Multimodale Registrierung und Analyse von Infrarot- und Tiefenkameradaten

Multimodal registration and analysis of infrared and depth camera data Bachelorarbeit von Larissa Yasmin Werner Tag der Einreichung: 04. Oktober 2022

1. Gutachten: Prof. Dr.-Ing. Christoph Hoog Antink, KIS*MED

2. Gutachten: Maurice Rohr, M.Sc., KIS*MED





TECHNISCHE UNIVERSITÄT DARMSTADT

Fachbereich Elektrotechnik und Informationstechnik KIS*MED - Künstlich Intelligente Systeme der Medizin

Studiengang Medizintechnik

Erklärung zur Abschlussarbeit gemäß § 22 Abs. 7 APB TU Darmstadt

Hiermit erkläre ich, Larissa Yasmin Werner, dass ich die vorliegende Arbeit gemäß § 22 Abs. 7 APB der TU Darmstadt selbstständig, ohne Hilfe Dritter und nur mit den angegebenen Quellen und Hilfsmitteln angefertigt habe. Ich habe mit Ausnahme der zitierten Literatur und anderer in der Arbeit genannter Quellen keine fremden Hilfsmittel benutzt. Die von mir bei der Anfertigung dieser wissenschaftlichen Arbeit wörtlich oder inhaltlich benutzte Literatur und alle anderen Quellen habe ich im Text deutlich gekennzeichnet und gesondert aufgeführt. Dies gilt auch für Quellen oder Hilfsmittel aus dem Internet.

Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Mir ist bekannt, dass im Falle eines Plagiats (§ 38 Abs. 2 APB) ein Täuschungsversuch vorliegt, der dazu führt, dass die Arbeit mit 5,0 bewertet und damit ein Prüfungsversuch verbraucht wird. Abschlussarbeiten dürfen nur einmal wiederholt werden.

Bei einer Thesis des Fachbereichs Architektur entspricht die eingereichte elektronische Fassung dem vorgestellten Modell und den vorgelegten Plänen.

Darmstadt, 04. Oktober 2022

L. Y. Werner

Inhaltsverzeichnis

1	Einle 1.1 1.2	eitung Motivation	8 8 9
	1.5		10
2	The	oretische Grundlagen	11
	2.1	Bildgebende Verfahren	11
		2.1.1 Thermographie	11
		2.1.2 Tiefenkamera	12
	2.2	Bildregistrierung	13
		2.2.1 Registrierungskriterien	13
		2.2.2 Registrierungsverfahren	18
3	Verv	vandte Arbeiten	25
	3.1	Morphologischer Gradient	25
	3.2	Histogram Matching	27
4	Met	hodik	29
-	4 1	Verwendete Daten	29
	4.2	Programmiersprache und verwendete Bibliotheken	31
	4.3	Tiefenbilder	31
	110	4.3.1 Lochkamera Modell	32
		4.3.2 Algorithmus zur Erzeugung künstlicher Tiefenbilder	35
	4.4	Erzeugung von Thermalbildern aus RGB-Videoframes	40
	4.5	Vorverarbeitung der Bilder	42
		4.5.1 Konvertierung zu 8-bit Bildern	42
		4.5.2 Invertierung	42
		4.5.3 Binärbilder	43
		4.5.4 Morphologische Filter	46

7	Fazi	t		94		
	6.3	Intensi	itätsbasierte Registrierung mit Mutual Information	90		
	6.2	Merkn	nalsbasierte Registrierung mit SIFT	89		
	6.1	Gemei	nsamkeiten in den multimodalen Daten	86		
6	Disk	ussion		86		
		5.3.3	vergieich der Transformationsmatrizen	80		
		5.3.2	Anniichkeit der registrierten Bilder	76		
		5.3.1	Visuelle Bewertung	73		
	5.3	Intensi	ItatsDasierte Kegistrierung	73		
	F 0	5.2.2	Registrierung mit SIFT in Bildausschnitten	71		
		5.2.1	Globale Registrierung mit SIFT	69		
	5.2	Merkn	alsbasierte Registrierung	69		
		5.1.3	Statistische Merkmale	67		
		5.1.2	Korrelation in 2D-Histogrammen	63		
		5.1.1	Oberflächendarstellung	61		
	5.1	Zusam	menhang zwischen Tiefen- und Temperaturwerten	61		
5	Ergebnisse					
		4.8.2	Bewertung anhand einer Referenztransformationsmatrix	59		
		4.8.1	Bewertung anhand von Ähnlichkeitsmetriken	59		
	4.8	Bewer	tung der Registrierung	58		
		4.7.3	Intensitätsbasierte Registrierung	58		
		4.7.2	Merkmalsbasierte Registrierung	57		
		4.7.1	Manuelle Registrierung	54		
	4.7	Registi	rierungsverfahren	54		
	4.6	Zusam	menhang zwischen Tiefen- und Temperaturwerten	52		
		4.5.7	Intensitätstransformationen	49		
		4.5.6	Adaptive Equalization	47		
		4.5.5	Histogram Matching	47		

Abbildungsverzeichnis

2.1	Beispiele von 2D-Transformationen	16
4.1	Ablauf der Arbeit	30
4.2	Modell einer Lochkamera	32
4.3	Darstellung eines 3D-Modells in Meshlab	36
4.4	Perspektive der Tiefenbilder mit unterschiedlicher Entfernung der Kamera	38
4.5	Abtastung eines Dreiecks	39
4.6	Erzeugen von Thermal Grauwertbildern	41
4.7	Vergleich der Histogramme von Original- und konvertierten Bildern	43
4.8	Invertierung der Bilder	44
4.9	Bilder nach globalem Schwellenwert	45
4.10	Bilder nach adaptivem lokalen Schwellenwert	46
4.11	Morphologische Gradientenbilder	47
4.12	Bilder nach Histogram Matching	48
4.13	Bilder nach CLAHE	49
4.14	Intensitätstransformierte Bilder	51
4.15	Auswahltool für Landmarke	55
4.16	Manuell registrierte Bilder	56
4.17	Lage der gewählten Punkte für den Vergleich der Matrizen im Bild	60
5.1	3D-Darstellung der Intensitätswerte	62
5.2	2D-Histogramme der Intensitätswerte der gesamten Bilder	64
5.3	Bildausschnitte von Augen- und Nasenpartie	65
5.4	2D-Histogramme der Intensitätswerte der Bildausschnitte	65
5.4	2D-Histogramme der Intensitätswerte der Bildausschnitte	66
5.5	Statistische Merkmale von Patches	68
5.6	Ahnlichkeit von Formen in Patches	68
5.7	Beispiele einer globalen Registrierung mit SIFT	69
5.7	Beispiele einer globalen Registrierung mit SIFT	70

5.8	Beispiele einer lokalen Registrierung mit SIFT		71
5.8	Beispiele einer lokalen Registrierung mit SIFT		72
5.9	Beispiele einer intensitätsbasierten Registrierung	5	73
5.10	Vergleich der intensitätsbasierten Registrierung mit morphologischen Gra-		
	dientenbildern und adaptiven Schwellenwertbildern	. 7	74
5.11	Beispiele für eine präzise intensitätsbasierte Registrierung mit morphologi-		
	schen Gradientenbildern		75
5.12	Beispiele für eine unpräzise intensitätsbasierte Registrierung mit morpho-		
	logischen Gradientenbildern	. 7	76
5.13	Ähnlichkeit der transformierten Bilder mit NMI	. 7	78
5.13	Ähnlichkeit der transformierten Bilder mit NMI	. 7	79
5.14	Bewertung der Registrierung mit der manuellen Referenztransformations-		
	matrix	. 8	81
5.14	Bewertung der Registrierung mit der manuellen Referenztransformations-		
	matrix	. 8	82
5.15	Vergleich der mittleren quadratischen Distanzen der intensitätsbasierten		
	Registrierung	. 8	83
5.16	Vergleich von manueller Registrierung und intensitätsbasierter Registrie-		
	rung mit morphologischen Gradientenbildern	. 8	84
5.16	Vergleich von manueller Registrierung und intensitätsbasierter Registrie-		
	rung mit morphologischen Gradientenbildern	. 8	85

Tabellenverzeichnis

4.1	Bearbeitungsschritte der Intensitätstransformationen	50
5.1	Mittelwerte der Ähnlichkeitswerte der transformierten Bilder	77
5.2	Mittelwert und Standardabweichung der mittleren quadratischen Distan-	
	zen aus dem Vergleich mit der manuellen Referenzmatrix	83

1 Einleitung

1.1 Motivation

Mit Hilfe bildgebender Verfahren ist es heutzutage möglich, Krankheiten bei Patient*innen zu diagnostizieren. Einige bildgebende Verfahren sind aus der medizinischen Diagnostik oder Therapie nicht mehr wegzudenken. Bei kontaktlosen Diagnoseverfahren werden keine Sensoren an die Patient*innen angebracht, wodurch die Wahrscheinlichkeit eines Ausschlags oder einer Entzündung reduziert werden kann [1]. Um kontaktlose Verfahren zu testen und eine Einschätzung über die Wirksamkeit dieser geben zu können, müssen typischerweise große Datenmengen analysiert werden. Die zu untersuchenden Datensätze werden dabei häufig nicht selbst erzeugt, sondern es werden bereits vorhandene Datensätze verwendet, um neue Algorithmen zu testen und zu validieren. Bei solchen nicht selbst aufgenommenen Datensätzen hat man keinen Einfluss darauf, wie die Daten strukturiert sind. Ein Vergleich von Algorithmen zur kontaktlosen Vitalparameteruntersuchung mit Hilfe von vorhandenen Datensätzen ist wünschenswert. Diese Arbeit soll eine Grundlage für einen solchen Vergleich schaffen.

Vitalparameter sind wichtig für ein frühzeitiges Erkennen einer Veränderung des Gesundheitszustandes eines*einer Patient*in. Zu den Vitalparametern zählen Herzfrequenz, Blutdruck, Atemfrequenz und Körpertemperatur [2]. Um die Vitalparameter zu messen, werden meistens Sensoren verwendet, welche einen direkten Körperkontakt zu den Patient*innen benötigen [2]. Ein kontaktlose Verfahren kann die Messung für den*die Patient*in angenehmer gestalten und ist zudem auch hygienischer.

In dieser Arbeit wird die Verarbeitung von Daten zum Vergleich zweier Algorithmen zur kontaktlosen Messung der Atemfrequenz thematisiert. Die Atemfrequenz ist ein wichtiger Vitalparameter, der den Gesundheitszustand der Lunge und des Herzens beschreibt [3]. Hierbei wird ein Verfahren, welches mit einer Infrarotkamera arbeitet, und ein weiteres, das die Atemfrequenz durch eine Tiefenkamera ermittelt, verwendet. Es gibt verschiedene

Möglichkeiten, die Atemfrequenz mittels einer Infrarotkamera zu bestimmen. [4] beschreibt einen Ansatz für die kontaktlose Messung der Atemfrequenz. Dazu werden kleine Temperaturänderungen im Gesicht untersucht. Diese Temperaturänderungen treten im Bereich der Nase und des Munds auf. Eine Inspiration von kalter Luft aus der Umgebung sowie eine Exspiration von warmer Luft aus den Lungen kann beobachtet werden. Dadurch kann in bestimmten Atemintervallen die Atemfrequenz ermittelt werden. In [2] wird ein Deep-Learning-basierter Ansatz beschrieben, mit dem die Atemfrequenz mittels einer Infrarot-Thermographie bestimmt werden kann. Dafür wird mittels Objekterkennung die Brustregion der Versuchsperson detektiert. Der Algorithmus nutzt aus, dass durch die Atmung Brustkorbbewegungen hervorgerufen werden. Diese können mit einem Optical-Flow-Algorithmus bestimmt werden, woraus schließlich die Atemfrequenz berechnet wird. Sowohl Tiefenkamera als auch Infrarot-Thermographie können dazu verwendet werden, die durch die Atmung induzierten Brustkorbbewegungen aufzunehmen, um daraus die Atemfrequenz zu ermitteln. Um sowohl die Temperaturwerte der Infrarotkamera als auch die Tiefenwerte der Tiefenkamera in einem Bild zu erfassen, müssen die Bilder fusioniert werden. Jedem Temperaturpunkt soll also ein Tiefenwert zugeordnet werden. Dafür ist der erste Schritt eine Registrierung der beiden Bilddaten.

Diese Arbeit wird sich daher mit der Registrierung zwischen Tiefenkamera- und Thermaldaten befassen. Bei der Bildregistrierung handelt es sich um den Prozess des Überlagerns von Bildern, die zu unterschiedlichen Zeiten, von unterschiedlichen Betrachterpunkten oder von unterschiedlichen Verfahren aufgenommen werden [5]. Bei einer erfolgreichen Registrierung können Erkenntnisse darüber gewonnen werden, welche Punkte in beiden Bildern übereinstimmen. Da Tiefen- und Thermalbilder von unterschiedlicher Bildmodalität sind, wird die Registrierung als multimodal bezeichnet.

1.2 Zielsetzung

Ziel dieser Arbeit ist es, Tiefen- und Thermalbilder multimodal zu registrieren. Dabei soll zum einen der Zusammenhang zwischen Tiefen- und Temperaturwerten untersucht und zum anderen verschiedene Registrierungsverfahren getestet werden. Daraus soll eine Möglichkeit gefunden werden, mit der die Daten registriert werden können. Dieses Verfahren kann damit allein aufgrund der Daten und ohne extern angebrachtes künstliches Objekt Tiefen- und Thermalbilder aufeinander abbilden, sodass bekannt ist, welcher Temperaturwert welchem Tiefenwert entspricht. Wenn ein Verfahren für die Registrierung von Tiefenund Thermalbildern gefunden wird, kann dieses für den Vergleich der Algorithmen zur Respirationsmessung mit Kameras weiter getestet werden.

Die Arbeit konzentriert sich auf zwei Ansätze. Dazu zählt die Analyse der Daten für eine mögliche Registrierung anhand von statistischen Merkmalen und das Testen von verschiedenen Registrierungsverfahren.

Bei der Analyse der Daten soll ein Zusammenhang von Tiefen- und Temperaturwerten untersucht werden. Dafür werden statistische Merkmale in kleinen Bereichen der Bilder betrachtet, um darüber eine Zuordnung der zugehörigen Punkte in den Bildern zu erlangen. Dafür werden die Bilder mittels eines Gitters in kleine Bereiche, sogenannte "Patches", eingeteilt und jedem Patch werden statistische Merkmale zugewiesen. Falls es eine Gemeinsamkeit in den statistischen Merkmalen von Tiefen- und Thermalbildern gibt, sollen diese anschließend einander zugeordnet werden. Dadurch können die einzelnen Bereiche so geometrisch verschoben werden, dass die zugehörigen Bereiche übereinander liegen. Die Registrierung besteht in diesem Fall dann aus mehreren einzelnen Transformationsmatrizen.

Außerdem werden verschiedene bestehende Registrierungsverfahren getestet. Damit soll überprüft werden, ob es Registrierungsverfahren gibt, die die Daten registrieren können. Außerdem soll ein Vergleich dieser Registrierungsverfahren vorgenommen werden. Dabei werden die Bilder vor der Registrierung auf verschiedene Weise mit Filtern und Intensitätstransformationen bearbeitet, um markante Merkmale in den Bildern hervorzuheben. Anschließend soll untersucht werden, ob eine solche Vorverarbeitung der Bilder zu besseren Registrierungsergebnissen führt.

1.3 Aufbau der Arbeit

Zu Beginn der Arbeit werden die theoretischen Grundlagen der bildgebenden Verfahren von Thermographie und Tiefenkamera sowie die allgemeinen Grundlagen der Bildregistrierung beschrieben. Die grundlegenden Begriffe der Registrierung und verschiedene Registrierungsansätze werden vorgestellt. Im anschließenden Kapitel wird auf zwei konkrete Verfahren eingegangen, die ein ähnliches Vorgehen wie in dieser Arbeit beschreiben. Danach werden die verwendeten Methoden der Arbeit beschrieben. Es wird erklärt, wie das Vorgehen umgesetzt wird und wie die Ergebnisse bewertet werden. Daran anschließend werden im nächsten Kapitel die Ergebnisse vorgestellt. Diese werden dann in Kapitel 6 interpretiert. Zum Schluss folgen eine kurze Zusammenfassung und ein Ausblick.

2 Theoretische Grundlagen

2.1 Bildgebende Verfahren

Für einen kurzen Überblick der zu registrierenden Daten werden in diesem Abschnitt die grobe Funktionsweise einer Thermographie- und einer Tiefenkamera vorgestellt.

2.1.1 Thermographie

Die Thermographie ist ein bildgebendes Verfahren, mit dem die Temperaturverteilungen an Oberflächen dargestellt werden [6]. Das Verfahren arbeitet kontaktlos und nicht-invasiv.

Jeder Körper, der eine Temperatur oberhalb des absoluten Nullpunktes hat, strahlt Wärme ab. Ein Körper strahlt am meisten ab, wenn er alle auftreffende Strahlung absorbiert. Ein solcher Körper wird als "schwarzer Körper" bezeichnet. Der Absorptionsgrad wird wie folgt bestimmt [6]:

$$\alpha = \frac{\Phi_A}{\Phi_0} = \frac{\text{absorbierte Strahlungsleistung}}{\text{auftreffende Strahlungsleistung}}$$
(2.1)

Ein schwarzer Körper hat damit einen Absorptionsgrad von $\alpha = 1$ [6].

Dössel trifft folgende Annahme für einen menschlichen Körper: "Die Haut verhält sich im IR näherungsweise wie ein "Schwarzer Körper", d.h. $\alpha \approx 0, 98$." [6]. Das Spektrum der Strahlungsdichte der Haut ändert sich mit der Temperatur des Körpers. Die Temperatur der Haut kann daher durch die Messung der abgestrahlten Infrarot-Leistung insgesamt oder in einem ausgewählten Infrarot-Spektralbereich bestimmt werden. Die von den Organen des Menschen produzierte Wärme wird über Blut und Wärmeleistung abtransportiert.

Wenn die Wärmestrahlung gemessen werden soll, sind Luftbewegungen, d.h. Konvektion, zu vermeiden, um das Ergebnis der Messung möglichst wenig zu beeinflussen [6].

Für die Thermographie werden Thermographiekameras verwendet, welche die Wärmestrahlung von Körpern erfassen und dadurch die Temperaturverteilung an der Körperoberfläche visualisieren. Die Wärmestrahlung äußert sich in Form von Infrarot-Strahlung und ist mit dem menschlichen Auge nicht sichtbar [7]. Je nach Temperatur eines Körpers wird die Wärmestrahlung mit unterschiedlicher Wellenlänge innerhalb des Infrarot-Bereichs ausgestrahlt. Diese Infrarotstrahlung wird von Thermographiekameras detektiert und visualisiert [7]. Die entstehenden Thermalbilder können als Grauwertbilder dargestellt werden, wobei jeder Grauwert einem Temperaturwert entspricht. Zur besseren Visualisierung erfolgt die Darstellung häufig mit "Falschfarben-Bildern", um kleine Temperaturunterschiede im Bild hervorheben zu können.

2.1.2 Tiefenkamera

Eine Tiefenkamera misst die Entfernung von Objekten zur Kamera und speichert diese in einem Bild [8]. Jedem Pixel in dem aufgenommenen Tiefenbild ist ein Wert zugeordnet, der diese Distanz repräsentiert. Diese Entfernung zur Kamera wird als Tiefe interpretiert.

Es gibt verschiedene Tiefenkameras, die mit unterschiedlichen Verfahren die Tiefe aufzeichnen.

Tiefe kann aufgezeichnet werden, indem strukturiertes Licht von einer Quelle auf eine Szene projiziert wird. Das ausgestrahlte Lichtmuster ist dabei bekannt. Das Lichtmuster ändert sich durch Objekte mit einer Tiefe in der Szene. Diese Änderung des Lichtmusters durch die Szene wird analysiert. Daraus kann schließlich die Tiefe berechnet werden [8].

Eine weitere Möglichkeit Tiefe aufzunehmen sind Stereo-Tiefenkameras, welche mit zwei nah aneinanderliegenden Sensoren arbeiten. Bekannt ist hierbei die Distanz zwischen den Sensoren. Beide Sensoren nehmen visuelle Merkmale auf. Aus der bekannten Distanz zwischen den Sensoren kann die Tiefeninformation gewonnen werden. Das Verfahren arbeitet ähnlich wie das menschliche Tiefenempfinden mit zwei Augen [8].

Des Weiteren gibt es Tiefenkameras, bei welchen die bekannte Variable die Geschwindigkeit des Lichts ist. Die Szene wird mit Licht bestrahlt und es wird gemessen, wie lang das Licht braucht, um wieder zurück zum Sensor der Kamera zu gelangen. Aus dieser Zeit wird die Tiefeninformation gewonnen [8].

2.2 Bildregistrierung

Eine Bildregistrierung beschreibt den Prozess des Überlagerns von zwei oder mehreren Bildern, welche dasselbe Objekt oder dieselbe Szene abbilden, sodass diese Bilder passgenau aufeinander liegen. Es wird also versucht, die Bilder möglichst exakt aufeinander abzubilden. Dazu muss eine geometrische Transformation gefunden werden, welche Punkte in dem einen Bild Punkten in dem anderen Bild zuordnet [9]. Im Folgenden werden die notwendigen Grundbegriffe und Definitionen, welche zum Verständnis der vorliegenden Arbeit in Bezug auf die Bildregistrierung benötigt werden, kurz erklärt.

2.2.1 Registrierungskriterien

Die folgende Einteilung der Registrierungskriterien basiert hauptsächlich auf den Klassifikationskriterien für Registrierungsmethoden von Maintz und Viergever [10].

Registrierungsdimensionalität

Die Registrierungsdimensionalität gibt an, welche Dimensionen die zu registrierenden Bilder haben. Werden nur geometrische Dimensionen betrachtet, kann zwischen 2D/2D, 2D/3D und 3D/3D unterschieden werden [10].

Eine 3D/3D-Registrierung beschäftigt sich mit der Registrierung von Volumen. Es werden also ganze dreidimensionale Datensätze für die Registrierung verwendet. Eine 2D/2D-Registrierung liegt vor, wenn entweder 2D-Bildmodalitäten oder separate Schichtbilder eines Volumens registriert werden sollen. Eine Registrierung zwischen 2D/2D oftmals einfacher und schneller zu berechnen als eine Registrierung im Fall einer 3D/3D-Registrierung. Gründe hierfür sind die geringere Anzahl an Parametern und die wesentlich kleinere Menge an Daten, die verwendet werden. Eine 2D/3D-Registrierung wird benötigt, wenn räumliche Daten direkt projektiven Daten oder einzelne Schichtbilder räumlichen Daten angeglichen werden sollen [10]. In der vorliegenden Arbeit werden aus 3D-Modellen 2D-Tiefenbilder erzeugt und mit 2D-Thermographiebildern registriert. Aus einer 2D/3D-Registrierung wird damit eine 2D/2D-Registrierung.

Basis der Registrierung

Nach [10] werden Registrierungsverfahren in extrinsische und intrinsische Verfahren unterteilt. Bei extrinsischen Verfahren wird ein künstliches Objekt in die Szene integriert. Künstliche Objekte können dabei invasiv in den*die Patient*in oder auch nicht-invasiv in die Szene/an den*die Patient*in angebracht werden. Solche Objekte können beispielsweise bestimmte Schrauben sein. Diese künstlichen Objekte können von den Modalitäten gut erfasst und detektiert werden und dienen daher als Marker. Anhand der Objekte können die Kamerasysteme kalibriert werden. Es wird eine Transformation gefunden, welche das eine Kamerasystem in das andere transformiert. Damit ist die Registrierung vergleichsweise einfach und schnell und kann meistens automatisiert werden [10]. Intrinsische Registrierungsverfahren arbeiten ausschließlich mit den aufgezeichneten Bildinformationen, welche von dem*der Patient*in generiert werden. Die intrinsische Registrierung kann mit bestimmten Merkmalen, auch Landmarken genannt, arbeiten, Segmentierungsergebnisse verwenden oder auf Pixelintensitäten beruhen. Aufgrund der Datengrundlage wird für diese Arbeit eine intrinsische Registrierung gewählt.

Art der Transformation

Es wird zwischen Bewegtbild und Referenzbild unterschieden. Das Bewegtbild soll im Registrierungsverfahren dem Referenzbild angepasst werden [10]. Um die Bilder zu registrieren, wird eine Funktion benötigt, welche das Bewegtbild in das Koordinatensystem des Referenzbildes abbildet. Diese Funktion wird auch als "Abbildungsfunktion" bezeichnet und ist für die Transformation des Bewegtbildes zuständig [5]. Es gibt verschiedene Transformationsmodelle, welche sich in der Anzahl ihrer Freiheitsgrade unterscheiden.

Die einfachste Transformationsart ist eine rigide Transformation. Bei dieser Transformation sind nur Translationen und Rotationen erlaubt [10]. [5] führt eine zusätzliche Art von Transformationen ein. Diese ist eine Ähnlichkeitstransformation, welche einer rigiden Transformationen mit einem zusätzlichen Skalierungsfaktor entspricht. Eine rigide Transformationen hat drei Freiheitsgrade, eine Ähnlichkeitstransformation vier. Diese beiden Transformationen können mittels zweier Punkte [u, v] und [x, y] berechnet werden. Dabei entsprechen u und v Punkten im Bewegtbild und x und y Punkten im Referenzbild. Nach [5] lassen sich die Punkte im Bewegtbild wie folgt aus den Punkten im Referenzbild berechnen:

$$u = s(x\cos(\varphi) - y\sin(\varphi)) + t_x \tag{2.2}$$

$$v = s(x\sin(\varphi) + y\cos(\varphi)) + t_y \tag{2.3}$$

Dabei steht *s* für den Skalierungsfaktor einer Ähnlichkeitstransformation. Bei einer rigiden Transformation gilt s = 1. φ beschreibt den Rotationswinkel und t_x beziehungsweise t_y stehen für eine Translation in x- und y-Richtung.

Wenn durch eine Transformation parallele Geraden auf parallele Geraden abgebildet werden, wird die Transformation affine Transformation genannt. Bei einer affinen Transformation gibt es zusätzlich zu Translation, Rotation und Skalierung die Möglichkeit einer Scherung. Für eine affine Transformation werden drei nicht-kollineare Punkte benötigt. Bei dieser Transformationsart gibt es sechs Freiheitsgrade [5]. Nach [5] kann eine affine Transformation folgendermaßen berechnet werden:

$$u = a_0 + a_1 x + a_2 y \tag{2.4}$$

$$v = b_0 + b_1 x + b_2 y \tag{2.5}$$

Werden bei der Transformation lediglich Geraden auf Geraden abgebildet ohne parallel sein zu müssen, heißt die Transformation projektiv. Vier Punkte sind nötig, um eine projektive Transformation zu bestimmen [10]:

$$u = \frac{a_0 + a_1 x + a_2 y}{1 + c_1 x + c_2 y} \tag{2.6}$$

$$v = \frac{b_0 + b_1 x + b_2 y}{1 + c_1 x + c_2 y} \tag{2.7}$$

Außerdem gibt es nach Maintz und Viergever elastische oder deformierbar genannte Transformationen [10]. Hier werden Geraden auf Kurven abgebildet.

Transformationsdomäne

Die Transformationsdomäne gibt den Bereich an, auf den die Transformation angewendet wird. Hierbei wird zwischen einer lokalen und einer globalen Transformation unterschieden [10]. Globale Transformation bedeutet, dass die Transformation auf das gesamte Bild angewendet wird. Bei einer lokalen Transformation bezieht sich diese nur auf Teilbereiche des Bildes.

Abbildung 2.1 zeigt die unterschiedlichen Transformationsarten in Kombination mit lokaler und globaler Transformationsdomäne.



Abbildung 2.1: Beispiele von 2D-Transformationen aus [10]

Interaktion

Unter der Interaktion einer Bildregistrierung ist der Grad des nötigen manuellen Eingriffs des*der Benutzer*in während der Registrierung zu verstehen [10]. Es wird zwischen den drei Interaktions-Stufen interaktiv, halbautomatisch und automatisch unterschieden.

Bei einem interaktiven Verfahren wird der Eingriff des*der Benutzer*in für die Registrierung benötigt. Bilder müssen beispielsweise durch Verschieben, Drehen und/oder Skalieren manuell registriert werden, um die geeigneten Transformationsparameter zu ermitteln. Im Gegensatz dazu ist bei automatischen Verfahren kein Benutzer*innen-Eingriff erforderlich. Bei halbautomatischen Verfahren ist ein Eingriff des*der Benutzer*in möglich. Dadurch soll eine Verbesserung der Registrierung erreicht werden. So können die Bilder beispielsweise manuell vorregistriert oder auch die Registrierung visuell nachkontrolliert werden. Angestrebt wird in der vorliegenden Arbeit eine automatische Registrierung.

Optimierungsprozess

Um die optimalen Transformationsparameter der Registrierung zu bestimmen, können entweder einzelne Punkte oder die gesamten Bilder verglichen werden [10]. Wenn einzelne Punkte verwendet werden, muss herausgefunden werden, welche Punkte in den Bildern zusammengehören. Aus den Punktkorrespondenzen kann dann die Transformationsmatrix bestimmt werden. Es ist zu beachten, dass für verschiedene Transformationen eine Mindestanzahl an zusammengehörenden Punkte gefunden werden muss.

Wenn die Bilder anhand ihrer Bildinformationen verglichen werden, wird ein Maß benötigt, welches angibt, wie ähnlich sich die Bilder sind. Dieses Maß wird als Ähnlichkeitsmetrik oder Kostenfunktion bezeichnet [10]. Im Lauf der Registrierung muss der Wert der Ähnlichkeitsmetrik optimiert werden. Die Bilder werden also solange transformiert bis sie sich maximal ähneln. In der Regel wird das Maximum oder Minimum der Ähnlichkeitsmetrik, abhängig von der verwendeten Ähnlichkeitsmetrik, in einem iterativen Prozess optimiert und damit die besten Transformationsparameter gefunden [10].

Modalitäten

Nach [10] wird bei einer Registrierung zwischen einer monomodalen und einer multimodalen Registrierung unterschieden. Im Fall einer monomodalen Registrierung stammen die zu registrierenden Bilder aus den gleichen bildgebenden Verfahren. Im Gegensatz dazu stammen bei einer multimodalen Registrierung die Daten von unterschiedlichen bildgebenden Verfahren. Multimodale Registrierungen sind oftmals kompliziert, da durch die unterschiedlichen bildgebenden Verfahren die aufgenommen Bilder sehr unterschiedlich aussehen und somit zumindest eine automatische Registrierung deutlich erschwert wird. In dieser Arbeit ist die Registrierung multimodal, da die Daten zweier verschiedener bildgebender Verfahren registriert werden sollen.

Subjekt der Registrierung

Das Subjekt einer Registrierung ist der zu registrierende Gegenstand in den Bildern [10]. Bei medizinischen Daten handelt es sich dabei um den*die Patient*in. Eine weitere Einteilung kann getroffen werden, je nachdem ob die Bilder von dem*derselben Patient*in oder von unterschiedlichen Patient*innen stammen. Ersteres wird als intrasubjektiv und letzteres als intersubjektiv bezeichnet [10]. Für die vorliegende Arbeit wird davon ausgegangen, dass in beiden Bildern dieselbe Person abgebildet ist. Es handelt sich hierbei also um eine intrasubjektive Registrierung.

Objekt der Registrierung

Nach [10] wird unter dem Objekt einer Registrierung der aufgenommene Bereich des Körpers verstanden. So kann das Objekt in verschiedene Regionen wie beispielsweise Kopf, Thorax oder Abdomen unterteilt werden. Diese Regionen können anschließend weiter spezifiziert werden. Im Bereich des Kopfes kann als Objekt somit das Gehirn, der Schädel, das Auge oder auch die Zähne definiert werden. In der vorliegenden Arbeit geht es nicht speziell um ein genaues Objekt. Da in den aufgenommenen Daten jedoch immer der Kopf abgebildet ist, ist dieser somit auch das Objekt der Registrierung. Der Fokus auf ein bestimmtes Registrierungsobjekt ist aber nicht ausschlaggebend für die Arbeit.

2.2.2 Registrierungsverfahren

Im folgenden Abschnitt findet eine Einteilung der intrinsischen Registrierungsverfahren statt. Es gibt eine sehr große Anzahl von konkreten Ansätzen unterschiedlicher Registrierungsmethoden. In den meisten Fällen können diese in eine der folgenden Kategorien eingeteilt werden:

- Intensitätsbasierte Verfahren
- Merkmalsbasierte Verfahren
- Deep-Learning-basierte Verfahren
- Kombination der vorherigen Verfahren

Zunächst wird auf die intenstiätsbasierten und merkmalsbasierten Verfahren näher eingegangen, da sie für die Arbeit von größerer Relevanz sind.

Intensitätsbasierte Verfahren

Intensitätsbasierte Verfahren arbeiten mit den Intensitätswerten der Bilder. Der Registrierungsvorgang bekommt zwei Bilder als Eingabe. Das eine Bild ist das Referenzbild und das andere das Bewegtbild. Das Bewegtbild soll bei der Registrierung mit einer geometrischen Transformation mit dem Referenzbild passgenau überlagert werden [11]. Eine Metrik dient als Maß des Unterschieds oder der Ähnlichkeit der zu registrierenden Bilder. Im Registrierungsprozess gilt es die optimalen Transformationsparameter zu finden, welche das Bewegtbild auf das Referenzbild abbildet. Dazu wird je nach Metrik das Maß des Unterschieds minimiert oder das Ähnlichkeitsmaß maximiert [11].

[11] beschreibt, dass zum Finden der optimalen Transformationsparameter bei intensitätsbasierten Verfahren in einem iterativen Prozess ausgehend von einer Initialschätzung Bildtransformationen berechnet werden. Das Bewegtbild wird mit diesen geschätzten Transformationsparametern transformiert. Anschließend wird mit der Ähnlichkeitsmetrik berechnet, wie exakt die Registrierung mit den aktuellen Transformationsparametern ist, also wie ähnlich sich die Bilder nach der Transformation sind [11]. Nach einer Transformation wird eine Interpolation zwischen den Bildpixeln benötigt, da die Bildpixel oft nicht exakt übereinander liegen. Durch die Interpolation wird dann die Pixelposition jedes Pixels berechnet. Der Optimierungsprozess wird so lange iterativ fortgesetzt, bis entweder keine bessere Transformation, also kein besserer Wert der Ähnlichkeitsmetrik, gefunden wird oder die Maximalanzahl an Iterationen durchlaufen ist [11].

Bei der Wahl der Ähnlichkeitsmetrik muss bedacht werden, dass diese Metriken einen Ähnlichkeitswert anhand der Intensitätswerte berechnen. Da es bei multimodalen Daten im Normalfall keinen linearen Zusammenhang zwischen den Intensitätswerten gibt, können viele Ähnlichkeitsmetriken in diesem Fall nicht verwendet werden.

Im Folgenden werden drei Ähnlichkeitsmetriken vorgestellt.

Methode des kleinsten quadratischen Fehlers Ein einfaches Ähnlichkeitsmaß ist die Methode des kleinsten quadratischen Fehlers (engl. Mean Squared Error, *MSE*). Diese Methode basiert auf den Pixelintensitäten. Im Registrierungsprozess wird die quadratische Differenz der Intensitätswerte der zu registrierenden Bilder minimiert. Hierbei berechnet sich der Abstand wie folgt [9]:

$$MSE(A,B) = \frac{1}{N} \sum_{i}^{N} (A_i - B_i)^2$$
(2.8)

N steht für die Anzahl der betrachteten Pixel, A_i und B_i jeweils für das *i*-te Pixel in Bild A und B. Für eine maximale Ähnlichkeit der Bilder muss der Wert minimiert, der quadratische Fehler also kleiner werden [9].

Diese Metrik geht davon aus, dass die Intensitätswerte in beiden Bildern sich sehr ähnlich sind. Für multimodale Daten ist diese Metrik damit ungeeignet, da nicht davon ausgegangen werden kann, dass die Intensitätswerte ähnlich sind [9].

Strukturelle Ähnlichkeit Als weitere Ähnlichkeitsmetrik kann der Index struktureller Ähnlichkeit (engl. Structural Similarity Index, *SSIM*) verwenden werden. Das Verfahren der strukturellen Ähnlichkeit von Wang, Bovik, Sheikh u. a. basiert auf der Annahme, dass eine Änderung der strukturellen Informationen eines Bildes ein gutes Maß ist, um eine Distortion eines Bildes festzustellen [12]. Dabei wird die Ähnlichkeitsmessung in die Bereiche Helligkeit mit Helligkeitsfunktion l(x, y), Kontrast mit Kontrastfunktion c(x, y)und Struktur mit Strukturfunktion s(x, y) unterteilt. Die Helligkeit wird über die mittlere Intensität μ der Bilder verglichen. Der Kontrastvergleich geschieht über den Vergleich der Standardabweichung σ der Intensitätswerte vom Mittelwert. Die Strukturvergleichsfunktion wird auf die Signale, welche mit der eigenen Standardabweichung normalisiert sind, angewendet. Bestehend aus den drei Vergleichsfunktionen ergibt sich der strukturelle Ähnlichkeitsindex *SSIM*(x, y) [12]:

$$SSIM(x,y) = [l(x,y)]^{\alpha} \times [c(x,y)]^{\beta} \times [s(x,y)]^{\gamma}$$
(2.9)

Mutual Information Bei multimodalen Daten ist die repräsentative Metrik die Mutual Information (MI) [5]. Diese Metrik setzt keine Beziehung zwischen den Intensitätswerten voraus und misst die statistische Abhängigkeit von zwei Datensätzen. Je höher die Abhängigkeit der Datensätze, desto größer ist der Wert der Mutual Information. Im Fall der Registrierung sind beide Bilder korrekt registriert, wenn die Bilder maximal voneinander abhängen und somit der Wert der MI maximal ist [5].

Alle Formeln dieses Abschnittes zur Mutual Information entstammen [13]. Die Mutual Information basiert auf Entropien, welche in der Kommunikationstheorie verwendet werden, um den Informationsgehalt einer Nachricht beurteilen zu können. Ereignisse $e_1, ... e_m$ treten mit einer Wahrscheinlichkeit $p_1, ..., p_m$ auf. Es wird ausgenutzt, dass eine geringe Auftretenswahrscheinlichkeit eines Ereignisses für einen hohen Informationsgehalt spricht [13].

Unter einer Entropie wird der durchschnittliche Informationsgehalt einer Nachricht verstanden. Die Shannon Entropie H ist folgendermaßen definiert:

$$H = \sum_{i} p_{i} \log \frac{1}{p_{i}} = -\sum_{i} p_{i} \log p_{i}$$
(2.10)

Damit kann die Entropie H(A) eines Bildes A definiert werden als:

$$H(A) = -\sum_{i} p_i \log p_i \tag{2.11}$$

Der Informationsgehalt zweier Bilder kann über die Verbundentropie ${\cal H}(A,B)$ berechnet werden:

$$H(A,B) = -\sum_{i,j} p(i,j) \log p(i,j)$$
(2.12)

Die Mutual Information I(A, B) kann sowohl über die Einzelentropien H(A) und H(B) und die Verbundentropie H(A, B) als auch über eine der Einzelentropien und deren bedingten Entropie H(A|B) beziehungsweise H(B|A) definiert werden. Die bedingte Entropie gibt dabei die Grauwertverteilung in einem Bild abhängig von der Grauwertverteilung im anderen Bild an.

Beispielsweise wird die Grauwertverteilung in Bild A abhängig von der Grauwertverteilung in Bild B folgendermaßen bestimmt:

$$H(A|B) = H(A,B) - H(B)$$
 (2.13)

Der Wert der Mutual Information kann also definiert werden mittels folgender Gleichung:

$$I(A,B) = H(A) + H(B) - H(A,B) = H(A) - H(A|B) = H(B) - H(B|A)$$
(2.14)

Die Mutual Information Metrik kann sowohl bei monomodalen als auch bei multimodalen Registrierungsproblemen verwendet werden. Bei einer Registrierung mit multimodalen Daten liefert die Mutual Information Metrik oftmals gute Ergebnisse [5].

Merkmalsbasierte Verfahren

Merkmalsbasierte Verfahren arbeiten mit charakteristischen Eigenschaften der Bilder. Unter charakteristischen Eigenschaften können sowohl markante Punkte, Linien und Ecken als auch ganze Segmente verstanden werden [14]. Die charakteristischen Eigenschaften werden im folgenden als Merkmale oder Landmarke bezeichnet. Wenn gute Merkmale detektiert werden können, kann aus den korrespondieren Punkten und ihren Pixelpositionen die Transformation für die Registrierung bestimmt werden, die die Punkte aus dem Bewegtbild auf die dazugehörigen Punkte aus dem Referenzbild abbildet. Eine merkmalsbasierte Registrierung besteht im Normalfall aus den folgenden Hauptschritten [14]:

- 1. Merkmalsdetektion: Die Merkmale müssen in beiden Bildern entweder manuell oder automatisch detektiert werden.
- 2. Merkmalsbeschreibung: Die extrahierten Merkmale werden für den Matching-Schritt beschrieben. Dieser Schritt ist nur bei der automatischen merkmalsbasierten Registrierung notwendig.
- 3. Merkmals-Matching: Es werden die Übereinstimmungen zwischen den Merkmalen beider Bilder gefunden.
- 4. Transformations-Modell: Durch die gefundenen Matches werden die Transformationsparameter bestimmt.
- 5. Resamplingprozess: Das Bewegtbild wird mittels der berechneten Transformation auf das Referenzbild abgebildet.

Merkmalsdetektion und -extraktion kann unterteilt werden in eine Detektion von Ecken und eine Detektion von kleinen Bereichen [14]. Ecken beschreiben dabei die Position im Bild, an der sich die Grauwerte signifikant ändern. Es gibt verschiedene Detektionsverfahren, welche Ecken in Bildern detektieren. Beispiele hierfür sind der Harris Ecken Detektor oder der Gaußsche Laplace Detektor.

Die Detektion von kleinen Bereichen bezieht sich auf Flächen, welche sich in Farbe und Grauwert von der Umgebung unterscheiden. Ein klassischer Detektor für solche Flächen ist SIFT (Scale-Invariant Feature Transform) [14]. Im folgenden Abschnitt wird die Funktionsweise des SIFT Merkmalsdetektors kurz erläutert.

SIFT Das Verfahren von Lowe [15] ist robust gegenüber Skalierungsänderungen, Translation, Rotation und geometrischen Verzerrungen.

Um Skalierungsinvarianz zu garantieren, wird für die Merkmalsdetektion ein Skalenraum-Filter verwendet. Dabei wird eine Approximation eines Laplacian of Gaussian (LoG) genutzt, um zu detektieren, an welchen Stellen eine große Änderung im untersuchten Bild auftritt. LoG wird dabei durch einen Difference of Gaussians (DoG) approximiert [15]. Hierbei wird das Bild mit einem Gaußschen Filter weichgezeichnet und von dem Originalbild abgezogen. Dieser Prozess wird für verschiedene Stufen einer Gaußschen Pyramide durchgeführt. Anschließend werden lokale Extrema im Bild gesucht. Ein Extremum beschreibt dann einen potentiellen signifikanten Merkmalspunkt [15].

Um genauere Merkmale zu erhalten, werden alle potentiellen Merkmale mit niedrigem Kontrast aussortiert. Dies geschieht mittels einer Taylorreihenentwicklung des Skalenraums. Wenn die Intensität an einem Extremwert kleiner als ein bestimmter Schwellenwert ist, wird das Merkmal verworfen [15]. DoG reagiert stark auf Kanten, weshalb diese auch herausgefiltert werden. Hierfür wird eine 2×2 Hesse-Matrix verwendet, um das Verhältnis der Hauptkrümmung zu bestimmen. Ist dieses Verhältnis ebenfalls über einem Schwellenwert, wird das Merkmal verworfen [15].

Nach [15] wird für die Rotationsinvarianz jedem gefundenen Merkmals-Punkt seine Orientierung zugewiesen. Die Umgebung der gefundenen Merkmals-Punkte wird in kleine Bereiche unterteilt. Für jeden Bereich wird die Magnitude und Richtung mittels des Gradienten berechnet. Es wird ein Histogramm der Orientierung erstellt. Das Maximum im Histogramm und alle Werte über 80 % werden verwendet, um die Orientierung zu berechnen. Dadurch werden Schlüsselpunkte mit gleicher Lage und Skalierung, aber unterschiedlicher Richtung erzeugt. Um jeden Schlüsselpunkt herum wird die Umgebung in kleine Blöcke unterteilt, in denen ein Histogramm der Orientierung erzeugt wird. Durch diese Histogramme wird der Merkmals-Punkt mit seiner Umgebung als Vektor beschrieben.

Die gefundenen Merkmale aus beiden Bildern werden zugeordnet, indem ihr nächster Nachbar gefunden wird [15].

Vergleich der Registrierungsverfahren

In diesem Abschnitt werden die Verwendungszwecke der Registrierungsverfahren sowie deren Vor- und Nachteile beschrieben. Die Einordnung geschieht analog zur Arbeit von Zitova [5]. Merkmalsbasierte Verfahren können verwendet werden, wenn die Informationen über die Strukturen des Bildes signifikanter sind als die Informationen, welche aus den Bildintensitäten gewonnen werden können. Multimodale Daten können damit leichter registriert werden unter der Annahme, dass die Strukturen in beiden Modalitäten gut erkennbar sind [5]. Der Hauptnachteil von merkmalsbasierten Verfahren ist, dass es schwierig sein kann, Merkmale zu detektieren und zu beschreiben. Außerdem können die Merkmalsdeskriptoren zu sensitiv sein, wodurch es schwierig ist, Übereinstimmungen zwischen den gefundenen Merkmalen zu finden [5].

Nach Zitova werden intensitätsbasierte Verfahren bevorzugt verwendet, wenn es keine markanten Merkmale in den zu registrierenden Bildern gibt. Im Falle einer multimodalen Registrierung ist die Wahl einer geeigneten Ähnlichkeitsmetrik schwierig, da es oft keinen Zusammenhang zwischen den Intensitätswerten unterschiedlicher Modalitäten gibt. Der Hauptnachteil von intensitätsbasierten Verfahren ist deren Zeitkomplexität. Diese wächst mit der Anzahl an Freiheitsgraden der erwarteten geometrischen Transformation. Um das Finden der optimalen Transformationsparameter zu beschleunigen, können pyramidale Bildrepräsentationen und verschiedene Optimierungsalgorithmen verwendet werden [5].

In [5] werden ebenfalls Hybrid-Methoden erwähnt, die die Vorteile beider Verfahren ausnutzen. Ein Beispiel hierfür ist, zu Beginn einer Registrierung mit einem merkmalsbasierten Verfahren die beiden Bilder grob zu registieren und anschließend das Ergebnis mit einem intensitätsbasierten Verfahren zu optimieren.

Des Weiteren können auch Deep-Learning-basierte Verfahren für die Registrierung verwendet werden. Deep-Learning-basierte Ansätze können intensitäsbasierte Verfahren unterstützen, für die direkte Bestimmung von Transformationsparametern verwendet werden, die Zeitkomplexität reduzieren oder auch Merkmale in Bildern mit Erscheinungsdifferenzen detektieren [14].

3 Verwandte Arbeiten

Bislang gibt es keine bestehende Arbeit zu der Registrierung zwischen Tiefen- und Thermalbildern. Es gibt Arbeiten, die sich mit der multimodalen Registrierung zwischen Thermalbildern und sichtbaren Bildern (RGB-Bildern/Grauwertbildern) beschäftigen sowie Arbeiten, welche eine multimodale Registrierung zwischen Tiefenbildern und sichtbaren Bildern untersuchen. Im folgenden Kapitel werden zwei Arbeiten vorgestellt, welche eine Registrierung von Thermalbildern und sichtbaren Bildern anhand einer Vorverarbeitung der Bilder untersuchen.

3.1 Morphologischer Gradient

Im folgenden Abschnitt wird ein Verfahren von Zeng, Adu, Liu u. a. vorgestellt [16]. Das Verfahren basiert auf morphologischen Operationen und einem modifizierten SIFT Algorithmus. Bei dem Registrierungsprozess wird eine Registrierung zwischen Thermalbildern und sichtbaren Bildern untersucht. Die Unterschiede in den Grauwerten dieser Bilder stellen eine große Herausforderung für die Registrierung dar, da es schwierig ist, die gleichen Merkmale aus beiden Bildern zu detektieren. Bei dem Verfahren wird ausgenutzt, dass die Konturen der Szene in sowohl den Thermalbilddaten als auch den sichtbaren Bilddaten enthalten sind. Durch die Vorverarbeitung mit dem morphologischen Gradienten werden die Kanten der beiden Grauwertbilder hervorgehoben. Dadurch werden auch die Histogramme der Bilder sehr ähnlich.

[16] beschreibt das Prinzip von morphologischen Operationen. Unter Morphologie in der Bildverarbeitung wird die Analyse von Bildstrukturen verstanden. Dabei werden die Bilder als Mengen betrachtet und mit verschiedenen morphologischen Operationen die Form dieser Mengen analysiert. Die Strukturen in den Bildern werden ebenfalls von kleineren Mengen beschrieben, welche als Strukturelemente (SE) bezeichnet werden. Zu den einfachsten morphologischen Operationen zählen Erosion und Dilatation, auf denen viele morphologische Methoden basieren. Bei einer Erosion wird geprüft, ob ein Strukturelement vollständig in die vom Bild definierte Menge passt. Die Erosion in einem Grauwertbild kann als punktweises Minimum verstanden werden. Der Wert der Erosion eines Pixels in einem Grauwertbild ist damit der minimale Wert in der Umgebung des Strukturelements [16]. Durch eine Erosion werden dunkle Strukturen im Bild vergrößert und hellere Strukturen verkleinert. Somit kann durch die Erosion beispielsweise weißes Rauschen aus Bildern eliminiert werden. Gleichzeitig wird die Struktur insgesamt verkleinert. Für einen Punkt f(x, y) in einem Grauwertbild, kann eine Erosion folgendermaßen berechnet werden [16]:

$$dst(x,y) = \min_{i,j} src(x+i,y+j)$$
(3.1)

dst beschreibt dabei den erodierten Wert an Position (x, y) und (i, j) stehen für die Punkte in der Umgebung des Strukturelements.

Die Dilatation ist die entgegengesetzte Operation zur Erosion. Hier wird geprüft, ob das Strukturelement die Bildmenge berührt. Die Dilatation in Grauwertbildern beschreibt das punktweise Maximum. Der Wert der Dilatation eines Pixels in einem Grauwertbild ist damit der maximale Wert in der Umgebung des Strukturelements [16]. Hierdurch werden helle Strukturen im Bild vergrößert und dunkle Strukturen verkleinert.

Für einen Punkt f(x, y) in einem Grauwertbild kann eine Dilatation folgendermaßen berechnet werden [16]:

$$dst(x,y) = \max_{i,j} src(x+i,y+j)$$
(3.2)

dst beschreibt dabei den dilatierten Wert an Position (x, y) und (i, j) stehen für die Punkte in der Umgebung des Strukturelements.

Der morphologische Gradient beschreibt die Differenz zwischen Dilatation und Erosion. Die Formel zur Berechnung des morphologischen Gradienten lautet [16]:

$$g = \max_{i,j} f(x+i, y+j) - \min_{i,j} f(x+i, y+j)$$
(3.3)

Dabei steht g für das morphologische Gradientenbild, f ist das Originalbild, f(x, y) ist ein Punkt im Bild und (i, j) sind die Punkte in der Umgebung des Strukturelements. Es entsteht ein Bild mit hervorgehobenen Kanten.

Nachdem die Bilder mittels der morphologischen Gradientenmethode vorverarbeitet wurden, werden die Merkmale in den Gradientenbildern mit einem modifizierten SIFT Algorithmus, welcher C_SIFT genannt wird, extrahiert. Die Gradientenmethode in SIFT zum Berechnen der Orientierung eines Merkmals wird in [16] bei C_SIFT durch eine Schwerpunktsmethode ersetzt. Der Schwerpunkt beschreibt ein nicht zentriertes Moment in der Bildverarbeitung und kann wie folgt berechnet werden [16]:

$$m_{pq} = \sum_{x,y} x^p y q I(x,y) \tag{3.4}$$

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}\right) \tag{3.5}$$

$$\theta = a \tan 2(m_{01}, m_{10}) \tag{3.6}$$

wobe
i m_{pq} ein Bildmoment bezeichnet, $\sum_{x,y}$ für die Pixel in
einer bestimmten Region zentriert auf den Merkmals-Punkt steht,
C die Koordinaten des Schwerpunkts des Bildes beschreibt und
 θ der Winkel zwischen Merkmalspunkt und Schwerpunkt ist.

Durch die Schwerpunktsmethode wird jedem Merkmal eine Hauptrichtung zugewiesen, wodurch eine Rotationsinvarianz gegeben ist [16].

Die Merkmale werden anschließend mit dem BRIEF Merkmalsdeskriptor [17] beschrieben. Offensichtliche falsche Übereinstimmungen werden mit dem RANSAC Algorithmus [18] herausgefiltert.

3.2 Histogram Matching

Krishnan, Balasubramanian und Jeyakumar stellen ein Verfahren zur Registrierung von Thermalbildern und sichtbaren Bildern vor [19]. Es wird beschrieben, dass durch die multimodalen Daten die Intensitätsverteilungen der Bilder sehr unterschiedlich sind, wodurch eine intensitätsbasierte Registrierung erschwert wird. Merkmalsbasierte Verfahren, die mit SIFT [15], SURF [20] oder ORB [21] arbeiten, detektieren keine guten Merkmalspunkte in den Thermalbildern, da diese keine guten beziehungsweise scharfen Gesichtsmerkmale bereitstellen [19]. Als Vorverarbeitungsschritt werden die Histogramme der Bilder angeglichen. Anschließend werden die Transformationsparameter der Registrierung bestimmt, indem die Ähnlichkeit der Bilder maximiert wird. Durch die Histogramme der Bilder wird deutlich, dass beide Bilder eine sehr unterschiedliche Intensitätsverteilung haben. Für die unangeglichenen Bilder führt ein intensitätsbasiertes Registrierungsverfahren zu einer falschen Registrierung. Daher werden zu Beginn die Histogramme angeglichen. Dabei wird die Wahrscheinlichkeit eines Grauwerts für ein Thermalbild als $p_{f_1}(l_i), l_i \in [0, 255]$ und für ein sichtbares Bild als $p_{f_2}(l_i), l_i \in [0, 255]$ definiert. Daraus folgt deren kummulative Wahrscheinlichkeitsdichte $C_{f_{1/2}}(l_i)$ [19]. Um die Histogramme anzugleichen, wird eine Abbildungsfunktion gesucht, die die kummulativen Wahrscheinlichkeitsdichten angleicht. Es soll gelten [19]:

$$C_{f_1}(l_i) \approx C_{f_2}(l_i) \tag{3.7}$$

Jedes Pixel des Bildes, welches angeglichen werden soll, wird mit der Abbildungsfunktion Q(i) transformiert, wodurch ein Bild mit angeglichenem Histogramm entsteht. Nach der Vorverarbeitung durch das Anpassen der Histogramme (Histogram Matching), wird in [19] ein intensitätsbasiertes Registrierungsverfahren durch Maximierung der Ähnlichkeit der Bilder verwendet. Das angeglichene Thermalbild wird als Referenzbild behandelt und das sichtbare Bild als Bewegtbild. Als Ähnlichkeitsmetrik wird die Mutual Information Metrik verwendet. Die Performance wird anhand der Werte der Mutual Information Metrik und des strukturellen Ähnlichkeitsindex bewertet [19]. Die Ergebnisse des Verfahrens machen deutlich, dass die Methode für eine angemessene Registrierung von Thermalbildern und sichtbaren Bildern verwendet werden kann. In [19] wird darauf eingegangen, dass bereits Arbeiten existieren, die ebenso gute Ergebnisse mit merkmalsbasierten Verfahren erzielen können. Diese Arbeiten konzentrieren sich jedoch nicht speziell auf Thermalbilder und sichtbare Bilder, welche menschliche Gesichter abbilden.

4 Methodik

Einen groben Überblick über den Ablauf der Arbeit zeigt Abbildung 4.1. In der obersten Ebene (gelb) ist der Datensatz, aus dem die verwendeten Daten dieser Arbeit stammen, abgebildet. Die zweite Ebene (grün) beschreibt den Prozess der Erzeugung der für die Registrierung benötigten Daten. Dafür werden mittels des gegebenen Datensatzes aus 3D-Modellen künstliche Tiefenbilder und aus Thermalvideos Grauwert-Thermalbilder generiert. Diese Bilder werden in Ebene 3 (blau) mit verschiedenen Vorverarbeitungsmöglichkeiten bearbeitet, um Merkmale in den zu registrierenden Bildern hervorzuheben. Es soll untersucht werden, ob eine Vorverarbeitung der Bilder zu besseren Registrierungsergebnissen führt. In Ebene 4 (rosa) wird zum einen ein Zusammenhang zwischen den erzeugten Bildern untersucht und zum anderen verschiedene Registrierungsverfahren getestet. Ebene 5 (orange) zeigt die einzelnen Methoden, die für die Analyse der Daten verwendet werden, sowie die verschiedenen getesteten Registrierungsverfahren. Zuletzt werden in der untersten Ebene (rot) die Ergebnisse der Registrierung ausgewertet.

4.1 Verwendete Daten

Wie in Abbildung 4.1 zu sehen ist, wird als Datengrundlage für diese Arbeit der Datensatz BP4D+ verwendet (Ebene 1, Abbildung 4.1). BP4D+ erweitert den Datensatz BP4D und beinhaltet multimodale Daten, die spontane Emotionen am Menschen aufzeichnen. Die Emotionen wurden an 140 Versuchspersonen unterschiedlicher Ethnien erprobt. Darunter waren 82 weibliche und 58 männliche Versuchspersonen [22]. Die Versuchspersonen wurden nach Geschlecht als F001-F082 und M001-M058 kodiert. Die aufgenommenen Daten bestehen aus Gesichtsausdruck, Thermaldaten, 2D- und 3D-Sequenzen und physiologischen Daten. Pro Versuchsperson gibt es zehn Aufgaben (T1-T10), mit denen unterschiedliche Emotionen hervorgerufen werden sollen [22].



Abbildung 4.1: Überblick über den Ablauf der Arbeit: Die sechs verschiedenen Ebenen sind farbig markiert. Ebene 1 zeigt den verwendeten Datensatz, Ebene 2 beschreibt die Erzeugung von Daten, Ebene 3 stellt eine Vorverarbeitung der Daten dar, Ebene 4 steht für die zwei untersuchten Ansätze, Ebene 5 definiert die verwendeten Methoden in diesen Ansätzen genauer und Ebene 6 zeigt die Bewertung der Ergebnisse.

Die 3D-Sequenzen und deren korrespondierende 2D-Textur-Sequenzen wurden von einem dreidimensionalen bildgebenden System aufgezeichnet, welches aus einem 3D Stereosensor und einem 2D Videosensor bestand. Der 3D Stereosensor setzte sich aus zwei monochromen Stereokameras zusammen, die vertikal gleich ausgerichtet waren. Der 2D Videosensor war eine 2D RGB-Farbkamera, welche zwischen den beiden Stereokameras platziert war. Aus jedem Frame der aufgezeichneten Videos wurde mittels einer passiven Stereo-Photogrammetrie ein 3D-Modell erzeugt [22]. Die Bildfrequenz des dreidimensionalen bildgebenden Systems lag bei 25 frames per second (fps).

Der Thermalsensor wurde mit einem Stativ an dem dreidimensionalen System befestigt. Mit dieser Thermalkamera wurden Thermalvideos mit einer Auflösung von 640×480

Pixeln pro Bild aufgenommen. Die Bildfrequenz wurde für eine bessere Synchronisierung aller Sensoren im System auf 25 fps eingestellt [22].

Für die Registrierung zwischen Thermal- und Tiefenbildern werden die 3D-Gesichts-Modelle sowie die Thermalvideodaten verwendet. Aus den 3D-Modellen werden künstliche Tiefenbilder generiert. Die Thermalvideos werden in einzelne Frames unterteilt und diese RGB-Frames in Grauwertbilder, welche die passenden Temperaturwerte speichern, umgewandelt. Diese beiden Verfahren werden in den folgenden Unterkapiteln genauer ausgeführt.

4.2 Programmiersprache und verwendete Bibliotheken

Der Hauptteil dieser Arbeit ist in Python implementiert. Dabei sind die Bilder zweidimensionale Numpy-Arrays, in denen jeder Eintrag im Array dem Intensitätswert an dieser Pixelposition im Bild entspricht. Für die Erzeugung der Tiefenbilder aus den 3D-Modellen wird dabei open3d¹ zum Einlesen der Dreiecksgitter verwendet. Die Objekterkennung von in den Bildern integrierten Temperaturwerten zur Erzeugung von Thermal-Grauwertbildern aus RGB-Bildern wird mit pytesseract² ermöglicht. Eine anschließende Vorverarbeitung der Bilder wird mit OpenCV³ und Scikit-Image (Skimage)⁴ umgesetzt. Auch bietet Open-CV Methoden für die Berechnung von Bildmomenten sowie für eine merkmalsbasierte Registrierung. Eine intensitätsbasierte Registrierung in Python wird mittels des Simple Insight Segmentation and Registration Toolkit (SimpleITK)⁵ getestet. Für eine manuelle Registrierung und eine weitere intensitätsbasierte Registrierung werden verschiedene Methoden der Programmiersprache Matlab verwendet.

4.3 Tiefenbilder

Für die Erzeugung von künstlichen Tiefenbildern wird ein Lochkamera-Modell verwendet. Daher wird im nächsten Abschnitt zuerst das Modell einer Lochkamera näher beschrieben, bevor der finale Algorithmus erklärt wird.

```
<sup>1</sup>http://www.open3d.org/
<sup>2</sup>https://pypi.org/project/pytesseract/
<sup>3</sup>https://opencv.org/
<sup>4</sup>https://scikit-image.org/
<sup>5</sup>https://simpleitk.org/
```

4.3.1 Lochkamera Modell

Ein Lochkamera-Modell ist ein einfaches Kamera-Modell mit einer Kamera ohne Linse und einer kleinen Öffnung. Lichtstrahlen passieren diese kleine Öffnung. Dadurch wird ein invertiertes Bild der Szene auf die gegenüberliegende Seite der Kamera projiziert [23].

In diesem Lochkamera-Modell werden die mathematischen Beziehungen einer Transformation von 3D-Weltkoordinaten auf eine 2D-Bildebene beschrieben [24]. Die Weltkoordinaten werden durch ein eigenes dreidimensionales kartesisches Koordinatensystem beschrieben. Das Kamerakoordinatensystem ist ebenso ein eigenes kartesisches dreidimensionales Koordinatensystem und entspricht im Normalfall nicht der Ausrichtung des Weltkoordinatensystems. Es wird angenommen, dass das Kamerakoordinatensystem irgendwo im Weltkoordinatensystem liegt. Die aufgespannte Bildebene wird durch ein zweidimensionales Koordinatensystem beschrieben.

Analog zu [25] wird im Folgenden angenommen, dass die virtuelle Bildebene vor der Kamera liegt und das aufrechte Bild der Szene abbildet. Die reale Bildebene liegt hinter der Kamera. Es wird untersucht, wo der Strahl von der Kamera zu einem Punkt in der 3D-Welt auf die virtuelle Bildebene auftrifft. Der Ursprung der Kamera, in dem alle Strahlen konvergieren, wird Kamerazentrum oder Zentrum der Projektion genannt [25]. Abbildung 4.2 veranschaulicht das Modell einer Lochkamera.



Abbildung 4.2: Modell einer Lochkamera: Es wird untersucht, wo die Strahlen von Kamerazentrum zu Objekt auf die virtuelle Bildebene treffen. Die Beziehung von Weltkoordinaten und deren korrespondierenden Pixeln in der virtuellen Bildebene kann durch eine 3×4 -Kameramatrix P beschrieben werden [24]. Die Formel für die Transformation von Weltkoordinaten auf die Bildebene wird durch die Gleichung x = PX beschrieben [24]. Dabei ist P die Kameramatrix, X ist ein homogener Punkt in Weltkoordinaten der Form $X = (X_w, Y_w, Z_w, 1)^T$ und x entspricht den transformierten Koordinaten auf der Bildebene.

Die Transformationsmatrix ist aus zwei verschiedenen Transformationen zusammengesetzt:

- 1. Die rigide Transformation vom 3D-Weltkoordinatensystem zum 3D-Kamerakoordinatensystem. Diese Transformation wird durch die extrinsischen Matrix [R|t]beschrieben [24].
- 2. Die Projektionstransformation, welche die 3D-Weltkoordinaten aus dem Blick der Kamera auf die virtuelle 2D-Bildebene projiziert. Diese Transformation wird durch die intrinsische Matrix *K* beschrieben [24].

Zusammengesetzt ergibt sich folgende Gleichung für die Kameramatrix P:

$$P = K[R|t] \tag{4.1}$$

Extrinsische Kameraparameter

Die extrinsische Kameramatrix beschreibt die Position und Orientierung der Kamera in Weltkoordinaten. Diese 3×4 -Matrix besteht aus zwei Komponenten [26]. Eine Rotationsmatrix R und ein Translationsvektor t.

Oftmals wird die extrinsische Matrix als quadratische homogene Matrix mit einer unteren Zeile von (0, 0, 0, 1) definiert. Dadurch kann die Matrix in eine Rotation gefolgt von einer Translation geteilt werden [26]:

$$\begin{bmatrix} \frac{R}{0} & t\\ \hline 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & t_1\\ 0 & 1 & 0 & t_2\\ \hline 0 & 0 & 1 & t_3\\ \hline 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & 0\\ r_{2,1} & r_{2,2} & r_{2,3} & 0\\ \hline r_{3,1} & r_{3,2} & r_{3,3} & 0\\ \hline 0 & 0 & 0 & 1 \end{bmatrix}$$
(4.2)

Die extrinsische Matrix beschreibt, wie Punkte aus dem Weltkoordinatensystem in das Kamerakoordinatensystem transformiert werden. Der Translationsvektor t kann dabei als Position des Weltkoordinatenursprungs in Kamerakoordinaten interpretiert werden. Die Rotation R hingegen beschreibt die Richtung der Weltkoordinatenachsen in Kamerakoordinaten. Nach [26] ist es wichtig zu beachten, dass mit der extrinsischen Matrix die Weltkoordinaten zu den Kamerakoordinaten transformiert werden. Oftmals ist es natürlicher, die Kameraposition und -orientierung direkt zu definieren, als festzulegen, wie Punkte im Weltkoordinatensystem in das Kamerakoordinatensystem transformiert werden. Im Folgenden wird beschrieben, wie die extrinsischen Parameter anhand der Kameraposition bestimmt werden können. Der Spaltenvektor C beschreibt die Lage des Kamerazentrums in Weltkoordinaten. Rc ist eine Rotationsmatrix, welche die Orientierung der Kamera im Bezug auf die Weltkoordinaten definiert. Die Ausrichtung der Kamera wird dann von der Transformationsmatrix [Rc|C] beschrieben, welche mit einer unteren Zeile von (0, 0, 0, 1) zu einer quadratischen homogenen Matrix ergänzt wird [26].

$$\begin{bmatrix} \frac{R}{0} & t \\ \hline 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{Rc}{0} & C \\ \hline 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{Rc^T}{0} & -Rc^TC \\ \hline 0 & 1 \end{bmatrix}$$
(4.3)

Da die Inverse einer Rotationsmatrix deren Transponierte ist und die Invertierung eines Translationsvektors diesen negiert, kann die Gleichung umgeformt und eine Beziehung zwischen den Rotationsmatrizen R und Rc sowie eine Beziehung zwischen den Translationsvektoren t und C aufgestellt werden [26]. Es lässt sich erkennen, dass $R = Rc^T$ und t = -RC ist.

Intrinsische Kameraparameter

Die intrinsische Kameramatrix transformiert einen 3D-Punkt im Kamerakoordinatensystem in ein zweidimensionales Bildkoordinatensystem. Dabei werden interne Kameraparameter verwendet, die nicht von der Ausrichtung der Kamera in den Weltkoordinaten abhängen. Die Bildebene wird mit einer Distanz f zum Kamerazentrum platziert [27]. Diese Distanz wird Brennweite genannt. Ein projizierter Punkt auf der Bildebene wird durch die Pixelpositionen x und y beschrieben. Nach [27] berechnen sich diese zu $x = f \frac{X_c}{Z_c}$ und $y = f \frac{Y_c}{Z_c}$.

Sind die Pixel im Bildsensor nicht quadratisch, wird zwischen zwei verschiedenen Brennweiten fx und fy unterschieden [27]. Falls das optische Zentrum (cx, cy) der Kamera nicht mit dem Zentrum der Bildkoordinaten übereinstimmt, muss dieses ebenfalls in der

intrinsischen Matrix berücksichtigt werden. Außerdem kann ein Versatz zwischen x- und y-Achse des Kamerasensors bestehen. Dieser Versatz wird ebenfalls in der intrinsischen Matrix berücksichtigt [27]. So ergibt sich folgende Matrix *K*:

$$K = \begin{bmatrix} fx & \gamma & cx \\ 0 & fy & cy \\ 0 & 0 & 1 \end{bmatrix}$$
(4.4)

Die Gleichung zur Berechnung der Pixelpositionen lautet somit:

$$\begin{bmatrix} u'\\v'\\w'\end{bmatrix} = \begin{bmatrix} fx & \gamma & cx\\0 & fy & cy\\0 & 0 & 1\end{bmatrix} \begin{bmatrix} X_c\\Y_c\\Z_c\end{bmatrix}$$
(4.5)

Die Bildkoordinaten u und v können anschließend nach [27] mit den Gleichungen $u = \frac{u'}{w'}$ und $v = \frac{v'}{w'}$ ausgedrückt werden.

Die vollständige Gleichung für die Lochkamera kann wie folgt zusammengesetzt werden:

$$\begin{bmatrix} u'\\v'\\w' \end{bmatrix} = \begin{bmatrix} fx & \gamma & cx\\0 & fy & cy\\0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & t_1\\r_{2,1} & r_{2,2} & r_{2,3} & t_2\\r_{3,1} & r_{3,2} & r_{3,3} & t_3 \end{bmatrix} \begin{vmatrix} X_w\\Y_w\\Z_w\\1 \end{vmatrix}$$
(4.6)

4.3.2 Algorithmus zur Erzeugung künstlicher Tiefenbilder

Zu Beginn der Arbeit werden aus dem gegebenen Datensatz künstliche Tiefenbilder generiert (Ebene 2, Abbildung 4.1). Dazu werden die 3D-Modell-Daten verwendet. Diese beinhalten ein dreidimensionales Gitter, welches aus kleinen Dreiecken besteht. Ein Dreieck wird dabei von drei Punkten im Raum, im folgenden als Vertex bezeichnet, aufgespannt. Jedes 3D-Modell des Datensatzes besteht aus 30000 - 50000 Vertices. Eine Visualisierung eines 3D-Modells in Meshlab⁶ ist in Abbildung 4.3 zu sehen. Die Abbildung soll veranschaulichen, woraus sich die 3D-Modelle zusammensetzen. Das erste Bild zeigt das 3D-Modell mit Material. Das Material ist eine zum 3D-Modell zugehörige Datei, welche die Oberflächeneigenschaften des Objekts für eine Visualisierung beschreibt, und wird für

⁶Meshlab ist eine Software für die Bearbeitung von 3D Dreiecksgittern.

den weiteren Verlauf nicht benötigt. Im zweiten Bild sind Material und Gitter zu sehen und das letzte Bild zeigt nur das Gitter.



(a) 3D-Modell mit Material



(b) 3D-Modell mit Material und Gitter



(c) 3D-Modell mit Gitter

Abbildung 4.3: Darstellung eines 3D-Modells in Meshlab, Subjekt F001

In einem ersten Ansatz zur Erzeugung der Tiefenbilder wurde eine Methode geprüft, welche sich an einem Z-Buffering-Verfahren orientierte. Hier wird als Tiefeninformation der z-Wert der einzelnen Punkte des 3D-Modells angenommen. Dieser Algorithmus ist im Vergleich zu den anderen getesteten Algorithmus deutlich schneller, hat aber den Nachteil, dass keine Einstellung der Perspektive möglich ist. Durch die Annahme, dass die Tiefe dem z-Wert der Punkte im Gitter entspricht, ist bei diesem Verfahren die Perspektive mit einer unendlich weit entfernten Kamera gleichzusetzen. Ein weiterer Ansatz ist von der Methode des Raycastings inspiriert. Hier wird eine Position der Lochkamera und eine Position der Bildebene im Raum gewählt. Von der Kamera aus werden Strahlen durch die Bildebene auf das 3D-Modell projiziert. Es wird ermittelt, an welcher Position der Strahl die Bildebene trifft, um die Pixelposition im Tiefenbild herauszufinden. Anschließend wird überprüft, ob der Strahl einen Punkt im Gitter trifft. Hierdurch muss zuerst überprüft werden, ob der Strahl auf eine von einem Dreieck aufgespannte Ebene trifft. Wenn die Ebene getroffen wird, muss außerdem getestet werden, ob der Strahl innerhalb oder außerhalb des Dreiecks die Ebene schneidet. Das muss für alle Dreiecke getestet werden, da nicht bekannt ist, welche Punkte am nächsten an der Kamera liegen. Die Entfernung zum nächsten getroffenen Punkt in der Ebene wird als Tiefeninformation an der vorherigen ausgerechneten Pixelposition gespeichert. Die Berechnung des Schnittpunkts zwischen Strahl und Ebene und der anschließende Test, ob der Schnittpunkt innerhalb oder außerhalb des Dreiecks liegt, ist sehr zeitintensiv. Der Algorithmus wurde an kleineren Gittern mit neun Dreiecken getestet. Da die Laufzeit hier so lang war, dass sich weitere Tests nicht lohnen würden, wurden auch keine weiteren Test mit dieser Methode an den
3D-Gesichtsmodellen durchgeführt.

Von der Raycasting-Methode und dem Lochkamera-Modell inspiriert, wurde anschließend eine Art umgekehrtes Raycasting implementiert. Für das Generieren der Tiefenbilder werden die 3D-Gitterdaten eingelesen⁷. Dadurch können sowohl Informationen über die Koordinaten der Vertices, als auch Informationen über die Konstellation der Vertices, welche ein Dreieck bilden, gewonnen werden. Zu Beginn muss die Ausrichtung der Kamera festgelegt werden. Dafür müssen die Koordinatenachsen von Gitter (Weltkoordinaten) und Kamera betrachtet werden.

In den 3D-Modellen zeigt bei einer Betrachtung des Gesichts von vorne die x-Achse nach rechts, die y-Achse nach oben und die z-Achse nach vorne. Das 3D-Modell liegt in z-Richtung etwas hinter dem Koordinatenursprung. Bei einer frontalen Ausrichtung der Kamera auf das Gesicht zeigt die x-Achse der Kamera nach rechts, die y-Achse nach unten und die z-Achse nach vorne. Damit sind die y- und z-Achsen der Koordinatenachsen von Kameraund Weltkoordinatensystem jeweils entgegengesetzt. Für die Kamera-Rotationsmatrix *Rc*, ausgedrückt in Weltkoordinaten, wird daher Folgendes angenommen:

$$Rc = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$
(4.7)

Wenn die Translation nicht weiter definiert wird, wird die Kamera mittig zwischen maximaler und minimaler x-Komponente sowie mittig zwischen maximaler und minimaler y-Komponente platziert. Die z-Komponente der Translation gibt die Entfernung zum 3D-Gitter an. Der erste Frame wird als Referenz für die Berechnung der Kameraposition verwendet. Für alle weiteren Frames wird anschließend die Kameraposition des ersten Frames verwendet. Würde man für jeden Frame die Kameraposition neu berechnen, würde durch die schwankenden minimalen und maximalen x- und y-Werte der 3D-Modelle die Kameraposition zwischen den Frames leicht variieren.

Abbildung 4.4 zeigt eine Variation des Abstands von Kamera zu 3D-Gitter. Es wird deutlich, dass die Entfernung der Kamera einen deutlichen Einfluss auf die Perspektive des künstlichen Tiefenbildes hat. Für die Erzeugung der Tiefenbilder wird eine Distanz von z = 1000 mm gewählt, da diese der Perspektive der Thermalbilder am ähnlichsten ist.

Des weiteren wird eine Bildhöhe- und breite gewählt, die als Werte die Maße 640×480 Pixel hat, da das den Bildmaßen der Thermalbilder entspricht. Festgelegt wird außerdem

⁷Das Einlesen geschieht mittels der "read_triangle_mesh()"-Methode von open3d.



(a) Abstand z = 100 (b) Abstand z = 300 (c) Abstand z = 500 (d) Abstand z = 1000

Abbildung 4.4: Perspektive der Tiefenbilder mit unterschiedlicher Entfernung der Kamera zum 3D-Modell in mm, Subjekt F001

die maximale Tiefe des Gitters. Für den besten Kontrast sollte die maximale Tiefe möglichst mit der Distanz von Kamera zum am weitesten entfernten Vertex im Gitter übereinstimmen. Hier wird als Wert eine Maximaltiefe von 250 mm festgelegt. Das vorläufige Tiefenbild wird als Numpy-Array mit den gegebenen Bildmaßen erzeugt. Das Tiefenbild speichert zu Beginn die Maximalgittertiefe addiert mit der Entfernung der Kamera zum 3D-Gitter.

Berechnung der Abtastung der Dreiecke Durch ein Abtasten der bestehenden Vertices würde es zu einer Unterabtastung des 3D-Gitters kommen. Das Tiefenbild hätte Löcher. Daher wird jedes Dreieck des 3D-Gitters einzeln abgetastet, um genügende Sample-Punkte zu erhalten.

Die Sample-Länge wird berechnet, indem die Breite der Bounding Box aller Dreiecke, also die maximale x-Koordinate des 3D-Modells minus die minimale x-Koordinate, durch die doppelte Bildbreite geteilt wird. Damit erhält man eine Abtastrate in x-Richtung. Dieser Prozess wird auch mit der Höhe der Bounding Box und der Bildhöhe wiederholt, um eine Abtastrate in y-Richtung zu erhalten. Die Sample-Länge ist dann der maximale Wert von den beiden errechneten Sample-Längen.

Für jedes Dreieck wird nun ausgerechnet, welche Sample-Punkte für das Dreieck benötigt werden. Es wird an allen drei Kanten des Dreiecks mit der zuvor berechneten Sample-Länge die Anzahl der Sample-Punkte auf der jeweiligen Kante berechnet. Es wird der Index der Kante bestimmt, welche die wenigsten und die meisten Sample-Punkte hat. Dadurch ist klar, welche Kante die längste Kante eines Dreiecks und welche die kürzeste Kante ist. Es wird davon ausgegangen, dass ein Dreieck aus zwei langen und einer kurzen Kante besteht. Die Anzahl der Sample-Punkte wird bei beiden langen Kanten auf die gleiche Anzahl gesetzt, damit die Fläche des Dreiecks anschließend besser abgetastet werden kann. Die Sample-Punkte an der längsten Kante werden berechnet. Wenn die zweitlängste Kante nicht genau gleich viele Sample-Punkte wie die längste Kante hat, wird die Anzahl der Sample-Punkte angepasst. Dadurch haben die beiden längsten Kanten gleich viele Sample-Punkte. Zwischen den gegenüberliegenden Punkten an den langen Kanten wird jetzt die Fläche des Dreiecks abgetastet. Dadurch ist die ganze Fläche jedes Dreiecks ausreichend abgetastet. Das Verfahren zum Abtasten der Fläche eines Dreiecks ist an einem Beispiel in Abbildung 4.5 zu sehen.



Abbildung 4.5: Abtastung eines Dreiecks des 3D-Modells

Jeder der erzeugten Sample-Punkte wird in einen homogenen Punkt umgewandelt. Über die Kameramatrix wird ausgerechnet, an welcher Pixelposition (u, v) in der Bildebene der Strahl von Sample-Punkt zu Kamera die virtuelle Bildebene schneidet. Des weiteren wird die Entfernung von Sample-Punkt zu Kamera berechnet. Die Entfernung ist dabei die Länge des Vektors, der das Kamerazentrum und den Sample-Punkt verbindet.

$$dist = \sqrt{(sp_x - c_x)^2 + (sp_y - c_y)^2 + (sp_z - c_z)^2}$$
(4.8)

Dabei steht sp für den Sample-Punkt und c beschreibt das Kamerazentrum.

Sollte diese Entfernung kleiner sein als die momentan gespeicherte Entfernung im Tiefenbild an dieser Pixelposition, wird die neue kleinere Entfernung gespeichert. Der Punkt liegt also näher an der Kamera als der Punkt der davor gespeicherten Entfernung. Anschließend kann beispielsweise noch ein Gauß-Rauschen auf den Hintergrund gelegt werden, um die künstlich erzeugten Tiefenbilder realistischer darzustellen.

4.4 Erzeugung von Thermalbildern aus RGB-Videoframes

Der nächste Schritt ist die Erzeugung von Grauwert-Thermalbildern aus den gegebenen Thermalvideos des Datensatzes (Ebene 2, Abbildung 4.1). Der Datensatz liefert für jede Versuchsperson pro Aufgabe ein Thermalvideo. Die einzelnen Frames der Videos liegen damit als RGB-Daten vor. Um aus den RGB-Bildern Grauwertbilder, bestehend aus den passenden Temperaturwerten, zu erzeugen, wird ein Algorithmus verwendet, welcher die Grauwertbilder aus den RGB-Bildern anhand einer auf den Frames abgebildeten Colormap rücktransformiert. Die Schwierigkeit hierbei ist, dass die Colormap, mit der die Thermalvideos zur besseren Visualisierung eingefärbt sind, auch Teil des Bildes ist und damit sowohl die Farbanzeige als auch die Temperaturanzeige aus Pixeln mit unterschiedlichen RGB-Werten bestehen.

Zu Beginn wird das Thermalvideo in einzelne Frames unterteilt. Aus diesen RGB-Bildern werden verschiedene Bereiche des Bildes ausgeschnitten. Zum einen wird aus dem Bild der Bereich der Colormap weggeschnitten, damit am Ende nur das wirkliche Bild ohne Colormap und Temperaturanzeige bleibt. Außerdem wird die Farbanzeige der Colormap ausgeschnitten. Da die Videos mit derselben Kamera aufgenommen sind, verändert sich die Position der Colormap nicht. Zusätzlich wird der Bereich ausgeschnitten, der die höchste Temperatur anzeigt, sowie der Bereich mit der niedrigsten Temperatur. Aus den Bildausschnitten mit den Temperaturwerten muss nun zunächst die angezeigte Zahl aus dem Bild ermittelt werden. Hier wird kein fester Wert gesetzt, da die Temperaturskala in den Videos immer leicht variiert. Für das Auslesen der Zahlen aus dem Bild wird das Objekterkennungstool "Tesseract" verwendet. Damit die Objekterkennung erfolgreich ist, müssen die Bildausschnitte jeweils noch vorverarbeitet werden. Die Bildausschnitte werden mittels eines experimentell ermittelten globalen Schwellenwerts zu einem Binärbild verarbeitet. Falls die Objekterkennung keine sinnvollen Temperaturwerte erkennt und die erkannte Zahl somit nicht im Bereich zwischen 18.0 °C und 42.0 °C liegen sollte, wird als niedrigste Temperatur 26.0 °C und als höchste Temperatur 37.5 °C angenommen. Sind die minimalen und maximalen Temperaturwerte der Colormap ermittelt, werden als nächstes die RGB-Werte der Colormap gespeichert. In x-Richtung sind die RGB-Werte der Colormap nicht identisch, sondern schwanken leicht. Daher wird hier die mittlere Spalte verwendet. Die Sample-Temperaturwerte werden aus dem Bereich von minimaler bis maximaler Temperatur und der Anzahl an Pixeln in y-Richtung der Colormap berechnet:

samples =
$$\frac{t_{max} - t_{min}}{\text{colormap height} - 1}$$
 (4.9)

In einem Numpy-Array werden pro y-Pixel der Blau-, Grün- und Rotwert sowie die ausgerechnete Temperatur der Colormap abgespeichert. Nun wird über das gesamte ausgeschnittene Thermalbild iteriert. An jedem Pixel wird der RGB-Wert ausgelesen. Die RGB-Werte von Colormap und Thermalbild werden wie Punkte im dreidimensionalen Raum betrachtet. Es wird die euklidische Distanz zwischen den RGB-Werten eines Pixels des Thermalbilds und den RGB-Werten der Referenz-Colormap mittels

$$dist = (bmap - b)^{2} + (gmap - g)^{2} + (rmap - r)^{2}$$
(4.10)

berechnet. Dabei stehen *bmap*, *gmap* und *rmap* für die RGB-Werte der Colormap. *b*, *g* und *r* repräsentierten die RGB-Werte eines Pixels im Thermalbild. Gespeichert wird für dieses Pixel dann der zugehörige Temperaturwert mit der geringsten euklidischen Distanz. Der Vergleich mit der euklidischen Distanz erfolgt, da es Pixel geben kann, die nicht genau mit einem RGB-Referenzwert der Colormap übereinstimmen. Der Algorithmus erzeugt also ein Grauwertbild mit Temperaturwerten zwischen minimaler und maximaler ausgelesener Temperatur.

Abbildung 4.6 zeigt die einzelnen Thermalbilder während der Erzeugung der Thermal Grauwertbilder am Beispiel von Subjekt F001. Links ist das Original RGB-Bild, rechts das erzeugte Grauwertbild.



(a) RGB-Frame aus Thermalvideo



(b) Thermal Grauwertbild

Abbildung 4.6: Vergleich eines Thermal RGB-Bildes und eines erzeugten Thermal Grauwertbildes, Subjekt F001

4.5 Vorverarbeitung der Bilder

Die künstlich erstellten Tiefenbilder sowie die Grauwertbilder der Thermographiedaten werden in verschiedenen Vorverarbeitungsschritten bearbeitet (Ebene 3, Abbildung 4.1). Eine Vorverarbeitung beinhaltet in dieser Arbeit das Anwenden von verschiedenen Filtern und Intensitätstransformationen auf die Bilder. Dazu zählen eine Konvertierung, eine Invertierung, verschiedene Segmentierungsansätze, eine Bearbeitung mit einem morphologischen Gradienten, ein Histogram Matching, eine Adaptive Equalization und drei verschiedene Intensitätstransformationen. Mit den verschiedenen Vorververarbeitungsmöglichkeiten wird untersucht, ob so verschiedene Merkmale in den Bildern hervorgehoben werden können und dadurch auch die Registrierung verbessert werden kann.

4.5.1 Konvertierung zu 8-bit Bildern

Einige Methoden von OpenCV und Skimage arbeiten mit 8-bit Bildern mit Werten zwischen 0 und 255. Um diese Methoden nutzen zu können, müssen die Tiefen- und Thermalbilder vorher umskaliert werden, damit der Wertebereich zwischen 0 und 255 liegt. Für die Skalierung werden Maximalwert und Minimalwert des Bildes benötigt. Jeder Wert im konvertierten Bild berechnet sich durch folgende Skalierung:

$$arr_{conv}[i][j] = \frac{arr[i][j] - arr_{min}}{arr_{max} - arr_{min}} \cdot 255$$
(4.11)

Hierbei muss beachtet werden, dass es zu einem Informationsverlust der Bilder kommen kann. Die Tiefenbilder haben einen Wertebereich von ca. 1000 - 1300 mm. Der Wertebereich der Thermalbilder reicht von ca. 25-40 °C. Der Wertebereich der Thermalbilder wird also durch die Konvertierung gestreckt, wohingegen der Wertebereich der Tiefenbilder gestaucht wird. Abbildung 4.7 stellt die Histogramme der Thermal- und Tiefenbilder vor und nach der Konvertierung dar.

4.5.2 Invertierung

Bei der Visualisierung eines Bildes entsprechen die Intensitätswerte der einzelnen Pixel verschiedenen Grauwerten. Dabei wird ein kleiner Wert als dunkel und ein großer Wert als hell betrachtet. Der kleinste Wert ist somit schwarz und der hellste weiß. Die Tiefenbilder



Abbildung 4.7: Vergleich der Histogramme der Originalbilder und der konvertierten Tiefenund Thermalbilder, Subjekt F001

speichern die Entfernung zur Kamera in mm. Punkte mit der größten Entfernung im Hintergrund haben die höchsten Werte, der Punkt mit der geringsten Entfernung, im Normalfall die Nase, den niedrigsten Wert. Das Tiefenbild zeigt damit ein dunkles Gesicht mit hellem Hintergrund. Die Thermalbilder speichern die Temperatur in °C. Die Punkte, die den, im Gegensatz zum kälteren Hintergrund, warmen Körper beschreiben, speichern daher höhere Werte als Punkte des Hintergrunds. Somit wird hier der Hintergrund dunkel und der Körper hell dargestellt. In den Histogrammen der Bilder aus Abbildung 4.7 lässt sich gut erkennen, welche Intensitätswerte zum Hintergrund gehören und welche zum Körper. Für eine bessere Vergleichbarkeit werden vor der Anwendung einiger Filter eines oder beide Bilder invertiert, damit in beiden Fällen der Hintergrund oder der Körper entweder hell oder dunkel ist. Um ein Bild zu invertieren, werden die Intensitätswerte des Bildes von dessen Maximalwert subtrahiert. Ein Pixel des invertierten Bildes berechnet sich somit durch $img_{neg}[y][x] = maxintensity - img[y][x]$. Dadurch werden die Intensitätswerte negiert. Kleine Werte werden groß und umgekehrt. Abbildung 4.8 zeigt die invertierten Bilder anhand des Beispiels von Subjekt F001.

4.5.3 Binärbilder

Eine Idee für die Vorverarbeitung der Bilder ist aus diesen Binärbilder zu generieren, da in beiden Bildern die Konturen zu erkennen sind. Dadurch sollen die Bilder segmentiert werden.



(a) Invertiertes Tiefenbild



(b) Invertiertes Thermalbild

Abbildung 4.8: Invertiertes Tiefen- und Thermalbild, Subjekt F001

Bei einer Segmentierung werden die Bilder in verschiedene nicht-überlappende zusammengehörende Regionen unterteilt. Diese Regionen haben eine gleiche Charakteristik wie Textur oder Intensität [28]. In medizinischen Bildern werden durch Segmentierungsmethoden im optimalen Fall zusammengehörende anatomische Strukturen gefunden. Bei Schwellenwertansätzen werden die Intensitätswerte des Bildes in binäre Regionen unterteilt. Der Schwellenwert ist der Wert, an dem die beiden Bereiche geteilt werden. Er trennt die beiden gewünschten Klassen voneinander. Alle Pixel mit Intensitätswerten oberhalb dieses Schwellenwerts werden dann in die eine Klasse und die restlichen Pixel in die andere Klasse eingeteilt [28].

In dieser Arbeit wurden verschiedene Schwellenwertverfahren getestet. Im Folgenden wird auf einen globalen und einen adaptiven lokalen Schwellenwert genauer eingegangen. Zusätzlich wurden ein Otsu und ein Voronoi Schwellenwertverfahren getestet, welche jedoch kein nennenswertes Ergebnis erbrachten und deshalb auch nicht weiter ausgeführt werden.

Globaler Schwellenwert

Ein globaler Schwellenwert gilt für das gesamte Bild. Die Pixel des Bildes werden anhand des Schwellenwerts in zwei Klassen unterteilt. An jedem Pixel wird überprüft, ob der Intensitätswert über oder unter dem Schwellenwert liegt. Wenn der Intensitätswert kleiner ist, wird der Wert an der Stelle auf 0 gesetzt, sonst auf den Maximalwert⁸ [29].

⁸Für das globale Schwellenwertverfahren wird die OpenCV-Methode "threshold()" verwendet.

Bevor das globale Schwellenwertverfahren angewendet wird, werden sowohl Tiefen- als auch Thermalbild zu 8-bit Bildern konvertiert. Das Tiefenbild wird zusätzlich zur Konvertierung invertiert, damit für die Erzeugung der Binärbilder ein einheitlicher Schwellenwert verwendet werden kann. Verschiedene globale Schwellenwerte werden getestet. Beim Erzeugen der Testdaten wird ein Schwellenwert von 40 verwendet, da er bei einzelnen Testbildern die besten Ergebnisse erzielte.

Abbildung 4.9 zeigt global binarisierte Bilder am Beispiel von Subjekt F001. Es fällt auf, dass die Umrisse der Person in beiden Bildern unterschiedlich verlaufen. Die Binarisierung mit globalem Schwellenwert des Tiefenbildes ist einfach, da es trotz künstlich erzeugtem Rauschen im Hintergrund einen deutlichen Sprung zwischen Hintergrund und Versuchsperson gibt. Bei den Thermalbildern ist dieses Verfahren nicht so leicht, da der Körper Wärme abstrahlt. Wie gut zu erkennen ist, ist es schwierig einen globalen Schwellenwert zu finden. Im Bereich des Halses wird hier die Wärmestrahlung, welche als Rauschen im Bild zu erkennen ist, fälschlicherweise zum Körper gezählt. Hingegen werden die Haare der Versuchsperson nicht hinzugezählt. Wenn der Schwellenwert erhöht wird, werden zwar die Haare als Körper wahrgenommen, jedoch wird dann das Rauschen im unteren Bereich des Bildes zu groß. Wird der Schwellenwert kleiner, wird das Rauschen um den Hals weniger, aber ebenso die Haare. Für das Thermalbild wird somit kein optimaler globaler Schwellenwert gefunden.



(a) Global binarisiertes Tiefenbild



(b) Global binarisiertes Thermalbild

Abbildung 4.9: Binärbilder mit globalem Schwellenwert, Subjekt F001

Adaptiver lokaler Schwellenwert

Wenn ein globaler Schwellenwert keine guten Ergebnisse liefert, kann ein adaptives lokales Schwellenwertverfahren eingesetzt werden [29]. Das Bild wird in kleine Regionen

unterteilt. Für jede Region wird ein individueller Schwellenwert bestimmt⁹.

Vor der Verwendung des lokalen Schwellenwertverfahrens werden beide Bilder mit einem Gaussfilter geglättet, um Hintergrundrauschen zu verringern. Ebenfalls werden beide Bilder zu 8-bit Bildern konvertiert, da die Methode ein 8-bit Bild als Übergabeparameter erwartet. Eine anschließende Filterung mit einem Medianfilter wird angewendet, um weiteres Rauschen zu reduzieren.

Abbildung 4.10 zeigt die Tiefen- und Thermalbilder von Subjekt F001 nach der Filterung mit dem adaptiven lokalen Schwellenwertverfahren.



(a) Lokal binarisiertes Tiefenbild

(b) Lokal binarisiertes Thermalbild

Abbildung 4.10: Binärbilder mit adaptivem lokalen Schwellenwert, Subjekt F001

4.5.4 Morphologische Filter

Die theoretischen Grundlagen zu morphologischen Operationen finden sich in Kapitel 3.1. Dieser Ansatz zur Vorverarbeitung der Bilder basiert auf der dort vorgestellten Arbeit [16]. Bei diesem Vorverarbeitungsschritt wird eine Bearbeitung der Bilder mit dem morphologischen Gradienten vorgenommen. Hierzu werden beide Bilder erst mit einem Gaussfilter geglättet, um Rauschen zu reduzieren. Anschließend wird ein Gradientenbild erzeugt¹⁰. Wie schon erwähnt, beschreibt der morphologische Gradient die Differenz zwischen Dilatation und Erosion eines Bildes. Dadurch werden die Konturen der Versuchsperson im Bild hervorgehoben.

⁹Für das adaptive lokale Schwellenwertverfahren wird die OpenCV-Methode "adaptiveThreshold()" verwendet.

¹⁰Für die morphologischen Gradientenbilder wird die OpenCV-Methode "morphologyEx()" verwendet

Abbildung 4.11 zeigt die Tiefen- und Thermalbilder von Subjekt F001 nach der Anwendung des morphologischen Gradienten.





(a) Morphologisches Gradientenbild Tiefenbild (b) Morphologisches Gradientenbild Thermalbild

Abbildung 4.11: Tiefen- und Thermalbild als morphologische Gradientenbilder, Subjekt F001

4.5.5 Histogram Matching

Das Histogram Matching Verfahren ist in Kapitel 3.2 beschrieben. Die Idee dafür basiert auf [19]. Bei diesem Vorverarbeitungsschritt soll das Histogramm eines Bildes an das Histogramm eines Referenzbildes angepasst werden¹¹.

Vor dem Histogram Matching werden beide Bilder zu 8-bit Bildern konvertiert, um einen groben Ausgleich der Histogramm-Wertebereiche zu erlangen. Das Verfahren wird sowohl mit den invertierten als auch mit den nicht-invertierten Bildern und den daraus entstehenden verschiedenen Kombinationen getestet. Für den weiteren Verlauf wird das invertierte konvertierte Tiefenbild dem konvertierten Thermalbild angepasst.

Abbildung 4.12 zeigt die Tiefen- und Thermalbilder von Subjekt F001 mit angepasstem Tiefenbild-Histogramm an das Histogramm des Thermalbildes.

4.5.6 Adaptive Equalization

Um die Gesichtsmerkmale in den Bildern hervorzuheben, wird eine "Contrast Limited Adaptive Histogram Equalization" vorgenommen. Dabei soll der Kontrast im Bild vergrößert

¹¹Für das Histogram Matching wird die Skimage-Methode "match_histograms()" verwendet.







(b) Thermal-Referenzbild

Abbildung 4.12: Tiefen- und Thermalbild angepasst mit Histogram Matching, Subjekt F001

werden, ohne den Rauschanteil zu erhöhen.

Bei einer "Histogram Equalization" wird das Histogramm des Grauwertbildes betrachtet. Auffällige Peaks im Histogramm sollen ausgeglichen werden. Intensitätswerte, welche häufig vorkommen (Peaks), sollen auf Intensitätswerte verteilt werden, welche nicht so häufig im Histogramm vertreten sind [30]. Dadurch soll die kummulative Verteilungsfunktion linearisiert werden. Das Histogramm wird somit von Minimal- zu Maximalwert gestreckt. Der Kontrast im Bild wird hierdurch erhöht, gleichzeitig aber auch das Rauschen im Bild [30]. Bei der "Adaptive Histogram Equalization" wird das Bild in mehrere Bereiche unterteilt und für jeden Bereich eine Histogram Equalization vollzogen. Dadurch kann lokal der Kontrast erhöht werden. Nach [30] ist das Problem der "Adaptive Histogram Equalization", dass es die Neigung hat, Rauschen in homogenen Bereichen des Bildes deutlich zu verstärken. Eine Variante der "Adaptive Histogram Equalization" ist die "Contrast Limited Adaptive Histogram Equalization". Durch eine Limitierung der Kontrastverstärkung wird verhindert, dass das Rauschen im Bild deutlich verstärkt wird. Hier wird das Histogramm an einem bestimmten vordefinierten Wert abgeschnitten, bevor die kummulative Verteilungsfunktion berechnet wird [30].

Auch hier werden die Bilder vor der Anwendung der "Contrast Limited Adaptive Histogram Equalization" (CLAHE)¹² zu 8-bit Bildern konvertiert. Diese Methode wird sowohl mit invertierten als auch nicht-invertierten Bildern getestet. Für weitere Schritte werden invertierte und anschließend konvertierte Tiefen- und Thermalbilder verwendet. Die Bilder sehen sich dadurch zwar nicht im Gesamten ähnlicher, aber es wird versucht,

¹²Für die "Contrast Limited Adaptive Histogram Equalization" (CLAHE) werden hier die OpenCV-Methoden "createCLAHE()" und "apply(image)" verwendet.

nur den Bereich von Augen und Nase anzugleichen, da das markante Merkmale in den Gesichtern der Versuchspersonen sind. Abbildung 4.13 zeigt ein Beispiel solcher Bilder an Subjekt F001.



(a) Invertiertes Tiefenbild bearbeitet mit CLAHE



(b) Invertiertes Thermalbild bearbeitet mit CLAHE

Abbildung 4.13: Invertiertes Tiefen- und Thermalbild bearbeitet mit einer Contrast Limited Adaptive Equalization (CLAHE), Subjekt F001

4.5.7 Intensitätstransformationen

Intensitätstransformationen werden verwendet, um die Intensitätswerte der Bilder zu manipulieren und dadurch deren Aussehen zu verändern. Beispielsweise kann so auch der Kontrast erhöht werden. Intensitätstransformationen finden in der räumlichen Domäne statt. Räumliche-Domänen-Methoden beziehen sich auf die Bildebene. Die Bildpixel werden Pixel für Pixel direkt manipuliert [31]. Dieser Prozess kann mit g(x, y) = T[f(x, y)] beschrieben werden [32]. Dabei steht f(x, y) für das Bild, T ist die Operation, welche auf Bild f angewendet wird, und g(x, y) ist das transformierte Bild.

Die Invertierung von Bildern ist am Anfang dieses Kapitels beschrieben, weshalb an dieser Stelle nicht weiter darauf eingegangen wird.

Die Intensitätstransformationen werden im Folgenden in drei Kategorien eingeteilt. Unterschieden wird zwischen logarithmischen, invers logarithmischen und gamma Intensitätstransformationen.

Die logarithmische Intensitätstransformation wird durch die Gleichung $s = c \cdot \log(1+r)$ beschrieben [32]. s ist die transformierte Intensität, r ist die Intensität des Originalbildes und c ist eine Skalierungskonstante mit $c = \frac{maxintensity}{\log(1+maxintensity)}$. Bei der inversen logarithmischen Intensitätstransformation wird jeder Intensitätswert mit der Gleichung $s = \exp(r^{\frac{1}{c}}) - 1$ manipuliert.

Gamma Intensitätstransformationen können durch $s = c \cdot r^{\gamma}$ beschrieben werden [32]. Für einen Wert $\gamma < 1$ können damit Wurzelfunktionen, für $\gamma > 1$ Potenzfunktionen für die Intensitätstransformation verwendet werden. Je nachdem wie γ gewählt wird, kann das Bild somit gebleicht oder verdunkelt werden.

Auch hier werden verschiedene Kombinationen getestet. Im weiteren Verlauf werden drei Kombinationen betrachtet. In Tabelle 4.1 werden der Zweck und die genauen Bearbeitungsschritte für diese Kombinationen vorgestellt.

	Variante 1	Variante 2	Variante 3
Zweck	Kontrasterhöhung in	Kontrasterhöhung in	Kontrasterhöhung in
	beiden Bildern, Be-	beiden Bildern, Be-	beiden Bildern, ge-
	reich von Augen und	reich von Augen und	samtes Bild soll farb-
	Mund soll farblich an-	Mund soll farblich an-	lich angepasst werden
	gepasst werden	gepasst werden	
Tiefenbild	Konvertierung, Gam-	Invertierung, In-	Invertierung, In-
	ma Transformation	vers logarithmische	vers logarithmische
	mit $\gamma = 0.7$	Transformation, Kon-	Transformation, Kon-
		vertierung, CLAHE	vertierung, CLAHE
Thermalbild	Invers logarithmische	Invertierung, Kon-	Invers logarithmische
	Transformation	vertierung, Gamma	Transformation
		Transformation mit	
		γ = 0.6, Glättung	
		durch Gaussfilter	

Tabelle 4.1: Bearbeitungsschritte der Intensitätstransformationen angewendet auf Tiefenund Thermalbilder

Abbildung 4.14 stellt die drei Varianten intensitätstransformierter Bilder an Beispiel von Subjekt F001 dar.



(a) Variante 1 Tiefenbild



(c) Variante 2 Tiefenbild



(e) Variante 3 Tiefenbild



(b) Variante 1 Thermalbild



(d) Variante 2 Thermalbild



(f) Variante 3 Thermalbild

Abbildung 4.14: Intensitätstransformierte Tiefen- und Thermalbilder, Subjekt F001: Dargestellt werden immer in einer Zeile die zusammengehörigen bearbeiteten Tiefen- und Thermalbilder. (a) zeigt das konvertierte Tiefenbild mit einer gamma Transformation mit $\gamma = 0.7$, (b) und (f) zeigen das invers logarithmisch transformierte Thermalbild, (c) und (e) stellen das invertierte, invers logarithmisch transformierte, konvertierte und anschließend mit einer CLAHE bearbeitete Tiefenbild dar, (d) zeigt das Thermalbild, welches invertiert, konvertiert, gamma transformiert mit $\gamma = 0.6$ und anschließend mit einem Gaussfilter geglättet wurde.

4.6 Zusammenhang zwischen Tiefen- und Temperaturwerten

Im weiteren Verlauf werden die multimodalen Daten analysiert, um einen möglichen Zusammenhang zwischen Tiefen- und Temperaturwerten herauszufinden (Ebene 4, Abbildung 4.1). Dafür werden bereits registrierte Bilder verwendet. Für die Analyse der Intensitätswerte werden unter anderem 3D-Darstellungen der Oberfläche und 2D-Histogramme der Bilder und Bildausschnitten erstellt. Außerdem werden verschiedene statistische Merkmale der Bilder untersucht (Ebene 5, Abbildung 4.1).

Für die Untersuchung der statistischen Merkmale werden die registrierten Bilder in kleine Bereiche ("Patches") unterteilt. Aus den Bildern werden mit einer übergebenen Patch-Größe die entsprechenden Patches berechnet und in einem 4-dimensionalen Array gespeichert. Dadurch sind die Patches geordnet und es gibt eine klare Zuordnung, welche Patches aneinander grenzen. Des Weiteren wird auch eine Methode implementiert, die die Patches wieder zu einem Bild zusammenfügt. Für jedes Patch werden verschiedene statistische Merkmale berechnet. Dazu zählen:

- Mittelwert: Der Durchschnittswert der Intensitätswerte
- Median: Der Wert, der genau in der Mitte der Intensitätswerte liegt
- Standardabweichung: Streuung der Intensitätswerte um den Mittelwert
- Varianz: Quadrat der Standardabweichung
- Kurtosis: Beschreibt, ob die Intensitätswertverteilung flacher oder spitzer als eine Normalverteilung verteilt ist
- Schiefe: Beschreibt die Art und Stärke der Asymmetrie einer Intensitätswertverteilung
- verschiedene Bildmomente

Bildmomente In [33] wird die Verwendung von Bildmomenten für die Detektion von Formen in segmentierten Bildern beschrieben. Ein Bildmoment kann dabei im Generellen folgendermaßen berechnet werden [33]:

$$m_{p,q} = NF \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} B(x,y) x^p y^q$$
(4.12)

Dabei steht B(x, y) für eine zweidimensionale Bildfunktion, N bezeichnet die Zeilenund M die Spaltenanzahl. Mit Hilfe der Schwerpunktskoordinaten können diese nicht translationsinvarianten Momente zu zentralen Momenten verarbeitet werden. Die Schwerpunktskoordinaten berechnen sich nach [33] zu $\bar{x} = \frac{m_{1,0}}{m_{0,0}}$ und $\bar{y} = \frac{m_{0,1}}{m_{0,0}}$.

Mit diesen Schwerpunktskoordinaten lassen sich die zentralen Momente definieren. Diese sind translationsinvariant, aber nicht skalierungsinvariant [33]:

$$\mu_{p,q} = \sum_{x=1}^{M} \sum_{y=1}^{N} B(x,y) (x-\bar{x})^p (y-\bar{y})^q$$
(4.13)

Um neben der Translationsinvarianz ebenso eine Skalierungsinvarianz zu erhalten, können die Momente normiert werden. Die normalisierten Momente berechnen sich wie folgt [33]:

$$\eta_{p,q} = \frac{\mu_{p,q}}{m_{0,0}}^{lpha} \quad \text{mit} \quad \alpha = \frac{p+q}{2} + 1$$
(4.14)

Hu-Momente sind Kombinationen aus normalisierten Momenten. Die daraus entstehenden Momente sind translations-, skalierungs- und rotationsinvariant. Die sieben Hu-Momente sind folgermaßen definiert [33]:

$$\begin{split} h_1 &= \eta_{2,0} + \eta_{0,2} \\ h_2 &= [\eta_{2,0} - \eta_{0,2}]^2 + 4[\eta_{1,1}]^2 \\ h_3 &= [\eta_{3,0} - 3\eta_{1,2}]^2 + [\eta_{0,3} - 3\eta_{2,1}]^2 \\ h_4 &= [\eta_{3,0} + \eta_{1,2}]^2 + [\eta_{0,3} - 3\eta_{2,1}]^2 \\ h_5 &= [\eta_{3,0} - 3\eta_{1,2}][\eta_{3,0} + \eta_{1,2}] \{[\eta_{3,0} + \eta_{1,2}]^2 - 3[\eta_{0,3} + \eta_{2,1}]^2\} \\ &+ [3\eta_{2,1} - \eta_{0,3}][\eta_{0,3} + \eta_{2,1}] \{3[\eta_{3,0} + \eta_{1,2}]^2 - [\eta_{0,3} + \eta_{2,1}]^2\} \\ h_6 &= [\eta_{2,0} + \eta_{0,2}] \{[\eta_{3,0} + \eta_{1,2}]^2 - [\eta_{0,3} + \eta_{2,1}]^2\} \\ &+ 4\eta_{1,1}[\eta_{3,0} + \eta_{1,2}][\eta_{0,3} + \eta_{2,1}] \\ h_7 &= [3\eta_{2,1} - \eta_{0,3}][\eta_{3,0} + \eta_{1,2}] \{[\eta_{3,0} + \eta_{1,2}]^2 - 3[\eta_{0,3} + \eta_{2,1}]^2\} \\ &+ [3\eta_{1,2} - \eta_{3,0}][\eta_{0,3} + \eta_{2,1}] \{3[\eta_{3,0} + \eta_{1,2}]^2 - [\eta_{0,3} + \eta_{2,1}]^2\} \end{split}$$

Mit diesen Bildmomenten können Formen in segmentierten Bildern detektiert werden. Das Detektieren von Formen wurde sowohl in den gesamten Bildern als auch in kleinen Bereichen des Bildes getestet¹³.

4.7 Registrierungsverfahren

In diesem Abschnitt werden die getesteten Registrierungsverfahren vorgestellt (Ebene 4, Abbildung 4.1). Die Registrierungsverfahren wurden sowohl an den unbearbeiteten als auch an den bearbeiteten Bildern getestet. Beim Registrierungsprozess wird das Tiefenbild als Bewegtbild und das Thermalbild als Referenzbild angenommen. Für einen Vergleich der Verfahren werden zuerst von allen 140 Subjekten der erste Frame der ersten Aufgabe (T1) manuell mit Matlab registriert. Die daraus entstehenden Transformationsmatrizen dienen als Referenzmatrizen, mit denen die Transformationsmatrizen der zu testenden Verfahren verglichen werden. Es wird davon ausgegangen, dass die optimalen Transformationsparameter der Registrierung nur für den ersten Frame der Videos gefunden werden müssen, da sich die Kamerapositionen während des Videos im Normalfall nicht verändern sollte. Das Testen der Registrierungsverfahren an allen Frames wäre zu zeitintensiv, da allein die Erzeugung der Tiefenbilder sehr zeitaufwändig ist. Die Annahme wird an zwei Videos (F001 und M001, Aufgabe T1) überprüft. Dafür werden alle Tiefenbilder-Frames mit der ausgerechneten Transformation des ersten Frames transformiert. Anschließend wird ein übereinandergelagertes Bild von Tiefen- und Thermalbild pro Frame gespeichert und diese übereinandergelagerten Bilder zu einem Video zusammengefügt. Die beiden Videos von registriertem Tiefen- und Thermalbild werden visuell bewertet.

Die getesteten Registrierunsverfahren beinhalten eine manuelle Registrierung, eine merkmalsbasierte Registrierung und eine intensitätsbasierte Registrierung (Ebene 5, Abbildung 4.1) und werden im Folgenden genauer beschrieben.

4.7.1 Manuelle Registrierung

Um eine Referenztransformationsmatrix zu erhalten, welche die Bilder einigermaßen gut registriert, wird eine manuelle Registrierung der ersten Frames von Aufgabe T1 von allen Subjekten durchgeführt. Für die manuelle Registrierung wird jeweils ein erzeugtes Tiefen-

¹³Die Bildmomente werden mit der OpenCV-Methode "moments()", Hu-Momente mit der OpenCV-Methode "HuMoments()" und die Ähnlichkeit von Formern mit der OpenCV-Methode "matchShapes()" berechnet.

und Thermalbild in Matlab geladen. Anschließend werden Landmarken in Bewegt- und Referenzbild gewählt¹⁴. Die Koordinaten dieser Landmarken werden gespeichert und daraus eine Transformationsmatrix berechnet¹⁵. Dabei muss angegeben werden, welche Art von Transformation verwendet werden soll. Hier wird eine affine Transformation gewählt. Diese wird genutzt, da die Tiefenbilder durch das Anpassen der Kameraentfernung eine ähnliche Perspektive wie die Thermalbilder haben.

Für eine affine Transformation werden mindestens drei Punkte benötigt, welche nicht voneinander abhängig sind. Bei der manuellen Registrierung werden acht Punkte gewählt, die sowohl das Gesicht des Subjekts von außen (also die Umrisse der Person), als auch die Gesichtsmerkmale definieren. Dafür werden die Augeninnenwinkel, die untere Abgrenzung der Nase, der mittlere Punkt der Lippen, das Kinn, ein Punkt am oberen Rand des Kopfes und auf beiden Seiten jeweils ein Punkt (oftmals im Bereich der Ohren) gewählt. Abbildung 4.15 zeigt die gewählten Landmarken an dem Beispiel von Subjekt M001.



Abbildung 4.15: Auswahltool für eine manuelle Registrierung mit Landmarken, Subjekt M001: Für jede Versuchsperson werden acht Landmarken gewählt. Aus diesen Landmarken wird eine affine Transformationsmatrix für die manuelle Registrierung berechnet.

Die erzeugte Matrix ist eine 3×3 Transformationsmatrix. Im weiteren Verlauf wird die Transformationsmatrix in Python benötigt. Daher wird die Matlab-Matrix als Textdatei gespeichert und in Python eingelesen. Das Bewegtbild kann mit einer 3×3 -Matrix verzerrt

¹⁴Das Wählen von Landmarken wird mit der Matlab-Funktion "cpselect" durchgeführt.

¹⁵Die Transformationsmatrix wird mit der Matlab-Funktion "fitgeotrans" berechnet.

werden¹⁶. Da in OpenCV eine affine Transformationsmatrix als 2×3 -Matrix definiert ist, eine affine Transformationsmatrix jedoch einer perspektivischen 3×3 -Transformationsmatrix mit der untersten Zeile als (0, 0, 1) entspricht, kann auch eine affine 3×3 -Matrix für die Verzerrung verwendet werden. Da die Matlab-Matrix anders definiert ist als die für die Verzerrung benötigte Matrix, muss die Matlab-Matrix transponiert werden. Die transponierte Matrix wird als Numpy-Array gespeichert. Anschließend wird für eine visuelle Beurteilung der Registrierung ein übereinandergelagertes Bild bestehend aus Thermalbild und verzerrtem Tiefenbild gespeichert¹⁷.

Ein übereinandergelagertes Bild der manuellen Registrierung in Matlab von Subjekt M001 mit verschiedenen Darstellungsformen zeigt Abbildung 4.16. Das erste Bild zeigt die beiden Originalbilder übereinandergelagert. Das zweite und dritte Bild werden jeweils mit einer Schachbrettmuster-Maske bearbeitet, sodass abwechselnd Tiefen- und Thermalbild abgebildet ist. Dabei werden für das zweite Bild die Bilder mit einer Intensitätstransformation nach Variante 3 bearbeitet, da dadurch die Gesichtsmerkmale des Tiefenbildes besser zur Geltung kommen und der Hintergrund ebenfalls dunkel ist. Das dritte Bild stellt die übereinandergelagerten morphologischen Gradientenbilder dar. Dadurch sind die Konturen der Person besser zu erkennen.



Abbildung 4.16: Manuell registrierte Bilder, Subjekt M001: Es wurden unterschiedliche Darstellungsformen derselben manuellen Registrierung gewählt. Dargestellt sind links die übereinandergelagerten Originalbilder, in der Mitte die intensitätstransformierten Bilder nach Variante 3 als Schachbrettdarstellung und rechts die morphologischen Gradientenbilder ebenfalls als Schachbrettdarstellung.

¹⁶Das Verzerren eines Bildes wird mit OpenCV-Methode "warpPerspective()" durchgeführt.

¹⁷Das Tiefenbild wird dafür mit der OpenCV-Methode "warpPerspective()" verzerrt und mit der OpenCV-Methode "addWeighted()" übereinandergelagert.

4.7.2 Merkmalsbasierte Registrierung

An einzelnen Bildern werden mehrere Merkmalsdetektoren ausprobiert. Dabei werden bestehende Verfahren von Matlab mit der "Matlab Registration Estimator App" sowie Verfahren in Python mit OpenCV und Skimage verwendet. Eine Merkmalsdetektion wird mit SURF, SIFT, ORB, FAST, BRIEF und HARRIS versucht. Oftmals werden keine oder nur wenige Merkmale gefunden. Für weitere Tests wird ausschließlich der SIFT-Algorithmus von OpenCV genutzt, da hier die meisten Merkmale gefunden werden und somit die Ergebnisse am vielversprechendsten sind.

Globale Registrierung mit SIFT

Eine globale Registrierung mit SIFT wird an den unbearbeiteten und bearbeiteten Bildern durchgeführt. Dafür werden die Bilder eingelesen und zu 8-bit-Bildern konvertiert. Der SIFT Merkmalsdetektor- und deskriptor wird erzeugt. Es werden Merkmale in den zu registrierenden Bildern gesucht und versucht, passende Merkmale zuzuordnen¹⁸. Beim Matching-Prozess gibt es verschiedene Vorgehensweisen. [34] beschreibt zwei Matching-Verfahren. Unterschieden wird zwischen einem "Brute-Force Matcher" und einem "FLANN based Matcher". In dieser Arbeit wird der "FLANN based Matcher" mit dem Vorgehen aus [35] verwendet. Im Normalfall werden hier nur Übereinstimmungen in Betracht gezogen, die als gut angesehen werden. Nach [15] sind das die Merkmale, die am nächsten aneinander liegen und dadurch ein bestimmtes Verhältnis der Distanzen erfüllen. Bei den vorliegenden Daten werden mit diesem Kriterium jedoch entweder keine oder nicht ausreichend viele Übereinstimmungen gefunden, weshalb alle Merkmale berücksichtigt werden müssen. Die Transformationsmatrix¹⁹ und ein Bild der zugeordneten Merkmale werden erzeugt. Die Matrix ist in diesem Fall eine perspektivische Transformationsmatrix und keine affine Matrix.

Registrierung in Bildausschnitten

Zusätzlich zu den gesamten Bildern werden lokale Bildausschnitte betrachtet. Dafür werden die Tiefen- und Thermalbilder einzelner Subjekte so ausgeschnitten, dass nur die Gesichtsmerkmale von Augen und Nase zu sehen sind. In diesen Bildausschnitten wird

¹⁸Das Detektieren von Merkmalen und das anschließende Matching geschieht mit der OpenCV-Methode "detectAndCompute()".

¹⁹Die Transformationsmatrix wird mit "findHomography()" berechnet.

erneut eine Registrierung mit dem SIFT-Algorithmus getestet. Es wird untersucht, ob die Bilder nach der Vorverarbeitung im genannten Bildausschnitt so ähnlich sind, dass eine merkmalsbasierte Registrierung mit SIFT verbessert werden kann.

4.7.3 Intensitätsbasierte Registrierung

Eine intensitätbasierte Registrierung mit der Mutual Information Metrik als Ähnlichkeitsmaß wird sowohl in Matlab als auch in Python mit SimpleITK getestet. Hierbei schneidet die Registrierung in Matlab besser ab, da sie einerseits in mehreren Durchläufen mit denselben Bildern dieselben Ergebnisse erzielt und außerdem die Registrierung an sich besser funktioniert. Die Registrierung mit SimpleITK in Python funktioniert mit den gegebenen Daten nicht so zuverlässig wie in Matlab.

Für die intensitätsbasierte Registrierung in Matlab wird die entsprechende Matlab-Funktion in Python aufgerufen. Es werden die gespeicherten Tiefen- und Thermalbilder verwendet. Die RGB-Bilder werden zu Grauwertbildern umgewandelt. Anschließend werden Optimierer und Metrik erzeugt²⁰. Hierfür wird als Modus "multimodal" gewählt. Das entspricht einer Mutual Information Metrik und einem "One Plus One Evolutionary" Optimierer. An einzelnen Testbildern wird eine bessere Registrierung erzeugt, indem der "InitialRadius" und die "MaximumIterations" des Optimierers angepasst werden. InitialRadius wird daher auf InitialRadius/3.5 und MaximumIterations auf 300 gesetzt wie in [36]. Die Transformationsmatrix wird berechnet²¹. Es wird eine affine Transformation gewählt. Genau wie bei der manuellen Registrierung wird auch hier die Transformationsmatrix als Textdatei gespeichert, von Python eingelesen, transponiert und als Numpy-Array mit zusätzlich übereinandergelagertem Bild der Registrierung abgespeichert.

4.8 Bewertung der Registrierung

Die Ergebnisse der Registrierung werden auf verschiedene Weise bewertet (Ebene 6, Abbildung 4.1). Zum einen kann eine visuelle Betrachtung der übereinandergelagerten registrierten Bilder erfolgen. Dabei können die Bilder als Ganzes übereinandergelegt werden oder immer abwechselnde Kacheln eines der beiden Bilder als Schachbrettmusterdarstellung betrachtet werden. Durch die visuelle Beurteilung kann eine Einschätzung gegeben

²⁰Optimierer und Metrik werden mit der Matlab-Funktion "imregconfig" erzeugt.

²¹Das Berechnen der Transformationsmatrix geschieht mittels der Matlab-Funktion "imregtform".

werden, ob die Registrierung gut ist und auch Vergleiche zwischen den Registrierungsverfahren und Filtern gezogen werden. Im Folgenden werden zwei weitere Möglichkeiten zur Evaluation der Registrierungsergebnisse vorgestellt.

4.8.1 Bewertung anhand von Ähnlichkeitsmetriken

Das Problem an der visuellen Betrachtung der Registrierung ist, dass es nur eine qualitative und keine quantitative Bewertung ist. Daher kann diese subjektiv ausfallen und ist bei einer großen Menge an Daten schwierig zu verwenden.

Für eine quantitative Bewertung der Registrierung kann wie in [19] eine Ähnlichkeitsmetrik zur Bewertung der Registrierung verwendet werden. Die Ähnlichkeitsmetrik gibt dabei an, wie ähnlich sich die Bilder sind. Es wird davon ausgegangen, dass die Registrierung genauer ist, je ähnlicher sich die Bilder sind. In den Grundlagen zur Bildregistrierung wurde bereits beschrieben, dass nicht jede Ähnlichkeitsmetrik bei multimodalen Daten, aufgrund der Unterschiede in den Bildintensitäten, genutzt werden kann. Für die Untersuchung der Ähnlichkeit werden hier die normalisierte Mutual Information (NMI), die strukturelle Ähnlichkeit (SSIM) sowie die Methode des kleinsten quadratischen Fehlers (MSE) verwendet, wobei hauptsächlich die NMI-Metrik relevant ist, da diese gut für multimodale Daten verwendet werden kann. Die Originalbilder werden auf ihre Ähnlichkeit untersucht. Es wird die Ähnlichkeit der Bilder nach der Transformation des Bewegtbildes mit der berechneten Transformationsmatrix der verschiedenen Registrierungsverfahren berechnet.

4.8.2 Bewertung anhand einer Referenztransformationsmatrix

Eine weitere Methode zur qualitativen Auswertung besteht darin, die bei der Registrierung entstandene Transformationsmatrix mit einer "optimalen" Transformationsmatrix zu vergleichen. Da die Originalbilder manuell registriert sind, kann ein Vergleich zwischen den Transformationsmatrizen der verschiedenen Registrierungsverfahren gezogen werden, wobei die manuelle Transformationsmatrix als "optimal" angesehen wird. Für den Vergleich muss überprüft werden, wie ähnlich sich diese Transformationsmatrizen sind. Mit der Ähnlichkeit der Matrix ist gemeint, wie ähnlich diese Matrizen Punkte transformieren. Für eine ähnliche Transformation wird davon ausgegangen, dass die Distanz zwischen einem Punkt p transformiert mit der Referenzmatrix R und demselben Punkt p transformiert mit Matrix A möglichst klein sein soll. Am ähnlichsten wären die Matrizen also, wenn die

Punkte nach der Transformation an derselben Stelle lägen. Es wird die Transformation von neun Punkten im Bild untersucht. Abbildung 4.17 zeigt die gewählten Punkte.



Abbildung 4.17: Lage der gewählten Punkte für den Vergleich der Matrizen im Bild: Die Punkte werden mit Referenz- und Vergleichsmatrix transformiert. Aus den transformierten Punkten wird die mittlere quadratische Distanz als Ähnlichkeitswert bestimmt.

Die z-Komponente der Punkte wird auf 1 gesetzt. Jeder dieser neun Punkte wird anschließend sowohl mit der Referenztransformationsmatrix aus der manuellen Registrierung und mit der zu untersuchenden Transformationsmatrix transformiert:

$$\begin{bmatrix} x_T \\ y_T \\ z_T \end{bmatrix} = \begin{bmatrix} m_{1,1} & m_{1,2} & m_{1,3} \\ m_{2,1} & m_{2,2} & m_{2,3} \\ m_{3,1} & m_{3,2} & m_{3,3} \end{bmatrix} \times \begin{bmatrix} x_P \\ y_P \\ z_P \end{bmatrix}$$
(4.15)

Daraus entstehen jeweils zwei transformierte Punkte $p_{T,R}$ und $p_{T,A}$. Um zu berechnen, wie weit die Punkte auseinanderliegen, wird die Distanz zwischen den Punkten berechnet:

$$dist = \sqrt{(x_{T,R} - x_{T,A})^2 + (y_{T,R} - y_{T,A})^2 + (z_{T,R} - z_{T,A})^2}$$
(4.16)

Eine kleine Distanz zwischen den Punkten spricht für eine ähnliche Transformationsmatrix. Die neun berechneten Distanzen werden quadriert. Anschließend wird der Mittelwert der quadrierten Distanzen als Ähnlichkeitswert genutzt.

5 Ergebnisse

In diesem Kapitel werden die Ergebnisse der verwendeten Methoden dargestellt. Zunächst wird der Zusammenhang zwischen Tiefen- und Temperaturwert beschrieben. Anschließend wird auf die verschiedenen getesteten Registrierungsverfahren eingegangen.

5.1 Zusammenhang zwischen Tiefen- und Temperaturwerten

5.1.1 Oberflächendarstellung

Um einen Zusammenhang zwischen Tiefen- und Temperaturwerten zu untersuchen, werden die 3D-Oberflächendarstellungen einiger Versuchspersonen erstellt. Gegenübergestellt werden dabei die Intensitätswertverteilungen von Tiefen- und Thermalbild. Dabei beschreibt die x-Achse die Bildhöhe und die y-Achse die Bildbreite. Mit der z-Achse werden die Intensitätswerte als Höhenverteilung dargestellt. Bei den Tiefenbildern sind hier also die Tiefenwerte und bei den Thermalbildern die Temperaturwerte abgebildet. Durch die Höhenverteilung wird visuell überprüft, ob in manchen Bereichen der Bilder ein ähnliches Verhalten der Intensitätswerte festgestellt werden kann. Abbildung 5.1 zeigt einige Beispiele für die dreidimensionale Darstellung der Intensitätswerte der Bilder. Dabei werden die invertierten Tiefenbilder für einen besseren visuellen Vergleich verwendet. Es werden vier Versuchspersonen für die Darstellung ausgewählt. Davon sind zwei Versuchspersonen weiblich und zwei männlich.

In den 3D-Oberflächendarstellungen der Tiefenbilder gibt es keinen fließenden Übergang zwischen den Tiefenwerten des Gesichts der Versuchsperson und dem Hintergrund. Damit kann klar zwischen Hintergrund und Versuchsperson unterschieden werden. Innerhalb der Gesichter ist bei den invertierten Tiefenbildern ein deutlicher Verlauf vom nächsten Wert (Nase) zum am weitesten entfernten Wert zu sehen. Im Gegensatz dazu ist in den Thermalbildern kein klarer Verlauf der Temperaturwerte in den Gesichtern der Versuchspersonen zu sehen. Es kann zwar ebenfalls zwischen warmer Person und kaltem Hintergrund unterschieden werden, jedoch gibt es hier keine klare Kante, welche beide Bereiche voneinander abtrennt. Auch gibt es keinen Punkt im Gesicht der Versuchspersonen, welcher deutlich am wärmsten oder am kältesten ist. Abbildung 5.1b zeigt, dass die Mitte des Gesichts beispielsweise nicht unbedingt immer am wärmsten sein muss. Hier ist der Bereich um die Nase eindeutig kälter als Stirn und Kinn. Die verschiedenen Abbildungen machen somit deutlich, dass es keine allgemeinen warmen oder kalten Regionen in den Gesichtern der Versuchspersonen gibt.

Insgesamt kann kein eindeutiger Zusammenhang zwischen den Tiefen- und Temperaturwerten aus den 3D-Darstellungen der Intensitätswerte der dargestellten Versuchspersonen hergeleitet werden. Im Bereich des Hintergrunds gibt es die Übereinstimmung, dass dieser sowohl am kältesten als auch am weitesten entfernt ist, jedoch gibt es im Bereich des Gesichts keine derartigen Gemeinsamkeiten.



Abbildung 5.1: 3D-Darstellung der Intensitätswerte von Tiefen- und Thermalbildern

5.1.2 Korrelation in 2D-Histogrammen

Histogramme des gesamten Gesichts

Neben der dreidimensionalen Darstellung der Intensitätswerte werden zweidimensionale Histogramme erstellt, um eine Gemeinsamkeit der Intensitätswertverteilungen zu detektieren. Hierfür werden die Originalbilder verwendet. Diese werden vorher registriert, um eine möglichst genaue Übereinstimmung zwischen Tiefen- und zugehörigem Temperaturwert zu erlangen. Außerdem werden die Bilder etwas zugeschnitten, um den Hintergrund zu verkleinern, da vor allem der Gesichtsbereich untersucht werden sollte. In den Histogrammen soll die Korrelation von Tiefen- und Temperaturwerten gezeigt werden. Die x-Achse enthält die Tiefeninformation und die y-Achse die Information über die Temperatur. Für die Korrelationsdiagramme werden ebenfalls vier Versuchspersonen genutzt. Darunter sind ebenfalls zwei weibliche und zwei männliche Versuchspersonen.

Die 2D-Histogramme in Abbildung 5.2 zeigen bei allen ausgewählten Versuchspersonen eine große Korrelation vom am weitesten entfernten Wert und einem kalten Wert. Auch lässt sich erkennen, dass es bei diesem Tiefenwert auch einen leichten Verlauf von kalt zu warm gibt. Es gibt also hier keine Veränderung in den Tiefenwerten, aber eine Veränderung in den Temperaturwerten. Außerdem lässt sich eine leichte Korrelation in sehr nahen Werten und warmen Werten erkennen. Diese Korrelation unterscheidet sich in jeder Abbildung leicht.

Histogramme eines Bildausschnitts im Gesicht

Bei mehreren Versuchspersonen werden nochmals kleinere Ausschnitte des Bildes untersucht, um einen möglichen Zusammenhang der Korrelation zwischen nahen und warmen Werten einschätzen zu können. Die Bildausschnitte von sechs ausgewählten Versuchspersonen in Abbildung 5.3 zeigen den Bereich von Augen und Nase.

Passend dazu sind in Abbildung 5.4 die 2D-Histogramme dieser Bildausschnitte zu sehen. Bei den Korrelationsdiagrammen muss beachtet werden, dass die Tiefenbilder immer gleich aussehen. Das liegt daran, dass die Nase im Tiefenbild immer am nächsten ist und der Hintergrund immer am weitesten entfernt. Bei den Thermalbildern kann kein klares Verhalten der Temperaturwerte erkannt werden. Bei den 2D-Histogrammen fällt auf, dass insgesamt die höchste Korrelation in den Bildausschnitten bei den am weitesten entfernten Werten besteht. Es gibt jedoch keinen allgemeinen Zusammenhang.



(c) Subjekt M001

(d) Subjekt M009

Abbildung 5.2: 2D-Histogramme der Intensitätswerte der gesamten Tiefen- und Thermalbilder



Abbildung 5.3: Bildausschnitte von Augen- und Nasenpartie: Pro Subjekt ist links das Tiefenbild und rechts das Thermalbild desselben Bildausschnitts abgebildet. Die Bildausschnitte werden für eine genauere Analyse der Tiefenund Temperaturwerte innerhalb des Gesichts verwendet.



Abbildung 5.4: 2D-Histogramme der Intensitätswerte der ausgeschnittenen Augen- und Nasenpartie der Tiefen- und Thermalbilder



(e) Subjekt M009



Abbildung 5.4: 2D-Histogramme der Intensitätswerte der ausgeschnittenen Augen- und Nasenpartie der Tiefen- und Thermalbilder

5.1.3 Statistische Merkmale

Zusätzlich werden die Gemeinsamkeiten von Patches untersucht. Dafür werden ebenfalls registrierte Bilder verwendet. Es wird daher davon ausgegangen, dass die Patches in beiden Bildern den gleichen Bereich eines Bildes abbilden. Anschließend wird der Zusammenhang der verschiedenen Patches untersucht. Dafür wird über die Patches der Bilder iteriert und jedes Patch auf den Wert des zu untersuchenden Merkmals gesetzt. Die Patches werden also "eingefärbt". Es wird untersucht, ob eine Gemeinsamkeit der statistischen Werte zu erkennen ist. Da davon ausgegangen wird, dass die statistischen Werte in den Patches nicht identisch sind, ergibt sich durch das Einfärben ein guter Überblick über die Gemeinsamkeiten, da beim Visualisieren in Python beide Bilder separat skaliert werden. Dadurch muss aber beim Betrachten der Bilder bei gleichfarbigen Patches beachtet werden, dass diese nicht denselben statistischen Wert haben. Abbildung 5.5 zeigt die eingefärbten Patches mit den Werten des jeweiligen statistischen Merkmals an den Originalbildern. Es wurden Patches der Größe 10×10 Pixel verwendet. Standardabweichung und Varianz zeigen die Konturen der Personen. Eindeutige Gemeinsamkeiten von verschiedenen statistischen Merkmalen, anhand derer ein Matching für eine Registrierung durchgeführt werden könnte, werden nicht erkannt.

Auch bei der Betrachtung von Bildmomenten kann keine weitere Erkenntnis gewonnen werden, weshalb diese auch nicht separat dargestellt sind. Da die Hu-Momente zum Detektieren von Formen in Bildern verwendet und dabei auf segmentierte Bilder angewendet werden, wird ebenfalls die Ähnlichkeit der Formen von den einzelnen Patches berechnet und die Bilder dementsprechend eingefärbt. Da eine allgemeine Segmentierung des Thermalbildes sich als schwierig herausstellt, werden hierfür die morphologischen Gradientenbilder und die mit einem adaptiven Schwellenwert segmentierten Bilder verwendet. Abbildung 5.6 stellt die Ähnlichkeit der Patches anhand der OpenCV-Methode "matchShapes()" dar. Hierfür werden Patches der Größe 20×20 Pixel verwendet. Je kleiner die berechnete Distanz, desto ähnlicher sind sich die Formen in diesem Patch. Es ist klar zu sehen, dass in Teilen des Bildes sehr ähnliche Formen detektiert werden, in anderen jedoch nicht. Eine eindeutige Ähnlichkeit bestimmter Formen wird auch hier nicht erkannt.



Abbildung 5.5: Originalbilder mit eingefärbten Patches in den Werten verschiedener statistischer Merkmale, Subjekt F009: In den registrierten Bildern liegen die zusammengehörenden Patches an derselben Stelle. Eine ähnliche Färbung der Patches deutet auf einen Zusammenhang der Tiefen- und Thermalbilder hin.



(a) Morphologische Gradientenbilder



(b) Adaptiver Schwellenwert

Abbildung 5.6: Ähnlichkeit von Formen in Patches, Subjekt F009: Berechnet wird die Ähnlichkeit der Formen in den einzelnen Patches von registrierten Bildern. Ein kleiner Wert zeigt eine hohe Ähnlichkeit der Formen.

5.2 Merkmalsbasierte Registrierung

5.2.1 Globale Registrierung mit SIFT

Abbildung 5.7 zeigt die globale Registrierung mit dem SIFT-Algorithmus an den unbearbeiteten Originalbildern sowie den vorverarbeiteten Bildern mit verschiedenen Filtern. Dargestellt werden die Ergebnisse der Registrierung von Versuchsperson F001. Wie zu erkennen ist, werden bei dieser Versuchsperson keine richtigen Zuordnungen der Merkmale gefunden. Auch bei anderen Versuchspersonen kann das Verfahren keine guten Ergebnisse erzielen. In den Bildern werden nicht die gleichen Merkmale detektiert, wodurch auch keine guten Übereinstimmungen gefunden werden können. Auch durch Vorverarbeitung der Filter kann kein gutes Matching erzielt werden. Dadurch werden zwar teilweise eine größere Anzahl an Merkmalen gefunden, jedoch werden diese ebenfalls nicht richtig zugeordnet.



(c) Invertiertes Thermalbild

(d) Binärbilder

Abbildung 5.7: Globale Registrierung mit SIFT, Subjekt F001: Gezeigt ist links das Tiefenund rechts das Thermalbild. Mit grün sind die gefunden Matches dargestellt.



(e) Adaptiver Schwellenwert



(f) Morphologische Gradientenbilder



(g) Histogram Matching



(i) Intensitätstransformation Variante 1





(j) Intensitätstransformation Variante 2



(k) Intensitätstransformation Variante 3

Abbildung 5.7: Globale Registrierung mit SIFT, Subjekt F001: Gezeigt ist links das Tiefenund rechts das Thermalbild. Mit grün sind die gefunden Matches dargestellt.

5.2.2 Registrierung mit SIFT in Bildausschnitten

Für die Registrierung in kleinen Bildausschnitten werden die Ausschnitte aus Abbildung 5.3 verwendet. Abbildung 5.8 zeigt die mit SIFT gefundenen Matches in den Bildausschnitten. Es werden nur die Vorverarbeitungsmöglichkeiten abgebildet, welche interessante Ergebnisse erbringen. Dargestellt wird jeweils das Originalbild, das morphologische Gradientenbild und das intensitätstransformierte Bild nach Variante 2.

Bei Versuchsperson F001 werden gute Übereinstimmungen der Gesichtsmerkmale an den Originalbildern gefunden. Bei den anderen Versuchspersonen werden entweder gar keine Übereinstimmungen gefunden oder die gefundenen Merkmale falsch zugeordnet. Bei den morphologischen Gradientenbildern werden insgesamt bei allen Versuchspersonen deutlich mehr Matches gefunden. Diese sind aber nur bei Versuchsperson F001 und M009 gut. Bei den restlichen Versuchspersonen stimmen die zugeordneten Merkmale nicht überein. Die intensitätstransformierten Bilder erbringen keine guten Ergebnisse. Bei Versuchsperson F001 werden zwar vier Übereinstimmungen gefunden, davon sind aber nur zwei richtig zugeordnet. Bei den anderen Versuchspersonen kann der SIFT-Algorithmus mit diesen bearbeiteten Bildern auch keine besseren Ergebnisse erzeugen.



Abbildung 5.8: Lokale Registrierung mit SIFT: Jede Reihe zeigt von links nach rechts den untersuchten Bildausschnitt des Originalbildes, des morphologischen Gradientenbildes und des intensitätstransformierten Bildes nach Variante 2 derselben Versuchsperson. Gezeigt ist links das Tiefen- und rechts das Thermalbild. Mit grün sind die gefunden Matches dargestellt.



(p) Subjekt M013

(q) Subjekt M013

(r) Subjekt M013

Abbildung 5.8: Lokale Registrierung mit SIFT: Jede Reihe zeigt von links nach rechts den untersuchten Bildausschnitt des Originalbildes, des morphologischen Gradientenbildes und des intensitätstransformierten Bildes nach Variante 2 derselben Versuchsperson. Gezeigt ist links das Tiefen- und rechts das Thermalbild. Mit grün sind die gefunden Matches dargestellt.
5.3 Intensitätsbasierte Registrierung

5.3.1 Visuelle Bewertung

Abbildung 5.9 stellt die übereinandergelagerten Bilder der intensitätsbasierten Registrierung in Matlab an Versuchsperson F001 dar. Deutlich zu erkennen ist, dass die registrierten Bilder ohne Vorverarbeitung bei dieser Versuchsperson nicht gut übereinander passen. Ebenfalls ist zu sehen, dass die Registrierung mit den morphologischen Gradientenbildern und den adaptiven Schwellenwertbildern die Originalbilder am besten aufeinander abbilden. Dabei ist jedoch die Registrierung mit den morphologischen Gradientenbildern etwas präziser. Dieser Effekt kann auch bei anderen Versuchspersonen gezeigt werden.



Abbildung 5.9: Intensitätsbasierte Registrierung mit Matlab, Subjekt F001

Auffällig bei der visuellen Bewertung der Registrierung mit morphologischen Gradientenbildern und adaptiven Schwellenwertbildern ist, dass bei einer guten visuellen Registrierung mit morphologischen Gradientenbildern die Registrierung mit adaptiven Schwellenwertbildern in vielen Fällen ebenfalls nicht schlecht ist. Jedoch liegen bei dem Überlagern der registrierten Bilder die Umrisse der Tiefenbilder oftmals etwas zu weit außen, wodurch die Umrisse der beiden Bilder nicht deckungsgleich sind. Abbildung 5.10 verdeutlicht dieses Verhalten am Beispiel von Subjekt F008.



(a) Morphologische tenbilder



(d) Adaptiver Schwellenwert



Gradien- (b) Morphologische Gradien- (c) Morphologische tenbilder





Gradientenbilder



(f) Adaptiver Schwellenwert

Abbildung 5.10: Vergleich der intensitätsbasierten Registrierung mit morphologischen Gradientenbildern und adaptiven Schwellenwertbildern, Subjekt F008

(e) Adaptiver Schwellenwert

Da die Originalbilder nach Verzerrung des Tiefenbildes mit der Transformationsmatrix der Registrierung mit morphologischen Gradientenbilder visuell am passgenausten übereinanderliegen, werden weitere Beispiele für die Registrierung mit dieser Vorverarbeitung gezeigt. Abbildung 5.11 gibt Beispiele für ein gutes Überlagern der Bilder anhand der Registrierung mit morphologischen Gradientenbilder, Abbildung 5.12 zeigt Beispiele für ein ungenaues Überlagern. Pro Versuchsperson sind jeweils die vollständig übereinandergelagerten Originalbilder und die übereinandergelagerten Bilder mit Schachbrettmuster einmal mit den intentisätstransformierten Bildern nach Variante 3 sowie den morphologischen Gradientenbildern abgebildet. Dabei wird das Tiefenbild jeweils mit der berechneten Transformationsmatrix, die sich aus der intensitätsbasierten Registrierung mit den morphologischen Gradientenbildern ergibt, verzerrt. Bei gut übereinandergelagerten Bildern stimmen sowohl die Konturen der Person, als auch die Gesichtsmerkmale überein. Die übereinandergelagerten Originalbilder geben eine gute Orientierung, ob die Bilder gut aufeinander passen. Mit den Schachbrettmuster-Bildern kann überprüft werden, ob auch die Gesichtsmerkmale zusammenpassen. Anhand der Schachbrettmuster-Bilder mit morphologischen Gradientenbildern ist der Verlauf der Konturen der Versuchspersonen besonders deutlich zu erkennen. Sind diese gut übereinandergelagert, ist eine fast kontinuierliche Linie sichtbar. Auch die Gesichtsmerkmale können so gut überprüft werden. Da es sich um multimodale Daten handelt und die Bilder daher sehr unterschiedlich aussehen, muss dies auch in den übereinandergelagerten Bildern berücksichtigt werden.

Bei einer visuellen Auswertung der intensitätsbasierten Registrierung mit morphologischen Gradientenbildern kann für ca. 57 % aller 140 Versuchspersonen eine akzeptable Registrierung erreicht werden. Es muss beachtet werden, dass dieser Wert einer subjektiven Beurteilung entstammt.



Abbildung 5.11: Beispiele für eine präzise intensitätsbasierte Registrierung mit Matlab anhand morphologischer Gradientenbilder



Abbildung 5.12: Beispiele für eine unpräzise intensitätsbasierte Registrierung mit Matlab anhand morphologischer Gradientenbilder

5.3.2 Ähnlichkeit der registrierten Bilder

Abbildung 5.13 stellt die Auswertung der Ähnlichkeit der transformierten Bilder anhand der Ähnlichkeitsmetrik "Normalized Mutual Information" dar. Auf der x-Achse sind alle Versuchspersonen aufgetragen. Der erste Frame von Aufgabe T1 wird von jeder Versuchsperson mit den Originalbildern und den vorverarbeiteten Bildern registriert. Die daraus entstehenden Transformationsmatrizen werden jeweils auf die Originalbilder angewendet, um die Ähnlichkeit nach der Transformation zu überprüfen. Auf der y-Achse ist der passende Ähnlichkeitsmetrik-Wert aufgetragen. NMI kann Werte zwischen 1 und 2 annehmen. Je höher der Wert, desto ähnlicher sind sich die Bilder. Als Referenz sind jeweils die unregistrierten Bilder in blau abgebildet. Die orangene Menge zeigt die Ähnlichkeit nach der manuellen Registrierung. Diese beiden Mengen sind in jedem Diagramm dieselben und dienen als Vergleich für die intensitätsbasierte und merkmalsbasierte Registrierung. Es ist zu erkennen, dass die manuelle Registrierung bereits zu einer deutlichen Verbesserung der Ähnlichkeit führt. Zur Vollständigkeit wird die Ähnlichkeit nach der merkmalsbasierte ne Registrierung als rote Punktmenge dargestellt. Es wird erneut ersichtlich, dass die

merkmalsbasierte Registrierung keine guten Ergebnisse erbringt. Die Ähnlichkeit nach der merkmalsbasierten Registrierung ist bei allen vorverarbeiteten Bildern und den Originalbildern kleiner als die Originalbilder vor der Registrierung. Die grüne Punktmenge stellt die intensitätsbasierte Registrierung dar. Es ist zu sehen, dass die intensitätsbasierte Registrierung in allen Fällen eine verbesserte Ähnlichkeit gegenüber den unregistrierten Bildern bringt. Ebenso auffällig ist, dass die intensitätsbasierte Registrierung in vielen Fällen sogar bessere Ähnlichkeitswerte erzielt als die manuelle Registrierung. In den Diagrammen zu den morphologischen Gradientenbildern und den adaptiven Schwellenwertbildern ist zu erkennen, dass die intensitätsbasierte Registrierung nicht durchschnittlich bessere Ähnlichkeitswerte als die manuelle Registrierung nicht durchschnittlich bessere

Einen Vergleich der mittleren Ähnlichkeitswerte für jede Vorverarbeitungsmöglichkeit zeigt Tabelle 5.1. Erneut ist zu erkennen, dass die Ähnlichkeitswerte nach einer merkmalsbasierten Registrierung mit SIFT erheblich kleiner als bei den anderen Registrierungsverfahren, sogar kleiner als bei den nicht registrierten Bildern, sind. Die größte Ähnlichkeit wird in den Originalbildern nach einer intensitätsbasierten Registrierung mit den intensitätstransformierten Bildern nach Variante 2 detektiert. Die intensitätsbasierte Registrierung mit den morphologischen Gradientenbildern erzielt sehr ähnliche Ähnlichkeitswerte wie eine manuelle Registrierung.

Vorverarbeitungsoption	nicht re-	manuell	intensitäts	ba siert kmalsbasiert
	gistriert			
Original	1.0426	1.1104	1.1328	1.0363
Invertiertes Tiefenbild	-	-	1.1324	1.0344
Invertiertes Thermalbild	-	-	1.1326	1.0328
Binärbilder	-	-	1.1186	1.0313
Adaptiver Schwellenwert	-	-	1.1054	1.0382
Morphologische Gradientenbilder	-	-	1.1125	1.0299
Histogram Matching	-	-	1.1343	1.0370
CLAHE	-	-	1.1297	1.0356
Intensitätstransformation Variante 1	-	-	1.1327	1.0355
Intensitätstransformation Variante 2	-	-	1.1345	1.0364
Intensitätstransformation Variante 3	-	-	1.1334	1.0364

Tabelle 5.1: Vergleich der Ähnlichkeit der Bilder anhand der Normalized Mutual Information Metrik: Für jede Vorverarbeitungsmöglichkeit ist der Mittelwert der Ähnlichkeitswerte dargestellt.







(f) Morphologische Gradientenbilder

Abbildung 5.13: Auswertung der Ähnlichkeit der transformierten Bilder anhand der Normalisierten Mutual Information Ähnlichkeitsmetrik: Ein hoher Wert bedeutet eine höhere Ähnlichkeit der Bilder.









(i) Intensitätstransformation Variante 1



(j) Intensitätstransformation Variante 2



(k) Intensitätstransformation Variante 3

Abbildung 5.13: Auswertung der Ähnlichkeit der transformierten Bilder anhand der Normalisierten Mutual Information Ähnlichkeitsmetrik: Ein hoher Wert bedeutet eine höhere Ähnlichkeit der Bilder.



5.3.3 Vergleich der Transformationsmatrizen

Abbildung 5.14 stellt die mittleren quadratischen Distanzen zwischen den transformierten Punkten der manuellen Referenztransformationsmatrix und der zu untersuchenden Transformationsmatrix der intensitätsbasierten Registrierung, wie in Kapitel 4.8.2 beschrieben, dar. Auf der x-Achse sind die Versuchspersonen aufgetragen. Pro Versuchsperson wird die Registrierung von dem ersten Frame der ersten Aufgabe (T1) untersucht, da für diese Bilder eine manuelle Registrierung vorhanden ist. Jeder Punkt in den Diagrammen gibt die mittlere quadratische Distanz zwischen den Matrizen an. Mit den mittleren quadratischen Distanzen soll berechnet werden, wie ähnlich sich die Transformationsmatrizen von manueller Referenzregistrierung und intensitätsbasierter Registrierung sind. Sollten die Punkte nach der Transformation mit den beiden Matrizen exakt aufeinanderliegen, ist die mittlere quadratische Distanz 0. Je höher die mittlere quadratische Distanz ist, desto weiter entfernt sind die transformierten Punkte voneinander. Eine kleine mittlere quadratische Distanz wird als gut angenommen und spricht für eine ähnliche Transformation wie die Transformation der manuellen Referenzregistrierung.

Um einen besseren Vergleich zwischen den einzelnen Filtern zu erlangen, wird der Mittelwert der mittleren quadratischen Distanzen jeder Vorverarbeitungsmethode in einem Balkendiagramm in Abbildung 5.15 visualisiert. Außerdem sind für den Vergleich der Vorverarbeitungsmöglichkeiten sowohl Mittelwert als auch Standardabweichung in Tabelle 5.2 dargestellt. Die Mittelwerte geben an, wie gut die Registrierung insgesamt in Bezug auf die manuelle Referenzregistrierung funktioniert. Ein kleiner Wert spricht für eine gute Registrierung. Bei der Vorverarbeitung mit dem morphologischen Gradienten sind die Mittelwerte der mittleren quadratischen Distanzen am kleinsten. Nach diesem Bewertungskriterium hat dieses Verfahren insgesamt die ähnlichste Transformationsmatrix zur manuellen Referenzmatrix. Auch die Vorverarbeitung mit einem adaptiven Schwellenwert kann Ergebnisse erzielen, die gut an die manuelle Referenztransformation herankommen. Eine Registrierung mit Binärbildern folgt diesen beiden Vorverarbeitungsmethoden bei der Betrachtung des Mittelwertes. Die Standardabweichung bei diesen beiden Methoden ist jedoch im Vergleich zu der Registrierung mit den morphologischen Gradientenbildern sehr hoch. Die restlichen Vorverarbeitungsmethoden zeigen sowohl hohe Mittelwerte als auch hohe Standardabweichungen. Es wird versucht einen Schwellenwert zu finden, ab dem eine Registrierung akzeptiert werden kann. Wird für diesen Schwellenwert eine mittlere quadratische Distanz von 200 gewählt, erhält man ca. 54 % gut registrierte Daten für die Vorverarbeitung mit einem morphologischen Gradienten und ca. 40 % für die Vorverarbeitung mit einem adaptiven Schwellenwert.



Abbildung 5.14: Bewertung der Registrierung mit der manuellen Referenztransformationsmatrix: Neun Punkte werden sowohl mit der manuellen Referenzmatrix als auch der Vergleichsmatrix transformiert. Zwischen den Punkten wird die mittlere quadratische Distanz berechnet und pro Versuchsperson dargestellt. Eine kleine mittlere quadratische Distanz spricht für eine ähnliche Registrierung zur manuellen Registrierung.

81



Abbildung 5.14: Bewertung der Registrierung mit der manuellen Referenztransformationsmatrix: Neun Punkte werden sowohl mit der manuellen Referenzmatrix als auch der Vergleichsmatrix transformiert. Zwischen den Punkten wird die mittlere quadratische Distanz berechnet und pro Versuchsperson dargestellt. Eine kleine mittlere quadratische Distanz spricht für eine ähnliche Registrierung zur manuellen Registrierung.

82



Vergleich der Mittelwerte der mittleren quadratischen Distanzen

Abbildung 5.15: Vergleich der mittleren quadratischen Distanzen der intensitätsbasierten Registrierung: (von links nach rechts) Original, Invertiertes Tiefenbild, Invertiertes Thermalbild, Morphologische Gradientenbilder, Binärbilder, Adaptiver Schwellenwert, Histogram Matching, CLAHE, Intensitätstransformation 1, Intensitätstransformation 2, Intensitätstransformation 3

Vorverarbeitungsoption	Mittelwert	Standardabweichung
Original	8495.42	18541.89
Invertiertes Tiefenbild	7905.94	16754.01
Invertiertes Thermalbild	8315.53	16757.30
Binärbilder	3267.90	6812.29
Adaptiver Schwellenwert	1856.72	6321.60
Morphologische Gradientenbilder	664.45	1572.92
Histogram Matching	8680.70	19465.99
CLAHE	14313.94	24830.19
Intensitätstransformation Variante 1	5353.92	6549.62
Intensitätstransformation Variante 2	9867.83	14883.43
Intensitätstransformation Variante 3	8631.03	12178.87

Tabelle 5.2: Mittelwert und Standardabweichung der mittleren quadratischen Distanzen aus dem Vergleich mit der manuellen Referenzmatrix

In Abbildung 5.16 werden Beispiele für den Vergleich von manueller Referenzregistrierung und intensitätsbasierter Registrierung mit morphologischen Gradientenbildern gezeigt. Dabei wird pro gezeigter Versuchsperson zuerst die manuelle Registrierung, danach die intensitätsbasierte Registrierung und zuletzt ein übereinandergelagertes Bild der beiden Registrierungen mit den eingezeichneten neun Punkten, die für die Auswertung der Matrizen verwendet werden, dargestellt. Die weißen Sternchen zeigen die gewählten untransformierten neun Punkte, die roten Dreiecke beschreiben die transformierten Punkte der manuellen Registrierung und die blauen Kreise die transformierten Punkte der intensitätsbasierten Registrierung. Wie zu erkennen ist, sind die Registrierungen bei Versuchsperson F014 sehr ähnlich. Die mittlere quadratische Distanz liegt hier bei 61.36 und ist damit klein. Bei Versuchsperson M007 ist die Registrierung zwar auch ähnlich, es ist aber eine Rotation bei der intensitätsbasierten Registrierung sichtbar und daher sind die Registrierungen sich etwas unähnlicher als bei Versuchsperson F014. Die mittlere quadratische Distanz beträgt 156.64 und ist damit zwar immer noch nicht sehr hoch, aber höher als bei Versuchsperson F014. Bei Versuchsperson M023 kann eine deutliche Verzerrung bei der intensitätsbasierten Registrierung festgestellt werden. Das äußert sich auch in einer deutlich höheren mittleren quadratischen Distanz von 1617.36. Die intensitätsbasierte Registrierung von Versuchsperson F049 zeigt eine große Translation in y-Richtung. Die beiden Registrierungen passen damit nicht zusammen. Die mittlere quadratische Distanz ist hier sehr hoch und liegt bei 13607.32.



Abbildung 5.16: Vergleich von manueller Registrierung und intensitätsbasierter Registrierung mit morphologischen Gradientenbildern: Dargestellt sind pro Subjekt die manuelle Registrierung, die intensitätsbasierte Registrierung und die übereinandergelagerten Bilder dieser beiden Registrierungen mit den neun transformierten Punkten zur Auswertung der mittleren quadratischen Distanz.



(j) Manuell, Subjekt M007

(k) Intensitätsbasiert, Subjekt (l) *MSE* M007 M007

MSE = 156.64, Subjekt M007

Abbildung 5.16: Vergleich von manueller Registrierung und intensitätsbasierter Registrierung mit morphologischen Gradientenbildern: Dargestellt sind pro Subjekt die manuelle Registrierung, die intensitätsbasierte Registrierung und die übereinandergelagerten Bilder dieser beiden Registrierungen mit den neun transformierten Punkten zur Auswertung der mittleren quadratischen Distanz.

6 Diskussion

6.1 Gemeinsamkeiten in den multimodalen Daten

In den Ergebnissen zeigt sich auf unterschiedliche Weise, dass kein allgemeiner Zusammenhang zwischen den multimodalen Daten von Tiefenkamera und Thermographie gezeigt werden kann. Die dreidimensionale Darstellung der Intensitätswerte beider Bilddaten zeigt deutlich, dass in den Tiefenbildern der Hintergrund durch eine klare Kante von dem Gesicht, also dem aufgenommen Objekt, getrennt ist. Im Gegensatz dazu gibt es in den Thermalbildern keine solche klare Abgrenzung. Hier ist es also nicht so leicht, das Bild anhand der Intensitätswerte in zwei Kategorien (Hintergrund und abgebildete Versuchsperson) zu unterteilen. Dieser Effekt wird bereits bei dem Vorverarbeitungsschritt mit einem globalen Schwellenwert beobachtet. Die Tiefenbilder sind durch die klare Abgrenzung der zwei Ebenen leicht zu segmentieren, da es einen Sprung in den Intensitätswerten an der Grenze von Hintergrund zu Versuchsperson gibt.

Der Verlauf der Intensitätswerte in den Thermalbildern stammt von der Wärmeabstrahlung der abgebildeten Person. Die Versuchsperson ist warm und strahlt Wärme in die kalte Umgebung ab. Diese Wärmeabstrahlung ist auf den Thermalbildern dann als Rauschen zu sehen. Durch diese Wärmeabstrahlung fällt eine globale Segmentierung des Thermalbildes sehr schwer, da kein Schwellenwert gefunden wird, der das Bild in die beiden eindeutigen Kategorien von Hintergrund und Versuchsperson einteilen kann.

Eine Gemeinsamkeit der Daten kann in dieser Darstellung nur für den Hintergrund festgestellt werden. Dieser ist in den Tiefenbildern immer am weitesten entfernt und hat damit in den Originalbildern den höchsten Wert und in den invertierten Tiefenbildern dementsprechend den niedrigsten Wert. Gleichzeitig ist in den Thermalbildern der Hintergrund immer kälter als die Versuchsperson. In den Gesichtern selbst gibt es keinen allgemeinen Zusammenhang. Bei den Tiefenbildern ist das Verhalten der Werte bekannt. Der nächste Punkt ist im Normalfall die Nase. In den Originalbildern ist dies daher der niedrigste Wert, in den invertierten Tiefenbildern der größte Wert. Bei den Thermalbildern gibt es kein eindeutiges Verhalten. Die Werte sind immer unterschiedlich verteilt. Es können keine generellen Aussagen dahingehend getroffen werden, dass ein Punkt im Gesicht immer wärmer ist als ein anderer. Auch bei einer einzelnen Person verändert sich die Intensitätswertverteilung im Gesicht pro Frame. Dies liegt unter anderem an der Durchblutung der Haut. Die lokale Durchblutungsregulation wird hauptsächlich über die Änderung der Gefäßweite reguliert [37]. Diese Änderung kann durch verschiedene Mechanismen beeinflusst werden. Dazu zählen auch verschiedene Hormone und das vegetative Nervensystem [37]. Durch Stress, verschiedene Emotionen oder körperliches Empfinden kann die Durchblutung beeinflusst werden, wodurch diese von jeder Versuchsperson individuell verschieden ist.

Auch die 2D-Histogramme der registrierten Bilder zeigen eine Änderung der Temperaturwerte von kälteren zu wärmeren Werten bei dem am weitesten entfernten Tiefenwert. Da bekannt ist, dass dieser Wert dem Hintergrund entspricht, ist auch hier diese Änderung der Temperaturwerte auf die Wärmeabstrahlung der Versuchspersonen zurückzuführen. Die Korrelation zwischen nahen Werten und warmen Werten kann keinem genauen Gesichtsmerkmal zugeordnet werden. Die Korrelation der registrierten Bilder zeigt das gleiche Ergebnis wie die 3D-Darstellung der Intensitätswerte. Eine Gemeinsamkeit kann nur im Hintergrund gefunden werden.

Beim Betrachten der 2D-Histogramme der Bildausschnitte von Augen und Nase der Versuchspersonen fällt auf, dass die höchste Korrelation in den weiter entfernten Tiefenwerten in den Bildausschnitten besteht. Dieser Bereich beschreibt wahrscheinlich die Haut, welche die Nase umgibt. In den Grauwertbildern der Bildausschnitte ist zu sehen, dass die Tiefenbilder der Versuchspersonen immer ähnlich aussehen. Die Nase wird dunkel dargestellt und die Augen hell. Die umliegende Haut hat einen Grauwert, der zwischen den Grauwerten von Nase und Augen liegt. Die Tiefenwertverteilung ist allgemein jedoch immer ähnlich. Im Gegensatz dazu sehen die Thermalbilder oftmals sehr unterschiedlich aus. Bei manchen Versuchspersonen hat die Nase im Grauwertbild helle, bei anderen dunkle Strukturen. Bei einigen, jedoch nicht allen, Versuchspersonen sind zusätzlich dunkle Bereiche auf der Haut, welche die Durchblutung der Wangen abbilden, sichtbar. Wenn sich die Durchblutung der Haut ändert, führt das zu einer Änderung der Temperatur und einer damit eingehenden Änderung der Thermalbilder. Da die Tiefenbilder nicht abhängig von der Durchblutung der Versuchsperson sind, verändern sich diese auch nicht bei einer Änderung der Durchblutung. Für die Gesamtheit der multimodalen Daten lässt sich kein allgemeiner Zusammenhang feststellen. Die 2D-Histogramme zeigen, dass die Wärmeverteilung in den Gesichtern der Versuchspersonen immer unterschiedlich ist

und damit allgemein nicht mit der immer gleich bleibenden Tiefenwertverteilung der Tiefenbilder korreliert.

Das Untersuchen von statistischen Merkmalen in Patches zeigt beim Verwenden der Standardabweichung, dass in beiden Bildern die Konturen der Bilder detektiert werden. In den Tiefenbildern werden die Umrisse der Person dadurch deutlich sichtbar, die Gesichtsmerkmale werden hier allerdings weniger deutlich erkannt. In den Thermalbildern werden die Gesichtsmerkmale stärker hervorgehoben, die Abgrenzung zum Hintergrund ist jedoch schwächer. Dieser Effekt geht wieder auf die Grenze von Versuchsperson zu Hintergrund zurück. Beim Tiefenbild ist hier ein deutlicher Sprung in den Intensitätswerten, wodurch der Umriss klar detektiert wird. Beim Thermalbild gibt es keine klare Abgrenzung, weshalb die detektierten Konturen hier auch nicht so stark sind.

Eine gewisse Ähnlichkeit ist in der Standardabweichung der Patches im Bereich der Nase zu sehen. Für einen Registrierungsansatz ist dies aber zu ungenau. Aus der Betrachtung der restlichen statistischen Merkmale können keine weiteren Erkenntnisse abgeleitet werden. Die Berechnung der Hu-Momente in den Bildern zeigen auch keinen weiteren Zusammenhang der Intensitätswerte. Mit der "matchShapes()"-Methode kann zwar eine gewisse Ähnlichkeit in manchen Bereichen des Bildes erkannt werden, dabei handelt es sich aber nicht um die Bereiche, die von besonderem Interesse sind. Eine Ähnlichkeit der Formen ist vor allem im Hintergrund und der Stirn zu sehen. Beides sind Bereiche in den Bildern, bei denen so gut wie keine Konturen zu sehen sind. Es ist also nicht verwunderlich, dass es bei einer weißen Fläche eine Ähnlichkeit in den Formen gibt. An den interessanten Bereichen des Bildes sind die Formen nach dieser Methode nicht ausreichend ähnlich, um mit den Hu-Momenten eine merkmalsbasierte Registrierung in Betracht zu ziehen. Tiefen- und Thermalbild sehen sich nach Bearbeitung mit einem lokalen adaptiven Schwellenwert also immer noch sehr unähnlich.

Die Daten stimmen nur in Bereichen überein, die für die Registrierung nicht von Bedeutung sind. Dies liegt möglicherweise daran, dass es sich um Bilder von menschlichen Gesichtern handelt. Bei Objekten oder Szenen, in denen es eine einheitliche Wärmeverteilung gibt, würden eventuell bessere Zusammenhänge detektiert werden. Alle diese Verfahren zur Betrachtung der Gemeinsamkeit in den Intensitätswerten zeigen die Herausforderung einer Registrierung von multimodalen Daten.

6.2 Merkmalsbasierte Registrierung mit SIFT

Eine globale merkmalsbasierte Registrierung mit SIFT mit den verwendeten Daten ist nicht zielführend. Dabei können verschiedene Vorverarbeitungsschritte die globale Registrierung nicht verbessern. Es ist zu erkennen, dass keine gleichen Merkmale gefunden werden. Daher können auch keine passenden Übereinstimmungen gefunden werden. Die Vorverarbeitung mit einem morphologischen Gradienten bringt ebenfalls keine Verbesserung. Auch wenn in dieser Arbeit SIFT und nicht C SIFT wie in [16] für die merkmalsbasierte Registrierung verwendet wird, kann an den verwendeten Daten für Tiefen- und Thermalbilder gezeigt werden, dass eine merkmalsbasierte Registrierung sich als sehr schwierig herausstellt. Entweder liegt dies an den Tiefen- und Thermalbildern im Generellen oder an den Eigenschaften des hier verwendeten Datensatzes. Problematisch an den 3D-Modellen, aus denen die Tiefenbilder generiert sind, ist, dass hier immer nur das Gesicht abgebildet ist. Da der Oberkörper fehlt, gibt es hier eine Kante im Tiefenbild, welche im Thermalbild so nicht vorliegt. Die zu registrierenden Bilder sehen damit von der Form her nicht gleich aus. Auch wird mit den verwendeten Vorverarbeitungsfiltern keine Möglichkeit gefunden, die Bilder global in ihrem Aussehen anzugleichen. Die Unähnlichkeit in den Intensitätswerten und der Form erschweren die globale merkmalsbasierte Registrierung mit SIFT. Auch wurden in [16] keine Bilddaten von menschlichen Gesichtern verwendet, was ebenfalls einen Unterschied für den Registrierungsprozess darstellt. Da die Tiefenbilder nicht von der Durchblutung der Haut und damit von der Temperatur abhängen, ist eine Temperaturänderung in den Tiefenbildern auch nicht dargestellt. Bei sichtbaren Bildern wären Durchblutungs- beziehungsweise Temperaturänderungen in der Haut durch eine Rötung dieser zu sehen. Daher würden hier die Daten sicherlich stärker korrelieren. Sowohl in sichtbaren Bildern als auch in Thermalbildern wäre damit eine Änderung zu sehen, wenn sich die Temperatur verändern würde.

Auch eine merkmalsbasierte Registrierung mit SIFT in den Bildausschnitten von Augen und Nase kann nicht verlässlich eingesetzt werden. Es wird gezeigt, dass bei den untersuchten Versuchspersonen keine allgemeine Vorverarbeitung dafür sorgt, dass die Bilder in den Bildausschnitten immer ähnlich aussehen. Am besten schneiden die morphologischen Gradientenbilder ab. Hier werden zum einen die meisten Merkmale gefunden und diese in zwei von sechs Versuchspersonen einigermaßen richtig zugeordnet. Da die morphologischen Gradientenbilder nur von der Änderung der Intensitätswerte, aber nicht von den Intensitätswerten an sich abhängen, können die Konturen der Gesichter und der Gesichtsmerkmale gut dargestellt werden. Es ist eindeutig zu sehen, dass es sich um dieselbe Person handelt. Trotzdem sehen die Umrisse auch von bereits registrierten Bildern nicht identisch aus. Das erschwert die merkmalsbasierte Registrierung. Obwohl mit dem menschlichen Auge klar zu sehen ist, welche Umrisse aufeinander passen, fällt das dem SIFT-Algorithmus sehr schwer. Insgesamt ist zu sehen, dass eine lokale Registrierung in den gewählten Bildausschnitten nicht verlässlich eingesetzt werden kann. Eine automatische Detektion der Bildausschnitte funktioniert bis jetzt nicht, da nicht bekannt ist, wo sich die Augen- und Nasenpartie im Thermalbild befindet, und es keine Möglichkeit gibt, anhand der Intensitätswerte diesen Bildausschnitt zu detektieren. Die Bildausschnitte müssten damit also manuell ausgewählt werden. Hier bleibt die Frage, ob in diesem Fall dann nicht eine manuelle Registrierung mit Wählen von Landmarken einfacher wäre.

Da eine merkmalsbasierte Registrierung nicht funktioniert, kann damit die Aussage aus [19] bestätigt werden, dass in den Thermalbildern von menschlichen Gesichtern das Detektieren von guten Merkmalen schwierig ist. Auch in Tiefenbildern ist dies nicht leicht, da vor allem innerhalb des Gesichts weniger Kontraste vorliegen.

6.3 Intensitätsbasierte Registrierung mit Mutual Information

Visuelle Bewertung der registrierten Bilder Aus der visuellen Betrachtung geht hervor, dass eine intensitätsbasierte Registrierung mit den Originalbildern nicht zielführend ist. Auch ein Histogram Matching ähnlich zu [19] führt zu keinen guten Ergebnissen. Es kann also analog zu [19] gezeigt werden, dass eine intensitäsbasierte Registrierung mit den Originalbildern nicht funktioniert. Anhand der verwendeten Tiefen- und Thermalbilder kann jedoch nicht bestätigt werden, dass ein Angleichen der Bilder mit einem Histogram Matching auch bei einer Registrierung zwischen Tiefen- und Thermalbildern zielführend ist. Am besten funktioniert eine Registrierung mit den morphologischen Gradientenbildern. Hier kann es zu sehr guten Registrierungsergebnissen kommen. Die Methode funktioniert jedoch nicht bei allen Versuchspersonen. Aus einer visuellen Beurteilung können für ca. 57 % der untersuchten Versuchspersonen die Registrierungsergebnisse als gut angesehen werden. In mehr als der Hälfte der Fälle scheint die Registrierung mit morphologischen Gradientenbildern damit zu funktionieren. Die deutliche Verbesserung der intensitätsbasierten Registrierung ist wahrscheinlich darauf zurückzuführen, dass hier nur die wichtigen Gesichtskonturen zu sehen sind. Bei diesen Bildern liegt weniger Fokus auf den Intensitätswerten als auf den Formen. Auch die Registrierung mit adaptiven Schwellenwerten kann gute Ergebnisse liefern. Visuell passender ist jedoch meistens die Registrierung mit den morphologischen Gradientenbildern. Beide Bilder zeigen ausschließlich die Formen der Versuchspersonen. In den morphologischen Gradientenbildern werden feinere Linien

dargestellt und in den Tiefenbildern auch mehr Gesichtsmerkmale hervorgehoben als mit einer Vorverarbeitung durch einen adaptiven Schwellenwert. Das ist ein möglicher Grund für die bessere Registrierung mit den morphologischen Gradientenbildern als mit den adaptiven Schwellenwertbildern. Trotzdem sehen die Bilder nach Bearbeitung mit dem morphologischen Gradienten nicht identisch aus. Hier besteht wieder die Problematik, dass in den Tiefenbildern lediglich der Kopf ohne Rumpf dargestellt ist. Wenn die Tiefenbilder die vollständige Person mit Oberkörper zeigen würden, könnte die Registrierung eventuell verbessert werden, da die Konturen dann noch besser übereinstimmen würden.

Ähnlichkeit der registrierten Bilder Bei der Auswertung der Ähnlichkeit der registrierten Bilder wird erwartet, dass eine passende Registrierung mit einem hohen Ähnlichkeitswert einhergeht. Da bei einer visuellen Betrachtung der registrierten Bilder eine Vorverarbeitung mit morphologischen Gradientenbildern, gefolgt von der Bearbeitung mit einem adaptiven Schwellenwert, die besten Ergebnisse erzielt, wird erwartet, dass diese beiden Methoden auch die höchsten Ähnlichkeitswerte erlangen. Dies ist jedoch nicht der Fall. Mit dem Vergleich der Ähnlichkeitsmetriken wird gezeigt, dass eine manuelle Registrierung eine deutliche Steigerung der Ähnlichkeit im Vergleich zu den unregistrierten Daten bringt. Das bestätigt, was visuell bei den manuell registrierten Bildern beobachtet werden kann. Auch bei der merkmalsbasierten Registrierung wird mit der Ähnlichkeitsmetrik gezeigt, dass hier keine signifikante Verbesserung der Ähnlichkeit vorliegt. Es kann hier also bestätigt werden, was bei der visuellen Betrachtung der merkmalsbasierten Registrierung mit SIFT auffällt. Die Bilder sind sich nach der merkmalsbasierten Registrierung oftmals unähnlicher als davor. Auffällig ist, dass die intensitätsbasierte Registrierung in den meisten Fällen höhere Ähnlichkeitwerte erhält als die manuelle Registrierung. Visuell betrachtet ist die manuelle Registrierung jedoch häufig sehr viel besser als eine intensitätsbasierte Registrierung. Die Ähnlichkeitswerte widersprechen somit in vielen Fällen der visuellen Bewertung. Bei der Registrierung mit morphologischen Gradientenbildern und adaptiven Schwellenwertbildern ist deutlich zu erkennen, dass hier die Ähnlichkeitswerte im Durchschnitt nicht über den Werten der manuellen Registrierung liegen. Hier sind die Ähnlichkeitswerte sehr ähnlich zu den Werten der manuellen Registrierung. Das bedeutet, dass die Originalbilder nach der manuellen Registrierung und nach der intensitätsbasierten Registrierung mit den morphologischen Gradientenbildern und adaptiven Schwellenwertbildern nah beieinanderliegende Ähnlichkeitswerte besitzen. Daraus folgt, dass bei dem Vergleich der Ähnlichkeiten mit der Normalized Mutual Information nicht davon ausgegangen werden kann, dass der höchste Ähnlichkeitswert gleichzeitig die beste Registrierung bedeutet. Das Verhalten deutet darauf hin, dass ein ähnlicher Wert zur entsprechenden manuellen Registrierung auf eine gute Registrierung hindeutet.

Ähnlichkeit der Transformationsmatrizen Bei der Betrachtung der Ähnlichkeit der Transformationsmatrizen anhand des Vergleichs von neun transformierten Punkten und deren mittleren quadratischen Distanzen wird erwartet, dass bei einer hohen mittleren quadratischen Distanz eine große Abweichung der intensitätsbasierten Registrierung von der manuellen Registrierung festzustellen ist. Außerdem wird erwartet, dass aufgrund der visuellen Betrachtung die mittlere quadratische Distanz der intensiätsbasierten Registrierung mit morphologischen Gradientenbildern am kleinsten ist. Der Vergleich der Mittelwerte der mittleren quadratischen Distanzen bestätigt die visuelle Betrachtung. Die mittlere quadratische Distanz ist bei der Registrierung mit den morphologischen Gradientenbildern am kleinsten und hat auch die niedrigste Standardabweichung. Damit kann gesagt werden, dass dieses Verfahren am besten funktioniert. Auch kann gezeigt werden, dass die Registrierung mit adaptiven Schwellenwertbildern am zweitnächsten an die manuelle Registrierung herankommt. Hier muss aber beachtet werden, dass dieses Verfahren eine relativ große Standardabweichung hat. Damit funktioniert das Verfahren zum einen schlechter und zum anderen auch nicht so verlässlich wie das Verfahren mit den morphologischen Gradientenbildern.

Es ist zu beachten, dass nur die mittlere quadratische Distanz zur manuellen Registrierung dargestellt wird. Die manuelle Registrierung ist nicht perfekt. Eine gute intensitätsbasierte Registrierung kann visuell besser als die manuelle Registrierung sein. Auch ist an Beispiel von Versuchsperson M007 aus Abbildung 5.16l zu sehen, dass nicht immer klar gesagt werden kann, ob die manuelle Registrierung oder die intensitätsbasierte Registrierung passender ist. Bei der intensitätsbasierten Registrierung ist hier in den registrierten Bildern eine Rotation des Bewegtbildes im Vergleich zur manuellen Registrierung zu sehen. Diese Rotation sorgt dementsprechend für eine Abweichung der Transformationsmatrizen. Sowohl manuelle Registrierung als auch intensitätsbasierte Registrierung sind visuell betrachtet stimmig. Es kann nicht klar gesagt werden, welche der beiden Methoden bei dieser Versuchsperson die besseren Ergebnisse erzielt. Es darf daher nicht davon ausgegangen werden, dass die kleinste Distanz automatisch die beste Registrierung bedeutet. Es sagt nur aus, dass diese Registrierung am nächsten an die manuellen Registrierung kommt.

Insgesamt betrachtet sagt dieses Maß auch das aus, was visuell beobachtet wurde. Die intensitätsbasierte Registrierung mit den morphologischen Gradientenbildern ist am nächsten an der manuellen Registrierung, gefolgt von der Registrierung mit den adaptiven Schwellenwertbildern. Die Registrierung mit den globalen Binärbildern steht an dritter Stelle. Es fällt auf, dass in allen drei dieser Bearbeitungsschritte die Filter eine Auswirkung auf die Form und nicht die Intensitätswerte haben. Die Beschränkung auf die wichtigen Merkmale wie Umrisse und Gesichtskonturen verbessert die intensitätsbasierte Registrierung offensichtlich. Da die Konturen nicht identisch sind, hat die intensitätsbasierte Registrierung trotzdem Schwierigkeiten, immer gute Ergebnisse zu erzielen. Teilweise haben die Tiefenbilder durch die 3D-Modelle sehr "ausgefranste" Ränder, wodurch eine intensitätsbasierte Registrierung zusätzlich erschwert wird.

Die Evaluation der intensitätsbasierten Registrierung zeigt sowohl visuell als auch mit der Bewertung der Ähnlichkeit der Transformationsmatrizen, dass eine Vorverarbeitung mit dem morphologischen Gradienten im Vergleich zu den Originalbildern eine deutliche Verbesserung der Registrierungsergebnisse mit den verwendeten Daten bringt. Es wird deutlich, dass eine quantitative Bewertung der Registrierungsergebnisse nicht einfach ist. Deshalb ist eine qualitative Bewertung durch visuelles Betrachten der Registrierung wichtig. Auch kann bislang keine Möglichkeit gefunden werden, herauszufinden, wann genau die intensitätsbasierte Registrierung mit morphologischen Gradientenbildern scheitert. Die Registrierungsergebnisse geben hier keine eindeutige Antwort.

7 Fazit

In dieser Arbeit wird der Zusammenhang zwischen den multimodalen Daten von Tiefenkamera und Thermographie analysiert. Zusätzlich werden verschiedene Vorverarbeitungsschritte durchgeführt und mehrere Registrierungsverfahren getestet, um die Daten möglichst gut zu registrieren. Die verwendeten Daten entstammen dem Datensatz "BP4D+". Um einen Zusammenhang zwischen Tiefen- und Temperaturwerten zu erkennen, werden verschiedene Ansätze gewählt. Dazu zählen das Erstellen von 3D-Darstellungen der Intensitätswertverteilungen und 2D-Histogrammen zur Untersuchung der Korrelation zwischen den Werten sowohl in den gesamten Bildern als auch in Bildausschnitten von Nase und Augen. Des Weiteren werden mehrere statistische Merkmale und Bildmomente in kleinen Bereichen des Bildes untersucht. Dabei wird kein signifikanter Zusammenhang der Intensitätswerte festgestellt. Bei den verwendeten Daten gibt es die Gemeinsamkeit, dass der Hintergrund weit entfernt und kalt ist, im Gesicht der Versuchspersonen gibt es jedoch keinen allgemeinen Zusammenhang. Hier ist die Durchblutung und damit auch die Temperatur beziehungsweise das Thermalbild jeder Versuchsperson individuell. Da die Tiefenwerte nicht von der Durchblutung oder Temperatur insgesamt abhängen, wird auch keine Temperaturänderung detektiert. Die Tiefenbilder sehen damit immer gleich aus, während die Thermalbilder sich dauerhaft optisch ändern und das nicht nur von Person zu Person, sondern auch innerhalb der aufgezeichneten Videos.

Diesen Effekt gleichen am besten die Vorverarbeitungsschritte aus, die nicht die Intensitätswerte verändern, sondern hauptsächlich die abgebildeten Formen segmentieren. Dazu zählen morphologische Gradientenbilder, adaptive Schwellenwertbilder und Binärbilder. Bei den Binärbildern ist jedoch anzumerken, dass ein globaler Schwellenwert für die Thermalbilder sehr schwierig zu finden ist. Da die Versuchspersonen Wärme abstrahlen, umgibt die Personen viel Rauschen. Dadurch ist es schwierig, das Bild in zwei Kategorien einzuteilen. Ein adaptiver Schwellenwert umgeht das Problem. Hier werden jedoch bei den Tiefenbildern nur sehr wenige Merkmale im Gesicht detektiert. Mit den morphologischen Gradientenbildern funktioniert eine Beschreibung der Konturen am besten. Wichtig zu erwähnen ist, dass keine Vorverarbeitung gefunden wird, mit der Tiefen- und Thermalbilder insgesamt ähnlicher gemacht werden können. Des Weiteren besteht das Problem, dass die abgebildeten Versuchspersonen in den Tiefen- und Thermalbildern keine identische Form haben, da die Tiefenbilder aus 3D-Modellen generiert sind und diese nur den Kopf ohne Körper darstellen. Dadurch ist in den Tiefenbildern immer eine Kante abgebildet, die in den Thermalbilder nie zu finden sein wird.

Die getesteten Registrierungsverfahren können eingeteilt werden in globale merkmalsbasierte Registrierung mit SIFT, lokale merkmalsbasierte Registrierung mit SIFT in Bildausschnitten, manuelle Registrierung in Matlab und intensitätsbasierte Registrierung mit Mutual Information. Dabei werden bei einer globalen merkmalsbasierten Registrierung mit SIFT keine guten Merkmale gefunden und diese auch nicht richtig zugeordnet. Auch eine lokale Registrierung kann keine deutlich besseren Ergebnisse erzielen. In der lokalen Registrierung in dem getesteten Bildausschnitt kann die Registrierung mit morphologischen Gradientenbildern teilweise gute Merkmale zuordnen. Das Verfahren wird jedoch aufgrund von unzuverlässigen Ergebnissen nicht weiter untersucht. Eine manuelle Registrierung der Daten wird in Matlab umgesetzt. Hier kann durch das Wählen von acht Landmarken in jedem Bild eine gute Registrierung erreicht werden. Generell ist es bei der manuellen Registrierung nicht leicht, genau die richtigen zugehörigen Punkte zu finden, da die Bilder sehr unterschiedlich aussehen. Daher ist die manuelle Registrierung nicht perfekt, aber für einen Vergleich als Referenz ausreichend. Das manuelle Registrieren ist jedoch sehr zeitaufwändig. Bei der intensitätsbasierten Registrierung kann eine Vorverarbeitung mit einem morphologischen Gradienten gute Ergebnisse erzielen. Bei einer guten intensitätsbasierten Registrierung ist diese genauer als die manuelle Registrierung. Damit ist dieser Ansatz eine vielversprechende Möglichkeit einer automatischen Registrierung. Mit diesem Verfahren lassen sich visuell betrachtet ca. 57 % und mit der Auswertung der Ähnlichkeit der Transformationsmatrizen ca. 54 % der 140 Testdaten automatisch registrieren. Eine visuelle Nachkontrolle ist aber in jedem Fall erforderlich, da eben nicht alle Daten automatisch registriert werden können. Dabei bleibt die Frage offen, ob es einen bestimmten Grund für das Scheitern der Registrierung bei manchen Versuchspersonen gibt oder dieser mehr oder weniger willkürlich ist. Eine Registrierung mit adaptiven Schwellenwertbildern kann teilweise auch gute Ergebnisse erzielen, insgesamt ist die Registrierung mit morphologischen Gradientenbildern aber genauer. Mögliche Gründe für die falsch registrierten Daten könnten einerseits die künstlich generierten Tiefenbilder und andererseits das sehr unterschiedliche Aussehen der Bilder aufgrund der verschiedenen Modalitäten sein. Die vielversprechendste Methode für die besten Ergebnisse ist eine intensitätsbasierte Registrierung mit der Mutual Information Metrik und einer Vorverarbeitung der Bilder mit dem morphologischen Gradienten. Insgesamt funktioniert eine manuelle Registrierung mit Landmarken trotzdem zuverlässiger.

Es wird deutlich, dass die quantitative Bewertung der Registrierung eine große Herausforderung darstellt und nicht klar gezeigt werden kann, warum und wann die Registrierung genau fehlschlägt.

Es gibt mehrere Aspekte, die in möglichen weiteren Arbeiten untersucht werden könnten. Zum einen sollte getestet werden, ob die Registrierung mit anderen Daten besser funktioniert. Vielversprechender könnten Daten sein, in denen das Subjekt etwa gleich aussieht und im Tiefenbild nicht ein Teil der Person fehlt. Außerdem könnte getestet werden, ob die Registrierung fehlschlägt, weil es sich um abgebildete menschliche Gesichter handelt oder ob die Registrierung bei anderen abgebildeten Szenen ebenfalls nicht immer funktioniert. Auch könnte angelehnt daran überprüft werden, ob es in dem Fall dann einen Zusammenhang zwischen Tiefen- und Temperaturwerten gibt. Des Weiteren müsste untersucht werden, wie die Registrierung besser bewertet werden kann und in welchen Fällen die Registrierung scheitert.

Abschließend lässt sich sagen, dass das Ziel einer Registrierung zwischen Tiefenkameraund Thermographiedaten nur teilweise umgesetzt wird. Dadurch wird die Schwierigkeit einer Registrierung zwischen multimodalen Daten deutlich. Es können eindeutige Verbesserungen der Registrierungsqualität erreicht und damit vielversprechende Möglichkeiten zur automatischen Registrierung gefunden werden. Eine visuelle Nachkontrolle ist hier aber auf jeden Fall erforderlich, um falsche Registrierungsergebnisse auszuschließen.

Literatur

- [1] A. K. AlZubaidi, Y. Ethawi, G. M. Schmölzer, S. Sherif, M. Narvey und M. Seshia, "Review of Biomedical Applications of Contactless Imaging of Neonates Using Infrared Thermography and Beyond", *Methods and Protocols*, 2018. DOI: 10.3390/ mps1040039.
- S. Lyra, L. Mayer, L. Ou u. a., "A Deep Learning-Based Camera Approach for Vital Sign Monitoring Using Thermography Images for ICU Patients", *Sensors*, Jg. 21, S. 1495, 2021. DOI: 10.3390/s21041495.
- U. Koehler, O. Hildebrandt, F. Magnet, J. Storre und W. Grimm, "Die Atemfrequenz – ein vernachlässigter Vitalparameter", DMW - Deutsche Medizinische Wochenschrift, Jg. 142, S. 130–134, 2017. DOI: 10.1055/s-0042-120485.
- [4] G. Sun, Y. Nakayama, S. Dagdanpurev u. a., "Remote sensing of multiple vital signs using a CMOS camera-equipped infrared thermography system and its clinical application in rapidly screening patients with suspected infectious diseases", *International Journal of Infectious Diseases*, Jg. 55, S. 113–117, 2017. DOI: 10.1016/j.ijid.2017.01.007.
- [5] B. Zitova, "Mathematical Approaches for Medical Image Registration", in *Encyclope*dia of Biomedical Engineering, Elsevier, 2019, S. 21–32, ISBN: 978-0-12-805144-3.
 DOI: 10.1016/B978-0-12-801238-3.99990-2.
- [6] O. Dössel, Bildgebende Verfahren in der Medizin, Von der Technik zur medizinischen Anwendung. Springer-Verlag Berlin Heidelberg New York, 2000, ISBN: 3540660143.
- [7] A. K. Abbas, K. Heimann, K. Jergus, T. Orlikowsky und S. Leonhardt, "Neonatal non-contact respiratory monitoring based on real-time infrared thermography", *BioMedical Engineering OnLine*, Jg. 10, S. 93, 2011. DOI: 10.1186/1475-925X-10-93.
- [8] BEGINNER'S GUIDE TO DEPTH (UPDATED), 2019. Adresse: https://www. intelrealsense.com/beginners-guide-to-depth/(besucht am 10.08.2022).

- [9] M. Schmidt, "Verbesserung nicht-rigider Registrierungsverfahren durch lokale Abbildungen", Diplomarbeit, Universität Koblenz Landau, 2009.
- [10] J. B. A. Maintz und M. A. Viergever, "A Survey of Medical Image Registration", *Medical Image Analysis*, Jg. 2, Nr. 1, S. 1–37, 1997.
- [11] A. Vieten, *Monomodale und multimodale Registrierung von autoradiographischen und histologischen Bilddaten* (Schriften des Forschungszentrums Julich. Reihe Lebenswissenschaften, Life sciences). Forschungszentrum Jülich, Zentralbibliothek, 2005, ISBN: 978-3-89336-390-2.
- Z. Wang, A. Bovik, H. Sheikh und E. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", *IEEE Transactions on Image Processing*, Jg. 13, Nr. 4, S. 600–612, 2004. DOI: 10.1109/TIP.2003.819861.
- J. Pluim, J. Maintz und M. Viergever, "Mutual-Information-Based Registration of Medical Images: A Survey", *IEEE Transactions on Medical Imaging*, Jg. 22, Nr. 8, S. 986–1004, 2003. DOI: 10.1109/TMI.2003.815867.
- [14] X. Zhang, C. Leng, Y. Hong, Z. Pei, I. Cheng und A. Basu, "Multimodal Remote Sensing Image Registration Methods and Advancements: A Survey", *Remote Sens*, Jg. 13, S. 5128, 2021. DOI: 10.3390/rs13245128.
- [15] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, Jg. 60, S. 91–110, 2004. DOI: 10.1023/B: VISI.0000029664.99615.94.
- [16] Q. Zeng, J. Adu, J. Liu, J. Yang, Y. Xu und M. Gong, "Real-time adaptive visible and infrared image registration based on morphological gradient and C_SIFT", *Journal* of *Real-Time Image Processing*, Jg. 17, S. 1103–1115, 2020. DOI: 10.1007/s11554-019-00858-x.
- [17] M. Calonder, V. Lepetit, C. Strecha und P. Fua, "BRIEF: Binary Robust Independent Elementary Features", in *Computer Vision – ECCV 2010*, Springer Berlin Heidelberg, 2010, S. 778–792, ISBN: 978-3-642-15560-4 978-3-642-15561-1. DOI: 10.1007/ 978-3-642-15561-1_56.
- [18] M. A. Fischler und R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", *Communications of the ACM*, S. 381–395, 1981. DOI: 10.1145/358669.358692.
- P. T. Krishnan, P. Balasubramanian und V. Jeyakumar, "Histogram Matched Visible and Infrared Image Registration for Face Detection", in *IEEE EUROCON 2021* - 19th International Conference on Smart Technologies, 2021. DOI: 10.1109/ EUROCON52738.2021.9535615.

- [20] H. Bay, T. Tuytelaars und L. Van Gool, "SURF: Speeded Up Robust Features", in *Computer Vision – ECCV 2006*, Springer Berlin Heidelberg, 2006, S. 404–417. DOI: 10.1007/11744023_32.
- [21] E. Rublee, V. Rabaud, K. Konolige und G. Bradski, "ORB: An efficient alternative to SIFT or SURF", in 2011 International Conference on Computer Vision, 2011, S. 2564–2571. DOI: 10.1109/ICCV.2011.6126544.
- [22] Z. Zhang, J. M. Girard, Y. Wu u. a., "Multimodal Spontaneous Emotion Corpus for Human Behavior Analysis", in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2016, S. 3438–3446, ISBN: 978-1-4673-8851-1. DOI: 10.1109/CVPR.2016.374.
- [23] What Is Camera Calibration? Adresse: https://de.mathworks.com/help/ vision/ug/camera-calibration.html#bu0nh2_(besucht am 21.07.2022).
- [24] Pinhole Camera Model | HediVision. Adresse: https://hedivision.github. io/Pinhole.html (besucht am 21.07.2022).
- [25] N. Krishna, Image Formation and Pinhole Model of the Camera, 2022. Adresse: https://towardsdatascience.com/image-formation-and-pinholemodel-of-the-camera-53872ee4ee92 (besucht am 21.07.2022).
- [26] K. Simek, Dissecting the Camera Matrix, Part 2: The Extrinsic Matrix, 2012. Adresse: https://ksimek.github.io/2012/08/22/extrinsic/ (besucht am 21.07.2022).
- [27] S. Mallick, *Geometry of Image Formation*, 2020. Adresse: https://learnopencv. com/geometry-of-image-formation/ (besucht am 21.07.2022).
- [28] D. L. Pham, C. Xu und J. L. Prince, "A Survey of Current Methods in Medical Image Segmentation", *Image Segmentation*,
- [29] OpenCV: Image Thresholding. Adresse: https://docs.opencv.org/4.x/d7/ d4d/tutorial_py_thresholding.html (besucht am 23.08.2022).
- [30] D. P. Sharma, "Intensity Transformation using Contrast Limited Adaptive Histogram Equalization", *International Journal of Engineering Research*, Jg. 2, Nr. 4, S. 282–285, 2013.
- [31] M. S. Vidya, A. H. Shastry und Y. Mallya, "4 Automated detection of intracranial hemorrhage in noncontrast head computed tomography", in *Advances in Computational Techniques for Biomedical Image Analysis*, Academic Press, 2020, S. 71–98, ISBN: 978-0-12-820024-7. DOI: 10.1016/B978-0-12-820024-7.00004-9.

- [32] Python | Intensity Transformation Operations on Images, 2019. Adresse: https: //www.geeksforgeeks.org/python-intensity-transformationoperations-on-images/ (besucht am 23.08.2022).
- [33] S. Stiene, "Konturbasierte Objekterkennung aus Tiefenbildern eines 3D-Laserscanners", Masterarbeit, Universität Osnabrück, 2006.
- [34] OpenCV: Feature Matching. Adresse: https://docs.opencv.org/3.4/dc/ dc3/tutorial_py_matcher.html (besucht am 13.09.2022).
- [35] OpenCV: Feature Matching + Homography to find Objects. Adresse: https://docs. opencv.org/3.4/d1/de0/tutorial_py_feature_homography.html (besucht am 13.09.2022).
- [36] Register Multimodal MRI Images. Adresse: https://de.mathworks.com/ help/images/registering-multimodal-mri-images.html (besucht am 13.09.2022).
- [37] N. Menche, Biologie Anatomie Physiologie, kompaktes Lehrbuch für Pflegeberufe. Elsevier Urban & Fischer, 2016, ISBN: 9783437268038.