



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Masterthesis

Florian Kletz

**Thermografische Gesichtsanalyse mit
einem multimodalen Kamerasystem**

*Fakultät Technik und Informatik
Studiendepartment Informatik*

*Faculty of Engineering and Computer Science
Department of Computer Science*

Florian Kletz

**Thermografische Gesichtsanalyse mit
einem multimodalen Kamerasystem**

Masterthesis eingereicht im Rahmen der Masterprüfung

im Studiengang Master of Science Informatik
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr. Kai von Luck
Zweitgutachter: Dr. Florian Vogt

Eingereicht am: 26. September 2017

Florian Kletz

Thema der Arbeit

Thermografische Gesichtsanalyse mit einem multimodalen Kamerasystem

Stichworte

Thermografie, maschinelles Sehen, Bildverarbeitung, Bildkalibrierung, Bildsynchronisierung, Bildregistrierung, Gesichtserkennung, Infrarot, Gesicht, Analyse, Multimodal, Fusion, Mensch-Computer-Interaktion, affective computing, emotionaler Zustand, Gesichtstemperatur

Kurzzusammenfassung

Dieses Dokument analysiert, designed, realisiert und evaluiert ein System zur thermografischen Gesichtsanalyse. Dazu werden Kameras verschiedener Modalitäten so miteinander verknüpft, dass Aufnahmen des visuellen Spektrums mit Aufnahmen einer Thermografiekamera angereichert werden. Hierzu sind mehrere Schritte wie eine Kamerakalibrierung, eine Bildsynchronisierung, eine Bildregistrierung und eine Gesichtserkennung nötig. Nach diesen Schritten ist es möglich, konkrete Temperaturwerte im Gesicht abzufragen. Die verschiedenen Module werden dazu analysiert und im Rahmen dieser Arbeit umgesetzt. Abschließend erfolgt eine Evaluierung der Module sowie der vorgeschlagenen Verfahren. Als Ergebnis geht hervor, dass eines der Verfahren geeignet ist, um eine thermografische Gesichtsanalyse durchzuführen.

Florian Kletz

Title of the paper

Thermal Facial Analysis with a Multimodal Camera System

Keywords

Thermography, computer vision, image processing, image calibration, image synchronization, image registration, facial recognition, infrared, face, analysis, multimodal, fusion, human-computer interaction, affective computing, emotional state, non invasive Temperature sensors, facial temperature

Abstract

This document analyses, designs, realizes and evaluates a system for the thermal facial analysis. To achieve this, cameras of different modalities are merged in such a way that images of the visual spectrum are enriched with data of a thermographic camera. To archive this, several process steps such as camera calibration, image synchronization, image registration and facial recognition are addressed. A combination of tools and algorithms makes it possible of retrieving temperature measures on precise surface areas of the human head. All modules are analysed and implemented within the scope of this work. Finally, all modules and a proposed procedure are evaluated. The result shows that one of the procedure is suitable to perform a thermographic facial analysis.

Inhaltsverzeichnis

1. Einleitung	1
1.1. Zielsetzung	2
1.2. Beitrag der Thermografie zu Human-Computer Interaction und Affective Computing	3
1.3. Gliederung	5
2. Analyse	6
2.1. Infrarotthermografie	6
2.1.1. Grenzen der Infrarotthermografie	8
2.1.2. Visualisierung von Thermografieaufnahmen	8
2.2. Gesichtserkennung	10
2.2.1. Gesichtserkennung auf Thermografiebildern	11
2.2.2. Gesichtserkennung auf visuellen Bildern	14
2.2.3. Einordnung und Zusammenfassung der Gesichtserkennung	20
2.3. Gesichtsatlas	23
2.4. Fusion multimodaler Aufnahmen	25
2.4.1. Kamerakalibrierung	25
2.4.2. Bildregistrierung	30
2.4.3. Zusammenfassung der Fusion multimodaler Aufnahmen	45
2.5. Verwandte Arbeiten	48
2.5.1. THERMAL FACE RECOGNITION IN AN OPERATIONAL SCENARIO	48
2.5.2. INFRARED HUMAN FACE AUTO LOCATING BASED ON SVM AND A SMART THERMAL BIOMETRICS SYSTEM	49
2.5.3. ADVANCED SURVEILLANCE SYSTEMS: COMBINING VIDEO AND THERMAL IMAGERY FOR PEDESTRIAN DETECTION	49
2.5.4. IMAGE FUSION OF VISIBLE AND THERMAL IMAGES FOR FRUIT DETECTION	50
2.5.5. NON-RIGID VISIBLE AND INFRARED FACE REGISTRATION VIA REGULARI- ZED GAUSSIAN FIELDS CRITERION	51
2.6. Szenarien- und Anforderungsanalyse	52
2.6.1. EMOTIONBIKE	52
2.6.2. Anforderungen	53
2.6.3. Abgrenzungen	54

3. Design & Realisierung	55
3.1. Designentscheidungen	56
3.2. Hardwaredesign	56
3.2.1. INFRA TEC VARIOCAM HD 875	56
3.2.2. ALLIED VISION MANTA G-235C	58
3.2.3. MICROSOFT KINECT FOR XBOX ONE	59
3.2.4. Setup	60
3.3. Softwaredesign	63
3.3.1. Architektur	64
3.3.2. Bilder- und Datenakquirierung	67
3.3.3. Gesichtserkennung	70
3.3.4. Temperaturextraktion	73
3.3.5. Datenstruktur und Datenausgabe	76
3.4. Softwarefusion	77
3.4.1. Kamerakalibrierung	77
3.4.2. Synchronisierung und Ausdünnung	82
3.4.3. Bildregistrierung	84
3.4.4. Zusammenfassung Softwarefusion	93
4. Evaluierung	95
4.1. Bildregistrierung	96
4.1.1. Bildregistrierung anhand von Silhouetten	96
4.1.2. Bildregistrierung anhand von Kanten	102
4.1.3. Zusammenfassung Bildregistrierung	106
4.2. Gesichtserkennung	107
4.2.1. Geschwindigkeit	108
4.3. ROI-Extraktion und Temperaturermittlung	109
4.3.1. Geschwindigkeit	110
5. Schlussbetrachtung	111
5.1. Zusammenfassung	111
5.2. Ausblick	114
A. Anlagen	116
A.1. Anlage 1 – Kamerakalibrierung	116
A.2. Anlage 2 – DLIB Landmarking Punkte	118
A.3. Anlage 3 – Datenstruktur	120
A.4. Anlage 4 – Registrierung	121
A.5. Anlage 4 – IRB Temperaturdaten	122

Tabellenverzeichnis

2.1. Einordnung der Komplexität für die Gesichtserkennung	21
2.2. Einordnung der Bildregistrierung	45
3.1. Technische Daten der INFRA TEC VARIOCAM HD 875	57
3.2. Technische Daten der ALLIED VISION MANTA G-235C	58
3.3. Technische Daten der MICROSOFT KINECT FOR XBOX ONE	59
3.4. Eigenschaften des verwendeten Kalibrierungsmusters	79
4.1. Quantitatives Registrierungsergebnis der Bildregistrierung anhand von Silhou- etten	96
4.2. Quantitatives Registrierungsergebnis der Bildregistrierung anhand von Kanten	102
A.1. Intrinsische Parameter der INFRA TEC VARIOCAM HD 875	116
A.2. Intrinsische Parameter der ALLIED VISION MANTA G-235C	116
A.3. Intrinsische Parameter der MICROSOFT KINECT FOR XBOX ONE	117
A.4. DLIB Annotierung der Identifikatoren 0 - 67	118

Abbildungsverzeichnis

1.1. Geräte zur Oberflächentemperaturmessung	3
2.1. Der Infrarotbereich im elektromagnetischen Spektrum	7
2.2. Verschiedene Falschfarbendarstellungen derselben Aufnahme	9
2.3. <i>Face detection</i> auf einem Thermografiebild	12
2.4. Beispielhafte <i>false positive</i> Annotierung eines <i>face landmarking</i> Algorithmus auf Thermografiebildern	13
2.5. „ <i>Face detection divided into approaches</i> “	15
2.6. Rotatorische <i>pitch</i> - (X-Achse), <i>yaw</i> - (Y-Achse) und <i>roll</i> -Bewegungen (Z-Achse) eines Kopfes	18
2.7. Anatomie des Kopfes mit dargestellten Arterien, Venen und Muskeln, sowie hervorgehobenen ROI	23
2.8. Zusammenspiel der Kameraparameter	28
2.9. Exemplarische Darstellung häufig verwendeter 2D-Kalibrierungsmuster	28
2.10. Schematische Darstellung von radial und tangential verzeichneten Abbildungsfehlern	30
2.11. Exemplarische Darstellung zur <i>Multi-Temporal-Fusion</i>	32
2.12. Exemplarische Darstellung eines Verfahrens zur <i>Fusion for image restoration</i>	32
2.13. Exemplarische Darstellung eines Verfahrens zur <i>Multi-Focus-Fusion</i>	33
2.14. Exemplarische Darstellung eines Verfahrens zur <i>Multi-Modal-Fusion</i>	33
2.15. Visualisierungen von Multikamerasystem-Aufbauten und deren Bezugsgrößen	35
2.16. Visualisierungen der Bildwinkel in einem Multikamerasystem	36
2.17. Darstellung der Fusion mittels Polarisationsstrahlteiler	38
2.18. Übersicht der 2D-Transformationen eines Rechtecks	42
2.19. Darstellung des Registrierungsprozesses	51
3.1. Front- und Rückansicht der INFRA TEC VARIOCAM HD 800 SERIE	57
3.2. VARIOCAM und MANTA-Kamera im Standalone-Aufbau	60
3.3. VARIOCAM und KINECT im EMOTIONBIKE Aufbau	61
3.4. Schichtendarstellung der einzelnen Elemente, die softwaretechnisch verarbeitet werden	63
3.5. Schematischer Ablauf mit den benötigten Teilkomponenten inkl. der verwendeten Frameworks	65
3.6. Verteilungssicht des Gesamtsystems (exklusive sonstiger EMOTIONBIKE Komponenten)	68
3.7. Exemplarische Darstellung des <i>face landmarking</i> erzeugt mittels DLIB	71

3.8.	Darstellung der berechneten ROI (teilweise nur rechte Gesichtshälfte) basierend auf den mittels DLIB erzeugten <i>face landmarking</i> -Punkten eines Gesichtes . . .	72
3.9.	Rechtecküberlagerung der definierten ROI zur Iteration	75
3.10.	Beispielbilder der grafischen Ausgaben	76
3.11.	Schachbrettmuster im visuellen- und im Thermografiebild, vor und während der Bestrahlung mit einem Halogenbaustrahler	78
3.12.	Visualisierung der extrinsischen Kameraparameter für die beiden verwendeten Szenarien	81
3.13.	Schematische Darstellung der Bildsynchronisierung	82
3.14.	Silhouettenerzeugung mittels Kontrast- und Helligkeitsmanipulation	85
3.15.	Ablauf der „Intensity-Based Automatic Image Registration“	86
3.16.	Schematischer Ablauf der Bildregistrierung anhand von Silhouetten	88
3.17.	Kantenerzeugung auf dem Thermografiebild	90
3.18.	Kantenerzeugung auf dem visuellen Bild	90
3.19.	Kantenregistrierung mittels „ <i>Robust Point Set Registration Using Gaussian Mixture Models</i> “	91
3.20.	Korrespondierende Punkte des visuellen Bildes vor und nach der Anwendung des „ <i>Robust Point Set Registration Using Gaussian Mixture Models</i> “-Verfahrens	92
4.1.	Exemplarischer Vergleich der Bildregistrierung anhand von Silhouetten zwischen zwei repräsentativen Aufnahmen als Schachbrettüberlagerungsdarstellung	97
4.2.	Darstellung der 2×3 -dimensionalen Transformationsmatrix für die affine Transformation mit den Durchschnittswerten und der Standardabweichung für die drei Probanden	99
4.3.	Darstellung der Zeitaufwände für die verschiedenen Schritte der <i>Bildregistrierung anhand von Silhouetten</i>	100
4.4.	Exemplarischer Vergleich der Bildregistrierung anhand von Kanten zwischen zwei repräsentativen Aufnahmen als Schachbrettüberlagerungsdarstellung	103
4.5.	Darstellung der 3×3 -dimensionalen Transformationsmatrix für die projektive Transformation mit den Durchschnittswerten und der Standardabweichung für die drei Probanden	104
4.6.	Darstellung der Zeitaufwände für die verschiedenen Schritte der <i>Bildregistrierung anhand von Kanten</i>	105
4.7.	Position der Nasenspitze	107
4.8.	Darstellung der Zeitaufwände für die verschiedenen Schritte der Gesichtserkennung	108
4.9.	Überlagerung der Thermografiebilder mit den ROI	109
A.1.	DLIB Annotierung der Identifikatoren 0 - 67	119

Listings

3.1.	Anweisung zur Extraktion von 30 Einzelbildern pro Sekunde aus einem Video	70
3.2.	Pseudocode des RAY-CASTING ALGORITHMUS	75
3.3.	OPENCV Methodensignatur zur Kontrast- und Helligkeitskorrektur	86
3.4.	Pseudocode für die Registrierung mittels Silhouetten	89
A.1.	Basis Datenstruktur als JSON	120
A.2.	ROI Element als JSON	120
A.3.	Verwendete Parameter für die <i>Registrierung auf Silhouetten</i>	121
A.4.	Verwendete Parameter für die <i>Registrierung auf Kanten</i>	121
A.5.	Aufbau der ASCII-Datei der Temperaturdaten	122

1. Einleitung

Mit dem Aufkommen der künstlichen Intelligenz auf der DARTMOUTH CONFERENCE im Jahr 1956 befasst sich die Informatik mit der Interpretation von Kamerabildern, welche das Forschungsgebiet des maschinellen Sehens (*Computer Vision*) begründen. Der Themenbereich zählt auch heute noch zu den aktiv betrachteten. Grund hierfür ist die zunehmend kostengünstigere Verfügbarkeit von verschiedenen optischen Geräten und deren universelle Einsetzbarkeit, aber auch die kontinuierliche Weiterentwicklung von Hardware und Algorithmen zur Bildverarbeitung. Die gestiegenen Verarbeitungskapazitäten heutiger Computer tragen den übrigen Teil dazu bei. Durch diese Aspekte ergibt sich fortwährend eine Vielzahl von neuen zivilen, wie auch polizeilichen und militärischen Anwendungsfeldern, in denen maschinelles Sehen nützlich ist. Die Themenbereiche sind vielfältigster Natur und beinhalten beispielsweise Personen- und/oder Gegenstandserkennung, Verkehrsüberwachung oder Sicherheitssysteme im Allgemeinen.

Viele dieser Systeme nutzen in diesem Kontext hauptsächlich Bildinformationen des visuellen Spektrums und sind so sehr nahe an den Bildinformationen, die auch ein Mensch verarbeitet. Besonders in Sicherheitssystemen werden auch andere Spektren des Lichts genutzt, wie beispielsweise das Infrarotspektrum, das einen Bereich abdeckt, der für das menschliche Auge nicht vollständig sichtbar ist. Dies ermöglicht den Systemen ein Sehen bei Dunkelheit oder eingeschränkter Beleuchtung.

Neben den rein sichtbaren Merkmalen, die Lebewesen oder Gegenstände aufweisen – die durch das klassische maschinelle Sehen erkennbar sind –, gibt es eine Reihe von weiteren Eigenschaften, die für eine tiefer gehende Analyse von Objekten von Interesse sind. Bei Lebewesen können dies Vitalwerte sein, bei Gegenständen beispielsweise die Materialbeschaffenheit. I. d. R. werden diese medizinischen oder technischen Eigenschaften durch invasive oder minimal-invasive Verfahren, jedoch mindestens durch direkten Kontakt gewonnen.

Der englischsprachige Ausdruck „*functional infrared thermal imaging*“ (fITI) umschreibt dazu ein Verfahren, das Temperaturinformationen mittels Infrarotthermografie nichtinvasiv aufzeichnet und extrahiert [Ioannou u. a. (2014)]. Diese Infrarotthermografie misst dazu die

von den Objekten ausgestrahlte Wärmestrahlung, die für das menschliche Auge, genau wie reguläre Infrarotstrahlung, nicht sichtbar ist.

Verfahren des maschinellen Sehens können mit den zuvor aufgeführten zusätzlichen Eigenschaften angereichert werden. Durch diese neue Informationsquelle können viele weitere Anwendungsfälle behandelt werden, die über das klassische maschinelle Sehen hinausgehen.

Einer dieser Anwendungsfälle ist die Temperaturmessung, um die lokale Körpertemperatur einer Person zu ermitteln. Der Schwerpunkt dieser Arbeit liegt auf der Gesichtstemperatur und der daraus resultierende Analyse des Gesichts. Neben der Feststellung der Temperatur und der damit einhergehenden Möglichkeit der medizinischen Diagnostik bieten Temperaturen – besonders jene des Gesichts – einen Indikator für den sog. *mental state* einer Person. Diese Indikatoren lassen sich nutzen, um personenbezogenen Approximationen von Emotions- bzw. Gefühlszuständen wie Stress, Angst, Erschrockenheit etc. ermitteln. Computersysteme können diese Informationen weiterverarbeiten, um sich auf Basis dieser individuell auf den Benutzer einzustellen. Dies Individualisierung wird als *affective computing* bezeichnet und bildet einen Teil des Forschungsbereichs der *Human-Computer Interaction (HCI)* [Picard (1999)].

Systeme die *affective computing* betreiben, versuchen Empathie zu simulieren, indem sie ähnlich wie Menschen auf Emotionen, Persönlichkeitsmerkmale, Fähigkeiten usw. des Gegenübers eingehen. Mit der kontinuierlichen Temperaturüberwachung einer Person bieten sich Möglichkeiten diese Systeme mit Daten anzureichern. Die eingangs erwähnte Approximation von Emotions- bzw. Gefühlszuständen kann dazu Dialogsysteme individualisieren, indem Inhalt und Menge von Informationen variiert werden. Assistenzsysteme (beispielsweise Fahrassistenzsysteme) können in diesem Zusammenhang eine teilweise oder eine komplette Entlastung des Fahrers vornehmen, sobald dieser mit der Situation überlastet zu sein scheint (beispielsweise aufgrund von Stress).

Die automatisierte Temperaturmessung des Gesichts bzw. einzelner Bereiche dessen mittels Thermografiekamera stellt bisher einen selten betrachteten Anwendungsfall dar. Diese Arbeit widmet sich diesem Thema unter Zuhilfenahme von weiteren Kameras des visuellen Spektrums, deren Aufnahmen mit den ermittelten Temperaturdaten angereichert werden, um so thermografische Analysen auf den visuellen Bildern zu ermöglichen.

1.1. Zielsetzung

Ziel dieser Arbeit ist es, ein System zu entwerfen und realisieren, dass aus Aufnahmen eines menschlichen Gesichts robust Temperaturwerte ermittelt. Hierzu soll eine Kombination von Kameras verschiedener Modalitäten genutzt werden, um sowohl die visuellen Informationen

einer gewöhnlichen Kamera bereitzustellen, als auch die korrespondierenden Temperaturwerte. Die Aufnahmen dieser Geräte sollen miteinander angereichert werden, um ihnen eine zusätzliche Dimension zu verleihen. Um präzise Bereiche des Gesichts zu analysieren, sind sog. *region of interest* (ROI) zu definieren und aus den Aufnahmen zu extrahieren. Die Temperaturwerte sollen dann weiteren Forschergruppen zur Interpretation zur Verfügung gestellt werden.

1.2. Beitrag der Thermografie zu Human-Computer Interaction und Affective Computing

Es gibt bereits eine Vielzahl von Anwendungsfällen im täglichen Leben, bei denen Temperaturinformationen eine bedeutende, wenn auch nicht immer offensichtliche Rolle spielen. So werden bei der industriellen Fertigung von Bauteilen und Lebensmitteln Temperaturmesssysteme eingesetzt, um die Produktion zu überwachen. Auch spielt die Temperaturmessung in der Veterinär- und Humanmedizin eine entscheidende Rolle zur Erkennung von Krankheiten und Infektionen oder bei der Tumorfrüherkennung [Lahiri u. a. (2012); Ioannou u. a. (2014)]. Die Temperaturen für diese Anwendungsfelder werden meist durch invasive Sensoren, die in das Objekt eindringen, dieses direkt berühren, oder mittels sog. *Pyrometer* (die eine punktgenaue, berührungslose Temperaturmessung ermöglichen) ermittelt [Gebhart u. a. (1980)]. Seit



(a) Sensoren



(b) Pyrometer

Quelle: U.S. NAVY © - bearbeitet



(c) Thermografiekamera

Abbildung 1.1.: Geräte zur Oberflächentemperaturmessung

einigen Jahren werden die Vorteile von Temperaturinformationen in der Baubranche unter der Bezeichnung *Bauthermografie* genutzt, um Leckagen in der Baustruktur oder Materialschäden zu erkennen. Dies wird häufig unter Zuhilfenahme von Thermografiekameras (die im Vergleich zu Pyrometern, Temperaturwerte größerer Bereiche liefern können) ermittelt. Ähnliche Systeme werden von Polizei, Feuerwehr und Militär genutzt, um bei Dunkelheit oder unzureichender Sicht Personen zu erkennen und dadurch aufzufinden [Rudol und Doherty (2008)]. Exemplarische Darstellungen von verschiedenen Geräten zur Temperaturmessung finden sich in Abbildung 1.1.

Im Kontext von HCI und *affektive computing* bieten Temperaturinformationen eine neue Möglichkeit für Personen mit Computersystemen zu interagieren. So kann durch die Temperatur der emotionale Zustand einer Person approximiert werden. Durch diese zusätzlichen Informationen über die Person hat das System ein besseres Verständnis über diese. Dadurch kann sich ein System situativ auf die Gegebenheiten einstellen und der Situation angepasste Informationen bereitstellen. Eine Weiterführung hiervon sind die sog. *Companion-Systemen* bzw. von *Companion-Technologie*, die von [SFB 62 (2009)] wie folgt definiert wurde:

„Companion-Systeme sind kognitive technische Systeme, die ihre Funktionalität konsequent und vollständig auf den individuellen Nutzer ausrichten, indem sie sich an seinen Fähigkeiten, Vorlieben, Anforderungen und aktuellen Bedürfnissen orientieren, sich auf seine Situation und emotionale Befindlichkeit einstellen, stets verfügbar, kooperativ und vertrauenswürdig sind und ihrem jeweiligen Nutzer als kompetente, partnerschaftliche Dienstleister gegenüberreten.“ – (SFB 62, 2009)

Die emotionalen Befindlichkeiten, die aus den Temperaturwerten interpretiert werden können, stellen also einen möglichen Beitrag für diese Systeme dar.

Auch für andere Bereiche der HCI bietet die Thermografie Anwendungsfälle. Ein Beispiel hierzu ist der Einsatz von Thermografie als biometrisches Merkmal in der Authentifizierung. Möglich macht dies die physikalische Eigenschaft, dass die Arterien- und Venenstruktur des Gesichts jedes Menschen – selbst bei eineiigen Zwillingen – unterschiedlich ist [Kong u. a. (2005)]. Mit einer hochauflösenden Thermografiekamera können diese Blutgefäße sichtbar gemacht werden und so die Person identifiziert bzw. wiedererkannt werden.

1.3. Gliederung

Diese Masterthesis besteht aus fünf Teilen, von denen dieser Teil die ‚**Einleitung**‘ bildet.

Im zweiten Teil ‚**Analyse**‘ werden die Anforderungen eines Systems analysiert, das eine thermografische Gesichtsanalyse mit einem multimodalen Kamerasystem ermöglicht. Dieser Teil analysiert dazu sowohl Anforderungen an die Software als auch an die Hardware und fasst diese am Ende nochmals zusammen.

Der dritte Teil ‚**Design & Realisierung**‘ stellt den Praxisanteil dieser Arbeit dar. In diesem werden die aus dem Analyseteil gewonnenen Erkenntnisse umgesetzt, um ein Systemdesign zu beschreiben und zu implementieren. Dazu wird neben der verwendeten Hardware auf die Algorithmen und Verfahren sowie die Tools eingegangen, die verknüpft werden, um das System zu implementieren.

Der vierte Teil ‚**Evaluierung**‘ bewertet die Leistungsfähigkeit der im Zuge dieser Arbeit entwickelten Komponenten. Die Bereiche Bildregistrierung, Gesichtserkennung und die ROI- bzw. Temperaturextraktion stehen hier besonders im Fokus.

Der letzte Teil ‚**Schluss**‘ rekapituliert die Ergebnisse der in den vorangegangenen Teilen durchgeführten Analyse und Entwicklung sowie deren Evaluation. Dazu fasst dieser Teil die Arbeit zusammen und bieten einen Ausblick für Anknüpfungspunkte.

2. Analyse

In diesem Teil der Arbeit werden die Anforderungen eines Systems analysiert, das eine thermografische Gesichtsanalyse mit einem multimodalen Kamerasystem ermöglicht. Dazu werden in verschiedenen Kapiteln Software- wie auch Hardwareaspekte betrachtet.

Gegliedert ist dieser Teil in sechs Kapitel. Das erste Kapitel behandelt die ‚Infrarotthermografie‘, das die Eigenschaften dieser Technologie zusammenfasst. Das Kapitel ‚Gesichtserkennung‘ grenzt die verschiedenen Arten der Gesichtserkennung voneinander ab und definiert die Unterschiede der Verfahren. Ebenso weist es auf die Grenzen und Herausforderungen der Gesichtserkennung hin – sowohl im visuellen Spektrum als auch im Thermografiespektrum – und ordnet die für diese Arbeit nötigen Verfahren ein. Das Kapitel ‚Gesichtsatlas‘ analysiert die anatomischen Gegebenheiten des menschlichen Gesichts. Basierend auf den Arterien- und Venenverläufen werden die sog. *regions of interest* (ROI), in denen eine thermografische Gesichtsanalyse durchgeführt werden kann, auf wissenschaftlicher Grundlage definiert. Das Kapitel ‚Fusion multimodaler Aufnahmen‘ analysiert und diskutiert, wie Aufnahmen verschiedener optischer Geräte miteinander fusioniert werden können. Untergliedert wird dieses Kapitel in die Abschnitte zur ‚Kamerakalibrierung‘, die eine baulich bedingte Linsenverzeichnung beseitigt, sowie zur ‚Bildregistrierung‘ in der eine Aufnahme so transformiert wird, um einer anderen – basierend auf den Bildinhalten – nahezu zu entsprechen. Dieser Aspekt ist nötig, um im Zusammenspiel mit der zuvor behandelten Gesichtserkennung die ROI auch auf andere Modalitäten zu übertragen. Als vorletztes Kapitel werden **verwandte Arbeiten** aufgeführt, aus denen wertvolle Informationen und Erkenntnisse mit in diese Arbeit einfließen. Abgeschlossen wird der Analyseteil mit einer Beschreibung des Szenarios, in dem dieses System primär verwendet werden soll, sowie einer Zusammenfassung der Anforderungen und Abgrenzungen des Themas dieser Arbeit (‚Szenarien und Anforderungsanalyse‘).

2.1. Infrarotthermografie

Die technologische Grundlage der Infrarotthermografie liegt in der physikalischen Eigenschaft, dass jeder Körper oberhalb des absoluten Nullpunkts (0 K, $-273,15\text{ }^{\circ}\text{C}$) eine elektromagnetische Strahlung aussendet. Dabei befindet sich die Infrarotstrahlung auf einer Wellenlänge

2. Analyse

zwischen $1 \mu\text{m}$ und $100 \mu\text{m}$ und ist somit zwischen dem sichtbaren Licht und den Mikrowellen angesiedelt, wie in Abbildung 2.1 illustriert. Zur Messung der Infrarotstrahlung eines Körpers wird die von ihm ausgestrahlte Energie durch eine spezielle Linse auf einen Infrarotdetektor innerhalb der Thermografiekamera fokussiert. Die Sensorelektronik übernimmt die Bildverarbeitung, wodurch nach Bestimmung der Strahlungsintensität die Temperatur des aussendenden Körpers ermittelt wird [Schuster und Kolobrodov (2004); InfraTec (2004, 2015b)].

Klassische Infrarotkameras zeichnen hingegen nicht die von den Körpern ausgestrahlte Wärmestrahlung auf, sondern – ähnlich wie Kameras des visuellen Spektrums – die Reflexionen einer separaten Lichtquelle (Infrarotstrahler) bzw. des noch vorhandenen Restlichts. Dieses Spektrum ist für den Menschen nur eingeschränkt und einige Tiere mit bloßem Auge nicht sichtbar, lässt sich aber mit einer Infrarotkamera sichtbar machen.

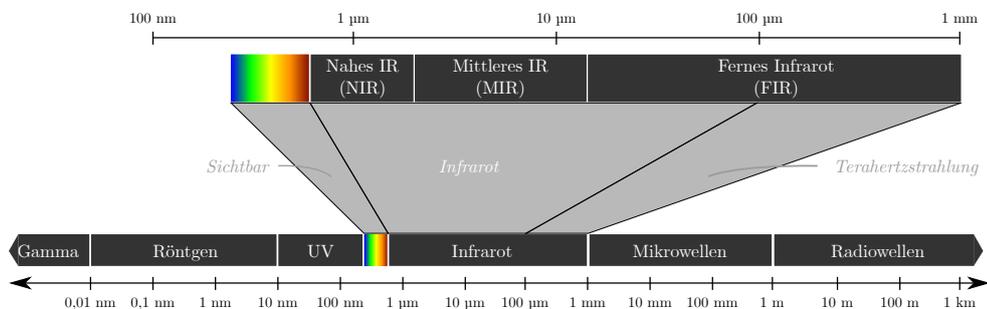


Abbildung 2.1.: Der Infrarotbereich im elektromagnetischen Spektrum

Als Temperaturmessmethode bietet sich die Infrarotthermografie aus zweierlei Gründen an. Zum einen handelt es sich um ein berührungsloses Messverfahren, bei dem kein direkter Kontakt zum Messobjekt vorhanden sein muss. Zum anderen handelt es sich um ein hochsensitives Messverfahren das es ermöglicht, die Temperaturen ganzer Flächen und Bereiche aufzuzeichnen. Grund hierfür ist die physikalische Gegebenheit, die aus der Gleichung 2.1 (Stefan-Boltzmann-Gesetz) hervorgeht.

$$I = \sigma \cdot \varepsilon \cdot T^4 \quad (2.1)$$

Diese besagt, dass die Strahlungsintensität (I) mit der Temperatur (T) in der vierten Potenz – unter Einbeziehung der Stefan-Boltzmann-Konstante (σ) sowie des Emissionskoeffizienten (ε) des zu messenden Materials – zusammenhängt. Dies bedeutet, dass eine Verdopplung der Temperatur des Messobjekts eine Vervierfachung der Strahlung bewirkt. Daraus resultiert, dass selbst kleinste Temperaturveränderungen registriert werden können [Spyra (2010)].

2.1.1. Grenzen der Infrarotthermografie

Trotz vieler Vorteile der Infrarotthermografie gibt es auch klare physikalische Grenzen und externe Faktoren, die beachtet werden müssen, wenn ein solches System eingesetzt wird. Diese Einflussfaktoren werden im Nachfolgenden kurz genannt und beschrieben:

Verdeckung Die größte Einschränkung bei der Nutzung eines solchen Systems ist eine Verdeckung des zu messenden Objekts. Eine Verdeckung unterscheidet sich bei der Infrarotthermografie grundlegend von der anderer bildgebender Verfahren. So kann die Infrarotstrahlung beispielsweise Glas nicht vollständig durchdringen, sondern reflektiert die Strahlung, was wiederum bei Nichtbetrachten zu Fehlmessungen führen kann. Andere Materialien wie beispielsweise dünne, blickdichte Polyethylenfolien können von der Infrarotstrahlung durchdrungen werden, obwohl diese nicht lichtdurchlässig sind. Hier spielen besonders die Größen *Reflexionsgrad*, *Absorptionsgrad* und *Transmissionsgrad* (häufig mit Φ_ρ , Φ_α und Φ_τ abgekürzt) des Materials eine zentrale Rolle [Zimmermann (2012)].

Externe Temperatureinflüsse Beachtet werden muss ebenso, dass das zu messende Objekt nicht von sog. Störstrahlungen beeinflusst wird. Dies können Sonnenstrahlen, aber auch andere in der Nähe des Messobjekts befindliche Wärmequellen sein, die dieses von außen erwärmen. Es gilt daher, Störstrahlung möglichst zu vermeiden. Ebenso können durch eine Änderung der Umgebung bzw. der Umgebungstemperatur ohne vorherige Akklimatisierung Temperaturveränderungen hervorgerufen werden.

2.1.2. Visualisierung von Thermografieaufnahmen

Mit dem bildgebenden Verfahren der Thermografie werden keine visuellen Informationen aufgezeichnet. Vielmehr wird die Intensität der Infrarotstrahlung, die von einem einzelnen Punkt ausgeht, als Temperatur an diesem Punkt gedeutet.¹ Dabei wird die für das menschliche Auge unsichtbare Infrarotstrahlung in elektrische Signale umgewandelt. Damit das menschliche Auge die Informationen, die mittels dieses Verfahrens aufgezeichnet werden, erfassen kann, wird den einzelnen Punkten in Abhängigkeit von der gemessenen Temperatur ein Farbwert zugewiesen. Ein Verfahren zur Zuweisung dieser Farbwerte wird als Falschfarbendarstellung oder farbcodierte Darstellung bezeichnet. Die Farbe gibt dabei keinen Aufschluss über die tatsächliche Temperatur. Sie dient lediglich zur kontrastreicheren Abgrenzung für das menschliche Auge und eine Analyse durch den Menschen [Schuster und Kolobrodov (2004)].

¹vgl. Gleichung 2.1

2. Analyse

Die Bilder in Abbildung 2.2 beinhalten beispielsweise alle dieselben Informationen. Für die verschiedenen Fotos wurde lediglich jeweils eine andere Farbcodierung gewählt. Die Legende eines Bildes gibt Aufschluss darüber, bei welchem Farbton es sich um welche Temperatur handelt. Dabei ist zu beachten, dass für das menschliche Auge verschiedene Farbtöne bzw. Nuancen der Abstufung nicht erkennbar sind; für einen Computer hingegen schon. Somit kann eine – für das menschliche Auge – gleichmäßige Färbung trotzdem verschiedene Temperaturen repräsentieren. Die Farben hierfür können dynamisch oder statisch festgelegt werden, wie bereits in [Kletz und Kleimann (2017)] beschrieben. Je nach Wahl der Farbcodierung kann der gesamte Bildbereich farblich codiert werden (wie in Abbildung 2.2a dargestellt) oder es können jene Temperaturen, die einen Schwellenwert überschreiten, farblich hervorgehoben werden (wie teilweise in Abbildung 2.2b zu sehen).

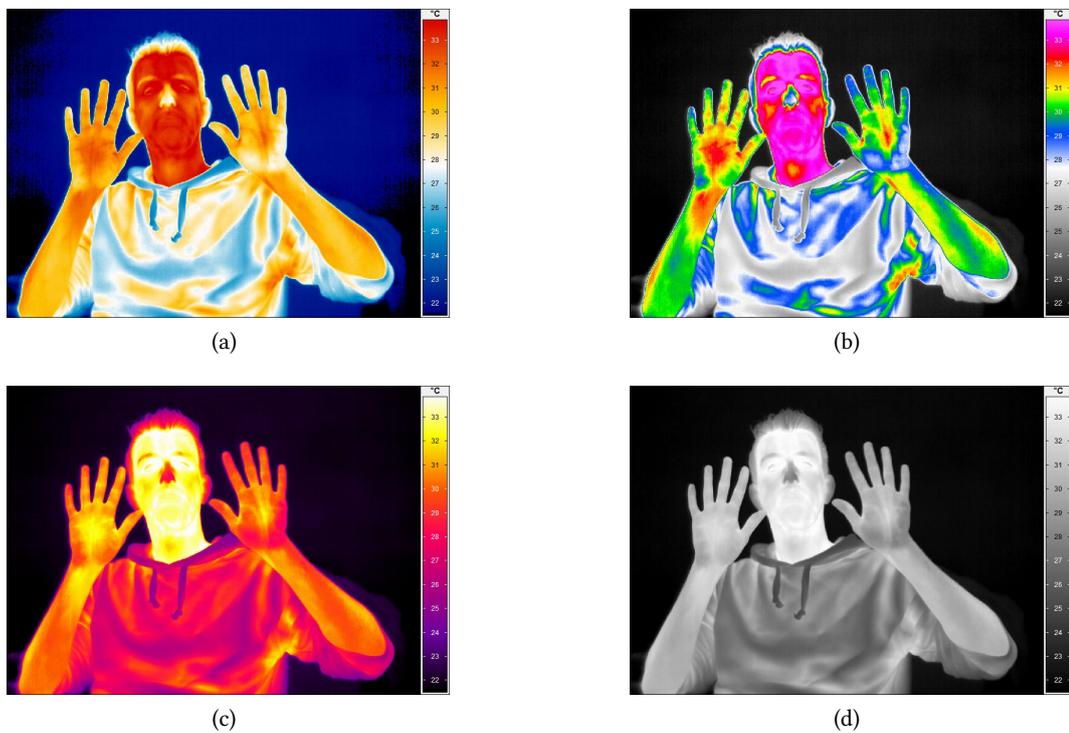


Abbildung 2.2.: Verschiedene Falschfarbendarstellungen derselben Aufnahme. Die tatsächliche Temperatur ergibt sich aus der Färbung und kann mittels der Legende bestimmt werden.

2.2. Gesichtserkennung

Die Gesichtserkennung ist eine der bedeutendsten Fähigkeiten, die Menschen besitzen, und daher seit langer Zeit ebenso ein Forschungsgebiet der Informatik. Forschungen zur (semi-) automatisierten Gesichtserkennung reichen bis in die 1960er Jahre zurück [Bledsoe (1966)].

Mit dem deutschen Begriff *Gesichtserkennung* und dem englischen Begriff *face recognition* können verschiedene Ausprägungen gemeint sein oder interpretiert werden. Diese bieten allesamt Antworten auf die Fragestellungen, ob ein Gesicht präsent ist und wer auf dem Bild zu erkennen ist, auf den Ausdruck der Person oder aber auf die Gesichtsform und die Position des Gesichts. Die Verfahren arbeiten i. d. R. anhand von Bildern, die im visuellen Spektrum aufgezeichnet sind. Im Nachfolgenden werden die – zum Teil auch in der Literatur unterschiedlich verwendeten – Bezeichnungen genauer voneinander abgegrenzt und erklärt bzw. definiert.

Face detection Unter der *face detection* wird der erste Schritt bei der Gesichtserkennung verstanden. Dieser besteht i. d. R. aus zwei aufeinanderfolgenden Schritten. Zu Beginn wird bestimmt, ob sich auf dem zu analysierenden Bild Gesichter befinden bzw. wie viele Gesichter erkennbar sind. Im zweiten Schritt wird die Position dieser bestimmt. Ziel ist es, alle Gesichter in einem Bildausschnitt unabhängig von deren Position, Skalierung, Alter, Geschlecht, Gesichtsausdruck, Gesichtshaarung usw. zu erkennen. Die Verfahren zur *face detection* sind typischerweise sehr zeitaufwendig, da in deren Rahmen das gesamte Bild nach Informationen durchsucht werden muss [Jain und Li (2005)]. Mit diesem Prozess werden daher nur die groben Positionen festgestellt. Technisch können diese durch verschiedene Verfahren ermittelt werden. Hierzu zählen Eigenschaften wie Hautfarbe (wenn Farbbilder genutzt werden), Bewegungen (in Videos), Gesichts- bzw. Kopfform, Aussehen des Gesichts oder aber Kombinationen dieser Eigenschaften.

Face tracking Unter *face tracking* kann einerseits das Synonym der allgemeinen *Gesichtserkennung* gemeint sein, andererseits aber auch das Nachverfolgen eines Gesichts im Bewegtbild. Auch hierbei gibt es verschieden starke Ausprägungen. Erkannt werden können einfache Bewegungen, beispielsweise in Richtung einer Kamera, wobei trotz der Größenänderung die Person immer als dieselbe Person (wieder-)erkannt wird. Aber auch Personen in einer Menschenmenge können mittels des Verfahrens von anderen Personen eindeutig unterschieden werden. Bei dem Prozess des *face tracking* spielen viele Faktoren der *face identification* zusammen. Typischerweise handelt es sich hier jedoch nicht um Systeme zur Authentifizierung und sind daher fehlertoleranter.

Face identification Mit der *face identification* lassen sich die erkannten Gesichter eindeutig einer bestimmten Person zuordnen und sie somit verlässlich wiedererkennen bzw. miteinander abgleichen. Genutzt werden können solche Verfahren zur Authentifizierung in Sicherheitssystemen wie beispielsweise dem EASYPASS²-System an Grenzkontrollen deutscher Flughäfen. Grundsätzlich sollte eine Verifikation von Gesichtern für Sicherheitssysteme jedoch immer nur als unterstützende Maßnahme eingesetzt werden [BSI (2003)].

Face landmarking Unter *face landmarking* wird eine Ausprägung der *feature extraction*, also der Lokalisierung von Merkmalen, verstanden. Dieses Verfahren ist eine Verfeinerung des *face detection*-Prozesses und basiert auf dessen Ergebnissen. Mit den für dieses Verfahren verfügbaren Algorithmen können die Organe/Merkmale des Kopfes erkannt werden. Hierzu zählen i. d. R. Lippen, Augen, Nase sowie Augenbrauen. In einigen Fällen sind ebenfalls die Ohren oder das Kinn bzw. die Umrisse des Kopfes im Allgemeinen inbegriffen. Der Detailgrad kann bei dem Verfahren stark variieren. So kann beispielsweise der Mund grob erkannt und mit einer geometrischen Figur (typischerweise rechteckig oder oval) umrandet werden. Genauere Verfahren können hingegen ein aus mehreren Punkten bestehendes Polygon an den Konturen der Lippen, der Nase, der Augen usw. platzieren.

2.2.1. Gesichtserkennung auf Thermografiebildern

Gesichtserkennung auf Thermografiebildern ist – im Gegensatz zur Gesichtserkennung auf visuellen Bildern – ein bisher nur rudimentär betrachteter Forschungsbereich der *Computer Vision*. Dies lässt sich vor allem auf den relativ hohen Preis der Thermografiegeräte zurückzuführen, der wiederum zu einer recht geringen Verbreitung führt. Ebenso führen Exportbeschränkungen sowie eine primäre militärische Nutzung von Thermografiekameras bis dato zu einer eingeschränkten zivilen Verbreitung. Ein weiterer Aspekt ist, dass die Auflösung von Thermografiekameras z. T. deutlich geringer ist als bei vergleichbaren visuellen Kameras [Wilder u. a. (1996)]. Diese Hürden flachen jedoch immer weiter ab. So sind heutzutage bereits einfache Thermografiekameras für unter 300 € im Handel erhältlich und das zivile Nutzungsspektrum vergrößert sich. Auch sind mittlerweile höhere Auflösungen möglich als noch vor wenigen Jahren. Dennoch liegen die Kosten von Thermografiekameras weiterhin über jenen von handelsüblichen Kameras des visuellen Spektrums, was maßgeblich an den hohen Rohstoffpreisen der verbauten Bauteile liegt [Reese u. a. (2012)]. Aufgrund der zunehmenden Verbreitung wird dieser Technologie in den nächsten Jahren wahrscheinlich eine größere Aufmerksamkeit gewidmet.

²<http://www.easypass.de>

Neben diesen Faktoren spielen aber auch Besonderheiten der visualisierten Temperaturdaten eine entscheidende Rolle bei der Gesichtserkennung mittels Thermografiebildern. So unterscheidet sich die Farbgebung von visuellen Bildern eklatant von den verschiedenen Falschfarbendarstellungen³. Dies bedeutet, dass robust funktionierende Gesichtserkennungssysteme für visuelle Bilder auf Basis der Falschfarbendarstellungen einer Thermografiekamera nur bedingt funktionieren [Kletz und Kleimann (2017)]. Dies schließt jedoch eine Gesichtserkennung auf Thermografiebildern nicht grundsätzlich aus. So kann man Gesichtserkennungssysteme mit Thermografiebildern der entsprechenden Falschfarbendarstellungen trainieren. Ein Verfahren dazu ist beispielsweise in [Xiaoyu u. a. (2006)]⁴ erläutert.

Beim Anlernen mit Thermografiebildern ist allerdings darauf zu achten, dass neben den verschiedenen Posen und Emotionen des Gesichts ebenso die Komplexität hinzukommt, dass diese Posen und Emotionen bei verschiedenen Temperaturen/Durchblutungen des Gesichtes aufgenommen werden müssen, um später bei jeglichen Temperaturen eine robuste Erkennung zu ermöglichen.



Abbildung 2.3.: *Face detection* auf einem Thermografiebild mittels VIOLA-JONES-METHODE [Viola und Jones (2001)] der Programmbibliothek OPENCV

Einige Verfahren, die auf visuellen Bildern funktionieren, ermöglichen allerdings (auch ohne das zusätzliche Anlernen von Thermografiebildern) eine Gesichtserkennung auf selbigen. Dies wird in Abbildung 2.3 exemplarisch dargestellt und in [Reese u. a. (2012)] dokumentiert. Bedingt ist es durch die Wahl der Falschfarbendarstellung und die dadurch entstehende Ähnlichkeit zu visuellen Bildern. Auch spielt das Szenario – wie auch bei visuellen Systemen – eine große Rolle. So kann bei schlechter oder stark variierender Beleuchtung eine Gesichtserkennung auf Thermografiebildern bessere Ergebnisse liefern, da diese beleuchtungsunabhängig sind

³vgl. Abschnitt 2.1.2 *Visualisierung von Thermografieaufnahmen*

⁴vgl. Abschnitt 2.5.2 *INFRARED HUMAN FACE AUTO LOCATING BASED ON SVM AND A SMART THERMAL BIOMETRICS SYSTEM*

und auch bei völliger Dunkelheit dasselbe Bild wie bei normaler Beleuchtung aufzeichnen [Torresan u. a. (2004)].⁵ Eine Über- oder Unterbelichtung, wie bei visuellen Aufnahmen, kann dadurch bei Thermografiebildern nicht auftreten. Befindet sich jedoch beispielsweise hinter der Person eine Wärmequelle mit einer dieser bzw. deren Kopftemperatur annähernd gleichenden Temperatur, so können diese Flächen auf dem Thermografiebild ineinander übergehen, was eine Erkennung erschwert und zu *false negatives* führt. Auch darf sich die Person nicht hinter einer Glasscheibe befinden, da die thermische Strahlung von Glas reflektiert wird und somit für die Thermografiekamera undurchsichtig erscheint. Dieses Problem besteht ebenso beim Tragen einer Brille.

Eine *face detection*⁷ auf Thermografiebildern ist somit für gewisse Anwendungsfälle eine durchaus praktikable Möglichkeit, um robust und valide die ungefähre Position des Kopfes bzw. der Person zu bestimmen. Da jedoch die einzelnen Organe des Gesichts sich im Vergleich zu visuellen Aufnahmen deutlich von der restlichen Fläche des Gesichts abheben (kein Schattenwurf unter der Nase oder unter den Augenbrauen; dafür – aufgrund des Temperaturunterschieds – eine anders gefärbte Nase), ist eine Bestimmung von detaillierteren Positionen/ROI problematisch und fehleranfällig. Die gängigen Verfahren können so beispielsweise keine Augen erkennen bzw. nicht die korrekte Position der Augen im Gesicht bestimmen. Ein *face land-*

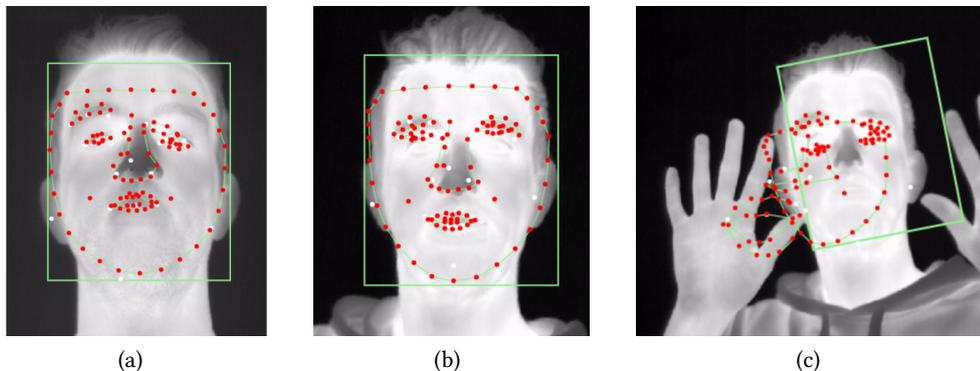


Abbildung 2.4.: Beispielhafte *false positive* Annotierung eines *face landmarking* Algorithmus⁶ auf Thermografiebildern: (a) Rechte Augenbraue und Mund falsch positioniert; (b) Augenbraue statt Auge annotiert; (c) Mund wurde außerhalb des Kopfes erkannt und positioniert.

⁵vgl. Abschnitt 2.5.3 ADVANCED SURVEILLANCE SYSTEMS: COMBINING VIDEO AND THERMAL IMAGERY FOR PEDESTRIAN DETECTION

⁶Exemplarisch generiert mittels des BETAFACE Algorithmus (www.betafaceapi.com/demo.html)

⁷vgl. Abschnitt 2.2 Face detection

*marking*⁸ und damit eine genaue Rückgabe von ROI ist somit nicht ohne Weiteres möglich. Viele Algorithmen erkennen keine Gesichtsmerkmale oder positionieren diese innerhalb des Gesichts falsch. Exemplarisch wird dies in Abbildung 2.4 gezeigt. *Face identification*⁹ hingegen ist in der Theorie durchaus möglich. So ist bekannt, dass die Adernstruktur des Gesichts bei jedem Menschen – selbst bei eineiigen Zwillingen – einzigartig ist [Kong u. a. (2005)]. Mit einer hochauflösenden Thermografiekamera kann diese sichtbar gemacht werden, wodurch Personen eindeutig identifiziert bzw. wiedererkannt werden können.

2.2.2. Gesichtserkennung auf visuellen Bildern

Die Gesichtserkennung auf visuellen Bildern ist ein seit Jahrzehnten betrachtetes Themengebiet, das aufgrund seiner vielfältigen Einsatzzwecke auch heutzutage noch immer ein interessantes Forschungsgebiet darstellt. Im Laufe der Jahre wurden speziell für die Gesichtserkennung eine Vielzahl von Algorithmen und Methoden entwickelt und weiterentwickelt. Trotzdem gibt es auch heute noch keine Universallösung, die für alle Anwendungsfälle gleich effizient und stabil funktioniert. Allerdings hilft es, das Umfeld, in dem eine Gesichtserkennung stattfinden soll, vorab zu kennen bzw. dieses festzulegen. So bieten verschiedenste Institutionen Wettbewerbe für Gesichtserkennungsimplementierungen an, um zu bestimmen, wie gut sich diese für die verschiedenen Szenarien eignen. Ein bekannter Vertreter dieser Implementierungen ist der sog. FACE RECOGNITION VENDOR TEST (FRVT) [NIST (2000–2017)] bzw. die FACE RECOGNITION GRAND CHALLENGE (FRGC) der US-amerikanischen Bundesbehörde NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST)¹⁰, die die Ergebnisse proprietärer Algorithmen kommerzieller Anbieter vergleicht. In Deutschland werden seit dem Jahr 2002 Gesichtserkennungsalgorithmen im Rahmen des Projekts BIOFACE [BSI (2003)] des BUNDESAMTS FÜR SICHERHEIT IN DER INFORMATIONSTECHNIK¹¹ untersucht.

Die Szenarien, die bei der Auswahl eines Algorithmus zur Gesichtserkennung eine Rolle spielen, können in drei wesentliche Merkmale unterteilt werden. Das erste Merkmal bildet die eingesetzte Kamera. Eigenschaften wie der Aufzeichnungswinkel der Szene spielen ebenso eine Rolle wie die Bildauflösung der Kamera. Das zweite Merkmal bildet der Bildausschnitt. Hier spielen Faktoren wie die Anzahl und die Größe der zu erkennenden Gesichter eine Rolle. Ebenso ist der Hintergrund von Bedeutung (Laborumgebung oder dynamischer Hintergrund). Das dritte Merkmal ist die Metaebene. Hierzu zählen Faktoren wie die Geschwindigkeit, mit der

⁸vgl. Abschnitt 2.2 Face landmarking

⁹vgl. Abschnitt 2.2 Face identification

¹⁰<http://www.nist.gov>

¹¹<http://www.bsi.de>

ein Gesicht erkannt werden soll/muss, oder das Auftreten bzw. die Anzahl von *false positive*- oder *false negative*-Fehlern.

Techniken zur Gesichtserkennung

Um automatisiert eine Gesichtserkennung durchzuführen, bieten sich eine Vielzahl von verschiedenen Herangehensweisen und Algorithmen an. Häufig wird hierbei auf erster Ebene zwischen den sog. *FEATURE-BASED APPROACHES* d. h. merkmalsbasierten Ansätzen und den

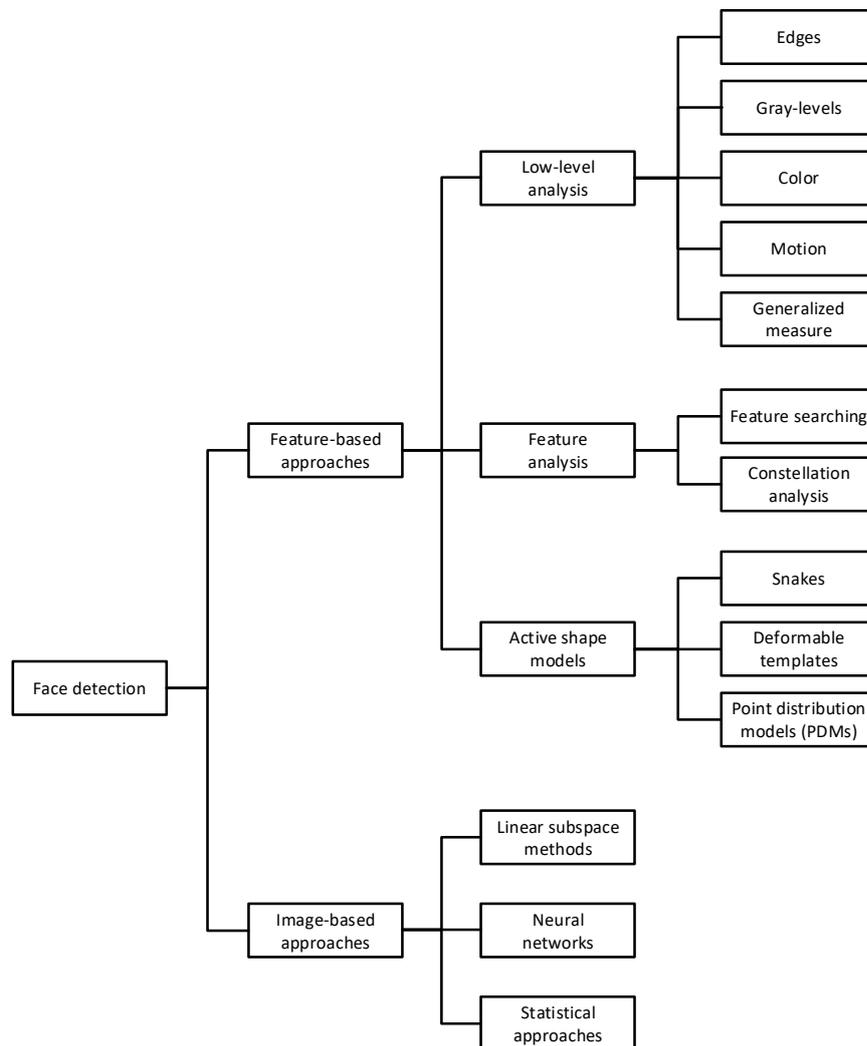


Abbildung 2.5.: „Face detection divided into approaches“ nach **Hjelmås und Low (2001)**

IMAGE-BASED APPROACHES, also bildbasierten Ansätzen unterschieden. In Abbildung 2.5 findet sich eine Hierarchie verschiedener Vorgehensweisen, die beispielsweise in [Hjelmås und Low (2001); Zhao u. a. (2003); Vijayakumari (2013); Tan u. a. (2006)] ausführlich beschrieben und verglichen werden und deshalb hier nur kurz vorgestellt werden. In [Jain und Li (2005)] finden sich darüber hinaus eine Vielzahl von Anwendungen, die diese Methoden nutzen.

Merkmalsbasierte Ansätze Die *merkmalsbasierten Ansätze* nutzen allgemeine Merkmale des menschlichen Gesichts wie dessen Größe, Verhältnisse und Abstände zwischen Augen, Nase, Mund und sonstigen sichtbaren Eigenschaften. Ebenso kann das Verhältnis von hellen und dunklen Gesichtsbereichen zueinander genutzt werden (beispielsweise Schattenwurf unter der Nasenpartie im Kontrast zu einer hellen Nasenspitze oder Schattenwurf auf die Augenhöhlen) oder die Eigenschaft ausgenutzt werden, dass Augenbrauen, Pupillen und Lippen i. d. R. dunkler erscheinen als der Rest des Gesichts. Je nach genutztem Medium kann auch die Hautfarbe (bei Farbbildern) oder die Bewegung (bei Videos) eine Gesichtserkennung ermöglichen. Diese Arten werden als *low-level Analysen* bezeichnet.

Bei der *Merkmalanalyse* werden auffällige Merkmale des Kopfes detektiert und davon ausgehend das Bild weiter durchsucht (z. B. in vertikaler Richtung), um weniger auffällige Merkmale zu erkennen. Basierend auf den Referenzabständen (den sog. *anthropometrischen Messungen*) von zwei erkannten Merkmalen werden weitere Merkmale gesucht. Hierdurch ist es möglich, dass Merkmale nur auf dem Kopf und nicht etwa im Hintergrund der Szene erkannt werden bzw. dass diese leicht verworfen werden können. Statt nur von einem Merkmal auszugehen, gibt es auch die Möglichkeit, das Gesicht in Bereiche aufzuteilen, die unter verschiedenen Blickwinkeln dieselbe Eigenschaft haben. Diese haben ebenso den Vorteil, dass so mögliche irrtümlich erkannte Merkmale (z. B. aus dem Hintergrund) schnell verworfen werden können.

Als dritte Gruppierung der merkmalsbasierten Ansätze gibt es die *aktiven Formmodelle*. Algorithmen dieser Gruppierung werden typischerweise genutzt, um die Kopfumrisse zu ermitteln. Der *Schlangenansatz* besteht darin, die ungefähren Umrisse zu wählen und diese anschließend durch Vergrößern oder Verkleinern der eigentlichen Kopfform anzugleichen. Dieses Verfahren kann zudem auf die anderen Merkmale des Gesichts angewandt werden. Mit dem Ansatz der *verformbaren Muster* werden die Schwierigkeiten, die der Schlangenansatz im Finden von Konturen bei schwachem Kontrast hat, gemildert. Dazu werden ähnlich wie beim Schlangenansatz Muster auf das Gesicht gelegt und diese vergrößert und verkleinert sowie verzerrt, um den Umrissen von Merkmalen des Gesichts zu entsprechen. Die letzte Variante dieser Gruppe sind die *verteilten Punktmodelle*, bei denen Punkte auf statistischer Grundlage Formen der Gesichtsmerkmale bilden. Diese lassen sich besser an die tatsächlichen

Gesichtsmerkmale anpassen als die anderen Methoden dieser Gruppe [Yow und Cipolla (1997); Hjelmås und Low (2001)].

Lernbasierte Ansätze Die *lernbasierten Ansätze* begegnen den Schwierigkeiten der merkmalsbasierten Ansätze im Hinblick auf die unvorhergesehenen Darstellungen des Gesichts und die auftretenden Umwelteinflüsse, beispielsweise im Hintergrund der Person. Diese Art der Ansätze werden auch als *bildbasierte Ansätze* bezeichnet. Alle Verfahren dieser Kategorie nutzen Wissen, das zuvor angelernt wurde. Dazu werden die Systeme mit einer großen Anzahl von Bildern trainiert, die sowohl Gesichter („positiv Beispiele“) zeigen als auch Bilder ohne jegliche Gesichter („negativ Beispiele“). Dazu nutzen viele dieser Verfahren ein Suchfenster, das das Bild Stück für Stück nach bekannten Mustern absucht [Belhumeur u. a. (1997); Hjelmås und Low (2001); Rowley u. a. (1998); Vijayakumari (2013)].

Grenzen und Herausforderungen der Gesichtserkennung auf visuellen Bildern

Unabhängig von der Wahl des Algorithmus zur Gesichtserkennung gibt es physikalische Grenzen, durch die eine Gesichtserkennung verhindert wird und Herausforderungen, die eine Gesichtserkennung stark einschränken und behindern können [Singh und Om (2013); Jain und Li (2005)]. Die größten Einschränkungen sowie die Herausforderungen werden im Nachfolgenden kurz betrachtet.

Feststellung der Kopfpose (Rotation) Die *Feststellung der Kopfpose* ist elementar entscheidend dafür, ob bzw. wie effizient ein Algorithmus ein Gesicht erkennen kann. Grundsätzlich unterscheiden viele Algorithmen bei ihrer Erkennungsstrategie, ob es sich um eine Erkennung der frontalen (dem sog. *passport photograph scenario*) oder der Profilsansicht des Kopfes handelt. Neben diesen beiden klar definierten Ansichten eines Kopfes gibt es zudem noch viele Zwischenstufen, wie der Kopf zur aufzeichnenden Kamera ausgerichtet sein kann. Klassischerweise werden die Rotationen der verschiedenen Achsen des Kopfes mit *pitch* (X-Achse), *yaw* (Y-Achse) und *roll* (Z-Achse) bezeichnet (vgl. Abbildung 2.6).

Algorithmen können Teile dieser Bewegungen ausgleichen. Allerdings sind diesen trotzdem Grenzen gesetzt. Spätestens, wenn sich das Gesicht vollständig aus dem Bereich dreht, ist offensichtlich keine Gesichtserkennung mehr möglich. Ausgelöst werden kann dies auch durch den Winkel, in dem die Kamera die Person aufzeichnet (beispielsweise, wenn es sich um eine Überwachungskamera handelt, die an der Decke befestigt ist). Bei der rotatorischen Bewegung handelt es sich um eine sog. formerhaltende Transformation („rigid“) des Kopfes, da dessen Form hier unverändert bleibt und sich lediglich der Blickwinkel auf diesen verändert.

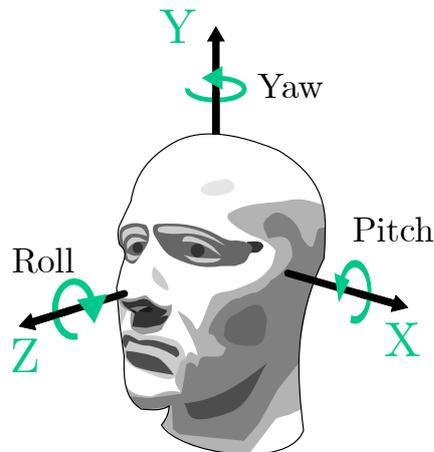


Abbildung 2.6.: Rotatorische *pitch*- (X-Achse), *yaw*- (Y-Achse) und *roll*-Bewegungen (Z-Achse) eines Kopfes.

Quelle: CHRKL © ⓘ - bearbeitet

Feststellen der Kopfpose (Translation) Neben der bereits beschriebenen rotatorischen Bewegung des Kopfes kann auch die translatorische Bewegung des Kopfes eine Herausforderung für den Gesichtserkennungsalgorithmus sein. Einige Algorithmen (vornehmlich jene des *face tracking*) verwenden einen Teilbereich der gesamten Aufnahme, um eine Gesichtserkennung durchzuführen. Dieser umfasst typischerweise die Position des Kopfes, der auf dem vorangegangenen Frame erkannt wurde, zuzüglich eines Schwellenwertes (der angenommenen maximalen Distanz einer Bewegung). Ist dieser Schwellenwert zu gering bzw. die Kopfbewegung größer als durch den Schwellenwert vorgesehen, kommt es zu *false negatives* oder zu rechen- und/oder zeitintensiven Anpassungen (neuer Durchlauf mit veränderten Schwellenwert, Durchsuchung der gesamten Aufnahme, etc.). Bei der translatorischen Transformation handelt es sich ebenso um eine formerhaltende Transformation, ähnlich jener der rotatorischen Transformation.

Umgang mit Verdeckung Ein Gesicht kann situativ bedingt teilweise oder ganz verdeckt werden. Die Verdeckung kann durch Gegenstände oder Personen die in das Bildfeld zwischen der Kamera und der Person gelangen ausgelöst werden. Neben der Verdeckung durch Dritte kann auch eine Verdeckung durch die Person selber geschehen. Diese Art der Verdeckung wird als Selbstverdeckung bezeichnet und kann beispielsweise durch Extremitäten geschehen. So können Selbstverdeckungen durch Hände und Arme geschehen, aber auch durch die Nase. Bei suboptimalen Blickwinkeln der Kamera kann die Nase große Flächen des Gesichts (wie

beispielsweise eine Wange oder ein Auge) verdecken, sodass diese Merkmale von der Kamera nicht aufgezeichnet werden können. Einige Algorithmen sind in der Lage, mit den vorhandenen Fragmenten weiterhin eine Gesichtserkennung durchzuführen. Sobald das Gesicht vollständig verdeckt ist, ist dies nicht mehr möglich.

Umgang mit Verdeckung mittels Brille Eine andere Art der Verdeckung der Augenpartie können *Brillen und Sonnenbrillen* darstellen. Die meisten Algorithmen sind in der Lage, trotz dieser eine Gesichtserkennung durchzuführen. Eine genaue Ortung der Augen ist hingegen schon weniger trivial. Sobald eine zu stark abgedunkelte oder gar reflektierende Sonnenbrille getragen wird, ist diese nicht mehr möglich.

Umgang mit mangelndem und zu starkem Kontrast und Helligkeit Eine der größten Herausforderungen, die an bildgebende Verfahren im Allgemeinen gestellt wird, ist die der ausreichenden Beleuchtung und des korrekten Kontrastes. Dieser Aspekt spielt bei der Gesichtserkennung eine maßgebliche Rolle. Ist das Gesicht stark unterbelichtet oder stark überbelichtet bzw. der Kontrast zu hoch oder zu gering, kann dies zu Problemen bei der Erkennung führen oder diese zum Scheitern bringen. Minimale Defizite der Beleuchtung und des Kontrasts können die meisten Algorithmen jedoch ausgleichen.

Umgang mit Bewegungsunschärfe Die bereits beschriebenen translatorischen und rotatorischen Bewegungen des Kopfes können je nach Geschwindigkeit in einer Bewegungsunschärfe resultieren. Bedingt wird dies maßgeblich durch eine schnelle Bewegung des Kopfes (oder der Kamera) bei langer Verschlusszeit der Kamera, die wiederum bei nicht ausreichender Beleuchtung nötig ist. Durch Unschärfe eines Bildes können Details verloren gehen. Dies kann in der Folge die Gesichtserkennung erschweren oder verhindern.

Präsenz von Gesichtsverformung aufgrund von Mimik Die Mimik ist ein für den Menschen bedeutendes nonverbales Kommunikationsmittel [Ellgring (2010)], das durch Muskelkontraktionen im Gesicht entsteht. Diese stellen dabei für die Gesichtserkennung eine besondere Herausforderung dar, da sie den Abstand zwischen Gesichtsmerkmalen oder die Position dieser auf dem Gesicht maßgeblich ändern und gleichzeitig das Gesicht weiterhin erkannt werden soll. Besonders für das *face landmarking* ist es besonders essenziell, die Mimik bzw. die Position der Gesichtsmerkmale – auch bei starkem Einsatz von Mimik – zuverlässig zu erkennen. Die Mimik stellt aufgrund der flüssigen („non-rigid“) Verformbarkeit eine anspruchsvollere Herausforderung dar als eine starre („rigid“) Änderung der Pose, die beispielsweise aufgrund translatorischer Bewegung entsteht.

Umgang mit Schärfentiefe Unter *Schärfentiefe* versteht man die Ausdehnung des scharfen Bereichs im Objektraum. Diese Eigenschaft wird von der Größe der Blendenöffnung und/oder der Länge der Brennweite bestimmt. Wird bei der Kamera eine kleine Blendenöffnung und/oder eine kurze Brennweite verwendet, herrscht eine große Schärfentiefe vor, was bedeutet, dass große Teile des Bildes scharfgestellt bzw. fokussiert sind. Bei großer Blendenöffnung und/oder langer Brennweite ist hingegen nur ein kleiner Bereich scharfgestellt. Nach Möglichkeit sollte daher das gesamte zu analysierende Objekt scharfgestellt aufgezeichnet werden, damit die Algorithmen die bestmöglichen Bedingungen haben, um eine Gesichtserkennung durchzuführen. Je nach verwendeter Hardware und Messaufbau kann es möglich sein, dass somit nur Teile des Bildes verwendet werden können.

Unterscheidung mehrerer Personen Eine weitere Herausforderung, die bei der Gesichtserkennung eine entscheidene Rolle spielt, ist die der Unterscheidung von mehreren Personen in derselben Szene. So können auf einer Aufnahme durchaus mehrere Personen abgebildet sein. Dabei sollte stets die korrekte Anzahl von Personen detektiert werden (weder *false negatives* noch *false positives*). Bei Systemen, die ein *face tracking* nutzen, kommt hierbei die Komplexität hinzu, die jeweilige Person zu verfolgen und nicht mit einer anderen Person, die ebenfalls im Bild ist, zu verwechseln. Eine Möglichkeit um die Personen untereinander unterschieden zu können, bietet der Einsatz von *face identification*-Algorithmen.

2.2.3. Einordnung und Zusammenfassung der Gesichtserkennung

In diesem Kapitel wurde die Gesichtserkennung näher betrachtet. Dazu werden die eingangs definierten Abgrenzungen, die im Rahmen der Gesichtserkennung vorgenommen werden können, hier wieder aufgegriffen. Dies geschieht im Rahmen einer Einordnung, die sich nach dem Aufwand, der Realisierbarkeit und der Güte der benötigten Daten – basierend auf der durchgeführten Analyse – richtet.

Die Einordnung findet für die Kategorien *face detection* (zusammenfassend für *Präsenzerkennung* und *Lokalisationsextraktion*), *face identification*, *face landmarking* sowie die Resistenz gegenüber Bewegungen (*Bewegungsresistenz*), d. h. bezüglich der Frage, wie tolerant ein System gegenüber Veränderungen der Kopfpose (translatorisch sowie rotatorisch) ist, statt. Die Zuordnung wird hierzu für die betrachteten Kameraarten vorgenommen (Thermografiekamera, Kamera im visuellen Spektrum und eine Kamera im visuellen Spektrum zuzüglich Tiefeninformationen). Bewertet wird nach den Stufen *leicht*, *mittel*, *schwer* und *sehr schwer*. Zu beachten ist dabei allerdings, dass dies nur eine grobe Einordnung ist, da es beim maschinellen Sehen

2. Analyse

immer stark auf die jeweilige Szene, den dazugehörigen Hintergrund sowie die umgebenden Faktoren ankommt.¹² Dargestellt sind die Einordnungen in Tabelle 2.2.3.

	Face detection	Face identification	Face landmarking	Bewegungs- resistenz
Thermografiekamera	○	+	--	-
Kamera im visuellen Spektrum	+	-	+	○/-
Kamera im visuellen Spektrum zzgl. Tiefen- informationen	+	○	+	+

+ = leicht; ○ = mittel; - = schwer; -- = sehr schwer

Tabelle 2.1.: Einordnung der Komplexität für die Gesichtserkennung

Aus der Tabelle ergibt sich, dass eine *face detection* mit einer Thermografiekamera zwar sehr wohl möglich ist (dies wurde ebenfalls in Abschnitt 2.2.1 gezeigt), dies allerdings nicht so robust funktioniert wie in vergleichbaren Aufnahmen im visuellen Spektrum. Der maßgebliche Grund hierfür ist, dass die *face detection*-Algorithmen für Aufnahmen im visuellen Spektrum optimiert sind. Dies bedingt ebenso, dass die beiden hier betrachteten Kameras des visuellen Spektrums in dieser Kategorie besser abschneiden.

Bei der *face identification* schneidet die Kamera im visuellen Spektrum ohne Tiefeninformationen mit der im Vergleich niedrigsten Wertung ab. Der Grund dafür ist, dass mit den anderen Modalitäten weitere Informationen vorliegen und dadurch eine robustere Identifikation ermöglicht wird (durch Tiefeninformation ermöglicht dies eine Unterscheidung zwischen Foto und realer Person und die Adern- und Venenstruktur ermöglichen eine hochgenaue Identifikation in der Thermografie¹³).

Das *face landmarking*, das für die präzise Zuordnung/Erkennung der Gesichtsmerkmale nötig ist, ist auf Aufnahmen einer Thermografiekamera gegenwärtig nur unter Umständen möglich. Dies ist maßgeblich bedingt durch die Falschfarbendarstellung und die daraus resultierenden abweichenden Kontraste zwischen den Gesichtsmerkmalen (verglichen mit Aufnahmen

¹² vgl. Abschnitt 2.1.1 Grenzen der Infrarot-Thermoografie und Abschnitt 2.2.2 Grenzen und Herausforderungen der Gesichtserkennung auf visuellen Bildern

¹³ vgl. Abschnitt 2.2.1 Gesichtserkennung auf Thermografiebildern

im visuellen Spektrum). Ebenso steht dieser Aspekt bisher nicht im Fokus des Forschungsbereichs um das maschinelle Sehen, weshalb bisher keine explizit auf diesen Anwendungsfall ausgerichteten Algorithmen verfügbar sind. In Bezug auf visuelle Bilder hingegen ist – auch aufgrund der Vielzahl von Anwendungsfällen – ein *face landmarking* hingegen weit verbreitet und robust nutzbar. Tiefeninformationen können diese Verfahren darüber hinaus wirksam mit ihren Informationen anreichern und dadurch beispielsweise robuster machen sowie beschleunigen. Als letzter Punkt wird die *Bewegungsresistenz* betrachtet. Hier wird die visuelle Kamera mit Tiefeninformationen am höchsten eingeordnet. Der Grund hierfür ist, dass durch diese zusätzlichen Informationen die Kopfpose weit mehr variieren kann als auf vergleichbaren planaren Aufnahmen. Bei der Thermografiekamera bedingen bereits die suboptimale *face detection* und das *-landmarking* das schlechtere abschneiden im Vergleich zu den anderen Modalitäten.

Zusammenfassend lässt sich sagen, dass im Rahmen dieser Thesis das präzise *face landmarking* eines der bedeutendsten Aspekte ist. Mittels der dadurch erreichten präzisen Lokalisierung von Gesichtsmerkmalen ist es möglich, an definierten Punkten des Gesichts Temperaturen zu messen, was u. a. das Hauptziel dieser Thesis darstellt. Ebenso ist es nötig, dass kleinere natürliche Kopfbewegungen tolerierbar sind, weshalb diesen beiden Aspekten die höchste Aufmerksamkeit in der Betrachtung gewidmet wird. Aufgrund der zuvor durchgeführten Einordnung folgt daher, dass das *face landmarking* anhand von Aufnahmen im visuellen Spektrum stattfinden muss. Zu begründen ist dies damit, dass im Thermografiespektrum die bisher verfügbaren Algorithmen nicht die gewünschte Robustheit aufweisen.

2.3. Gesichts atlas

Um eine thermografische Gesichtsanalyse durchführen zu können, ist ein Verständnis der Gesichtsanatomie notwendig, auf das in diesem Kapitel näher eingegangen wird. Medizinisch definiert sich das Gesicht aus der Vorderfläche des Kopfes, die im Wesentlichen aus den Gesichtsmerkmalen Mund, Nase und Augen besteht. Das Fundament des Gesichts besteht aus den Knochen des Gesichtsschädels und den Nasenknorpeln. Über diesen liegen die Weichteile der Muskulatur und Fettkörper, die von Arterien und Venen sowie Nerven durchzogen sind. Darüber liegt die Gesichtshaut [Anderhuber u. a. (2012); Lynch (2006)].

In diesem Kontext sind besonders die Arterien und Venen von Interesse, da der Blutdurchfluss dieser (neben den Schweißdrüsen) maßgeblich für die Temperatur der Gesichtsoberfläche verantwortlich ist. Die Abbildung 2.7 illustriert die Gesichtsarterien und -venen sowie die Gesichtsmuskulatur.

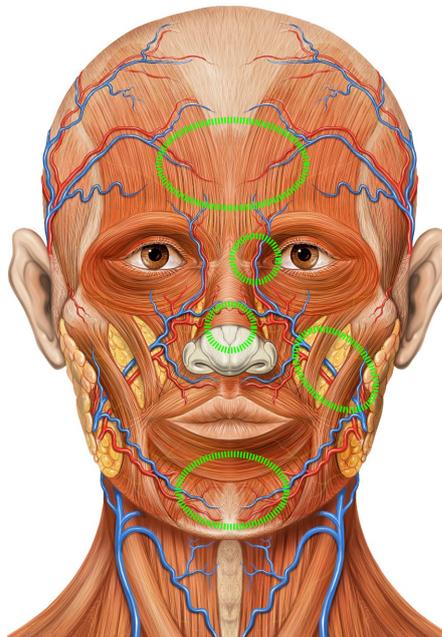


Abbildung 2.7.: Anatomie des Kopfes mit dargestellten Arterien (rot), Venen (blau) und Muskeln (braun) sowie hervorgehobenen ROI (teilweise nur in der rechten Gesichtshälfte dargestellt) (grün)

Hintergrundgrafik: Lynch (2006) © ⓘ

Verschiedene Forschungen und Ausarbeitungen haben sich bereits mit den Bereichen im Gesicht – den ROI – befasst, die sich für eine Temperaturmessung besonders eignen. Aus diesem Grund behandelt diese Arbeit nicht die Auswahl und die Eignungsfeststellung der

ROI, sondern die technische Extraktion dieser Bereiche bzw. im Weiteren die Extraktion der dortigen Temperaturwerte.

Neben Bestimmung der Temperaturwerte der ROI und der dadurch ermittelten absoluten Temperatur können diese Werte in den verschiedenen Regionen auch Aufschluss über den sog. *mental state* bzw. den emotionalen Zustand einer Person liefern. So zeigen verschiedene Forschungen, dass basierend auf der Änderung der Gesichtstemperatur u. a. Stress, Angst, Erschrockenheit, Freude, Schmerz etc. erkannt werden können [Ioannou u. a. (2014); Spyra (2010); Lahiri u. a. (2012)].

Die ROI definieren sich maßgeblich aus den Lagen der Arterien und Venen innerhalb des Gesichts. So definiert [Ioannou u. a. (2014)] – in Form einer Zusammenfassung verschiedener Forschungen – die folgenden ROI, die ebenfalls in Abbildung 2.7 eingezeichnet sind:

- Regio frontalis (Stirn)
- Regio depressor supercilii (Bereich zwischen Nase und Augen)
- Regio nasalis (Nasenspitze)
- Regio buccalis (Wangenregion)
- Regio mentalis (Kinnregion)

Da das Gesicht bei gesunden Menschen über eine nahezu symmetrische Temperaturänderung verfügt, ist es medizinisch unerheblich, welche Gesichtshälfte bei der Temperaturmessung gesunder Menschen betrachtet wird. So ändert sich die Temperatur i. d. R. in beiden Seiten des Gesichts gleichmäßig. Verändert werden kann dieses Verhalten durch Krankheiten wie beispielsweise das sog. *Harlekin-Syndrom* oder Tumore im Gesicht [Spyra (2010)]. Diese können dafür sorgen, dass Bereiche des Gesichts sowohl eine erhöhte Temperatur aufweisen (vornehmlich durch erhöhten Blutdurchfluss) aber auch eine verminderte Temperatur aufgrund von verminderten Blutdurchfluss eine Folge hiervon sein. Angewandt wird dieses Verfahren auch in der Brustkrebsfrüherkennung und dortigen Auffälligkeiten in der asymmetrischen Wärmeverteilung der Körperhälften [Gautherie und Gros (1980)].

2.4. Fusion multimodaler Aufnahmen

In diesem Kapitel wird analysiert, welche Schritte nötig sind, um Aufnahmen, die unter verschiedenen Modalitäten aufgezeichnet werden, präzise miteinander zu fusionieren. Unter einer Fusion wird der Vorgang bzw. das Ergebnis des Vorgangs der Überlagerung mehrerer Bilder oder sonstiger Informationen verstanden, bei dem ein neues oder ein Bild mit zusätzlichen Informationen erzeugt wird [Yufeng (2011); Olesch (2014)]. Bei der Fusion von Aufnahmen kann es sich sowohl um Aufnahmen einer Modalität handeln, aber auch um welche verschiedener Modalitäten. In diesem Kontext werden die nötigen Schritte einer Fusion von Kamerabildern betrachtet, die im sichtbaren Spektrum sowie im Spektrum der Infrarotthermografie generiert werden was einer Fusion multimodaler Aufnahmen entspricht.

Im Wesentlichen sind für den Prozess zwei Schritte nötig, die nachfolgend erläutert werden. Der erste – als ‚Kamerakalibrierung‘ bezeichnete Schritt – befasst sich mit der Umrechnung der sog. Weltkoordinaten in die sog. Kamerakoordinaten, mit denen es u. a. möglich ist, die geometrische Verzerrung (genauer die Verzeichnung) der Kameras zu eliminieren. Dieser Schritt ist nötig, um die Präzision des nachfolgenden Schrittes zu erhöhen. Der zweite Schritt befasst sich mit der sog. Registrierung der Aufnahmen bzw. Messgrößen, hier als ‚Bildregistrierung‘ bezeichnet. Darüber hinaus wird in diesem Kapitel auf die Transformationsmöglichkeiten von Bildern eingegangen, die eine der Aufnahmen so verzerren können, dass sie den Inhalten der zweiten Aufnahme entspricht und dadurch die Fusion ermöglicht. Im Anschluss

Die nachfolgenden Abschnitte setzen ein Grundverständnis der Funktion der Lochkamera und der Linsenkameras sowie der Unterschiede zwischen beiden voraus. Die beschriebenen mathematischen Prinzipien beruhen auf der Theorie der Lochkamera und werden hier nicht weiter betrachtet.

2.4.1. Kamerakalibrierung

Mit dem Prozess der Kamerakalibrierung werden sog. *intrinsische* und *extrinsische* Parameter einer Kamera bestimmt. Diese charakterisieren die innere sowie äußere Orientierung der verwendeten Kamera. Mithilfe dieser Parameter wird die Abbildung eines 3D-Punktes aus der Welt in einem 2D-Punkt im Bild ermöglicht. Somit wird die baulich bedingte Verzeichnung eines Kameraobjektivs beseitigt bzw. das Bild entzerrt. Je mehr Wissen in das Abbildungsmodell eingebracht wird, desto höher ist die Qualität der späteren Kalibrierung [Stubing (2010)].

Kameraparameter

Dieser Absatz befasst sich mit den zuvor genannten Kameraparametern. Dazu wird nachfolgend in intrinsische sowie extrinsische Kameraparameter unterschieden.

Intrinsische Kameraparameter Die *intrinsischen Kameraparameter* beschreiben die Eigenschaften der Kamera und unterscheiden sich von Kamera zu Kamera. Diese werden benötigt für die Definition der Abbildung eines Punktes aus dem 3D-Kamerakoordinatensystem (metrisch) im 2D-Bildkoordinatensystem (Pixel). Dabei wird auf die nachfolgenden intrinsischen Parameter Bezug genommen:

- Brennweite des Objektivs (f)
- Position des Bildhauptpunktes (c)
- radiale Verzerrungskoeffizienten (k_1, k_2)
- tangentielle Verzerrungskoeffizienten (p_1, p_2)
- Verzerrungsparameter s

Die Parameter werden i. d. R. experimentell ermittelt, indem Merkmale (bzw. die Position dieser Merkmale) aus den aufgezeichneten Bildern extrahiert werden. Wie bei jedem experimentellen Verfahren ist es durchaus möglich, dass die Kameraparameter sich von Versuch zu Versuch leicht unterscheiden. Dies hängt maßgeblich mit den jeweiligen Objekten auf den Bildausschnitten zusammen und lässt sich nicht vermeiden [Stubing (2010)].

Extrinsische Kameraparameter Die *extrinsischen Kameraparameter* beschreiben die bei Bewegung der Kamera veränderliche Lokation (Position des optischen Zentrums der Kamera) und die Orientierung (Ausrichtung der optischen Achse) der Kamera im Weltkoordinatensystem. Die sog. Freiheitsgrade beschreiben die Transformation der Eigenschaften der Kamera in das Weltkoordinatensystem. Gemäß der hier verwendeten Abbildung gibt es sechs extrinsische Parameter:

- Translation in x-Richtung: (Tx)
- Translation in y-Richtung: (Ty)
- Translation in z-Richtung: (Tz)
- Rotation um x-Achse in bestimmtem Winkel (Rx) oder (ω)
- Rotation um y-Achse in bestimmtem Winkel (Ry) oder (φ)
- Rotation um z-Achse in bestimmtem Winkel (Rz) oder (κ)

Diese Parameter beschreiben demnach sowohl die Verschiebung als auch die Drehung der Kamera zum Weltkoordinatensystem. So wird im Rahmen der Gleichung für die Bildkoordinaten $(u,v,1)$ die Rotationsmatrix (R) multipliziert mit dem jeweiligen Punkt (hier X,Y,Z) multipliziert und anschließend mit der Verschiebung um den Translationsvektor vom Mittelpunkt t addiert.

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = R \cdot \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + t \quad (2.2)$$

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \cdot \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} t1 \\ t2 \\ t3 \end{pmatrix} \quad (2.3)$$

Die Ermittlung der extrinsischen Parameter kann mit denselben Aufnahmen erfolgen wie die Ermittlung der intrinsischen Parameter. Das Zusammenspiel zwischen den intrinsischen und extrinsischen Parametern ist in Abbildung 2.8 dargestellt. Dabei ist darauf zu achten, dass die Ergebnisse sich von Objektiv zu Objektiv unterscheiden und dieses daher nicht gewechselt werden sollte, solange die ermittelten Parameter zugrunde gelegt werden.

Die Notwendigkeit der Parameter, die mathematischen Details und die Transformationen hierzu werden u. a. in [Stubing (2010), Süße und Rodner (2014)] ausführlich erläutert.

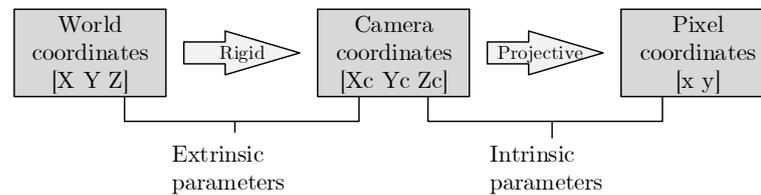


Abbildung 2.8.: Zusammenspiel der Kameraparameter nach [MathWorks (2017a)]

Kalibrierungsprozess

Mittels der Kalibrierung werden die Kameraparameter bestimmt. Grundsätzlich kann eine Kalibrierung mit verschiedenen Verfahren erfolgen, von denen die sog. *Testfeldkalibrierung* das wohl am weitesten verbreitete Verfahren ist [Hartley und Zisserman (2003)]. Neben diesen Verfahren besteht die Möglichkeit der sog. *Selbstkalibrierung*, bei der eine Kalibrierung anhand der Bildinhalte der Szene erfolgt. Für Verfahren der Testfeldkalibrierung wird ein sog. *Kalibrierungsmuster* verwendet, das je nach verwendetem Framework in der Darstellung variiert. Ziel ist es, dass aus dem Muster mindesten fünf Punkte extrahiert werden können [Stubing (2010)]. Mit diesen Punkten ist es möglich, die 3D-Lageberechnung durchzuführen. Bedingt wird dies durch die Eigenschaft, dass die Kalibrierungsmuster Merkmale besitzen, die dem System zur Berechnung zur Verfügung gestellt werden. Die Kalibrierungsmuster bestehen typischerweise aus einfachen geometrischen Figuren, die gleichmäßig verteilt angeordnet sind. Unterschieden wird in 2D-Kalibrierungsmuster und 3D-Kalibrierungsmuster. Letztere bestehen i. d. R. aus mehreren 2D-Kalibrierungsmustern desselben Aufbaus, die in einem fest

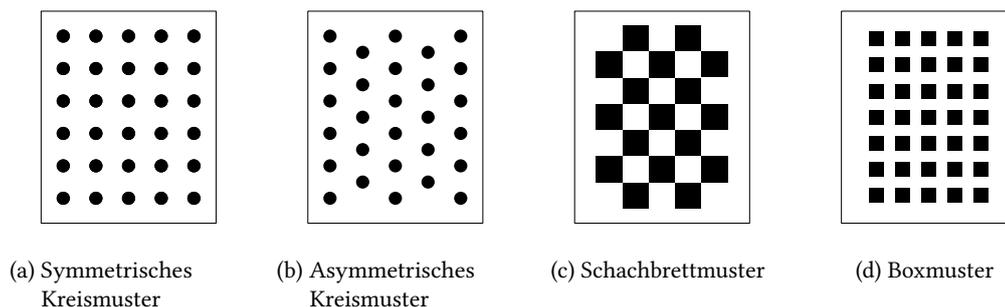


Abbildung 2.9.: Exemplarische Darstellung häufig verwendeter 2D-Kalibrierungsmuster. Die Anzahl, Größe und Abstände der geometrischen Figuren sowie die Größe des Testfelds können dabei beliebig variieren.

vorgegebenen Winkel zueinander angebracht sind. Bei der Herstellung ist auf eine sehr hohe Verarbeitungsqualität zu achten. 2D-Kalibrierungsmuster bestehen aus schwarz-weißen, gut zu lokalisierenden Punktmerkmalen, die auf eine feste Platte aufgeklebt oder aufgemalt werden. Häufig verwendete 2D-Muster, von denen auch invertierte Varianten bestehen, finden sich in Abbildung 2.9. Darüber hinaus gibt es noch viele weitere Muster bzw. Abwandlungen dieser Muster. So beinhalten einige von diesen beispielsweise noch grafische Informationen über die Orientierung des Musters, vergleichbar mit jenen von QR-Codes.

Das zur Kalibrierung genutzte System wird mit Informationen über das Kalibrierungsmuster bestückt. Bei den Kreismustern und Boxmustern sind dies beispielsweise die Anzahl der horizontal und vertikal angeordneten Kreise/Boxen, der Kreis- bzw. Boxdurchmesser sowie die horizontalen und vertikalen Abstände zueinander. Beim Schachbrettmuster sind es hingegen die Quadratgröße sowie die Anzahl der horizontalen und vertikalen Quadrate.

Zur Kalibrierung wird das Kalibrierungsmuster in verschiedenen Winkeln vor der Kamera bewegt und dabei aufgezeichnet. Die geometrischen Formen lassen sich aufgrund des starken Kontrastes einfach mittels Bilderkennungsverfahren im Bild erkennen. Mit dem Wissen, dass alle Formen dieselbe Größe und denselben Abstand zueinander aufweisen, kann die genaue Rotation des Kalibrierungsmusters erkannt und nicht korrekte Verzeichnungen des Bildes detektiert werden. Durch dieses Verfahren ist es möglich, die Kameraparameter zu ermitteln [Süße und Rodner (2014)].

Linsenverzeichnung

Kameras und deren Linsen erzeugen eine Vielzahl von optischen Effekten. Dies wird als Linsenverzeichnung bezeichnet. Durch diese kommt es i. d. R. zu Abbildungsfehlern, die bestmöglich mathematisch beseitigt werden müssen. Die optischen Effekte können je nach Linsenbauart erwünscht sein (beispielsweise bei einem sog. Fischaugenobjektiv) oder aufgrund von Fehlern unerwünscht auftreten. Bedingt werden Abbildungsfehler u. a. durch unsymmetrisch hergestellte Linsen oder baulich bedingte unpräzise Hardware der Kamera (Mittelpunkt des Kamerasensors befindet sich nicht exakt auf der optischen Achse der Kamera), aber auch Timingfehler zwischen Grabber- und Kamera-Hardware erzeugen diese Fehler. Durch diese Verzeichnung wird die Position der Bildpixel auf der Bildebene geändert, was es zu vermeiden gilt.

Die Linsenverzeichnung gruppiert sich in zwei Arten, die in Abbildung 2.10 schematisch dargestellt sind. Eine Art der Verzeichnung ist die sog. *radiale Linsenverzeichnung*. Bei dieser wird zwischen der tonnenförmigen Verzeichnung – bei der der Abstand zur optischen Achse vergrößert wird – und der kissenförmigen – bei der der Abstand zur optischen Achse

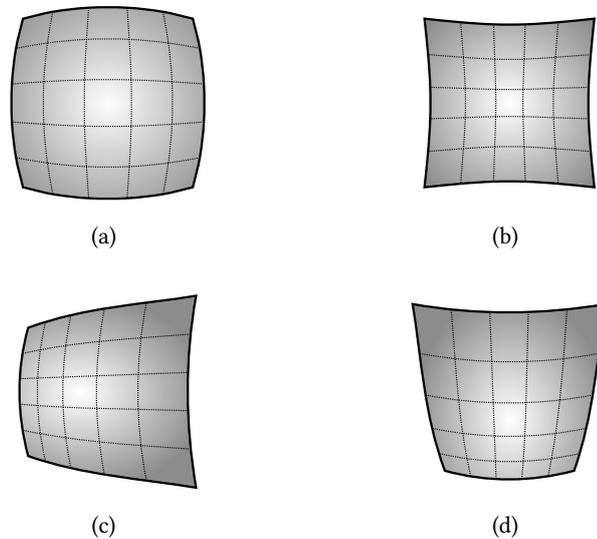


Abbildung 2.10.: Schematische Darstellung von radial (a)+(b) und tangential (c)+(d) verzeichneten Abbildungsfehlern. (a) tonnenförmige-/positive Verzeichnung (b) kissenförmige/negative Verzeichnung (c) horizontale Verzeichnung (d) vertikale Verzeichnung.

Quelle (a)+(b): HK KNG © ⓘ - bearbeitet

verringert wird – unterschieden. Die andere Art der Verzeichnung ist die sog. *tangentiale Linsenverzeichnung*. Diese tritt auf, wenn der Chip in der Kamera nicht eben montiert ist und dadurch die Lichtstrahlen nicht orthogonal auf den Chip treffen können. Die Auswirkungen der tangentialen Linsenverzeichnung sind im Vergleich dazu deutlich geringer als jene der radialen Linsenverzeichnung.

2.4.2. Bildregistrierung

Unter der Bildregistrierung zweier Bilder wird ein Vorgang in der digitalen Bildverarbeitung verstanden, dessen Ziel es ist, eine *plausible* Transformation zu finden und anzuwenden, um zu einem Bild eine möglichst genaue Übereinstimmung in einem zweiten Bild zu finden [Olesch (2014)]. Typischerweise werden dazu ein Bild als Ausgangsbild (das unverändert bleibt) und ein oder mehrere andere Bilder (die transformiert werden sollen) verwendet. Die zu transformierenden Bilder werden auch als Objektbilder bezeichnet. Gelegentlich kommen auch Verfahren zum Einsatz, bei denen zwei Bilder so transformiert werden, dass sie einem dritten (neuen) Bild entsprechen.

Ihren Ursprung hat die Bildregistrierung vornehmlich in der medizinischen Bildverarbeitung im Rahmen der Kombination verschiedener bildgebender Verfahren der Medizin zur Diagnostik. So werden häufig Bilder von Magnetresonanztomographie (MRT), Röntgen, Magnetresonanztomographie (MRA) und anderen Verfahren miteinander registriert, um die Informationen des Bildes zu erhöhen [Lahiri u. a. (2012); Stathaki (2011); MathWorks (2017b)]. Ein weiterer häufiger Anwendungsfall ist jener der militärischen und zivilen satellitengestützten Fernerkundung, bei der verschiedene modale (z. B. visuelle Fotos und Infrarotfotos) aufgezeichnet werden, um den Informationsgehalt der sog. Orthofotos zu erhöhen [Zitova und Flusser (2003); MathWorks (2017b)].

Um in einem Multikamera-Aufbau eine hochpräzise Ausrichtung der Bilder zu erlangen, ist eine Bildregistrierung ebenfalls unerlässlich. Der nachfolgende Absatz gliedert sich dazu in die *Arten der Bildregistrierung*, in dem eine Abgrenzung zwischen den verschiedenen Arten vorgenommen wird, dem Absatz der *physikalischen Aspekte*, die bei einer Bildregistrierung zu beachten sind, sowohl an die Anordnung des Messsystems, die Objektive sowie die Kamera als auch an das Messobjekt gestellt werden. Die *Methoden zur tatsächlichen Sensordatenfusion* werden in dem darauffolgenden Absatz erläutert. Abgeschlossen wird dieser Abschnitt mit den *Transformationsarten*, mit denen die Objektbilder transformiert werden, um dem jeweiligen Referenzbild zu entsprechen.

Arten der Bildregistrierung

Als Bildregistrierung wird der Prozess des Überlagerns von zwei oder mehr Bildern derselben Szene (a) aus verschiedenen Blickwinkeln, (b) zu verschiedenen Zeiten, (c) zur selben Zeit aus demselben Blickwinkel, (d) in verschiedenen Fokusbereichen oder (e) durch verschiedene Bildsensoren bezeichnet. In diesem Absatz werden die verschiedenen Arten der Registrierung erläutert und voneinander abgegrenzt [Stathaki (2011)].

Multi-View-Fusion Mit der *Multi-View-Fusion* wird der Prozess bezeichnet, bei dem Bilder überlagert werden, die aus verschiedenen Blickwinkeln aufgezeichnet wurden. Dies ist typischerweise der Fall, wenn sich zwei Kameras an verschiedenen Positionen im Raum befinden und dasselbe Objekt abbilden. Ziel dieser Art der Fusion kann sowohl die Erweiterung des Bildfelds (ähnlich dem als *stitching* bekannten Verfahren) als auch die 3D-Rekonstruktion eines Objekts sein. Mathematisch basiert diese Art der Fusion auf dem Modell der Epipolargeometrie, das die geometrische Beziehung der Bildpunkte in der Szene darstellt.

Multi-Temporal-Fusion Die *Multi-Temporal-Fusion* ermöglicht es, Veränderungen in einem Bild zu erkennen. Dies erfolgt durch das Aufzeichnen desselben Bildausschnitts von derselben Kameraposition, jedoch zu verschiedenen Zeitpunkten. Mittels diesem Verfahren können beispielsweise durch Subtraktion der Bildinhalte selbige voneinander abgezogen werden und so nur die Veränderungen zwischen beiden dargestellt werden. Ein bekannter Anwendungsfall hiervon ist das Abgleichen von Satellitenbildern vor und nach Großschadensereignissen, um beispielsweise zerstörte Infrastruktur zu identifizieren. Exemplarisch werden dazu zwei Aufnahmen in Abbildung 2.11 gezeigt.

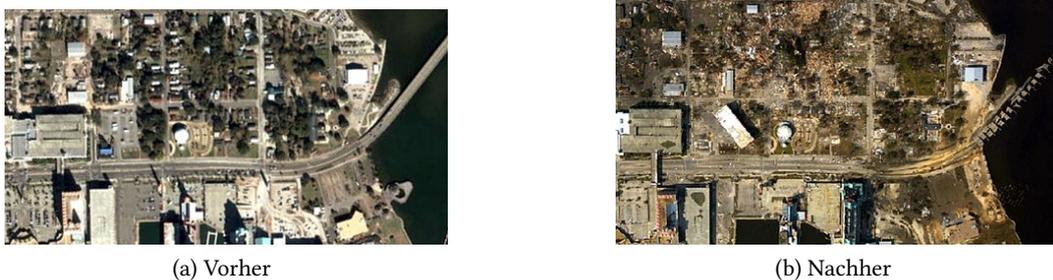


Abbildung 2.11.: Exemplarische Darstellung zur *Multi-Temporal-Fusion* anhand Satellitenaufnahmen vor und nach Hurrikan Katrina 2005 in Mississippi.

Quelle NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION © - Ausschnitt

Fusion for image restoration Mit dem Verfahren der *Fusion for image restoration* lassen sich aus zwei oder mehr beispielsweise weichgezeichneten und mit Rauschen behafteten Bildern scharfgezeichnete und mit minimiertem Rauschen behaftete Bilder erzeugen. Hierzu werden zwei oder mehr Bilder derselben Szene, die aus demselben Blickwinkel und auch sonst mit gleichen Modalitäten aufgezeichnet wurden, benötigt. Ermöglicht wird dies durch die Annahme, dass jedes Bild „korrekte“ Anteile hat und diese somit aus mehreren Bildern zu einem Gesamtbild zusammengesetzt werden können. Exemplarisch wird dies in Abbildung 2.12

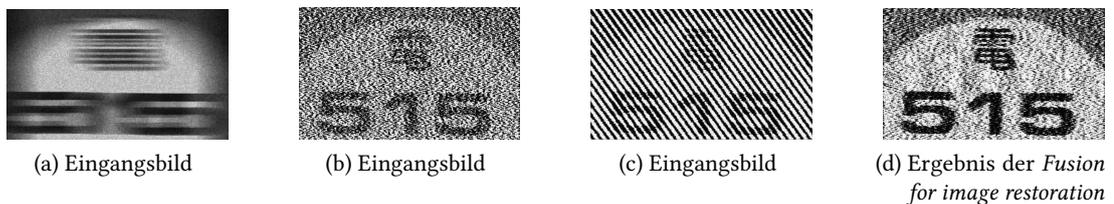


Abbildung 2.12.: Exemplarische Darstellung eines Verfahrens zur *Fusion for image restoration*: (a)-(c) Eingangsbilder, (d) Ergebnis des Algorithmus COOPERATIVE NEURAL FUSION REGULARIZATION (CNFR) aus [Xia und Kamel (2007)].

Quelle (Xia und Kamel, 2007) - Ausschnitt

gezeigt. Ein Teilbereich hiervon ist ebenso die sog. *Dekonvolution* also die Umkehrung von mathematischen Faltungsoperationen. Erweitert werden kann dieses zu einem Verfahren, dass aus vielen Bildern niedriger Auflösung eine hohe und qualitativ bessere Auflösung erzeugt.

Multi-Focus-Fusion Mit dem Verfahren der *Multi-Focus-Fusion* lässt sich aus mehreren Bildern einer Szene, die alle verschiedenen Tiefen des Bildes fokussiert haben, ein Bild erzeugen, auf dem alle Tiefen des Bildes scharfgestellt sind. Die Aufnahmen hierzu müssen aus demselben Blickwinkel aufgezeichnet werden und die Objekte der Szene dürfen sich nicht bewegen oder verändern. Beispielhaft in Abbildung 2.13 dargestellt.



Abbildung 2.13.: Exemplarische Darstellung eines Verfahrens zur *Multi-Focus-Fusion*: (a)+(b) Eingangsbilder, (c) Ergebnis des Algorithmus FREQUENCY SELECTIVE WEIGHTED MEDIAN FILTER (FSWM) aus [Aslantas und Kurban (2009)].

Quelle (Aslantas und Kurban, 2009) - Ausschnitt

Multi-Modal-Fusion Mittels der *Multi-Modal-Fusion* lassen sich Bilder verschiedener bildgebender Verfahren überlagern. Beispiele hiervon können Bilder, die durch Magnetresonanztomographie (MRT), Röntgen, Infrarotfotografie, Thermografiebilder, multispektrale Aufnahmen oder durch andere Verfahren erzeugte Bilder sein. Derartige Bilder müssen ebenso aus demselben Blickwinkel aufgezeichnet sein und dasselbe Objekt abbilden. Diese Art der Bildregistrie-

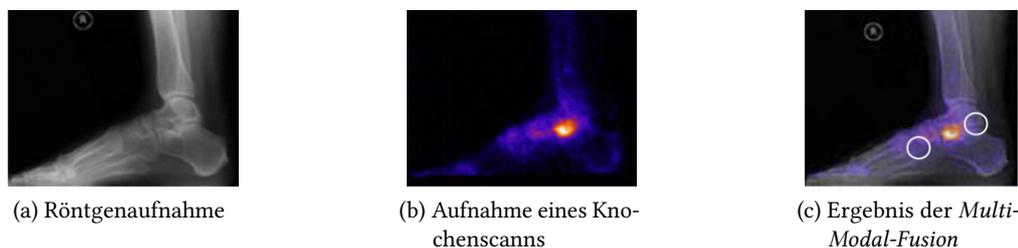


Abbildung 2.14.: Exemplarische Darstellung eines Verfahrens zur *Multi-Modal-Fusion*: (a)+(b) Eingangsbilder, (c) Ergebnis des Algorithmus der von [Bhatnagar u. a. (2015)] entwickelt wurde. Quelle (Bhatnagar u. a., 2015) - Ausschnitt

rung wird häufig im Rahmen bildgebender Verfahren der Medizin genutzt, um verschiedene Befunde zu kombinieren, wie in Abbildung 2.14 dargestellt.

Aus der Analyse der Arten der Bildregistrierung ergibt sich schon jetzt, dass das zu entwickelnde System der Problemstellung der *Multi-Modal-Fusion* entspricht. Die Fusion erfolgt hierbei zwischen Aufnahmen des visuellen Spektrums sowie Aufnahmen die mittels Infrarotthermografie erzeugt werden. Ebenso trifft auch die Art der *Multi-View-Fusion* auf diese Problemstellung zu, da es sich um mehrere physikalisch getrennte Kameras handelt (siehe auch nächsten Abschnitt). Als weitere Art der Fusion betrachtet das System auch Aspekte der *Multi-Temporal-Fusion*, da es sich um eine Beobachtung über die Zeit handelt.

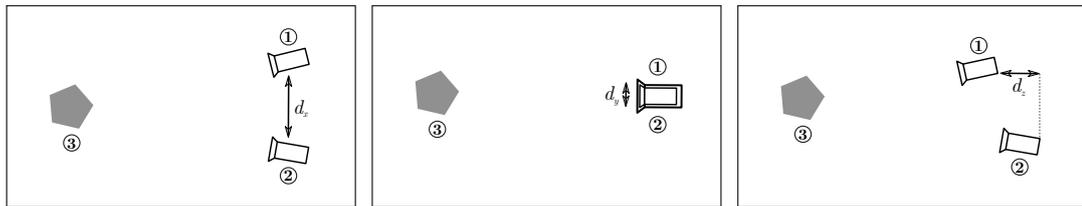
Physikalische Aspekte eines Multikamerasystems

Bei der Nutzung eines Multikamerasystems ergibt sich gleichzeitig eine Reihe von physikalischen Aspekten, die betrachtet werden müssen. Dazu gliedert sich dieser Absatz in die Aspekte *Anordnung*, *Linse*, *Kamera* sowie *Objekt*. Die *Anordnung des Messsystems* beschreibt, wie die Kameras im Verhältnis zueinander und zum Objekt, das betrachtet werden soll, räumlich positioniert sein können. Im darauffolgenden Absatz wird kurz auf die üblichen *Linsenarten* eingegangen. Der Absatz *Kamera* geht noch einmal auf die verschiedenen Spektren ein, die mit den verschiedenen Kameras aufgezeichnet werden können. Abgeschlossen wird dieser Absatz mit der Benennung der Eigenschaften und Aspekte, die an ein Messobjekt gestellt werden, das durch das Multikamerasystem betrachtet wird.

Anordnung des Messsystems Aus den Arten der Bildregistrierung ergeben sich mögliche physikalische Anordnungen für die Kameras eines Multikamerasystems. Eine Illustration dieser findet sich in Abbildung 2.15. Sie zeichnen sich maßgeblich durch die Entfernungen zueinander sowie zum analysierenden Objekt aus.

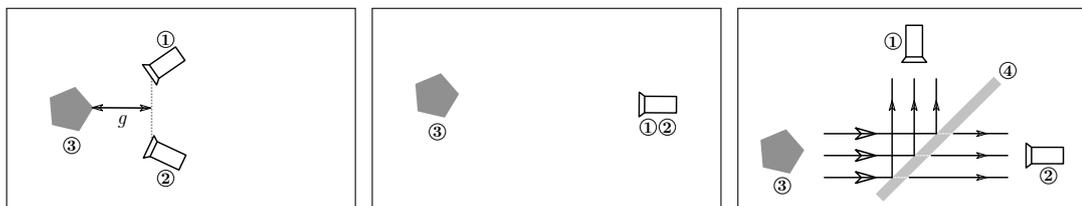
Baulich bedingt können Kameras bzw. die Bildsensoren von zwei oder mehr Kameras nicht auf derselben optischen Achse liegen, weshalb es nötig ist, diese versetzt anzubringen. In Abbildung 2.15a-2.15c werden die Anordnungen der Kameras zueinander im dreidimensionalen Raum betrachtet. Darüber hinaus spielt der Abstand zum Messobjekt (die sog. Gegenstandsweite g) eine entscheidende Rolle (Abbildung 2.15d). Wird ein nicht planares Objekt aus einer verhältnismäßig dichten Gegenstandsweite betrachtet, kommt es je nach Beschaffenheit des Objekts zur sog. Parallaxe, bei der sich die Bildinhalte aufgrund der Blickwinkel unterscheiden. Um diesen Herausforderungen entgegenzuwirken, ist es möglich, ein zu analysierendes Objekt auf derselben optischen Achse aufzuzeichnen. Eine Möglichkeit dazu ist in Abbildung 2.15e

2. Analyse



(a) Versatz der Kameras in der X-Achse um die Distanz d (d_x) (b) Versatz der Kameras in der Y-Achse um die Distanz d (d_y) (c) Versatz der Kameras in der Z-Achse um die Distanz d (d_z)

Ebenso sind Kombinationen dieser Achsenverschiebungen möglich.



(d) Variation der Gegenstandsweite (g) (e) Kameras an derselben Position mit zeitlichem Versatz t (f) Umlenkung von Bildinformationen z. B. mittels ④ Spiegels

Abbildung 2.15.: Visualisierungen von Multikamerasystem-Aufbauten und deren Bezugsgrößen: ① Thermografiekamera, ② visuelle Kamera, ③ zu analysierendes Objekt

dargestellt. Im Rahmen dieser wurden die Kameras ausgetauscht und das Objekt mit einem zeitlichen Versatz (t) aufgezeichnet. Dies ist jedoch nur möglich, wenn es sich um ein statisches Messobjekt handelt, das seine Eigenschaften nicht verändert bzw. dessen Veränderungen marginal sind. Ein anderer Ansatz der Anordnungen ist in [Abbildung 2.15f](#) dargestellt. Hier werden (Teile der) Bildinformationen mit einem Spiegel umgelenkt. Auf diese Weise wird beiden verwendeten Kameras ermöglicht, zeitgleich auf der gleichen optischen Achse aufzuzeichnen.

Objektive Neben der Anordnung des Messsystems spielen auch die Linsen der Kameras eine entscheidende Rolle in Bezug auf das Messsystem. Optische Linsen haben eine Vielzahl von Eigenschaften, die je nach Modell oder Bauart variieren können. An dieser Stelle wird allerdings nur auf den Bildwinkel (α) einer Linse eingegangen.

Aus der Kombination des Bildwinkels und der Gegenstandsweite ergibt sich der auszunehmende Bereich einer Kamera. Dabei basiert der Wert technisch auf der Brennweite (f) und der Größe des Bildsensors bzw. dessen Format. Vereinfacht kann gesagt werden, dass der aufzunehmende Bereich umso größer ist und die Objekte auf dem Bild umso kleiner erscheinen,

je größer der Bildwinkel ist. Analog dazu gilt: je kleiner der Bildwinkel einer Linse ist, desto größer wirken die Objekte auf dem Bild, während der Ausschnitt sich verkleinert.

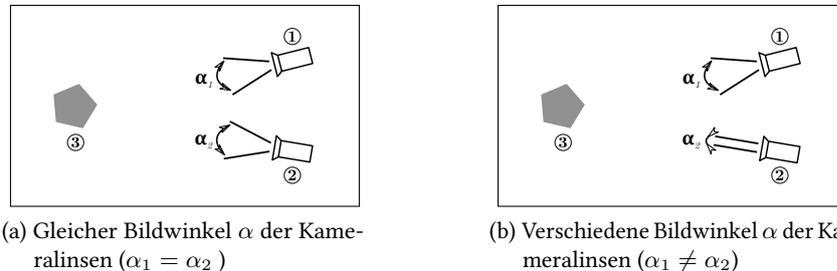


Abbildung 2.16.: Visualisierungen der Bildwinkel in einem Multikamerasystem: ① Thermografiekamera, ② visuelle Kamera, ③ zu analysierendes Objekt

Unterscheiden sich die Bildwinkel (wie in Abbildung 2.16b) der verwendeten Kameras, ist bei einer Registrierung der beiden Aufnahmen eine starke Anpassung eines der Bilder nötig, damit dieses dem anderen entspricht, oder es kann nur der kleinere der beiden durch die Kameras aufgezeichneten Bereiche für den Registrierungsprozess verwendet werden.

Kamera Die verwendeten Kameras spielen offensichtlich ebenso eine Schlüsselrolle. So können diese – wie bereits mehrfach erwähnt – in verschiedenen Modalen arbeiten, die sich nicht nur auf das sichtbare Spektrum begrenzen. Als weitere Spektren können in diesem Rahmen sowohl für das menschliche Auge nicht sichtbare Bereiche wie die Infrarotthermografie als auch Tiefeninformationen – die mittels Musterprojektion, Lichtfeldkamera, Time-of-Flight (ToF) o. ä. Kameras und Systemen gewonnen werden können – gemeint sein. Für diese Systeme werden an die Anordnung ggf. noch weitere Anforderungen gestellt als bereits die eingangs erwähnten.

Messobjekt Ebenso werden an das eigentliche zu analysierende/ aufzuzeichnende Objekt je nach Aufbau und verwendeter Hardware besondere Anforderungen gestellt. Darüber hinaus kann allerdings auch das Objekt die eigentlichen Anforderungen definieren. Mittels Kameras wird i. d. R. ein dreidimensionaler Raum (die Welt) in einen zweidimensionalen Raum (das Bild) übertragen, wodurch Informationen über das Objekt verloren gehen können. Eine besondere Rolle – besonders bei einem multimodalen Kamerasystem – spielt die physikalische Beschaffenheit des Objekts, aber auch der Umgebung. So gibt es Wellenlängen und Aggregatzustände (hier besonders gasförmig), die von Kameras nicht aufgezeichnet werden können oder aber die Aufnahme stören, da sie das Objekt anders erscheinen lassen bzw. sogar verfälschen. Daneben

spielen auch Bewegungen des Objekts eine entscheidende Rolle. Je nach Ausleuchtung der Szene im Zusammenhang mit der Schnelle der Bewegungen eines Objektes kann bei Kameras eine irreversible Bewegungsunschärfe auftreten, die die Aufnahmen ggf. für weitere Verarbeitungsschritte unbrauchbar macht. Ebenso spielen die Tiefe und die Bewegung in der Z-Achse des Objekts eine bedeutende Rolle für das Kamerasystem. Je nach Tiefenschärfe der Kamera muss eine kontinuierliche (Re-)Fokussierung des Objekts erfolgen.

Methoden zur Multisensor-Datenfusion

Aus der vorhergehenden Analyse der verschiedenen Arten der Bildregistrierung ergibt sich, dass die Ansätze der *Multi-View-Fusion* und *Multi-Modal-Fusion* näher betrachtet werden müssen. Ebenso spielt der Aspekt der *Multi-Temporal-Fusion* – wenn auch in anderer Form – eine entscheidende Rolle, um später Auswertungen über die Zeit durchführen zu können. Zusammen mit den physikalischen Aspekten, die im vorherigen Absatz betrachtet wurden, ergeben sich drei Methoden, die eine (Kamera-)Sensordatenfusion ermöglichen. Diese werden nachfolgend beschrieben.

Fusion mittels Polarisationsstrahlteilers Eine Möglichkeit, eine *Multi-Modal-Fusion* durchzuführen – die aus der Analyse der Anforderungen an den physikalischen Aufbau erkenntlich wird – ist die Nutzung eines sog. Polarisationsstrahlteilers oder kurz Strahlteilers. Dieser ermöglicht es, das gesamte Spektrum oder je nach Beschaffenheit auch nur bestimmte Spektralbereiche umzulenken und so von einem anderen Ort im Messaufbau aufzuzeichnen. In einem multimodalen Kamerasystem mit einer Thermografiekamera kann sich die physikalische Eigenschaft von Glas zunutze gemacht werden. Glas kann als Strahlteiler für die Thermografiestrahlung eingesetzt werden, da es diese Strahlung reflektiert¹⁴, sichtbares Licht, dies hingegen nahezu unverändert durchdringt. Werden die Kameras sinnvoll angeordnet, besteht die Möglichkeit, ein Objekt mit zwei unterschiedlichen Kameras aus demselben Blickwinkel zu betrachten. Der Aufbau ist exemplarisch in Abbildung 2.17 dargestellt. Durchgeführt wird eine solche Registrierung zwischen einer Kamera im visuellen Spektrum und einer Thermografiekamera beispielsweise in [Ó Conaire u. a. (2005)]. Eine Registrierung ist in diesem Fall nur sehr rudimentär nötig, da lediglich eine Überlagerung der beiden Bilder erfolgen muss. An den Aufbau werden hingegen hohe Anforderungen bezüglich der Präzision gestellt. So müssen die Kameras exakt ausgerichtet sein und je nach Objektiv in der korrekten Entfernung zur Glasplatte bzw. dem Strahlteiler montiert werden. Dies bedeutet je nach verwendeten Kameras

¹⁴vgl. Abschnitt 2.1.1 Grenzen der Infrarot-Thermoografie

einen hohen Platzanspruch für das Multikamerasystem. Auch kann durch den Strahlteiler die Bildqualität (sowohl der Thermografiekamera als auch der Kamera im visuellen Spektrum) gemindert werden. Der Grund hierfür ist eine Absorption der Strahlung durch einen nicht vollkommen reinen Strahlteiler.

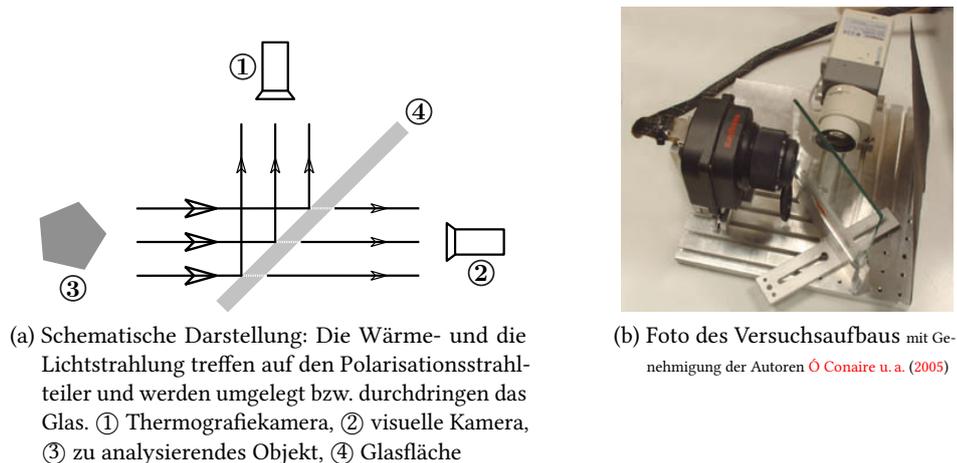


Abbildung 2.17.: Darstellung der Fusion mittels Polarisationsstrahlteiler

Bei der Nutzung von zwei visuellen Kameras kann anstelle einer Glasfläche ein sog. Einwegspiegel oder halbdurchlässiger Spiegel verwendet werden. Die reflektierende Fläche des Strahlteilers reflektiert das Licht und lenkt es somit zu einer Kamera um. Die zweite Kamera zeichnet durch die lichtdurchlässige Rückseite des Strahlteilers auf und hat somit eine nicht umgelenkte optische Achse, die den Strahlteiler orthogonal durchdringt.

Direktes Überlagern Eine andere Möglichkeit zur multimodalen Fusion ist es, die Aufnahmen miteinander zu überlagern, ohne dass sie einander angepasst werden müssen. Typischerweise ist dies der Fall, wenn es sich um Kameras mit ähnlichen Eigenschaften handelt (gleicher Bildwinkel, gleiche Zoomstufe, etc.). Gleichzeitig müssen die beiden Bildsensoren – wie in der Analyse bereits ermittelt – baulich nah beieinanderliegen und ein relativ weit entferntes Objekt aufzeichnen. Durch die bauliche Kompaktheit und die Betrachtung eines weit entfernten Objekts ist der Versatz der beiden Kameras zu vernachlässigen, weshalb auch u. a. der sog. Parallaxeffekt nicht auftritt. Dieser Fall tritt beispielsweise bei der Fernerkundung mittels Satelliten auf. Aufgrund der mehreren Kilometer umfassenden Höhe der Satelliten tritt die Distanz weniger Zentimeter zwischen den Kameras in den Hintergrund. Bei der Analyse von verhältnismäßig nahen Objekten ist dies hingegen meist nicht möglich [Bahnsen (2013)].

Pixelgenau basierend auf Bildinhalten Besonders bei einem Multikamerasystem, bei dem die Kameras bzw. die Bildsensoren nicht auf derselben Position liegen und das zu betrachtende Objekt weniger weit (typischerweise nur wenige Meter) entfernt ist als bei der Fernerkundung, müssen andere Verfahren genutzt werden. Hier bietet sich eine pixelgenaue an, d. h. eine Überlagerung basierend auf den Bildinhalten. Unterschieden wird hierbei zwischen *merkmal-* oder *flächenbasierten* sowie *manuell* gewählten Punkten. Basierend auf diesen kann eines der Bilder mittels geometrischer Transformation so verzerrt werden, dass es deckungsgleich mit dem anderen ist. Auf die drei genannten Arten von Punkten wird dazu im Nachfolgenden eingegangen:

- **Merkmalsbasiert**

Bei einer reinen *Muti-View-Fusion* – also der Fusion mehrerer Bilder, die aus verschiedenen Blinkwinkeln, jedoch innerhalb desselben Spektrums aufgezeichnet wurden – bietet sich ein Schritt mit den Namen *feature detection*, also die Erkennung von Merkmalen in den Bildern, an. Dieser Schritt wird als erster von vier Schritten in [Zitova und Flusser (2003)] beschrieben. Die Merkmale bestehen aus Elementen im Bild, die gut erkennbar sind und dadurch ohne Probleme wiedergefunden werden können. Beispiele hierfür sind Ecken, Kanten oder sonstige typische Bereiche (Blobs) im Bild. Bekannte Algorithmen aus der Bildverarbeitung für die Merkmalerkennung sind SCALE-INVARIANT FEATURE TRANSFORM (SIFT) [Lowe (2004)], SPEEDED UP ROBUST FEATURES (SURF) [Bay u. a. (2008)] oder MAXIMALLY ATABLE EXTREMAL REGIONS (MSER) [Matas u. a. (2004)]. Die mit diesen oder anderen Algorithmen gefundenen Punkte werden im zweiten Schritt, dem *feature matching*, mit den korrespondierenden Punkten im anderen Bild zusammengeführt. Hierdurch kann eines der Bilder geometrisch transformiert werden, damit dieses dem anderen weitgehend entspricht.

Da sich der Hauptanwendungsfall für diese Art von Algorithmen auf unimodale Bildspektren (typischerweise auf die Modalität des sichtbaren Spektrums) beschränkt, sind die Algorithmen nicht ohne Weiteres für multimodale Bildspektren einsetzbar. Bedingt wird dies durch die grundlegend verschiedenen Bildinformationen, die aufgezeichnet werden.¹⁵ Besonders auffällige Merkmale im sichtbaren Spektrum können dieselbe Temperatur haben, was auf den Thermografiebildern als eine nahezu homogene Fläche dargestellt wird. Ebenso können einheitliche Flächen wie beispielsweise ein seit kurzer Zeit eingeschalteter Heizkörper im sichtbaren Spektrum deutlich andere Eigenschaften aufweisen als auf Thermografiebildern, auf denen der Wärmefluss sichtbar wird. Dennoch gibt es

¹⁵siehe auch Abschnitt 3.4.1 Schachbrett

Gemeinsamkeiten, die auch bei Aufzeichnungen verschiedener Modalitäten in beiden erkennbar sind. In [Aguilera u. a. (2012)] wird hierzu ein Ansatz präsentiert, in dem der SIFT-Algorithmus modifiziert wird, damit auf den unterschiedlichen Darstellungen der verschiedenen Modalitäten identische Merkmale besser erkennbar sind.

• Flächenbasiert

Analog zu dem Vorgehen, bei dem einzelne Merkmale im Bild erkannt und aufeinander abgebildet werden, gibt es die Möglichkeit, komplette Bilder aufeinander abzubilden. Dazu wird eines der Bilder beliebig transformiert¹⁶ und durch eine sog. Bewertungsfunktion ermittelt, zu welchem Grad diese Transformation dem Zielbild entspricht. Dieser Ansatz ist jedoch besonders rechenintensiv, da er einer *Brute Force* Methode nahekommt. Iterativ wird hierbei versucht, die Bilder immer besser aufeinander auszurichten [Zitova und Flusser (2003)]. Für die Ausrichtung der Bilder gibt es grundlegend die folgenden drei Vorgehensweisen: *correlation-like methods*, *fourier methods* oder *mutual information methods*. Die Vorgehensweise der „Ausrichtung anhand von Ähnlichkeiten“ in den Bildern (*mutual information*) bietet sich laut den Autoren am besten für Aufnahmen verschiedener Modalitäten an [Zitova und Flusser (2003); Soldan (2014)].

• Manuell

Ohne die automatische Detektion von Merkmalen oder die Nutzung der flächenbasierten Ansätze gibt es die Möglichkeit der Festlegung von Punkten. Basierend auf diesen findet dann eine Transformation des Bildes statt. Dieses Verfahren wird beispielsweise *control point registration* genannt [MathWorks (2017a)]. Hier werden in einem der beiden zu registrierenden Bildern Punkte festgelegt und in dem anderen Bild die dazu korrespondierenden Punkte ermittelt.

Soll eine Reihe von Aufnahmen gemeinsam registriert werden, genügt es in der Vielzahl der Fälle, die Registrierung mit der ersten Aufnahme durchzuführen. Für die nachfolgenden Aufnahmen kann dieselbe Zuordnung erfolgen wie bei der ersten Aufnahme. Dies reduziert den Rechen- und Verarbeitungsaufwand enorm, da die rechenintensive Ermittlung der korrekten Registrierung eingespart werden kann. Möglich ist dies allerdings nur, wenn ein starrer Abstand zwischen den aufnehmenden Kameras vorliegt, der sich während der Aufnahmen nicht verändert. Ebenso darf das zu messende Objekt nicht in der Distanz zur Kamera variieren, da hierdurch die Skalierung der Aufnahmen zueinander ggf. nicht mehr korrekt wäre.

¹⁶vgl. Abschnitt 2.4.2 Transformation

Transformation

Um die Bilder aufeinander abbilden zu können, müssen entweder beide Bilder in ein drittes Zielbild transformiert werden oder es muss eine Aufnahme so transformiert werden, dass diese der anderen nahezu entspricht. Unterschieden wird dabei zwischen *linearer 2D- und 3D-Transformation* und *nichtlinearer 2D- und 3D-Transformation*.

Lineare 2D- und 3D-Transformation Lineare Transformationen können sowohl auf 2D-Objekte als auch auf 3D-Objekte angewandt werden. Sie lassen sich in formerhaltende Transformationen (*rigid*) – bei denen die Objektform erhalten bleibt – und formverändernde Transformationen (*non-rigid*) – bei denen die Objektform verloren geht, unterscheiden. Nachfolgend werden die Regeln für die verschiedenen Arten der Transformationen anhand von Punkten beschrieben. Diese können durch Kombination zu komplexeren Transformationen führen. Die Reihenfolge der Punkte ist dabei jedoch nicht beliebig.

formerhaltende Transformationen (rigid):

- Translation

Verschieben eines Punktes (x, y) um den Vektor (t_x, t_y) :

$$(x', y') = (x + t_x, y + t_y)$$

- Rotation

Drehung eines Punktes (x, y) um den Winkel θ :

$$(x', y') = (x * \cos(\theta) - y * \sin(\theta), x * \sin(\theta) + y * \cos(\theta))$$

- Spiegelung

$$(x', y') = (x * u, y * v)$$

Sonderfall der Skalierung, wobei $((u = 0 \wedge v = -y) \vee (u = -x \wedge v = 0)) \vee$

$$(u = -x \wedge v = -y).$$

formverändernde Transformationen (non-rigid):

- Skalierung

Skalierung eines Punktes (x, y) um die Faktoren u und v :

$$(x', y') = (x * u, y * v)$$

- Spiegelung

Sonderfall der Skalierung wobei $u = \{w | w \leq 0 \wedge w \neq -x\} \wedge$

$$v = \{w | w \leq 0 \wedge w \neq -y\}.$$

- Scherung

Verschiebung eines Punktes (x, y) um den Faktor m :

X-Scherung: $(x', y') = (x, y + m_y * x)$; Y-Scherung: $(x', y') = (x + m_x * y, y)$.

- perspektivische Projektion

In der Geometrie und im maschinellen Sehen wird darüber hinaus noch von weiteren Transformationsarten gesprochen, die auf den zuvor genannten Arten aufbauen, aber bestimmte Eigenschaften aufweisen. Diese finden sich illustriert ebenso in Abbildung 2.18:

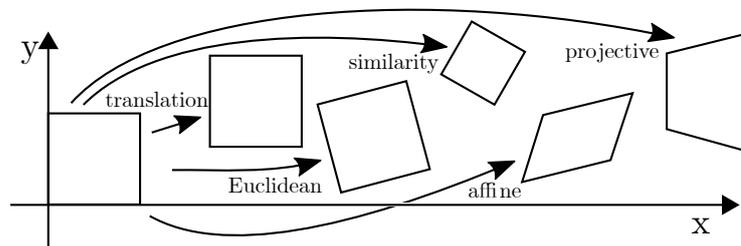


Abbildung 2.18.: Übersicht der 2D-Transformationen eines Rechtecks

- Projektive Transformation

Kombination von Rotation, Translation, Skalierung, Scherung und perspektivische Projektion (acht Freiheitsgerade).

- Affine Transformation

Kombination von Rotation, Translation, Skalierung und Scherung mit der Eigenschaft, dass parallele Linien parallel bleiben, rechte Winkel aber nicht erhalten bleiben müssen (sechs Freiheitsgerade).

- Similarity Transformation

Kombination von Rotation, Translation und Skalierung, die eine Teilmenge der affinen Transformationen ohne die Scherung darstellt (vier Freiheitsgerade).

- Euklidische Transformation

Kombination von Translation, Rotation und Spiegelung, die wiederum eine Teilmenge der Similarity Transformationen darstellt (drei Freiheitsgerade).

- Isometrische Transformation

Kombination von Translation und Rotation, die wiederum eine Teilmenge der euklidischen Transformationen darstellt (drei Freiheitsgerade).

Die hier aufgeführten Abbildungen können auch mittels sog. *Transformationsmatrizen* dargestellt werden. Dafür wird der Pixelkoordinatenvektor zu einem sog. *homogenen Vektor* erweitert, bei dem jedem Punkt eine zusätzliche Koordinate h an dritter Stelle zugeordnet wird. Diese zusätzliche Koordinate ist $h = 1$ [Soldan (2014)]. Der Punkt (x, y) lautet in der Vektorschreibweise somit $(x, y, 1)^\top$. Die Matrixschreibweise bietet sich an, da i. d. R. größere Teile (Objekte oder Bilder) transformiert werden müssen und nicht einzelne Punkte.

Eine Transformationsmatrix am Beispiel der Skalierung wird durch Gleichung 2.4 dargestellt.

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} \quad (2.4)$$

Gleichung 2.4 wird überführt in die homogene Koordinatenschreibweise des 2D-Raumes. Diese ist in Gleichung 2.5 dargestellt.

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} s_x & 0 & \mathbf{0} \\ 0 & s_y & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (2.5)$$

Die oben genannten 2D-Konzepte lassen sich ebenso auf Transformationen in 3D übertragen. Hierzu wird einem Punkt im dreidimensionalen Raum (x, y, z) ebenso eine zusätzliche Koordinate hinzugefügt, deren Wert analog zu den 2D-Koordinaten ebenfalls $h = 1$ beträgt [Szeliski (2011)]. Die Vektorschreibweise lautet daher $(x, y, z, 1)^\top$. Die Gleichung 2.4 bzw. 2.5 auf 3D erweitert wird durch Gleichung 2.6 dargestellt.

$$\begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{pmatrix} s_x & 0 & 0 & \mathbf{0} \\ 0 & s_y & 0 & \mathbf{0} \\ 0 & 0 & s_z & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (2.6)$$

Zu beachten ist jedoch, dass einige Arten der Transformation (wie beispielsweise die Rotation) in der 3D-Transformation weniger trivial sind, da es in der 3D-Welt mehrere Freiheitsgrade gibt [Szeliski (2011)].

Nichtlineare 2D- und 3D-Transformation Für verschiedene Anwendungsfälle ist es nötig, nur Teile eines Bildes zu verzerren. Die Transformation soll folglich nur bereichsweise wirken. Geometrische Transformationen in diesem Kontext können beispielsweise kubisch oder zylindrisch sein, während verschiedene Bereiche unterschiedlich stark verzerrt werden. Ebenso kann ein Bild nach einem vorgegebenen Netz verzerrt werden (*mesh warping*). Bei den nicht linearen Transformationen erhöht sich jedoch die Parameteranzahl im Vergleich zu den linearen Transformationen deutlich.

Laut [Soldan (2014)] lassen sich nichtlineare Transformationen umgehen, indem zuvor eine Kamerakalibrierung¹⁷ durchgeführt und dadurch die Verzeichnung der Linsen ausgeglichen wird. Die verzeichnungsfreien Bilder lassen sich anschließend mittels linearer Transformation zusammenführen.

¹⁷vgl. Abschnitt 2.4.1 Kamerakalibrierung

2.4.3. Zusammenfassung der Fusion multimodaler Aufnahmen

Aus den Erkenntnissen des Kapitels 2.4 über die Fusion multimodaler Aufnahmen ergeben sich zwei Tatsachen. Zum einen zeigt sich, dass eine Kamerakalibrierung mit einem geeigneten Kalibrierungsmuster sinnvoll erscheint, um die Verzeichnung der Linse bzw. der optischen Bauteile auszugleichen. Für den Kalibrierungsprozess kann dafür auf bewährte Verfahren und Algorithmen bzw. Tools zurückgegriffen werden. Um das Kalibrierungsmuster auch in anderen Modalen nutzen zu können, ist jedoch ein Muster zu wählen, das die Eigenschaften in den benötigten Modalen liefert. Die Analyse zeigt, dass es hierfür in der Literatur bereits verschiedene Ansätze gibt.

Bezüglich der eigentlichen Bildregistrierung wurden verschiedene Aspekte betrachtet. Neben der Definition bzw. der Erläuterung der verschiedenen Arten der Bildregistrierung wurden die physikalischen Aspekte eines Kamerasystems am Beispiel einer Kamera im visuellen Spektrum und einer Thermografiekamera evaluiert. Zusammen mit den Methoden zur Multisensor-Datenfusion, die im darauffolgenden Absatz betrachtet wurden, ergibt sich eine Reihe von Möglichkeiten, wie die Registrierung durchgeführt werden kann. Eine Einordnung findet tabellarisch in Tabelle 2.2 statt, eine Erläuterung im anschließenden Text.

Verfahren	Elemente der Registrierung		Eignung
Strahlteiler	-	-	○
Direktes Überlagern	-	-	○
Auf Bildinhalten	Merkmalbasiert	Merkmalsextraktion	○
		Kantenextraktion	+
		Silhouettenextraktion	+
	Flächenbasiert	Kalibriermuster	○
		Manuell	-
Tiefeninformationen	-	-	○

○ = bedingt oder nicht geeignet; + = geeignet

Tabelle 2.2.: Einordnung der Bildregistrierung

Die Einordnung findet vor dem Hintergrund der generischen Anwendbarkeit auf verschiedene Szenarien statt. Demnach werden Verfahren, die einen speziellen Aufbau voraussetzen, als bedingt geeignet klassifiziert.

Konkret bedeutet dies, dass sowohl das Verfahren, bei dem ein *Strahlteiler*¹⁸ (oder bei der Nutzung einer Thermografiekamera ein Umlenken mittels Glas) gewählt wird, das Verfahren, bei dem ein *direktes Überlagern*¹⁹ möglich ist, als auch das Verfahren das *Tiefeninformationen* verwendet, als nicht geeignet klassifiziert werden. Zu begründen ist dies damit, dass hierfür hohe Anforderungen an den Aufbau des Kamerasystems gestellt werden und dass, sofern bereits vorhandene Aufnahmen genutzt werden sollen, diese auf eine andere Art und Weise aufgezeichnet wurden. Dennoch stellen diese drei Möglichkeiten gute Verfahren dar um eine präzise Bildregistrierung zur ermöglichen. Sie scheiden daher nicht aufgrund von Präzisionsmängeln aus, sondern aufgrund von hohen Anforderungen an die verwendeten Geräte sowie deren Aufbau.

Aus der Analyse der Methoden zur Multisensor-Datenfusion von Bildinhalten geht hervor, dass sich diese grundlegend in die drei Untergruppen *merkmalbasiert*, *flächenbasiert* und *manuell* aufteilen lassen. Bei der *merkmalbasierten* Fusion²⁰ – also der Fusion basierend auf detektierten Merkmalen/Organen des Gesichts – zeigte sich, dass in einem multimodalen Kontext eine *Merkmalextraktion* in den meisten Fällen ungeeignet erscheint. Gründe hierfür sind die unterschiedlichen Darstellungsarten in den verschiedenen Modalen und dadurch ein Mangel an übereinstimmenden Punkten, die aufeinander abgebildet werden können.

Um diesem Problem zu begegnen, gibt es Verfahren, mit denen dennoch Merkmale der verschiedenen Modalitäten aufeinander abgebildet werden können. Eines davon ist die *Kantenextraktion*. Für diese werden Kantendarstellungen der zu registrierenden Bilder benötigt, die mit einem zusätzlichen Algorithmus generiert werden müssen. Sie stellen dabei eine Abstraktionsebene der Bildinhalte dar. Mittels dieser ist es – wie bereits einige Arbeiten gezeigt haben (Kong u. a. (2007); Ma u. a. (2015)) – möglich, ähnliche Merkmale auf den Aufnahmen zu detektieren. Da dennoch kaum exakt gleiche Merkmale erkannt werden können, müssen die Registrierungsalgorithmen entsprechend angepasst werden. Dieses Verfahren erscheint dabei prinzipiell für die gewünschten Szenarien bzw. im Rahmen einer generischen Anwendbarkeit als geeignet.

Ebenso geeignet erscheint ein Verfahren, bei dem eine *Silhouettenextraktion* durchgeführt wird. Die hierfür benötigten Silhouetten müssen mittels einer ähnlichen Vorgehensweise erzeugt werden wie jene der Kantenextraktion. Verwendet werden kann hierfür beispielsweise das sog. *Hintergrundentfernungsverfahren*. Eine Registrierung findet dann auf der Silhouette statt. Auf Thermografiebildern lässt sich die dafür benötigte Hintergrundentfernung anhand der Temperatur herstellen. Der Hintergrund weist i. d. R. eine geringere Temperatur auf als

¹⁸vgl. Abschnitt 2.4.2 Fusion mittels Polarisationsstrahlteilers

¹⁹vgl. Abschnitt 2.4.2 Direktes Überlagern

²⁰vgl. Abschnitt 2.4.2 Merkmalsbasiert

der Kopf einer Person. Temperaturen, die unter einem festzulegenden Schwellenwert liegen, können somit aus der Aufnahme entfernt werden und dadurch – den wärmeren – Vordergrund freistellen. Auf den visuellen Bildern kann die Extraktion der Silhouette des Gesichts sowohl mittels eines *face landmarking* erfolgen als auch durch zusätzlich aufgezeichnete Tiefeninformationen, die Vordergrund und Hintergrund voneinander trennen.

Da bei der *flächenbasierten* Registrierung²¹ möglichst planare Ebenen im Bild vorhanden sein müssen, schneidet dieses Verfahren in Bezug auf ein Szenario der Gesichtsregistrierung als nicht geeignet aus. Alternativ könnte die Registrierung anhand eines bzw. mehrerer Referenzbilder stattfinden und dann auf andere Aufnahmen der Serie angewandt werden. Hierfür bietet sich ein Kalibrierungsmuster an, das mit beiden Modalitäten aufgezeichnet wurde. Da typischerweise aufgezeichnete Bildserien nicht über ein Kalibrierungsmuster verfügen, ist dieser Ansatz auch nicht für generische Anwendungsfälle anwendbar. Erschwerend kommt hinzu, dass eine Registrierung nur für die Entfernung (Gegenstandsweite) exakt möglich ist, die in den Referenzbildern das Kalibrierungsmuster zur Kamera hat. Ändert sich die Gegenstandsweite zur Kamera, kommt es zu Abweichungen bei der Registrierung der Bilder.

Das Verfahren der *manuellen* Zuordnung²² von Bildinhalten in den beiden Modalitäten entfällt ebenso als Möglichkeit. Grundsätzlich ist dieses Verfahren zwar durchaus möglich, sofern es sich um die Registrierung von einzelnen Bildpaaren handelt. Für das Verarbeiten von mehreren Bildpaaren ist für dieses Verfahren jedoch ein sehr hoher zeitlicher Aufwand nötig, den es zu vermeiden gilt.

Insgesamt wurde in der Analyse ermittelt, dass eine Registrierung bestenfalls anhand der Bildinhalte stattfinden sollte. Hier zeigt sich, dass besonders die Verfahren, die die verschiedenen Modalitäten auf einer Abstraktionsebene wieder zusammenbringen – hier *Kantenextraktion* und *Silhouettenextraktion* genannt – am geeignetsten erscheinen, um eine präzise Registrierung von Aufnahmen von Gesichtern durchzuführen.

²¹vgl. Abschnitt 2.4.2 **Flächenbasiert**

²²vgl. Abschnitt 2.4.2 **Manuell**

2.5. Verwandte Arbeiten

Im Folgenden werden einige Arbeiten vorgestellt, die sich mit einer ähnlichen Fragestellung beschäftigen wie diese Thesis. Ausgewählt wurden sowohl Arbeiten, die Thermografiebilder und visuelle Bilder miteinander kombinieren, wie auch Arbeiten die sich grundsätzlich mit der Fusion multimodaler Bildern befassen.

2.5.1. THERMAL FACE RECOGNITION IN AN OPERATIONAL SCENARIO

In der Studie **THERMAL FACE RECOGNITION IN AN OPERATIONAL SCENARIO** von **Socolinsky und Selinger (2004)** der EQUINOX CORPORATION²³ wurde erforscht, inwieweit eine Fusion von visuellen Bildern und Thermografiebildern eine *face detection*²⁴ verbessern kann. Besonderes Augenmerk dieser Studie ist es, dass die Ergebnisse in verschiedenen Beleuchtungssituationen und Hintergründen betrachtet wurden. Hierzu wurde eine Vielzahl von Variationen der Aufnahmen für die Versuche genutzt. Möglich war dies durch die Vielzahl von Probanden (385), die unter verschiedensten Bedingungen aufgezeichnet wurden (im Innen- und Außenbereich inkl. unruhigem Hintergrund wie Fußgänger und Fahrzeuge, aber auch direkt nach körperlicher Anstrengung und verschiedensten Akklimatisationszeiträumen nach größeren Temperaturunterschieden). Dadurch gilt diese Studie als größte jemals veröffentlichte Studie zur Gesichtserkennung mit der Nutzung von Thermografiebildern in Bezug auf Anzahl der Bilder und Probanden.

Ebenso nutzten die Autoren eine im Vergleich zu anderen Veröffentlichungen verhältnismäßig hochauflösende Thermografiekamera mit der Auflösung 320×240 Pixel im Infrarotspektrum zwischen 8μ und 12μ . Ein weiterer interessanter Aspekt ist, dass das in der Studie genutzte System vollkommen autonom arbeitet und somit keine manuelle Datenvorbehandlung/Normalisierung (wie Zuschneiden der Bilder oder Änderung der Helligkeit) stattfand.

Im Rahmen der Studie wurde die Genauigkeit der *face detection* für die Algorithmen PRINCIPAL COMPONENT ANALYSIS (PCA), LINEAR DISCRIMINANT ANALYSIS (LDA) und die Eigenentwicklung von EQUINOX – sowohl auf Bildern im visuellen Spektrum und auf den Thermografiebildern als auch nach einer Fusion beider Arten von Bildern – gemessen und jeweils miteinander verglichen. Die Ergebnisse zeigen, dass nach der Fusion der Bilder die Erkennungsquote bei allen verwendeten Algorithmen verbessert wird. Besonders stark steigen die Quoten bei der Verwendung im Außenbereich. Für die Kalibrierung der beiden Kameras wurde die sog. *Einpunkt-Kalibrierung* verwendet. Die tatsächlich gemessenen Temperaturen spielen

²³<http://www.equinoxsensors.com>

²⁴vgl. Abschnitt 2.2 *face detection*

bei dieser Studie keine Rolle und werden daher weder ausgewertet noch in den Prozess der *face detection* mit einbezogen.

2.5.2. INFRARED HUMAN FACE AUTO LOCATING BASED ON SVM AND A SMART THERMAL BIOMETRICS SYSTEM

Das Paper **INFRARED HUMAN FACE AUTO LOCATING BASED ON SVM AND A SMART THERMAL BIOMETRICS SYSTEM** von **Xiaoyu, Jihong, Pingjiang und Zhihong (2006)** beschreibt die Problematik der automatisierten Temperaturmessung des menschlichen Kopfes mit einer Thermografiekamera. Ausschlaggebend für diese Forschung war die SARS-Pandemie (*severe acute respiratory syndrome*) 2003 in China mit tausenden Toten und der Wunsch nach einer automatisierten, berührungslosen Temperaturmessung von Passagieren an Flughäfen.

Die *face detection* wurde hier ausschließlich mittels Thermografiebildern, d. h. ohne ein zusätzliches visuelles Bild durchgeführt. Dazu wurde mit eigenen Datensätzen eine *support vector machine* (SVM) trainiert und diese anschließend genutzt. Der Vorteil dieses Systems ergibt sich daraus, dass eine erhöhte Temperatur grundsätzlich an beliebiger Stelle im betrachteten Bildausschnitt auftreten kann (in diesem Fall Temperaturen zwischen 30 – 40 °C). Eine genauere Festlegung auf bestimmte Bereiche/ROI des Gesichts erfolgt nicht. Durch die *face detection* wird nur sichergestellt, dass auf der aktuellen Aufnahme genau eine Person dargestellt ist. Aus der Arbeit geht darüber hinaus hervor, dass eine absolute Temperatur zu einem Zeitpunkt, zu dem kein näheres Wissen über eine Person besteht, nur bedingt aussagekräftig ist. Interessanter ist die Fragestellung nach der relativen Temperaturänderung über die Zeit. Gleichwohl kann ein solches System dennoch dabei unterstützen, potenziell Infizierte zu detektieren und eine genauere medizinische Untersuchung zu veranlassen.

2.5.3. ADVANCED SURVEILLANCE SYSTEMS: COMBINING VIDEO AND THERMAL IMAGERY FOR PEDESTRIAN DETECTION

Im Paper **ADVANCED SURVEILLANCE SYSTEMS: COMBINING VIDEO AND THERMAL IMAGERY FOR PEDESTRIAN DETECTION** von **Torresan, Turgeon, Ibarra-Castanedo, Hebert und Maldague (2004)** wird ein Sicherheits- und Überwachungssystem für den Außenbereich vorgeschlagen, mit dem Fußgänger robust – unabhängig von der Beleuchtungssituation – erkannt und getrackt werden können. Hierzu wurde das Bild einer visuellen Kamera mit dem einer Thermografiekamera kombiniert, um die Vorteile beider Spektren zu nutzen.

Technisch wurde eine Thermografiekamera mit einer Auflösung von 320×256 Pixeln und eine visuelle Kamera mit einer Auflösung von 640×480 Pixeln verwendet. Da es sich bei der

Fußgängererkennung nicht um eine hochpräzise Erkennung auf Pixelebene handelt, wurde ein sog. „*region-based approach*“ gewählt. Daher wurde nur geometrisch mit intrinsischen Parametern²⁵ kalibriert. Da die Erkennung vor dem Zusammenführen beider Bilder parallel (d. h. sowohl auf dem Thermografiebild als auch auf dem visuellen Bild) stattfindet, können je nach Spektrum verschiedene (nicht zwangsläufig dieselben/dieselbe Anzahl von) Personen erkannt werden. Somit ist das System weitaus robuster als die alleinige Verwendung einer der beiden Kameras. Die tatsächliche Temperatur der Personen ist in dieser Arbeit nebensächlich und dient vielmehr der Unterscheidung und Abgrenzung des Hintergrunds aufgrund einer unterschiedlichen Temperatur im Vergleich zum Menschen. Gleichzeitig ermöglicht der Temperaturunterschied ein Freistellen der Person vom Hintergrund.

2.5.4. IMAGE FUSION OF VISIBLE AND THERMAL IMAGES FOR FRUIT DETECTION

In der Veröffentlichung **IMAGE FUSION OF VISIBLE AND THERMAL IMAGES FOR FRUIT DETECTION** von **Bulanon, Burks und Alchanatis (2009)** wird ein komplettes Systemsetup und die Durchführung einer Fusion zweier Aufnahmen beschrieben. Das System teilt die Rahmenbedingungen und das Ziel der Fusion. Dieses ist nah an dem hier zu betrachtenden Szenario, weshalb diese Arbeit hier aufgeführt wird. In dem Aufbau wird ein multimodales Kamerasystem aus zwei verschiedenen Kameras genutzt (bestehend aus einer Thermografiekamera sowie einer Kamera aus dem sichtbaren Spektrum). Die verwendeten Kameras zeichnen Bilder von Orangenbäumen in einer Orangenplantage aus unmittelbarer Nähe auf. Die Aufnahmen werden dann nach einer Vorverarbeitung (Angleichen des Kamerasichtfeldes, Konvertierung in Graustufen etc.) registriert. Anschließend werden die Bildinhalte in ein neues Bild zusammengeführt. Ziel ist es die Detektion der Orangen – und damit ein automatisiertes Ernten – durch die Verwendung der beiden genutzten Spektren zu optimieren. Die tatsächliche Temperatur der Orangen, der Äste, Blätter usw. spielt in dem Szenario keine Rolle; es wird sich nur die physikalische Eigenschaft zunutze gemacht, dass die Orangen bei Sonnenbestrahlung eine höhere Temperatur (und daher eine kontraststärkere Färbung in der Falschfarbendarstellung) annehmen als die Umgebung.

Für die Registrierung der beiden Aufnahmen befindet sich vor den Orangenbäumen ein 0,5m großer quadratischer Rahmen aus PVC-Rohren. Mithilfe dieses Rahmens, der auf beiden Aufnahmen sichtbar ist, findet die Ausrichtung der Registrierung statt, d. h. die Bestandteile des Baums werden beim Registrierungsprozess nicht betrachtet. Dies ermöglicht eine Registrierung der Bilder unabhängig von der Variation der Bildinhalte (beispielsweise veränderter Schattenwurf, Bewegung durch Wind oder Veränderung der Szene durch Wachstum), sofern

²⁵vgl. Abschnitt 2.4.1 Kameraparameter

der Rahmen unter allen Bedingungen unverändert bleibt. Technisch findet der Registrierungsprozess mit der MATLAB IMAGE PROCESSING TOOLBOX statt, in der die Aufnahmen mit linearen 2D-Transformationen aufeinander ausgerichtet werden.

2.5.5. NON-RIGID VISIBLE AND INFRARED FACE REGISTRATION VIA REGULARIZED GAUSSIAN FIELDS CRITERION

Die Arbeit **NON-RIGID VISIBLE AND INFRARED FACE REGISTRATION VIA REGULARIZED GAUSSIAN FIELDS CRITERION** von **Ma, Zhao, Ma und Tian (2015)** befasst sich mit der Problematik der Registrierung von multimodalen Aufnahmen von nicht planaren Objekten, die mittels

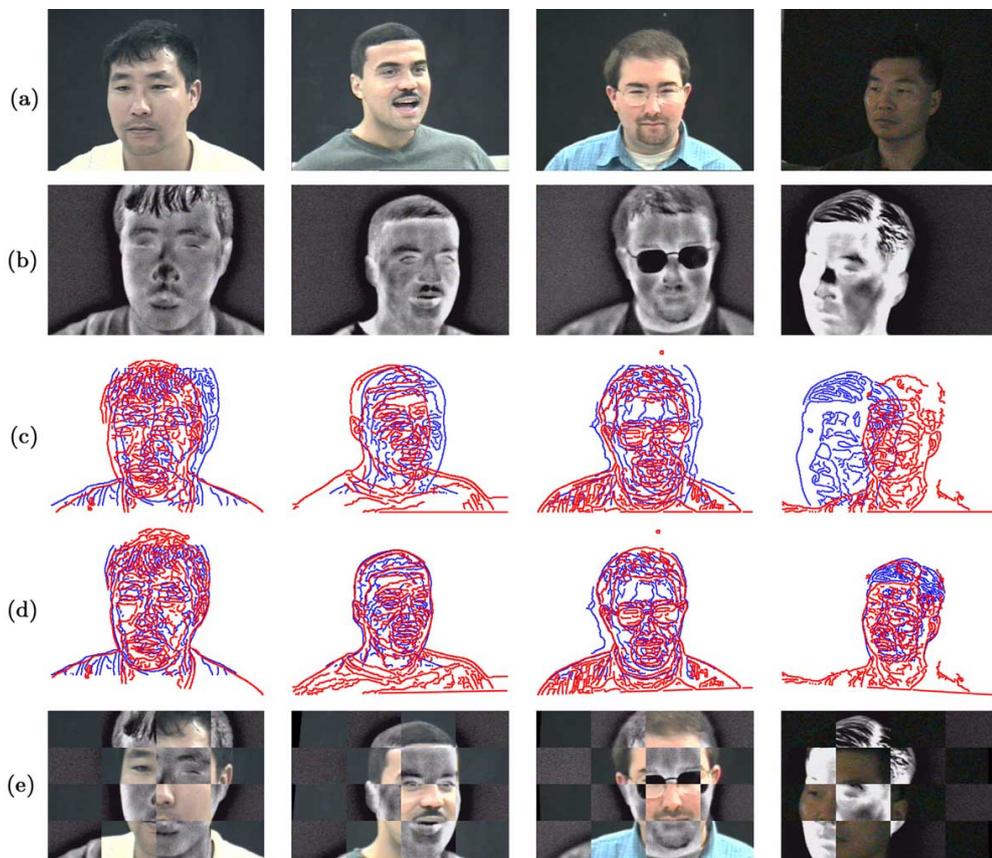


Abbildung 2.19.: Darstellung des Registrierungsprozesses: (a) visuelles Spektrum, (b) Thermografiebild, (c) Kantenextraktion und einfache Überlagerung (rot = visuelles Bild; blau = Thermografiebild), (d) registrierte Kantenextraktion, (e) Überlappung als Schachbrettmuster nach erfolgreicher Registrierung.

mit Genehmigung der Autoren **Ma u. a. (2015)**, auf Basis von Bildern der UTK-IRIS THERMAL/VISIBLE FACE DATABASE

Thermografiekameras und Kameras des visuellen Spektrums aufgezeichnet wurden. Das von den Autoren vorgestellte Verfahren demonstriert die Registrierung anhand von Aufnahmen von Gesichtern, die aus kurzer Distanz aufgezeichnet wurden. An den Registrierungsaspekt werden in dieser Veröffentlichung ähnliche Anforderungen gestellt wie jene, denen im Rahmen dieser Arbeit begegnet werden muss.

Für die Registrierung der beiden Aufnahmen werden von diesen die Kanten der Bildinhalte extrahiert. Erzeugt werden diese mittels des weitverbreiteten CANNY-EDGE-DETECTOR-ALGORITHMUS [Canny (1986)]. Der eigentliche Registrierungsprozess findet dadurch nicht auf dem Thermografiebild bzw. der Aufnahme im visuellen Spektrum statt, sondern auf den daraus extrahierten Kantendarstellungen. Hierbei wird die Eigenschaft ausgenutzt, dass auf den Bildern eine merkmalsbasierte Registrierung – aufgrund der wenig übereinstimmenden Merkmale beider Modalitäten – äußerst kompliziert ist, in der Kantendarstellung jedoch weit mehr gleiche bzw. ähnliche Merkmale dargestellt werden. Dabei stellen die Autoren allerdings auch klar, dass auch die durch die Kanten dargestellten Merkmale im Gesicht sich nur unzureichend aufeinander abbilden lassen und somit die meisten Abbildungen auf den Konturen des Kopfes stattfinden. Diese Kanten werden durch eine abgewandelte Variante eines sog. *Gaussian mixed model* aufeinander abgebildet, das ausführlich in der Arbeit beschrieben wird. Der Registrierungsprozess der Autoren sowie die dazugehörigen Aufnahmen sind für einige Beispiele in Abbildung 2.19 dargestellt.

2.6. Szenarien- und Anforderungsanalyse

In den nachfolgenden Abschnitten wird u. a. das Szenario erläutert, in dem das zu entwickelnde Softwaresystem eingesetzt wird. Dabei handelt es sich um das Forschungsprojekt EMOTIONBIKE, das im Nachfolgenden kurz umrissen wird. Darauf folgen die Anforderungen an das System, die sich aus der hier vorgenommenen Analyse entwickelt haben und aus den verwandten Arbeiten gelernt werden konnten. Diese Anforderungen werden technisch und funktional an das System gestellt, damit dieses zur weiteren Analyse genutzt werden kann. Ebenso werden – als Abgrenzung dieser Arbeit – die Aspekte aufgeführt, die in diesem Rahmen nicht adressiert werden sollen.

2.6.1. EMOTIONBIKE

Das EMOTIONBIKE²⁶ (EMOBIKE) ist ein Forschungsprojekt der HAW HAMBURG, in dessen Rahmen neue Wege zur HCI und der Companion-Technologie untersucht werden. Im Mittelpunkt

²⁶<https://emotionbike.org>

steht die personenzentrierte Messdatenerhebung, die mittels verschiedener Sensoren durchgeführt wird. Zum Einsatz kommen in der aktuellen Ausprägung kamerabasierte Sensoren, Biosensoren und Sensoren an den zum Versuchsaufbau gehörenden Hardwarekomponenten, die Informationen über den emotionalen sowie physischen Zustand einer Person ermitteln. Technisch besteht das EmotionBike aus einem klassischen Ergometer (vgl. Abbildung 3.3), an dem eine Vielzahl von Veränderungen vorgenommen wurden, sowie einem Fernseher, der die virtuelle Welt projiziert, in der die Person interagiert. Die Laborumgebung erlaubt es, sich mittels Fahrrad frei in dieser zu bewegen. Dazu werden die Tret- und Lenkbewegungen in die virtuelle Welt übertragen. In dieser können verschiedene Ereignisse simuliert werden, um bei der Person eine körperliche Reaktion hervorzurufen, die mittels der Sensorik detektiert wird. Der Companion-Ansatz besteht hierbei darin, je nach körperlichem oder emotionalem Zustand unterschiedlich auf die Person einzuwirken [Müller u. a. (2015); Müller u. a. (2016)].

Das EMOTIONBIKE-Projekt wurde mit einer Thermografiekamera erweitert, um die thermischen Veränderungen während der hervorgerufenen Ereignisse zu ermitteln. Die Integration dieser zusätzlichen Kamerakomponente sowie die Kommunikation mit den anderen Komponenten wurde bereits in [Kletz und Kleimann (2017)] dokumentiert und wird im Zuge dieser Thesis weiter vertieft.²⁷ Der Projektaufbau und die darin erzeugten Thermografieaufnahmen stellen die Datengrundlage dar, auf der das in dieser Thesis vorgestellte Verfahren zur thermografischen Gesichtsanalyse entwickelt werden soll.²⁸

2.6.2. Anforderungen

Aus der Analyse ergeben sich eine Reihe von Anforderungen an das zu entwickelnde System. Diese werden im Folgenden genannt und kurz erläutert.

- **Ausgabe von gemessenen Temperaturwerten auf festgelegten ROI im Gesicht**

Das primäre Ziel ist es, Temperaturwerte eines Gesichts punkt- und/oder bereichsgenau abzufragen. Daraus ergibt sich die Anforderung, ebenjene Temperaturen mittels einer geeigneten Schnittstelle auszugeben und damit anderen Personen oder Systemen zur Weiterverarbeitung zur Verfügung zu stellen.

- **Definition der Datenstruktur**

Für die Ausgabe der Temperaturwerte gilt es, eine Datenstruktur zu definieren und diese zu dokumentieren, um eine Weiternutzung zu ermöglichen.

²⁷ vgl. Abschnitt 3.2.4 EMOTIONBIKE Aufbau

²⁸ vgl. Abschnitt 1.1 Zielsetzung

- **Face landmarking auf visuellen Aufnahmen**

Für eine präzise Lokalisation von Gesichtsmarkmalen ergibt sich aus der Analyse der Gesichtserkennung, dass ein Verfahren benötigt wird, das präzise Positionen des Gesichtes auf dem Bild ausgeben kann. Um möglichst pixelgenau die Position der Gesichtsmarkmale zu extrahieren, wird daher eine Gesichtserkennung in der Ausprägung des *face landmarkings* benötigt.

- **Registrierung von visuellen und Thermografieaufnahmen**

Bereits in der Analyse zeigt sich, dass eine detaillierte Gesichtserkennung auf Thermografiebildern nicht robust durchführbar ist. Zudem werden visuelle Informationen für eine spätere Anzeige der Gesichter benötigt. Daher ist es nötig, Aufnahmen, die mit einer visuellen Kamera und einer Thermografiekamera entstehen, miteinander zu registrieren und somit die Aufnahmen mit all ihren Informationen (visuelle Details und Temperaturinformationen) zu überlagern.

2.6.3. Abgrenzungen

Neben den Anforderungen gibt es auch sog. Nicht-Ziele, von denen hier die hauptsächlichsten aufgeführt werden:

- **Analyse der Temperaturen**

Eine Interpretation der extrahierten Temperaturwerte ist nicht Gegenstand der Arbeit.

- **Evaluierung der Eignung für HCI**

Da keine Analyse der Temperaturen stattfindet, erfolgt auch keine Evaluierung, inwieweit die Daten sich für die HCI oder ein *Affective-Computing*-System eignen.

- **Verbesserung der Objekterkennung auf Thermografiebildern**

Ebenso ist es nicht Gegenstand dieser Arbeit, ein Verfahren zu entwickeln, um eine robuste und präzise Objekterkennung oder Gesichtserkennung auf Thermografiebildern zu ermöglichen.

3. Design & Realisierung

Dieses Teil bildet den Praxisanteil der Arbeit. Hierzu werden die Erkenntnisse und theoretischen Grundlagen, die im Analyseteil gewonnen wurden, umgesetzt und implementiert. Ziel dieses Teils ist es die eingesetzte Hard- und Software zu beschreiben sowie die entwickelten Verfahren zu erläutern.

Gegliedert ist dieser Teil in vier Kapitel. Das Kapitel ‚**Designentscheidungen**‘ nennt kurz allgemeine Grundsatzentscheidungen, die getroffen werden und dann in den nachfolgenden Kapiteln wieder aufgegriffen werden. Das Kapitel ‚**Hardwaredesign**‘ beschreibt neben den technischen Merkmalen der zur Verfügung stehenden optischen Geräten die typischen Hardware-Setups, in denen das entwickelte System eingesetzt werden soll. Die Kapitel ‚**Softwaredesign**‘ und ‚**Softwarefusion**‘ gehen auf die Softwarekomponenten ein, die im Rahmen dieser Arbeit verwendet werden. Dazu wird im erstgenannten Kapitel die Architektur des Gesamtsystems sowie die dafür genutzten Module erläutert. Ein Abschnitt widmet sich dazu explizit der Bilder- bzw. Datenakquirierung der verschiedenen Kamertypen mittels der durch die Hersteller zur Verfügung gestellten *Software Development Kit* (SDK). Der Abschnitt Gesichtserkennung befasst sich mit der Detektion von Gesichtsmerkmalen, dem *face landmarking*. Mittels dieser Informationen werden die ROI, die für die Temperaturmessung interessant sind, extrahiert. Das Kapitel Softwarefusion beschreibt die softwaretechnische Realisierung des Registrierungsvorgangs der multimodalen Aufnahmen. Der Abschnitt Kamerakalibrierung beschreibt eine Vorbedingung des Registrierungsvorgangs, bei dem die bauliche Linsenverzeichnung der Kameras beseitigt wird, um deren Präzision zu erhöhen. Gefolgt wird dies von einem Abschnitt, der die (zeitliche) Synchronisierung und die nötige Ausdünnung der Aufnahmen für den Fall beschreibt, dass mit verschiedenen Startzeitpunkten und/oder Geschwindigkeiten aufgezeichnet wird. Der Abschnitt Bildregistrierung beschreibt schließlich das eigentliche Verfahren, in dem die multimodalen Aufnahmen registriert werden, um eine möglichst große Übereinstimmung der Aufnahmen zu erreichen, die wiederum benötigt wird, um eine genaue thermografische Gesichtsanalyse durchzuführen.

3.1. Designentscheidungen

Als eine grundlegende Designentscheidung wird festgelegt, dass für die thermografische Gesichtsanalyse, eine Kombination aus zwei unabhängig operierenden Kameras genutzt wird. Begründet wird dies aus den Erkenntnissen des Abschnitts 2.2.1 die gezeigt haben, dass ein *face landmarking* mit frei verfügbaren Algorithmen auf Thermografiebildern nicht fehlerfrei möglich ist. Die verwendeten Kameras arbeiten dazu in verschiedenen Modalen, um sowohl die sichtbaren Informationen des visuellen Spektrums als auch die Temperaturdaten des Thermografiespektrums zur Verfügung zu haben.

3.2. Hardwaredesign

Das Hardwaredesign beschreibt die für die Szenarien verwendete Hardware mit ihren technischen Eigenschaften. Kern dieses Kapitels bilden die drei verwendeten Kamerasysteme. Diese bestehen zum einen aus der Thermografiekamera, die in beiden Szenarien zum Einsatz kommt, und zum anderen aus zwei verschiedenen Kamerasystemen, die Bildinformationen im visuellen Spektrum aufzeichnen. Die Kameras im visuellen Spektrum werden für die Gesichtserkennungskomponente genutzt, die in späteren Schritten mit den Thermografiedaten der Thermografiekamera angereichert wird. In diesem Kapitel wird zudem auf die Versuchsaufbauten eingegangen, in denen die Kamerasysteme eingesetzt werden. Dazu werden sowohl die physikalischen Rahmenbedingungen genannt als auch Möglichkeiten zur Überführung in reelle Einsatzszenarien aufgezeigt.

3.2.1. INFRA TEC VARIOCAM HD 875

In diesem Abschnitt wird auf die durch die HAW HAMBURG beschaffte Thermografiekamera aus dem Hause INFRA TEC¹ eingegangen. Es handelt sich um das Modell VARIOCAM HD 875 (vgl. Abbildung 3.1), das zur Gruppe der qualitativ hochwertigen Geräte gezählt werden kann. Bedingt wird dies durch eine mögliche Auflösung von 1.024×768 IR-Pixeln² und eine Aufnahmegeschwindigkeit von 30 Hz im Vollbildmodus. Als Detektor nutzt die Kamera einen ungekühlten Mikrobolometer-Focal-Plane-Array-Detektor. Die Messgenauigkeit der Kamera beträgt $\pm 1 \text{ K}$ im Temperaturbereich zwischen $(0 - 100)^\circ\text{C}$ und ist daher u. a. ideal für die Messung von Körpertemperaturen und alltäglichen Gegenständen. Eine vollständige Auflistung der technischen Spezifikationen sind in [InfraTec (2015a)] aufgeführt.

¹<http://www.infratec.de/>

²Infrarot-Pixeln

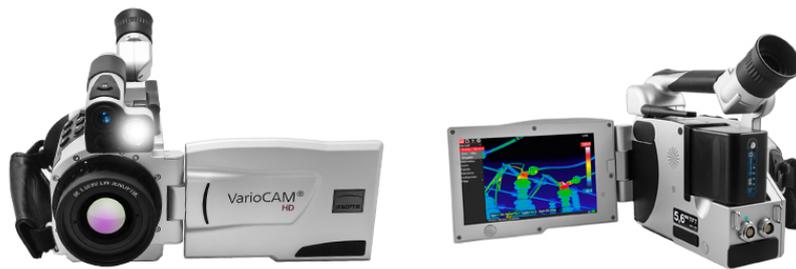


Abbildung 3.1.: Front- und Rückansicht der INFRA TEC VARIOCAM HD 800 SERIE Quelle: ©InfraTec

Als Schnittstelle verfügt die Kamera über einen *GigE-Vision-Interface-Standard*-Anschluss, der auf dem Gigabit-Ethernet-Standard arbeitet. Diese Schnittstelle ermöglicht es, mittels einer zusätzlichen Netzwerkkarte auf die Daten der Kamera sowohl via proprietärer Software als auch über ein SDK zuzugreifen. Trotz des Ethernet-Standards sind manuelle Konfigurationen der Netzwerkkarte nötig. Eine Anleitung hierfür findet sich in [InfraTec (2014)]. Als zusätzliches optisches Gerät verfügt das verwendete Modell über eine Kamera die Bilder im sichtbaren Spektrum aufzeichnet. Aufgezeichnet können Daten dieser Kamera jedoch nur lokal im Gerät. Ein direkter Zugriff auf diese Daten ist bisher mittels SDK nicht vorgesehen. Eine automatische Bildregistrierung zwischen den beiden Aufnahmen ist ebenfalls nicht Bestandteil der Hardwarekomponente.

	Kamera	Linse
Auflösung	$1024 \times 768px$	-
max. Frequenz	$30Hz$	-
Temperatur Auflösung	$< 0,05 K$	-
Spektralbereich	$7,5 - 14 \mu m$	-
Brennweite	-	$30mm$
Bildfeld	-	$1.0/30mm$ $(32,4 \times 24,6)^\circ$
Physikalische Abmessungen $(B \times H \times T)$	$15,5cm \times 12,5cm \times 21cm$	

Tabelle 3.1.: Technische Daten der INFRA TEC VARIOCAM HD 875 [InfraTec (2015a)]

SDK Die Kamera verfügt über ein SDK, mit dem es möglich ist, eigene Applikationen zu schreiben, mit denen auf die Funktionen der Kamera zugegriffen werden kann. Technisch wird dafür vom Hersteller eine *Dynamic Link Library* (DLL) angeboten, die in *Borland Delphi 7.0* erstellt wurde. Eingebunden werden kann diese DLL u. a. im Microsoft .NET-Ökosystem oder auch in einer C++ Umgebung. Die Funktionen des SDK lassen sich in zwei Gruppen einteilen. Eine Gruppe bilden Funktionen, mit deren Hilfe die Kamera gesteuert wird. Hierzu zählen beispielsweise Fokussierungseinstellungen oder das Lesen und Setzen von Kameraparametern. Eine zweite Gruppe bilden Funktionen, mit denen die Temperaturrohdaten von der Kamera abgerufen werden. Eine ausführliche technische Beschreibung hierzu befindet sich in [InfraTec (2014)].

3.2.2. ALLIED VISION MANTA G-235C

Neben der Thermografiekamera wird noch eine weitere Kamera benötigt, die im sichtbaren Spektrum arbeitet. Für diesen Aufbau wird eine MANTA G-235C-Kamera des Herstellers ALLIED VISION³ mit einem 1" 16mm/F1.4-Objektiv der Firma KOWA⁴ als Videokamera verwendet, die – wie die Thermografiekamera – über den *GigE-Vision-Interface-Standard* angesprochen wird. Das ausgewählte Modell wird mittels *Power over Ethernet* (PoE) mit Strom versorgt, weshalb in dem konkreten Projektaufbau ein separater PoE-Splitter verwendet wird. Anhand des Bildes dieser Kamera wird die Gesichtserkennung durchgeführt.

	Kamera	Linse
Auflösung	1936 × 1216px	-
max. Frequenz	50,7Hz	-
Diagonaler Bildwinkel		22.72°
Verzeichnung	-	-1,0 %
Linsenfassung	C-Mount	C-Mount
Physikalische Abmessungen (B × H × T)	2,9cm × 4,4cm × 8,63cm	4,3cm × 4,3cm × 5,43cm ×

Tabelle 3.2.: Technische Daten der ALLIED VISION MANTA G-235C [Allied Vision (2017)]

³<http://www.alliedvision.com>

⁴<http://www.kowa.eu/lenses>

SDK Wie die Thermografiekamera verfügt auch die visuelle Kamera über ein SDK, mit dem diese angesprochen werden kann. Das SDK mit dem Namen VIMBA bietet *Application Programming Interfaces* (APIs) für die Nutzung mit den Programmiersprachen C und C++ sowie für jene des .NET-Ökosystems an. Mit diesen ist es möglich, Bilder von der Kamera zu akquirieren und diese dann durch separate Algorithmen weiterzuverarbeiten [Allied Vision (2013)]. Da die Kamera und das Objektiv keine Autofokussierungsmöglichkeit bieten, lässt sich diese Funktion auch nicht mit dem SDK durchführen. Die Fokussierungseinstellung muss daher manuell an der Hardware vorgenommen werden und kann sich nicht dynamisch ändern.

3.2.3. MICROSOFT KINECT FOR XBOX ONE

Als eine weitere Kamera, die u. a. im sichtbaren Lichtspektrum arbeitet, wird die KINECT FOR XBOX ONE (Modell 1520) aus dem Hause MICROSOFT⁵ verwendet. Neben der Aufzeichnung des sichtbaren Lichtspektrums ist diese zusätzlich in der Lage, Bilder im Infrarotspektrum aufzuzeichnen und kann so auch bei schwacher Beleuchtung operieren. Des Weiteren verfügt sie über ToF-Tiefensensoren und ist dadurch dazu fähig, dreidimensionale Aufnahmen der Szenerie aufzuzeichnen.

Die verschiedenen Optiken verfügen über verschiedene Auflösungen. So verfügt die Farbkamera über eine Auflösung von 1080p (1920 × 1080 Pixel) bei einer Frequenz von ebenfalls 30Hz, die Infrarotkamera und die Tiefensensoren verfügen hingegen nur über 512 × 424 Pixel bei einer Frequenz von 30Hz. Die Kommunikation mit der Kinect-Kamera erfolgt über USB 3.0 und den sog. Kinect-Adapter, der gleichzeitig als Stromversorger für die Kamera fungiert.

	Farbkamera	Infrarotkamera	Tiefensensor
Auflösung	1920 × 1080px	512 × 424px	512 × 424px
max. Frequenz	30Hz ⁶	30Hz	30Hz
Bildwinkel	84.1° × 53.8°		70° × 60°
Physikalische Abmessungen (B × H × T)	24,9cm × 6,6cm × 6,7cm		

Tabelle 3.3.: Technische Daten der MICROSOFT KINECT FOR XBOX ONE [Microsoft (2017)]

⁵<http://developer.microsoft.com/windows/kinect/>

⁶30Hz bei guten Lichtverhältnissen, 15Hz bei schlechten Lichtverhältnissen

SDK Das sog. *Kinect für Windows SDK 2.0* bildet das SDK für die Kinect und stellt hierfür Anwendungen der Programmiersprachen C++, C#, Visual Basic und anderen .NET-Programmiersprachen eine DLL bereit, über die die Funktionen abgerufen werden können. Ebenso ist ein Akquirieren der Bilder der verschiedenen Spektren möglich, die die Kinect verarbeitet. Des Weiteren bietet das SDK bereits viele Implementierungen wie eine Körperskeletterkennung oder Gestenerkennung, die für diesen Anwendungsfall allerdings nicht benötigt werden.

3.2.4. Setup

Das Hardwaresetup besteht aus zwei verschiedenen Aufbauten, da es sich um zwei verschiedene Szenarien handelt, in denen das System genutzt werden kann. Das erste davon wird im Unterabschnitt ‚Standalone-Aufbau‘ behandelt und ist ein vergleichsweise einfacher, aber dennoch vielfältig nutzbarer Aufbau. Das zweite Szenario hat einen komplexeren Aufbau, da es im Verbund mit anderen Sensoren verwendet wird. Dieses wird im Absatz ‚EMOTIONBIKE-Aufbau‘ betrachtet.

Standalone-Aufbau

Bei diesem Aufbau wird die Thermografiekamera in Kombination mit der MANTA-Kamera verwendet und frontal auf das zu analysierende Objekt ausgerichtet (in diesem Fall auf eine Person). Um den baulich bedingten Abstand der Linsen möglichst gering zu halten, wird die

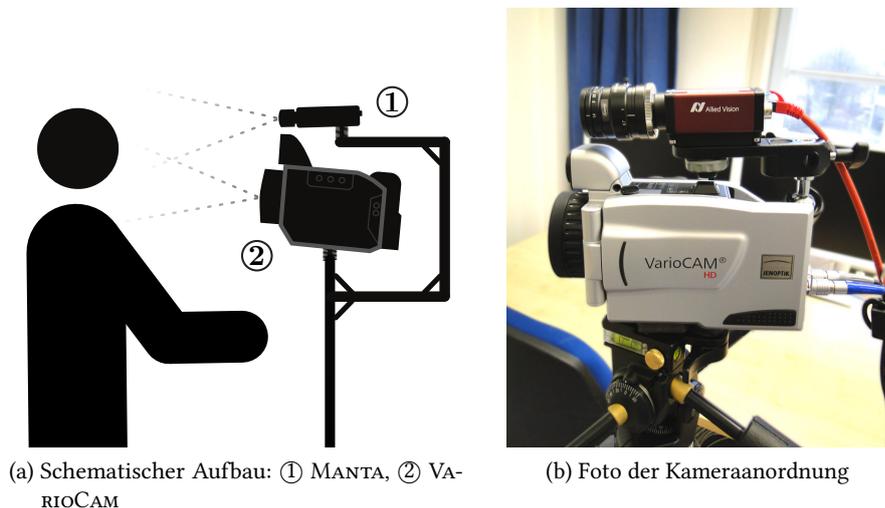


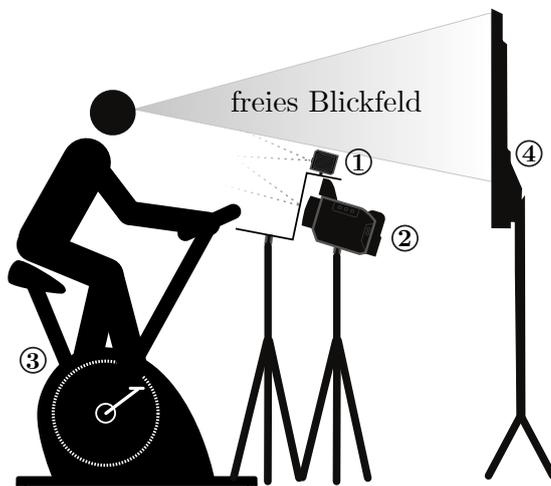
Abbildung 3.2.: VARIOCAM und MANTA-Kamera im Standalone-Aufbau

MANTA-Kamera mit einem Klemmstativ am Stativ der Thermografiekamera befestigt und lotrecht über dieser angebracht. Hierdurch wird versucht, den X-Achsen-Versatz möglichst zu eliminieren. Bezüglich der Z-Achse wird ebenso versucht, diese nahezu in eine Linie zu bringen. Der Abstand auf der Y-Achse ist im Vergleich zu den anderen Achsen am größten. Die Sichtachse der Kameras wird auf das Gesicht der zu analysierenden Person ausgerichtet, sodass ein möglichst frontales Aufzeichnen von diesem ermöglicht wird.

Dieses Szenario stellt durch den Aufbau ein sog. *passport scenario* dar, bei dem typischerweise Porträtaufnahmen genutzt werden. Aufgrund seiner orthogonalen Anordnung zum Probanden bietet es die Möglichkeit, den größtmöglichen Kopfbereich zu betrachten. Das Szenario eignet sich besonders zur Diagnostik von Temperaturen verschiedener Gesichtsbereiche.

EMOTIONBIKE-Aufbau

In diesem Aufbau wird die Thermografiekamera in einen komplexen, bereits bestehenden Aufbau eingebracht. Dieser verfügt bereits über eine KINECT-Kamera, weshalb diese darin genutzt wird, um die visuellen Bilder zu akquirieren. Auch bei diesem Aufbau wird versucht, eine bauliche Nähe zwischen den verschiedenen Kameras zu schaffen. Dazu wird die Thermografiekamera auf einem zusätzlichen Stativ montiert, das hinter den anderen Sensoren positioniert



(a) Schematischer Aufbau: ① KINECT, ② VARIOCAM, ③ Ergometer, ④ TV



(b) Foto der Kamera- und Sensorenanordnung

Abbildung 3.3.: VARIOCAM und KINECT im EMOTIONBIKE Aufbau. (Dargestellt sind hier nur die benötigten Komponenten. Teilweise zusätzlich dargestellte Komponenten sind nur nachrichtlich aufgeführt.)

wird. In die bestehenden Halterungen wird eine Öffnung integriert, die es ermöglicht, auch Person unterschiedlicher Körpergrößen zu beobachten, indem die Kamera je nach Person nachjustiert wird. Anders als beim ‚Standalone-Aufbau‘ ist es hier nicht möglich, die Verschiebung in der X-Achse gering zu halten. Bedingt wird dies durch die baulich nicht zentrale Anordnung der visuellen Kinect-Kamera (vgl. Abbildung 3.3b). Da die beiden Kameras ebenfalls vertikal übereinander angeordnet sind, entsteht zusätzlich ein Abstand auf der Y-Achse. Die Distanz der Kameras auf der Z-Achse wird durch den Aufbau ebenso möglichst gering gehalten. Eine Verdeckung des Probanden sowie seines Sichtfelds wird in diesem Aufbau ausgeschlossen.

Dieser Aufbau stellt einen Anwendungsfall nach, bei dem der Proband nicht primär auf den Kameraaufbau ausgerichtet ist und nicht einmal direktes Wissen über die Existenz dieses haben muss. Die (Kamera-)Sensoren sind als kontinuierliches Beobachtungssystem ausgelegt, das einen Probanden bei seiner Tätigkeit beobachtet und die aufgezeichneten Daten auswertet. Gleichzeitig beeinflussen sie den Probanden, indem sie diesen auf verschiedene Wege ent- oder belasten. Genutzt werden kann dies daher als ein Companion-System, das sich auf die Befindlichkeiten des Probanden einstellen. Ein Beispiel hierfür ist die Überwachung eines Fahrzeugführers, bei der ein Autopilotensystem eingreift, sobald es medizinische/emotionale Auffälligkeiten bei diesem feststellt.

3.3. Softwaredesign

Das Softwaredesign beschreibt die im Rahmen dieser Arbeit realisierten Softwarekomponenten, deren Ziel es ist, eine thermografische Gesichtsanalyse durchzuführen. Die Software folgt dabei dem Designprinzip der Modularität, sodass die Teilkomponenten austauschbar sind.

Die im Kapitel 3.2 beschriebenen Kameras verfügen über SDKs, die mittels den Programmiersprachen C++ bzw. C# angesprochen werden können. Aus diesem Grund fiel die Wahl für die Programmierung der Komponenten auf C++. Neben den Geschwindigkeitsvorteilen, die durch die hardwarenahe Programmierung gewonnen werden, ist somit auch ein Verarbeiten der Aufnahmen ohne vorherige Umwandlung möglich.

Um die Komplexität gering zu halten, werden ferner erprobte Algorithmen, Tools und Frameworks innerhalb des Softwarestacks eingesetzt und sinnvoll miteinander kombiniert. Deren Auswahl werden die Ergebnisse der Analyse in Teil 2 zugrunde gelegt.

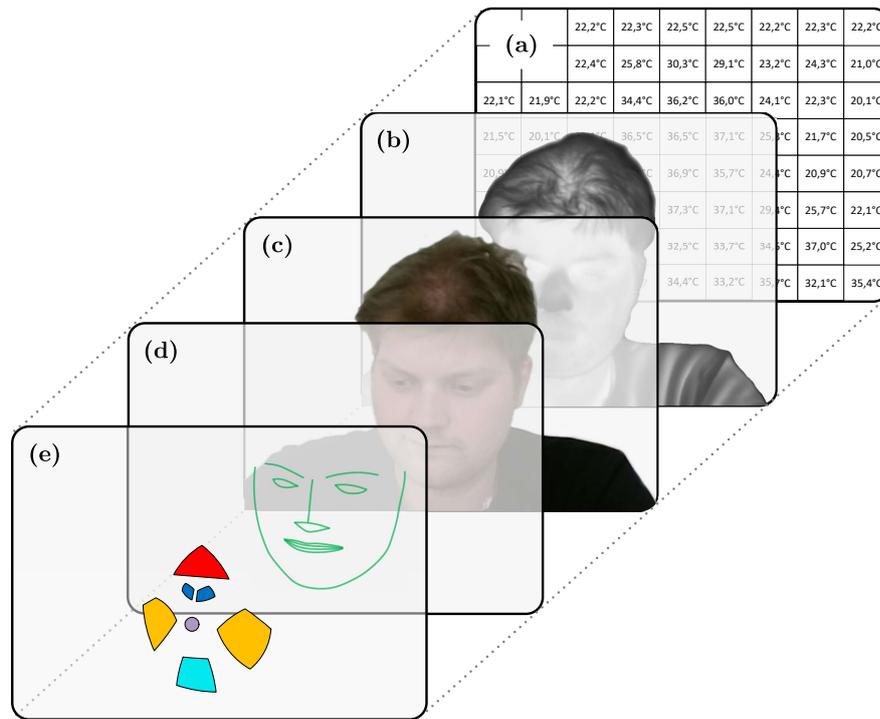


Abbildung 3.4.: Schichtendarstellung der einzelnen Elemente, die softwaretechnisch verarbeitet werden: (a) Temperaturrohdaten; (b) Thermografiebild; (c) visuelles Bild; (d) Gesichtserkennung; (e) ROI-Extraktion

Wie schematisch in Abbildung 3.4 dargestellt ist, bildet die Software aus den Rohdaten, die mittels der Kameras aufgezeichnet wurden, mehrere Schichten. Diese stehen in direktem Zusammenhang zueinander bzw. bilden sich jeweils aus den darunterliegenden Schichten. So bildet sich aus den Temperaturrohdaten (Abbildung 3.4a) das Thermografiebild⁷ (Abbildung 3.4b). Dieses gilt es mit dem visuellen Bild (Abbildung 3.4c) zu registrieren. Auf dem transformierten Bild (als Ergebnis der Registrierung) wird das *face landmarking* durchgeführt, um die Pixelpositionen der Gesichtsmerkmale zu bestimmen (Abbildung 3.4d). Mit diesen Positionen lassen sich schlussendlich die ROI lokalisieren (Abbildung 3.4e), über die die tatsächliche Temperaturextraktion erfolgt.

Eine genaue Beschreibung der einzelnen Verfahren wird in diesem sowie im nachfolgenden Kapitel dargestellt. So wird in diesem Kapitel auf die darunterliegende Architektur sowie einzelne Module eingegangen. Hierzu zählen die Aspekte der Bild- und Datenakquirierung der verschiedenen Kameras, der Themenbereich der Gesichtserkennung inkl. der Bestimmung der ROI sowie die pixelgenaue Temperaturextraktion und deren Rückgabe. Das darauffolgende Kapitel befasst sich mit der eigentlichen Softwarefusion der verschiedenen Kameraaufnahmen.

3.3.1. Architektur

Die Architektur beschreibt die verschiedenen Module, die benötigt werden, um eine thermografische Gesichtsanalyse in einem multimodalen Kamerasystem durchzuführen. Konkret besteht die Software aus den folgenden Modulen, deren Ablauf in Abbildung 3.5 schematisch dargestellt wird:

Bilder- und Datenakquirierung Dieses Modul ist für die Akquirierung der Daten von den eingesetzten Kameras zuständig. Diese wiederum werden durch die von den Herstellern zur Verfügung gestellten SDKs angesprochen und arbeiten autark voneinander. Gesteuert werden die Aufnahmezeitpunkte durch das in der Analyse erwähnte EMOTIONBIKE-Projekt. Eine detailliertere Beschreibung ist in Unterabschnitt [3.3.2 Bilder- und Datenakquirierung](#)⁵ zu finden.

Bilddatenvorverarbeitung bzw. Bilddatenerzeugung Dieses Modul steht stark mit dem zuvor genannten in Verbindung und ist dafür zuständig, aus den Temperaturrohdaten der Thermografiekamera ein grafisches Bild zu erzeugen. Dazu wird jeder Temperatur ein eindeutiger Farbwert zugewiesen und das Bild pixelweise zusammengesetzt. Die Aufnahmen

⁷vgl. Abschnitt [2.1.2 Falschfarbendarstellung](#)

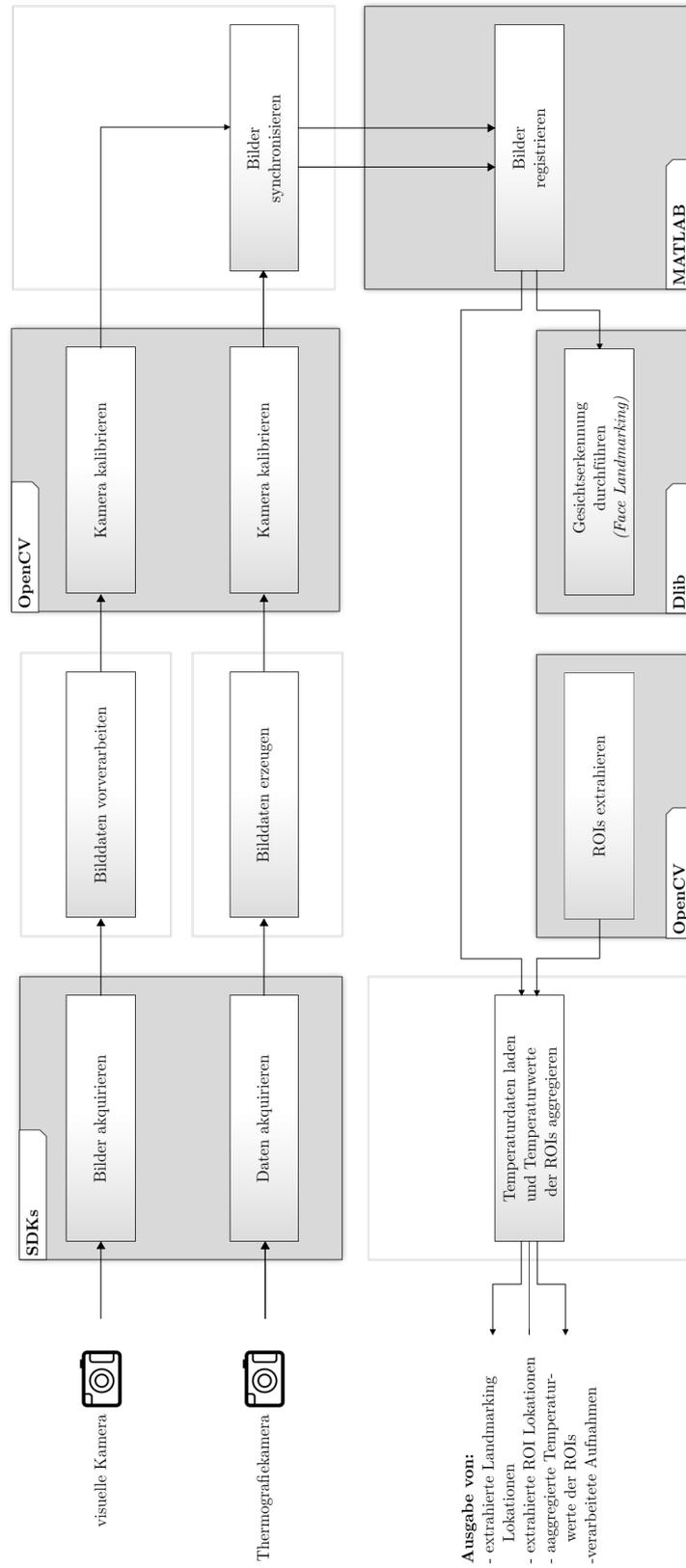


Abbildung 3.5.: Schematischer Ablauf mit den benötigten Teilkomponenten inkl. der verwendeten Frameworks

der visuellen Kamera werden – sofern sie als Video aufgezeichnet wurden – wieder in Einzelbilder zerlegt.

Kamerakalibrierung Die Kamerakalibrierung wird benötigt, um die hardwarebedingte Linsenverzeichnung der verwendeten Kameras zu reduzieren. Hier genügt es, die Linsenverzeichnung einmalig für die verwendete Kamera zu bestimmen und dann auf die Aufnahmen anzuwenden. Das technische Verfahren hiervon ist in Abschnitt [3.4.1 Kamerakalibrierung](#)⁶ des nächsten Kapitels erläutert.

Bildsynchronisierung Für den Fall, dass die aufgezeichneten Daten nicht in derselben Frame-Rate aufgezeichnet wurden und/oder der Start der Aufnahmen nicht zeitgleich erfolgte, ist es nötig, diese für die Bildregistrierung zu synchronisieren. Das Modul Bildsynchronisierung wird im Abschnitt [3.4.2 Synchronisierung und Ausdünnung](#)⁶ behandelt.

Bildregistrierung Die Bildregistrierung bildet das größte und wichtigste Modul dieser Reihe. In diesem Modul werden die Aufnahmen der verschiedenen Kameras miteinander registriert, d. h. eine der beiden Aufnahmen wird so transformiert, dass sie der anderen in Bezug auf die Positionen von Bildinhalten (hier dem Gesicht) nahezu entspricht. Je präziser das Ergebnis dieses Moduls ist, um so präzisere Ergebnisse können die nachfolgenden Module liefern. Diesem Modul widmet sich der Abschnitt [3.4.3 Bildregistrierung](#)⁶.

Gesichtserkennung Dieses Modul liefert die Pixelpositionen der einzelnen Gesichtsmarkierungen. Diese werden benötigt, um die Lage des Gesichtes in der registrierten Aufnahme (bzw. im transformierten visuellen Bild) zu bestimmen und dem nachfolgenden Modul zu übergeben. Das Verfahren wird in Abschnitt [3.3.3 Gesichtserkennung](#)⁶ erläutert.

ROI-Extraktion Basierend auf dem Ergebnis des Gesichtserkennungsmoduls werden arithmetisch die zuvor definierten ROI pixelgenau bestimmt. Dazu bilden verschiedene Bezugspunkte des Gesichtes die Flächen, die die ROI repräsentieren. Erläutert wird dies in Unterabschnitt [3.3.3 Extraktion von Gesichtsbereichen \(ROI\)](#)⁶ anhand von Gesichtsmarkierungen.

Temperaturwertextraktion bzw. -aggregation Um die Temperaturwerte der ROI zu extrahieren, iteriert dieses Modul über alle Pixelpositionen die bei der ROI-Extraktion bestimmt wurden. Ebenso findet hier eine mathematische Aggregation der Temperaturwerte für die einzelnen ROI statt. Das Verfahren wird in Abschnitt [3.3.4 Temperaturextraktion](#)⁶

erläutert. Das Modul stellt gleichzeitig anderen Programmen die Temperaturwerte zur Weiterverarbeitung zur Verfügung.

Die eingangs erwähnte Austauschbarkeit der Module bezieht sich hier vornehmlich auf die Komponenten der Bildakquirierung, die Kamerakalibrierung, die Gesichtserkennung sowie die Definition der ROI.

Technisch findet die Akquirierung der Daten auf einem Computer statt, die Verarbeitung zur Analyse auf einem weiteren. Angestoßen wird der Aufnahmebeginn sowie das Ende vom EMOTIONBIKE-System über eine ACTIVEMQ-Nachricht. Die Hardwareausstattung des Systems dient gleichzeitig auch zur Datenablage. Die Verteilungsansicht ist in Abbildung 3.6 dargestellt. Die Abbildung zeigt dabei bewusst nur jene Komponenten, die für das hier entwickelte System benötigt werden. Sonstige Komponenten des EMOTIONBIKE werden hier nicht weiter betrachtet.

3.3.2. Bilder- und Datenakquirierung

Wie bereits in Abschnitt 3.2 beschrieben, verfügen alle hier verwendeten Kameras über ein SDK, mit dem eine Kommunikation mit den Kameras ermöglicht wird, sowie Daten bzw. Bilder von diesen abgerufen werden können.

INFRA TEC VARIOCAM HD 875

Beim SDK der INFRA TEC VARIOCAM HD 875 handelt es sich um eine „nicht verwaltete WIN32 API“, weshalb ein dynamisches Laden dieser nötig ist, um sie im .NET-Ökosystem zu verwenden. Das dynamische Laden der benötigten Funktion wird mittels des Interface `Irgrabinterface` vorgenommen, das ebenfalls Teil des SDK ist. Mit dessen Hilfe werden die Methoden der Hauptklasse zur Verfügung gestellt. Für die zwei verschiedenen Setups – die in Abschnitt 3.2.4 beschrieben sind – werden zwei verschiedene Implementierungsansätze gewählt, die im Nachfolgenden kurz umrissen werden. Eine detailliertere Dokumentation findet sich dazu in [Kletz und Kleimann (2017)].

Innerhalb des EMOTIONBIKE-Ökosystems wird für die Kommunikation der einzelnen Komponenten der Message Broker ACTIVE MQ⁸ verwendet. Die Implementation dazu findet sich ebenfalls in [Kletz und Kleimann (2017)]. Diese erlaubt es, mit den anderen Komponenten über einen zentralen Controller zu kommunizieren, der u. a. auch Beginn und Ende von Aufnahmesequenzen anstößt. In dieser Konstellation werden die Temperaturrohdaten der Kamera für jeden Frame abgerufen und mittels des SDK in das proprietäre Format der IRB-Files umgewandelt.

⁸<http://activemq.apache.org/>

3. Design & Realisierung

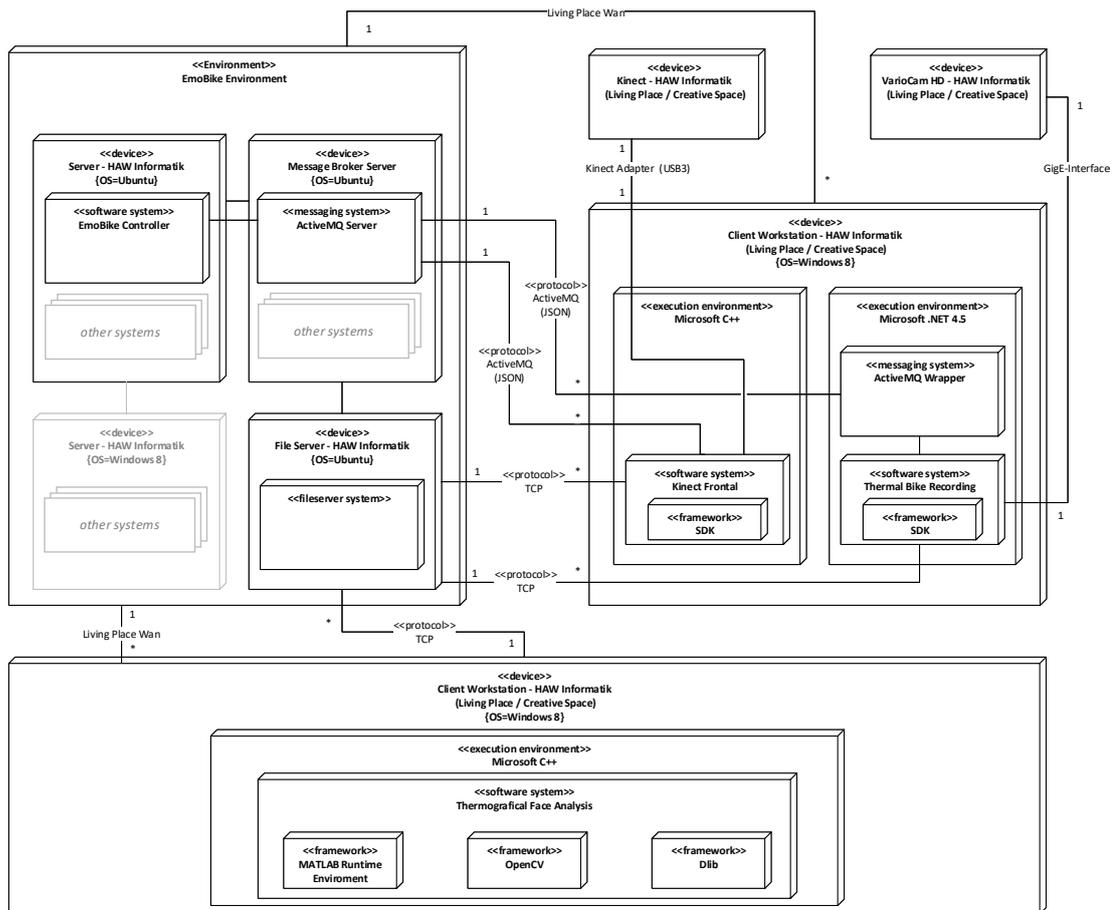


Abbildung 3.6.: Verteilungssicht des Gesamtsystems
(exklusive sonstiger EMOTIONBIKE Komponenten)

Die Umwandlung in das proprietäre Format wurde gewählt, um die Temperaturrohdaten inkl. Metadaten zu persistieren. Alternativ besteht auch die Möglichkeit, die Temperaturrohdaten als ASCII-Datensätze oder als umgewandelte Bilddateien (JPEG, BMP, PNG etc.) auszugeben. Details dazu finden sich in [InfraTec (2014)]. Die IRB-Files können jederzeit wieder in die zuvor genannten Formate konvertiert werden.

Innerhalb des Standalone-Aufbaus ist keine Kommunikation mit externen Komponenten über das Netzwerk nötig, da die Verarbeitung auf einem System durchgeführt wird. Die Temperaturrohdaten werden hier ebenfalls pro Frame von der Kamera abgerufen. Diese Daten werden jedoch nicht persistiert, sondern direkt weiterverarbeitet. Dazu wird pixelweise (jedem Pixel ist ein Temperaturwert zugeordnet) über alle Temperaturen iteriert. Anschließend werden diese in Abhängigkeit von den anderen eingefärbt⁹ und schließlich zu einem Bitmap zusammengesetzt und anderen Komponenten zur Weiterverarbeitung (zusammen mit den jeweiligen Temperaturdaten) bereitgestellt. Eine ausführliche Dokumentation der technischen Realisierung dazu findet sich ebenfalls in [Kletz und Kleimann (2017)].

ALLIED VISION MANTA G-235C

Das VIMBA-SDK der ALLIED VISION MANTA G-235C Kamera bietet die Möglichkeit, die Aufnahmen sowohl synchron als auch asynchron von der Kamera abzurufen. Das Bildformat ist in der Standardkonfiguration vom Typ `System.Drawing.Image`¹⁰ und kann daher in beliebige Formate der Ausgabe umgewandelt werden und so an weitere Komponenten übergeben werden.

MICROSOFT KINECT FOR XBOX ONE

Die Bilderakquirierung mittels der KINECT-Kamera ist bereits eine Komponente des EMOTION-BIKE-Ökosystems, die mittels *Kinect für Windows SDK 2.0* auf die bildgebenden Komponenten der MICROSOFT KINECT FOR XBOX ONE zugreift. Die Komponente erzeugt aus den einzelnen Bildframes der visuellen Kamera ein mp4-codiertes Video mit einer Bildrate von 30Hz. Die Komponente ist in [Müller u. a. (2016)] beschrieben. Da Video ein durchaus übliches Verfahren ist, um Daten visueller Medien abzuspeichern, wird im nächsten Absatz ein simples Verfahren erläutert, mit dessen Hilfe die Einzelbilder wieder aus dem Video extrahiert werden können. Dies ist nötig, um Aufnahmen verschiedener optischer Geräte, die unterschiedliche Bildraten aufzeichnen, zu synchronisieren.¹¹

⁹vgl. Abschnitt 2.1.2 Falschfarbendarstellung

¹⁰<https://msdn.microsoft.com/library/system.drawing.image>

¹¹vgl. Abschnitt 3.4.2 Synchronisierung

Extraktion von Einzelbildern Die Zerlegung wird mittels des Tools FFMPEG¹² durchgeführt. Die Befehlsfolge dazu ist in Auflistung 3.1 dargestellt.

```
ffmpeg -i frontalvideo.mp4 -r 30 -f image2 image-%07d.png
```

Listing 3.1: Anweisung zur Extraktion von 30 Einzelbildern pro Sekunde aus einem Video

Mittels dieser Folge wird das Video, das mit $30Hz$ aufgezeichnet wurde, in je 30 Frames pro Sekunde zerlegt. Als Ausgabeformat wurde das verlustfreie Format PORTABLE NETWORK GRAPHICS (PNG) gewählt.

3.3.3. Gesichtserkennung

Als eindeutiges Ergebnis geht aus der Analyse im Bereich der Gesichtserkennung hervor, dass zur Extraktion bzw. Lokalisation einzelner Gesichtsbereiche ein *face landmarking*-Verfahren nötig ist.

Nach Testung und Recherche verschiedener Frameworks wurde in der jetzigen Ausprägung die Softwarebibliothek DLIB¹³ gewählt. DLIB ist eine plattformübergreifende Open-Source-Softwarebibliothek, die in C++ von Davis King veröffentlicht wurde [King (2009)]. Ausschlaggebend für die Wahl dieser Bibliothek war deren freie Verfügbarkeit. Durch ihre Quelloffenheit ermöglicht sie zusätzlich die Nutzung als sog. *white-box*. Aufgrund der jahrelangen Weiterentwicklung und der Tatsache, dass die Bibliothek im Hardwarenahen C++ geschrieben wurde, ist sie zudem äußerst performant.

Die Softwarebibliothek bietet eine Vielzahl von Funktionen und Methoden, von denen in diesem Kontext lediglich die Komponente des *face landmarking* genutzt und betrachtet wird. Wie in der Analyse bereits erwähnt, ist die Vorbedingung für ein *face landmarking* eine erfolgreiche *face detection* auf der Aufnahme. DLIB arbeitet daher ebenfalls in diesen zwei Schritten.

Zur *face detection* nutzt die Bibliothek den HISTOGRAM OF ORIENTED GRADIENTS (HOG) Algorithmus zusammen mit einem linearen Klassifizierer, einer Bildpyramide sowie einem *Sliding-Window*-Schema. Annotiert wird das Ergebnis der *face detection* mittels vier Koordinaten als Rechteck bzw. Pixelpositionen, der sog. *bounding box*.

Der *face landmarking*-Algorithmus nutzt diese *bounding box*, um aus ihr die einzelnen Gesichtsmerkmale zu extrahieren. Dafür lässt sich der Algorithmus mit verschiedenen Modellen bestücken. Eines dieser Modelle ist das sog. `shape_predictor_68_face`

¹²<http://www.ffmpeg.org/>

¹³<http://dlib.net>

_landmarks . dat, das 68 Punkte des Gesichts extrahiert. Mittels dieser Punkte bzw. Pixelkoordinaten des Bildes ist es möglich, verschiedene Merkmale des Gesichts zu extrahieren, wozu in dem hier genutzten Modell die folgenden zählen: Gesichtsumrandung, Augenbrauen, Augen, Außenkontur der Lippen, Mundöffnung, Nasenbein sowie Nasenvorhof/Dome. Diese 68 Punkte sind in Abbildung A.1 illustriert und finden sich aufgelistet in Anlage A.2. Aus den Punkten lassen sich Polygone zusammensetzen, die die einzelnen Gesichtsmerkmale repräsentieren. Exemplarisch ist dies in Abbildung 3.7 für ein Bild dargestellt.



Abbildung 3.7.: Exemplarische Darstellung des *face landmarking* erzeugt mittels DLIB

Das hier verwendete Modell wurde mittels Maschinenlernalgorithmus auf dem iBUG 300W DATASET¹⁴ trainiert. Dieses beinhaltet je 300 Aufnahmen von Gesichtern im Innenraum und Außenbereich. Dabei handelt es sich nicht um biometrische Fotos von Gesichtern, sondern um reale Alltagssituationen. Diese Datengrundlage eignet sich daher besonders für das hier verwendete Szenario, bei dem eine gewisse Kopfbewegung vorkommen kann¹⁵ und dennoch bestenfalls keine *false positives* oder *false negatives* auftreten sollen.

Aus Kombinationen der detektierten Gesichtsmerkmale lassen sich arithmetisch weitere Gesichtsbereiche bzw. ROI – wie jene die in der Analyse des Gesichtsatlas¹⁶ festgelegt wurden – berechnen.

¹⁴<https://ibug.doc.ic.ac.uk/resources/300-w/>

¹⁵vgl. Abschnitt 2.2.2 Kopfpose

¹⁶vgl. Abschnitt 2.3 Gesichtsatlas

Extraktion von Gesichtsbereichen (ROI) anhand von Gesichtsmerkmalen

Aus der Analyse im Kapitel ‚Gesichtsatlas‘ ergibt sich, dass für die Temperaturmessung fünf ROI (bzw. sieben bei Unterscheidung nach Gesichtshälfte) im Fokus stehen. Durch das zuvor beschriebene *face landmarking* mit der DLIB Bibliothek lassen sich aus diesen Landmarking-Punkten bzw. deren Pixelpositionen die ROI berechnen und extrahieren. Dazu werden ausgewählte Landmarking-Punkte, zu neuen geometrischen Formen verbunden, welche die ROI repräsentieren. Die Polygone entsprechen dabei den Bereichen des Gesichtes welche im Analyseabschnitt 2.3 evaluiert wurden. Abbildung 3.8 illustriert die im Nachfolgenden beschriebene Lokalisierung der ROI, basierend auf den Bezugspunkten des *face landmarking*-Ergebnisses.

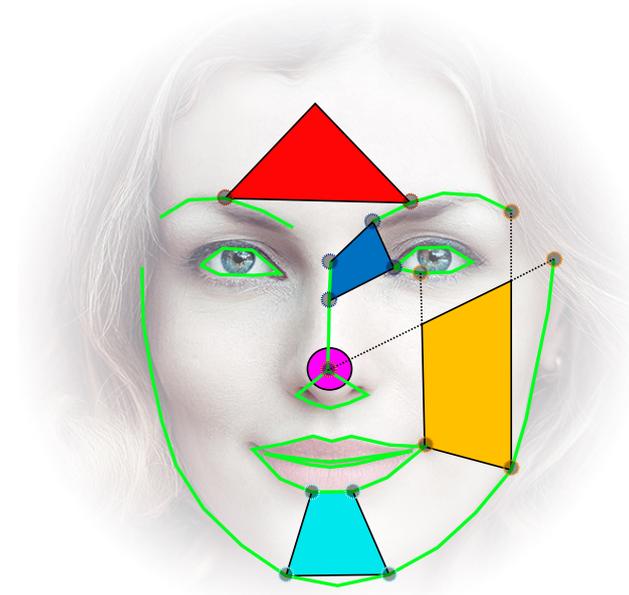


Abbildung 3.8.: Darstellung der berechneten ROI (teilweise nur rechte Gesichtshälfte) basierend auf den mittels DLIB erzeugten *face landmarking*-Punkten eines Gesichtes

rot = Regio frontalis (Stirn); **blau** = Regio depressor supercilii (Bereich zwischen den Augen) [*nur rechts*];

lila = Regio nasalis (Nasenspitze); **orange** = Regio buccalis (Wangenregion) [*nur rechts*]; **cyan** = Regio

mentalis (Kinnregion) Hintergrundgrafik: ©©

Nachfolgend werden die ROI aufgelistet und beschrieben aus welchen Landmarking-Punkten diese zusammengesetzt werden. Die verwendeten Nummerierungen beziehen sich auf die Bezugspunkte des *face landmarking*-Algorithmus, die in Anlage A.2 aufgelistet sind.

Regio frontalis (Stirn) Die ROI der Stirn wird als gleichschenkliges-rechtwinkliges Dreieck extrahiert. Dazu bilden die Punkte 19 und 24 die Hypotenuse des Dreiecks. Die Höhe (h) des Dreiecks ergibt sich aus den Winkeln α , β und γ , wobei gilt: $\alpha = \beta = 45^\circ$ und $\gamma = 90^\circ$.

Regio depressor supercilii (Bereich zwischen den Augen) Die ROIs zwischen den Augen lassen sich für die linke und rechte Gesichtshälfte unterscheiden. In der linken Hälfte bilden die Punkte 21, 27, 28, 39 und in der rechten Hälfte die Punkte 22, 27, 28, 42 jeweils ein Viereck. Die beiden ROIs werden als zwei getrennte Bereiche betrachtet.

Regio nasalis (Nasenspitze) Die Nasenspitze bildet die am einfachsten zu berechnende ROI. Der Punkt 30 bildet den Mittelpunkt eines Kreises. Der Radius (r) des Kreises ist gleich der Länge der Distanz zwischen den Punkten 29 und 30 und deckt dadurch die komplette Nasenspitze ab.

Regio buccalis (Wangenregion) Für die Wangenregion ist eine etwas komplexere Bestimmung der ROI nötig. Dazu bildet (für die rechte Gesichtshälfte) der Schnittpunkt der Geraden aus den Punkten 47 + 54 sowie 30 + 16 den Punkt (\mathcal{A}). Der Schnittpunkt der Geraden 16 + 30 sowie 26 + 12 stellt den Punkt (\mathcal{B}) dar. Die Punkte \mathcal{A} , \mathcal{B} , 12 und 54 bilden schließlich ein Viereck, das die ROI definiert. Analog dazu bildet sich die Region für die linke Gesichtshälfte aus den Schnittpunkten der Geraden 40 + 4 sowie 0 + 30 als (\mathcal{A}') und die Geraden 4 + 17 sowie 0 + 30 als (\mathcal{B}'). Die Punkte \mathcal{A}' , \mathcal{B}' , 4 und 48 bilden daraus das Viereck der linken Gesichtshälfte.

Regio mentalis (Kinnregion) Die ROI des Kinns bildet aus den Punkten 7, 9, 56 und 58 ein Viereck bzw. ein Trapez.

3.3.4. Temperaturextraktion

Die durch die INFRA TEC VARIOCAM HD 875 aufgenommenen Rohdaten können in verschiedenen Datenformaten ausgegeben bzw. in diese konvertiert werden. Neben der grafischen Ausgabe als Bilddateien lassen sich die Temperaturwerte auch pixelgenau als ASCII-Datei ausgeben. Diese Datei besteht neben einem Header aus den Temperaturwerten als 1024×768 -Matrix. Der genaue Aufbau der Datei inkl. der Headerinformationen ist in Anlage A.5 aufgeführt. Für die pixelgenaue Bestimmung der Temperatur kann somit mittels (x, y) -Tupel auf den konkreten Wert zugegriffen werden.

Da die ROI nicht präzise einer genauen Pixelposition zugeordnet sind, sondern sich aus einer Fläche mehrerer Pixelpositionen zusammensetzen, ist eine Aggregation der Temperaturwerte

des Bereichs nötig. Für Temperaturwerte bieten sich die Aggregationsfunktionen MAXIMUM, MINIMUM, MITTELWERT sowie MEDIAN besonders an. Mit der MITTELWERT-Funktion bietet sich die Möglichkeit, eine Gleichverteilung über die gesamte ROI-Temperatur zu erlangen. Sollte es jedoch bei der Bestimmung der ROI zu Fehlern bzw. Abweichungen der Positionierung kommen, ist es möglich, dass weitere Teile einbezogen werden. Im Extremfall kann dies dazu führen, dass Bereiche außerhalb des Gesichts einbezogen werden, was entweder zu einem starken Abfall der mittleren Temperatur (wenn hinter/neben dem Kopf eine kältere Temperatur vorherrscht) oder zu einer Steigerung dieser (wenn eine sehr warme Wärmequelle hinter/neben dem Kopf besteht) führt. Ebenfalls wird ein Abfall der Temperatur durch Gesichtsbehaarung (Bart, Augenbrauen sowie Kopfhare) verursacht, sobald diese in die ROI reichen.

Die Nutzung der MEDIAN-Funktion erweist sich in solchen Fällen für geeigneter, da sie Ausreißer von Temperaturen nicht in das Ergebnis einbezieht. So können bei kleineren Abweichungen der Position der ROI bzw. im Fall, dass Haare mit in die ROI gelangen, Temperaturwerte deutlich robuster aggregiert werden.

Aus den bereits genannten Gründen scheidet die MINIMUM-Funktion für eine plausible und repräsentative Temperaturermittlung der ROI aus.

Die MAXIMUM-Funktion hingegen ist äußerst interessant für die Temperaturermittlung einer gesamten ROI. Hierdurch lassen sich u. a. die Temperaturwerte der Arterien-/Venenbereiche extrahieren, ohne die genaue Lage zu kennen. Da der Blutdurchfluss der Blutgefäße maßgeblich für die Temperatur des umgebenden Gewebes verantwortlich ist ([Spyra (2010)]), ist es durch die Maximum-Funktion möglich, konstant (z. B. über die Zeitdimension) den wärmsten Punkt bzw. die wärmste Temperatur der ROI zu ermitteln, was wiederum die mutmaßliche Lage der Arterien und Venen ist.

Technisch sind – unabhängig von der Aggregationsfunktion – alle Temperaturwerte auf den Pixelkoordinaten zu extrahieren, die innerhalb der definierten ROI liegen. Da es sich bei diesen nicht um Rechtecke handelt, ist ein einfaches Iterieren mittels einer verschachtelten Schleife nicht möglich. Vielmehr ist für die definierten Bereiche das sog. „*point in polygon problem*“ zu lösen, das in zweiwertiger Logik bestimmt, ob eine gegebene Pixelposition innerhalb des Polygons (respektive ROI) liegt oder nicht. Somit ist es möglich, über einen größeren Bildbereich zu iterieren, der einem über dem Polygon (für die hier definierten ROIs Dreieck, Viereck und Kreis) liegendem Rechteck entspricht. Die Ausmaße der Rechtecke ergeben sich bei den Drei- und Vierecken aus den Extremen der jeweiligen Eckpunkte, beim Kreis aus dem Mittelpunkt zuzüglich des Radius der X- und Y-Achse. Das Verfahren zur Überlagerung mit einem Rechteck wird auch als „*minimal umgebendes Rechteck*“ (MUR) bezeichnet. Für die hier genutzten ROI ist dies in Abbildung 3.9 dargestellt. Bei der Iteration über die jeweiligen Pixelpositionen

ist schließlich jeweils zu prüfen, ob diese innerhalb des Polygons liegen, um dann mit dem extrahierten Temperaturwert die Aggregation vorzunehmen.

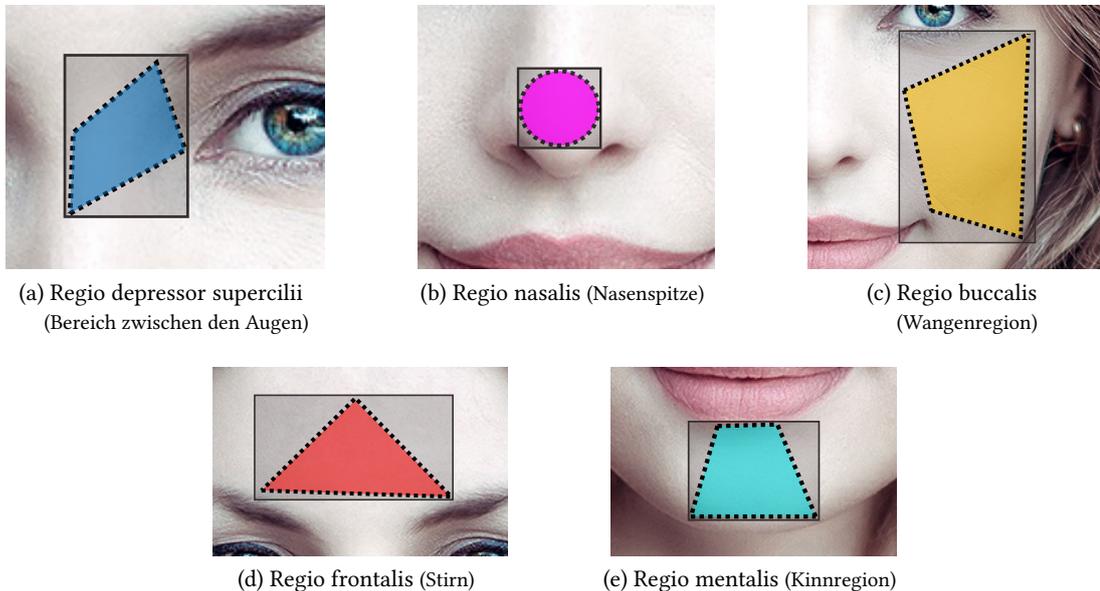


Abbildung 3.9.: Rechtecküberlagerung (grau) der definierten ROI (farbig) zur Iteration. Mittels Iteration auf den Rechtecken lassen sich anhand des RAY-CASTING-ALGORITHMUS die Pixelpositionen innerhalb der ROI bestimmen. Daraus lassen sich in der Folge die Temperaturwerte extrahieren. Hintergrundgrafik: ©©

Für die Lösung des *point in polygon problem* wird hier eine Implementierung des RAY-CASTING-ALGORITHMUS genutzt [Shimrat (1962)], bei dem für den zu überprüfenden Punkt ein Strahl in horizontaler (alternativ vertikaler) Richtung ausgesandt wird und die Anzahl der Schnitte mit den Polygonkanten gezählt wird. Diese Anzahl (gerade/ungerade) gibt Aufschluss über die Lage des Punktes. Ein Pseudocode des Algorithmus ist in Auflistung 3.2 dargestellt.

```
1 foreach Kante in Polygon:
2     if schnitt_mit_kante(Punkt,Kante) then
3         schnitte++
4
5 if ist_gerade(schnitte) then
6     return true //Punkt liegt innerhalb des Polygons
7 else
8     return false //Punkt liegt außerhalb des Polygons
```

Listing 3.2: Pseudocode des RAY-CASTING ALGORITHMUS

3.3.5. Datenstruktur und Datenausgabe

Um die strukturierten Daten anderen Anwendungen zur Verfügung zu stellen, wird das kompakte Datenformat `JAVASCRIPT OBJECT NOTATION (JSON)`¹⁷ zur Ausgabe verwendet.

Drittanwendungen werden darüber Information über Metadaten und die ROI inkl. Temperaturwerte übergeben. Eine vollständige Darstellung hiervon findet sich in Anlage A.3. Konkret beinhalten diese Daten als globale Metadaten Pixelangaben über die Breite und Höhe der Aufnahme. Die ROI werden als Liste übergeben, die neben der Bezeichnung die aggregierten Temperaturwerte in Maximal-, Minimal-, Durchschnitts- sowie Mediantemperatur beinhaltet. Ebenso werden für jedes ROI Metadaten ausgegeben. So werden die Punktpositionen (in X-/Y-Angaben) der Polygone ausgegeben, aus denen die ROI bestehen. Dies ermöglicht den Anwendungen, sofern dies erforderlich ist, diese zu visualisieren oder auch weitere Auswertungen mit diesen durchzuführen. Des Weiteren wird die Anzahl der Pixel zurückgeben die für diese ROI berücksichtigt wurden. Abschließend wird noch die Temperatureinheit ausgegeben um die Werte in einen Kontext zu setzen.

In der aktuellen Ausprägung werden die JSON-Datenstrukturen persistiert. Eine Übertragung, beispielsweise über einen Message-Broker, ist aber ebenso denkbar.

Darüber hinaus werden verschiedene Arten von Bildern zur Verfügung gestellt, die weiterverarbeitet werden können. Diese sind in Abbildung 3.10 dargestellt. Hierzu zählt jeweils das Ursprungsbild als visuelles (3.10a) sowie als Thermografiebild (3.10b), das transformierte visuelle Bild (3.10c), eine Überlagerung des transformierten visuellen Bildes mit dem Thermografiebild (3.10d), selbiges mit Überlagerung der eingezeichneten ROI (3.10e) und das transformierte visuelle Bild überlagert mit den ROI (3.10f). Auch diese Daten werden persistiert ausgegeben.



Abbildung 3.10.: Beispielbilder der grafischen Ausgaben

¹⁷<http://www.json.org>

3.4. Softwarefusion

Das folgende Kapitel beschreibt die eigentliche Softwarefusion, die den elementaren Kern dieser Arbeit ausmacht. Sie hat das Ziel, die vorliegenden Aufnahmen des visuellen Spektrums mit jenen des Thermografiespektrums zu fusionieren. Dadurch wird erreicht, dass die visuellen Bilder mit den Temperaturdaten angereichert werden, was wiederum ermöglicht, in Bezug auf die visuellen Bilder – unter Zuhilfenahme der bereits erläuterten Gesichtserkennung – computergestützt die Temperaturwerte für die festgelegten Bereiche abzufragen.

Dazu wird auf die genutzten und entwickelten Hard- und Softwarekomponenten detaillierter eingegangen. Das Kapitel ist dabei in derselben Chronologie gegliedert wie diese auch die spätere Anwendung nutzt. Begonnen wird mit dem Baustein der Kamerakalibrierung, indem neben dem eigens hierfür entwickelten Kalibrierungsmuster auch auf den Kalibrierprozess sowie dessen Ergebnisse eingegangen wird. Da die Bildakquirierung mit verschiedenen Bildwiederholungsraten erfolgt, müssen die Einzelbilder der unterschiedlichen Modalitäten zueinander synchronisiert werden. Dabei muss die Menge der Aufnahmen ausgedünnt werden. Das hier genutzte Verfahren wird daher ebenso betrachtet. Abgeschlossen wird das Kapitel mit der eigentlichen Bildregistrierung, für die die beiden zuvor genannten Bausteine als Vorbedingung benötigt werden. Im folgenden Kapitel wird auf die Verarbeitung der Aufnahmen eingegangen und die technische Implementation dazu erläutert.

3.4.1. Kamerakalibrierung

Um die in Kapitel 2.4.1 beschriebene Linsenverzeichnung der optischen Geräte und dessen Linsen zu entfernen, wird ein Kalibrierungsmuster benötigt. Eine Möglichkeit besteht dazu in der Verwendung eines zweidimensionalen Schachbrettmusters. Um dieses auch zur Kalibrierung der optischen Geräte verschiedener Modalitäten zu nutzen, kann kein einfaches gedrucktes Schachbrettmuster verwendet werden, da es mittels der Thermografiekamera als eine homogene Fläche dargestellt wird – wie in Abbildung 3.11e (und 3.11f) dargestellt. Aus diesem Grund muss das Schachbrett den multimodalen Gegebenheiten angepasst werden, auf die im nachfolgenden Abschnitt näher eingegangen wird.

Schachbrett

Die Grundidee, um das abgewandelte Schachbrettmuster – sowohl im sichtbaren Spektrum als auch im Thermografiebereich – zur Kamerakalibrierung nutzen zu können, beruht auf dem Konzept, dass Teile dieses Musters erwärmt werden. Durch die Erwärmung bleibt das Muster im visuellen Spektrum unverändert und kann dadurch weiterhin genutzt werden. Anders verhält

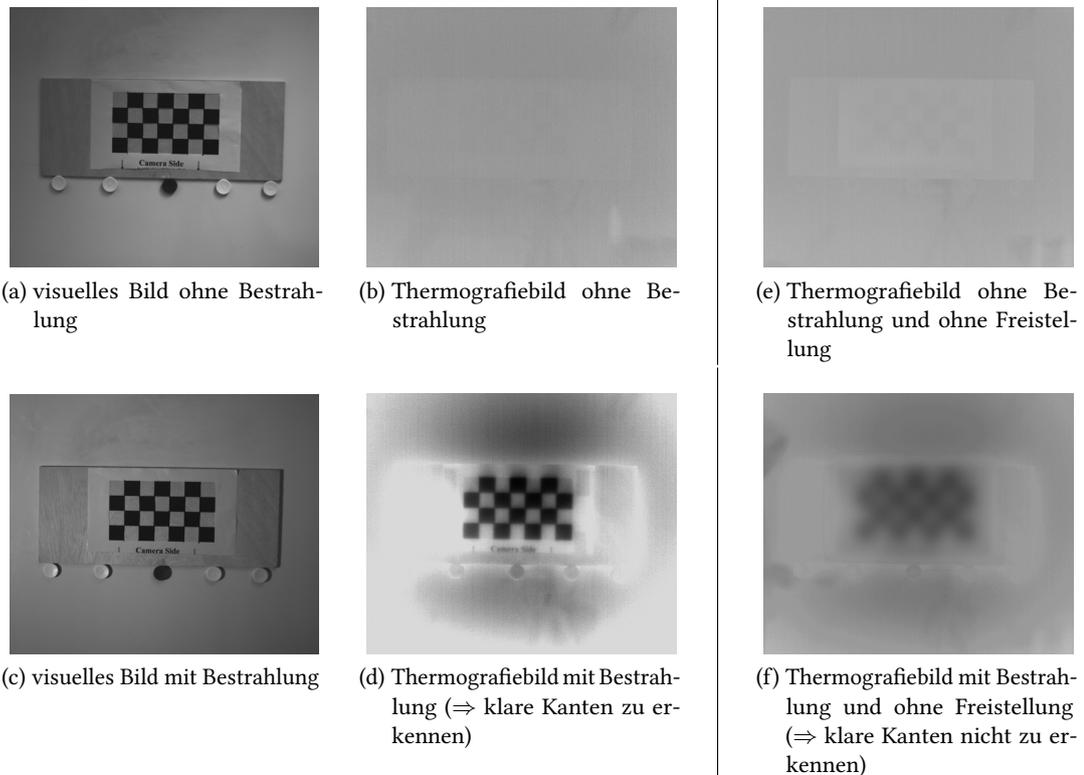


Abbildung 3.11.: Schachbrettmuster im visuellen- (a)+(c) und im Thermografiebild (b)+(c) vor und während der Bestrahlung mit einem Halogenbaustrahler. Bilder (e)+(f) zeigen zum Vergleich ein Schachbrettmuster als Thermografiebild, das nicht freigestellt wurde, ebenfalls vor und während der Bestrahlung.

sich das Muster im Thermografiespektrum. Hier werden die erwärmten Bereiche sichtbar bzw. hervorgehoben.

Die Herausforderung liegt darin, durch die Erwärmung die Integrität des Schachbrettmusters im visuellen Spektrum zu erhalten bzw. in das Thermografiespektrum zu überführen und sichtbar zu machen. Hierzu gibt es bereits verschiedene erprobte Verfahren, wie in [Ursine u. a. (2012)] beschrieben. Hier wird vorgeschlagen, entweder die weißen oder die schwarzen Flächen des Schachbretts mit Kupferplatten zu bestücken. Da Kupfer gut Wärme leitet und für einen gewissen Zeitraum speichert, lassen sich die jeweiligen Kupferplatten unter Laborbedingungen mittels Heizstrahler oder durch Bestrahlung mit beispielsweise einem Halogenstrahler erwärmen. Außerhalb einer Laborumgebung genügen im Sommer bereits Sonnenstrahlen, um ein Erwärmen der Plättchen zu ermöglichen, oder im Winter kühler Wind, um die Plättchen abzukühlen. Ein anderes Verfahren findet sich in [Ellmauthaler u. a. (2013)]. Für dieses werden

Miniatur-Glühbirnen auf dem Brett angeordnet. Diese können hierbei sowohl im aus- als auch im eingeschalteten Zustand im visuellen Spektrum detektiert werden. Sobald sie mit Strom versorgt werden, erzeugen sie neben der Leuchtenergie auch Wärmenergie. Die entstehende Wärme wird im Thermografiespektrum sichtbar und ermöglicht mit diesen Bezugspunkten die Kalibrierung der Kamera.

Schachbrettmuster	
Feldgröße	$30 \times 30mm$
Spalten	7
Zeilen	4
Zu detektierende Punkte	18

Tabelle 3.4.: Eigenschaften des verwendeten Kalibrierungsmusters

Bereits in [Kletz und Kleimann (2017)] wurde ein Schachbrett nach dem Verfahren aus [Saponaro u. a. (2015)] entwickelt. Bei diesem sind die weißen Felder ausgeschnitten und das modifizierte Muster auf eine helle Holzplatte aufgeklebt. Das auf diese Weise präparierte Kalibrierungsmuster wird mittels eines Halogenstrahlers erwärmt. Aufgrund der Beschaffenheit des Musters erwärmen sich die schwarzen (noch vorhandenen) Flächen des Schachbretts dabei stärker als die dahinterliegende Holzplatte. Hierdurch lassen sich nach der Bestrahlung die einzelnen Felder auch im Thermografiespektrum deutlich voneinander separieren, wodurch eine Kalibrierung ermöglicht wird. Das entwickelte Schachbrettmuster sowie das Verhalten eines regulären/nicht modifizierten Schachbrettmusters sind in Abbildung 3.11 dargestellt. Die Parameter des Schachbrettmusters und dessen Eigenschaften sind in Tabelle 3.4 aufgeführt.

Kalibrierungsprozess

Der Kalibrierungsprozess der Kamera muss einmalig mit jedem verwendeten optischen Gerät durchgeführt werden. Um beide hier verwendeten Kameras unter gleichen Bedingungen kalibrieren zu können, wurde ein Tool entworfen, mit dem mit beiden Kameras zeitgleich Aufnahmen des Schachbrettmusters aufgezeichnet werden. Hierdurch entsteht zusätzlich der Vorteil, dass die Bilder nachbearbeitet werden können, falls beispielsweise der Kontrast zwischen den hellen und dunklen Flächen nicht ausreicht, oder fehlerhafte Aufnahmen – beispielsweise nicht im Bildbereich beider Kameras – aussortiert werden können. Die Aufnahmen

werden im verlustfreien PNG-Format gespeichert, um möglichst keine Bildinformationen zu verlieren. Für eine gute Kalibrierung werden 10 – 20 Aufnahmen empfohlen [MathWorks (2017a)].

Eine Kalibrierung kann durch verschiedene Frameworks unterstützt werden. Ein häufig genutztes Framework zur Kalibrierung ist die COMPUTER VISION SYSTEM TOOLBOX¹⁸ als Erweiterung der Software MATLAB des Herstellers THE MATHWORKS¹⁹. Ein ähnlich verbreitetes Framework ist OPENCV²⁰, das ebenso die Möglichkeit der Kamerakalibrierung anbietet. Da die Ergebnisse der Kalibrierung in beiden Frameworks aus- und eingelesen werden können, besteht eine Interoperabilität.

Nachfolgend wird der Kalibrierungsvorgang mit der COMPUTER VISION SYSTEM TOOLBOX beschrieben, da diese viele Auswertungsmöglichkeiten bietet. Für die Kalibrierung wurden je 22 Aufnahmen ausgewählt, die möglichst gleichverteilt über den gesamten Bildbereich beider Kameras aufgezeichnet wurden. Das Schachbrettmuster wurde während der Aufzeichnung mehrfach um die eigene Achse gedreht sowie perspektivisch geneigt, um robustere Ergebnisse zu liefern. Die Verteilung der Schachbrettmuster während der Aufnahmen und die relativen Positionen der Kameras zueinander lassen sich Abbildung 3.12 entnehmen. Fokussierung oder Zoomeinstellungen sind während des gesamten Aufzeichnungsvorgangs unverändert.

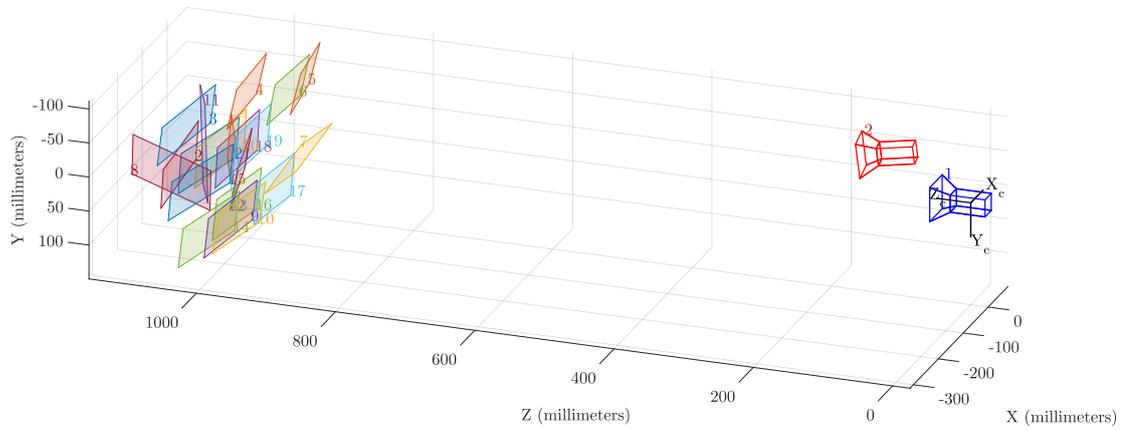
Mittels dieses Verfahrens der Kamerakalibrierung werden die intrinsischen und extrinsischen Kameraparameter ermittelt. Mit diesen ist es möglich, die potenziell vorhandene Linsenverzeichnung auszugleichen. Ein Ausgleichen ist nötig, da eine Bildregistrierung erschwert wird, falls beide Kameras eine unterschiedliche Linsenverzeichnung aufweisen. Die tatsächlich ermittelten Kameraparameter für die verschiedenen Kameras finden sich in Anlage A.1.

¹⁸<https://mathworks.com/products/computer-vision>

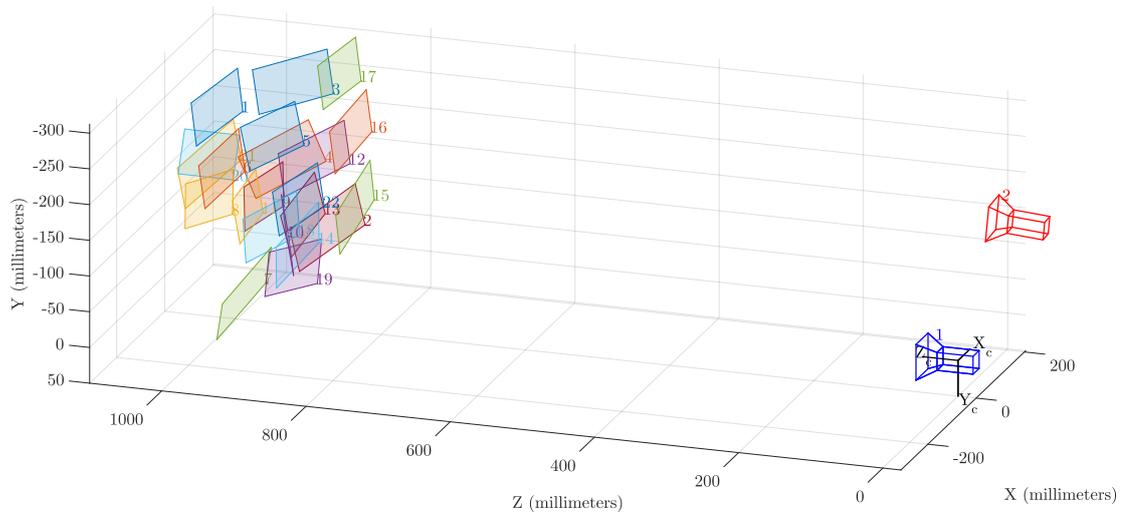
¹⁹<https://mathworks.com/>

²⁰<http://opencv.org/>

3. Design & Realisierung



(a) VARIOCAM (blau) und MANTA (rot)



(b) VARIOCAM (rot) und KINECT (blau)

Abbildung 3.12.: Visualisierung der extrinsischen Kameraparameter für die beiden verwendeten Szenarien

3.4.2. Synchronisierung und Ausdünnung

Um Aufnahmen miteinander zu registrieren ist es nötig, dass diese zum annähernd selben Zeitpunkt aufgezeichnet werden. Die Toleranz d. h. wie groß der zeitliche Abstand zwischen den Aufnahmen sein darf, ergibt sich aus dem jeweiligen Szenario. Handelt es sich um Aufnahmen einer statischen Szene kann zwischen den Aufnahmen ein größerer zeitlicher Abstand erlaubt sein, als bei einer Szene die stark in Bewegung ist. Wird dies nicht beachtet – und Aufnahmen verwendet die einen zu großen zeitlichen Versatz haben – kann dies die Registrierung zum scheitern bringen, oder aber zu Bewegungsartefakten als sog. Geisterbilder („Ghosting“) kommen.

Optische Geräte – ganz gleich welcher Modalität – können prinzipiell in verschiedenen Bildraten aufzeichnen. Diese Bildraten müssen einander angeglichen werden um den zeitlichen Versatz bzw. Bewegungsartefakte zu verhindern. Dazu ist zu beachten, dass die maximale Verarbeitungsrate des Systems nur maximal genauso hoch sein kann wie beim optischen Gerät mit der niedrigsten Bildrate. Zeichnet ein optisches Gerät beispielsweise mit einer Bildrate von 5Hz auf, ein zweites in einer Bildrate von 30Hz , so kann das System, das diese beiden Bilddaten verarbeiten soll, mit einer Bildrate von maximal 5Hz arbeiten. In diesem Fall müssen von dem optischen Gerät mit einer höheren Bildrate die Bilder aussortiert werden, für die es keine Entsprechung des anderen optischen Geräts gibt. Bei der Aussortierung ist neben der Anzahl der Einzelbilder ebenso auf einen zeitgleichen Beginn der Bildsequenzen zu achten, um nicht fälschlicherweise mit einem sog. Offset eines der optischen Geräte zu arbeiten. Illustriert wird dies in Abbildung 3.13. Dieser Schritt ist eine Grundvoraussetzung, damit eine robuste Bildregistrierung²¹ zwischen zwei (oder mehr) Bildern gelingen kann.

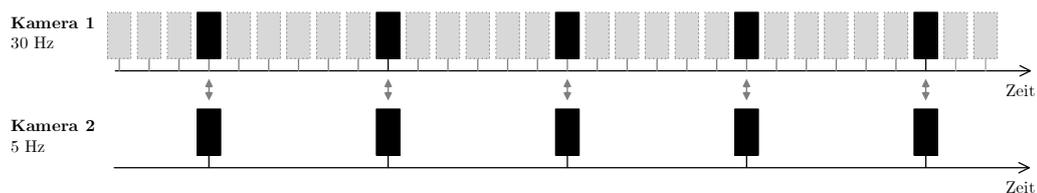


Abbildung 3.13.: Schematische Darstellung der Bildsynchronisierung: Grau dargestellte Frames der „Kamera 1“ können verworfen werden

²¹vgl. Abschnitt 2.4.2 Bildregistrierung

Wie bereits in Kapitel 3.2 gezeigt, verfügen die hier verwendeten Kameras teilweise über verschiedene maximale Bildfrequenzen. Diese gilt es nach dem oben beschriebenen Verfahren anzugleichen. Da auch eine Auswertung im Pre-Processing ermöglicht werden soll, besteht zudem die Möglichkeit, zuvor (in verschiedener Bildrate) abgespeicherte Bilder zu laden, auf die dasselbe Verfahren anwendbar ist.

Um eine genaue zeitliche Zuordnung der Einzelbilder zu erlangen, bieten sich verschiedene Verfahren an. Werden die Bilder in einem verteilten System akquiriert, bietet sich die Nutzung eines Zeitstempels mit einer synchronisierten Uhr an. Hierdurch lassen sich auch im Nachhinein die Zuordnungen der Einzelbilder einfach mittels Zeitstempel gestalten. Wird ein Standalone-System betrieben, an das die verschiedenen optischen Systeme angeschlossen sind, bietet es sich an, die Bilder innerhalb eines Threads zu akquirieren, um den möglichen Versatz auf ein Minimum zu reduzieren [Hennes u. a. (2015)].

Synchronisation der INFRATEC VARIOCAM HD 875 mit der MICROSOFT KINECT FOR XBOX ONE

Die visuelle Kamera des EMOTIONBIKE-Szenarios zeichnet den Videostream konstant mit $30Hz$ auf. Die Thermografiekamera mit $16Hz$ Einzelbilder. Die Startzeitpunkte der Aufzeichnungen finden zeitstempelbasiert synchronisiert auf einer Maschine statt, weshalb ein eindeutiger Startzeitpunkt definiert ist, von dem nachfolgende Zeitpunkte berechnet werden können.

Durch die verschiedenen Frameraten ergibt sich, dass die aufgezeichneten Daten beider Kameras auf $2Hz$ ausgedünnt werden müssen. Für die Daten im visuellen Spektrum bedeutet dies, dass 14 von 15 Frames für eine Registrierung verworfen werden müssen; bei der Thermografiekamera sind hiervon sieben von acht Frames betroffen. Zu beachten ist jedoch, dass die verworfenen Frames trotzdem angezeigt werden können, allerdings die präzise Registrierung nicht möglich ist. Auch stellt dieses Schema den höchstmöglichen Wert dar, der weiter ausgedünnt werden kann.

In der Praxis zeigt sich allerdings, dass es bei der visuellen Kamera durchaus zu Aussetzern bei der Bildakquirierung kommt. Durch die Zeitstempel kann jedoch eine Synchronisierung von mindestens einer Sekunde garantiert werden.

3.4.3. Bildregistrierung

Nach den Schritten der Kamerakalibrierung und der Synchronisierung der Bilder ist es möglich, die Bilder miteinander zu registrieren.

Als Teilergebnis der Analyse geht hervor, dass sowohl die *Silhouettenextraktion* als auch die *Kantenextraktion* geeignet erscheinen, um zwei multimodale Bilder miteinander zu registrieren. Dieses Teilergebnis wird hier aufgegriffen und in den nachfolgenden zwei Verfahren dokumentiert, auf die diese Eigenschaften zutreffen.

Durch die Erzeugung der Silhouetten oder durch Extraktion von Kanten ist es möglich, die Bildinformationen der multimodalen Aufnahmen in eine unimodale Darstellung zu überführen und somit die Anzahl der Variablen zu minimieren. Die Registrierung der Bilder kann dann auf den unimodalen Repräsentationen der eigentlichen Aufnahmen erfolgen. Die in diesem Zuge ermittelte Transformationsmatrix lässt sich unverändert auf die multimodalen Aufnahmen anwenden und somit diese Bilder registrieren.

Grundsätzlich muss für die Registrierung entweder das eine Bild so geometrisch transformiert werden, sodass es dem anderen entspricht, oder es müssen beide Aufnahmen so transformiert werden, dass sie einem neuen, dritten Bild entsprechen. Für beide Verfahren wurde die Designentscheidung getroffen, das visuelle Bild zu transformieren, um dem Thermografiebild zu entsprechen. Grund hierfür ist, dass hierdurch das Thermografiebild und die dazugehörigen Temperaturdaten unverändert genutzt werden können, was besonders die spätere Temperaturbestimmung enorm vereinfacht. Die auf dem visuellen Bild stattfindende Gesichtserkennung ist i. d. R. auch nach einer durch die Registrierung bedingten Transformation weiterhin möglich.

Bildregistrierung anhand von Silhouetten

Das erste hier vorgestellte Verfahren führt eine Bildregistrierung anhand von Silhouetten der eigentlichen Aufnahmen durch. Dazu gliedert sich dieser Abschnitt in die Silhouettenerzeugung sowie die Silhouettenregistrierung.

Silhouettenerzeugung Zur Silhouettenerzeugung eignen sich eine ganze Reihe von Verfahren, die alle das Ziel haben, den Vordergrund (hier die Person) vom Hintergrund zu separieren. Bei einfachen Szenen genügt es teilweise bereits, eine Farbe oder einen Farbbereich zu entfernen, um das gewünschte Objekt freizustellen bzw. den Hintergrund zu entfernen. Wesentlich häufiger werden Verfahren genutzt, die sich Bewegungen des Objektes zunutze machen. Diese

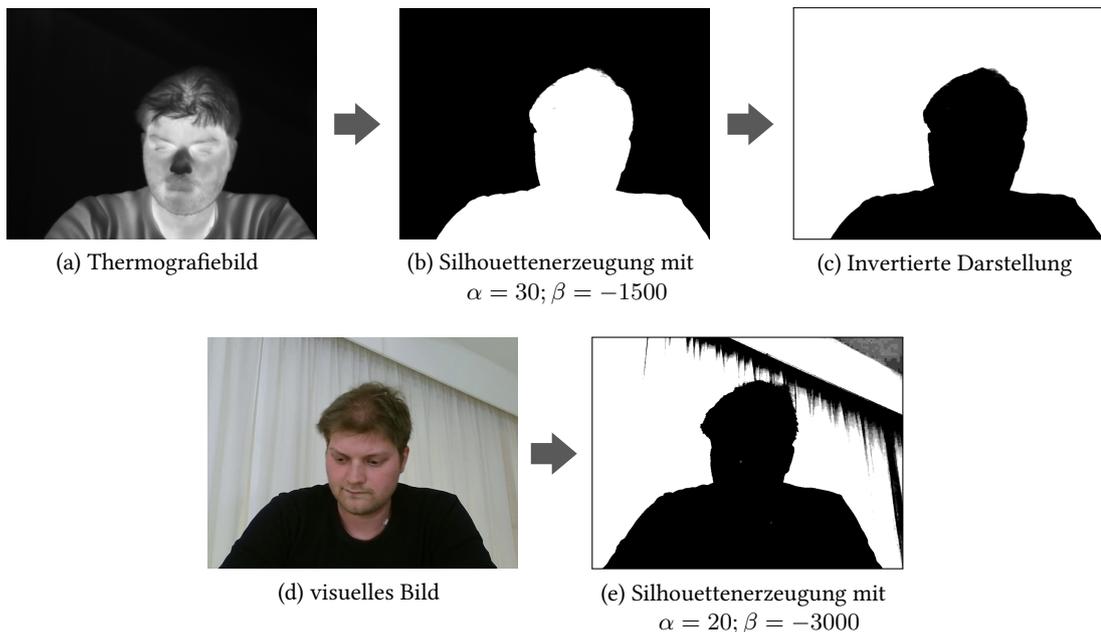


Abbildung 3.14.: Silhouettenerzeugung mittels Kontrast- und Helligkeitsmanipulation: (a)+(d) zeigen die Aufnahmen, auf denen die Silhouetten erzeugt werden sollen. Das Ergebnis ist in den Abbildungen (b)+(e) mit den dazugehörigen Parametern dargestellt. Abbildung (b) muss noch invertiert werden, um der Farbcodierung in (e) zu entsprechen. Dies ist in Abbildung (c) dargestellt.

nutzen dazu mehrere aufeinanderfolgende Frames (beispielsweise aus einem Video) und werfen die starren Elemente der Szene. Die bewegend Objekte stellen somit den Vordergrund der Szene dar, der gesucht wird. Dasselbe Verfahren nutzen auch Bewegungserkennungsalgorithmen, um festzustellen, ob neue Objekte in das Kamerasichtfeld eintreten. Ebenso ist es möglich, Silhouetten mittels Kantendetektionsalgorithmen zu extrahieren²² oder aber die Entfernungsinformationen einer Tiefenkamera zu verwenden. In diesem Kontext wird jedoch ein wesentlich unkomplizierteres Verfahren gewählt. Dazu werden mittels Kontrast- und Helligkeitsmanipulation die Bildinhalte so verändert, dass eine Silhouette erzeugt wird. Exemplarisch ist dies in Abbildung 3.14 dargestellt. Technisch wird dies mittels der `OPENCV`-Funktion `convertTo()` realisiert, deren Methodensignatur in Auflistung 3.3 dargestellt ist.

Die Werte für $alpha(\alpha)$ und $beta(\beta)$ wurden durch explorative Versuche ermittelt und betragen für die Thermografiebilder $\alpha = 30; \beta = -1500$ und für die visuellen Bilder $\alpha =$

²²vgl. Abschnitt 3.4.3 Kantenerzeugung

20; $\beta = -3000$. Da die Thermografiebilder in einer Falschfarbendarstellung vorliegen, bei der die wärmsten Bereiche weiß und die kühleren Bereiche schwarz dargestellt werden, muss die extrahierte Silhouette noch invertiert werden. Dies ist in Abbildung 3.14c dargestellt. Zusätzliche Artefakte, die bei der Silhouettenerzeugung mit erzeugt werden (wie beispielsweise in Abbildung 3.14e), können vernachlässigt werden, sofern sie nicht die Form der Person verändern.

```
input_image.convertTo(output_image, output_type,  $\alpha$ ,  $\beta$ );
```

Listing 3.3: OPENCV Methodensignatur zur Kontrast- und Helligkeitskorrektur

Silhouettenregistrierung Die erzeugten Silhouetten müssen im nächsten Schritt miteinander registriert werden. Die Registrierung soll automatisch vorgenommen werden. Aus diesem Grund wurde zur Bilderregistrierung für die MATLAB IMAGE PROCESSING TOOLBOX²³ gewählt. Der in MATLAB geschriebene Code lässt sich u. a. in eine C, C++ oder .NET-Bibliothek

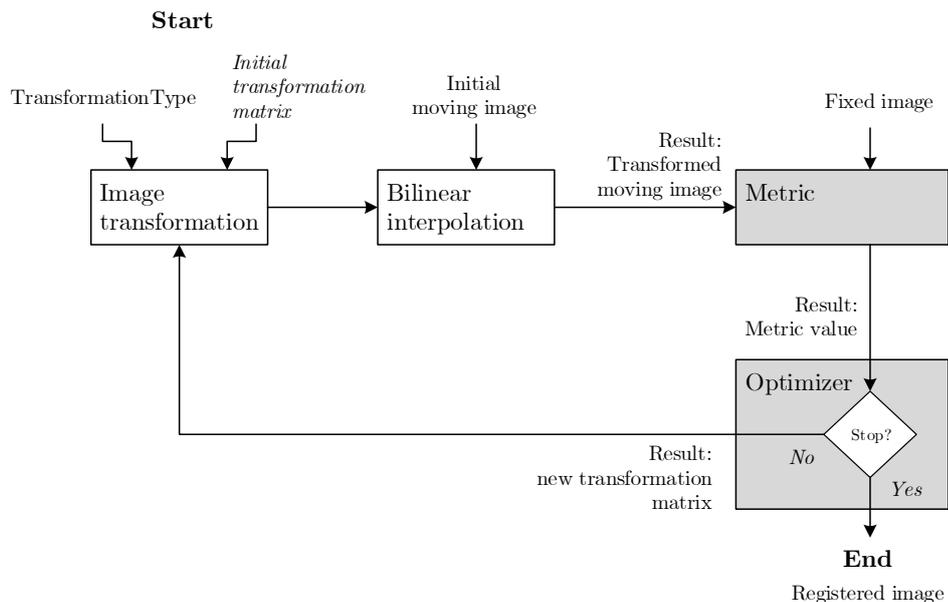


Abbildung 3.15.: Ablauf der „Intensity-Based Automatic Image Registration“ nach (MathWorks, 2017b) (*moving image*: das zu transformierende Bild; *fixed image*: das Referenzbild)

²³<https://mathworks.com/products/image>

exportieren und somit auch in anderen Softwarelösungen nutzen. Der hier beschriebene MATLAB-Code ist in einer C++ Bibliothek in das Hauptprojekt eingebunden, um im hardwarenahen C++ eine möglichst hohe Geschwindigkeit zu erlangen.

Innerhalb von MATLAB gibt es eine Reihe von Methoden, die sich mit der Bildregistrierung beschäftigen. Ebenso gibt es Implementierungen zahlreicher Algorithmen, die verwendet werden können. Eine vollständige Übersicht hiervon findet sich in [MathWorks (2017b)]. Für die Registrierung der Silhouetten wird das sog. „*Monomodal-Intensity*“-Verfahren aus der Gruppe der „*intensity-based registration*“ verwendet, deren Ablauf in Abbildung 3.15 dargestellt ist. Die Funktion `imregtform()` errechnet dann die Transformationsmatrix unter Zuhilfenahme verschiedener Parameter für das zu transformierende Bild. Die verwendeten Parameter finden sich in Anlage A.4. Die Transformationsmatrix wird im sog. „*geometric transformation object*“ gespeichert und stellt eine 3×3 -dimensionale Matrix mit sechs Freiheitsgraden als Transformationsinformationen für eine *affine* Transformation dar, wie in Gleichung 3.1 dargestellt.

$$T = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{pmatrix} \quad (3.1)$$

Der eigentliche Registrierungsprozess findet – wie in den Abbildungen 3.16c - 3.16f zu sehen – anhand der Silhouetten statt. Bei diesem wird die Silhouette des visuellen Bildes („*moving image*“) transformiert, um der Silhouette des Thermografiebildes („*fixed image*“) zu entsprechen. Das Ergebnis des Registrierungsprozesses ist eine Transformationsmatrix, die Informationen enthält, wie das zu registrierende Bild transformiert werden muss, um den anderen zu entsprechen. Nach dem Anwenden dieser mit der Methode `imwarp()` können die beiden Aufnahmen überlagert werden. Die grafische Überlagerung stellt eine Möglichkeit dar, um die Güte der Registrierung zu bewerten. Das Ergebnis der Überlagerung der Silhouette des Thermografiebildes und der transformierten Silhouette des visuellen Bildes ist in Abbildung 3.16f als Differenzdarstellung zu sehen. Mit dieser Art der Darstellung werden die Bereiche hervorgehoben (hier mit der Farbe Schwarz), die sich vollständig überlagern.

Da die Silhouettendarstellungen eine Entsprechung des visuellen bzw. Thermografiebildes sind, ist es möglich, die Transformationsmatrix, die auf der Silhouettendarstellung ermittelt wurde, auch auf die tatsächlichen Bilder anzuwenden. Dadurch wird das visuelle Bild so transformiert, dass es dem Thermografiebild nahezu entspricht. Das Ergebnis ist in Abbildung 3.16h dargestellt. Die grafische Überlagerung der Aufnahmen ist für die Weiterverwendung nicht nötig; jedoch kann sie genutzt werden, um die Qualität des Registrierungsprozesses zu bewerten.

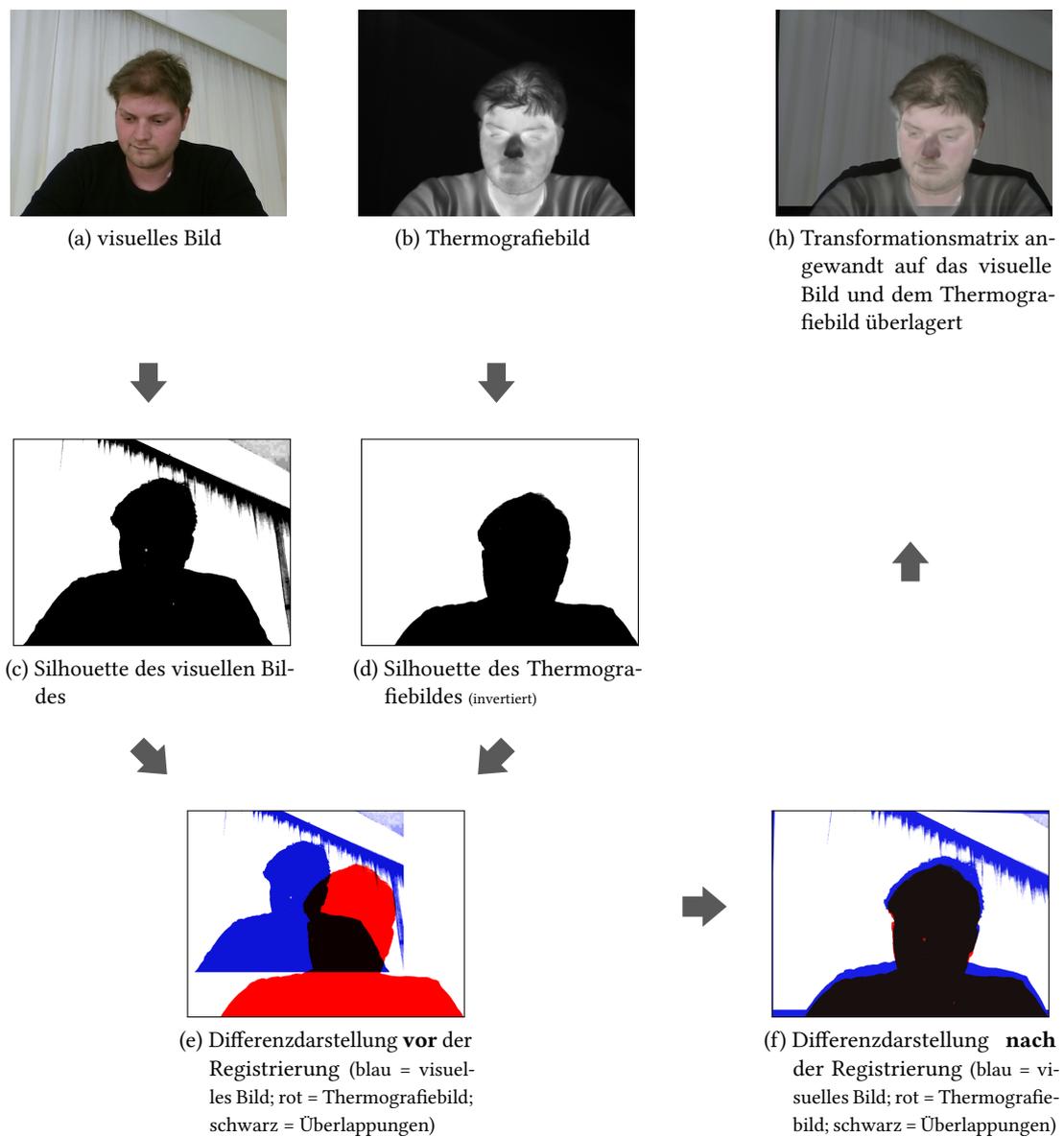


Abbildung 3.16.: Schematischer Ablauf der Bildregistrierung anhand von Silhouetten: (a)+(b) zeigen die zu registrierenden Aufnahmen, (c)+(d) die Entsprechungen der Silhouetten, die mittels Kontrasterhöhung erzeugt wurden. (e) zeigt die Differenzdarstellung der Silhouetten aus (c)+(d) im realen Größenverhältnis. (f) stellt das Ergebnis der Registrierung dar, bei dem (c) transformiert wurde, um (d) zu entsprechen. Dieselbe Transformationsmatrix der Silhouetten ist auf das Bild (a) angewendet und überlagert das Bild (b). Das Ergebnis ist in Abbildung (h) dargestellt.

Der Ablauf der Silhouettenregistrierung ist sowohl als Pseudocode in Auflistung 3.4 aufgeführt als auch in Abbildung 3.16 dargestellt.

```
1 VIS = bild_laden(visuelles_Bildes)
2 TIR = bild_laden(Thermografiebild)
3
4 VIS_S = erzeugen_der_Silhouette(VIS)
5 TIR_S = erzeugen_der_Silhouette(TIR)
6
7 //Berechnung der Transformationsmatrix (anhand der Silhouetten)
8 TRANSFORMATIONSMATRIX = imregtform(VIS_S, TIR_S)
9
10 //Transformation der Silhouette (des visuellen Bildes)
11 VIS_S_REG = imwarp(VIS_S, TRANSFORMATIONSMATRIX)
12 //Transformation des visuellen Bildes
13 VIS_REG = imwarp(VIS, TRANSFORMATIONSMATRIX)
14
15 //Überlagerung der Silhouetten
16 imshowpair(VIS_S_REG, TIR_S)
17 //Überlagerung der Bilder
18 imshowpair(VIS_REG, TIR)
```

Listing 3.4: Pseudocode für die Registrierung mittels Silhouetten

Bildregistrierung anhand von Kanten

Das zweite vorgestellte Verfahren führt eine Bildregistrierung anhand von Kanten der eigentlichen Aufnahmen durch. Dazu gliedert sich dieser Abschnitt in die Kantenerzeugung sowie die Kantenregistrierung.

Kantenerzeugung Die Kantenerzeugung verfolgt dasselbe Ziel wie das Verfahren der Silhouettenerzeugung: die multimodalen Informationen der Aufnahmen in eine unimodale Repräsentation zu überführen. Auf dieser wird auch hierbei die Registrierung durchgeführt.

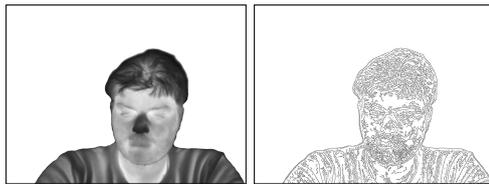
Als weitere Möglichkeit – neben der Registrierung anhand von Silhouetten – bietet sich dafür die Registrierung anhand von Kanten an. Um diese Kanten zu extrahieren, bietet sich wiederum eine Vielzahl von Methoden, Verfahren und Algorithmen an. Bekannte Vertreter der Kantenextraktion sind beispielsweise der LAPLACE-FILTER [Marr und Hildreth (1980)] oder der CANNY-ALGORITHMUS [Canny (1986)]. Angewendet werden können diese auf mindestens drei



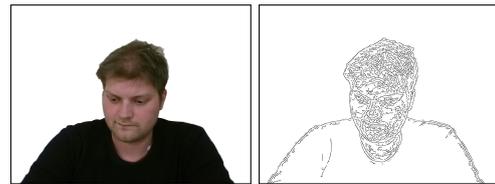
(a) Vorder- und Hintergrund



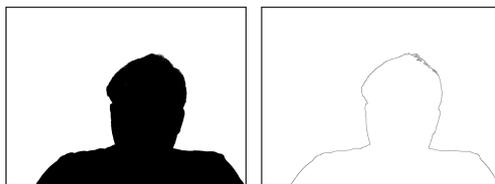
(a) Vorder- und Hintergrund



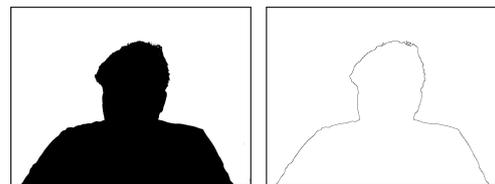
(b) Vordergrund



(b) Vordergrund



(c) Kontur



(c) Kontur

Abbildung 3.17.: Kantenerzeugung auf dem Thermografiebild

Abbildung 3.18.: Kantenerzeugung auf dem visuellen Bild

Kantenerzeugung mittels der `OPENCV`-Implementierung des `CANNY`-Algorithmus, auf den drei verschiedenen Varianten. Als verwendete Parameter wurde für diese Darstellung $lowThreshold = 6$ genutzt.

Varianten. So kann die Kantenextraktion auf den gesamten Bildbereich angewandt werden (um sowohl Vordergrund und Hintergrund zu extrahieren), nur auf den Bereich des Bildvordergrunds (hier das Gesicht) oder aber auf die Kontur des zu registrierenden Objektes. Die Varianten sind für das Thermografiebild in [Abbildung 3.17](#) und für das visuelle Bild in [Abbildung 3.18](#) dargestellt. Erzeugt wurden diese Darstellungen mit der `OPENCV`-Implementierung des `CANNY-ALGORITHMUS`. Für die Freistellung des Vordergrunds wurde das Verfahren aus dem vorherigen Abschnitt²⁴ genutzt und die Inhalte voneinander subtrahiert. Für die Konturextraktion wurden ebenso die Silhouetten des vorherigen Abschnitts genutzt.

²⁴vgl. Abschnitt [3.4.3 Silhouettenerzeugung](#)

Das Ergebnis des Algorithmus ist – neben der grafischen Darstellung – eine zweidimensionale Liste von Punkten (x, y) , die verbunden die extrahierten Linienzüge/Polygone bilden. Diese Punkte sind die Datengrundlage, auf der die spätere Registrierung erfolgt.

Um eine zügige Registrierung zu ermöglichen und die Komplexität gering zu halten, wird sich in dieser Ausprägung für die Variante der Kantenerzeugung auf der Kontur entschieden.

Kantenregistrierung Um die Punktlisten – die sich aus den Kanten ergeben – miteinander zu registrieren, wird ein Verfahren benötigt, das einerseits tolerant in Bezug auf unterschiedliche Größen der Listen ist, andererseits aber auch mit nicht hundertprozentig korrespondierenden Punkten zurechtkommt. Der Grund hierfür ist, dass die extrahierten Kanten der verschiedenen Modale zwar ähnlich sind, jedoch i. d. R. weder genaue Übereinstimmungen aufweisen, noch dieselbe Anzahl von Punkten zurückliefern.

Das hier vorgestellte Verfahren ist in drei Schritte unterteilt. Im ersten Schritt werden die Punkte (hier des visuellen Bildes) iterativ so transformiert, dass sie den Punkten des Thermografiebildes entsprechen bzw. diesen bestmöglich angenähert sind. Im zweiten Schritt wird die Transformationsmatrix zwischen den Ursprungspunkten und den transformierten Punkten berechnet. Diese wird dann im dritten Schritt auf das visuelle Bild angewandt. Das daraus resultierende Bild sollte dem Thermografiebild nahezu entsprechen und somit die Bildregistrierung erfolgreich abschließen.

Der erste Schritt wird durch ein Verfahren namens „Robust Point Set Registration Using Gaussian Mixture Models“ durchgeführt, das in der konkreten Ausführung von [Ma u. a. \(2015\)](#)

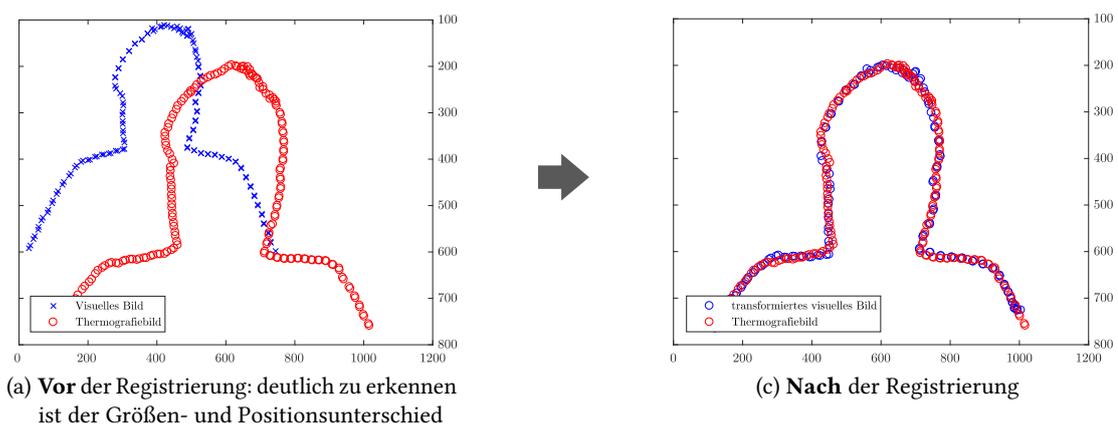


Abbildung 3.19.: Kantenregistrierung mittels „Robust Point Set Registration Using Gaussian Mixture Models“ (Aufgrund der Menge an Punkten sind in diesen Darstellungen nur 5 % der Punkte dargestellt).

veröffentlicht wurde. Implementiert ist dieses Verfahren in MATLAB-Code, der keine weiteren Bibliotheken benötigt. Neben einigen Konfigurationsparametern benötigt das Verfahren nur die zwei Punktlisten. Dies sind hier einerseits die Punktliste der Kontur des visuellen Bildes und andererseits die Punktliste der Kontur des Thermografiebildes. Die erste Liste wird während des Registrierungsprozesses so verändert, dass sie der zweiten Liste nahezu entspricht, und dann zurückgegeben. Die Punkte der Listen sind in Abbildung 3.19 dargestellt. In Abbildung 3.19c ist zu erkennen, dass die Integrität der Punkte erhalten bleibt und somit keine 1:1-Beziehung zwischen den Punkten vorherrscht. Das Ergebnis ist jedoch das gewünschte: eine Überdeckung der verschiedenen Punktlisten.

Im zweiten Schritt ist es nötig, die Transformationsmatrix zu ermitteln. Diese bildet sich aus den transformierten Punkten des visuellen Bildes und den nicht-transformierten Punkten des selbigen. Diese Punkte sind in Abbildung 3.20 dargestellt. Das verwendete Verfahren aus dem ersten Schritt bietet diese Funktion nicht.

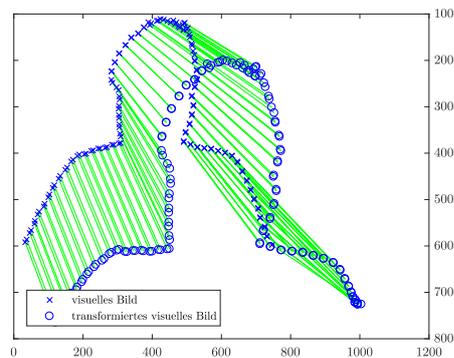


Abbildung 3.20.: Korrespondierende Punkte des visuellen Bildes vor und nach der Anwendung des „Robust Point Set Registration Using Gaussian Mixture Models“-Verfahrens (Aufgrund der Menge an Punkten sind in diesen Darstellungen nur 5 % der Punkte dargestellt).

Mit den beiden Punktlisten lässt sich die Transformationsmatrix ermitteln. Dieses Verfahren wird als Ermittlung der sog. Homography bezeichnet, wofür u. a. OPENCV- und MATLAB-Implementierungen bereitstehen. In OPENCV heißt die Methode dafür `findHomography()` und bietet drei verschiedene Algorithmen zur Bestimmung selbiger. Diese unterscheiden sich in der Menge und der Wahl der verwendeten Punkte. In diesem Stadium wird sich für die Verwendung aller Punkte entschieden. Alternativ stehen noch die Algorithmen RANSAC und LEAST-MEDIAN zur Verfügung, die nur mit einer Teilmenge der Punkte arbeiten. Alle Algorithmen liefern hierbei eine 3×3 -dimensionale Transformationsmatrix mit neun Freiheitsgraden

(*Homography*) – wie in Gleichung 3.2 dargestellt – mit der eine *perspektivische* Transformation vorgenommen werden kann.

$$H = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad (3.2)$$

Im dritten Schritt kann die eigentliche Transformationsmatrix einerseits auf die Punklisten angewandt werden, um zu überprüfen, ob die Punkte korrekt transformiert werden, und andererseits auf das zu registrierende Bild, um dieses – analog zu dem Verfahren der Registrierung anhand der Silhouetten – zu transformieren. Nach diesem Schritt ist das visuelle Bild so transformiert, dass es dem Thermografiebild nahezu entspricht. Nach der erfolgreichen Bildregistrierung können die im vorherigen Kapitel beschriebenen Komponenten mit diesem Bild weitere Schritte durchführen.

3.4.4. Zusammenfassung Softwarefusion

Die Softwarefusion mit ihren drei Themenbereichen *Kamerakalibrierung*, *Synchronisierung* und *Ausdünnung* sowie *Bildregistrierung* stellt den bedeutendsten Aspekt der Softwarekomponente dar. Diese drei aufeinanderfolgenden Schritte ermöglichen es, die Aufnahmen der visuellen Kamera mit jenen der Thermografiekamera möglichst genau zu überlagern, wodurch die visuellen Bilder mit den Temperaturinformationen der Thermografieaufnahmen angereichert werden können.

Die Kamerakalibrierung sorgt im ersten Schritt dafür, dass die Linsenverzeichnung beseitigt wird, was wiederum nötig ist, um bei der späteren Bildregistrierung die Genauigkeit zu erhöhen. Die Synchronisierung und Ausdünnung ist dann für die korrekte Zuordnung der einzelnen Bildframes zueinander verantwortlich. Auch dies ist nötig – besonders bei Bewegungen innerhalb des Aufnahmebereichs – um die Registrierung zu ermöglichen.

Die eigentliche Registrierung ermöglicht schlussendlich, dass eine Aufnahme – hier die der visuellen Kamera – so transformiert wird, dass sie dem Thermografiebild nahezu entspricht. Da das Bestimmen der Transformationsmatrix i. d. R. ein sehr zeitaufwendiger Vorgang ist, kann dies auf jedem Bildpaar vorgenommen werden oder auf einem (oder mehreren) designierten Bildpaaren. Die ermittelte Transformationsmatrix kann dann auf alle Bildpaare der Reihe angewandt werden, sofern sich beispielsweise der Aufbau während dieser Zeit nicht verändert. Die nachfolgenden Komponenten bauen auf dem Transformationsergebnis auf. So

3. Design & Realisierung

wird die Gesichtserkennung bereits auf dem transformierten Bild durchgeführt und nicht auf der tatsächlichen Aufnahme. Auf diese Weise wird der Vorteil ausgenutzt, dass sich die Gesichtsmerkmale nach der Transformation auf beiden Aufnahmen exakt überlagern und so die Temperatur an jeder Stelle abgefragt werden kann.

4. Evaluierung

In diesem Kapitel wird evaluiert, wie leistungsfähig die einzelnen im Zuge dieser Arbeit entwickelten Komponenten funktionieren. Besonders im Fokus stehen hierbei die Komponenten der Bildregistrierung und der Gesichtserkennung. Zusätzlich wird auf die ROI-Extrahierung und die Temperaturermittlung eingegangen.

Die in diesem Kapitel erläuterten Messergebnisse beziehen sich auf verschiedene Versuchsreihen, die im Rahmen des EMOTIONBIKE-Forschungsprojekts aufgezeichnet wurden. Dazu wurden willkürlich drei Probanden ausgewählt, von denen die Daten jeweils eines sog. *measures* verwendet wurde. Konkret werden die folgenden Aufnahmereihen verwendet:

- E5-P35-S2128-M1941 (im Nachfolgenden als P35 referenziert)
- E5-P40-S2161-M2004 (im Nachfolgenden als P40 referenziert)
- E5-P47-S2207-M2099 (im Nachfolgenden als P47 referenziert)

Für die gewählten Probanden wurde das Kamerasetup jeweils leicht verändert, um dieses an die jeweilige Körpergröße anzupassen. Dadurch können diese als jeweils eigene Versuchsreihen betrachtet werden.

Für die drei Probanden liegen die Thermografieaufnahmen in der Auflösung 1024×768 vor und die visuellen Aufnahmen in einer Auflösung von 800×600 . Da sie in der dynamischen Situation des EMOTIONBIKE aufgezeichnet wurden, weisen Teile der Datensätze eine starke Bewegung des Probanden bzw. dessen Kopfes auf (siehe Kapitel 4.2). Auch sind verschiedene Gesichtsausdrücke und Emotionen in den Datensätzen vorhanden. Aufgrund dieser Datengrundlage stellen die ausgewählten Probanden angemessene Kandidaten zur Evaluierung der hier umgesetzten Softwarelösung dar.

Dieser Teil der Arbeit gliedert sich dazu in das Kapitel ‚Bildregistrierung‘ – in dem sowohl das Verfahren der Bildregistrierung anhand von Silhouetten als auch jenes der Bildregistrierung anhand von Kanten evaluiert wird – das Kapitel ‚Gesichtserkennung‘, das die Grundlage für die Gesichtsanalyse bildet, sowie in das Kapitel ‚ROI-Extrahierung und die Temperaturermittlung‘. Die Kapitel betrachten sowohl die Präzision und Genauigkeit der einzelnen Verfahren als auch die dafür benötigten Zeitaufwände bzw. die Geschwindigkeiten.

4.1. Bildregistrierung

Zur Evaluierung und Bewertung der Bildregistrierung bietet sich neben der visuellen Einordnung durch den Menschen auch das Betrachten der Transformationsmatrizen an. Dieses Kapitel gliedert sich vor diesem Hintergrund in zwei Abschnitte für die beiden im Rahmen dieser Arbeit entwickelten Verfahren zur Bildregistrierung; einerseits zur Bildregistrierung anhand von Silhouetten und andererseits zur Bildregistrierung anhand von Kanten.

4.1.1. Bildregistrierung anhand von Silhouetten

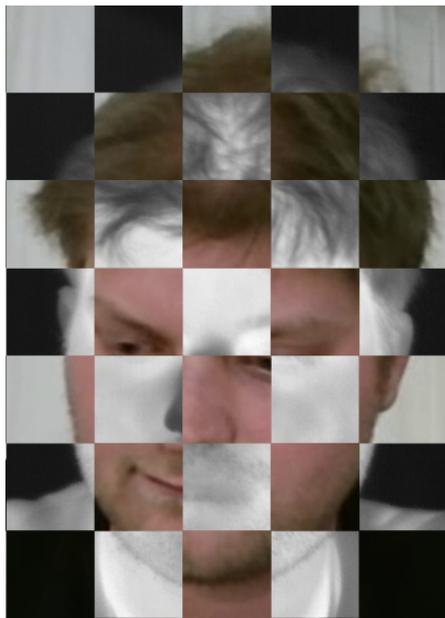
Zur Bewertung der Bildregistrierung anhand von Silhouetten wird zunächst die Anzahl der erfolgreichen Registrierungen der Gesamtanzahl der Aufnahmen – die für die Evaluierung herangezogen wurden – gegenübergestellt. Dies ist in Tabelle 4.1 aufgeführt. So wurden von den 3×25 exemplarisch ausgewählten Bildpaaren durchschnittlich 85 % erfolgreich registriert. Als erfolgreich werden alle Registrierungen betrachtet, deren Transformationsmatrix den Medianwerten zzgl. einem Schwellenwert entspricht. Die als nicht erfolgreich klassifizierten Registrierungen wiesen typischerweise Werte mit dem Faktor 100 zum Median auf und sind somit als eindeutige Ausreißer zu werten. Aufgrund der eindeutigen Auffälligkeit innerhalb der Transformationsmatrix können diese automatisch erkannt und verworfen werden. Alternativ kann auf diesen Bildern auch die Transformationsmatrix eines vorherigen Bildpaars (oder Mittelwerte/Mediane vorheriger Bildpaare) angewandt werden. Keine erfolgreiche Registrierung bedeutet daher in diesem Kontext nur, dass auf den Bildinhalten (hier den Silhouetten) eines Bildpaares mit dem vorgeschlagenen Verfahren eine fehlerhafte Registrierung – also ein *false positive* – erfolgt ist.

Proband	Erfolgreich registriert	Betrachtete Aufnahmen	Prozentsatz
P35	19	25	76 %
P40	23	25	92 %
P47	24	25	88 %

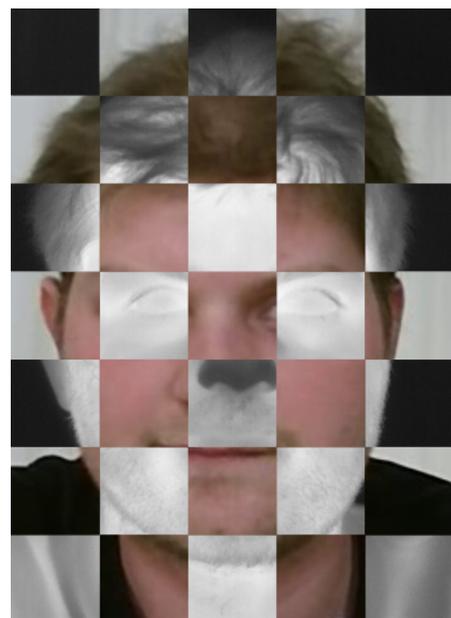
Tabelle 4.1.: Quantitatives Registrierungsergebnis der Bildregistrierung anhand von Silhouetten

4. Evaluierung

Die als „erfolgreich registriert“ klassifizierten Bildpaare werden im Weiteren detaillierter betrachtet. Hierfür bietet sich – besonders für die Auswertung durch den Menschen – die Überlagerung der transformierten Aufnahmen an. So zeigt Abbildung 4.1 das visuelle Bild, das mittels Schachbrettüberlagerung durch das Thermografiebild überlagert ist. Anhand dieses Verfahrens lässt sich besonders gut die Kongruenz der Gesichtsmerkmale visualisieren. In der Abbildung ist dazu ein hoher Kongruenzanteil (Abbildung 4.1a), sowie ein weniger hoher Kongruenzanteil (Abbildung 4.1b) dargestellt. Anhand dessen ist zu erkennen, dass bei beiden exemplarisch gezeigten Aufnahmen die Köpfe nahezu exakt übereinander abgebildet sind. Bedingt wird dies durch die zur Registrierung verwendeten Silhouetten und deren gutes Registrierungsergebnis. Die Abweichungen, die vor allem an der Nase deutlich erkennbar sind, sind hauptsächlich auf die physikalische Gegebenheit zurückzuführen, dass das Gesicht keine planare Ebene darstellt und somit eine Parallaxe entsteht. Je weiter die Kameras physikalisch voneinander entfernt platziert sind (bei gleichbleibender oder kürzerer Entfernung



(a) Registrierte Aufnahmen mit hohem Kongruenzanteil aller Gesichtsmarkmale



(b) Registrierte Aufnahmen mit weniger hohem Kongruenzanteil der Gesichtsmarkmale (das linke Ohr und der Nasenvorhof weisen einen Versatz auf).

Abbildung 4.1.: Exemplarischer Vergleich der Bildregistrierung anhand von Silhouetten zwischen zwei repräsentativen Aufnahmen als Schachbrettüberlagerungsdarstellung. Zu sehen ist die visuelle Aufnahme, überlagert mit dem Thermografiebild.

zum Objekt)¹, desto größer wird dieser Effekt. Gleichzeitig kann dies verringert werden, desto größer wird dieser Effekt. Gleichzeitig kann er verringert werden, wenn die Kameras physikalisch näher beieinanderstehen bzw. wenn sie ein weiter entferntes Objekt betrachten oder das betrachtete Objekt planar ist.

Neben der bereits eingangs betrachteten Ausreißererkenung lässt sich die Transformationsmatrix noch detaillierter betrachten. Abbildung 4.2 illustriert diese dazu analog zur ursprünglichen Matrixdarstellung, bei der die ersten beiden Spaltenvektoren die Rotation, Skalierung, Scherung etc. darstellen und der dritte Spaltenvektor die Translation. Dazu sind in der Abbildung sowohl die Mittelwerte der jeweiligen Elemente als auch die Standardabweichung für diese aufgeführt. Als Datengrundlage dienen hier die Werte der als erfolgreich klassifizierten Registrierungen der drei verschiedenen Probanden.

Aus diesen geht hervor, dass die Transformationsmatrix sich selbst bei Betrachtung der Durchschnittswerte bei jedem hier betrachteten Probanden unterscheidet. Auch zeigt sich, dass die Standardabweichung um teilweise mehrere hundert Pixel vom Durchschnittswert abweicht. Der Grund hierfür sind unter anderem die starken Bewegungen der Probanden während der Versuche und die damit verbundene Veränderung der Distanz zu den Kameras. Dies erklärt zudem die nicht einheitliche Standardabweichung für die einzelnen Elemente der jeweiligen Probanden, da diese sich untereinander nicht identisch bewegt haben.

Besonders auffällig sind die hohen Pixelwerte des dritten Spaltenvektors der Matrix. Dieser stellt die Translation dar und lässt sich mit den unterschiedlichen Auflösungen der verwendeten Kameras und der damit einhergehenden Verschiebung in den optischen Mittelpunkt erklären. Dies erklärt ebenso die hohen Werte in den Elementen A_{11} und A_{22} , die sich unter anderem aus der nötigen Skalierung ergeben.

Die konkreten Werte für Skalierung, Rotation und Scherung lassen sich jedoch aus der Transformationsmatrix nicht ablesen.

¹vgl. Kapitel 2.4.2 - Abschnitt [Anordnung des Messsystems](#)

4. Evaluierung

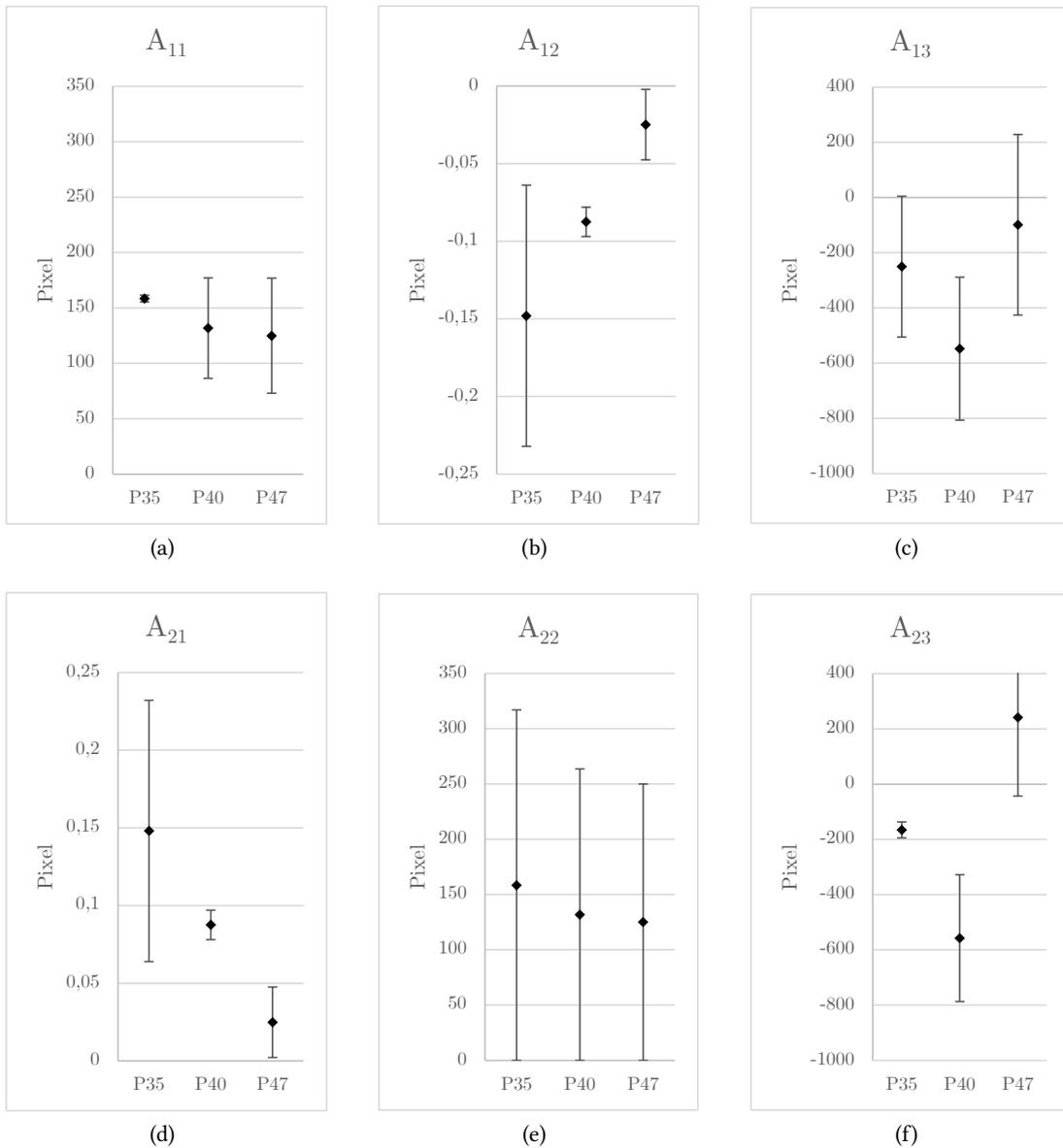


Abbildung 4.2.: Darstellung der 2×3 -dimensionalen Transformationsmatrix für die affine Transformation mit den Durchschnittswerten und der Standardabweichung für die drei Probanden. Vertikale Achse: Pixeldistanz; horizontale Achse: Probandenbezeichnung; Raute: Durchschnittswert; Fehlerindikatoren: Standardabweichung (zu beachten sind die teilweise unterschiedlich gewählten Achsenskalierungen).

Die einzelnen Diagramme sind analog zur Matrixdarstellung angeordnet, bei der der letzte Zeilenvektor (aufgrund von unveränderlichen Werten) nicht dargestellt wird:

$$\begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ 0 & 0 & 1 \end{pmatrix} \quad (4.1)$$

Geschwindigkeit

Neben der Betrachtung der Genauigkeit ist ein weiterer bedeutender Aspekt eines Software-systems die Geschwindigkeit, in der diese Genauigkeit erreicht wird. Bei der hier betrachteten Bildregistrierung anhand von Silhouetten lässt sich dies in drei Teilschritte unterteilen: die Silhouettenerzeugung, in der aus dem Thermografiebild und dem visuellen Bild die Silhouettenrepräsentierung erzeugt wird, den eigentlichen Registrierungsvorgang sowie das Anwenden der Transformation auf das visuelle Bild. Die einzelnen Zeitmessungen für die verschiedenen Schritte und die hier betrachteten Probanden sind in Abbildung 4.3 – sowohl als einzelne Werte als auch mit dem dazugehörigen Durchschnittswert – dargestellt.

Die Silhouettenerzeugung (Abbildung 4.3a) benötigt durchschnittlich über alle Probanden hinweg 19 Millisekunden. Dies ist vor allem auf einige Ausreißer zurückzuführen. So liegt der Median hierfür bei lediglich 12 Millisekunden.

Der eigentliche Registrierungsvorgang (Abbildung 4.3b) und die dortige Ermittlung der Transformationsmatrix ist der mit Abstand teuerste Schritt der Bildregistrierung anhand von Silhouetten sowie des Gesamtsystems. Hier beträgt die Zeitdauer im Durchschnitt 6,8 Sekunden. Auch der Median unterscheidet sich hier nicht gravierend von diesem Wert.

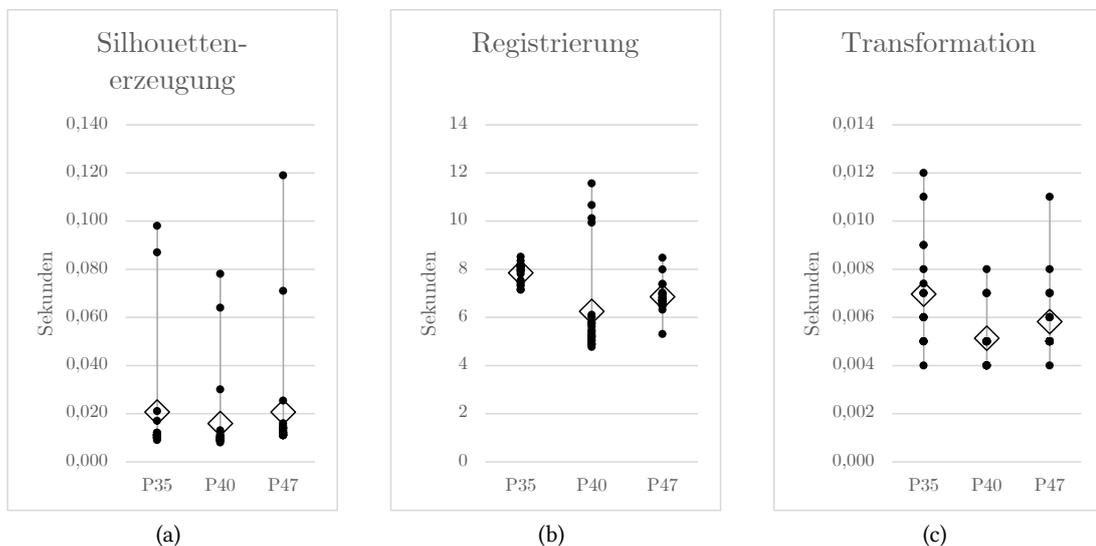


Abbildung 4.3.: Darstellung der Geschwindigkeiten für die verschiedenen Schritte der Bildregistrierung auf Silhouetten. Vertikale Achse: Sekunden; horizontale Achse: Probandenbezeichnung; Raute: Durchschnittswert; Punkt: Einzelwert (zu beachten sind die unterschiedlich gewählten Achsenskalierungen).

Das abschließende Anwenden der ermittelten Transformationsmatrix auf das visuelle Bild und die damit verbundene Transformation (Abbildung 4.3c) benötigen im Durchschnitt lediglich sechs Millisekunden (im Median fünf Millisekunden).

Dies bedeutet, dass die komplette Bildregistrierung anhand von Silhouetten im Durchschnitt genau sieben Sekunden benötigt, wenn das Verfahren auf jedes Bildpaar der Serie angewendet wird. Dieser Wert kann drastisch gesenkt werden, wenn die Silhouettenerzeugung und der eigentliche Registrierungsvorgang nicht auf jedes Bildpaar angewandt werden. Dazu kann eine einmalig ermittelte Transformationsmatrix auch auf andere Bildpaare derselben Reihe angewandt werden, ohne dass jeweils eine Neuberechnung dieser vonnöten ist. Somit entstehen diese Zeitaufwände nur einmalig. Für jedes Bildpaar sind dann nur noch die Zeitaufwände der Transformation einzuplanen. Parallelisierung sowie eine Erhöhung der Rechenleistung (ggf. unter der Nutzung von CUDA-Prozessoren) würden diese Werte weiter verringern.

4.1.2. Bildregistrierung anhand von Kanten

Zur Bewertung der Bildregistrierung anhand von Kanten wird ebenfalls zunächst die Anzahl der erfolgreich registrierten Bildpaare betrachtet. Dazu wurden dieselben Bildpaare wie im Abschnitt *Bildregistrierung anhand von Silhouetten* verwendet, um die Ergebnisse vergleichbar zu machen. Hieraus geht als erste Erkenntnis hervor, dass bei dieser Art der Registrierung nur durchschnittlich 60 % der Aufnahmen erfolgreich miteinander registriert werden konnten. Tabelle 4.2 stellt dazu die Werte separiert für jeden Probanden dar. Auch hier wurden jene Bildpaare als erfolgreich registriert klassifiziert, deren Werte dem Median zzgl. einem Schwellenwert entsprechen. Außerdem gilt hier gleichermaßen, dass die fehlerhaft registrierten Aufnahmen durch das Anwenden der Transformationsmatrix aus einem vorherigen Bildpaar (das als erfolgreich registriert klassifiziert wurde) als erfolgreich registriert gelten.

Proband	Erfolgreich registriert	Betrachtete Aufnahmen	Prozentsatz
P35	13	25	52 %
P40	13	25	52 %
P47	19	25	76 %

Tabelle 4.2.: Quantitatives Registrierungsergebnis der Bildregistrierung anhand von Kanten

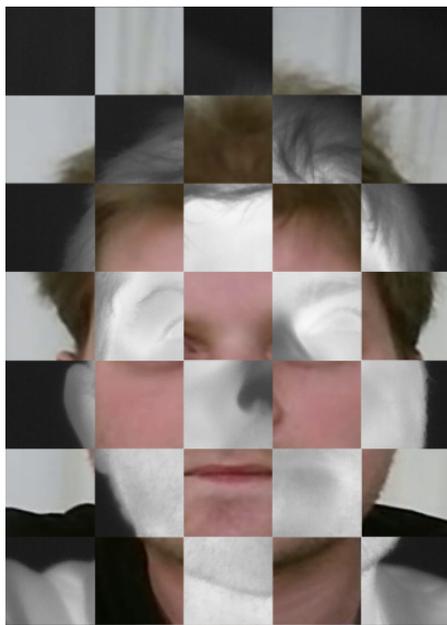
In Abbildung 4.4 sind zwei exemplarische Ergebnisse der Bildregistrierung dargestellt. Auf diesen zeigt sich, dass die Kopfform sowie die einzelnen Gesichtsmerkmale zwar nah beieinander positioniert sind, jedoch nicht an das Ergebnis der *Bildregistrierung anhand von Silhouetten* herankommen (vgl. Abbildung 4.1). Ebenso ist zu erkennen, dass kein gleichmäßiger Versatz während der Registrierung entsteht. So zeigt Abbildung 14 beispielsweise eine nahezu korrekte Überdeckung der Ohren, jedoch einen großen Versatz im Kinnbereich dieses Probanden. Grund hierfür ist die projektive Transformation, die das visuelle Bild teilweise zu stark transformiert und somit verzerrt, sodass keine hundertprozentige Überdeckung erreicht wird.

Der maßgebliche Grund für die teilweise zu stark verzerrten visuellen Aufnahmen ist der Schritt zur Ermittlung der Homography zwischen den beiden Punktlisten.² Hier zeigt sich, dass die nichtlineare 2D-Transformation des „*Robust Point Set Registration Using Gaussian Mixture Models*“-Verfahrens zur Registrierung der Punktlisten keine eindeutige lineare 2D-Transformation ermitteln kann, die auf die Aufnahmen angewendet werden kann. Das bedeutet, dass zwar die einzelnen Elemente der Punktliste stark korrelierend aufeinander abgebildet

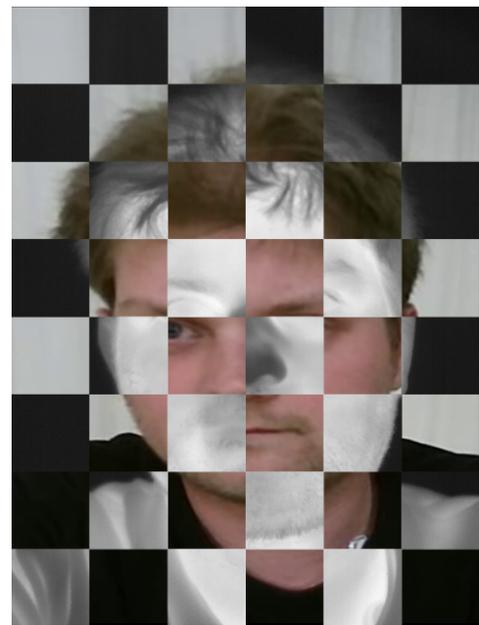
²vgl. Kapitel 3.4.3 - Abschnitt *Kantenregistrierung*

4. Evaluierung

werden können, sich anhand dessen aber keine eindeutige Transformationsmatrix ermitteln lässt, die auf alle Punkte gleichermaßen angewandt werden kann. Dies bedeutet gleichzeitig, dass sich keine Homography für das gesamte Bild ermitteln lässt, sondern lediglich eine Näherung zur ursprünglichen Punkttransformation.



(a) Registrierte Aufnahmen (Gesichtsmerkmale weisen einen Versatz auf)



(b) Registrierte Aufnahmen (Gesichtsmerkmale weisen einen Versatz auf)

Abbildung 4.4.: Exemplarischer Vergleich der Bildregistrierung anhand von Kanten zwischen zwei repräsentativen Aufnahmen als Schachbrettüberlagerungsdarstellung. Zu sehen ist die visuelle Aufnahme überlagert mit dem Thermografiebild.

Auch für dieses Verfahren werden die Werte der Transformationsmatrizen näher betrachtet. Die Messgrößen sind dazu in Abbildung 4.5 für jedes Matricelement und jeden Probanden dargestellt. Diese zeigen – im Vergleich zu jenen der *Bildregistrierung anhand von Silhouetten* –, dass die Matricelemente A_{11} , A_{12} , A_{21} und A_{22} (die Skalierung, Rotation und Scherung angeben) vergleichbar kleine Werte erhalten haben. Auch ist die Standardabweichung vergleichsweise gering. Die Elemente, die die Translation angeben (A_{13} und A_{23}), weisen hingegen ähnliche Werte auf wie das zuvor betrachtete Verfahren. Der bei einer Homography veränderliche dritte Zeilenvektor gibt die perspektivische Transformation an.

4. Evaluierung

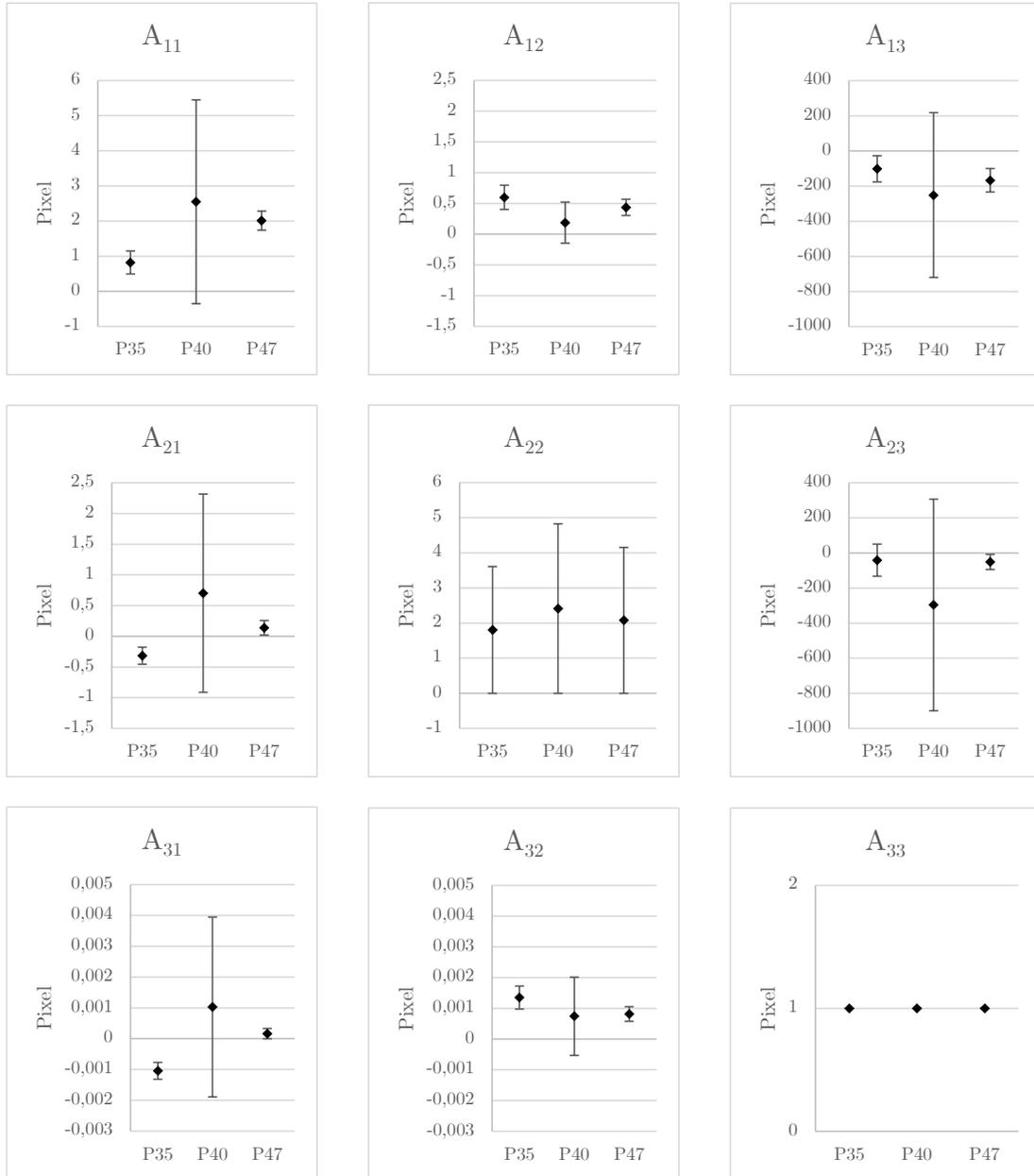


Abbildung 4.5.: Darstellung der 3×3 -dimensionalen Transformationsmatrix für die projektive Transformation mit den Durchschnittswerten und der Standardabweichung für die drei Probanden. Vertikale Achse: Pixeldistanz; horizontale Achse: Probandenbezeichnung; Raute: Durchschnittswert; Fehlerindikatoren: Standardabweichung (zu beachten sind die teilweise unterschiedlich gewählten Achsenskalierungen).

Die einzelnen Diagramme sind analog zur Matrixdarstellung angeordnet:

$$\begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \quad (4.2)$$

Geschwindigkeit

Auch für dieses Verfahren wird die Geschwindigkeit betrachtet. Dazu werden vier verschiedene Schritte unterschieden, die bei diesem Verfahren benötigt werden. Differenzieren lässt sich hierbei in die Schritte der Silhouetten- bzw. Kantenerzeugung, in denen aus dem Thermografiebild sowie dem visuellen Bild die Punktlisten der Kantenrepräsentierungen erzeugt werden, den Registriervorgang der Punktlisten, der Ermittlung der Homography sowie das Anwenden der Transformation. Dazu werden auch hier sowohl die Durchschnittswerte als auch die jeweiligen Einzelwerte für jeden Probanden und für jeden Schritt dargestellt. Die Diagramme dazu finden sich in [Abbildung 4.6](#).

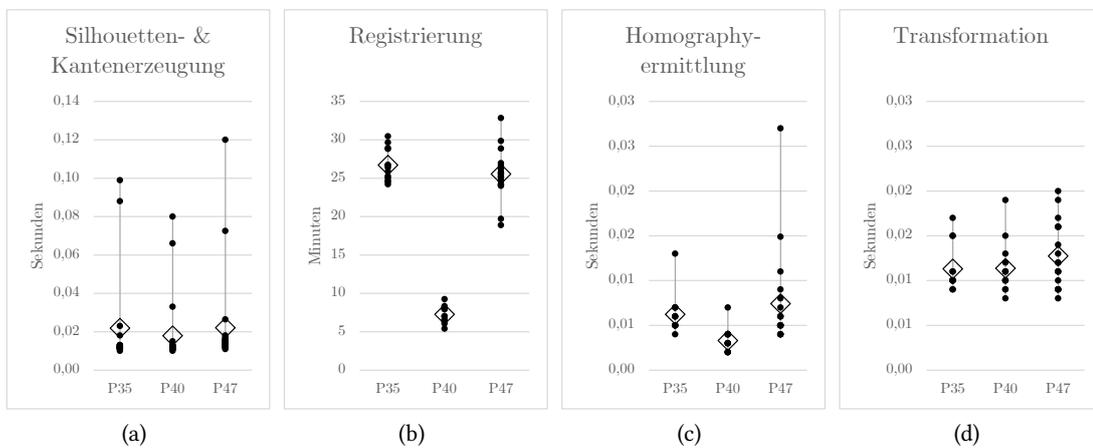


Abbildung 4.6.: Darstellung der Geschwindigkeiten für die verschiedenen Schritte der Bildregistrierung anhand von Kanten. *Vertikale Achse: Sekunden/Minuten; horizontale Achse: Probandenbezeichnung; Raute: Durchschnittswert; Punkt: Einzelwert* (zu beachten sind die unterschiedlich gewählten Achsenskalierungen).

So zeigt sich, dass die Silhouetten- bzw. Kantenerzeugung ([Abbildung 4.6a](#)) in durchschnittlich 20 Millisekunden (Median: 12 Millisekunden) erfolgt.

Der Registriervorgang ([Abbildung 4.6b](#)), bei dem versucht wird, die Punktlisten aufeinander abzubilden, ist vergleichsweise extrem zeitaufwendig und liegt bereits im Minutenbereich. So zeigt sich, dass im Durchschnitt – über alle Probanden hinweg – eine Dauer von 20 Minuten für das Registrieren eines Bildpaares benötigt wird. In Spitzen dauert eine Registrierung bis zu 30 Minuten. Gleichzeitig ist jedoch auch ersichtlich, dass der gleiche Vorgang für einen anderen Probanden auch schon nach 5 Minuten abgeschlossen ist. Dies ist durch die teilweise stark abweichenden Mengen an Punkten zu erklären. So wurden bei Probanden mit weniger Kopfhaaren auch weniger Details als Silhouette extrahiert und dadurch Punkte generiert. Bei

Personen mit längeren Haaren (vor allem solchen, die aus der Silhouette herausstehen) erhöht sich gleichzeitig der Detailgrad und somit die Menge der Punkte, die es zu registrieren gilt. Dies zeigt, dass ein Minimieren der Punkte dieses Verfahren deutlich beschleunigen und seine Dauer möglicherweise auch in den Sekundenbereich absenken kann.

Die Homographyermittlung (Abbildung 4.6c) lässt sich nach der Registrierung in durchschnittlich fünf Millisekunden durchführen.

Die eigentliche Transformation (Abbildung 4.6d), bei der die ermittelte Homography auf das visuelle Bild angewandt wird, lässt sich in durchschnittlich 30 Millisekunden durchführen.

4.1.3. Zusammenfassung Bildregistrierung

Aus der Evaluierung für den Bereich der Bildregistrierung geht eindeutig hervor, dass beim Verwenden von mehr Details nicht zwangsläufig das Registrierungsergebnis verbessert wird. So zeigt sich, dass der Ansatz, der nur die Silhouettendarstellung betrachtet, deutlich genauere Registrierungen ermöglicht als das Verfahren, das viele Punkte – hier jene der Kontur – betrachtet, um anhand dieser eine Registrierung durchzuführen.

Bei den Geschwindigkeitsmessungen zeigt sich, dass der Schritt der Registrierung einen Flaschenhals für die Bildregistrierung darstellt, da er bei beiden Verfahren deutlich mehr Zeit benötigt als die anderen Schritte. Diese absoluten Zeiten können aber nicht zuletzt durch das Reduzieren der Komplexität, d. h. beispielsweise durch vorherige Reduzierung der Größe der genutzten Bilder bzw. der Punktlisten oder durch andere Programmier Techniken, weiter gesenkt werden.

4.2. Gesichtserkennung

Um die Qualität der Gesichtserkennung zu evaluieren, ist es vorteilhaft, Informationen zum Szenario zu kennen, in dem diese stattfinden soll. Im hier verwendeten Szenario werden die Probanden frontal durch die visuelle Kamera gefilmt, wobei es durch die Fahrradfahrbewegung mitunter zu starken Bewegungen des Kopfes kommt. Dazu illustriert Abbildung 4.7 die Position der Nasenspitze auf jedem Frame für jeden Probanden. Dies zeigt, dass große Bildbereiche der 800×600 Pixel großen Aufnahmen durch das Gesicht eingenommen werden. Ebenso zeigt sich, dass die Probanden (zum Teil aufgrund der verschiedenen Versuchsdurchläufe) eine unterschiedlich starke Bewegung aufweisen. Für das Szenario ist zudem zu beachten, dass die Probanden nicht direkt in die Kamera schauen, da sie hauptsächlich die visuelle Komponente des EMOTIONBIKE betrachten. Die Probanden schauen dadurch über die Kamera hinweg.

Für die Evaluierung wurden zusammen 7.762 Aufnahmen der drei Probanden betrachtet. Von diesen konnte auf 6.827 eine Gesichtserkennung stattfinden, was einer Quote von 88 % entspricht. Bei den 935 Aufnahmen, auf denen kein Gesicht erkannt wurde, wurde der Kopf so

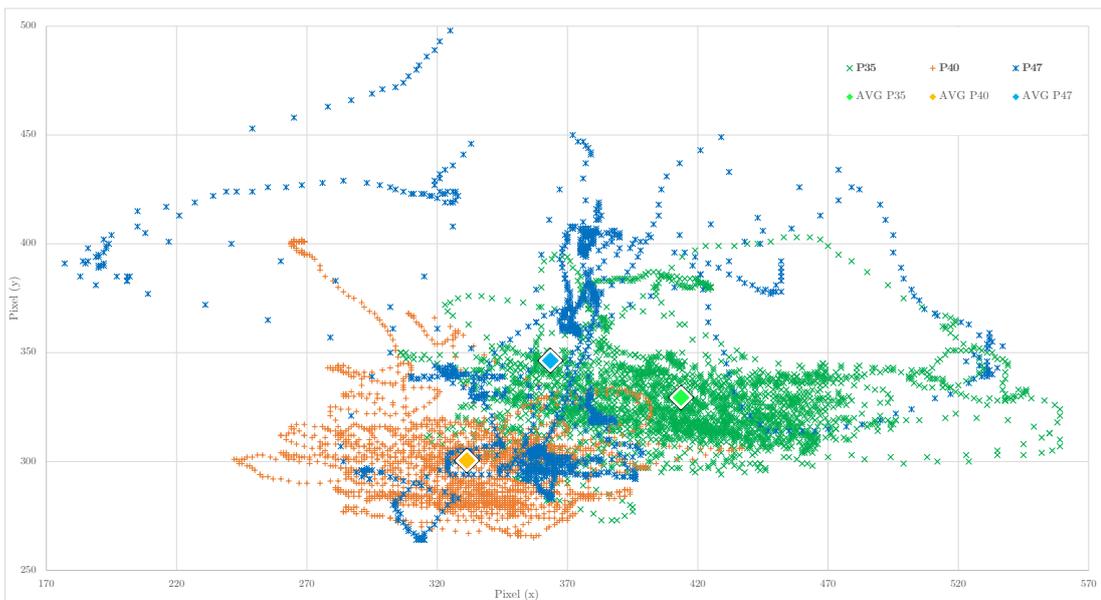


Abbildung 4.7.: Position der Nasenspitze (DLIB-Punkt 30)³ innerhalb der 6.827 näher betrachteten Aufnahmen für die drei Probanden inkl. Hervorhebung deren durchschnittlicher Position

³vgl. Anlage 2 – DLIB Landmarking Punkte

stark in den rotatorischen *Pitch*- und *Yaw*-Bewegungen verdreht, dass das Gesicht nicht mehr in Richtung der Kamera (oder der visuellen Komponente des EMOTIONBIKE) zeigt und dadurch nicht mehr alle Gesichtsmerkmale im Bild sind.

Auf die Genauigkeit des hier verwendeten Verfahrens wird nicht näher eingegangen, da dies bereits in King (2009) dokumentiert ist.

4.2.1. Geschwindigkeit

Wie bereits im Rahmen der Analyse erwähnt, gliedert sich die Gesichtserkennung in die Schritte *face detection* und *face landmarking*. Bei der Geschwindigkeitsbewertung wird daher ebenfalls zwischen diesen beiden Kategorien unterschieden, die auch in Abbildung 4.8 wieder aufgegriffen werden.

Bei der Geschwindigkeitsmessung der 6.827 Aufnahmen für die *face detection* (Abbildung 4.8a) zeigt sich, dass durchschnittlich 101 Millisekunden benötigt werden. Dabei nutzt das Verfahren in der aktuellen Ausprägung keine Informationen über den vorherigen Frame, sondern durchsucht jeweils das gesamte Bild erneut.

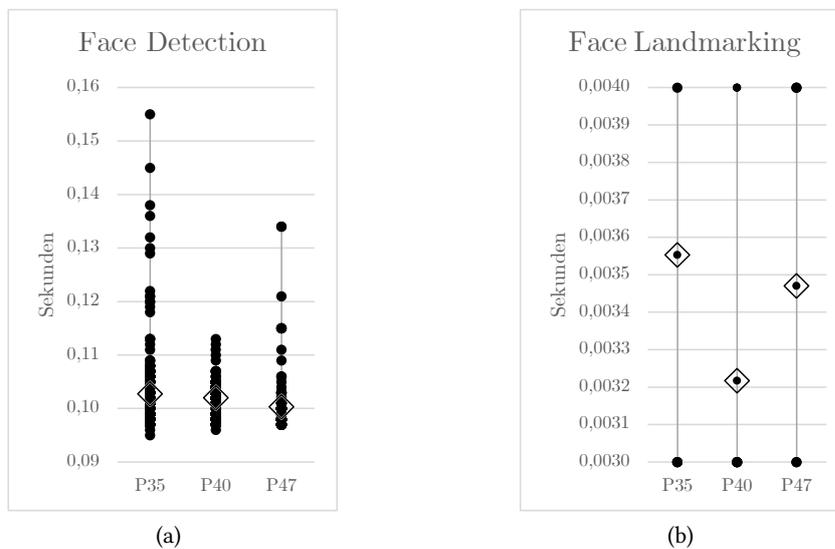


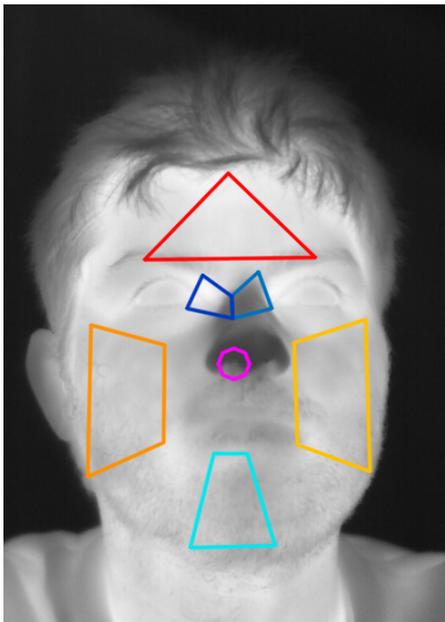
Abbildung 4.8.: Darstellung der Geschwindigkeiten für die verschiedenen Schritte der Gesichtserkennung. Vertikale Achse: Sekunden; horizontale Achse: Probandenbezeichnung; Raute: Durchschnittswert; Punkt: Einzelwert (zu beachten sind die unterschiedlich gewählten Achsenskalierungen).

Das anschließende *face landmarking* (Abbildung 4.8b) benötigt hingegen nur eine verschwindend geringe Zeit. So beläuft sich der entsprechende Zeitwert durchschnittlich auf 3 Millisekunden pro Aufnahme.

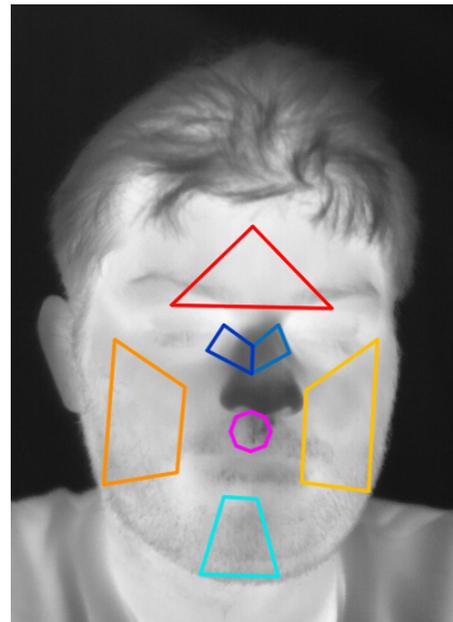
4.3. ROI-Extraktion und Temperaturermittlung

Um eine thermografische Gesichtsanalyse durchführen zu können, müssen die ROI im Gesicht festgelegt werden, innerhalb derer die Temperaturen bestimmt werden sollen. Zur Evaluierung der ROI wird dazu an dieser Stelle deren Position genauer betrachtet.

Bereits bei der Festlegung der ROI und der Auswahl der Bezugspunkte, aus denen die Flächen geformt werden, wurde darauf geachtet, dass diese nicht zu nah an anderen Gesichtsmerkmalen liegen. Dies soll verhindern, dass aufgrund von Präzisionsdefiziten bei der Bildregistrierung, falsche Bereiche die Temperaturmessergebnisse verfälschen. Abbildung 4.9 stellt dazu die festgelegten ROI auf zwei repräsentativen Aufnahmen dar. Das dazugehörige *face landmarking*



(a) Thermografiebild mit hohem Kongruenzanteil aller ROIs



(b) Exemplarische Darstellung eines Versatzes bei der ROI „Regio nasalis (Nasenspitze)“

Abbildung 4.9.: Überlagerung der Thermografiebilder mit den ROI, die durch das *face landmarking* auf dem korrespondierenden visuellen Bild ermittelt wurden (nach der Registrierung mit dem dazugehörigen Thermografiebild.

Hier: *Bildregistrierung anhand von Silhouetten*)

und die darauf basierende Generierung der ROI wurde anhand des visuellen Bildes ermittelt und das korrespondierende Thermografiebild in der Darstellung damit überlagert. Diese Darstellung dient somit lediglich zur Bewertung der Lage der ROI durch den Menschen und wird für die Temperaturermittlung nicht benötigt.

Aus der Darstellung geht hervor, dass – analog zu den Evaluierungen der Bildregistrierungsverfahren – die Nase den geringsten Kongruenzanteil aufweist. Der Grund hierfür ist, dass diese bereits bei der Bildregistrierung nicht korrekt mit dem korrespondierenden visuellen Bild übereinstimmt. Dieses Defizit zieht sich daher bis zu diesem Punkt durch das Verfahren. Aufgrund der bereits gezeigten Unterschiede bei der Bildregistrierung von Bildpaar zu Bildpaar unterscheidet sich auf diese Weise auch der Überdeckungsanteil. So zeigt Abbildung 4.9a einen deutlich höheren Kongruenzanteil als Abbildung 4.9b.

Des Weiteren zeigt sich in Abbildung 4.9b, dass die „Regio frontalis (Stirn)“ teilweise die Augenbrauen beinhaltet. Auch dies ist auf die Bildregistrierung zurückzuführen. Dies bedeutet jedoch nicht automatisch, dass diese ROI für die Temperaturermittlung verworfen werden muss. Je nach Messmethode kann das Eintreten von anderen Gesichtsmerkmalen in die ROI ignoriert werden. Bei der Bestimmung der Median- oder Maximaltemperatur der ROI spielen die (hier kühleren) Augenbrauen keine Rolle.

4.3.1. Geschwindigkeit

Sowohl die Definition der ROI als auch die Extraktion der Temperaturwerte innerhalb dieser haben auf die Verarbeitungsgeschwindigkeit nahezu keine Auswirkung. So ergeben sich die Polygone aus bereits vorhandenen Punkten, die während des *face landmarking* ermittelt wurden (zzgl. der arithmetischen Berechnung von Schnittpunkten einiger dieser Punkte). Die Temperaturwerte besitzen die Datenstruktur eines C++ Vektors. Durch die pixelgenaue Position der ROI kann auf diesen Vektor ebenfalls pixelgenau zugegriffen werden und die dort extrahierten Werte können aggregiert werden. Auch dies erfolgt im unteren Millisekundenbereich, weshalb hierzu keine detaillierte Auswertung vorgenommen wird.

5. Schlussbetrachtung

In diesem Teil der Arbeit wird das zuvor Erarbeitete nochmals rekapituliert und zusammengefasst. Dazu wird auf alle Teile und alle im Rahmen dieser Arbeit entwickelten Komponenten und Module eingegangen. Abgeschlossen wird dieser Teil und damit die Arbeit mit einem Ausblick, in dem beschrieben wird, wie das System in Hinblick auf Geschwindigkeit und Präzision verbessert werden kann und welche Anknüpfungsmöglichkeiten diese Arbeit bietet.

5.1. Zusammenfassung

Ziel dieser Arbeit war es, eine thermografische Gesichtsanalyse mit einem multimodalen Kamerasystem durchzuführen. Dazu wurden die Grundlagen analysiert und auf Grundlage dieser ein Systemdesign vorgeschlagen und implementiert, das abschließend in Bezug auf seine Eignung evaluiert wurde.

Teil 1 beginnt mit der Einleitung, die die Motivation, die Kontexteinordnung sowie die Gliederung der Arbeit beinhaltet. Dazu wird sowohl auf die Zielsetzung als auch darauf, wie Thermografie einen Beitrag zum Forschungsfeld *Human-Computer Interaction* und *Companion-Systemen* leistet, eingegangen.

Im 2. Teil wurde die Analyse vorgenommen, in der auf die technologischen Grundlagen der Infrarotthermografie und der Gesichtserkennung eingegangen wurde. In Bezug auf Letztere wurde dazu zwischen Thermografiebildern und visuellen Bildern unterschieden. Es wurde gezeigt, dass eine genaue Form der Gesichtserkennung – hier als *face landmarking* bezeichnet – nicht auf Aufnahmen des Thermografiespektrums möglich ist. Dies begründet die Notwendigkeit der Nutzung einer weiteren Kamera, die die Aufnahmen im visuellen Spektrum aufzeichnet. Um eine Gesichtsanalyse durchzuführen zu können, musste sich mit der grundlegenden Gesichtsanatomie auseinandergesetzt werden. Aus den hieraus ermittelten Erkenntnissen wurden die ROI definiert, anhand derer eine thermografische Analyse erfolgen soll. Den zweiten Baustein der Analyse bilden die Fusionsmöglichkeiten multimodaler Aufnahmen. In diesem wurde analysiert, wie die Kameras kalibriert werden können, um die baulich bedingte Linsenverzeichnung zu beseitigen, und wie Aufnahmen grundsätzlich miteinander registriert

werden können. Für die weitere Analyse wurden verwandte Arbeiten aufgeführt, die sich mit ähnlichen Fragestellungen befassen wie diese Arbeit. Die Erkenntnisse aus diesen fließen in die spätere Realisierung mit ein. Am Ende dieser Analyse wurde noch das vornehmliche Szenario betrachtet, für das das System optimiert ist. Ebenso wurden die Anforderungen nochmals zusammengefasst sowie Nicht-Ziele abgegrenzt.

Im Teil 3 wurden die Erkenntnisse der Analyse in ein Design einer Systemlösung eingebracht, das ebenso implementiert wurde. Dazu wurde auf die verwendete Hardware sowie dessen physikalisches Setup eingegangen. Im anschließenden Softwaredesign wurden neben der Architektur auch die einzelnen Module der Software beschrieben. Dazu wurde auf die Bilder bzw. Datenakquirierung der Kameras, auf die Gesichtserkennung im visuellen Bild sowie auf die Temperaturextraktion eingegangen. Zudem wurde die Datenstruktur definiert, mit der die extrahierten Temperaturwerte ausgegeben werden. Große Aufmerksamkeit wurde in Design und Realisierung dem Aspekt der Softwarefusion gewidmet, dessen Ziel die Bildregistrierung ist. Dazu wurden in diesem Kapitel die Kameras kalibriert sowie die Aufnahmen der verschiedenen Kameras rechnerisch so miteinander synchronisiert, dass Bildpaare entstanden sind. Diese beinhalten Aufnahmen, die zum annähernd selben Zeitpunkt aufgezeichnet wurden und so Kandidaten für eine Bildregistrierung sind. Für die abschließende Bildregistrierung wurden zwei verschiedene Verfahren vorgeschlagen, von denen eines mit Silhouetten und das andere mit Kanten arbeitet. In beiden Fällen arbeiten die Bildregistrierungsalgorithmen mit Entsprechungen der tatsächlich aufgezeichneten Bilder, um deren multimodale Darstellungen in eine unimodale Repräsentation zu überführen. Auf diesen wurden jeweils die nötigen Transformationen ermittelt und unverändert auf die reellen Aufnahmen angewandt, wodurch diese registriert wurden. Dieser Teil beinhaltet somit sämtliche Verarbeitungsschritte vom Aufbau der Hardware über das Akquirieren der Aufnahmen bis hin zur Bildregistrierung und der Temperaturextraktion in den zuvor definierten ROI.

Teil 4 evaluiert das im vorherigen Teil umgesetzte Hard- und Softwaresystem. Für die Bildregistrierung wurden beide vorgeschlagenen Verfahren detailliert auf die Genauigkeit und auf die Verarbeitungsgeschwindigkeit hin untersucht. Hier zeigte sich, dass das Verfahren der *Bildregistrierung anhand von Silhouetten* eine deutlich höhere Anzahl von erfolgreichen Registrierungen aufweist. Auch zeigte sich, dass die Genauigkeit signifikant höher ist als beim Verfahren der *Bildregistrierung anhand von Kanten*. Quantitativ bedeutet dies, dass das Verfahren mit Silhouetten eine Quote von 85 % erreicht, jenes mit Kanten hingegen lediglich 60 %. Die qualitative Bewertung der Genauigkeit ließ sich im Rahmen der Evaluation nicht in Zahlen bemessen, doch zeigte sich auf den Aufnahmen, dass auch hier das Verfahren auf

Silhouetten deutlich bessere Ergebnisse erzielt. Bei der Evaluierung der Geschwindigkeiten der Registrierungsvorgänge zeigte sich, dass die Überführung der reellen Aufnahmen in die Repräsentationen – die für das Registrieren verwendet werden – einen verschwindend geringen zeitlichen Aufwand von im Median 19 Millisekunden benötigte. Auch die abschließende Transformation der visuellen Bilder benötigte im Maximum 12 Millisekunden beim Verfahren mit Silhouetten (affine Transformation) und maximal 20 Millisekunden beim Verfahren mit Kanten (perspektivische Transformation). Den zeitlich kostspieligen Anteil spiegelte die eigentliche Registrierung bzw. die damit verbundene Ermittlung der nötigen Transformation wider. So benötigt dieser Teil unter Verwendung der Silhouetten durchschnittlich 6,8 Sekunden, bei der Verwendung der Kanten durchschnittlich ca. 20 Minuten. Gleichzeitig wurde allerdings auch deutlich, dass letzterer Wert hauptsächlich durch die Anzahl der verwendeten Kanten (bzw. Punkte dieser) im Mittel so in die Höhe schnellte. Eine Reduzierung der verwendeten Punkte und dadurch eine Reduzierung der Eingangsparameter ließe den Wert beträchtlich sinken. Ebenso wurde aufgezeigt, dass die Bildregistrierung nicht auf jedem Bildpaar neu ermittelt werden muss, sondern dies auch einmalig pro Kamerasetup erfolgen kann. Diese einmalig berechnete Transformationsmatrix kann auf alle Bildpaare der Reihe angewandt werden und benötigt somit nur die Zeitaufwände, die die Transformation des visuellen Bildes verbraucht.

Neben der Bildregistrierung wurde das Modul für die Gesichtserkennung ebenso evaluiert. Hier zeigte sich, dass mit dem vorgeschlagenen Verfahren eine Gesichtserkennung auf 88 % der betrachteten Aufnahmen des visuellen Spektrums erreicht wird. Diese Zahl begründet sich maßgeblich durch das Abwenden des Gesichts auf den restlichen 12 % der Aufnahmen. Da die hier betrachtete Gesichtsanalyse prinzipiell frontale Aufnahmen benötigt, eigneten sich diese 12 % auch nicht für die Weiterverarbeitung. Innerhalb der Evaluation dieses Teilbereichs zeigte sich zudem, dass die Probanden starke, nicht homogene Bewegungen auf dem betrachteten Szenario des EMOTIONBIKE verursachten. Dies stellte für dieses Verfahren jedoch keine erkennbaren Probleme dar. Bei der Betrachtung der Geschwindigkeiten zeigte sich, dass die Gesichtserkennung durchschnittlich 103 Millisekunden benötigt. Dies inkludiert sowohl die Dauer für die *face detection* als auch für das *face landmarking*.

Bei der Analyse der ROI-Extraktion zeigte sich, dass die verschiedenen ROI robust an den jeweiligen verorteten Bereichen eingezeichnet werden konnten. Die ROI reichen nur in wenigen Fällen in andere Bereiche. Eine Ausnahme stellte hierbei die „Regio nasalis (Nasenspitze)“ dar. Aufgrund des Parallaxe-Effekts, der durch die verschiedenen Blickwinkel der Kameras verursacht wurde, ist das Registrierungsergebnis der Nase nicht so präzise wie bei den anderen Gesichtsmarkern, weshalb es hier zu Defiziten kommt, die eine Auswertung dieser ROI ggf. verhindern.

Aus der Evaluierung geht somit als Erkenntnis hervor, dass sich die Kombination aus den vorgeschlagenen Verfahren und Tools als geeignet erweist, um eine thermografische Gesichtsanalyse mit einem multimodalen Kamerasystem durchzuführen. Zu den vorgeschlagenen Verfahren zählt das Verfahren zur *Bildregistrierung anhand von Silhouetten*, zusammen mit dem hier genutzten Tool zur präzisen Gesichtserkennung in der Ausprägung des *face landmarking*. Anhand dessen wiederum die ROI ermittelt wurden und auf diesen anschließend die Temperaturwerte aggregiert wurden.

5.2. Ausblick

Die im Rahmen dieser Thesis entwickelten Softwarelösungen und theoretischen Bestandteile stehen für zukünftige Projekte den Laboren und Einrichtungen der HAW HAMBURG zur Verfügung. Hierdurch wird einer Fortführung und Erweiterung der hier beschriebenen Thematik ermöglicht.

Ein Aspekt zur Weiterführung des Themas befasst sich mit der Analyse der physiologischen Gegebenheiten von Personen bzw. des menschlichen Gesichts. So ermöglicht das hier entwickelte System eine Auswertung der Temperaturen, was wiederum eine tief greifende Analyse sowohl hinsichtlich des emotionalen Zustands, als auch hinsichtlich medizinischer Diagnosen ermöglicht. Einordnen lassen sich diese Systeme in den Bereich der *Consumer Health Informatics* (CHI) welche ein Teilgebiet der medizinischen Informatik darstellen. Diese Systeme stellen Gesundheitsressourcen medizinischen Laien zur Verfügung und können beispielsweise durch Temperaturinformationen angereichert werden, welche der Person in verarbeiteter oder unverarbeiteter Form zugänglich gemacht werden. Eine Bewegung die unter dem Namen *quantified self* zusammengefasst wird, bilden dafür Einsatzzwecke zur personenzentrierten Analyse und Optimierung eines selbst.

Ein zweiter Aspekt zur Weiterführung des Themas ist die Optimierung des hier vorgeschlagenen Systems im Hinblick auf Geschwindigkeit wie auch Genauigkeit der Registrierung und der dadurch präziseren Temperaturermittlung. Ein Bestandteil hiervon ist die Gesichtserkennung direkt auf dem Thermografiebild und somit die dortige *feature extraction* direkt auf diesem zu ermöglichen. Die Nutzung einer zweiten, visuellen Kamera, könnte dadurch entfallen, was bereits im Abschnitt 2.2.1 aufgezeigt wurde.

Neben diesen ganz offensichtlichen Fortführungsmöglichkeiten die dieses System ermöglicht, besteht eine Möglichkeit darin, die Übertragbarkeit sowie Generalisierung zu evaluieren und dahin gehend anzupassen. In der aktuellen Ausprägung ist das System für die Nutzung in einer labor- bzw. laborähnlichen Umgebung konzipiert und entwickelt worden. Interessante

Fragestellungen ergeben sich besonders hinsichtlich der Anwendbarkeit in dynamischen und unkontrollierbaren Umgebungen. Denkbar wäre ein System, das auch für mehrere Personen im Bildfeld die Temperaturwerte liefert. Des Weiteren besteht die Möglichkeit die Thermografiedaten mit Daten anderer Sensoren zu kombinieren. Dadurch wird erreicht, dass ein vollumfänglicheres Bild erreicht wird und neue Aspekte mit in den jeweiligen Anwendungsfall einfließen.

In Anbetracht der voranschreitenden Verbreitung und Computerisierung der Umwelt mit der daraus resultierenden Rechnerallgegenwart (*Ubiquitous Computing*) kann man davon ausgehen, dass auch Temperaturdaten – sofern die positiven Aspekte der Allgemeinheit bewusst werden und diese am Markt nachgefragt werden – Einzug in das tägliche Leben erhalten. Erste Ansätze zeigen sich diesbezüglich bereits in den aufkommenden Thermografiekameras für Heimanwender, welche die Infrarotthermografie einem breiten Markt zugänglich macht.

A. Anlagen

A.1. Anlage 1 – Kamerakalibrierung

Nachfolgend werden die ermittelten intrinsischen Parameter der drei betrachteten Kameras aufgeführt, welche im Abschnitt [3.4.1 Kamerakalibrierung](#) ermittelt wurden und im Abschnitt [2.4.1 Kamerakalibrierung](#) theoretisch beschreiben wurden.

INFRA^{TEC} VARIOCAM HD 875

Focal length (pixels):	[1744.6374 +/- 55.2407	;	1747.4619 +/- 52.4195]
Principal point (pixels):	[4513.7398 +/- 22.0249	;	470.1978 +/- 35.5575]
Skew:	[-5.2404 +/- 5.9035]		
Radial distortion:	[-0.2633 +/- 0.1093	;	-3.5177 +/- 2.3999	;
		25.4689 +/- 16.7740]		
Tangential distortion:	[-0.0111 +/- 0.0062	;	-0.0041 +/- 0.0036]

Tabelle A.1.: Intrinsische Parameter der INFRA^{TEC} VARIOCAM HD 875

ALLIED VISION MANTA G-235C

Focal length (pixels):	[2677.5597 +/- 53.1694	;	2678.4514 +/- 51.3007]
Principal point (pixels):	[846.6444 +/- 47.3997	;	706.2746 +/- 44.5021]
Skew:	[-4.1570 +/- 5.7865]		
Radial distortion:	[-0.1039 +/- 0.0860	;	-1.3522 +/- 2.0823	;
		5.9583 +/- 14.7131]		
Tangential distortion:	[-0.0014 +/- 0.0028	;	-0.0083 +/- 0.0028]

Tabelle A.2.: Intrinsische Parameter der ALLIED VISION MANTA G-235C

MICROSOFT KINECT FOR XBOX ONE

Focal length (pixels):	[742.9419 +/- 19.2653	;	740.2612 +/- 18.7166]
Principal point (pixels):	[477.1537 +/- 9.6715	;	479.0556 +/- 8.6288]
Skew:	[-5.0754 +/- 1.8266]		
Radial distortion:	[0.0300 +/- 0.0389	;	0.0693 +/- 0.3859	;
		0.4645 +/- 1.3464]		
Tangential distortion:	[0.0198 +/- 0.0036	;	-0.0126 +/- 0.0048]

Tabelle A.3.: Intrinsische Parameter der MICROSOFT KINECT FOR XBOX ONE

A.2. Anlage 2 – DLIB Landmarking Punkte

Für die Gesichtserkennung werden im Rahmen dieser Arbeit ein *face detection*, sowie deren Verfeinerung als *face landmarking* benötigt. Das hierfür gewählte Tool DLIB¹ ermöglicht beide Schritte. Das *face landmarking* arbeitet dazu auf der `shape_predictor_68_face_landmarks.dat` welches 68 Punkte des Gesichts extrahiert. Diese Punkte sind Merkmalen im Gesicht zugeordnet. Dazu findet sich in der nachfolgenden Tabelle A.4 sowie in Abbildung A.1 die Zuordnung der Punkte.

Bezeichnung		Intervall der Punkte	geom. Figur
Gesichtsaußenseiten		[0,16]	Polygonzug
Augenbrauen	Links	[17,21]	Polygonzug
	Rechts	[22,26]	Polygonzug
Nase	Bein	[27,30]	Polygonzug
	Vorhof/Dome	[30,35]	Hexagon
Auge	Links	[36,41]	Hexagon
	Rechts	[42,47]	Hexagon
Lippe	Äußere Kontur	[48,59]	Polygon
	Mundöffnung	[60,67]	Polygon

Tabelle A.4.: DLIB Annotierung der Identifikatoren 0 - 67

¹<http://dlib.net>

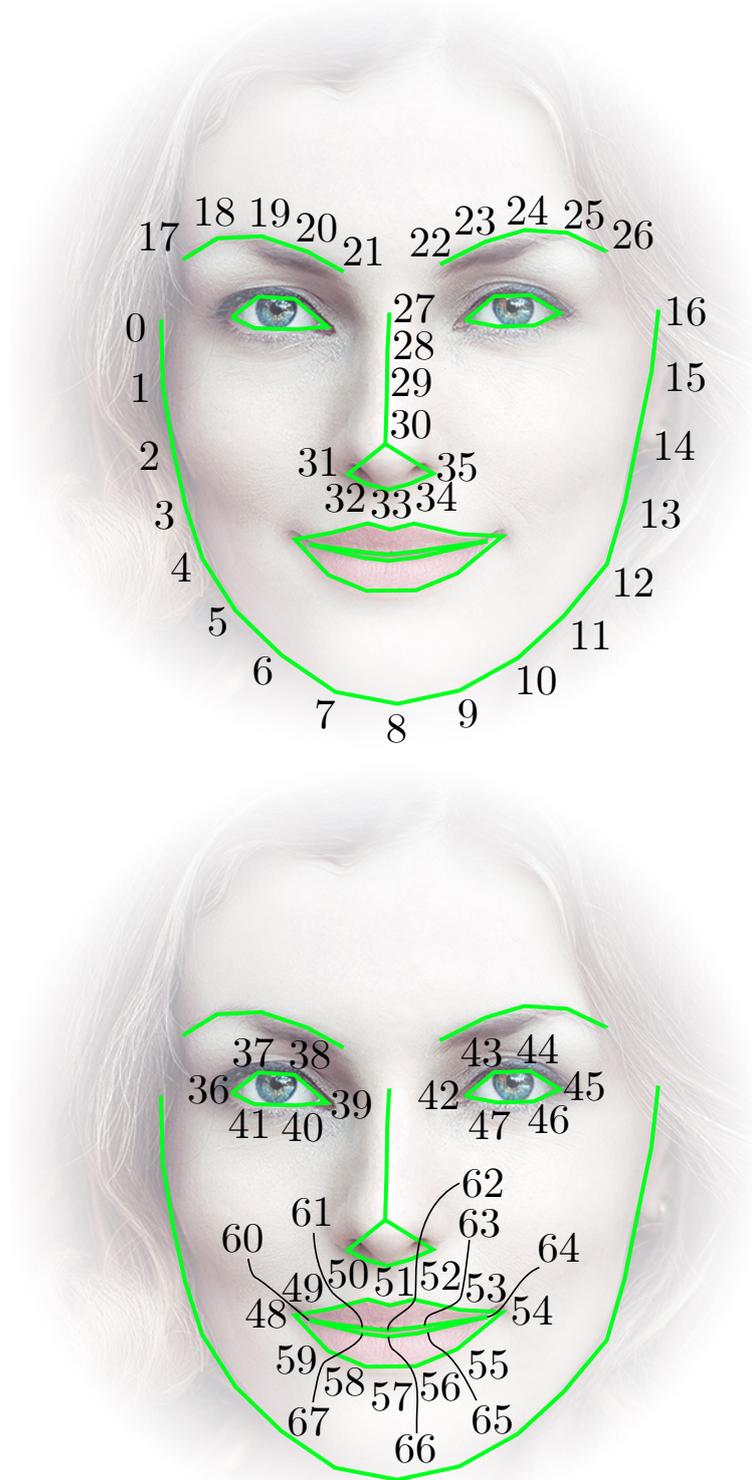


Abbildung A.1.: DLIB Annotierung der Identifikatoren 0 - 67

Hintergrundgrafik: ©©

A.3. Anlage 3 – Datenstruktur

```
1 {
2   "ROIs" :
3   [
4     // Liste von ROI Elementen (siehe Listing A.2: ROI Element)
5   ],
6   "Metadata" :
7   [
8     {
9       "Width" : " 1024 " ,
10      "Hight" : " 768 "
11    }
12  ]
13 }
```

Listing A.1: Basis Datenstruktur als JSON

```
4 {
5   "Name" : " Stirn " ,
6   "Values" :
7   [
8     { "Min" : 23.36 } ,
9     { "Max" : 23.97 } ,
10    { "Average" : 23.00 } ,
11    { "Median" : 23.64 }
12  ] ,
13   "Unit" : " celsius " ,
14   "RoiMetadata" :
15   [
16     { "IncludedPixel" : 5463 } ,
17     { "Vertices" :
18       [
19         { "x" : 399 , "y" : 400 } ,
20         { "x" : 546 , "y" : 390 } ,
21         { "x" : 467 , "y" : 321 }
22       ]
23     }
24  ]
25 }
```

Listing A.2: ROI Element als JSON

A.4. Anlage 4 – Registrierung

Registrierung auf Silhouetten

```
1 [optimizer, metric] = imregconfig('monomodal');
2 optimizer.GradientMagnitudeTolerance = 1.00000e-04;
3 optimizer.MinimumStepLength = 1.00000e-05;
4 optimizer.MaximumStepLength = 6.25000e-02;
5 optimizer.MaximumIterations = 100;
6 optimizer.RelaxationFactor = 0.500000;
```

Listing A.3: Verwendete Parameter für die *Registrierung auf Silhouetten*

Registrierung auf Kanten

```
1 iter_num = 1;
2 anneal = 0.93;
3 sigma0 = 0.08;
4 N0 = 10;
5 eta = 1;
6 is_grad = 1;
7 fgt = 0;
8 beta = 0.5;
9 lambda = 0.01;
```

Listing A.4: Verwendete Parameter für die *Registrierung auf Kanten*

A.5. Anlage 4 – IRB Temperaturdaten

Nachfolgende Auflistung stellt einen exemplarischen Aufbau der ASCII-Datei dar, die mittels dem SDK² von der INFRA^{TEC} VARIOCAM HD 875 erzeugt wird. Diese Datei enthält sowohl Metainformationen (gekennzeichnet durch das Schlüsselwort „[Settings]“), als auch einen Temperaturwert für jeden Pixel (gekennzeichnet ab dem Schlüsselwort „[Data]“). Diese Passage beinhaltet in diesem Fall eine Matrix mit 1024×768 , Tabulator getrennten, Elementen.

```
1 [Settings]
2 Version=3
3 ImageWidth=1024
4 ImageHeight=768
5 ShotRange=16.85;36.85
6 CalibRange=-40.00;229.00
7 TempUnit=°C
8
9 [Data]
10 24,12    24,10    ...
11 ...
```

Listing A.5: Aufbau der ASCII-Datei der Temperaturdaten

²vgl. Abschnitt 3.2.1 INFRA^{TEC} VARIOCAM HD 875

Literaturverzeichnis

- [Aguilera u. a. 2012] AGUILERA, Cristhian ; BARRERA, Fernando ; SAPPÀ, Angel D. ; TOLEDO, Ricardo: A Novel SIFT-Like-Based Approach for FIR-VS Images Registration. In: *Proc. Quantitative InfraRed Thermography* (2012). – URL <http://dx.doi.org/10.21611/qirt.2012.149>
- [Allied Vision 2013] ALLIED VISION: *Vimba .NET API - Programmer's Manual*. Taschenweg 2a, 07646 Stadtroda, Germany: Allied Vision Technologies GmbH (Veranst.), 2013. – Dokumentation
- [Allied Vision 2017] ALLIED VISION: Manta Data sheet G-235. Taschenweg 2a, 07646 Stadtroda, Germany, 2017. – Forschungsbericht. – URL <https://www.alliedvision.com/en/products/machine-vision-cameras/detail/Manta/G-235/action/pdf.html>. Dokumentation
- [Anderhuber u. a. 2012] ANDERHUBER, Friedrich ; PERA, Franz ; STREICHER, Johannes: *Waldeyer - Anatomie des Menschen - Lehrbuch und Atlas in einem Band*. Berlin : Walter de Gruyter, 2012. – ISBN 978-3-110-22863-2
- [Aslantas und Kurban 2009] ASLANTAS, V. ; KURBAN, R.: A comparison of criterion functions for fusion of multi-focus noisy images. In: *Optics Communications* 282 (2009), Nr. 16, S. 3231 – 3242. – URL <https://doi.org/10.1016/j.optcom.2009.05.021>. – ISSN 0030-4018
- [Bahnsen 2013] BAHNSEN, Chris: Thermal-visible-depth image registration. In: *Unpublished Master Thesis, Aalborg University, Aalborg, Denmark* (2013). – Masterthesis
- [Bay u. a. 2008] BAY, Herbert ; ESS, Andreas ; TUYTELAARS, Tinne ; VAN GOOL, Luc: Speeded-up robust features (SURF). In: *Computer vision and image understanding* 110 (2008), Nr. 3, S. 346–359. – URL <http://dx.doi.org/10.1016/j.cviu.2007.09.014>
- [Belhumeur u. a. 1997] BELHUMEUR, P. N. ; HESPANHA, J. P. ; KRIEGMAN, D. J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. In: *IEEE Transactions*

- on pattern analysis and machine intelligence* 19 (1997), Nr. 7, S. 711–720. – URL <https://doi.org/10.1109/34.598228>
- [Bhatnagar u. a. 2015] BHATNAGAR, Gaurav ; WU, Q.M. J. ; LIU, Zheng: A new contrast based multimodal medical image fusion framework. In: *Neurocomputing* 157 (2015), Nr. Supplement C, S. 143 – 152. – URL <https://doi.org/10.1016/j.neucom.2015.01.025>. – ISSN 0925-2312
- [Bledsoe 1966] BLEDSOE, Woodrow W.: The model method in facial recognition. In: *Panoramic Research Inc., Palo Alto, CA, Rep. PR1* 15 (1966)
- [BSI 2003] BSI: Studie BioFace: Vergleichende Untersuchungen von Gesichtserkennungssystemen. In: *Bundesamt für Sicherheit in der Informationstechnik* (2003). – URL <https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Studien/BioFace/BioFaceIIBericht.pdf>
- [Bulanon u. a. 2009] BULANON, D.M. ; BURKS, T.F. ; ALCHANATIS, V.: Image fusion of visible and thermal images for fruit detection. In: *Biosystems Engineering* 103 (2009), Nr. 1, S. 12–22. – URL <http://doi.org/10.1016/j.biosystemseng.2009.02.009>. – ISSN 1537-5110
- [Canny 1986] CANNY, J.: A Computational Approach to Edge Detection. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8 (1986), Nov, Nr. 6, S. 679–698. – URL <https://doi.org/10.1109/TPAMI.1986.4767851>. – ISSN 0162-8828
- [Ellgring 2010] ELLGRING, Johann H.: Nonverbale Kommunikation. (2010)
- [Ellmauthaler u. a. 2013] ELLMAUTHALER, A. ; SILVA, E. A. B. da ; PAGLIARI, C. L. ; GOIS, J. N. ; NEVES, S. R.: A novel iterative calibration approach for thermal infrared cameras. In: *2013 IEEE International Conference on Image Processing*, URL <http://doi.org/10.1109/ICIP.2013.6738450>, Sept 2013, S. 2182–2186. – ISSN 1522-4880
- [Gautherie und Gros 1980] GAUTHERIE, Michel ; GROS, Charles M.: Breast thermography and cancer risk prediction. In: *Cancer* 45 (1980), Nr. 1, S. 51–56. – URL <http://dx.doi.org/10.1002/cncr.2820450110>. – ISSN 1097-0142
- [Gebhart u. a. 1980] GEBHART, J.R. ; KINCHEN, B.E. ; STRANGE, R.R.: *Optical pyrometer and technique for temperature measurement*. sep 1980. – URL <https://www.google.com/patents/US4222663>. – US Patent 4,222,663

- [Hartley und Zisserman 2003] HARTLEY, Richard ; ZISSERMAN, Andrew: *Multiple view geometry in computer vision*. Cambridge university press, 2003. – URL <https://doi.org/10.1017/CBO9780511811685>. – ISBN 978-0-511-81168-5
- [Hennes u. a. 2015] HENNES, Maria ; URBAN, Steffen ; WURSTHORN, Sven: Zur Synchronisierung von Multi-Sensor-Systemen–Grundlagen und Realisierungen. (2015). – URL <http://dx.doi.org/10.13140/RG.2.1.2006.5761>
- [Hjelmås und Low 2001] HJELMÅS, Erik ; LOW, Boon K.: Face Detection: A Survey. In: *Computer Vision and Image Understanding* 83 (2001), Nr. 3, S. 236–274. – URL <http://dx.doi.org/10.1006/cviu.2001.0921>. – ISSN 1077-3142
- [InfraTec 2004] INFRA TEC: *Einführung in Theorie und Praxis der Infrarot-Thermografie*. Gostritzer Strasse 61 - 63, 01217 Dresden, Germany: InfraTec GmbH Infrarotsensorik und Messtechnik (Veranst.), 2004
- [InfraTec 2014] INFRA TEC: *Zugriff auf Echtzeit-Digitaldaten mit InfraTec-SDK*. Gostritzer Straße 61 - 63, 01217 Dresden, Germany: InfraTec GmbH Infrarotsensorik und Messtechnik (Veranst.), 2014. – Dokumentation
- [InfraTec 2015a] INFRA TEC: *Benutzerhandbuch VarioCam HD*. Gostritzer Straße 61 - 63, 01217 Dresden, Germany: InfraTec GmbH Infrarotsensorik und Messtechnik (Veranst.), 2015. – Benutzerhandbuch
- [InfraTec 2015b] INFRA TEC: *Thermografie Theorie - Physikalische Grundlagen*. Gostritzer Strasse 61 - 63, 01217 Dresden, Germany: InfraTec GmbH Infrarotsensorik und Messtechnik (Veranst.), 2015. – URL <http://www.infratec.de/de/thermografie/thermografie-wissen/theorie.html>
- [Ioannou u. a. 2014] IOANNOU, Stephanos ; GALLESE, Vittorio ; MERLA, Arcangelo: Thermal infrared imaging in psychophysiology: Potentialities and limits. In: *Psychophysiology* 51 (2014), 10, Nr. 10, S. 951–963. – URL <http://doi.org/10.1111/psyp.12243>. – ISSN 1469-8986
- [Jain und Li 2005] JAIN, Anil K. ; LI, Stan Z.: *Handbook of face recognition*. New York, NY : Springer, 2005. – URL <http://dx.doi.org/10.1007/978-0-85729-932-1>
- [King 2009] KING, Davis E.: Dlib-ml: A Machine Learning Toolkit. In: *Journal of Machine Learning Research* 10 (2009), S. 1755–1758

- [Kletz und Kleimann 2017] KLETZ, Florian ; KLEIMANN, Jorin: Verwendung einer Thermografiekamera im Human-computer interaction Kontext / Fakultät Technik und Informatik, Department Informatik, Hochschule für Angewandte Wissenschaften Hamburg. Berliner Tor 7, 20099 Hamburg, 2017. – Forschungsbericht. – URL <http://users.informatik.haw-hamburg.de/~ubicomp/projekte/master2017-proj/kletz-kleimann.pdf>. Projektbericht
- [Kong u. a. 2005] KONG, Seong G. ; HEO, Jingu ; ABIDI, Besma R. ; PAIK, Joonki ; ABIDI, Mongi A.: Recent advances in visual and infrared face recognition-a review. In: *Computer Vision and Image Understanding* 97 (2005), Nr. 1, S. 103 – 135. – URL <http://doi.org/10.1016/j.cviu.2004.04.001>. – ISSN 1077-3142
- [Kong u. a. 2007] KONG, Seong G. ; HEO, Jingu ; BOUGHORBEL, Faysal ; ZHENG, Yue ; ABIDI, Besma R. ; KOSCHAN, Andreas ; YI, Mingzhong ; ABIDI, Mongi A.: Multiscale Fusion of Visible and Thermal IR Images for Illumination-Invariant Face Recognition. In: *International Journal of Computer Vision* 71 (2007), Feb, Nr. 2, S. 215–233. – URL <https://doi.org/10.1007/s11263-006-6655-0>. – ISSN 1573-1405
- [Lahiri u. a. 2012] LAHIRI, B.B. ; BAGAVATHIAPPAN, S. ; JAYAKUMAR, T. ; PHILIP, John: Medical applications of infrared thermography: A review. In: *Infrared Physics & Technology* 55 (2012), Nr. 4, S. 221–235. – URL <http://www.sciencedirect.com/science/article/pii/S1350449512000308>. – ISSN 1350-4495
- [Lowe 2004] LOWE, David G.: Distinctive image features from scale-invariant keypoints. In: *International journal of computer vision* 60 (2004), Nr. 2, S. 91–110. – URL <http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94>
- [Lynch 2006] LYNCH, Patrick J.: *Head anatomy anterior view*. 2006. – CC BY 2.5
- [Ma u. a. 2015] MA, Jiayi ; ZHAO, Ji ; MA, Yong ; TIAN, Jinwen: Non-rigid visible and infrared face registration via regularized Gaussian fields criterion. In: *Pattern Recognition* 48 (2015), Nr. 3, S. 772–784. – ISSN 0031-3203
- [Marr und Hildreth 1980] MARR, David ; HILDRETH, Ellen: Theory of edge detection. In: *Proceedings of the Royal Society of London B: Biological Sciences* 207 (1980), Nr. 1167, S. 187–217
- [Matas u. a. 2004] MATAS, Jiri ; CHUM, Ondrej ; URBAN, Martin ; PAJDLA, Tomáš: Robust wide-baseline stereo from maximally stable extremal regions. In: *Image and vision computing*

- 22 (2004), Nr. 10, S. 761–767. – URL <http://dx.doi.org/10.1016/j.imavis.2004.02.006>
- [MathWorks 2017a] MATHWORKS: *Computer Vision System Toolbox™ User's Guide*. 3 Apple Hill Drive, Natick, MA 01760-2098, USA: The MathWorks, Inc. (Veranst.), 2017. – URL https://mathworks.com/help/pdf_doc/vision/vision_ug.pdf. – Handbuch
- [MathWorks 2017b] MATHWORKS: *Image Processing Toolbox™ User's Guide*. 3 Apple Hill Drive, Natick, MA 01760-2098, USA: The MathWorks, Inc. (Veranst.), 2017. – URL https://mathworks.com/help/pdf_doc/images/images_tb.pdf. – Handbuch
- [Microsoft 2017] MICROSOFT: Kinect-Hardware. One Microsoft Way Redmond, WA 98052-6399, USA, 2017. – Forschungsbericht. – URL <https://developer.microsoft.com/windows/kinect/hardware>. Dokumentation
- [Müller u. a. 2016] MÜLLER, L. ; BERNIN, A. ; GHOSE, S. ; GOZDZIELEWSKI, W. ; WANG, Qi ; GRECOS, C. ; LUCK, K. von ; VOGT, F.: Physiological data analysis for an emotional provoking exergame. In: *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, URL <http://dx.doi.org/10.1109/SSCI.2016.7850042>, Dec 2016, S. 1–8
- [Müller u. a. 2015] MÜLLER, Larissa ; ZAGARIA, Sebastian ; BERNIN, Arne ; AMIRA, Abbes ; RAMZAN, Naeem ; GRECOS, Christos ; VOGT, Florian: *EmotionBike: A Study of Provoking Emotions in Cycling Exergames*. S. 155–168. In: CHORIANOPOULOS, Konstantinos (Hrsg.) ; DIVITINI, Monica (Hrsg.) ; BAALSRUD HAUGE, Jannicke (Hrsg.) ; JACCHERI, Letizia (Hrsg.) ; MALAKA, Rainer (Hrsg.): *Entertainment Computing - ICEC 2015: 14th International Conference, ICEC 2015, Trondheim, Norway, September 29 - October 2, 2015, Proceedings*. Cham : Springer International Publishing, 2015. – URL http://dx.doi.org/10.1007/978-3-319-24589-8_12. – ISBN 978-3-319-24589-8
- [NIST 2000–2017] NIST: Face Recognition Vendor Test (FRVT). In: *National Institute of Standards and Technology (NIST) (2000-2017)*. – URL <https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt>
- [Ó Conaire u. a. 2005] Ó CONAIRE, Ciarán ; COOKE, Eddie ; O'CONNOR, Noel E. ; MURPHY, Noel ; SMEATON, Alan F.: Fusion of infrared and visible spectrum video for indoor surveillance. (2005)

- [Olesch 2014] OLESCH, Janine: *Grundlagen der Bildregistrierung*. S. 73–142. In: *Bildregistrierung für die navigierte Chirurgie: Spezialisierte Ansätze zur Anwendung in der navigierten Leberchirurgie*. Wiesbaden : Springer Fachmedien Wiesbaden, 2014. – URL https://doi.org/10.1007/978-3-658-05655-1_5. – ISBN 978-3-658-05655-1
- [Picard 1999] PICARD, Rosalind W.: Affective Computing for HCI. In: *HCI (1)*, 1999, S. 829–833
- [Reese u. a. 2012] REESE, Kristopher ; ZHENG, Yufeng ; ELMAGHRABY, Adel: A comparison of face detection algorithms in visible and thermal spectrums. In: *Int'l Conf. on Advances in Computer Science and Application*, 2012
- [Rowley u. a. 1998] ROWLEY, H. A. ; BALUJA, S. ; KANADE, T.: Neural network-based face detection. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (1998), Jan, Nr. 1, S. 23–38. – URL <https://doi.org/10.1109/34.655647>. – ISSN 0162-8828
- [Rudol und Doherty 2008] RUDOL, P. ; DOHERTY, P.: Human Body Detection and Geolocalization for UAV Search and Rescue Missions Using Color and Thermal Imagery. In: *2008 IEEE Aerospace Conference*, URL <http://doi.acm.org/10.1109/AERO.2008.4526559>, March 2008, S. 1–8. – ISSN 1095-323X
- [Saponaro u. a. 2015] SAPONARO, P. ; SORENSEN, S. ; RHEIN, S. ; KAMBHAMETTU, C.: Improving calibration of thermal stereo cameras using heated calibration board. In: *2015 IEEE International Conference on Image Processing (ICIP)*, URL <http://doi.org/10.1109/ICIP.2015.7351702>, Sept 2015, S. 4718–4722
- [Schuster und Kolobrodov 2004] SCHUSTER, Norbert ; KOLOBRODOV, Valentin G.: *Infrarotthermographie*. John Wiley & Sons, 2004. – URL <http://doi.org/10.1002/9783527624645>
- [SFB 62 2009] SFB 62, Sonderforschungsbereich Transregio 6.: *Sonderforschungsbereich zur Innovation im Mensch-Technik Dialog: Companion-Systeme*. 2009. – URL <http://www.sfb-trr-62.de>
- [Shimrat 1962] SHIMRAT, M.: Algorithm 112: Position of Point Relative to Polygon. In: *Commun. ACM* 5 (1962), Aug, Nr. 8. – URL <http://doi.acm.org/10.1145/368637.368653>. – ISSN 0001-0782
- [Singh und Om 2013] SINGH, R. ; OM, H.: An overview of face recognition in an unconstrained environment. In: *2013 IEEE Second International Conference on Image Informa-*

- tion Processing (ICIIP-2013), URL <http://dx.doi.org/10.1109/ICIIP.2013.6707679>, Dec 2013, S. 672–677
- [Socolinsky und Selinger 2004] SOCOLINSKY, D. A. ; SELINGER, A.: Thermal face recognition in an operational scenario. In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004*. Bd. 2, URL <http://dx.doi.org/10.1109/CVPR.2004.1315275>, June 2004, S. II–1012–II–1019 Vol.2. – ISSN 1063-6919
- [Soldan 2014] SOLDAN, S.: *Sensordatenfusionsansätze in der Thermografie zur Verbesserung der Messergebnisse*, Dissertation, 2014. – URL <http://d-nb.info/1057849677>
- [Spyra 2010] SPYRA, Peer S.: *Einsatz der kontaktfreien Infrarotthermographie zum Nachweis von Veränderungen der Hauttemperatur bei Interventioneller Schmerztherapie am sympathischen Nervensystem*. Von-Melle-Park 3, 20146 Hamburg, Universität Hamburg, Dissertation, 2010. – URL <http://ediss.sub.uni-hamburg.de/volltexte/2010/4916>
- [Stathaki 2011] STATHAKI, Tania: *Image fusion: algorithms and applications*. Academic Press, 2011. – URL <http://doi.org/10.5772/691>
- [Stubing 2010] STUBING, Sascha: *Entwicklung eines Verfahrens zur Lagekorrektur von Kalibrierungsmustern in Bildsequenzen*. 2010. – URL https://www2.informatik.hu-berlin.de/~stuebing/public/Studienarbeit_Sascha_Stuebing_2010.pdf. – Studienarbeit
- [Süße und Rodner 2014] SÜSSE, Herbert ; RODNER, Erik: *Kamerakalibrierung*. S. 339–361. In: *Bildverarbeitung und Objekterkennung: Computer Vision in Industrie und Medizin*. Wiesbaden : Springer Fachmedien Wiesbaden, 2014. – URL http://dx.doi.org/10.1007/978-3-8348-2606-0_15. – ISBN 978-3-8348-2606-0
- [Szeliski 2011] SZELISKI, Richard: *Computer vision: algorithms and applications*. Springer Science & Business Media, 2011. – URL <http://dx.doi.org/10.1007/978-1-84882-935-0>
- [Tan u. a. 2006] TAN, Xiaoyang ; CHEN, Songcan ; ZHOU, Zhi-Hua ; ZHANG, Fuyan: Face recognition from a single image per person: A survey. In: *Pattern Recognition* 39 (2006), Nr. 9, S. 1725–1745. – URL <https://doi.org/10.1016/j.patcog.2006.03.013>. – ISSN 0031-3203

- [Torresan u. a. 2004] TORRESAN, Helene ; TURGEON, Benoit ; IBARRA-CASTANEDO, Clemente ; HEBERT, Patrick ; MALDAGUE, Xavier P.: *Advanced surveillance systems: combining video and thermal imagery for pedestrian detection*. 2004. – URL <http://dx.doi.org/10.1117/12.548359>
- [Ursine u. a. 2012] URSINE, W ; CALADO, F ; TEIXEIRA, G ; DINIZ, H ; SILVINO, S ; ANDRADE, R de: Thermal/visible autonomous stereo visio system calibration methodology for non-controlled environments. In: *11th International Conference on Quantitative Infrared Thermography*, URL <http://doi.org/10.21611/qirt.2012.261>, 2012
- [Vijayakumari 2013] VIJAYAKUMARI, V: Face recognition techniques: A survey. In: *World journal of computer application and technology* 1 (2013), Nr. 2, S. 41–50. – URL <http://dx.doi.org/10.3745/JIPS.2009.5.2.041>
- [Viola und Jones 2001] VIOLA, Paul ; JONES, Michael: Robust Real-time Object Detection. In: *International Journal of Computer Vision*, URL <https://doi.org/10.1023/b:visi.0000013087.49260.fb>, 2001
- [Wilder u. a. 1996] WILDER, J. ; PHILLIPS, P. J. ; JIANG, Cunhong ; WIENER, S.: Comparison of Visible and Infra-red Imagery for Face Recognition. In: *Proceedings of the 2Nd International Conference on Automatic Face and Gesture Recognition (FG '96)*. Washington, DC, USA : IEEE Computer Society, 1996 (FG '96), S. 182–187. – URL <http://dx.doi.org/10.1109/AFGR.1996.557262>. – ISBN 0-8186-7713-9
- [Xia und Kamel 2007] XIA, Y. ; KAMEL, M. S.: Novel Cooperative Neural Fusion Algorithms for Image Restoration and Image Fusion. In: *IEEE Transactions on Image Processing* 16 (2007), Feb, Nr. 2, S. 367–381. – URL <http://dx.doi.org/10.1109/TIP.2006.888340>. – ISSN 1057-7149
- [Xiaoyu u. a. 2006] XIAOYU, W. ; JIHONG, C. ; PINGJIANG, W. ; ZHIHONG, H.: Infrared Human Face Auto Locating Based on SVM and A Smart Thermal Biometrics System. In: *Sixth International Conference on Intelligent Systems Design and Applications* Bd. 2, URL <http://dx.doi.org/10.1109/ISDA.2006.253759>, Oct 2006, S. 1066–1072. – ISSN 2164-7143
- [Yow und Cipolla 1997] YOW, Kin C. ; CIPOLLA, Roberto: Feature-based human face detection. In: *Image and vision computing* 15 (1997), Nr. 9, S. 713–735. – URL [https://doi.org/10.1016/S0262-8856\(97\)00003-6](https://doi.org/10.1016/S0262-8856(97)00003-6)

- [Yufeng 2011] YUFENG, Zheng: *Image fusion and its applications*. 2011
- [Zhao u. a. 2003] ZHAO, Wenyi ; CHELLAPPA, Rama ; PHILLIPS, P J. ; ROSENFELD, Azriel: Face recognition: A literature survey. In: *ACM computing surveys (CSUR)* 35 (2003), Nr. 4, S. 399–458. – URL <http://dx.doi.org/10.1145/954339.954342>
- [Zimmermann 2012] ZIMMERMANN, Thomas ; ARRAY (Hrsg.): *Lehrbuch der Infrarotthermografie : allgemeine Grundlagen der Thermodynamik, Grundlagen der Strahlungsphysik, Infrarot-Geräte-Technologie (für normative Stufe 1 und 2)*. Stuttgart : Fraunhofer IRB-Verl., 2012. – 163–164 S. – URL <http://d-nb.info/1025946545/04>
- [Zitova und Flusser 2003] ZITOVA, Barbara ; FLUSSER, Jan: Image registration methods: a survey. In: *Image and vision computing* 21 (2003), Nr. 11, S. 977–1000. – URL [http://dx.doi.org/10.1016/S0262-8856\(03\)00137-9](http://dx.doi.org/10.1016/S0262-8856(03)00137-9)

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne fremde Hilfe selbständig verfasst und nur die angegebenen Hilfsmittel benutzt habe.

Hamburg, 26. September 2017

Florian Kletz