



Technische Universität Ilmenau

Fakultät für Informatik und Automatisierung

Fachgebiet Neuroinformatik und Kognitive Robotik

Konzepterstellung und Merkmalsuntersuchung für die Entwicklung eines Systems zur Analyse von Gesichtsmotorik-Übungen mittels Tiefeninformation

Masterarbeit zur Erlangung des akademischen Grades Master of Science

Birant Sibel Olgay

Betreuer: Dipl.-Ing. Cornelia Lanz, FG Neuroinformatik und Kognitive Robotik/
Lehrstuhl Digitale Bildverarbeitung, FSU Jena

Verantwortlicher Hochschullehrer:

Prof. Dr. H.-M. Groß, FG Neuroinformatik und Kognitive Robotik

Die Masterarbeit wurde am 13.04.2012 bei der Fakultät für Informatik
und Automatisierung der Technischen Universität Ilmenau eingereicht.

Danksagung

An dieser Stelle möchte ich mich bei Herrn Prof. Horst-Michael Groß für die Mühen bei der Themensuche meiner Masterarbeit bedanken. Ebenso möchte ich Cornelia Lanz für die Betreuung der Arbeit und die Unterstützung beim labeln des Datensatzes über die Feiertage danken.

Vielen Dank an meine Kommilitonen und an die Mitarbeiter des Fachgebietes NIKR, die sich als Probanden für meine Arbeit zur Verfügung gestellt haben, vor allem an Erik Schaffernich für die beratenden Gespräche.

Meinen Eltern möchte ich für die beruhigende Gewissheit Ihrer Unterstützung für alles, was ich mir zum Ziel gesetzt habe, danken. Mein Vater war mir dabei immer ein großes Vorbild im wissenschaftlichen Bereich.

Danke an meine Freunde, mit denen ich eine tolle Studienzeit hatte, besonders hervorheben möchte ich Alexander und Thomas.

Meinem Freund Felix möchte ich besonders für das tiefe Vertrauen in meine wissenschaftlichen Fähigkeiten danken. Vielen Dank auch an Stefan und Daniela für das geduldige Lesen meiner Arbeit.

Erklärung: „Hiermit versichere ich, dass ich diese Masterarbeit selbstständig verfasst und nur die angegebenen Quellen und Hilfsmittel verwendet habe. Alle von mir aus anderen Veröffentlichungen übernommenen Passagen sind als solche gekennzeichnet.“

Ilmenau, 13.04.2012

.....

Birant Sibel Olgay

Kurzfassung

Patienten mit Störungen der mimischen oder orofazialen Muskulatur müssen regelmäßig Gesichtsmotorik-Übungen durchführen, um einen Erfolg in der logopädischen Therapie zu erreichen. Ein trainingsunterstützendes Tiefenbildkamerasystem könnte Feedbackkomponenten liefern, um die Möglichkeiten im Heimtraining zu verbessern. Dafür wird die Klassifikationsmöglichkeit von Gesichtsmotorik-Übungen in Tiefenbildaufnahmen anhand von statischen Bildern innerhalb einer Offline-Klassifikation untersucht. Eine Merkmalsextraktion von Normalenvektoren und Punktsignaturen wurde in MATLAB umgesetzt. Dabei werden deren Parameter und Ortsdefinitionen innerhalb der Gesichtsdaten vielfach variiert. Am berechneten Merkmalsraum wird ein Merkmalsranking bezüglich der Mutual Information durchgeführt. Anhand der paarweisen Mutual Information wird eine Redundanzuntersuchung vorgenommen, um gegebenenfalls redundante Merkmale auszuschließen oder über bestimmte Merkmalsgruppen eine Hauptkomponentenanalyse durchzuführen. Der resultierende Merkmalsraum wurde für die Klassifikation genutzt und an einem Datensatz getestet, der mittels Kinectsensor aufgenommen wurde. Der Datensatz enthält RGB- und Tiefenbilder von elf Probanden und jeweils 19 Gesichtsmotorik-Übungen. Durch die Klassifikation mittels Naive Bayes Klassifikator und einer 11-fachen Kreuzvalidierung wird dabei eine Erfolgsrate von 53% und mittels Hold-Out Methode von 65% erreicht. Ebenso wird die Hälfte der Klassen mit geringerer Balanced Error Rate als 20% erkannt, weshalb von einer Klassifikationsmöglichkeit der Gesichtsmotorik-Übungen mittels Tiefeninformationen des Kinectsensors ausgegangen werden kann.

Abstract

Patients with orofacial myofunctional diseases, dysphagia or speech disorders based on amissed articulation have to regularly practice facial muscle exercises for a successful therapy. A development of a training support system with range data sensors could improve the potential for the patients practicing at home. Thus the capabilities of an offline classification of facial muscle exercises based on static depth images are examined. Surface normals and point signatures are implemented in MATLAB as well as extracted with different parameters of computation and various positions. For all extracted features, firstly a feature ranking with a mutual information criterion is executed, secondly a redundancy analysis is done with paired mutual information. Redundant features are removed or rather the dimension of a group of features is reduced by principle component analysis. The accuracy of classification with a Naive Bayes model and Hold-Out Validation achieves 65%, with 11-fold cross-validation 53%. Furthermore, for half of the exercises the Balanced Error Rate for recognition is smaller than 20%. The classification is tested on a dataset, which contains eleven individuals and nineteen different facial muscle exercises. The results prove the possibility of recognition of facial muscle exercises in depth images.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Zielstellung	2
2	Gundlagen	5
2.1	Die Kamerasysteme	5
2.1.1	Das Tiefenbild	6
2.1.2	Das Prinzip der Abstandsmessung mit Time of Flight-Sensoren	6
2.1.3	Das Messprinzip des Kinectsensors	8
2.2	Sprach- und Schluckstörungen	10
2.2.1	Dysarthrie	10
2.2.2	Dysglossie	11
2.2.3	Dyslalie	11
2.2.4	Dysphagie	12
2.2.5	Orofaziale Dysfunktion	13
2.3	Informatische Grundlagen	13
2.3.1	Merkmalsselektion	14
2.3.2	Klassifikation	18
3	State of the Art	25
3.1	Gesichts- und Mimikerkennung in Tiefenbildaufnahmen	25
3.2	Trainingsunterstützung der logopädischen Therapie	29

4	Verwendete Verfahren	33
4.1	Messaufbau	33
4.1.1	Auswahl der Kamera	33
4.1.2	Bedingungen des Messaufbaus	34
4.2	Katalog der Gesichtsmotorik-Übungen	36
4.2.1	Auswahlkriterien der Übungen	36
4.2.2	Ausgewählte Übungen	37
4.3	Erstellung der Basisprimitive	42
4.3.1	Das Facial Action Coding System	42
4.3.2	Anpassen des FACS an die bestehende Problematik	43
4.4	Erstellung des Datensatzes	45
4.5	Auswahl und Definition der Features	47
4.5.1	Normalenvektoren	49
4.5.2	Punktsignaturen	51
4.6	Konzept der Merkmalsselektion	53
4.7	Konzept der Klassifikation	56
5	Umsetzung	59
5.1	Framework zur Klassifikation der Tiefenbilder	59
5.2	Normalenvektoren	61
5.2.1	Allgemeine Umsetzung	61
5.2.2	Verschiedene Umsetzungen der Normalenvektoren als Merkmale	67
5.3	Punktsignaturen	76
5.3.1	Punktsignaturen mit dynamischem Radius	77
5.3.2	Punktsignaturen mit statischem Radius	84
5.4	Dimensionsreduktion	85
5.4.1	Merkmalsranking mittels Mutual Information	86
5.4.2	Redundanzuntersuchung	88
6	Ergebnisse	93
6.1	Ergebnisse der Klassifikation	93

6.2	Ergebnisse mit verschobenen Markerpositionen	98
7	Diskussion	101
7.1	Extraktion und Selektion der Merkmale	101
7.2	Einfluss der Markerposition	104
7.3	Diskussion der Klassifikation	106
7.3.1	Unterschiede durch Validierungsarten	106
7.3.2	Klassifikationsleistungen bezüglich einzelner BP	107
7.3.3	Abhängigkeit der Klassifikationsleistung vom Datensatz	113
8	Zusammenfassung und Ausblick	115
A	Ergänzende Unterlagen	121
A.1	Dysarthriearten	121
A.2	Basisprimitive	122
A.3	Übungskatalog der Gesichtsmotorik-Übungen	123
A.4	Ergebnisse der Klassifikation	134
	Literaturverzeichnis	144

Abkürzungsverzeichnis

AD	Action Descriptor
AU	Action Unit
BER	Balanced Error Rate
BP	Basisprimitive
DCT	Discrete Cosine Transform
FACS	Facial Action Coding System
FN	False Negative
FPR	False Positive Rate
IR	Infrarot
kNN	k nächste Nachbarn
LMS	Least-Mean-Squares
MFT	Myofunktionelle Therapie
MI	Mutual Information
PCA	Principal Component Analysis
ROC	Receiver Operating Characteristics
ROI	Region of Interest
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TOF	Time of Flight
TPR	True Positive Rate
TP	True Positive

Kapitel 1

Einleitung

1.1 Motivation

Bei Sprach- und Schluckstörungen bzw. Schwächen im Bereich der orofazialen¹ und mimischen Muskulatur werden Gesichtsmotorik-Übungen als Teiltherapie durchgeführt. Um beispielsweise eine Stärkung oder Streckung der mimischen oder orofazialen Muskulatur durch eine logopädische Therapie zu erreichen, müssen Patienten die Übungen zwischen den Therapiestunden zu Hause regelmäßig durchführen und trainieren. Dabei fehlt es oftmals an Motivation und an Möglichkeiten einer objektiven Kontrolle der korrekten Übungsdurchführung. Ein Kamerasystem, welches Farb- und Tiefenbilder der Patienten während der Übungsdurchführung aufnimmt und ein Echtzeit-Feedback nach bestimmten Evaluationskriterien über die Qualität der Übungsdurchführung ausgibt, könnte dabei Abhilfe schaffen. Gerade für Kinder und Jugendliche wird somit eine Möglichkeit des selbstständigen Übens geschaffen, da sie oftmals bei den Übungen auf die Unterstützung und Zeit der Eltern angewiesen sind. Ebenso kann die Integration der Übungen in ein Computerprogramm am häuslichen Computer neue motivierende Komponenten schaffen. Dafür könnten grundlegende Konzepte von bestehenden Computerprogrammen um eine Echtzeitauswertung der Übungen erweitert werden. Eine Kombination des therapeutischen Ansatzes mit einer virtuellen Spielumgebung ist ebenso denkbar. Ein Vorteil dabei wäre eine mögliche Erweiterung des Programms

¹Von lateinisch os (Mund) und facies (Gesicht), deutsch: den Mund und das Gesicht betreffend

durch die Echtzeitveränderung von Avataren², die eine Abbildung der Patienten als Ersatz für ihr Spiegelbild liefern. Somit ist keine zwangsläufige Auseinandersetzung des Patienten mit seinem Spiegelbild für die Übungsdurchführung notwendig. Gerade bei Störungen der Gesichtsmuskulatur kommt es oftmals zu sichtbaren Veränderungen, wie zum Beispiel bei starken Asymmetrien durch eine halbseitige Lähmung. Diese visuell sichtbaren Beeinträchtigungen können zu einem gestörten Selbstwertgefühl der Patienten bezüglich ihres Aussehens führen. Durch Integration von Avataren können Patienten somit je nach Tagesform zwischen ihrem Spiegelbild und der Darstellung durch einen Avatar wählen. Um grundlegende Schritte für ein solches Programm zu schaffen, müssen zunächst einige Voruntersuchungen getätigt werden. Es gilt zu untersuchen, ob eine Übungserkennung anhand von Tiefeninformationen mittels ausgewählter Merkmale überhaupt möglich ist. Erst anschließend ist ein solches System schrittweise auf das gewünschte Ziel zu erweitern. Die Ziele, die innerhalb der vorliegenden Masterarbeit gesetzt werden, sind im folgenden Abschnitt genauer erläutert.

1.2 Zielstellung

Eines der Ziele dieser Arbeit ist, einen Katalog von Gesichtsmotorik-Übungen im Hinblick auf betrachtete Krankheitsbilder und das entsprechende Therapieziel zu erstellen. Ebenso müssen die Übungen für Aufnahmen mit Probanden und für eine Untersuchung bestimmter Merkmale geeignet sein. Nach der Auswahl einer geeigneten Tiefenbildkamera und der Erstellung eines Aufnahmeszenarios sollen Daten von Probanden aufgenommen werden. Aus diesen Daten soll durch einige Vorverarbeitungsschritte ein Datensatz erstellt werden, an dem im weiteren Verlauf der Arbeit Untersuchungen stattfinden können. Aus verschiedenen Ansätzen aus der Mimik- und der Gesichtserkennung sollen Merkmale gewählt werden, welche bezüglich der Klassifikationsmöglichkeiten der Gesichtsmotorik-Übungen untersucht werden. Diese Analyse soll anhand von statischen Tiefenbildern der Gesichtsmotorik-Übungen geschehen. Für diese Zwecke wird ein bereits vorhandenes MATLAB-Framework aus dem Fachgebiet für Neu-

²Ein Avatar ist eine Kunstfigur, welche den Benutzer in der virtuellen Welt repräsentiert und ist oft im Comicstil gehalten.

roinformatik und Kognitive Robotik der TU Ilmenau für eine Offline-Klassifikation³ erweitert und ergänzt. Das Hauptziel ist die Implementierung der ausgewählten und definierten Merkmale in MATLAB. Ebenso ist die Erweiterung der Vorverarbeitungsschritte, der Klassifikation und der Evaluierung für die Übungserkennung ein Ziel dieser Arbeit. Um aus den extrahierten und ausgewählten Merkmalen einen geeigneten Merkmalsraum für die Klassifikation zu schaffen, soll eine Merkmalsuntersuchung und anschließende Selektion durchgeführt werden. Es ist vorgesehen, die erzielte Klassifikation nach entsprechenden Leistungsbewertungsverfahren auszuwerten.

Zu Beginn dieser Masterarbeit werden in Kapitel 2 einige grundlegende Definitionen und Erklärungen der technischen, medizinischen und informatischen Zusammenhänge aufgeführt, die das Verständnis fördern sollen. Kapitel 3 gibt einen kurzen Überblick über die bisher genutzten Ansätze aus der Mimik- und Gesichtserkennung in Tiefendaten. Ebenso werden bekannte Hilfsmittel vorgestellt, welche das Training von Gesichtsmotrik-Übungen innerhalb einer logopädischen Therapie unterstützen. Anschließend sind alle für diese Arbeit verwendeten Verfahren in Kapitel 4 mit den entsprechenden Auswahlkriterien und den Ergebnissen dargestellt. Dies umfasst den Aufbau des Messsystems, die Erstellung des Übungskatalogs nach entsprechenden Kriterien und das Umsetzungs- und Untersuchungskonzept der ausgewählten Merkmale. Die Implementierung der Merkmale und die Umsetzung der Arbeitsschritte aller zur Merkmalsuntersuchung genutzten Verfahren werden in Kapitel 5 erläutert. Die erzielten Ergebnisse dieser Arbeit werden in Kapitel 6 dargestellt und in Kapitel 7 diskutiert. Abschließend wird in Kapitel 8 ein Resümee der Arbeit dargelegt. Es werden einige Verbesserungsvorschläge des Klassifikationsprozesses sowie einige Möglichkeiten der Vertiefung des Themas in weiterführenden Arbeiten vorgeschlagen.

³Die Erkennung erfolgt erst nach der Aufnahme und bietet keine Echtzeitklassifikation.

Kapitel 2

Gundlagen

Dieses Kapitel gliedert sich in drei Teile und vermittelt grundlegende Informationen, um das Verständnis dieser Arbeit zu erleichtern. Zu Beginn werden in Abschnitt 2.1 die technischen Prinzipien der Tiefenbildkameras erläutert. Da bei Sprach- und Schluckstörungen Gesichtsmotorik-Übungen als Teiltherapie genutzt werden, sind die medizinischen Hintergründe kurz dargelegt (Abschnitt 2.2). Informationstechnische Ansätze, die in der vorliegenden Arbeit genutzt werden, sind in Abschnitt 2.3 erklärt.

2.1 Die Kamerasysteme

Für die Aufnahme von Tiefeninformationen können beispielsweise die zwei folgenden Messsysteme genutzt werden, welche ohne weitere Vorverarbeitungsschritte Tiefeninformationen ausgeben können: Zum einen die *Kinect* der *Xbox 360* der *Microsoft Corporation*, welcher seit November 2010 auf dem Markt ist und zusammen mit *PrimeSense*¹ entwickelt wurde [VILLAROMAN et al., 2011]. Zum anderen der *PMD[vision]®CamCube 3.0* der Firma *PMD Technologies*², welcher im Juni 2010 erschien. Die Systeme sind in Abbildung 2.1 dargestellt.

¹ www.primesense.com

² www.pmdtec.com



(a) *PMD[vision]® CamCube 3.0*,
siehe www.pmdtec.com



(b) *Xbox 360 Kinect*, siehe www.xbox.com/en-US/kinect

Abbildung 2.1: Mögliche Kamerasysteme zur Aufnahme der Tiefendaten.

2.1.1 Das Tiefenbild

Das Tiefenbild ist ebenso wie das Grauwert- oder Intensitätsbild als Matrix aufgebaut, jedoch mit dem Unterschied, dass die Bildmatrix in ihren Elementen keine Grauwerte der jeweiligen Umgebung beinhaltet, sondern Distanzwerte. Die Distanzwerte des dargestellten Objekts werden relativ zur Position des Messsystems angegeben. Die Spalten und Zeilen der Matrix können als X- und Y-Koordinaten des kartesischen Systems interpretiert werden. Somit ist die Bildmatrix $I(x,y)$ eine diskrete Funktion über die Werte x und y und deren Ergebnisse stellen die Distanzwerte über die Z-Komponente dar. In Abbildung 2.2 wird ein Beispiel eines Tiefenbildes gezeigt, welches ein Gesicht darstellt. Die Farben codieren zur besseren Interpretation des Bildes die Tiefenwerte.

2.1.2 Das Prinzip der Abstandsmessung mit Time of Flight-Sensoren

Bei Nutzung des Time of Flight (TOF) Messprinzips wird die Zeit gemessen, die ein Lichtimpuls bestimmter Länge nach dem Aussenden zu einem Objekt und der Reflektion zum entsprechenden Detektor benötigt. Mit Hilfe der gemessenen Zeit kann bei bekannter Geschwindigkeit des Lichts ein Rückschluss auf die Distanz des Objektes zum Sensor erfolgen. Bei der direkten Zeitmessung wird ein hoch energetischer Lichtimpuls von einem Emitter ausgesendet und nach Reflektion an einem Objekt von einem Detektor empfangen. Beide Elemente sind mit einem Timer verbunden, welcher die benötigte Zeit misst. Dieses in [KAHLMANN et al., 2006] beschriebene Messprinzip,

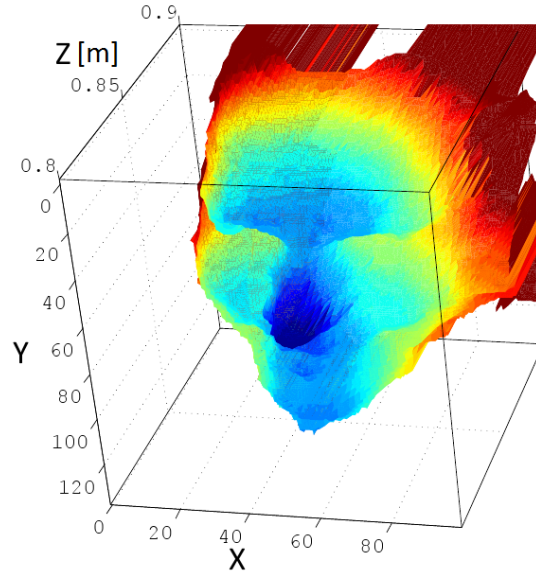


Abbildung 2.2: Tiefenbild eines Gesichts, welches mit MATLAB dargestellt wurde. Die Farben codieren die Distanzwerte, welche an der Z-Achse in Meter abgetragen sind.

die Pulsed Modulation³, setzt allerdings sehr schmale Lichtimpulse und einen genauen Zeitmesser voraus. Nachteilig wirkt auch die Streuung des Lichtes bei der Reflexion, welche die Messung verfälscht. Die meisten Sensoren nutzen bisher Infrarotlicht von lichtemittierenden Dioden mit einer Wellenlänge um 850nm [RAPP, 2007]. Bei der Continuous Wave Modulation⁴ wird, wie in [RAPP, 2007] beschrieben, amplitudenmoduliertes, kontinuierliches Licht mit konstanter Frequenz emittiert. Dabei weist das reflektierte Licht bei Detektion eine Phasenverschiebung zum Originalsignal auf. Es erfolgt daher keine Zeitmessung, sondern eine indirekte Messung der Phasenverschiebung φ_d , welche über die Distanz d entsteht und proportional zur zurückgelegten Distanz des Lichtes ist. Setzt man in die Gleichung der skalaren gleichmäßig beschleunigten Bewegung die Phasenlaufzeit τ_{ph} ein und ersetzt die Kreisfrequenz ω_m des modulierten Lichtes durch die Frequenz ν_m , ergibt sich der Zusammenhang in Gleichung 2.1.

$$d = \frac{\varphi_d c}{4\pi\nu_m} \quad \text{mit} \quad \tau_{ph} = \frac{\varphi_d}{\omega_m} \quad \text{und} \quad \omega_m = 2\pi\nu_m \quad (2.1)$$

³sendet ein gepulstes Signal

⁴sendet ein kontinuierliches sinusförmiges Signal

In diesem Fall ist c die konstante Lichtgeschwindigkeit des modulierten Lichts im Medium Luft. Die Phasenverschiebung kann nicht direkt gemessen werden, daher erfolgt eine Schätzung über die Berechnung der Korrelationsfunktion innerhalb der Kamera zwischen dem modulierten eingehenden Signal und dem Referenzsignal. Das Referenzsignal ist mit dem emittierten Signal phasengleich. Die berechnete Korrelationsfunktion enthält Informationen über die Parameter des reflektierten Signals. Der Gleichanteil verursacht durch das Umgebungslicht, die Amplitude A und die Phase φ_d können wie in [RAPP et al., 2008] und [PLAUE, 2006] beschrieben aus der Funktion berechnet werden. Da die Korrelationsfunktion nur an N diskreten Punkten berechnet werden kann, werden die gesuchten Parameter über eine Regression geschätzt, wofür mindestens drei Abtastpunkte gewählt werden müssen.

Durch diese Schätzungen können in Echtzeit Amplituden- und Phasenwerte als Rohdaten ausgegeben werden, wobei die Phasenwerte nach Gleichung 2.1 in Distanzwerte umwandelbar sind. Beim *CameCube 3.0* wird das Verfahren der Continuous Wave Modulation genutzt. Die Berechnungen der Rohdaten finden hier direkt auf dem genutzten Chip innerhalb der Kamera statt [PLAUE, 2006].

2.1.3 Das Messprinzip des Kinectsensors

Das Messprinzip des Kinectsensors basiert auf der Abstandsmessung mittels Triangulation, bei der normalerweise zwei Kameras benötigt werden. Wie in [MENNA et al., 2011] beschrieben, wird ein einzelner Infrarot (IR)-Lichtstrahl der Wellenlänge 830nm durch ein Beugungsgitter in mehrere Lichtstrahlen aufgeteilt. Durch die Beugung des Lichtstrahls werden gemeinsame Interferenzen zwischen verschiedenen Wellen gleicher Frequenz, aber unterschiedlicher Amplitude und Phase, gebildet. Durch diesen Effekt entsteht ein konstantes und kontinuierliches Speckle-Muster⁵, welches mittels IR-Projektor ausgesendet wird. Die Reflektionen werden an den Objekten mit Hilfe einer IR-Kamera detektiert. Zusätzlich sind innerhalb der Kamera Reflektionen als Referenzdaten des Speckle-Musters bei einem bestimmten Abstand Z_{Ref} gespeichert,

⁵Auch Lichtgranulation genannt; bezeichnet ein Muster aus vielen zufällig verteilten Punkten, welche durch das Interferenzphänomen entstehen.

welche als Referenzbeobachtung genutzt und als zweite Kamera interpretiert werden können. Die Echtzeitaufnahme und das virtuelle Referenzbild bilden zwei Stationen und führen mittels Triangulationsberechnungen zu Distanzwerten, wie in [KHOSHELMHAM und ELBERINK, 2012] erläutert ist.

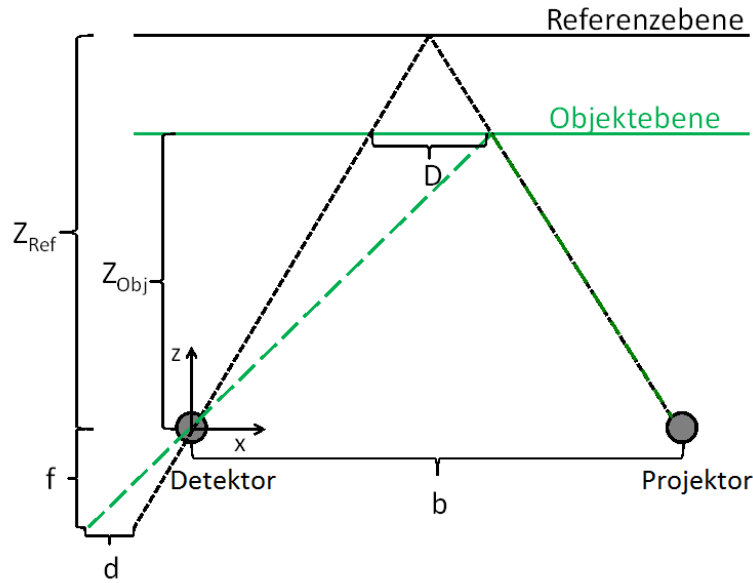


Abbildung 2.3: Abstandsmessung bei der Triangulation im Kinectsensor; Darstellung der geometrischen Zusammenhänge. Vgl. [KHOSHELMHAM und ELBERINK, 2012].

Jeder Punkt des detektierten Speckle-Musters hat innerhalb der Referenzdaten einen korrespondierenden Punkt. Durch die Berechnung der Korrelation zwischen der Referenz und der Beobachtung werden beide IR-Muster miteinander verglichen [MENNA et al., 2011]. Sobald der Abstand der Objektebene Z_{Obj} , die parallel zur Bildebene (XY-Ebene) verläuft, von Z_{Ref} abweicht, ergeben sich Verschiebungen D zwischen den korrespondierenden Punkten des Musters in X-Richtung. Diese Verschiebungen sind in Abbildung 2.3 dargestellt und können durch die Bildkorrelation als Disparitätswerte d berechnet werden. Durch die in Gleichung 2.2 dargestellten geometrischen Zusammenhänge der Triangulation kann aus den gemessenen und bekannten Werten für jeden Speckle des Musters ein Distanzwert Z_{Obj} berechnet werden. Die Länge b ist

dabei der konstante Abstand zwischen der IR-Kamera und dem IR-Projektor, während f die Fokaldistanz der IR-Kamera ist.

$$Z_{Obj} = \frac{Z_{Ref}}{Z_{Ref} \frac{d}{fb} + 1} \quad \text{mit} \quad \frac{D}{b} = \frac{Z_{Ref} - Z_{Obj}}{Z_{Ref}} \quad \text{und} \quad D = \frac{d Z_{Obj}}{f} \quad (2.2)$$

2.2 Sprach- und Schluckstörungen

Sprach- und Schluckstörungen sind durch ihre Ursachen und Symptome oft aneinander gekoppelt und werden häufig gemeinsam therapeutisch behandelt. Im folgenden Abschnitt werden Krankheitsbilder vorgestellt, bei denen als Therapie oder als Teiltherapie Gesichtsmotorik-Übungen genutzt werden.

2.2.1 Dysarthrie

Dysarthrie ist eine erworbene Störung der Sprechmotorik und eventuell anderer am Sprechen beteiligter Systeme, durch die Probleme bei der Artikulation, der Phonation⁶ oder der Sprechatmung (Respiration) auftreten können [JÖRG, 1997]. Die Störung wird durch Läsionen an zerebralen Zentren, Bahnen oder Kerngebieten von motorischen Neuronen ausgelöst, welche beim Sprechen beteiligt sind [BRAUN, 1999]. Man unterscheidet Dysarthrien nach Ort und Ausmaß der organischen Schädigung. Auslöser können Entzündungen, Tumore, Durchblutungsstörungen oder hirnatrophische Prozesse an oben genannten Systemen sein [FRANKE, 2008]. Oft sind Dysarthrien Folgeerscheinungen verschiedener neurologischer Krankheiten, welche je nach Schädigungsort als Auslöser der Sprachstörungen angenommen werden. In Tabelle A.1 in Abschnitt A.1 werden die unterschiedlichen Dysarthriefformen mit einigen beispielhaften auslösenden Pathologien aufgezeigt.

Als Symptom der Dysarthrie gelten Störungen, die eine mangelhafte Artikulation her-

⁶Durch Kehlkopfaktivitäten verursachte Lautbildung mit Hilfe des von der Atmung bereitgestellten Luftstroms.

vorrufen, wie beispielsweise eine geringe Muskelkontrolle der im Ansatzrohr ablaufenden Bewegungen zur Lautbildung und -formung. Sind die gestörten motorischen Systeme nicht nur für die Sprechbewegung, sondern auch für das Schlucken verantwortlich, kommt es zu einer begleitenden Dysphagie, einer Schluckstörung, welche in Unterabschnitt 2.2.4 genauer erklärt wird. Bei der Dysarthrie ist die Ausführung motorischer Befehle gestört, wohingegen das Sprachverständnis, die Wortfindung und das Gedächtnis nicht betroffen sind. Kommen diese Symptome ebenfalls vor, kann es sich nicht nur um eine reine Sprechstörung, sondern um eine begleitende Aphasie⁷ handeln. Zur Kräftigung der Muskulatur oder um muskelathropischen Verläufen entgegen zu wirken, werden, als Teiltherapie der Dysarthrien, zur Verbesserung der Artikulation oftmals Gesichtsmotorik-Übungen eingesetzt.

2.2.2 Dysglossie

Laut [BRAUN, 1999] versteht man unter der Dysglossie einerseits eine Artikulationsstörung in Folge von organischen Schädigungen der peripheren Artikulationsorgane und andererseits Artikulationsstörungen, welche indirekt durch Lähmung der Artikulationsorgane aufgrund einer Schädigung der peripheren Hirnnerven auftreten. Organische Schäden können infolge einer Operation oder durch Unfällen entstehen. Motorik-Übungen können zur Verbesserung der Artikulation genutzt werden.

2.2.3 Dyslalie

In [FRANKE, 2008] ist die Dyslalie als fehlerhafte bzw. fehlende Lautbildung oder als das Ersetzen eines Lautes definiert. Je nachdem, bei welchem Laut Probleme entstehen, werden Dyslalien in der Fachsprache unterschiedlich benannt. Nur Dyslalien, welche durch eine Dysglossie oder motorische Schwächen verursacht werden, können durch Gesichtsmotorik-Übungen verbessert werden. Wird die Lautfehlbildung durch eine auditive Schwäche oder durch anderweitige Auslöser hervorgerufen wie beispielsweise durch entwicklungsbedingte Ursachen, müssen andere Maßnahmen zur therapeutischen Behandlung ergriffen werden.

⁷Neurologisch verursachte Störung des Assoziationsvermögens.

2.2.4 Dysphagie

Störungen des Schluckvorgangs bezeichnet man allgemein als Dysphagien, welche die in [PROSIEGEL und WEBER, 2010] ausführlich beschrieben werden. Neurologische Störungen verursacht durch einen Schlaganfall, Morbus Parkinson, Chorea Huntington, Multiple Sklerose, Morbus Wilson oder schwere Schädelhirntraumata können beispielsweise Auslöser einer Dysphagie sein. Je nachdem, in welchem Bereich die neurologischen Störungen vorkommen, sind andere Phasen des Schluckens betroffen. Nicht neurologische Auslöser können Folgen von Tumoren oder Fehlbildungen im Hals- oder Nasenbereich sein.

Der Schluckvorgang unterteilt sich in vier verschiedene Phasen: Die orale Vorbereitungsphase umfasst die Nahrungszerkleinerung und die Bolusformung⁸. In der oralen Transportphase wird der Bolus in den Mundteil des Rachens geschoben. In der pharyngealen Phase wird der Bolus durch den Rachen transportiert, wobei die Atemwege geschützt werden. In der letzten, ösophagealen Phase wird der Bolus bis zum Mageneingang gebracht. Die ersten beiden Phasen können willentlich ausgeführt werden, sind aber als Reflexketten einzuordnen. Die beiden letzten Phasen werden dagegen als rein reflektorische Prozesse angesehen.

Folgen von Dysphagien können Unterernährung, Dehydratation und Respirationspneumonien durch das Einatmen bakteriell verunreinigten Materials sein. Einbußen der Lebensqualität entstehen bei entsprechenden Maßnahmen, wie zum Beispiel der Nutzung von Ernährungssonden.

Therapiemaßnahmen werden in drei Gruppen unterteilt: Die restituierenden Maßnahmen sollen die gestörten Funktionen wieder herstellen, die kompensatorische Therapie schafft Ersatzbewegungen, um die Störungen ausgleichen zu können. Die adaptiven Maßnahmen ermöglichen unter Einsatz externer Hilfsmittel, beispielsweise durch Nahrungsmittelanpassung, eine Verbesserung der bisherigen Situation. Für restituierende und kompensatorische Maßnahmen können unter anderem oralmotorische Übungen sowie Übungen aus der myofunktionellen Therapie im orofazialen Bereich genutzt werden.

⁸Ein Bolus ist eine kleine Menge zerkleinerter Nahrung oder Flüssigkeit.

2.2.5 Orofaziale Dysfunktion

In [BRAUN, 1999] wird eine grundlegende Übersicht über die orofaziale Dysfunktion gegeben. In [CLAUSNITZER, 2004] und [KITTEL, 1998] sind entsprechende Therapiemaßnahmen vorgestellt. Die orofaziale Dysfunktion ist eine Muskelfunktionsstörung der orofazialen Muskulatur, die in den inneren und den äußeren Bereich eingeteilt wird. Der innere Funktionskreis umfasst die Zungen- und Mundbodenmuskulatur und die Muskulatur des weichen Gaumens. Der äußere Funktionskreis der orofazialen Muskulatur beinhaltet die Mund- und Kaumuskeln. Myofunktionelle Störungen an diesen Muskelgruppen werden in zwei Störungsgruppen unterteilt: Zum einen in die Tonusstörungen und zum anderen in die orofazialen Bewegungsstörungen, aus denen beispielsweise falsche Ruhelagen und Schluckbewegungen von Lippen und Zunge resultieren. Die Ursachen des Krankheitsbildes sind nicht immer geklärt, da es genetische, organische und funktionelle Einflüsse gibt. Die Symptomatik kann jedoch aus Haltungs- und Bewegungsgewohnheiten, Dysglossien oder Zahn- und Kieferanomalien entstehen. Oft ist die Myofunktionelle Therapie (MFT) an eine kieferorthopädische Behandlung gekoppelt. Symptome wie das Zungenpressen, das Offenhalten des Mundes und die Mundatmung, sowie die Hypersalivation⁹ können Folgestörungen wie die unphysiologische Atmung, fehlerhafte Lautbildung und Stimmstörungen verursachen. Beim Zungenpressen stößt die Zunge beim Schluckvorgang gegen die Zähne oder liegt interdental, weswegen Speichel oder Nahrung nicht in Richtung Rachen geschoben werden und der Schluckreflex ausbleibt. Als Begleit- oder Folgeerscheinung können Dyslalien, Dysphagien und Dysarthrien auftreten. Als Therapiemaßnahmen werden Motorik-Übungen im orofazialen Bereich angewendet, welche daher den Namen MFT trägt.

2.3 Informatische Grundlagen

Die Erkennung von Gesichtsmotorik-Übungen ist ein Klassifikationsproblem, für das bestimmte Schritte abzuarbeiten sind, welche in Abbildung 2.4 dargestellt sind. Nach der Definition der benötigten Klassen, der Aufnahme und Vorverarbeitung der Daten

⁹hohe Speichelabsonderung

müssen entsprechende Merkmale extrahiert werden. Aus den Features¹⁰ werden mittels Merkmalsselektion (siehe Unterabschnitt 2.3.1) die für das bestehende Klassifikationsproblem relevanten Merkmale ausgewählt. Diese stehen dann als Eingangsdaten für den Klassifikator bereit. Die Grundlagen der Merkmalsselektion und der Klassifikation werden im Folgenden genauer erläutert.



Abbildung 2.4: Arbeitsschritte für die maschinelle Erkennung von Gesichtsmotorik-Übungen.

2.3.1 Merkmalsselektion

Für jedes Klassifikationsproblem können zahlreiche Merkmale extrahiert werden, die nicht immer eine Aussage über das gestellte Problem enthalten. A priori Wissen spielt daher eine große Rolle, damit nur Merkmale extrahiert werden, bei denen ein Informationswert über die Klassenzugehörigkeit der Daten zu erwarten ist [DUDA et al., 2001]. Beispielsweise lohnt es sich nicht, die Körpergröße als Merkmal für die Erkennung der Gesichtsmotorik-Übungen zu extrahieren. Oft kann trotz Berücksichtigung des Vorwissens keine Aussage über die Relevanz eines Merkmals getroffen werden und es entsteht ein hochdimensionaler Merkmalsraum. Allerdings ist ein kleinerer Merkmalsraum wünschenswert, damit weniger freie Parameter für die Modellbildung existieren. Die richtige Zuordnung der Klasse ist aufgrund einer kleineren Dimension durch eine Minimasuche des Klassifikationsfehlers einfacher zu erreichen. Das Training und die Zuordnung der Klassen in der Einsatzphase ist weniger arbeitsintensiv und somit schneller. Bleibt die Anzahl der Daten im Datensatz gleich, aber die Featureanzahl steigt deutlich, steigt auch die Wahrscheinlichkeit des Overfittings¹¹ des Klassifikators und die Generalisierungsfähigkeit nimmt ab. Daher ist eine Dimensionsredukti-

¹⁰englisch, für Merkmale

¹¹englisch, für Überanpassung; Wenn der Klassifikator zu stark an die Trainingsdaten angepasst wird, einen kleinen Bias-Fehler, aber einen hohen Varianz-Fehler bei unterschiedlichen Beispieldaten aufweist, und damit keine gute Generalisierungseigenschaft besitzt.

on oftmals nötig. Das Ziel der Featureselektion ist, irrelevante¹² Merkmale aus dem Merkmalsraum auszuschließen sowie redundante¹³ Merkmale zu finden und diese gegebenenfalls zu vermeiden [LIU und MOTODA, 2008].

Für die Featureselektion gibt es zahlreiche Ansätze, von denen an dieser Stelle nur einige grob erläutert werden. In [GUYON und ELISSEEFF, 2003] wird ein zusammenfassender Überblick über die Möglichkeiten gegeben. Zum einen gibt es Filteransätze, bei denen klassifikatorunabhängig mit Hilfe bestimmter Relevanzkriterien eine Rangfolge der Features festgelegt wird. Beispiele solcher Relevanzkriterien sind die lineare Korrelation und die Transinformation¹⁴ zwischen den einzelnen Features und der Klassenfunktion. Die einfachste Art eines Filters ist das Ranking, welches die k besten Merkmale bezüglich eines der Kriterien auswählt und alle anderen vernachlässigt. Die Transinformation als Relevanzkriterium wird im Folgenden genauer erläutert, da sie in dieser Arbeit Anwendung findet. Eine weitere Möglichkeit sind Wrapperansätze, bei denen ein beliebiger Klassifikator mit unterschiedlichen Subsets trainiert. Anhand der Leistungsbewertung mittels Fehlerrate wird das beste Subset ausgewählt. Entweder werden alle möglichen Subsets untersucht oder es wird nach bestimmten Suchstrategien vorgegangen. Bei der sequentiellen Vorwärtsselektion wird mit einem leeren Subset begonnen. Alle Merkmale werden einzeln in das Subset aufgenommen und das Modell mit diesem trainiert. Nur das Merkmal, das die beste Fehlerrate des Klassifikators hervorgerufen hat, wird dauerhaft aufgenommen. Das Subset wird schrittweise solange vergrößert, bis der Fehler der Klassifikation nicht mehr verringert wird. Ähnlich verläuft die sequentielle Rückwärtsselektion, bei der mit allen Merkmalen im Subset begonnen und sukzessiv verkleinert wird. Bei den Embedded-Ansätzen wird ein Klassifikator trainiert, welcher seine Struktur an das Klassifikationsproblem anpasst. Anhand dieser Struktur können nach der Trainingsphase Aussagen über die Merkmale getroffen werden.

¹²Merkmale, die keinen Informationswert für das Klassifizierungsproblem tragen.

¹³Merkmale, deren Information über das Klassifikationsproblem ebenso von anderen Merkmalen getragen wird. Die Klassifikationsleistung wird durch das Merkmal nicht verbessert.

¹⁴Gemeinsame Information zweier Zufallsvariablen, wird im folgenden Abschnitt genauer erläutert.

Transinformation

Die Transinformation wird auch Mutual Information (MI)¹⁵ genannt und ist ein statistisches Maß für die Menge der Information, die eine Zufallsvariable über eine andere enthält. Die MI kann über die Entropieformulierung erklärt werden, was in Abbildung 2.5 dargestellt wird und in [ZHANG und HANCOCK, 2011] und [BATTITI, 1994] wie im Folgenden erläutert wird.

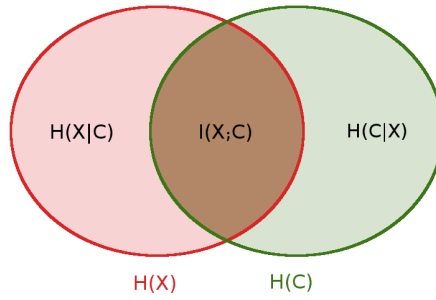


Abbildung 2.5: Zusammenhang zwischen Entropien $H(\mathbf{X})$ und $H(\mathbf{C})$ der Zufallsvariablen \mathbf{X} und \mathbf{C} und der Transinformation $I(\mathbf{X}; \mathbf{C})$; Vgl. [DUDA et al., 2001, S. 633]

Die MI $I(\mathbf{X}; \mathbf{C})$ wird als Schnittmenge zwischen den beiden Entropien $H(\mathbf{X})$ und $H(\mathbf{C})$ der Zufallsvariablen \mathbf{X} und \mathbf{C} dargestellt. Beide Zufallsvariablen sind definiert als $\mathbf{X} = (x_1, x_2, \dots, x_n)$ und $\mathbf{C} = (c_1, c_2, \dots, c_m)$ und können mehrere Werte annehmen. Die Entropie wird definiert als Maß der Unsicherheit oder als mittlerer Informationsgehalt einer Zufallsvariable und ist in Gleichung 2.3 dargestellt. Die bedingte Entropie ist das Maß der Unsicherheit unter Kenntnis einer anderen Zufallsvariable und ist in Gleichung 2.4 zu sehen.

$$H(\mathbf{X}) = - \sum_{i=1}^n P(x_i) \log(P(x_i)) \quad (2.3)$$

$$H(\mathbf{X}|\mathbf{C}) = - \sum_{i=1}^n \sum_{j=1}^m P(x_i|c_j) \log(P(x_i|c_j)) \quad (2.4)$$

¹⁵englisch, für gemeinsame Information

$P(x_i)$ ist dabei die Wahrscheinlichkeit, mit der die Zufallsvariable X den Wert x_i angenommen hat und $P(x_i|c_j)$ ist die bedingte Wahrscheinlichkeit, dass \mathbf{X} den Wert x_i annimmt, wenn die Zufallsvariable \mathbf{C} den Wert c_j angenommen hat. Die bedingte Entropie $H(\mathbf{X}|\mathbf{C})$ hat die Eigenschaft, dass sie nicht größer als $H(\mathbf{X})$ werden kann. Sind die Zufallsvariablen statistisch unabhängig voneinander, nehmen beide Entropien die gleichen Werte an.

$$I(\mathbf{X}; \mathbf{C}) = H(\mathbf{X}) - H(\mathbf{X}|\mathbf{C}) \quad (2.5)$$

Aus der definierten Entropie und der bedingten Entropie über die gleiche Zufallsvariable unter Kenntnis einer zweiten, kann die MI zwischen beiden berechnet werden. In Gleichung 2.5 und der Abbildung 2.5 wird der Zusammenhang verdeutlicht, wobei ebenso daraus ersichtlich wird, dass es sich bei der MI um eine symmetrische Größe handelt.

In [GUYON und ELISSEEFF, 2003] wird die MI zudem über den Abstand der Wahrscheinlichkeitsverteilungen der beiden Zufallsvariablen interpretiert und bildet dann die Kullback-Leiber-Divergenz in der Gleichung 2.6.

$$I(X; C) = \int_x \int_c P(x, c) \log \left(\frac{P(x, c)}{P(x) P(c)} \right) dx dc \quad (2.6)$$

$P(x, c)$, als gemeinsame Wahrscheinlichkeit bzw. Verbundwahrscheinlichkeit und die Einzelwahrscheinlichkeiten der Zufallsvariablen sind dabei unbekannt. Die Transformation kann daher nur über verschiedene Ansätze geschätzt werden. Zum einen kann eine Kerneldichteschätzung vorgenommen werden, bei der die Verteilungen durch überlagerte Gaußkurven beschrieben werden [GROSS, 2011]. Zum anderen können die Verteilungen über eindimensionale bzw. zweidimensionale Histogramme geschätzt werden, wobei sich die Kullback-Leiber-Divergenz diskretisiert, was zur Gleichung 2.7 führt [GUYON und ELISSEEFF, 2003].

$$I(X; C) = \sum_x \sum_c P(x, c) \log \left(\frac{P(x, c)}{P(x) P(c)} \right) \quad (2.7)$$

Bei diesem Ansatz ist zu beachten, dass die Verteilungsschätzung von der Anzahl und dementsprechend von der Breite der Säulen im Histogramm abhängig ist und diese ausreichend groß gewählt werden müssen. Die MI enthält das Maß an Informationen, die eine Zufallsvariable über eine weitere enthält. Sind beide Variablen statistisch unabhängig voneinander, ist die MI gleich Null. Berechnet man die MI über die gleichen Variablen, wird die MI maximal. Der maximale Wert ist dabei abhängig von der Anzahl der Samples im Datensatz, was durch die histogrammbasierte Verteilungsschätzung zu erklären ist.

2.3.2 Klassifikation

Wie in [WALTER, 2004] erläutert, werden bei der Übungserkennung den Eingangsdaten, also den Bildern der aufgenommenen Übungen, diskrete Ausgabewerte zugeordnet, welche als Klassen bezeichnet werden. Ausgewählte Merkmale werden für die Klassifikation aus jedem der Eingangsbilder extrahiert und elementweise zu einem Merkmalsvektor zusammengefasst. Anhand dieses Vektors, welcher für die Klassifikation bereitgestellt wird, kann die Zuordnung der Klassen erfolgen.

Für jedes Klassifikationsproblem muss ein spezieller Klassifikator erzeugt werden. Dafür wird das Klassifikator-Modell in der Lernphase mit Hilfe eines Trainingsdatensatzes trainiert. Dieser Datensatz besteht aus Merkmalsvektoren der Trainingsbilder mit bereits zugeordneten Klassen. Anhand der Verteilungen der einzelnen Merkmale im Merkmalsvektor über die verschiedenen Klassen wird der Klassifikator so moduliert, dass er die Klassen möglichst gut voneinander trennt. Wenn die realen Klassen innerhalb der Lernphase bekannt sind, handelt es sich um Supervised Learning¹⁶. Eager Learner¹⁷ erstellen zunächst ein Generalisierungsmodell, bevor sie für unbekannte Daten genutzt werden können. Die Trainingsphase kann somit unter Umständen rechenaufwändig sein. Bei Lazy Learner¹⁸ hingegen ist die Vorbereitungsphase durch das Speichern der Trainingsdaten beendet. Erst bei der Klassifikation werden die un-

¹⁶englisch, für überwachtes Lernen

¹⁷englisch, für eifrige Klassifikatoren

¹⁸englisch, für träge Klassifikatoren

	ermittelte Klassenzugehörigkeit	
tatsächliche Klassenzugehörigkeit	Klasse 1	Klasse 2
Klasse 1	True Positive (TP)	False Negative (FN)
Klasse 2	False Positive (FP)	True Negative (TN)

Tabelle 2.1: Konfusionsmatrix eines binären Klassifikationsproblems.

bekannten Daten mit den Modelldaten verglichen und entsprechend zugeordnet, was die Einsatzphase aufwändiger macht.

Im weiteren Verlauf dieses Abschnitts sollen drei Klassifikatoren näher erläutert werden, welche in der vorliegenden Arbeit genutzt werden. Die Leistungsbewertung der Klassifikatoren kann auf verschiedene Arten erfolgen, welche grundlegend in [LEBING, 2011] und [GROSS, 2011] vorgestellt werden. Die Bewertung setzt allerdings das Wissen über die tatsächliche Klassenzuordnung für einen Testdatensatz voraus. Eine gute Übersicht liefert eine Konfusionsmatrix, welche bei einem Klassifikationsproblem mit n Klassen die Größe $n \times n$ annimmt und alle möglichen Kombinationen aus tatsächlicher und vom Klassifikator zugeordneter Klasse angibt. Bei einem binären Klassifikationsproblem setzt sich die Matrix aus vier Werten zusammen, die in Tabelle 2.1 aufgezeigt sind. Diese Matrix kann auf beliebig große Multiklassenprobleme erweitert werden. Im Falle der aufgezeigten Tabelle ist die Klasse 1 als positiv und die Klasse 2 als negativ zu bezeichnen. True Positive (TP)¹⁹ bezeichnet die Anzahl der richtigerweise zu Klasse 1 zugeordneten Daten. False Negative (FN)²⁰ umfasst dagegen alle Daten, die fälschlicherweise in die Klasse 2 geordnet wurden, und bildet mit TP die Summe der untersuchten Daten, welche die reale Klassenzugehörigkeit der Klasse 1 aufweisen. Die zweite Zeile ist analog zu interpretieren. Aus den aufgezeigten Werten können weitere Evaluationswerte berechnet werden. Dies betrifft zum einen die True Positive Rate (TPR)²¹, die auch Sensitivität oder Recall eines Klassifikators genannt wird. Ebenso kann die Trefferrate des Klassifikators für eine Klasse angeben. Zum anderen kann daraus die False Positive Rate (FPR) berechnet werden, welche in diesem Fall

¹⁹englisch, für richtigerweise positiv

²⁰englisch, für fälschlicherweise Negative

²¹englisch, für Rate der richtigerweise Positiven

auch Fehlalarmrate genannt wird. Die Balanced Error Rate (BER)²² bildet den Mittelwert über die Anzahl fehlerhafter Klassifikationen beider Klassen und berücksichtigt die Dysbalance von Datensätzen. Bei der Berechnung der BER gibt P die Anzahl aller Daten an, welche die tatsächliche Klassenzugehörigkeit 1 besitzen und N die Anzahl aller Daten mit der tatsächlichen Klassenzugehörigkeit 2 an [GROSS, 2011]. Die genannten Größen sind in Gleichung 2.8 dargestellt.

$$TPR = \frac{TP}{TP + FN}, \quad FPR = \frac{FP}{FP + TN} \quad \text{und} \quad BER = \frac{1}{2} \left(\frac{FN}{P} + \frac{FP}{N} \right) \quad (2.8)$$

Bewertungsmaße, die sich auf alle Klassen beziehen, sind die Erfolgs- und Fehlerrate, welche jeweils die Anzahl aller richtig bzw. falsch klassifizierten auf die Anzahl aller untersuchten Daten normiert [LEBING, 2011].

Für die Bewertung eines binären Klassifikationsproblems ist ebenso das Maß der Receiver Operating Characteristics (ROC)²³ geeignet, bei der die TPR gegen die FPR in einem Graphen abgetragen werden und den ROC-Raum aufspannen. Anhand des Verlaufs kann eine Aussage über die Leistungsfähigkeit eines Klassifikators getroffen werden.

Naive Bayes Klassifikator

Bayes Klassifikatoren sind statistische und lernende Klassifikatoren und gehören zu den Eager Learnern. Ihre Ausgabe ist nicht die direkte Zuweisung der Daten zu diskreten Klassen, sondern eine Wahrscheinlichkeitsaussage der Klassenzugehörigkeit für jede Klasse. Anhand der höchsten Wahrscheinlichkeiten können anschließend die Klassen zugeordnet werden. Die Berechnung dieser Wahrscheinlichkeiten wird im Folgenden aus [HAN und KAMBER, 2006] und [DUDA et al., 2001] zusammenfassend erläutert. Aus den Testdaten werden Featurevektoren $\mathbf{X} = (x_1, x_2, \dots, x_n)$ mit n Einzelmerkmalen extrahiert und können m möglichen Klassen C_1, C_2, \dots, C_m zugeordnet werden, wobei

²²englisch, für balancierte Fehlerrate

²³englisch, für Grenzwertoptimierungskurve

$n, m \in \mathbb{N}$. Die Berechnung der Wahrscheinlichkeiten erfolgt über das Bayestheorem, welches für jede a posteriori Wahrscheinlichkeit $P(C_i|\mathbf{X})$ berechnet wird. Durch die Erweiterung des Nenners um $P(C_i)$ und eine anschließende Umformung ergibt sich das Bayestheorem, welches in Gleichung 2.9 dargestellt ist. Die a posteriori Wahrscheinlichkeit gibt an, mit welcher Wahrscheinlichkeit die zum Featurevektor \mathbf{X} gehörigen Daten der Klasse C_i zugeordnet sind.

$$P(C_i|\mathbf{X}) = \frac{P(C_i \cap \mathbf{X})}{P(\mathbf{X})} = \frac{\frac{P(C_i \cap \mathbf{X})}{P(C_i)} P(C_i)}{P(\mathbf{X})} = \frac{P(\mathbf{X}|C_i) P(C_i)}{P(\mathbf{X})}, \quad i \in \mathbb{N} \quad (2.9)$$

Dabei ist $P(C_i)$ die Eintrittswahrscheinlichkeit der Klasse C_i , die auch a priori Klassenwahrscheinlichkeit genannt wird. $P(\mathbf{X})$ ist die Verteilungsdichte des Featurevektors \mathbf{X} und $P(\mathbf{X}|C_i)$ ist die Verteilungsdichte des Featurevektors, wenn die Klasse C_i vorgegeben ist. Aus allen a posteriori Wahrscheinlichkeiten über die Klassen wird die maximale Wahrscheinlichkeit gesucht. $P(\mathbf{X})$ ist klassenunabhängig und damit konstant, wodurch für die Maximierung nur der Zähler ($P(\mathbf{X}|C_i) P(C_i)$) betrachtet wird. Wenn die a priori Klassenwahrscheinlichkeit unbekannt ist, wird sie über alle Klassen gleich angenommen und kann aus dem Quotienten der Anzahl der Daten in Klasse C_i im Trainingsset und der Anzahl aller Daten des Trainingssets geschätzt werden. $P(\mathbf{X}|C_i)$ muss somit für alle Klassen berechnet werden, was bei vielen Merkmalen sehr aufwändig ist. Bei dem Naive Bayes Klassifikator wird daher die Annahme getroffen, dass alle Merkmale klassenbedingt unabhängig sind, also bei bekannter Klasse eine statistische Unabhängigkeit aufweisen. Mit Hilfe dieser Annahme kann eine Vereinfachung der Maximierung getroffen werden, die in Gleichung 2.10 dargestellt ist. Dabei wird $P(x_j|C_i)$ für jedes Element des Featurevektors aus dem Trainingsdatenset geschätzt, oder bei kontinuierlichen Merkmalen als normalverteilt angenommen.

$$P(\mathbf{X}|C_i) = \prod_{j=1}^n P(x_j|C_i), \quad i, j \in \mathbb{N} \quad (2.10)$$

Die Annahmen, die getroffen werden, können in der Realität meistens nicht vollständig

erfüllt werden. Trotzdem erzielt der Naive Bayes Klassifikator häufig gute Klassifikationsergebnisse.

K-Nächste-Nachbarn Klassifikator

Der k nächste Nachbarn (kNN) Klassifikator gehört zu den Lazy Learnern und ist ein nicht parametrischer und verteilungsfreier Klassifikator und in [HAN und KAMBER, 2006] erläutert. Während der Trainingsphase werden die $n \in \mathbb{N}$ Einzelmerkmale der Featurevektoren $\mathbf{X} = (x_1, x_2, \dots, x_n)$ der Trainingsdaten in einem n -dimensionalem Merkmalsraum gespeichert. Das Lernen des Klassifikators erfolgt über die Analogie zu den bekannten Trainingsdaten. Während der Einsatzphase wird der Testfeaturevektor mit den gespeicherten Trainingsdaten verglichen und die k ähnlichsten Featurevektoren werden gesucht, wobei $k \in \mathbb{N}$. Die Ähnlichkeit wird an metrischen Kriterien, wie beispielsweise der euklidischen Distanz, gemessen, wenn die Features ausschließlich numerisch beschrieben sind. Diese metrischen Distanzwerte geben die Ähnlichkeit der Merkmalsvektoren an und sind damit das Kriterium für die Nachbarschaftsbeziehungen. Bevor die Ähnlichkeitsberechnung durchgeführt werden kann, ist allerdings eine Skalierung der Featurevektoren in einen gleichen Wertebereich notwendig, um eine Vergleichbarkeit zu schaffen. Sobald alle k Nachbarn bestimmt sind, kann den Testdaten die Klasse zugeordnet werden, die am häufigsten in seiner Nachbarschaft im Merkmalsraum vorkommt. Bei kontinuierlichen Klassenlabeln²⁴ kann die Klasse durch Mittelwertbildung über die Nachbarschaft bestimmt werden. Für den Spezialfall $k = 1$ wird den Testdaten die Klasse des nächsten Nachbarn im Merkmalsraum zugeordnet.

Stützvektormethode

Mit Hilfe der Stützvektormethode wird der Klassifikator, der im englischen Support Vector Machine (SVM) genannt wird, modelliert. SVMs gehören zu den Eager Learnern und lösen nach einer Normierung der Merkmale in den gleichen Wertebereich binäre Klassifikationsprobleme. Sie haben zwar im Vergleich zu anderen Klassifikatoren langsame Trainingszeiten, allerdings können sie auch für hoch-dimensionale Merkmals-

²⁴In diesem Fall bezeichnen die Label die unterschiedlichen Klassen

räume genutzt werden und sind weniger anfällig für Overfitting [HAN und KAMBER, 2006].

Alle Trainingsdaten sind durch einen Featurevektor $\mathbf{X} = (x_1, x_2, \dots, x_n)$ im n -dimensionalen Merkmalsraum mit der entsprechend zugeordneten Klasse C_i dargestellt. In [DUDA et al., 2001] und [GÖRZ und ROLLINGER, 2003] wird die Funktionsweise von SVMs erläutert und in dieser Arbeit zusammenfassen beschrieben. Bei SVMs müssen alle Merkmale innerhalb des Featurevektors numerisch sein. Es wird bei den SVMs eine Hyperebene im mehrdimensionalen Raum berechnet, die beide Klassen voneinander trennt. Nimmt man an, dass die beiden Klassen linear voneinander trennbar sind, ergibt sich eine separierende Hyperebene zu $\mathbf{W} \mathbf{X} + w_0 = 0$. Dabei ist \mathbf{W} der Normalenvektor der Ebene und ordnet den einzelnen Merkmalen Gewichte zu, weswegen er Gewichtsvektor genannt wird. Die Variable w_0 gibt den Abstand zum Ursprung des Merkmalsraums an. Die Ebene separiert beide Klassen im gesamten Merkmalsraum voneinander, sodass eine Klasse unter der Ebene und die andere über der Ebene liegt. Diese Aussagen lassen sich zu einer Bedingung in Gleichung 2.11 als Beispiel für einen 2-dimensionalen Merkmalsraum zusammenfassen.

$$C_i = (w_0 + w_1 x_1 + w_2 x_2) \geq 1, \quad 1 \leq i \leq n \quad (2.11)$$

Die strikte Einhaltung dieser Nebenbedingung wird durch die Nutzung des Fehlergewichts aufgehoben, jedoch wird die Nichteinhaltung bestraft. Um die Hyperebene mit den besten Generalisierungseigenschaften zu finden, soll der Abstand zwischen den Klassenbeispielen und der Ebene maximal sein. Dieser ist mit $\frac{1}{\|\mathbf{W}\|}$ gegeben, wobei $\|\mathbf{W}\|$ die euklidische Norm des Gewichtsvektors ist. Die Trainingsdaten, die am nächsten an der Hyperebene liegen, werden Stützvektoren und der Abstand zwischen den Stützvektoren und der Hyperebene Margin genannt. Die Lage und Position der Hyperebene ist allein von den Stützvektoren abhängig, während alle anderen Trainingsdaten keinen Einfluss haben. Die Berechnung ist somit ein Optimierungsproblem und kann mit Hilfe einer Lagrangefunktion aus den aufgezählten Bedingungen formuliert werden. Die Formulierung der Optimierungsfunktion soll hier nicht näher betrachtet werden

und kann in [GÖRZ und ROLLINGER, 2003, S. 541] nachgelesen werden. Ist keine lineare Trennbarkeit der Klassen möglich, werden alle Trainingsdaten in einen höherdimensionalen Merkmalsraum transformiert. In diesem Merkmalsraum kann eine lineare Hyperebene nach beschriebener Optimierung mit maximalen Margin gefunden und in den originalen Merkmalsraum projiziert werden. Im Originalmerkmalsraum wird somit eine nicht-lineare Hyperebene gebildet, welche die Daten optimal trennt. Da Projektionen sehr rechenaufwändig sind, nutzt man den sogenannten „Kerneltrick“. Alle Datenpunkte in der zu optimierenden Lagrangefunktion kommen nur als Skalarprodukt vor. Damit ist ihre Berechnung von der Transformationsfunktion unabhängig. Als Trick verwendet man nur Kernelfunktionen²⁵, die sich in höheren Dimensionen wie Skalarprodukte verhalten. Somit muss zur Trennfunktionsbestimmung keine wirkliche Projektion durchgeführt werden. Die Parameter der Kernelfunktion sind allerdings zusammen mit dem Fehlergewicht vor der Berechnung der Hyperebene zu setzen.

²⁵Transformationsfunktion, welche die Daten virtuell in eine höhere Dimension transformiert, in der die lineare Trennbarkeit gegeben ist.

Kapitel 3

State of the Art

Dieses Kapitel stellt einen Überblick über häufig genutzte Ansätze aus der Gesichts- und Mimikererkennung mittels Tiefenbildern auf. Zudem werden bereits vorhandene Medien für die Unterstützung des Motoriktrainings innerhalb der logopädischen Therapie vorgestellt.

3.1 Gesichts- und Mimikererkennung in Tiefenbilddaufnahmen

Bisher sind noch keine Veröffentlichungen zur Unterscheidung von Gesichtsmotorik-Übungen anhand von Tiefeninformationen bekannt. Seit den 90er Jahren werden allerdings Tiefendaten mit TOF-Kameras aufgenommen und in der Gesichtsdetektion genutzt. Mimikbewegungen sowie Gesichtsmotorik-Übungen äußern sich beide durch Formveränderungen der Gesichtsoberfläche. Die Erkennung von Gesichtsmotorik-Übungen kann daher als Teilgebiet oder Ergänzung der Mimikererkennung¹ betrachtet werden. Für die Auswahl der zu verfolgenden Ansätze wird demnach auf bereits vorhandene Ideen aus der Mimikererkennung zurückgegriffen. Im einem Teilbereich der Mimikererkennung hat es sich durchgesetzt, Emotionen in sechs verschiedene Basisemotionen zu klassifizieren [EKMAN und FRIESEN, 2003]. Bei der Gesichtserkennung werden Profile definiert, um Gesichtsdaten anhand dieser Informationen den einzelnen Personen

¹Im englischen als „Facial Expression Recognition“ bezeichnet.

zuordnen zu können. Es kann geprüft werden, ob ein solches Profil auch Aussagen über verwendete Motorik-Übungen enthält. Die Ansätze der Gesichtserkennung aus Tiefendaten sind daher ebenso zu betrachten, wie die Ansätze der Mimikerkennung. Im folgenden Abschnitt wird daher ein Überblick über die bisherigen Verfahren beider Teilgebiete gegeben.

Bei dreidimensionalen Daten werden häufig normierte Distanzen zwischen Referenzpunkten als Merkmale verwendet. Diese Referenzpunkte können automatisch mit verschiedenen Algorithmen detektiert, oder manuell in die Daten als sogenannte Labelpunkte gesetzt werden. In [YIN et al., 2006] und [YIN et al., 2008] wird der 3D Datensatz *BU-3DFE Dataset*² beschrieben, der neben Bildern zu Emotionen bereits 83 solcher Label bereitstellt. Ebenso zu erwähnen ist das *Bosphorus-Dataset* aus [SAVRAN et al., 2008]. Beide werden genutzt, um eine Vergleichbarkeit zwischen den entwickelten Erkennungsalgorithmen zu schaffen. In den Veröffentlichungen [SOYEL und DEMIREL, 2008], [TEKGÜÇ et al., 2009] und [TANG und HUANG, 2008] werden reine Distanzwerte als Merkmale zur Klassifikation verwendet. Alle wurden am *BU-3DFE Dataset* getestet. Durch eine manuelle Selektion werden in [SOYEL und DEMIREL, 2008] viele der möglichen Distanzen ausgeschlossen und somit die Dimension des Merkmalsraumes reduziert. Durch die Annahme der Gesichtssymmetrie werden nur Distanzwerte in einer Hälfte des Gesichts genutzt. Ebenso werden nur speziell die Gesichtsbereiche betrachtet, in denen die Emotionsbildung verläuft bzw. in denen die Oberflächenveränderungen auftreten. Aus 23 resultierenden Referenzpunkten wird durch Mittelungen ein Distanzvektor aus sechs Elementen gebildet, durch welchen mittels Probabilistisches Neuronales Netzwerk klassifiziert wird. Im Ansatz [TEKGÜÇ et al., 2009] werden alle Referenzpunkte genutzt, aus denen durch alle Kombinationen ein großer Distanzpool erstellt wird. Mit Hilfe eines genetischen Algorithmus werden nur relevante Distanzen aus dem Pool ausgewählt. Vergleichbar ist es in [TANG und HUANG, 2008], in der durch eine automatische Merkmalsselektion die Dimension des Merkmalraums verringern wird. Als Selektionskriterium wird die Maximierung der mittleren relativen Entropie der Merkmale genutzt. Durch den Algorithmus in [LI und SOARES, 2011]

²Binghamton University 3D Facial Expression Dataset

werden nur die zwei Klassen, „fröhlich“ und „neutral“, unterschieden. Die normierten Distanzen werden jedoch um weitere distanzbasierte Merkmale erweitert. Anhand der berechneten Abstände zwischen den Referenzpunkten werden Gesichtsbereiche eindeutig definiert, über die dann Tiefenwerthistogramme gebildet werden. Der Merkmalsvektor besteht aus normierten Distanzen und aus distanzabhängigen Parametern, die das gebildete Histogramm beschreiben.

Eine Erweiterung der normierten Distanzen sind die sogenannten Punktsignaturen, welche erstmals in [CHUA und JARVIS, 1997] als Ansatz für die Gesichtserkennung vorgestellt werden. Die Punktsignaturen sind eindimensionale Darstellungen eines dreidimensionalen Zusammenhangs. Um einen Referenzpunkt innerhalb der Tiefendaten wird eine Kugel gelegt. An der Schnittkurve zwischen der Kugel und des Tiefendatenreliefs werden die Tiefendaten ausgelesen und entsprechend verrechnet. Es ergibt sich ein Tiefenprofil um den Referenzpunkt, welches entlang der Schnittkurve verläuft. Genauere Erklärungen folgen in Unterabschnitt 4.5.2. Aus den Punktsignaturen an bestimmten festgelegten Gitterpunkten und den jeweiligen acht direkten Nachbarn wird ein Profil eines Gesichts entwickelt. Ebenso werden die Punktsignaturen in [WANG et al., 2002] zur Gesichtsidentifikation verwendet, wobei sie in diesem Ansatz als dreidimensionale Ergänzungen von zweidimensionalen Informationen genutzt werden. Zudem werden vor der Berechnung nur vier relevante Referenzpunkte bestimmt. In [MPIPERIS et al., 2007] werden die Punktsignaturen außerdem an Punkten eines Gitters berechnet. Zusätzlich werden in diesem Ansatz die Radien der Kugeln variiert, sodass mehrere Punktsignaturen für jeden Referenzpunkt existieren. Vor der Gesichtserkennung wird eine Selektion der berechneten Punktsignaturen mit Hilfe eines Relevanzkriteriums getroffen.

Die Krümmungsanalyse wurde bisher oftmals im Teilgebiet der Mimikerkennung als lokales Feature verwendet. An jedem der Punkte der vorhandenen Tiefendaten wird innerhalb einer definierten Nachbarschaft eine Fläche eingepasst. Den Flächen, wird über die Berechnung der Weingarten-Matrix und ihrer Eigenwerte eine Krümmungskategorie zugeordnet [SUN et al., 2008, WANG et al., 2006]. Für jeden Pixel der Daten wird ein Krümmungsattribut berechnet. Im Ansatz [SUN et al., 2008] werden acht verschiedene

Krümmungsattribute genutzt, welchen wiederum zu Darstellungszwecken eine Farbe zugeordnet wird. Alle Pixel enthalten somit eine Farbcodierung für ein Krümmungsattribut und werden als „Label-Map“ über das gesamte Gesicht abgebildet. Hingegen wird in [WANG et al., 2006] das Gesicht zusätzlich in sieben für die Mimik aussagekräftige Regionen geteilt, in denen Histogramme über die Krümmungsattribute ermittelt werden.

Durch die Bildung von Normalenvektoren, an jedem Punkt der Tiefendaten, kann ebenfalls eine merkmalsbasierte Darstellung der Daten erfolgen. In [CHEW et al., 2010] werden die Normalen an jedem Punkt der Gesichtsdaten für die Zuordnung zu einem der Referenzgesichter aus dem in [CHANG et al., 2003] vorgestellten Datenset genutzt. Zunächst wird jedes Bild in eine Standardform transformiert. Anschließend werden für jedes korrespondierende Punktpaar der zu vergleichenden Bilder die Winkel zwischen den Normalen berechnet. Der minimale Median über alle Winkel der Bildpaare führt zur Identifikation des Bildes zu einer im Datensatz existierenden Referenz. Die Point Cloud Library ist ein offenes „Stand-Alone-Projekt“, welches für die Datenverarbeitung von dreidimensionalen Punktwolken ein Framework in C++ bereitstellt [RUSU und COUSINS, 2011]. In diesem Framework sind unter anderem Funktionen für die Merkmalschätzung zu finden, mit welchen für jeden Punkt eine Krümmungsanalyse durchgeführt und die Normalenvektoren in einer bestimmten Nachbarschaft berechnet werden [RUSU und COUSINS, 2012].

Aus den aufgeführten Verfahren werden innerhalb dieser Arbeit zwei Ansätze ausgewählt und für die bestehende Problematik Erkennung der Gesichtsmotorik-Übungen angewendet und untersucht.

3.2 Trainingsunterstützung der logopädischen Therapie

Bei vielen therapeutischen Ansätzen mittels Gesichtsmotorik-Übungen ist es für einen schnellen und anhaltenden Erfolg wichtig, dass bestimmte Bewegungsübungen nicht nur in den Therapiestunden beim Logopäden, sondern auch selbstständig zu Hause durchgeführt werden. Die bisher vorhandenen Möglichkeiten für eine Trainingsunterstützung in der Therapie bei Sprach- und Schluckstörungen sollen im Folgenden vorgestellt werden. Einige Logopäden stellen unterstützende schriftliche Anleitungen und Bildmaterialien zur Verfügung. Somit können der Ablauf der Übungen und wichtige Durchführungshinweise zu Hause nachgeschlagen werden. Beispiele für kommerziell erhältliche Materialien zur autodidaktischen Trainingsunterstützung für Patienten mit Morbus Parkinson sind in [BENECKE, 1998] aufgeführt.

Einen weiteren Ansatz zur Trainingsunterstützung gibt es in Form von Computerprogrammen, wie beispielsweise *CoMuZu- Computer-Mund-Zunge* [SCHMIED und SCHMIED, 2007], mit dem eine Unterstützung bei der MFT von Kindern und Jugendlichen erreicht werden soll. Der Therapeut kann Übungen aus Übungskatalogen, sortiert nach Kategorien und Schwierigkeitsstufen, auswählen und für den Patienten freischalten. Zusätzlich kann er Informationen oder Erinnerungen zu den entsprechenden Übungen anfügen. Für jeden Patienten werden die Einstellungen auf einem Datenträger gespeichert, über den der Patient am Heimrechner auf seine Profildaten zugreifen kann. Der Patient erhält Anleitungen der jeweiligen Übungen durch Video- und Tondateien sowie eine schriftliche Anleitung. In Tagebuchoptionen sollen die Patienten beispielsweise Übungszeiten, Bemerkungen oder Fragen zu den Übungen notieren. Nach Rückgabe des Datenträgers ist dem Therapeut die Möglichkeit gegeben, die erstellten Listen auszuwerten. Je nach Angaben zur Häufigkeit der Übungsdurchführung werden, automatisch oder durch den Therapeuten, virtuelle Belohnungen in Form von Medaillen, Pokalen oder speziell Judo-Gürtel vergeben. Die unterstützende Arbeit am Computer stellt eine Herausforderung zum Handeln dar, gilt somit bei Jugendlichen als Motiva-

tionsfaktor und wird von diesen und dem Therapeuten als sinnvolle Unterstützung der MFT angesehen. Nachteilig bei dem Programm *CoMuZu* ist, dass es keine Analysemöglichkeit der Übungsdurchführung während der Übungsphasen gibt. Die Nutzer erhalten somit keine Rückkopplung, da keine Evaluation der Übungsdurchführung möglich ist. Die Logomedien stellen weitere Materialien für die Trainingsunterstützung bei logopädischen Behandlungen zur Verfügung und bestehen aus drei verschiedenen Zweigen [VMS GESELLSCHAFT FÜR THERAPEUTISCHE MEDIEN MBH, 2009]:

- *LogoMem* ist eine Audio-CD-Reihe, welche auditive Anleitungen zur Übung oft benötigter Sprachanwendungen bietet. Durch häufiges Hören und Wiederholen der Kommunikationsstücke im entsprechenden Kontext wird das sprechbasierte Memorieren³ verbessert.
- Die Video- und DVD-Reihe der Logomedien wird als *LogoVid*-Reihe bezeichnet, wobei es sich um videobasierte Anleitungen für logopädische Übungen handelt. Die Videos sind ebenso wie die Audio-CDs je nach therapeutischem Ziel sortiert. Für die in dieser Arbeit beschriebenen Ziele gibt es die Videos „Artikulation 1-3“ und „Mund- und Gesichtsmotorik“, welche bei Störungen der Sprech- und Schluckmotorik eingesetzt werden können. Die entsprechenden Übungen werden im Video vorgeführt und durch auditive und schriftliche Erklärungen ergänzt. Nachteilig ist, dass Art der Übungen, Anzahl der Wiederholungen und Durchführungsgeschwindigkeit durch das übungsbegleitende Videomaterial vorgegeben sind und damit nicht auf den einzelnen Nutzer angepasst werden können.
- Bei der *LogoCeedee* handelt es sich um ein PC-Lernprogramm, welches die Verbindung zwischen Sprache und Bildern fördert. Es steigert das Assoziationsvermögen bei Menschen, bei denen dies durch neuronale Veränderungen gestört ist, wie beispielsweise bei einigen Schlaganfallpatienten. Dieses Programm ist somit eher für Menschen mit aphasischen Störungen von Nutzen.

Die vorgestellten trainingsunterstützenden Medien enthalten bisher keinerlei bewerten-

³Auswendiglernen

de Komponenten. Der Nutzer bekommt nach einer Übungsdurchführung kein Feedback über die Ausführungsqualität der Übung. Somit gibt es für die Patienten keinen Anhaltspunkt, ob die Übung richtig ausgeführt wurde. Der Trainingsplan kann bisher in keinem der aufgeführten Ansätze dynamisch angepasst werden. Führt ein Patient eine Übung regelmäßig und gewissenhaft durch, könnte eine Erweiterung des Therapieplans durch neue Übungen erfolgen. Ebenso könnten die Haltedauer und die Durchführungshäufigkeiten individuell an die Fortschritte der Patienten angepasst werden. Eine motivierende Komponente für Kinder und Jugendliche gibt es bisher nur im Rahmen des integrierten Belohnungssystems bei *CoMuZu*. Diese könnte durch dynamische Therapieplanung und Feedback-Komponenten in einem spielerischen Kontext noch motivierender auf die Kinder und Jugendlichen wirken.

Die Möglichkeit, den Nutzer-Avatar ändern zu können, wäre für eine Spielumgebung eine motivierende Komponente. Oft haben Patienten aufgrund einer neuronalen Störung sichtbare Veränderungen im Gesicht, die zu einem unbefriedigenden Selbstwertgefühl bezüglich des eigenen Aussehens führen. Übungen, die vor dem Spiegel oder einer Kamera mit direkter Bildwiedergabe durchgeführt werden, sind daher oftmals nicht förderlich. Daher könnte eine Nutzung eines Avatars, als Ersatz für das eigene Spiegelbild, Hemmungen bei Patienten abbauen. Als System, welches eine Echtzeitveränderung eines Avatars mit entsprechender Mimik aus Tiefendaten erzeugt, ist das Spiel „Avatar“ für die Xbox 360 mit dem Kinectsensor zu nennen. „Avatar“ erschien am 25. Juli 2011 und ist Teil der *Kinect Fun Labs*⁴. Ziel der Anwendung ist es, Gestik und Mimik der Spieler in eine virtuelle Welt zu übertragen, sodass ein Online-Chatroom entsteht. Laut der Kritik in [VIDEOGAMEZONE, 2011] ist die Mimikübertragung der Spieler in die virtuelle Welt noch nicht ausgereift, da die Mimik verzögert und noch zu ungenau dargestellt wird. Als langfristiges Ziel ist eine Kombination aus einem Trainingssystem mit integrierter Echtzeitevaluation der Übungen und einer spielerischen Umgebung, speziell mit der Avatar-Nutzung, gesetzt. Im Folgenden werden allerdings zunächst die Möglichkeiten einer Übungserkennung untersucht, um die Grundlage für ein System mit Feedbackkomponenten zu schaffen.

⁴Kostenlose Applikationen für Xbox LIVE Nutzer

Kapitel 4

Verwendete Verfahren

Im Folgenden wird eine Auswahl des Messsystems zur Datenaufnahme getroffen und die vorbereitenden Schritte für die Klassifikation erläutert. Die Kriterien der Übungsauswahl und der Katalog an Gesichtsmotorik-Übungen werden vorgestellt. Nach einer genauen Definition der Klassen, wird das Konzept für eine Merkmalsselektion und die anschließende Klassifikation beschrieben.

4.1 Messaufbau

Im Folgenden wird aus den vorhandenen Tiefenbildkameras, welche in Abschnitt 2.1 erläutert wurden, ein Messsystem ausgewählt, mit welchem der Datensatz aufgenommen wird. Ebenso werden aufgrund der technischen Zusammenhänge des gewählten Systems einige Bedingungen für den Messaufbau während der Datensatzaufnahme getroffen.

4.1.1 Auswahl der Kamera

Sowohl der Kinectsensor als auch der *CamCube 3.0* können Tiefeninformationen aufnehmen. Der *CamCube 3.0* erfasst simultan zur Tiefenbildaufnahme ein Grauwertbild, wohingegen der Kinectsensor ein Farbbild aufnimmt. Die zusätzliche Aufnahme von Farbinformationen kann für spätere Erweiterungen der Merkmalsextraktion auf zweidimensionale Daten von Vorteil sein. Zusätzlich unterscheiden sich die beiden Systeme

stark im Kaufpreis. Der Kinectsensor liegt derzeit bei 99,99 Euro¹, wobei der *Cam-Cube 3.0* circa 7750 Euro kostet. Da die Kamera ein Teil des Messaufbaus für eine Trainingsunterstützung der Patienten im häuslichen Bereich ist, darf der Anschaffungspreis keine finanzielle Belastung für den potentiellen Nutzer darstellen. Dieser Punkt ist vor allem zu beachten, da es sich bei der Zielgruppe um Kinder und Jugendliche sowie Senioren handelt. Der Kinectsensor besitzt den Vorteil, dass er neben der Therapieunterstützung anderweitig in Kombination mit der Spielkonsole *Xbox 360* genutzt werden kann. Ebenso ist möglich, dass die Patienten oder Personen im näheren Umfeld bereits einen Kinectsensor besitzen. Ein höherer Preis kann außerdem zu Hemmungen der Patienten im Umgang mit dem System führen.

4.1.2 Bedingungen des Messaufbaus

Der Distanzmessbereich des Kinectsensors wird in [KHOSHELHAM und ELBERINK, 2012] beschrieben und beträgt 0.5m bis 5m. Dieser wird durch die Geometrie des Sensors und Messparameter begrenzt. Die Messung der Motorik-Übungen sollte so nah wie möglich am Kinectsensor erfolgen. Zum einen wird die Tiefenauflösung mit abnehmender Distanz zur Kamera quadratisch erhöht und es können kleinere Tiefenabstände erfasst werden. Zum anderen wird die Standardabweichung in Tiefenrichtung mit zunehmender Distanz zum Sensor quadratisch zunehmen. Im Bezug auf diese beiden Größen sollen für eine bessere Qualität der Aufnahmen die Abstände zur Kamera so klein wie möglich gehalten werden. Die Standardabweichung in Z-Richtung liegt bei einem Abstand von 5m bei 4cm, wobei die kleinste detektierbare Tiefendistanz bereits größer als 7cm ist. Objekte, die weiter als 5m von der Kamera entfernt liegen werden nur noch gering aufgelöst und mit erhöhter Unsicherheit gemessen. Der Kopf sollte sich während der Messungen im Zentrum der aufgenommenen RGB- und Tiefenbilder befinden, da aufgrund der Linsenkrümmungen der Kameras in den Bildern radiale Verzerrungen entstehen, welche im Zentrum am geringsten sind. Zu beachten ist außerdem, dass die Verzerrung der IR-Kamera über die Entfernung zunimmt.

¹Kaufpreis am 20.03.2012 inkl. Adventurespiel im Versandhaus Amazon, www.amazon.de.

Befindet sich ein Objekt in zu geringer Distanz zum Kinectsensor, können aufgrund der geometrischen Anordnung des IR-Projektors und der IR-Kamera im Messsystem Lücken im aufgenommenen Tiefenbild entstehen. Da Projektor und Kamera mit Abstand b auseinander liegen, nehmen beide zwar die gleiche Szene, allerdings um einige Pixel verschoben, auf. Dieser Versatz ist immer gleich und wird bei der Berechnung mittels extrinsischer Parameter verrechnet. Bei zu geringer Distanz ergeben sich jedoch Verdeckungen, die nicht ausgeglichen werden können [KHOSHELHAM und ELBERINK, 2012]. Bei der Gesichtsmotorik-Übungsaufnahme ist die Nase für diesen Problembereich ein kritisches Gebiet. Erfolgt die Frontalaufnahme im Kamerazentrum zu dicht an der Kamera, liegt ein seitlicher Bereich der Nase im Schatten und wird von der IR-Kamera nicht erfasst. Der andere seitliche Bereich wird von der Nase verdeckt, daher vom Speckle-Muster nicht getroffen und somit ebenfalls bei der Messung nicht erreicht, wodurch sich Lücken im Tiefenbild bilden. Lückenlose Tiefenbilder des Gesichts lassen sich nach praktischen Erfahrungen ab einer Distanz von 70cm aufnehmen. Um den Probanden oder Patienten einige Zentimeter Bewegungsfreiheit zu gewährleisten, ist ein Messbereich von 75-95cm einzuhalten. Bei zu starkem Licht im Aufnahmebereich ist der Kontrast des IR-Lichts zu gering und wird nicht komplett von der Kamera aufgenommen, was ebenfalls zu Lücken in der Aufnahme führt und zu vermeiden ist. Der Sensor sollte auf einem Stativ oder einem kleinen Podest auf Kopfhöhe positioniert werden. Somit liegt aus oben genannten Gründen der Kopf im Bildzentrum und es gibt keine glatten spiegelnden Flächen, wie beispielsweise eine Tischplatte direkt unter dem Aufnahmesystem, welche ebenfalls nicht messbare Bereiche hervorrufen [KHOSHELHAM und ELBERINK, 2012]. Die Punktdichte des Speckle-Musters beschreibt die Anzahl der Speckles pro Flächeneinheit. Sie ist bei Objekten mit geringer Distanz zum Sensor kleiner als bei entfernten Objekten. Trotzdem wird eine Nahaufnahme des Gesichts bevorzugt, damit möglichst viele Speckles im Gesichtsbereich liegen und nicht im Hintergrund, welcher für diese Untersuchung keine Informationen trägt.

4.2 Katalog der Gesichtsmotorik-Übungen

Es gibt eine Vielzahl von Gesichtsmotorik-Übungen, die von Logopäden zu therapeutischen Zwecken eingesetzt werden. Ebenso werden in bestimmten Lehrbüchern und Selbsthilfebüchern zahlreiche Übungen vorgestellt, von denen einige ausgewählt worden sind. Um eine Auswahl aus diesen Übungen treffen zu können, die innerhalb dieser Arbeit verwendet werden sollen, werden im Folgenden einige Auswahlkriterien definiert. In Unterabschnitt 4.2.2 folgt eine Vorstellung aller ausgewählten Übungen.

4.2.1 Auswahlkriterien der Übungen

Es werden nur Gesichtsmotorik-Übungen in Betracht gezogen, die für die in Abschnitt 2.2 vorgestellten Krankheitsbilder als Teiltherapie Anwendung finden. Ebenso werden Übungen betrachtet, welche Mimikbewegungen stärken, da oftmals neben den Krankheitsbildern in Abschnitt 2.2, welche durch neurologische Störungen verursacht werden, Mimikschwächen oder sogar Fazialisparesen² einhergehen. Für den Übungskatalog ist es erstrebenswert, dass er möglichst aus wenigen und einfachen Übungen aufgebaut ist, die abwechslungsreiche Trainingsmöglichkeiten bieten, indem sie für verschiedene Bereiche im Gesicht eingesetzt werden können. Jede der ausgewählten Übungen soll krankheitsübergreifend therapeutische Wirkung erzielen. Dabei ist es sinnvoll, dass einfache, statische Übungen gewählt werden, die einzeln und in Kombination mit anderen Übungszuständen oder Ruhelagen zu dynamischen und komplexen Wechselübungen kombiniert werden können. Dabei ist erforderlich, dass die Übungen mit Hilfe des Kinectsensors erfasst werden. Die Zustände und Bewegungen müssen daher bei Übungsdurchführung von außen sichtbar sein. Übungen aus der myofunktionellen Therapie im orofazialen Bereich, die nur innerhalb der Mundhöhle verlaufen und keine äußere Veränderungen verursachen, sind für diese Zwecke nicht geeignet. Um die Aufnahme so einfach wie möglich zu halten, wird nur ein Sensor für die Übungsaufnahme genutzt. Veränderungen bei der Übungsdurchführung müssen in einer Frontalansicht des Gesichts zu erkennen sein. Übungen, bei denen zusätzlich eine laterale Aufnahme

²Gesichtslähmung, als Folge von Verletzungen oder Funktionsstörungen des Fazialisnervs (Gesichtsnerv)

für eine Evaluation notwendig ist, werden ausgeschlossen. Verdeckungen in den Tiefenbildaufnahmen müssen vermieden werden, da der Gesichtsbereiches für eine Klassifikation gut sichtbar sein muss. Somit wird auf Übungen, für die Material benötigt wird, verzichtet. Als Beispiel sind Lippen- und Zungenübungen zu nennen, bei denen Spatel als Widerstände und Knöpfe oder Bonbons als Transportobjekt im Mundraum Verwendung finden. Die nach diesen Kriterien ausgewählten Übungen werden im folgenden Abschnitt näher beschrieben.

4.2.2 Ausgewählte Übungen

Die folgenden Abschnitte geben kurze Durchführungshinweise und Ziele für die in dieser Arbeit betrachteten Gesichtsmotorik-Übungen. Der Übungskatalog wurde in Kommunikation mit mit Logopäden³ sowie der Informationszusammenfassung aus den Büchern [BENECKE, 1998], [CLAUSNITZER, 2004], [PROSIEGEL und WEBER, 2010], [THIELE, 1997] und [KITTEL, 1998] erstellt. Eine Zusammenfassung zum Nachschlagen der genutzten Übungen mit beigefügten Seitenangaben der jeweiligen Quellen ist in tabellarischer Form im Abschnitt A.3 dargestellt.

Sofern nicht anders vorgegeben, sind statische Halteübungen fünf bis zehn Sekunden zu halten. Dynamische Übungen hingegen sind weniger als fünf Sekunden in den jeweiligen Lagen zu halten. Bei diesen Übungen sollen die Wechselbewegungen zwischen den verschiedenen Zuständen eine Wirkung erzielen.

Während jeder Übungsdurchführung ist zu beachten, dass keine Ausgleichsbewegungen, sogenannte Synkinesien, in den Bewegungsablauf integriert werden, da diese meistens nur sehr mühevoll wieder abtrainiert werden können. Die Gesichtsregionen, die nicht an der Übung beteiligt sind, sollten daher bei Durchführung der Übungen möglichst entspannt sein. Auf eine aufrechte Sitzposition der Probanden und Patienten ist Wert zu legen.

³Ressortleiterin Eva Schillikowski der sprachtherapeutischen Abteilung der m&i-Fachklinik Bad Liebenstein und Frau Irina Stangenberger aus Ilmenau/ Arnstadt

Augen- und Stirnübungen

„Das Stirnrunzeln“ wird durch das Hochziehen der Augenbrauen verursacht, wodurch sich die Stirn quer zur Muskulatur in Falten legt. Sie kann als Halteübung einige Sekunden in diesem Zustand gehalten werden oder in Kombination mit der Ruhelage der Augenbrauen als dynamische Übung funktionieren. Für die Übung „Augenbrauen zusammenziehen“ werden die Augenbrauen in die Mitte zusammen geschoben. Auch diese Übung kann mit der Ruhelage der Augenbrauen kombiniert werden. Für das Trainieren des Augenschlusses wird die Übung „Augen schließen“ durchgeführt. In diesem Fall ist es wichtig, dass beide Augen gleichzeitig geschlossen werden. Dabei soll das Schließen nur durch die Augenlidbewegung geschehen und keinesfalls durch das Hochziehen der Wangen. Das Zupressen, welches in der Übung „Augen zusammenpressen“ erwünscht ist, soll an dieser Stelle vermieden werden. Bei der Pressbewegung ist ebenfalls ein gleichzeitiges Schließen der Augen zu erreichen.

Lippenübungen

Lippenübungen bzw. Orbikularisübungen⁴ aktivieren allesamt den Kreismuskel des Mundes. Als erste Lippenübung ist „Lippen spitzen“ zu nennen. Ausgangspunkt für diese Übung ist die sogenannte LM-Haltung. Die LM-Haltung ist die Kombination aus der L- und der M-Haltung, welche jeweils die Ruhelage der Lippen bzw. der Zunge beschreibt. Die L-Haltung bezeichnet die Ruhelage der Zunge und wird am besten durch die Haltung beschrieben, die die Zunge bei der Aussprache des Buchstaben „L“ einnimmt. Ebenso ist die M-Haltung zu betrachten, da sich die Lippen in entspanntem Zustand ähnlich der Aussprache des Buchstaben „M“ befinden. Aus dieser Lage werden die Lippen sternförmig zusammengezogen. Die Zahnreihen bleiben dabei locker aufeinander und die Lippen geschlossen. Als Hilfestellung kann auch das Zusammenschieben der Mundwinkel durchgeführt werden. In [CLAUSNITZER, 2004] werden die spitzen Lippen auch als „Kirschmund“ bezeichnet. Eine sehr ähnliche Übung, die bei vielen kaum von den „Spitzen Lippen“ zu unterscheiden ist, wird „Kussmund“ genannt. Sie

⁴Der Musculus obicularis oris, in diesem Zusammenhang oftmals als Orbikularis abgekürzt, bildet die muskuläre Grundlage der Lippen [KIRSCH et al., 2011, S. 580].

wird bei gleicher Bewegung allerdings ein wenig lockerer ausgeführt. Als Wechselübung kann die Übung mit dem „Breitziehen der Lippen“, der „O-Form“ oder der Ruhelage kombiniert werden. Durch das Ausführen als Halteübung kann eine Tonussteigerung im Orbikularismus erreicht werden, da er beim Üben immer wieder kontrahiert wird. Es können Gewebedehnungen bzw. Oberlippenverlängerungen und Lippenrundungen erzielt werden. Sie fördert die Unabhängigkeit der inter- und extraoralen Muskulatur. Das „Lippen Breitziehen“ wird auch „Mundwinkelzugübung“ genannt. Dabei werden die Mundwinkel seitwärts gezogen und in diesem Zustand gehalten. Die Mundwinkel führen dabei keine Bewegung nach oben oder unten aus. Die Übung kann wie oben bereits erwähnt als Wechselübung dynamisch fungieren. Eine mögliche Erweiterung ist, die Mundwinkel einzeln nach außen zu ziehen und eine dynamische Übung aus dem Wechsel der Seiten zu gestalten. Die Ausgangslage ist die M-Haltung. Die Lippen sind daher geschlossen und sollen während des Übungsverlaufes locker geschlossen bleiben. Bei der richtigen Durchführung wird die Ringmuskulatur gestreckt und es kommt bei verminderter Wangen- und Lippenmuskulatur zur Tonusstärkung. Daher sollten selbst minimale Öffnungen zwischen den seitlichen Bereichen der Lippen vermieden werden. Die Vokalübungen sind ein wichtiger Bestandteil der Lippenübungen und werden unter anderem oft in der Teiltherapie von Dysphagie und Morbus Parkinson genutzt. Sie können ebenfalls mit weiteren Übungen zu Wechselübungen kombiniert werden. Die Übung „A formen“ wird bei weit geöffnetem Mund durchgeführt, wobei die Wangenmuskulatur komplett entspannt ist. Für die Durchführung als Halteübung kann es hilfreich sein, einen A-Laut zu bilden und den Zustand einige Sekunden zu halten. Hierbei muss der Laut nur in den ersten Sekunden gebildet werden und kann beim weiteren Halten verstummen. Die Nutzung der Lautbildung als Unterstützung der Bewegungsformung ist bei allen Vokalübungen hilfreich. Die Übung „O formen“ ist eine Möglichkeit, die Oberlippenmuskulatur zu aktivieren und zu verlängern. Die Übung zielt auf die zentralen Zirkulärfasern der Muskulatur. Auch diese Art der Vokalübung wird in der M-Haltung begonnen. Der Mundspalt wird von allen Seiten gleichzeitig und gleichmäßig geschlossen, bis nur noch eine kleine Öffnung vorhanden ist. Bei der Durchführung sind die Zahnreihen nicht geschlossen. Ganz ähnlich ist die Vokalübung „U formen“. Bei „I formen“ werden beide Mundwinkel nach außen gezogen, wobei der

Mund leicht geöffnet ist. Das Breitziehen der Lippen soll hier ohne Anspannung der Halsmuskeln geschehen. Die Zahnreihen sind ebenfalls nicht geschlossen.

Zungenübungen

Bei den Zungenübungen sind mehrere verschiedene Variationen zu betrachten. Sie unterscheiden sich hauptsächlich in der Streckrichtung der Zunge und in der aktivierten Muskulatur. Zungenübungen verbessern die eigene Wahrnehmung über die einzelnen Muskelzustände der Zunge. Bei „Zunge herausstrecken“ wird die Zunge bei weit geöffnetem Mund horizontal herausgestreckt. Dabei darf die Zungenspitze nicht nach oben bzw. unten zeigen. Die Zunge soll so weit wie möglich gestreckt werden, bis ein Ziehen in der Muskulatur spürbar ist. Diese Position sollte gehalten werden, ohne dass die Zunge auf der Unterlippe oder der unteren Zahnreihe aufliegt. Diese Zungenprotraktion⁵ führt zu einer Stärkung des Zungenmittelteils und der transversalen Zungenmuskulatur, ebenso wird die Binnenmuskulatur der Zunge stabilisiert. Wird die Zunge bei weit geöffnetem Mund in Richtung Nase gestreckt, handelt es sich um die Übung „Zunge zur Nase“. Zu beachten ist, dass der Unterkiefer die Zunge nicht nach oben drückt, sondern die Bewegung allein durch die Zunge bewirkt wird. Dadurch wird die Zungenspitzenhebung aktiviert und gestärkt. In Kombination mit der Übung „Zunge herausstrecken“ wird sie zu einer Wechselübung und häufig „Winken“ genannt. Bei „Zunge in den Mundwinkel“ wird die Zunge ebenfalls ausgestreckt. Im Gegensatz zu den erwähnten Zungenübungen wird sie nun seitlich bewegt und die Spitze in den Mundwinkel gelegt. Dabei soll der Weg zum Mundwinkel direkt gefunden werden. Zur dynamischen Übung kann sie durch Seitenwechsel abgewandelt werden. Als Erweiterung kann statt des direkten Weges eine Kreisbewegung genutzt werden, um die Zungenspitze zum anderen Mundwinkel zu führen. Die Übung steigert die Flexibilität und Mobilität der Longitudinalfasern der Zunge. Die Muskeln zur Seitwärts- und Drehbewegung der Zunge werden ebenfalls aktiviert. Alle der bisher genannten Zungenübungen können durch die unterschiedliche Streckrichtung gut als Wechselübung miteinander kombiniert werden. Eine Kombination aus Zungen- und Wangenübung ist

⁵Zungenprotraktion bezeichnet das Vorschieben der Zunge

die Übung „Zunge in Wange bohren“, die in dieser Arbeit auch als „Zunge in Wange boxen“ bezeichnet wird. Die Zunge wird zwischen den Zahnreihen, welche folglich nicht aufeinander liegen, mit der Zungenspitze von innen kräftig gegen die Wange gepresst, wobei sich eine von außen sichtbare Wangenbeule bildet. Das Pressen der Zunge gegen die Wange steigert wie „Zunge in den Mundwinkel“ die Flexibilität und Mobilität der Longitudinalfasern, wobei die Wange als Widerstand wirkt. Als zusätzlicher Effekt wird die Wangenmuskulatur gestreckt und ein wenig gestärkt.

Wangenübungen

Bei der Wangenübung „Wangen aufblasen“ wird im Mundraum der Luftdruck maximal erhöht, wobei keine Luft durch den Lippenspalt entweichen sollte. Dabei hebt sich das Gaumensegel und die Wangenmuskulatur wird gestreckt und gedehnt, was zur Korrektur bei Verkrampfungen, also zur Entspannungshilfe, dienen kann. Dieser Zustand soll einige Sekunden gehalten werden und führt bei häufigem Üben zur Verbesserung des Lippenschlusses. Patienten und Probanden dürfen während der Übung ruhige Nasenatmung durchführen. Durch geringe Krafteinwirkung während der Übungsdurchführung, beispielsweise durch die flachen Hände, die von außen auf die Wangen gebracht werden, wird der Schwierigkeitsgrad der Übung erhöht, was allerdings im Laufe dieser Arbeit aufgrund der entstehenden Verdeckungen nicht betrachtet wird. Es können Tonussteigerungen im Orbikularis⁶ und Velum⁷ erreicht werden. Kontrahiert man zusätzlich zum Luftdruck eine der beiden Wangenmuskeln, wird nur die gegenüberliegende Wange aufgeblasen. „Eine Wange aufblasen“ verbessert daher durch häufiges Üben die Kontraktionsfähigkeit der Wange. Beide Wangen können als dynamische Übung abwechselnd kontrahiert und aufgeblasen werden. Bei der dynamischen Übung hilft ein leichter Schluss der Zahnreihen, um eine Kieferbewegung während des Seitenwechsels als Ausgleichsbewegung zu vermeiden.

⁶Deutsche Kurzform für Musculus Orbicularis Oris, dem Ringmuskel des Mundes, welcher im Lippenbereich liegt.

⁷Kurzform für Velum palatinum, Gaumensegel

4.3 Erstellung der Basisprimitive

Mit dem Ziel, dass die Motorik-Übungen in einer Fortsetzung dieser Arbeit als dynamische Kombinationsübungen aus den einzelnen Zuständen der Übungen genutzt werden, werden sie in zu erkennende Bewegungseinheiten unterteilt. Sind die Übungen auf eine Art von Bewegung beschränkt, so werden sie selbst als Bewegungseinheit betrachtet. Als Vorbild für die Aufteilung und Begründung der Bewegungseinheiten gilt das Das Facial Action Coding System (FACS)⁸, welches daher in seinem Aufbau und der Zielstellung im Folgenden genauer erläutert wird. Anschließend wird nach einer Analyse eine Anpassung des FACS auf die in dieser Arbeit bestehende Problematik vorgenommen, wobei ein System für die vorliegende Masterarbeit formuliert wird.

4.3.1 Das Facial Action Coding System

Das Facial Action Coding System (FACS) ist ein deskriptives Dokumentationssystem für Gesichtsbewegungen. Es wurde seit 1978 von Ekman et al. entwickelt und seitdem zur schriftlichen Dokumentation von Gesichtsausdrücken genutzt [UNIVERSITÄT DES SAARLANDES, 2009]. In der Emotionspsychologie wird das System angewendet, um reale von unechten Emotionen besser unterscheiden zu können. Ebenso findet es in der Computeranimation Anwendung, um eine möglichst realistische Darstellung von Emotionen erzeugen zu können. Die Bewegungen des Gesichts werden in grundlegende Primitive zerlegt und als Action Unit (AU)⁹ bezeichnet. Diese Primitive wurden auf anatomischer Grundlage der mimischen Muskulatur gebildet und für die kleinsten möglichen Bewegungen im Gesicht genutzt [EKMAN et al., 2003b]. Eine AU beschreibt eine oder mehrere zusammengefasste Muskelbewegungen. Die verschiedenen Einheiten sind durch Zahlen codiert. Es gibt unterschiedliche AUs, welche sich in Einheiten des oberen und unteren Gesichtsbereich unterteilen. Der untere Bereich wird zusätzlich, basierend auf den Bewegungsrichtungen, in fünf verschiedene Gruppen aufgeteilt. Zu den codierten Bewegungen der mimischen Muskulatur werden die Intensitäten der Bewegungsausführung bei maximaler Durchführung durch die Buchstaben A bis E

⁸englisch, für Gesichtsbewegungen-Kodierungssystem

⁹englisch, für Bewegungseinheiten

angegeben [EKMAN et al., 2003a]. Die Action Descriptor (AD)¹⁰ sind von den AU zu unterscheiden. Sie beschreiben ebenso Zustände des Gesichts, wurden jedoch nicht auf muskulärer Basis entwickelt, weswegen sie gröber definiert sind. Die ADs sind in der Zahlencodereihe der AUs enthalten. Als Beispiel ist das Zeigen der Zunge als AU 19 zu nennen, wodurch jegliches Zeigen der Zunge codiert, jedoch keinerlei Bewegungsrichtung der Zunge berücksichtigt wird. Das FACS enthält zusätzliche Codes, um Kopf- und Augenbewegungen eindeutig zu beschreiben. Verdeckungen des Gesichts werden ebenso durch entsprechende Codes dokumentiert. Die AUs werden im Einzelnen in [EKMAN et al., 2003a] mit den jeweiligen innervierten Muskeln, den sichtbaren Veränderungen und entsprechenden Intensitätsstufen vorgestellt. Diese vollständige Zusammenfassung und das enthaltende Lehrmaterial, welches kostenpflichtig erworben werden kann, dient als Referenz für alle codierten Dokumentationen. Da die Mimik eines Menschen vollständig mit dem FACS codiert und in Basisprimitive eingeteilt werden kann, dient es als Basis für die Bewegungsprimitive der zu klassifizierenden Übungen. Die komplette Übernahme des Systems für diese Arbeit ist nicht möglich, die Gründe dafür sind im folgenden Kapitel erläutert.

4.3.2 Anpassen des FACS an die bestehende Problematik

Bei Gesichtsmotorik-Übungen handelt es sich nicht um natürliche Mimikbewegungen, daher sind einige der Übungsbewegungen nicht im FACS enthalten. Alle sichtbaren Zungenbewegungen werden bisher, wie bereits erwähnt, nur durch AU19 codiert. Zungenbewegungen sind jedoch ein Teil der myofunktionellen Übungen, die in dieser Arbeit betrachtet werden, weswegen sie detaillierter unterschieden werden müssen.

Eine weitere Problematik umfasst den Bereich der Wangenübungen. Die Wangen werden nicht durch Aktivität der Wangenmuskulatur aufgeblasen, sondern indirekt durch das Erzeugen eines Innendrucks und dem zusätzlichen Lippenschluss. Für das Aufblasen nur einer Wange ist zusätzlich die Wangenmuskulatur auf der jeweils anderen Seite aktiv. Drückt die Zunge von innen gegen die Wange, tritt die Bewegung bzw. Formveränderung ebenso an anderer Stelle auf als die Innervation der Muskulatur.

¹⁰Bewegungsbeschreibungen

Demnach muss das FACS für die Beschreibung der Motorik-Übungen im Gesicht um einige neue Bewegungseinheiten erweitert werden und an einigen Stellen bezüglich der Ursache der Bewegung detaillierter unterschieden werden. Einheiten, die nicht benötigt werden, können vernachlässigt werden. Es wird ein neues Codierungssystem aus Basisprimitive (BP) erstellt, welches auf der Idee und dem Aufbau des FACS basiert. Bei den Zungenbewegungen ist bisher keine Einteilung der Bewegungen bezüglich innervierter Muskulatur in AU vorgenommen worden. Daher werden diese mit Hilfe von [KIRSCH et al., 2011, S. 619-628] auf anatomischer Basis analysiert und in entsprechende BP aufgeteilt. Die beschriebenen, indirekten Bewegungen werden nicht auf muskulärer Basis, sondern auf deskriptiver Bewegungsbasis in BP eingeteilt. Die Zuordnung der AUs und der BP zu den Stirn-, Augen- und Lippenübungen sind in Tabelle A.2 aufgeführt. Die Erweiterungen des FACS und die BP der Zungen- und Wangenübungen sind in Tabelle A.3 aufgelistet. Eine Kurzform der Ergebnisse ist in Tabelle 4.1 dargestellt. Das ergänzende Adjektiv der BP „bilateral“ weist auf dessen Zweiseitigkeit hin. Das BP kann auf beiden Seiten einzeln durchgeführt werden.

Übungsabk.	Basisprimitive	Übungsabk.	Basisprimitive
Stirn	BP 1 (bilateral)	Zunge	BP 11
Brauen	BP 2	ZungeO	BP 12
Auge	BP 3 (bilateral)	ZungeLi	BP 13 (bilateral)
Augenpressen	BP 4 (bilateral)	ZungeRe	
Kuss	BP 5	BoxenLi	BP 14 (bilateral)
Breit	BP 6	BoxenRe	
AForm	BP 7	Wangen	BP 15
OForm	BP 8	WangeLi	BP 16 (bilateral)
IForm	BP 9	WangeRe	
UForm	BP 10		

Tabelle 4.1: Übersicht der Basisprimitive

Anzumerken ist, dass die Aufteilung der BP fast vollständig den eigentlichen Übungen entspricht, allerdings die Grundlage für dynamische und kombinierte Übungen liefert. Ebenfalls kann das System durch Intensitätsangaben, wie es beim FACS genutzt wird,

erweitert werden. Einige bilaterale BP werden für eine Übungsdurchführung nur gemeinsam genutzt, wie es bei den Augen- und Stirnübungen der Fall ist. Die BP15 stellt einen Spezialfall dar, da sie nicht aus der Kombination der bilateralen BP16 gebildet wird. Die Wangenmuskulatur ist bei der Durchführung nicht auf beiden Seiten angespannt, sondern komplett entspannt und bildet somit eine eigenständige BP.

4.4 Erstellung des Datensatzes

Alle Gesichtsmotorik-Übungen, welche im Übungskatalog im Unterabschnitt 4.2.2 genauer beschrieben sind, wurden von elf Probanden vor dem Kinectsensor durchgeführt, welcher die Übungsdurchführungen erfasste. Der erstellte Datensatz enthält keine Verdeckungen durch Fremdobjekte und wurde mit gesunden Probanden, die keine Schwächen der mimischen und orofazialen Muskulatur aufweisen, erstellt. Die Aufnahme erfolgte mit dem in Abschnitt 4.1 ausgewählten Messsystem und unter den definierten Bedingungen. Alle ausgewählten Probanden hatten keine Erfahrungen in der Ausführung von Gesichtsmotorik-Übungen. Die einzelnen Übungen wurden daher vor der Aufnahme kurz vorgeführt und trainiert. Den Probanden wurden, nach den in Unterabschnitt 4.2.2 aufgeführten Anleitungen, zu jeder Übung entsprechende Durchführungshinweise gegeben. Jeder Proband hat eine Einverständniserklärung unterschrieben, die besagt, dass die Daten anonymisiert im Rahmen dieser und anschließender Arbeiten verwendet werden dürfen.

Der Kinectsensor wird durch eine Aufnahmeeinheit auf einem Laptop mit dem Betriebssystem openSUSE gesteuert. Dabei werden ca. jede Sekunde statische RGB- und Tiefenbilder der Übungsdurchführung gespeichert. Die RGB-Bilder werden im JPEG-Format mit einer Bittiefe von 24 und die Tiefenbilder im PNG-Format mit einer Bittiefe von 16 gespeichert. Beide Bildarten sollen möglichst zeitsynchron aufgenommen werden, was anhand des dreizehnstelligen Unix-Zeitstempels¹¹ im Dateinamen überprüft wird. Korrespondierende RGB- und Tiefenbilder unterscheiden sich in

¹¹der Unix-Zeitstempel eine mindestens zehnstellige Zahl, die das Datum mit der sekundengenauen Uhrzeit seit dem 1. Januar, 1970 um 00:00 Uhr codiert, siehe www.unixtimestamp.de.

ihrer Aufnahmezeit maximal um 16 Millisekunden. Für jeden Probanden werden die RGB-Bilder sortiert. Nur die Bilder, welche die maximale Übungsdurchführung festhalten, werden im weiteren Verlauf der Arbeit genutzt. Das bedeutet, dass alle Bilder einer Übungssequenz, welche den Weg zur Halteposition der Übung und von dieser weg, nicht berücksichtigt werden. Dabei werden für jede Übung und jeden Proband ungefähr sieben Bilder ausgewählt, bei denen die Übung möglichst exakt ausgeführt wurde. Alle weiteren Bilder, welche die Übungsübergänge und die Anfangs- und Endstadien der Übungsdurchführung darstellen, werden aussortiert. In den ausgewählten RGB-Bilder werden mit Hilfe des Tools *AAMLabelTool 2*¹² 58 spezielle Punkte des Gesichts mit sogenannten Labelpunkte¹³ manuell markiert. Die markierten Punkte sind entsprechend Abbildung 4.1 definiert und durchnummeriert. Das Setzen der Labelpunkte in die RGB-Bilder aller Probanden wurde von zwei verschiedenen Personen durchgeführt. Neben den definierten Punkten werden zu jedem Bild die dargestellte Übung, eine anonymisierte Probandenidentifikationsnummer und das Geschlecht des Probanden zugeordnet. Ebenso wird vermerkt, ob Verdeckungen der Positionen der Labelpunkte durch beispielsweise die Zunge vorhanden sind. Alle Labelpositionen und

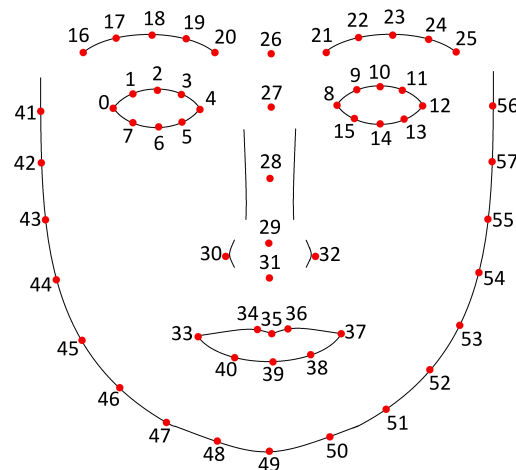


Abbildung 4.1: Alle mit dem AAMLabelTool 2 eingefügten Labelpunkte mit entsprechender Nummerierung an der definierten Position.

¹²Labeltool, welches vom Fachgebiet Kognitive Robotik und Neuroinformatik bereitgestellt wurde.

¹³In diesem Zusammenhang speziell definierte Punkte

zugeordneten Attribute werden innerhalb einer Textdatei gespeichert. Die Tiefenbilder sind aufgrund der Geometrie des Kinectsensors mit den Intensitätsbildern nicht komplett deckungsgleich. Die Tiefenbilder werden daher mit Hilfe von intrinsischen und extrinsischen Parametern transformiert und mit den jeweiligen RGB-Bildern registriert. Die im RGB-Bild gesetzten Label können somit für das Tiefenbild genutzt werden, um bestimmte Gesichtsregionen eindeutig zu finden. Lücken, die sich eventuell im Gesichtsbereich des Tiefenbildes befinden, werden mit Hilfe eines Medianfilters in der direkten Nachbarschaft der Lücke geschlossen. Ein Beispieltiefenbild des Datensatzes der Übung „Kuss“ ist in Abbildung 4.2 dargestellt, in welchem die Farben den Tiefenwert codieren und die Label in schwarz eingezeichnet sind.

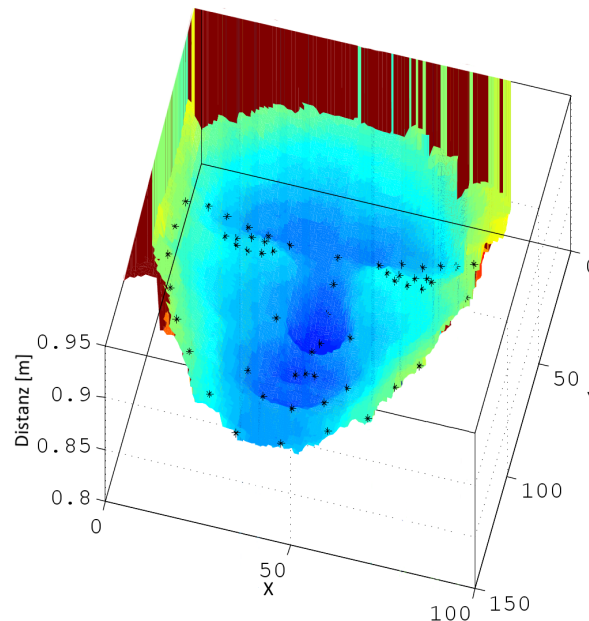


Abbildung 4.2: Tiefenbilddarstellung der Übung „Kuss“ mit den im RGB-Bild gesetzten Labelpunkte in schwarz.

4.5 Auswahl und Definition der Features

Für die Untersuchung der aufgenommenen Daten bezüglich der Erkennung der BP, welche in Unterabschnitt 4.3.2 beschrieben wurden, müssen entsprechend aussagekräftige Merkmale aus den Bilddaten extrahiert werden. In Abschnitt 3.1 wurden einige

Merkmale aus dem Bereich der Gesichts- und Mimikerkennung in Tiefendaten vorgestellt. Aus diesen werden repräsentative Merkmale gesucht, die auf ihre Relevanz zur BP-Erkennung untersucht werden. Normierte Distanzen in 3D-Daten sind skalierungs-, translations- und rotationsinvariant. Diese Eigenschaft bewirkt, dass die Kopfposition keinen Einfluss auf die Merkmalsberechnung, und somit auf das Klassifizierungsergebnis hat. Zu beachten ist, dass die Rotationsinvarianz um die X- und Y-Achse, welche die Bildebene aufspannen, nur für kleine Winkel gilt. Die Winkel sind auf einen Bereich begrenzt, in dem die Gesichtsinformation vollständig in der Frontalaufnahme enthalten ist. Aufgrund dieser Vorteile distanzbasierter Merkmale sollen Punktsignaturen an aussagekräftigen Punkten untersucht werden. Punktsignaturen gelten als „Fingerabdruck“ eines Gesichts, da das berechnete Profil zwischen den einzelnen Individuen charakteristisch variiert. Zu untersuchen ist, ob dieses Schema für die Unterscheidung der BP anpassbar ist. Gesichtsformen variieren zwischen Individuen und vor allem zwischen ethnischen Gruppen. Es ist allerdings anzunehmen, dass grobe Proportionen erhalten bleiben. Die Nase ist beispielsweise im Höhenprofil immer mindestens als lokales Extremum gekennzeichnet. Bestimmte Veränderungen bei Motorik-Übungen können über die Individuen als gleich angenommen werden, wie zum Beispiel die Veränderung des Höhenprofils über den Mundbereich durch das Herausstrecken der Zunge. Weiterhin sind Unterschiede zwischen den Übungsprofilen bei maximaler Übungsdurchführung zu erwarten. Ebenso ist anzunehmen, dass die Normalenvektoren an jedem Punkt der 3D-Gesichtsdaten eine gute Repräsentation der Gesichtsform geben. Es wird angenommen, dass bei unterschiedlichen Personen ähnliche Verformungen bei gleichen Übungen auftreten, was allerdings zu untersuchen bleibt. Als Beispiel ist zu nennen, dass die Wangen beim Aufblasen bei jedem Individuum gewölbt sein werden. Aus den genannten Gründen werden als Merkmale für die Klassifikation in die BP zunächst Punktsignaturen an bestimmten Punkten des Gesichts und die Normalen an allen Gesichtspunkten extrahiert.

4.5.1 Normalenvektoren

Der Normalenvektor als lokales Feature soll an jedem einzelnen Punkt bestimmt werden, der zum Gesichtsbereich zugeordnet wird. Ein Normalevektor kann allerdings nur mittels seiner Umgebung berechnet werden. Eine einfache Methode ist die Berechnung der Normalen an den Knotenpunkten innerhalb eines Meshs¹⁴, beispielsweise eines Meshs aus Dreieckelementen. In diesem Fall werden die Normalenvektoren an den Vertices¹⁵ aus dem Mittelwert der Normalenvektoren der drei benachbarten Flächen berechnet. Innerhalb dieser Arbeit sollen Daten betrachtet werden, die als Tiefenbild vorliegen und außer zu Darstellungszwecken wird kein Mesh genutzt. Für jeden einzelnen Punkt der Tiefenbildmatrix muss daher eine Umgebung geschaffen werden, in der ein Normalenvektor für den betrachteten Punkt geschätzt werden kann, wie es in [RUSU und COUSINS, 2012] durchgeführt wird. Zur Umgebungsdefinition werden die Datenpunkte der k -Nachbarschaft hinzugezogen. Die Variable k gibt die Anzahl der Punkte an, die zur Nachbarschaft gezählt werden. Die Nachbarauswahl wird allein an der Anordnung der Pixel in der Tiefenbildmatrix und durchgeführt. Die Tiefenwerte sind für die Nachbarschaftsauswahl irrelevant. Legt man ein Koordinatensystem so in die Tiefendaten, dass die XY-Ebene der Bildmatrix entspricht, sind die Nachbarn in der X- und Y-Richtung äquidistant. In der Z-Richtung, der Tiefe, variiert der Abstand. In diesem Satz aus Datenpunkten soll eine Ebene gebildet werden, deren Flächennormale eine Approximation des gesuchten Normalenvektors darstellt. Für die Ebenenregression stehen $k + 1 > 3$ Punkte zur Verfügung, da die kleinste nutzbare Nachbarschaft in der Bildmatrix die 4er-Nachbarschaft ist. Aus diesem Grund gibt es keine eindeutige Ebenengleichung. In dieser Arbeit soll die Berechnung der Regressionsebene nach der Least-Mean-Squares (LMS)-Methode¹⁶, bezogen auf die Distanzen der einzelnen Nachbarschaftsdatenpunkte zur Ebene, erfolgen. Für die Regressionsebene werden die Koeffizienten der Koordinatengleichung berechnet, aus denen der Normalenvektor in

¹⁴Englisch für Masche, wird i.A. verwendet, um ein Netz aus Knotenpunkten und Kanten zu beschreiben.

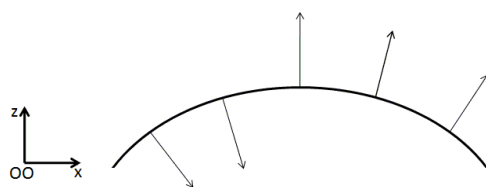
¹⁵Knoten

¹⁶Methode der kleinsten mittleren Quadrate

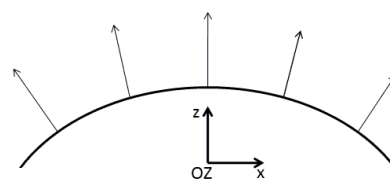
kartesischen Koordinaten abgelesen werden kann (siehe Gleichung 4.1).

$$0 = b_1x + b_2y + b_3z + b_4 \quad \text{mit} \quad \mathbf{n} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \quad (4.1)$$

Zu beachten ist, dass der gewählte Normalenvektor nur eine Lösungsmöglichkeit aus vielen möglichen linear abhängigen Vektoren ist. Zwei Richtungen des Normalenvektors sind möglich. Zum einen die Richtung des positiven Normaleneinheitsvektors \mathbf{e}_n zum anderen die Richtung des negativen Normaleneinheitsvektors $-\mathbf{e}_n$. Das globale Koordinatensystem liegt aus praktischen Gründen in der linken oberen Ecke der Bildmatrix, wobei die Y-Richtung nach unten zeigt. Damit entsprechen die Spalten- und Zeilenzahlen den X- und Y-Werten. Bei einer Berechnung der Normalen an den jeweiligen Bildpunkten wird die Richtung des Normalenvektors entsprechend des Bezugskoordinatensystems, welches im Folgenden „OO“ genannt wird, bestimmt. Anschließend wird der Normalenvektor in ein lokales Koordinatensystem am betrachteten Punkt verschoben. Erfolgt die Berechnung über das Kreuzprodukt zweier Vektoren, so sollte sie immer im mathematisch positiven Sinne geschehen, um eine einheitliche Richtung der Normalen zu erreichen. In Abbildung 4.3a ist das Ergebnis einer möglichen Berechnung bezüglich des beschriebenen Koordinatensystem „OO“ dargestellt.



(a) Normalen mit Bezugssystem OO



(b) Bezugssystem der Normalenberechnung im Objektzentrum (OZ)

Abbildung 4.3: Verschiedene Bezugssysteme für die Normalenberechnung; Die Richtungen der Normalen ist abhängig vom Bezugssystem

Für eine bessere Interpretationsmöglichkeit der Ergebnisse der Normalen ist die in Abbildung 4.3b dargestellte Weise jedoch besser geeignet. Dafür wird das Bezugs-

koordinatensystem „OO“ in ein Koordinatensystem im Objektzentrum, welches „OZ“ genannt wird, verschoben. Die Länge der Vektoren ist für eine Formbeschreibung des Gesichtsreliefs irrelevant, daher werden Einheitsnormalenvektoren gebildet. Nach der Normierung der kartesischen Koordinaten kann die Länge vernachlässigt werden, da sie bei allen Einheitsvektoren eins ist. Die Normalenvektoren werden durch den Polwinkel θ und den Azimutwinkel ϕ der sphärischen Koordinaten eindeutig beschrieben.

4.5.2 Punktsignaturen

Die Art der Punktsignaturberechnung soll nach dem Schema bestimmt werden, wie es in [CHUA und JARVIS, 1997] beschrieben und in [WANG et al., 2002], [CHUA et al., 2000] und [MPIPERIS et al., 2007] angewendet wurde. Die Berechnung der Punktsignaturen soll allerdings nur an bestimmten Referenzpunkten durchgeführt werden. Die Definition der Punktsignatur wird im Folgenden näher erläutert.

Um einen definierten Referenzpunkt \mathbf{p}^0 der Tiefendaten des Gesichts wird eine Kugel mit dem Radius r gelegt. Die Schnittkurve C zwischen den Gesichtsdaten und der entsprechenden Kugel wird berechnet. In Abbildung 4.4 sind die Gesichtsdaten grau und die Kugel für die Punktsignaturberechnung rot dargestellt. Die Schnittkurve C wird durch die Kante der Kugel am Relief dargestellt und zeigt deutlich, dass die Kurve im aufgezeigten Fall nicht in einer Ebene verläuft. Man kann erkennen, dass die Kurve aus Sicht des Normalenvektors am Mittelpunkt der Kugel keinen Kreis ergibt.

Da es sich bei den Gesichtsdaten nicht um eine kontinuierliche Kurve handelt bzw. nicht mit einem Mesh gerechnet wird, soll die Schnittkurve mit einem Radiusintervall berechnet werden, wobei sich eine diskrete Schnittkurve ergibt. Am Punkt \mathbf{p}^0 und seiner Umgebung wird eine Fläche eingepasst, um einen Normalenvektor \mathbf{n}_1 des Punktes \mathbf{p}^0 approximieren zu können. Dies soll nach dem gleichem Verfahren wie in Unterabschnitt 4.5.1 geschehen. Um den berechneten Einheitsnormalenvektor $\mathbf{e}_{\mathbf{n}_1}$ wird ein Vektor $\mathbf{e}_{\mathbf{n}_2}$, der senkrecht zu $\mathbf{e}_{\mathbf{n}_1}$ steht, mit diskretem Winkel $\Delta\alpha$ gedreht, um die Schnittkurve abzutasten. Zu beachten ist, dass, obwohl die Abtastung mit einem konstanten Winkel $\Delta\alpha$ erfolgt, die Schnittkurve nicht in äquidistanten Abständen ab-

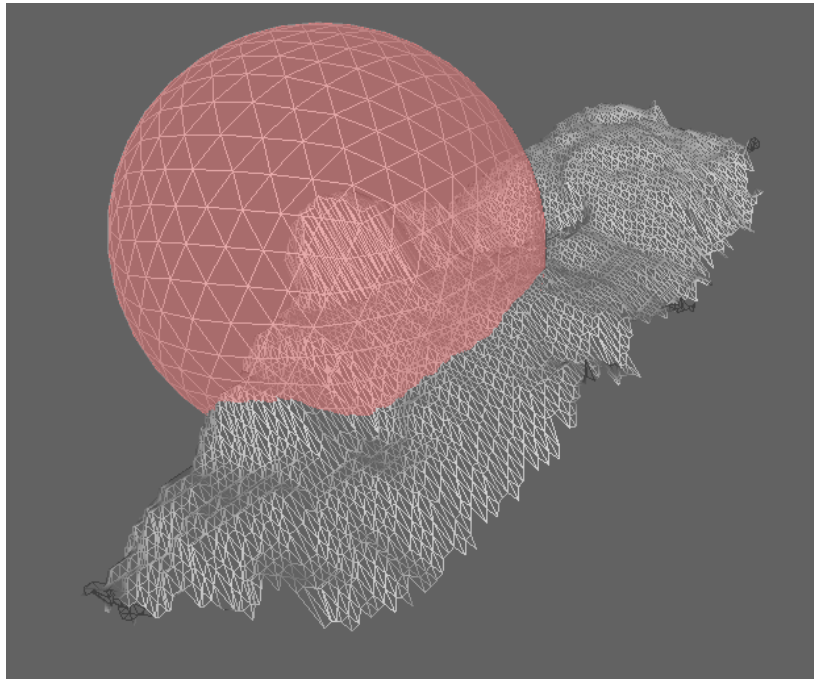


Abbildung 4.4: Schnittkurve C zwischen Gesichtsrelief und der Kugel um den Referenzpunkt \mathbf{p}^0 . Die Gesichtsdaten sind für eine bessere Darstellung als Mesh dargestellt.

getastet wird, da sich C nicht zwangsläufig in einer planaren Fläche befindet. Dieser Zusammenhang ist in der Darstellung in Abbildung 4.5a und b verdeutlicht.

Da die Punktsignaturen durch die Abtastung entlang des Winkels $\Delta\alpha$ periodische Funktionen ergeben, ist es für einen Profilvergleich von Bedeutung, den Start- und Endpunkt der Signatur eindeutig über alle Bilder gleich zu definieren. Durch eine Phasendifferenz bei zwei zu vergleichenden Signaturen werden ähnliche Kurven sehr unterschiedliche Werte aufweisen. Eine mögliche Phasendifferenz müsste daher für eine Vergleichbarkeit korrigiert werden oder ist von Beginn an zu vermeiden. In Abbildung 4.5b werden die Schnittkurve und die angedeutete Kugel in der Draufsicht auf den Normaleneinheitsvektor $\mathbf{e}_{\mathbf{n}_1}$ dargestellt. Der Vektor $\mathbf{e}_{\mathbf{n}_1}$ bildet mit dem Einheitsvektor $\mathbf{e}_{\mathbf{n}_2}$, der den Startpunkt der Abtastung angibt, ein neues Koordinatensystem. Der Vektor $\mathbf{e}_{\mathbf{n}_2}$ liegt in der Ebene, welche in die Schnittkurve C eingepasst und in \mathbf{p}^0 verschoben wird. Zwischen der berechneten Ebene, auf welcher der Referenzpunkt \mathbf{p}^0 liegt, und den ausgewählten Schnittpunkten sollen die Distanzwerte berechnet und über den Winkel aufgetragen werden. Um eine Aussage über die Lage der Punktsi-

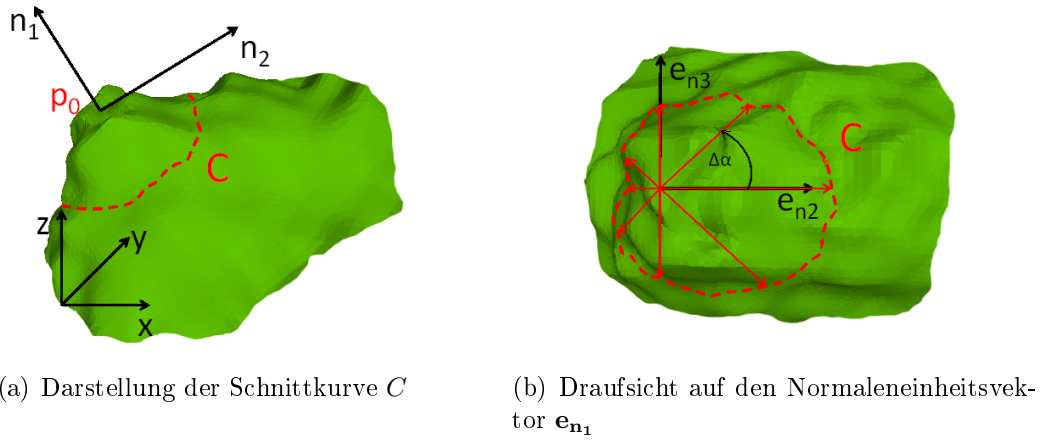


Abbildung 4.5: Abtastung der Schnittkurve C mit $\Delta\alpha$ um den Normaleneinheitsvektor \mathbf{e}_{n_1} mit dem Startvektor \mathbf{n}_2

gnatur im Bezug auf die Referenzebene geben zu können, werden allen Punkten entsprechende Vorzeichen zugeordnet. Definitionsgemäß liegen alle Punkte mit positiven Distanzen „über“ der Referenzebene, also näher an der Kamera, mit der die Daten aufgenommen wurden, als der Referenzpunkt. Alle Punkte mit negativen Distanzen liegen „unter“ der Referenzebene und weiter von der Kamera entfernt. Um keine irrelevanten Daten zu erheben, sollte die Punktsignatur so berechnet werden, dass alle Schnittpunkte im Gesichtsbereich liegen. Ebenso könnten, um Randeffekte zu vermeiden, nur bestimmte Teile bzw. Kugelsegmente zur Schnittpunktberechnung genutzt werden.

4.6 Konzept der Merkmalsselektion

Wird der Datensatz nach dem in Abschnitt 4.4 beschriebenen Verfahren aufgenommen, ergibt sich mindestens eine Anzahl von 1463 Bildern. Normalenvektoren und Punktsignaturen werden an definierten Stellen im Gesicht bzw. an vorher bestimmten Bereichen berechnet. Dabei bestehen Punktsignaturen an einem Referenzpunkt aus mehreren diskreten Funktionswerten und sollen zusätzlich mit unterschiedlichen Parametern berechnet werden. Genauere Beschreibungen der Berechnungen und der Parameterv-

riationen erfolgen im Abschnitt 5.2 und Abschnitt 5.3. Durch beide Featurearten wird der Merkmalsraum hochdimensional und kann größer als 1000 Dimensionen enthalten. Wie bereits in Unterabschnitt 2.3.1 beschrieben, ist somit eine Merkmalsselektion durchzuführen, um eine bessere Generalisierung der Klassifikatoren zu ermöglichen. Ebenso verläuft die Klassifikation nach einer Selektion weniger rechenaufwändig und die Wahrscheinlichkeit des Overfittings sinkt.

Um die Dimension des Merkmalsraums zu verringern, wird ein Filteransatz gewählt. Die Merkmale werden durch ein entsprechendes Relevanzkriterium in eine Rangfolge gebracht und nur signifikante Merkmale ausgewählt. Die MI, welche in Abschnitt 2.3.1 erläutert wurde, ist ein Maß für die Information, die eine Zufallsvariable über eine zweite besitzt, und somit als Relevanzkriterium geeignet. Sie bildet im Gegensatz zum linearen Korrelationskoeffizienten nicht nur lineare Zusammenhänge ab und ist daher zu bevorzugen [BATTITI, 1994]. Die MI soll zwischen allen Merkmalen und der entsprechenden Zielfunktion gebildet werden. Je höher die MI ist, umso mehr Informationen trägt das Merkmal über die Klassenverteilung. Bei einem niedrigen MI-Wert besteht kaum oder gar keine statistische Abhängigkeit der beiden Größen. Mit der MI als Relevanzkriterium kann ein Ranking als einfache Möglichkeit der Filter umgesetzt werden.

Im Zusammenhang der Merkmalsselektion wird innerhalb dieser Arbeit zwischen Featurearten, Featureblöcken und Einzelfeatures differenziert. Eine Featureart umfasst alle Merkmale, die auf die gleiche Weise berechnet werden. Einzelne Werte des Featurevektors werden Merkmal oder Einzelmerkmal genannt. Mehrere Einzelfeatures werden zu Featureblöcken zusammengefasst, welche beispielsweise alle Werte einer Punktsignatur an einem Referenzpunkt mit einem Parametersatz umfassen. Mittels Merkmalsselektion sollen verschiedene Featureblöcke, welche sich beispielsweise durch verschiedene Parameter unterscheiden, auf ihre Relevanz für die Klassifikation untersucht werden. Der Informationswert von Kombinationen verschiedener Features ist in den MI-Werten nicht enthalten. Ein Merkmal muss eine statistische Abhängigkeit zu den Klassenlabel besitzen, um als relevant eingestuft zu werden [ZHANG und HANCOCK, 2011]. Durch das Ranking der berechneten MI-Werte allein kann keine Aussage über den

Informationswert eines Featureblocks getroffen werden. Werden die besten Merkmale im Ranking gewählt, bezieht sich das allein auf Einzelfeatures. In dieser Arbeit soll ein MI-Grenzwert festgelegt werden. Liegt ein gewisser Prozentsatz an Einzelfeatures aus einem Block über dem MI-Grenzwert, wird der komplette Featureblock für den weiteren Verlauf genutzt. Der MI-Mittelwert eines Blocks wird an dieser Stelle als Grenzkriterium ausgeschlossen, da er zu stark von vereinzelt Ausreißern in dem kleinen MI-Wertebereich beeinflusst wird. Um einen Grenzwert festzulegen, kann ein Vergleichswert zwischen gleichverteiltem bzw. normalverteiltem Rauschen und der Labelfunktion berechnet werden. Alle Merkmale, die weniger Informationen über die Klassen tragen als Rauschen, sind für das Klassifikationsproblem zu vernachlässigen. Die Hälfte der Merkmale aus einem Block sollen mindestens doppelt so große Informationswerte aufweisen wie das Maximum aus den MI-Werten des Rauschens. Für eine Reproduzierbarkeit der Grenzwertberechnung sollten nach dem zentralen Grenzwertsatz mindestens 30 normalverteilte und 30 gleichverteilte Rauschkanäle erzeugt werden. Folglich kann eine Normalverteilung der MI-Werte angenommen werden. Durch die Berechnung der Mittelwerte über die Kanäle sind im Vergleich zu den Einzelinformationswerten der Rauschkanäle zwei stabilere Werte zu erwarten. Das Maximum aus beiden Mittelwerten der verschiedenen Rauscharten soll für eine Grenzwertfestlegung genutzt werden.

Durch das vorgestellte Konzept wird eine Merkmalsteilmenge mit kleinerer Dimension als die Originalmenge gefunden. Dabei sind keine Einbußen in der Leistung der Klassifikation zu erwarten, da nur irrelevante und redundante Merkmale entfernt werden sollen. Zu beachten ist jedoch, dass durch das Verfahren keinesfalls eine Aussage über die optimale Teilmenge getroffen wird. Die berechnete Teilmenge verbessert die Klassifikationsleistung, gibt aber keine Aussage über die beste Lösungsmöglichkeit. Die optimale Teilmenge des Originalmerkmalsraums würde die besten Klassifikationsergebnisse aus allen möglichen Teilmengen erzielen.

4.7 Konzept der Klassifikation

Die Klassifikation soll mit den in Abschnitt 4.5 ausgewählten Merkmalen durchgeführt werden. Dazu werden sie in jedem Bild des Datensatzes extrahiert und gemeinsam in den Wertebereich $[-1, 1]$ skaliert. Die Klassen werden als BP definiert und entsprechen innerhalb dieser Arbeit den eigentlichen Übungen, da nur Halteübungen und statische Bilder betrachtet werden. Um eine bessere Interpretationsmöglichkeit über die Leistungsfähigkeit der Merkmale zu schaffen, sollen drei verschiedene Klassifikatoren genutzt werden. Zu beachten ist jedoch, dass die Evaluierung der Merkmale mit der Leistungsfähigkeit der Klassifikatoren zusammenhängt. Bei kleiner Erfolgsrate der Klassifikation kann somit nicht entschieden werden, ob die Merkmale für diese Anwendung nicht ausreichend signifikant sind oder ob der Klassifikator bzw. dessen Parameter ungünstig gewählt wurden. Mit der Wahl des kNN- und des Naive Bayes Klassifikator sowie von SVM werden mehrere Arten an Klassifikatoren abgedeckt.

Für die Klassifikation mittels SVMs soll die *Library for Support Vector Machines* (LibSVM)¹⁷ genutzt werden, welche in [CHANG und LIN, 2011] beschrieben wird und unter MATLAB eingesetzt werden kann. Bei SVM handelt es sich um einen binären Klassifikator, weswegen für Multiklassenprobleme mehrere SVMs in Kombination verwendet werden. Für die Nutzung der SVM auf Multiklassenprobleme gibt es zwei bekannte Methoden: Zum einen die Methode one-against-rest¹⁸, mit der für das bestehende Problem mit x Klassen auch x verschiedene SVM berechnet werden. Zum anderen gibt es die one-against-one¹⁹-Methode, bei der $\frac{x(x-1)}{2}$ verschiedene SVMs kombiniert werden, um eine Trennbarkeit aller Klassen zu schaffen, was bei 19 verschiedenen Klassen 171 kombinierten SVMs entspricht [BRAMER, 2006]. LibSVM nutzt die one-against-one-Methode. Die Parameterwahl für die SVMs soll durch die von LibSVM bereitgestellte Funktion zur Parametersuche erfolgen.

Für den kNN-Klassifikator soll die Ähnlichkeitsberechnung mittels euklidischer Di-

¹⁷ <http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/>, letzter Abruf am 23.03.2012

¹⁸englisch, für Einer gegen die Restlichen.

¹⁹englisch, für Einer gegen Einen.

stanz erfolgen. Der Parameter der Nachbarschaftsanzahl wird auf $k = 5$ gesetzt. Diese haben während der Erstellung der vorliegenden Masterarbeit sehr gute Ergebnisse geliefert. Die Validierung der Modelle soll nach zwei verschiedenen Methoden durchgeführt werden, welche in [KOHAVI, 1995] erläutert sind. Zum einen soll die Hold-Out Methode genutzt werden, bei welcher der Datensatz zufällig in zwei Teile, in die Trainings- und die Testdaten, geteilt wird. Bei ungünstiger Teilung kann es vorkommen, dass in den Trainingsdaten eine Klasse gar nicht oder nur selten vertreten ist, was zu Klassifikationsfehlern führt. Daher soll die Methode der Stratifikation genutzt werden, sodass in den jeweiligen Teilmengen möglichst die gleiche Klassenverteilung wie im Originalset vorliegt. Damit das Ergebnis auch bei der Nutzung des statistischen Bayes-Klassifikators reproduzierbar ist, sollen die Klassifikationsergebnisse über mindestens 30 Durchgänge gemittelt werden. Zusätzlich soll eine n -fache Kreuzvalidierung durchgeführt werden. Dabei wird der Datensatz ebenfalls in n gleich große Teile getrennt. Die Trainingsphase erfolgt dabei mit $n - 1$ Teilen, wobei an der nicht genutzten Datenmenge getestet wird. Dieser Vorgang wird für alle n Teile wiederholt. Anschließend wird über alle Durchgänge gemittelt.

Für die n -fache Kreuzvalidierung soll zusätzlich eine weitere Bedingung gelten. Der Datensatz enthält für jede Person und Übung mindestens sieben Bilder. Diese sieben Darstellungen der gleichen Übung wurden allerdings in derselben Messreihe aufgenommen. Teilweise wird die Übung zwischen den Bildern gehalten und nicht erneut aus der Ruhelage ausgeführt, dadurch können innerhalb dieser Bilder sehr starke Ähnlichkeiten auftreten. Damit ein Auswendiglernen des Klassifikators verhindert wird, soll die Datenmenge bei der Kreuzvalidierung personenweise aufgeteilt werden. Somit wird in jedem Durchgang an Bildern einer Person getestet, die in den Trainingsdaten nicht vertreten ist. Daraus folgt eine elf-fache Kreuzvalidierung, bei denen in allen elf Teilen die Stratifikation näherungsweise erfüllt ist und die Aufteilung nicht zufällig geschieht.

Die Evaluierung der Klassifikationen soll möglichst übersichtlich gestaltet werden. Daher ist die Erstellung eines Überblick über die Erfolgs- und Fehlerrate der Klassifikatoren bezüglich aller Übungen vorteilhaft. Das Klassifikationsergebnis kann mittels Konfusionsmatrizen dargestellt werden.

Kapitel 5

Umsetzung

Die Umsetzung des Klassifikationsproblems erfolgt in MATLAB und wird durch die Erweiterung eines im Fachgebiet für Kognitive Robotik und Neuroinformatik der TU Ilmenau entwickelten Frameworks durchgeführt. Zunächst werden die einzelnen Komponenten des Frameworks erläutert. Den Hauptteil des Kapitels stellt die Beschreibung der Implementierung der ausgewählten Merkmale dar. Die Vorgänge zur Berechnung der Merkmale werden im Abschnitt 5.2 für Normalenvektoren und im Abschnitt 5.3 für Punktsignaturen beschrieben. Die Durchführungsschritte der anschließenden Merkmalsselektion werden erläutert und mit der Darstellung der Ergebnisse der Klassifikation beendet.

5.1 Framework zur Klassifikation der Tiefenbilder

Das Framework ist in drei Abschnitte geteilt. Zu Beginn werden Vorverarbeitungsschritte vorgenommen, um die entsprechende Klassifikation vorzubereiten. Es folgt die Extraktion der gewünschten Merkmale aus den Tiefenbildern. Abschließend wird anhand der erstellten Merkmalsvektoren die Trainings- und Testphase des Klassifikators durchgeführt, um die Klassifikation evaluieren zu können. Alle benötigten Teilschritte sind als Funktionen in MATLAB umgesetzt worden und in das Framework eingebunden.

Für einen Durchlauf der Klassifikation werden alle Tiefenbilder des Datensatzes ein-

gelesen. Es besteht die Möglichkeit, in vorbereitenden Schritten Datensatzteilmengen zu erstellen. Das Klassifikationsproblem kann neu definiert werden, indem nur Bilder bestimmter Klassen berücksichtigt bzw. die Anzahl der Klassen variiert werden kann. Dabei können zusätzlich mehrere Übungen zu einer Klasse zusammengefügt werden. Für die Hauptbetrachtung in dieser Arbeit wird diese Möglichkeit allerdings nicht genutzt. Werden Merkmale betrachtet, die nur Teilbereiche des Gesichts abbilden, können Übungen, die im betrachteten Bereich gleich oder ähnlich sind, zusammengefasst werden, um die Klassifikationsleistung nicht zu verfälschen. Betrachtet man beispielsweise ein Merkmal, welches örtlich auf den Mundbereich beschränkt ist, können alle Augen- und Stirnübungen zu einer Klasse „Neutraler Mundbereich“ zusammengefasst werden. Das Verbinden mehrerer Übungen in einer Klasse wird unter Berücksichtigung der Datensatzbalance durchgeführt. Ein stark unbalancierter Datensatz kann bei ungünstiger Auswahl der Lernstrategie bei der Klassifikation zur Begünstigung der stark repräsentierten Klasse führen [GROSS, 2011]. Der Datensatz wird in einem solchen Fall durch die Methode des Undersampling¹ ausbalanciert. Dabei werden aus der überrepräsentierten Klasse zufällig Daten aus der Datenmenge entfernt. Es wird angenommen, dass dabei eventuell repräsentative Daten für eine untersuchte Klasse unbeachtet bleiben [GROSS, 2011]. Nach der Erstellung der Datenmenge werden in allen Tiefenbildern die gemessenen Disparitätswerte in Meterangaben umgerechnet.

Bei der darauffolgenden Merkmalsextraktion werden für alle Bilder der ausgewählten Datensatzteilmenge einige der implementierten Features ausgewählt. Die Berechnung erfolgt nach den in Abschnitt 5.2 und Abschnitt 5.3 beschriebenen Regeln. Bei allen Featurefunktionen werden dabei für jedes Bild Featureblöcke ausgegeben, die zu einem Featurevektor zusammengesetzt werden. Das Ergebnis über alle Bilder ist die Merkmalsmatrix, deren Spalten die Merkmale und Zeilen die Beobachtungen darstellen. In jeder Zeile der Merkmalsmatrix befindet sich somit ein Featurevektor eines bestimmten Tiefenbildes. Bei einer $m \times n$ Matrix hat der berechnete Merkmalsraum für die Klassifikation n Dimensionen und es werden m verschiedene Bilder betrachtet, wobei $m, n \in \mathbb{N}$ sind.

¹englisch, für Unterabtastung

Es stehen zwei Möglichkeiten der Validierung und drei verschiedene Klassifikatoren zur Verfügung, welche in Abschnitt 4.7 beschrieben wurden. Es können nacheinander beide Validierungen in einem Durchlauf der Funktionen durchgeführt werden. Als Ergebnisse werden die über die Durchläufe gemittelten Erfolgsraten und die gemittelte Konfusionsmatrix in Prozentangaben ausgegeben. Zu beachten ist, dass für die Ausgabe gerundet wird. Kleine Unstimmigkeiten beider Größen zueinander sind daher auf die Rundungsfehler zurückzuführen.

5.2 Normalenvektoren

Bei Normalenvektoren zur Formbeschreibung des Gesichts handelt es sich um lokale Merkmale, da sie an jedem der definierten Punkte im Tiefenbild einzeln berechnet werden. In diesem Abschnitt wird zunächst eine allgemeine Beschreibung der Berechnungsschritte an einem Punkt erfolgen und anschließend die genaue Definition der verschiedenen Berechnungsorte getroffen.

Die Berechnung von Normalenvektoren wird nur im Gesichtsbereich vorgenommen. Eine Definition über die Gesichtslage innerhalb der Tiefenbilder ist über die Labelpunkte gegeben, die vor der Bearbeitung manuell in die RGB-Bilder gesetzt wurden. Mit Hilfe dieser Label kann eine Gesichtsmaske berechnet werden, die den Gesichtsbereich vom Hintergrund des Bildes segmentiert, sodass in folgenden Schritten nur Gesichtsdaten berücksichtigt werden.

5.2.1 Allgemeine Umsetzung

Der Datenpunkt $\mathbf{p}^0 = (p_1^0, p_2^0, p_3^0)^T$ wird innerhalb seiner Nachbarschaft betrachtet. Für die Nachbarschaftsdefinition gibt es unterschiedliche Varianten, wobei sie in dieser Arbeit fest definiert wird. Der Einfluss der Variation der Nachbarschaftsparameter auf die gegebene Problemstellung bleibt offen. Wie bereits im Unterabschnitt 4.5.1 erklärt, wird die Nachbarschaft nicht mit einer metrischen Distanz in der dreidimensionalen

Umgebung berechnet, sondern in der zweidimensionalen Betrachtung der XY-Ebene des Tiefenbildes. Sie wird durch die Annahme vereinfacht, dass die Nachbarn in der Tiefenbildmatrix ebenfalls die nächsten Nachbarn in der dreidimensionalen Betrachtung sind. Für die Berechnung der Normalenvektoren wird die Achter-Nachbarschaft gewählt, welche auch Moore-Nachbarschaft genannt wird. Die Datenpunkte innerhalb der Nachbarschaft werden durch entsprechende Ortsvektoren beschrieben und im weiteren Verlauf als Nachbarschaftspunktmenge $NB = \{\mathbf{p}^0, \mathbf{p}^1, \dots, \mathbf{p}^k\}$ mit $k = 8$, welche den betrachteten Punkt \mathbf{p}^0 einschließt, bezeichnet. In die NB wird eine Ebene eingepasst, deren Flächennormale den Normalenvektor am Punkt (p_1^0, p_2^0, p_3^0) approximiert. Da die NB mehr als drei Punkte umfasst, ist die Suche nach der Ebene ein überbestimmtes Problem und hat unendlich viele mögliche Lösungen. Als Lösungsansatz wird eine Regressionsebene in der NB nach der Methode der kleinsten Quadrate gesucht. Die Abstände der Datenpunkte zur Ebene sollen in der Summe minimal sein. Damit höhere Distanzen bei der Minimumsuche stärker bestraft werden als kleinere, wird die quadratische Summe der Abstände betrachtet.

Eine Ebene ist in ihrer Normalform durch einen Punkt, dem sogenannten Aufpunkt \mathbf{a} auf der Ebene, sowie ihrem Normalenvektor \mathbf{n} eindeutig definiert (siehe Gleichung 5.1).

$$\mathbf{n} (\mathbf{x} - \mathbf{a}) = 0 \quad \text{mit} \quad \mathbf{a} = \frac{1}{k} \sum_{i=0}^k \mathbf{p}^i \quad (5.1)$$

Der Vektor \mathbf{a} beschreibt dabei den Abstand der Ebene zum Ursprung und kann durch den Mittelwert über alle Punkte in der NB bestimmt werden. Alle Punkte der NB werden daher bezüglich des Aufpunktes angegeben. Durch Berechnung des Einheitsnormalenvektors \mathbf{e}_n und einer Umformung ergibt sich die Hessesche Normalform Gleichung 5.2.

$$\mathbf{x} \cdot \mathbf{e}_n - g = 0 \quad \text{mit} \quad \mathbf{a} \cdot \mathbf{e}_n = g \quad (5.2)$$

Die Hessesche Normalform ist für alle Punkte auf der Ebene erfüllt, wobei g den Abstand der Ebene zum Ursprung beschreibt und größer oder gleich Null sein muss. Ebenso kann die Hessesche Normalform durch Einsetzen der Datenpunkte als Ebenenpunkte \mathbf{x} für die Berechnung von orthogonalen Abständen d zwischen Punkten und der Ebene genutzt werden. In der Gleichung 5.3 ist die Abstandsgleichung in Normalform und kartesischer Form angegeben. Die kartesische Form gilt hier nur für \mathbb{R}^3 und es wird mit dem Normalenvektor \mathbf{n} gerechnet.

$$d = \mathbf{p} \cdot \mathbf{e}_n - g \quad \text{und} \quad d = \frac{p_1 n_1 + p_2 n_2 + p_3 n_3 - p_0}{\sqrt{n_1^2 + n_2^2 + n_3^2}} \quad (5.3)$$

Haben d und g bei der Abstandsberechnung die gleichen Vorzeichen, befindet sich der untersuchte Punkt auf der anderen Seite der Ebene als der Ursprung. Bei unterschiedlichem Vorzeichen liegen der Ursprung und der Punkt im gleichen Halbraum.

Nach der Methode der kleinsten Quadrate wird das Minimum der Summe der quadrierten Abstände der Punkte in der NB zu der gesuchten Ebene gefunden. Der Aufpunkt der Ebene kann durch die Mittelwertbildung über alle Punkte aus der NB berechnet werden. Durch Einsetzen aller Punkte aus der NB in die Hessesche Normalform ergibt sich somit der zu minimierende Ausdruck in Gleichung 5.4.

$$\min \left\{ \sum_{i=0}^k ((\mathbf{p}_i - \mathbf{a}) \cdot \mathbf{e}_n)^2 \right\} \quad (5.4)$$

Fasst man die Datenpunkte nach Abzug des Mittelwertes für alle Punkte der NB zu einer $(k+1) \times 3$ Matrix A zusammen, ergibt sich ein linear überbestimmtes Gleichungssystem, welche immer durch die Singulärwertzerlegung (SVD)² lösbar ist. Die SVD löst überbestimmte Gleichungssysteme nach der Methode der Ausgleichsrechnung. Bei bestimmten Gleichungssystemen ist die optimale Lösung gleich der exakten. Die SVD hängt stark mit der (PCA)³ zusammen. Die SVD erweitert diese auf nicht

²im englisch als „Singular Value Decomposition (SVD)“ bezeichnet

³englisch: Principal Component Analysis (PCA)

quadratische Matritzen, wobei für alle symmetrischen Matritzen die Singulärwerte dem Quadrat der Eigenwerte entsprechen. Dementsprechend kann geschrieben werden, dass eine PCA einer quadratischen Matrix, welche nur nicht negative Eigenwerte hat, ebenfalls eine SVD ist [GRAMLICH, 2004].

Durch diese Herangehensweise ist eine weitere Interpretationsmöglichkeit des Problems möglich, wie es auch in [RUSU und COUSINS, 2012] formuliert ist. Die Regressionsebene soll also, um die Summe der Distanzen so gering wie möglich zu halten, entlang der größten Varianzen der Punktmenge aufgespannt werden. Dementsprechend handelt es sich bei der Normalen der Fläche um den Eigenvektor des geringsten Eigenwertes, welcher zu den aufspannenden Vektoren orthogonal ist.

Die SVD ist in Gleichung 5.5 definiert.

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T \quad (5.5)$$

$\mathbf{A} \in \mathbb{R}^{m \times n}$ ist eine nicht quadratische Matrix und $\mathbf{U} \in \mathbb{R}^{m \times m}$ und $\mathbf{V} \in \mathbb{R}^{n \times n}$ sind zueinander orthogonale Matritzen. $\mathbf{S} \in \mathbb{R}^{m \times n}$ mit dem Rang n ist eine Diagonalmatrix, welche n Singulärwerte $\sigma_1, \sigma_2, \dots, \sigma_n$ mit absteigender Größe auf der Hauptdiagonalen enthält [OVE, 2011].

Wie in [KOVESI] und [OVE, 2011] beschrieben, ergibt sich die Berechnung der Regressionsebene nach SVD durch den Normalenvektor der Ebene als dritter Spaltenvektor aus \mathbf{U} . Die beiden ersten Spaltenvektoren repräsentieren die Komponenten entlang der größten Varianz und spannen die Regressionsebene auf. Ebenso ist die minimale quadrierte Summe der Distanzen zwischen den Punkten und der besten Ebene gleich dem Quadrat des kleinsten Singulärwertes σ_3^2 . In Abbildung 5.1 ist ein Beispiel eines Ergebnisses der SVD dargestellt. Die Punktwolke bildet sich aus dem betrachteten Punkt \mathbf{p}^0 und seinen acht Nachbarn, welche rot eingezeichnet sind. Die Regressionsebene ist grau dargestellt und zeigt die Ebene, bei der die Summe der Distanzen (blau) im Quadrat minimal ist. Der resultierende Normalenvektor wird zu Darstellungszwecken an den untersuchten Punkt verschoben und grün dargestellt.

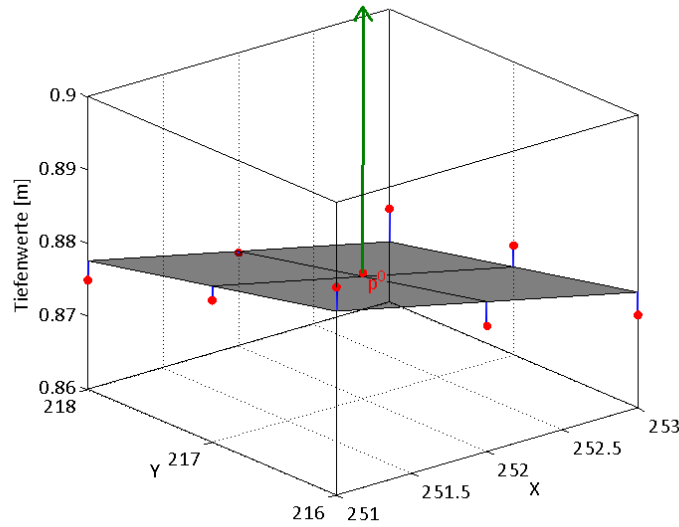


Abbildung 5.1: Der Punkt \mathbf{p}^0 und seine acht Nachbarn (rot). Die orthogonalen Distanzen zur Ebene (blau) sollen in der quadratischen Summe minimal sein. Durch die SVD ergeben sich die Regressionsebene (grau) und deren Flächennormalenvektor (grün).

Die Verschiebung der Normalen ist nur für Darstellungszwecke nützlich. Zwischen dem Bezugskoordinatensystem in der Bildecke und dem lokalen Koordinatensystemen jeweils an dem betrachteten Punkt liegt keine Rotation vor. Es handelt sich dabei um eine reine Translationsverschiebung. Da die Normalen nur durch den Polar- und Azimutwinkel der sphärischen Koordinaten angegeben werden, ist keine Verschiebung in den Punkt (p_1^0, p_2^0, p_3^0) notwendig. Die Winkel bleiben bei Translationsverschiebungen gleich. Um die Darstellungsweise auf die in Unterabschnitt 4.5.1 beschriebene Art zu erhalten, müsste das Bezugskoodinatensystem wie erläutert in das Objektzentrum verschoben werden. Hier erfolgt jedoch nur eine virtuelle Verschiebung. Dazu werden aus allen Bilddaten des Gesichts die Koordinaten des Objektzentrums bestimmt. Dafür wird der Mittelpunkt der Gesichtsdarstellung in der XY-Ebene und der Z-Wert der Gesichtsdaten, welcher am weitesten von der Kamera entfernt liegt, bestimmt. Der Normalenvektor \mathbf{n}_1 in Richtung $\mathbf{e}_{\mathbf{n}1}$ und der korrespondierende Vektor \mathbf{n}_2 in Richtung $-\mathbf{e}_{\mathbf{n}1}$ werden vorübergehend in kartesischen Koordinaten in das lokale Koordinatensystem verschoben. In Abbildung 5.2 sind beide gestrichelt in grün dargestellt. Zwischen

den zwei Punkten, die durch diese beiden Vektoren als Ortsvektoren beschrieben werden, und dem Objektzentrum OZ wird die euklidische Distanz berechnet (siehe Gleichung 5.6).

$$\|n_1 - OZ\|_2 \quad \text{und} \quad \|n_2 - OZ\|_2 \quad (5.6)$$

Die größere Distanz bestimmt den zugehörigen Vektor, der für eine bessere Darstellung gewählt wird. In Abbildung 5.2 liegt der Punkt, welcher durch n_2 beschrieben wird, weiter vom Objektzentrum entfernt und zeigt in die zu wählende Richtung für den Normalenvektor.

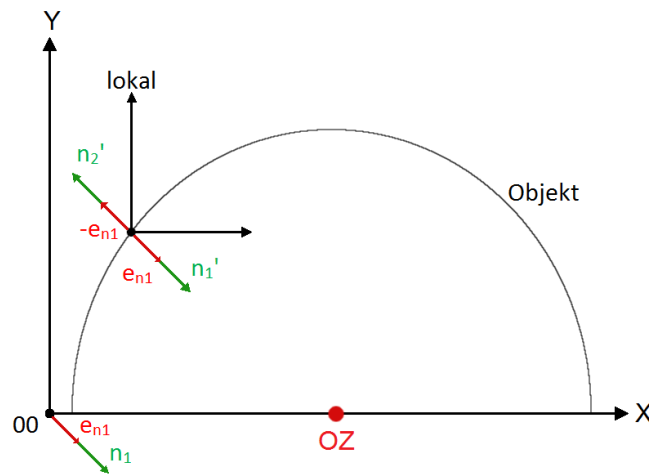


Abbildung 5.2: Die Berechnung der Normalenvektoren erfolgt bezüglich des Koordinatensystems "OO". Die Richtung entlang des Normalenvektors wird mit Hilfe des Objektzentrums bestimmt.

Nach Berechnung des Normaleneinheitsvektors kann der Vektor durch den Polar- und Azimutwinkel der sphärischen Koordinaten eindeutig beschrieben werden.

Durch die Berechnung des Normalenvektors mittels Regressionsebene innerhalb einer Nachbarschaft entstehen Randprobleme. Sobald sich Datenpunkte der NB außerhalb des Gesichts befinden, wird die Normale der Regressionsebene den Normalenvektor nicht mehr gut approximieren, da die Werte im Hintergrund sehr viel höhere Tiefen-

werte besitzen. Tiefenwerte, die aufgrund von Verdeckungen oder Schatten bei der Aufnahme nicht ermittelt werden können, werden mit dem größten Disparitätswert von 2047 belegt (siehe Unterabschnitt 2.1.3) und bei der Umwandlung in Meter nicht berücksichtigt. Im Randbereich kommt es beispielsweise durch Haare zu solchen Effekten. Daher sollten die Werte der NB, welche nicht im Gesichtsbereich liegen, aus der NB entfernt werden. Alle NBs für Punkte des Randbereichs beinhalten somit weniger Elemente als NBs für Punkte im Zentrum des Gesichts.

5.2.2 Verschiedene Umsetzungen der Normalenvektoren als Merkmale

Durch die Variationen der Gesichtsformen und -größen über die Probanden und die intraindividuellen Abstände zur Kamera werden die Gesichtsdaten nicht durch die gleiche Anzahl an Pixeln repräsentiert. Es muss jedoch die Voraussetzung gelten, dass die Merkmale in verschiedenen Bildern an korrespondierenden Stellen berechnet werden. Die Featurevektoren müssen für jedes Bild gleiche Dimensionen besitzen, um eine Vergleichbarkeit zu ermöglichen. Eine alleinige pixelbasierte Berechnung ist daher für diese Zwecke nicht ausreichend.

Die Berechnung setzt eine vorherige Standardisierung der Gesichtsdaten oder genaue Ortsdefinitionen, welche die Stellen der Berechnung eindeutig beschreiben, voraus. Erfolgt die Berechnung nur an definierten Punkten, ist für jedes Gesicht der Ort und die Anzahl der Vektoren gleich und der Featurevektor gleich dimensional. Eine weitere Möglichkeit ist die Berechnung an allen Punkten des Gesichts, wobei die ausgegebenen Merkmale die Verteilungsbeschreibung der Normalenwerte, beispielsweise über ein Histogramm, angeben. Für die Nutzung eines Histogramms zur Beschreibung der Häufigkeitsverteilungen ist wiederum eine Diskretisierung der Winkelwerte oder eine Zuordnung der Winkel zu bestimmten Kategorien, welche die Histogrammbalken bilden, notwendig. Die Berechnung über bestimmte Bereiche mit einer anschließenden Standardisierung durch Abtastung oder Interpolation der berechneten Werte auf eine konstante Anzahl an Merkmalen ist ebenso möglich.

Normalenvektoren an Markerpunkten

Eine eindeutig definierte Pixelmenge, die in jedem Bild korrespondierende Punkte markiert, besteht aus den 58 Labelpunkten, die in den vorbereitenden Schritten manuell gesetzt wurden. Werden die Koordinaten dieser Pixel genutzt, um die Normalenvektoren zu berechnen, sind eindeutige Ortsdefinitionen vorhanden. Dabei ist in den Berechnungsschritten keine Diskretisierung oder Standardisierung der Normalenvektoren notwendig und die berechneten Winkel werden als Merkmale angegeben. Da einige Labelpunkte über die verschiedenen Übungen ihre Positionen nicht verändern, enthalten sie eventuell keine Information über die Klassenfunktion. Aus diesem Grund wird ein zweiter Featureblock berechnet, der eine Teilmenge des ersten Featureblocks darstellt und die manuell selektierten Normalenvektoren an dynamischen Labelpunkten enthält. Die 41 dynamischen Labelpunkte bestimmen die Markerpositionen, die durch das Ausführen einiger Motorik-Übungen relativ zu anderen Markerpositionen verschoben werden. Da einige Labelpunkte die Gesichtsregion begrenzen, treten die bereits erwähnten Randprobleme bei der Berechnung der Normalenvektoren auf und die NB wird verringert. Zusätzlich können die Labelpunkte selbst durch Auf- oder Abrunden der Positionen auf ganzzahlige Spalten- und Zeilenwerte außerhalb der bestimmten Gesichtsmaske liegen und werden ebenfalls aus der NB entfernt. Für die übrigen Punkte in der NB wird die Regressionsrechnung vorgenommen.

Der erste Featureblock besteht für jedes der m Bilder aus 116 Einzelmerkmalen, da jeder Normalenvektor durch zwei Winkelwerte repräsentiert wird. Die Featurevektoren werden in der Reihenfolge der Labelnummerierungen mit den jeweiligen Winkeln des Normalenvektors an der entsprechenden Markerposition zusammengesetzt (siehe Gleichung 5.7).

$$\mathbf{x}^m = (\varphi_1, \theta_1, \varphi_2, \theta_2, \dots, \varphi_l, \theta_l), \quad \forall m \quad \text{und} \quad l = 58 \quad \text{oder} \quad l = 41 \quad (5.7)$$

Um eine Vergleichbarkeit der Merkmale verschiedener Berechnungsmethoden zu ermöglichen, müssen alle Featurewerte in einen Wertebereich skaliert werden. Negative

Vorzeichen der Werte besitzen einen Informationswert und sollen bei der Skalierung nicht verloren gehen. Daher werden alle Features, welche negative Werte enthalten, in den Wertebereich $[-1, 1]$ und alle Features, welche nur positive Werte enthalten, in den Bereich $[0, 1]$ skaliert. Der Definitionsbereich der Winkel ist in der Abbildung 5.3 verdeutlicht. Der Normalenvektor n wird durch die Winkel θ und φ eindeutig beschrieben. φ ist dabei von $[-180^\circ, 180^\circ]$ definiert. Innerhalb der sphärischen Koordinatendarstellung ist θ von $[0^\circ, 180^\circ]$ definiert. Da jedoch die Blickrichtung der Kamera aus der Richtung $-\mathbf{e}_z$ auf die XY-Ebene festgelegt ist, sind Beschränkungen im dreidimensionalen Bereich vorhanden. Verdeckte Gegenstände bzw. Objektrückseiten können in dieser Darstellung nicht dargestellt werden, daher ist θ auf den Bereich $[90^\circ, 180^\circ]$ beschränkt.

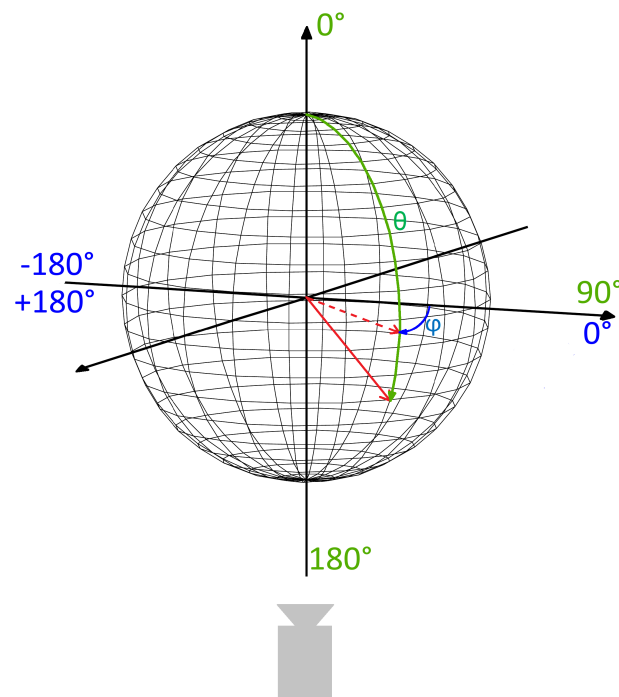


Abbildung 5.3: Definition der Normalenwinkel in sphärischen Koordinaten innerhalb eines Tiefenbildes. Der Normalenvektor ist rot und seine Projektion in die XY-Ebene gestrichelt dargestellt. Die Blickrichtung der Kamera ist aus $-\mathbf{e}_z$ auf die XY-Ebene gerichtet. Der Polarwinkel θ ist daher im Bereich $[90^\circ, 180^\circ]$ und der Azimutwinkel φ in $[-180^\circ, 180^\circ]$ definiert.

Normalenvektoren innerhalb definierter Spalten

Die Berechnung der Normalenvektoren erfolgt über einen definierten Ortsbereich, welcher durch eine Spalte eindeutig festgelegt ist. Die jeweiligen ausgewählten Spalten sind in ihrer Lage durch bestimmte Labelpunkte bestimmt. Diese Definition reicht für die pixelbasierte Berechnung nicht aus, da die Pixelanzahl innerhalb der Spalten variiert. Es ist daher eine anschließende Standardisierung notwendig. Die Verläufe der beiden Winkel werden über die Spalten dargestellt. Nicht alle Spalten des Gesichtsbereichs werden gewählt, da zu viele Einzelfeatures entstehen würden. Ein Beispieltiefenbild enthält 6147 Pixel mit Tiefeninformationen des Gesichts, daher würden sich mindestens 12294 Einzelmerkmale ergeben, je nachdem, ob die Werte der Winkelverläufe durch Interpolation noch vermehrt werden. Es werden manuell einige Spalten ausgewählt, die über Bereiche verlaufen, in denen Übungsveränderungen visuell sichtbar sind. Daraus ergeben sich drei Spalten, die durch die Labelpunkte 6, 29, 14 festgelegt werden. Eine genaue Darstellung der Lage der Spalten ist in Abbildung 5.4 grün dargestellt.

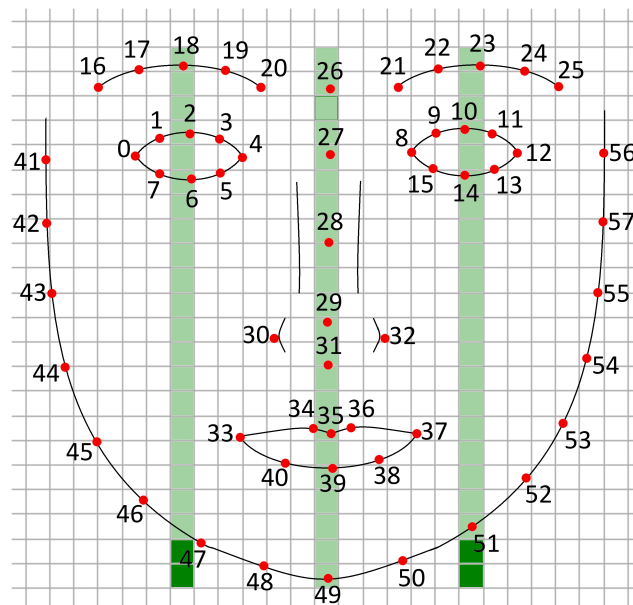


Abbildung 5.4: Das Gitter deutet die Spalten und Zeilen im Tiefenbild an. Die ortsbestimmenden Spalten der Normalenvektoren sind grün eingezeichnet. Dunkelgrün kennzeichnet die Tiefenwerte, die keine Gesichtsinformation tragen.

Die Anzahl der Pixel innerhalb der Spalten wird über das Minimum und das Maximum der Zeilen innerhalb der Gesichtsmaske bestimmt. Es ergeben sich einige der berücksichtigten Werte innerhalb der Spalten, welche in Abbildung 5.4 dunkelgrün markiert werden, die nicht mehr im Gesichtsbereich liegen. Die Winkelwerte an diesen Stellen werden mit den Konstanten, $\theta = 180^\circ$ und $\varphi = 0^\circ$ belegt. Sie sind so gewählt, wie es bei nicht ausgelenkten Normalenvektoren der Fall ist. Da es sich dabei nur um wenige Pixel handelt, wird keine Verfälschung der Merkmale durch das Besetzen der Stellen mit konstanten Werten angenommen. Über die Winkelverläufe erfolgt eine Interpolation mit definierter Anzahl an Werten, um alle Winkelverläufe über die Spalten mit gleich vielen und somit standardisierten Werten zu bilden. Die Anzahl der Werte, über die interpoliert wird, wird nacheinander auf 40, 60 oder 80 gesetzt, sodass sich über drei Durchläufe drei verschiedene Featureblöcke bilden. Ist die gewählte Zahl kleiner als die Anzahl der wahren Winkelwerte über einer Spalte, wird dieser Verlauf abgetastet. Um keinen Informationsverlust durch eine zu geringe Interpolationskonstante zu verursachen, wird diese variiert. Später soll durch die Merkmalsselektion einer der Featureblöcke ausgewählt werden. Jeder Featurevektor eines Blockes setzt sich für alle m Bilder, wie in Gleichung 5.8 gezeigt, zusammen.

$$\mathbf{x}^m = (\varphi_1^1, \dots, \varphi_i^1, \theta_1^1, \dots, \theta_i^1, \varphi_1^2, \dots, \varphi_i^2, \theta_1^2, \dots, \theta_i^2, \varphi_1^3, \dots, \varphi_i^3, \theta_1^3, \dots, \theta_i^3), \quad \forall m \quad (5.8)$$

mit $i = 40, 60 \text{ oder } 80$

Histogrammberechnung über diskrete Winkelverteilungen

Bei der Berechnung der Winkelverteilungen über alle Pixel der Gesichtsdaten kann eine einheitliche Datenmenge durch die Bildung eines Histogramms über die Winkelverteilungen erreicht werden. Dafür werden die Winkel ϕ und θ diskretisiert und voneinander unabhängig betrachtet. Informationsträger sind nicht die Winkel selbst, sondern die Häufigkeiten der Winkel über alle Pixel. Da θ über 90° und φ über 360° variieren kann, werden für beide Winkel zehn Intervalle eingeführt, über die Histogramme gebildet werden, um aussagekräftige Histogramme zu erhalten.

Bei der Normalenberechnung an jedem Punkt werden die X- und Y- Werte jeweils als dimensionslose Längeneinheiten betrachtet, da die Spalten- und Zeilenzahl der jeweiligen Pixel betrachtet wird. Im Vergleich dazu ist die Auflösung in Z-Richtung erheblich kleiner, da die Tiefenwerte in Meter angegeben werden und über ein Gesicht um ungefähr $0.025m$ variieren. Für eine Darstellung muss die Z-Achse daher sehr gestreckt werden, damit überhaupt ein Relief des Gesichts erkennbar ist. Als Folge dieser Art der Berechnung besitzt der Polarwinkel θ meistens nur sehr geringe Abweichungen von 180° . Bei einer linearen Aufteilung der zehn Winkelintervalle für θ würden daher fast alle der Winkel im ersten Intervall, welches durch $171^\circ \leq \theta < 180^\circ$ begrenzt würde, liegen. Das Histogramm hätte somit kaum Informationswert. Die Intervalle des Polarwinkels werden folglich nicht alle gleich groß gewählt, sondern durch eine empirisch ermittelte Funktion gebildet. Da der Polarwinkel in diesem Zusammenhang von $[90^\circ, 180^\circ]$ definiert ist und die Intervalle beginnend bei 90° immer kleiner werden sollen, wird der Hilfswinkel $\alpha = 180 - \theta$ eingeführt, für den die Intervallgrenzen nach der Gleichung 5.9 bestimmt werden. Dabei ist p der Stauchfaktor und k der Grad des Polynoms. Beide Größen werden empirisch bestimmt, um eine lineare Aufteilung der Grenzen von α , in 10° -Schritten, in eine nichtlineare Intervallgrenzbestimmung α' umzurechnen. Eine direkte Berechnung ist mittels Gleichung 5.10 möglich.

$$\alpha' = p \alpha^k \quad \text{mit} \quad k = 2,5 \quad \text{und} \quad p = 0.001 \quad (5.9)$$

$$\theta' = 180 - p (180 - \theta)^k \quad (5.10)$$

Der Definitionsbereich des Winkels φ wird linear in zehn Intervalle geteilt, welche daher jeweils eine Größe von 36° haben. Für beide Winkel sind die zehn Intervalle in Tabelle 5.1 aufgelistet, über die für jeden der beiden Winkel Histogramme berechnet werden.

Intervall	$\varphi [^\circ]$	Intervall	$\theta [^\circ]$
9	$\varphi < -144$	10	$\theta < 103.16$
7	$-144 \leq \varphi < -108$	9	$103.16 \leq \theta < 122.76$
5	$-108 \leq \varphi < -72$	8	$122.76 \leq \theta < 139.00$
3	$-72 \leq \varphi < -36$	7	$139.00 \leq \theta < 152.11$
1	$-36 \leq \varphi < 0$	6	$152.11 \leq \theta < 162.32$
2	$0 \leq \varphi < 36$	5	$162.32 \leq \theta < 169.88$
4	$36 \leq \varphi < 72$	4	$169.88 \leq \theta < 175.07$
6	$72 \leq \varphi < 108$	3	$175.07 \leq \theta < 178.21$
8	$108 \leq \varphi < 144$	2	$178.21 \leq \theta < 179.68$
10	$144 \leq \varphi$	1	$179.68 \leq \theta$

Tabelle 5.1: Definierte Winkelintervalle der Normalenvektoren zur Histogrammbestimmung.

Durch die Nutzung der Histogramme ist die Länge der Information zwar standardisiert, allerdings geht der Ortsbezug der Normalenvektoren im Gesicht vollständig verloren. Daher werden spezielle Regionen im Gesicht definiert, über die jeweils Histogramme gebildet werden, um eine örtliche Zuordnung der Winkelverteilungen herstellen zu können. Dabei sind die Gesichtsregionen so definiert, dass einzelne Übungsbewegungen in diesen visuell sichtbar sind, und die Form dieser Regionen von Interesse⁴ einen Informationswert bezüglich der Übungsklassifikation haben. Die Regionen der Wangen und des Mundes werden beispielsweise einzeln betrachtet. Die Festlegung der Regionen wurde rein visuell vorgenommen und die Position über die Lage der Labelpunkte definiert. Alle zehn Gesichtsregionen sind in Abbildung 5.5 dargestellt und entsprechend nummeriert.

Jedes der gebildeten Histogramme wird auf die Anzahl der untersuchten Pixel innerhalb der jeweiligen Regionen, also auf die relative Häufigkeit, normiert. Die Summe der Häufigkeiten über ein Histogramm ergibt somit eins und alle Häufigkeiten befinden sich im Wertebereich $[0, 1]$. Eine Vergleichbarkeit mit anderen Merkmalsarten ist gegeben. Über jede Region werden für beide Winkel die Histogramme über zehn Intervalle gebildet, was einen Merkmalsvektor mit 200 Elementen ergibt. Der Aufbau des

⁴englisch: Region of Interest (ROI)

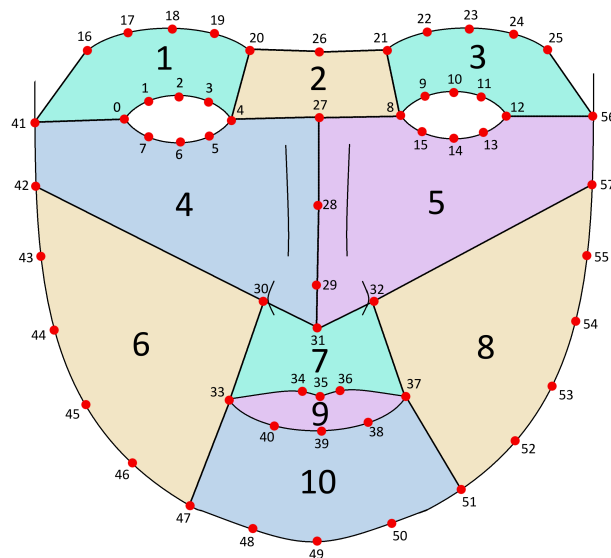


Abbildung 5.5: Eingezeichnete Regionen, in denen jeweils Histogramme über die Intervalle der Normalenwinkel berechnet werden.

ausgegebenen Featurevektors ist in Tabelle 5.1 angegeben.

$$\mathbf{x}^m = (\varphi_{hist_1}, \varphi_{hist_2}, \dots, \varphi_{hist_k}, \theta_{hist_1}, \theta_{hist_2}, \dots, \theta_{hist_k}) \quad \forall m \quad \text{mit} \quad k = 10 \quad (5.11)$$

Im weiteren Verlauf werden die Histogramme der beiden Winkel über eine Region als Featureblock bezeichnet. Histogramme über die definierten Winkelintervalle sind für ein Beispieltiefenbild der Übung „BoxenLi/BP14Li“ über die Region acht in Abbildung 5.6 dargestellt, wobei die Intervalle über die Intervalle in Tabelle 5.1 definiert sind.

Histogrammberechnung über diskrete Winkelkombinationen

Bei den bisherigen Betrachtungen wurden die Winkel der Normalenvektoren einzeln betrachtet. Mit dem folgenden Ansatz werden die Winkel in Kombination untersucht. Dafür werden die Normalen an allen Gesichtspunkten berechnet. Die Winkel werden in jeweils zehn Intervalle unterteilt, welche in Tabelle 5.1 angegeben sind. Eine Kombina-

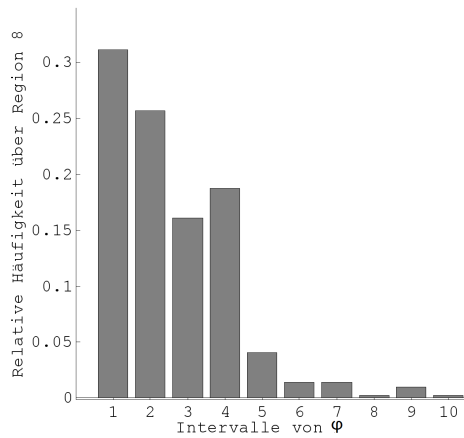
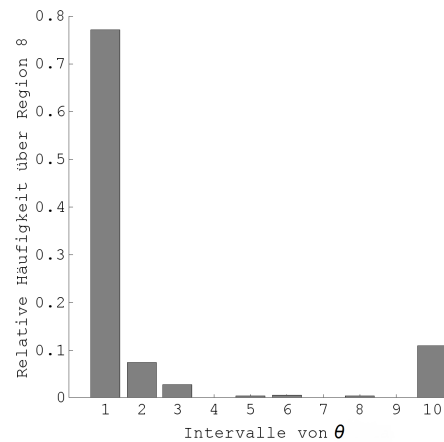
(a) Histogramm über die φ (b) Histogramm über die θ

Abbildung 5.6: Histogramme über die definierten Winkelintervalle für ein Beispielbild der Übung „BoxenLi“ über die Region

tion zweier Intervalle der beiden Winkel miteinander ergibt jeweils einen Oberflächenabschnitt der Einheitskreisoberfläche. Da es für beide Winkel jeweils zehn Intervalle gibt, existieren 10×10 Kombinationsmöglichkeiten. Oberflächenteile, die durch Kombinationen von Winkelintervallen entstehen, sind in Abbildung 5.7 durch die farbigen Markierungen angedeutet, wobei in der Darstellung in Abbildung 5.7 beide Winkel linear unterteilt wurden. Die Größe der Oberflächenanteile ist allerdings nicht für alle Kombinationen gleich. Die Aufteilung der Intervalle des Winkels θ verläuft nicht linear, wie bereit im vorherigem Ansatz beschrieben wurde. Diesen einzelnen Anteilen der Oberfläche werden bestimmte Nummern zugeordnet. Die Häufigkeiten, mit denen jede Kombination der Winkel über die Gesichtsdaten vorkommt, werden als Histogramme abgetragen, sodass in jedem Histogramm beide Winkel berücksichtigt werden.

Um den Ortsbezug herzustellen, werden zehn Histogramme über die einzelnen Gesichtsbereiche gebildet, die in Abbildung 5.5 dargestellt sind. Es ergeben sich zehn Histogramme, die jeweils Mengenangaben über 100 verschiedene Kombinationen der Winkel enthalten. Der ausgegebene Featurevektor enthält daher 1000 Einzelmerk-

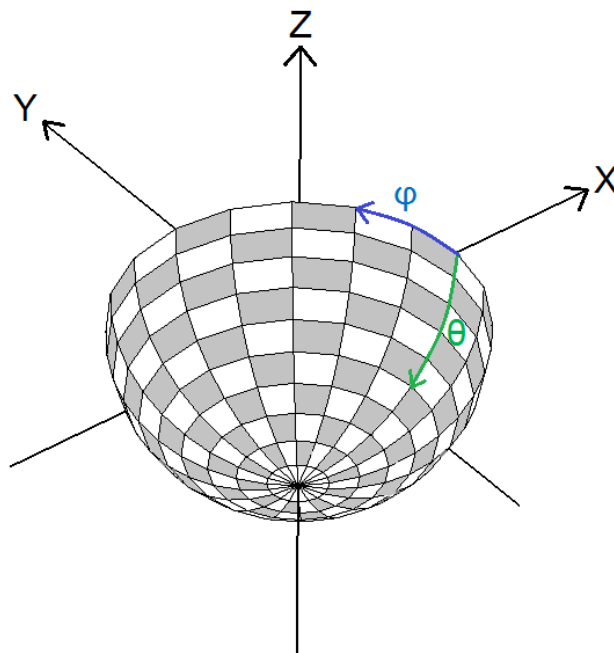


Abbildung 5.7: Die Oberflächenbereiche, die bei Kombination der Intervalle der beiden Winkel des Normalenvektors entstehen, sind durch weiße und graue Abgrenzungen gekennzeichnet. In der Darstellung sind die Intervallgrenzen beider Winkel linear bestimmt.

male und ist für jedes Bild m , wie in Gleichung 5.12 gezeigt, aufgebaut.

$$\mathbf{x}^m = (hist_1, hist_2, \dots, hist_k) \quad \forall m \quad \text{mit} \quad k = 10 \quad (5.12)$$

5.3 Punktsignaturen

Die Punktsignaturen wurden bereits im Unterabschnitt 4.5.2 vorgestellt und definiert. Im Folgenden wird die Berechnung von zwei Varianten der Punktsignaturen vorgestellt, die sich in der Radiusbestimmung der Kugel unterscheiden, mit dem die Schnittkurve bestimmt wird. Die allgemeine Berechnungsart ist in beiden Versionen gleich und wird daher nur einmal im Unterabschnitt 5.3.1 erläutert.

5.3.1 Punktsignaturen mit dynamischem Radius

Durch die Angabe von zwei Labelpunkten wird die dynamische Punktsignatur eindeutig bestimmt. Der erste Punkt gibt den Referenzpunkt \mathbf{p}^0 der Punktsignatur an. Die euklidische Distanz zwischen beiden Punkten gibt den Radius r an. Anschließend wird die Schnittkurve einer Kugel mit Radius r um den Punkt \mathbf{p}^0 und den Gesichtsdaten berechnet. Befinden sich Gesichtsdaten auf der Kugel, werden sie als Schnittpunkte gespeichert. Die Gesichtsdaten bestehen aus diskreten Werten, weswegen die Kugeloberfläche oftmals genau zwischen den einzelnen Pixeln liegt und keine Schnittpunkte existieren. In der Schnittkurve C können daher große Lücken entstehen. Der Radius der Kugel wird deshalb zu einem Intervall um r erweitert. Der maximale Berechnungsfehler des Radius wird abgeschätzt und als Intervallgrenze genutzt, sodass die Kugeloberfläche nicht mehr berührungslos zwischen einzelnen Pixel liegen kann. Die Unsicherheit der Ortsangabe durch die dimensionslosen X- und Y-Werte beträgt jeweils 0.5 Pixel. Die Unsicherheit in Z-Richtung wird sehr groß mit $0.035m$ angenommen, wie es bei einem Abstand zur Kamera von 5m der Fall ist (siehe Unterabschnitt 4.1.2). Die maximale Unsicherheit der Tiefendaten ist somit $\sqrt{0.5^2 + 0.5^2 + 0.035^2} \approx 0.7$ und führt zu einem Radiusintervall von $[r - 0.7, r + 0.7]$ für die Schnittkurvenberechnung. Die Schnittbedingung wird dabei spalten- und zeilenweise über alle Gesichtsdaten geprüft. Die gespeicherten Schnittpunkte sind daher nicht entlang der Schnittkurve sortiert, was erst mit Hilfe eines weiteren Verarbeitungsschritt nachgeholt wird.

In alle Schnittpunkte der Kurve C wird eine Ebene gelegt, um den Normalenvektor am Referenzpunkt \mathbf{p}^0 zu schätzen. Die Approximation des Normalenvektors erfolgt nach gleicher Berechnungsart wie die Normalenvektorbestimmung, welche in Unterabschnitt 5.2.1 beschrieben ist. Mit dem Unterschied, dass die Regressionsebene in die berechnete Schnittpunktmenge eingepasst wird. Die berechnete Ebene wird entlang des approximierten Normalenvektors \mathbf{n}^1 in den Referenzpunkt \mathbf{p}^0 verschoben und bildet die Referenzebene E_{Ref} , zu der die Distanzbestimmung erfolgt. Um eine Abtastung der Schnittkurve zu erreichen, werden alle n Schnittpunkte in E_{Ref} projiziert und dort um den Normalenvektor abgetastet. Dafür werden mittels Abstandsgleichung der

Hesseschen Normalform die orthogonalen Abstände aller Schnittpunkte zu E_{Ref} berechnet (siehe Unterabschnitt 5.2.1). Das Verschieben der Schnittpunkte entlang des Normalenvektors um den berechneten Abstand erfolgt nach Gleichung 5.13 und bildet die projizierte Schnittkurve C' .

$$\mathbf{p}^i = \mathbf{p}^i - (d \mathbf{e}_{\mathbf{n}1}) \quad i = 1, \dots, n \quad n \in \mathbb{N} \quad (5.13)$$

Der zweite angegebene Referenzpunkt bestimmt die Startrichtung der Abtastung der Schnittkurve, indem ein Vektor zwischen den Referenzpunkten \mathbf{p}^0 und \mathbf{p}^1 gebildet wird. Dieser Vektor wird in E_{Ref} projiziert und bildet dann einen Ortsvektor \mathbf{p}^{rot} bezüglich des lokalen Koordinatensystems am Punkt \mathbf{p}^0 . Dieser Ortsvektor wird durch die Rotationsmatrix \mathbf{R}_{rot} um die Normale \mathbf{n}^1 mit dem Winkel α gedreht. Die Rotationsmatrix ist in Gleichung 5.14 angegeben.

$$\mathbf{R}_{rot}(\alpha) = \begin{pmatrix} \cos(\alpha) + n_1^2(1 - \cos(\alpha)) & n_1n_2(1 - \cos(\alpha)) - n_3\sin(\alpha) & n_1n_3(1 - \cos(\alpha)) + n_2\sin(\alpha) \\ n_2n_1(1 - \cos(\alpha)) + n_3\sin(\alpha) & \cos(\alpha) + n_2^2(1 - \cos(\alpha)) & n_2n_3(1 - \cos(\alpha)) - n_1\sin(\alpha) \\ n_3n_1(1 - \cos(\alpha)) - n_2\sin(\alpha) & n_3n_2(1 - \cos(\alpha)) + n_1\sin(\alpha) & \cos(\alpha) + n_3^2(1 - \cos(\alpha)) \end{pmatrix} \quad (5.14)$$

Damit die Drehung des Ortsvektors um den Ursprung des lokalen Koordinatensystems erfolgt, werden die Punkte zunächst in dieses verschoben und nach der Rotationsberechnung wieder zurückgerechnet, wie es in Gleichung 5.15 gezeigt ist.

$$\mathbf{p}^{rot}/(\alpha) = \mathbf{R}_{rot}(\alpha) (\mathbf{p}^{rot} - \mathbf{p}_{rot}) + \mathbf{p}_{rot} \quad (5.15)$$

Die Rotation erfolgt für jeden Abtastwinkel α , wobei die verschieden stark rotierten Ortsvektoren am Referenzpunkt um den Normalenvektor gedreht sind. Durch den mi-

nimalen Abstand zwischen der Schnittkurve C' in der Ebene E_{Ref} und dem entsprechend rotierten Ortsvektor \mathbf{p}^{rot} wird der abgetastete Schnittpunkt in C' gefunden. Dabei wird die Distanz zwischen der gefundenen Stelle in C' und den korrespondierenden Punkten der Kurve C als Funktionswert der Punktsignatur von α gespeichert.

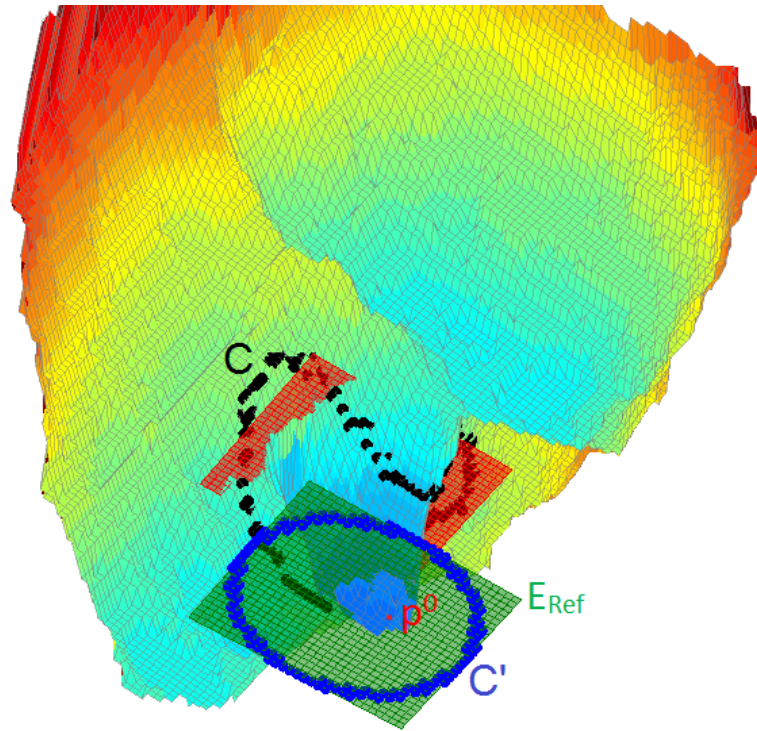


Abbildung 5.8: Tiefendaten eines Gesichts mit farblicher Kodierung der Distanzen zur Kamera. In schwarz ist die Schnittkurve C gekennzeichnet, deren Projektionen in die Ebene E_{Ref} (grün) werden durch C' (blau) angegeben. E_{Ref} ist die verschobene Regressionsebene (rot) der Schnittpunktmenge von C . Die Kurve C' wird in der E_{Ref} abgetastet und die korrespondierenden Distanzen zwischen C und C' berechnet.

Der Winkel α , mit dem die Schnittkurve abgetastet wird, ist eine konstante Größe und für eine äquidistante Kreisabtastung ein ganzzahliger Teiler von 360° . Der Winkel muss eine bestimmte radiusabhängige Größe überschreiten, um redundante Distanzen zu vermeiden. Aufgrund der diskreten Tiefendaten würden sonst bei zu kleinem Abtastwinkel mehrmals die gleichen Pixel abgetastet. Innerhalb der Masterarbeit ist der Winkel unter Berücksichtigung dieser Bedingungen mit $\alpha = 15^\circ$ festgelegt.

Der Radius wird, wie bereits beschrieben, durch die euklidische Distanz beider Referenzpunkte bestimmt. Da die Referenzpunkte aus den Labelpunkten gewählt wurden, die manuell in die Bilder eingezeichnet wurden, handelt es sich um reproduzierbare und eindeutig bestimmte Punkte. Allerdings können die Markerpositionen durch Übungsdurchführungen relativ zu anderen Markerpositionen verschoben werden. Beispielsweise befindet sich der Mundwinkel bei den Übungen „Kussmund“ und „Lippen breitziehen“ in unterschiedlichen Abständen zur Nasenspitze. Da die Radiusbestimmung in jedem Bild neu erfolgt, um unabhängig von interindividuellen Varianzen zu sein, variiert dieser dynamisch über die verschiedenen Übungen. Daher sind die untersuchten Orte abhängig von den Motorik-Übungen.

Die Punktsignaturen werden nur an speziell ausgewählten Referenzpunkten berechnet. Eine manuelle Vorauswahl von fünf Referenzpunkten erfolgt in Abhängigkeit über den vermuteten Informationswert für die Klassifikation. Punktsignaturen sollen keine irrelevanten Daten enthalten, weswegen sie nicht über den Gesichtsbereich hinauslaufen und Distanzen im Hintergrund berechnen. Die Referenzpunkte sind daher in Zentrumnähe des Gesichts gewählt. Durch die Lage der Referenzpunkte auf der vertikalen Symmetrielinie des Gesichts werden ähnliche Bereiche in beiden Gesichtshälften untersucht. Dies ist notwendig, da keine Bewegungssymmetrie der Übungen angenommen werden kann. Gerade bei muskulären Erkrankungen und Störungen, die mittels der ausgewählten Übungen therapiert werden können, sind Asymmetrien der Gesichtshälften bei Bewegungen typisch. Liegt der Referenzpunkt nicht auf der vertikalen Mittellinie des Gesichts, wird ebenso die korrespondierende Punktsignatur in der anderen Gesichtshälfte berechnet. Um eine Vergleichbarkeit der Signaturen über verschiedene Übungen zu ermöglichen, werden möglichst stabile Referenzpunkte gewählt, deren Lage bei verschiedenen Übungen robust ist, aber deren Umgebungen über die Übungen veränderlich sind. Es wurden fünf verschiedene Referenzpunkte gewählt, denen jeweils zur Radiusbestimmung und Startrichtungsdefinition der Abtastung ein zweiter zugeordnet wird. Die gewählten Signaturen sind in Abbildung 5.9 als Kreise im \mathbb{R}^2 angedeutet, deren Radius durch beide Punkte bestimmt ist.

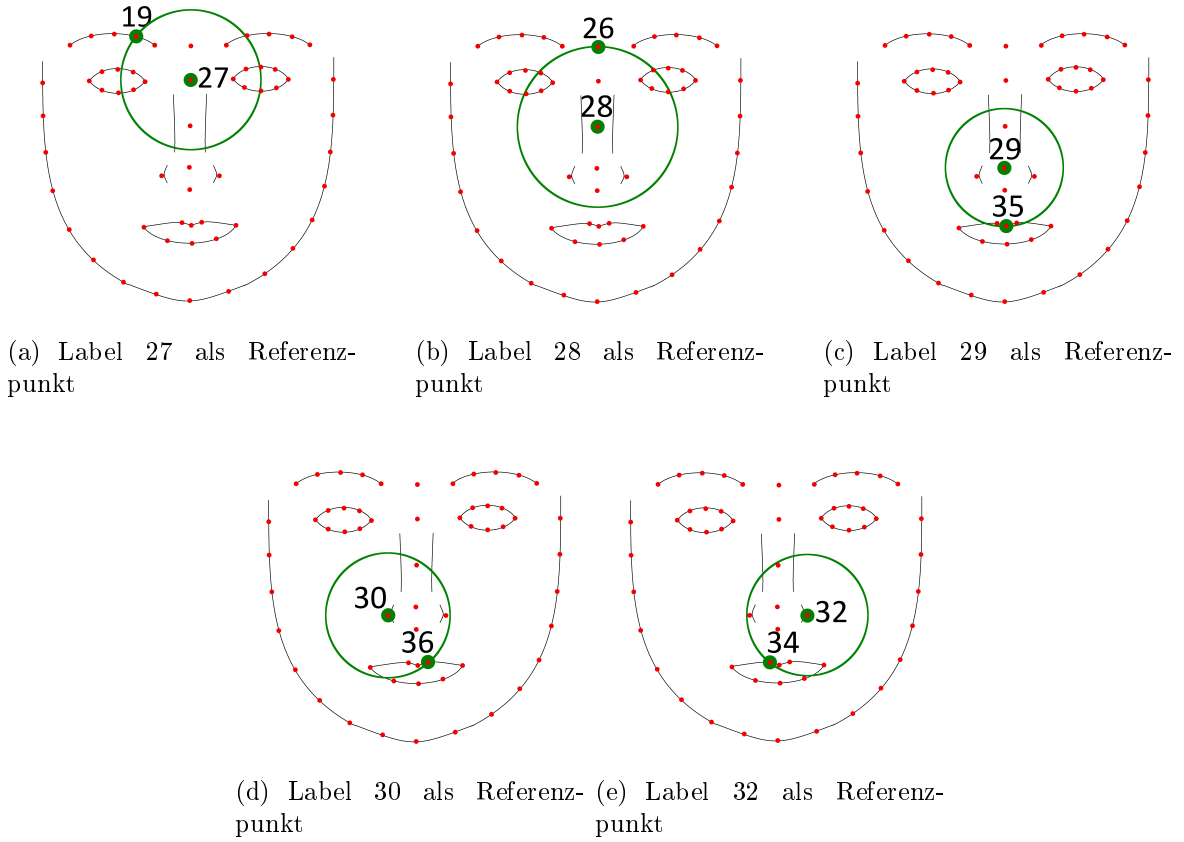


Abbildung 5.9: Alle manuell ausgewählten Punktsignaturen mit dynamischem Radius sind im \mathbb{R}^2 als Kreise angedeutet. In rot sind alle Labelpositionen eingezeichnet, welche als Referenzpunkte in Frage kommen. Die ausgewählten Punkte sind hervorgehoben.

Jede der Punktsignaturen wird durch einen Winkel von 15° abgetastet, was 24 Stützstellen für jede der Funktionen ergibt. Am Funktionsverlauf der Punktsignatur wird eine diskrete Kosinustransformation⁵ vorgenommen. Zum einen wird dabei die Dimension einer Signatur gesenkt und zum anderen stellen die Einzelmerkmale nicht einzelne Distanzen der Signatur dar, sondern beschreiben den Verlauf der Signatur. Die DCT ist eine reelle orthogonale Transformation, welche eine diskrete Funktion vom Ortsbereich in den Ortsfrequenzbereich transformiert. Dabei wird das Signal als Folge gewichteter trigonometrischer Funktionen mit unterschiedlichen Frequenzen dargestellt. Die

⁵englisch: Discrete Cosine Transform (DCT)

Basisfunktionen mit den kleinsten Koeffizienten werden für eine Datenreduktion als erstes entfernt und verändern das Signal nur geringfügig. Anhand einer Beispielpunktsignatur am Referenzpunkt 29 wird eine DCT vorgenommen und mit unterschiedlich vielen Koeffizienten zurück in den Ortsbereich transformiert. Das Ergebnis ist zusammen mit der Originalsignatur in Abbildung 5.10 dargestellt. Erkennbar ist, dass eine Rücktransformation ab 14 Koeffizienten einen sehr ähnlichen Funktionsverlauf besitzt wie die originale Punktsignatur. Mit zehn Koeffizienten ist der Verlauf bereits gut angenähert, daher wird mit zwölf DCT-Koeffizienten in jedem Fall eine gute Darstellung der Signatur erreicht. Die zwölf größten Koeffizienten werden nach der Skalierung als Featureblock ausgegeben.

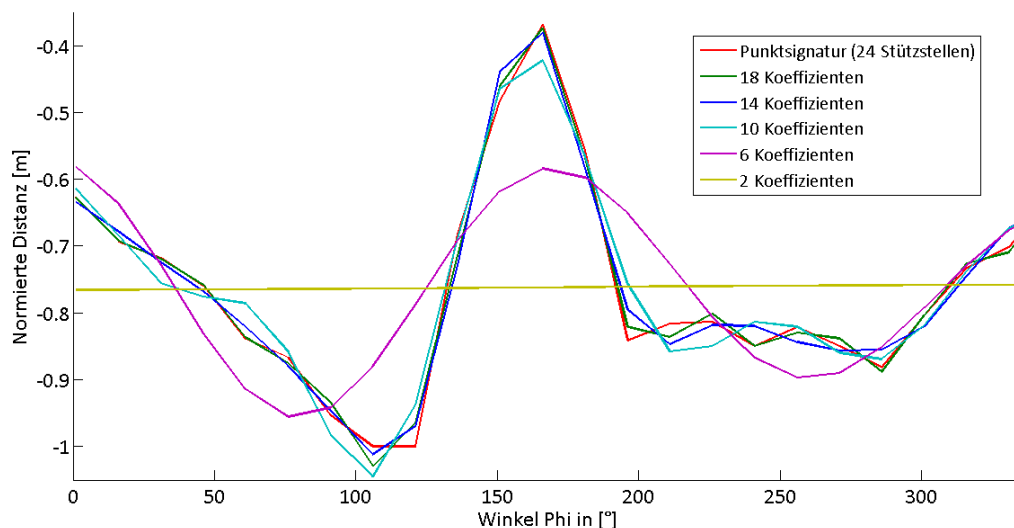


Abbildung 5.10: Die dynamische Punktsignatur am Punkt 29 ist in rot dargestellt. Rücktransformationen der diskreten Kosinustransformation mit unterschiedlich vielen Koeffizienten werden farblich unterschieden.

Ob durch das Entfernen der zwölf kleinsten Koeffizienten ein Verlust von Informationen über die Klassenverteilung des Klassifikationsproblems stattgefunden hat, wird in einer Untersuchung geprüft. Für alle fünf ausgewählten Punkte werden die Punktsignaturen zum einen durch den Distanzverlauf über 24 Stützstellen dargestellt und zum anderen durch eine durch die DCT auf zwölf Stützstellen reduzierte Form. Die MI zwischen

den einzelnen Features der Featureblöcke wird paarweise⁶ berechnet und die einzelnen MI-Werte in einer Matrix dargestellt, welche in Abbildung 5.11 zu erkennen ist.

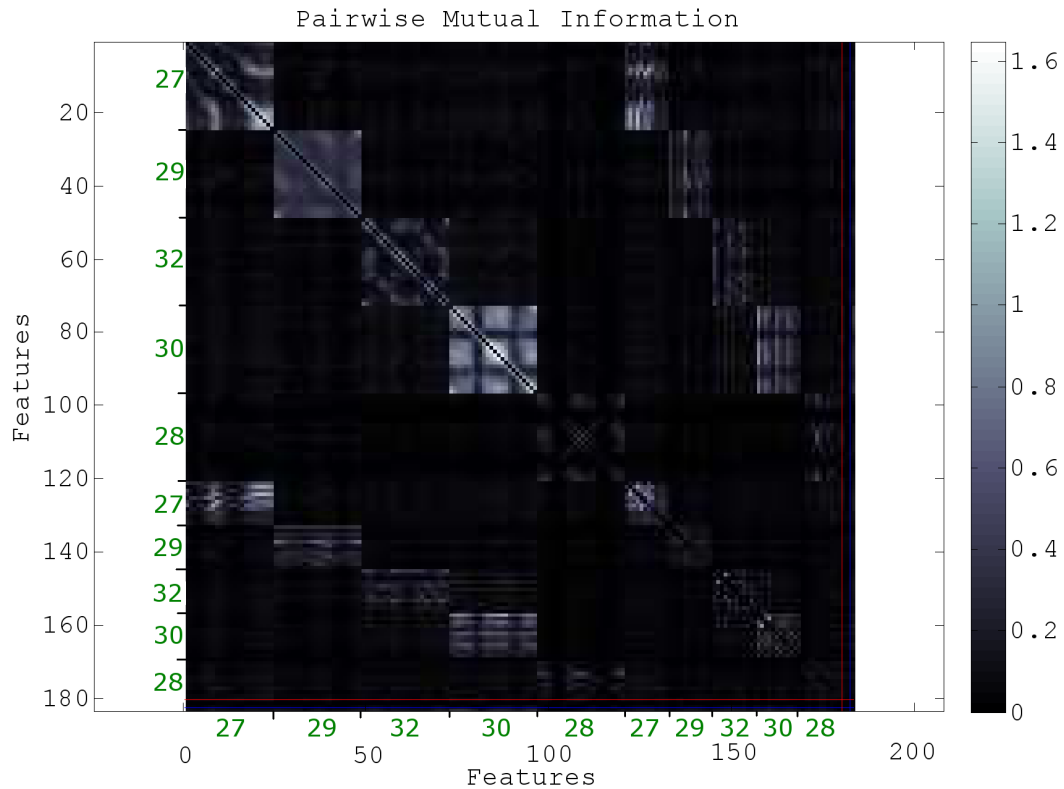


Abbildung 5.11: MI zwischen den Einzelmerkmalen der Punktsignaturen. Die Referenzpunkte sind für jeden Featureblock in grün angegeben.

Auf der X- und Y-Achse sind jeweils alle Features der zehn Blöcke hintereinander aufgereiht, beginnend bei den nicht reduzierten Punktsignaturen. Die Referenzpunkte, an denen die jeweiligen Blöcke berechnet werden, sind in grün abgetragen. Je höher die MI zwischen zwei verschiedenen Merkmalen ist, umso größer ist die statistische Abhängigkeit der beiden zueinander. Bei einem hohen MI-Wert zwischen zwei Merkmalen tragen beide ähnliche Informationen und sind daher als redundant anzusehen. Die Punktsignaturen weisen innerhalb ihres Funktionsverlaufs Redundanzen auf, was durch eine Blockbildung entlang der Hauptdiagonalen zu erkennen ist. Neben diesen Redundanzen, die später weiter diskutiert werden, ist deutlich eine erhöhte MI zwi-

⁶englisch: Pairwise Mutual Information

schen den Einzelmerkmalen der Punktsignaturen am gleichen Punkt mit unterschiedlicher Anzahl an Stützstellen zu erkennen. Dies führt zu einer hellen Rechteckbildung innerhalb der oberen und unteren Dreiecksmatrix. Daher besteht ein statistischer Zusammenhang zwischen den originalen und den reduzierten Punktsignaturen und eine Reduzierung der Stützstellen kann vorgenommen werden.

5.3.2 Punktsignaturen mit statischem Radius

Bei den Punktsignaturen mit einem statischen Radius erfolgt die Berechnung auf ähnliche Weise wie zuvor bei Punktsignaturen mit dynamischem Radius. Die Berechnung unterscheidet sich allein in der Methode der Radiusbestimmung. Die Länge des Radius ist konstant, da er durch die verschiedenen Bewegungen der Gesichtsmotorik-Übungen nicht beeinflusst wird. Der Radius wird so gewählt, dass Umgebungen erfasst werden, die durch Motorik-Übungen veränderlich sind. Der Radius berechnet sich aus der Gleichung 5.16 als ein Vielfaches der Länge EW , welche den Mittelwert beider Augenbreiten⁷ des Gesichts darstellt. Mittels Multiplikator m wird der Radius variiert.

$$r = m \cdot EW \quad m \in \mathbb{R} \tag{5.16}$$

Die Größe EW wird dabei jeweils individuell in jedem Bild bestimmt. Innerhalb eines Bildes ist der Radius translations- und rotationsinvariant, wobei die Translation in Z-Richtung der Skalierung entspricht. Da die Augenbreite im gleichen Verhältnis wie das komplette Gesicht durch Abstandsregulierung zur Kamera skaliert wird, ist der Radius skalierungsinvariant. Die Translation in X- und Y-Richtung verändert die Längeneinheit EW nicht, solange die Tiefendaten das Gesicht vollständig abbilden. Die Unabhängigkeit der Längeneinheit von der Rotation des Gesichts ist durch zwei Einschränkungen gegeben. Die Rotation um die X- und Y-Achse nimmt solange keinen Einfluss auf die Radiusbestimmung, bis durch Verdeckungen oder Eigenschattierungen die Label zur Positionsdefinition nicht mehr in den Tiefendaten sichtbar sind. Die Ro-

⁷englisch: Eye Width (EW)

tationsinvarianz um diese beiden Achsen ist daher für kleine Winkel γ und β gegeben. Die Abbildung 5.12 stellt diese Zusammenhänge dar.

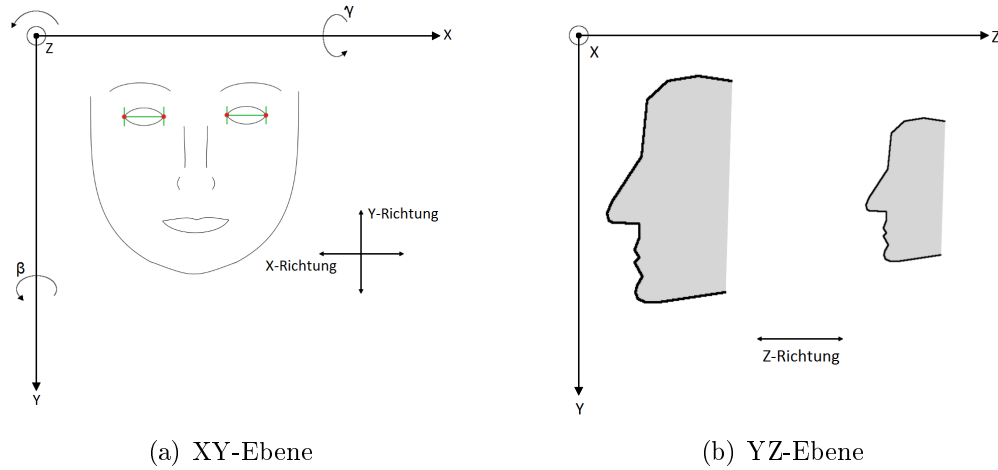


Abbildung 5.12: a) Translationsinvarianz des Radius in X- und Y-Richtung und Rotationsinvarianz um die Achse Z, b) Translationsinvarianz in Z-Richtung und Rotationsinvarianz bezüglich kleiner Winkel γ und β um die X- und Y-Achse.

Trotz der labelunabhängigen Radiusbestimmung wird für die Punktsignatur mit statischem Radius ein zweiter Referenzpunkt angegeben, um wie bei den dynamischen Radien den Startpunkt der Abtastung eindeutig festzulegen. Die Punktsignatur mit statischem Radius wird an den gleichen fünf Punkten berechnet wie die Punktsignaturen mit dynamischem Radius, welche in der Abbildung 5.9 dargestellt werden. Dabei werden für jeden der Punkte drei Signaturen mit unterschiedlichen Radien berechnet, welche durch den Multiplikator variiert werden. Dieser wird nacheinander mit $m = \{1, 1.5, 2\}$ gesetzt. Alle Punktsignaturen werden zunächst berechnet und anschließend mit einem Verfahren, welches in Abschnitt 4.6 beschrieben wurde, auf ihren Informationswert getestet. Wie in [MPIPERIS et al., 2007] wird daher eine Vorauswahl aller berechneten Signaturen anhand eines Kriteriums getroffen.

5.4 Dimensionsreduktion

Alle in Abschnitt 5.2 und Abschnitt 5.3 beschriebenen Features bilden zusammen einen Featurevektor und spannen einen Merkmalsraum mit 2718 Dimensionen auf. Eine zu-

sammenfassende Übersicht ist in Tabelle 5.2 gegeben. Diese hohe Anzahl der Dimensionen ist für das gestellte Problem mit der Datensatzgröße von 1480 Bildern aus den in Abschnitt 4.6 beschriebenen Gründen zu groß. Eine Dimensionsreduktion mittels Filteransatz und Featureranking wird nach dem Konzept, welches in Abschnitt 4.6 vorgestellt wurde, durchgeführt. Dabei erfolgt zusätzlich eine Redundanzuntersuchung der einzelnen Features, wobei eventuell redundante Featureblöcke ausgeschlossen werden. Eine Hauptkomponentenanalyse wird unter Umständen ebenfalls durchgeführt.

Featureart	Dimension	Anzahl Featureblöcke
Punktsignaturen (dynamisch) an 5 Punkten	5×12	5
Punktsignaturen (statisch) an 5 Punkten mit jeweils 3 Radian	15×12	15
Normalenvektor über zwei Labelsets der Größe 57 und 41	$2 \times 57 + 2 \times 41$	2
Winkelverläufe der Normalen über drei Spalten mit 40, 60 oder 80 Stützstellen pro Spalte	$3 \times 2 \times 40 + 3 \times 2 \times 60 + 3 \times 2 \times 80$	3
Histogramme über beide Winkel der Normalen mit jeweils 10 Intervallen über 10 ROI	$(10 + 10) \times 10$	10
Histogramme über 100 Winkelkombinationen in 10 ROI	10×10	10
Summe der Dimensionen über die Featurearten	2718	

Tabelle 5.2: Überblick über die Featurearten mit den resultierenden Merkmalsdimensionen, der Anzahl der resultierenden Featureblöcke und der Dimension des gesamten Merkmalsraumes.

5.4.1 Merkmalsranking mittels Mutual Information

Für alle Einzelmerkmale wird ein Featureranking mittels MI durchgeführt. Dabei wird für alle der MI-Wert mit der Klassenfunktion berechnet und absteigend sortiert. Die Grenzwertberechnung erfolgt durch zwei verschiedene künstliche Merkmale. Gleichverteiltes und normalverteiltes Rauschen werden als künstliches Merkmal genutzt, mit denen der MI-Wert mit der Zielfunktion berechnet wird. Dieser Vorgang wird für jeweils

hundert künstliche Merkmale durchgeführt und der Mittelwert über alle berechneten Werte bestimmt. Das Maximum aus beiden Mittelwerten der MI-Werte von gleichverteiltem und normalverteiltem Rauschen wird verdoppelt und bildet den Grenzwert für alle folgenden Betrachtungen. Mit den verwendeten Parametern ergibt sich der Grenzwert zu einem MI-Wert von $G = 0.0814$. Die Ergebnisse des Rankings werden featureblockweise bestimmt, wofür zunächst einzelnen Blöcke genauer definiert werden.

Jede Punktsignatur umfasst für jeden Referenzpunkt zwölf Funktionswerte, die einen Featureblock bilden. Die Winkelverläufe der Normalenvektoren über bestimmte Spalten werden über alle drei Spalten zu einem Block zusammengefasst, woraus sich drei Blöcke mit unterschiedlicher Stützstellenabtastung ergeben. Die Normalenwinkel über ein Labelset werden im Gesamten als Block zusammengefasst, weshalb sich zwei Featureblöcke bilden. Die Histogramme über die Winkelintervalle werden über jede der zehn Regionen als Block zusammengefasst. Für die Kombination der Winkelintervalle ist daher jedes Histogramm einzeln zu betrachten, während bei den Histogrammen über einzelne Winkelintervalle jeweils zwei Histogramme über eine Region als Block zusammengefasst werden.

Das Filterkriterium für Featureblöcke ist die Größe des Medians der MI eines Featureblocks, welcher größer als der festgelegte Grenzwert sein muss. Das bedeutet, dass mehr als 50% der Einzelfeatures eines Blocks über dem festgelegten Grenzwert liegen müssen, um selektiert zu werden. Alle Blöcke, deren Mediane unterhalb der Grenze liegen, werden nicht selektiert und im weiteren Verlauf dieser Arbeit nicht weiter betrachtet. Die selektierten Merkmalsblöcke sind in Form eines Balkendiagramms in Abbildung 5.13 gezeigt. Zu erkennen ist, dass mindestens die Hälfte der Einzelfeatures aller selektierten Blöcke über dem Grenzwert liegen, welcher mit einer roten Linie eingezeichnet ist. Die Winkelhistogramme der Normalenvektoren in den Regionen sechs bis zehn bilden die ersten 100 Merkmale. Die Regionen sind jeweils im Wechsel mit blau und grau gekennzeichnet, beginnend bei Region sechs. Die Farbabstufungen grenzen die einzelnen Histogramme über die zwei Winkel voneinander ab, wobei für jede Region zuerst das Histogramm über φ und dann das über θ abgetragen ist. Alle

Punktsignaturen mit dynamischem Radius sind in grün und durch Farbabstufungen gekennzeichnet. Ab Merkmal 161 folgen die Punktsignaturen mit statischem Radius, wobei die farblichen Gruppierungen der Blöcke Signaturen am gleichen Referenzpunkt kennzeichnen. Eine Punktsignatur am Punkt 28 mit dem Multiplikator $m = 1$ erfüllt das Kriterium nicht. Die Punktsignaturen werden immer in der Reihenfolge 27, 29, 32, 30 und 28 der Referenzpunkte dargestellt.

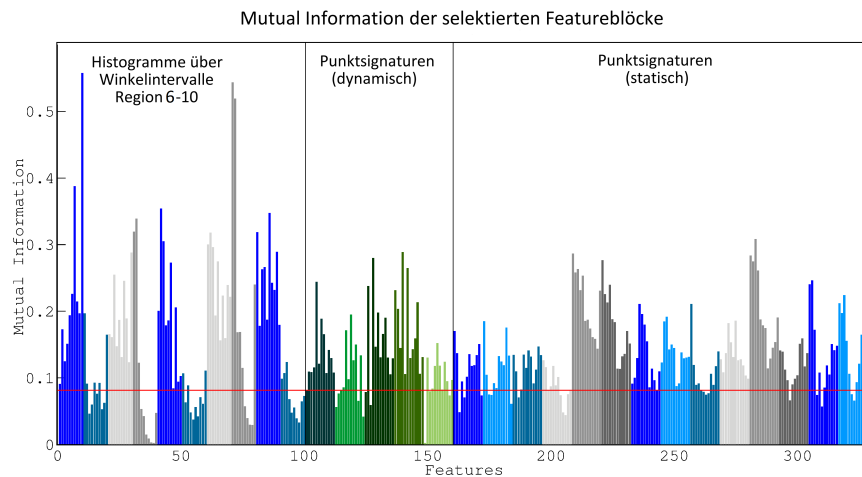


Abbildung 5.13: MI der selektierten Featureblöcke, welche das Filterkriterium erfüllen. Von links nach rechts: Die Histogramme über die Winkelintervalle in den Regionen sechs bis zehn, die Punktsignaturen mit dynamischem Radius und 14 von 15 Punktsignaturen mit statischem Radius.

5.4.2 Redundanzuntersuchung

Um redundante Merkmale, die im Vergleich zu anderen keinen Mehrwert besitzen, zu finden und aus dem Merkmalsraum zu entfernen, wird zwischen allen Einzelmerkmalen der bisher selektierten Features paarweise die MI berechnet. Die Ergebnisse werden in einer MI-Matrix in Abbildung 5.14 abgebildet. Zu beachten ist, dass die MI symmetrisch ist und an der Hauptdiagonalen gespiegelt wird. Die Betrachtung einer der Dreiecksmatrizen ist daher ausreichend. Die Hauptdiagonale müsste zusätzlich dazu die höchsten MI-Werte tragen, da es die MI-Werte eines Features mit sich selbst darstellt. Da diese Werte nicht von Interesse sind, werden sie nicht berechnet und in schwarz dargestellt. Grüne Linien grenzen unterschiedliche Featurearten voneinander ab. Das

erste Quadrat (oben links) beinhaltet alle Histogramme über die Winkelintervalle in den Regionen sechs bis zehn. Dabei wird über jede Region zuerst das Histogramm über φ und dann über θ aufgereiht. Das zweite Quadrat enthält alle Punktsignaturen mit dynamischen Radien mit der Referenzpunktfolgenfolge 27, 29, 32, 30 und 28. Die Reihenfolge im letzten Quadrat für die Punktsignaturen mit statischen Radien ist gleich, wobei an jedem Referenzpunkt mehrere Signaturen mit unterschiedlichen Radien in der Reihenfolge $m = 1$, $m = 1.5$ und $m = 2$ liegen. Am Referenzpunkt 28 werden nur die zwei größeren der Signaturen gebildet.

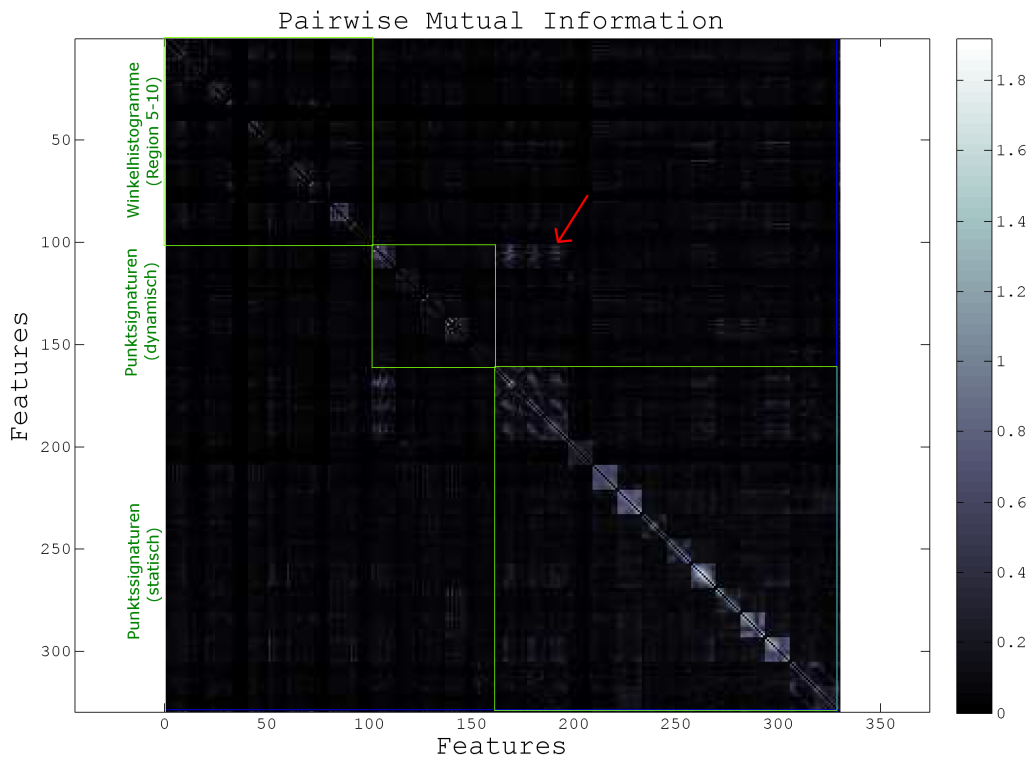


Abbildung 5.14: Redundanzprüfung durch paarweise MI zwischen den selektierten Featureblöcken. Die grünen Linien trennen verschiedene Featurearten voneinander. Hohe MI-Werte weisen auf statistische Abhängigkeiten hin. Durch den roten Pfeil sind Abhängigkeiten zwischen statischen und dynamischen Punktsignaturen am Referenzpunkt 27 gekennzeichnet. Die hellen Blöcke entlang der Hauptdiagonalen weisen auf featureblockinterne Redundanzen hin.

Alle Einzelfeatures der Signaturen mit dem Referenzpunkt 27 bilden untereinander höhere MI-Werte als die Umgebung. Diese Abhängigkeiten sind in Abbildung 5.14 zu erkennen, da die ersten statischen Punktsignaturen ein 36×36 Feature großes erhell-

tes Quadrat bilden. Zusätzlich sind außerhalb der Hauptdiagonalen erhellt 12×36 Feature große Rechtecke zu erkennen, von denen eins in Abbildung 5.14 mit einem roten Pfeil gekennzeichnet ist. Ein Entfernen von redundanten Blöcken ist daher möglich. Würde einer der Blöcke, die untereinander Redundanzen aufweisen, zwischen den eigenen Einzelfeatures keine Abhängigkeiten besitzen, würde im Vergleich zu den anderen jeder der Einzelfeatures einen Informationsmehrwert zur Klassifikation beitragen und wäre somit den anderen vorzuziehen. Da jedoch alle Signaturen am Punkt 27 innerhalb der Featureblöcke Abhängigkeiten aufweisen, werden weitere Quantile⁸ als Grenzwertkriterium genutzt. Bei der dynamischen Signatur liegen alle Einzelfeatures über dem Grenzwert, weshalb auch das 90%-Quantil über dem Grenzwert liegt. Bei den statischen Punktsignaturen wird dieses Kriterium nicht erfüllt. Bei diesen liegen 25% unter dem Grenzwert. Unterstützend dazu ist auch der Mittelwert über die MI-Werte der Einzelfeatures bei der Punktsignatur mit dynamischem Radius mit 0.1349 größer als bei den statischen Signaturen. Alle statischen Punktsignaturen am Punkt 27 mit $m = 1$, $m = 1.5$ und $m = 2$ werden somit aus der selektierten Featuremenge entfernt. Nach dieser Redundanzprüfung zwischen den unterschiedlichen Featureblöcken und dem vorherigen Merkmalsranking ergibt sich ein Merkmalsraum mit 328 Dimensionen, was 12.1% der Dimensionsmenge vor der Merkmalsselektion entspricht.

Die erhöhten MI-Werte, die sich blockweise entlang der Hauptdiagonalen abzeichnen, weisen auf blockinterne statistische Abhängigkeiten hin. Somit ist innerhalb der Blöcke zusätzlich eine weitere Dimensionsreduktion notwendig, damit jedes Einzelfeature Informationsmehrwert für die Klassifikation enthält. Daher wird auf alle Featureblöcke, welche einen Mittelwert über alle MI-Werte aus allen Kombinationen der internen Einzelfeatures einen Wert von 0.25 überschreiten, eine PCA durchgeführt. Der Schwellwert des Mittelwerts ist ein visuell festgelegter Wert, ab dem sich deutlich ein heller Block in der MI-Matrix abzeichnet. Für diese Untersuchung werden die Histogramme über die einzelnen Winkelintervalle einzeln betrachtet, womit sich über die fünf Regionen zehn verschiedene Blöcke ergeben. Durch diese Festlegung erfolgt eine Dimensionsreduktion mittels PCA für das Histogramm über die Intervalle des Winkels φ der zehnten

⁸Bezeichnet statistische Lagemaße, wobei der Median als 50%-Quantil bezeichnet wird.

ROI sowie der dynamischen Punktsignatur am Referenzpunkt 27 und an allen übrigen statischen Punktsignaturen. Die Anzahl der genutzten Dimensionen wird durch eine quantitative Analyse ermittelt, welche mit Gleichung 5.17 beschrieben wird.

$$\frac{\sum_{i=0}^{u-1} \lambda_i}{\sum_{j=0}^{n-1} \lambda_j} \geq \sigma \quad (5.17)$$

Der Prozentsatz der Datenvarianz, die im reduzierten Datensatz erhalten bleiben soll, wird durch σ repräsentiert und 0.9 gewählt. Das gewährleistet eine 90%ige Datenvarianz. In Gleichung 5.17 wird die Summe der ersten u absteigend sortierten Eigenwerte λ_i auf die Summe aller n Eigenwerte λ_j des n -dimensionalen Merkmalsraums normiert. Die gesuchte Merkmalsdimension m ist ausreichend, um eine Datenvarianz von σ zu erhalten [GROSS, 2011]. Die Dimension der Merkmale wird daher bei den transformierten Daten im neuen Koordinatensystem auf u Dimensionen reduziert. Nach der PCA ist die Datenreduktion abgeschlossen und es ergibt sich ein Merkmalsraum mit 183 Dimensionen, was 6.6% des Merkmalsraums der Ausgangssituation entspricht.

Kapitel 6

Ergebnisse

Die Ergebnisse der Klassifikation werden in den folgenden Abschnitten durch verschiedene Größen und Darstellungsweisen aufgezeigt. Die Klassifikationsleistungen beruhen auf den in Kapitel 5 beschriebenen Verfahren. Alle Berechnungen der Merkmale basieren auf der manuellen Setzung der Markerpunkte im RGB-Bild. Um die Stabilität der Verfahren gegenüber den Positionen der Labelpunkte zu testen, werden die manuell gesetzten Positionen zufällig verschoben. Die Ergebnisse dieser Untersuchung werden in Abschnitt 6.2 dargestellt.

6.1 Ergebnisse der Klassifikation

Die Erkennung von 19 Klassen erfolgte mit drei verschiedenen Klassifikatoren: Zum einen mit dem kNN- und dem Naive Bayes Klassifikator und zum anderen mit SVMs mit radialen Basisfunktionskernen. Die Nachbarschaft des kNN-Klassifikator wurde mit $k = 5$ festgelegt. Die Klassifikation wurde mit drei verschiedenen Merkmalsräumen durchgeführt. Sie wurde an allen extrahierten Merkmalen, welche einen Merkmalsraum mit 2718 Dimensionen aufspannen, mit allen Merkmalen, die das Filterkriterium erfüllen und am resultierenden Merkmalsraum nach der Dimensionsreduktion durchgeführt. Dabei wird deutlich, dass die Erfolgsrate bei allen Klassifikatoren mit der 11-fachen Kreuzvalidierung über die Dimensionsreduktion steigt oder gleich bleibt. Bei der Klassifikation mittels Hold-Out Methode wird die Erfolgsrate des Naive Bayes

Klassifikators erhöht, bei den anderen beiden Methoden gehen diese leicht zurück. Festzuhalten ist jedoch, dass eine Dimensionsreduktion um 56% zum vorherigen Schritt bei dem kNN-Klassifikator und den SVMs eine Einbuße der Leistung von weniger als 10% verursacht. In Tabelle 6.1 sind die Erfolgsraten bezüglich aller Klassifikationen angegeben. Hingegen wurde die Klassifikation mit SVM bei 2718 Dimensionen nicht durchgeführt, da für jeden Durchgang der Klassifikation mit 19 Klassen 171 SVMs trainiert werden müssten, was aufgrund der Größe der vorliegenden Datenmenge zu aufwändig ist (siehe Abschnitt 4.7). Die Erfolgsraten mittels Hold-Out Methode bei 2718 Dimensionen werden über drei Durchläufe, bei den reduzierten Merkmalsräumen über 30 Durchläufe, gemittelt, um eine stabile Angabe der Größen zu erreichen.

Dimension	kNN		Naive Bayes		SVM	
	HO	11-fache KV	HO	11-fache KV	HO	11-fache KV
2718	49%	16%	52%	21%	-	-
328	49%	24%	61%	53%	68%	29%
183	44%	28%	65%	53%	62%	39%

Tabelle 6.1: Erfolgsraten über alle betrachteten Klassifikatoren und Validierungsarten. Reduzierung des Merkmalsraumes durch Ranking und Dimensionsreduktion mittels Redundanzentfernung und PCA.

Die Erfolgsraten aller Klassifikatoren werden durch den ersten Schritt der Merkmalsselektion mittels Filterkriterium gesteigert oder bleiben gleich. Die Dimensionsreduktion durch Entfernen von redundanten Merkmalen und die anschließende PCA steigern nur die Ergebnisse der 11-fachen Kreuzvalidierung.

Die vollständigen Konfusionsmatrizen der Klassifikationen sind in Abschnitt A.4 dargestellt. Alle Erfolgsraten sind mit Standardabweichungen über mehrere Durchläufe in Tabelle 6.2 dargestellt. Zwischen den Erfolgsraten der Klassifikation mittels Hold-Out Methode und der 11-fachen Kreuzvalidierung ist eine deutliche Differenz der Erfolgsrate zu erkennen. Die Erfolgsraten mittels 11-facher Kreuzvalidierung sind für jeden Klassifikator kleiner.

Das Ergebnis wird durch die BER bezüglich aller Klassen ebenfalls bestätigt, welche in Tabelle 6.3 aufgelistet sind. Bei beiden Validierungsmethoden besitzt die Klassifikation

	Hold-Out Methode	11-fache Kreuzvalidierung
kNN	$44 \pm 1.5\%$	$28 \pm 4.7\%$
Naive Bayes	$65 \pm 2\%$	$53 \pm 10.1\%$
SVM	$62 \pm 2\%^*$	$39 \pm 9.4\%$
* Berechnung der Standardabweichung nur für die Vollständigkeit, da die Erfolgsrate nur über drei Durchläufe gemittelt wurde.		

Tabelle 6.2: Erfolgsraten der Klassifikation in 19 Klassen mittels resultierendem Merkmalsraum. Bei der Hold-Out Methode wurde für kNN und Bayes über 30, bei SVM über 3 Durchläufe gemittelt. Bei der 11-fachen Kreuzvalidierung sind die Datensets personenweise aufgeteilt.

	Hold-Out Methode	11-fache Kreuzvalidierung
kNN	53%	81%
Naive Bayes	35%	49%
SVM	40%	64%

Tabelle 6.3: BER der drei Klassifikatoren über 19 Klassen bei Nutzung der Hold-Out Methode und der 11-fachen Kreuzvalidierung und Nutzung des resultierenden Merkmalsraums.

mittels Naive Bayes Klassifikator die höchste Erfolgsrate sowie die niedrigsten BER und zeigt somit hinsichtlich der zwei untersuchten Größen die besten Klassifikationsleistungen. Beide Ergebnisse werden daher bezüglich der einzelnen Klassen genauer betrachtet und in Tabelle 6.4 aufgezeigt.

Für jede Klasse wird die TPR, also die Menge der richtig klassifizierten Daten in Bezug auf die Anzahl aller zu der Klasse gehörigen Daten, angegeben. Für die Betrachtungen wird pro Klasse ein binäres Klassifikationsproblem betrachtet, wobei die jeweils restlichen Klassen zu einer zusammengefasst werden. Als Folge ergibt sich eine stark unbalancierte binäre Aufteilung. Um diese Dysbalance für die Interpretation der Ergebnisse auszugleichen, wird zusätzlich für jedes binäre Problem die BER berechnet. Es ergibt sich für beide Validierungsarten bezüglich der BER ein ähnlicher Verlauf über die einzelnen BPs, was in Abbildung 6.1 bildlich veranschaulicht ist. Ebenfalls wird erneut deutlich, dass die Klassifikation mittels 11-facher Kreuzvalidierung bezüglich jeder Klasse eine schlechtere Klassifikationsleistung erreicht.

Bei der 11-fachen Kreuzvalidierung werden die einzelnen Datenblöcke personenweise aufgeteilt, wobei keine Bilder der Testpersonen im Trainingsset enthalten sind. Die

BP	Kürzel	Naive Bayes, Hold-Out		Naive Bayes, 11-fache KV	
		TPR [%]	BER [%]	TPR[%]	BER [%]
1	Stirn	71	16	45	29
2	Brauen	46	28	24	39
3	Augen	34	34	13	45
4	Augenpressen	50	27	31	37
5	Kuss	58	22	46	29
6	Breit	83	9	71	16
7	AForm	81	11	79	12
8	OForm	43	30	28	37
9	IForm	72	15	76	13
10	UForm	47	29	27	38
11	Zunge	57	28	47	29
12	ZungeO	79	11	66	18
13l	ZungeLi	61	20	46	27
13r	ZungeRe	58	22	37	32
14l	BoxenLi	84	8	80	11
14r	BoxenRe	84	8	69	17
15	Wangen	82	9	72	14
16l	WangeLi	80	10	73	14
16r	WangeRe	83	6	80	11

Tabelle 6.4: TPR und BER bezüglich der einzelnen Klassen bei Klassifikation mittels Naive Bayes Klassifikator. Für die Berechnungen wurden jeweils binäre Klassifikationsprobleme der Klasse gegenüber den restlichen Daten angenommen.

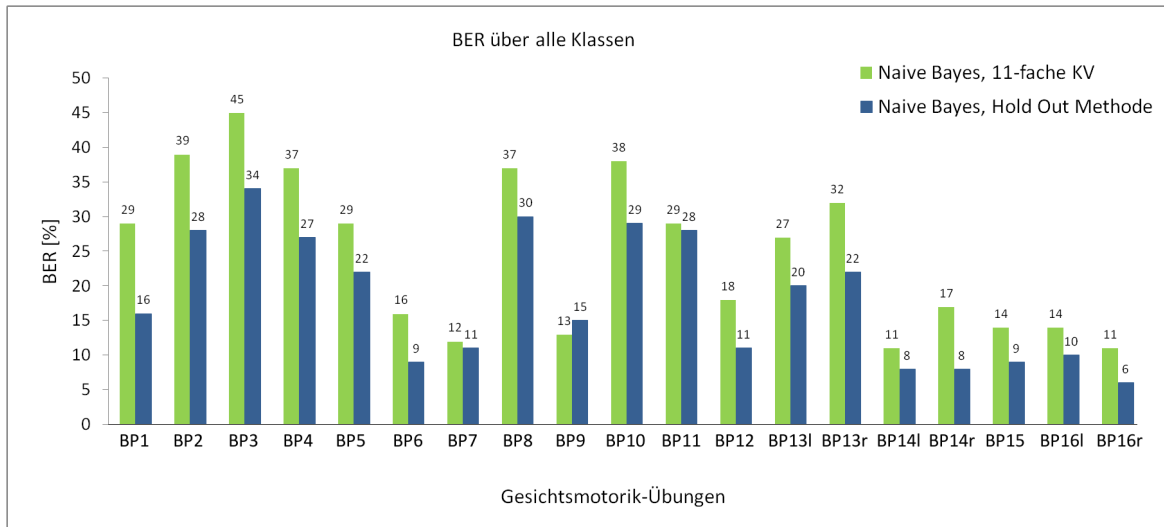


Abbildung 6.1: BER der 19 Klassen mit Klassifikation durch den Naive Bayes Klassifikator über die 11-fache Kreuzvalidierung und der Hold-Out Methode über 30 Durchläufe.

Erfolgsraten der einzelnen Validierungsschritte der Kreuzvalidierung sind in Tabelle 6.5 aufgelistet. Dabei ist auffällig, dass für alle drei Klassifikatoren die Erfolgsraten über die einzelnen Testpersonen deutliche Differenzen aufweisen und die gleichen Verläufe über die einzelnen Probanden zeigen.

Proband	SVM	kNN	Naive Bayes
1	55 %	29 %	68 %
2	50 %	28 %	64 %
3	45 %	28 %	59 %
4	44 %	27 %	58 %
5	43 %	27 %	57 %
6	39 %	23 %	57 %
7	37 %	22 %	55 %
8	34 %	21 %	53 %
9	31 %	20 %	44 %
10	27 %	17 %	42 %
11	24 %	16 %	33 %

Tabelle 6.5: Erfolgsraten einzelner Durchläufe der 11-fachen Kreuzvalidierung über alle Probanden.

6.2 Ergebnisse mit verschobenen Markerpositionen

Die extrahierten Merkmale, welche für die Klassifikation der BP genutzt werden, sind in ihrer Ortsdefinition abhängig von den in den Vorverarbeitungsschritten manuell gesetzten Labelpunkten. Die Positionsdetektion der Referenzpunkte der Merkmale soll zukünftig vollautomatisch durchgeführt werden, weswegen die Robustheit der Merkmale gegenüber Verschiebungen der Markerpositionen an einem Beispiel untersucht wird. Eine automatische Detektion der Positionen unterliegt statistischen Schwankungen. Die manuell gesetzten Markerpositionen werden daher zufällig in der Bildebene in X- und Y-Richtung mit normalverteilter Distanz um den Mittelwert Null verschoben. Dafür wird in beide Richtungen eine Standardabweichung von $\sigma_x = 3$ und $\sigma_y = 3$ Pixel gewählt.

	Hold-Out	11-fache KV
kNN	45 \pm 1.7%	23 \pm 4.7%
Naive Bayes	66 \pm 1.7%	54 \pm 10%
SVM	63 \pm 0.8%	39 \pm 9.4%
* Berechnung der Standardabweichung nur für die Vollständigkeit, da die Erfolgsrate nur über drei Durchläufe gemittelt wurde.		

Tabelle 6.6: Erfolgsrate aller Klassifikationen bei künstlich verschobenen Markerpositionen mit $\sigma_x = 3$ und $\sigma_y = 3$ Pixel. Bei der Hold-Out Methode sind die Ergebnisse für kNN- und Naive Bayes Klassifikator über 30, sonst über 3 Durchläufe gemittelt.

	Hold-Out	11-fache KV
kNN	59%	80%
Naive Bayes	36%	49%
SVM	40%	65%

Tabelle 6.7: BER aller Klassifikationen bei künstlich verschobenen Markerpositionen mit $\sigma_x = 3$ und $\sigma_y = 3$ Pixel. Bei der Hold-Out Methode sind die Ergebnisse für kNN- und Naive Bayes Klassifikator über 30, sonst über 3 Durchläufe gemittelt.

Die Unterschiede zwischen den Klassifikationsleistungen mit verschobenen Markerpositionen zu den Originalpositionen bleiben sehr gering für die getesteten zufälligen

Veränderung der Labelpositionen in kleinen Bereichen. In Abbildung 6.2 sind die Klassifikationsleistungen basierend auf manueller Positionierung der Markerpositionen rot und die Klassifikationsleistungen mit zufällig verschobenen Positionen der Markerpunkte grün dargestellt. Für jede Parameterkombination befindet sich die Differenz der Erfolgsrate innerhalb der berechneten Unsicherheiten der Erfolgsraten bei Nutzung der exakten Labelpositionen. Die maximale Differenz der Erfolgsrate liegt bei 5% und bei der BER bei 6%.

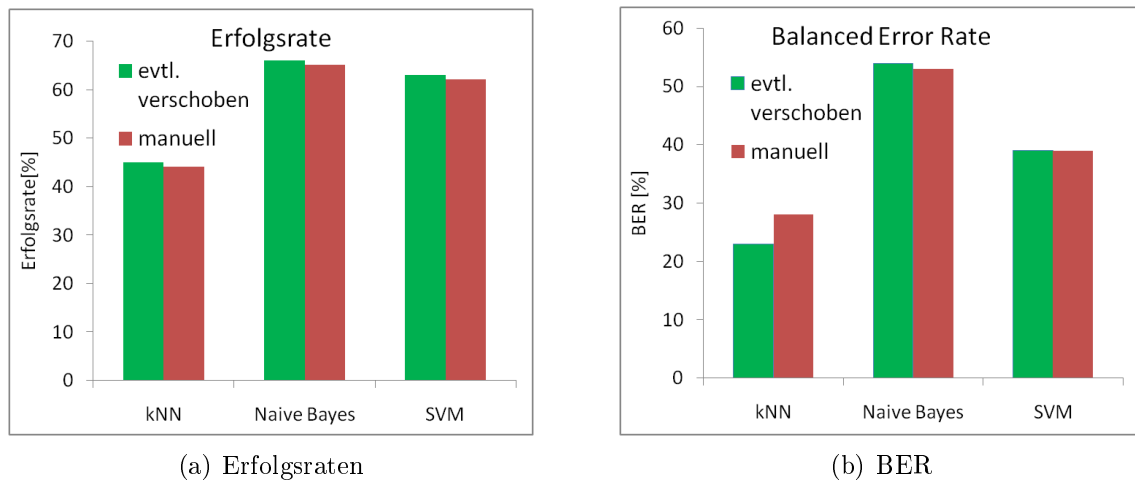


Abbildung 6.2: Die Erfolgsraten und BER der Klassifikation basierend auf manuell gesetzten Markerpunkten sind rot, die Positionen mit $\sigma_x = 3$ und $\sigma_y = 3$ Pixel bestimmten Verschiebungen grün dargestellt.

Kapitel 7

Diskussion

Die Ergebnisse der Klassifikation werden zum einen bezüglich der durchgeführten Merkmalsselektion und -extraktion und zum anderen im Bezug auf die unterschiedlich genutzten Klassifikatoren und Validierungsmethoden diskutiert. Klassifikationsleistungen einzelner Übungen werden detaillierter betrachtet.

7.1 Extraktion und Selektion der Merkmale

Bei der Durchführung der Merkmalsreduktion mittels Featureranking ist auffällig, dass von allen Berechnungsarten der Normalenvektoren nur Histogramme über die Winkelintervalle in den ROI sechs bis zehn das Filterkriterium erfüllen. Weitere Variationen der Ortsdefinition oder Standardisierung weisen nur geringe MI-Werte auf. Daher sind entweder die Berechnungen ungeeignet oder die ausgewählten Orte zeigen keine signifikanten Veränderungen über ausreichend viele Übungen. Betrachtet man das Verhältnis der Übungen, so beeinflussen bei richtiger Durchführung nur vier der 19 Übungen den Stirn- und Augenbereich. Ein Merkmal im Augenbereich enthält im Vergleich zu einem im Mundbereich, der von 15 Übungen beeinflusst wird, weniger Information über die Klassenfunktion und hat somit einen geringeren MI-Wert. Der Referenzpunkt der einzigen Punktsignatur, die das Filterkriterium bezüglich des MI-Grenzwertes nicht erfüllt, liegt ebenso in der oberen Gesichtshälfte. Berechnungsorte wie die definierten Spalten oder die Markerpositionen werden nicht ausreichend über die BPs verändert

und somit im Featureranking aussortiert. Weitere Punktsignaturen im Stirnbereich erfüllen jedoch das Kriterium. Bei Veränderungen des Tiefenbildes durch Übungen im Stirn- und Augenbereich handelt es sich im Vergleich zu weiteren Regionen um geringe Veränderungen. Die Normalenvektoren werden über eine Nachbarschaft geschätzt und unterliegen einem Glättungseffekt. Kleine Veränderungen im Pixelbereich werden daher von Normalenvektoren nicht erfasst, die eventuell von günstig gelegenen Punktsignaturen erhoben werden. Aus diesen Gründen erfüllt keine Berechnungsart der Normalenvektoren das Filterkriterium im oberen Gesichtsbereich und es werden nur Punktsignaturen selektiert.

Eine Berechnung der Normalenvektoren in einer kleineren Nachbarschaft würde den Glättungseffekt senken. Bei Berechnungen der Normalenvektoren an den Knoten innerhalb eines Meshs wird für diesen der Mittelwert aus allen Flächennormalen der angrenzenden Flächen gebildet. In einem Dreiecksmesh bedeutet das, dass die Flächen eindeutig von jeweils drei Punkten aufgespannt werden. Der Normalenvektor wird über Mittelung der angrenzenden Flächennormalen bestimmt. Kleinere Tiefendistanzen werden somit ebenfalls erfasst. Anliegendes Rauschen hat durch die Mittelwertbildung weniger Einfluss auf die Normalenberechnung als bei einer einfachen Dreier-Nachbarschaft zur Berechnung. Zusätzlich wird der Rauscheffekt durch die Histogrammbildung abgeschwächt, da die Winkel der Normalenvektoren nur indirekte Merkmale darstellen und durch die Histogrammhäufigkeiten repräsentiert werden.

Die Histogramme über die Winkelkombinationen werden ebenfalls über zehn ROI berechnet, erfüllen jedoch in keiner der Regionen das Filterkriterium. Viele der 100 Winkelbereiche pro Histogramm werden trotz nicht linearer Aufteilung der Winkelintervalle für den Polarwinkel θ in den untersuchten Regionen nicht besetzt. Eine Vielzahl der Einzelmerkmale trägt somit keine Information über die Zielfunktion und das Filterkriterium für den Featureblock wird nicht erfüllt. Die Winkelverteilungen könnten durch eine verbesserte Normalenberechnung aussagekräftiger werden. Die Berechnung der Normalenvektoren erfolgt, wie in Unterabschnitt 5.2.1 beschrieben, über dimensionslose Größen zur Beschreibung der X- und Y-Richtung, was zu sehr kleinen Winkeln

θ führt. Eine Umrechnung der Ortsangaben in Meter, wie es kürzlich in der Veröffentlichung [KHOSHELHAM und ELBERINK, 2012] vorgestellt wurde, könnte dieses Problem beheben. Die Umrechnung erfolgt nach Gleichung 7.1 und Gleichung 7.2, wobei x und y die Tiefenbildkoordinaten des Objektes und x_0 und y_0 der Ursprung des genutzten Koordinatensystems sind. Die Parameter δx und δy korrigieren die Verzerrung des Bildes aufgrund der Linsenkrümmung der Kamera.

$$X_{Obj} = -\frac{Z_{Obj}}{f}(x - x_0 + \delta x) \quad (7.1)$$

$$Y_{Obj} = -\frac{Z_{Obj}}{f}(y - y_0 + \delta y) \quad (7.2)$$

Durch diese Berechnung sind Normalenwinkel zu erwarten, die am Gesichtsrand 45° überschreiten und in verschiedenen Gesichtsbereichen über die Übungen differenzierbare Winkelhistogramme bilden.

Normalenvektoren an einem Punkt werden mittels Umgebungsdaten approximiert. An den Gesichtsrandern wird die Nachbarschaftsdatenmenge verkleinert, um keine Verfälschung der Normalen durch Hintergrunddaten zu verursachen. Diese Verkleinerung der Nachbarschaftsmenge kann allerdings zu einer Verschlechterung der Approximation führen, da Bildrauschen einen höheren Einfluss auf die Berechnung nimmt. Eine Lösung dafür wäre ein kleiner Randbereich, der für die Berechnung der Normalen unberücksichtigt bleibt. Die Menge der berechneten Normalenvektoren wird dabei zwar kleiner, allerdings verfälschen keine Hintergrunddaten oder Approximationsungenauigkeiten die Berechnungen.

7.2 Einfluss der Markerposition

Um die Stabilität der Klassifikation gegenüber verschobenen Markerpositionen zu testen, werden die manuell ausgewählten Positionen der Label unabhängig voneinander in X- und Y-Richtung verschoben. Dafür wird die Verschiebung in beide Richtungen mit einer Standardabweichung von $\sigma_x = 3$ und $\sigma_y = 3$ Pixel um die manuell gesetzten Positionen gewählt. Die Labelposition befindet sich also mit einer Wahrscheinlichkeit von 68% innerhalb eines 7×7 Pixel großen Quadrates um die manuell gesetzte Position. Die Labelpunkte werden manuell oftmals nicht in die Mitte der Pixel gesetzt. Die Position in den Zeilen und Spalten der Bildmatrix wird durch Auf- und Abrunden bestimmt. Daher kann je nach Position das Rechteck, welches durch die Standardabweichungen bestimmt wird, durch Rundungen vergrößert werden. In Abbildung 7.1 werden diese beispielhaft dargestellt.

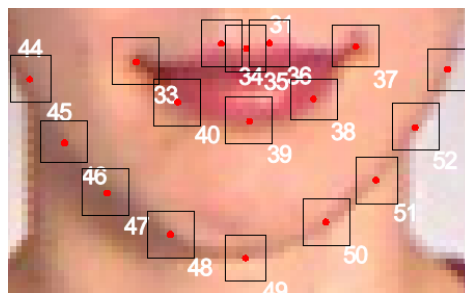


Abbildung 7.1: RGB-Bildausschnitt mit manuell gesetzten Labels (rot) und den Bereichen, in denen mit einer Wahrscheinlichkeit von 68% die verschobenen Marker liegen (schwarz).

Die Merkmalsselektion und Redundanzuntersuchung wurde aus den vorherigen Schritten übernommen, um einen vergleichbaren Merkmalsraum zu nutzen. Die PCA wurde auf die gleichen Featureblöcke neu berechnet, da sich die Datenwolke der berechneten Merkmale verändert hat, wodurch sich neue Hauptkomponenten ergeben. Der resultierende Merkmalsraum wird von 183 Dimensionen aufgespannt. Die Ergebnisse der Klassifikation sind durch die Erfolgsraten mit Standardabweichung und BER über alle Klassifikatoren und beide Validierungsarten in Tabelle 6.6 und Tabelle 6.7 auf Seite 98 festgehalten. Als positives Resultat wird deutlich, dass die Klassifikationsergebnisse über die beschriebenen Verschiebungen stabil sind, da sich die Erfolgsraten und BER

nur geringfügig ändern. Allerdings muss beachtet werden, dass automatische Detektionsverfahren auch größere statistische Abweichungen haben könnten. Eine Überprüfung der Robustheit der Klassifikation in Abhängigkeit von der Standardabweichung der Verschiebung wäre daher von Vorteil. Mit dieser kann ein Grenzwert für die Unsicherheit angegeben werden, mit dem das in dieser Arbeit vorgestellte Verfahren noch eine robuste Erkennung der BP durchführen kann. Ein solcher Grenzwert würde die Auswahl automatischer Detektionsverfahren erleichtern.

Bei kleinen Unsicherheiten sind die Ergebnisse der Untersuchung stabil, da die Labelabhängigkeit des Merkmalsraumes durch die Merkmalsselektion bereits gesunken ist. Die Merkmalsblöcke der Normalenvektoren, für die eine starke Abhängigkeit von Markerpositionen zu erwarten sind, erfüllen das Filterkriterium nicht und werden bereits in vorverarbeitenden Schritten aussortiert. Für die Darstellung des Winkelverlaufs über definierte Spalten wird pro Spalte nur ein Marker für eine eindeutige Ortsdefinition genutzt, weswegen die Berechnung stark von den Verschiebungen des Markers abhängig und keine gute Vergleichbarkeit zwischen unterschiedlichen Bildern gegeben ist. Die Berechnung der Normalenvektoren an den Labelpositionen unterliegt ebenfalls starken Schwankungen, weshalb beide Methoden im Hinblick auf spätere Verwendung mittels automatischer Detektion ungeeignet sind. Bei Berechnung mittels Histogrammbildung sind die Regionen durch die Labelpositionen bestimmt. Jedoch wird in diesen Ansätzen der Verlauf der Häufigkeit der Winkel betrachtet und einzelne Verschiebungen der Markerpositionen beeinflussen nicht die berechneten Winkel selbst, sondern die Größe der ROI. Es kann also dazu führen, dass die betrachtete Region kleiner wird oder Hintergrundinformationen in die Betrachtung einfließen und diese verfälschen. Die Labelpositionen haben daher nur einen indirekten Einfluss auf die Histogramme.

Die Punktsignaturen sind sehr stark von den Markerpositionen abhängig und können bei starken Verschiebungen nicht als robust angenommen werden, da die Referenzpunkte der einzelnen Signaturen allein von einer Markerposition bestimmt wird. Eine Verbesserung könnte durch die Berechnung der Referenzpunkte aus dem Mittelpunkt mehrerer Markerpositionen erreicht werden. Die Verschiebung eines einzelnen Punk-

tes hat somit weniger starken Einfluss auf die Lage der Signatur. Ebenso könnte die Bestimmung der statischen Radien über größere Mittelwertbildung genutzt werden. Hier werden zwar bereits zwei Distanzen gemittelt; der Einfluss der Verschiebungen könnte gegebenenfalls vermindert werden, wenn sich die statistischen Verschiebungen gegenseitig aufheben.

7.3 Diskussion der Klassifikation

Die veränderlichen Klassifikationsleistungen über die Validierungsmethoden werden in Unterabschnitt 7.3.1 beschrieben. Die Erfolgsrate einzelner Klassen wird in Unterabschnitt 7.3.2 und die einzelner Probanden in Unterabschnitt 7.3.3 diskutiert.

7.3.1 Unterschiede durch Validierungsarten

Betrachtet man die Tabelle 6.2 und Tabelle 6.3, wird deutlich, dass die Erfolgsrate bei der 11-fachen Kreuzvalidierung niedriger und die BER höher im Vergleich zur Klassifikation mittels Hold-Out Methode ist. Im Datensatz sind immer mindestens sieben verschiedene Bilder der gleichen Person bei einer Übung vorhanden. Die Übungen wurden allerdings oftmals nicht neu ausgeführt, sondern es wurden mehrere Bilder aus einer Halteübungseinheit aufgenommen. Daher sind die Bilder in Position, Kopfhaltung und Übungsdurchführung sehr ähnlich und könnten als Kopien aufgefasst werden. Befinden sich einige der Bilder im Trainingsdatensatz des Klassifikators, werden die sehr ähnlichen Bilder im Testdatensatz eher als richtig klassifiziert. Dieser Effekt kann wie das Random-Oversampling¹ interpretiert werden, bei dem die Daten aus unterrepräsentierten Klassen zufällig dupliziert werden, um einen balancierten Datensatz zu erreichen. Allerdings steigt dabei das Risiko des Overfittings. Die Klassifikation mittels 11-facher Kreuzvalidierung erzielt somit realistischere Ergebnisse. Innerhalb der vorliegenden Arbeit sind bei dieser Validierungsmethode keine Bilder der Person, an der getestet wird, im Trainingsdatensatz vorhanden. Eine Überprüfung der erzielten Ergebnisse mittels Hold-Out Methode müsste mit einem angepassten Datensatz ge-

¹englisch, für zufällige Überabtastung

schehen. Dabei können die Bilder der Übungen zwar aus der gleichen Aufnahmeserie stammen; die Übungen sollten jedoch siebenmal neu positioniert und aus der Ruhelage ausgeführt werden. Man erreicht dadurch größere, aber auch realistischere Intravariabilitäten in den Übungsdurchführungen und ein Auswendiglernen des Klassifikators würde vermieden.

7.3.2 Klassifikationsleistungen bezüglich einzelner BP

Die Klassifikation in 19 Klassen mittels Naive Bayes Klassifikator ergibt mittels 11-facher Kreuzvalidierung die besten Klassifikationsergebnisse und wird im Folgenden genauer betrachtet. Für jede Klasse wird ein binäres Klassifikationsproblem gegenüber allen restlichen Klassen angenommen und Erfolgsrate und BER in Prozent angegeben. Die Ergebnisse sind in Tabelle 6.4 aufgelistet. Einen kompletten Überblick über die gesamte Klassifikation bildet die Konfusionsmatrix in Tabelle A.20 im Abschnitt A.4. Einige der Klassen werden mit einer BER von weniger als 20% erkannt, was im Vergleich mit anderen Übungen sehr gut ist. Hervorzuheben sind die Wangenübungen (inkl. Wangenboxen), welche mit elf bis 19% BER klassifiziert werden. Auffällig ist dabei, dass Daten mit realer Zugehörigkeit zu den Wangenübungen kaum zu anderen Klassen zugeordnet werden. Die Übungen werden allerdings untereinander einige Male falsch klassifiziert. So werden Übungen, bei denen einseitig eine Wange aufgeblasen ist, als beidseitige Wangenübung eingestuft. Ebenfalls gibt es Falschklassifikationen zwischen einseitigen Wangenübungen und den Übungen, bei denen die Zunge in die gleiche Wange gedrückt wird. Der Wangenbereich wird durch die ROI, in denen die Normalenvektoren berechnet werden, und durch verschiedene Punktsignaturen untersucht.

Schlechtere Ergebnisse bezüglich der beiden untersuchten Größen zeigt die Erkennung der bilateralen BP 3, also das lockere Schließen der Augen, welche eine Erfolgsrate von 13% und einen BER von 45% besitzt. Die Tabelle 7.1 zeigt einen Ausschnitt aus der gesamten Konfusionsmatrix über die Stirn- und Augenübungen.

Erkennbar sind hohe Werte neben der Hauptdiagonalen des Konfusionsmatrixaus-

	ermittelte Klassenzugehörigkeit			
tatsächliche Klassenzugehörigkeit	BP1	BP2	BP3	BP4
BP1	45	11	8	15
BP2	11	24	17	24
BP3	24	23	13	15
BP4	15	22	8	31

Tabelle 7.1: Auszug der Stirn- und Augenübungen aus der Konfusionsmatrix der Klassifikation mittels Naive Bayes Klassifikator über 11-fache Kreuzvalidierung. Prozentuale Angaben normiert auf die Anzahl genutzter Daten pro Klasse.

schnitts, welche die Prozentsätze der falsch klassifizierten Objekte bezüglich der Anzahl der Daten der realen Klassenzugehörigkeit angeben. So werden beispielsweise 24% der getesteten Daten der BP 2 (Zusammenziehen der Brauen) als BP 4 (beide Augen zusammenpressen) klassifiziert. Alle dargestellten BP weisen Bewegungen im Augen- und Stirnbereich auf, können allerdings nicht ausreichend voneinander differenziert werden. Zu beachten ist dabei, dass die Tiefenbilder keine Informationen über den Öffnungszustand der Augen tragen. Ebenso sind weniger starke Veränderungen des Stinbereichs durch Stirn- und Augenübungen im Tiefenbild zu sehen als in der unteren Gesichtshälfte bei den restlichen Motorik-Übungen. Der genutzte Merkmalsraum enthält zudem im Verhältnis zum unteren Gesichtsbereich kaum Merkmale, die Informationen über den Zustand der Stirn- und Augenbereiche enthalten. Die einzigen Merkmale, welche Informationen über den Stirnbereich erfassen, sind eine dynamische Punktsignatur am Referenzpunkt 27 sowie zwei statische am Referenzpunkt 28. Die drei Punktsignaturen sind in Abbildung 7.2 als Kreise angedeutet.

Ein Großteil der Veränderungen im Augen- und Stirnbereich sind im Tiefenbild nicht erkennbar. Die kleinste Tiefendistanz, welche im Aufnahmebereich detektierbar ist, liegt etwa bei 1-1.5mm [KHOSHELHAM und ELBERINK, 2012]. Falten zwischen den Augenbrauen, welche sich beim Zusammenschieben der Brauen in die Mitte ausprägen, sind im Tiefenbild zu erkennen, falls die Ausprägung tiefer als die kleinste detektierbare Tiefendistanz ist. Die Veränderungen, welche bei BP 1 auftreten (Augenbrauen hochziehen) und quer zur Muskulatur verlaufen, haben bei allen Probanden eine zu geringe Tiefe. Die Faltenbildung ist in der Tiefenbilddaufnahme nicht sichtbar und ist somit kein Informationsträger für die Klassifikation. Für die Augen- und Stirnübun-

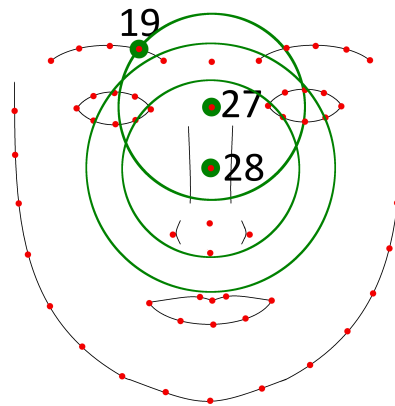


Abbildung 7.2: Der Stirnbereich wird durch anliegende Punktssignaturen untersucht. Eine Punktssignatur mit dynamischem Radius liegt an Referenzpunkt 27. Zwei Punktssignaturen mit unterschiedlichen statischen Radien liegen am Referenzpunkt 28.

gen ist was schlechte Klassifikationsergebnis aufgrund der fehlenden Informationen zu erwarten.

Zungenübungen werden meistens als solche klassifiziert, jedoch werden die Zungenstreckrichtungen oftmals nicht erkannt und als BP 11 eingeordnet, bei der die Zunge gerade herausgestreckt ist. In Tabelle 7.2 ist ein Ausschnitt aus der gesamten Konfusionsmatrix in Tabelle A.20 über die Zungenübungen gezeigt.

tatsächliche Klassenzugehörigkeit	ermittelte Klassenzugehörigkeit			
	BP11	BP12	BP13l	BP13r
BP11	47	17	1	2
BP12	18	66	0	0
BP13 l	21	1	46	0
BP13 r	23	1	4	37

Tabelle 7.2: Auszug von Zungenübungen aus der Konfusionsmatrix der Klassifikation mittels Naive Bayes Klassifikator über 11-fache Kreuzvalidierung. Prozentuale Angaben normiert auf die Anzahl eingehender Daten pro Klasse.

Die neunte ROI umfasst den Lippenbereich und umschließt bei geöffnetem Mund die komplette Mundhöhle. Die Normalenvektoren weisen eindeutige Veränderungen auf, wenn die Zunge herausgestreckt wird. Da der Ortsbezug über die Region allerdings verloren geht, kann die Zeigerichtung der Zungenspitze oft nicht erkannt werden. Die BP 12, bei welcher die Zungenspitze nach oben zeigt, wird teilweise mit BP 11 vertauscht. BP 13 wird bei beiden Streckrichtungen oftmals als BP 11 oder BP 7, bei der ein „A“ geformt wird, klassifiziert. Diese beiden Varianten sind auf die unterschiedliche Art der Durchführung der Übung „Zunge in Mundwinkel“ zurückzuführen. Wird die Zungenspitze direkt aus der L-Haltung bei geöffnetem Mund zum Mundwinkel geführt, ist die Zunge kaum sichtbar und die Erhebung am Mundwinkel im Tiefenbild ist nur sehr klein. Somit ähnelt die Haltung sehr stark der BP 7, was eine solche Falschklassifikation erklärt. Ist die Zunge während der Seitwärtsbewegung weit herausgestreckt, ist sie im Tiefenbild deutlich erfasst. Ebenso könnte man die unterschiedliche Ausführung durch verschieden starke Intensitätsstufen beschreiben. Die Ausführungsarten sind in Abbildung 7.3 verdeutlicht.

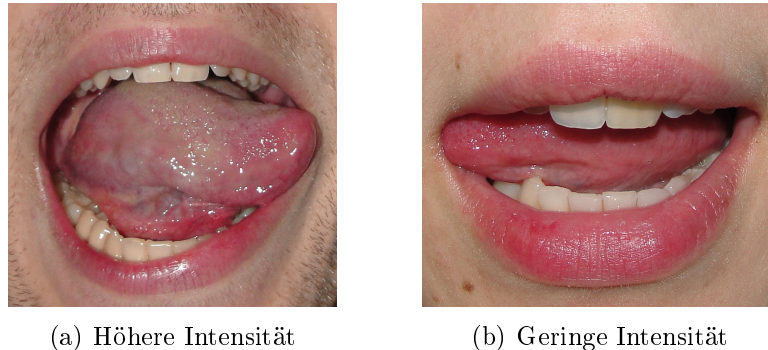


Abbildung 7.3: Unterschiedliche Intensitätsstufen oder Ausführungsarten der BP 13.

Die Punktsignaturen verlaufen meistens nicht direkt über die Mundöffnung. Alle Punktsignaturen mit dynamischem Radius tangieren die Oberlippe durch die Radiusbestimmung oder verlaufen im äußeren Drittel über die Mundöffnung. Dass die Punktsignaturen an den Referenzpunkten 29, 30 und 32 über den mittleren Teil der Mundöffnung verlaufen, ist möglich, aber nicht für jede Gesichtsform anzunehmen. Punktsignaturen, welche so definiert sind, dass sie immer quer bzw. längs über den Mund verlaufen,

würden die Klassifikation an dieser Stelle deutlich verbessern. Beispielsweise könnten Signaturen am Mundwinkel als Referenzpunkt gewählt werden. Der Referenzpunkt wäre dann zwar nicht robust, aber durch eine Ergänzung mit einem dynamischen Radius, der immer durch die Mundmitte verläuft, würde ein Informationsmehrwert erzielt. Eine Möglichkeit für die Berechnung des zweiten Referenzpunktes bestünde in der Mittelwertbildung beider Mundwinkelpositionen. Ebenso könnten zusätzliche Informationen über Farbhistogramme aus den korrespondierenden RGB-Bildern extrahiert werden und einen Informationsmehrwert für die Klassifikation der Zungenübungen liefern.

Weiterhin sollen drei BP gesondert betrachtet werden; der entsprechende Ausschnitt aus der Konfusionsmatrix ist in Tabelle 7.3 dargestellt. Diese Gruppe umfasst das BP 5, 8 und 10, bei der die Lippen zu einem Kussmund, zu einem „O“ oder zu einem „U“ geformt sind. Diese BP werden häufig untereinander, aber selten mit anderen BP vertauscht.

	ermittelte Klassenzugehörigkeit		
tatsächliche Klassenzugehörigkeit	BP5	BP8	BP10
BP5	46	9	34
BP8	23	28	27
BP10	41	17	27

Tabelle 7.3: Auszug aus der Konfusionsmatrix der Übungen „OForm“, „UForm“ und „Kussmund“. Klassifikation erfolgt mittels Naive Bayes Klassifikator über 11-fache Kreuzvalidierung. Prozentuale Angaben normiert auf die Anzahl eingehender Daten pro Klasse.

Alle drei BP haben relativ hohe BER von 29% bis 38% und im Vergleich zu anderen niedrige Erfolgsraten. Die Falschklassifikationen zwischen diesen drei Übungen waren zu erwarten, da sie sehr ähnliche Formen aufweisen. Das geformte „O“ und „U“ können selbst im RGB-Bild ohne eine Lautbildung schwer auseinander gehalten werden (siehe Abbildung 7.4).

Die BP 5 kann von beiden anderen nur durch den Lippenschluss unterschieden werden. Oftmals ist die Öffnung bei den BP 8 und 10 sehr klein und somit nicht im Tiefenbild zu erkennen. Mittels Kinectsensor kann eine kleine Mundöffnung aufgrund von

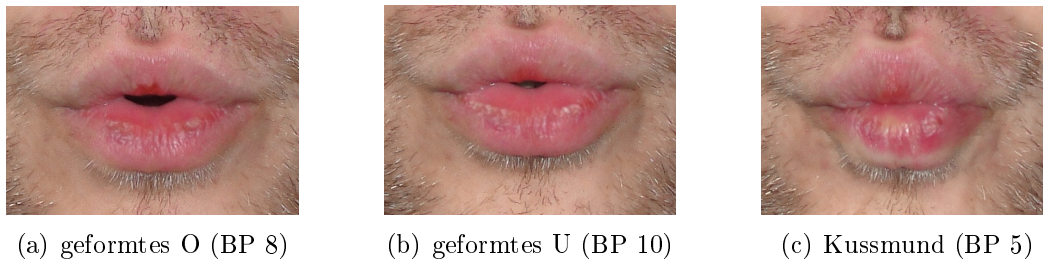


Abbildung 7.4: Direkter Vergleich der ähnlich geformten Lippen zu einem „O“, „U“ und zu einem Kussmund.

Verdeckungen und Schattierungen nicht erfasst werden. Eine beispielhafte Tiefenbildaufnahme der Durchführung der Vokalübung, bei der die Lippen zu einem „O“ geformt werden, ist in Abbildung 7.5 dargestellt. In der Aufnahme ist keine Öffnung zwischen den Lippen erkennbar.

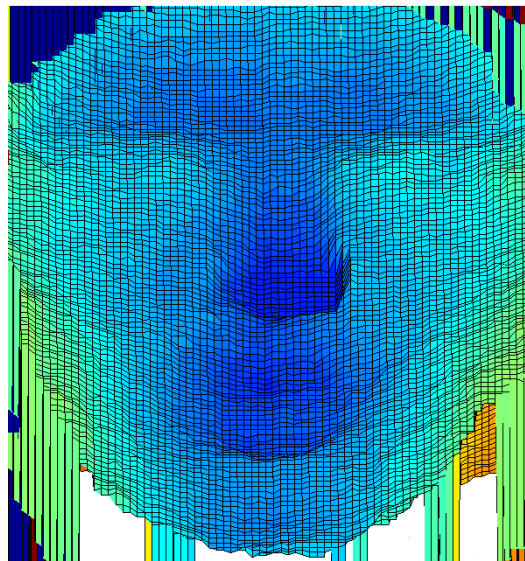


Abbildung 7.5: Tiefenbilddarstellung der Gesichtsmotorik-Übung „OForm“, bei welcher die BP 8 aktiviert ist. Obwohl eine Öffnung zwischen den Lippen vorhanden ist, wird diese mittels Kinectsensor nicht erfasst.

Im Gegensatz dazu werden die restlichen Vokalübungen sehr gut mit einer BER von 12% und 13% klassifiziert. Ebenso wird die Übung „Lippen breitziehen“ mit einer BER von 16% klassifiziert. Eine Wechselübung zwischen den gegenläufigen Vokalübungen,

beispielsweise zwischen „IForm“ und „OForm“, würde daher aufgrund der Integration von Vorwissen gut klassifiziert werden. Das Gleiche gilt für den dynamischen Wechsel zwischen breitgezogenen Lippen und einem Kussmund. Durch das Einbinden dieses Vorwissens bei Wechselübungen kann eine Klassifikation der Übungen in die BP mit einer guten Erfolgsrate funktionieren.

7.3.3 Abhängigkeit der Klassifikationsleistung vom Datensatz

Anhand der Erfolgsraten in Tabelle 6.5 über die einzelnen Schritte der 11-fachen Kreuzvalidierung wird eine Abhängigkeit der Klassifikationsleistung von verschiedenen Personen deutlich. Auch wenn bei einem kleinen Datensatz über einige Probanden gute Ergebnisse erzielt werden, kann es nicht auf alle Gruppen übertragen werden. Ein sicheres Ergebnis für die Klassifikation wird nur bei Steigerung der Anzahl der Probanden im Datensatzes erreicht, sodass der Klassifikator in den einzelnen Validierungsschritten über mehr Personen trainiert wird. Keine der Testpersonen hat myofunktionelle Einschränkungen im Gesichtsbereich. Zudem befinden sich alle Probanden in der Altersgruppe von 20 bis 35 Jahren. Daher werden stärkere Schwankungen der Erfolgsrate zwischen den verschiedenen Personen erwartet, wenn der Datensatz durch einzelne Probanden erweitert wird, die deutlich jünger oder älter sind. Ein stabiles Ergebnis kann daher nur über eine große und repräsentative Gruppe über Alter, Geschlecht, ethnische Zugehörigkeit usw. erreicht werden.

Kapitel 8

Zusammenfassung und Ausblick

Bei Störungen der mimischen oder orofazialen Muskulatur werden Gesichtsmotorik-Übungen als Teiltherapie eingesetzt. Gerade Patienten mit orofazialer Dysfunktion oder Beeinträchtigung des Schluck- und Sprachvorgangs verursacht durch neurologische Störungen müssen diese Übungen regelmäßig durchführen, um einen therapeutischen Erfolg zu erzielen. Es gibt bisher keine trainingsunterstützenden Module, welche dem Patienten ein Feedback über die Qualität der Übungsdurchführung geben. Durch ein Tiefenkamerasystem könnte auch zu Hause eine Übungsüberwachung ermöglicht werden, welche durch Integration von spielerischen Komponenten zusätzlich motivierende Faktoren aufweisen kann. Um grundlegende Schritte für ein solches System zu schaffen, wurde in der vorliegenden Masterarbeit zunächst die Klassifikationsmöglichkeit ausgewählter Übungen mittels Tiefeninformationen anhand von bestimmten Merkmalen untersucht. Die wesentlichen Schritte der Mustererkennung sind dafür auf die bestehende Problematik angewendet worden, was in Abbildung 8.1 zusammengefasst dargestellt ist.

Zunächst wurde ein Übungskatalog mit 19 Gesichtsmotorik-Übungen mit Hinblick auf bestimmte Krankheitsbilder und den Anwendungsbereich erstellt. Diese Übungen sind in Anlehnung an das FACS in entsprechende BP eingeteilt worden. Die BP geben gleichzeitig die Definition der Klassen, in die klassifiziert wird. Nach der Auswahl des

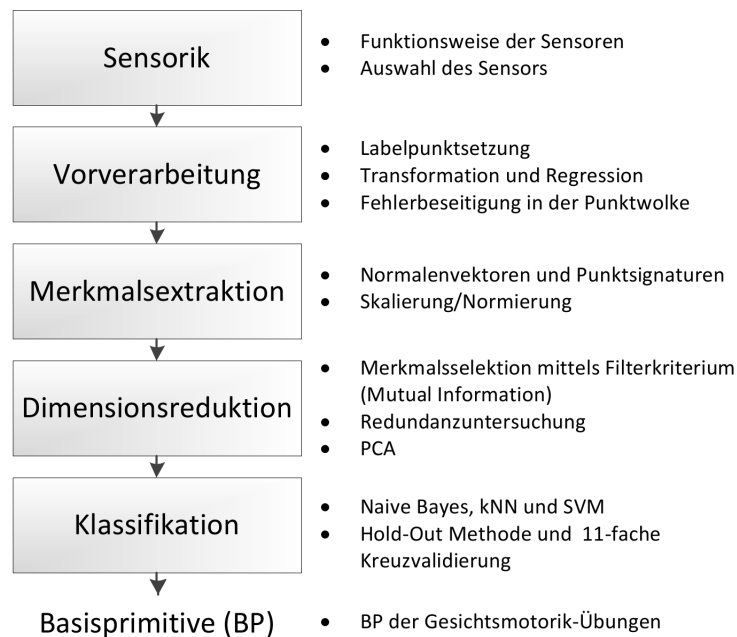


Abbildung 8.1: Zusammenfassender Überblick über die Module der Klassifikation der Gesichtsmotorik-Übungen und den darin durchgeführten Schritten in der vorliegenden Arbeit.

Sensors wurde ein Aufnahmeszenario festgelegt und ein Datensatz aus 1480 Bildern von elf verschiedenen Personen aufgenommen, welcher nach entsprechenden Vorverarbeitungsschritten für die Merkmalsextraktion bereitgestellt wird.

Es wurden Normalenvektoren und Punktsignaturen in unterschiedlichen Varianten in MATLAB umgesetzt und aus den Bildern extrahiert. Dabei ist die Normalenvektorberechnung über verschiedene Möglichkeiten der Ortsdefinition für die Berechnung variiert worden. Zusätzlich wurden unterschiedliche Standardisierungsverfahren der Merkmalsvektoren auf eine einheitliche Dimension eingesetzt. Die Grundberechnung der Normalenvektoren blieb dabei über alle Variationen gleich. Die Normalenvektoren sind an der manuell gesetzten Labelpunktmenge berechnet und die Verläufe der sphärischen Koordinaten der Normalenvektoren über definierte Spalten der Tiefenbildmatrix erstellt worden. Zudem wurden Histogramme in zehn definierten ROI über Winkelintervalle und -kombinationen ermittelt. Bei der Umsetzung der Punktsigna-

turen wurden zwei Möglichkeiten zur Radiusbestimmung unterschieden. Der Radius ist entweder über eine statische Größe aus der Augenbreite der jeweiligen Probanden abgeleitet oder durch die euklidische Distanz zwischen zwei angegebenen Labelpunkten gebildet worden. Zusätzlich wurden die Ortsdefinition und die Radiusgröße über alle Bilder mehrfach variiert. Daraus ergab sich ein hochdimensionaler Merkmalsraum, dessen Dimension mittels Merkmalsselektion vermindert wurde. Durch Featureranking und einem Kriterium basierend auf der MI wurde eine Merkmalsteilmenge selektiert. An dieser selektierten Teilmenge wurde eine Redundanzuntersuchung anhand der paarweisen MI vorgenommen und eventuell blockweise redundante Merkmale entfernt. Haben die zusammengehörigen Werte der einzelnen Berechnungsarten, der Featureblöcke, untereinander hohe Redundanzen aufgezeigt, wurde eine PCA durchgeführt, um den Informationswert einzelner Merkmale zu steigern und die Dimension des Merkmalsraumes weiter zu senken.

Die Klassifikation wurde mit SVMs, dem kNN- und dem Naive Bayes Klassifikator durchgeführt, welche mittels Hold-Out Methode und der 11-fachen Kreuzvalidierung evaluiert wurden. Aufgrund des Datensatzaufbaus sind die etwas niedrigeren Erfolgsraten der 11-fachen Kreuzvalidierung gegenüber der Hold-Out Methode als realistischer anzusehen und wurden daher weiter betrachtet. Anhand einer Erfolgsrate von 53% ergab sich das beste Klassifikationsergebnis für alle 19 Klassen am reduzierten Merkmalsraum mittels Naive Bayes Klassifikator. Bei der Untersuchung der TPR und der BER bezüglich der einzelnen Klassen ergab sich, dass Augen- und Stirnübungen durch sehr kleine Unterschiede im Tiefenbild schlechter klassifiziert werden. Die Ergebnisse der Klassifikation der Wangenübungen erwiesen sich mit einer BER kleiner als 20% als die besten. Sehr ähnliche Übungen, wie die Lippenübung „OForm“ und „UForm“, sind bei alleiniger Nutzung der Merkmale nicht zu trennen, da sie selbst im RGB-Bild zu starke Ähnlichkeiten aufweisen. Im Tiefenbild wird zudem die kleine Öffnung zwischen den Lippen bei beiden Übungen nicht erfasst, wodurch eine zusätzliche Ähnlichkeit zu der Übung „Kussmund“ auftritt. Allerdings könnten diese im Wechsel mit bestimmten Halteübungen durch Integration von Vorwissen bessere Klassifikationsergebnisse erzielen. Entsprechende Gegenspieler, wie beispielsweise die Übungen „AForm“, „IForm“

und „Lippen breitziehen“, besitzen ebenfalls eine BER kleiner als 20%. Es ist festzuhalten, dass eine Klassifikation mittels Tiefendaten in die BP mit einigen Einschränkungen gut funktioniert. Durch weitere Anpassung der Merkmale, des Datensatzes und entsprechender Parameter der Merkmalsselektion und Klassifikation ist eine deutliche Steigerung der Erfolgsrate und eine Verminderung der BER zu erwarten.

Sehr schlecht klassifizierte Übungen, zum Beispiel das lockere Augenschließen, können aus dem Datensatz ausgeschlossen werden, um die Klassifikationsrate zu verbessern. Im Hinblick auf diese Übung, die kaum eine Veränderung des Tiefenbildes verursacht und für eine bessere Erkennung der Streckrichtung der Zunge ist eine Erweiterung mit zweidimensionalen Informationen aus Farbbildern denkbar. Um die Informationswerte der einzelnen Features zu steigern, können Untersuchungen bezüglich veränderlicher Parameter für die Extraktion durchgeführt werden. Vor der Berechnung der Normalenvektoren sollten zunächst die Ortsangaben innerhalb der Bildebene in Meter umgerechnet werden, um besser interpretierbare Winkel zu erhalten. Dies wurde bereits in Abschnitt 7.1 erläutert und in [KHOSHEHAM und ELBERINK, 2012] durchgeführt. Ebenso könnte der Einfluss der Nachbarschaftsgröße auf die Klassifikationsergebnisse bzw. auf die MI-Werte der berechneten Merkmale untersucht werden, wie es in [CHEW et al., 2010] vorgestellt wird. Dadurch kann eine optimale Nachbarschaftsgröße als Kompromis zwischen Rauscheinfluss und Glättungseffekt bei der Berechnung der Normalen gefunden werden. Bei der Berechnung der Punktsignaturen kann die Abtastrate der Schnittkurve erhöht werden, um keinen Informationsverlust durch einen zu klein gewählten Winkel zu verursachen. Redundante Daten durch mehrfaches Abtasten gleicher Pixel können durch eine anschließende Dimensionsreduktion mittels DCT des Funktionsverlaufes der Punktsignatur ausgeglichen werden. Ebenso könnte die Berechnung an jedem Datenpunkt des Tiefenbildes erfolgen, um erst mittels Featureranking die besten für die entsprechende Klassifikation zu selektieren.

Die Featureselektion kann für eine bessere Dimensionsreduktion aus verschiedenen Ansätzen kombiniert werden und somit bessere Merkmalsräume für die Klassifikation erzielen. Für die hochdimensionalen Merkmalsräume, wie sie in diesem Fall vorlie-

gen, wäre eine kombinierte Lösung aus Filterverfahren und anschließender Suche eines optimalen Subsets am bereits vorselektierten Merkmalsraum möglich. Die Generalisierbarkeit der Klassifikatoren könnte durch optimale Parameter verbessert werden. Dafür ist beispielsweise die Anzahl der Nachbarn und die Art der metrischen Distanz bei dem kNN-Klassifikator zu untersuchen. Ebenfalls ist eine gewichtete Kombination der drei bereits genutzten Klassifikatoren mit einer Mehrheitsentscheidung möglich, um die Erfolgsrate zu steigern. Untersuchungen für eine optimale Klassifikatorstruktur sind in einer weiterführenden Arbeit vorzunehmen. Die Robustheit der Klassifikation kann hinsichtlich möglicher Verdeckungen im Gesicht, wie beispielsweise durch Brillen und Haarsträhnen, untersucht werden, um eventuelle Vorgaben für ein Trainingssystem geben zu können. Ebenso sollte das Klassifikationsergebnis über einen repräsentativen Datensatz stabil sein. Daher ist eine Erweiterung des bestehenden Datensatzes durch Probanden unterschiedlichen Alters, Geschlechts, ethnischer Herkunft oder ähnliches, notwendig.

Weitere Schritte für die Entwicklung eines trainingsunterstützenden Systems sind die Ausarbeitung und Umsetzung von Evaluationskriterien für die Durchführung der Gesichtsmotorik-Übungen. Dabei ist die Ausweitung der Untersuchungen auf dynamische Bilder sinnvoll, um a priori Wissen über die Ausgangslage der Übung zu erhalten. Im Hinblick auf die dynamische Übungsbewertung ist eine Erweiterung des BP-Systems mit Durchführungsintensitäten sinnvoll.

Zusätzlich ist für eine automatische Feedbackkomponente ein Zusammenführen der Klassifikationsansätze mit einer automatischen Detektion von Labelpunkten aus dem Teilgebiet der Mimikerkennung notwendig. Dafür ist zunächst die Untersuchung der Robustheit des vorhandenen Systems auf Detektionsunsicherheiten voran zu stellen. Eine Integration von Projekten der Computeranimation wäre für die Erweiterung eines bewertenden Moduls um Avatare vorstellbar.

Anhang A

Ergänzende Unterlagen

A.1 Dysarthriearten

Dysarthrieform	Lokalisation der Läsion	Bemerkung	mögliche Ursache
kortikal	motorische Hirnrindengebiete, motorische Ursprungsgebiete beider Hemispären	selten, da beidseitige Felder betroffen sind, meist mit gekoppelter Aphasie	Zerebrale Gefäßsyndrome, Schlaganfall
bulbär, hypoton, peripher	Hirnstamm, Motorische Hirnnerven in Medulla oblongata und Pons	hypotones Sprechbewegungsmuster, Bulbärsprache	Bulbärparalyse, ALS (amyotrophe Lateralsklerose)
zerebellar, ataktisch	Kleinhirn (Cerebellum)	Fehlkoordinierte Sprechbewegung, ataktische Sprechmotorik	Multiple Sklerose
extrapyramidal	extrapyramidales System, Stammganglien	hyperkinetische und rigide, hypokinetische	Chorea Huntington, Morbus Parkinson
pyramidal suprabulbär supranuklear pseudobulbär	Pyramidenbahn	schwerfällig, harte bis spastische Artikulation	Pseudobulbärparalyse, hirnatrophi- scher Prozess

Tabelle A.1: Dyarthriearten, geordnet nach Orten der verursachenden Läsionen¹

¹Vgl. Tabellen in [BRAUN, 1999, S. 150] und [JÖRG, 1997, S. 126] und Definitionen in [FRANKE, 2008, S. 59ff]

A.2 Basisprimitive

Nr.	Übungs- label	AU	Innervierte Muskeln	Basisprimitive
X: Weitere unbekannte AU				
1	Stirn	1+ 2	Venter frontalis (Teil des Occipitofrontalis)	BP1 (bilateral)- Augenbraue heben
2	Brauen	4	Corrugator supercilii, Procerus, Depressor supercilii (Teil des Orbicularis Oculi)	BP2 - Augenbrauen zusammenschieben
3	Auge	43	Orbicularis Oculi (pars palpebralis), Entspannung des Levator palpebrae superioris	BP3 (bilateral)- Auge entspannt geschlossen
4	Augenpressen	44+6	Orbicularis Oculi (pars palpebralis und pars orbitalis)	BP4 (bilateral)- Auge zusammengepresst
5	Kuss	18	Orbicularis Oris, Incisivus labii superioris und inferioris	BP5 - Lippen spitzen
6	Breit	14+ 20	Buccinator, Risorius	BP6 - Mundwinkel nach außen ziehen
7	AForm	25F+ 27	Entspannung Mentalis und Orbicularis Oris, Digastricus Pterygoideus medialis	BP7 - A formen
8	OForm	18+ 25+ X	Orbicularis Oris, Incisivus labii superioris und inferioris, Entspannung mentalis	BP8 - geöffneter runder Mund
9	IForm	25+ 14	Entspannung mentalis, (Buccinator), Risorius	BP9 - geöffneter Mund und Mundwinkel nach außen
10	UForm	25+ X	Entspannung Mentalis und Orbicularis Oris	BP10 - U formen

Tabelle A.2: Basisprimitive für die Klassifizierung der Augen-, Stirn- und Lippenübungen; Anpassung des FACS an die ausgewählten Übungen.

Nr.	Übungs- label	AU	Innervierte Muskeln	Basisprimitive
k.A: keine Angabe				
11	Zunge	19	Beidseitig Genioglossus, Transversus linguae, Verticalis linguae	BP11- Zunge gerade herausstrecken
12	ZungeO	19	Beidseitig Genioglossus, beide longitudinalis superioris	BP12- Zunge raus und nach oben
13	ZungeLi	19	Beidseitig Genioglossus, je einseitig longitudinalis superioris	BP13(bilateral)- Zunge in Mundwinkel
14	ZungeRe	19		
15	BoxenLi	k.A	Mentalis, Orbicularis Oris, einseitig Styloglossus, je einseitig longitudinalis superioris	BP14(bilateral)- Wange ausbeulen
16	BoxenRe	k.A		
17	Wangen	k.A	Orbicularis Oris und innerer Druck durch Luft	BP15- Wangen aufblasen
18	WangenLi	k.A	Orbicularis Oris, Buccinator einseitig, innerer Druck durch Luft	BP16(bilateral)- Einseitig Wange aufblasen
19	WangenRe	k.A		

Tabelle A.3: Basisprimitive für die Klassifizierung der Zungen- und Wangenübungen; Anpassung des FACS an die ausgewählten Übungen.

A.3 Übungskatalog der Gesichtsmotorik-Übungen

Im Folgenden ist ein tabellarischer Überblick über die ausgewählten Übungen gegeben. Das entsprechende Basisprimitiv sowie der Übungsname und das jeweilige Kürzel der Übungen sind beigefügt. Neben Durchführungshinweisen und Zielen, die mit den

Übungen erreicht werden sollen, sind die Bücher aufgeführt, in denen die Übungen angegeben sind. Die Informationen sind aus den angegebenen Büchern zusammengetragen. Das Kürzel „BL“ steht für Übungen, die von Mitarbeitern der Fachklinik Bad Liebenstein aus dem Bereich der Sprachtherapie empfohlen wurden.


BP 1	Stirn runzeln/ <i>Stirn</i>	
Quellen	BL [BENECKE, 1998, S.17]	
Durchführung	Stirnrunzeln durch Hochziehen der Augenbrauen- statisch; Im Wechsel mit der Ruhelage- dynamisch	
Hinweise	-	
Ziele	-	

Tabelle A.4: Übungsübersicht Teil 1: Stirnrunzeln


BP 2	Augenbrauen zusammenziehen/ <i>Brauen</i>	
Quellen	BL; [BENECKE, 1998, S.17]	
Durchführung	Zusammenschieben der Augenbrauen- statisch; Im Wechsel mit der Ruhelage- dynamisch	
Hinweise	-	
Ziele	-	

Tabelle A.5: Übungsübersicht Teil 2: Augenbrauen zusammenziehen


BP 3		Augen schließen/ <i>Augen</i>
Quellen	BL	
Durchführung	Beide Augen gleichzeitig und locker schließen	
Hinweise	<ul style="list-style-type: none"> • Gleichzeitiges Schließen der Augen • Nur Augenlidbewegung • Keine weiterer Muskulatur zur Schließung 	
Ziele	-	

Tabelle A.6: Übungsübersicht Teil 3: Augen schließen


BP 4		Augen zusammenpressen/ <i>Augenpressen</i>
Quellen	BL	
Durchführung	Beide Augen gleichzeitig zusammendrücken	
Hinweise	-	
Ziele	-	

Tabelle A.7: Übungsübersicht Teil 4: Augen zusammenpressen


BP 5	Lippen spitzen oder Kussmund/ <i>Kuss</i>	
Quellen	BL; [THIELE, 1997, S.108]; [KITTEL, 1998, S.81]; [CLAUSNITZER, 2004, S.59]; [BENECKE, 1998, S.22]; [PROSIEGEL und WEBER, 2010, S.154]	
Durchführung	Ausgangspunkt ist die LM-Haltung, Lippen spitz formen, dabei Lippen zusammengepresst- statisch; Im Wechsel mit LM-Haltung, mit breit gezogenen Lippen oder mit „OForm“- dynamisch	
Hinweise	<ul style="list-style-type: none"> • Lippen bleiben geschlossen • Zahnreihen leicht geschlossen und Zunge in Ruhelage • Hilfe: Mundwinkel in die Mitte schieben 	
Ziele	<ul style="list-style-type: none"> • Als Halteübung zur Tonussteigerung • Gewebedehnung bei Mundschlussproblemen • Stärkung der Muskulatur • Unabhängigkeitsförderung der inter- und extraoralen Muskulatur • Lippenrundung 	

Tabelle A.8: Übungsübersicht Teil 5: Lippen spitzen


BP 6	Lippen breitziehen oder Mundwinkelzugübung/ <i>Breit</i>	
Quellen	BL; [THIELE, 1997, S.30]; [KITTEL, 1998, S.81]; [CLAUSNITZER, 2004, S.67]; [BENECKE, 1998, S.22]; [PROSIEGEL und WEBER, 2010, S.154]	
Durchführung	Ausgangspunkt ist die M-Haltung, Mundwinkel seitwärts ziehen- statisch; Im Wechsel mit anderen Lippenhaltungen bzw. Orbikularisübungen- dynamisch	
Hinweise	<ul style="list-style-type: none"> •Lippen bleiben geschlossen •Mundwinkel nicht nach oben/unten ziehen •Erweiterung, einzelnes seitliches Ziehen der Mundwinkel 	
Ziele	<ul style="list-style-type: none"> •Streckung der Ringmuskulatur der Orbikularismuskeln •Tonussteigerung im Gürtel bei verminderter Wangen- Lippenmuskulatur 	

Tabelle A.9: Übungsübersicht Teil 6: Lippen breitziehen

BP 7	Vokalübung- „A“ formen/ <i>AForm</i>	
Quellen	BL, [THIELE, 1997, S.7]; [BENECKE, 1998, S.21]	
Durchführung	Mund weit öffnen, „A“ sagen und Form halten- statisch; Im Wechsel mit der M-Haltung oder anderen Vokalübungen- dynamisch	
Hinweise	Wangenmuskulatur bleibt bei Übungsdurchführung entspannt	
Ziele	<ul style="list-style-type: none"> •Streckung der Orbikularismuskeln •Tonussteigerung im Gürtel bei verminderter Wangen- und Lippenmuskulatur 	

Tabelle A.10: Übungsübersicht Teil 7: „A“ formen


BP 8	Vokalübung- „O“ formen/ <i>OForm</i>	
Quellen	BL; [KITTEL, 1998, S.82]; [THIELE, 1997, S.169]; [BENECKE, 1998, S.21]	
Durchführung	Mundspalt von allen Seiten gleichförmig schließen bis kleiner offener Kreis- statisch; Im Wechsel mit Ausgangslage (M-Haltung) oder anderen Vokal- bzw. Orbikularisübungen- dynamisch	
Hinweise	<ul style="list-style-type: none"> •Zahnreihen nicht schließen •Beide Lippen aufeinander zu bewegen 	
Ziele	<ul style="list-style-type: none"> •Aktivierung und Verlängerung der Oberlippenmuskulatur •Anwendung bei insuffizienter Muskulatur •Verstärkung der zentralen Zirkulärfasern 	

Tabelle A.11: Übungsübersicht Teil 8: „O“ formen


BP 9		Vokalübung- „I“ formen/ <i>IForm</i>
Quellen	BL; [PROSIEGEL und WEBER, 2010, S.154]; [THIELE, 1997, S.84,148]	
Durchführung	Mundwinkel mit leicht geöffnetem Mund nach außen ziehen und „I“ sagen, Form halten- statisch; Im Wechsel mit Ruhelage oder anderen Vokalübungen- dynamisch	
Hinweise	<ul style="list-style-type: none"> •Zahnreihen nicht schließen •Breitziehen der Lippen ohne Anspannung der Halsmuskulatur 	
Ziele	<ul style="list-style-type: none"> •Aktivierung und Verlängerung der Oberlippenmuskeln •Anwendung bei insuffizienter Muskulatur •Verstärkung der zentralen Zirkulärfasern 	

Tabelle A.12: Übungsübersicht Teil 9: „I“ formen


BP 10		Vokalübung- „U“ formen/ <i>UForm</i>
Quellen	BL; [PROSIEGEL und WEBER, 2010, S.154]; [BENECKE, 1998, S.21]	
Durchführung	Mit leicht geöffnetem Mund „U“ sagen und Form halten- statisch; Im Wechsel mit der Ruhelage und anderen Vokalübungen- dynamisch	
Hinweise	-	
Ziele	-	

Tabelle A.13: Übungsübersicht Teil 10: „U“ formen


BP 11	Zunge herausstrecken/ <i>Zunge</i>	
Quellen	BL; [BENECKE, 1998, S.22]; [CLAUSNITZER, 2004, S.98]; [KITTEL, 1998, S.71]; [THIELE, 1997, S.302]; [PROSIEGEL und WEBER, 2010, S.154]	
Durchführung	Zunge so weit wie möglich gerade und horizontal nach draußen strecken und haltenstatisch; Im Wechsel mit der L-Haltung oder weiteren Streckrichtungen-dynamisch	
Hinweise	<ul style="list-style-type: none"> • Weit geöffneter Mund • Strecken der Zunge bis ein Ziehen im Muskel spürbar • Zunge nicht auf Lippen oder Zahnreihe auflegen • Zungenspitze nicht nach oben/unten zeigen • Verstärkung der Anspannung durch Widerstand möglich 	
Ziele	<ul style="list-style-type: none"> • Aktivierung und Stärkung des Zungenmittelteils • Stabilisierung der Binnenmuskulatur • Förderung der Zungenwahrnehmung für unterschiedliche Lagen • Stärkung der transversalen Zungenmuskeln • Förderung der Zungenprotraktion 	

Tabelle A.14: Übungsübersicht Teil 12: Zunge gerade herausstrecken

BP 12	Zunge zur Nase/ <i>ZungeO</i>	
Quellen	BL; [BENECKE, 1998, S.23]; [KITTEL, 1998, S.64]; [CLAUSNITZER, 2004, S.100]; [PROSIEGEL und WEBER, 2010, S.155]	
Durchführung	Zunge so weit wie möglich rausstrecken, dabei die Zungenspitze zur Nase- statisch; Im Wechsel mit der L-Haltung oder als Koordinationsübung mit anderen Streckrichtungen-dynamisch	
Hinweise	<ul style="list-style-type: none"> • Unterkiefer soll Zunge nicht hochschieben • Weit geöffneter Mund 	
Ziele	<ul style="list-style-type: none"> • Aktivierung der Seitwärts- und Drehbewegung der Zunge • Flexibilität- und Mobilitätssteigerung der Longitudinalfasern 	

Tabelle A.15: Übungsübersicht Teil 12: Zunge zur Nase


BP 13	Zunge in linken bzw. rechten Mundwinkel/ <i>ZungeLi/ ZungeRe</i>	
Quellen	BL; [BENECKE, 1998, S.22]; [CLAUSNITZER, 2004, S.98]; [PROSIEGEL und WEBER, 2010, S.155]	
Durchführung	Ausstrecken der Zunge zum rechten/linken Mundwinkel und halten- statisch; Wechsel der Seiten- dynamisch	
Hinweise	<ul style="list-style-type: none"> •Temposteigerung möglich •Zungenspitze in Mundwinkel legen •Direkter Weg zum Mundwinkel •Variation durch Kreisbewegung 	
Ziele	<ul style="list-style-type: none"> •Aktivierung der Seitwärts- und Drehbewegung der Zunge •Flexibilität- und Mobilitätssteigerung der Longitudinalfasern 	

Tabelle A.16: Übungsübersicht Teil 13: Zunge in den Mundwinkel


BP 14	Zunge in linke bzw. rechte Wange bohren/ <i>BoxenLi/BoxenRe</i>	
Quellen	BL; [BENECKE, 1998, S.23]; [THIELE, 1997, S.267]; [CLAUSNITZER, 2004, S.52]	
Durchführung	Zunge zwischen den Zähnen hindurch von innen gegen die Wange pressen- statisch; Wechsel der Seiten- dynamisch	
Hinweise	<ul style="list-style-type: none"> •Einseitige Halteübung •Retrales Pressen der Zunge mit aller Kraft •Evtl. bei Wechsel der Seite Pausen in Ruhelage einlegen 	
Ziele	<ul style="list-style-type: none"> •Muskeltraining bei Ausfall der Zunge •Geringer Streckeffekt auf die Wange •Flexibilitäts- und Mobilitätssteigerung der Longitudinalfasern 	

Tabelle A.17: Übungsübersicht Teil 14: Zunge in Wange bohren


BP 15	Wangen aufblasen/ <i>Wangen</i>	
Quellen	[THIELE, 1997, S.22]; [BENECKE, 1998, S.24]; [CLAUSNITZER, 2004, S.56]; [PROSIEGEL und WEBER, 2010, S.154]	
Durchführung	Luftdruck im Mund erhöhen, dabei keine Luft durch den Lippenspalt entweichen lassen und halten- statisch	
Hinweise	<ul style="list-style-type: none"> •Ruhige Nasenatmung •Lippen zusammenpressen •Ziehen in Wangenmuskulatur wird spürbar •Bei lockerer Wangenmuskulatur Dehn- bzw. Entspannungsübung 	
Ziele	<ul style="list-style-type: none"> •Hebung des Gaumensegels •Verbesserung des Lippenschluss •Strecken/Dehnen der Wangenmuskulatur •Korrektur bei Verkrampfungen •Bei Eigenmanipulation kann Tonussteigerung des Orbikularis und des Velums erreicht werden 	

Tabelle A.18: Übungsübersicht Teil 16: Beide Wangen aufblasen


BP 16	Linke bzw. rechte Wange aufblasen/ <i>WangeLi</i> / <i>WangeRe</i>	
Quellen	[CLAUSNITZER, 2004, S.56]; [PROSIEGEL und WEBER, 2010, S.153]; [THIELE, 1997, S.23]	
Durchführung	Aufblasen einer Wange durch Kontraktion der anderen- statisch; Wechsel der Seiten, dynamisch	
Hinweise	<ul style="list-style-type: none"> •Bei Durchführung ruhige Nasenatmung •Leichter Schluss der Zahnreihen vermeidet Kieferbewegung beim Seitenwechsel, kein pressen 	
Ziele	Stärkung der Wangenkontraktionsfähigkeit	

Tabelle A.19: Übungsübersicht Teil 15: Eine Wange aufblasen

A.4 Ergebnisse der Klassifikation

Im Folgenden sind die vollständigen Konfusionsmatritzen der Klassifikationen mittels SVM, kNN- und Naive Bayes Klassifikator aufgezeigt. Die Klassifikationen wurden mit zwei verschiedenen Validierungsarten durchgeführt: Zum einen durch eine Hold-Out Methode, wobei die Hälfte der Daten zur Validierung genutzt wurden, und zum anderen durch eine 11-fache Kreuzvalidierung. Bei der 11-fachen Kreuzvalidierung wurde der Datensatz personenweise aufgeteilt.

Vorhergesagte Klassenzugehörigkeit																				
	BP1	BP2	BP3	BP4	BP5	BP6	BP7	BP8	BP9	BP10	BP11	BP12	BP13 Li	BP13 r	BP14 l	BP14 r	BP15	BP16 l	BP16 r	
Wahre Klassenzugehörigkeit	BP1	45	11	8	15	3	1	0	4	0	0	0	0	0	1	10	0	0	0	
	BP2	11	24	17	24	2	0	0	3	0	0	0	0	0	6	6	0	3	0	
	BP3	24	23	13	15	0	3	0	0	1	0	5	0	0	5	9	0	0	1	
	BP4	15	22	8	31	1	12	0	0	1	0	1	0	0	6	1	0	0	0	
	BP5	1	0	4	3	46	0	0	9	0	34	0	0	0	3	1	0	0	0	
	BP6	1	5	4	6	0	71	0	0	9	0	0	0	0	0	0	0	1	3	
	BP7	2	4	0	0	0	0	79	5	0	1	4	0	1	5	0	0	0	0	
	BP8	1	9	1	0	23	0	6	28	0	27	0	0	0	0	4	0	0	1	
	BP9	1	0	0	1	0	10	0	0	76	0	6	0	0	6	0	0	0	0	
	BP10	0	9	1	0	41	0	0	17	0	27	0	1	0	0	1	3	0	0	
	BP11	0	0	0	4	0	1	8	3	16	1	47	17	1	2	0	0	0	0	
	BP12	0	0	1	1	0	1	0	8	5	0	18	66	0	0	0	0	0	0	
	BP13 l	0	1	0	1	0	1	18	0	9	0	21	1	46	0	0	0	0	0	
	BP13 r	2	0	0	4	0	0	19	0	8	0	23	1	4	37	0	1	0	0	
	BP14 l	1	2	5	5	0	0	0	1	0	0	0	1	0	0	80	0	1	2	
	BP14 r	3	1	0	1	0	0	0	8	0	0	0	0	0	0	0	69	1	0	
	BP15	0	1	0	1	1	7	0	0	1	0	0	0	0	0	1	1	72	9	
	BP16 l	0	0	0	3	0	7	0	1	0	0	0	0	0	0	8	0	7	73	
	BP16 r	1	3	3	0	0	0	0	0	1	0	0	0	0	0	0	3	8	0	

Tabelle A.20: Konfusionsmatrix der Klassifikation mittels Naive Bayes Klassifikator nach Dimensionsreduktion (Merkmalsselektion und PCA). Validierung mittels 11-facher Kreuzvalidierung.

		Vorhergesagte Klassenzugehörigkeit																	
	BP1	BP2	BP3	BP4	BP5	BP6	BP7	BP8	BP9	BP10	BP11	BP12	BP13 l	BP13 r	BP14 l	BP14 r	BP15	BP16 l	BP16 r
Wahre Klassenzugehörigkeit	BP1	71	9	9	6	1	1	1	0	1	0	0	0	0	0	0	0	0	0
	BP2	15	46	16	15	1	1	1	0	1	0	0	0	0	1	1	1	0	0
	BP3	21	22	34	14	0	0	1	1	2	0	1	0	0	2	2	0	0	0
	BP4	10	21	11	50	0	7	0	0	0	0	0	0	0	1	0	0	0	0
	BP5	0	0	1	2	58	0	0	9	0	28	0	0	0	1	0	0	0	0
	BP6	3	2	1	4	0	83	0	0	6	0	0	0	0	0	0	1	0	2
	BP7	4	1	0	1	0	0	81	3	0	1	1	0	4	4	0	0	0	0
	BP8	1	1	0	1	12	0	3	43	0	35	0	0	0	0	0	3	0	1
	BP9	0	1	1	2	0	8	0	0	72	0	8	2	1	4	0	0	0	0
	BP10	0	2	2	0	23	0	0	24	0	47	0	0	0	0	1	1	0	0
	BP11	0	0	2	1	0	1	10	3	9	0	57	7	4	6	0	0	0	0
	BP12	0	0	0	1	0	0	0	2	2	0	15	79	1	0	0	0	0	0
	BP13 l	0	0	0	1	0	0	15	1	4	0	9	1	61	9	0	0	0	0
	BP13 r	0	0	1	3	0	0	16	0	1	0	11	1	9	58	0	0	0	0
	BP14 l	4	1	3	3	1	0	0	0	0	0	0	0	0	0	84	0	1	3
	BP14 r	3	1	2	1	1	0	0	5	0	0	0	0	0	0	0	84	0	4
BP15	0	0	0	2	1	2	0	0	1	0	0	0	0	0	0	1	82	6	
BP16 l	0	0	1	2	0	0	0	1	0	0	0	0	0	0	7	0	80	0	
BP16 r	0	1	3	1	0	0	0	0	0	0	0	0	0	0	1	4	7	0	83

Tabelle A.21: Klassifikation mittels Naive Bayes Klassifikator nach Dimensionsreduktion (Merkmalselektion und PCA).
Validierung nach Hold-Out Methode über 30 Durchläufe gemittelt.

		Vorhergesagte Klassenzugehörigkeit																		
		BP1	BP2	BP3	BP4	BP5	BP6	BP7	BP8	BP9	BP10	BP11	BP12	BP13 Li	BP13 r	BP14 l	BP14 r	BP15	BP16 l	BP16 r
Wahre Klassenzugehörigkeit	BP1	15	4	7	4	3	17	0	10	1	1	4	8	0	3	3	1	13	3	1
	BP2	5	3	8	6	2	10	0	10	2	5	2	6	5	5	8	3	11	2	10
	BP3	7	11	12	4	3	8	1	1	1	4	1	4	1	5	4	3	16	0	13
	BP4	8	7	6	5	5	11	2	6	2	1	5	5	4	1	1	6	15	2	8
	BP5	3	0	3	0	24	0	1	27	0	19	3	0	0	3	4	4	5	3	4
	BP6	4	0	3	1	3	24	4	0	22	0	20	6	3	3	4	0	4	1	0
	BP7	1	0	0	0	2	2	35	0	4	0	13	1	18	17	2	0	4	1	0
	BP8	5	0	3	1	16	1	4	23	0	15	4	3	6	0	6	6	3	1	3
	BP9	1	0	0	4	0	24	3	1	36	0	17	4	1	6	0	0	1	1	0
	BP10	1	0	1	1	23	0	0	15	0	22	5	3	1	1	10	4	6	1	4
	BP11	3	0	2	1	4	18	9	4	10	4	18	6	5	4	2	0	3	5	0
	BP12	8	5	8	4	4	16	1	6	8	1	10	29	0	0	0	0	1	0	0
	BP13 l	0	4	3	5	0	5	22	5	9	1	7	3	16	5	0	0	3	8	4
	BP13 r	1	4	1	1	5	2	14	6	6	1	14	4	12	14	2	1	2	7	0
	BP14 l	1	6	4	4	7	0	1	7	1	11	2	2	0	0	43	0	2	6	0
	BP14 r	0	4	1	3	13	0	0	17	0	10	0	1	0	0	4	38	8	1	0
	BP15	5	2	12	11	7	1	2	6	1	0	8	0	0	4	2	0	26	7	6
BP16 l	1	1	3	3	0	1	3	8	6	3	10	1	1	8	6	1	25	15	1	
BP16 r	0	1	7	7	3	0	1	8	1	1	7	0	0	1	1	1	11	3	44	

Tabelle A.22: Konfusionsmatrix der Klassifikation mittels kNN Klassifikator ($k = 5$) nach Dimensionsreduktion (Merkmalsselektion und PCA). Validierung mittels 11-facher Kreuzvalidierung.

		Vorhergesagte Klassenzugehörigkeit																Wahre Klassenzugehörigkeit	
		BP1	BP2	BP3	BP4	BP5	BP6	BP7	BP8	BP9	BP10	BP11	BP12	BP13 li	BP13 r	BP14 l	BP14 r		BP15
BP1	41	10	10	5	2	6	0	2	0	1	4	1	2	2	1	4	5	2	1
BP2	13	26	14	3	1	5	1	3	1	3	1	3	4	3	2	3	5	3	5
BP3	6	10	45	10	2	2	0	2	1	4	1	1	2	1	3	3	3	1	4
BP4	6	3	12	42	3	4	2	1	2	1	4	3	3	1	1	1	7	0	3
BP5	2	1	2	1	55	0	0	14	0	15	1	1	0	1	5	1	1	1	0
BP6	4	0	1	0	0	47	4	1	19	1	12	2	4	3	1	1	1	1	0
BP7	2	0	0	0	1	3	64	1	3	1	5	1	5	9	1	1	1	2	0
BP8	2	2	2	1	16	0	4	32	0	18	4	1	4	1	4	6	1	1	2
BP9	1	1	0	1	0	20	6	1	42	0	10	3	5	6	0	0	0	1	0
BP10	1	1	1	1	14	0	1	18	0	45	3	1	0	1	5	4	1	1	4
BP11	2	1	2	1	2	14	4	1	6	2	41	7	6	8	1	1	2	1	1
BP12	3	2	1	1	2	6	1	5	5	2	8	55	2	3	0	0	2	2	1
BP13 l	2	1	4	1	1	6	9	2	8	0	9	1	25	22	2	0	3	3	3
BP13 r	2	2	1	1	2	3	8	2	5	1	10	2	17	38	0	1	2	2	1
BP14 l	3	2	3	2	4	0	2	8	1	15	1	1	1	1	41	8	3	3	0
BP14 r	1	0	2	1	7	1	0	9	0	6	0	2	0	0	4	60	2	1	2
BP15	4	2	4	4	1	5	2	3	2	2	4	4	2	1	3	4	43	7	6
BP16 l	6	1	3	2	1	3	2	4	1	2	4	2	5	7	3	2	14	34	4
BP16 r	3	4	4	4	2	0	0	3	1	4	4	0	1	2	1	2	4	1	60

Tabelle A.23: Konfusionsmatrix der Klassifikation mittels kNN Klassifikator ($k = 5$) nach Dimensionsreduktion (Merkmalsselektion und PCA). Validierung mittels Hold-Out Methode über 30 Durchläufe.

		Vorhergesagte Klassenzugehörigkeit																			
		BP1	BP2	BP3	BP4	BP5	BP6	BP7	BP8	BP9	BP10	BP11	BP12	BP13 Li	BP13 r	BP14 l	BP14 r	BP15	BP16 l	BP16 r	
Wahre Klassenzugehörigkeit	BP1	28	1	20	11	0	7	1	7	3	0	6	7	3	0	0	0	4	1	0	
	BP2	3	3	29	27	0	6	0	8	3	0	3	0	5	0	6	3	3	0	0	
	BP3	8	7	27	8	1	5	0	12	1	0	1	1	8	3	5	1	8	1	1	
	BP4	12	12	4	29	2	8	0	11	2	0	6	1	4	1	4	4	1	0	0	
	BP5	3	0	3	4	28	0	0	27	0	23	0	0	3	0	1	9	0	1	0	
	BP6	1	1	3	4	0	41	1	0	23	0	14	6	3	0	0	0	3	0	1	
	BP7	1	0	1	0	1	2	42	2	0	0	10	0	21	17	0	0	0	2	0	
	BP8	4	3	3	0	22	0	3	30	0	20	1	4	5	0	4	3	0	0	0	
	BP9	3	0	0	3	0	11	0	0	60	0	18	1	1	1	0	0	0	1	0	
	BP10	1	1	4	0	27	0	0	41	0	18	0	1	0	0	4	1	1	0	0	
	BP11	0	1	4	2	2	10	14	3	19	0	29	4	8	0	1	0	0	0	2	
	BP12	1	3	0	1	0	11	0	1	9	0	10	64	0	0	0	0	0	0	0	
	BP13 l	0	1	3	8	0	1	26	5	8	0	8	1	21	11	0	0	1	5	0	
	BP13 r	0	2	1	5	1	0	16	5	11	0	7	0	23	24	1	0	0	4	0	
	BP14 l	0	5	2	2	6	0	1	1	0	0	0	0	0	0	79	0	0	2	0	
	BP14 r	3	7	3	1	7	0	0	13	0	4	0	0	0	0	3	51	7	0	1	
	BP15	0	1	7	4	1	1	1	2	1	1	4	0	4	0	2	1	1	55	8	6
	BP16 l	1	1	1	3	0	0	1	8	0	0	1	0	4	3	6	0	0	18	51	0
	BP16 r	1	4	6	4	0	1	0	3	0	0	1	1	3	0	1	4	7	0	0	62

Tabelle A.24: Konfusionsmatrix der Klassifikation mittels SVM (one-against-one) nach MDimensionsreduktion (Merkmalsselektion und PCA). Validierung mittels 11-facher Kreuzvalidierung.

		Vorhergesagte Klassenzugehörigkeit																			
		BP1	BP2	BP3	BP4	BP5	BP6	BP7	BP8	BP9	BP10	BP11	BP12	BP13 li	BP13 r	BP14 l	BP14 r	BP15	BP16 l	BP16 r	
Wahre Klassenzugehörigkeit		BP1	69	5	9	3	2	2	1	0	1	0	6	0	0	0	1	0	0	1	1
		BP2	7	47	26	6	1	1	2	1	1	0	2	0	1	0	0	0	0	2	2
		BP3	2	16	55	11	0	0	0	4	0	3	2	0	3	0	0	0	1	4	2
		BP4	6	8	11	60	0	2	1	0	2	0	3	1	4	0	0	1	1	0	1
		BP5	1	0	2	0	64	0	0	12	0	19	0	0	0	0	1	2	0	0	0
		BP6	4	3	3	0	0	66	0	0	12	0	9	2	2	0	0	0	0	0	1
		BP7	2	0	1	1	1	2	68	0	1	0	6	0	11	8	0	0	0	0	0
		BP8	3	1	1	2	10	0	3	43	0	28	2	0	2	3	2	3	0	0	0
		BP9	2	1	1	0	0	22	3	0	56	0	7	2	5	1	0	0	0	0	0
		BP10	0	1	1	0	19	0	0	29	0	48	0	0	0	0	2	1	0	0	0
		BP11	3	0	3	0	1	6	8	0	8	1	63	3	1	1	0	0	0	1	0
		BP12	0	0	0	0	2	3	0	0	5	1	8	80	0	1	0	1	0	0	0
		BP13 l	2	2	4	1	0	0	12	1	4	0	8	0	40	24	0	0	1	1	1
		BP13 r	2	5	3	1	0	2	8	0	6	1	6	1	17	48	0	0	0	0	0
		BP14 l	3	1	2	1	2	0	0	2	0	3	0	0	1	0	80	1	1	2	1
		BP14 r	0	0	1	3	0	0	0	6	0	3	0	0	0	0	5	81	0	0	3
BP15	2	4	0	1	0	2	0	1	0	1	2	0	2	1	0	1	76	5	4		
BP16 l	3	1	2	0	0	2	3	3	4	1	0	0	1	4	6	0	10	62	0		
BP16 r	1	5	4	1	0	0	0	3	0	0	2	0	2	0	2	0	6	0	76		

Tabelle A.25: Konfusionsmatrix der Klassifikation mittels SVM (one-against-one) nach Dimensionsreduktion (Merkmalsselektion und PCA). Validierung mittels Hold-Out Methode über 3 Durchläufe.

Abbildungsverzeichnis

2.1	Kamerasysteme zur Aufnahme von Tiefendaten.	6
2.2	Tiefenbilddarstellung eines Gesichts	7
2.3	Triangulationsprinzip im Kinectsensor	9
2.4	Arbeitsschritte der Erkennung von Gesichtsmotorik-Übungen	14
2.5	Entropien zweier Zufallsvariablen und Transinformation	16
4.1	Nummerierung der genutzten Labelpunkte	46
4.2	Tiefenbilddarstellung der Übung „Kuss“ mit eingezeichneten Markern .	47
4.3	Normalenrichtungen bei unterschiedlichen Bezugssystemen	50
4.4	Schnittkurve C der Punktsignaturberechnung	52
4.5	Abtastung der Schnittkurve C	53
5.1	Regressionsebene innerhalb einer achter Nachbarschaft	65
5.2	Richtungsbestimmung der Normalenvektoren	66
5.3	Definition der Normalenwinkel in sphärischen Koordinaten	69
5.4	Spaltendefinition als Ortsbestimmung der Normalenvektoren	70
5.5	Zehn Gesichtsregionen zur Ortsbestimmung der Histogrammbildung . .	74
5.6	Beispielhistogramme über diskrete Winkelintervalle	75
5.7	Oberflächenbereiche des Einheitskreises bei Intervallkombination der Normalenwinkel	76
5.8	Berechnungsschritte der Punktsignatur	79
5.9	Ortsdefinition der Punktsignaturen	81
5.10	DCT-Rücktransformation einer Punktsignatur mittels unterschiedlicher Koeffizientenanzahl	82

5.11	Paarweise MI der Punktsignaturen (dynamisch)	83
5.12	Translations- und Rotationsinvarianz des Radius	85
5.13	MI der selektierten Featureblöcke	88
5.14	Paarweise MI zwischen den selektierten Featureblöcke	89
6.1	BER aller 19 Klassen	97
6.2	Einfluss der Positionsverschiebung der Labelpunkte	99
7.1	RGB-Bildausschnitt mit Labelpunkte und deren Positionsunsicherheit .	104
7.2	Merkmale im Stirnbereich	109
7.3	Unterschiedliche Intensitätsstufen der BP 13	110
7.4	Ähnlichkeiten der BP 5, 8 und 10.	112
7.5	Tiefenbild der Übung „OForm“	112
8.1	Zusammenfassender Überblick über die Module der Klassifikation der Gesichtsmotorik-Übungen	116

Tabellenverzeichnis

2.1	Konfusionsmatrix eines binären Klassifikationsproblems.	19
4.1	Übersicht der Basisprimitive	44
5.1	Winkelintervalle zur Histogrammbestimmung	73
5.2	Featurearten und Featureblöcke	86
6.1	Erfolgsratenänderung durch Dimensionsreduktion	94
6.2	Erfolgsraten der Klassifikation in 19 Klassen	95
6.3	BER der Klassifikatoren über 19 Klassen mit unterschiedlichen Klassi- fikatoren und Validierungsarten	95
6.4	TPR und BER bezüglich einzelner Klassen	96
6.5	Erfolgsraten einzelner Durchläufe der 11-fachen Kreuzvalidierung	97
6.6	Erfolgsraten bei verschobenen Markerpositionen	98
6.7	BER aller Klassifikationen bei künstlich verschobenen Markerpositionen	98
7.1	Auszug der Konfusionsmatrix von Stirn- und Augenübungen	108
7.2	Auszug der Konfusionsmatrix über Zungenübungen	109
7.3	Auszug der Konfusionsmatrix über die Übungen „OForm“, „UForm“ und „Kussmund“	111
A.1	Dyarthriearten	121
A.2	Basisprimitive der Gesichtsmotorik-Übungen Teil 1	122
A.3	Basisprimitive der Gesichtsmotorik-Übungen Teil 2	123
A.4	Übungsübersicht Teil 1: Stirnrunzeln	124

A.5	Übungsübersicht Teil 2: Augenbrauen zusammenziehen	124
A.6	Übungsübersicht Teil 3: Augen schließen	125
A.7	Übungsübersicht Teil 4: Augen zusammenpressen	125
A.8	Übungsübersicht Teil 5: Lippen spitzen	126
A.9	Übungsübersicht Teil 6: Lippen breitziehen	127
A.10	Übungsübersicht Teil 7: „A“ formen	127
A.11	Übungsübersicht Teil 8: „O“ formen	128
A.12	Übungsübersicht Teil 9: „I“ formen	129
A.13	Übungsübersicht Teil 10: „U“ formen	129
A.14	Übungsübersicht Teil 12: Zunge gerade herausstrecken	130
A.15	Übungsübersicht Teil 12: Zunge zur Nase	131
A.16	Übungsübersicht Teil 13: Zunge in den Mundwinkel	132
A.17	Übungsübersicht Teil 14: Zunge in Wange bohren	132
A.18	Übungsübersicht Teil 16: Beide Wangen aufblasen	133
A.19	Übungsübersicht Teil 15: Eine Wange aufblasen	134
A.20	Konfusionsmatrix der Klassifikation mittels Naive Bayes Klassifikator und 11-facher Kreuzvalidierung	135
A.21	Konfusionsmatrix Klassifikation mittels Naive Bayes Klassifikator und Hold-Out Methode	136
A.22	Konfusionsmatrix der Klassifikation mittels KNN Klassifikator ($k = 5$) und 11-facher Kreuzvalidierung	137
A.23	Konfusionsmatrix der Klassifikation mittels KNN Klassifikator ($k = 5$) und Hold-Out Methode	138
A.24	Konfusionsmatrix der Klassifikation mittels SVM (one-againsts-one) und 11-facher Kreuzvalidierung	139
A.25	Konfusionsmatrix der Klassifikation mittels SVM (one-against-one) und Hold-Out Methode	140

Literaturverzeichnis

- [BATTITI, 1994] BATTITI, ROBERTO (1994). *Using Mutual Information for Selecting Features in Supervised Neural Net Learning*. Neural Networks, IEEE Transactions on, 5(4):537–550.
- [BENECKE, 1998] BENECKE, PETRA (1998). *Logopädie: Sprechübungen für Parkinson-Patienten*. Pharmacia und Upjohn.
- [BRAMER, 2006] BRAMER, MAX.A. (2006). *Artificial Intelligence in Theory and Practice: IFIP 19th World Computer Congress, TC-12 IFIP AI 2006 Stream, August 21-24, 2006, Santiago, Chile*. IFIP (Boston). Springer.
- [BRAUN, 1999] BRAUN, O. (1999). *Sprachstörungen bei Kindern und Jugendlichen: Diagnostik, Therapie, Förderung*. Kohlhammer Pädagogik. Kohlhammer.
- [CHANG und LIN, 2011] CHANG, C.C. und C. LIN (2011). *LIBSVM: A Library for Support Vector Machines*. ACM Transactions on Intelligent Systems and Technology (TIST), 2(3):27.
- [CHANG et al., 2003] CHANG, KYONG I., K. W. BOWYER und P. J. FLYNN (2003). *Face Recognition Using 2D and 3D Facial Data*. In: *ACM Workshop on Multimodal User Authentication*, S. 25–32.
- [CHEW et al., 2010] CHEW, WEI J., K. P. SENG und L.-M. ANG (2010). *Surface Normal Angle for 3D Face Recognition*. In: *Computer Applications and Industrial Electronics (ICCAIE), 2010 International Conference on*, S. 477 –480.
-

- [CHUA und JARVIS, 1997] CHUA, CHIN S. und R. JARVIS (1997). *Point Signature: a New Representation for 3d Object Recognition*. In: *International Journal of Computer Vision*, Bd. 25, S. 63–85.
- [CHUA et al., 2000] CHUA, CHIN-SENG, F. HAN und Y.-K. HO (2000). *3D Human Face Recognition Using Point Signature*. In: *Proc. Fourth IEEE Int Automatic Face and Gesture Recognition Conf*, S. 233–238.
- [CLAUSNITZER, 2004] CLAUSNITZER, VOLKMAR (2004). *Orofaziale Muskelfunktions-therapie (OMF)*, Bd. 2., durchges. Aufl. Modernes Lernen Borgmann.
- [DUDA et al., 2001] DUDA, RICHARD O., P. E. HART und D. G. STORK (2001). *Pattern Classification*. John Wiley and Sons, 2., illustriert Aufl.
- [EKMAN und FRIESEN, 2003] EKMAN, PAUL und W. V. FRIESEN (2003). *Unmasking the Face: a Guide to Recognizing Emotions from Facial Clues*. Malor Books.
- [EKMAN et al., 2003a] EKMAN, PAUL, W. V. FRIESEN und J. C. HAGER (2003a). *Facial Action Coding System - The Manual*. <http://face-and-emotion.com/dataface/facs/manual/TitlePage.html>. Letzter Abruf: 05.03.2012.
- [EKMAN et al., 2003b] EKMAN, PAUL, W. V. FRIESEN und J. C. HAGER (2003b). *Facial Action Coding System Investigator's Guide - HTML Demonstration Version*. <http://face-and-emotion.com/dataface/facs/guide/FACSIVTi.html>. Letzter Abruf: 05.03.2012.
- [FRANKE, 2008] FRANKE, ULRIKE (2008). *Logopädisches Handlexikon*. Uni-Taschenbücher. Ernst Reinhardt Verlag, 8. , ergänzte Aufl.
- [GRAMLICH, 2004] GRAMLICH, GÜNTER (2004). *Anwendungen der linearen Algebra*. Mathematik-Studienhilfen. Fachbuchverl. Leipzig im Carl-Hanser-Verl.
- [GROSS, 2011] GROSS, HORST-MICHAEL (2011). *Skript zur Vorlesung Angewandte Neuroinformatik*.
-

- [GÖRZ und ROLLINGER, 2003] GÖRZ, G. und C.-R. ROLLINGER (2003). *Handbuch der Künstlichen Intelligenz*. Oldenbourg Wissenschaftsverlag, 4. Aufl.
- [GUYON und ELISSEEFF, 2003] GUYON, ISABELLE und A. ELISSEEFF (2003). *An Introduction to Variable and Feature Selection*. The Journal of Machine Learning Research, 3:1157–1182.
- [HAN und KAMBER, 2006] HAN, JIAWEI und M. KAMBER (2006). *Data Mining: Concepts and Techniques*. The Morgan Kaufmann series in data management systems. Morgan Kaufmann, 2., illustriert Aufl.
- [JÖRG, 1997] JÖRG, J. (1997). *Neurologische Therapie*. Springer-Verlag, 3., überarb. Aufl.
- [KAHLMANN et al., 2006] KAHLMANN, T., F. REMONDINO und H. INGENSAND (2006). *Calibration for Increased Accuracy of the Range Imaging Camera Swissranger*. ISPRS Commission V Symposium 'Image Engineering and Vision Metrology', S. 136 –141.
- [KHOSHELHAM und ELBERINK, 2012] KHOSHELHAM, KOUROSH und S. O. ELBERINK (2012). *Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications*. Sensors, 12(2):1437–1454.
- [KIRSCH et al., 2011] KIRSCH, JOACHIM, C. A. MAY, D. LORKE, A. WINKELMANN, W. SCHWAB, G. HERRMANN und R. FUNK (2011). *Taschenlehrbuch Anatomie*. Thieme.
- [KITTEL, 1998] KITTEL, ANITA M. (1998). *Myofunktionelle Therapie*. Idstein: Schulz-Kirchner, 3. Aufl.
- [KOHAVI, 1995] KOHAVI, RON (1995). *A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection*. In: *International Joint Conference on Artificial Intelligence*, Bd. 14, S. 1137–1145.
-

- [KOVESI] KOVESI, P. D. *MATLAB and Octave Functions for Computer Vision and Image Processing*. Centre for Exploration Targeting, School of Earth and Environment, The University of Western Australia. Available from: [http://www.csse.uwa.edu.au/\\$\sim\\$pk/research/matlabfns/](http://www.csse.uwa.edu.au/\simpk/research/matlabfns/), letzter Abruf: 08.04.2012.
- [LEBING, 2011] LEBING, CHRISTIN (2011). *Data Mining zur Identifikation potentieller Kunden*. Masterarbeit, Hochschule Wismar, Fakultät für Wirtschaftswissenschaften.
- [LI und SOARES, 2011] LI, CHAO und A. SOARES (2011). *Automatic Facial Expression Recognition Using 3D Faces*. International Journal of Engineering Research & Innovation, 3(1):30–34.
- [LIU und MOTODA, 2008] LIU, HUAN und H. MOTODA (2008). *Chapman & Hall/CRC Data Mining and Knowledge Discovery Series*. In: *Computational Methods of Feature Selection*, S. 419. Chapman & Hall/CRC.
- [MENNA et al., 2011] MENNA, F., F. REMONDINO, R. BATTISTI und E. NOCERINO (2011). *Geometric Investigation of a Gaming Active Device*. In: *Proceedings of SPIE*, Bd. 8085, S. 80850G.
- [MPIPERIS et al., 2007] MPIPERIS, I., S. MALASIOTIS und M. STRINTZIS (2007). *3D Face Recognition by Point Signatures and Iso-contours*. In: *Proceedings of the Fourth conference on IASTED International Conference: Signal Processing, Pattern Recognition, and Applications*, S. 328–332. ACTA Press.
- [OVE, 2011] OVE, EDLUND (2011). *Some Notes on Least Squares, QR-factorization, SVD and Fitting*. http://www.math.ltu.se/courses/c0002m/least_squares.pdf. Letzter Abruf: 24.03.2012.
- [PLAUE, 2006] PLAUE, MATTHIAS (2006). *Technical Report: Analysis of the PMD Imaging System*. Technischer Bericht, Interdisciplinary Center for Scientific Computing, Univ. of Heidelberg.
-

- [PROSIEGEL und WEBER, 2010] PROSIEGEL, MARIO und S. WEBER (2010). *Dysphagie: Diagnostik und Therapie*, Bd. 14 d. Reihe *Praxiswissen Logopädie*. Springer-Verlag Berlin-Heidelberg, 1. Aufl.
- [RAPP et al., 2008] RAPP, H., M. FRANK, M. HAMPRECHT und B. JÄHNE (2008). *A Theoretical and Experimental Investigation of the Systematic Errors and Statistical Uncertainties of Time-Of-Flight-Cameras*. International Journal of Intelligent Systems Technologies and Applications, 5(3):402–413.
- [RAPP, 2007] RAPP, HOLGER (2007). *Experimental and Theoretical Investigation of Correlating TOF-Camera Systems*. Diplomarbeit, University of Heidelberg.
- [RUSU und COUSINS, 2011] RUSU, RADU B. und S. COUSINS (2011). *3D is here: Point Cloud Library (PCL)*. In: *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China.
- [RUSU und COUSINS, 2012] RUSU, RADU B. und S. COUSINS (2012). *PCL-Module Features*. http://docs.pointclouds.org/trunk/group__features.html. Letzter Abruf: 27.02.2012.
- [SAVRAN et al., 2008] SAVRAN, ARMAN, N. ALYÜZ, H. DİBEKLİOĞLU, O. ÇELİK-TUTAN, B. GÖKBERK, B. SANKUR und L. AKARUN (2008). *Bosphorus Database for 3D Face Analysis*. In: SCHOUTEN, BEN, N. JUUL, A. DRYGAJLO und M. TISTARELLI, Hrsg.: *Biometrics and Identity Management*, Bd. 5372 d. Reihe *Lecture Notes in Computer Science*, S. 47–56. Springer Berlin / Heidelberg.
- [SCHMIED und SCHMIED, 2007] SCHMIED, STEFANIE und U. SCHMIED (2007). *CoMuZu- Computer Mund Zunge*. www.comuzu.de. Letzter Abruf: 27.02.2012.
- [SOYEL und DEMIREL, 2008] SOYEL, H. und H. DEMIREL (2008). *3D Facial Expression Recognition with Geometrically Localized Facial Features*. In: *Proc. 23rd Int. Symp. Computer and Information Sciences ISCIS '08*, S. 1–4.
- [SUN et al., 2008] SUN, YI, M. REALE und L. YIN (2008). *Recognizing Partial Facial Action Units Based on 3D Dynamic Range Data for Facial Expression Recognition*.
-

- In: *Proc. 8th IEEE Int. Conf. Automatic Face & Gesture Recognition FG '08*, S. 1–8.
- [TANG und HUANG, 2008] TANG, HAO und T. S. HUANG (2008). *3D Facial Expression Recognition Based on Automatically Selected Features*. In: *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition Workshops CVPRW '08*, S. 1–8.
- [TEKGÜÇ et al., 2009] TEKGÜÇ, UMUT, H. SOYEL und H. DEMIREL (2009). *Feature Selection for Person-independent 3D Facial Expression Recognition Using NSGA-II*. In: *Feature selection for person-independent 3D facial expression recognition using NSGA-II*, S. 35–38.
- [THIELE, 1997] THIELE, ERHARD (1997). *Myofunktionelle Therapie: Katalog der Übungen zur neuromotorischen Funktionsregulation*. In: *Theorie und Praxis der MFT*, Bd. 3. Hüthig.
- [UNIVERSITÄT DES SAARLANDES, 2009] UNIVERSITÄT DES SAARLANDES (2009). *E-learning-Kurs About FACS*. <http://www.uni-saarland.de/fak5/orga/Kurs/home.htm#inhalt>. Letzter Abruf: 27.03.2012.
- [VIDEOGAMEZONE, 2011] VIDEOGAMEZONE (2011). *Avatar Kinect für Xbox 360 im Hands-On-Test Lustige Features aber schlechte Erkennung von Mimik und Gestik*. www.videogamezone.de. Letzter Abruf: 27.03.2012.
- [VILLAROMAN et al., 2011] VILLAROMAN, NORMAN, D. ROWE und B. SWAN (2011). *Teaching Natural User Interaction Using OpenNI and the Microsoft Kinect Sensor*. In: *Proceedings of the 2011 conference on Information technology education, SIGITE '11*, S. 227–232, New York, NY, USA. ACM.
- [VMS GESELLSCHAFT FÜR THERAPEUTISCHE MEDIEN MBH, 2009] VMS GESELLSCHAFT FÜR THERAPEUTISCHE MEDIEN MBH (2009). *Logomedien*. www.logomedien.de. Letzter Abruf: 27.03.2012.
-

-
- [WALTER, 2004] WALTER, JÖRG ANDREA (2004). *Datamining: Methoden integrativer Datenpräsentation*. Cuvillier Verlag, Göttingen.
- [WANG et al., 2006] WANG, JUN, L. YIN, X. WEI und Y. SUN (2006). *3D Facial Expression Recognition Based on Primitive Surface Feature Distribution*. In: *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Bd. 2, S. 1399–1406.
- [WANG et al., 2002] WANG, YINGJIE, C.-S. CHUA und Y.-K. HO (2002). *Facial Feature Detection and Face Recognition from 2D and 3D Images*. In: *Pattern Recognition Letters*, Bd. 23, S. 1191–1202.
- [YIN et al., 2008] YIN, LIJUN, X. CHEN, Y. SUN, T. WORM und M. REALE (2008). *A High-Resolution 3D Dynamic Facial Expression Database*. In: *Proc. 8th IEEE Int. Conf. Automatic Face & Gesture Recognition FG '08*, S. 1–6.
- [YIN et al., 2006] YIN, LIJUN, X. WEI, Y. SUN, J. WANG und M. J. ROSATO (2006). *A 3D Facial Expression Database for Facial Behavior Research*. In: *Proc. 7th Int. Conf. Automatic Face and Gesture Recognition FGR 2006*, S. 211–216.
- [ZHANG und HANCOCK, 2011] ZHANG, ZHIHONG und E. R. HANCOCK (2011). *Mutual Information Criteria for Feature Selection*. In: *Similarity-based Pattern Recognition*, S. 235–249. Springer-Verlag.
-

Thesen

- Gegenwärtig existieren für Gesichtsmotorik-Übungen keine unterstützenden und somit therapiefördernden Systeme mit Feedbackkomponenten bezüglich der Durchführungsqualität.
- Eine Offline-Klassifikation von Gesichtsmotorik-Übungen mittels Punktsignaturen und Normalenvektoren anhand von statischen Tiefenbildern des Kinectsensors ist möglich.
- Kombinationen von Punktsignaturen an Gesichtsbereichen, welche bei Übungsdurchführungen verändert werden, bilden Distanzprofile einzelner Motorik-Übungen.
- Normalenvektoren dienen als Oberflächenbeschreibung des Gesichts; Eine Histogrammbildung über Intervalle der sphärischen Koordinaten in definierten Ortsbereichen, den ROIs, standardisiert den Merkmalsraum auf eine einheitliche Dimension.
- Der zusammengestellte Katalog der Gesichtsmotorik-Übungen bildet eine Grundlage für Trainingspläne bei verschiedenen orofazialen und mimischen Muskelstörungen.
- Die Codierung der Übungen über das im Rahmen dieser Arbeit entwickelte BP-System bildet, nach einer Erweiterung durch Intensitätsstufen der Übungen, die Basis für eine Evaluierung der Übungsdurchführungen.

Ilmenau, 13.04.2012

.....

Birant Sibel Olgay