalverarbeitung

Datenbanken

Für die Experimente und Untersuchungen der verschiedenen ROIs wurden mehrere Datenbanken verwendet, um mögliche Überanpassungen auf eine Datenbank zu vermeiden und aussagekräftigere Ergebnisse zu generieren. Die **PURE** Datenbank wurde für die ersten Untersuchungen der Hautsegmentierung verwendet. Diese Datenbank enthält wenig schnelle Bewegungen und liegt in einer verlustfreien Kompression vor. Sie wurde daher für die Vergleiche der verschiedenen Schwellwerte ausgewählt, um einen Großteil möglicher Fehlerquellen (Bewegung, Tracking, Beleuchtungsänderungen, ...) während der Implementierung und Auswertung auszuschließen. Im Anschluss wurde der Einfluss der Hautsegmentierungs-ROI auf den größeren und vielfältigeren Datenbanken **BioVid** und **MMSE-HR** getestet und mit den anderen Verfahren verglichen. Die verwendeten Datenbanken sind in Kapitel 4.3 beschrieben. Die Generierung und Auswertung der Grundwahrheiten sind in Kapitel 4.2 beschrieben.

Algorithmen

Um die Auswirkung der ROIs auf die nachfolgenden Verarbeitungsschritte zu testen, wurden verschiedene Algorithmen mit unterschiedlichen Ansätzen für die verschiedenen Verarbeitungsschritte der Herzratenschätzung ausgewählt und implementiert. Unseres Wissens wurde keiner der Algorithmen mithilfe der **PURE** Datenbank erstellt oder auf ihr getestet.

Blöcher, Schneider, Schinle und Stork [Blö+17] verwenden die Stirnregion des Gesichts als ROI (siehe Abb. 4.23b) über die für jedes Frame die RGB-Werte gemittelt werden. Nach der Normalisierung der Farb-Zeitsignale wird mithilfe einer ICA (siehe Kapitel 2.2.3) unter Verwendung der Jade-Algorithmus-Implementierung das PPG-Signal berechnet und anschließend bandpassgefiltert (0,75-4 Hz). Der Ansatz verwendet die Hilbert-Transformation (siehe Kapitel 2.2.3), um den Phasenwinkel des Pulssignals

zu berechnen. Die Herzfrequenz wird durch die Erkennung der Phasensprünge, die die Schläge repräsentieren, berechnet und die Herzfrequenz aus den IBIs bestimmt.

de Haan und Jeanne [dJ13] verwenden ein nicht näher beschriebenes *einfaches Hautauswahlverfahren* zur Definition der ROI. Ihr Verfahren zur Generierung des PPG-Signals verwendet einen chrominanzbasierten Ansatz, um die RGB-Kanäle zum PPG-Signal zu kombinieren. Das Ziel ist den Effekt von spiegelnden Reflexionen, welche durch Bewegungen des Kopfes eine starke Varianz in der Helligkeit verschiedener Hautregionen erzeugen, zu eliminieren. Der Ansatz ist in Kapitel 2.2.3 beschrieben. Das PPG-Signal wird anschließend mittels einer FFT in den Frequenzbereich transformiert und die maximale Leistungsspitze des Spektrums zur Bestimmung der Herzfrequenz verwendet.

Feng, Po, Xu, Li und Ma [Fen+15] verwenden eine Version der Wangen ROI. Zur Berechnung des PPG-Signals wird ein adaptives Grün/Rot-Differenzfarbmodell verwendet (siehe Kapitel 2.2.3). Nach einem längeren (20s) Fenster, welches für eine erste Spektralanalyse genutzt werden die letzten 4 Sekunden des Signals erneut gefiltert. Dazu wird ein Bandpass mit mehreren kleineren Durchlassbereichen (± 10 BPM) um die Frequenz mit der höchsten Signalstärke und ihre nächsten zwei Harmonischen verwendet. Die Herzfrequenz wird dann aus den **IBIs** der Signalspitzen berechnet. Es wurde zusätzlich eine modifizierte Version des Algorithmus (**Feng_mod**) verwendet, bei der die zusätzlichen Durchlassbänder der harmonischen Frequenzen nicht genutzt wurden. Dadurch wurde die Anzahl der falsch positiven Peaks stark reduziert und die Schätzrate verbessert.

Lewandowska, Rumiński, Kocejko und Nowak [Lew+11] verwenden die *FaceMid* ROI (siehe Abb. 4.23a). Die über jeden Frame gemittelten RGB-Kanalsignale werden zunächst bandpassgefiltert (0,5-3,7 Hz). Nach einer PCA (siehe Kapitel 2.2.3) der drei Farbsignale wird die Ausgabe der ersten Hauptkomponente zur Berechnung der Herzfrequenz verwendet, indem die maximale Leistungsspitze des Spektrums in der Fourier-Transformation des Signals gefunden wird.

Poh, McDuff und Picard [PMP11] verwendet die *FaceMid* ROI. Es werden alle drei RGB-Kanäle extrahiert. Nach einer Glättung des Signales und einer ICA (siehe Kapitel 2.2.3) wird das Ausgangssignal mit dem maximalen

Leistungsspeak im gefilterten Spektrums als PPG-Signal definiert. Dieses Signal wird bandpassgefiltert (0,7-4 Hz) und die Herzfrequenz wird aus den IBIs der Signalspitzen berechnet. Um Bewegungsartefakte zu minimieren, werden die IBIs vor der Herzratenberechnung mittels eines modifizierten *noncausal of variable threshold* (NC-VT) Algorithmus [Vil+97] gefiltert.

Rapczynski, Werner und Al-Hamadi [RWA16] verwenden eine farbbaasierte Hautsegmentierung als ROI. Der normalisierte Grün-Kanal (siehe Kapitel 2.2.3) wird als PPG-Signal verwendet. Das Signal wird dann mit einem adaptiven Bandpass mit einem kleinen Durchlassbereich (± 15 BPM) in Abhängigkeit von der letzten Herzfrequenzschätzung gefiltert (siehe Kapitel 3.2). Ein graphenbasierter Algorithmus bestimmt eine optimale Peaksequenz, welche die Sequenz der letzten Messung mit minimaler Abweichung fortsetzt (siehe Kapitel 3.3). Aus den IBIs der ermittelten Peaks wird die Herzfrequenz errechnet .

Wang, Brinker, Stuijk und Haan [Wan+17] verwenden die *FaceMid* ROI. Die für jedes Frame gemittelten RGB-Farbsignale werden zunächst Fouriertransformiert. Anschließend wird das PPG-Signal aus dem Frequenzbereich mit einer spezifischen Kombinationen der Farbkanal und der berechneten Gewichte rekonstruiert. Eine ausführliche Beschreibung des Verfahrens ist in Kapitel 2.2.3 zu finden. Die maximale Spektralleistungsspitze des PPG-Signals wird dann zur Berechnung der Herzfrequenz verwendet.

Li, Chen, Zhao und Pietikainen [Li+14] stellten einen Ansatz für eine Beleuchtungskompensation vor. Die Methode basiert auf der abstandsregulierten Level-Set-Evolution. Dabei wird der Hintergrundbereich segmentieren und ein Referenzsignal zur Beleuchtungskompensation generiert. Die Herzfrequenz wird dann, unter Berücksichtigung des Referenzsignals, anhand der maximalen Leistungsspektraldichte ermittelt. Der Algorithmus wurde ohne die in der Veröffentlichung enthaltene Verwerfung der verrauschtesten 5% der Signalfester verwendet, um eine bessere Vergleichbarkeit zu ermöglichen.

4 Experimentelle Ergebnisse

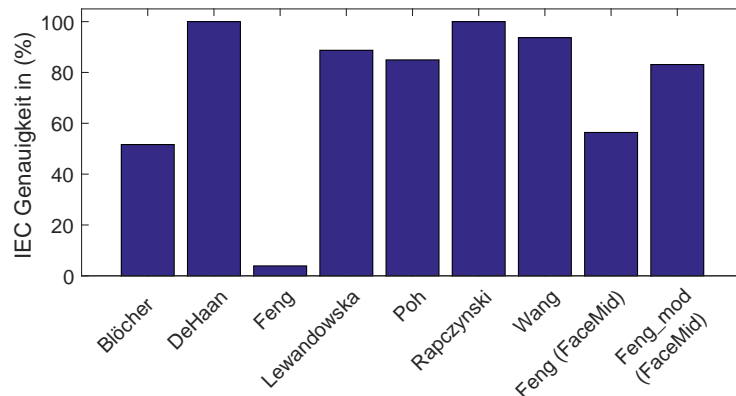


Abbildung 4.24: IEC Mittelwert über alle 10 Probanden für jeden Algorithmus auf dem *Steady*-Teil des *PURE* Datensatzes.

4.5.4 Ergebnisse

Die verschiedenen implementierten Verfahren wurden sowohl mit den landmarkenbasierten als auch auf den hautfarbenbasierten ROI getestet. Neben dem arithmetischen Mittel des Fehlers und der Standardabweichung wird die Güte der Herzratenschätzung für die einzelnen Versuchsreihen mithilfe der IEC Genauigkeit angegeben (siehe Kapitel 4.2.2).

Implementierung

Die Algorithmen wurden nur unter Verwendung des *Steady* Teils der *PURE* Datenbank implementiert, um die grundsätzliche Funktionsweise der Herzratenschätzung sicherzustellen und eine mögliche Überanpassung an die Datenbank während des Testens der Implementationen zu verhindern.

Abbildung 4.24 zeigt die Ergebnisse der verschiedenen Algorithmen auf dem *Steady* Teil des *PURE* Datensatzes. Der in [Fen+15] vorgestellte ROI-Ansatz von **Feng** konnte nicht zuverlässig SURF-Punkte in den Gesichtern der Probanden finden und verfolgen. Die ROI auf den Wangen war mit diesem Ansatz sehr instabil. Dies führte zu deutlich schlechteren Ergebnissen als die anderen Verfahren (siehe Abb. 4.24). Weitere Versuche an Testvideos

Tabelle 4.5: IEC-Genauigkeit (in %) von ROIs (Spalten) und Algorithmen (Zeilen) für den *PURE* Datensatz. Beste ROI für jeden Algorithmus fett markiert.

	FaceMid	Forehead	Haut ($t=0.2$)	Haut ($A=30\%$)	Mittelwert	Std. Abweichung
Blöcher [Blö+17]	59.3	52.8	65.1	66.0	58.3	6.7
DeHaan [dJ13]	90.3	88.0	89.7	93.8	86.2	4.6
Feng_mod	80.9	82.5	81.0	83.3	77.2	6.0
Feng [Fen+15]	56.2	61.2	59.9	60.1	55.2	6.4
Lewandowska [Lew+11]	68.9	60.5	83.3	54.2	61.8	13.7
Poh [PMP11]	78.7	81.8	82.4	82.6	75.9	7.8
Rapczynski [RWA16]	78.0	86.1	90.0	90.1	85.6	5.6
Wang [Wan+17]	88.2	88.3	83.8	90.3	85.5	2.4
Mittelwert	75.1	75.2	79.4	77.5		
Std. Abweichung	11.7	13.5	10.3	14.2		

mit besserer Qualität ergaben eine stabilere ROI. Daher wurde dieser ROI-Ansatz für die weitere Analyse auf der *PURE* Datenbank ausgeschlossen.

Schwellwerte der Hautdetektion

Um den Effekt der Schwellenwerte zu untersuchen, wurde die Herzfrequenzschätzung mit unterschiedlichen Werten für den Wahrscheinlichkeitschwellenwert t und den Flächenschwellenwert A durchgeführt. Der Schwellenwert t wurde für die Werte von 0, 1 bis 0, 9, in 0, 1-Schritten, getestet. Der Schwellenwert A wurde für die Werte von 10% bis 90%, in 10%-Schritten, getestet. Die Randfälle wurden ausgelassen ($t = 0$, $t = 1$, $A = 0\%$ und $A = 100\%$), da sie entweder keine oder alle Pixel enthalten.

Die Abbildungen 4.25 und 4.26 zeigen die Ergebnisse für die verschiedenen Schwellenwerte. Für die meisten Algorithmen lag der optimale Wert bei $t = 0, 2$. Die Wahl eines hohen Schwellenwerts reduziert die Anzahl der Pixel und hat negative Auswirkungen auf die Schätzung. Der Blöcher-Ansatz ist ein Ausreißer, da er scheinbar einen besseren Initialisierungsschritt erhalten hat, kombiniert mit der Tatsache, dass alle implementierten Algorithmen die zuletzt erfolgreich geschätzte Herzfrequenz ausgeben, wenn sie während eines Schrittes kein Ergebnis berechnen können. Die Flächenschwelle A führt zu einer stabilen Fehlerrate ab $A = 20\%$.

4 Experimentelle Ergebnisse

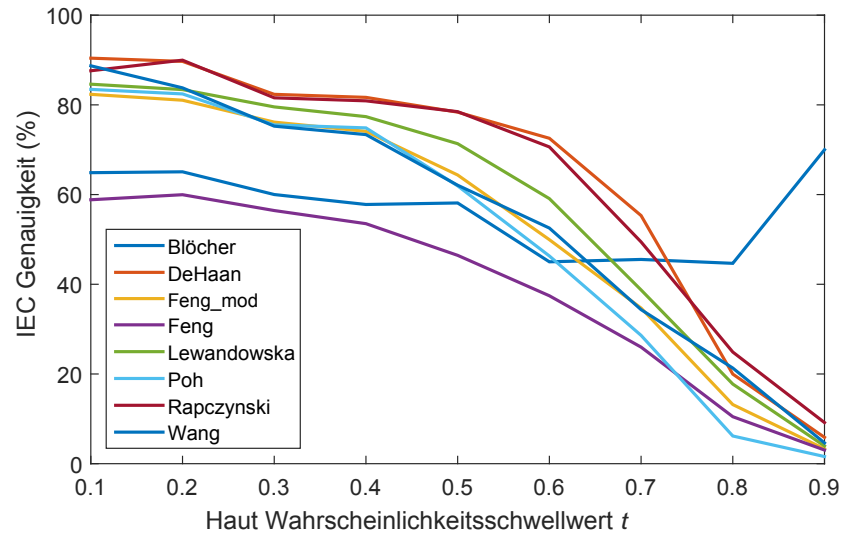


Abbildung 4.25: IEC Genauigkeit der einzelnen Verfahren auf der PURE Datenbank für verschiedene Werte des Wahrscheinlichkeitsschwellwertes t .

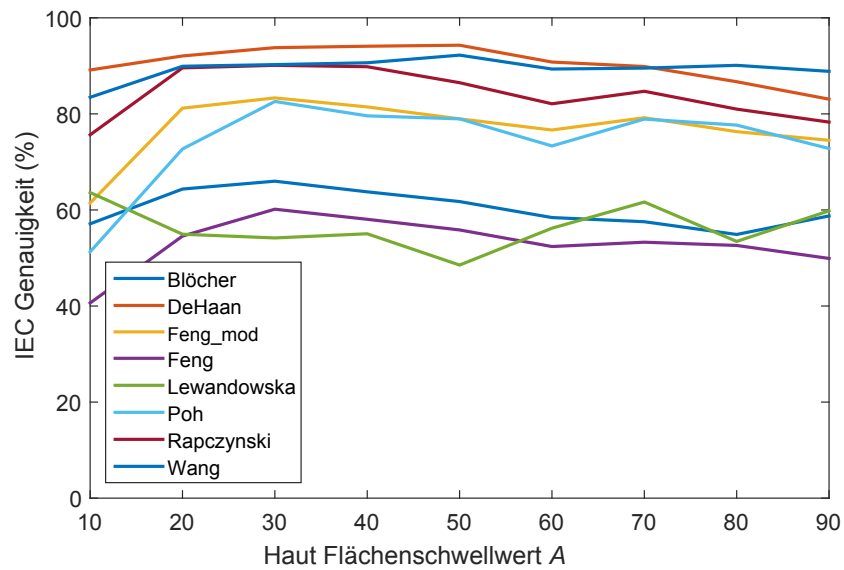


Abbildung 4.26: IEC Genauigkeit der einzelnen Verfahren auf der PURE Datenbank für verschiedene Werte des Flächenschwellwertes A .

Tabelle 4.5 zeigt die IEC-Genauigkeit für ausgewählte Kombinationen sowie die Mittelwerte und Standardabweichungen für die Algorithmen und die *FaceMid* und *Forehead* ROIs, sowie die Hautdetektion mit den Schwellwerten $t = 0.2$ und $A = 30\%$.

Gewichtete Hautdetektion

Aufgrund der notwendigen Parameterwahl bei den beiden Schwellwertverfahren, wurde die gewichtete Mittelwertbildung auf Grundlage der Hautfarbenwahrscheinlichkeit, als parameterfreie Alternative auf der *BioVid Heat Pain Database Part C* und der *MMSE-HR* Datenbank getestet.

Die Ergebnisse für alle Algorithmen wurden für die genannten Datenbanken mit den folgenden ROIs berechnet: *BoundingBox*, *FaceMid*, *Forehead* und *Haut(gewichtet)*. Jede Herzfrequenz-Schätzung wurde einmal pro Sekunde unter Verwendung eines gleitenden Fensters auf die Daten angewendet. Der Startpunkt (0-30 Sek.) und die Breite des Fensters (4-30 Sek.) waren abhängig vom verwendeten Algorithmus.

Der *Feng-ROI*-Ansatz hatte Schwierigkeiten, über einen längeren Zeitraum stabile SURF-Punkte in den Gesichtern der Testpersonen zu finden und zu verfolgen. Die Videos in der Originalarbeit waren nur 20 Sekunden lang, verglichen mit bis zu 2 Minuten in der *MMSE-HR* und etwa 25 Minuten im *BioVid*-Datensatz. Dies verursachte eine Drift der ROIs, da sich durch die konstante affine Transformation kleine Fehler über die Zeit aufsummieren. Aus diesem Grund waren die ROIs auf den Wangen sehr instabil und unzuverlässig mit einer mittleren IEC von 14,4% (*BioVid*) und 14,5% (*MMSE-HR*) über alle Algorithmen. Ein periodisches Zurücksetzen der ROIs oder andere korrigierende Maßnahmen könnten den Ansatz in zukünftigen Vergleichen stabilisieren. Aufgrund dieser Problematik wurde die *Feng-ROI* nicht in die folgenden Ergebnistabellen aufgenommen.

Bei der Analyse der Ergebnisse gab es eine deutliche Abweichung beim *Li*-Algorithmus. Die IEC-Genauigkeit war auf der *MMSE-HR* Datenbank deutlich schlechter, als auf dem *BioVid*-Datensatz. Dies könnte durch eine, für den Datensatz, ungünstig gewählte Schrittweite verursacht werden, welche und zu einer schlechten iterativen Hintergrundkorrektur führen könnte.

Ergebnisse ROI Die Tabellen 4.6 und 4.7 zeigen die Ergebnisse der Herzrhythmus-Schätzung für die jeweiligen Datenbanken in Abhängigkeit der verwendeten ROI und Algorithmen. Die *Forehead* und *Haut(gewichtet)* ROIs erzielen

4.5 Region of Interest (ROI)

Tabelle 4.6: Ergebnisse der Kombinationen von ROIs (Spalten) und Algorithmen (Zeilen) für den *BioVid* Datensatz. IEC-Genauigkeit in %. Original-ROI der Algorithmen fett.

	Bounding Box	FaceMid	Forehead	Haut gewichtet	Mittelwert	Std. Abweichung
Blöcher [Blö+17]	53.2	57.8	59.6	51.1	55.4	3.9
DeHaan [dJ13]	67.9	70.5	71.7	67.8	69.5	1.9
Feng (mod) [Fen+15]	67.3	80.0	87.2	88.3	80.7	9.7
Feng [Fen+15]	46.2	58.7	65.6	66.3	59.2	9.3
Li [Li+14]	34.3	49.0	66.2	45.2	48.7	13.2
Lewandowska [Lew+11]	61.0	71.8	80.3	90.2	75.8	12.4
Poh [PMP11]	67.5	75.3	81.8	80.8	76.4	6.6
Rapczynski [RWA16]	84.9	89.7	92.6	93.3	90.1	3.8
Wang [Wan+17]	84.0	85.5	84.7	86.3	85.1	1.0
Mittelwert	62.9	70.9	76.6	74.4		
Std. Abweichung	16.5	13.6	11.3	17.6		

Tabelle 4.7: Ergebnisse der Kombinationen von ROIs (Spalten) und Algorithmen (Zeilen) für den *MMSE-HR* Datensatz. IEC-Genauigkeit in %. Original-ROI der Algorithmen fett.

	Bounding Box	FaceMid	Forehead	Haut gewichtet	Mittelwert	Std. Abweichung
Blöcher [Blö+17]	45.8	45.1	40.0	64.4	48.9	10.7
DeHaan [dJ13]	77.2	75.3	79.1	71.2	75.7	3.4
Feng (mod) [Fen+15]	74.6	70.8	72.2	77.7	73.8	3.0
Feng [Fen+15]	66.1	65.2	65.9	69.8	66.7	2.1
Li [Li+14]	8.5	8.0	21.0	10.5	12.0	6.1
Lewandowska [Lew+11]	62.5	62.4	68.0	79.6	68.1	8.1
Poh [PMP11]	81.1	80.0	84.9	82.3	82.0	2.1
Rapczynski [RWA16]	89.5	89.3	87.3	91.1	89.3	1.6
Wang [Wan+17]	86.2	86.5	91.3	88.4	88.1	2.4
Mittelwert	65.7	64.7	67.7	70.6		
Std. Abweichung	25.3	25.2	23.3	24.1		

bei beiden Datensätzen im Mittel die besten Ergebnisse. Dies deutet, wie bei den vorherigen Ergebnissen der Hautschwellwerte, auf einen Qualität-vor-Quantität-Ansatz bei der Auswahl der ROI-Pixel hin. Weniger *bessere* Pixel erreichen eine höhere Genauigkeit als mehr *unzuverlässige* Pixel.

Bei Betrachtung der Algorithmen mit hoher IEC-Genauigkeit, mit mindestens einem erzielten Wert von $> 80\%$ (*Feng(mod)*, *Lewandowska*, *Poh*, *Rapczynski*, *Wang*), ist eine Verbesserung der Herzratenschätzung bei Reduzierung der nicht-Haut Pixel zu erkennen. So profitieren die auf statistischen Methoden aufbauenden Verfahren, wie *Lewandowska* (PCA) und *Poh* (ICA), stark von der Verwendung einer genaueren ROI. Die IEC-Genauigkeiten steigen bei den beiden Ansätzen, auf der *BioVid*, um etwa 13% bis 31% und der *MMSE-HR*, um etwa 4% bis 17%.

Es sind zudem große Unterschiede zwischen den Ergebnissen auf unterschiedlichen Datenbanken zu erkennen. In einigen Fällen konnten wahrscheinlich fehlende Signalverarbeitungs- und Korrekturschritte einiger Algorithmen nicht von einer genaueren ROI ausgeglichen werden. So haben die *Feng(mod)* und *Lewandowska* Ansätze auf der *MMSE-HR* Datenbank mehr als 10% niedrigere IEC-Genauigkeiten erreicht, als auf den bewegungsärmeren *BioVid* Daten. Die anderen Algorithmen (*Poh*, *Rapczynski*, *Wang*) erreichten eine stabilere und von den Daten und ROIs unabhängig hohe IEC-Genauigkeiten. Die parameterfreie gewichtete Hautdetektion erreichte im Mittel leicht bessere Genauigkeitswerte, als die *Forehead* ROI. Für einzelne Algorithmen-Datenbanken Kombinationen erzielte die *Forehead* ROI jedoch leicht bessere Werte.

Ergebnisse Algorithmen Die Algorithmen von *Rapczynski* und *Wang* schnitten in beiden Datensätzen am besten ab und erreichten mit allen ROIs IEC Genauigkeiten von mindestens 84% bis zu 93,3%. Beide waren bei Nutzung verschiedener ROIs vergleichbar stabil, mit Standardabweichungen von 1,0% bis 3,8%. Die Methode von *Wang* zeigte im Vergleich mit den anderen Ansätzen eine hohe Unabhängigkeit (Stdabw. von 1,0%) von der gewählten ROI auf dem *MMSE* Datensatz. Das beste Gesamtergebnis wurde mit *Rapczynski + Haut(gewichtet)* auf der *BioVid* erzielt (93,3%). Der *Rapczynski*-Ansatz wurde auf Grundlage der *BioVid*-Daten entwickelt und publiziert, sodass eine positiver Bias für diesen Datensatz angenommen werden kann. Auf

dem *MMSE-HR* Datensatz erreichte diese Kombination eine Genauigkeit von 91,1%.

4.6 Adaptiver Bandpass

Um den Einfluss des in Kapitel 3.2 beschriebenen dynamischen Bandpasses auf die Genauigkeit der geschätzten Herzrate zu untersuchen, wurde dieser mit einem statischen Bandpass, mit gleichen Cutoff-Frequenzen (0,5 – 4Hz) verglichen. Teile der Ergebnisse dieses Kapitels wurden vorab in [RWA16] publiziert.

Für die Schätzung der Herzfrequenz wurden zwei Ansätze verwendet, ein spektraler auf Basis einer FFT und der auf Peakanalyse basierende IBI-Graph-Ansatz (siehe Kapitel 3.3). Die Werte für die gültigen Peakabstände zueinander d wurden aus den maximalen und minimalen Grenzfrequenzen des Bandpasses abgeleitet und auf 0,3 bis 2 Sekunden begrenzt. Bei der Herzratenschätzung mittels FFT wurde der Spektralanteil mit dem höchsten Peak innerhalb des Bereiches 0,5 – 4Hz als gefundene Herzrate definiert. Für die Peakanalyse wurden zunächst alle Peaks des Signales identifiziert. Aus diesen wird vom IBI-Graph Verfahren eine Untermenge als optimaler Pfad ausgegebenen. Aus den gemittelten zeitlichen Abständen der neuen Peaks wird die Herzrate bestimmt und in BPM ungerechnet. Die Generierung und Auswertung der Grundwahrheiten sind in Kapitel 4.2 beschrieben. Die Güte der Herzratenschätzung wurde für die einzelnen Versuchsreihen mithilfe der IEC Genauigkeit angegeben (siehe Kapitel 4.2.2).

Die Kombinationen aus Bandpass und Herzratenschätzung wurden auf der **BioVid Heat Pain Database** getestet (siehe Kapitel 4.3). Als ROI wurde die gewichtete Hautwahrscheinlichkeit (siehe Kapitel 3.1) verwendet. Aus den RGB-Signalen wurde mittels des **normG** Verfahrens (siehe Kapitel 2.2.3) das PPG-Signal berechnet.

Tabelle 4.8 zeigt die Ergebnisse der Kombinationen aus dynamischem und statischen Bandpass und den zwei Herzratenschätzungen. Der dynamische Bandpass hat im Vergleich zu dem statischen auf die spektrale Herzratenschätzung kaum einen Einfluss. Die IEC-Genauigkeit (siehe Kapitel

4.2.2) ist bei der Verwendung des statischen Bandpasses um 1% größer, während die der mittlere Fehler μ von 3,40 auf 2,77 BPM und die Standardabweichung σ von 10,78 auf 9,72 BPM sinkt. Bei der Verwendung einer Peakbasierten Methode wie dem IBI-Graph Verfahren führt der dynamische Bandpass zu einer deutlichen Steigerung der Genauigkeit, welche von 23,2% auf 93,3% steigt und die Erkennungsrate des spektralen Ansatz übertrifft. Auch der mittlere Fehler und dessen Standardabweichung sinken deutlich. Dies lässt sich auf die starke Reduzierung der Peaks in den dynamisch gefilterten PPG-Signalen zurückführen, welche die Bestimmung der *korrekten* Herzschläge erleichtert. Da für die spektrale Herzratenschätzung die Anzahl und Position der Peaks keine direkte Rolle spielt, kann dieser Ansatz auch in einem größeren Frequenzbereich gute Ergebnisse liefern.

Tabelle 4.8: Einfluss des dynamischen und statischen Bandpasses auf zwei spektrale (FFT) und auf Peakanalyse basierende (IBI-Graph) Herzratenschätzung.

Methode	IBI-Graph		FFT	
	dyn.	statisch	dyn.	statisch
Bandpass				
IEC-Genauigkeit in %	93,3	23,2	86,3	87,3
μ Fehler in BPM	1,04	-14,83	3,40	2,77
σ Fehler in BPM	5,66	20,32	10,78	9,72

4.7 Pulserkennung durch LSTM

Das in Kapitel 3.4 vorgestellte Model zur Identifikation von Pulsschlägen aus videobasierten Farbsignalen wurde implementiert und an mehreren Datenbanken trainiert und validiert. Es wurden die *BioVid* Datenbank für die Trainingsdaten und die *BioVidEmo*, *BP4D+* und *PURE* Daten für die Auswertung verwendet. Die beiden Teile der *BioVidEmo* (D und E) werden weitergehend zusammengefasst als *BioVidEmo* bezeichnet. Die *PURE* und *BP4D+* Datenbanken hatten aufgrund der kurzen Gesamtlängen zu wenige Daten für das Training eines erfolgreichen Modells. Die Daten wurden in Sequenzen von 10 Sekunden Länge mit insgesamt 250 Bildern aufgeteilt. Dabei wurden die Zeitfenster jeweils um 1 Sekunden versetzt. Die Gesichtspunkte wurden wie in Kapitel 4.1 beschrieben gefunden und die mittleren

RGB-Werte des Gesichtes mittels der gewichteten Haut-ROI berechnet (siehe 3.1).

Modellparameter Das Modell wurde für die Videosequenzen von 10 Sekunden Länge und 25 FPS implementiert. Dabei wurden sowohl die RGB, der normierte Grünkanal und das mit dem CHROM Verfahren erzeugte Signal als Eingangssequenzen verwendet. Für die Schichten wurden mehrere Netze mit unterschiedlich vielen Neuronen trainiert ($n = 8, 16, 32, 48, 64, 96, 128$). Die Batchgröße während des Trainings betrug 1024, die Kernelgröße der Convolution-Schichten $k = [5, 1]$ und die MaxPooling Parameter $p = 3$ und $s = 2$. Die Anzahl der Neuronen in den Dense-Schichten am Ende des Netzes betragen $n_x = 200$ und $n_y = 100$.

Training Für das Training wurden 10% der *BioVid*-Sequenzen zufällig als Validierungsset definiert, sowie der *Adam-Optimizer* und *binary chrossentropy* als *Loss* Funktion verwendet. Die *Learning Rate* wurde zu Trainingsbeginn auf 0,001 gesetzt und halbiert, wenn der Validierungsloss über 2 Epochen kein neues Minimum erreicht hat. Das Training wurde abgebrochen, wenn über 5 Epochen kein neues Minimum des Validierungslosses erreicht wurde. Es wurden maximal 100 Epochen trainiert. Zusätzlich zu den in Abbildung 3.8 dargestellten Schichten wurden während des Trainings zwischen alle Schichten *Dropout*-Schichten mit 25% Ausfallquote hinzugefügt. Aufgrund der asymmetrischen Verteilung der Zielsignale (ca. 10-15 Peaks in 250 Bildern) wurden zudem die Modelle mit unterschiedlichen Gewichtungsfaktoren ($w = 1, 3, 8, 16, 24$) trainiert. Dabei wurden die Zeitpunkte der Pulsschläge bei der Fehlerberechnung während des Trainings stärker gewichtet. Zur Evaluierung wurde das Modell mit dem niedrigsten Validierungsloss während des Trainings weiterverwendet.

Fehlerberechnung Für die Fehlerberechnung wurden die Peaks der Ausgangssequenz mit der *SciPy find_peaks* Funktion bestimmt, welche einen minimalen Abstand von 12 Bildern haben und eine Höhe von minimal 0,3 erreichen. Sequenzen welche weniger als 5 Peaks hatten und somit unter der minimalen Herzrate von 30 BPM lagen, wurden automatisch als fehlerhaft

4 Experimentelle Ergebnisse

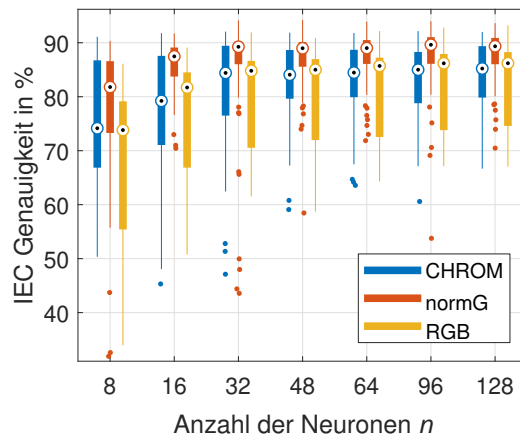


Abbildung 4.27: Boxplot der IEC Genauigkeit in Abhängigkeit der Neuronen n .

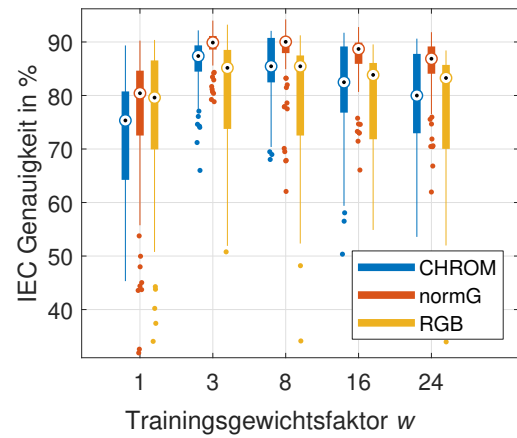


Abbildung 4.28: Boxplot der IEC Genauigkeit in Abhängigkeit der Trainingsgewichtsfaktor w .

klassifiziert. Aus den IBIs der Pulse der Ausgangssequenz und dem korrespondierenden Zielvektor wurden sowohl die geschätzte Herzrate als auch die Grundwahrheit bestimmt und die Fehler für die Sequenz berechnet.

Ergebnisse Das Modell wurden jeweils fünfmal für jede Parameterkombination trainiert und ausgewertet. Dies resultierte in 525 trainierten Netzen und insgesamt 1575 Validierungen auf den drei Testdatenbanken. Die Abbildung 4.27 zeigt die Verteilung der IEC Genauigkeit der Validierungen als Boxplot, in Abhängigkeit der verwendeten Neuronen n . Die Genauigkeit nimmt bei einer Erhöhung der Neuronen zu. Die Netze mit *normG* als Input erreichen dabei die besten Resultate. Die Abbildung 4.28 zeigt analog die Verteilung der IEC Genauigkeit in Abhängigkeit des verwendeten Trainingsgewichtsfaktors w . Die Nutzung des Gewichtsfaktors ($w > 1$) zeigt eine deutliche Verbesserung der IEC Genauigkeit auf den Validierungsdaten. Die Tabelle 4.9 zeigt die IEC Genauigkeit und die mittleren Messfehler für die unterschiedlichen Datenbanken für die besten Trainingsgewichte und Neuronenanzahlen für verschiedene PPG-Eingangssignale. Die besten IEC Einzelergebnisse sind in Tabelle 4.10 dargestellt und lagen bei der *PURE* Datenbank bei 93,4%, der *BioVidEmo* bei 93,2% und bei der *BP4D+* bei 92,1%.

Tabelle 4.9: Mittelwerte und Standardabweichungen der IEC Genauigkeit und des Messfehlers, über alle trainierten LSTM Modelle mit $n \geq 16$ und $3 \geq w \geq 8$ für verschiedene Eingangssignale und Datenbanken.

Trainingsdatenbank	IEC in %			Fehler in BPM		
	CHROM	normG	RGB	CHROM	normG	RGB
BioVidEmo	86.3 ± 3.7	89.5 ± 2.5	90.1 ± 1.5	1.00 ± 7.90	0.79 ± 7.23	1.05 ± 7.33
BP4D+	90.8 ± 1.1	90.2 ± 0.9	86.6 ± 1.1	0.06 ± 7.31	0.39 ± 7.74	0.40 ± 8.94
PURE	84.7 ± 2.3	91.4 ± 1.9	72.0 ± 2.6	-0.52 ± 12.40	-0.86 ± 8.83	3.73 ± 18.98

Tabelle 4.10: IEC Genauigkeit und mittlere Fehler μ und Standardabweichungen σ der zehn LSTM Modelle mit den besten IEC Raten (mit Anzahl der Neuronen n und dem Trainingsgewichtsfaktors w) und gleicher Test- und Trainingsdatenbank.

Trainingsdatenbank	PPG	n	w	IEC in %	$\mu \pm \sigma$ in BPM
PURE	normG	64	3	93,4	$-1,73 \pm 9,69$
BioVidEmo	RGB	128	3	93,2	$0,20 \pm 5,48$
BioVidEmo	normG	96	3	93,2	$-0,13 \pm 5,59$
PURE	normG	48	3	92,9	$-1,72 \pm 9,34$
PURE	normG	32	8	92,3	$0,41 \pm 7,06$
BP4D+	CHROM	32	8	92,1	$1,08 \pm 6,83$
BP4D+	CHROM	32	3	92,0	$-0,26 \pm 6,77$
BioVidEmo	RGB	96	3	92,0	$-0,01 \pm 6,59$
PURE	normG	128	8	91,9	$0,17 \pm 6,87$
BP4D+	CHROM	48	8	91,8	$0,99 \pm 7,21$

4.8 Atmung

Der in Kapitel 3.5 beschriebene Algorithmus wurde in Kombination mit mehreren Pre- und Post-Processing-Schritten getestet und die Ergebnisse mit vier anderen Algorithmen aus der Literatur auf zwei verfügbaren Datenbanken verglichen. Dies war nach unserem Kenntnisstand die erste umfassende Untersuchung und Vergleich der bisher entwickelten Methoden zur Atemratenschätzung auf einer größeren Datenmenge. Die Ergebnisse dieses Kapitels wurden in den Arbeiten [FRA20] und [FRA21] publiziert.

4.8.1 PPG-Signal Generierung

Region of Interest

In Kapitel 4.5.4 wurde umfassend der Einfluss der ROI auf die kontaktlose Herzratenschätzung getestet. Die beiden ROIs (*Forehead* und *Haut(gewichtet)*), mit den besten Ergebnissen, wurden für die weitere Schätzung der Atemrate verwendet, da eine genauere Herzratenschätzung auch zu einer besseren Ableitung des Pulssignals und damit auch zu einer genaueren Atemratenschätzung führen sollte. Die verwendeten ROIs sind in den Kapiteln 2.2.2 & 4.1 ausführlich beschrieben. Zusätzlich wird der von Van Gastel, Stuijk und De Haan [VSD16] vorgeschlagene Algorithmus verwendet, welche als ROI das Gesicht in 30 Unterregionen unterteilt und in Kapitel 2.2.2 beschrieben ist.

Signalverarbeitung

Für die Experimente wurde die Fensterlänge auf 30 Sekunden festgelegt. Zum Testen und Bewerten wurde eine Länge von 60 Sekunden als Benchmark hinzugefügt, um die Leistung der Algorithmen bei längeren Fenstern zu analysieren. Die Schrittweite wurde für beide Fenstergrößen auf zehn Sekunden festgelegt. Damit ist das Kriterium erfüllt, dass bei einer minimalen Atemrate von sechs Schlägen pro Minute mindestens einer im Fenster

vollständig erfasst wird. Die folgenden, für den Vergleich verwendeten Signalextraktionsverfahren, sind in Kapitel 2.2.3 ausführlich beschrieben.

Poh, McDuff und Picard [PMP11] verwendeten eine ICA (siehe Kapitel 2.2.3) basierend auf dem *Joint Approximation Diagonalization of Eigen-matrices* (JADE) Verfahren [Car99] als Grundlage für die Atemratenschätzung.

Sun, Hu, Azorin-Peris, Greenwald, Chambers und Zhu [Sun+11] verwendete jeden RGB-Kanal als Inputkanal und nutzte eine SCICA (siehe Kapitel 2.2.3) zur Ableitung der Atemrate.

Van Gastel, Stuijk und De Haan [VSD16] berechneten die Pixeldifferenzen von 30 Unterregionen der ROI, um die Gewichte der Linearkombination für das Pulssignal abzuleiten. Diese Gewichte werden dann auf die Pixeldifferenzen angewendet, die ausschließlich Atemfrequenzen enthalten und somit das Atmungssignal erzeugen. Für die Berechnung der Gewichte wurde in unserer Implementierung die chrominanzbasierte Methode [dJ13] verwendet, da diese in der Originalarbeit für Videos die besten Ergebnisse im sichtbaren Lichtspektrum erzielte.

Sanyal und Nundy [SN18] verwendeten die Stirn als ROI und transformierten die Pixel-RGB-Werte in den HSV-Farbraum (siehe Kapitel 2.2.3), wobei nur der Farbtonkanal und die Pixel mit Werten im Bereich $[0, 0, 1]$ verwendet wurden, um das PPG-Signal zu mitteln. Das Signal wird im Anschluss bandpassgefiltert und daraus die Atemrate bestimmt.

4.8.2 Datenbanken

Für eine bessere Reproduzierbarkeit wurde die videobasierte Atemratemessung auf zwei verfügbare Datenbanken der **BP4D+** und der **AtemDB** getestet. Weitere Details zu den Datenbanken sind in Kapitel 4.3 zu finden.

BP4D+ Die erste verwendete Datenbank ist die **BP4D+** von Zhang u. a. [Zha+16]. Da sich die Probanden uneingeschränkt bewegen und sprechen, ist die Datenbasis für die berührungslose Überwachung der Atemrate anspruchsvoll. Dies liegt auch daran, dass die Datenbank nur natürliche und keine künstlichen Atemzyklen enthält.

Da die Signale der Atemgrundwahrheiten zeitweise sehr stark von Artefakten und Störungen beeinflusst waren, konnten nicht alle Signalfenster verwendet werden (siehe Abb. 4.30). So kann zum Beispiel ein zu locker angelegter Brustgurt, während einer Bewegung des Oberkörpers zu einer starken Varianz oder einem Verlust des Signals führen. Daher wurde ein Verfahren für die Bewertung der zugrundeliegenden Atemgrundwahrheiten erarbeitet.

Zuerst wurden die Signale mit einem Butterworth-Bandpass mit Grenzfrequenzen von 0,1 Hz und 0,5 Hz gefiltert und anschließend auf einen Wertebereich von [-1 bis 1] normiert. Die Minima und Maxima wurden mit einem Peak-Detektor bestimmt. Signale wurden verworfen, wenn die Standardabweichung der Differenz zwischen den Atemintervallen höher als 1 Sekunde oder die Standardabweichung der Signalhöhen der Minima eines Signals größer als 0,2 war. Zusätzlich wurden alle Videos verworfen, welche kürzer als die minimale Fensterlänge von 30 Sekunden waren. Die verbleibenden 269 Signalfenster wurden für die Validierung der Algorithmen verwendet.

Nachfolgend ist ein je Beispiel für die respiratorischen Signalklassen dargestellt. Abb. 4.29 zeigt eine Atemgrundwahrheit mit guter Qualität und Abb. 4.30 zeigt ein verworfenes Signal, bei dem die Grundwahrheit nicht bestimmt werden kann.

AtemDB Als zweite Datenbank wurde die **AtemDB** Datenbank verwendet, welche speziell für die Forschung an videobasierter Atemratenschätzung erstellt wurde. Diese enthält sowohl spontane Atmung der Probanden als auch vorgegebene Atemmuster von 10, 15 und 20 BPM.

Im Gegensatz zur **BP4D+** wurden die Teilnehmer gebeten, sich nicht stark zu bewegen, wodurch die Datenbank gut als Benchmark verwendet werden kann. Die Grundwahrheiten für die Herz- und Atemraten wurden analog zur **BP4D+** verarbeitet.

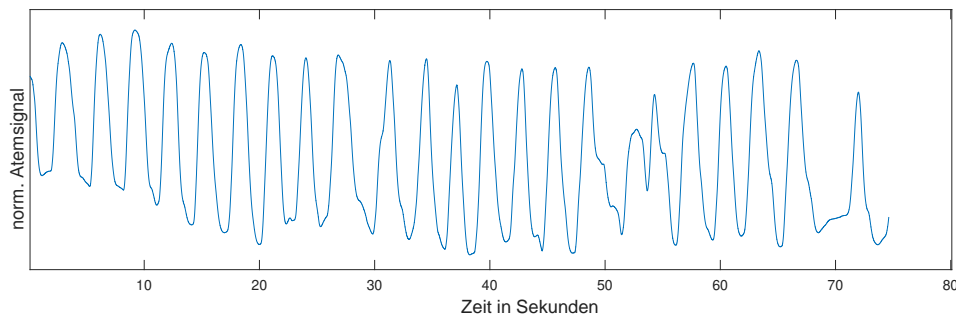


Abbildung 4.29: Beispiel für ein normiertes Atemsignal (dimensionslos) der BP4D+, bei dem die Grundwahrheit bestimmt werden kann.

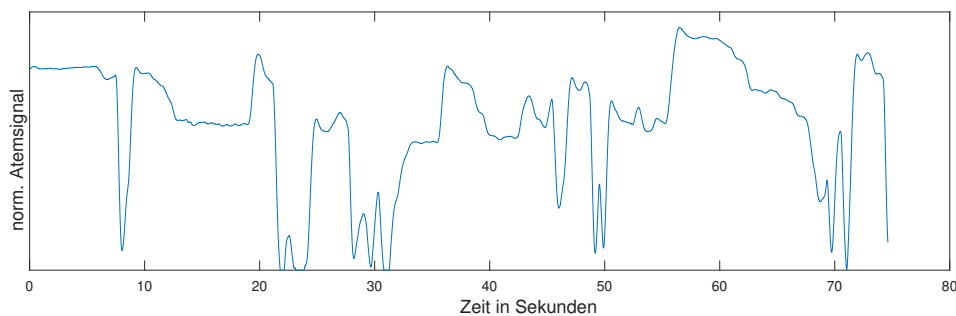


Abbildung 4.30: Beispiel für ein verworfenes normiertes Atemsignal (dimensionslos) der BP4D+, bei dem die Grundwahrheit nicht ermittelt werden konnte.

4.8.3 Ergebnisse

Die verschiedenen Modulationen wurden in zwei Kombinationen entwickelt und getestet. **FuseMod** wurde zunächst in [FRA20] vorgestellt und nutzt die in Kapitel 3.5.2 vorgestellten AM, BM und FM Modulationen. Dieses Verfahren wurde in [FRA21], als **FuseMod2** bezeichnet, weiterentwickelt und um die RF-Modulationen erweitert.

Ein Überblick über die Ergebnisse des **FuseMod**-Algorithmus ist in den Tabellen 4.11 - 4.14 beschrieben. Für jede Datenbank und ROI Kombination sind die *DR* und mittlerer Fehler angegeben. Die Ergebnisse werden zudem auf die verschiedenen PPG-Signalverarbeitungsmethoden und Mittlungsmethoden aufgeschlüsselt. Die Fensterlänge ist dabei auf 30 Sekunden beschränkt.

4 Experimentelle Ergebnisse

Tabelle 4.11: Ergebnisse für **FuseMod (Haut(gewichtet)-ROI, 30s-Fenster)** unter Verwendung verschiedener PPG- und Mittlungsmethoden auf der **BP4D+**.

PPG Signal	Mittelwert				Median			
	G	hue	CHROM	normG	G	hue	CHROM	normG
DR in %	53.11	60.81	63.63	61.06	62.97	70.84	72.16	71.00
μ in BPM	-0.52	-0.39	-0.22	-0.48	-0.56	-0.44	-0.26	-0.49
σ in BPM	3.90	3.38	3.35	3.47	4.23	3.62	3.52	3.78

Tabelle 4.12: Ergebnisse für **FuseMod (Forehead-ROI, 30s-Fenster)** unter Verwendung verschiedener PPG- und Mittlungsmethoden auf der **BP4D+**.

PPG Signal	Mittelwert				Median			
	G	hue	CHROM	normG	G	hue	CHROM	normG
DR in %	46.39	59.67	56.10	52.86	54.19	67.22	67.88	61.99
μ in BPM	-0.69	-0.53	-0.34	-0.72	-0.65	-0.40	-0.28	-0.63
σ in BPM	4.01	3.57	3.65	3.88	4.44	3.83	3.93	4.35

Tabelle 4.13: Ergebnisse für **FuseMod (Haut(gewichtet)-ROI, 30s-Fenster)** mit verschiedenen PPG- und Mittlungsmethoden auf der **AtemDB**.

PPG Signal	Mittelwert				Median			
	G	hue	CHROM	normG	G	hue	CHROM	normG
DR in %	51.21	56.69	65.86	67.39	63.18	68.41	81.27	78.47
μ in BPM	-1.50	-1.57	-1.13	-1.41	-1.80	-1.69	-1.15	-1.39
σ in BPM	3.74	3.45	2.90	3.06	4.19	3.89	3.05	3.33

Tabelle 4.14: Ergebnisse für **FuseMod (Forehead-ROI, 30s-Fenster)** mit verschiedenen PPG- und Mittlungsmethoden auf der **AtemDB**.

PPG Signal	Mittelwert				Median			
	G	hue	CHROM	normG	G	hue	CHROM	normG
DR in %	44.08	43.82	53.89	51.97	49.68	54.14	63.69	62.55
μ in BPM	-1.48	-1.90	-1.58	-1.65	-1.86	-2.10	-1.78	-1.81
σ in BPM	4.19	4.13	3.91	3.78	4.70	4.64	4.35	4.26

Signalverarbeitung Es wurden verschiedene Ansätze der PPG Signalverarbeitung untersucht. Die Ergebnisse zeigen eine deutliche Verbesserung der Erkennungsleistung für *hue*, *CHROM* und *normG* im Vergleich zum Grün-Kanal. Dies spiegelt sich in einer signifikanten Erhöhung des *DR* von bis zu 19 %, einer Reduzierung des mittleren Fehlers μ und der Standardabweichung σ wider. Dieser Anstieg ist bei der *BP4D+* Datenbank über beide Mittlungsmethoden und ROIs hinweg zu beobachten. Auf der *AtemDB* sinkt die Leistung des *hue* im Vergleich zu *CHROM* und *normG*, wobei *CHROM*, mit Ausnahme der Kombination Forehead/*BP4D+* immer die höchsten Erkennungsraten erreicht (siehe Tab. 4.11 - 4.14).

Von den Mittlungsmethoden erzielte der Median auf beiden Datenbanken bessere Erkennungsraten als der Mittelwert. Dieses Ergebnis kann darauf zurückgeführt werden, dass die einzelnen Modulationen zeitweise starken Schwankungen unterliegen, und durch den Median besser ausgeglichen werden können.

ROIs Bei der Betrachtung der ROI erzielt die gewichtete Hautdetektion deutlich bessere Ergebnisse als die *Forehead* ROI. Der Unterschied zwischen den beiden ROIs ist bei der *AtemDB* deutlicher als bei der *BP4D+*. Bei der Haut-ROI wurden auf der *AtemDB*, welche deutlich weniger Bewegungen enthält als die *BP4D+*, wie zu erwarten in allen Signalkombinationen deutlich bessere Ergebnisse erzielt. Für die *Forehead*-ROI war dieser Effekt nicht zu beobachten. Die *DR* ist in einigen Fällen sogar gesunken. Die Stirnregion der Probanden ist in der *AtemDB* häufiger durch systematische Störungen aufgefallen. So zeigen die Aufnahmen der Probanden Verdeckungen durch Haare, oder Lichtreflexionen auf der Stirn. Darüber hinaus trug eine Testperson ein Kopftuch, was die sichtbare Haut der *Forehead*-ROI ebenfalls reduzierte.

Artefaktreduktion Die Artefaktreduktion ist ein wesentlicher Bestandteil des entwickelten Algorithmus. Die Tabellen 4.15 und 4.16 stellen die Ergebnisse mit und ohne zusätzliche Artefaktreduktion für beide Datenbanken dar.

4 Experimentelle Ergebnisse

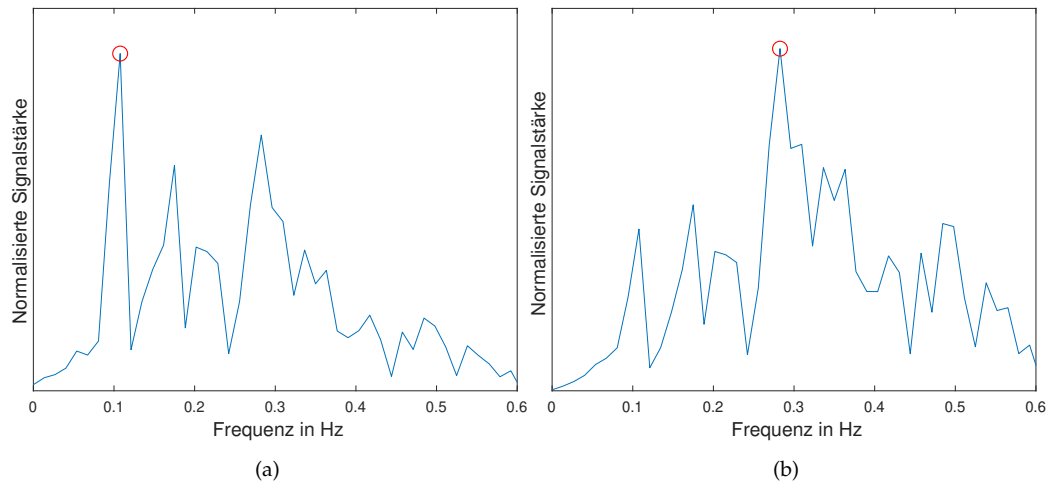


Abbildung 4.31: Vergleich der Frequenzspektren eines AM-Signales (a) ohne und (b) mit Artefaktreduktion mit einer Grundwahrheitsfrequenz von 0,28 Hz.

Insbesondere die *BP4D+* enthält in großem Umfang niederfrequente, durch Bewegung verursachte Störsignale, auch Traube-Hering-Mayer-Wellen genannt. Diese können die gesuchte Grundwahrheitsfrequenz im Spektrum überlagern und nicht gefiltert werden, ohne plausible Atemfrequenzen zu entfernen. Dies ist am hohen mittleren Fehler μ zu erkennen, welcher eine starke negative Tendenz in den niederfrequenten Bereich zeigt (siehe Tabellen 4.15 und 4.16). Zudem zeigen die FDR-Raten, dass ohne die Artefaktreduktion mehr als die Hälfte der untersuchten Fenster auf der *BP4D+* eine dominante Störfrequenz im unteren Bereich des beobachteten Spektrums haben. Durch die Differenzierung des Signals während der Artefaktreduktion kann dieser Anteil auf etwa 2 % reduziert werden, womit die DR um etwa 40 % erhöht und der mittlere Fehler μ deutlich verringert wird. Der Einfluss der Artefaktreduktion auf das Frequenzspektrum ist in Abb. 4.31 am Beispiel einer AM dargestellt. Die Problematik ist analog für BMs und FMs.

Die *AtemDB* enthält weniger bewegungsinduzierte Störfrequenzen, aber auch hier kann der DR je nach PPG-Verfahren zwischen 6 % und 14 % erhöht werden. Die Falscherkennungen im unteren niederfrequenten Bereich liegen bei Verwendung der Artefaktreduktion unter 1 %.

Tabelle 4.15: Ergebnisse mit und ohne Artefaktreduktion für FuseMod (30s-Fenster) auf der BP4D+.

Algorithmus	Artefakt- reduktion	DR in %	μ in BPM	σ in BPM	FDR in %
FuseMod (hue, median)	✓	70.84	-0.44	3.62	2.32
FuseMod (hue, median)	✗	29.66	-7.10	5.63	52.94
FuseMod (CHROM, median)	✓	72.16	-0.26	3.52	1.16
FuseMod (CHROM, median)	✗	31.40	-6.93	5.65	50.37
FuseMod (normG, median)	✓	71.00	-0.49	3.78	2.57
FuseMod (normG, median)	✗	27.26	-7.44	5.56	54.68

Tabelle 4.16: Ergebnisse mit und ohne Artefaktreduktion für FuseMod (30s-Fenster) auf der AtemDB.

Algorithmus	Artefakt- reduktion	DR in %	μ in BPM	σ in BPM	FDR in %
FuseMod (hue, median)	✓	68.41	-1.69	3.89	0.64
FuseMod (hue, median)	✗	62.17	-2.93	4.37	21.53
FuseMod (CHROM, median)	✓	81.27	-1.15	3.05	0.76
FuseMod (CHROM, median)	✗	67.13	-2.76	4.36	12.99
FuseMod (normG, median)	✓	78.47	-1.39	3.33	0.89
FuseMod (normG, median)	✗	69.17	-2.40	4.25	15.16

4 Experimentelle Ergebnisse

Tabelle 4.17: Ergebnisse der **FuseMod**-Implementierungen (30 und 60 Sekunden Fenster) und den **Vergleichsalgorithmen** (30 Sekunden und vorgeschlagene Fensterlänge aus der jeweiligen Originalpublikation) auf der **BP4D+**.

Algorithmus	Fenster	ROI	DR in %	μ in BPM	σ in BPM
FuseMod (hue, median)	30 s	Haut	70.84	-0.44	3.62
FuseMod (hue, median)	60 s	Haut	71.40	-0.20	3.07
FuseMod (CHROM, median)	30 s	Haut	72.16	-0.26	3.52
FuseMod (CHROM, median)	60 s	Haut	71.78	-0.25	2.96
FuseMod (normG, median)	30 s	Haut	71.00	-0.49	3.78
FuseMod (normG, median)	60 s	Haut	68.22	-0.03	3.12
Poh [PMP11]	30 s	Haut	37.37	-3.78	5.63
Poh [PMP11]	60 s	Haut	37.01	-4.10	5.53
Poh [PMP11]	30 s	Forehead	32.45	-3.66	5.85
Poh [PMP11]	60 s	Forehead	36.26	-3.61	5.43
Sun [Sun+11]	30 s	Haut	20.30	5.39	5.56
Sun [Sun+11]	34 s	Haut	17.77	5.32	5.58
Sun [Sun+11]	30 s	Forehead	23.07	5.27	5.56
Sun [Sun+11]	34 s	Forehead	17.18	5.31	5.75
VanGastel [VSD16]	30 s	VanGastel	22.66	0.19	9.29
VanGastel [VSD16]	8 s	VanGastel	31.27	1.00	9.51
Sanyal [SN18]	30 s	Haut	21.16	-6.33	4.36
Sanyal [SN18]	20 s	Haut	17.99	-5.35	5.00
Sanyal [SN18]	30 s	Forehead	19.09	-6.56	4.32
Sanyal [SN18]	20 s	Forehead	12.56	-5.91	4.76

4.8.4 Vergleich

Um die entwickelte Methode zur Atemratenerkennung zu validieren, wurden vier andere Algorithmen aus der Literatur nach implementiert. Für eine bessere Vergleichbarkeit wurden, die in der entsprechenden Originalarbeit vollgeschlagene Fensterlänge und zusätzlich eine einheitliche Fensterlänge von 30 Sekunden verwendet. Zudem wurden die beiden **FuseMod** Ansätze mit der längeren 60 Sekunden Fensterlänge aus [PMP11] validiert. Für alle Fenster wurde eine Schrittweite von 10 Sekunden verwendet.

Die Ergebnisse aller Algorithmen sind in den Tabelle 4.17 und 4.18 dargestellt. Das **FuseMod** Verfahren erreicht auf der **BP4D+** *DR* zwischen 68,2% bis 72.2%. Dabei bringt eine Verlängerung des Fensters keine signifikan-

Tabelle 4.18: Ergebnisse der **FuseMod**-Implementierungen (30 und 60 Sekunden Fenster) und den **Vergleichsalgorithmen** (30 Sekunden und vorgeschlagene Fensterlänge aus der jeweiligen Originalpublikation) auf der **AtemDB**.

Algorithmus	Fenster	ROI	DR in %	μ in BPM	σ in BPM
FuseMod (hue, median)	30 s	Haut	68.41	-1.69	3.89
FuseMod (hue, median)	60 s	Haut	78.47	-1.08	3.07
FuseMod (CHROM, median)	30 s	Haut	81.27	-1.15	3.05
FuseMod (CHROM, median)	60 s	Haut	87.68	-0.65	2.33
FuseMod (normG, median)	30 s	Haut	78.47	-1.39	3.33
FuseMod (normG, median)	60 s	Haut	84.87	-0.72	2.46
FuseModV2 (normG, median)	60 s	Haut	90.09	-0.56	1.96
FuseModV2 (CHROM, median)	60 s	Haut	90.09	-0.53	2.03
Poh [PMP11]	30 s	Haut	55.54	0.85	4.89
Poh [PMP11]	60 s	Haut	59.13	0.88	4.42
Poh [PMP11]	30 s	Forehead	47.01	1.18	5.35
Poh [PMP11]	60 s	Forehead	50.55	1.09	5.03
Sun [Sun+11]	30 s	Haut	18.22	4.15	8.90
Sun [Sun+11]	34 s	Haut	16.91	3.63	8.66
Sun [Sun+11]	30 s	Forehead	25.86	3.33	8.83
Sun [Sun+11]	34 s	Forehead	16.51	3.53	8.98
VanGastel [VSD16]	30 s	VanGastel	35.16	5.15	9.01
VanGastel [VSD16]	8 s	VanGastel	32.27	6.34	9.26
Sanyal [SN18]	30 s	Haut	52.99	-2.31	4.15
Sanyal [SN18]	20 s	Haut	41.66	0.21	3.99
Sanyal [SN18]	30 s	Forehead	53.12	-2.03	4.25
Sanyal [SN18]	20 s	Forehead	34.09	-0.06	4.65

te Verbesserung. Auf der *AtemDB* erreicht **FuseMod** eine *DR* von bis zu 84.9% und **FuseModV2** 90,1%. Der *FuseMod*-Algorithmus steigert seine *DR* auf der *AtemDB* bei Verwendung von 60-Sekunden-Fenstern auf allen PPG-Signalen.

Die Vergleichsalgorithmen erzielten auf der herausfordernden *BP4D+* deutlich schlechtere Ergebnisse, als das vorgestellte **FuseMod** Verfahren. So erreichte Poh, McDuff und Picard [PMP11] (*BP4D+*, 30s, Haut) als bester Vergleichsalgorithmus eine *DR* von 37.37%. Alle Algorithmen, mit Ausnahme von Sun, Hu, Azorin-Peris, Greenwald, Chambers und Zhu [Sun+11], verbessern ihre *DR* auf der *AtemDB*. Insbesondere Sanyal [SN18] erzielt eine um mehr als 30 % höhere *DR* und erreicht fast den besten Vergleichsalgorithmus von Poh, McDuff und Picard [PMP11]. Dies lässt sich wahrscheinlich auf die deutlich geringeren Bewegungen im Datensatz zurückführen. Auch bei den Vergleichsalgorithmen erzielt die Haut-ROI bessere Ergebnisse als die *Forehead*-ROI. Bei den Vergleichsalgorithmen verbessert ein längeres Fenster die *DR* und reduziert den mittleren Fehler. Alleine bei dem Algorithmus von Sun, Hu, Azorin-Peris, Greenwald, Chambers und Zhu [Sun+11] verringert sich die Genauigkeit bei einer Änderung des Fensters von 30 auf 34 Sekunden. Jedoch sind die zugrundeliegenden Genauigkeiten bereits niedrig (18,2% und 25,9%) und weisen eine hohe Standardabweichung des Fehlers auf.

4.9 Lebenderkennung

Das in Kapitel 3.6 vorgestellte Modell zur Unterscheidung von Masken und lebendem Gewebe wurde auf der *HKBU 3D Mask Attack with Real World Variations Database (HKBU-MARs)* [Liu+16] Version 1+ validiert. Teile der Ergebnisse dieses Kapitels wurden vorab in [Rap+18b] und [RLA19] publiziert.

Die Datenbank besteht aus 170 Videos von 12 Personen. Von jedem Probanden stehen 10 Videos ohne Masken und 5 mit verschiedenen Masken zur Verfügung. Die Videos eines Probanden ohne Maske, waren aus Gründen des Datenschutzes, nicht verfügbar. Alle Videos wurden mit einer *Logitech*



Abbildung 4.32: Beispiele der HKBU Datenbank. Probanden ohne und mit verschiedenen Masken.

C920 Webcam in einer Auflösung von 1280×720 Pixeln unter der vorhandenen Raumbelichtung, mit Frontalansicht und neutralem Gesichtsausdruck aufgenommen. Die Videos sind in *MJPEG* encodiert. Dabei ist zu beachten, dass das verwendete Kompressionsverfahren nicht verlustfrei ist und zu Informationsverlusten führt, wie im Abschnitt zu den Videoeigenschaften 4.4 beschrieben wurde.

Mit den aus den PPG-Signalen berechneten Peak Features (siehe Kapitel 3.6.2) wurden verschiedene Klassifikatoren mithilfe von maschinellem Lernen trainiert. Dazu wurden die Videos des **HKBU** Datensatzes in zwei Klassen (Gesicht oder Maske) kategorisiert.

Die Daten sind mit einem 5-fachen Crossvalidation Verfahren trainiert worden. Die genutzte Datenbank liegt nur in RGB vor, daher wird für die quantitative Validierung der Unterscheidung von Masken und menschlichen Gesichtern nur durch die Vitalparameterschätzung untersucht. Die Analyse der 2,5D-Gesichtserkennung wurde zunächst auf einem vorläufigen Testdatensatz quantitativ als Proof-of-Concept durchgeführt.

Als Klassifikatoren wurden Support-Vector-Maschinen (SVM), k-Nearest-Neighbor (KNN) und Entscheidungsbäume verwendet. Die Klassifikatoren wurden in *Matlab* implementiert, mit unterschiedlichen Parametern getestet und miteinander verglichen. Bei den SVM wurde die Art des Kernels (Linear, Quadratisch, Kubisch, Gaussian), die Kernel-Scale im Fall des Gaussian Kernels (0.75, 3, 12) und die Box-Constraint (0.1, 1, 10) Parameter variiert.

4 Experimentelle Ergebnisse

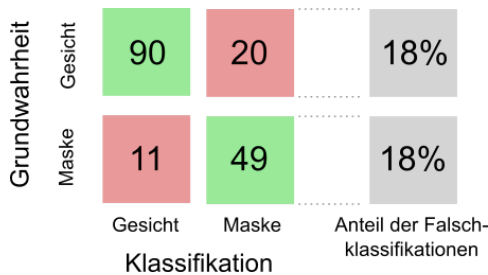


Abbildung 4.33: Konfusionsmatrix des Tree-Klassifikators

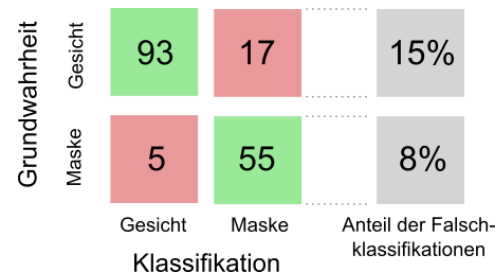


Abbildung 4.34: Konfusionsmatrix des SVM-Klassifikators

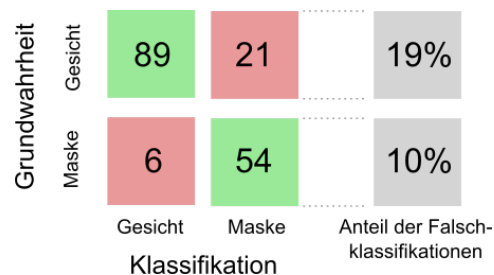


Abbildung 4.35: Konfusionsmatrix des KNN-Klassifikators

Für die Entscheidungsbäume wurde die maximale Anzahl an Knoten im Baum (4, 10, 100) variiert. Für die KNN Klassifikatoren wurden unterschiedliche Distanzmetriken (Euklidisch, Quadratisch Invers, Cosine) und Anzahl der betrachteten Nachbarn (1, 10, 100) untersucht, mit gewichteten und ungewichteten Distanzmetriken.

Für die unterschiedlichen Klassifikatoren werden im Folgenden diejenigen mit den besten Ergebnissen näher beschrieben. Die Abbildungen 4.33, 4.34 und 4.35 zeigen die Konfusionsmatrixen der drei besten Klassifikatoren in ihren jeweiligen Gruppen. Die Matrizen stellen die korrekten und inkorrekten Zuordnungen der Videos zu den Klassen *Gesicht* und *Maske* dar. Zusätzlich wurden für beide Gruppen der Anteil der falsch klassifizierten Videos angegeben. Der beste Baum (Knoten: 4) zeigte eine leicht niedrigere Gesamtgenauigkeit (81,8%) wobei beide Klassen gleichmäßig gut erkannt wurden. Bei der Erhöhung der Knoten fand keine Änderungen bei der Gesamtgenauigkeit statt, jedoch eine Verschiebung der Fehler zu einer höheren Anzahl der Falsch-Positiven Klassifikationen der Masken als Gesichter statt,

was für die geplante Anwendung einen Nachteil darstellt. Der KNN Klassifikator (Nachbarn: 10, Cosinus Distanz, ungewichtet) lieferte eine etwas bessere Gesamterkennungsrate (84,1%). Die beste SVM (Gaussian Kernel, Scale 12) hatte insgesamt die höchste Klassifikationsrate (87,1%). Beim SVM und dem KNN ist zudem eine leichte Tendenz zu sehen, die Masken mit höherer Genauigkeit korrekt zu klassifizieren als menschliche Gesichter.

4.10 Messung im MRT

Neben den multispektralen Versuchen zur Isolierung günstiger Wellenlängen für die Messung des PPG-Signales, wurde die Verwendung einer monochromatischen rauscharmen Kamera für die Vitalparameterschätzung in einem MRT Gerät untersucht, in welchem üblicherweise, bedingt durch die hohen Feldstärken, speziell abgeschirmte Messtechnik eingesetzt werden muss.

Versuchsaufbau und Datenaufnahme

Abbildung 4.36 zeigt den verwendeten Versuchsaufbau. Der Proband lag bei der Messung im MRT. Der Kopf wurde mithilfe einer üblichen, mit Schaumstoff ausgelegten, Fixierung stabilisiert. Diese wurde mit einem verstellbaren Spiegel ausgestattet, über den Patienten während einer Untersuchung Bilder oder Videos gezeigt werden können. Die Grundwahrheit des Pulses wurde über einen an das MRT angeschlossenen Fingersensor gemessen. Die LED-Beleuchtung innerhalb der MRT-Röhre und die übliche Raumbeleuchtung waren während der Messungen eingeschaltet.

Aufgrund der großen Feldstärken, welche durch ein MRT-Gerät erzeugt werden, können elektrische Geräte nicht in der Nähe eines MRTs verwendet werden. Daher wurde eine in der Wand vorhandene Öffnung an der Kopfseite des MRTs für die Aufnahmen verwendet. Aufgrund der Höhe der Öffnung wurde das Gesicht des Probanden über ein Spiegelsystem, bestehend aus zwei Spiegel zwischen dem MRT und der Öffnung und dem Spiegel an der Fixierung des Kopfes, gefilmt. Für die Synchronisierung der Daten wurde ein Triggersignal mit 25Hz von einer Triggerbox generiert und

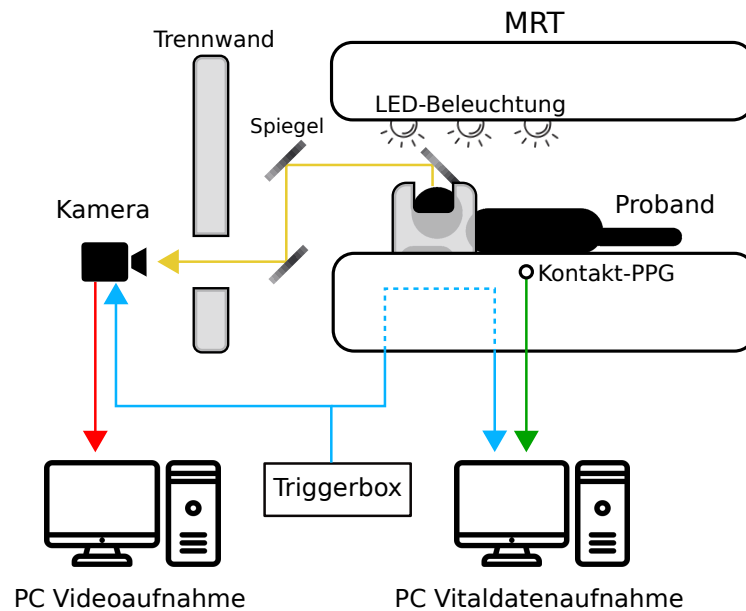


Abbildung 4.36: Messaufbau für die Herzratenschätzung im MRT. (Blau: Triggersignal, Rot: Videodaten, Grün: Vitaldaten)

sowohl an die Kamera als auch an den externen Anschluss des MRTs angeschlossen. Das Triggersignal wurde zusammen mit den Vitalparameterdaten auf einem an das MRT angeschlossenen PC aufgezeichnet. Die Videodaten wurden auf einem an die Kamera angeschlossenen PC aufgezeichnet.

Es wurde eine monochromatische rausch-arme Kamera (IDS UI-3080CP Rev. 2 mit Fujinon HF8XA-5M Objektiv) verwendet und die Videos in der Auflösung von 1000 x 1000 Pixeln aufgenommen. Die Videos wurden während der Aufnahme mittels des HuffYUV-Codec verlustfrei gespeichert und nachträglich mit dem x264 Codec und einem $CRF = 0$ zur Archivierung enkodiert. Für den Versuch lagen die Probanden in einem MRT vom Typ Siemens MAGNETOM Skyra mit einer Bohrung von 70 cm und einem kurzen Magneten mit einer maximalen Magnetfeldstärke von 3 Tesla. Während der Messungen war das MRT Gerät im Bereitschaftsbetrieb.

Es wurden zehn Messungen von acht Probanden (4 weiblich, 4 männlich) im Alter von 21-31 Jahren durchgeführt. Die Probanden sollten wie bei einer üblichen Untersuchung ruhig im MRT liegen und wurden 5 Minuten

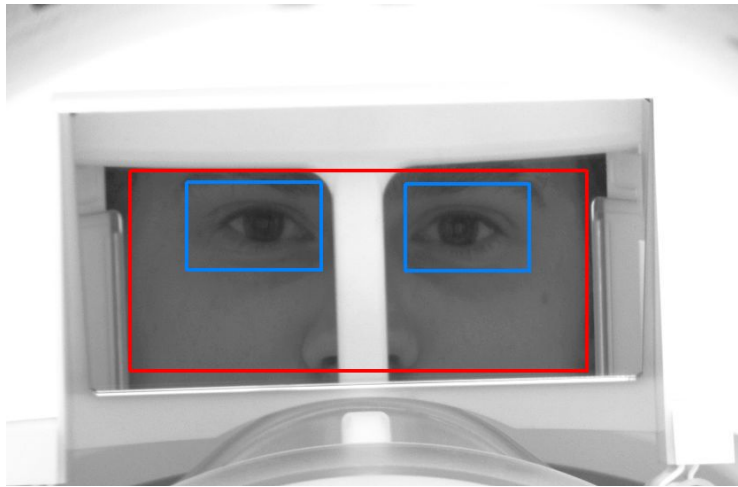


Abbildung 4.37: Ausschnitt eines MRT-Videos (Proband 2) mit Darstellung der ROI (rot) und den Augenbereichen (blau). Helligkeit und Kontrast wurden für diese Darstellung angepasst.

lang aufgenommen. Die Messung einer Probandin konnte nicht ausgewertet werden, da aufgrund des Größenunterschiedes des Fingerpulsoximeters und der Finger der Probandin kein dauerhafter Sensorkontakt zur Haut bestand. Dadurch war über einen Großteil der Messung keine Pulsgrundwahrheit messbar. Die letzten beiden Probanden wurden jeweils einmal mit und ohne der im MRT integrierten LED-Beleuchtung aufgezeichnet.

Datenverarbeitung

Aufgrund des hohen Verdeckungsgrades des Gesichtes und der fehlenden RGB Farbinformationen konnte bei dieser Messung nicht auf die üblichen Methoden zur ROI Bestimmung zurückgegriffen werden. Zusätzlich waren die genauen Kopfpositionen der Probanden aufgrund unterschiedlicher physiologischer Faktoren (Länge Hals, Größe Kopf) für jeden Probanden unterschiedlich. Um dies zu kompensieren, wurde der Spiegel entsprechend neu ausgerichtet. Daher wurde die ROI für jede Messung per Hand definiert. Abbildung 4.37 zeigt ein Beispiel der verwendeten ROI-Bereiche. Im ersten Bild des Videos wurde manuelle ein Rechteck definiert, mit dem Ziel einen möglichst großer Ausschnitt des sichtbaren Gesichtes abzudecken.

4 Experimentelle Ergebnisse

Tabelle 4.19: IEC Genauigkeit, Mittelwert und Std. Abweichung der im MRT durchgeführten Herzratenschätzung.

Proband LED	1 💡	2 💡	3 💡	4 💡	5 💡	6 💡	6 💡	7 💡	7 💡
IEC (in %)	100,0	91,0	100,0	96,9	97,4	92,4	95,7	92,3	96,7
μ (in BPM)	0,05	-2,74	0,05	-0,60	2,44	1,60	0,67	2,04	0,28
σ Fehler (in BPM)	1,24	7,97	0,72	3,99	1,47	2,43	2,50	3,72	3,41

Zusätzlich wurden zwei Bereiche um die Augen definiert, um den Einfluss der Augenbewegungen zu minimieren. Das PPG-Signal wurde für jedes Bild im Video aus dem Mittelwert der Pixel-Farbwerte in der ROI, ohne die Augenbereiche, gebildet.

Die Herzrate wurde einmal pro Sekunde bestimmt. Dazu wurde der adaptive Bandpass und das IBI-Graph Verfahren verwendet (siehe Kapitel 3.3 und 3.2). Der adaptive Bandpass wurde mit einem Zeitfenster von 10 Sekunden und einer Framerate von 25 FPS initialisiert. Für den ersten Zeitschritt wurde ein 30 Sekunden langes und für die folgenden Schritte ein 10 Sekunden langes Zeitfenster ausgewertet.

Das PPG-Signal innerhalb des Zeitfensters wurde bei jeder Auswertung zunächst normiert. Darauf folgend wurde ein Polynom vierten Grades gefittet und vom Signal abgezogen, um niederfrequente Anteile und Signaldrift zu unterdrücken. Das Signal wurde anschließend mit einem adaptiven Bandpass gefiltert. Für die Auswertung wurden zunächst die Maxima mit einer minimalen Amplituden-Prominenz von 0,2 und einem Mindestabstand von 10 Bildern bestimmt. Die gefundenen Maxima wurden mittels des IBI-Graph Verfahren bewertet und aus der gefundenen Abfolge der Maxima, die Herzrate bestimmt.

Ergebnisse

Tabelle 4.19 zeigt die Ergebnisse der im MRT durchgeführten Herzratenschätzung. Bei allen Messungen mit eingeschalteten LEDs lag die IEC Genauigkeit (siehe Kap. 4.2.2), mit einer Ausnahme, bei über 95%. Die Versuchsperson 2 hat während der Messung seinen Kopf der Länge nach in

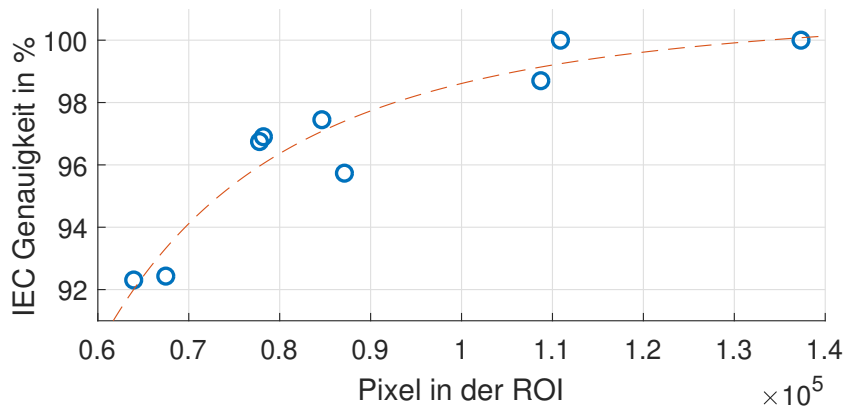


Abbildung 4.38: Die IEC Genauigkeit in Abhängigkeit der Größe der ROI in Pixel (gefittete Potenzreihe 2. Ordnung der Daten in rot).

der Fixierung bewegt, wodurch Artefakte im PPG-Signal entstanden sind auf und dadurch die Messung erschwert haben. Ohne die durch die Bewegungsartefakte beeinflussten Zeitfenster erreicht die Herzratenschätzung bei Proband 2 eine IEC Genauigkeit von 98,7%. Die Messungen mit abgeschalteten LEDs hatten eine geringere IEC Genauigkeit von etwa 3% und einen höheren mittleren Fehler bei der Herzratenschätzung.

Die Varianz in den Körpergrößen der Probanden, der Spiegel- und Augenposition im Bild führte zu unterschiedlich großen ROIs bei der Auswertung. Abbildung 4.38 zeigt die Abhängigkeit der IEC Genauigkeit von der Größe der ROI, mit dem korrigierten Wert für Proband 2. Es zeigt sich, dass die Genauigkeit bei verringerter Pixelanzahl abnimmt.

4.11 Multispektrale Messung

Die meisten Verfahren im Stand der Technik nutzen RGB Kameras, um die Vitalparameter zu schätzen, da insbesondere der Grünkanal günstige spektrale Eigenschaften für die Messung des PPG-Signales aufweist. Um die Eigenschaften weiterer Spektralbereiche zu untersuchen, wurden in mehreren Versuchen multispektrale Messungen in verschiedenen Wellenlängen durchgeführt. Zum einen, eine allgemeine Untersuchung im sichtbaren und

nah-infrarotem Licht im Bereich von 400 - 950 nm (siehe Kapitel 4.11.1) und zum anderen eine genauere Betrachtung der PPG-Signale im nah-infraroten Spektrum mit einer höheren spektralen Auflösung (siehe Kapitel 4.11.2). Zudem wurde die Verwendung einer rausch-armen monochromatischen Kamera für die Vitalparamterschätzung in einem MRT Gerät untersucht (siehe Kapitel 4.10). Teilergebnisse der multispektralen Messungen wurden in den Arbeiten [Rap+16b] und [Rap+18a] publiziert. Die experimentellen Versuche wurden an der TU Ilmenau durchgeführt. Die Beschreibungen der zwei Versuchsaufbaue und der Datenaufnahme wurden von den Co-Autoren der TU Ilmenau geschrieben und für diese Arbeit übernommen.

4.11.1 Multispektrale Messung 450 - 950 nm

Versuchsaufbau und Datenaufnahme

Abbildung 4.39 zeigt den verwendeten Versuchsaufbau. Zwei Halogenlampen mit breitem Lichtspektrum wurden links und rechts der Kamera in einem Winkel von $\pm 45^\circ$ Grad positioniert. Eine „Smart Spectral Imager 2.0“ [Ros+15] Multispektralkamera, mit einem CMOS-Sensor, einer Auflösung von 1024 x 1160 Pixeln und acht Spektralkanälen von 400 - 950 nm mit einer Halbwertsbreite von 50 nm, wurde für die Messungen verwendet. Die Videos wurden mit einer Bildwiederholrate von 60 FPS aufgezeichnet.

Das Kamerasystem hat einen integrierten FPGA, mit der sich eine Anpassung der Integrationszeit für jeden Spektralkanal durchführen lässt. Dadurch kann ein Korrekturkoeffizient für jeden Kanal definiert werden, um einen Intensitätsausgleich zwischen den Spektralkanälen zu erreichen, um die unterschiedlichen Eigenschaften der einzelnen Komponenten (Licht, Filter, Sensor) bei verschiedenen Wellenlängen auszugleichen. Weißes Acrylglas wurde für die Kalibrierung der einzelnen Kanäle verwendet, da es ein stabiles Reflexionsverhalten im VIS und NIR-Bereich aufweist. Die mittlere Intensität des Kalibrationsobjekts in jedem Spektralkanal wurde durch die Anpassung der Integrationszeit angeglichen.

Es wurde eine Datenbank mit 9 Aufnahmen von 5 Personen erstellt. Für jede Messung wurde das Gesicht des Probanden über alle Spektralkanäle

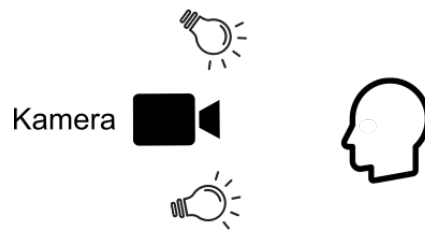


Abbildung 4.39: Verwendeter Versuchsaufbau mit 8-kanalige Multispektralkamera und Halogen-Lampen.

ein Video mit 10 Sekunden Länge aufgenommen. Die unterschiedlichen Spektralkanäle wurden nacheinander aufgezeichnet. Die Probanden wurden aufgefordert, sich während der Aufnahme möglichst nicht zu bewegen. Die Gesichter der Probanden wurden mittig in der Aufnahme platziert. Die Puls-Grundwahrheit des Probanden wurde während den Aufnahmen mit einem Fingerpulsoximeter gemessen.

Datenverarbeitung

Für die Schätzung der Herzrate wurde zunächst eine ROI für jedes Video definiert. Da sich die Probanden während der Aufnahmezeit nicht bewegt haben, wurde die ROI manuell für jede Aufnahme, für die gesamte Aufnahmedauer definiert. Dabei wurden die Ränder der Videobilder um 100 Pixel in jeder Richtung beschnitten, um den Großteil des Hintergrundes zu entfernen. Die verbliebenen Pixelwerte wurden für jedes Bild gemittelt, um das PPG-Signal aus den mittleren Intensitäten der Videobilder zu bestimmen.

Für jedes Zeitsignal wurde anschließend ein Polynom 1. Ordnung mittels des *Least Squares*-Verfahrens bestimmt, um den Offset und linearen Trend des Signales zu entfernen. Im Anschluss wurde das Spektrum des normierten Signales berechnet. Dazu wurde die spektrale Leistungsdichte des Signales mithilfe eines Periodigramms für die Frequenzen von 30 BPM (0,5 Hz) bis 200 BPM (3,33 Hz), mit einer Schrittweite von 1 BPM (1/60 Hz) ermittelt. Dabei wurde ein Hammingfenster derselben Länge, wie das zu analysierende Signal, verwendet.

4 Experimentelle Ergebnisse

Der maximale Peak des Leistungsdichtespektrums wird als die geschätzte Herzrate angenommen. Dabei stellt der Fehler die Differenz des Peaks zum Ground-Truth-Wert in BPM dar. Die absoluten Fehler werden über alle Probanden gemittelt, um so eine Aussage zu den einzelnen Frequenzbändern zu ermöglichen.

Ergebnisse

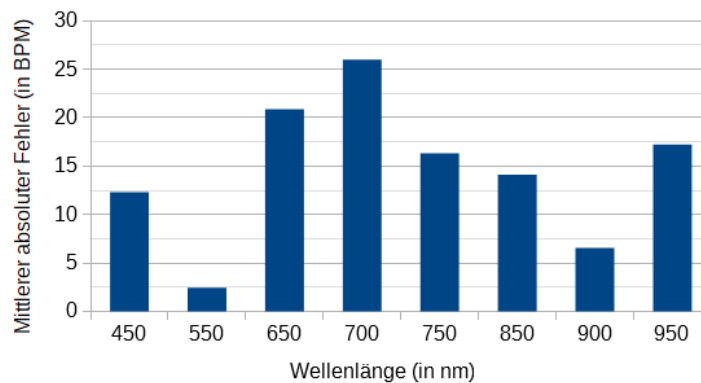


Abbildung 4.40: Mittlerer absoluter Fehler in Abhängigkeit der Wellenlänge.

Für jedes vermessene Band wurde der mittlere absolute Fehler ermittelt. Abbildung 4.40 zeigt den mittleren absoluten Fehler für die gemessenen Wellenlängen. Es sind zwei Minima bei der Fehlermessung zu erkennen. Diese liegen bei den Wellenlängen 550 nm und 900 nm. Der mittlere absolute Fehler bei 550 nm beträgt 2,44 BPM, und bei 900 nm 6,56 BPM. Bei Betrachtung der Einzelmessungen (siehe Tabelle 4.20) ist zu sehen, dass ein Proband (Messung 8) bei der Messung mit 900 nm ein signifikant schlechteres Ergebnis als die anderen Probanden aufweist. Dies könnte durch einen Fehler bei der Messung erklärt werden, etwa von kleinen Bewegungen des Kopfes, welche zu Artefakten im PPG-Signal führen können oder von Fehlern bei der Ground-Truth Kontrollmessung der Herzrate. Ohne Proband 8 ergibt sich ein mittlerer Fehler auf 2.75 BPM bei 900 nm.

Tabelle 4.20: Absolute Fehler in BPM (Fehler >10 BPM hervorgehoben)

Wellenlänge (in nm)	Messung								
	1	2	3	4	5	6	7	8	9
450	3	31	0	1	5	6	41	22	2
550	3	5	1	0	1	0	4	6	2
650	57	9	6	52	2	4	18	38	2
700	3	3	38	6	84	25	2	39	34
750	2	56	42	12	31	3	1	0	0
850	3	0	3	66	40	1	7	5	2
900	0	4	3	1	6	2	2	37	4
950	2	4	2	49	2	2	54	37	3

4.11.2 Multispektrale Messung 675 - 950 nm

Die Ergebnisse aus Abschnitt 4.11.1 zur multispektralen Messung der Haut für den Einsatz in der Herzratenschätzung wurden als Grundlage für weitergehende Untersuchungen genommen. Um den Einfluss der Wellenlänge auf die Messungen im NIR-Bereich genauer zu spezifizieren, wurde ein zweiter Versuchsaufbau mit höherer spektraler Auflösung konzipiert und Messungen mit einer größeren Anzahl an Probanden durchgeführt.

Versuchsaufbau und Datenaufnahme

Für den Versuch wurde eine Hyperspektralkamera (Ximea MQ022HG-IM-SM5X5-NIR) mit einem 35 mm TECHSPEC Objektiv verwendet. Die Kamera hat 25 spektrale Bänder mit einer Schrittweite von 10 nm über einen Messbereich von 675 - 950 nm und eine Auflösung von 409 x 217 Pixel je spektralem Band. Die spektralen Bänder bis 920 nm haben eine Halbwertsbreite von 10 nm. Darüber steigt die Halbwertsbreite herstellungsbedingt auf bis zu 20 nm an. Die Kamera wurde über eine USB 3.0 Schnittstelle verbunden und angesteuert.

Der Versuchsaufbau ist in Abbildung 4.41 dargestellt. Aufgrund der niedrigen Auflösung der Kamera wurde ein Abschnitt (etwa 4,5 x 2,5 cm) auf die Innenseite des Handgelenkes aufgezeichnet, um gezielt viel Haut und

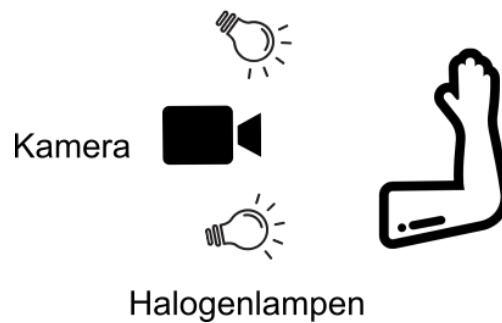


Abbildung 4.41: Experimentaler Aufbau mit Hyperspektralkamera und zwei Halogenlampen.

damit möglichst viele Pulsinformationen zu erfassen. Die Probanden legten ihren Arm quer zur Kamera auf ein Kissen. Dabei wurde diese von zwei Halogenlampen beleuchtet. Alle spektralen Bänder wurden gleichzeitig aufgenommen. Die Messungen wurden in einem dunklen Raum durchgeführt, um den Einfluss durch weitere Beleuchtungsquellen zu verhindern. Je Versuchsperson wurde ein 10 Sekunden langes Video mit 60 Hz mithilfe der *MATLAB Image Acquisition Toolbox* aufgenommen. Es wurden von 38 (11 Frauen, 27 Männer) Probanden Video- und Pulsdaten aufgezeichnet. Davon waren 5 europäischer und 33 asiatischer Abstammung. Die Probanden wurden gebeten, ihren Körper während der Aufnahme möglichst nicht zu bewegen. Die Herzschlag-Grundwahrheit wurde mithilfe eines Fingerpulsoximeters (Pulox PO-200) aufgezeichnet. Der gemessene Puls wurde während der Videoaufnahme abgelesen.

Datenverarbeitung

Aus den 38 Messungen wurde für jedes der 25 Bänder ein PPG Signal generiert. Aus jedem der 950 Videos wurde ein PPG-Signal berechnet, indem die mittlere Intensität für jedes Frame bestimmt wurde. Um Störungen aufgrund von Randeffekten, ausgelöst durch kleine Bewegungen des Armes, in den Videos zu minimieren wurde die Pixel bei der Mitteilung gewichtet. Dazu wurde eine normierte multivariate Normalverteilung mit den

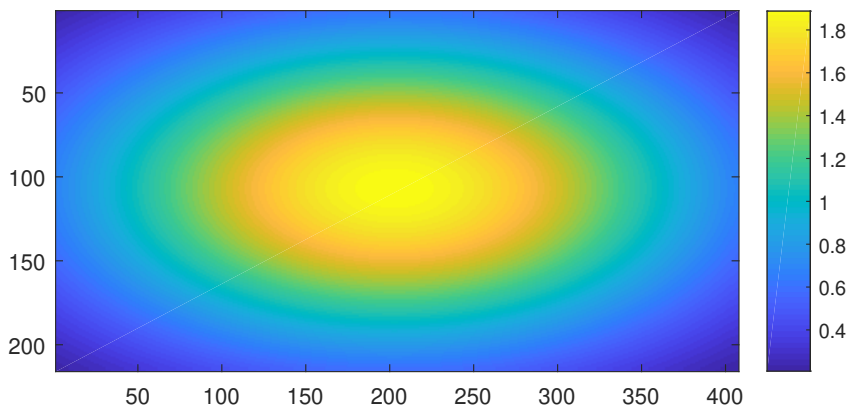


Abbildung 4.42: Normierte multivariate Normalverteilung zur gewichteten Mittelwertbildung in Pixel.

Standardabweichungen 141 in x -Richtung und 71 in y -Richtung verwendet (siehe Abb. 4.42).

Das resultierende PPG-Signal wurde zuerst mit einer gefitteten Gerade erster Ordnung *detrended* und danach standardisiert. Im Anschluss wird das Signal mit einem FIR Bandpassfilter 128er-Ordnung und den Grenzfrequenzen 0,5Hz und 3 Hz gefiltert und mit einem Hamming-Fester gewichtet. Für das gefilterte Signal wird die spektrale Leistungsdichte innerhalb der Bandpassgrenzen mithilfe eines Periodogramms in 1 BPM-Schritten kalkuliert (siehe Abb. 4.43).

In dem Spektrum werden zwei Peaks bestimmt. Der Peak, welcher am nächsten an der gemessenen Grundwahrheit liegt, wird als **GW-Peak** definiert. Die Breite des Peaks wird als der Abstand der Flanken auf halber Höhe bestimmt. Der dadurch definierte Bereich wird als das **GW-Band** bezeichnet. Das restliche Spektrum wird als Rauschen definiert. Der höchste Peak außerhalb des GW-Bandes wird als **Rausch-Peak** definiert und seine Höhe und Breite analog zum GW-Peak bestimmt.

4 Experimentelle Ergebnisse

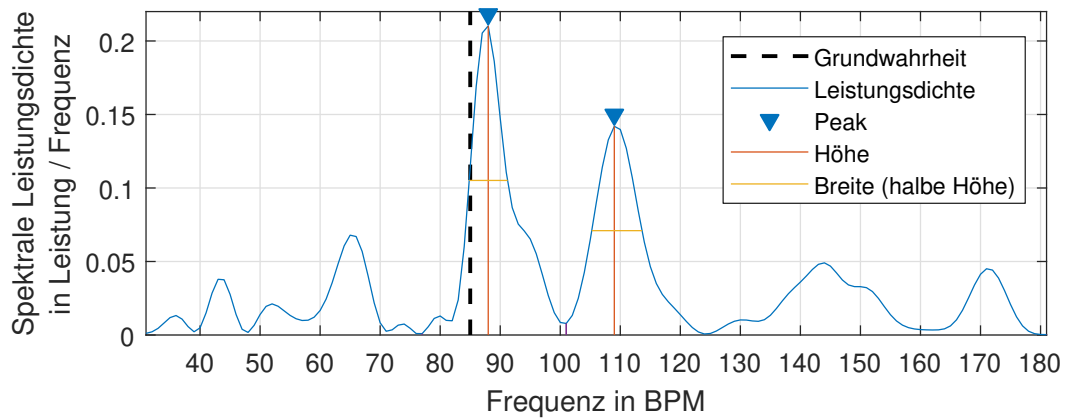


Abbildung 4.43: Beispiel der berechneten spektralen Leistungsdichte, der Grundwahrheit und den gefundenen Peaks mit Höhe und Breite.

Ergebnisse

Verschiedene Fehlermaße wurden verwendet, um den Einfluss der Wellenlänge auf die kamerabasierte Herzratoschätzung zu bestimmen. Der Fehler e (in BPM) der Herzratoschätzung wurde für alle Bänder in den Messungen bestimmt. Dazu wurde die Differenz der im Spektrum gefundenen Frequenz f des GW-Peaks und der gemessenen Herzratosgrundwahrheit hr gebildet. Darüber hinaus wurden das Verhältnis der Peakhöhen zwischen dem GW-Peak und dem Rauschpeak bestimmt, sowie der Anteil der Leistungsdichte des GW-Bandes vom Gesamtsignal.

Absoluter Fehler Der Fehler der Herzratoschätzung wurde für jeden Probanden über alle Wellenlängen gemittelt und ist in Abb. 4.44 dargestellt. Die Genauigkeit der Messung variiert stark zwischen den unterschiedlichen Probanden. Dem können unterschiedliche Fehlerquellen zugrunde liegen. Aufgrund der geringen örtlichen Auflösung der vermessenen Hautpartie können Artefakte im Signal bereits durch kleine rhythmische Bewegungen erzeugt werden. Es finden sich bei den Probanden 9, 15, 17, 24 und 30 dominante Störfrequenzen im niedrigen Messbereich $< 1\text{Hz}$ und bei den Probanden 19, 24, 31 dominante Störsignale über der gemessenen Grundwahrheit (siehe Abb. 4.45). Zudem liegt die gemessene Grundwahrheit

bei den Probanden 13, 17, 19, 24 und 30 mit signifikantem Abstand von dem nächsten Spektralpeak, was auf einen Fehler bei dem Ablesen der Grundwahrheit hindeuten kann (siehe Abb. 4.45). Die acht oben genannten Probanden wurden aufgrund der beschriebenen Fehlerquellen und Abweichungen daher für die weiteren Betrachtungen nicht verwendet.

Zur Untersuchung des Einflusses der verwendeten Wellenlängen wurde der Fehler der Herzratoschätzung für die einzelnen Wellenlängen über die verbliebenen Probanden gemittelt (siehe Abb. 4.44). Der ermittelte Fehler bei allen Wellenlängen niedrig und zeigt keine systematische Abhängigkeit von dieser. Der Einfluss von äußeren Störfaktoren, wie zum Beispiel der Messfehler des verwendeten Pulsoximeters (± 3 BPM), überlagert die Messergebnisse bei der Auswertung über den Fehler der Herzrate.

Signalleistung Um den Einfluss der Wellenlänge systematischer zu erfassen, wurde daher der Anteil der Pulsfrequenz von der Leistung des gesamten PPG-Signales bestimmt. Dazu wurde der Anteil der Leistung des *GW-Bandes* im Signal bestimmt (siehe 4.47). Weiterhin wurde das Verhältnis der Höhe des *GW-Peaks* zum *Rausch-Peak* in Abhängigkeit der Wellenlänge berechnet (siehe Abb. 4.48). Bei beiden Auswertungen ist eine deutliche Abhängigkeit der Stärke des Pulssignales von der Wellenlänge erkennbar. Dabei ist die Signalstärke im Bereich um 850 nm am größten.

Bildintensität Aufgrund der verwendeten Sensortechnik (Mosaik-Aufbau) kann für die unterschiedlichen Spektralbereiche keine Helligkeitskorrektur durchgeführt werden, da alle Bänder gleichzeitig auf demselben Bildsensor und mit denselben Einstellungen belichtet und ausgelesen werden. Die Intensität der aufgenommenen Videos variiert daher über die verschiedenen Wellenlängen. Sie steigt zunächst, bis zu dem Bereich um 820 nm, an und fällt dann wieder ab (siehe Abb. 4.49).

Die Intensität korreliert jedoch nicht mit der gemessenen Signalstärke in Abb. 4.47 und 4.48. Um einen möglichen Einfluss auf die Messung der Signalstärke auszuschließen, wurde eine mögliche Abhängigkeit des Anteils der Pulsfrequenz von der Intensität untersucht (siehe Abb. 4.50). Auch hier konnte keine systematische Abhängigkeit gefunden werden.

4 Experimentelle Ergebnisse

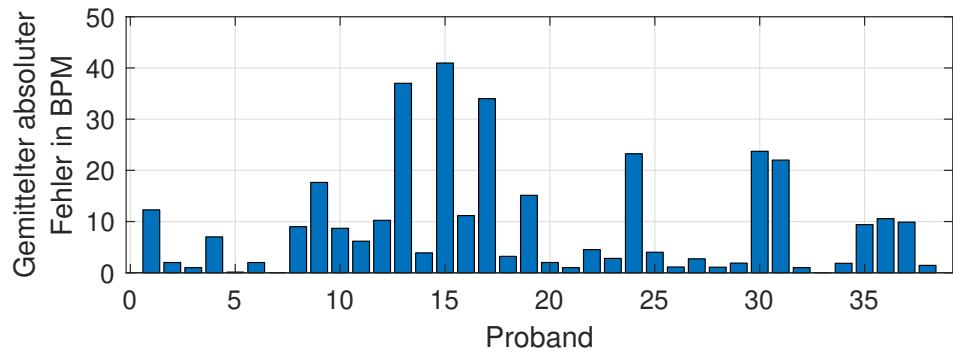


Abbildung 4.44: Gemittelter absoluter Fehler der Messung der Herzraten für alle Probanden.

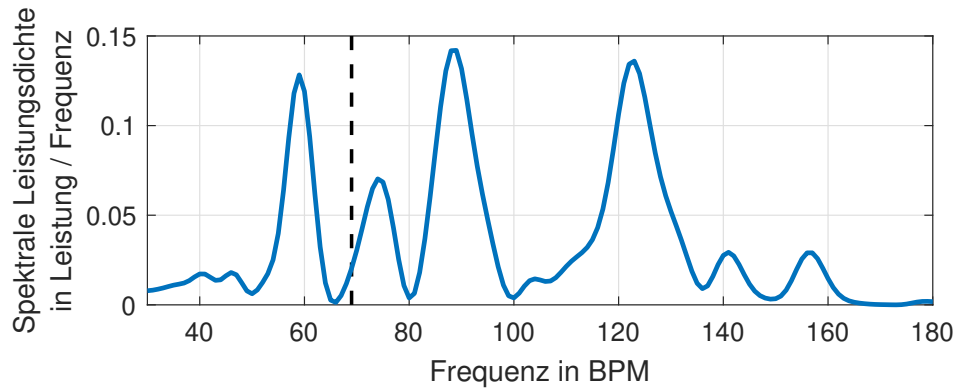


Abbildung 4.45: Periodogramm mit dominanten Störfrequenzen und fehlerhafter Messung der Grundwahrheit (Proband 24, 860nm, Grundwahrheit schwarz).

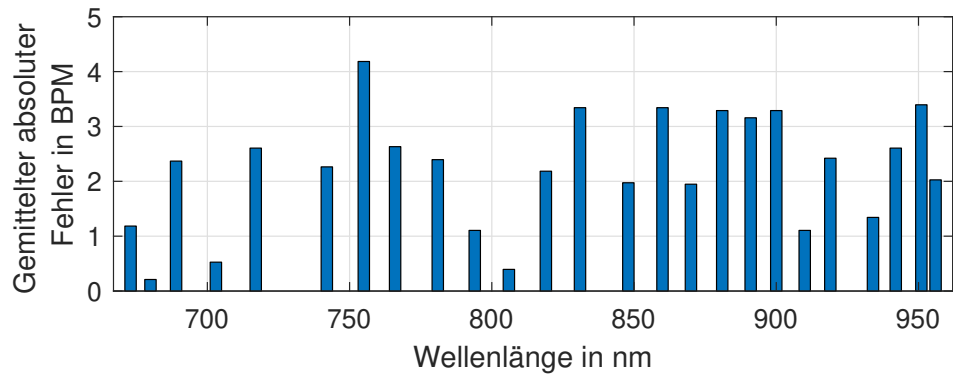


Abbildung 4.46: Gemittelter absoluter Fehler der Messung der Herzraten in Abhängigkeit der Wellenlänge.

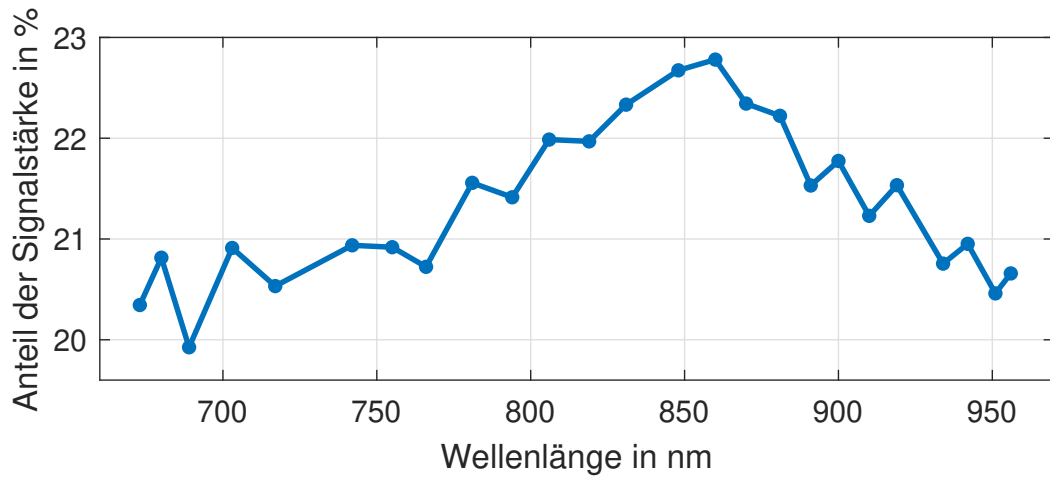


Abbildung 4.47: Gemittelter Anteil der Leistung der Pulsfrequenz vom PPG-Signal in Abhängigkeit der Wellenlänge.

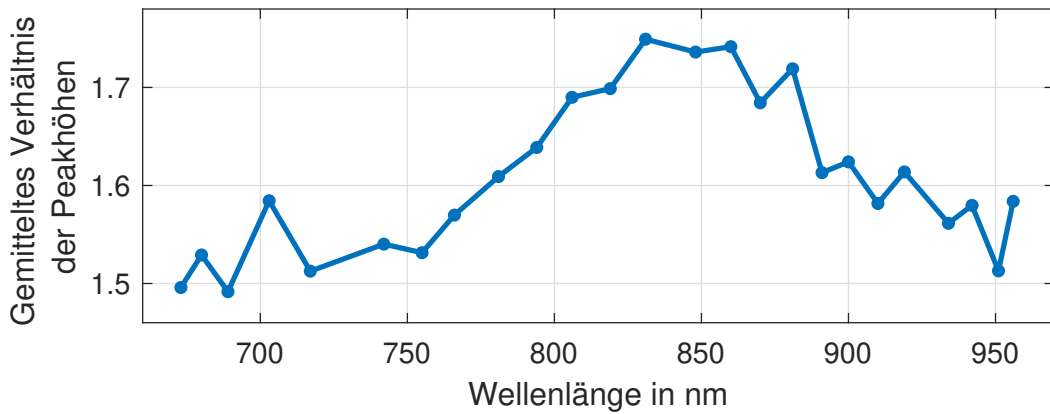


Abbildung 4.48: Gemitteltetes Verhältnis der Höhe des *GW-Peaks* zum *Rausch-Peak* in Abhängigkeit der Wellenlänge.

4 Experimentelle Ergebnisse

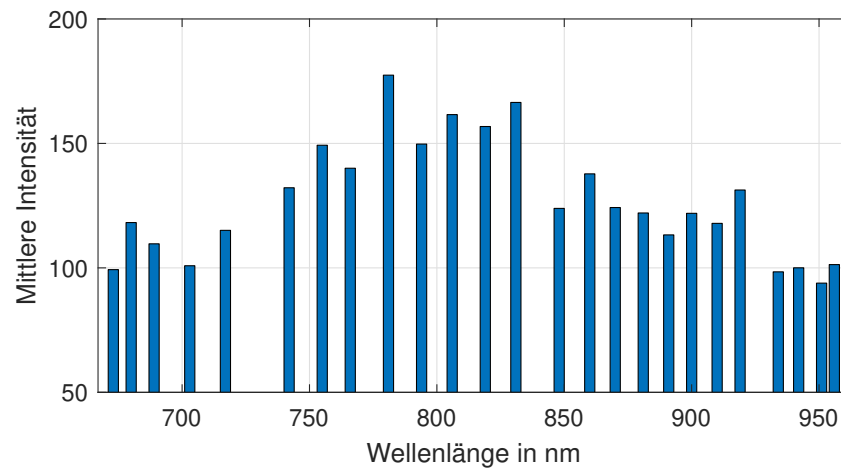


Abbildung 4.49: Mittlere Intensität (Helligkeit) in Abhängigkeit der Wellenlänge.

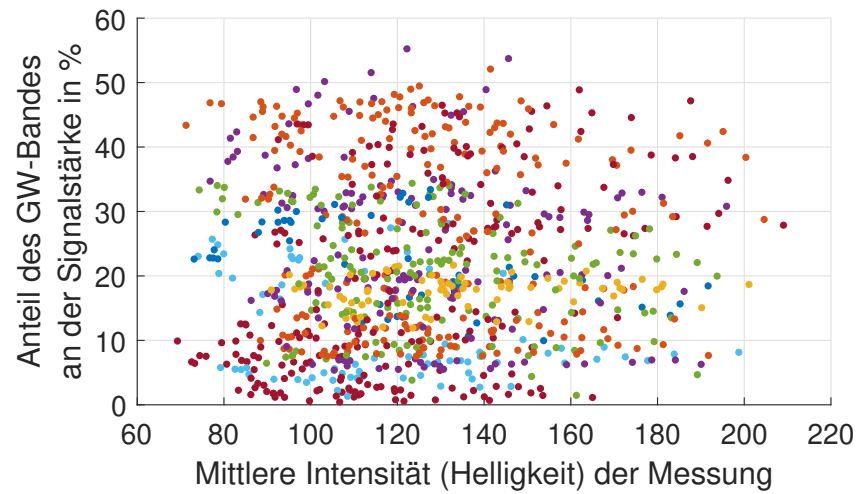


Abbildung 4.50: Anteil der Pulsfrequenz in Abhängigkeit der Intensität (Probanden farblich gruppiert).

5 Zusammenfassung und Diskussion

Die Ergebnisse der im Kapitel 4 vorgestellten Experimente und Daten werden in diesem Kapitel zusammengefasst. Auf dieser Grundlage werden die Verbesserungen der Störsicherheit der Vitalparametermessung durch die unterschiedlichen Methoden diskutiert. Dabei werden zunächst verschiedene Aspekte der Videokompression und -speicherung behandelt. Danach werden sowohl der Einfluss der Region of Interest (ROI), Signalextraktion, als auch der Herz- und Atemratenmessung betrachtet. Die Nutzung der Vitalparameterschätzung für die Lebenderkennung sowie der Messung im MRT werden im Anschluss diskutiert. Abschließend wird die Auswirkung der Wellenlänge im Nahinfrarotspektrum (NIR) auf die Schätzung der Herzfrequenz zusammengefasst.

5.1 Videoeigenschaften

Im Kapitel 4.4 wurden der Einfluss verschiedener Parameter der Videoaufnahme und -codierung getestet und untersucht. Diese können genutzt werden, um Vorschläge für die Encodierungs- und Archivierungsoptionen zukünftiger Datensätze für die videobasierte Herzfrequenzschätzung zu erarbeiten und so die Qualität der Videodaten zu verbessern. In den folgenden Abschnitten werden die Einflüsse der einzelnen Aspekte bewertet und abschließend eine Reihe von, aus den Daten abgeleiteten, Empfehlungen für die Aufnahme und Archivierung vorgestellt.

5.1.1 Wahl des CRF-Wertes

Von allen getesteten Parametern hat die Wahl des *CRF* den größten Einfluss. Je nach verwendetem Codec können schon kleine Unterschiede einen stark nachteiligen Einfluss auf das extrahierte PPG-Signal haben. Zudem ist die Wahl des *CRF* bei der Nutzung verschiedener Codecs nicht vergleichbar. Während die Ergebnisse im niedrigen *CRF*-Bereich (0-7) bei der Nutzung von *x264* nur geringe Einbußen in der Signalqualität zeigen, fällt die IEC-Genauigkeit bei *x265* bereits bei kleinen Werten unter 5 stark ab. Weiterhin führt die Wahl des *CRF* auf verschiedene Datenbanken auch zu unterschiedlich starker Verringerung der Genauigkeit der Vitalparametermessung (siehe Abbildungen 4.8 und 4.9). Dabei können Unterschiede, wie Auflösung und Videoinhalt (Bewegung, Hintergrund, ...), zu einer anderen Wahl der dynamischen Kompressionsverfahren und -parameter bei der Encodierung führen, was in einer unterschiedlichen Art und Stärke der Informationsreduktion im Video resultiert.

Zu beachten sind insbesondere die voreingestellten *CRF* Werte von 23 für *x264* und 28 für *x265*. Diese reduzieren die Qualität der PPG-Signale dermaßen, dass die Videos für die Herzfrequenzschätzung praktisch unbrauchbar werden (siehe Abbildung 5.1 und Abschnitt 4.4.2). Wie in Kapitel 4.3 am Beispiel der *COHFACE* Datenbank beschrieben, können durch fehlende Expertise und Verwendung der Voreinstellungen bei der Encodierung ganze Datenbanken für die Forschung in der kamerabasierten Vitalparametermessung unbrauchbar werden.

Während theoretisch ein optimaler *CRF* Wert in Abhängigkeit von der Qualität, Auflösung und dem Inhalt des zu encodierenden Datensatzes gefunden werden könnte, ist ein $CRF = 0$ die sicherste Option in Bezug auf die PPG-Qualität. Der Wert sollte nur erhöht werden, wenn die Dateigröße ein Problem darstellt und andere Optionen zur Reduzierung des Speicherplatzes bereits genutzt wurden.

Auffällig bei den Messungen für unterschiedliche *CRF* Werte mit dem *x264* Codec (siehe z.B. Abbildungen 5.1, 4.8 und 4.9) ist ein Abfall und eine plötzlich folgende Steigerung der Genauigkeit bei niedrigeren *CRF* Werten zu beobachten. Dieser Effekt tritt in verschiedenen Datenbanken auf und zeigt sich auch bei der Fehlerberechnung der Herzrate in [MBE17, Tabelle 1]

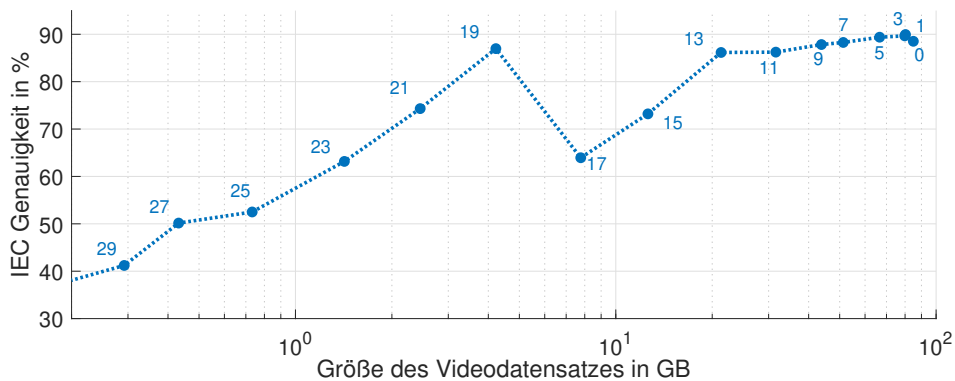


Abbildung 5.1: Mittlere IEC-Genauigkeit für verschiedene CRF-Werte (Zahlen) und die Datensatzgröße für den **x264** codecs (YUV₄₂₀) auf dem MMSE-Datensatz.

bei der dort getesteten *stationären Aufgabe* und der *Zufallsbewegungsaufgabe* bei Verwendung des *x264*-Codecs. Der Fehler nimmt dort bei einem $CRF=6$ zunächst zu und anschließend bei $CRF=9$ kurzzeitig wieder ab.

Die genaue Ursache dieses Effekts konnte aufgrund der Komplexität des verwendeten Codecs nicht abgeschätzt werden. Wahrscheinlich haben die unterschiedlichen Kompressionsverfahren des Codecs, wie zum Beispiel die Größe der verwendeten Quantisierungstabellen, oder die Abmaße und Variabilität des Block-Matchings, verschieden starken Einfluss auf die Vitalparametermessung. Bei verschiedenen CRF Werten kommt es zu einer dynamischen Auswahl und Parameterwahl dieser Verfahren, was zu systematischen Unterschieden beim Kompressionsfehler führen kann. Hinweise auf diese diskreten *Sprünge* in der Kompression sind in der mittleren Abweichung der encodierten Farbwerte in der Abbildung 4.7 zu erkennen.

5.1.2 Einfluss von räumlicher Unterabtastung

Es wurden zwei Formen der Informationsreduktion durch räumliche Unterabtastung untersucht, Farbunterabtastung und den Einfluss der Bildauflösung. In beiden Fällen konnte gezeigt werden, dass die Genauigkeit der Herzfrequenzschätzung bei Reduzierung der Farb- und Pixelinformation

konstant gehalten werden kann. Insbesondere die Methoden der Unterabtastung, welche die neuen Pixelinformationen durch ein einzelnes Quellpixel, ohne Filterung oder Mittelwertbildung, bestimmen (*YUV₄₂₀* und *neighbor*), zeigen stabilere IEC-Ergebnisse trotz Informationsreduktion.

Farbunterabtastung

Bei Einsatz der Farbunterabtastung *YUV-422* mit dem *x264* Codec konnten nur geringe Unterschiede in der IEC-Genauigkeit festgestellt werden (siehe Abb. 4.10). Gleichzeitig führten die zusätzlichen Farbinformationen im *YUV444*-Format zu einer Verdopplung der Dateigröße (siehe Abb. 4.10). Bei Verwendung des *x265*-Codecs waren die Größenunterschiede für die farbunterabgetasteten Videos deutlich geringer (+14% bei $CRF = 0$) (siehe Abb. 4.11). In beiden Fällen erzielte *YUV₄₂₀* bessere Ergebnisse als *YUV444*. Insgesamt erreichte der *x264* Codec etwas besser Ergebnisse als *x265*.

Die bessere Leistung des *YUV₄₂₀*-Formats, trotz weniger Farbinformationen, könnte mit einer besseren Optimierung im Kodierungsprozess (Standard-Pixelformat) oder einem hohen PPG-Informationsgehalt im Y-Kanal erklärt werden. Alternativ könnte bereits eine Farbunterabtastung hardwareseitig in den verwendenden Kameras durchgeführt worden sein, wobei die Farbinformationen später bei der Konvertierung in das *PNG*-Format hochgerechnet wurden. Diese Effekte sollten in Zukunft weiter untersucht werden.

Auflösung

Die Abbildungen 4.18 und 4.19 zeigen, dass die Anzahl der Gesichtspixel ohne negativen Einfluss auf die IEC-Genauigkeit bis zu einer Grenze reduziert werden kann. Wenn das Bild auf weniger als 10.000 Bounding-Box-Pixel reduziert wird, erzielt nur der *nearest neighbor* Algorithmus weiterhin stabile Ergebnissen (siehe Abb. 4.18), während die Genauigkeit bei den *area* und *bilinear* Skalierungsmethoden beginnt abzufallen.

Abbildungen 5.2 und 5.3 zeigen zwei Beispiele für PPG-Signale mit unterschiedlichen Auflösungen und Skalierungsalgorithmen. Während die PPG-Signale bei geringer Skalierung fast identisch sind (siehe Abb. 5.2),

5.1 Videoeigenschaften

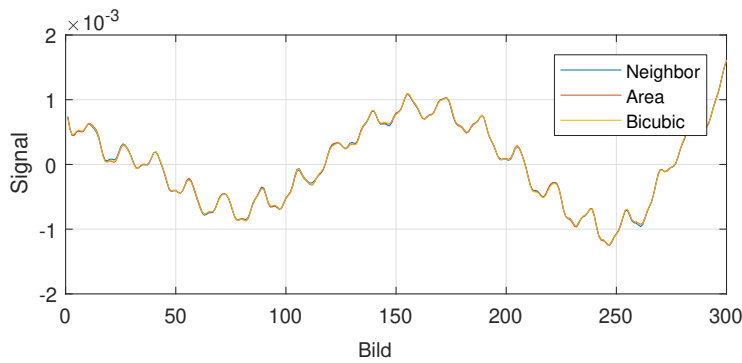


Abbildung 5.2: Beispiel PPG-Signal ($normG$) aus dem MMSE-HR Datensatz (videoID: F005/T10) der ersten 300 Bilder mit einer Auflösung von 976x1306 Pixel unter Verwendung verschiedener Skalierungsalgorithmen.

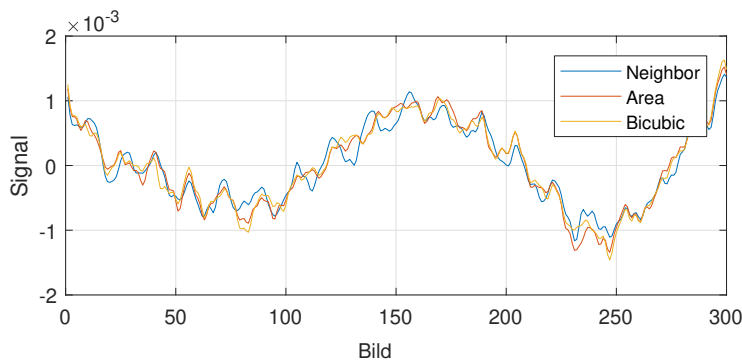


Abbildung 5.3: Beispiel PPG-Signal ($normG$) aus dem MMSE-Datensatz (videoID: F005/T10) der ersten 300 Bilder mit einer Auflösung von 130x174 Pixel unter Verwendung verschiedener Skalierungsalgorithmen.

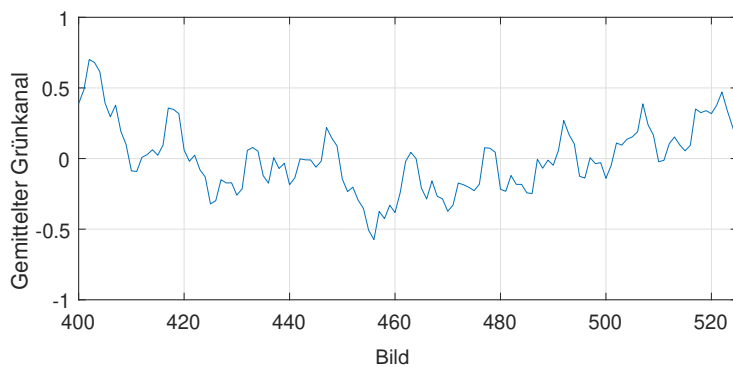


Abbildung 5.4: Beispiel des gemittelten Grün-Kanal Signales (MMSE-HR, videoID: F005/T10, x265, CRF=0).

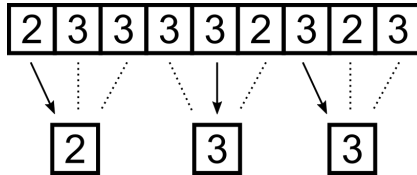


Abbildung 5.5: Beispiel für die Skalaierung mit dem *nearest neighbor* Verfahren.

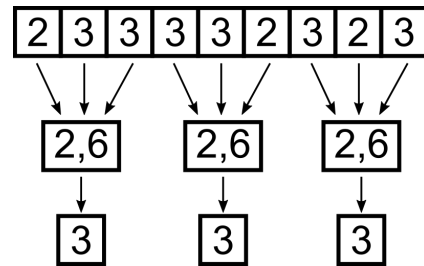


Abbildung 5.6: Beispiel für die Skalaierung mit dem *area* oder *bicubic* Verfahren.

unterscheiden sich die Signale in der kleinsten berechneten Auflösung erheblich voneinander. Der *nearest neighbor* Algorithmus weist dabei eine wesentlich deutlichere Peak-Prominenz als die beiden anderen Methoden auf. Besonders an den Flanken der Signale (siehe Abb. 5.3, im Bereich 100-150) ist die Peakhöhe des *area*- und *bicubic*-Skalierungsalgorithmus deutlich geringer.

Eine mögliche Erklärung für diesen Effekt ist, dass der Herzschlag im PPG-Signal eine geringere Amplitude als die Farbquantisierungsschwellen des Videoformates hat. Die Abbildung 5.4 zeigt ein Beispiel eines gemittelten Grünkanales aus dem MMSE-HR Datensatz. Die Spitzen der Herzschläge haben eine Höhe von > 1 und sind damit kleiner als die Farbquantisierungsschritte des Videos. Das PPG-Signal wird demnach nur durch die Mittlung der Farbwerte von genügend Hautpixeln erkennbar, welche von einem Quantisierungsschritt zum anderen hin und her wechseln.

Im Falle der Videocodierung werden die Ergebnisse der Mittelwertbildung durch Reduzierung der Videoauflösung jedoch nicht als Gleitkommazahl, sondern als ganzzahliger Pixelwert (0-255) im neuen Video gespeichert. Dadurch erzielt der Mittelwert einer Teilmenge der ursprünglichen Pixelwerte ein besseres Ergebnis, als der Mittelwert einer interpolierten und quantisierten Teilmenge von Pixelwerten, bei der diese Information verloren geht. Ein Beispiel wäre das Array [2 3 3 3 3 2 3 2 3] mit einem Mittelwert von 2,66. Reduziert man die *Auflösung* um den Faktor 3 und nimmt eine zufällige Teilmenge von jeweils drei Werten (*nearest neighbor*), wie in Abbildung 5.5 gezeigt, so wäre das Ergebnis [2 3 3] (nicht notwendigerweise

in dieser Reihenfolge) mit einem Mittelwert von ebenfalls 2,66. Wenn die neuen Werte gemittelt und gerundet werden (*area, bicubic*), wäre das neue Ergebnis jedoch [3 3 3] mit einem anderen Mittelwert als die Ausgangsreihe (siehe Abbildung 5.6).

Daher sollte jede lokale Filterung oder Mittelwertbildung vermieden werden, um diese Subpixel-Farbinformationen bei der Neuquantisierung vor Rundungen zu schützen. Dies gilt auch für Bildtransformationen, Drehungen und ähnliche Operationen. Wenn zum Beispiel mehrere Bildoperationen zur Berechnung der ROI notwendig sind, sollte die ROI auf das ursprüngliche Bild abgebildet werden, anstatt das transformierte Bild zu verwenden.

5.1.3 Einfluss der zeitlichen Unterabtastung

Der Einfluss der zeitlichen Abtastung wurde durch die Änderung der Bildwiederholrate untersucht. Die Abbildungen 4.20, 4.21 und 4.22 zeigen Genauigkeit in Abhängigkeit der Bildwiederholrate. Für niedrigere Bildwiederholraten ist diese zunächst leicht abfallend, unter einem Wert von 6 bis 8 FPS fällt die Genauigkeit jedoch stark ab.

Dies lässt sich durch eine Unterschreitung der minimal notwendigen Abtastfrequenz nach dem Nyquist-Shannon-Abtasttheorem erklären. Da der Bereich der Herzfrequenz bei einem gesunden Menschen zwischen 30 und 200 BPM (0,5 - 3,3 Hz) liegt, darf die Abtastfrequenz (hier Bildwiederholrate) nicht unter dem doppelten der Herzfrequenz liegen, um die Pulsinformationen aus den Bildern wieder zu rekonstruieren. Daher kommt es im Bereich um 7 FPS zu einem unterschreiten der notwendigen Abtastfrequenz für hohe Herzraten, was den Einbruch der Genauigkeit zur Folge hat.

Höhere Bildwiederholraten fügen laut Blackford und Estep [BE15] und Sun, Hu, Azorin-Peris, Kalawsky und Greenwald [Sun+12] keine PPG-Information hinzu, erhöhen jedoch die Dateigröße. Daher sollten für die Bestimmung der Herzrate auch niedrigere Bildwiederholraten um 10 FPS ausreichend Informationen enthalten. Bei Bestimmung der Herzrate durch die Auswertung des PPG-Spektrums verringert eine niedrigere Abtastfrequenz, bei gleicher Signallänge, jedoch die Auflösung der Frequenzanteile und kann so die Genauigkeit der Messung beeinträchtigen.

5.1.4 Empfehlungen für die Kompression der Videodaten

Die Kompression der Videodaten ist ein sehr wichtiger Faktor für kamera-basierte Vitalparameterschätzung. Jedes Experiment oder Datensatz kann durch die Nutzung einer problematischen Kompressionsmethode oder die Verwendung eines ungünstig gesetzten Standardparameters zunichtegemacht werden, wodurch viel Arbeitsaufwand, Zeit und Geld verloren gehen können. Andererseits ist die Reduzierung der Dateigröße durch bedachte Videokompression möglich und erleichtert, bei richtiger Anwendung, durch die Nutzung und den Austausch von Daten um dieses Forschungsgebiet voranzubringen.

Aus den Ergebnissen der Experimente wurden Vorschläge abgeleitet, um die Qualität der Videodaten für die kamerabasierte Herzfrequenzschätzung zu erhöhen, ohne die PPG-Informationen zu verfälschen.

- Hardware
 - Es sollten, wenn möglich, Industriekameras mit einem geringen Signal-Rausch-Verhältnis verwendet werden, um alle Aspekte des Aufnahmeprozesses zu kontrollieren. Dies ermöglicht genaue Einstellungen der meisten Kameraparameter und Zugriff auf unkomprimierte Bilddaten.
 - Bei der Verwendung von Consumer-Produkten sollte mit besonderer Vorsicht vorgegangen werden. Bestimmte Geräte (Webcams, Camcorder usw.) geben Videodaten nur komprimiert an den Nutzer aus. Diese wird hardwareseitig durchgeführt und kann häufig nicht umgangen werden.
 - Es sollten geeignete Beleuchtung und dazu passende Verschlusszeiten gewählt werden, um einen hohen Dynamikbereich von Farbe und Helligkeit zu erreichen.
- Aufnahme
 - Die Bildwiederholrate sollte zwischen 10-30 Hz gewählt werden.
 - Während der Aufnahme sollten die Videodaten im unkomprimierten RGB-*avi*-Format, mit dem HuffYUV-Codec oder als PNG-Bilder gespeichert werden.
 - Die Verwendung von hohen Farbtiefen sollte die Erkennung von kleinen Farbänderungen verbessern.

- Encodierung
 - Die Daten sollten mit dem x264 und einem CRF=0 encodiert werden (spart >80% Speicherplatz im Vergleich zu unkomprimierten Daten). **Der Standard-CRF sollte nicht verwendet werden.**
 - Chroma-Subsampling (YUV₄₂₀) kann genutzt werden, um 50% Dateigröße zu sparen (Standardeinstellung in x264).
 - Die Auflösung *kann* reduziert werden, wobei > 50.000 Gesichtspixel beibehalten werden sollten. Dabei sollte das *nearest-neighbor* downsampling verwendet werden, um den Verlust von Subpixel-Farbinformationen zu vermeiden und Speicherplatz zu sparen.

5.2 Region of Interest (ROI)

In Kapitel 4.5 wurde der Einfluss verschiedener ROIs auf die Herzratenerkennung untersucht. Dabei wurden Regionen aus dem Stand der Technik mit neuen Ansätzen einer auf Hautfarbenerkennung basierenden Methode verglichen. Die Haut-ROI wurde mit festen Schwellwerten für eine minimale Fläche (A), einer minimalen Wahrscheinlichkeit (t) oder einer gewichteten Mittlung der Pixel gebildet.

Durch die Mannigfaltigkeit der Trainingsdaten sind die verwendeten Hautfarbenwahrscheinlichkeiten auf einer Vielzahl an Datenbanken verwendbar. Die ROI ist zudem nicht von einer stabilen Lokalisation der Gesichtslandmarken abhängig und somit stabiler bei Bewegungen des Kopfes und Verdeckungen. Durch die Verwendung einer statischen Look-Up-Table ergeben sich Herausforderungen bei stark verfärbten, zum Beispiel geröteten, Hautpartien. Auch wird teilweise das Weiß der Augen und Zähne bei einigen Probanden als Haut erkannt. Dies kann durch die Mimik des Probanden, wie reden oder blinzeln, dominante Störfrequenzen in das PPG-Signal tragen.

In den Ergebnissen ist eine Dominanz der Haut-ROIs gegenüber anderen statischen ROIs verschiedener Gesichtsbereiche zu erkennen. So zeigt die Tabelle 5.1 eine deutliche Verbesserung der Genauigkeit der Herzratenschätzung und Tabellen 4.11 - 4.14 der Atemratenschätzung. Die Ergebnisse in Abhängigkeit des Schwellwertes t (siehe Abb. 4.25) deuten

Tabelle 5.1: Mittlere IEC Genauigkeit und Standardabweichung (in %) von ausgewählten Algorithmen (min IEC >70%) für verschiedene Datenbanken und Regions of Interest, basierend auf den Werten aus den Tabellen 4.5 - 4.7. (Für die PURE Datenbank wurde $t = 0.2$ für die Spalte *Haut* verwendet.)

Datenbank		Bounding Box	FaceMid	Forehead	Haut
PURE	μ	-	79,3	81,2	85,0
	σ	-	9,8	9,6	3,5
BioVid	μ	72,1	78,8	83,1	84,5
	σ	9,0	7,0	6,4	8,4
MMSE-HR	μ	78,5	77,4	80,5	81,7
	σ	8,8	9,2	8,3	6,6

darauf hin, dass bei der Wahl der ROI die Qualität der gewählten Pixel wichtiger ist als die Anzahl der Pixel. So erreicht die Herzratenschätzung die besten Ergebnisse, wenn nur die 10-20% der Pixel mit der höchsten Hautfarbenwahrscheinlichkeiten verwendet werden.

Für die minimale Größe der ROI, welche für die Schätzung verwendet werden sollte, konnte der Wert von etwa 10.000 Pixel identifiziert werden. Ab dieser Schwelle ist die Genauigkeit, sowohl bei den Experimenten zur Reduktion der Videoauflösung in Kapitel 4.4, als auch bei der Messung der Herzrate im MRT in Kapitel 4.10, abgefallen.

Die Mittelwerte und Standardabweichungen in Tabelle 4.5 deuten darauf hin, dass die Wahl der ROI einen geringeren Einfluss auf die Genauigkeit der Herzfrequenzschätzung hat als die Wahl des verwendeten Algorithmus. Die Mittelwerte der IEC-Genauigkeit haben für die ROIs eine deutlich kleinere Verteilung (67,5 - 79,4%) als die Mittelwerte für die Algorithmen (55,2 - 86,2%). Auch die Standardabweichungen für die einzelnen ROIs sind deutlich niedriger als die der Algorithmen. Ähnliche Ergebnisse sind in Tabellen 4.6 und 4.7 für den Vergleich mit der gewichteten Haut-ROI zu sehen.

5.3 Signalextraktion

Die Ergebnisse aus mehreren Experimenten mit verschiedenen Signalverarbeitungsverfahren wurden in der Tabelle 5.2 zusammengefasst. Zusätzlich wurden in der Tabelle verschiedene Eigenschaften der Algorithmen, wie die PPG-Signalverarbeitungsmethode, die Bandpassfrequenzen, Art der Bestimmung der Herzrate und verwendete Korrekturmethode aufgeführt. Während die PURE Datenbank eingeschränkte und kontrollierte Bewegungen enthält, haben die *BioVid* und *MMSE-HR* Datenbanken keine Einschränkungen für die Bewegung der Probanden. Dabei wurde für die PURE Datenbank auch explizit der *Steady* Teil getestet, in welchem sich die Probanden nicht bewegen.

Tabelle 5.2: Zusammenfassung der IEC Ergebnisse aus der Abbildung 4.24 und den Tabellen 4.5-4.7 und 4.9 für ausgewählte Algorithmen und deren PPG-Signalverarbeitungsmethode, Bandpassfrequenzen, Bestimmung der Herzrate (HR) und verwendeten Korrekturmethode.

Algorithmus	PPG Methode	Filter (in Hz)	HR Methode	Korrektur	Mittlere IEC Genauigkeit (in %)			
					PURE (<i>Steady</i>)	PURE	BioVid	MMSE-HR
Poh	ICA(RGB)	0,7-4	IBI	IBI NC-VT	85	76	76	82
Lewandowska	PCA(RGB)	0,5-3,7	Spektrum	-	89	62	76	68
DeHaan	CHROM	0,66-3	Spektrum	-	100	86	70	76
Feng_mod	aRG	adaptiv	IBI	-	83	77	81	74
Rapczynski	normG	adaptiv	IBI	IBI-Graph	100	85	90	89
Wang	IFFT(CHROM)	0,5-4	Spektrum	gewichtet	94	85	85	88
LSTM	CHROM	-	IBI	-	-	85	86	91
LSTM	normG	-	IBI	-	-	91	90	90
LSTM	RGB	-	IBI	-	-	72	90	87

Im Weiteren werden die Ansätze der *klassischen* Signalverarbeitung und die auf LSTM-Netzen basierenden Methoden getrennt betrachtet.

5.3.1 Klassische Verfahren

Auffällig ist das starke Abfallen der Genauigkeit bei den Verfahren, welche auf statistischen Signalverarbeitungsmethoden basieren. So zeigen *Lewandowska* und *DeHaan* bei den Datensätzen mit viel Bewegung (*BioVid* und *MMSE-HR*) im Mittel mit 72% und 73% deutlich schlechtere Ergebnisse, als auf den *PURE (Steady)* Daten. Da *Wang* ebenfalls den CHROM Ansatz

der Signalgenerierung nutzt, ist die sinkende Genauigkeit bei *DeHaan* auf eine nicht ausreichende Bewegungskompensation zurückzuführen, welche bei *Wang* durch eine Gewichtung der einzelnen Frequenzen vor der IFFT erreicht wird. Die Ergebnisse von *Poh* sind ebenfalls etwas stabiler, was vermutlich auf das verwendete Korrekturverfahren zurückzuführen ist.

Ähnliche Effekte konnten in Kapitel 4.8 bei den Untersuchungen zur Atemratenschätzung betrachtet werden. Verfahren, welche die RBG-Kanäle durch statistische Methoden kombinieren, wie *Poh*, oder *Sun*, zeigten deutlich schlechtere Ergebnisse, als das vorgestellte *FuseMod* Verfahren. Auch verwenden diese Algorithmen, wie *Sanyal*, einen statischen Bandpass gefolgt von einer spektralen Analyse. Dies führt, wie bei der Herzrate, zu schlechteren Ergebnissen. Ebenfalls fehlt in den Ansätzen eine Fehlerkorrektur, welche beim *FuseMod* durch die Unterdrückung der niederfrequenten Störungen erreicht wird. Der Ansatz von *VanGastel* lässt sich an dieser Stelle nur schlecht vergleichen, weil größere Unterschiede bei der Bestimmung der ROI vorliegen, deren Einfluss schwer bestimmbar ist.

Weiterhin zeigen adaptive Bandpässe ebenfalls gute Resultate bei der Bewegungskompensation. Der modifizierte *Feng_mod* Algorithmus erreicht ohne weitere Korrekturverfahren mit *Poh* vergleichbare Ergebnisse. Bei *Rapczynski* zeigt die Kombination eines flexiblen Bandpasses und einer folgenden Korrektur der Peakauswahl noch deutlich bessere Ergebnisse. Es liegt daher nahe, dass eine statische Bandpassfilterung nicht zur Kompensation von stärkeren Störsignalen ausreicht.

Die besten Algorithmen wurden unter Verwendung von Datensätzen mit starker Bewegung des Gesichts implementiert. Während *Rapczynski* den BioVid-Datensatz verwendet hat, wurde *Wang* mit Videodaten von Personen, welche auf einem Laufband liefen vorgestellt. Diese beiden Algorithmen erreichten durchgehend hohe Genauigkeiten (>85% IEC-Fehler). Die anderen Ansätze wurden auf Daten mit minimaler Bewegung des Kopfes entwickelt und konnten die starken Gesichts- und Kopfbewegungen in den hier verwendeten Datensätze kompensieren.

5.3.2 LSTM Netze

Die Signalverarbeitung mittels eines trainierten LSTM Modells wurde in Kapitel 4.7 untersucht. Die Abbildung 4.27 zeigt, dass das Verwenden von mehr Neuronen im Modell zu besseren Ergebnissen führt. Dabei ist jedoch zukünftig die Gefahr des Overfittings zu beachten. Ebenso hat die Verwendung von Trainingsgewichten (siehe Abb. 4.28) zu einer Verbesserung der Ergebnisse geführt, da die ungleiche Verteilung der nicht-Peak/Peak Samples in den Zieldaten kompensiert werden konnte.

Die Abbildung 5.7 zeigt die IEC Genauigkeiten der LSTM Modelle aus Tabelle 4.9 für verschiedene PPG Eingangssignale. Auf den verschiedenen Testdatenbanken werden dabei deutliche Unterschiede sichtbar. Auf der *BioVidEmo* erzielen die Modelle mit *normG* und *RGB* bessere Signale als das *CHROM* Verfahren, bei welchem bereits mehrere Vorverarbeitungsschritte, wie eine Bandpassfilterung, durchgeführt wurden. Bei der *BP4D+* erzielt das *CHROM* Verfahren die besten Resultate. Auf der *PURE* Datenbank erreichen die Modelle mit *RGB* deutlich schlechtere Ergebnisse als alle anderen Kombinationen und *normG* die besten.

Dies kann vermutlich auf die verwendeten Trainingsdaten zurückgeführt werden. Da die Modelle auf der *BioVid* trainiert wurden, haben die *BioVidEmo* Daten dasselbe Setting, Beleuchtung und Probanden. Dadurch ist weniger Generalisierung der Daten von dem Model erforderlich und die Eingangssignale mit dem höchsten Informationsgehalt erzielen die besten Ergebnisse. Die *BP4D+* Daten haben ein mit der *BioVid* vergleichbares Setting und Beleuchtung, unterscheiden sich jedoch im Inhalt und den Probanden. Daher scheint eine aufwendigere Vorverarbeitung der RGB Daten mit dem *CHROM* Verfahren zu einer besseren Generalisierung zu führen. Die

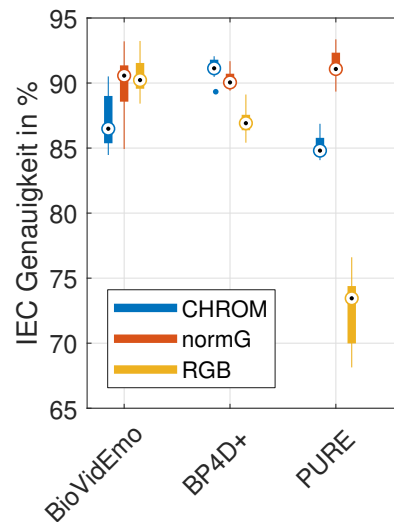


Abbildung 5.7: Boxplot der IEC Genauigkeit der LSTM Modelle aus Tabelle 4.9.

stärksten Unterschiede in den Daten liegen bei der *PURE* Datenbank vor. Die Videos haben deutliche Unterschiede beim verwendeten Setting, der Beleuchtung und den Videoeigenschaften (Codec, Auflösung, etc.). Durch die stark abweichenden Lichtverhältnisse scheint das Modell Schwierigkeiten bei der Verwendung der unverarbeiteten RGB Signale zu haben. Gleichzeitig scheint eine zu starke Vorverarbeitung, wie durch das *CHROM* Verfahren, zu viele Informationen aus den Signalen zu entfernen.

5.4 Atmung

In Kapitel 4.8 wurden die Ergebnisse der Atemratenschätzung dargestellt. Durch die Untersuchung verschiedener Algorithmen, PPG-Ansätze, Artefaktreduktionstechniken, Fusionsverfahren sowie weiterer Vor- und Nachbearbeitungsschritte wurde eine umfassende Untersuchung im Bereich der gesichtsvideobasierten Atemfrequenzerkennung im sichtbaren Lichtspektrum realisiert, die sowohl bestehende als auch neu entwickelte Methoden umfasst.

Es wurde speziell eine Datenbank (**AtemDB**) aufgenommen, welche explizit für die Validierung der kamarabasierten Atemratenschätzung erstellt wurde. Das Ziel war die Generierung hochqualitativer Daten mit wenigen Bewegungsartefakten, um einen Benchmark für die Grundlagen der Atemratenschätzung zu schaffen. Dies war notwendig, da es nur sehr wenige frei zugängliche Datenbanken gab, welche synchronisierte Video- und Atemsignale mit ausreichend hoher Qualität aufweisen konnten. Die Aufnahmen enthalten dabei sowohl *freies Atmen* als auch angeleitetes Atmen und decken ein breites Spektrum möglicher Atemfrequenzen ab. Die Atemratenschätzung wurde neben der **AtemDB** auch auf der **BP4D+** evaluiert, um bessere Aussagen über die Generalisierungsfähigkeit der untersuchten Verfahren zu treffen. Die **BP4D+** enthält Videos mit starkem Bewegungsanteil, während in der **AtemDB** die Probanden angehalten wurden, sich nicht zu bewegen.

Die erreichten Erkennungsraten (**DR**) des vorgestellten **FuseMod** Algorithmus liegen mit bis zu 90,1% weit über den Erkennungsraten des bestplatzierten Vergleichsalgorithmus von 59.1%. Von den verschiedenen Mittlungsme-

thoden des **FuseMod** erzielte der Median auf beiden Datenbanken bessere Erkennungsraten als der Mittelwert. Dies liegt an starken kurzfristigen Schwankungen der unterschiedlichen Modulationen, welche durch den Median besser ausgeglichen werden. Darüber hinaus wurde ein Verfahren entwickelt, die durch Bewegung und Traube-Hering-Mayer-Wellen verursachte Artefakte durch die Differenzierung des Signals vor der FFT-Analyse stark reduziert. Für die Länge des Zeitfensters zeigen die Daten, dass die Verwendung kürzerer Fenster für bewegte Daten und längerer Fenster für unbewegte Daten von Vorteil ist. Insgesamt erwiesen sich 30-Sekunden-Fenster als gut geeignete Länge.

Da die Güte des verwendeten PPG-Signals eine entscheidende Rolle für die weitere Analyse zur Atemratenschätzung spielt, sind die in Kapitel 5.3 beschriebenen Vor- und Nachteile der unterschiedlichen Signalverarbeitungsverfahren auch für die Atemratenschätzung gültig. Dies zeigt sich in den deutlich schlechteren Detektionsraten der Vergleichsalgorithmen. Die Verwendung von statistischen Verfahren oder fehlende Fehlerkorrektur der Daten führen bei den anderen Verfahren zu deutlich schlechteren Ergebnissen. Dies wird insbesondere an den großen Unterschieden in der DR (ca. -20%) zwischen der bewegungsreichen **BP4D+** und der statischen **AtemDB** deutlich (siehe Tabellen 4.17 und 4.18).

5.5 Lebenderkennung

In Kapitel 3.6.2 wurde ein System zur Kombination von kamerabasierter Vitalparameterschätzung aus RGB-Daten und 2,5D Tiefendaten zur Verhinderung der Überwindung von Gesichtserkennung vorgestellt und in Kapitel 4.9 getestet.

Das Modell zu Lebenderkennung ist als Proof-of-Concept konzipiert worden und erreicht auf Grundlage von kamerabasierter Vitalparameterschätzung eine Erkennungsrate von 87,1% bei der Unterscheidung von lebensechten 3D-Masken und menschlichen Gesichtern. Auffällig ist ein beobachtbarer Bias in Richtung Falsch-Negativer (Video: Gesicht, Klassifikation: Maske) Fehler. Diese treten im Verhältnis von 3,5:1 bis 1,2:1 mit Tendenz zur Klassifizierung als Maske auf (siehe Abb. 4.33-4.35). Dies ist vermutlich auf

die verwendeten Daten zurückzuführen, welche ein 2:1 Verhältnis von Gesichts/Masken Videos aufweisen und zu einer asymmetrischen Gewichtung der Klassen in den Trainingsdaten führen.

Dieses erste Modell zeigt, dass die Vitalparameterschätzung für die Identifikation von Menschen in sicherheitsrelevanten Settings eingesetzt werden könnte.

5.6 Messung im MRT

In Kapitel 4.10 wurde die Vitalparamterschätzung mittels einer monochromen Kamera in einem MRT untersucht. Die Ergebnisse zeigen, dass eine Messung der Herzrate auch mit monochromen Kameras mit hoher Genauigkeit und bis zu drei Spiegeln möglich ist.

Weiterhin sind im Versuch am MRT durch die große Distanz (3 m) und notwendige Kopffixierung die ROI entsprechend klein und große Teile des Gesichtes verdeckt. Dadurch zeigten sich eine Reihe von Herausforderungen, wie durch Blinzeln verursachte Störfrequenzen, welche bei einer kleinen ROI größeres Gewicht haben. So konnten die Bewegungen eines Probanden nicht durch die weitere Signalverarbeitung kompensiert werden. Durch das Fehlen von RGB Informationen und den starken Verdeckungen des Gesichtes wären daher neue Ansätze für die Generierung und Verfolgung der ROI im MRT notwendig, welche auch in Messungen im NIR Bereich Verwendung finden würden.

5.7 Multispektrale Messung

In Kapitel 4.11.1 wurde ein Experiment zur kamerabasierten kontaktlosen Schätzung der Herzrate im Nahinfrarotspektrum beschrieben. Trotz des geringeren Absorptionsverhaltens im nahinfraroten Bereich waren robuste Messungen möglich. Der Fehler der Messungen im Nahinfrarotbereich (900 nm, 2,75 BPM) war vergleichbar mit den Messungen im sichtbaren Spektrum (550 nm, 2,44 BPM).

Das erste durchgeführte Experiment basierte auf einer kleinen Datenbasis von neun Probanden. Um die Genauigkeit der Messungen im Nahinfrarotbereich weiter zu untersuchen, wurde ein weiterer Datensatz mit einem angepassten Kamera- und Beleuchtungsaufbau und 40 Personen durchgeführt. Dieser wurde in Kapitel 4.11.2 beschrieben.

Die Auswirkung der Wellenlänge auf die Schätzung der Herzfrequenz wurde anhand verschiedener Benchmark-Parameter bewertet. So wurde eine Korrelation zwischen der Wellenlänge und des Leistungsanteils der Herzfrequenzen gefunden (siehe Abb. 4.47). Der höchste Anteil der Herzfrequenzen lag im Bereich von 850 nm bis 875 nm. Dieser Effekt spiegelt sich auch in der Prominenz der Herzfrequenzen im Spektrum wieder (siehe Abb. 4.48). Die höchsten Werte wurden dafür im Bereich von 825 nm bis 875 nm ermittelt. Dies deutet darauf hin, dass die Wellenlänge für die Herzfrequenzschätzung im NIR-Spektrum um 850 nm bis 875 nm gewählt werden sollte.

6 Ausblick

Im Folgenden werden verschiedene Ansätze für zukünftige Forschung der kamerabasierten Vitalparameter beschrieben. Dabei werden weitere Verbesserungen der Signalverarbeitungskette, sowie neue Ansätze für die Messungen erläutert, wie die Messung der Herzratenvariabilität (HRV) oder Sauerstoffsättigung.

Neben den diskutierten Ansätzen ist der Forschungstransfer in verschiedenen Anwendungsszenarien, wie Sport, der Fahrerzustandserfassung in Fahrzeugen oder im Kontext der Mensch-Maschine-Interaktion, von großer Wichtigkeit. Die praktische Umsetzung der Forschung außerhalb des Labors zeigt neue Szenario-spezifische Herausforderungen an die Robustheit und Genauigkeit.

6.1 Videoeigenschaften

Es werden mehr Datensätze mit einer höheren Varianz der Bildinhalte (Probanden, Bewegung, Setting, ...) benötigt, um die in Kapitel 5.1 aufgestellten Hypothesen zu validieren und allgemeine Schlussfolgerungen über den genauen Einfluss der Videokompression auf die kamerabasierte Herzfrequenzschätzung zu ziehen. So sollte der Einfluss anderer Videoparameter, die über den Rahmen dieser Arbeit hinausgehen, getestet werden. Ein mögliches Beispiel wäre, den Parameter CRF_{max} gleich dem CRF-Wert zu setzen, um die Verringerung der Qualität bei Bewegung zu verhindern.

Die einzelnen Kompressionsmethoden, welche in modernen Videocodecs zum Einsatz kommen, sollten getrennt voneinander untersucht und der Einfluss auf die Herzfrequenzschätzung analysiert werden. So könnte eine Analyse der Schwankungen der Herzfrequenzgenauigkeit bei höheren

CRF-Werten unter Verwendung des x264-Codecs, wie in Abb. 4.8 und 4.9 dargestellt, zu deutlich kleineren Videodateien mit gut erhaltenen PPG-Informationen führen, wenn der beobachtete Effekt verstanden und zuverlässig reproduziert werden könnte.

Langfristig könnten so dedizierte PPG-Codecs entwickelt werden, die auf Erhaltung von PPG-Informationen spezialisiert sind. Dazu könnten verschiedene Ansätze, wie die Reduzierung der Videobitrate in ausschließlich nicht essenziellen Bereichen des Bildes, verwendet werden. Dies kann zum Beispiel durch den Einsatz von Gesichts- oder Hauterkennung während des Encodierverfahrens erreicht werden.

6.2 Region of Interest

Ein wichtiger Punkt, der in der bisherigen Forschung der ROIs noch nicht berücksichtigt wurde, sind Settings mit höherer Dynamik bezüglich Licht und Probandentätigkeiten. Weitere Datensätze werden benötigt, um die bisherigen Methoden für solch anspruchsvollere Situationen weiterzuentwickeln. Dazu könnte zum Beispiel die Vitalparametererkennung in sich bewegenden Fahrzeugen verwendet werden.

Eine weitere Möglichkeit zur Verbesserung der ROIs ist die Erforschung von dynamischen ROIs. Die im Stand der Technik vorgestellten ROIs basieren häufig nur auf Gesichtsländern oder der Hautfarbe. So werden äußere Faktoren wie die Kopffrotation oder die Richtung der Beleuchtung bei der Wahl der ROI Pixel nicht berücksichtigt. Diese und andere Faktoren könnten mit Verfahren aus der Literatur kombiniert werden, welche Regionen mit einem hohen SNR identifizieren. So könnte die Wahl der ROI Pixel mit einer Segmentierung des Gesichtes kombiniert werden, um dynamisch die ROI anzupassen und die Signalqualität zu verbessern. Dabei kann auch die Verwendung von tiefen neuronalen Netzen für die ROI Bestimmung betrachtet werden. Neben dem Erreichen von höheren Genauigkeiten sollte aber die Verarbeitungszeit nicht aus den Augen gelassen werden, um die entwickelten Ansätze realistisch einsetzbar zu lassen.

Die vorgestellte Hautfarbendetektion benötigt RGB Bilder für die Segmentierung der ROI Pixel. Für die präsentierten Experimente im Nahinfrarotspektrum sind diese daher nicht einsetzbar. Daher müssen für diese Daten landmarkenbasierte Verfahren genutzt werden, welche schlechtere Ergebnisse erzielen. Durch eine Kombination aus RGB Hauterkennung und landmarkenbasierten Verfahren, könnten diese Limitationen umgangen werden. So könnte ein auf maschinellem Lernen basierendes Verfahren entwickelt werden, welches anhand von RGB und NIR Daten trainiert wird und Hautpixel in NIR Bildern segmentieren kann, ohne auf RGB Bilder angewiesen zu sein. Alternativ könnte auch ein Transfer der Haut ROI anhand der Gesichtlandmarken auf die NIR entwickelt werden, wenn Daten für beide Spektralbereiche zur Verfügung stehen.

6.3 Signalverarbeitung

Viele verschiedene Ansätze der PPG Signalverarbeitung wurden in der Literatur vorgestellt. Diese weisen zum Teil stark unterschiedliche Vor- und Nachteile bezüglich verschiedener Parameter auf, welche noch nicht komplett nachvollzogen werden können. Weitere Untersuchungen und Vergleichsstudien auf mehr Daten sind nötig, um die genauen Abhängigkeiten modellieren und verstehen zu können. Wie bei den ROIs wurden auch für die Signalverarbeitungsmethoden bisher kaum Settings mit höherer Dynamik bezüglich der Beleuchtung und Probandentätigkeiten verwendet.

Die in Kapitel 4.7 untersuchten neuronalen LSTM Netze zeigten bereits eine hohe Genauigkeit und sollten weiter untersucht werden. Insbesondere sollten die erkennbaren Biases durch mehr Trainingsdaten ausgeglichen werden. Durch die Modifizierung von Auflösung und Videoqualität können weitere Daten für das Training generiert werden. Ebenfalls können die Eingangs- und Ausgangssignale optimiert werden. So könnten zum Beispiel mehrere PPG Varianten in das Netz gegeben werden, oder Zielpulsschläge statt als diskrete Impulse als (Normal-) Verteilungen modelliert werden, welche von den Netzen leichter erreicht werden können und eine bessere Aussage über die Trainingserfolge liefern.

6.4 Lebenderkennung

Das in Kapitel 3.6.2 vorgestellte Proof-of-Concept kann in zukünftigen Arbeiten durch eine Vielzahl an Ansätzen noch verbessert werden. Durch die Nutzung von Videos ohne verlustbehaftete Kompression und mit längerer Dauer kann die Erkennungsrate möglicherweise noch weiter gesteigert werden. Zudem könnte eine Datenbank mit sowohl 2,5D Tiefendaten, RGB-Videos und synchronisierten Biodaten aufgenommen und als zukünftiger Benchmark genutzt werden, um eine bessere Aussage über die Genauigkeit des Gesamtsystems treffen zu können.

Die momentan sich in der Entwicklung befindlichen *Curved Displays* stellen zukünftig eine weitere Herausforderung dar, wenn dreidimensional geformte flexible Bildschirme kostengünstig produziert und leicht zugänglich sein werden. Mögliche Weiterentwicklungen des Systems könnten weitere Körperpartien beinhalten, um höhere Sicherheitsstandards zu ermöglichen. Das Vergleichen der Pulssignale, des Gesichtes und einer Hand, durch die Kamera oder ein traditionelles Pulsoximeter würde ein Umgehen weiter erschweren.

6.5 Sauerstoffsättigung

Für die generelle Weiterentwicklung der kamerabasierten Vitalparametermessung ist zudem das mit Sauerstoff gesättigte Hämoglobin des arteriellen Blutes interessant. Oximeter zur Messung der Sauerstoffsättigung nutzen die unterschiedlichen Lichtabsorptionsverhalten des Hämoglobins aus und messen die Eigenschaften der Haut bei einer roten und einer nahinfraroten Wellenlänge, um Schlüsse auf die arterielle Sauerstoffsättigung (SpO_2) zu ziehen. Die kamerabasierte kontaktlose Messung des PPG-Signals im nahinfraroten Spektrum eröffnet dabei die Möglichkeit zur kontaktfreien Detektion der arteriellen Sauerstoffsättigung. So sollte für eine zukünftige kontaktlose Schätzung der arteriellen Sauerstoffsättigung untersucht werden, wie Videodaten des roten und nahinfraroten Spektrums kombiniert werden können, um eine Messung des SpO_2 Wertes zu ermöglichen.

6.6 Herzratenvariabilität

Die Herzfrequenzvariabilität (HRV) ist ein wichtiges Vitalzeichen, das direkte Rückschlüsse auf den körperlichen Zustand eines Patienten zulässt und als eines der wichtigsten Vitalparameter gilt. Für eine Messung der HRV ist eine zeitlich genaue Bestimmung der Herzschläge (Peaks) im PPG-Signal erforderlich. Eine Herausforderung besteht darin, dass bereits kleinste Bewegungen und/oder Gesichtsbewegungen und/oder Gesichtsausdrücke der Probanden zu Artefakten im PPG-Signal führen können, die größtenteils auf das wechselnde Reflexionsverhalten der beobachteten Hautbereiche zurückzuführen sind. Daher kann die hohe Messgenauigkeit der Herzfrequenz im Stand der Technik nur durch eine starke zeitliche Filterung erreicht werden.

Um Modelle zu trainieren, die invariant gegenüber diesen Bewegungsartefakten sind, eignen sich zum Beispiel tiefe neuronale Netze. Unter Verwendung von Verfahren zur 3D Kopfposeschätzung und der Action-Unit Erkennung (Gesichtsmuskelbewegungen), könnte in Zukunft ein System trainiert werden, um aus Videodaten bewegungsinvariante PPG-Signale zu bestimmen. Für den Ansatz können Informationen über die detektierten Hautregionen in den Bildern generiert und für die Bewegungskompensation verwendet werden. Diese Daten können als Input für ein auf zeitliche Signalverarbeitung optimiertes neuronales Netz dienen, wie es in Kapitel 3.4 vorgestellt wurde. Zudem können dabei neue ROI Segmentationsverfahren auf CNN-Basis entwickelt werden, um eine genauere Detektion der Haut in den Videobildern zu erzielen und auf diese Weise Störartefakte weiter zu reduzieren.

Literatur

- [ACW16] Robert Amelard, David A Clausi und Alexander Wong. »Spatial probabilistic pulsatility model for enhancing photoplethysmographic imaging systems«. In: *Journal of biomedical optics* 21.11 (2016), S. 116010 (siehe S. 30).
- [AP81] R. Rox Anderson und John A. Parrish. »The Optics of Human Skin«. In: *Journal of Investigative Dermatology* 77.1 (1981), S. 13–19. ISSN: 0022-202X. DOI: <https://doi.org/10.1111/1523-1747.ep12479191>. URL: <https://www.sciencedirect.com/science/article/pii/S0022202X15461251> (siehe S. 8, 9).
- [BE15] Ethan B Blackford und Justin R Estep. »Effects of frame rate and image resolution on pulse rate measured using multiple camera imaging photoplethysmography«. In: *Medical Imaging 2015: Biomedical Applications in Molecular, Structural, and Functional Imaging*. Bd. 9417. International Society for Optics und Photonics. 2015, S. 94172D (siehe S. 141).
- [Bec+17] Christoph Becker u. a. »Camera-based measurement of respiratory rates is reliable«. In: *European Journal of Emergency Medicine* (2017), S. 1. DOI: 10.1097/mej.0000000000000476 (siehe S. 29).
- [Bel] Fabrice Bellard. *FFMPEG*. <http://ffmpeg.org/>. Zugriff: 16. August 2020 (siehe S. 13, 71).
- [Blö+17] Timon Blöcher u. a. »An online PPGI approach for camera based heart rate monitoring using beat-to-beat detection«. In: *Sensors Applications Symposium (SAS), 2017 IEEE*. IEEE. 2017, S. 1–6 (siehe S. 19, 25, 91, 95, 99).

- [BTVo6] Herbert Bay, Tinne Tuytelaars und Luc Van Gool. »SURF: Speeded Up Robust Features«. In: *Computer Vision – ECCV 2006*. Hrsg. von Aleš Leonardis, Horst Bischof und Axel Pinz. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, S. 404–417. ISBN: 978-3-540-33833-8 (siehe S. 20).
- [Car99] Jean-Francois Cardoso. »High-Order Contrasts for Independent Component Analysis«. In: *Neural Computation* 11.1 (1999), S. 157–192. DOI: 10.1162/089976699300016863 (siehe S. 107).
- [Cas+18] Denisse Castaneda u. a. »A review on wearable photoplethysmography sensors and their potential future applications in health care«. In: *International Journal of Biosensors & Bioelectronics* 4.4 (2018). DOI: 10.15406/ijbsbe.2018.04.00125 (siehe S. 10, 18).
- [Cha+16] Peter H. Charlton u. a. »An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram«. In: *Physiological Measurement* 37.4 (2016), S. 610–626. DOI: 10.1088/0967-3334/37/4/610 (siehe S. 29).
- [Cha+18] Peter H. Charlton u. a. »Breathing Rate Estimation From the Electrocardiogram and Photoplethysmogram: A Review«. In: *IEEE Reviews in Biomedical Engineering* 11 (2018), S. 2–20. ISSN: 1941-1189. DOI: 10.1109/RBME.2017.2763681 (siehe S. 29, 49, 55, 64).
- [Cho+14] Kyunghyun Cho u. a. »Learning phrase representations using RNN encoder-decoder for statistical machine translation«. In: *arXiv preprint arXiv:1406.1078* (2014) (siehe S. 35).
- [Cho+15] Karan Chopra u. a. »A comprehensive examination of topographic thickness of skin in the human face«. In: *Aesthetic surgery journal* 35.8 (2015), S. 1007–1013 (siehe S. 9).
- [CJ10] Pierre Comon und Christian Jutten. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. 1st. USA: Academic Press, Inc., 2010. ISBN: 0123747260 (siehe S. 27).

- [CM18] Weixuan Chen und Daniel J. McDuff. »DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks«. In: *CoRR* abs/1805.07888 (2018). arXiv: 1805.07888. URL: <http://arxiv.org/abs/1805.07888> (siehe S. 31, 32).
- [Cor+18] Juan Abdon Miranda Correa u. a. »Amigos: A dataset for affect, personality and mood research on individuals and groups«. In: *IEEE Transactions on Affective Computing* (2018) (siehe S. 65).
- [CVS16] Peter H. Charlton, Mauricio Villarroel und Francisco Salguiero. »Waveform Analysis to Estimate Respiratory Rate«. In: *Secondary Analysis of Electronic Health Records* (2016), S. 377–390. DOI: 10.1007/978-3-319-43742-2_26 (siehe S. 29).
- [DJ07] M.E. Davies und C.J. James. »Source separation using single channel ICA«. In: *Signal Processing* 87.8 (2007). Independent Component Analysis and Blind Source Separation, S. 1819–1832. ISSN: 0165-1684. DOI: <https://doi.org/10.1016/j.sigpro.2007.01.011>. URL: <https://www.sciencedirect.com/science/article/pii/S0165168407000151> (siehe S. 28).
- [dJ13] Gerard de Haan und Vincent Jeanne. »Robust pulse rate from chrominance-based rPPG«. In: *IEEE Transactions on Biomedical Engineering* 60.10 (2013), S. 2878–2886 (siehe S. xxiii, 19, 23, 92, 95, 99, 107).
- [Ell87] Johann Heinrich Ellgring. »FACS [Facial Action Coding System]«. In: 1987 (siehe S. 67).
- [Fen+15] Litong Feng u. a. »Motion-resistant remote imaging photoplethysmography based on the optical properties of skin«. In: *IEEE Transactions on Circuits and Systems for Video Technology* 25.5 (2015), S. 879–891 (siehe S. 10, 20, 23, 89, 92, 94, 95, 99).
- [FFM] FFMPEG Wiki. *FFMPEG H.264 Video Encoding Guide*. Zugriff: 16. August 2020. URL: <https://trac.ffmpeg.org/wiki/Encode/H.264> (siehe S. 74).
- [FRA20] Marc-Andre Fiedler, Michal, Rapczynski und Ayoub Al-Hamadi. »Fusion-Based Approach for Respiratory Rate Recognition From Facial Video Images«. In: *IEEE Access* (**Impact Factor: 3.4**)

- 8 (2020), S. 130036–130047. DOI: 10.1109/ACCESS.2020.3008687 (siehe S. v, 19, 48, 106, 109).
- [FRA21] Marc-Andre Fiedler, Michal, Rapczynski und Ayoub Al-Hamadi. »Facial Video-Based Respiratory Rate Recognition Interpolating Pulsatile PPG Rise And Fall Times«. In: *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. 2021, S. 545–549. DOI: 10.1109/ISBI48211.2021.9434132 (siehe S. vi, 48, 106, 109).
- [FRSo1] Karl Pearson F.R.S. »LIII. On lines and planes of closest fit to systems of points in space«. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.11 (1901), S. 559–572. DOI: 10.1080/14786440109462720 (siehe S. 28).
- [GMR16a] Otkrist Gupta, Dan McDuff und Ramesh Raskar. »Real-Time Physiological Measurement and Visualization Using a Synchronized Multi-Camera System«. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2016, S. 46–53 (siehe S. 20).
- [GMR16b] Otkrist Gupta, Dan McDuff und Ramesh Raskar. »Real-time physiological measurement and visualization using a synchronized multi-camera system«. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2016, S. 46–53 (siehe S. 19, 30).
- [Goe+17] Matthias Goebeler u. a., Hrsg. *Basiswissen Dermatologie*. ger. Springer-Lehrbuch. 1 Online-Ressource (XVII, 335 Seiten), Illustrationen. Berlin, Heidelberg: Springer, [2017]. ISBN: 978-3-662-52811-2 (siehe S. 9).
- [HAM17] Guillaume Heusch, André Anjos und Sébastien Marcel. »A reproducible study on remote heart rate measurement«. In: *arXiv* (Sep. 2017). URL: <https://arxiv.org/abs/1709.00962> (siehe S. 65).
- [He+16] Kaiming He u. a. »Deep residual learning for image recognition«. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, S. 770–778 (siehe S. 35).

- [Her+17] Alberto Hernando u. a. »Finger and forehead PPG signal comparison for respiratory rate estimation based on pulse amplitude variability«. In: *2017 25th European Signal Processing Conference (EUSIPCO)* (2017). DOI: 10.23919/eusipco.2017.8081575 (siehe S. 29).
- [HNM15] Mohammad A. Haque, Kamal Nasrollahi und Thomas B. Moeslund. »Heartbeat Signal from Facial Video for Biometric Recognition«. In: *Image Analysis*. Hrsg. von Rasmus R. Paulsen und Kim S. Pedersen. Cham: Springer International Publishing, 2015, S. 165–174. ISBN: 978-3-319-19665-7 (siehe S. 19).
- [Hua+15] Lei Huand u. a. »Robst skin detectioniion in real-world images«. In: *Journal of Visual Communication and Image Representation* 29.1481 (Mai 2015), S. 147–152 (siehe S. 21).
- [Hue08] M Huelsbusch. »An image-based functional method for optoelectronic detection of skin-perfusion«. In: *RWTH Aachen (in German)* (2008) (siehe S. 23).
- [ICM16] Luca Iozzia, Luca Cerina und Luca T Mainardi. »Assessment of beat-to-beat heart rate detection method using a camera as contactless sensor«. In: *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the. IEEE.* 2016, S. 521–524 (siehe S. 19).
- [Int11a] International Electrotechnical Commission. »IEC 60601 Medical electrical equipment - Part 2-27: Particular requirements for the basic safety and essential performance of electrocardiographic monitoring equipment«. In: (2011) (siehe S. 63).
- [Int11b] International Telecommunication Union (ITU). »Recommendation BT.601, Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios«. In: (2011) (siehe S. 16).
- [Int15] International Telecommunication Union (ITU). »Recommendation BT 709, Parameter values for the HDTV standards for production and international programme exchange«. In: (2015) (siehe S. 16).

- [JH91] Christian Jutten und Jeanny Herault. »Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture«. In: *Signal Processing* 24.1 (1991), S. 1–10. ISSN: 0165-1684. DOI: [https://doi.org/10.1016/0165-1684\(91\)90079-X](https://doi.org/10.1016/0165-1684(91)90079-X). URL: <https://www.sciencedirect.com/science/article/pii/016516849190079X> (siehe S. 28).
- [JR02] Michael J Jones und James M Rehg. »Statistical color models with application to skin detection«. In: *International Journal of Computer Vision* 46.1 (2002), S. 81–96 (siehe S. 21).
- [JR99] Michael J. Jones und James M. Rehg. »Statistical color models with application to skin detection«. In: *Computer Vision and Pattern Recognition (CVPR)*. Bd. 1. IEEE, 1999, S. 274–280. DOI: 10.1109/CVPR.1999.786951 (siehe S. 22, 37, 38, 89).
- [Kar+13] Walter Karlen u. a. »Multiparameter Respiratory Rate Estimation From the Photoplethysmogram«. In: *IEEE Transactions on Biomedical Engineering* 60.7 (2013), S. 1946–1953. DOI: 10.1109/tbme.2013.2246160 (siehe S. 29).
- [Kaw+14] Michal Kawulok u. a. »Self-Adaptive Skin Segmentation in Color Images«. In: *Proceedings of the 19th Iberoamerican Congress (CIARP)*. Bd. 8827. Springer. Puerto Vallarta, Mexico, Nov. 2014, S. 96–103. DOI: 10.1007/978-3-319-12568-8_12 (siehe S. 21).
- [Kaw13] Michal Kawulok. »Fast propagation-based skin regions segmentation in color images«. In: *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. IEEE, 2013, S. 1–7. DOI: 10.1109/FG.2013.6553733 (siehe S. 21).
- [KD14] Neslihan Kose und Jean-Luc Dugelay. »Mask spoofing in face recognition and countermeasures«. In: *Image and Vision Computing* 32.10 (2014), S. 779–789 (siehe S. 56).
- [Koe+11] Sander Koelstra u. a. »Deap: A database for emotion analysis; using physiological signals«. In: *IEEE transactions on affective computing* 3.1 (2011), S. 18–31 (siehe S. 66).
- [Lai+20] Juho Laitala u. a. »Robust ECG R-peak detection using LSTM«. In: *Proceedings of the 35th annual ACM symposium on applied computing*. 2020, S. 1104–1111 (siehe S. 46).

- [Láz+14] Jesús Lázaro u. a. »Respiratory rate influence in the resulting magnitude of pulse photoplethysmogram derived respiration signals«. In: (Sep. 2014), S. 289–292 (siehe S. 29).
- [Lee+07] Han-Wook Lee u. a. »The Periodic Moving Average Filter for Removing Motion Artifacts from PPG Signals«. In: (2007) (siehe S. 55).
- [Lew+11] Magdalena Lewandowska u. a. »Measuring pulse rate with a webcam - A non-contact method for evaluating cardiac activity«. In: *Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on*. IEEE. 2011, S. 405–410 (siehe S. 19, 92, 95, 99).
- [Li+10a] Chunming Li u. a. »Distance regularized level set evolution and its application to image segmentation«. In: *IEEE transactions on image processing* 19.12 (2010), S. 3243–3254 (siehe S. 26).
- [Li+10b] Jin Li u. a. »Comparison of respiratory-induced variations in photoplethysmographic signals«. In: *Physiological Measurement* 31.3 (Okt. 2010), S. 415–425. DOI: 10.1088/0967-3334/31/3/009 (siehe S. 29).
- [Li+14] Xiaobai Li u. a. »Remote heart rate measurement from face videos under realistic situations«. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014, S. 4264–4271 (siehe S. 26, 93, 99).
- [Li+18] X. Li u. a. »The OBF Database: A Large Face Video Database for Remote Physiological Signal Measurement and Atrial Fibrillation Detection«. In: *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*. 2018, S. 242–249. DOI: 10.1109/FG.2018.00043 (siehe S. 70).
- [Liu+16] Siqi Liu u. a. »A 3D mask face anti-spoofing database with real world variations«. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2016, S. 100–106 (siehe S. 116).

- [LUO92] Lars-G. Lindberg, Hakan Ugnell und P. A Oberg. »Monitoring of respiratory and heart rates using a fibre-optic sensor«. In: *Medical & Biological Engineering & Computing* 30.5 (1992), S. 533–537. DOI: 10.1007/bf02457833 (siehe S. 55).
- [Mad+11] K. Venu Madhav u. a. »Estimation of respiration rate from ECG, BP and PPG signals using empirical mode decomposition«. In: *2011 IEEE International Instrumentation and Measurement Technology Conference* (2011). DOI: 10.1109/imtc.2011.5944249 (siehe S. 11).
- [MBE17] Daniel J McDuff, Ethan B Blackford und Justin R Estep. »The impact of video compression on remote cardiac pulse measurement using imaging photoplethysmography«. In: *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on*. IEEE. 2017, S. 63–70 (siehe S. 136).
- [Mer+11] David J. Meredith u. a. »Photoplethysmographic derivation of respiratory rate: a review of relevant physiology«. In: *Journal of Medical Engineering & Technology* 36.1 (2011), S. 1–7. DOI: 10.3109/03091902.2011.638965 (siehe S. 11, 29).
- [MGP14] Daniel McDuff, Sarah Gontarek und Rosalind W Picard. »Improvements in remote cardiopulmonary measurement using a five band digital camera«. In: *IEEE Transactions on Biomedical Engineering* 61.10 (2014), S. 2593–2601 (siehe S. 30).
- [MHP11] Jukka Määttä, Abdenour Hadid und Matti Pietikäinen. »Face spoofing detection from single images using micro-texture analysis«. In: *2011 international joint conference on Biometrics (IJCB)*. IEEE. 2011, S. 1–7 (siehe S. 56).
- [Mon+17] Hamed Monkaresi u. a. »Automated detection of engagement using video-based estimation of facial expressions and heart rate«. In: *IEEE Transactions on Affective Computing* 8.1 (2017), S. 15–28 (siehe S. 19).
- [Mor+18] Jermana Moraes u. a. »Advances in Photoplethysmography Signal Analysis for Biomedical Applications«. In: *Sensors* 18.6 (Sep. 2018), S. 1894. DOI: 10.3390/s18061894 (siehe S. 55).

- [MPS11] Luis F Corral Martinez, Gonzalo Paez und Marija Strojnik. »Optimal wavelength selection for noncontact reflection photoplethysmography«. In: *22nd Congress of the International Commission for Optics: Light for the Development of the World*. Bd. 8011. International Society for Optics und Photonics. 2011, S. 801191 (siehe S. 31).
- [NFFo6] Meir Nitzan, Igor Faib und Haim Friedman. »Respiration-induced changes in tissue blood volume distal to occluded artery, measured by photoplethysmography«. In: *Journal of Biomedical Optics* 11.4 (2006), S. 040506. DOI: 10.1117/1.2236285 (siehe S. 11, 29).
- [Nil+07] Lena M. Nilsson u. a. »Combined photoplethysmographic monitoring of respiration rate and pulse: a comparison between different measurement sites in spontaneously breathing subjects«. In: *Acta Anaesthesiologica Scandinavica* 0.0 (2007). DOI: 10.1111/j.1399-6576.2007.01375.x (siehe S. 18).
- [Nil13] Lena M. Nilsson. »Respiration Signals from Photoplethysmography«. In: *Anesthesia & Analgesia* 117.4 (2013), S. 859–865. DOI: 10.1213/ane.0b013e31828098b2 (siehe S. 18).
- [Nis+16] Humaira Nisar u. a. »Contactless heart rate monitor for multiple persons in a video«. In: *Consumer Electronics-Taiwan (ICCE-TW), 2016 IEEE International Conference on*. IEEE. 2016, S. 1–2 (siehe S. 19, 20).
- [Niu+19] Xuesong Niu u. a. »Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation«. In: *IEEE Transactions on Image Processing* 29 (2019), S. 2409–2423 (siehe S. 35).
- [Pau+17] Michael Paul u. a. »A camera-based multispectral setup for remote vital signs assessment«. In: *EMBECE & NBC 2017*. Springer, 2017, S. 968–971 (siehe S. 31).
- [PBCo5] Son Lam Phung, Abdesselam Bouzerdoum und Douglas Chai. »Skin segmentation using color pixel classification: analysis and comparison«. In: *IEEE transactions on pattern analysis and machine intelligence* 27.1 (2005), S. 148–154 (siehe S. 40).

- [PMP₁₀] Ming-Zher Poh, Daniel J McDuff und Rosalind W Picard. »Non-contact, automated cardiac pulse measurements using video imaging and blind source separation.« In: *Optics express* 18.10 (2010), S. 10762–10774 (siehe S. 19).
- [PMP₁₁] Ming-Zher Poh, Daniel J McDuff und Rosalind W Picard. »Advancements in noncontact, multiparameter physiological measurements using a webcam.« In: *IEEE Transactions on Biomedical Engineering* 58.1 (2011), S. 7–11 (siehe S. 19, 30, 92, 95, 99, 107, 114–116).
- [PS₀₀] Ioannis Pavlidis und Peter Symosek. »The imaging issue in an automatic face/disguise detection system.« In: *Proceedings IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications (Cat. No. PR00640)*. IEEE. 2000, S. 15–24 (siehe S. 56).
- [Rag+₁₃] R Raghavendra u. a. »3d face reconstruction and multimodal person identification from video captured using smartphone camera.« In: *2013 IEEE International Conference on Technologies for Homeland Security (HST)*. IEEE. 2013, S. 552–557 (siehe S. 56).
- [Ram+₁₂] M. Raghu Ram u. a. »A Novel Approach for Motion Artifact Reduction in PPG Signals Based on AS-LMS Adaptive Filter.« In: *IEEE Transactions on Instrumentation and Measurement* 61.5 (2012), S. 1445–1457. DOI: 10.1109/tim.2011.2175832 (siehe S. 55).
- [Rap+₁₄] Michal Rapczynski u. a. »Simultaneous multi-camera calibration based on phase-shift measurements on planar surfaces.« In: *IEEE International Instrumentation and Measurement Technology Conference, I2MTC 2014, Proceedings, Montevideo, Uruguay, May 12-15, 2014*. IEEE, 2014, S. 175–180. DOI: 10.1109/I2MTC.2014.6860728. URL: <https://doi.org/10.1109/I2MTC.2014.6860728> (siehe S. v).
- [Rap+_{16a}] Michal Rapczynski u. a. »Der Einfluss von Hautfarbensegmentierung auf die kontaktfreie Schätzung von Vitalparametern.« In: *22. Workshop Farbbildverarbeitung - Ilmenau*. 2016 (siehe S. vi, 19, 89).

- [Rap+16b] Michal Rapczynski u. a. »Multispektrale Vermessung der Haut zur Verbesserung kontaktloser Herzratenschätzung«. In: 22. *Workshop Farbbildverarbeitung - Ilmenau*. 2016 (siehe S. vi, 124).
- [Rap+18a] Michal Rapczynski u. a. »A Multi-Spectral Database for NIR Heart Rate Estimation«. In: *2018 25th IEEE International Conference on Image Processing (ICIP)*. Okt. 2018, S. 2022–2026. DOI: 10.1109/ICIP.2018.8451104 (siehe S. v, 124).
- [Rap+18b] Michal Rapczynski u. a. »How the Region of Interest Impacts Contact Free Heart Rate Estimation Algorithms«. In: *2018 25th IEEE International Conference on Image Processing (ICIP)*. Okt. 2018, S. 2027–2031. DOI: 10.1109/ICIP.2018.8451846 (siehe S. v, 19, 89, 116).
- [Rap+21] Michal Rapczynski u. a. »A Baseline for Cross-Database 3D Human Pose Estimation«. In: *Sensors (Impact Factor: 3.6)* 21.11 (2021), S. 3769. DOI: 10.3390/s21113769. URL: <https://doi.org/10.3390/s21113769> (siehe S. v).
- [RJ13] D.N. Rutledge und D. Jouan-Rimbaud Bouveresse. »Independent Components Analysis with the JADE algorithm«. In: *TrAC Trends in Analytical Chemistry* 50 (2013), S. 22–32. ISSN: 0165-9936. DOI: <https://doi.org/10.1016/j.trac.2013.03.013>. URL: <https://www.sciencedirect.com/science/article/pii/S0165993613001222> (siehe S. 25).
- [RLA19] Michal Rapczynski, Christopher Lang und Ayoub Al-Hamadi. »Verhinderung der Überwindung von Gesichtserkennung durch kamerabasierte Vitalparameterschätzung«. In: 24. *Workshop Farbbildverarbeitung*. 2019 (siehe S. vi, 116).
- [Ros+15] Maik Rosenberger u. a. »Nearfield sensing and actuation for multispectral imaging systems«. In: *2015 IEEE Sensors Applications Symposium (SAS)*. IEEE. 2015, S. 1–6 (siehe S. 124).
- [RPR12] Ajita Rattani, Norman Poh und Arun Ross. »Analysis of user-specific score characteristics for spoof biometric attacks«. In: *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE. 2012, S. 124–129 (siehe S. 56).

- [Rui+14] Luis M. Ruiz u. a. »Heart rate variability using photoplethysmography with green wavelength«. In: *2014 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)* (2014). DOI: 10.1109/ropec.2014.7036296 (siehe S. 10).
- [RWA16] Michal Rapczynski, Philipp Werner und Ayoub Al-Hamadi. »Continuous Low Latency Heart Rate Estimation from Painful Faces in Real Time«. In: *23th International Conference on Pattern Recognition (ICPR)*. 2016 (siehe S. v, 19, 22, 42, 73, 93, 95, 99, 101).
- [RWA19] Michal Rapczynski, Philipp Werner und Ayoub Al-Hamadi. »Effects of Video Encoding on Camera Based Heart Rate Estimation«. In: *IEEE Transactions on Biomedical Engineering (Impact Factor : 4.5)* 66.12 (2019), S. 3360–3370. DOI: 10.1109/TBME.2019.2904326 (siehe S. v, 13, 70).
- [SA14a] Frerk Saxen und Ayoub Al-Hamadi. »Color-based skin segmentation: an evaluation of the state of the art«. In: *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2014, S. 4467–4471 (siehe S. 37).
- [SA14b] Frerk Saxen und Ayoub Al-Hamadi. »Superpixels for Skin Segmentation«. In: *20. Workshop Farbbildverarbeitung*. 2014, S. 153–159 (siehe S. 21).
- [Sch+14] Marcus Schmidt u. a. »A real-time QRS detector based on higher-order statistics for ECG gated cardiac MRI«. In: *Computing in Cardiology Conference (CinC), 2014*. IEEE. 2014, S. 733–736 (siehe S. 63).
- [SFM18] Radim Spetlík R., Vojtech Franc und Jirí Matas. »Visual heart rate estimation with convolutional neural network«. In: *Proceedings of the British Machine Vision Conference, Newcastle, UK*. 2018, S. 3–6 (siehe S. 34).
- [SJR10] Janis Spigulis, Dainis Jakovels und Uldis Rubins. »Multi-spectral skin imaging by a consumer photo-camera«. In: *Multimodal Biomedical Imaging V*. Bd. 7557. International Society for Optics und Photonics. 2010, S. 75570M (siehe S. 30).

- [SMG14] Ronny Stricker, Steffen Müller und Horst-Michael Gross. »Non-contact video-based pulse rate measurement on a mobile service robot«. In: *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*. IEEE. 2014, S. 1056–1062 (siehe S. 22, 68).
- [SN18] Shourjya Sanyal und Koushik Kumar Nundy. »Algorithms for Monitoring Heart Rate and Respiratory Rate From the Video of a Users Face«. In: *IEEE Journal of Translational Engineering in Health and Medicine* 6 (2018), S. 1–11. DOI: 10.1109/jtehm.2018.2818687 (siehe S. 27, 49, 107, 114–116).
- [Sol+12] M. Soleymani u. a. »A Multimodal Database for Affect Recognition and Implicit Tagging«. In: *Affective Computing, IEEE Transactions on* 3.1 (Jan. 2012), S. 42–55. DOI: 10.1109/T-AFFC.2011.25 (siehe S. 65).
- [SP15] Anderson Santos und Helio Pedrini. »Human Skin Segmentation Improved by Saliency Detection«. In: *16th International Conference on Computer Analysis of Images and Patterns (CAIP)*. Bd. 9257. 2015. DOI: 10.1007/978-3-319-23117-4_13 (siehe S. 21).
- [Spi+07] Janis Spigulis u. a. »Simultaneous recording of skin blood pulsations at different vascular depths by multiwavelength photoplethysmography«. In: *Applied Optics* 46.10 (2007), S. 1754. DOI: 10.1364/ao.46.001754 (siehe S. 9).
- [Spi17] Janis Spigulis. »Multispectral, fluorescent and photoplethysmographic imaging for remote skin assessment«. In: *Sensors* 17.5 (2017), S. 1165 (siehe S. 30).
- [Sul+12] Gary J Sullivan u. a. »Overview of the high efficiency video coding(HEVC) standard«. In: *IEEE Transactions on circuits and systems for video technology* 22.12 (2012), S. 1649–1668 (siehe S. 13, 71).
- [Sun+11] Yu Sun u. a. »Motion compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise«. In: *Journal of biomedical optics* 16.7 (2011), S. 077010–077010 (siehe S. 107, 114–116).

- [Sun+12] Yu Sun u. a. »Noncontact imaging photoplethysmography to effectively access pulse rate variability«. In: *Journal of biomedical optics* 18.6 (2012), S. 061205 (siehe S. 30, 141).
- [Svi+18] Nina Sviridova u. a. »Photoplethysmogram at green light: Where does chaos arise from?«. In: *Chaos, Solitons & Fractals* 116 (2018), S. 157–165. DOI: 10.1016/j.chaos.2018.09.016 (siehe S. 10).
- [SZ14] Karen Simonyan und Andrew Zisserman. »Very Deep Convolutional Networks for Large-Scale Image Recognition«. In: *CoRR abs/1409.1556* (2014). URL: <http://arxiv.org/abs/1409.1556> (siehe S. 31).
- [Tar+17] Elizabeth A Tarbox u. a. »Motion correction for improved estimation of heart rate using a visual spectrum camera«. In: *SPIE Commercial+ Scientific Sensing and Imaging*. International Society for Optics und Photonics. 2017, S. 1021607–1021607 (siehe S. 19).
- [TD91] Carlo Tomasi und T Kanade Detection. »Tracking of point features«. In: *Int. J. Comput. Vis* (1991), S. 137–154 (siehe S. 21).
- [Ver+17] Wim Verkruysse u. a. »Calibration of contactless pulse oximetry«. In: *Anesthesia and analgesia* 124.1 (2017), S. 136 (siehe S. 30).
- [Vil+14] Mauricio Villarroel u. a. »Continuous non-contact vital sign monitoring in neonatal intensive care unit«. In: *Healthcare Technology Letters* 1.3 (Jan. 2014), S. 87–91. DOI: 10.1049/htl.2014.0077 (siehe S. 28).
- [Vil+97] J Vila u. a. »Time-frequency analysis of heart-rate variability«. In: *IEEE Engineering in Medicine and Biology Magazine* 16.5 (1997), S. 119–126 (siehe S. 93).
- [Vol+17] Mikhail V. Volkov u. a. »Video capillaroscopy clarifies mechanism of the photoplethysmographic waveform appearance«. In: *Scientific Reports* 7.1 (2017). DOI: 10.1038/s41598-017-13552-4 (siehe S. 10).

- [VSD16] Mark Van Gastel, Sander Stuijk und Gerard De Haan. »Robust respiration detection from remote photoplethysmography«. In: *Biomedical Optics Express* 7.12 (März 2016), S. 4941. DOI: 10.1364/boe.7.004941 (siehe S. 20, 106, 107, 114, 115).
- [VSN08] Wim Verkruyssen, Lars O. Svaasand und J. Stuart Nelson. »Remote plethysmographic imaging using ambient light«. In: *Optics express* 16.26 (2008), S. 21434–21445. URL: <http://www.opticsinfobase.org/abstract.cfm?URI=oe-16-26-21434-1> (besucht am 21.01.2014) (siehe S. 10, 18).
- [Wan+17] Wenjin Wang u. a. »Robust heart rate from fitness videos«. In: *Physiological Measurement* 38.6 (2017), S. 1023 (siehe S. 24, 93, 95, 99).
- [Wei+12] Lan Wei u. a. »Automatic webcam-based human heart rate measurements using laplacian eigenmap«. In: *Asian Conference on Computer Vision*. Springer, 2012, S. 281–292 (siehe S. 19).
- [Wer+13] Philipp Werner u. a. »Towards Pain Monitoring: Facial Expression, Head Pose, a new Database, an Automatic System and Remaining Challenges«. In: *Proceedings of the British Machine Vision Conference*. BMVA Press, 2013, S. 119.1–119.13. DOI: 10.5244/C.27.119 (siehe S. 66).
- [Wer+14] Philipp Werner u. a. »Automatic Heart Rate Estimation from Painful Faces«. In: *IEEE International Conference on Image Processing 2014 (ICIP 2014)*. Paris, France, Okt. 2014, S. 1947–1951 (siehe S. 19, 20).
- [WMS05] Fokko P Wieringa, Frits Mastik und Antonius FW van der Steen. »Contactless multiple wavelength photoplethysmographic imaging: a first step toward SpO₂ camera technology«. In: *Annals of biomedical engineering* 33.8 (2005), S. 1034–1041 (siehe S. 30).
- [YLZ19] Zitong Yu, Xiaobai Li und Guoying Zhao. »Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks«. In: *preprint arXiv: 1905.02419* (2019) (siehe S. 32).

- [Zha+16] Zheng Zhang u. a. »Multimodal spontaneous emotion corpus for human behavior analysis«. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, S. 3438–3446 (siehe S. 68, 107).
- [Zha20] Ayoub Al-Hamadi; Maria Nisser; Gunther Notni; Thomas Pertsch; Michal Rapczynski; Jan Sperrhake; Chen Zhang. »Verfahren und Vorrichtung zur kontaktfreien Bestimmung von zeitlichen Farb- und Intensitäts-veränderungen bei Objekten«. DE. 102020108064A1. 2020 (siehe S. vi).
- [ZRF94] Lingeng Zhao, Stanley Reisman und Thomas W. Findley. »Derivation of respiration from electrocardiogram during heart rate variability studies«. In: *Computers in Cardiology* (1994). DOI: 10.1109/cic.1994.470251 (siehe S. 10, 29).