



UNIVERSITÄT ZU LÜBECK
INSTITUT FÜR MEDIZINISCHE INFORMATIK

Aus dem Institut für Medizinische Informatik
der Universität zu Lübeck
Direktor: Prof. Dr. rer. nat. habil. Heinz Handels

**“Statistische Appearance-Modelle
basierend auf
probabilistischen Korrespondenzen
für die medizinische Bildanalyse”**

Inauguraldissertation
zur
Erlangung der Doktorwürde
der Universität zu Lübeck

Aus der Sektion Informatik/Technik

vorgelegt von
Julia Krüger
aus Schwerin

Lübeck, 25. Oktober 2018

1. Berichterstatter: Prof. Dr. rer. nat. habil. Heinz Handels

2. Berichterstatter: Prof. Dr. rer. nat. Thorsten M. Buzug

Tag der mündlichen Prüfung: 08.06.2018

Zum Druck genehmigt. Lübeck, den 22.10.2018

Zusammenfassung

Die modellbasierte Bildanalyse spielt eine große Rolle in der medizinischen Bildverarbeitung. Die Modellierung der Form und Erscheinung eines bestimmten Organs basiert dabei auf der statistischen Analyse einer Menge von Trainingsdaten, welche das gesuchte Organ repräsentieren. Um die gesuchte Struktur zu beschreiben, wird eine mittlere Repräsentation und ein Satz an Variationsmoden bestimmt, welche die Variationsbreite des Organs abdecken. Einer der wichtigsten Aspekte bei der Generierung klassischer statistischer Form- und Erscheinungsmodelle ist die Determinierung Eins-zu-Eins-korrespondierender Landmarken in allen Trainingsdaten. Zur gleichen Zeit stellt diese unter Umständen sehr aufwendige Korrespondenzbestimmung die Schwachstelle der klassischen Ansätze dar, da die Qualität des generierten Modells von der Genauigkeit der Landmarkenpositionen und der Existenz aller Landmarken in allen Bildern abhängt.

Die vorliegende Dissertation stellt einen probabilistischen Modellansatz für statistische Appearance-Modelle unter der Nutzung von Korrespondenzwahrscheinlichkeiten vor. Die Grundlage der vorangestellten Problemformulierung bildet eine Repräsentation von Bildern durch eine Menge unabhängiger D-dimensionaler Merkmalsvektoren, welche Positions- und Appearance-Informationen kombinieren. Im Kontrast zu klassischen statistischen Form- oder Appearance-Modellen nutzt das probabilistische Modell Korrespondenzwahrscheinlichkeiten, wodurch eine vorangehende, potentiell aufwendige Definition von Eins-zu-Eins-korrespondierenden Landmarken unnötig wird. Dies eliminiert den Bedarf einer elaborierten Vorverarbeitung der Trainingsdaten und unterstützt des Weiteren ein arbiträres Abtasten der Test- und Trainingsdaten. Ein abgeleitetes globales Optimierungskriterium wird sowohl für die Modellgenerierung als auch -anpassung genutzt. Für die Optimierung der gesuchten Modellparameter des globalen Kriteriums, u. a. mittleres Modell und Variationsmoden, wird ein Maximum-A-posteriori-Verfahren eingesetzt, bei dem außerdem die probabilistischen Korrespondenzen mittels eines iterativen Expectation-Maximization-Ansatzes zwischen den Daten des Trainingsdatensatzes geschätzt werden. Dabei werden alle Modellparameter analytisch über das Kriterium optimiert. Notwendige Regularisierungsparameter werden datengetrieben bestimmt und angepasst, sodass kein manuelles Parametertuning stattfinden muss. Eine Steigerung der Robustheit während der Modelloptimierung liefern verschiedene in der Arbeit vorgestellte Multi-Level-Erweiterungen.

Im Gegensatz zu klassischen Modellen, welche die Organform explizit durch definierte Landmarkenpositionen modellieren und daher hauptsächlich für die Segmentierung der modellierten Organe in neuen Bilddaten eingesetzt werden, stellt die vorliegende Arbeit vielseitige Einsatzmöglichkeiten für die vorgestellte Methode vor. Da die Forminforma-

tion nicht mehr explizit durch die Positionen der Merkmalsvektoren repräsentiert ist, wird eine Repräsentation der Segmentierungsinformationen über zusätzliche Appearance-Merkmale eingeführt. Es können sowohl Kontur- oder Landmarkeninformationen über Distanzkartenwerte als auch Labelinformationen in der genutzten Bildrepräsentation dargestellt werden. Die Wahl, Art und Anzahl der Appearance- oder Segmentierungsmerkmale ist in keiner Weise eingeschränkt, wodurch einerseits die Multi-Organsegmentierung bereits in dem Framework beinhaltet ist und andererseits die Anwendung nicht auf Segmentierungsprobleme einschränkt ist. Die Nutzung probabilistischer Korrespondenzen verringert nicht nur die notwendigen Vorverarbeitungsschritte für die u. a. Landmarkenbestimmung, sondern liefert darüber hinaus automatisch Informationen über die Qualität der Modellanpassung. Dies kann Auskunft darüber geben, an welchen Bereichen bzw. für welche Fälle das Modell keine ausreichende Variabilität aufweist. Diese Information kann genutzt werden, um z. B. pathologische oder artefaktbelastete Regionen in medizinischen Bilddaten zu lokalisieren. Sind solche Bereiche identifiziert, können z. B. die Appearance-Werte des Modells verwendet werden, um artefaktbehaftete Bereiche im Bild zu rekonstruieren.

Die Flexibilität des Frameworks wird weiterhin dadurch verdeutlicht, dass die vorgestellte Maximum-A-posteriori-Formulierung samt Optimierungsverfahren eingesetzt werden kann, um einen Registrierungsansatz zweier Bilder, welcher auf der Nutzung der probabilistischen Korrespondenzen basiert, zu formulieren. Der Vorteil eines solchen probabilistischen Registrierungsalgorithmus ist die Robustheit gegenüber fehlerhaften/pathologischen Regionen, die ausschließlich in einem der beiden Bilder vorkommen. Außerdem werden solche Gebiete (z. B. Tumor- oder Schlaganfallgewebe) durch die genutzten Korrespondenzwahrscheinlichkeiten automatisch indiziert.

Die Evaluierung des präsentierten Modellansatzes zeigt, dass die Methode in der Lage ist, sowohl Daten mit hoher Formvariabilität (Hände in 2D Röntgenbildern) als auch Daten mit hoher Appearance-Variabilität (3D-MRT-Gehirndaten) zu modellieren und diese Modelle an neue ungesehene Bilder anzupassen. Weiterhin wird demonstriert, dass die Korrespondenzwahrscheinlichkeiten zur Klassifikation von gesundem und pathologischem (nicht im Modell vorhandenen) Gewebe eingesetzt werden kann. Die Evaluierung der Registrierung zeigt, dass das probabilistische Registrierungsverfahren mit etablierten Registrierungsansätzen verglichen werden kann und im Fall der evaluierten 2D MRT-Gehirndaten sogar signifikant bessere Ergebnisse liefert.

Abstract

Model based analysis plays an important role in medical image processing. Modeling shape and appearance of a certain organ is usually based on the statistical analysis of a set of training data, which represents the organ of interest. To describe the sought structure a mean representation and a set of variation modes are determined to represent the organ's variation as expressed by the training data set. One key aspect of the generation of classical shape and appearance models is the determination of one-to-one corresponding landmarks in all the training data. But the identification of these correspondences is the most challenging part of such methods. Unfortunately, the quality of the generated model is highly dependent on the accuracy of the landmark positions and furthermore implies the existence of all landmarks in all the processed image data.

To overcome these shortcoming, this dissertation proposes a probabilistic model approach for statistical appearance models using correspondence probabilities. The basis of the proposed approach is a sparse representation of images by a set of independent multi-dimensional feature vectors that combine position and appearance information. In contrast to one-to-one correspondences as employed in classical shape or appearance models, probabilistic correspondences between the derived multi-dimensional feature vectors are used to render the extensive preprocessing step of finding landmarks and correspondences unnecessary. This furthermore reduces the model's dependence on the landmark positions. Model generation as well as model fitting is expressed by optimizing a single global criterion. To determine its optimal model parameters (mean model, variation modes), a maximum a-posteriori approach is used. The probabilistic correspondences between the training data and the parameters are estimated through an iterative expectation maximization approach. All model parameters can be optimized analytically via the criterion. The data driven adjustment of all necessary regularization parameters removes the need for manual parameter tuning. An increase in robustness during model optimization is the result of several multi-level extensions presented in this dissertation.

In contrast to classical statistical models, which explicitly model the organ shape by distinct landmark positions and are therefore mainly used for the segmentation of the modeled organ in new image data, this work presents versatile application possibilities for the proposed method. Since the shape information is no longer explicitly represented by the positions of the feature vectors, a representation of the segmentation information using additional appearance features is introduced. Those can describe both contour or landmark information using distance map values or label information in the given image representation. The choice, type, and number of appearance or segmentation features is by no means limited, which, on the one hand, includes multi-organ segmentation already in the framework and, on the other hand, does not limit the application to segmenta-

tion problems. The use of probabilistic correspondences not only reduces the necessary preprocessing steps and landmark determination, but also automatically provides knowledge about the quality of model fitting. This can provide information about in which image regions or for which cases the model does not include sufficient variability. This information can be used to, e. g., locate pathological or artifact-corrupted regions in medical image data. If such areas are identified, the model’s appearance values can be used to reconstruct these corrupted areas in the image.

The flexibility of the framework is further demonstrated by the fact that the proposed maximum a-posteriori formulation and optimization method can be used to formulate a registration approach for two images based on the use of probabilistic correspondences. The advantage of such a probabilistic registration algorithm is the robustness against corrupted and pathological regions occurring exclusively in one of the two images. In addition, these areas (e. g., tumor or stroke tissue) can be automatically identified by analyzing the correspondence probabilities.

The evaluation of the presented model approach shows that the method is able to model both data with high shape variability (hands in 2D X-ray images) and data with high appearance variability (3D-MRI brain data), and to adapt these models to new unseen images. Furthermore, it is demonstrated that the correspondence probabilities can be used for the classification of healthy and pathological (not present in the model) tissue. The evaluation of the registration shows that the probabilistic registration procedure can be compared to established registration approaches and even provides significantly better results in the case of the evaluated 2D MRI brain data.

Inhaltsverzeichnis

1. Einführung	1
1.1. Motivation	1
1.2. Stand der Forschung	3
1.2.1. Statistische Formmodelle	3
1.2.2. Appearance-Modelle	5
1.2.3. Probabilistische Korrespondenzen	7
1.2.4. Atlas- und registrierungsbasierte Verfahren	8
1.3. Ziele der Arbeit	8
1.3.1. Zusammenfassung der mit dieser Arbeit erreichten Ziele	10
1.3.2. Publikationen	11
1.3.3. Gliederung der Arbeit	12
2. Objektrepräsentation und Korrespondenzwahrscheinlichkeiten	13
2.1. Bildrepräsentation aus unabhängigen Merkmalsvektoren	13
2.2. Korrespondenzwahrscheinlichkeiten zwischen Bildrepräsentationen	17
2.3. Expectation-Maximization-Ansatz zur Optimierung einer Transformation mittels Korrespondenzoptimierung	19
2.3.1. Normierung der Erwartungswerte	20
3. Probabilistische Appearance-Modelle	25
3.1. Modell auf Basis von zufälligen Observationen	25
3.1.1. Parameter	26
3.2. Maximum-A-posteriori-Ansatz zur Parameteroptimierung	33
3.2.1. Likelihood $p(\mathbf{S}_{\text{tr}} Q, \Theta)$	34
3.2.2. A-priori-Wahrscheinlichkeiten der Modellparameter $p(\Theta)$	35
3.2.3. A-priori-Wahrscheinlichkeit der instanzabhängigen Parameter $p(Q_k \Theta)$	40
3.2.4. Zusätzliche Symmetriebedingungen für die Modellgenerierung in $p(Q \Theta)$	41
3.3. Globales Kriterium	44
4. Modellgenerierung und Modellanpassung	45
4.1. Optimierung der Parameter anhand eines globalen Kriteriums	46
4.1.1. Expectation-Maximization-Algorithmus	46
4.1.2. Parameteroptimierung mithilfe partieller Ableitungen	48
4.1.3. Optimierung des mittleren Modells	49
4.1.4. Optimierung der Variationsmoden	50
4.1.5. Optimierung der Standardabweichung der Variationsmoden	52
4.1.6. Optimierung der instanzabhängigen Modengewichte	53
4.1.7. Optimierung der instanzabhängigen affinen Transformation	55

4.2. Regularisierungsparameter	62
4.2.1. Datengetriebene Anpassung der Merkmalsvarianzen Σ	63
4.2.2. Datengetriebene Anpassung der Varianzen der Nachbarschaftsbeziehungen für $p(\Theta)$	66
4.3. Initialisierung der Modellparameter	68
4.3.1. Initialisierung der Merkmalsvarianzen Σ	68
4.3.2. Initialisierung des mittleren Modells	68
4.3.3. Initialisierung der Nachbarschaftsbeziehungen für $p(\Theta)$	70
4.3.4. Zusammenfassung der Initialisierung	72
4.4. Modellanpassung	73
4.5. Zusammenfassung der Modellgenerierung und -anpassung	75
5. Erweiterungen des Modells	77
5.1. Multi-Level-Ansatz zum Inkrementieren der Merkmalsvektoranzahl	77
5.1.1. Initialisierung des nächsten Levels	78
5.2. Sukzessives Inkrementieren der Modenanzahl n	83
5.3. Sukzessives Dekrementieren der Merkmalsvarianzen Σ während der Modellanpassung	86
5.4. Modellierung lokaler Form- und Appearance-Informationen zur Erhöhung der Freiheitsgrade	87
6. Anwendungsbereiche für probabilistische Appearance-Modelle	91
6.1. Rekonstruktion von Bildinformationen aus spärlich verteilten Merkmalsvektoren	92
6.2. Objektsegmentierung	95
6.2.1. Rekonstruktion einer Objektkontur	97
6.3. Landmarkendetektion	97
6.3.1. Rekonstruktion von Landmarken	97
6.4. Initialisierung von $\mathbf{S}_{new}$	101
6.5. Rekonstruktion von fehlenden Bildkanälen in multispektralen Bilddaten	103
6.6. Klassifikation gesunder und pathologischer Strukturen	103
6.6.1. Korrespondenzwahrscheinlichkeiten als Grundlage für die Klassifikation	105
6.7. Modellgenerierung auf der Grundlage von partiellen Trainingsdaten	105
6.8. Bildrekonstruktion bei artefaktbehafteten Daten	106
7. Probabilistische Appearance-adaptierende Registrierung beliebigdimensionaler Bilder	109
7.1. Appearance-adaptierende Registrierung	111
7.2. Rekonstruktion eines Deformationsfeldes	113
7.2.1. Inverses Deformationsfeld	114
7.3. Registrierung von Bilddaten mit fehlenden Korrespondenzen	115

7.4. Probabilistische Registrierung als Nachverarbeitungsschritt für die Modellanpassung	116
8. Experimente	117
8.1. Merkmalsauswahl	117
8.1.1. Signalwerte	117
8.1.2. Gabor-Wavelets	117
8.1.3. Segmentierungsmerkmale: Distanzkarten	119
8.2. Positionen der Merkmalsvektoren	119
8.3. Metriken	122
8.3.1. Konturdistanzen	122
8.3.2. Punktdistanzen	122
8.3.3. Dice-/Jaccard-Koeffizienten	122
8.4. Experiment 1: Anwendung zur Handsegmentierung und Landmarkendetektion	123
8.5. Experiment 2: Detektion von Schlaganfallgewebe	128
8.6. Experiment 3: Probabilistische Registrierung von 2D MRT-Gehirnschichten	133
9. Diskussion	137
9.1. Zusammenfassung der Vorteile des vorgestellten Frameworks	137
9.2. Diskussion der Parameteroptimierung und der Parameterwahl	139
9.3. Diskussion der Ergebnisse	141
9.4. Einschränkungen des Ansatzes	142
9.5. Ausblick	143
A. Ableitung des globalen Kriteriums	145
A.1. Ableitung nach beliebigem Parameter x	145
A.2. Ableitung nach den Merkmalsvektoren des mittleren Modells $\bar{\mathbf{m}}_j$	149
A.3. Ableitung nach den Variationsmoden $\mathbf{V}$	150
A.4. Ableitung nach der Standardabweichung λ_p der Moden	152
A.5. Ableitung nach den Gewichten der Variationsmoden ω_k	153
A.6. Ableitung nach der Transformation $T_k = \{A_k, t_k\}$	155
Literatur	159

1. Einführung

1.1. Motivation

Die Segmentierung medizinischer Bilder ist ein zentrales Problem in der Analyse und Bearbeitung medizinischer Bilddaten. Sie bildet meist die Grundlage für weitergehende Verfahren zur computergestützten Diagnostik und Therapie. Automatische robuste Segmentierungsverfahren werden unter anderem benötigt bei der

- quantitativen Radiologie – automatische Anatomie-/Pathologiedetektion mit anschließender Volumenbestimmung,
- Strahlentherapie – Identifikation von betroffenem und gesundem Gewebe – oder
- computergestützten Chirurgie – Planung von chirurgischen Eingriffen mittels 3D Modellen der betroffenen Region/Organe.

Eine automatische Segmentierung der Daten gewährleistet dabei im Gegensatz zur manuellen Segmentierung stabilere und reproduzierbare Ergebnisse. Des Weiteren kann eine genaue manuelle Segmentierung der gesuchten Organe/Strukturen mit einem enormen Zeitaufwand verbunden sein.

Die größten Herausforderungen für automatische Segmentierungsverfahren bilden zum einen die hohe anatomische (und pathologische) Variabilität zwischen Patienten und zum anderen die schwankende Bildqualität der Daten. Bei der Entwicklung solcher Verfahren muss berücksichtigt werden, dass die Form und Erscheinung der gesuchten Objekte (Organe) zwischen Patienten oder sogar zwischen zeitlich auseinanderliegenden Aufnahmen desselben Patienten stark schwanken können. Abgesehen von unterschiedlicher Bildauflösung, verschiedener Positionierung der Patienten oder abweichenden Akquirierungsparametern leiden medizinische Bilddaten außerdem unter Rauschen oder Artefakten. Einfachere Segmentierungsverfahren wie rein signalwertbasierte oder gradientenbasierte Methoden stoßen daher schnell an ihre Grenzen, wenn es sich um schlecht definierte Organgrenzen handelt.

Die Idee der modellbasierten Segmentierungsverfahren ist die Nutzung von Vorwissen über das gesuchte Objekt. Einfache Methoden gehen dabei lediglich von einer z. B. “glatten Konturgrenze” als Modellannahme aus, wohingegen komplexere Modelle die Form oder Erscheinung eines betrachteten Organs modellieren. Die Herausforderung bei der Generierung dieses Vorwissens ist die heterogene Natur von medizinischen Strukturen, die oftmals eine hohe Variabilität in Form und Erscheinung aufweisen: Ungeachtet der Tatsache, dass ein bestimmtes Organ einem generellen “Bauplan” folgt, unterscheiden sich die individuelle Form und Erscheinung jedes Patienten voneinander. Um dieses Problem der “Individualität versus Generalität” einer gewissen anatomischen Struktur zu adressieren, können statistische Modelle eingesetzt werden [Heimann & Meinzer 2009]. Diese wurden entwickelt, um die statistischen Eigenschaften einer Gruppe von

ähnlichen Objekten (Organe, Knochen,...) zu beschreiben. Der Vorteil eines modellbasierten Verfahrens liegt in seiner Robustheit: Wird ein solches Modell, welches das gesuchte Organ mit allen möglichen Variationen beschreibt, auf ein neues Bild adaptiert, ist gewährleistet, dass die Ergebnissegmentierung einer Instanz dieses Objekts entspricht.

Die Grundlage eines jeden statistischen Modells stellt eine Trainingsdatenmenge von N Datensätzen dar, welche das zu modellierende Objekt beschreibt: $\mathbf{S}_{\text{tr}} = \{\mathbf{S}_k | k = 1, \dots, N\}$. In den klassischen Methoden werden die Datensätze als Sequenz von N_s Landmarken beschrieben: $\mathbf{S}_k = [\mathbf{s}_{k,1}, \mathbf{s}_{k,2}, \dots, \mathbf{s}_{k,N_s}]$. Jede Landmarke $\mathbf{s}_{k,i} \in \mathbb{R}^D$ repräsentiert dabei eine definierte Position im Objekt. Alle Landmarken $\mathbf{s}_{1,i}, \mathbf{s}_{2,i}, \dots, \mathbf{s}_{k,i}, \dots, \mathbf{s}_{N,i}$ mit demselben Index i in allen Trainingsdaten werden als korrespondierend angenommen. Cootes et al. führte den Begriff der *point distribution models* für diese Repräsentation mittels korrespondierender Landmarken ein [Cootes et al. 1992]. Neben der Beschreibung der Form (durch Landmarken) eines Organs kann ebenfalls die Erscheinung (engl. appearance) der betrachteten Struktur mit modelliert werden. Diese "Appearance"-Informationen werden in klassischen Modellen abhängig von der durch Landmarken repräsentierten Objektform abgetastet [Cootes et al. 2001].

Da mithilfe dieser Informationen die statistischen Form- bzw. Appearance-Eigenschaften des Objekts bestimmt werden, hängt die Qualität des generierten Modells vollständig von einer korrekten Positionierung der Landmarken ab. Allerdings stellt sich dabei die Frage, ob die exakte anatomische Landmarke in allen Bildern an gut definierten Positionen existiert – auf Grund von anatomischen (oder pathologischen) Unterschieden zwischen Patienten oder auf Grund von unterschiedlichen Bildakquirierungstechniken. Dieselbe Problematik ergibt sich bei der Adaption des Modells an neue unbekannte Bilder, da hier ebenfalls von einer Eins-zu-Eins-Korrespondenz zwischen Modell und neuem Bild ausgegangen wird.

Um die grundlegende Annahme der Eins-zu-Eins-Korrespondenzen in klassischen Modellen zu ersetzen, können probabilistische Korrespondenzen eingesetzt werden. Diese Korrespondenzwahrscheinlichkeiten bieten Robustheit gegenüber "ungenau" bestimmten "Landmarken" sowie gegenüber fehlenden Strukturen in Bilddaten.

Der zweite wichtige Aspekt – neben den korrekten Landmarken – der die Qualität der modellbasierten Anwendungen ausschlaggebend bestimmt, ist die Art der Anpassung des Modells an neue Bilddaten. Ist die Modellgenerierung in den meisten Fällen durch ein gut definiertes Verfahren abgedeckt, so müssen hingegen für die Modellanpassung unabhängige Methoden entwickelt werden, da zuerst zu definieren ist, wie eine "gute Anpassung" bestimmt ist.

Die vorliegende Arbeit präsentiert einen neuen vielseitig einsetzbaren Ansatz für "probabilistische Appearance-Modelle". Die Methode verfolgt den Ansatz der Korrespondenzwahrscheinlichkeiten, um die Abhängigkeit der Ergebnisse von der Landmarkenbestimmung zu verringern. Außerdem basieren sowohl die Korrespondenzbestimmung und die Modellgenerierung als auch die Modellanpassung auf der Minimierung desselben Optimierungskriteriums, im Gegensatz zum klassischen Ansatz, welcher aus verschiedenen Schritten und Teilmethoden zusammengesetzt ist.

1.2. Stand der Forschung

1.2.1. Statistische Formmodelle

Die ersten statistischen Formmodelle (SFM) wurden von Cootes et al. [Cootes *et al.* 1995] eingeführt. Dabei wird ein Objekt durch eine Sequenz von Landmarken, die dessen Form beschreiben, modelliert. Alternative Methoden, die Form eines Objekts darzustellen, umfassen unter anderem B-Splines [Tsagaan *et al.* 2002], “Fourier Surfaces” [Staib & Duncan 1996], “Spherical Harmonics” (SPHARMS) [Kelemen *et al.* 1999], Wavelets [Davatzikos *et al.* 2003; Yu *et al.* 2007] oder Level-Set-Ansätze [Malladi *et al.* 1995]. Die vorliegende Arbeit konzentriert sich auf die Bildrepräsentation mittels Landmarken (bzw. Punktmengen). Diese Repräsentation ist unkompliziert und gleichzeitig relativ uneingeschränkt einsetzbar.

Das landmarkenbasierte Modell besteht aus der mittleren Form der betrachteten Struktur und dessen möglichen Formvariationen, die in einem zugrundeliegenden Trainingsdatensatz beinhaltet sind. Eine übliche Methode, die Variationen zwischen Objekten zu bestimmen, ist eine Hauptkomponentenanalyse (engl. principal component analysis (PCA)). Die grundlegenden Schritte, ein statistisches Formmodell zu generieren, sind wie folgt:

Gegeben: Trainingsdatensatz $\mathbf{S}_{\text{tr}} = \{\mathbf{S}_k | k = 1, \dots, N\}$ mit N_s korrespondierenden Landmarken $\mathbf{s}_{k,i} \in \mathbb{R}^D$ pro Datensatz: $\mathbf{S}_k = [\mathbf{s}_{k,1}, \mathbf{s}_{k,2}, \dots, \mathbf{s}_{k,N_s}] \in \mathbb{R}^{D \cdot N_s}$

1. Eliminiere rigide oder affine Unterschiede zwischen den Daten – richte alle $\mathbf{S}_k$ zueinander aus.
2. Bilde mittleres Modell: $\bar{\mathbf{M}} = \frac{1}{N} \sum_{i=1}^{N_s} \mathbf{S}_k$ mit $\bar{\mathbf{m}}_i = \frac{1}{N} \sum_{k=1}^N \mathbf{s}_{k,i}$
3. Ermittle Variationsmoden z. B. mittels Hauptkomponentenanalyse (PCA):
 - Berechne Kovarianzen: $C = \frac{1}{N-1} \sum_{k=1}^N (\mathbf{S}_k - \bar{\mathbf{M}})(\mathbf{S}_k - \bar{\mathbf{M}})^T$
 - Berechne Eigenvektoren $\mathbf{V}_p \in \mathbb{R}^{D \cdot N_s}$ und Eigenwerte λ_p der Kovarianzmatrix und sortiere $\mathbf{V}_p$ absteigend nach λ_p , mit $p = 1, \dots, N - 1$
4. Entferne die Eigenvektoren (Moden) mit den kleinsten Eigenwerten, um Overfitting und Modellierung von Rauschen zu vermeiden.

Sind die Modellparameter ($\bar{\mathbf{M}}, \mathbf{V}_p, \lambda_p, p = 1, \dots, n$) gegeben, können neue Instanzen $\mathbf{M} \in \mathbb{R}^{D \cdot N_s}$ durch die Gewichtung (“Formparameter”) der ersten n Eigenvektoren mit $\omega_p \in \mathbb{R}$ erzeugt werden:

$$\mathbf{M} = \bar{\mathbf{M}} + \sum_{p=1}^n \mathbf{V}_p \cdot \omega_p, \text{ mit } \mathbf{m}_i = \bar{\mathbf{m}}_i + \sum_{p=1}^n \mathbf{v}_{p,i} \cdot \omega_p.$$

Um die Modellinstanz $\mathbf{M}$, welche einem neuen unbekannten Bild am genauesten entspricht, zu bestimmen, müssen im Bild zuerst Landmarkenkandidaten $\mathbf{S}_{new}$ definiert werden. An diese Kandidaten kann anschließend das Modell mittels Parameteroptimierung angepasst werden:

$$|\mathbf{S}_{new} - T(\bar{\mathbf{M}} + \sum_{p=1}^n \mathbf{V}_p \cdot \omega_p)| \rightarrow \min_{T, \omega_p}.$$

T stellt dabei eine weitere Anpassung mittels rigider (oder affiner) Transformation dar, um grobe lineare Unterschiede zwischen Modell und neuem Bild auszugleichen. Der Wertebereich von ω_p wird meist durch die Standardabweichung der Moden beschränkt: $|\omega_p| \leq 3\lambda_p$. Um die Qualität des formmodellbasierten Segmentierungsverfahrens zu maximieren, müssen zwei grundlegende Probleme gelöst werden:

Problem 1: Wo befinden sich die korrekten Landmarken im Trainingsdatensatz? Eine manuelle Definition der Landmarken in den Bilddaten ist durch eine potentiell sehr hohe Anzahl an notwendigen Landmarken nicht praktikabel. Außerdem ist die Reproduzierbarkeit von Ergebnissen für eine automatische Analyse von Daten wichtig. Dieses liefert eine manuelle Definition nicht. Daher wurden verschiedene automatische Methoden entwickelt, um diese korrespondierenden Landmarken zu bestimmen bzw. diese anschließend von einem Bild auf einen ganzen Trainingsdatensatz zu übertragen. Die meisten dieser Methoden führen eine Registrierung zwischen den betrachteten Formen durch. Die einfacheren Methoden nutzen Iterative-Closest-Point(ICP)-basierte Registrierungsansätze, um die Landmarken einer Referenzform auf alle anderen Formen des Trainingsdatensatzes zu übertragen [Josephson *et al.* 2005; Vos *et al.* 2004]. Dieser Ansatz wurde durch eine Reihe von Erweiterungen (korrekte Wahl der Referenzform, Erweiterung der Distanzdefinition für die “closest point”-Berechnung) verfeinert [Brett & Taylor 2000b]. Da die ursprünglich rigide Variante der ICP-Algorithmen nicht für sehr formvariable Probleme geeignet ist, wurden nichtrigide Registrierungsalgorithmen zur Korrespondenzbestimmung entwickelt [Subsol *et al.* 1998; Ferrarini *et al.* 2007; Brett & Taylor 2000a]. Eine weitere Methode ist die Registrierung einer Form (Landmarkenmenge) direkt auf ein (manuell segmentiertes) binäres Maskenbild, um die Landmarken zu übertragen [Shang & Dossel 2004; Zhao & Teoh 2005]. Heitz *et al.* führen außerdem eine gleichzeitige Segmentierung und Korrespondenzbestimmung unter Nutzung eines Templates durch [Heitz *et al.* 2005]. Die Qualität dieser Verfahren hängt vollständig von einer korrekten Registrierung der Templateform ab. Weitere Varianten zielen auf eine erhöhte Robustheit des Templates ab [Shang & Dossel 2004; Dam *et al.* 2008]. Außerdem wurden atlasbasierte Verfahren genutzt, um im Atlas vordefinierte Landmarken mittels optimierter Deformationsfelder auf andere Bilder zu übertragen [Frangi *et al.* 2001; Heitz *et al.* 2004; Rueckert *et al.* 2003; Andreopoulos & Tsotsos 2008]. Zhang *et al.* beschreiben eine Kombination aus registrierungsbasierter Korrespondenzsuche und

Formmodellgenerierung [Zhang & Cootes 2012]. Neben der direkten Registrierung der vorliegenden Formen (landmarken- oder maskenbasiert) wurden Methoden entwickelt, welche die Formen auf einen anderen gemeinsamen Parameterraum abbilden, z. B. einen Kreis bzw. eine Kugel um in diesem Raum die Korrespondenzen zu bestimmen [Floater & Hormann 2005]. Davies et al. führten die Nutzung der *minimum descriptor length* (MDL) ein, um die Definition der “korrekten” Landmarken mit der Korrespondenzbestimmung für das gesamte Trainingsset zu kombinieren [Davies et al. 2001, 2002; Gollmer & Buzug 2013]. Dabei wird eine Landmarkenrepräsentation angestrebt, welche das resultierende Modell so effizient wie möglich gestaltet. Die genaue Definition und Berechnung der MDL Kostenfunktion ist nicht trivial und relativ rechenintensiv.

Problem 2: Wo befinden sich die besten Landmarkenkandidaten im neuen Bild für die Modellanpassung? Um ein Modell an das neue Bild anzupassen, muss die Signalwertverteilung bzw. Erscheinung (engl. *appearance*) der Struktur modelliert werden. Die ersten Modelle von Cootes et al. wählen dafür die Anpassung an die stärksten lokal gefundenen Gradienten im Bild [Cootes et al. 1995]. Eine andere Methode ist die Modellierung der Appearance-Werte in der Nähe der Landmarken: Hierfür wird pro Landmarke ein Signalwertprofil senkrecht zur Objektkontur extrahiert. Anschließend wird über den Trainingsdatensatz das mittlere Profil samt Variationsmoden pro Landmarke bestimmt. Während der Modellanpassung wird nach den Landmarkenkandidaten im Bild gesucht, welche ähnliche Appearance-Profile aufweisen [Cootes & Taylor 1994; Behiels et al. 2002]. In [Cristinacce & Cootes 2006] lernt das gesuchte statistische Modell mit den “Constrained Local Models” zusätzlich die Variationen von Templatreionen in der lokalen Umgebung jeder Landmarke, um diese robuster an die Bilddaten anpassen zu können. Eine andere genutzte Möglichkeit ist die Extraktion von Gabor-Wavelet-Merkmalen unter den Landmarken [Jiao et al. 2003; Shen et al. 2003]. Eine Alternative für die Modellierung der Appearance-Merkmale mittels eines weiteren Modells, ist die Nutzung eines k-Nearest-Neighbors-Ansatzes (kNN). Dabei werden die Profile unter den Landmarken des Trainingsdatensatzes sowohl an den korrekten Positionen als auch an leicht geänderten Positionen als Datenbasis für den kNN-Ansatz gesampelt [de Bruijne et al. 2003].

Wenn bestimmt ist, wie die besten Landmarkenkandidaten definiert sind (starke Gradientenbeträge, bestimmtes Signalwertprofil, ...), besteht weiterhin die Frage nach dem Suchalgorithmus für diese. Um z. B. die wahrscheinlichsten Landmarkenkandidaten für eine aktuelle Positionierung des “Constrained Local Models” im Bild zu lernen, wird in [Lindner et al. 2015] “Random Forest Regression” eingesetzt. Insbesondere in 3D kann der Suchraum (das gesamte Bild) sehr groß sein. Daher nutzen die meisten Methoden eine lokale Suche im Bereich der aktuellen Landmarkenpositionen des Modells im Bild. Durch diese lokale Suche ist die initiale Positionierung des Modells im neuen Bild ausschlaggebend für die Qualität der Modellanpassung. Viele Methoden nutzen eine einfache manuelle Initialisierung (Platzierung des Schwerpunktes des Modells im

Bild). Hug et al. schlägt sogar eine Initialisierung mittel manueller Definition mehrerer (drei bis vier) Kontrollpunkte im Bild vor [Hug *et al.* 2000]. Automatische Initialisierungsmethoden umfassen unter anderem affine Vorregistrierung [Fripp *et al.* 2005] oder generalisierte Hough-Transformationen [Ecabert *et al.* 2008].

1.2.2. Appearance-Modelle

Statistische Formmodelle sind auf Forminformationen beschränkt. Dies ist ausreichend, wenn die betrachtete Struktur durch starke Kanten (hohe Gradientenwerte oder aussagekräftige Bildinformationen nahe der Objektgrenze) in den Bilddaten charakterisiert wird. Für Objekte mit mehreren Appearance-Merkmalen innerhalb des Objekts oder mit schlecht definierten Kanten, ist es von Interesse, zusätzlich die Appearance-Informationen im Inneren des Objekts zu modellieren. Dafür wurden die “Aktiven Appearance-Modelle” (AAM) von Cootes et al. [Cootes *et al.* 2001] eingeführt, die eine Erweiterung der SFMs darstellen und neben den Forminformationen außerdem Appearance-Informationen des betrachteten Objekts beinhalten.

Die Appearance-Modelle setzen ebenfalls die Existenz von Eins-zu-Eins-Korrespondenzen voraus. Um die Appearance-Informationen unabhängig von der Form jedes Trainingsbildes zu extrahieren und anschließend zu modellieren, werden alle Bilder auf die mittlere Form gewarpt. Ein AAM wird in drei wesentlichen Schritten generiert:

1. Identifikation der korrespondierenden Landmarken in allen Trainingsbildern und Generierung eines statistischen Formmodells,
2. Warpen aller Trainingsbilder zur mittleren Form, um die Formunterschiede zu eliminieren und Abtastung aller Signalwerte an korrespondierenden Positionen der gewarpten Bilder, um ein Appearance-Modell zu bauen und
3. Kombination beider Modelle (z. B. durch eine weitere PCA auf den (gewichteten) Variationsmoden von Form und Appearance).

Siehe [Cootes *et al.* 2001] für eine detailliertere Beschreibung der einzelnen Schritte. Für die Modellanpassung ist ebenfalls ein Warming-Schritt für die Formnormalisierung notwendig.

Das erste AAM von Cootes et al. modelliert alle Appearance-Werte (abhängig von einer bestimmten Bildauflösung) innerhalb der Form eines Modells. Weitere Arbeiten modellieren die Appearance-Werte auf spärliche Weise. Dennoch werden Form- und Appearance-Informationen weiterhin getrennt voneinander betrachtet und anschließend kombiniert (wie beim klassischen AAM) [Donner *et al.* 2007; Cristinacce & Cootes 2008] oder das Appearance-Modell basiert auf Merkmalsdeskriptoren, wobei die Positionen der Merkmalsvektoren durch komplexe lokale Deskriptoren ausgewählt werden [Cristinacce & Cootes 2006; Larsen *et al.* 2007].

Das Hauptproblem bei der Anwendung der AAMs ist die Anpassung des Modells an ein neues unbekanntes Bild. Da Form und Appearance zunächst einzeln abgetastet und

modelliert werden, muss bei der Modellanpassung von der Grauwertdifferenz zwischen Testbildsignalwerten und Modellinstanz auf eine Modifikation von Modellparametern (Form und Appearance-Koeffizienten) geschlossen werden. Im originalen Ansatz von Cootes et al. werden z. B. mittels linearer Regression die Updateschritte für die Modellparameter abhängig von dem aktuellen Fehler zwischen Bild und Modell gelernt [Cootes et al. 2001]. Optimierungsbasierte Ansätze hingegen sollen das Problem analytisch lösen z. B. mittels Gauß-Newton-Optimierung [Tzimiropoulos & Pantic 2014] bzw. mittels der optimierten “(fast) Simultaneous Inverse Compositional”-Methode [Tzimiropoulos & Pantic 2016].

Zusammenfassung der Probleme der klassischen statistischen Modelle:

- Aufwendige Positionierung der Landmarken und Korrespondenzbestimmung zwischen diesen,
- Abhängigkeit der Modellqualität von den Landmarkenpositionen,
- unabhängige Algorithmen für Korrespondenzbestimmung, Modellgenerierung, Modellinitialisierung und Modellanpassung,
- unabhängige Modellierung von Form- und Appearance-Werten,
- Abhängigkeit der Qualität der Modellanpassung von der Initialisierung des Modells im Bild.

1.2.3. Probabilistische Korrespondenzen

Die grundlegende Annahme der Eins-zu-Eins-Korrespondenzen zwischen allen Trainingsdaten und zwischen Modell und neuem Bild führt wie erwähnt zu Problemen, sollten Bereiche nicht eindeutig korrespondieren bzw. nicht vorhanden sein. Punktbasierte Registrierungsansätze wie ICP Algorithmen stoßen auf dasselbe Problem der potentiell nicht-existenten Punkte bzw. Korrespondenzen zwischen zwei Punktmengen. Daher wurden Methoden für punktbasierte Registrierungen entwickelt, die Korrespondenzwahrscheinlichkeiten anstelle von Eins-zu-Eins-Korrespondenzen nutzen. Diese wahrscheinlichkeitsbasierten Ansätze umfassen den Expectation-Maximization-Iterative-Closest-Point (EM-ICP) Algorithmus von Granger et al. [Granger & Pennec 2002], den Softassign [Gold et al. 1997] oder den Coherent Point Drift Algorithmus [Myronenko & Song 2010] und andere nichtrigide Ansätze [Chui & Rangarajan 2003; Jian & Vemuri 2011].

Werden die Bilder/Objekte nicht mehr von Merkmalen an genau definierten Positionen im Objekt repräsentiert, handelt es sich nicht mehr um Landmarken, sondern lediglich um *Punkt-* bzw. *Merkmalsvektormengen*.

In Hufnagel et al. [Hufnagel et al. 2008] wurde die Nutzung der probabilistischen Korrespondenzen zwischen den Punktmengen der Objekte des Trainingsdatensatzes vorgeschlagen, um ein Formmodell zu generieren. Die Idee dabei ist, die Ungewissheit um

die Punktpositionen mit zu modellieren und außerdem die Kosten für die Bestimmung der “korrekten” Landmarken zu reduzieren. Doch mit der Nutzung der probabilistischen Korrespondenzen kann der klassische PCA-basierte Ansatz, ein statistisches Modell zu generieren, nicht mehr angewendet werden. Daher wurde eine Maximum-A-posteriori (MAP) Methode entwickelt, welche die Modellparameter (mittleres Modell und Variationsmoden) abschätzt, mithilfe der Minimierung eines einzelnen globalen Kriteriums. Ein anderes Formmodell, das auf der iterativen Optimierung von probabilistischen Korrespondenzen und Punktpositionen basiert, wurde von Yu et al. [Yu et al. 2014] vorgeschlagen. Dabei handelt es sich um ein hierarchisches Modell, welches im ersten Level die Formmerkmale einer gesamten Population beschreibt, im zweiten Level die Merkmale von Untergruppen der Population (z. B. unterschiedliche Krankheiten) und im dritten Level die individuellen Formvariationen pro Gruppe modelliert. Der Ansatz schlägt eine gemeinsame Optimierung der Korrespondenzen, Punktpositionen und Modellparameter vor.

1.2.4. Atlas- und registrierungsbasierte Verfahren

Die Idee, Appearance-Modelle für die Segmentierung oder Klassifikation eines zuvor trainierten Objekts zu nutzen, ist ähnlich zu atlasbasierten Ansätzen. Sogenannte Atlanten können bei der Segmentierung von Bilddaten eingesetzt werden – dabei wird der Atlas (das Templatebild) auf ein neues unbekanntes Bild registriert und zusätzliche im Atlas vorhandene Informationen (z. B. Segmentierungslabel) werden anschließend auf das neue Bild übertragen. Einfachste Atlanten bestehen lediglich aus einem Templatebild (eine Instanz der Population) samt entsprechenden Segmentierungsinformationen. Aufwendigere Verfahren ermitteln hingegen das optimale Templatebild (durch Mittelung des Durchschnittsbildes über eine Population oder durch Bestimmung der ähnlichsten Kandidaten der Population zum neuen zu segmentierenden Bild). Bei noch komplexeren Verfahren wie [Allasonnière et al. 2007; Xu et al. 2014] wird zusätzlich zum optimalen/ mittleren Templatebild auch die Deformationsvariabilität einer zugrundeliegenden Population im Atlas ermittelt und gespeichert. Bei der Registrierung des Atlanten auf ein neues Bild sind die gesuchten Deformationen durch die im Atlas vorhandene Deformationsvariabilität eingeschränkt, was wiederum dem in dieser Arbeit vorgestellten modellbasierten Verfahren ähnelt.

Neben der Güte des Templatebildes bildet die Registrierung des (beliebig komplexen) Atlanten auf ein neues Bild einen für die Qualität der Segmentierung ausschlaggebenden Faktor. Sowohl die meisten atlasbasierten als auch modellbasierten Verfahren, die auf der Registrierung bzw. Anpassung des Atlanten/Modells auf neue Daten basieren, arbeiten mit der Annahme, dass alle Klassen bzw. Bildregionen in den neuen Daten existieren. Die Herausforderung der Registrierung bzw. Adaption eines Bildes mit Artefakt/Pathologie/Überdeckung auf ein Bild ohne “Störung” ist noch nicht umfassend gelöst. Die einfachsten Ansätze nutzten eine (semi-)manuelle Segmentierung der Pathologie zur Maskierung, um ausschließlich die Bereiche, welche in beiden Bildern vor-

handen sind, mittels klassischer Registrierungsverfahren zu registrieren [Rohlfing *et al.* 2003; Staring *et al.* 2007]. Des Weiteren kann eine bekannte “korrekte” Deformation zwischen beiden Bildern genutzt werden, um nichtkorrespondierende Regionen zu identifizieren. Daher wurden ebenfalls Methoden entwickelt, die das gemeinsame Segmentierungs-Registrierungsproblem lösen, indem Expectation-Maximization-Methoden zur abwechselnden Schätzung der Korrespondenzen/Maskierung und der Deformationen eingesetzt werden [Chitphakdithai & Duncan 2010; Kwon *et al.* 2014]. Parisot *et al.* löst beide Probleme hingegen simultan [Parisot *et al.* 2014]. Darüber hinaus existieren erste Ansätze, die versuchen, pathologische Prozesse spezifisch und aufwendig zu modellieren/simulieren [Zacharaki *et al.* 2008; Mang *et al.* 2012] oder mittels kombinierter merkmals- und intensitätsbasierter Verfahren die Robustheit gegenüber fehlenden Korrespondenzen zu erhöhen [Hartkens *et al.* 2002].

1.3. Ziele der Arbeit

Die Arbeit präsentiert einen neuen Ansatz für “probabilistische Appearance-Modelle” unter der Nutzung von Korrespondenzwahrscheinlichkeiten anstelle von Eins-zu-Eins-Korrespondenzen. Eines der Ziele ist es, die Kosten für die Vorverarbeitung und aufwendige Landmarkenbestimmung und Korrespondenzfindung zu minimieren bzw. ganz zu eliminieren. Außerdem wird eine vielseitig einsetzbare Methode angestrebt, die auf einer mathematisch geschlossenen Formulierung eines Optimierungskriteriums basiert und dadurch nicht aus verschiedenen Schritten zusammengesetzt ist, wie bei klassischen Modellgenerierungsmethoden ((1) Landmarken/Korrespondenz-Bestimmung, (2) SFM Generierung, (3) Warping, (4) AAM Generierung).

Die Methode basiert auf den Vorarbeiten von Hufnagel *et al.*, in denen ein erster Ansatz zur Generierung und Nutzung eines probabilistischen Formmodells entwickelt wurde [Hufnagel *et al.* 2008, 2009]. Im Rahmen der vorliegenden Dissertation wird das probabilistische Formmodell um Appearance-Merkmale erweitert. Um eine unabhängige Betrachtung von Form und Appearance zu vermeiden, wird zuerst eine spärliche merkmalsbasierte Bildrepräsentation eingeführt, die Positions- und Appearance-Merkmale in Merkmalsvektoren kombiniert. Dadurch wird die Form eines Objekts nicht mehr explizit durch Landmarkenpositionen modelliert, sondern implizit über Appearance-Werte, was zu einer flexiblen Einsatzmöglichkeit des Modellansatzes führt. Für die Optimierung der Modellparameter wird ein Maximum-A-posteriori-Verfahren eingesetzt, bei dem außerdem die probabilistischen Korrespondenzen mittels eines Expectation-Maximization-Ansatzes zwischen den Daten des Trainingsdatensatzes mit optimiert werden. Dabei wird wie bei den klassischen Modellen aus einem gegebenen Trainingsdatensatz aus Bildrepräsentationen eine mittlere Repräsentation samt Variationsmoden berechnet. Um die Modellparameter zu optimieren, wird ein globales Kriterium minimiert [Ehrhardt *et al.* 2014; Krüger *et al.* 2015a, 2017].

Ein weiterer Vorteil des vorgestellten Ansatzes ist die Anpassung des Modells an unbekannte Bilder mittels Minimierung des gleichen Optimierungskriteriums, das für

die Modellgenerierung genutzt wird.

Ein konzeptuelles Problem klassischer Form- und Appearance-Modelle ist eine fehlende Annahme über die zugrundeliegende Deformation, welche das Modell an eine Observation anpasst. Einige Ansätze führen topologische Bedingungen als Regularisierung für die Modelldeformation bei statistischen Formmodellen ein [Cootes *et al.* 2008; Fletcher *et al.* 2003]. Der vorgeschlagene MAP-basierte Ansatz bietet eine einfache Möglichkeit, Nebenbedingungen als nichtkonstante A-priori-Wahrscheinlichkeit des Modells einzuführen. Dafür wird das globale Kriterium durch einen Regularisierungsterm erweitert, der un plausible Transformationen des Modells bestraft [Krüger *et al.* 2016]. Um die Modellgenerierung robuster zu gestalten, wird eine Regularisierung notwendig, weil durch die Kombination der Positions- und der Appearance-Merkmale eine hohe Anzahl von Freiheitsgraden in das Modell eingeführt wird.

Die Einsatzmöglichkeiten der vorgestellten Methode sind vielseitig. Die ursprüngliche Intention statistischer Formmodelle ist die Segmentierung von Objekten. Da die Forminformation nicht mehr explizit durch die Positionen der Merkmalsvektoren repräsentiert wird, wurde eine Repräsentation der Segmentierungsinformationen über zusätzliche Appearance-Merkmale eingeführt. Es können Kontur- oder Landmarkeninformationen über Distanzkartenwerte in der genutzten Bildrepräsentation dargestellt werden. Die Wahl, Art und Anzahl der Appearance- oder Segmentierungsmerkmale ist in keiner Weise eingeschränkt, was einerseits die Multi-Objektsegmentierung bereits in dem Framework beinhaltet und andererseits die Anwendung nicht auf Segmentierungsprobleme einschränkt.

Die Nutzung probabilistischer Korrespondenzen verringert nicht nur die notwendigen Vorverarbeitungsschritte und eliminiert den Schritt der Landmarkenbestimmung, sondern liefert darüber hinaus Informationen über die Qualität der Modellanpassung. Das kann zum einen Auskunft darüber geben, an welchen Stellen bzw. für welche Fälle das Modell nicht ausreichend Variabilität aufweist. Zum anderen kann diese Information genutzt werden, um z. B. pathologische oder artefaktbelastete Regionen in medizinischen Bilddaten zu lokalisieren [Krüger *et al.* 2015b]. Sind solche Bereiche identifiziert, können die Appearance-Werte des Modells genutzt werden, um Bildartefakte im Bild zu rekonstruieren.

Wird ein gegebenes Form- oder Appearance-Modell an ein neues Bild angepasst, kann dies auch als “Registrierung” des Modells auf ein neues Bild betrachtet werden. Bei der klassischen Registrierung sind die Korrespondenzen zwischen beiden Bildern unbekannt und werden über die Optimierung der Deformation (während der Registrierung) gesucht. Dies wiederum ist ähnlich dem Modellgenerierungsprozess des vorgestellten Ansatzes, da im Gegensatz zu klassischen Modellansätzen hier die Korrespondenzen ebenfalls während der Optimierung gesucht werden. Das heißt, das in der Arbeit beschriebene Modellgenerierungsverfahren kann zum Registrierungsverfahren zweier Bilder vereinfacht werden. Der Vorteil eines solchen probabilistischen Registrierungsalgorithmus ist die Robustheit gegenüber fehlerhaften/pathologischen Regionen, die ausschließlich in einem der beiden Bilder vorkommen. Außerdem werden solche Gebiete (z. B. Tumor- oder Schlaganfallge-

webe) durch die genutzten Korrespondenzwahrscheinlichkeiten automatisch indiziert.

1.3.1. Zusammenfassung der mit dieser Arbeit erreichten Ziele

Die vorliegende Dissertation beschreibt die Entwicklung eines probabilistischen Appearance Modells unter Nutzung **probabilistischer Korrespondenzen** mit den Vorteilen:

- keine kostenaufwendige Vorverarbeitung und Bestimmung von Landmarken und Eins-zu-Eins-Korrespondenzen in den Trainingsdaten,
- Robustheit gegenüber nicht vorhandener korrespondierender Strukturen in den Bilddaten,
- zusätzliche Informationen über die Qualität des Modells bzw. der Modellanpassung – lokal geringe Korrespondenzwahrscheinlichkeiten indizieren “nicht im Modell vorhandene” Strukturen.

Die Methode basiert auf einer **geschlossenen mathematischen Formulierung** eines globalen Optimierungskriteriums mit den Vorteilen:

- keine Abhängigkeit der Ergebnisse von einzelnen unabhängigen aneinandergereihten Verfahren,
- Modellgenerierung und -anpassung werden mit demselben Verfahren durchgeführt,
- alle Parameter werden entweder über das Kriterium optimiert oder datengetrieben an die aktuelle Situation während der Optimierung angepasst – kein manuelles Tuning von Parametern.

Die Grundlage der Modellformulierung bildet eine Datenrepräsentation über eine **gemeinsame Definition von Positions- und Appearance-Merkmalen** in einer Merkmalsvektormenge mit den Vorteilen:

- uneingeschränkte Anzahl und Art von beschreibbaren Merkmalen,
- Bearbeitung beliebigdimensionaler Bilder,
- keine Beschränkung auf Segmentierungsproblematik.

Um die Robustheit der Korrespondenzbestimmung weiterhin zu steigern, werden in der Arbeit **verschiedene Erweiterungen** des Ansatzes zur Regularisierung vorgestellt:

- lokale Regularisierung durch Einführung von Nachbarschaftsannahmen zwischen Merkmalsvektoren,
- Multi-Level-Ansatz, bei dem die Anzahl der Merkmalsvektoren in Modell und Trainingsdaten sukzessive inkrementiert wird,

- sukzessives Inkrementieren der Anzahl der Moden.

Der Modellgenerierungsansatz wurde zu einer **probabilistischen Registrierung** zweier Bilder vereinfacht, wobei eine solche Registrierung folgende Vorteile liefert:

- Robustheit gegenüber pathologischen Bereichen, die nicht in beiden Bildern vorhanden sind, durch die Nutzung der Korrespondenzwahrscheinlichkeiten mit gleichzeitiger Indizierung solcher Bereiche,
- beliebigdimensionale Deformationsfelder: Im Gegensatz zur klassischen Registrierung werden nicht nur die Verschiebungsfelder der Bildposition bestimmt, sondern ebenfalls die Signalwertunterschiede beider Bilder ausgeglichen.

1.3.2. Publikationen

Die Ergebnisse der Dissertation wurden in folgenden Publikationen veröffentlicht. Die Grundlagen des probabilistischen Appearance-Modells wurden in [Ehrhardt *et al.* 2014] und [Krüger *et al.* 2015a] publiziert. In [Krüger *et al.* 2015b] wird neben der modellbasierten Segmentierung die Anwendung des Modellansatzes zur Klassifikation neuer Bilddaten in gesunde und pathologische Bildbereiche vorgestellt. Die Erweiterungen des Modells um lokale Nachbarschaftsannahmen und die Nutzung des Multi-Level-Ansatzes zur Steigerung der Robustheit wurden in [Krüger *et al.* 2016] veröffentlicht. Und [Krüger *et al.* 2017] fasst den kompletten Modellansatz umfassend zusammen.

1.3.3. Gliederung der Arbeit

Zuerst wird im Abschnitt 2 die genutzte Bildrepräsentation von den klassischen Form- und Appearance-Repräsentationen abgegrenzt und die probabilistischen Korrespondenzen zwischen zwei Mengen unabhängiger Merkmalsvektoren eingeführt. Außerdem wird der Expectation-Maximization(EM)-Ansatz zur Optimierung der Korrespondenzen und Parameter erläutert. Im Abschnitt 3 werden die genutzten Modellparameter und das zur Optimierung der Parameter eingesetzte globale Kriterium vorgestellt. Die Minimierung des Kriteriums sowohl zur Modellgenerierung als auch zur Modellanpassung wird in Abschnitt 4 beschrieben. Abschnitt 5 führt verschiedene Erweiterungen wie den Multi-Level-Ansatz ein, die zur robusteren Modellgenerierung führen und Abschnitt 6 stellt mehrere unterschiedliche Modellanwendungen vor. Die probabilistische Registrierung, die vom vorgestellten Modellansatz abgeleitet werden kann, wird in Abschnitt 7 beschrieben. Abschnitt 8 stellt Experimente und Ergebnisse für angewandte Modelle und Registrierungen unterschiedlicher Bilddaten vor, während im Abschnitt 9 das vorgestellte Framework abschließend diskutiert wird.

2. Objektrepräsentation und Korrespondenzwahrscheinlichkeiten

2.1. Bildrepräsentation aus unabhängigen Merkmalsvektoren

Die Grundlage der klassischen landmarkenbasierten statistischen Form- und Appearance-Modelle ist eine punktbasierte Datenrepräsentation. Dabei wird ein Bild $I : \Omega \rightarrow \mathbb{R}, \Omega \subset \mathbb{R}^{D_x}$ der Dimension D_x bzw. das Objekt im Bild I durch eine Menge von N_s Merkmalsvektoren $\mathbf{s}_i \in \mathbb{R}^{D_x}$ dargestellt. Im Gegensatz zur klassischen Formformulierung, bei der die Form als Vektor $[\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{N_s}]$ dargestellt wird und jeder Punkt über den Index i an eine feste Position (Landmarke) im Objekt gebunden ist, werden in dieser Arbeit die Merkmalsvektoren als unabhängige Mengen beliebiger Größe betrachtet. Die Bild-/Objektrepräsentation sei demnach definiert als die Menge $\mathbf{S} = \{\mathbf{s}_i | i = 1, \dots, N_s\}$. Abbildung 1 stellt die klassischen Form- bzw. Appearance-Repräsentationen der Repräsentation mittels unabhängiger Merkmalsvektoren gegenüber.

Grundlage einer statistischen Modellierung ist ein vorliegender Trainingsdatensatz von N Bildrepräsentationen $\mathbf{S}_{\text{tr}} = \{\mathbf{S}_k | k = 1, \dots, N\}$ des zu modellierenden Objekts. Die Trainingsmenge sollte dabei idealerweise die gesamte (Form-)Variationsbreite des Objekts abbilden. Das gewünschte Modell besteht im Regelfall aus einem mittleren Modell $\bar{\mathbf{M}} = \{\bar{\mathbf{m}}_j | j = 1, \dots, N_m\}$ (ebenfalls eine Menge aus Merkmalsvektoren mit $\bar{\mathbf{m}}_j \in \mathbb{R}^{D_x}$) und einer Anzahl an Variationsmoden $\mathbf{V} = \{\mathbf{V}_p | p = 1, \dots, n\}$ mit $\mathbf{V}_p = \{\mathbf{v}_{p,j} | j = 1, \dots, N_m\}$ und $\mathbf{v}_{p,j} \in \mathbb{R}^{D_x}$, welche die Variationen des Trainingsdatensatzes abbilden.

Neue Modellinstanzen werden durch das mittlere Modell und einer Linearkombination der Moden erzeugt:

$$\mathbf{M} = \bar{\mathbf{M}} + \sum_{p=1}^n \omega_p \cdot \mathbf{V}_p. \quad (1)$$

Die Addition der Mengen $\bar{\mathbf{M}}$ und $\mathbf{V}_p$ sei als Addition der Merkmalsvektoren mit identischem Index j definiert:

$$\mathbf{m}_j = \bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_p \cdot \mathbf{v}_{p,j}. \quad (2)$$

Dies bedeutet, dass die Mengen $\bar{\mathbf{M}}$ und $\{\mathbf{V}_p, p = 1, \dots, n\}$ dieselbe Anzahl an Merkmalsvektoren aufweisen. $\omega_p, p = 1, \dots, n$ stellt eine Gewichtung der jeweiligen Variationsmode dar und ist somit für die individuelle Manipulation des Modells verantwortlich.

Im Gegensatz zur klassischen individuellen Modellierung von Form und Appearance, werden Positions- und Appearance-Merkmale in dieser Arbeit abhängig voneinander betrachtet. Jeder Merkmalsvektor umfasst die Position $\mathbf{x}_i \in \Omega$ und eine beliebige An-

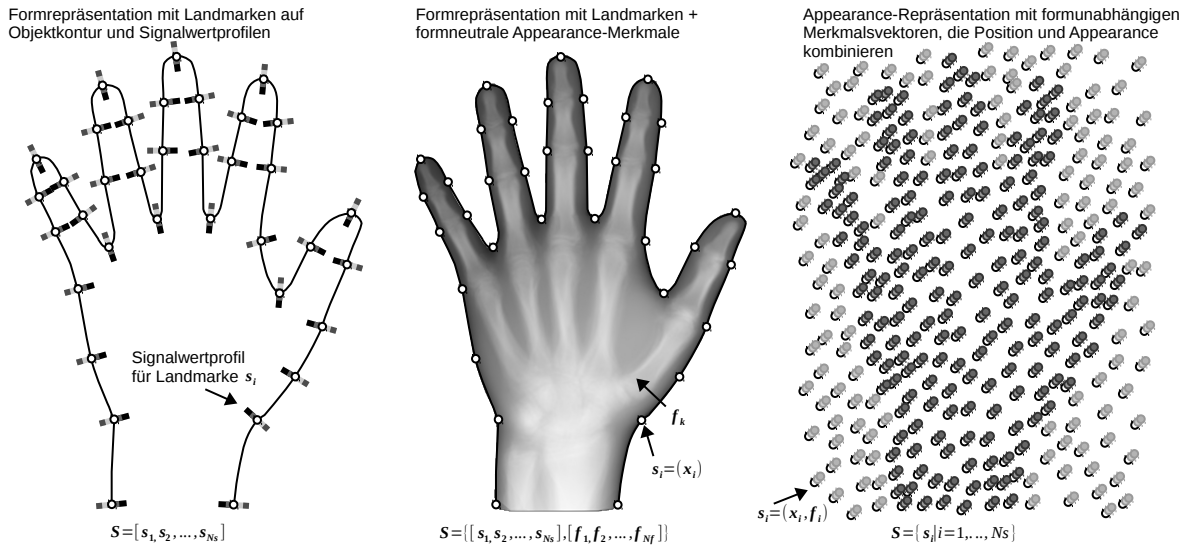


Abbildung 1: Gegenüberstellung von Bildrepräsentationen für statistische Modellierungen: Klassische Formmodelle basieren auf Landmarken, welche auf der Kontur des gesuchten Objekts lokalisiert sind. Das Modell wird an ein darunterliegendes Bild über Signalwertprofile angepasst. Bei klassischen Appearance-Modellen werden zusätzlich zu diesen Landmarken (für das Formmodell) die gesamten Signalwerte innerhalb der Objektkontur gesampelt, sodass zwei unterschiedliche Vektoren (Landmarken und Appearance-Werte) vorliegen. Der vorgestellte probabilistische Modellansatz basiert auf einer Menge an Vektoren, bei der pro Merkmalsvektor Positions- und Appearance-Werte gemeinsam modelliert werden.

zahl D_f an Appearance-Merkmalen $\mathbf{f}_i$, die zu der Bildposition $\mathbf{x}_i$ korrespondieren. Ein Merkmalsvektor ist dabei definiert als $\mathbf{s}_i = (\mathbf{x}_i, \mathbf{f}_i)$ mit $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,D_x})$ und $\mathbf{f}_i = (f_{i,1}, \dots, f_{i,D_f})$ und hat die Dimension $D = D_x + D_f$. Dabei ist zu beachten, dass die Dimension D beliebig groß sein kann. Analog zu $\mathbf{s}_i \in \mathbb{R}^{D=D_x+D_f}$ gilt auch $\bar{\mathbf{m}}_j \in \mathbb{R}^D$, $\mathbf{m}_j \in \mathbb{R}^D$ und $\mathbf{v}_{p,j} \in \mathbb{R}^D$. Je nach Wahl bzw. Generierung der Appearance-Merkmale kann das Eingabebild von nun an als "mehrkanal Bild" $\mathbf{F} = [I^{f_1}, \dots, I^{f_{D_f}}] : \Omega \rightarrow \mathbb{R}^{D_f}, \Omega \subset \mathbb{R}^{D_x}$, welches alle Appearance-Bilder zusammenfasst, betrachtet werden.

Der Frage nach der Positionierung der Merkmalsvektoren und nach der Wahl der Appearance-Merkmale wird in Abschnitt 8 nachgegangen. Zunächst sei nur bemerkt, dass im Gegensatz zu klassischen Form- und Appearance-Modellen die Positionen der Merkmalsvektoren unabhängig von der Form des zu untersuchenden Objekts gewählt werden. Das heißt, die naivste Definition der Merkmalsvektoren wäre eine gleichmäßige oder randomisierte Abtastung des Bildes unter Verwendung der Signalwerte als Appearance-Merkmale.

Bei der Abtastung von Daten ist es in den meisten Fällen sinnvoll einen Durchschnitts- bzw. Mindestabstand zwischen jeweils zwei Vektoren einzuhalten, um eine gleichmäßige Repräsentation der Daten zu gewährleisten. Bei einer gerasterten Abtastung, würde dieser Abstand lediglich der Schrittweite der Abtastung entsprechen. In der Arbeit wird

der Mindestabstand zwischen den Vektoren mit sp (engl. spacing) bezeichnet.

Da die Merkmalsvektoren beliebig gewählt werden können und vor der Modellgenerierung keine Korrespondenzen zwischen den Bildrepräsentationen des Trainingsdatensatzes bekannt sind, muss im Folgenden die Frage nach der Korrespondenz zwischen unabhängigen Merkmalsvektormengen geklärt werden.

Beispielbox 1: Bildrepräsentation durch Merkmalsvektoren

Zur Veranschaulichung werden im Laufe der vorliegenden Arbeit immer wieder generierte Appearance-Modelle der Hand verwendet.

Daten: Grundlage dieser Modelle bilden 120 2D Röntgenbilder der *Digital Hand Atlas Database*^a. 60 dieser Bilder werden zum Training der jeweiligen Modelle und 60 weitere Bilder als Testfälle genutzt. Es werden Bilder von 16- bis 18-jährigen männlichen und weiblichen Jugendlichen ausgewählt. Alle Bilder werden auf einen Pixelabstand von $0.5 \times 0.5 \text{ mm}$ downgesampelt. Dabei ist zu beachten, dass die unterschiedlichen Bilder nicht gleichmäßig beschnitten sind, was zu unterschiedlichen Bildgrößen führt. Die durchschnittliche Bildgröße liegt bei $332.90 (\pm 43.83) \times 460.13 (\pm 35.84)$ Pixel. Der Ursprung des Koordinatensystems aller Bilder wird daher in die jeweilige Mitte der Bilder verschoben. Abbildung 2 zeigt fünf Beispiele der verwendeten Daten. Artefakte (‘‘schwarze Boxen’’ oder Beschriftungen im Bild) sind nicht entfernt und unterschiedliche Signalwert-Verteilungen ebenfalls nicht ausgeglichen worden.

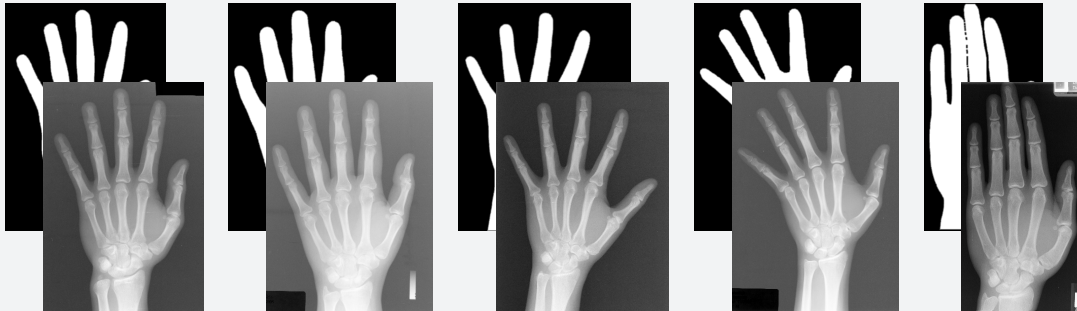


Abbildung 2: Fünf Beispiele der verwendeten Daten samt erzeugter Handmasken.

Wahl der Appearance-Merkmale: Als Appearance-Merkmale wird für die einfachen Beispielmolelle als erstes Merkmal die Gauß-geglätteten Signalwerte der Bilder an den entsprechenden Positionen genutzt:

$$f_{i,1} = (G_{\sigma=6mm} * I_k)(\mathbf{x}_i). \quad (3)$$

Als zweites Merkmal wird die euklidische Distanz vom jeweiligen Merkmalsvektor zur Handkontur in jedem Bild eingesetzt. Dafür wurden in allen Bildern die Hände vorher segmentiert. Zu jedem Bild I_k liegt also die Handmaske $I_k^{Mask} \in \{0, 1\}$ vor, mit $I_k^{Mask}(\mathbf{x}) = 1$, wenn $\mathbf{x}$ innerhalb der Hand liegt und $I_k^{Mask}(\mathbf{x}) = 0$ andernfalls. Das zweite Appearance-Merkmal sei hier demnach definiert als:

$$f_{i,2} = d(\mathbf{x}_i, I_k^{Mask}) = \begin{cases} + \min_{\mathbf{x} \in \Omega} d(\mathbf{x}_i, \mathbf{x}) & \text{if } I_k^{Mask}(\mathbf{x}_i) = 0 \wedge I_k^{Mask}(\mathbf{x}) = 1 \\ - \min_{\mathbf{x} \in \Omega} d(\mathbf{x}_i, \mathbf{x}) & \text{if } I_k^{Mask}(\mathbf{x}_i) = 1 \wedge I_k^{Mask}(\mathbf{x}) = 0, \end{cases} \quad (4)$$

mit $d(.,.)$ als euklidische Distanz zwischen zwei Vektoren. Dieses Merkmal wird in nachfolgenden Beispielboxen zur Segmentierung der Daten eingesetzt.

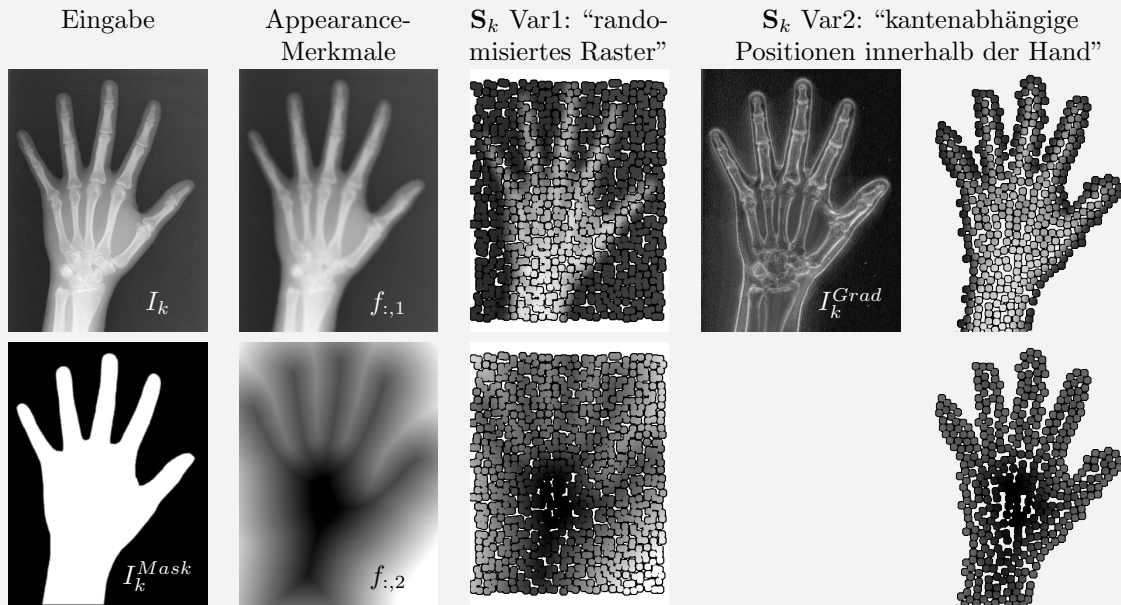


Abbildung 3: Beispiel für die Bildrepräsentation: Die erste Spalte stellt die Eingabedaten dar, die zweite die beiden genannten Appearance-Merkmale für jede mögliche Position im Bild und die letzten beiden Spalten zeigen die beiden beschriebenen Varianten, die Positionen der Merkmalsvektoren zu wählen.

Wahl der Positionen der Merkmalsvektoren: Für die Wahl der Positionen $\mathbf{x}_i$ der Merkmalsvektoren seien hier zwei Varianten genannt:

Variante 1: Das gesamte Bild wird abgetastet. Hierbei wird ein Raster mit einem Punktabstand (*sp*, engl. spacing) zwischen den Merkmalsvektoren von 6 *mm* genutzt. Die Positionen wurden um dieses Raster mit $\pm 2\text{mm}$ randomisiert, um eine höhere Variation zu gewährleisten. Diese Variante der Positionswahl hat den Vorteil, dass

sie sehr einfach und komplett unabhängig vom betrachteten Objekt ist. Ein Nachteil hingegen ist eine große Anzahl an “nicht interessierenden” Merkmalsvektoren im z. B. Bildhintergrund, welche nicht die Hand selber beschreiben. Je nach Bildgröße werden 884.55 ± 83.91 Merkmalsvektoren erzeugt.

Variante 2: Da auf Grundlage der Bildrepräsentationen Modelle generiert werden sollen und die Haltung (z. B. Fingerspreizung) der Hände zwischen den Daten zum Teil stark variiert, was zu unterschiedlich großen “Fingerzwischenräumen” führt, ist es sinnvoll, nur Merkmalsvektoren innerhalb bzw. in der Nähe der Hand zu wählen. Dies wird über das zweite Appearance-Merkmal (Distanz zur Handkontur) erreicht: $\forall i f_{i,2} < 5mm$. Damit in allen Trainingsdaten die Merkmalsvektoren an ähnlichen Strukturen zu finden sind, werden hier die Positionen abhängig von ihrer absteigenden “Distinctiveness”/Aussagekraft (Abschnitt 8.2) gewählt – unter der Bedingung, dass zwischen den Positionen ein Mindestabstand (von $sp = 6 mm$) herrscht. Als Distinctiveness-Werte wird in diesem Falle ein Gradientenbetragsbild I_k^{Grad} gewählt (siehe Abbildung 3). Das bedeutet, dass zuerst Merkmalsvektoren auf den Knochenstrukturen und an der Handkontur gewählt werden (siehe Abschnitt 8.1 und 8.2). Es ist zu beachten, dass diese Wahl der Positionen zwar abhängig von dem betrachteten Bildinhalt erfolgt, aber in keiner Weise gewährleistet, dass Merkmalsvektoren an identischen Positionen im Objekt in allen Trainingsdaten vorzufinden sind. Abhängig von der Handgröße und dem Bildausschnitt werden auf diese Weise unterschiedlich viele Merkmalsvektoren erzeugt mit $N_s(k) = 450.15 \pm 57.7$.

Parameterbelegung

Input-/Trainingsdaten

N	=	60
D	=	4 ($D_f = 2, D_x = 2$)
f_i	:	(1) Gauß-geglättete Signalwerte: $(G_{\sigma=6mm} * I_k)(x_i)$ (2) Distanzwerte zur Handmaske: $d(x_i, I_k^{Mask})$
x_i Var1	:	Randomisiertes Raster mit Punktabstand $sp = 6mm$
$N_s^{Var1}(k)$	:	884 ± 83
x_i Var2	:	kantenmerkmalsabhängig mit Mindestabstand $sp = 6 mm$, innerhalb der Hand ($\forall i f_{i,2} < 5mm$)
$N_s^{Var2}(k)$	=	450 ± 57

^a<http://www.ipilab.org/BAASWeb>

2.2. Korrespondenzwahrscheinlichkeiten zwischen Bildrepräsentationen

Es seien zwei Bildrepräsentationen $\mathbf{S} = \{\mathbf{s}_i | i = 1, \dots, N_s\}$ und $\mathbf{M} = \{\mathbf{m}_j | j = 1, \dots, N_m\}$ mit potentiell unterschiedlicher Anzahl an Merkmalsvektoren gegeben. Es sei unbekannt, welche Merkmalsvektoren beider Mengen miteinander korrespondieren. Des Weiteren wird angenommen, dass die Positionierung der Merkmalsvektoren “ungenau” ist und daher keine Eins-zu-Eins-Korrespondenz zwischen den Mengen vorliegt. Daher sind die

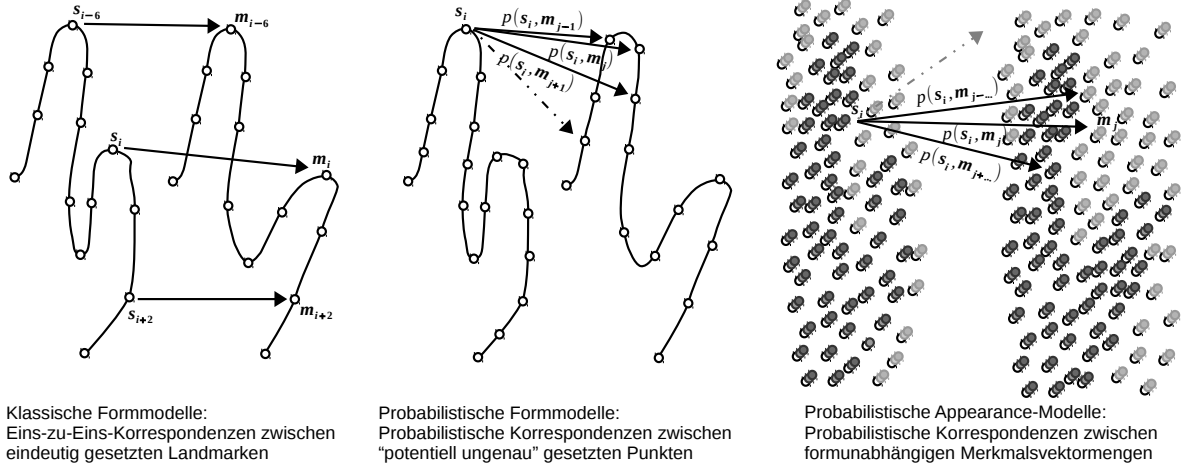


Abbildung 4: Gegenüberstellung der Korrespondenzannahmen: in klassischen Modellansätzen wird von einer Eins-zu-Eins-Korrespondenz zwischen möglichst exakt lokalisierten Landmarken ausgegangen. Probabilistische Formmodelle gehen nicht von einer akkuraten Positionierung der Merkmalsvektoren aus, sondern berechnen stattdessen die Korrespondenzwahrscheinlichkeiten zwischen "potentiell ungenau" bestimmten Merkmalsvektoren. Bei den probabilistischen Appearance-Modellen wird die Korrespondenzschätzung um die Appearance-Merkmale erweitert und des Weiteren wird von einer formunabhängigen Positionierung der Merkmalsvektoren ausgegangen.

Korrespondenzwahrscheinlichkeiten zwischen beiden Repräsentationen über die paarweisen Wahrscheinlichkeiten so definiert, dass Merkmalsvektor $\mathbf{s}_i$ aus $\mathbf{S}$ eine verrauschte Observation des Merkmalsvektors $\mathbf{m}_j$ aus $\mathbf{M}$ ist. Dies kann nach Granger et al. [Granger & Pennec 2002] als Gauß-Verteilung modelliert werden:

$$p(\mathbf{s}_i|\mathbf{m}_j) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} (\mathbf{s}_i - \mathbf{m}_j)^T \Sigma^{-1} (\mathbf{s}_i - \mathbf{m}_j) \right), \quad (5)$$

wobei $\Sigma \in \mathbb{R}^{D \times D}$ die Kovarianzmatrix repräsentiert und sowohl Positions- als auch Appearance-Merkmalvarianzen beschreibt. Da jeder Merkmalsvektor aus $\mathbf{S}$ zu jedem anderen Vektor aus $\mathbf{M}$ korrespondieren kann, setzt sich die Wahrscheinlichkeit für $\mathbf{s}_i$ zusammen als:

$$p(\mathbf{s}_i|\mathbf{M}) = \frac{1}{N_m} \sum_{j=1}^{N_m} p(\mathbf{s}_i|\mathbf{m}_j). \quad (6)$$

Siehe [Granger & Pennec 2002] für die genaue Herleitung. Des Weiteren werden die Merkmalsvektoren einer Repräsentation als unabhängig voneinander betrachtet und die

Korrespondenzwahrscheinlichkeit von $\mathbf{S}$ zu $\mathbf{M}$ kann zusammengefasst werden als:

$$p(\mathbf{S}|\mathbf{M}) = \prod_{i=1}^{N_s} \frac{1}{N_m} \sum_{j=1}^{N_m} p(\mathbf{s}_i|\mathbf{m}_j). \quad (7)$$

$p(\mathbf{S}|\mathbf{M})$ kann demnach als Maß betrachtet werden, welches anzeigt, wie stark beide Bildrepräsentationen miteinander korrespondieren. Dies wird im Folgenden zur Optimierung gesuchter Parameter eingesetzt.

Abbildung 4 stellt die klassische Eins-zu-Eins-Korrespondenz der probabilistischen Korrespondenzdefinition visuell gegenüber.

2.3. Expectation-Maximization-Ansatz zur Optimierung einer Transformation mittels Korrespondenzoptimierung

Gegeben seien erneut zwei Bildrepräsentationen $\mathbf{S} = \{\mathbf{s}_i | i = 1, \dots, N_s\}$ und $\mathbf{M} = \{\mathbf{m}_j | j = 1, \dots, N_m\}$ mit unbekannten Korrespondenzen und eine beliebige Transformation ϕ , welche $\mathbf{M}$ manipuliert. Die Annahme sei, dass ϕ sowohl die Positionen (z. B. mittels affiner Transformation) als auch die Appearance-Werte der Vektoren in $\mathbf{M}$ anpassen kann. $\mathbf{m}_j \circ \phi = (\mathbf{x}_j, \mathbf{f}_j) \circ \phi = (\mathbf{x}_j \circ \phi^{\mathbf{x}}, \mathbf{f}_j \circ \phi^{\mathbf{f}})$ sei dabei als Transformation von $\mathbf{m}_j$ definiert. Ziel sei es, die optimalen Parameter für ϕ zu finden, sodass $\mathbf{M}$ unter Transformation ϕ optimal zu $\mathbf{S}$ "passt":

$$p(\mathbf{S}|\mathbf{M}, \phi) \rightarrow_{\phi} \max = \prod_{i=1}^{N_s} \frac{1}{N_m} \sum_{j=1}^{N_m} p(\mathbf{s}_i|\mathbf{m}_j \circ \phi) \rightarrow_{\phi} \max. \quad (8)$$

Die Berechnung von $p(\mathbf{s}_i|\mathbf{m}_j \circ \phi)$ findet analog zu Gleichung 5 statt. Granger et al. haben zur Optimierung einer affinen Transformation einen Expectation-Maximization-Ansatz hinzugezogen [Granger & Pennec 2002], da es keine geschlossene Lösung für die Maximierung von $p(\mathbf{S}|\mathbf{M}, \phi)$ gibt. Der Ansatz lässt sich auf jede Art der Deformation anwenden. Die Grundidee besteht darin, den Erwartungswert von $\mathbf{S}$ bei gegebenem $\mathbf{M}$ und ϕ zu maximieren:

$$\mathbb{E}(\log p(\mathbf{S}|\mathbf{M}, \phi)) = \frac{1}{N_m} \sum_{i=1}^{N_s} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij}) \log p(\mathbf{s}_i|\mathbf{m}_j \circ \phi). \quad (9)$$

Die Matrix $\mathbf{H} \in \mathbb{R}^{N_s \times N_m}$ mit $H_{ij} \in \{0, 1\}$ beschreibt dabei die unbekannten Korrespondenzen zwischen $\mathbf{S}$ und $\mathbf{M}$ ($H_{ij} = 1$, wenn $\mathbf{s}_i$ zu $\mathbf{m}_j$ korrespondiert und $H_{ij} = 0$, wenn nicht). $\mathbb{E}(H_{ij})$ repräsentiert dementsprechend den Erwartungswert, dass $\mathbf{s}_i$ zu $\mathbf{m}_j$

korrespondiert mit gegebenem ϕ :

$$\mathbb{E}(H_{ij}) = \frac{p(\mathbf{s}_i | \mathbf{m}_j \circ \phi)}{\sum_{j=1}^{N_m} p(\mathbf{s}_i | \mathbf{m}_j \circ \phi)}. \quad (10)$$

Um die Likelihood in Gleichung 8 iterativ zu optimieren, wird ein Expectation-Maximization-Algorithmus eingesetzt. Dabei wird zwischen einem E(xpectation)-Schritt und einem M(aximation)-Schritt bis zur Konvergenz alterniert. Im E-Schritt sind die Parameter ϕ als fix angenommen und es werden die Korrespondenzwahrscheinlichkeiten $\mathbb{E}(H_{ij})$ neu berechnet. Im M-Schritt wird die Likelihood mit Hinblick auf die Parameter ϕ optimiert – unter Annahme fixer Korrespondenzwahrscheinlichkeiten. In Abschnitt 4.1.1 wird ein solcher EM-Ansatz genutzt, um die Parameter des probabilistischen Modells bei unbekannten Korrespondenzen zwischen den Trainingsdaten zu optimieren.

2.3.1. Normierung der Erwartungswerte

Da in Gleichung 10 die Erwartungswerte über die Summe aller j in $\mathbf{M}$ normiert werden, ergibt sich $\sum_{j=1}^{N_m} \mathbb{E}(H_{ij}) = 1, \forall i$. Folglich kann interpretiert werden, dass jeder Vektor $\mathbf{s}_i$ in $\mathbf{S}$ korrespondierende Merkmalsvektoren in $\mathbf{M}$ hat. Dem liegt die Annahme zugrunde, dass $\mathbf{S}$ keine Outlier beinhaltet und daher jeder Vektor $\mathbf{s}_i$ mit gleichem Erwartungswert durch die gesamte (transformierte) Menge $\mathbf{M}$ abgebildet werden kann. $\mathbf{M}$ darf hingegen Outlier beinhalten, die “gar nicht” zu $\mathbf{S}$ korrespondieren. Soll $\mathbf{M}$ im Gegensatz zu $\mathbf{S}$ keine Outlier aufweisen, wird hingegen mithilfe der Summe über die Merkmalsvektoren in $\mathbf{S}$ normiert:

$$\mathbb{E}(H_{ij}) = \frac{p(\mathbf{s}_i | \mathbf{m}_j \circ \phi)}{\sum_{i=1}^{N_s} p(\mathbf{s}_i | \mathbf{m}_j \circ \phi)}, \quad (11)$$

mit dem Ergebnis $\sum_{i=1}^{N_s} \mathbb{E}(H_{ij}) = 1, \forall j$.

Bei idealen Eins-zu-Eins-Korrespondenzen zwischen beiden Mengen wären beide Summen – $\sum_{i=1}^{N_s} \mathbb{E}(H_{ij})$ und $\sum_{j=1}^{N_m} \mathbb{E}(H_{ij})$ – unabhängig von der Normierungsrichtung gleich Eins.

Outlier: Eine einseitige Normierung der Erwartungswerte wie in Gleichung 10 (bzw. 11) führt zwar dazu, dass $\mathbf{S}$ (Gl. 10) “vollständig” durch $\mathbf{M}$ abgebildet werden kann, allerdings werden Outlier in $\mathbf{M}$ auch komplett ignoriert und in keiner Weise in die weitere Berechnung/Optimierung von ϕ mit einbezogen. Soll dies abgefangen werden, können extreme Outlier (mit $\sum_{i=1}^{N_s} \mathbb{E}(H_{ij}) \ll 1$) in die “zweite Richtung” zur Korrespondenz gezwungen werden. Nach der Normierung durch Gleichung 10 gilt $\forall i : \sum_{j=1}^{N_m} \mathbb{E}(H_{ij}) = 1$. Anschließend wird die andere Richtung ($\forall j$) nach Outliern überprüft und die Erwar-

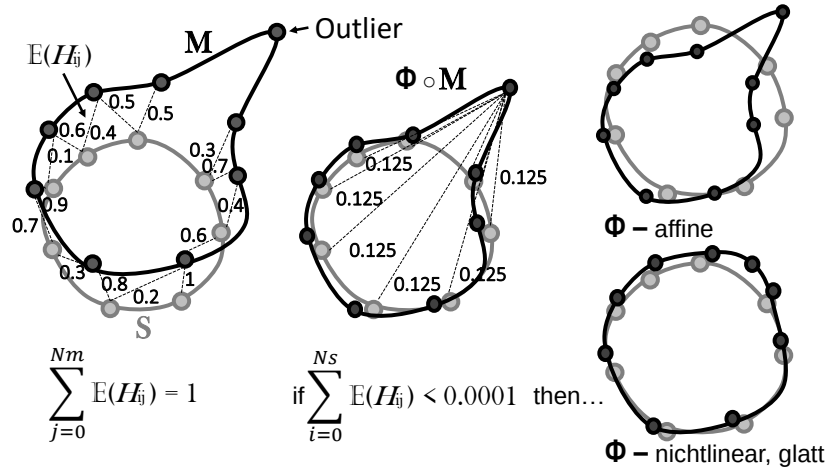


Abbildung 5: Normierung der Erwartungswerte $\mathbb{E}(H_{ij})$: links: Die Korrespondenzwahrscheinlichkeiten werden zwischen zwei Merkmalsvektormengen ($\mathbf{S}$ und $\mathbf{M}$) mit Gleichung 10 normiert. Ist zwischen zwei Punkten keine gestrichelte Linie eingezeichnet, ist der Erwartungswert gleich 0. Der Outlier in $\mathbf{M}$ (rechter oberer Punkt) hat zu keinem Vektor in $\mathbf{S}$ eine Korrespondenz ($\sum_{i=1}^{N_s} \mathbb{E}(H_{ij}) = 0$) und fließt daher nicht in die Berechnung der Transformation ϕ ein (abhängig von der gewählten Kovarianzmatrix Σ in Gleichung 5) – siehe mittlere Abbildung. Wird hingegen Gleichung 12 zusätzlich angewendet, wird der Outlier zur Korrespondenz “gezwungen”. Ist ϕ eine affine Transformation (rechts oben), kann die Einbeziehung der Outlier zu schlechteren Ergebnissen führen. Ist ϕ jedoch eine nichtlineare (hier nicht näher definierte) Deformation (rechts unten), kann dies dazu führen, dass $\mathbf{M}$ genauer an $\mathbf{S}$ angepasst wird. Das bedeutet, dass die Wahl der Normierung und die zusätzliche Outlier-Behandlung vom betrachteten Problem abhängig sind.

tungswerte gegebenenfalls geändert:

$$\forall j : \text{ if } \left(\sum_{i=1}^{N_s} \mathbb{E}(H_{ij}) < 0.00001 \right) \text{ then } \forall i : \mathbb{E}(H_{ij}) = \frac{1}{N_m N_s} \sum_{i=1}^{N_s} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij}). \quad (12)$$

Analog dazu für Gleichung 11 ($\forall j : \sum_{i=1}^{N_s} \mathbb{E}(H_{ij}) = 1$) :

$$\forall i : \text{ if } \left(\sum_{j=1}^{N_m} \mathbb{E}(H_{ij}) < 0.00001 \right) \text{ then } \forall j : \mathbb{E}(H_{ij}) = \frac{1}{N_m N_s} \sum_{i=1}^{N_s} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij}). \quad (13)$$

Demnach können die Gleichungen 10 und 12 (und analog Gleichungen 11 und 13) gemeinsam interpretiert werden als: $\mathbf{S}$ beinhaltet keine Outlier und jeder Vektor sollte gleichermaßen von $\mathbf{M} \circ \phi$ abgebildet werden können. Darüber hinaus sollte die Transformation keine “extremen” Outlier in $\mathbf{M}$ erzeugen – sollte ein Vektor in $\mathbf{M}$ “gar nicht” korrespondieren, so wird eine Korrespondenz zur gesamten Menge $\mathbf{S}$ “erzwungen” mit $\mathbb{E}(H_{ij}) = \frac{1}{N_m N_s} \sum_{i=1}^{N_s} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})$. Um die erste Bedingung (Gleichung 10 oder 11), dass

die Summe über den entsprechenden Index Eins sein soll, nicht zu verletzen, kann die jeweilige erste Normierung ein weiteres Mal angewendet werden: $\mathbb{E}(H_{ij}) = \frac{\mathbb{E}(H_{ij})}{\sum_{j=1}^{N_m} \mathbb{E}(H_{ij})}$ bzw. $\mathbb{E}(H_{ij}) = \frac{\mathbb{E}(H_{ij})}{\sum_{i=1}^{N_s} \mathbb{E}(H_{ij})}$. Abbildung 5 visualisiert den Effekt der Normierung mit Outliern in $\mathbf{M}$.

Beispielbox 2: Korrespondenzwahrscheinlichkeiten

Zur Veranschaulichung der probabilistischen Korrespondenzen wurde im Beispiel aus Abbildung 6 ein Modell $\mathbf{M}$ mit $N_m = 481$ Merkmalsvektoren an eine Observation $\mathbf{S}$ mit $N_s = 380$ Merkmalsvektoren mittels einer gegebenen Transformation ϕ angepasst. Dabei ändert ϕ sowohl die Positionen mittels affiner Transformation von $\mathbf{M}$ ($\phi^{\mathbf{x}}$) als auch die Appearance-Merkmale ($\phi^{\mathbf{f}}$). Wie die Parameter genau definiert sind und bestimmt werden, sei in diesem Beispiel nicht von Interesse.

In der ersten Zeile der Abbildung sind $\mathbf{S}$ und die (angepassten) Varianten von $\mathbf{M}$ durch das erste Appearance-Merkmal dargestellt. Die zweite Zeile visualisiert die Korrespondenzen für jeden Merkmalsvektor in $\mathbf{S}$ zum (angepassten) $\mathbf{M}$ ($N_m p(\mathbf{s}_i | \mathbf{M}, \phi^{\cdots})$) und in der dritten Zeile werden die inversen Korrespondenzen berechnet ($N_s p(\mathbf{m}_j | \mathbf{S}, \phi^{\cdots})$). Dafür werden Merkmalsvarianzen mit $\Sigma = \text{diag}(6, 6, 30, 20)$ genutzt. Dabei wird deutlich, dass die Korrespondenzwahrscheinlichkeiten sowohl mit der Positionsanpassung als auch mit der weiteren Appearance-Anpassung größer werden. Unter den Bildern sind die jeweiligen mittleren Wahrscheinlichkeiten für $N_m p(\mathbf{s}_i | \mathbf{M}, \phi^{\cdots})$ bzw. $N_s p(\mathbf{m}_j | \mathbf{S}, \phi^{\cdots})$ zu finden. Um zu verdeutlichen, welchen Anteil dabei die Positions- bzw. Appearance-Merkmale haben, wurden die Korrespondenzwahrscheinlichkeiten ebenfalls mit $\Sigma = \text{diag}(6, 6, \infty, \infty)$ bzw. $\Sigma = \text{diag}(\infty, \infty, 30, 20)$ berechnet. Es fällt auf, dass die probabilistischen Korrespondenzen für $\mathbf{S}$ ($N_m p(\mathbf{s}_i | \mathbf{M}, \phi^{\cdots})$) größer sind als für $\mathbf{M}$ ($N_s p(\mathbf{m}_j | \mathbf{S}, \phi^{\cdots})$). Der Grund ist die geringere Vektoranzahl in $\mathbf{S}$ als in $\mathbf{M}$ ($N_s = 380$, $N_m = 481$), wodurch ein $\mathbf{s}_i$ im Durchschnitt mit höherer Wahrscheinlichkeit zur $\mathbf{m}_j$ korrespondiert als in die Gegenrichtung. Sind die Werte für Σ ideal gewählt, ergibt sich die probabilistische Durchschnittswahrscheinlichkeit nach “erfolgreicher” Transformation zu ≈ 1 . Sind die Werte höher als 1, sind die Merkmalsvarianzen in Σ für die Daten zu hoch gewählt (z. B. ∞). Dies führt dazu, dass ein Merkmalsvektor einer Observation zu mehreren Vektoren der anderen Observation korrespondiert.

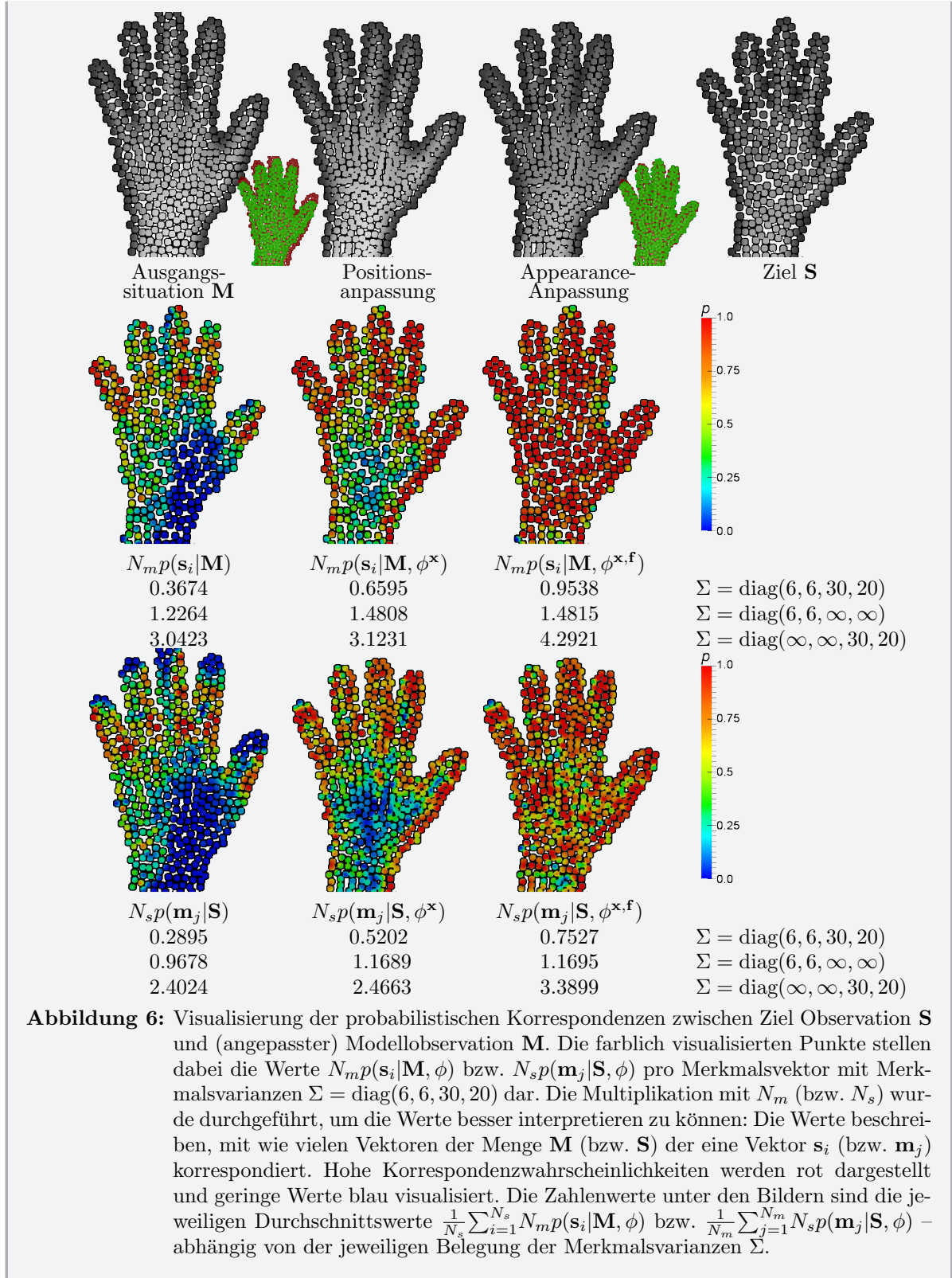
Parameterbelegung

Input-/Trainingsdaten

D = 4 ($D_f = 2$, $D_x = 2$)
 $\mathbf{f}_i$: (1) Gauß-geglättete Signalwerte: $(G_{\sigma=6mm} * I_k)(\mathbf{x}_i)$ (2) Distanzwerte zur Handmaske: $d(\mathbf{x}_i, I_k^{Mask})$
 $\mathbf{x}_i$: kantenmerkmalsabhängig mit Mindestabstand 6 mm, innerhalb der Hand ($\forall i, f_{i,2} < 5mm$)
 N_s = 380
 N_m = 481

Modellparameter

Σ = $\text{diag}(6, 6, 30, 20)$
 $\mathbf{m}_j \circ \phi$ = $(\mathbf{x}_j, \mathbf{f}_j) \circ \phi = (\mathbf{x}_j \circ \phi^{\mathbf{x}}, \mathbf{f}_j \circ \phi^{\mathbf{f}})$



3. Probabilistische Appearance-Modelle

3.1. Modell auf Basis von zufälligen Observationen

Liegt eine Trainingsmenge von N Bildern samt dazugehörigen Bildrepräsentationen $\mathbf{S}_{\text{tr}} = \{\mathbf{S}_k | k = 1, \dots, N\}$ vor, wird das Modell, welches diese Daten optimal abbildet, gesucht. Die probabilistische Sichtweise betrachtet die gegebenen Daten als Menge von Observationen, auf dessen Basis gewisse Populationsparameter – in diesem Fall “das Modell” – geschätzt werden sollen. Das heißt, folgende Wahrscheinlichkeit wird maximiert:

$$\text{Modell} = \operatorname{argmax}_{\text{Modell}} p(\text{Modell} | \mathbf{S}_{\text{tr}}). \quad (14)$$

Die Trainingsobservationen werden als zufällig durch das Modell generiert angenommen. Wie bei klassischen Ansätzen setzt sich das Modell aus einem mittleren Modell $\bar{\mathbf{M}}$ und einer Anzahl an n Variationsmoden $\mathbf{V} = \{\mathbf{V}_p | p = 1, \dots, n\}$ mit Standardabweichung $\{\lambda_p | p = 1, \dots, n\}$ zusammen. Laut Modellannahme werden neue Instanzen $\mathbf{M}$ durch ein solches Modell generiert, indem die Moden mittels Gewichte ω_p manipuliert werden. Größere lineare Anpassungen werden durch eine affine Transformation T durchgeführt:

$$\mathbf{M} = T^{-1}(\bar{\mathbf{M}} + \sum_{p=1}^n \omega_p \mathbf{V}_p) \text{ mit } \omega_p \sim N(0, \lambda_p). \quad (15)$$

ω_p wird als normalverteilt mit der Standardabweichung λ_p angenommen.

Um die optimalen Parameter zu finden, werden nun diejenigen Modellinstanzen $\mathbf{M}_k$ gesucht, welche die gegebene Trainingsdaten $\mathbf{S}_k$ am besten beschreiben $\forall k$:

$$\mathbf{S}_k \approx \mathbf{M}_k \text{ mit } \mathbf{M}_k = T_k^{-1}(\bar{\mathbf{M}} + \sum_{p=1}^n \omega_{k,p} \mathbf{V}_p) \quad (16)$$

Um zu bestimmen, ob zwei Merkmalsvektormengen ($\mathbf{S}_k$ und $\mathbf{M}_k$) einander ähnlich sind, müssen Korrespondenzen zwischen beiden bekannt sein. Diese sind zu Beginn der Modellgenerierung nicht bekannt und müssen gemeinsam mit den Parametern optimiert werden (siehe Abschnitt 2.3 und 4.1.1). Die probabilistischen Korrespondenzen werden über die Gleichung 5 berechnet. Dabei bestimmen die Merkmalsvarianzen Σ , welche Vektoren in beiden Mengen wie stark korrespondieren. Die Kovarianzmatrix Σ wird daher zu den Parametern hinzugefügt. Die Abbildungen 7 und 8 stellen den Unterschied zwischen der klassischen Modellgenerierung in einem Schritt mit bekannten Korrespondenzen (Abbildung 7) und der iterativen probabilistischen Modellgenerierung über die gemeinsame Optimierung der Modellinstanzen und Korrespondenzen (Abbildung 8) optisch dar.

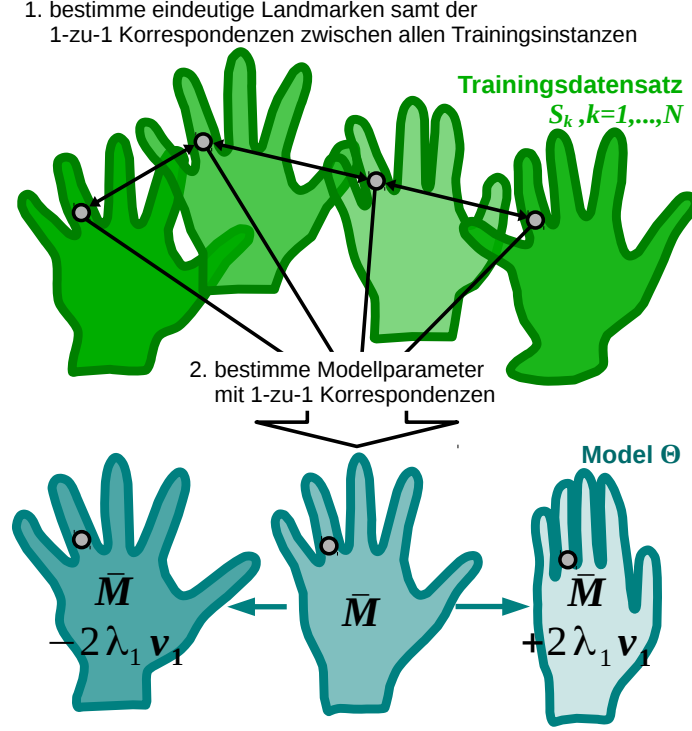


Abbildung 7: Modellgenerierung mit bekannten Eins-zu-Eins-Korrespondenzen: In klassischen Modellen sind die eindeutigen Korrespondenzen zwischen den Trainingsinstanzen und somit auch zum resultierenden Modell bekannt. Es können also in einem Schritt das mittlere Modell $\bar{M}$ und die Variationsmoden V (mittels PCA) bestimmt werden.

3.1.1. Parameter

Alle Parameter $\{\bar{M}, V_p, \lambda_p, \omega_{k,p}, T_k, n, \Sigma \mid p = 1, \dots, n, k = 1, \dots, N\}$ können in *Modellparameter* und *instanzabhängige Parameter* unterteilt werden. Als Modellparameter Θ werden diejenigen Parameter betrachtet, die für den gesamten Trainingsdatensatz und somit optimalerweise für die gesamte Population gleich sind:

$$\Theta = \{\bar{M}, V_p, \lambda_p, n, \Sigma \mid p = 1, \dots, n\}.$$

Als instanzabhängige Parameter $Q = \{Q_k \mid k = 1, \dots, N\}$ werden die Parameter definiert, welche die jeweilige Modellinstanz modifizieren:

$$Q_k = \{T_k, \omega_{k,p} \mid p = 1, \dots, n\}.$$

Im Folgenden werden alle Parameter genauer beschrieben.

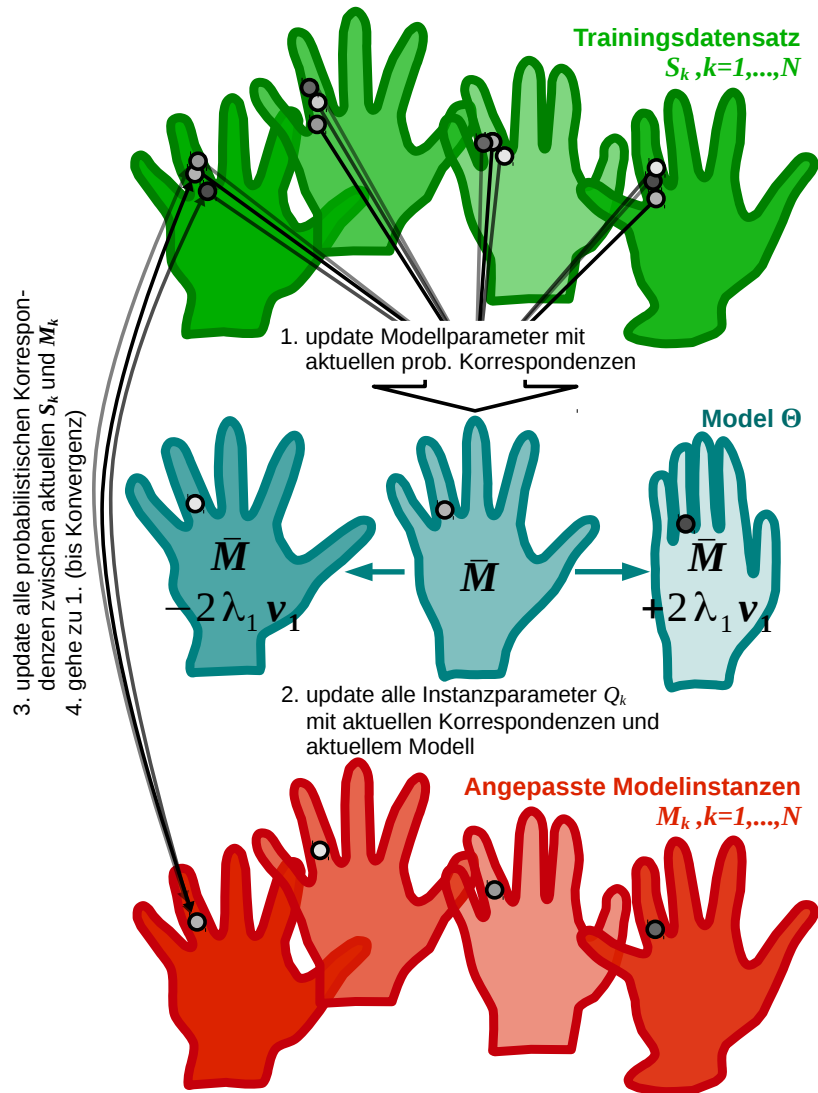


Abbildung 8: Modellgenerierung ohne bekannte Eins-zu-Eins-Korrespondenzen: Dem vorgestellten Modellierungsansatz sind keine Korrespondenzen bekannt, stattdessen werden Korrespondenzwahrscheinlichkeiten genutzt. Das heißt, das Modell kann nicht in einem Schritt berechnet werden, da die optimalen Modellparameter abhängig sind von den aktuellen probabilistischen Korrespondenzen zwischen Modell und Trainingsdatensatz. Um diese bestimmen zu können, werden nicht nur neue Modellparameter berechnet (1.) wie in klassischen Modellen, sondern es wird ebenfalls das aktuelle Modell auf die Trainingsinstanzen angepasst (2.). Anschließend können die Korrespondenzwahrscheinlichkeiten zwischen S_k und Modell (durch die Berechnung der Korrespondenzwahrscheinlichkeiten zwischen S_k und M_k) aktualisiert werden (3.) und die Optimierung der Modellparameter wird iterativ fortgesetzt bis zur Konvergenz.

Eingabe: $\mathbf{S}_k, k = 1, \dots, N$

- Die **Trainingsdaten** $\mathbf{S}_k = \{\mathbf{s}_{k,i} | i = 1, \dots, N_s(k)\}$, $\mathbf{s}_{k,i} = (\mathbf{x}_{k,i}, \mathbf{f}_{k,i}) \in \mathbb{R}^D$ liegen in Form der eingeführten Bildrepräsentation vor (Abschnitt 2). Die Anzahl $N_s(k)$ der Merkmalsvektoren kann zwischen den Trainingsdaten variieren.

Modellparameter: $\Theta = \{\bar{\mathbf{M}}, \mathbf{V}_p, \lambda_p, n, \Sigma | p = 1, \dots, n\}$:

- Das **mittlere Modell** $\bar{\mathbf{M}} = \{\bar{\mathbf{m}}_j | j = 1, \dots, N_m\}$, $\bar{\mathbf{m}}_j = (\mathbf{x}_j, \mathbf{f}_j) \in \mathbb{R}^D$ liegt in Form der eingeführten Bildrepräsentation vor und umfasst N_m Merkmalsvektoren (unabhängig von der Anzahl der Vektoren der Trainingsinstanzen $N_s(k)$).
- Die **Variationsmoden** $\mathbf{V}_p = \{\mathbf{v}_{p,j} | j = 1, \dots, N_m\}$, $p = 1, \dots, n$, mit $\mathbf{v}_{p,j} = (\mathbf{x}_j^\Delta, \mathbf{f}_j^\Delta) \in \mathbb{R}^D$ weisen ebenfalls die vektorielle Form auf und stellen die Offsets zwischen mittlerem Modell und Modellinstanz dar. Somit manipuliert eine Mode gleichzeitig Positions- und Appearance-Werte aller Merkmalsvektoren. Die Anzahl N_m und die Indizes j der Merkmalsvektoren $\mathbf{v}_{p,j}$ pro Mode p korrespondiert mit der Anzahl und den Indizes der Vektoren des mittleren Modells, sodass gilt: $\bar{\mathbf{M}} + \mathbf{V}_p := \{\bar{\mathbf{m}}_j + \mathbf{v}_{p,j} | j = 1, \dots, N_m\}$.
- Die **Anzahl der Variationsmoden** n bestimmt wie bei klassischen Modellen den Umfang der möglichen Variationen.
- Die **Standardabweichungen der Variationsmoden** $\lambda_p, p = 1, \dots, n$ beschreiben analog zu den Eigenwerten bei der klassischen Modellgenerierung den Einfluss der jeweiligen Mode p .
- Die **Kovarianzmatrix** $\Sigma \in \mathbb{R}^{D \times D}$ der Distanzen zwischen $\mathbf{s}_i$ und $\mathbf{m}_j$ regularisiert die Korrespondenzwahrscheinlichkeiten zwischen $\mathbf{s}_i$ und $\mathbf{m}_j$ (siehe Gleichung 5).

Trainingsinstanzabhängige Parameter: $Q_k = \{T_k, \omega_k\}$ für Instanz $\mathbf{S}_k, k = 1, \dots, N$

- Die **Gewichtung der Variationsmoden** $\omega_k = \{\omega_{k,p} | p = 1, \dots, n\}$ manipuliert sowohl die Positions- als auch die Appearance-Werte des Modells.
- Die **affine Transformation** $T_k : \mathbb{R}^{D_x} \rightarrow \mathbb{R}^{D_x}$ zwischen $\mathbf{S}_k$ und Modell sorgt während der Modellgenerierung für eine Ausrichtung der Trainingsdaten, sodass die Variationsmoden lediglich transformationsunabhängige Variationen beschreiben.

Es gibt zwei Möglichkeiten eine Transformation der Bildrepräsentationen $T_k(\mathbf{S}_k) : T_k \star \mathbf{s}_{k,i}, i = 1, \dots, N_s(k)$ durchzuführen. Dafür wird im Folgenden die $\star$ -Operation auf zwei unterschiedliche Weisen definiert:

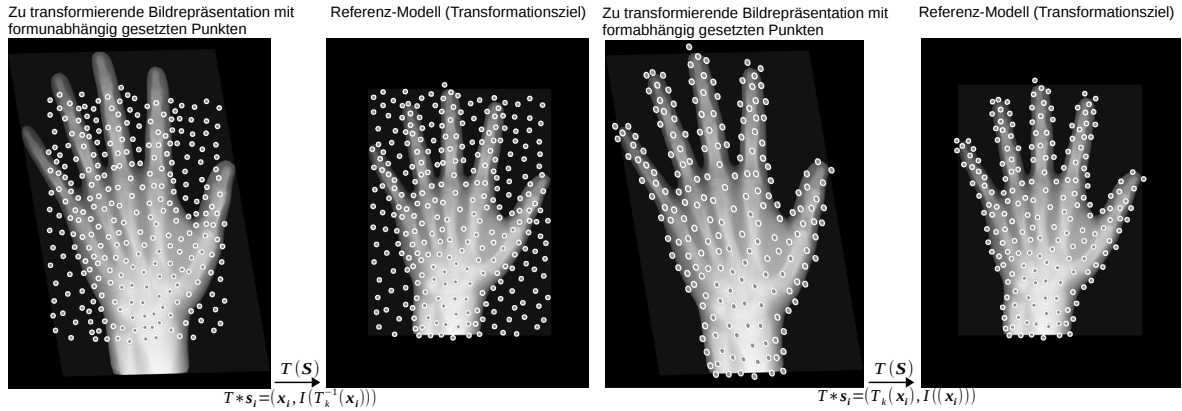


Abbildung 9: Gegenüberstellung beider Transformationsvarianten: Entweder wird das Bild “unter den Positionen der Merkmalsvektoren” transformiert (links), wobei die Positionsmerkmale der Merkmalsvektoren fix bleiben, oder die Positionen werden mit fixen Appearance-Merkmalen transformiert (rechts).

1. **Fixe Appearance-Werte (fixApp):** Die Merkmalsvektoren werden mit fixen Appearance-Merkmalen transformiert:

$$T_k \star s_{k,i} := (T_k(\mathbf{x}_{k,i}), \mathbf{f}_{k,i}). \quad (17)$$

Die Appearance-Merkmale bleiben gleich. Lediglich die Positionsmerkmale der Vektoren ändern sich. Dies ist sinnvoll, wenn die Positionen abhängig von Bildstrukturen gewählt worden sind (Abbildung 9, rechts).

2. **Fixe Positionswerte (fixPos):** Das Bild I_k (bzw. alle Appearance-Bilder $\mathbf{F} = [I_k^{f_1}, \dots, I_k^{f_{D_f}}]$) wird unter den Positionen der Merkmalsvektoren transformiert:

$$T_k \star s_{k,i} := (\mathbf{x}_{k,i}, \mathbf{F}_k(T_k^{-1}(\mathbf{x}_{k,i}))). \quad (18)$$

Die Position bleibt gleich und Appearance-Merkmale ändern sich (Abbildung 9, links).

Modellinstanzen: Sind die Parameter Θ und Q_k gegeben, so ist die entsprechende Modellinstanz definiert als:

$$\mathbf{M}_k = T_k^{-1}(\bar{\mathbf{M}} + \sum_{p=1}^n \omega_p \mathbf{V}_p) = \{\mathbf{m}_{k,j} | j = 1, \dots, N_m\} \quad (19)$$

mit

$$\mathbf{m}_{k,j} = T_k^{-1} \star (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}). \quad (20)$$

Da für die inverse Transformation von $T_k \star \mathbf{s}_{k,i} = (\mathbf{x}_{k,i}, \mathbf{F}_k(T_k^{-1}(\mathbf{x}_{k,i})))$ für die fixPos-Variante ebenfalls ein komplettes Bild der Modellinstanz nötig wäre ($T_k^{-1} \star \bar{\mathbf{m}}_j = (\mathbf{x}_j, \mathbf{F}_k^{\text{Modellbild}}(T_k(\mathbf{x}_j)))$), das Modell (die Modellinstanz) aber lediglich als spärliche Merkmalsvektormenge vorliegt, wird effektiv die Transformation T_k nicht auf das Modell angewendet, sondern auf die Trainings- bzw. Testinstanz $\mathbf{S}_k$, denn

$$T_k^{-1}(\bar{\mathbf{M}} + \sum_{p=1}^n \omega_{k,p} \mathbf{V}_p) \approx \mathbf{S}_k$$

entspricht

$$\bar{\mathbf{M}} + \sum_{p=1}^n \omega_{k,p} \mathbf{V}_p \approx T_k(\mathbf{S}_k).$$

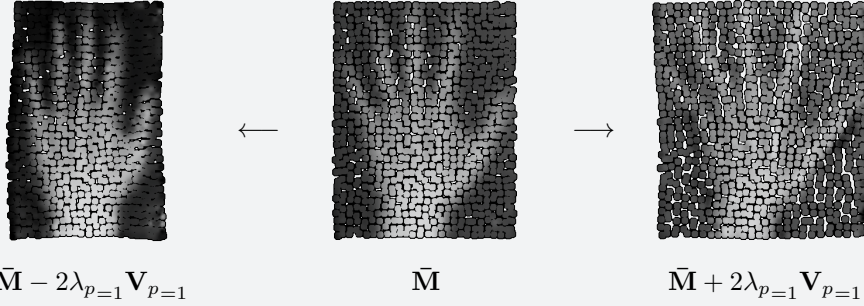
Für die fixApp-Transformationsvariante ($T_k \star \mathbf{s}_{k,i} = (T_k(\mathbf{x}_i), \mathbf{f}_i)$) ist die inverse Transformation definiert mit: $T_k^{-1} \star \bar{\mathbf{m}}_j = (T_k^{-1}(\mathbf{x}_j), \mathbf{f}_j)$.

Im Folgenden wird dennoch die Schreibweise $\mathbf{M}_k = T_k^{-1}(\bar{\mathbf{M}} + \boldsymbol{\omega}_k \mathbf{V})$ unabhängig von der Art der Transformation beibehalten.

Beispielbox 3: Modell

Es seien an dieser Stelle zwei Beispiele für Modelle mit den beschriebenen Parametern gegeben. Dabei wird das mittlere Modell und die gewichtete ($\pm$ zweifacher Standardabweichung) erste Mode für die Positionsänderungen und das erste Appearance-Merkmal visualisiert. Im ersten Beispiel werden die Positionen der Merkmalsvektoren komplett objektunabhängig gewählt. Im zweiten Beispiel werden die Merkmalsvektoren kantenmerkmalsabhängig gewählt und liegen nur innerhalb der Hand (siehe Beispielbox 1). Da die Merkmalsvektoren des Modells im ersten Beispiel keine “festen Rollen” besitzen, wird hier die fixPos-Transformationsvariante genutzt – die Positionen bleiben fix. Im zweiten Beispiel hingegen werden die Positionen durch die fixApp-Transformation geändert, da diese objektabhängig gewählt wurden.

Modell:



Modellanpassung:

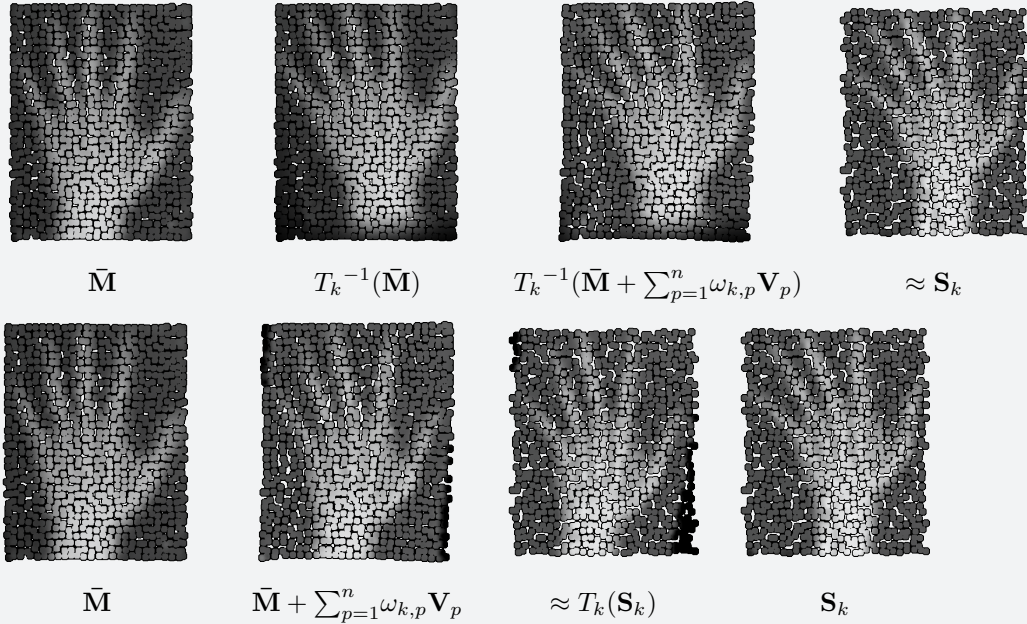


Abbildung 10: Visualisierung eines Modells samt Modellanpassung für Merkmalsvektoren mit Positionen auf “randomisiertem Raster”.

Parameterbelegung

Input-/Trainingsdaten

N	=	60
D	=	4 ($D_f = 2$, $D_x = 2$)
$\mathbf{f}_i$	:	(1) Gauß-geglättete Signalwerte: $(G_{\sigma=6mm} * I_k)(\mathbf{x}_i)$ (2) Distanzwerte zur Handmaske: $d(\mathbf{x}_i, I_k^{Mask})$
$\mathbf{x}_i$	:	Randomisiertes Raster mit Punktabstand 6mm
$N_s(k)$	=	884.55 ± 83.91
N_m	=	933

Modellparameter

n	=	15
$T_k \star \mathbf{s}_{k,i}$	=	$(\mathbf{x}_i, \mathbf{F}_k(T_k^{-1}(\mathbf{x}_i)))$

Der Vorteil des ersten Modells liegt in der Einfachheit der Initialisierung der Merkmalsvektorpositionen. Effektiv wird die Transformation nicht auf das Modell angewendet (wie in der ersten Zeile der Bilder für die Modellanpassung) sondern auf die Trainings- bzw. Testinstanz $\mathbf{S}_k$ (zweite Zeile), denn $T_k^{-1}(\bar{\mathbf{M}} + \sum_{p=1}^n \omega_{k,p} \mathbf{V}_p) \approx \mathbf{S}_k$ entspricht $(\bar{\mathbf{M}} + \sum_{p=1}^n \omega_{k,p} \mathbf{V}_p) \approx T_k(\mathbf{S}_k)$. Für das zweite Modell (mit objektabhängigen Merkmalsvektoren) ist es für die Parameterbestimmung nicht relevant, ob $\mathbf{S}_k$ oder das Modell transformiert wird.

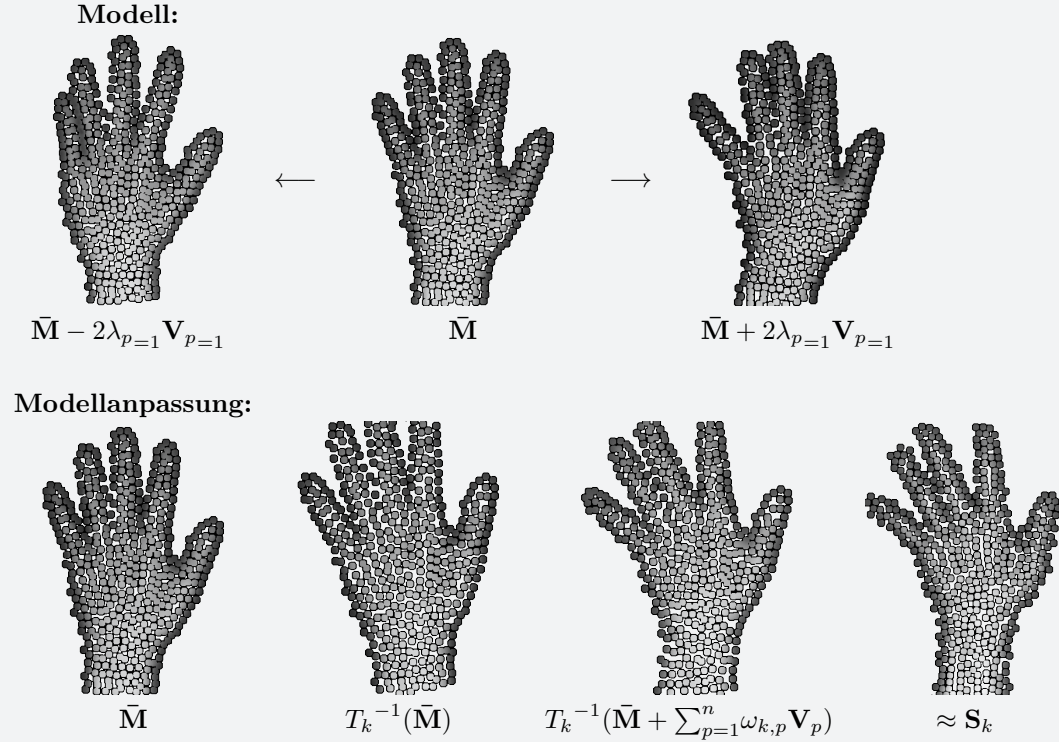


Abbildung 11: Visualisierung eines Modells samt Modellanpassung für “objektabhängige” Merkmalsvektoren.

Parameterbelegung

Input-/Trainingsdaten

N	$=$	60
D	$=$	4 ($D_f = 2$, $D_x = 2$)
$\mathbf{f}_i$	:	(1) Gauß-geglättete Signalwerte: $(G_{\sigma=6mm} * I_k)(\mathbf{x}_i)$ (2) Distanzwerte zur Handmaske: $d(\mathbf{x}_i, I_k^{Mask})$
$\mathbf{x}_i$	:	kantenmerkmalsabhängig mit Mindestabstand 6 mm, innerhalb der Hand ($\forall i, f_{i,2} < 5mm$)
sp	$=$	6 mm
$N_s(k)$	$=$	450 ± 57
N_m	$=$	481

Modellparameter

n	$=$	15
$T_k * \mathbf{s}_{k,i}$	$=$	$(T_k(\mathbf{x}_i), \mathbf{f}_i)$

Der Vorteil der zweiten Modellvariante liegt darin, dass sich die Merkmalsvektoren auch objektabhängig bewegen, wohingegen das erste Modell sich eher wie “Eigenfa-

ces” [Turk & Pentland 1991] verhält. Besonders bei der Hand, deren Form/Haltung stark variieren kann, ist eine objektabhängige Positionswahl sinnvoll, da Merkmalsvektoren im Bildhintergrund nicht zum Informationsgehalt der Hand beitragen. Daher wird diese Variante in folgenden Beispielen weiterverwendet. Darüber hinaus umfasst ein solches Modell weniger Vektoren und führt somit zu geringerem Rechenaufwand bei der Modellgenerierung.

3.2. Maximum-A-posteriori-Ansatz zur Parameteroptimierung

Das aufgestellte Problem, diejenigen Modellinstanzen $\mathbf{M}_k$ zu finden, welche die gegebene Trainingsdaten $\mathbf{S}_k$ am besten beschreiben $\forall k$

$$\mathbf{S}_k \approx \mathbf{M}_k \text{ mit } \mathbf{M}_k = T_k^{-1}(\bar{\mathbf{M}} + \sum_{p=1}^n \omega_{k,p} \mathbf{V}_p) \quad (21)$$

kann als Maximum-A-posteriori-Schätzung der Parameter Θ und $Q = \{Q_k | k = 1, \dots, N\}$ definiert werden:

$$p(Q, \Theta | \mathbf{S}_{\text{tr}}) \rightarrow \max_{Q, \Theta}. \quad (22)$$

Unter Nutzung des Bayes-Theorems wird folgendes Kriterium minimiert:

$$C(Q, \Theta) = -\log p(Q, \Theta | \mathbf{S}_{\text{tr}}) = -\log \left(\frac{p(\mathbf{S}_{\text{tr}} | Q, \Theta) p(Q | \Theta) p(\Theta)}{p(\mathbf{S}_{\text{tr}})} \right). \quad (23)$$

Die A-priori-Wahrscheinlichkeit der Trainingsinstanzen $p(\mathbf{S}_{\text{tr}})$ ist nicht abhängig von den Modellparametern und kann daher als konstant angenommen werden:

$$C(Q, \Theta) = -(\log p(\mathbf{S}_{\text{tr}} | Q, \Theta) + \log p(Q | \Theta) + \log p(\Theta)). \quad (24)$$

Im Folgenden wird auf alle Teilwahrscheinlichkeiten gesondert eingegangen. Die Modellparameter n und Σ werden im Rahmen des Abschnitts 3 als fix betrachtet. Die Wahl der Anzahl der Moden n wird in Abschnitt 5.2 diskutiert und da die Merkmalsvarianzen in Σ festlegen, welche Merkmalsvektoren miteinander korrespondieren (Gleichung 5), kann Σ nicht über das Kriterium $C(Q, \Theta)$ optimiert werden. Denn $C(Q, \Theta)$ würde minimal werden bei maximalen (unendlich großen) Werten in Σ , das heißt wenn “alle Vektoren der Trainingsdaten zu allen Vektoren des Modells” korrespondieren würden. In Abschnitt 4.2 wird die Bestimmung der Merkmalsvarianzen in Σ diskutiert.

3.2.1. Likelihood $p(\mathbf{S}_{\text{tr}}|Q, \Theta)$

Die Likelihood $p(\mathbf{S}_k|Q_k, \Theta)$ repräsentiert die Wahrscheinlichkeit, dass die Trainingsinstanzen über $\{Q_k\}$ vom Modell Θ generiert werden können. Die Instanzen untereinander werden als unabhängig angenommen:

$$p(\mathbf{S}_{\text{tr}}|Q, \Theta) = \prod_{k=1}^N p(\mathbf{S}_k|Q_k, \Theta). \quad (25)$$

Die Likelihood für eine Trainingsinstanz wird analog zur Gleichung 8 definiert und $p(\mathbf{s}_{k,i}|\mathbf{m}_{k,j})$ als Gauß-verteilt angenommen (Gleichung 5):

$$\begin{aligned} p(\mathbf{S}_k|Q_k, \Theta) &= \prod_{i=1}^{N_s(k)} \frac{1}{N_m} \sum_{j=1}^{N_m} p(\mathbf{s}_{k,i}|Q_k, \Theta) \\ &= \prod_{i=1}^{N_s(k)} \frac{1}{N_m} \sum_{j=1}^{N_m} p(\mathbf{s}_{k,i}|\mathbf{m}_{k,j}) \\ &= \prod_{i=1}^{N_s(k)} \frac{1}{N_m} \sum_{j=1}^{N_m} \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \Sigma^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \right), \end{aligned} \quad (26)$$

mit $\mathbf{m}_{k,j} = T_k^{-1} \star (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j})$ und $Q_k = \{T_k, \boldsymbol{\omega}_k\}$, $\Theta = \{\bar{\mathbf{M}}, \mathbf{V}_p, \lambda_p, n, \Sigma | p = 1, \dots, n\}$.

log von $p(\mathbf{S}_{\text{tr}}|Q, \Theta)$ für die Berechnung von $C(Q|\Theta)$:

$$\begin{aligned} &\text{argmin}_{Q, \Theta} - \log p(\mathbf{S}_{\text{tr}}|Q, \Theta) \quad (27) \\ &= \text{argmin}_{Q, \Theta} - \log \prod_{k=1}^N \prod_{i=1}^{N_s(k)} \left(\frac{1}{N_m} \sum_{j=1}^{N_m} p(\mathbf{s}_{k,i}|\mathbf{m}_{k,j}) \right) \\ &\quad N_m \text{ ist konstant} \\ &= \text{argmin}_{Q, \Theta} - \sum_{k=1}^N \sum_{i=1}^{N_s(k)} \log \left(\sum_{j=1}^{N_m} p(\mathbf{s}_{k,i}|\mathbf{m}_{k,j}) \right) \\ &= \text{argmin}_{Q, \Theta} - \sum_{k=1}^N \sum_{i=1}^{N_s(k)} \log \left(\sum_{j=1}^{N_m} \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \Sigma_j^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \right) \right) \\ &\quad \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} \text{ ist konstant} \\ &= \text{argmin}_{Q, \Theta} - \sum_{k=1}^N \sum_{i=1}^{N_s(k)} \log \left(\sum_{j=1}^{N_m} \exp \left(-\frac{1}{2} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \Sigma_j^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \right) \right) \end{aligned}$$

Mit der Minimierung von $-\log p(\mathbf{S}_{\text{tr}}|Q, \Theta)$ wird demnach die Distanz zwischen $\mathbf{S}_k$ und $\mathbf{M}_k$ minimiert.

3.2.2. A-priori-Wahrscheinlichkeiten der Modellparameter $p(\Theta)$

In den Vorarbeiten von Hufnagel et al. [Hufnagel *et al.* 2008] wurde die A-priori-Wahrscheinlichkeit des Modells als konstant angenommen. Damit sind “alle Modelle erlaubt” und unplausible topologische Verschiebungen der Merkmalsvektoren werden in keiner Weise bestraft. Sind die Korrespondenzen zwischen den Trainingsdaten (bzw. zwischen jeder Trainingsinstanz und dem Modell) bekannt bzw. können richtig bestimmt werden, wird die “Korrektheit” der Topologie dadurch gewährleistet, dass die Verschiebung/Variation jedes Merkmalsvektors über den gesamten Trainingsdatensatz modelliert (gemittelt) wird. Bei den probabilistischen Formmodellen liegen relativ wenige Punkte an definierten Positionen vor – z. B. ausschließlich auf der Objektgrenze – sodass die Korrespondenzbestimmung “relativ einfach” ist. Somit liefert die Mittelung der Formvariationen über den Trainingsdatensatz ausreichend Stabilität, um unplausible Topologien zu verhindern.

Im vorgestellten Appearance-Modell hingegen sind die Positionen der Merkmalsvektoren nicht mehr abhängig von der Objektform, wie es bei den Formmodellen der Fall ist. Allerdings heißt dies auch, dass die Korrespondenzbestimmung nicht mehr so “eindeutig” ist. In homogenen Gewebebereichen ist z. B. ohne topologische Annahme nicht eindeutig welche Vektoren miteinander korrespondieren. Durch die gemeinsame Definition der Positions- und Appearance-Merkmale pro Vektor kommt es außerdem zu einer wesentlich höheren Freiheit für das Modell, da die Distanz zwischen Trainingsinstanz und Modell bzw. Modellinstanz entweder durch eine Positionsanpassung der Modellvektoren minimiert werden kann oder durch eine Appearance-Merkmalanpassung.

Um die Korrespondenzbestimmung bzw. Modellgenerierung unter erhöhter Anzahl an Freiheitsgraden (gleichzeitige/gleichberechtigte Positions- und Appearance-Merkmaländerungen) und erhöhter Anzahl an Merkmalsvektoren mit “schlecht definierten Positionen” zu regularisieren, wird eine nicht konstante Wahrscheinlichkeit $p(\Theta)$ notwendig.

Es ist davon auszugehen, dass bereits eine Gewichtung der Appearance- und Positionsanpassung durch die Werte in Σ (Kovarianzmatrix in Gleichung 5) stattfindet. Die Matrix bestimmt, welche Merkmalsvektorkpaare als “stark korrespondierend” betrachtet werden – diejenigen mit ähnlichen Positionen oder mit ähnlichen Appearance-Merkmalen.

In Abbildung 12 werden die Probleme der gleichzeitigen Anpassung beschrieben: $\mathbf{M}_k(a)$ zeigt ein Modell, das hauptsächlich die Appearance-Werte anpasst und die Position der Modellvektoren an die jeweils nächstgelegenen Instanzvektoren (mit geringster Positionsdistanz) anpasst. Dieses Verhalten kommt der Idee der Eigenfaces nahe, welche die Position nicht ändern würden und ausschließlich mit Appearance-Wertänderungen arbeiten [Turk & Pentland 1991]. $\mathbf{M}_k(b)$ stellt ein Modell dar, das hingegen die Appearance-Merkmale kaum ändert und die Distanz zwischen $\mathbf{M}_k$ und $\mathbf{S}_k$ vor allem durch

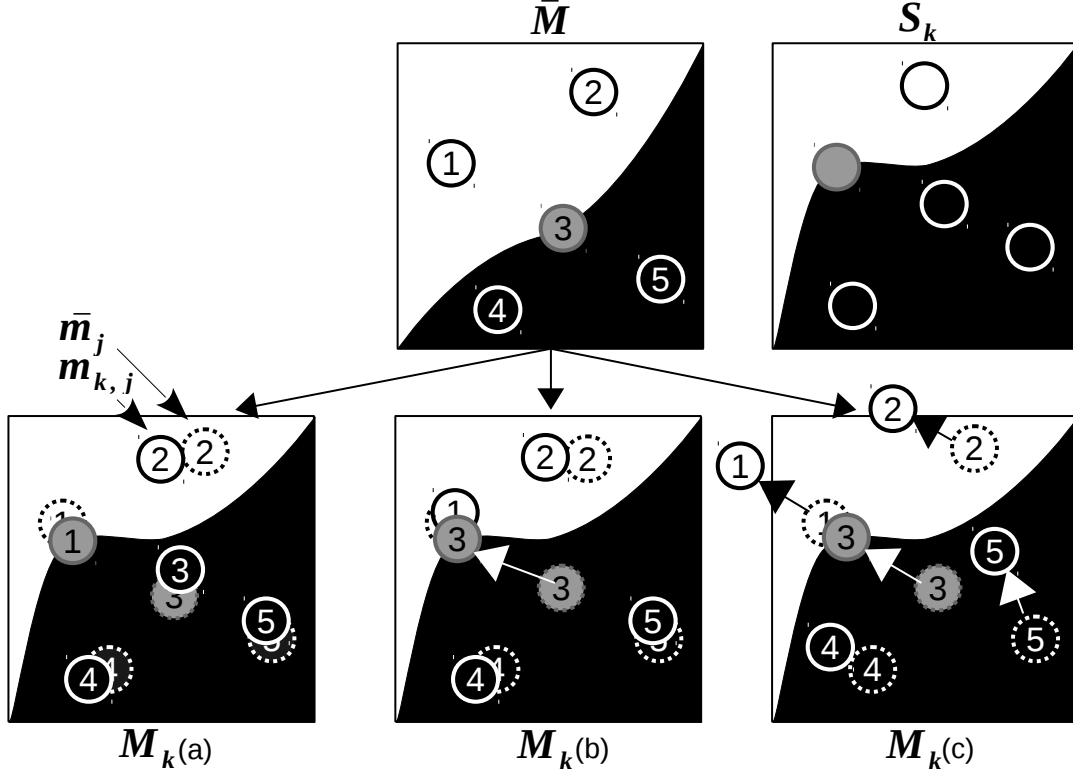


Abbildung 12: Gegenüberstellung von Appearance-Merkmal- und Positionsanpassung: Es sei ein Ausschnitt von fünf Merkmalsvektoren eines Modells und einer Trainingsinstanz mit entsprechender Modellinstanz für drei sich unterschiedlich verhaltende Modelle betrachtet. Im (mittleren) Modell $\bar{M}$ besitzt der Vektor ③ einen Appearance-Wert der “Grenzregion” zwischen beiden Geweben (Vektor ① und ② liegen im “weißen” und Vektor ④ und ⑤ im “schwarzen Gewebe”). S_k stellt exemplarisch eine Trainingsinstanz dar und $M_k(a)$ -(c) zeigen die entsprechenden Modellinstanzen korrespondierend zu S_k , die durch drei unterschiedliche Modelle erzeugt werden. Dabei indizieren die Merkmalsvektoren mit dem gepunkteten Rand die initiale Position des mittleren Modells ($\bar{\mathbf{m}}_j$) und die Merkmalsvektoren mit dem durchgezogenen Rand seien die adaptierten Modellvektoren ($\mathbf{m}_{k,j} = (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j})$, die affine Transformation sei hier nicht betrachtet): Das Modell zu $M_k(a)$ bevorzugt Appearance-Merkmaländerungen und adaptiert die Position der Modellvektoren zum nächstgelegenen Merkmalsvektor der Trainingsinstanz. In diesem Fall ändern Vektor ① und ③ ihre “Rollen” und werden zum “Grenzpunkt” bzw. “schwarzen Gewebe”. Das Modell zu $M_k(b)$ hingegen ändert die Appearance-Werte kaum und minimiert die Distanz zwischen M_k und S_k vor allem durch Positionsänderungen. Merkmalsvektor ③ bleibt daher ein “Grenzpunkt”. Es kann allerdings zur Kollision von Merkmalsvektoren kommen, da diese vollkommen unabhängig voneinander betrachtet werden (ohne Topologieannahmen – $p(\Theta)$ konstant). Außerdem liegen z. B. die Merkmalsvektoren ① und ② im homogenen Gewebe und besitzen daher keinen Anhaltspunkt für eine sinnvolle Positionsänderung. $M_k(c)$ beschreibt ein Modell mit Moden, welche die Distanzen zwischen Merkmalsvektoren relativ stabil halten (mit Nachbarschaftsannahme) und trotzdem große Positionsänderungen zulassen, falls erforderlich.

Positionsänderungen minimiert. Wenn $p(\Theta)$ konstant ist, kann dies zu großen Topologieveränderungen führen, da die Merkmalsvektoren komplett unabhängig voneinander betrachtet werden. Ziel sollte es sein, ein Modell zu generieren, das die Balance zwischen Appearance-Wert- und Positionsanpassungen findet, wobei die Merkmalsvektoren zu bestimmten Strukturen “gehören” und trotzdem ihre topologische Integrität beibehalten, wie in $\mathbf{M}_k(c)$. Die Appearance-Werte sollten lediglich im Rahmen der Appearance-Variabilität innerhalb bestimmter Strukturen/Gewebe adaptiert werden.

Um dies zu erreichen, können die Merkmalsvektoren nicht mehr unabhängig voneinander betrachtet werden und es muss eine Topologieannahme eingeführt werden. Zu diesem Zweck werden von nun an die Merkmalsvektoren des Modells in Abhängigkeit ihrer lokalen Nachbarn optimiert. Die Parameterwerte des “Modellpunktes” j – samt dazugehörenden Werten $\mathbf{v}_{p,j}$ und $\bar{\mathbf{m}}_j$ – sei also abhängig von ihren jeweiligen Nachbarn $(\mathbf{v}_{p,\eta}, \bar{\mathbf{m}}_\eta)$, $\eta \in \mathcal{N}(j)$. Dabei sei die Anzahl an Nachbarn $|\mathcal{N}(j)|$ pro Modellvektor j beliebig wählbar. $p(\Theta)$ kann bestimmt werden durch:

$$\begin{aligned} p(\Theta) = \prod_{p=1}^n p(\mathbf{V}_p, \lambda_p, \bar{\mathbf{M}}) &= \prod_{p=1}^n \prod_{j=1}^{N_m} \prod_{\eta \in \mathcal{N}(j)} p(\mathbf{v}_{p,j}, \mathbf{v}_{p,\eta}, \lambda_p, \bar{\mathbf{m}}_j, \bar{\mathbf{m}}_\eta) \\ &= \prod_{p=1}^n \prod_{j=1}^{N_m} \prod_{\eta \in \mathcal{N}(j)} p(\mathbf{v}_{p,j}, \bar{\mathbf{m}}_j, \lambda_p | \mathbf{v}_{p,\eta}, \bar{\mathbf{m}}_\eta) p(\mathbf{v}_{p,\eta}, \bar{\mathbf{m}}_\eta). \end{aligned} \quad (28)$$

Die Wahrscheinlichkeit $p(\mathbf{v}_{p,\eta}, \bar{\mathbf{m}}_\eta)$ wird als konstant angenommen.

Die Idee besteht darin, große lokale Topologieänderungen zu bestrafen. Es werden somit die Werte bevorzugt, welche die lokalen Abstände zwischen den Merkmalsvektoren des initialen Modells ($\bar{\mathbf{M}}$) relativ konstant halten.

$$\frac{|(\mathbf{m}_{\cdot,j} - \mathbf{m}_{\cdot,\eta}) - (\bar{\mathbf{m}}_j - \bar{\mathbf{m}}_\eta)|}{\|(\bar{\mathbf{m}}_j - \bar{\mathbf{m}}_\eta)\|}, \forall \eta \in \mathcal{N}(j) \quad (29)$$

soll also klein bleiben, wobei $\mathbf{m}_{\cdot,j}, \mathbf{m}_{\cdot,\eta}$ “alle erlaubten Modellinstanzvarianten” für den Vektor j samt Nachbarn repräsentieren. Um das Modell zu regularisieren, werden diese erlaubten Varianten mit der zweifachen Standardabweichung λ_p der Moden beschränkt: $\mathbf{m}_{\cdot,j} = \bar{\mathbf{m}}_j + \sum_{p=1}^n \pm 2\lambda_p \mathbf{v}_{p,j}$, das heißt, man setze $\omega_{l,p} = \pm 2\lambda_p$. Für die Bestimmung der Distanzvektoren

$$\mathcal{N}_{j,\eta}^0 = (\bar{\mathbf{m}}_j - \bar{\mathbf{m}}_\eta) \quad (30)$$

gibt es zwei Möglichkeiten: (1) Entweder sie werden einmal initialisiert mit initialem $\bar{\mathbf{M}}$ und bleiben während der Modelloptimierung konstant oder (2) sie werden mit den aktuellen Werten des mittleren Modells upgedatet. Ein Update der Distanzvektoren ist sinnvoll, wenn sich das initiale mittlere Modell noch stark während der Modelloptimierung ändert. Da die Merkmalsvektoren des Modells unabhängig von der Entfernung ihrer Nachbarn regularisiert werden sollen, werden die Distanzvektoren normalisiert mit

$$\|\mathcal{N}_{j,\eta}^0\| = \|(\bar{\mathbf{m}}_j - \bar{\mathbf{m}}_\eta)\|.$$

Analog zu $p(\mathbf{s}_{k,i}|\mathbf{m}_{k,j})$ kann nun $p(\mathbf{v}_{p,j}, \lambda_p, \bar{\mathbf{m}}_j|\mathbf{v}_{p,\eta}, \bar{\mathbf{m}}_\eta)$ als Gauß-verteilt modelliert werden:

$$\begin{aligned} p(\Theta) &= \prod_{j=1}^{N_m} \prod_{\eta \in \mathcal{N}(j)} \prod_{p=1}^n p(\mathbf{v}_{p,j}, \bar{\mathbf{m}}_j, \lambda_p | \mathbf{v}_{p,\eta}, \bar{\mathbf{m}}_\eta) \\ &= \prod_{j=1}^{N_m} \prod_{\eta \in \mathcal{N}(j)} \prod_{p=1}^n \frac{1}{(2\pi)^{D/2} \|\Sigma_{\mathcal{N}}\|} \\ &\quad \exp \left(-\frac{1}{2} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_\eta \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \right. \\ &\quad \left. \Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_\eta \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right). \end{aligned} \quad (31)$$

Die Schreibweise mit $\pm$ bedeutet hier und in allen folgenden Gleichungen, dass das gesamte Produkt einmal mit “−” und einmal mit “+” berechnet wird: $\prod((a \pm x) - (b \pm y)) = \prod(((a - x) - (b - y)) \cdot ((a + x) - (b + y)))$. $\Sigma_{\mathcal{N}} \in \mathbb{R}^{D \times D}$ beschreibt die erlaubte Variation der lokalen Distanzen zwischen mittlerem Modell und möglicher Modellinstanz. Über $p(\Theta)$ können sowohl die Positionen als auch die Appearance-Merkmale regularisiert werden. Soll aber ausschließlich die räumliche Topologie regularisiert werden, können in $\Sigma_{\mathcal{N}}$ die Werte für die Appearance-Merkmale sehr hoch gewählt werden – siehe Abschnitt 4.2.2 für die Bestimmung der Werte in $\Sigma_{\mathcal{N}}$.

Es ist zu beachten, dass $p(\Theta)$ keine absoluten Wertänderungen bestraft, sondern nur relative Distanzen zwischen lokalen Nachbarn regularisiert, um extreme topologische Änderungen zu vermeiden wie z. B. Reihenfolgeänderungen/Faltungen oder “Kollisionen” von Merkmalsvektoren. Dabei beschreibt $p(\Theta)$ eine lokale Glättung und keine strikte Bedingung, um z. B. Faltungseffekte komplett auszuschließen.

log von $p(\Theta)$ für die Berechnung von $C(Q|\Theta)$: Für die Minimierung sei $\frac{1}{(2\pi)^{Dx/2} \|\Sigma_{\mathcal{N}}\|}$ (siehe Gleichung 31) konstant:

$$\begin{aligned} &\text{argmin}_{Q,\Theta} - \log p(\Theta) \\ &= \text{argmin}_{Q,\Theta} - \log \prod_{j=1}^{N_m} \prod_{\eta \in \mathcal{N}(j)} \prod_{p=1}^n p(\mathbf{v}_{p,j}, \lambda_p | \bar{\mathbf{m}}_j, \bar{\mathbf{m}}_\eta, \mathbf{v}_{p,\eta}) \\ &= \text{argmin}_{Q,\Theta} - \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n \log p(\mathbf{v}_{p,j}, \lambda_p | \bar{\mathbf{m}}_j, \bar{\mathbf{m}}_\eta, \mathbf{v}_{p,\eta}) \end{aligned} \quad (32)$$

$$= \operatorname{argmin}_{Q, \Theta} \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n \left(\frac{1}{2} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_\eta \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \right. \\ \left. \Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_\eta \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right)$$

(Anmerkung zu $\pm$: die gesamte Summe wird jeweils einmal mit “−” und einmal mit “+” berechnet: $\sum((a \pm x) - (b \pm y)) = \sum(((a - x) - (b - y)) + ((a + x) - (b + y)))$.)

Beispielbox 4: Auswirkung der A-priori-Wahrscheinlichkeit

Um den Einfluss der A-priori-Wahrscheinlichkeit der Modellparameter $p(\Theta)$ zu visualisieren, wurde im folgenden Beispiel ein Modell mit und ein Modell ohne $p(\Theta)$ generiert. Die initialen lokalen Nachbarschaftsbeziehungen für $p(\Theta)$ werden in einem folgenden Beispiel beschrieben. Es ist ersichtlich, dass die erste Modellmode und die Modellinstanzen ohne $p(\Theta)$ ein weniger stabiles Verhalten aufweisen als unter Verwendung von $p(\Theta)$.

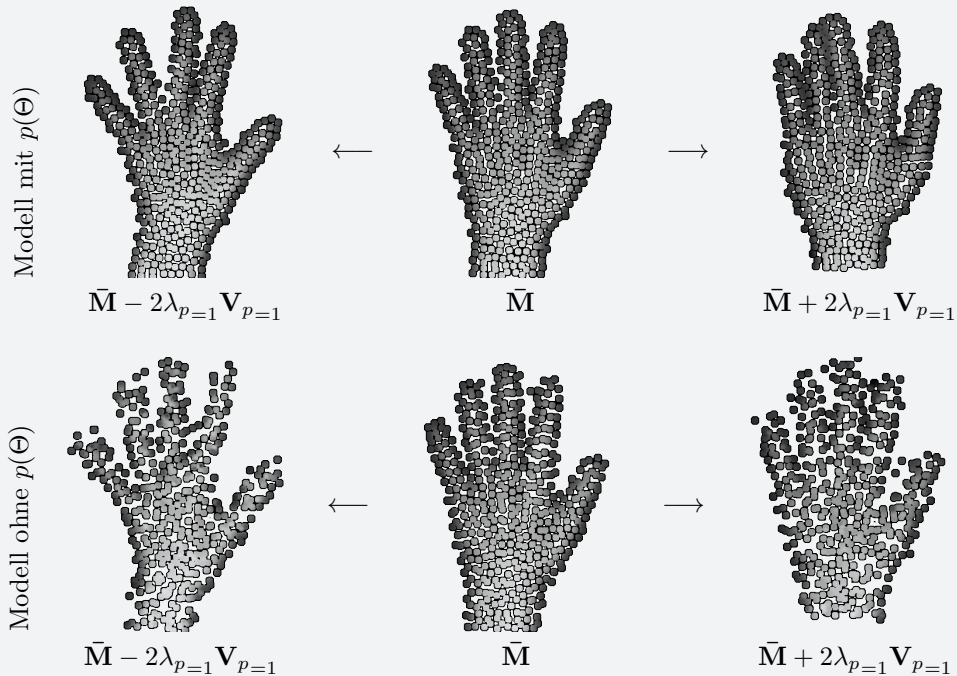
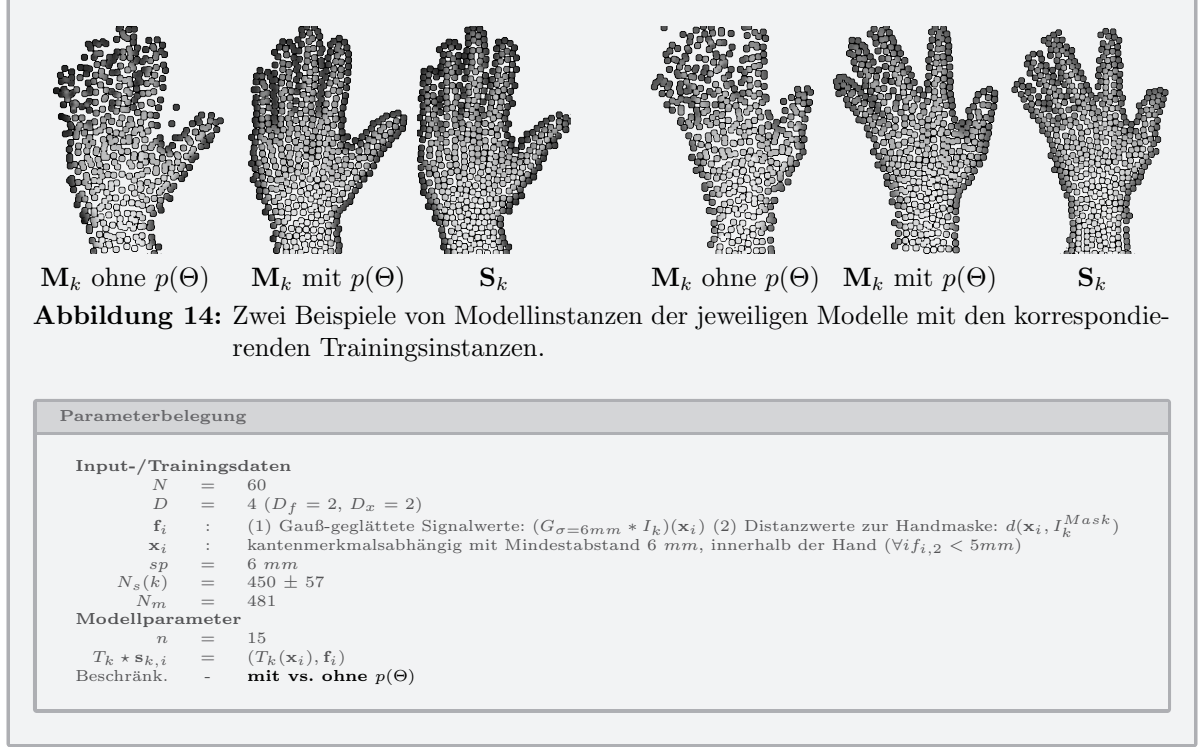


Abbildung 13: Gegenüberstellung von Modellen mit bzw. ohne der A-priori-Wahrscheinlichkeit $p(\Theta)$.



3.2.3. A-priori-Wahrscheinlichkeit der instanzabh\ddot{a}ngigen Parameter

$p(Q_k|\Theta)$

Die A-priori-Wahrscheinlichkeit der instanzabh\ddot{a}ngigen Transformation T_k wird als unabh\ddot{a}ngig von den Modellparametern und den Gewichten ω_k angenommen: $p(T_k|\Theta) = p(T_k)$. Au\ddot{a}erdem wird $p(T_k)$ als konstant betrachtet. Daher bleibt es $p(\omega_k|\Theta)$ zu betrachten. Die Gewichte ω_k der Variationsmoden werden als Gau\ss-verteilt mit Standardabweichung λ_p modelliert, um sie auf sinnvolle Werte zu beschr\ddot{a}nken:

$$p(Q_k|\Theta) = p(\omega_k|\Theta) = \prod_{p=1}^n \frac{1}{(2\pi)^{1/2}\lambda_p} \exp\left(-\frac{\omega_{k,p}^2}{2\lambda_p^2}\right). \quad (33)$$

log von $p(Q_k|\Theta)$ f\ddot{u}r die Berechnung von $C(Q|\Theta)$: n wird an dieser Stelle als fix betrachtet.

$$\begin{aligned} & \operatorname{argmin}_{Q,\Theta} -\log \prod_{k=1}^N p(Q_k|\Theta) \\ &= \operatorname{argmin}_{Q,\Theta} -\log \prod_{k=1}^N \prod_{p=1}^n \left(\frac{1}{(2\pi)^{1/2}\lambda_p} \exp\left(-\frac{\omega_{k,p}^2}{2\lambda_p^2}\right) \right) \end{aligned} \quad (34)$$

$$\begin{aligned}
&= \operatorname{argmin}_{Q, \Theta} - \sum_{k=1}^N \sum_{p=1}^n \log \left(\frac{1}{(2\pi)^{1/2} \lambda_p} \exp \left(-\frac{\omega_{k,p}^2}{2\lambda_p^2} \right) \right) \\
&= \operatorname{argmin}_{Q, \Theta} - \sum_{k=1}^N \sum_{p=1}^n \left(\log(1) - \log((2\pi)^{1/2}) - \log(\lambda_p) - \left(\frac{\omega_{k,p}^2}{2\lambda_p^2} \right) \right) \\
&= \operatorname{argmin}_{Q, \Theta} + \sum_{k=1}^N \sum_{p=1}^n \left(\log(\lambda_p) + \left(\frac{\omega_{k,p}^2}{2\lambda_p^2} \right) \right)
\end{aligned}$$

3.2.4. Zusätzliche Symmetriebedingungen für die Modellgenerierung in $p(Q|\Theta)$

Wird das Modell auf ein neues Bild adaptiert, sollten keine Einschränkungen für die Wahl der Transformation T_{new} oder der Moden-Gewichte ω_{new} gelten – ausgenommen der Beschränkung durch die Standardabweichungen λ_p . Während der Modellgenerierung hat sich aber gezeigt, dass es sinnvoll ist, Annahmen über die Verteilung der Gewichte und Transformationen über den Trainingsdatensatz zu treffen, da bis zu diesem Punkt im Kriterium nicht festgelegt ist, dass das mittlere Modell auch “in der Mitte” liegt. Bei dem klassischen Modellgenerierungsansatz wird dies durch eine initiale einmalige Mittelung der ausgerichteten korrespondierenden Landmarken und aller korrespondierenden Appearance-Werten über alle Trainingsdaten gewährleistet. In diesem Ansatz hingegen wird $\mathbf{M}$ in jeder Iteration optimiert. Bis hierher ist über die Likelihood mit $p(\mathbf{s}_{k,i}, \mathbf{m}_{k,j})$ lediglich festgelegt, dass die resultierende Modellinstanz $\mathbf{M}_k$ mit $\mathbf{m}_{k,j} = T_k^{-1} \star (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j})$ zur Trainingsinstanz $\mathbf{S}_k$ “passen soll” – nicht jedoch welchen Anteil dabei $\bar{\mathbf{m}}_j$, $\omega_{k,p}$, T_k und $\mathbf{v}_{p,j}$ haben sollen. Das heißt, ein “mittleres” Modell, das komplett außerhalb des Wertebereichs der Trainingsdaten liegt, kann durch die entsprechenden instanzabhängigen Parameter ausgeglichen werden. Dies führt zwar zu optimalen Ergebnissen die resultierenden $\mathbf{M}_k$ betreffend, jedoch leidet das generierte Modell an Overfitting an die Trainingsinstanzen, da die Modellparameter nur optimale Ergebnisse im Zusammenhang mit den optimalen instanzabhängigen Parametern liefern.

Um diesem Effekt entgegenzuwirken, kann eine Bedingung eingeführt werden – $p_{sym}(Q)$. Diese legt fest, dass die Gewichte und Transformationen über den gesamten Trainingsdatensatz einen Mittelwert von Null (bzw. Einheitsmatrix I_{D+1} für T_k) haben sollten. Dadurch ist gewährleistet, dass ein unsymmetrisches Modell nicht durch im selben Maße unsymmetrische instanzabhängige Parameter ausgeglichen werden kann:

$$p(Q|\Theta) = \left(\prod_{k=1}^N p(Q_k|\Theta) \right) \cdot p_{sym}(Q) \quad (35)$$

$$\begin{aligned}
\prod_{k=1}^N p(Q_k|\Theta) &= \left(\prod_{k=1}^N \prod_{p=1}^n \frac{1}{(2\pi)^{1/2} \lambda_p} \exp \left(-\frac{\omega_{k,p}^2}{2\lambda_p^2} \right) \right) \\
p_{sym}(Q) &= \frac{1}{(2\pi)^{1/2} \sigma_{\bar{\omega}}} \exp \left(-\frac{\|\frac{1}{N} \sum_{k=1}^N \omega_k\|^2}{2\sigma_{\bar{\omega}}^2} \right) \\
&\quad \cdot \frac{1}{(2\pi)^{1/2} \sigma_{\bar{T}}} \exp \left(-\frac{\|(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}\|_F^2}{2\sigma_{\bar{T}}^2} \right).
\end{aligned}$$

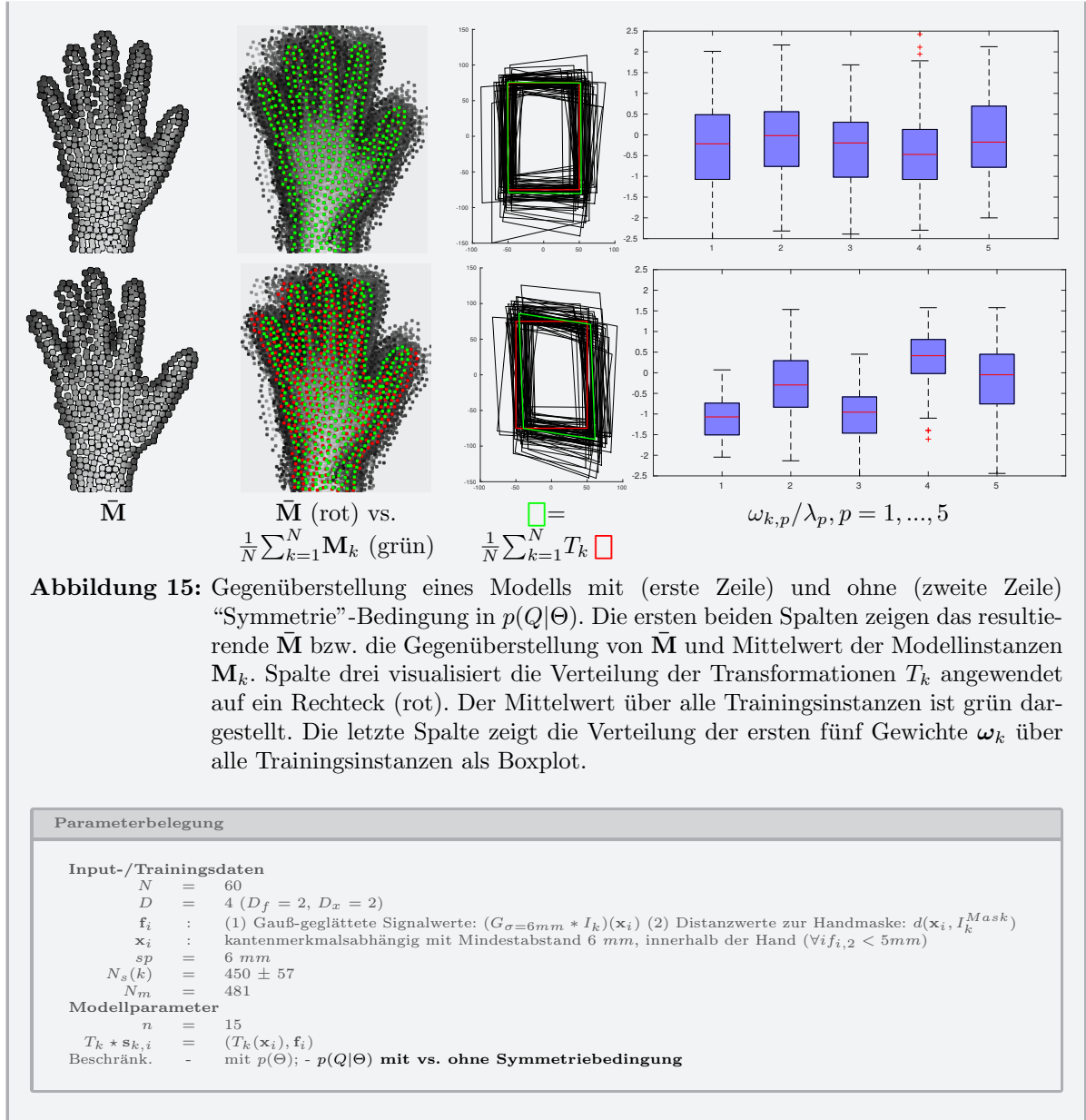
Auch hier sind $\sigma_{\bar{\omega}}$ und $\sigma_{\bar{T}}$ Regulierungsparameter, die in diesem Fall fest auf $\ll 0.001$ gesetzt werden können, um einen Mittelwert von Null zu erzwingen.

log von $p(Q|\Theta)$ für die Berechnung von $C(Q|\Theta)$:

$$\begin{aligned}
&\text{argmin}_{Q,\Theta} - \log p(Q|\Theta) \tag{36} \\
&= \text{argmin}_{Q,\Theta} - \left(\log \prod_{k=1}^N \prod_{p=1}^n \frac{1}{(2\pi)^{1/2} \lambda_p} \exp \left(-\frac{\omega_{k,p}^2}{2\lambda_p^2} \right) \right) \\
&\quad - \log \left(\frac{1}{(2\pi)^{1/2} \sigma_{\bar{\omega}}} \exp \left(-\frac{\|\frac{1}{N} \sum_{k=1}^N \omega_k\|^2}{2\sigma_{\bar{\omega}}^2} \right) \right) \\
&\quad - \log \left(\frac{1}{(2\pi)^{1/2} \sigma_{\bar{T}}} \exp \left(-\frac{\|(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}\|_F^2}{2\sigma_{\bar{T}}^2} \right) \right) \\
&= \text{argmin}_{Q,\Theta} + \left(\sum_{k=1}^N \sum_{p=1}^n \log(\lambda_p) + \left(\frac{\omega_{k,p}^2}{2\lambda_p^2} \right) \right) + \left(\frac{\|\frac{1}{N} \sum_{k=1}^N \omega_k\|^2}{2\sigma_{\bar{\omega}}^2} \right) \\
&\quad + \left(\frac{\|(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}\|_F^2}{2\sigma_{\bar{T}}^2} \right)
\end{aligned}$$

Beispielbox 5: Auswirkung der Symmetriebedingung

Im Folgenden wurden zwei Modelle generiert, um den Einfluss der “Symmetrie”-Bedingung in $p(Q|\Theta)$ zu visualisieren. Das erste Modell mit der Bedingung (erste Zeile) weist nach der Modellgenerierung ein mittleres Modell auf, das dem Mittelwert der Modellinstanzen (korrespondierend zu den Trainingsdaten) entspricht ($\bar{\mathbf{M}} \approx \frac{1}{N} \sum_{k=1}^N \mathbf{M}_k$). Im Gegensatz dazu weicht das $\bar{\mathbf{M}}$ (rote Punkte) des Modells ohne Bedingung (zweite Reihe) vom Mittelwert (grüne Punkte) ab. Außerdem sind die Verteilungen der instanzabhängigen Parameter (T_k und ω_k) nach der Anpassung des resultierenden Modells mit Bedingung annähernd mittelwertbefreit.



3.3. Globales Kriterium

Zusammenfassend ergibt sich folgendes Kriterium zur Beschreibung der optimalen Modellparameter:

$$\begin{aligned}
& \operatorname{argmin}_{Q, \Theta} C(Q, \Theta) \\
&= \operatorname{argmin}_{Q, \Theta} -\log p(Q, \Theta | \mathbf{S}_{\text{tr}}) \\
&= \operatorname{argmin}_{Q, \Theta} -\log \left(\frac{p(\mathbf{S}_{\text{tr}} | Q, \Theta) p(Q | \Theta) p(\Theta)}{p(\mathbf{S}_{\text{tr}})} \right), p(\mathbf{S}_{\text{tr}}) \text{ const.} \\
&= \operatorname{argmin}_{Q, \Theta} -\log(p(\mathbf{S}_{\text{tr}} | Q, \Theta) p(Q | \Theta) p(\Theta)) \\
&= \operatorname{argmin}_{Q, \Theta} -\log p(\mathbf{S}_{\text{tr}} | Q, \Theta) -\log p(Q | \Theta) -\log p(\Theta) \\
&= \operatorname{argmin}_{Q, \Theta} -\sum_{k=1}^N \left(\sum_{i=1}^{N_s(k)} \log \left(\sum_{j=1}^{N_m} \exp \left(-\frac{1}{2} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \Sigma_j^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \right) \right) \right) \\
&\quad + \left(\left(\sum_{k=1}^N \sum_{p=1}^n \log(\lambda_p) + \left(\frac{\omega_{k,p}^2}{2\lambda_p^2} \right) \right) + \left(\frac{\|\frac{1}{N} \sum_{k=1}^N \boldsymbol{\omega}_k\|^2}{2\sigma_{\bar{\omega}}^2} \right) \right. \\
&\quad \left. + \left(\frac{\|(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}\|_F^2}{2\sigma_{\bar{T}}^2} \right) \right) \\
&\quad + \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n \left(\frac{1}{2} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \right. \\
&\quad \left. \Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right)
\end{aligned} \tag{37}$$

4. Modellgenerierung und Modellanpassung

Ist eine Trainingsmenge an Bildrepräsentationen $\mathbf{S}_{\text{tr}} = \{\mathbf{S}_k | k = 1, \dots, N\}$ gegeben, wird das Modell, das diese Daten optimal abbildet, gesucht. Im Abschnitt 3 wurde zu diesem Zweck die Lösung des Problems über eine Maximum-A-posteriori-Schätzung der Modellparameter Θ und der instanzabhängigen Parameter $Q = \{Q_k | k = 1, \dots, N\}$ vorgestellt. Um die Parameter über

$$p(Q, \Theta | \mathbf{S}_{\text{tr}}) \rightarrow \max_{Q, \Theta}$$

zu optimieren, wurde ein globales Kriterium $C(Q, \Theta)$ (Gleichung 37) aufgestellt.

Das gesuchte Modell wird durch die populationsabhängigen Modellparameter Θ (Abschnitt 3.1.1) definiert. Das heißt als *Modellgenerierung* wird die Optimierung des globalen Kriteriums nach Θ bezeichnet. Durch die unbekannten Korrespondenzen zwischen Trainingsdatensatz und Modell müssen darüber hinaus gleichzeitig die instanzabhängigen Parameter Q_k optimiert werden – siehe Abbildung 8. Das heißt, die Modellgenerierung kann zusammengefasst werden als:

gegeben: N Trainingsinstanzen $\mathbf{S}_{\text{tr}} = \{\mathbf{S}_k | k = 1, \dots, N\}$

gesucht: das Modell Θ – mithilfe der Optimierung von

$$\operatorname{argmin}_{Q, \Theta} C(Q, \Theta) = \operatorname{argmin}_{Q, \Theta} -\log p(Q, \Theta | \mathbf{S}_{\text{tr}}).$$

Ist schließlich ein Modell gegeben, kann es an neue Instanzen $\mathbf{S}_{\text{new}}$ angepasst werden. Das Modell wird durch die instanzabhängigen Parameter $Q_{\text{new}} = \{T_{\text{new}}, \omega_{\text{new}}\}$ (Abschnitt 3.1.1) manipuliert:

$$\mathbf{M}_{\text{new}} = T_{\text{new}}^{-1}(\bar{\mathbf{M}} + \omega_{\text{new}} \mathbf{V}).$$

Als *Modellanpassung* wird demnach die Optimierung desselben globalen Kriteriums nach Q_{new} mit fixer Parameterbelegung für Θ bezeichnet. Dies kann zusammengefasst werden als:

gegeben: das Modell Θ und eine Instanz $\mathbf{S}_{\text{new}}$

gesucht: die Modellinstanz $\mathbf{M}_{\text{new}}$, die über die instanzabhängigen Parameter $Q_{\text{new}} = \{T_{\text{new}}, \omega_{\text{new}}\}$ definiert ist und $\mathbf{S}_{\text{new}}$ “am ähnlichsten” ist – mithilfe der Optimierung von

$$\operatorname{argmin}_{Q_{\text{new}}} C(Q_{\text{new}}, \Theta) = \operatorname{argmin}_{Q_{\text{new}}} -\log p(Q_{\text{new}}, \Theta | \mathbf{S}_{\text{new}}).$$

In Abschnitt 4.1 wird zunächst die Optimierung des Kriteriums nach den einzelnen Parametern beschrieben. Dabei wird zuerst (Abs. 4.1.1) auf die Optimierung durch einen

Expectation-Maximization-Ansatz eingegangen. Abschnitte 4.1.2 bis 4.1.7 gehen anschließend detailliert auf die für die Optimierung genutzten Ableitungen des Kriteriums nach den jeweiligen Parametern ein. In Abschnitt 4.2 werden ausführlich die Regularisierungsparameter (Σ, Σ_N) und ihre Bedeutung als wichtigste Parameter des Modells diskutiert und Abschnitt 4.3 erläutert die Initialisierung aller Parameter. Die Optimierung des Kriteriums in Hinblick auf die instanzabhängigen Parameter zum Zweck der Modellanpassung wird in Abschnitt 4.4 dargelegt. Am Ende des Kapitels (Abschnitt 4.5, Abbildung 23 und Abbildung 24) befindet sich eine Zusammenfassung des vorgestellten Modellgenerierungs- und Modellanpassungsansatzes.

4.1. Optimierung der Parameter anhand eines globalen Kriteriums

Durch die Optimierung des globalen Kriteriums $C(Q, \Theta)$ aus Gleichung 37 sollen Modell- und instanzabhängige Parameter bestimmt werden, welche die gegebene Trainingsmenge abbilden können, $\forall k$:

$$\mathbf{S}_k \approx \mathbf{M}_k \text{ mit } \mathbf{M}_k = T_k^{-1}(\bar{\mathbf{M}} + \sum_{p=1}^n \omega_{k,p} \mathbf{V}_p)$$

Eine probabilistische Formulierung der Ähnlichkeit zwischen einer Trainingsinstanz $\mathbf{S}_k$ und der entsprechenden Modellinstanz $\mathbf{M}_k$ wurde in Abschnitt 2.3 eingeführt ($p(\mathbf{S}_k|\mathbf{M}_k)$). Für die Berechnung von $p(\mathbf{S}_k|\mathbf{M}_k)$ werden neben den Modell- und instanzabhängigen Parametern (Θ, Q) ebenfalls die Korrespondenzwahrscheinlichkeiten $\mathbb{E}(H_{ij})$ benötigt (siehe Gleichung 10). H_{ij} sind dabei versteckte Variablen, welche die unbekannten Korrespondenzen zwischen den Merkmalsvektoren beschreiben. Die Schätzung des Erwartungswerts $\mathbb{E}(H_{ij})$ und der Parameter hängen wechselseitig voneinander ab, was bei der Optimierung von $C(Q, \Theta)$ berücksichtigt werden muss.

4.1.1. Expectation-Maximization-Algorithmus

Diese Art der Problemstellung, bei welcher die Eigenschaften der betrachteten Zufallsvariablen nur teilweise bekannt sind, kann durch einen Expectation-Maximization(EM)-Ansatz gelöst werden. Das Ziel von EM-Ansätzen ist die Maximierung einer ML- oder MAP-Schätzung zur Parameteroptimierung, welche von versteckten Variablen abhängen [Sundberg 1974; Dempster *et al.* 1977]. Aufgrund dieser versteckten/unbekannten Variablen kann das Problem nicht direkt gelöst werden, da die Optimierung der gesuchten Parameter typischerweise von der Verteilung der unbekannten Variablen abhängt und wiederum die Verteilung der unbekannten Variablen von den zu optimierenden Parametern abhängt. Daher ist ein iterativer Lösungsansatz notwendig, der beide Probleme alternierend löst – dabei werden abwechselnd zwei Schritte bis zur Konvergenz durchgeführt:

Expectation(E)-Schritt: Schätze die Erwartungswerte der log-Likelihood im Hinblick auf die versteckten Variablen. Die gesuchten Parametern werden dabei als fix betrachtet.

Maximization(M)-Schritt: Maximiere die MAP/ML-Schätzung in Abhängigkeit der gesuchten Parameter. Die geschätzten versteckten Variablen werden dabei als fix betrachtet.

Im vorliegenden Fall stellen die Korrespondenzen zwischen den jeweiligen $\mathbf{S}_k$ und $\mathbf{M}_k$ die versteckten Variablen dar und die zu optimierenden Parameter bilden Θ und $Q = \{Q_k | k = 1, \dots, N\}$. Da das Kriterium für die Parameteroptimierung über eine Maximum-Likelihood-Methode definiert wurde, bietet sich die Verwendung eines EM-Ansatzes an. Es wurde bewiesen, dass ein EM-Ansatz zu einem (lokalen) Minimum konvergiert [Wu 1983]. Eine Alternative wäre z. B. die Nutzung eines Gradientenabstieg- oder Gauß-Newton-Verfahrens für alle Parameter (versteckt und nicht versteckt). Diese konvergieren jedoch langsamer oder sind pro Iteration rechenaufwendiger (z. B. durch die Berechnung der Hesse-Matrix).

Für die Optimierung der Modell- und instanzabhängigen Parameter mittels globalem Kriterium bedeutet die Anwendung eines EM-Ansatzes die iterative Durchführung der beiden Schritte:

E-Schritt:

- Berechne Korrespondenzwahrscheinlichkeiten $\mathbb{E}(H_{ij})_k$ zwischen $\mathbf{S}_k$ und $\mathbf{M}_k$, $k = 1, \dots, N$ (Gleichung 5, 10 und 12)

M-Schritt:

- Berechne das Modell durch Optimierung von $C(Q|\Theta)$ nach jeweils $\bar{\mathbf{M}}$, $\mathbf{V}_p$, λ_p , $p = 1, \dots, n$ ($\text{argmin}_{\Theta} C(Q, \Theta)$)
- Berechne alle Modellinstanzen $\mathbf{M}_k$ durch Optimierung von $C(Q|\Theta)$ nach jeweils T_k , ω_k , $k = 1, \dots, N$ ($\text{argmin}_Q C(Q, \Theta)$)

Die Parameter Σ und Σ_N werden dabei nicht im M-Schritt optimiert. Denn $C(Q, \Theta)$ würde z. B. minimal werden bei maximalen (unendlich großen) Merkmalsvarianzen in Σ , das heißt, wenn “alle Merkmalsvektoren der Trainingsdaten zu allen Merkmalsvektoren des Modells” korrespondieren würden. In Abschnitt 4.2 wird die Bestimmung von Σ und Σ_N diskutiert.

Da im M-Schritt eine große Menge an Parametern optimiert werden muss und sich mit jeder Parameteränderung auch die aktuellen Korrespondenzwahrscheinlichkeiten ändern, ist es sinnvoll, nach jedem “partiellen” M-Schritt einen E-Schritt durchzuführen, um die Optimierung zu beschleunigen, $\forall k$:

E-Schritt: Berechne $\mathbb{E}(H_{ij})_k$

part. M-Schritt: Optimierte $\bar{\mathbf{M}}$

E-Schritt: Berechne $\mathbb{E}(H_{ij})_k$

part. M-Schritt: Optimierte $\mathbf{V}_p, \lambda_p, p = 1, \dots, n$

E-Schritt: Berechne $\mathbb{E}(H_{ij})_k$

part. M-Schritt: Optimierte T_k

E-Schritt: Berechne $\mathbb{E}(H_{ij})_k$

part. M-Schritt: Optimierte ω_k

E-Schritt: Die Berechnung der Korrespondenzwahrscheinlichkeiten für den E-Schritt wurde im Abschnitt 2.3 vorgestellt. Die Normierung der Erwartungswerte wird mit den Gleichungen 10 und 12 durchgeführt. Dies bedeutet, dass während der Modellgenerierung angenommen wird, dass die Merkmalsvektormengen der Trainingsdaten $\mathbf{S}_k$ keine Outlier beinhalten (Gleichung 10). Gleichzeitig soll das Modell keine $\mathbf{M}_k$ erzeugen, deren einzelne Vektoren gar nicht zum jeweiligen $\mathbf{S}_k$ korrespondieren (Gleichung 12). Während der Modellgenerierung ist dies zu empfehlen, damit alle Merkmalsvektoren für die Modellparameteroptimierung genutzt werden.

Die Optimierung aller Parameter in dem jeweiligen partiellen M-Schritt wird im Folgenden beschrieben.

4.1.2. Parameteroptimierung mithilfe partieller Ableitungen

Eine Methode, das Minimum einer Funktion zu bestimmen, ist die Nullstelle der ersten Ableitung zu finden. Das heißt, im Folgenden werden die partiellen Ableitungen des Kriteriums $C(Q, \Theta)$ nach allen Parametern (außer Σ und Σ_N) Null gleichgesetzt und anschließend zum jeweiligen Parameter umgestellt. Die ausführlichen Ableitungen und Umstellungen sind im Appendix der Arbeit zu finden.

Die partielle Ableitung von $C(Q, \Theta)$ nach einem beliebigen Parameter x ist gegeben durch:

$$\begin{aligned} & \frac{\partial}{\partial x} C(Q, \Theta) \\ &= \sum_{k=1}^N \left(\sum_{p=1}^n \frac{\partial}{\partial x} \left(\log(\lambda_p) + \left(\frac{\omega_{k,p}^2}{2\lambda_p} \right) \right) \right. \\ & \quad \left. + \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \underbrace{\mathbb{E}(H_{ij})_k}_{\text{fix im M-Schritt}} \Sigma^{-1}(\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \frac{\partial}{\partial x} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \right) \end{aligned} \quad (38)$$

$$\begin{aligned}
& - \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n - \left(\Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_\eta \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \right. \\
& \left. \frac{\partial}{\partial x} \left(\frac{(\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_\eta \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right) \\
& - \left(\frac{\frac{1}{N} \sum_{k=1}^N \boldsymbol{\omega}_k}{\sigma_{\bar{\omega}}^2} \right)^T \frac{\partial}{\partial x} \left(\frac{1}{N} \sum_{k=1}^N \boldsymbol{\omega}_k \right) \\
& - \left(\frac{(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}}{\sigma_{\bar{T}}^2} \right) \frac{\partial}{\partial x} \left(\frac{1}{N} \sum_{k=1}^N T_k - I_{D+1} \right)
\end{aligned}$$

An dieser Stelle sei darauf hingewiesen, dass im Kriterium in der Likelihood $(-\log p(\mathbf{S}_{\text{tr}} | Q, \Theta) = -\sum_{k=1}^N \sum_{i=1}^{N_s(k)} \log \left(\sum_{j=1}^{N_m} p(\mathbf{s}_{k,i} | \mathbf{m}_{k,j}) \right))$ über die Wahrscheinlichkeit $p(\mathbf{s}_{k,i} | \mathbf{m}_{k,j})$ summiert wird, welche in der Ableitung zum Erwartungswert $\mathbb{E}(H_{ij})_k = p(\mathbf{s}_{k,i} | \mathbf{m}_{k,j}) / \sum_{\iota=1}^{N_m} p(\mathbf{s}_{k,i} | \mathbf{m}_{k,\iota})$ wird. Des Weiteren sei angemerkt, dass $\mathbf{m}_{k,j}$ mit $\mathbf{m}_{k,j} = T_k^{-1} \star (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j})$ alle gesuchten Parameter “beinhaltet” und somit bei der Umstellung der Ableitung der Erwartungswert $\mathbb{E}(H_{ij})_k$ bei der Optimierung aller Parameter eine Rolle spielt.

4.1.3. Optimierung des mittleren Modells

Für die Optimierung des mittleren Modells kann die auf Null gesetzte Ableitung für jeden Vektor j geschlossen umgestellt werden zu

$$\begin{aligned}
& \text{“Mitte” der transformierten und variationsneutralen Trainingsdaten} \\
& \sum_{k=1}^N \left(\sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}) \right) \\
\bar{\mathbf{m}}_j = & \frac{\sum_{k=1}^N \sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \Sigma^{-1} + \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n \Sigma_{\mathcal{N}}^{-1} \left(\frac{1}{\|\mathcal{N}_{j,\eta}^0\|^2} \right)}{\sum_{k=1}^N \sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \Sigma^{-1} + \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n \Sigma_{\mathcal{N}}^{-1} \left(\frac{1}{\|\mathcal{N}_{j,\eta}^0\|^2} \right)} \quad (39) \\
& \text{“Mittelung” über die Nachbarn} \\
& \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n \Sigma_{\mathcal{N}}^{-1} \left(\frac{(-(\pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) + (\bar{\mathbf{m}}_\eta \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) + \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|^2} \right) \\
& + \frac{\sum_{k=1}^N \sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \Sigma^{-1} + \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n \Sigma_{\mathcal{N}}^{-1} \left(\frac{1}{\|\mathcal{N}_{j,\eta}^0\|^2} \right)}{\sum_{k=1}^N \sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \Sigma^{-1} + \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n \Sigma_{\mathcal{N}}^{-1} \left(\frac{1}{\|\mathcal{N}_{j,\eta}^0\|^2} \right)}
\end{aligned}$$

Der erste Teil der Gleichung bewirkt eine Mittelung über die “Reste” der Trainingsdaten, die nicht durch die aktuelle affine Transformation oder die aktuellen gewichteten Variationsmoden einander angeglichen wurden. Es kann also interpretiert werden als: das mittlere Modell liegt in der “Mitte” der transformierten (T_k) und variationsneutralen ($-\sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}$) Trainingsdaten. Die zweite Hälfte der Gleichung regularisiert das

mittlere Modell über die jeweiligen Nachbarn jedes Merkmalsvektors j . $\bar{\mathbf{m}}_j$ liegt demnach in der Mitte aller n Variationen von Vektor j (beschränkt durch $\pm 2 \cdot \lambda_p$) und außerdem in der Mitte der Variationen aller Nachbarvektoren $\eta \in \mathcal{N}(j)$. Der Divisor sorgt jeweils für die Normierung der Werte.

4.1.4. Optimierung der Variationsmoden

Die auf Null gesetzte Ableitung und anschließende Umstellung des Kriteriums nach $\mathbf{v}_{p,j}$ ergibt:

$$\frac{\partial}{\partial \mathbf{v}_{p,j}} C(Q, \Theta) = -\mathbf{a}_{p,j} + \sum_{q=1}^n (\mathbf{b}_{p,q,j} \Sigma^{-1} \mathbf{v}_{q,j})^T + \sum_{q=1}^n (\mathbf{b}'_{p,q,j} \Sigma_{\mathcal{N}}^{-1} \mathbf{v}_{q,j})^T \quad (40)$$

$$\mathbf{a}_{p,j} = (a_{p,j}^1, \dots, a_{p,j}^D) \quad (41)$$

“Fehler” zwischen mittlerem Modell und Trainingsdatensatz

$$= \sum_{k=1}^N \sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - \bar{\mathbf{m}}_j)^T \omega_{k,p}$$

Regularisierung über Nachbarmoden

$$+ \sum_{\eta \in \mathcal{N}(j)} \Sigma_{\mathcal{N}}^{-1} \left(\frac{(\bar{\mathbf{m}}_j - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \left(\frac{\pm 2 \cdot \lambda_p}{\|\mathcal{N}_{j,\eta}^0\|} \right)$$

$$\mathbf{b}_{p,q,j} = \sum_{k=1}^N \left(\sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \omega_{k,q} \omega_{k,p} \right)$$

$$\mathbf{b}'_{p,q,j} = \sum_{\eta \in \mathcal{N}(j)} \left(\frac{\pm 2 \cdot \lambda_p}{\|\mathcal{N}_{j,\eta}^0\|} \right)^2, \text{ for } p = q, \text{ else } b'_{p,q,j} = 0$$

Der erste Teil der Gleichung (in $\mathbf{a}_{p,j}$) ist zuständig für den Ausgleich des Fehlers zwischen Trainingsdaten und mittlerem Modell durch die gewichteten Variationsmoden. Die zweite Hälfte regularisiert die Modenwerte für einen Vektor j über die Werte seiner Nachbarn. Die Werte in $\mathbf{b}_{p,q,j}/\mathbf{b}'_{p,q,j}$ sorgen für die Normierung der Werte.

Da die Moden voneinander abhängig sind, müssen alle n Moden gleichzeitig optimiert werden. Die Ableitung nach allen Moden für einen Vektor j (bezeichnet als $\mathbf{V}_j$) kann als Matrixmultiplikation formuliert werden. Dabei wird $\mathbf{V}_j$ definiert als:

$$\mathbf{V}_j \in \mathbb{R}^{D \cdot n} = (v_{1,j}^1, \dots, v_{1,j}^D, \dots, v_{n,j}^1, \dots, v_{n,j}^D)^T \quad (42)$$

Durch die Umsortierung der Werte in $\mathbf{V}_j$ kann die Ableitung geschrieben werden als:

$$\frac{\partial}{\partial \mathbf{V}_j} C(Q, \Theta) = -\mathbf{A}_j + \mathbf{B}_j \mathbf{V}_j \quad (43)$$

mit

$$\mathbf{A}_j \in \mathbb{R}^{D \cdot n} = (a_{1,j}^1, \dots, a_{1,j}^D, \dots, a_{n,j}^1, \dots, a_{n,j}^D)^T \quad (44)$$

$$\mathbf{B}_j \in \mathbb{R}^{D \cdot n \times D \cdot n} = \begin{pmatrix} \mathbf{b}_{1,1,j} \Sigma^{-1} + \mathbf{b}'_{1,1,j} \Sigma_{\mathcal{N}}^{-1} & \dots & \mathbf{b}_{1,n,j} \Sigma^{-1} + \mathbf{b}'_{1,n,j} \Sigma_{\mathcal{N}}^{-1} \\ \dots & \mathbf{b}_{p,q,j} \Sigma^{-1} + \mathbf{b}'_{p,q,j} \Sigma_{\mathcal{N}}^{-1} & \dots \\ \mathbf{b}_{n,1,j} \Sigma^{-1} + \mathbf{b}'_{n,1,j} \Sigma_{\mathcal{N}}^{-1} & \dots & \mathbf{b}_{n,n,j} \Sigma^{-1} + \mathbf{b}'_{n,n,j} \Sigma_{\mathcal{N}}^{-1} \end{pmatrix}.$$

Nachdem $\frac{\partial}{\partial \mathbf{V}_j} C(Q, \Theta) = 0$ gesetzt wurde, ergibt sich $\mathbf{A}_j = \mathbf{B}_j \mathbf{V}_j$. Dies kann durch Lösung des linearen Gleichungssystems nach $\mathbf{V}_j$ aufgelöst werden.

Damit die Moden linear unabhängig voneinander sind, was unter Verwendung der PCA automatisch der Fall ist, muss die Nebenbedingung der Orthogonalität eingeführt werden:

$$\mathbf{V}^T \mathbf{V} = I_{n \times n}, \quad (45)$$

$$\text{mit } \mathbf{V} \in \mathbb{R}^{D \cdot N_m \times n}, \mathbf{V} = \begin{pmatrix} \mathbf{v}_{1,1} & \dots & \mathbf{v}_{n,1} \\ \dots & \mathbf{v}_{p,j} & \dots \\ \mathbf{v}_{1,N_m} & \dots & \mathbf{v}_{n,N_m} \end{pmatrix}, \mathbf{v}_{p,j} = (v_{p,j}^1, \dots, v_{p,j}^D), \text{ mit } I_{n \times n} \text{ als}$$

Einheitsmatrix der Größe $n \times n$.

Die Optimierung des Kriteriums nach $\mathbf{V}$ mit Nebenbedingung $\mathbf{V}^T \mathbf{V} = I_{n \times n}$ kann unter Nutzung der ‘‘Splitting Methode’’ für Probleme mit Orthogonalitätsnebenbedingung durchgeführt werden [Lai & Osher 2014]. Die Methode teilt ein Problem mit einer Nebenbedingung

$$\min_{\mathbf{V}} C(Q, \Theta) \quad \text{s.t.} \quad \mathbf{V}^T \mathbf{V} = I_{n \times n} \quad (46)$$

in zwei Nebenbedingungen auf:

$$\min_{\mathbf{V}} C(Q, \Theta) \quad \text{s.t.} \quad \mathbf{V} = \mathbf{P} \quad \text{s.t.} \quad \mathbf{P}^T \mathbf{P} = I_{n \times n}, \quad (47)$$

mit $\mathbf{P} \in \mathbb{R}^{D \cdot N_m \times n}$. Dabei wird die erste Nebenbedingung ($\mathbf{V} = \mathbf{P}$) mit dem Prinzip der Bregman-Iterationen gelöst. Hierbei wird zwischen zwei Schritten iteriert:

1. $\mathbf{V}^{it} = \operatorname{argmin}_{\mathbf{V}} C(Q, \Theta) + \frac{r}{2} \|\mathbf{V} - \mathbf{P} + \mathbf{N}^{it-1}\|_F^2$
2. $\mathbf{N}^{it} = \mathbf{N}^{it-1} + \mathbf{V}^{it} - \mathbf{P}$

r ist ein Regularisierungsfaktor und $\mathbf{N} \in \mathbb{R}^{D \cdot N_m \times n}$ stellt den aktuellen Fehler zwischen $\mathbf{V}$ und $\mathbf{P}$ dar mit $\mathbf{V} + \mathbf{N} = \mathbf{P}$. Das Prinzip ähnelt der ‘‘augmented Lagrangian Methode’’.

Um die Orthogonalitätsnebenbedingung einzuführen, wird die Matrix $\mathbf{P}$ mittels Singulärwert-Zerlegung (SVD) auf ihren orthogonalen Teil projiziert. Als $\mathbf{N}_j$ und $\mathbf{P}_j$ werden im Folgenden Untermengen von $\mathbf{N}$ und $\mathbf{P}$ bezeichnet – analog zu $\mathbf{V}_j$ und $\mathbf{V}$, siehe Gleichung 42. Die gesamte Optimierung der Moden unter Orthogonalitätsbedingung lässt sich zusammenfassen als:

Initialisiere $\mathbf{V}^0 = \mathbf{V}$, $\mathbf{P}^0 = \mathbf{V}^0$, $\mathbf{N}^0 = 0$

while nicht konvergiert **do**:

1. $\mathbf{V}^{it} = \operatorname{argmin}_{\mathbf{V}} (C(Q, \Theta) + \frac{r}{2} \|\mathbf{V} - \mathbf{P}^{it-1} + \mathbf{N}^{it-1}\|_F^2)$

- Neue Ableitung nach $\mathbf{V}_j$:

$$\frac{\partial}{\partial \mathbf{V}_j} = -\mathbf{A}_j + \mathbf{B}_j \mathbf{V}_j + r(\mathbf{V}_j - \mathbf{P}_j^{it-1} + \mathbf{N}_j^{it-1}) \quad (48)$$

- Löse folgendes Gleichungssystem nach $\mathbf{V}_j$ für alle j :

$$\mathbf{A}_j + r\mathbf{P}_j^{it-1} - r\mathbf{N}_j^{it-1} = (r + \mathbf{B}_j)\mathbf{V}_j \quad (49)$$

2. $\mathbf{P}_{temp}^{it} = \mathbf{V}^{it} + \mathbf{N}^{it-1}$; SVD auf $\mathbf{P}_{temp}^{it} = \mathbf{U}\mathbf{S}\mathbf{R}^T$

3. $\mathbf{P}^{it} = \mathbf{U}\mathbf{R}^T$

4. $\mathbf{N}^{it} = \mathbf{N}^{it-1} + \mathbf{V}^{it} - \mathbf{P}^{it}$

Der Wert von r regularisiert die erlaubte Größe des Unterschieds von $\mathbf{V}$ und der forciert orthogonalen Varianten $\mathbf{P}$ – d.h. wie stark $\mathbf{V}$ zu Orthogonalität gezwungen wird. Hier wurde $r = 500$ gewählt – siehe [Lai & Osher 2014] für die Diskussion von r .

4.1.5. Optimierung der Standardabweichung der Variationsmoden

Die auf Null gesetzte und umgestellte Ableitung des Kriteriums nach der “Standardabweichung” λ_p ergibt sich zum Polynom vierten Grades:

$$\frac{\partial}{\partial \lambda_p} C(Q, \Theta) = \lambda_p^0 \overbrace{\sum_{k=1}^N (-\omega_{k,p}^2) + \lambda_p^2 \sum_{k=1}^N (1)}^{\text{Mittelung über die Modengewichtungen}} \quad (50)$$

$$\left. \begin{aligned}
& + \lambda_p^3 \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \left(\Sigma_{\mathcal{N}}^{-1} \left(\frac{(\bar{\mathbf{m}}_j - \bar{\mathbf{m}}_\eta) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \right. \\
& \quad \left(\frac{((\pm 2 \cdot \mathbf{v}_{p,j}) - (\pm 2 \cdot \mathbf{v}_{p,\eta}))}{\|\mathcal{N}_{j,\eta}^0\|} \right) \\
& + \lambda_p^4 \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \left(\Sigma_{\mathcal{N}}^{-1} \left(\frac{((\pm 2 \cdot \mathbf{v}_{p,j}) - (\pm 2 \cdot \mathbf{v}_{p,\eta}))}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \right. \\
& \quad \left. \left(\frac{((\pm 2 \cdot \mathbf{v}_{p,j}) - (\pm 2 \cdot \mathbf{v}_{p,\eta}))}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right) .
\end{aligned} \right\} \begin{array}{l} \text{Beschränkung der erlaubten} \\ \text{Standardabweichung über Nachbarn} \end{array}$$

Der erste Teil (Polynom bis zum zweiten Grad) der Gleichung beschreibt die Standardabweichung der Moden über die Standardabweichung der Gewichte basierend auf einer Normalverteilung mit Mittelwert Null ($\lambda_p = \sqrt{1/N \sum_{k=1}^N \omega_{k,p}^2}$). Der Rest der Gleichung sorgt für die Beschränkung der erlaubten Standardabweichung mit der Regularisierung über die Modenwerte aller Modellvektoren und deren Nachbarn.

In der linearen Algebra werden die Eigenwerte einer Matrix manchmal als die Nullstellen sogenannter charakteristischer Polynome definiert. Wird die Definition “umgedreht”, können die Nullstellen eines Polynoms über die Eigenwertberechnung bestimmter “Companion” Matrizen berechnet werden. In “Numerical Recipes in C” [Press *et al.* 1992] wird in der Funktion `zrhqr(...)` die Nullstellenberechnung für Polynome über die Konstruktion der oberen Hessenberg-Matrix mit anschließender Eigenwertbestimmung für diese Matrix durchgeführt. Diese Methode kann angewendet werden, um die λ_p – die Nullstellen des oben definierten Polynoms – zu berechnen.

4.1.6. Optimierung der instanzabhängigen Modengewichte

Die Ableitung des Kriteriums nach den Modengewichten ω_k ist:

$$\begin{aligned}
\frac{\partial}{\partial \omega_{k,p}} C(Q, \Theta) &= \underbrace{\frac{\omega_{k,p}}{\lambda_p^2}}_{\text{Beschränkung von } \omega_{k,p} \text{ durch } \lambda_p} - \overbrace{\sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - \bar{\mathbf{m}}_j)^T \mathbf{v}_{p,j}}^{\text{Gewichtung der } \mathbf{V}, \text{ damit das Modell } \mathbf{S}_k \text{ beschreibt}} \\
&+ \underbrace{\sum_{q=1}^n \omega_{k,q} \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \mathbf{v}_{q,j}^T \Sigma^{-1} \mathbf{v}_{p,j}}_{\text{Abhängigkeit von } \omega_{k,p} \text{ von allen anderen } \omega_{k,q}} - \underbrace{\frac{\omega_{k,p}}{N^2 \sigma_{\bar{\omega}}^2} - \left(\frac{\sum_{l=1, l \neq k}^N \omega_{l,p}}{N^2 \sigma_{\bar{\omega}}^2} \right)}_{\text{Symmetriebedingung für } p(Q|\Theta)}
\end{aligned} \tag{51}$$

Der erste Anteil ($\frac{1}{\lambda_p}$) sorgt für die Beschränkung der Moden durch die aktuellen Standardabweichungen. Im nächsten Bereich der Gleichung werden die Gewichte so bestimmt,

dass das gewichtete Modell die gegebenen Trainingsdaten beschreibt und der dritte Abschnitt beinhaltet die Abhängigkeit einer Modengewichtung $\omega_{k,p}$ von allen anderen Modengewichtungen $\omega_{k,q}$, $q = 1, \dots, n$. Der letzte Bereich der Gleichung sorgt für die Symmetriebedingung. Der Mittelwert der Gewichte über die Trainingsdaten ($\frac{\sum_{l=1, l \neq k}^N \omega_{l,p}}{N^2 \sigma_{\bar{\omega}}^2}$) wird für die Optimierung von $\omega_{k,p}$ als fix angenommen. Wie in Abschnitt 3.2.4 beschrieben, ist diese Symmetriebedingung ausschließlich während der Modellgenerierung von Interesse. Bei der Modellanpassung an ein neues $\mathbf{S}_{new}$ sind die Gewichte ausschließlich durch λ_p beschränkt und nicht mehr durch die Gewichte der Trainingsdaten.

Da die Gewichte für alle Moden $p = 1, \dots, n$ abhängig voneinander sind, wird die Ableitung für den gesamten Vektor $\boldsymbol{\omega}_k$ als Matrixmultiplikation formuliert:

$$\frac{\partial}{\partial \omega_{k,p}} C(Q, \Theta) = \frac{\omega_{k,p}}{\lambda_p^2 - N^2 \sigma_{\bar{\omega}}^2} - d_{k,p} + \sum_{q=1}^n \omega_{k,q} c_{k,p,q} \quad (52)$$

$$\mathbf{C}_k \in \mathbb{R}^{n \times n} \quad \dots \quad c_{k,p,q} = \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \mathbf{v}_{q,j}^T \Sigma^{-1} \mathbf{v}_{p,j}$$

$$\mathbf{d}_k \in \mathbb{R}^{1 \times n} \quad \dots \quad d_{k,p} = \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - \bar{\mathbf{m}}_j)^T \mathbf{v}_{p,j} + \left(\frac{\sum_{l=1, l \neq k}^N \omega_{l,p}}{N^2 \sigma_{\bar{\omega}}^2} \right)$$

Das folgende lineare Gleichungssystem wird nach $\boldsymbol{\omega}_k$ gelöst:

$$\mathbf{d}_k = \left(\begin{pmatrix} \frac{1}{\lambda_1^2 - N^2 \sigma_{\bar{\omega}}^2} & & \\ & \ddots & \\ & & \frac{1}{\lambda_n^2 - N^2 \sigma_{\bar{\omega}}^2} \end{pmatrix} + \mathbf{C}_k \right) \boldsymbol{\omega}_k \quad (53)$$

Ein Mittelwert von Null für die Menge $\boldsymbol{\omega}_k$ des Trainingsdatensatz ist ausschließlich während der Modellgenerierung wichtig, weshalb die Symmetriebedingung während der Optimierung der Gewichte für die Modellanpassung an eine ungesehene Instanz $\mathbf{S}_{new}$ nicht genutzt wird:

$$\mathbf{d}_{new} \in \mathbb{R}^{1 \times n} \quad \dots \quad d_{new,p} = \sum_{i=1}^{N_s(new)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_{new} \Sigma^{-1} (T_{new} \star \mathbf{s}_{new,i} - \bar{\mathbf{m}}_j)^T \mathbf{v}_{p,j} \quad (54)$$

$$\mathbf{d}_{new} = \left(\begin{pmatrix} \frac{1}{\lambda_1^2} & & \\ & \ddots & \\ & & \frac{1}{\lambda_n^2} \end{pmatrix} + \mathbf{C}_{new} \right) \boldsymbol{\omega}_{new}.$$

4.1.7. Optimierung der instanzabhängigen affinen Transformation

Die Ableitung des Kriteriums nach T_k ergibt sich zu:

$$\begin{aligned} \frac{\partial}{\partial T_k} C(Q, \Theta) &= + \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} + \mathbf{v}_{p,j}))^T \quad (55) \\ &\quad \frac{\partial}{\partial T_k} (T_k \star \mathbf{s}_{k,i} - (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} + \mathbf{v}_{p,j})) \\ &\quad - \underbrace{\left(\frac{(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}}{\sigma_{\bar{T}}^2} \right) \frac{\partial}{\partial T_k} \left(\frac{1}{N} \sum_{k=1}^N T_k - I_{D+1} \right)}_{\text{Symmetriebedingung für } p(Q|\Theta)} \end{aligned}$$

Es wurden in Abschnitt 3.1.1 zwei verschiedene Varianten für T_k vorgestellt: **fixApp**: die Merkmalsvektoren werden mit fixen Appearance-Werten transformiert und **fixPos**: das Bild wird unter den fixen Positionen der Merkmalsvektoren transformiert. Für beide Varianten gelten unterschiedliche Ableitungen.

Variante fixApp – $T \star \mathbf{s}_i = (T(\mathbf{x}_i), \mathbf{f}_i)$: Die Transformation T wird in einen affinen Anteil $A \in \mathbb{R}^{D_x \times D_x}$ und eine Translation $t \in \mathbb{R}^{D_x}$ aufgeteilt, sodass gilt: $T \star \mathbf{s}_i = A\mathbf{s}_i + t = (A\mathbf{x}_i + t, \mathbf{f}_i)$. Die Optimierung der Transformation wird stabiler, wenn zuerst die Translation und anschließend der affine Anteil optimiert wird.

Da bei der **fixApp**-Transformation die Appearance-Werte der Merkmalsvektoren nicht geändert werden, beziehen sich die Variablen der folgenden Gleichungen lediglich auf den Koordinatenanteil der jeweiligen Vektoren $(\mathbf{s}_{k,i}, \bar{\mathbf{m}}_j, \mathbf{v}_{p,j})$. Um die Schreibweisen der Variablen beizubehalten, werden während der Optimierung von T die Anzahl der Appearance-Werte D_f auf Null gesetzt ($\mathbf{s}_{k,i}, \bar{\mathbf{m}}_j, \mathbf{v}_{p,j} \in \mathbb{R}^{D_x}$ und $\Sigma^{-1} \in \mathbb{R}^{D_x \times D_x}$). Der Einfluss der Appearance-Werte erfolgt über die Korrespondenzen $\mathbb{E}(H_{ij})$, die wie bisher unter Betrachtung aller Dimensionen (Positions- und Appearance-Werte) berechnet werden.

Die Ableitung erfolgt sowohl nach t als auch nach A :

$$t_k = \frac{-\Sigma^{-1}(A_k \sum_{i=1}^{N_s(k)} \mathbf{s}_{k,i}^T - \sum_{j=1}^{N_m} (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j})^T \sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k)}{\underbrace{\Sigma^{-1} N_s(k) - \frac{1}{N^2 \sigma_T^2}}_{\text{Symm-B für } p(Q|\Theta)}} \quad (56)$$

$$+ \underbrace{\left(\frac{\sum_{l=1, l \neq k}^N t_l}{N^2 \sigma_T^2} \right)}_{\text{Symm-B für } p(Q|\Theta)}$$

$$\dots \frac{1}{\underbrace{\Sigma^{-1} N_s(k) - \frac{1}{N^2 \sigma_T^2}}_{\text{Symm-B für } p(Q|\Theta)}}$$

$$A_k = F E^{-1} \quad (57)$$

$$E = \sum_{i=1}^{N_s(k)} \Sigma^{-1}(\mathbf{s}_{k,i} \mathbf{s}_{k,i}^T) - \underbrace{\left(\frac{1}{N^2 \sigma_T^2} \right)}_{\text{Symm-B für } p(Q|\Theta)} \quad (58)$$

$$F = \Sigma^{-1} \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \left((\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}) \mathbf{s}_{k,i}^T - t_k \mathbf{s}_{k,i}^T \right) + \underbrace{\left(\frac{\frac{1}{N} (\sum_{l=1, l \neq k}^N A_l) - I_D}{N \sigma_T^2} \right)}_{\text{Symm-B für } p(Q|\Theta)} \quad (59)$$

Da die umgestellten Ableitungen beider Parameter (t , A) vom jeweils anderen Parameter abhängig sind, erfolgt die Optimierung des Kriteriums nach der Transformation T_k iterativ, wobei zwischen der Berechnung der Translation t_k und der Berechnung des affinen Anteils A_k alterniert wird. Die Optimierung der Transformation wird stabiler, wird zuerst die Translation angepasst, da diese in den meisten Fällen den größeren Anteil an der Transformation belegt. Daher kann, wenn erwünscht, zuerst die Optimierung der Translation bis zur Konvergenz wiederholt werden. Um die Optimierung zu beschleunigen, werden zwischen den Schritten die probabilistischen Korrespondenzen upgedatet:

1. Update $\mathbb{E}(H_{ij})_k$ zwischen $T_k(\mathbf{S}_k)$ und $\bar{\mathbf{M}} + \omega_k \mathbf{V}$
(Anmerkung: das entspricht $\mathbb{E}(H_{ij})_k$ zwischen $\mathbf{S}_k$ und $\mathbf{M}_k$)
2. Optimierte t_k (zwischen $\mathbf{S}_k$ und $\bar{\mathbf{M}} + \omega_k \mathbf{V}$)

- (3. Gehe zu 1. bis Konvergenz)
3. Update $\mathbb{E}(H_{ij})_k$ zwischen $T_k(\mathbf{S}_k)$ und $\bar{\mathbf{M}} + \omega_k \mathbf{V}$
4. Optimierte A_k (zwischen $\mathbf{S}_k$ und $\bar{\mathbf{M}} + \omega_k \mathbf{V}$)
5. Gehe zu 1. bis Konvergenz

Variante fixPos – $T \star \mathbf{s}_i = (\mathbf{x}_i, \mathbf{F}(T^{-1}(\mathbf{x}_i)))$: Da bei der fixPos-Transformationsvariante die Appearance-Bilder $\mathbf{F} = [I^{f=0}, \dots, I^{f=D_f}]$ transformiert werden, enthält die Ableitung ebenfalls den Gradienten der Bilder $\nabla \mathbf{F}(T^{-1}(\mathbf{x}_{k,i}))$:

$$\begin{aligned} \frac{\partial}{\partial T_k} C(Q, \Theta) &= \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}))^T \\ &\quad \nabla \mathbf{F}_k(T_k^{-1}(\mathbf{x}_{k,i})) \frac{\partial}{\partial T_k} (T_k^{-1}(\mathbf{x}_{k,i})) \\ &\quad - \underbrace{\left(\frac{(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}}{\sigma_T^2} \right) \frac{\partial}{\partial T_k} \left((\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1} \right)}_{\text{Symmetriebedingung für } p(Q|\Theta)} \end{aligned} \quad (60)$$

Die Optimierung dieser Ableitung kann mittels Newton-Verfahren erfolgen. Da jedoch der Gradient des Bildes (bzw. aller Appearance-Merkmalbilder) in jeder Iteration des Newton-Verfahrens berechnet werden muss, ist diese Optimierung wesentlich langsamer als die Optimierung der fixApp-Transformationsvariante, mit welcher A und t geschlossen berechnet werden können. Mit einem Zwischenschritt kann die Berechnung aus der fixApp-Variante genutzt werden, um eine Transformation nach fixPos-Variante zu optimieren, da in beiden Varianten dieselben Werte in T gesucht sind. Eine optimale Transformation zwischen $\mathbf{S}$ und $\mathbf{M}$ mit $\mathbf{s}_i$ und $\mathbf{m}_j$ als Merkmalsvektoren wäre nach Variante fixPos:

$$T \star \mathbf{s}_i - \mathbf{m}_j \rightarrow_T \min \quad (61)$$

$$(\mathbf{x}_i, \mathbf{F}(T^{-1}(\mathbf{x}_i))) - (\mathbf{x}_j, \mathbf{F}(\mathbf{x}_j)) \rightarrow_T \min, \quad (62)$$

welches dasselbe ergibt wie:

$$(T(T^{-1}(\mathbf{x}_i)), \mathbf{F}(T^{-1}(\mathbf{x}_i))) - (\mathbf{x}_j, \mathbf{F}(\mathbf{x}_j)) \rightarrow_T \min. \quad (63)$$

Durch eine Substitution $\mathbf{x}_i' = T^{-1}(\mathbf{x}_i)$ ergibt sich:

$$(T(\mathbf{x}_i'), \mathbf{F}(\mathbf{x}_i')) - (\mathbf{x}_j, \mathbf{F}(\mathbf{x}_j)) \rightarrow_T \min. \quad (64)$$

Dies entspricht der fixApp-Transformationsvariante und kann somit durch die obere

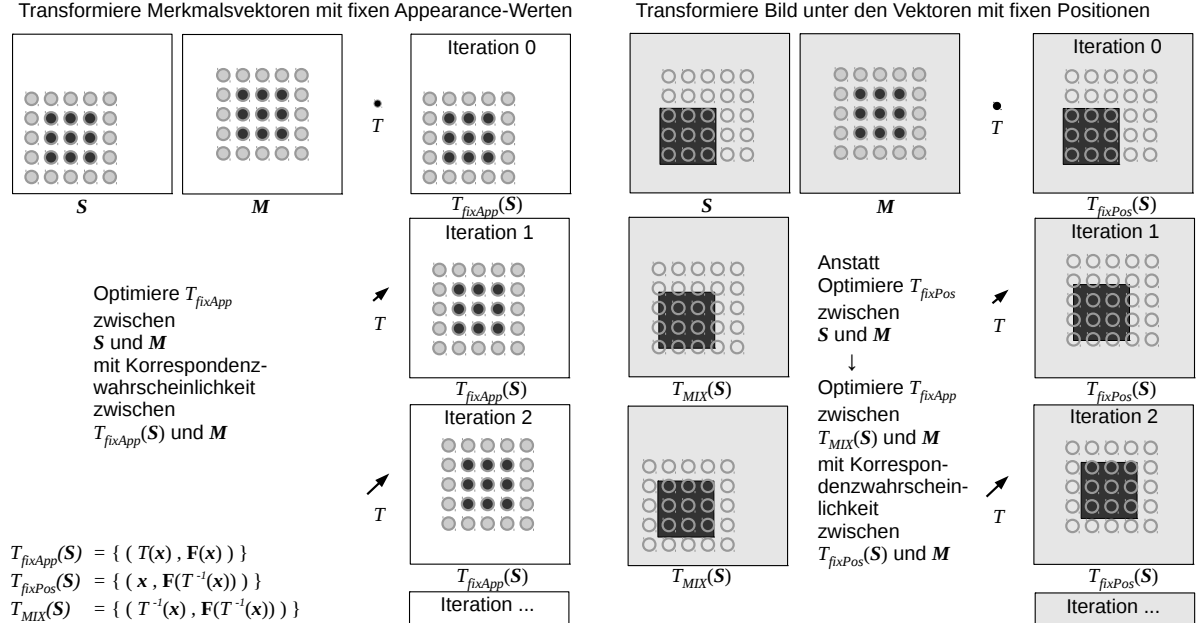


Abbildung 16: Optimierung der fixPos-Transformationsvariante über die Methode der fixApp-Variante: Werden die Positionen der Merkmalsvektoren mit fixen Appearance-Werten transformiert (links, Variante fixApp) können die Transformationsparameter geschlossen optimiert werden. Um die Transformation für die fixPos-Variante mit fixen Positionen und transformiertem Bild (rechts) ebenfalls mit geschlossener Parameterberechnung durchzuführen, wird ein Zwischenschritt eingeführt: die aktuell geschlossen berechnete Transformation (nach Variante fixApp) wird invers auf die Positionen angewendet ($T_{MIX}(S)$). Dies entspricht der gesuchten Verschiebung des Bildes unter den Merkmalsvektoren.

Optimierung (direkte Berechnung von t und A) gelöst werden. Abbildung 16 beschreibt die Optimierung über S' , die wie im vorangegangenen Paragraphen iterativ erfolgt:

1. Update $\mathbb{E}(H_{ij})_k$ zwischen $T_{k,fixPos}(S_k)$ und $\bar{M} + \omega_k V$
2. Substituiere S' mit $s'_i = (T^{-1}(x_i), F(T^{-1}(x_i)))$
3. Optimiere $t_{k,fixApp}$ zwischen S' und $\bar{M} + \omega_k V$
4. Update $\mathbb{E}(H_{ij})_k$ zwischen $T_{k,fixPos}(S_k)$ und $\bar{M} + \omega_k V$
5. Substituiere S' mit $s'_i = (T^{-1}(x_i), F(T^{-1}(x_i)))$
6. Optimiere $A_{k,fixApp}$ zwischen S' und $\bar{M} + \omega_k V$
7. Gehe zu 1. bis Konvergenz

Analog zur Optimierung der Gewichte ω_k , gilt die Symmetriebedingung, die in Abschnitt 3.2.4 vorgestellt wurde, ausschließlich während der Modellgenerierung. Für die

Adaption des generierten Modells an neue Bilddaten werden die mit geschweiften Klammern gekennzeichneten Abschnitte der vorangegangenen Gleichungen ignoriert.

Transformation zu $\bar{\mathbf{M}} + \omega_k \mathbf{V}$ oder Transformation zu $\bar{\mathbf{M}}$? Da eine Modellinstanz $\mathbf{M}_k$ definiert ist als $\mathbf{M}_k = T_k^{-1}(\bar{\mathbf{M}} + \omega_k \mathbf{V})$, wird folglich nach der Ableitung nach T_k die optimale Transformation zwischen $\mathbf{S}_k$ und $\bar{\mathbf{M}} + \omega_k \mathbf{V}$ gesucht. Für die Optimierung von ω_k gilt wiederum nach Abschnitt 4.1.6, dass die Gewichte gesucht werden, die den Abstand zwischen $\mathbf{S}_k$ und $T_k^{-1}(\bar{\mathbf{M}})$ minimieren. Das heißt, beide instanzabhängigen Parameter T_k und ω_k haben dieselbe “Priorität” und mit den vorliegenden partiellen Ableitungen nach beiden Parametern wird das Kriterium lediglich nach der optimalen Kombination aus beiden Parametern minimiert. Dies kann allerdings zu einem Overfitting des Modells an diese Kombination führen. Es wird durch die Optimierung nicht ausgeschlossen, dass eine Anwendung der so bestimmten Transformation auf das mittlere Modell (mit $\omega_{k,p} = 0$) zu einer Verschlechterung der Distanz von Modell zu $\mathbf{S}_k$ führt, d. h. $\text{dist}(\mathbf{S}_k, \bar{\mathbf{M}}) > \text{dist}(\mathbf{S}_k, \mathbf{M}_k)$ aber $\text{dist}(\mathbf{S}_k, \bar{\mathbf{M}}) < \text{dist}(\mathbf{S}_k, T_k^{-1}(\bar{\mathbf{M}}))$. Dies kann zwar zu optimalen Endergebnissen führen, jedoch werden die Modellgenerierung und auch Modellanpassung anfälliger für lokale Minima. Außerdem ist dadurch nicht auszuschließen, dass die generierten gewichteten Moden eine affine Transformation beschreiben.

Um dies zu verhindern und $\text{dist}(\mathbf{S}_k, \bar{\mathbf{M}}) > \text{dist}(\mathbf{S}_k, T_k^{-1}(\bar{\mathbf{M}})) > \text{dist}(\mathbf{S}_k, \mathbf{M}_k)$ zu gewährleisten, muss lediglich die Optimierung der Transformation zwischen $\mathbf{S}_k$ und $\bar{\mathbf{M}}$ stattfinden.

$$t_k = \frac{-\Sigma^{-1}(A_k \sum_{i=1}^{N_s(k)} \mathbf{s}_{k,i}^T - \sum_{j=1}^{N_m} \bar{\mathbf{m}}_j^T \sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k) + \underbrace{\left(\frac{\sum_{l=1, l \neq k}^N t_l}{N^2 \sigma_{\bar{T}}^2} \right)}_{\text{Symm-B für } p(Q|\Theta)}}{\underbrace{\Sigma^{-1} N_s(k)}_{\text{Symm-B für } p(Q|\Theta)} - \underbrace{\frac{1}{N^2 \sigma_{\bar{T}}^2}}_{\text{Symm-B für } p(Q|\Theta)}} \quad (65)$$

$$A_k = F E^{-1} \quad (66)$$

$$E = \sum_{i=1}^{N_s(k)} \Sigma^{-1}(\mathbf{s}_{k,i} \mathbf{s}_{k,i}^T) - \underbrace{\left(\frac{1}{N^2 \sigma_{\bar{T}}^2} \right)}_{\text{Symm-B für } p(Q|\Theta)}$$

$$F = \underbrace{\Sigma^{-1} \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k (\bar{\mathbf{m}}_j \mathbf{s}_{k,i}^T - t_k \mathbf{s}_{k,i}^T)}_{\text{Symm-B für } p(Q|\Theta)} + \underbrace{\left(\frac{\frac{1}{N} (\sum_{l=1, l \neq k}^N A_l) - I_D}{N \sigma_{\bar{T}}^2} \right)}_{\text{Symm-B für } p(Q|\Theta)}$$

Dies kann so interpretiert werden, dass “zuerst” die Transformation und anschließend die Gewichte der Moden optimiert werden. Daraus ergibt sich ebenfalls, dass die generierten Moden ausschließlich Form- (und Appearance-) Änderungen beschreiben, die nicht durch die affine Transformation ausgeglichen werden können.

Beispielbox 6: Optimierung der Transformation

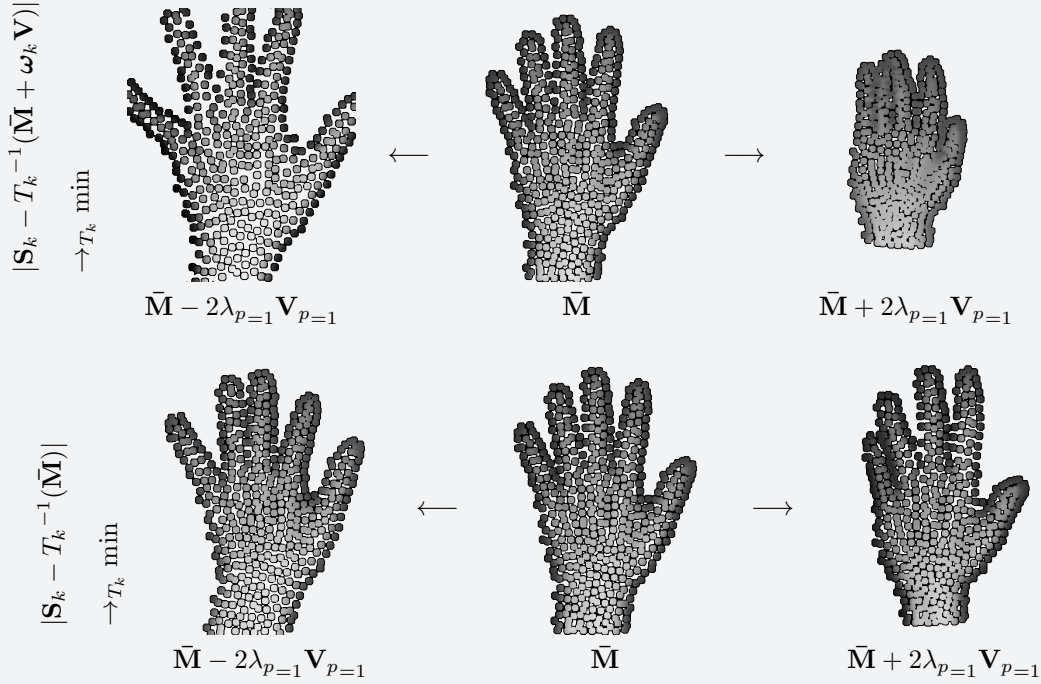


Abbildung 17: Gegenüberstellung von Modellen mit Optimierung der Transformation zu $\bar{\mathbf{M}} + \omega_k \mathbf{V}$ (oberes Modell) bzw. zu $\bar{\mathbf{M}}$ (unteres Modell).

Um den Einfluss der Optimierung der Transformation zu $\bar{\mathbf{M}} + \omega_k \mathbf{V}$ bzw. $\bar{\mathbf{M}}$ zu verdeutlichen, wurden zwei Modelle mit beiden Optimierungsvarianten generiert. Dabei wird deutlich, dass z. B. die erste Mode des Modells mit Transformation zu $\bar{\mathbf{M}} + \omega_k \mathbf{V}$ eine Skalierung der Merkmalsvektormenge erhält, die idealerweise durch die affine Transformation und nicht durch die Moden ausgeglichen werden sollte.

Eine quantitative Gegenüberstellung zeigt darüber hinaus, dass eine Transformationsoptimierung zu $\bar{\mathbf{M}}$ stabilere Ergebnisse liefert. In den oberen Boxplots sind die Handkonturdistanzen zwischen $\mathbf{S}_k$ und $\bar{\mathbf{M}}/T_k^{-1}(\bar{\mathbf{M}})/\mathbf{M}_k$ in *mm* für beide Modelle aufgetragen. Die Berechnung der Distanz wird in Abschnitt 8.3 beschrieben.

Für die Transformationsoptimierung zu $\bar{\mathbf{M}}$ gilt für das Beispiel Modell $\text{dist}(\mathbf{S}_k, \bar{\mathbf{M}}) > \text{dist}(\mathbf{S}_k, T_k^{-1}(\bar{\mathbf{M}})) > \text{dist}(\mathbf{S}_k, \mathbf{M}_k)$, bei dem zweiten Modell hingegen kommt es zu

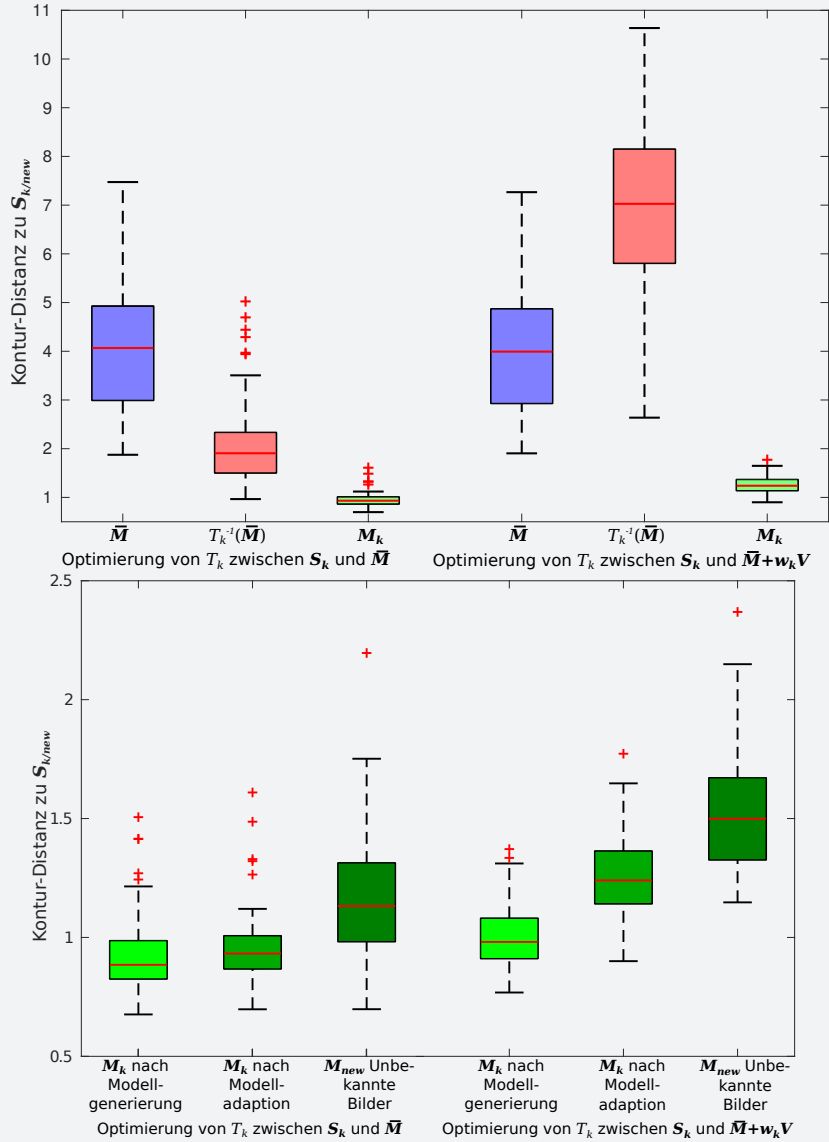


Abbildung 18: Quantitative Gegenüberstellung von Modellen mit Optimierung der Transformation zu $\bar{M} + \omega_k V$ bzw. zu $\bar{M}$.

einer deutlichen Verschlechterung der Distanzen, werden nur die optimierten Transformationen angewendet. Eine Verbesserung der Distanzen ergibt sich erst mit der Anwendung beider optimierter Parameter (T_k und ω_k).

Die unteren sechs Boxplots vergleichen die Endergebnisse beider Modelle: Die jeweils ersten Werte sind dabei die Distanzen zwischen S_k und M_k nach abgeschlossener Modellgenerierung. Das heißt, um das Modell zu generieren, wurden Modellparameter und instanzabhängige Parameter und dementsprechend die Modellinstanzen

$\mathbf{M}_k$ abwechselnd optimiert. “ $\mathbf{M}_k$ nach Modellgenerierung” zeigt also die Distanzen zu diesen resultierenden Modellinstanzen. Wird das generierte Modell anschließend erneut auf die Trainingsdaten $\mathbf{S}_k$ angepasst (ausschließliche Optimierung der instanzabhängigen Parameter), kann es dazu kommen, dass ein anderes lokales Minimum gefunden wird und daher die Werte für “ $\mathbf{M}_k$ nach Modellanpassung” abweichen. Die jeweils dritten Werte visualisieren die Distanzen für die Anpassung des Modells auf 60 unbekannte Bilder ($\mathbf{M}_{new}$). Die drei linken Verteilungen zeigen erneut das stabilere Verhalten: Die erneute Modellanpassung auf die bekannten Bilder des Trainingsdatensatzes zeigt ähnliche Werte wie direkt nach der Modellgenerierung, das heißt, es werden ähnliche instanzabhängige Parameter und ähnliche probabilistische Korrespondenzen gefunden (ähnliches lokales Minimum). Bei dem rechten Modell hingegen wird ersichtlich, dass die Optimierung nicht stabil genug ist, die instanzabhängigen Parameter zu finden, da die erneute Modellanpassung sowohl an bekannte ($\mathbf{M}_k$) also auch an unbekannte ($\mathbf{M}_{new}$) Bilder wesentlich schlechter ist.

Parameterbelegung

Input-/Trainingsdaten

N	=	60
D	=	4 ($D_f = 2, D_x = 2$)
$\mathbf{f}_i$	:	(1) Gauß-geglättete Signalwerte: $(G_{\sigma=6mm} * I_k)(\mathbf{x}_i)$ (2) Distanzwerte zur Handmaske: $d(\mathbf{x}_i, I_k^{Mask})$
$\mathbf{x}_i$	:	kantenmerkmalsabhängig mit Mindestabstand 6 mm, innerhalb der Hand ($\forall i f_{i,2} < 5mm$)
sp	=	6 mm
$N_s(k)$	=	450 ± 57
N_m	=	481

Modellparameter

n	=	15
$T_k * \mathbf{s}_{k,i}$	=	$(T_k(\mathbf{x}_i), \mathbf{f}_i), (\mathbf{S}_k - T_k^{-1}(\bar{\mathbf{M}} + \boldsymbol{\omega}_k \mathbf{V})) \rightarrow_{T_k} \min$ vs. $(\mathbf{S}_k - T_k^{-1}(\bar{\mathbf{M}})) \rightarrow_{T_k} \min$
Beschränk.	-	mit $p(\Theta)$; - $p(Q \Theta)$ mit Symmetriebedingung

4.2. Regularisierungsparameter

Die Optimierung der Modell- und instanzabhängigen Parameter wurde in Abschnitt 4.1 durch die Minimierung des globalen Kriteriums vorgestellt. Vier Parameter wurden dabei noch nicht betrachtet:

- Σ – Merkmalsvarianzen zwischen $\mathbf{S}_k$ und $\mathbf{M}_k$, die bestimmen, welche Merkmalsvektoren wie stark miteinander korrespondieren (Gleichung 5). Siehe Abschnitt 4.2.1 für die Berechnung.
- Σ_N – Varianzen der Nachbarschaftsbeziehungen für $p(\Theta)$ (Gleichung 31). Siehe Abschnitt 4.2.2 für die Berechnung.
- $\sigma_{\bar{\omega}}, \sigma_{\bar{T}}$ – Varianzen für die Abweichung der Mittelwerte der Modengewichte bzw. der Transformationen des gesamten Trainingsdatensatzes zu Null (bzw. zur Einheitsmatrix) (Abschnitt 3.2.4). Da die Mittelwerte Null für $\omega_{k,p}$ bzw. die Einheitsmatrix für T_k sein sollen, können diese beiden Werte klein ($\ll 0.001$) gewählt

werden. Die Werte bleiben während der gesamten Modellgenerierung fix. Für die Modellanpassung werden sie nicht genutzt, da die Symmetriebedingung dort nicht gilt.

Die vier Werte werden hier als “Regularisierungsparameter” bezeichnet, weil sie den Optimierungsprozess “regularisieren”: z. B. wie weit Merkmalsvektoren auseinanderliegen dürfen um zu korrespondieren (Σ) oder wie stark die Positionen der Merkmalsvektoren durch die Moden lokal variieren dürfen (Σ_N). Diese Werte müssen daher für die vorliegenden Daten bzw. für die gegebene Fragestellung, die durch das Modell beantwortet werden soll, angepasst werden. Die Bestimmung von Σ und Σ_N werden im Folgenden beschrieben.

4.2.1. Datengetriebene Anpassung der Merkmalsvarianzen Σ

Die Merkmalsvarianzen $\Sigma \in \mathbb{R}^{(D_x+D_f) \times (D_x+D_f)} = \mathbb{R}^{D \times D}$ bestimmen in $p(\mathbf{s}_{k,i} | \mathbf{m}_{k,j})$ (und somit in $\mathbb{E}(H_{ij})_k$), welche Merkmalsvektoren als wie stark korrespondierend abhängig von ihrem Abstand zueinander angenommen werden (Gleichung 5). Daher kann man Σ als wichtigsten Parameter des gesamten Modells betrachten, da die Qualität der optimalen Modell- und instanzabhängigen Parameter davon abhängt, wie gut korrespondierende Bereiche zwischen den Trainingsdaten bzw. zwischen entsprechender Trainings- und Modellinstanz gefunden werden können. Werden die Varianzen in Σ zu groß gewählt, korrespondieren zu viele Merkmalsvektoren miteinander und während der Modellgenerierung werden die Distanzen zwischen Trainingsdatensatz und Modell nicht kleiner mit den Iterationen, da die Merkmalsvektoren ja bereits als “gut passend” betrachtet werden, obwohl sie weit auseinanderliegen. Sind die Werte in Σ zu klein gewählt, werden nur die Vektorpaare, die bereits sehr nahe beieinanderliegen, als potentiell korrespondierend betrachtet. Das bedeutet, eigentlich korrespondierende Bereiche im Trainingsdatensatz, die weit auseinanderliegen, werden schlecht angepasst bzw. durch das Modell nicht korrekt modelliert. Die optimalen Werte für Σ sind zu Beginn der Optimierung groß – um potentiell weit auseinanderliegende Bereiche aneinander anzupassen – und werden kleiner mit fortschreitender Optimierung – um feinere Unterschiede zwischen den Bildrepräsentationen auszugleichen. In [Hufnagel 2010] wurden daher die Varianzen pro Iteration durch einen festen Faktor verkleinert. Als Startwert wurde ebenfalls ein fester Wert (z. B. $12mm$) gewählt. Granger führt eine Schätzung von Σ auf Grundlage der initialen Korrespondenzwahrscheinlichkeiten und der initialen Distanzen zwischen $\mathbf{S}_k$ und $\mathbf{M}_k$ ein [Granger & Pennec 2002]. Diese werden als untere Grenze für die ebenfalls pro Iteration mit festem Faktor verkleinerten Werte in Σ genutzt. Da diese Methoden von der Wahl der Startwerte für Σ und vom Faktor für das Dekrementieren abhängen, wird in dieser Arbeit eine rein datengetriebene Bestimmung für die Merkmalsvarianzen genutzt.

An dieser Stelle wird angenommen, dass die Positions- und Appearance-Merkmale unabhängig voneinander sind, sodass ausschließlich die Varianzen berechnet werden:

$$\Sigma \in \mathbb{R}^{D \times D} = \text{diag}(\sigma_{x^1}^2, \dots, \sigma_{x^{D_x}}^2, \sigma_{f^1}^2, \dots, \sigma_{f^{D_f}}^2) = \frac{1}{N} \sum_{k=1}^N \text{diag}(\text{Cov}(\mathbf{S}_k, \mathbf{M}_k)). \quad (67)$$

Um die Werte von Σ an die aktuellen Daten anzupassen und manuelles Parameter-Tuning zu verhindern, werden die Varianzen in jeder Iteration zwischen Trainingsdatensatz und Modell berechnet:

$$\text{Cov}(\mathbf{S}_k, \mathbf{M}_k) = \frac{1}{N} \sum_{k=1}^N \frac{\sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})}{\sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k}. \quad (68)$$

Dabei ist deutlich geworden, dass $\text{Cov}(\mathbf{S}_k, \mathbf{M}_k)$ zu schnell zu klein wird, da sich die Differenzen zwischen $\mathbf{S}_k$ und $\mathbf{M}_k$ idealerweise in jeder Iteration verringern. Eine zu schnelle Verringerung der Varianzwerte bewirkt, dass die Modellgenerierung nur noch die “perfekt” korrespondierenden Merkmalsvektorkpaare in Betracht zieht. Eine bessere Wahl für Σ ist die Varianzberechnung zwischen Trainingsdatensatz und mittlerem Modell:

$$\text{Cov}(\mathbf{S}_k, \bar{\mathbf{M}}) = \frac{1}{N} \sum_{k=1}^N \frac{\sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k (\mathbf{s}_{k,i} - \bar{\mathbf{m}}_j)^T (\mathbf{s}_{k,i} - \bar{\mathbf{m}}_j)}{\sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k}, \quad (69)$$

oder eine Kombination aus beiden:

$$\Sigma = \frac{1}{N} \sum_{k=1}^N \text{diag}(\text{Cov}(\mathbf{S}_k, \mathbf{M}_k) \cdot 0.5 + \text{Cov}(\mathbf{S}_k, \bar{\mathbf{M}}) \cdot 0.5). \quad (70)$$

Die Einbeziehung der Varianzen zum $\bar{\mathbf{M}}$ führt dazu, dass alle Vektoren betrachtet werden, die innerhalb der Variationsbreite der Trainingsdaten ($\mathbf{S}_k$) liegen.

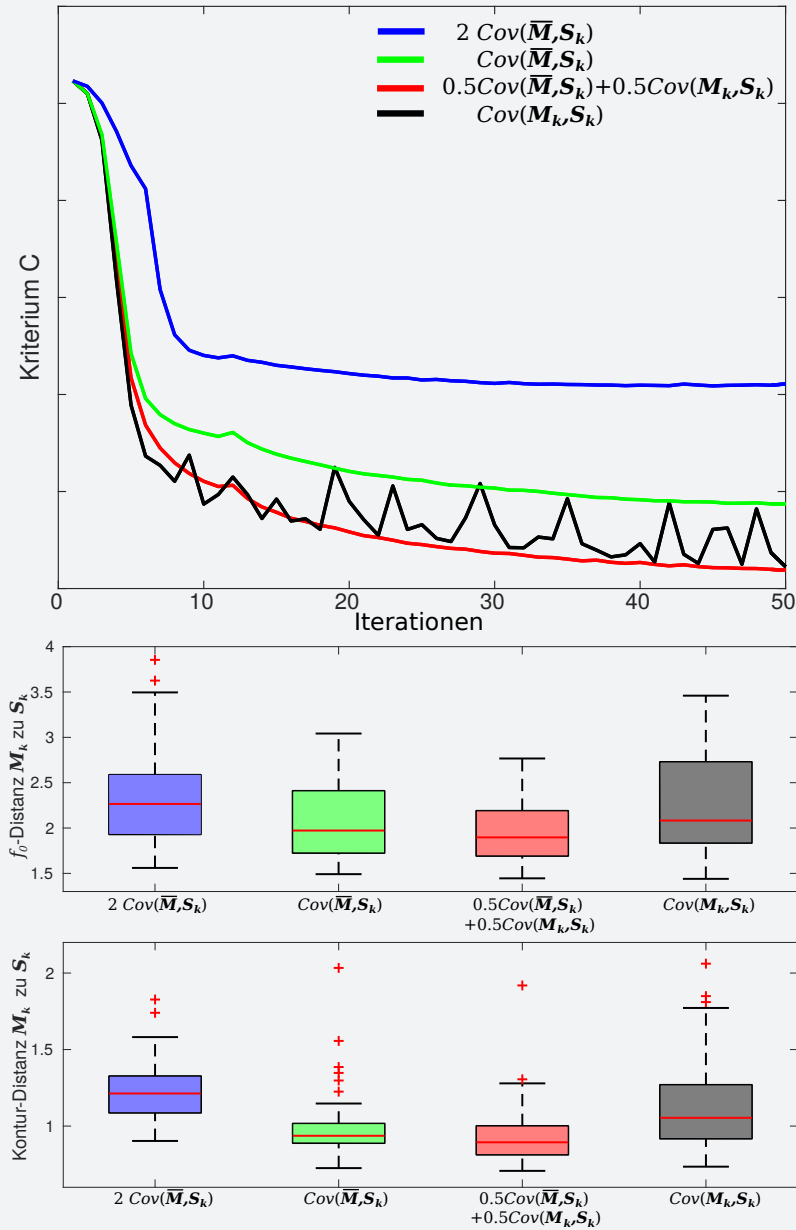
Beispielbox 7: Auswirkung der Merkmalsvarianzen Σ 

Abbildung 19: Quantitative Gegenüberstellung von vier Handmodellen mit unterschiedlichen Berechnungen der Merkmalsvarianzen Σ .

Für eine quantitative Auswertung wurden vier Handmodelle mit unterschiedlichen Werten für Σ generiert: das erste Modell (blau) mit $\Sigma = \frac{1}{N} \sum_{k=1}^N \text{diag}(2 \cdot \text{Cov}(\mathbf{S}_k, \bar{\mathbf{M}}))$, das zweite (grün) mit $\Sigma = \frac{1}{N} \sum_{k=1}^N \text{diag}(\text{Cov}(\mathbf{S}_k, \bar{\mathbf{M}}))$, das dritte (rot) mit $\Sigma = \frac{1}{N} \sum_{k=1}^N \text{diag}(\text{Cov}(\mathbf{S}_k, \mathbf{M}_k) \cdot 0.5 + \text{Cov}(\mathbf{S}_k, \bar{\mathbf{M}}) \cdot 0.5)$ und das vierte (schwarz) mit

$\Sigma = \frac{1}{N} \sum_{k=1}^N \text{diag}(\text{Cov}(\mathbf{S}_k, \mathbf{M}_k))$. Die obere Abbildung visualisiert die Entwicklung des globalen Kriteriums während der iterativen Optimierung der Modellparameter. Die unteren beiden Abbildungen stellen die Distanzen zwischen Trainingsdatensatz und generiertem Modell für beide Appearance-Merkmale dar (siehe Abschnitt 8.3 für die Berechnung der Konturdistanzen; für die f_0 -Distanz wurde lediglich die mittlere absolute Distanz zwischen Eingabebild I_k und aus $\mathbf{M}_k$ rekonstruiertem Bild (Abschnitt 6.1) berechnet). Aus den quantitativen Werten geht hervor, dass die Merkmalsvarianzen in Σ klein genug gewählt werden sollten, um während der Modellgenerierung kleine Distanzen zwischen Modell- und Ziel-/Trainingsinstanz anzustreben. Große Werte für Σ führen zwar zu einem stabilen Konvergenzverhalten, lassen dem generierten Modell aber zu viele Freiheiten. Sind die Werte in Σ jedoch zu klein gewählt (mit $\text{Cov}(\mathbf{S}_k, \mathbf{M}_k)$), kommt es zu lokalen Minima und $C(Q, \Theta)$ konvergiert nicht mehr stabil, weil zu jedem Vektor $\mathbf{s}_i$ nur noch Merkmalsvektoren $\mathbf{m}_j$ betrachtet werden, die bereits sehr ähnlich sind (fast alle $p(\mathbf{s}_i, \mathbf{m}_j) \approx 0$). Dieses widerspricht dem Gedanken der probabilistischen Korrespondenzbeziehungen.

Parameterbelegung

Input-/Trainingsdaten

N	=	60
D	=	4 ($D_f = 2, D_x = 2$)
$\mathbf{f}_i$	:	(1) Gauß-geglättete Signalwerte: $(G_{\sigma=6mm} * I_k)(\mathbf{x}_i)$ (2) Distanzwerte zur Handmaske: $d(\mathbf{x}_i, I_k^{Mask})$
$\mathbf{x}_i$	:	kantenmerkmalsabhängig mit Mindestabstand 6 mm, innerhalb der Hand ($\forall i, f_{i,2} < 5mm$)
sp	=	6 mm
$N_s(k)$	=	450 ± 57
N_m	=	481

Modellparameter

n	=	15
$T_k \star \mathbf{s}_{k,i}$	=	$(T_k(\mathbf{x}_i), \mathbf{f}_i)$ mit $(\mathbf{S}_k - T_k^{-1}(\tilde{\mathbf{M}})) \rightarrow_{T_k} \min$
Σ	=	$\frac{1}{N} \sum_{k=1}^N \text{diag}(2 \cdot \text{Cov}(\mathbf{S}_k, \tilde{\mathbf{M}}))$,
	=	$\frac{1}{N} \sum_{k=1}^N \text{diag}(\text{Cov}(\mathbf{S}_k, \tilde{\mathbf{M}}))$,
	=	$\frac{1}{N} \sum_{k=1}^N \text{diag}(\text{Cov}(\mathbf{S}_k, \mathbf{M}_k) \cdot 0.5 + \text{Cov}(\mathbf{S}_k, \tilde{\mathbf{M}}) \cdot 0.5)$,
	=	$\frac{1}{N} \sum_{k=1}^N \text{diag}(\text{Cov}(\mathbf{S}_k, \mathbf{M}_k))$
Beschränk.	-	mit $p(\Theta)$; - $p(Q \Theta)$ mit Symmetriebedingung

4.2.2. Datengetriebene Anpassung der Varianzen der Nachbarschaftsbeziehungen für $p(\Theta)$

In $p(\Theta)$ (Gleichung 31) beschreibt $\Sigma_{\mathcal{N}} \in \mathbb{R}^{D \times D}$ die erlaubten Variationen der lokalen Distanzen zwischen mittlerem Modell und möglicher Modellinstanz für die lokale Regularisierung. Da die Distanzen zwischen benachbarten Merkmalsvektoren mit den Distanzen im initialen Modell ($\mathcal{N}_{j,\eta}^0 = (\tilde{\mathbf{m}}_j^0 - \tilde{\mathbf{m}}_\eta^0)$) sowohl verglichen als auch normalisiert werden, beschreibt $\Sigma_{\mathcal{N}}$ unabhängig von der realen lokalen Distanz zwischen Nachbarn die erlaubten Varianzen (vergleiche: $p(\Theta) = \dots \frac{((\tilde{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\tilde{\mathbf{m}}_\eta \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \Sigma_{\mathcal{N}}^{-1} \dots$). Das heißt, $\Sigma_{\mathcal{N}}$ kann auf einen gewissen Prozentsatz gesetzt werden: $\Sigma_{\mathcal{N}} = \text{diag}(0.1, \dots, 0.1)$ würde ergeben, dass die initialen Vektordistanzen zwischen Nachbarn durch die Moden um 10% variieren dürfen – die Standardabweichung ist also 10%.

Um hier ebenfalls manuelles Parametertuning zu vermeiden (analog zu Σ), können die Varianzen der aktuellen Daten berechnet werden:

$$\Sigma_{\mathcal{N}} = \text{diag}(\text{Cov}(\bar{\mathbf{M}}, \bar{\mathbf{M}} \pm 2 \cdot \lambda \mathbf{V})) \quad (71)$$

$$\begin{aligned} &:= \text{diag}\left(\frac{1}{n \cdot N_m} \sum_{p=1}^n \sum_{j=1}^{N_m} \frac{1}{|\mathcal{N}(j)|}\right. \\ &\quad \left. \sum_{\eta \in \mathcal{N}(j)} \left\| \frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right\| \right). \end{aligned} \quad (72)$$

Im Gegensatz zu den Merkmalsvarianzen Σ werden die Werte in $\Sigma_{\mathcal{N}}$ größer, da die Moden zu Beginn Null sind und erst mit fortschreitender Optimierung größere Werte annehmen. Daher sind die mit Gleichung 71 bestimmten Werte zu Beginn zu klein (bzw. 0 mit $\mathbf{V} = \mathbf{0}$) und würden somit jegliche Änderungen der Nachbarschaftsdistanzen durch die Moden verhindern. Dies kann z. B. durch die Wahl einer unteren Grenze abgefangen werden:

$$\Sigma_{\mathcal{N}} = \min(\text{diag}(0.1, \dots, 0.1), \text{diag}(\text{Cov}(\bar{\mathbf{M}}, \bar{\mathbf{M}} \pm 2 \cdot \lambda \mathbf{V}))). \quad (73)$$

Außerdem ist es möglich, die berechneten Werte mit zusätzlichen Faktoren zu gewichten:

$$\Sigma_{\mathcal{N}} = \text{diag}(\text{faktor}_{x_1}, \dots, \text{faktor}_{x_{D_x}}, \text{faktor}_{f_1}, \dots, \text{faktor}_{f_{D_f}}) \text{diag}(\text{Cov}(\bar{\mathbf{M}}, \bar{\mathbf{M}} \pm 2 \cdot \lambda \mathbf{V})). \quad (74)$$

Je nach betrachtetem Objekt und dessen Variationsbreite können die Faktoren gewählt werden. Für die Koordinatenwerte (faktor_x) hat sich herausgestellt, dass eine Gewichtung mit Eins robuste Ergebnisse liefert. In den meisten Fällen soll vor allem eine topologische Konsistenz gesichert werden. Daher können die Faktoren für die Appearance-Merkmale (faktor_f) höher gewichtet werden (z. B. mit 10). Um die Appearance-Merkmale während der Regularisierung durch $p(\Theta)$ vollständig zu ignorieren, kann auch eine weit höhere Gewichtung gewählt werden ($\gg 10000$). Da die Werte durch Gleichung 71 an die aktuelle Variabilität des Modells angeglichen werden, ist die Qualität der Ergebnisse relativ robust gegenüber der Wahl der Faktoren. Diese können daher während der Optimierung des Kriteriums fix bleiben. Lediglich die Art der Gewichtung (geringer oder "gar nicht regularisierte" Appearance-Variabilität) muss festgelegt werden. Die untere Grenze (Gleichung 73) kann auch für alle Modelle z. B. mit 10% definiert werden.

4.3. Initialisierung der Modellparameter

Da während der Modellgenerierung alle Parameter iterativ optimiert werden, ist der Algorithmus relativ robust gegenüber der Initialisierung der einzelnen Parameter. Wie beschrieben, wird für alle Parameter pro Iteration das globale Optimum durch Ableitung des Kriteriums bestimmt. Das aktuelle globale Minimum des Kriteriums ist abhängig von den aktuellen probabilistischen Korrespondenzen zwischen Modell und Trainingsinstanzen. Für die Initialisierung ist es somit am wichtigsten, dass die initialen Korrespondenzen für die gesamten Trainingsdaten relativ gleich gut (bzw. “gleich schlecht”) sind. Damit ist gewährleistet, dass alle Trainingsinstanzen bzw. alle Merkmalsvektoren innerhalb der Trainingsinstanzen zur Optimierung der Modellparameter beitragen. Das initiale mittlere Modell $\bar{\mathbf{M}}$ (bzw. die Korrespondenzen aller $\mathbf{S}_k$ zu $\bar{\mathbf{M}}$) ist also ausschlaggebend für die Qualität des generierten Modells. Alle anderen Parameter können mit Null initialisiert werden. Die Merkmalsvarianzen in Σ werden mithilfe der Trainingsdaten initialisiert.

4.3.1. Initialisierung der Merkmalsvarianzen Σ

Da die Berechnung der Korrespondenzbeziehungen $\mathbb{E}(H_{ij})_k$ von den Merkmalsvarianzen Σ abhängig und die beschriebene Bestimmung von Σ (Abschnitt 4.2.1) wiederum von $\mathbb{E}(H_{ij})_k$ abhängig ist, muss eines von beidem als Erstes auf andere Weise initialisiert werden. Hier werden die Werte in Σ mit den Varianzen der Appearance-Merkmalen (Var^f) in den Trainingsdaten initialisiert. Die Varianzen für die Positionsmerkmale werden hingegen auf den Mindestabstand sp zwischen den Positionen der Merkmalsvektoren gesetzt. Das so initialisierte Σ wird lediglich für eine initiale Berechnung der Korrespondenzbeziehungen $\mathbb{E}(H_{ij})_k$ verwendet und ist daher nicht ausschlaggebend für die Modellqualität:

1. Setze $\Sigma = \text{diag}(sp^2, \dots, sp^2, \frac{1}{N} \sum_{k=1}^N Var^f(\mathbf{S}_k))$, mit $Var^f(\mathbf{S}_k) = \sum_{i=1}^{N_s(k)} (\frac{(\mathbf{f}_{k,i} - \bar{\mathbf{f}}_{k,i})^2}{N_s(k)})$
2. Berechne Korrespondenzen zwischen $\bar{\mathbf{M}}$ und $\mathbf{S}_k$: $\mathbb{E}(H_{ij})_k$
3. Berechne $\Sigma = \frac{1}{N} \sum_{k=1}^N \text{diag}(Cov(\mathbf{S}_k, \bar{\mathbf{M}}))$ (Abschnitt 4.2.1)

$\bar{\mathbf{M}}$ sei hier das initiale $\mathbf{S}_x$ – siehe nächsten Abschnitt.

4.3.2. Initialisierung des mittleren Modells

Die Wahl des initialen mittleren Modells ist ausschlaggebend für die Anzahl der Merkmalsvektoren im Modell N_m . Im einfachsten Fall wird $\bar{\mathbf{M}}$ mit einer Trainingsinstanz initialisiert: $\bar{\mathbf{M}} = \mathbf{S}_x$. Liegt diese Instanz $\mathbf{S}_x$ ungefähr in der Mitte des Trainingsdatensatzes, ist keine weitere Vorverarbeitung erforderlich. Die folgende Optimierung wird allerdings robuster, wenn zusätzlich die Merkmalswerte von $\bar{\mathbf{M}}$ über die Trainingsinstanzen “gemittelt” werden. Dafür wird $\bar{\mathbf{M}}$ zuerst mit affiner Transformation in die Mitte

des Trainingsdatensatzes gesetzt. Dies wird über eine paarweise Berechnung der T_k zwischen $\mathbf{S}_k$ und $\bar{\mathbf{M}}$ (Abschnitt 4.1.7) realisiert, wobei anschließend $\bar{\mathbf{M}}$ mit $\frac{1}{N} \sum_{k=1}^N T_k^{-1}$ transformiert wird. Guimond et. al haben gezeigt, dass auf diese Weise das $\bar{\mathbf{M}}$ (bzw. der bei Guimond et. al generierte Atlas) in den Schwerpunkt der Trainingsdaten konvergiert [Guimond *et al.* 2000]. Als nächster Schritt können die Merkmalsvektoren von $\bar{\mathbf{M}}$ "gemittelt" werden. Dies entspricht der Optimierung von $\bar{\mathbf{M}}$ (Abschnitt 4.1.3) ohne Betrachtung der Variationsmoden ($\mathbf{V} = 0$). Die Initialisierung von $\bar{\mathbf{M}}$ kann zusammengefasst werden als:

1. Setze $\bar{\mathbf{M}} = \mathbf{S}_x$, $T_k = id, \forall k$.
2. Transformiere $\bar{\mathbf{M}}$, sodass es mittig des Trainingsdatensatzes liegt:
 - a) Berechne Σ (Abschnitt 4.2.1).
 - b) Berechne Korrespondenzbeziehungen $\mathbb{E}(H_{ij})_k$ zwischen $T_k^{-1}(\bar{\mathbf{M}})$ und $\mathbf{S}_k$ (Abschnitt 1.2.3).
 - c) Optimierte alle Transformationen T_k zwischen $\mathbf{S}_k$ und $\bar{\mathbf{M}}$ (Abschnitt 4.1.7).
 - d) Transformiere $\bar{\mathbf{M}} = (\frac{1}{N} \sum_{k=1}^N T_k^{-1})(\bar{\mathbf{M}})$.
 - e) Gehe zu a) bis Konvergenz.
3. "Mittlere" $\bar{\mathbf{M}}$:
 - a) Berechne Σ (Abschnitt 4.2.1).
 - b) Berechne Korrespondenzbeziehungen $\mathbb{E}(H_{ij})_k$ zwischen $T_k^{-1}(\bar{\mathbf{M}})$ und $\mathbf{S}_k$ (Abschnitt 1.2.3).
 - c) Berechne "Mittelwert" der Trainingsinstanzen (= Optimierte $\bar{\mathbf{M}}$ mit $\mathbf{V} = 0$ – vgl. Abschnitt 4.1.3):

$$\bar{\mathbf{m}}_j = \frac{\sum_{k=1}^N (\sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \Sigma^{-1}(T_k \star \mathbf{s}_{k,i})) + \sum_{\eta \in \mathcal{N}(j)} \Sigma_{\mathcal{N}}^{-1} \left(\frac{\bar{\mathbf{m}}_{\eta} + \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|^2} \right)}{\sum_{k=1}^N \sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \Sigma^{-1} + \sum_{\eta \in \mathcal{N}(j)} \Sigma_{\mathcal{N}}^{-1} \left(\frac{1}{\|\mathcal{N}_{j,\eta}^0\|^2} \right)}.$$
 - d) Gehe zu a) bis Konvergenz.

Durch diese initiale Mittelung ist die Qualität des Modells (abgesehen von der Anzahl der Merkmale) unabhängig von der Wahl von $\mathbf{S}_x$. $\mathbf{S}_x$ kann zum Beispiel vollkommen zufällig bestimmt werden. Es kann auch die Instanz x mit

$$\operatorname{argmin}_x (N_{s(x)} - \frac{1}{N} \sum_{k=1}^N N_{s(k)})^2$$

gewählt werden, damit die Anzahl der Merkmalsvektoren im Modell möglichst nahe an dem Durchschnittswert der Vektoranzahl im Trainingsdatensatz liegt.

4.3.3. Initialisierung der Nachbarschaftsbeziehungen für $p(\Theta)$

Wie in Abschnitt 3.2.2 beschrieben, müssen für eine Bestimmung der A-priori-Wahrscheinlichkeit des Modells Nachbarschaftsannahmen über die Merkmalsvektoren des Modells getroffen werden. Zu diesem Zweck ist für jeden Merkmalsvektor des Modells $\bar{\mathbf{m}}_j$ die Nachbarschaftsmenge $\mathcal{N}(j)$ zu initialisieren. Dabei kann die Anzahl der Nachbarvektoren in $\mathcal{N}(j)$ beliebig gewählt werden und weiterhin darf die Anzahl für jedes j unterschiedlich sein. Um $\mathcal{N}(j)$ zu bestimmen, kann z. B. eine Delaunay-Triangulierung oder ein einfacher k-Nearest-Neighbors Algorithmus angewendet werden. Die optimale Nachbarschaftsanzahl ist abhängig von der Topologie der Merkmalspositionen im Modell und davon wie stark bzw. wie weit der Einfluss eines Vektors auf seine Umgebung sein soll. Außerdem steigt der Rechenaufwand der Optimierung der Modellparameter mit höherer Nachbarschaftsanzahl. Da hier nur eine leichte Regularisierung zwischen direkt benachbarten Vektoren erfolgen soll, wurde in dieser Arbeit eine durchschnittliche Anzahl von $|\mathcal{N}(j)| \approx 4$ (für 2D und $|\mathcal{N}(j)| \approx 6$ für 3D) gewählt.

Beispielbox 8: Initialisierung der Modellparameter

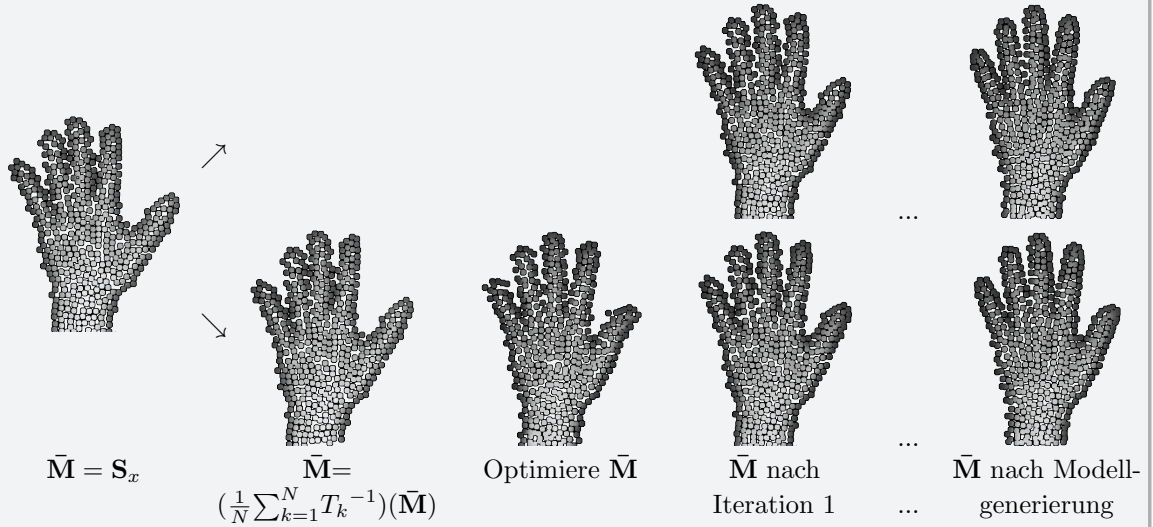


Abbildung 20: Gegenüberstellung zweier Modelle ohne (obere Zeile) und mit (untere Zeile) zusätzlicher Mittelung des initialen $\bar{\mathbf{M}}$, wobei $\mathbf{S}_x$ relativ in der Mitte des Trainingsdatensatzes liegt. Die letzte Spalte zeigt das resultierende $\bar{\mathbf{M}}$ nach vollendeter Modellgenerierung. Bei gut gewähltem $\mathbf{S}_x$ liefern beide Modelle robuste Ergebnisse.

Um die Bedeutung der Initialisierung des mittleren Modells $\bar{\mathbf{M}}$ zu zeigen, wurden vier Modelle generiert: zwei Modelle, bei denen die initiale Belegung ($\bar{\mathbf{M}} = \mathbf{S}_x$) mit einer Trainingsinstanz erfolgte, welche relativ in der Mitte des Trainingsdatensatzes liegt, und zwei Modelle mit einer sehr schlecht gewählten Trainingsinstanz. Für beide Belegungen wurde einmal das Modell ohne weitere Vorverarbeitung generiert und einmal mit zusätzlicher Mittelung der Transformation ($\bar{\mathbf{M}} = (\frac{1}{N} \sum_{k=1}^N T_k^{-1})(\bar{\mathbf{M}})$) und der Merkmalsvektoren (“Optimiere $\bar{\mathbf{M}}$ ”). Die Ergebnisse zeigen, dass die Initialisierung mit der zusätzlichen Mittelung gegenüber der Wahl von $\mathbf{S}_x$ robust ist.

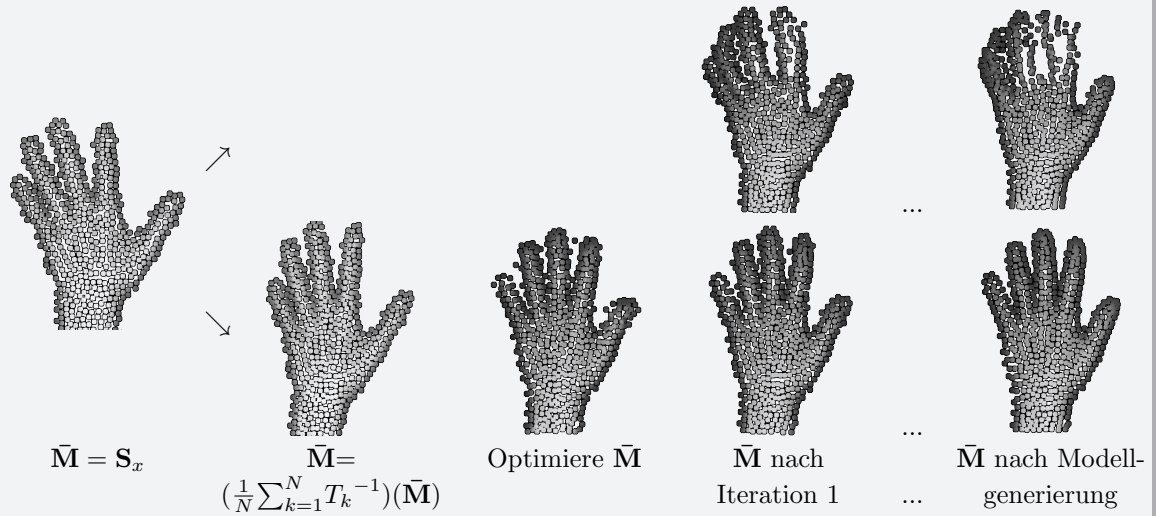


Abbildung 21: Gegenüberstellung zweier Modelle ohne (obere Zeile) und mit (untere Zeile) zusätzlicher Mittelung des initialen $\bar{\mathbf{M}}$, wobei $\mathbf{S}_x$ weit außerhalb der Mitte des Trainingsdatensatzes liegt. Die letzte Spalte zeigt das resultierende $\bar{\mathbf{M}}$ nach vollendeter Modellgenerierung. Bei schlecht gewähltem $\mathbf{S}_x$ (ohne zusätzliche Mittelung) ist ersichtlich, dass die Modellgenerierung bzw. die probabilistischen Korrespondenzen bereits nach einer Iteration ins lokale Minimum driften: kleiner Finger und Ringfinger von $\mathbf{S}_x$ liegen z.B. zu weit außerhalb der restlichen Trainingsdaten und werden somit gemeinsam zu den restlichen kleinen Fingern als korrespondierend angenommen – “verschmelzen” also. Die untere Zeile zeigt, dass durch die zusätzliche Mittelung die Modellgenerierung gegenüber der Wahl von $\mathbf{S}_x$ robust wird.

Initialisierung der Nachbarschaftsbeziehungen: Für den Handdatensatz sei hier beispielhaft die Initialisierung der Nachbarschaftsbeziehungen beschrieben:

Die initialen lokalen Nachbarschaftsbeziehungen für $p(\Theta)$ sind in der Abbildung rechts dargestellt. Es wurden durchschnittlich vier Nachbarn für jeden Merkmalsvektor gewählt mit einem Mindestwinkel zwischen den Nachbarn. Die Nachbarschaftsbeziehungen sind nicht symmetrisch. Da hier nur topologische Beziehungen zwischen den Merkmalsvektoren betrachtet werden, beziehen sich die Distanz- oder Winkelberechnungen lediglich auf den Koordinatenanteil der Vektoren:

```

minDist = 6 mm
while 1/N_m \sum_{j=1}^{N_m} |\mathcal{N}(j)| < 4
  for j = 1, ... N_m
    if |\mathcal{N}(j)| \leq 5
      for \eta = 1, ... N_m
        if (d(\bar{\mathbf{m}}_j, \bar{\mathbf{m}}_\eta) < minDist) \wedge \angle(\bar{\mathbf{m}}_\eta, \mathcal{N}(j)) > 30^\circ
          \mathcal{N}(j) = \{\mathcal{N}(j), \bar{\mathbf{m}}_\eta\}
        minDist = minDist \cdot 1.1
      end while
    end while
  end while

```

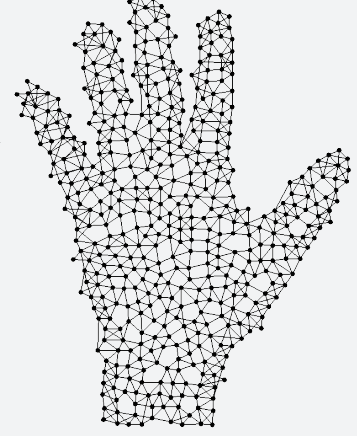


Abbildung 22: Nachbarschaftsbeziehungen

Parameterbelegung

Input-/Trainingsdaten

N	=	60
D	=	4 ($D_f = 2, D_x = 2$)
$\mathbf{f}_i$	:	(1) Gauß-geglättete Signalwerte: $(G_{\sigma=6mm} * I_k)(\mathbf{x}_i)$ (2) Distanzwerte zur Handmaske: $d(\mathbf{x}_i, I_k^{Mask})$
$\mathbf{x}_i$	:	kantenmerkmalsabhängig mit Mindestabstand 6 mm, innerhalb der Hand ($\forall i f_{i,2} < 5mm$)
sp	=	6 mm
$N_s(k)$	=	450 ± 57
N_m	=	481

Modellparameter

n	=	15
$T_k * \mathbf{s}_{k,i}$	=	$(T_k(\mathbf{x}_i), \mathbf{f}_i)$ mit $(\mathbf{S}_k - T_k^{-1}(\bar{\mathbf{M}})) \rightarrow_{T_k} \min$
Σ	=	$\frac{1}{N} \sum_{k=1}^N \text{diag}(\text{Cov}(\mathbf{S}_k, \mathbf{M}_k) \cdot 0.5 + \text{Cov}(\mathbf{S}_k, \bar{\mathbf{M}}) \cdot 0.5)$
$ \mathcal{N}(j) $	$\approx$	4
Beschränk.	-	mit $p(\Theta)$; - $p(Q \Theta)$ mit Symmetriebedingung

4.3.4. Zusammenfassung der Initialisierung

Die Initialisierung aller Parameter kann zusammengefasst werden als:

1. Setze $\mathbf{V} = 0, \{\lambda_p\} = 0, T_k = id, \omega_k = 0$
2. Setze $\bar{\mathbf{M}} = \mathbf{S}_x$
3. Initialisiere Nachbarschaftsbeziehungen $\mathcal{N}(j) \forall j$ in $\bar{\mathbf{M}}$
4. Initialisiere Σ
5. Transformiere $\bar{\mathbf{M}}$ so, dass es mittig des Trainingsdatensatzes liegt

IN N Trainingsinstanzen (mit D_f Appearance-Merkmalen) samt Positionen $\mathbf{x}_{k,i}$;
Anzahl der Moden n

INIT $\mathbf{S}_k = \{\mathbf{s}_{k,i} = (\mathbf{x}_{k,i}, \mathbf{f}_{k,i}) | i = 1, \dots, N_s(k)\}$;

Abs. 4.3

Initialisiere Parameter:

1. Setze $\mathbf{V} = 0, \{\lambda_p\} = 0, T_k = id, \boldsymbol{\omega}_k = 0$, setze $\bar{\mathbf{M}} = \mathbf{S}_x$
2. Initialisiere Nachbarschaftsbeziehungen $\mathcal{N}(j) \forall j$ in $\bar{\mathbf{M}}$ und die daraus resultierenden Distanzen $\mathcal{N}_{j,\eta}^0 = (\bar{\mathbf{m}}_j^0 - \bar{\mathbf{m}}_\eta^0)$ und $\Sigma_{\mathcal{N}}$ -Werte.
3. Initialisiere Σ
4. Transformiere und optimiere $\bar{\mathbf{M}}$

DO

Optimierung der Modellparameter Abs. 4.1

0. Update $\Sigma, \Sigma_{\mathcal{N}}$ (und berechne $\mathbf{M}_k, \mathbb{E}(H_{ij})_k$)
(Update die Distanzen $\mathcal{N}_{j,\eta}^0 = (\bar{\mathbf{m}}_j - \bar{\mathbf{m}}_\eta)$ für $p(\Theta)$) Abs. 4.2.1
1. Berechne $C(Q, \Theta)_{start}$
2. Optimierte $\bar{\mathbf{M}}$ (und berechne $\mathbf{M}_k, \mathbb{E}(H_{ij})_k$)
3. Optimierte $\mathbf{V}$ (und berechne $\mathbf{M}_k, \mathbb{E}(H_{ij})_k$)
4. Optimierte $\{\lambda_p\}$ (und berechne $\mathbf{M}_k, \mathbb{E}(H_{ij})_k$)

Optimierung der instanzabhängigen Parameter (Θ fix) $\leftarrow$ Modellanpassung

Abs. 4.4

5. Optimierte T_k (und berechne $\mathbf{M}_k, \mathbb{E}(H_{ij})_k$)
6. Optimierte $\boldsymbol{\omega}_k$ (und berechne $\mathbf{M}_k, \mathbb{E}(H_{ij})_k$)
7. Berechne $C(Q, \Theta)_{curr}$

WHILE $(C(Q, \Theta)_{start} - C(Q, \Theta)_{curr}) > \epsilon$

OUT $\Theta = \{\bar{\mathbf{M}}, \mathbf{V}_p, \lambda_p, n, \Sigma | p = 1, \dots, n\}$

Abbildung 23: Zusammenfassung des Modellgenerierungsprozesses.

6. Optimierte $\bar{\mathbf{M}}$ mit $\mathbf{V} = 0$

Die Anzahl der Moden n wird an dieser Stelle als fixer Inputparameter betrachtet.

4.4. Modellanpassung

Bei klassischen Form- und Appearance-Modellen stellt die Modellanpassung auf ein neues ungesehenes Bild ein Problem dar, welches unabhängig von der Modellgenerierung gelöst werden muss. Für die Optimierung der Parameter T_{new} und $\boldsymbol{\omega}_{new}$ für ein gegebenes $\mathbf{S}_{new}$ werden dabei ganz eigene Optimierungsverfahren entwickelt, siehe Abschnitt 1.2.1 und 1.2.2. Der beschriebene Ansatz hat dem gegenüber den großen Vorteil, dass die

```

IN   Eine Testinstanz (mit  $D_f$  Appearance-Merkmalen) samt Positionen  $\mathbf{x}_i$ ;
        $\Theta = \{\bar{\mathbf{M}}, \mathbf{V}_p, \lambda_p, n, \Sigma | p = 1, \dots, n\}$ 

INIT  $\mathbf{S}_{new} = \{\mathbf{s}_i = (\mathbf{x}_i, \mathbf{f}_i) | i = 1, \dots, N_s\}$ 
       Initialisiere Parameter: Setze  $T_{new} = id, \omega_{new} = 0$  Abs. 4.3

DO

       Optimierung der instanzabhängigen Parameter
       0. Update  $\Sigma$  (und berechne  $\mathbf{M}_{new}, \mathbb{E}(H_{ij})_{new}$ ) Abs. 4.2.1
       1. Berechne  $C(Q, \Theta)_{start}$  Abs. 4.1, 4.4
       2. Optimierte  $T_{new}$  (und berechne  $\mathbf{M}_{new}, \mathbb{E}(H_{ij})_{new}$ )
       3. Optimierte  $\omega_{new}$  (und berechne  $\mathbf{M}_{new}, \mathbb{E}(H_{ij})_{new}$ )
       4. Berechne  $C(Q, \Theta)_{curr}$ 

WHILE  $(C(Q, \Theta)_{start} - C(Q, \Theta)_{curr}) > \epsilon$ 

OUT  $\mathbf{M}_{new}$ 

```

Abbildung 24: Zusammenfassung der Modellanpassung.

Optimierung der instanzabhängigen Parameter und somit die Modellanpassung bereits mit demselben globalen Kriterium abgedeckt ist, welches zur Modellgenerierung genutzt wird. Das heißt, für die Modellanpassung werden lediglich die gefundenen Modellparameter Θ als fix angenommen und $C(Q_{new}, \Theta)$ wird mit Hinblick auf Q_{new} minimiert:

$$\operatorname{argmin}_{Q_{new}} C(Q_{new}, \Theta) = \operatorname{argmin}_{Q_{new}} -\log p(Q_{new}, \Theta | \mathbf{S}_{new}).$$

Die Symmetriebedingung, die in Abschnitt 3.2.4 vorgestellt wurde, gilt dabei während der Modellanpassung nicht, da die Werte für T_{new} nicht beschränkt und die Werte für ω_{new} ausschließlich durch die Standardabweichung der Moden beschränkt werden (siehe Abschnitt 4.1.7 und 4.1.6). In Abbildung 24 wird die Modellanpassung zusammengefasst. An dieser Stelle sei angenommen, dass die Merkmalsvektoren in $\mathbf{S}_{new}$ auf dieselbe Art definiert werden können, wie die Merkmalsvektoren im zugrundeliegenden Trainingsdatensatz. Eine ausführlichere Betrachtung der Initialisierung von $\mathbf{S}_{new}$ folgt in Abschnitt 6.4.

E-Schritt für Modellanpassung: Für die Modellgenerierung wurde die Normierung der Erwartungswerte (wie in Abschnitt 4.1.1 beschrieben) mit den Gleichungen 10 und 12 durchgeführt. Für die Modelloptimierung wird also angenommen, dass die Bildrepräsentationen des Trainingsdatensatzes keine Outlier beinhalten. Für die Modellanpassung hingegen kann angenommen werden, dass das gegebene Modell keine Outlier aufweist (da die zugrundeliegenden $\mathbf{S}_k$ keine Outlier enthielten). Ein neues unbekanntes Bild I_{new} könnte wiederum fehlerhafte Regionen oder “falsch abgetastete” Merkmalsvek-

toren in $\mathbf{S}_{new}$ besitzen. Das heißt, für die Modellanpassung sollte die Normierung über Gleichung 11 und 13 erfolgen – mit dem Ergebnis, dass $\sum_{i=1}^{N_s} \mathbb{E}(H_{ij}) = 1, \forall j$ ($= \mathbf{M}_{new}$ weist keine Outlier auf).

4.5. Zusammenfassung der Modellgenerierung und -anpassung

Die vollständige Generierung des Modells wird in Abbildung 23 und die Modellanpassung in Abbildung 24 zusammengefasst. Dabei sei darauf hingewiesen, dass:

1. lediglich die Anzahl der Moden n vor Modellgenerierung festgelegt werden muss,
2. alle Initialisierungen automatisch erfolgen,
3. die Modell- und instanzabhängigen Parameter durch eine analytische Lösung des globalen Kriteriums direkt optimiert werden,
4. für die Modellanpassung keine zusätzliche Methode nötig ist und
5. Regularisierungsparameter $(\Sigma, \Sigma_{\mathcal{N}})$ ohne manuelles Tuning pro Iteration an die aktuelle Datensituation angepasst werden.

5. Erweiterungen des Modells

Im vorangegangenen Abschnitt 4 wurde die Generierung des Modells über die Optimierung aller Parameter mittels des globalen Kriteriums vorgestellt. Dabei wird das Kriterium nach allen Parametern abgeleitet, Null gesetzt und nach dem jeweiligen Parameter umgestellt. Die Qualität des optimierten Modells hängt von den iterativ bestimmten probabilistischen Korrespondenzen zwischen Trainingsdatensatz und Modell ab. Das heißt, in jeder Iteration müssen

$$N_m \times \sum_{k=1}^N N_s(k)$$

Korrespondenzwerte upgedatet werden, um

$$n \times 2N_m \times D$$

(Modenanzahl $\times$ Vektoranzahl des Modells in $\bar{\mathbf{M}}$ und $\mathbf{V} \times$ Dimension eines Merkmalsvektors) Parameter zu optimieren. Um die Optimierung stabiler zu gestalten, werden im Folgenden verschiedene Ansätze vorgestellt, um die Anzahl der Parameter zu verringern bzw. mit verringerter Anzahl die Optimierung zu starten und anschließend die Parameteranzahl iterativ zu erhöhen (Abschnitt 5.1 und 5.2). In Abschnitt 5.3 wird das iterative Inkrementieren der Merkmalsvarianzen Σ diskutiert.

Bei modellbasierten Ansätzen ist außerdem die Anzahl der optimierten Parameter ($N_m \times D$) wesentlich größer als die Anzahl der Freiheitsgrade des resultierenden Modells (n). In Abschnitt 5.4 wird eine zusätzliche Erweiterung vorgestellt, um die Anzahl der Freiheitsgrade durch “lokal begrenzte” Moden zu erhöhen.

5.1. Multi-Level-Ansatz zum Inkrementieren der Merkmalsvektoranzahl

Die Anzahl der zu optimierenden Parameter und der Korrespondenzwahrscheinlichkeiten wird stark erhöht durch eine “hohe Anzahl” an Merkmalsvektoren in den Trainingsdaten (und somit hohe Anzahl an Merkmalsvektoren im Modell). Eine hohe Vektoranzahl ist außerdem gleichzusetzen mit einem geringen Abstand sp zwischen den Vektoren eines Modells bzw. der Trainingsinstanzen. Ist dieser Abstand sp wesentlich kleiner als die Positions- bzw. Formvariation des betrachteten Objekts, ist es wahrscheinlicher, dass falsche Korrespondenzen bestimmt werden: je kleiner der Merkmalsvektorabstand, desto kleiner die Werte der Merkmalsvarianzen in Σ für die Korrespondenzberechnung (Abschnitt 4.2.1) und demnach desto kleiner der Bereich in den Bildern, der als potentiell korrespondierend betrachtet wird. Daher ist es wahrscheinlicher, dass die Korrespondenzbestimmung in einem lokalen Minimum endet.

Durch die Erweiterung des Kriteriums um ein nichtkonstantes $p(\Theta)$ kommt es bereits zu einer lokalen Regularisierung der Merkmalsvektoren, da starke lokale Distanzände-

rungen benachbarter Vektoren bestraft werden (Abschnitt 3.2.2). Das Problem von lokalen Minima in den probabilistischen Korrespondenzbeziehungen ist damit jedoch nicht gelöst.

Ein ähnliches Problem tritt bei der elastischen Registrierung zweier Bilder auf, bei der die Anzahl der Deformationsparameter durch die Pixelanzahl der Bilder bestimmt wird. Ein weitverbreiteter Ansatz, die Herausforderung der hohen Anzahl an Eingabeparameter zu handhaben, ist die Lösung über einen Multi-Resolution- bzw. Multi-Level-Ansatz. Bei der Registrierung wird dabei die Auflösung (engl. resolution) der betrachteten Bilder zu Beginn der Optimierung verringert. Zum einen verringert sich dadurch die Pixelanzahl und zum anderen wird der Pixelabstand größer. So werden in den ersten Iterationen gröbere Unterschiede zwischen den Bildern ausgeglichen. Anschließend wird die Auflösung iterativ erhöht, um immer feinere Deformationen zu finden. Das bedeutet, große/grobe Deformationen werden mithilfe weniger Pixel gefunden (was die Optimierungszeit verringert) und mit höherer Auflösung (mehr Pixel) müssen lediglich die Restdistanzen ausgeglichen werden.

Analog hierzu wird im Folgenden ein Multi-Level-Ansatz vorgestellt. Dabei wird in jedem Level der Abstand sp zwischen den Merkmalsvektoren der Trainingsdaten und des Modells verringert und dementsprechend $N_s(k)$ bzw. N_m erhöht. Wird mit einem großen sp (relativ zur Positions- bzw. Form-Variation des betrachteten Objekts) gestartet, ist in den ersten Leveln sichergestellt, dass grobe/extreme Formvariationen abgebildet werden. In den höheren Leveln kommt es mit kleinem sp zur Anpassung an die feineren Form-/Appearance-Variationen.

Die "Verringerung" des Merkmalsvektorabstands sp erfolgt hierbei durch wiederholtes Abtasten der Eingabebilder des Trainingsdatensatzes mit einem jeweils kleineren Abstand sp (siehe Abschnitt 2). Das Modell des neuen Levels hat demnach ebenfalls mehr Merkmalsvektoren mit einem geringeren Abstand (siehe Initialisierung von $\bar{\mathbf{M}}$ durch $\bar{\mathbf{M}} = \mathbf{S}_x$ – Abschnitt 4.3).

5.1.1. Initialisierung des nächsten Levels

Um nachfolgend die Optimierung der Modellparameter in mehreren Leveln durchzuführen, müssen die Parameter eines Levels l mit den Parametern des vorangegangenen Levels $l - 1$ initialisiert werden. Analog zu dem Problem der nichtexistierenden bzw. unbekannten Korrespondenzen zwischen Trainingsinstanzen und Modell sind ebenfalls die Korrespondenzen zwischen zwei Modellen mit unterschiedlichem Merkmalsvektorabstand unbekannt. Daher werden ebenfalls probabilistische Korrespondenzen zwischen den Leveln genutzt: $p(\bar{\mathbf{m}}_j^l | \bar{\mathbf{m}}_j^{l-1})$. Gesucht sind die Parameter für Level l mit den gegebenen Parametern des Levels $l - 1$ mit $l = 2, \dots, N_l$:

$$\Theta^{l-1} = \{\bar{\mathbf{M}}^{l-1}, \mathbf{V}_p^{l-1}, \lambda_p^{l-1}, n, \Sigma^{l-1} | p = 1, \dots, n\}$$

werden (Abschnitt 4.3). Daher kann die Korrespondenzwahrscheinlichkeit zwischen zwei Merkmalsvektoren der mittleren Modelle unterschiedlicher Level über die Merkmalsvektoren mit demselben Index j dieser Trainingsinstanzen ($\mathbf{S}_x^l$ und $\mathbf{S}_x^{l-1}$) definiert werden:

$$p(\bar{\mathbf{m}}_j^l | \bar{\mathbf{m}}_j^{l-1}) := p(\mathbf{s}_{x,j}^l | \mathbf{s}_{x,j}^{l-1}) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_{l-1}|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} (\mathbf{s}_{x,j}^l - \mathbf{s}_{x,j}^{l-1})^T \Sigma_{l-1}^{-1} (\mathbf{s}_{x,j}^l - \mathbf{s}_{x,j}^{l-1}) \right). \quad (75)$$

Mit diesen berechneten Korrespondenzbeziehungen zwischen den Modellen können die Moden und das mittlere Modell des neuen Levels als gewichtete Mittelwerte initialisiert werden:

$$\mathbf{v}_{p,j}^l = \frac{\sum_{j=1}^{N_m^{l-1}} p(\bar{\mathbf{m}}_j^l | \bar{\mathbf{m}}_j^{l-1}) \cdot \mathbf{v}_{p,j}^{l-1}}{\sum_{j=1}^{N_m^{l-1}} p(\bar{\mathbf{m}}_j^l | \bar{\mathbf{m}}_j^{l-1})}, \quad (76)$$

$$\bar{\mathbf{m}}_j^l = \frac{\sum_{j=1}^{N_m^{l-1}} p(\bar{\mathbf{m}}_j^l | \bar{\mathbf{m}}_j^{l-1}) \cdot \bar{\mathbf{m}}_j^{l-1}}{\sum_{j=1}^{N_m^{l-1}} p(\bar{\mathbf{m}}_j^l | \bar{\mathbf{m}}_j^{l-1})}. \quad (77)$$

Des Weiteren sind die Korrespondenzwahrscheinlichkeiten zwischen den Trainingsdaten unterschiedlicher Level gegeben durch:

$$p(\mathbf{s}_{k,i}^l | \mathbf{s}_{k,i}^{l-1}) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_{l-1}|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} (\mathbf{s}_{k,i}^l - \mathbf{s}_{k,i}^{l-1})^T \Sigma_{l-1}^{-1} (\mathbf{s}_{k,i}^l - \mathbf{s}_{k,i}^{l-1}) \right). \quad (78)$$

Diese können hinzugezogen werden, um die probabilistischen Korrespondenzbeziehungen für das neue Level zu initialisieren:

$$\mathbb{E}(H_{ij}^l)_k = \frac{\sum_{\hat{i}=1}^{N_s(k)^{l-1}} \sum_{j=1}^{N_m^{l-1}} p(\bar{\mathbf{m}}_j^l | \bar{\mathbf{m}}_j^{l-1}) p(\mathbf{s}_{k,i}^l | \mathbf{s}_{k,\hat{i}}^{l-1}) \cdot \mathbb{E}(H_{i\hat{j}}^{l-1})_k}{\sum_{\hat{i}=1}^{N_s(k)^{l-1}} \sum_{j=1}^{N_m^{l-1}} p(\bar{\mathbf{m}}_j^l | \bar{\mathbf{m}}_j^{l-1}) p(\mathbf{s}_{k,i}^l | \mathbf{s}_{k,\hat{i}}^{l-1})}. \quad (79)$$

Die Abbildung 25 visualisiert die Initialisierung eines neuen Levels über die probabilistischen Korrespondenzen zwischen den Modellen bzw. zwischen den Trainingsdaten.

Diese Art der Initialisierung durch die Nutzung probabilistischer Korrespondenzen hat den Vorteil, dass keinerlei Informationen über die Positionen der Merkmalsvektoren beider aufeinanderfolgenden Level vorhanden sein muss. Allerdings werden die initialen Werte für $\mathbf{V}$ und $\bar{\mathbf{M}}$ auf relative einfache Weise über gewichtete Mittelwerte der lokal nächsten Merkmalsvektoren (mit $p(\bar{\mathbf{m}}_j^l | \bar{\mathbf{m}}_j^{l-1}) > 0$) berechnet. Dadurch auftretende “Glättungen” der Werte werden in folgenden Iterationen des aktuellen Levels der Modelloptimierung wieder korrigiert, da $\bar{\mathbf{M}}$ und $\mathbf{V}$ in jeder Iteration neu optimiert werden – in Abhängigkeit der aktuellen Korrespondenzwahrscheinlichkeiten. Daher ist die Berechnung der initialen Korrespondenzbeziehungen $\mathbb{E}(H_{ij}^l)_k$ mithilfe der alten Werte

```

IN    $N$  Trainingsinstanzen (mit  $D_f$  Appearance-Merkmalen) samt Positionen  $\mathbf{x}_{k,i}^l$  (für alle Level); Anzahl der Moden  $n$ , Anzahl der Level  $N_l$ 

FOR  $l = 1 : N_l$ 
  INIT  $\mathbf{S}_k^l = \{\mathbf{s}_{k,i}^l = (\mathbf{x}_{k,i}^l, \mathbf{f}_{k,i}^l) | i = 1, \dots, N_s(k)\}$ 
    IF  $l = 1$ : Abs. 4.3
      Setze  $\mathbf{V}^l = 0, \{\lambda_p\} = 0, T_k^l = id, \boldsymbol{\omega}_k^l = 0$ , setze  $\bar{\mathbf{M}}^l = \mathbf{S}_x^l$ 
      Initialisiere  $\Sigma^l$ , transformiere und optimiere  $\bar{\mathbf{M}}^l$ 
    IF  $l > 1$ : Abs. 5.1.1
      Initialisiere Parameter  $(\mathbf{V}^l, \bar{\mathbf{M}}^l)$  und Korrespondenzbeziehungen  $\mathbb{E}(H_{ij}^l)_k$ 
      für Level  $l$  mit Parametern von Level  $l - 1$ 
    Initialisiere Nachbarschaftsbeziehungen  $\mathcal{N}(j) \forall j$  in  $\bar{\mathbf{M}}^l$ 
  DO
    Optimierung der Modellparameter ... Abs. 4.1
    Optimierung der instanzabhängigen Parameter ...
  WHILE  $C(Q^l, \Theta^l)$  nimmt ab
OUT  $\Theta^{N_l} = \{\bar{\mathbf{M}}^{N_l}, \mathbf{V}_p^{N_l}, \lambda_p^{N_l}, n, \Sigma^{N_l} | p = 1, \dots, n\}$ 

```

Abbildung 26: Multi-Level-Ansatz: Zusammenfassung des Modellgenerierungsprozesses mit mehreren Leveln $l = 1, \dots, N_l$. Lediglich die Initialisierung der Modellparameter $(\mathbf{V}^l, \bar{\mathbf{M}}^l)$ (Hellgrau unterlegt) durch die Werte des vorangegangenen Levels weicht von der Modellgenerierung mit einem Level ab.

$\mathbb{E}(H_{ij}^{l-1})_k$ am wichtigsten für die Initialisierung des neuen Levels. Abbildung 26 fasst den Multi-Level-Ansatz für die Modellgenerierung zusammen. Die Optimierung der Parameter selbst bleibt zum vorangegangenen Abschnitt identisch.

Die Ausgabe des Multi-Level-Ansatzes bilden die Modellparameter des letzten Levels mit dem kleinsten Merkmalsvektorabstand sp – alle Parameter der anderen Level werden nicht gespeichert. Wird also das Modell auf ein neues Bild angepasst, wird $\mathbf{S}_{new}$ mit dem feinsten Merkmalsvektorabstand des letzten Levels abgetastet/erzeugt und das Modell entsprechend angepasst.

Beispielbox 9: Multi-Level-Modelle

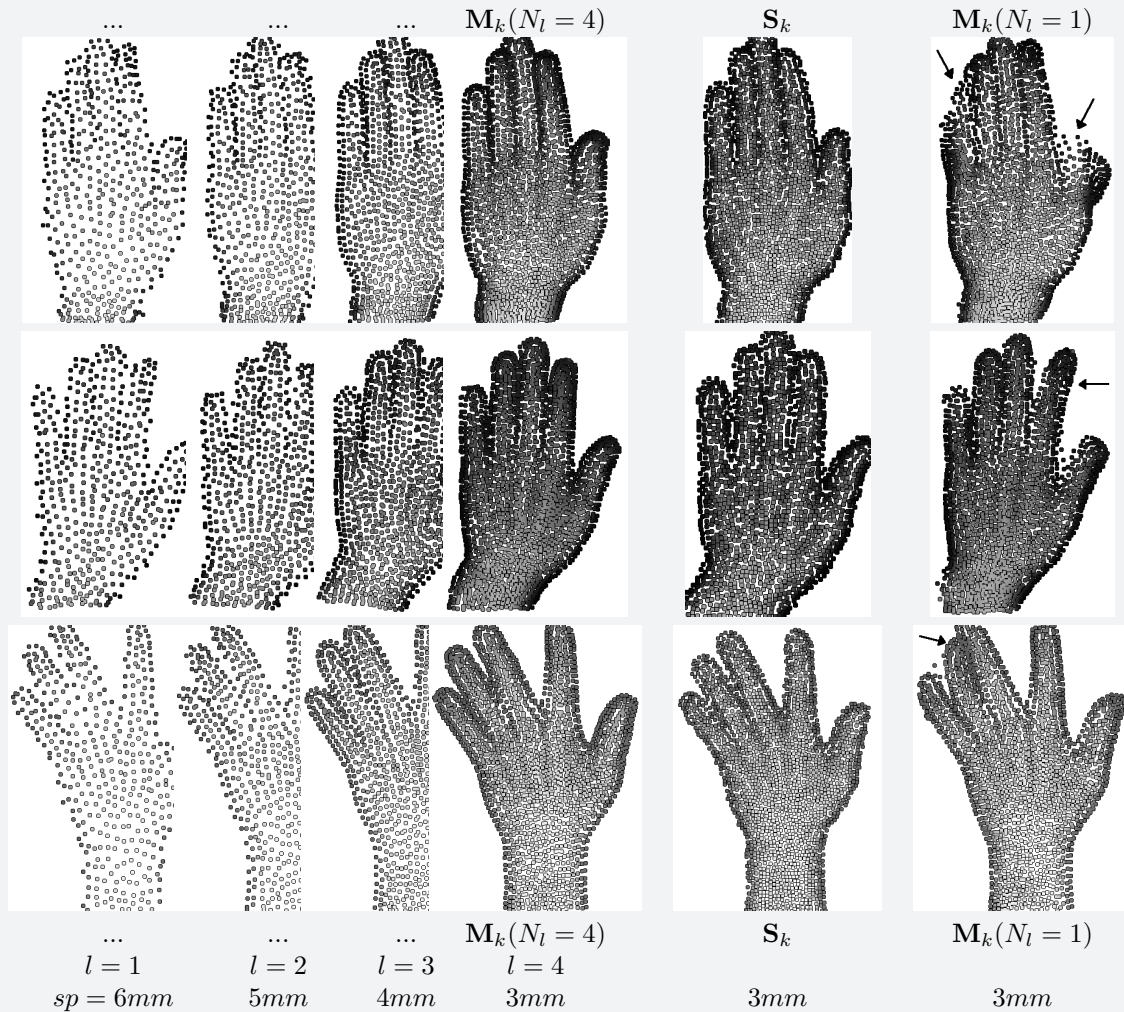


Abbildung 27: Gegenüberstellung zweier Modelle mit einem Level ($sp = 3mm$, rechts) und vier Levels ($sp = 6, 5, 4, 3mm$, links). Die mit Pfeilen markierten Stellen kennzeichnen Bereiche besonders hoher Variationen (im Gegensatz zum Rest der Trainingsdaten) in drei Trainingsbeispielen. Für diese Trainingsdaten stellt die Korrespondenzbestimmung mit kleinem Merkmalsvektorabstand eine besondere Herausforderung dar.

Um den Effekt der zu kleinen Merkmalsvektorabstände gegenüber der Nutzung des Multi-Level-Ansatzes zu demonstrieren, wurden zwei Modelle generiert: ein Modell mit $sp = 3mm$ mit einem Level ($N_l = 1$) und ein Modell mit vier Levels ($N_l = 4$), wobei die Abstände $sp = 6, 5, 4, 3mm$ sind. Die resultierenden Modelle haben demnach im letzten Level ($sp = 3mm$) dieselbe Anzahl an Merkmalsvektoren. In Abbildung 27 werden drei Beispiele für Trainingsinstanzen (Spalte: S_k) gezeigt, die anfällig für lokale Minima sind, da die Finger der Hände sehr dicht zusammen liegen und der

Rest der Trainingsdaten wesentlich mehr Fingerspreizung aufweisen oder die Hand im Gegensatz zum Großteil der Trainingsdaten wesentlich stärker im Handgelenk abgeknickt ist. Im ersten Modell (rechte Spalte $\mathbf{M}_k(N_l = 1)$) werden für diese Instanzen während der Parameteroptimierung falsche Korrespondenzen geschätzt und es kommt zur Generierung von Moden, die “unerwünschte” Variationen beinhalten, wie z. B. “Auseinanderlaufen” der Merkmalsvektoren, die zum Daumen gehören sollten, oder “Verschmelzung” von Ring- und Mittelfinger. Im zweiten Modell (die vier linken Spalten $\mathbf{M}_k(N_l = 4)$) führt die Vorschaltung der Parameteroptimierung für größere sp zu einem wesentlich stabileren Verhalten.

Parameterbelegung

Input-/Trainingsdaten	
N	= 60
D	= 4 ($D_f = 2, D_x = 2$)
$\mathbf{f}_i$	: (1) Gauß-geglättete Signalwerte: $(G_{\sigma=6mm} * I_k)(\mathbf{x}_i)$ (2) Distanzwerte zur Handmaske: $d(\mathbf{x}_i, I_k^{Mask})$
$\mathbf{x}_i$	: kantenmerkmalsabhängig mit Mindestabstand 3 mm, innerhalb der Hand ($\forall i f_{i,2} < 5mm$)
sp	= 3 mm vs. 3-6 mm
$N_s(k)$	= 450 $\pm$ 57 ($sp = 6 mm$) bis 1744 $\pm$ 230 ($sp = 3 mm$)
N_m	= 481 ($sp = 6 mm$) bis 1837 ($sp = 3 mm$)
Modellparameter	
n	= 15
$T_k \star \mathbf{s}_{k,i}$	= $(T_k(\mathbf{x}_i), \mathbf{f}_i)$ mit $(\mathbf{S}_k - T_k^{-1}(\bar{\mathbf{M}})) \rightarrow_{T_k} \min$
Σ	= $((4 - \min(it, 3))) \cdot \frac{1}{N} \sum_{k=1}^N \text{diag}(\text{Cov}(\mathbf{S}_k, \mathbf{M}_k) \cdot 0.5 + \text{Cov}(\mathbf{S}_k, \bar{\mathbf{M}}) \cdot 0.5)$
	für Modellanpassung
$ \mathcal{N}(j) $	≈ 4
N_l	= 1 vs. 4
Beschränk.	- mit $p(\Theta)$; - $p(Q \Theta)$ mit Symmetriebedingung

5.2. Sukzessives Inkrementieren der Modenanzahl n

Das vorgestellte Modell umfasst eine Vielzahl freier Parameter mit $N_m \cdot D$ für $\bar{\mathbf{M}}$ und $n \cdot N_m \cdot D$ für $\mathbf{V}$, die in jeder Iteration der Modellgenerierung gesamt optimiert werden. Mithilfe des Multi-Level-Ansatzes (durch das sukzessive Inkrementieren der Anzahl der Merkmalsvektoren N_m im Modell) kann bereits eine höhere Robustheit erreicht werden. Eine weitere Möglichkeit bildet das Inkrementieren der Modenanzahl n . Besonders bei gewünscht großem n ist es sinnvoll, die Optimierung der Parameter in den ersten Iterationen mit kleinem n zu starten. Da der Umfang der Variationsbreite von n abhängt, wird so in den ersten Iterationen sichergestellt, dass die groben Variationen im Trainingsdatensatz durch die Modellparameter abgedeckt sind und mit steigendem n immer mehr (feinere) Variationen hinzukommen. Nach dem Inkrementieren von n können die Werte der dazukommenden Mode in $\mathbf{V}$ mit 0 initialisiert werden.

Es sei beachtet, dass das Inkrementieren von n nicht dazu führt, dass im nächsten “ n -Level” ausschließlich die hinzukommenden Moden neu berechnet werden. In jeder Iteration werden alle Werte in $\mathbf{V}$ wie bisher optimiert. Die schrittweise Erhöhung von n führt lediglich dazu, dass in den ersten Iterationen (mit kleinem n) schneller die groben Variationen ohne Overfitting an einzelne Trainingsdaten gefunden werden und

die daraus folgenden Korrespondenzbeziehungen $\mathbb{E}(H_{ij})_k$ für die weiteren Iterationen ebenfalls gegenüber lokalen Minima robuster sind.

Das Problem der richtigen Anzahl von n , kann über die Werte der Standardabweichungen λ_p der Moden behandelt werden. Die Werte von λ_p beschreiben den Einfluss der jeweiligen Mode p . Wird der Wert für die letzte Mode (bei Sortierung der Moden nach λ_p) demnach zu klein, so ist diese entweder überflüssig oder beinhaltet nur Rauschen. In [Hufnagel *et al.* 2007] wurde die Optimierung der Modelle mit einem großen n ($n < N$) begonnen; fällt die Standardabweichung für die “kleinste Mode” unter einen bestimmten Wert, so wurde diese Mode entfernt. In der hier vorgestellten Methode hingegen wächst die Anzahl n . Es kann jedoch ebenfalls getestet werden, ob bzw. wie stark die Standardabweichung für die neu hinzugekommene Mode steigt (da mit $\mathbf{V}_{p=n} = 0$ und $\lambda_{p=n} = 0$ initialisiert wird). Anschließend kann bei keinem $\lambda_{p=n}$ bzw. kleiner Steigung der Werte die Inkrementierung von n abgebrochen werden. Die Definition von “klein” kann im Verhältnis zur größten ($\lambda_{p=1}$) oder mittleren ($\frac{1}{n-1} \sum_{p=1}^{n-1} \lambda_p$) Standardabweichung bestimmt werden.

Beispielbox 10: Die Anzahl der Moden n

Um die Auswirkung der Modenanzahl n zu demonstrieren, wurden mehrere Modelle mit unterschiedlichem n generiert. Die oberen Boxplots zeigen die Handkonturdistancen zwischen Trainingsdaten und generiertem Modell (bzw. adaptierter Modellinstanz) für unterschiedlich gewählte n (ohne Inkrementieren während der Parameteroptimierung). Es ist ersichtlich, dass ab $n = 25$ keine eindeutige Verbesserung der Distanzen auftritt. In den unteren Boxplots sind Modelle mit $n = 30/40/50$ sowohl ohne Inkrementieren (rot) als auch mit Inkrementieren (blau) gegenübergestellt, wobei die Optimierung mit Inkrementieren mit $n = 10$ begonnen wurde. Die Modelle mit Inkrementieren zeigen für jedes n die geringeren Handkonturdistancen nach vollendeter Parameteroptimierung. Außerdem wird bei der Betrachtung der resultierenden Standardabweichungen der Moden (untere Grafiken) deutlich, dass bei der n -inkrementierenden Optimierung (blau) die ersten Moden einen größeren Anteil an der gesamt Variation des Modells haben, während bei der Optimierung, die sofort mit der hohen Anzahl an Moden (rot; hier $n = 30$) beginnt, der Einfluss der ersten Moden nicht ganz so stark ist. Wird über jeweils p Moden summiert, so wird ersichtlich (Abb.29, unten rechts), dass das n -inkrementierende Modell mit $p < n$ immer einen höheren Informationsgehalt bzw. eine höhere Variationsbreite besitzt als das nichtinkrementierende Modell.

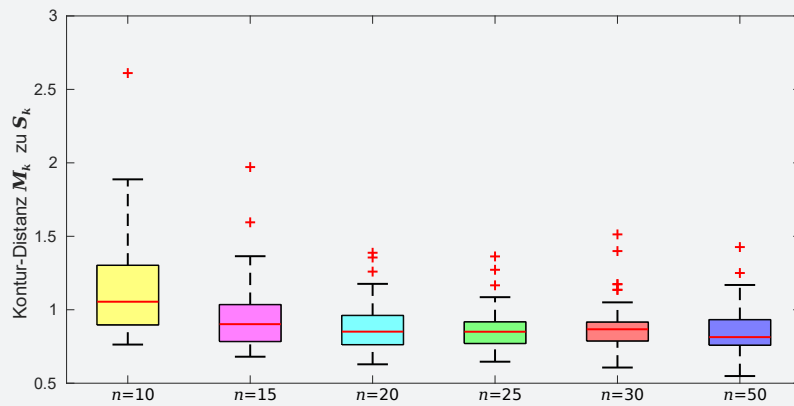


Abbildung 28: Handkonturdistancen zwischen Trainingsdaten und adaptierten Modellinstanzen für unterschiedlich gewählte n .

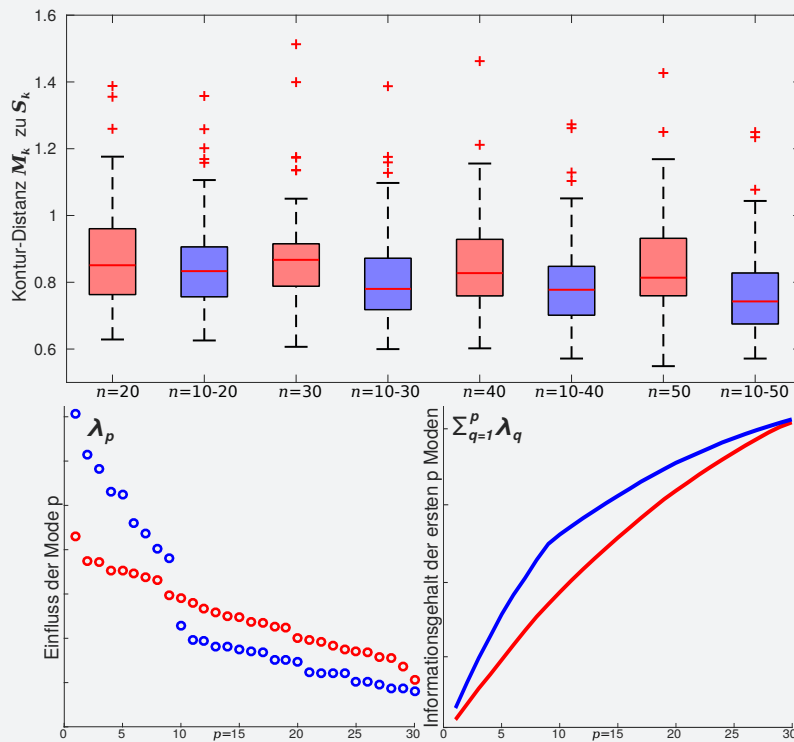


Abbildung 29: Gegenüberstellung von Distanzen (obere Boxplots) und Standardabweichungen der Moden (untere Grafiken) ohne Inkrementieren (rot) und mit Inkrementieren (blau) von n .

Parameterbelegung	
Input-/Trainingsdaten	
N	= 60
D	= 4 ($D_f = 2, D_x = 2$)
$\mathbf{f}_i$	: (1) Gauß-geglättete Signalwerte: $(G_{\sigma=6mm} * I_k)(\mathbf{x}_i)$ (2) Distanzwerte zur Handmaske: $d(\mathbf{x}_i, I_k^{Mask})$
$\mathbf{x}_i$	: kantenmerkmalsabhängig mit Mindestabstand 6 mm, innerhalb der Hand ($\forall i, f_{i,2} < 5mm$)
sp	= 6 mm
$N_s(k)$	= 450 ± 57
N_m	= 481
Modellparameter	
n	= 10 / 15 / 20 / 25 / 30 / 40 / 50 bzw. 10-20 / 10-30 / 10-40 / 10-50
$T_k \star \mathbf{s}_{k,i}$	= $(T_k(\mathbf{x}_i), \mathbf{f}_i)$ mit $(\mathbf{S}_k - T_k^{-1}(\bar{\mathbf{M}})) \rightarrow_{T_k} \min$
Σ	= $\frac{1}{N} \sum_{k=1}^N \text{diag}(\text{Cov}(\mathbf{S}_k, \mathbf{M}_k) \cdot 0.5 + \text{Cov}(\mathbf{S}_k, \bar{\mathbf{M}}) \cdot 0.5)$
$ \mathcal{N}(j) $	≈ 4
N_l	= 1
Beschränk.	- mit $p(\Theta)$; - $p(Q \Theta)$ mit Symmetriebedingung

5.3. Sukzessives Dekrementieren der Merkmalsvarianzen Σ während der Modellanpassung

Das iterative Inkrementieren der Anzahl der Merkmalsvektoren während der Modellgenerierung in Abschnitt 5.1 wird genutzt, um die korrekten Korrespondenzwerte während der Modellgenerierung robust über mehrere Level zu bestimmen. Für eine nachfolgende Anpassung des resultierenden Modells an neue Daten wird lediglich das letzte/feinste Level genutzt. Das heißt, die Modellanpassung wird nicht in einem Multi-Level-Ansatz durchgeführt. Dies ist auch nicht nötig, da bei der Modellanpassung die Freiheitsgrade bereits durch die Anzahl der Moden bzw. Modengewichte und die Anzahl der Parameter der affinen Transformation beschränkt sind.

Liegt jedoch ein Modell mit kleinem Merkmalsvektorabstand sp (relativ zur Formvariation des betrachteten Objekts) vor, so könnten die auf den Daten berechneten Merkmalsvarianzen Σ (Abschnitt 4.2) in den ersten Iterationen zu klein sein und “eigentlich korrespondierende” Merkmalsvektoren in $\mathbf{S}_{new}$ und $\bar{\mathbf{M}}$ bzw. $\mathbf{M}_{new}$ zu weit für die aktuellen Werte in Σ auseinanderliegen. Um also zu gewährleisten, dass sehr große Unterschiede ebenfalls betrachtet bzw. ausgeglichen werden, ist es sinnvoll die Werte für die berechneten Σ in den ersten Iterationen der Modellanpassung groß zu wählen, z. B. durch die Multiplikation mit einem Faktor. Ein Beispiel für einen solchen Faktor könnte $4 - \min(it, 3)$ in Iteration it sein. Nach $it = 3$ der Anpassung würde in diesem Fall der Faktor den Wert Eins annehmen und dementsprechend die ursprüngliche datengetriebene Bestimmung von Σ gelten (Abschnitt 4.2).

5.4. Modellierung lokaler Form- und Appearance-Informationen zur Erhöhung der Freiheitsgrade

Vorteil und gleichzeitig Einschränkung von modellbasierten Ansätzen ist die Beschränkung der möglichen Modellinstanzen durch die gegebenen Moden des Modells. Alle Merkmalsvektoren werden durch jede Mode beeinflusst. Finden bestimmte Bewegungen/Variationen zweier räumlich weit auseinanderliegender Regionen im Trainingsdatensatz häufig gemeinsam statt (z. B. Bewegung von Daumen und kleinem Finger), werden die Moden dies auch abbilden und eine unabhängige Bewegung beider Regionen wird mit größerer Wahrscheinlichkeit nicht modelliert.

Eine Lösung stellt die Einführung "lokaler" Moden, die ausschließlich Variationen von Teilbereichen der betrachteten Struktur modellieren, dar. Ein Anwendungsbeispiel wäre ein Modell aller Thoraxorgane, wobei ein "globales" Modell die Beziehungen zwischen den Organen beschreibt, während lokale Modelle für die Variationen jedes Organs gesondert zuständig sind.

Das globale Kriterium des vorgestellten Frameworks kann unkompliziert durch N_{loc} Kombination von Moden erweitert werden. Dabei sind bestimmte Moden $\mathbf{V}^{loc}$ (mit λ_p^{loc} und den jeweiligen Gewichten ω_k^{loc}) nur noch für eine Untermenge der Modellvektoren $j \in J^{loc}$ zuständig. Das globale Kriterium ergibt sich zu:

$$\begin{aligned}
 C(Q, \Theta) = & -\sum_{k=1}^N \left(\sum_{i=1}^{N_s(k)} \log \left(\sum_{j=1}^{N_m} \exp \left(-\frac{1}{2} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \Sigma_j^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \right) \right) \right) \\
 & + \left(\left(\sum_{loc=1}^{N_{loc}} \sum_{k=1}^N \sum_{p=1}^{n^{loc}} \log(\lambda_p^{loc}) + \left(\frac{(\omega_{k,p}^{loc})^2}{2(\lambda_p^{loc})^2} \right) \right) + \left(\frac{\frac{1}{N} \sum_{k=1}^N \|\omega_k^{loc}\|^2}{2\sigma_{\bar{\omega}}^2} \right) \right. \\
 & \left. + \left(\frac{\|(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}\|_F^2}{2\sigma_{\bar{T}}^2} \right) \right) \\
 & + \sum_{loc=1}^{N_{loc}} \sum_{j \in J^{loc}} \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^{n^{loc}} \left(\frac{1}{2} \right. \\
 & \left. \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p^{loc} \mathbf{v}_{p,j}^{loc}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p^{loc} \mathbf{v}_{p,\eta}^{loc})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \right. \\
 & \left. \Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p^{loc} \mathbf{v}_{p,j}^{loc}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p^{loc} \mathbf{v}_{p,\eta}^{loc})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right). \quad (80)
 \end{aligned}$$

Die Parameter werden demnach erweitert zu:

$$\Theta = \{\bar{\mathbf{M}}, \mathbf{V}^{loc}, \lambda_p^{loc}, n^{loc}, \Sigma, J^{loc} | p = 1, \dots, n^{loc}, loc = 1, \dots, N_{loc}\}$$

und

$$Q_k = \{T_k, \omega_k^{loc} | loc = 1, \dots, N_{loc}\}$$

für Instanz $\mathbf{S}_k, k = 1, \dots, N$. Die Optimierung aller Parameter findet weiterhin wie in Abschnitt 4.1 beschrieben statt. Die zusätzliche Information, die benötigt wird, ist lediglich die Untermenge an Merkmalsvektoren J^{loc} , die durch die jeweiligen Moden $\mathbf{V}^{loc}$ variiert werden. Diese Information wird nicht für alle Trainingsdaten benötigt, sondern ausschließlich für die Modellvektoren bzw. für die eine Trainingsinstanz $\mathbf{S}_x$ (mit $\bar{\mathbf{M}} = \mathbf{S}_x$, siehe Abschnitt 4.3 zu Initialisierung). Die neuen Modellinstanzen ergeben sich zu:

$$\mathbf{M}_k \text{ mit } \mathbf{m}_{k,j} = T_k^{-1} \star (\bar{\mathbf{m}}_j + \sum_{loc=1}^{N_{loc}} \sum_{p=1}^{n^{loc}} \omega_{k,p}^{loc} \mathbf{v}_{p,j}^{loc}), \text{ für } j \in J^{loc}. \quad (81)$$

Es ist zu beachten, dass die Regularisierung über die Nachbarn in $p(\Theta)$ weiterhin über alle Nachbarn jedes Merkmalsvektors stattfindet, auch wenn die Nachbarn nicht in der jeweiligen Untermenge J^{loc} liegen, was zu glatten Übergängen zwischen “partitionierten” Moden führt.

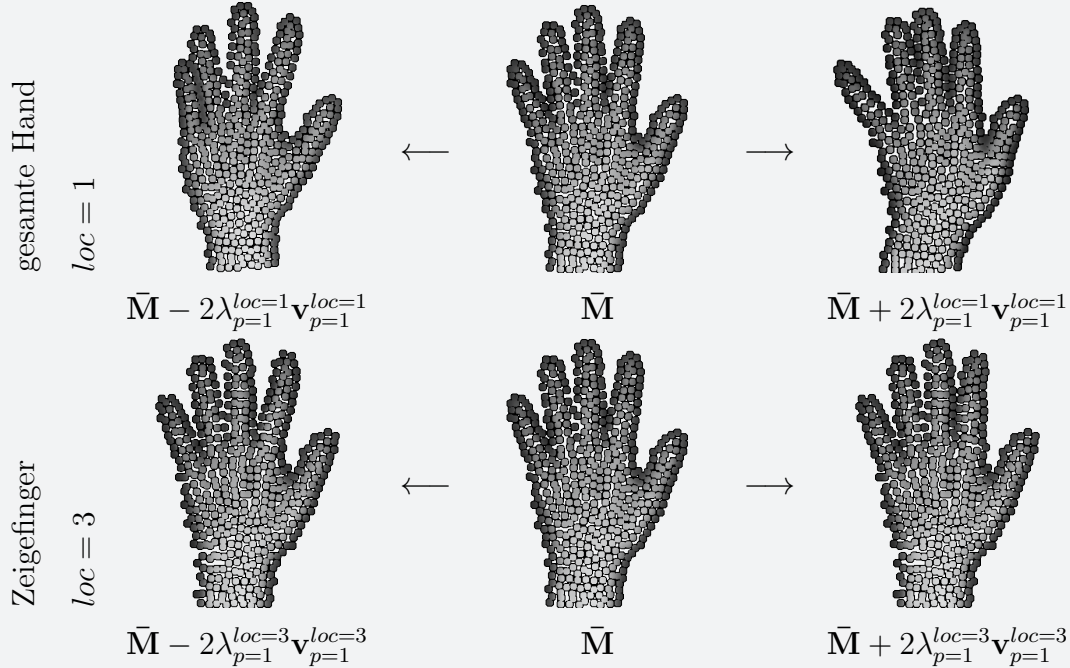
Optimierung von globalen Moden vor lokalen Moden: Mit dem oben gegebenen globalen Kriterium würden alle Moden abhängig voneinander optimiert werden. Analog zur Frage, ob T_k oder ω_k zuerst bzw. gleichberechtigt optimiert werden sollte (siehe Abschnitt 4.1.7), kann es auch in diesem Fall sinnvoll sein, bestimmte $\mathbf{V}^{loc}$ zuerst zu optimieren. Beim genannten Beispiel der Thoraxorgane würde es z.B. sinnvoll sein, wenn ein $\mathbf{V}^{loc=1}$ mit $j \in J^1 = [1, \dots, N_m]$ alle Merkmalsvektoren umfasst und somit für die “globalen” Variationen und die restlichen $\mathbf{V}^{loc=2, \dots, N_{loc}}$ entsprechend für jeweils ein Organ zuständig sind. Um zu gewährleisten, dass die lokalen Moden (für nur ein Organ) nicht die Aufgabe der globalen Variationsmoden übernehmen, würde $\mathbf{V}^{loc=1}$ unabhängig von den restlichen loc optimiert werden (mit $\mathbf{V}^{loc=2, \dots, N_{loc}}$ gleich Null in der Ableitung nach $\mathbf{V}^{loc=1}$). Daraus folgt, dass die globalen $\mathbf{V}^{loc=1}$ “zuerst” optimiert werden und lediglich die “Restdistanz” zwischen $\mathbf{S}_k$ und $T_k^{-1}(\bar{\mathbf{M}} + \omega_k^{loc=1} \mathbf{V}^{loc=1})$ durch $\mathbf{V}^{loc=2, \dots, N_{loc}}$ modelliert wird, da diese abhängig (mit $\mathbf{V}^{loc=1} \neq 0$) von allen loc optimiert werden.

Beispielbox 11: Lokale Moden

Das Handmodell kann z. B. in sieben Moden-Kombinationen unterteilt werden. Dabei würden die ersten Moden $\mathbf{V}^{loc=1}$ (mit $n^{loc=1} = 15$) alle Merkmalsvektoren umfassen. Das Modell wäre so mit allen vorangegangenen Beispielen identisch. Zusätzliche Moden $\mathbf{V}^{loc=2,\dots,7}$ (mit $n^{loc=2,\dots,7} = 3$) könnten für lokale Variationen der Handfläche bzw. aller Finger zuständig sein. Das rechte Bild visualisiert die Partitionierung des Modells bzw. der Trainingsinstanz $\mathbf{S}_x$. Da die Mengen $J^{loc=2,\dots,7}$ entsprechend dieser Partitionierung disjunkt sind, umfassen alle Moden des Modells $\{\mathbf{V}^{loc}\}$ genauso viele Parameter wie ein einziges globales $\mathbf{V}$ mit $n = 15 + 3 = 18$. Die Anzahl der Freiheitsgrade, die durch ω_k gegeben ist, steigt allerdings von 18 auf $\sum_{loc=1}^{N_{loc}} n^{loc} = 1 \cdot 15 + 6 \cdot 3 = 33$. Die folgenden Bilder zeigen die jeweils ersten Moden für $loc = 1, 3$ und 6 (gesamte Hand, Zeigefinger, kleiner Finger).



Abbildung 30: Partitionierung des Modells bzw. der Trainingsinstanz $\mathbf{S}_x$. Da die Mengen $J^{loc=2,\dots,7}$ entsprechend dieser Partitionierung disjunkt sind, umfassen alle Moden des Modells $\{\mathbf{V}^{loc}\}$ genauso viele Parameter wie ein einziges globales $\mathbf{V}$ mit $n = 15 + 3 = 18$. Die Anzahl der Freiheitsgrade, die durch ω_k gegeben ist, steigt allerdings von 18 auf $\sum_{loc=1}^{N_{loc}} n^{loc} = 1 \cdot 15 + 6 \cdot 3 = 33$. Die folgenden Bilder zeigen die jeweils ersten Moden für $loc = 1, 3$ und 6 (gesamte Hand, Zeigefinger, kleiner Finger).



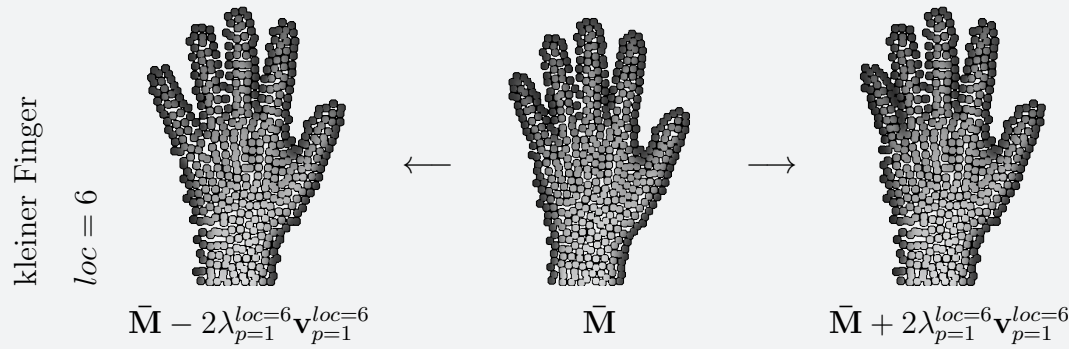


Abbildung 31: Visualisierung von drei Beispielen der sieben lokalen Moden – $loc = 1$ (gesamte Hand), 3 (Zeigefinger) und 6 (kleiner Finger).

Parameterbelegung

Input-/Trainingsdaten

N	=	60
D	=	4 ($D_f = 2, D_x = 2$)
$\mathbf{f}_i$	:	(1) Gauß-geglättete Signalwerte: $(G_{\sigma=6mm} * I_k)(\mathbf{x}_i)$ (2) Distanzwerte zur Handmaske: $d(\mathbf{x}_i, I_k^{Mask})$
$\mathbf{x}_i$	:	kantenmerkmalsabhängig mit Mindestabstand 6 mm, innerhalb der Hand ($\forall i, 2 < 5mm$)
sp	=	6 mm
$N_s(k)$	=	450 ± 57
N_m	=	481

Modellparameter

N_{loc}	=	7
n_{loc}	:	$n^{loc=1} = 15, n^{loc=2, \dots, 7} = 3$
$T_k * \mathbf{s}_{k,i}$	=	$(T_k(\mathbf{x}_i), \mathbf{f}_i)$ mit $(\mathbf{S}_k - T_k^{-1}(\bar{\mathbf{M}})) \rightarrow_{T_k} \min$
Σ	=	$\frac{1}{N} \sum_{k=1}^N \text{diag}(\text{Cov}(\mathbf{S}_k, \mathbf{M}_k) \cdot 0.5 + \text{Cov}(\mathbf{S}_k, \bar{\mathbf{M}}) \cdot 0.5)$
$ \mathcal{N}(j) $	$\approx$	4
N_l	=	1
Beschränk.	-	mit $p(\Theta)$; - $p(Q \Theta)$ mit Symmetriebedingung

6. Anwendungsbereiche für probabilistische Appearance-Modelle

Die klassische Anwendung für statistische Form- oder Appearance-Modelle ist die Segmentierung des modellierten Objekts in neuen Bildern. Dabei ist die gesuchte Objektsegmentierung dadurch gegeben, dass die Landmarken bei klassischen Formmodellen explizit auf der Objektkontur platziert sind. Da im vorgestellten Modell die Wahl der Merkmalsvektoren objektunabhängig erfolgen kann, muss die Information über die Objektkontur als ein Appearance-Wert pro Merkmalsvektor des Modells mit modelliert werden. Die Beschreibung der gesuchten Informationen über Appearance-Werte anstelle von Landmarkenpositionen hat außerdem den Vorteil, dass die Modellanwendung nicht nur auf Segmentierungsprobleme beschränkt ist. Es können beliebig viele weitere Merkmale modelliert werden, die auf das neue Bild übertragen werden sollen. Das Modell kann zur

- Segmentierung von Objektkonturen,
- Detektion von Landmarken,
- Multi-Objektsegmentierung bzw. Übertragung von beliebig vielen Labeln auf ein neues Bild oder
- Rekonstruktion von fehlenden Bildkanälen in multispektralen Bilddaten (z. B. fehlende Sequenz in MRT-Bildern)

eingesetzt werden, ohne dass zusätzliche Erweiterungen des Modellansatzes nötig sind (Abschnitt 6.2, 6.3, 6.5). Allerdings ist für z. B. die Segmentierungsanwendung eine komplette Rekonstruktion der spärlich verteilten Merkmalsvektoren des Modells für das gesamte Eingabebild notwendig (Abschnitt 6.1).

Darüber hinaus liefert die Nutzung der Korrespondenzwahrscheinlichkeiten weitere Vorteile für die Anwendung des Modells: Zum einen wird im Gegensatz zu klassischen Modellen nicht von Eins-zu-Eins-Korrespondenzen zwischen Modell und neuem Bild ausgegangen. Das bedeutet, die Modellanpassung ist relativ robust gegenüber nicht vorhandenen korrespondierenden Bereichen – auf Grund von physiologischen Unterschieden, Pathologien oder Artefakten im neuen Bild. Zum anderen liefern die probabilistischen Korrespondenzen zusätzliche Informationen über Unsicherheiten des Modells. Diese können z. B. für Klassifikationsprobleme von pathologischen Bereichen eingesetzt werden. Damit ergeben sich weitere mögliche Anwendungen (Abschnitt 6.6, 6.7 und 6.8) wie unter anderem:

- robuste Modellanpassung bei fehlenden Korrespondenzen bzw. “fehlerhaften Bereichen” im neuen Bild,
- Identifikation von lokalen Unsicherheiten des Modells,

- Klassifikation von pathologischen Regionen oder
- Rekonstruktion von “fehlerhaften Bereichen” in Bildern.

Im Folgenden werden unterschiedliche Anwendungen des vorgestellten Modells beschrieben.

6.1. Rekonstruktion von Bildinformationen aus spärlich verteilten Merkmalsvektoren

Die Grundlage des vorgestellten Modellansatzes bildet die spärliche merkmalsvektor-basierte Bildrepräsentation der Daten (Abschnitt 2). Je nach Anwendungsszenario ist jedoch nach einer kompletten Rekonstruktion der Bildinformationen für das Eingabebild I_{new} an allen Pixelpositionen gefragt. Damit kann die komplette Modellanpassung in drei große Schritte unterteilt werden: (1) Abtastung des neuen Eingabebildes I_{new} (bzw. mehrkanal Bildes $\mathbf{F}_{new}$) zu $\mathbf{S}_{new}$, (2) Modellanpassung auf $\mathbf{S}_{new}$ und schließlich (3) Rekonstruktion der merkmalsvektorbasierte Modellinstanz $\mathbf{M}_{new}$ zu einem kompletten (mehrkanal) Bild $\mathbf{F}_{new}^{rec}$:

$$\begin{aligned}
 \mathbf{F}_{new} : \Omega \rightarrow \mathbb{R}^{D_f^{in}}, \Omega \in \mathbb{R}^{D_x} &\xrightarrow{\text{Abtastung}} \mathbf{S}_{new} = \{\mathbf{s}_{new,i} | i = 1, \dots, N_s\} \\
 &\mathbf{s}_{new,i} = (\mathbf{x}_i \in D_x, \mathbf{f}_i \in D_f^{in}) \\
 \mathbf{S}_{new} &\xrightarrow{\text{Modellanpassung}} \mathbf{M}_{new} \\
 \mathbf{M}_{new} = \{\mathbf{m}_{new,j} | j = 1, \dots, N_m\} &\xrightarrow{\text{Rekonstruktion}} \mathbf{F}_{new}^{rec} : \Omega \rightarrow \mathbb{R}^{D_f^{out}}, \Omega \in \mathbb{R}^{D_x} \\
 \mathbf{m}_{new,j} &= (\mathbf{x}_j \in D_x, \mathbf{f}_j \in D_f^{out}).
 \end{aligned}$$

Die Merkmalsanzahl des Eingabebildes (D_f^{in}) und die des Ausgabebildes (D_f^{out}) kann je nach Anwendung unterschiedlich groß sein. In Abschnitt 6.2 wird dies beispielsweise für die Segmentierungsanwendung diskutiert.

Die Rekonstruktion der Merkmalsvektoren zu einem gesamten Bild kann durch Interpolation der Appearance-Werte zwischen den Positionen der Vektoren erfolgen. Die meisten einfacheren Verfahren wie lineare, polynomiale oder B-Spline-basierte Interpolationen wurden für die Rekonstruktion der Werte zwischen Vektoren mit regulärem Abstand (Werte auf einem Grit) entwickelt. Für nichtregulär verteilte Werte wurden Verfahren wie das “Kriging”/Gaußprozess-Regression [Stytz & Parrott 1993] oder die “Scattered Data B-Spline Interpolation” von Lee et al. [Lee *et al.* 1997] entwickelt. Da Kriging relativ zeitaufwändig ist, wird in dieser Arbeit die Scattered Data B-Spline Interpolation genutzt. Dabei stellen die Merkmalsvektoren die verstreuten Datenvektoren (engl. scattered data) dar. Die Qualität des rekonstruierten Bildes hängt von der Art der interpolierten Merkmale und vom Abstand zwischen den Merkmalsvektoren ab. In Abschnitt 6.2.1 wird die Rekonstruktion von der Kontur eines zu segmentierenden Objekts diskutiert. In der Beispielbox 12 erfolgt eine quantitative Auswertung, welche zeigt, dass

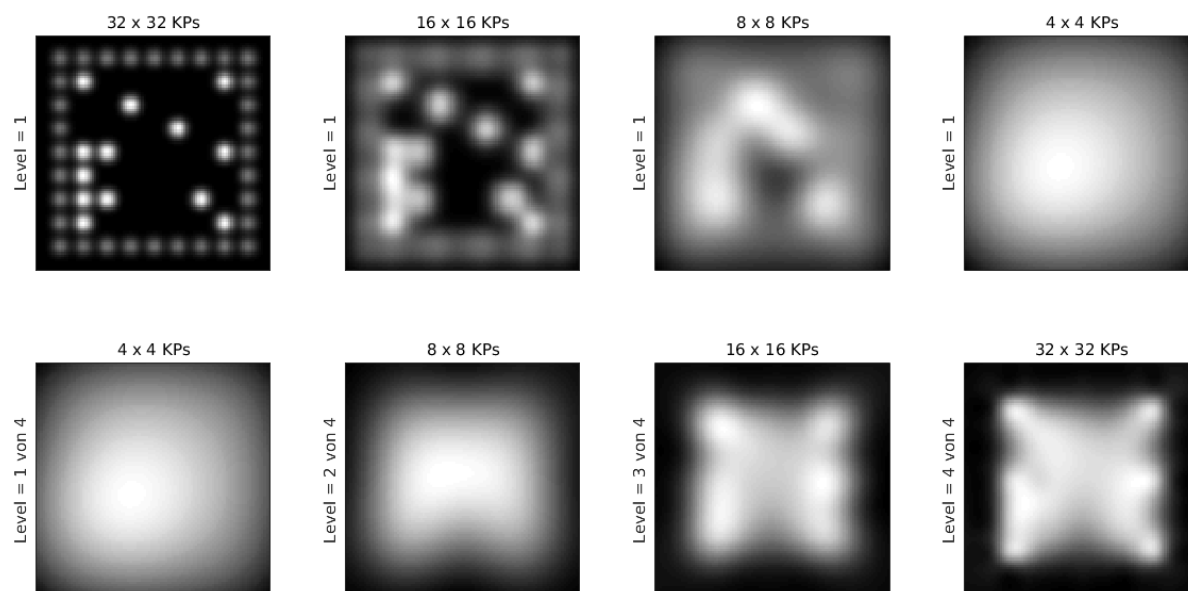


Abbildung 32: Interpolation von spärlich und unregelmäßig verteilten Datenpunkten mittels Multi-Level-B-Spline-Approximation: Die hellen Punkte (oben links), die ein Quadrat im Bild beschreiben sollen, sind unregelmäßig und teilweise sehr spärlich im Bild verteilt. Wird eine einfache B-Spline-Approximation auf diese Datenpunkte angewendet (obere Zeile), ist es nicht möglich die “richtige” Anzahl an Kontrollpunkten zu bestimmen – werden zu viele (mit zu geringem Abstand) genutzt (32×32 oder 16×16), so ist der Einflussbereich der jeweiligen Kontrollpunkte zu klein und die Lücken im Bild werden nicht geschlossen. Bei zu geringer Anzahl (4×4) wird die Interpolation jedoch zu grob. Es gibt also einen Tradeoff zwischen Glattheit und Genauigkeit der Approximation. Die Multi-Level-B-Spline-Approximation (untere Zeile) behandelt dieses Problem, indem in den ersten Leveln die “großen Lücken” mit wenigen Kontrollpunkten geschlossen werden (für eine glatte Approximation) und in höheren Leveln die Approximation genauer mit kleinerem Wirkungsbereich der Kontrollpunkte an die vorhandenen Datenpunkte angepasst wird (für die Genauigkeit). (Anmerkung: Es werden an dieser Stelle nur die Anzahl der Kontrollpunkte, die im Bildbereich liegen, genannt. Pro Dimension liegen jeweils drei weitere Kontrollpunkte außerhalb des Bildbereichs.)

die Qualität dieser Rekonstruktion ausreicht, um eine Objektkontur mit einer Genauigkeit von weit unter 1 *mm* zu bestimmen – mit einem Merkmalsvektorabstand von bis zu 6 *mm*. Alle Arten von Appearance-Merkmalen können unter Nutzung der Interpolation rekonstruiert werden.

Scattered Data B-Spline-Interpolation: Die von Lee et al. [Lee et al. 1997] vorgestellte Interpolation adressiert das Problem der Anpassung einer glatten Funktion/“Oberfläche” an unregelmäßig verstreute Datenpunkte. Die Methode basiert dabei auf der Multi-Level-B-Spline-Approximation. Die Standard-B-Spline-Interpolationen für Bilder nutzt eine Menge an N_c gewichteten Kontrollpunkten, mit homogenen Koordinaten (NURBS, “nicht-uniforme rationale B-Splines”). Sind für ein Bild I die Signalwerte an den Koordinaten $\{\mathbf{x}^{geg}\}$ gegeben, so ist die gesuchte Fläche, welche $I(\mathbf{x}) \forall \mathbf{x} \in \Omega$ interpolieren soll, über die Summe der mit rationalen B-Spline-Basisfunktionen (*ratB*) gewichteten Kontrollpunkte $\mathbf{P}_c, c = 1, \dots, N_c$ definiert:

$$I(\mathbf{x}^{geg}) = \sum_{c=1}^{N_c} \text{rat}B_{c,n}(\mathbf{x}^{geg}) \mathbf{P}_c,$$

mit $\text{rat}B_{c,n}(\mathbf{x}) = \frac{B_{c,n}(\mathbf{x})w_c}{\sum_{c=1}^{N_c} B_{c,n}w_c}$ als rationale B-Spline-Basis-Funktion, die mit w_c pro Kontrollpunkt c gewichtet wird. Die Glattheit und Genauigkeit der Approximation ist dabei durch den Grad n der B-Spline-Basisfunktion $B_{.,n}$ und die Anzahl bzw. den Abstand der Kontrollpunkte zueinander gegeben. In dieser Arbeit werden kubische ($n=3$) B-Spline-Funktionen genutzt. Da für die Approximation des Wertes an Koordinate $\mathbf{x}^{geg}$ jeweils $n + 1$ Kontrollpunkte pro Dimension genutzt werden, hat jeder Kontrollpunkt nur einen beschränkten “Einflussbereich” (siehe [Lee et al. 1997] für Details). Das bedeutet, je geringer der Abstand zwischen den Kontrollpunkten im Vergleich zu dem Abstand der gegebenen Daten $\mathbf{x}^{geg}$ ist, umso schlechter können die Signalwerte ($\forall \mathbf{x}$) zwischen den gegebenen Werten interpoliert werden. Bei größerem Abstand wird die Interpolation glatter. Für gegebene Daten, die sehr unregelmäßig im Bildraum verteilt sind (mit unregelmäßigen Abständen zueinander), ist es schwer bzw. nicht möglich einen “korrekten” Kontrollpunkt Abstand zu bestimmen. Abbildung 32 (obere Zeile) visualisiert das Problem des Tradeoffs zwischen Glattheit und Genauigkeit für solche Daten. Um dieses Problem zu lösen, schlagen Lee et al. einen Multi-Level-Ansatz vor, bei dem in den ersten Leveln die B-Spline-Approximation mit wenigen, weit auseinanderliegenden Kontrollpunkten begonnen wird, um “große Lücken” zu schließen, und anschließend in höheren Leveln die Fläche immer genauer an die vorhandenen Bildpunkte angepasst wird (Abbildung 32, untere Zeile) [Lee et al. 1997]. In [Krüger et al. 2014] wurde beispielsweise die Scattered Data B-Spline-Interpolation genutzt, um ein Deformationsfeld für ein komplettes Bild basierend auf nur lokal vorhandenen Deformationswerten zwischen Oberflächen zu interpolieren.

Diese Art der Interpolation der Daten hat den Vorteil, dass sie unabhängig von der

Verteilung der gegebenen Daten ist. Da die Merkmalsvektoren, der in der Arbeit verwendeten Bildrepräsentationen, ebenfalls beliebig im Bild verteilt sein können, bietet sich die Nutzung dieses Verfahrens zur Rekonstruktion der gesamten Bilder an. Ein sinnvoller minimaler Kontrollpunkt Abstand ist durch den minimalen Abstand der Merkmalsvektoren sp zueinander gegeben. In der Arbeit werden sechs Level mit Kontrollpunkt Abstand von $sp \cdot 2^{6-l}$ für Level $l = \{1, \dots, 6\}$ genutzt (jedes Level halbiert den Abstand, siehe [Lee *et al.* 1997]).

6.2. Objektsegmentierung

Um das Modell für Segmentierungszwecke zu nutzen, müssen die gesuchten Kontur- oder Landmarkeninformationen bei der Modellgenerierung als ein Merkmal mit modelliert werden. Das heißt, zu jedem Bild I_k im Trainingsdatensatz liegt die gegebene Maske $I_k^{Mask} \in \{0, 1\}$ vor, mit $I_k^{Mask}(\mathbf{x}) = 1$, wenn $\mathbf{x}$ innerhalb des Objekts liegt (oder auf der Kontur/Landmarke) und $I_k^{Mask}(\mathbf{x}) = 0$ andernfalls. Sind mehrere Objektmasken in I_k^{Mask} beschrieben, werden diese meist durch unterschiedliche Werte (Labelwerte $LV = \{0, 1, 2, 3, \dots\}$) pro Objekt in sogenannten Labelbildern repräsentiert. Diese Werte könnten z.B. direkt als Merkmale für jeden Merkmalsvektor genutzt werden ($f_{i,:}^{seg} = I_k^{Label}(\mathbf{x}_i)$). Die Genauigkeit der Rekonstruktion solcher Merkmale zu einem gesamten Bild hängt in diesem Fall vom Merkmalsvektorabstand ab. Um zu gewährleisten, dass auch Merkmalsvektoren, die nicht auf bzw. in der Nähe der gesuchten Kontur liegen, Informationen über deren Position beinhalten, wird die euklidische Distanz vom jeweiligen Vektor zur Kontur (Distanzkartenwerte) als Merkmal genutzt. Pro Objekt (Labelwert LV) im Bild wird ein Merkmal $f_{i,LV}^{seg}$ bestimmt mit:

$$f_{i,LV}^{seg} = d(\mathbf{x}_i, I_k^{Label}) = \begin{cases} + \min_{\mathbf{x} \in \Omega} d(\mathbf{x}_i, \mathbf{x}) & \text{if } I_k^{Label}(\mathbf{x}_i) \neq LV \wedge I_k^{Label}(\mathbf{x}) = LV \\ - \min_{\mathbf{x} \in \Omega} d(\mathbf{x}_i, \mathbf{x}) & \text{if } I_k^{Label}(\mathbf{x}_i) = LV \wedge I_k^{Label}(\mathbf{x}) \neq LV, \end{cases} \quad (82)$$

mit $d(.,.)$ als euklidische Distanz zwischen zwei Vektoren. Zu einem Merkmalsvektor können beliebig viele “Segmentierungsmerkmale” hinzukommen: $\mathbf{s}_i = (\mathbf{x}_i, \mathbf{f}_i, \mathbf{f}_i^{seg}) \in \mathbb{R}^{(D_x + D_f + D_f^{seg})}$.

Da es keine Beschränkung für die Merkmalsanzahl pro Vektor gibt, können auf diese Weise beliebig viele Distanzkarten mit modelliert werden und somit ist die Multi-Objektsegmentierung ohne weitere Methodenanpassung möglich.

Während der Modellgenerierung werden alle Merkmale wie bisher gemeinsam betrachtet, ohne dass am vorgestellten Algorithmus Änderungen auftreten. Wird das Modell an ein neues Bild angepasst, ist zu beachten, dass die gesuchten Merkmale $\mathbf{f}_i^{seg}$ für $\mathbf{S}_{new}$ nicht bekannt sind. Das heißt, das Modell mit Modellinstanz $\mathbf{M}_{new}$ mit $\mathbf{m}_{new,j} = (\mathbf{x}_j, \mathbf{f}_j, \mathbf{f}_j^{seg})$ wird an die Instanz $\mathbf{S}_{new}$ mit $\mathbf{s}_{new,i} = (\mathbf{x}_i, \mathbf{f}_i)$ angepasst. Auch dies ändert nicht die vorgestellte Modellanpassung, außer dass ω_{new} und T_{new} mithilfe eines Untervektors von $\mathbf{m}_{new,j}/\bar{\mathbf{m}}_j$ ohne $\mathbf{f}_j^{seg}$ berechnet werden. Die resultierende Modellinstanz

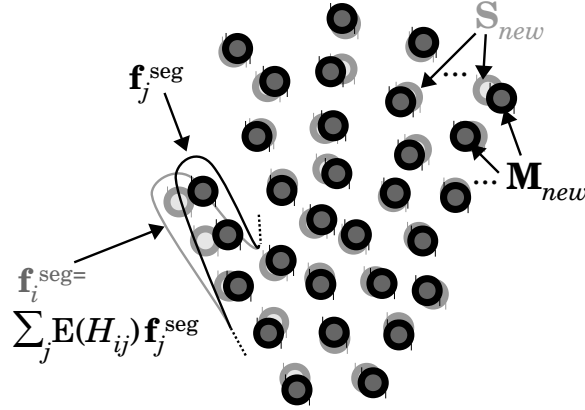


Abbildung 33: Berechnung der fehlenden Segmentierungsmerkmale für $\mathbf{S}_{new}$: Nach abgeschlossener Modellanpassung repräsentieren die Werte in $\mathbf{f}_j^{seg}$ die Segmentierungsergebnisse (schwarze Linie), welche durch die Modellinstanz $\mathbf{M}_{new}$ gegeben sind. Wird das Modell nicht perfekt an $\mathbf{S}_{new}$ angepasst (im Bereich des Daumens) – da die Variationsmoden dies nicht abdecken – können trotzdem genauere Segmentierungsergebnisse für $\mathbf{S}_{new}$ erreicht werden: Durch $\mathbf{f}_i^{seg} = \sum_{j=1}^{N_m} \mathbb{E}(H_{ij}) \mathbf{f}_j^{seg}$ werden die Segmentierungsmerkmale (graue Linie) für $\mathbf{S}_{new}$ aus $\mathbf{M}_{new}$ basierend auf den Korrespondenzbeziehungen rekonstruiert. In diesem Fall bedeutet dies, dass die Korrespondenzwahrscheinlichkeiten die “Daumen-Merkmalvektoren” korrekt zuordnen, auch wenn die Moden eine genauere Anpassung nicht zulassen.

$\mathbf{M}_{new} = T_{new}^{-1}(\bar{\mathbf{M}} + \omega_{new} \mathbf{V})$ umfasst wieder alle Merkmale.

Anschließend werden die fehlenden Werte $\mathbf{f}_i^{seg}$ in $\mathbf{S}_{new}$ unter Verwendung der Korrespondenzwahrscheinlichkeiten zwischen $\mathbf{S}_{new}$ und $\mathbf{M}_{new}$ und den Segmentierungsmerkmalen der Modellinstanz $\mathbf{f}_j^{seg}$ berechnet:

$$\mathbf{f}_i^{seg} = \sum_{j=1}^{N_m} \mathbb{E}(H_{ij}) \mathbf{f}_j^{seg}. \quad (83)$$

Diese Berechnung der $\mathbf{f}_i^{seg}$ hat den Vorteil, dass dadurch genauere Segmentierungsergebnisse erzielt werden können. In Abbildung 33 wird z. B. der Fall visualisiert, dass nach abgeschlossener Modellanpassung $\mathbf{M}_{new}$ im Bereich des Daumens nicht genau an $\mathbf{S}_{new}$ angepasst werden konnte – weil die Moden dies nicht zulassen. Dies bedeutet, die Segmentierungsmaske, die durch $\mathbf{f}_j^{seg}$ repräsentiert wird (schwarze Linie), segmentiert $\mathbf{S}_{new}$ nicht “perfekt”. Werden die Korrespondenzwahrscheinlichkeiten “zwischen den Merkmalsvektoren der Daumen” allerdings korrekt zugeordnet, können darüber die Werte $\mathbf{f}_i^{seg}$ (graue Linie) berechnet werden, die eine genauere Segmentierungsmaske für $\mathbf{S}_{new}$ repräsentieren.

Um die einzelnen Merkmalsvektoren als Segmentierungsergebnis für das gesamte Bild I_{new} interpretieren zu können, erfolgt als letzter Schritt eine Rekonstruktion einer kompletten Distanzkarte bzw. der Maske/des Labelbildes $I_{new}^{Mask/Label,rec}$ (Abschnitt 6.2.1). Im

Abschnitt 8.4 werden Experimente durchgeführt, welche die Segmentierungsergebnisse für die Hand sowohl mit den aus $\mathbf{f}_j^{seg}(\mathbf{M}_{new})$ rekonstruierten Masken als auch mit den aus $\mathbf{f}_i^{seg}(\mathbf{S}_{new})$ rekonstruierten Masken evaluieren.

In Abbildung 34 (obere Hälfte) wird der Segmentierungsprozess zusammengefasst.

6.2.1. Rekonstruktion einer Objektkontur

Sind die Merkmals- bzw. Maskenwerte des kompletten Bildraums gefragt, müssen die Werte zwischen den gegebenen Merkmalsvektoren interpoliert werden. Da bei der Segmentierung vor allem nach der Objektkontur gefragt wird, ist die Interpolation der Distanzwerte ausschlaggebend. Der Verlauf dieser Werte in einem Bild ist linear. Dadurch können Distanzwerte sehr gut durch Interpolation (Scattered Data B-Spline Interpolation, siehe Abschnitt 6.1) rekonstruiert werden. Werden die gerichteten Werte der Distanzkarten genutzt, die sich negativ innerhalb und positiv außerhalb der gesuchten Maske befinden (Gleichung 84), ergibt sich die gesuchte Objektkontur aus dem Nulldurchgang der rekonstruierten Werte. Daraus entsteht der Vorteil, dass die Merkmalsvektoren nicht auf der Kontur liegen müssten, um diese dennoch exakt beschreiben zu können.

6.3. Landmarkendetektion

Mithilfe der Distanzkarten können nicht nur Objektkonturen, sondern auch Landmarken im Bild modelliert werden. In dem Fall würden die Eingabebilder ($I_k^{Landmarken}$) lediglich an den Positionen der gesuchten Landmarken den Wert Eins annehmen. Die resultierenden Werte der Distanzkarten nehmen dementsprechend nur positive Werte und Null an den Landmarkenpositionen selbst an:

$$f_{i,:}^{seg} = d(\mathbf{x}_i, I_k^{Landmarken}) = \begin{cases} + \min_{\mathbf{x} \in \Omega} d(\mathbf{x}_i, \mathbf{x}) & \text{if } I_k^{Landm.}(\mathbf{x}_i) = 0 \wedge I_k^{Landm.}(\mathbf{x}) = 1 \\ 0 & \text{if } I_k^{Landm.}(\mathbf{x}_i) = 1 \end{cases} \quad (84)$$

Auf diese Weise können beliebig viele Landmarken durch ein zusätzliches Merkmal beschrieben werden. Liegen die Landmarken im Vergleich zum Merkmalsvektorabstand sp weit auseinander, ist die gemeinsame Beschreibung aller Landmarken durch eine Distanzkarte sinnvoll, um die Merkmalsdimension pro Vektor so klein wie möglich zu halten. In diesem Fall liegen "genug" Merkmalsvektoren zwischen den jeweiligen Landmarken, um diese ausreichend genau zu beschreiben. Bei geringerem Abstand zwischen den Landmarken, können mehrere Distanzkarten genutzt werden, um jeweils Teilmengen (mit größerem Abstand) der Landmarken zu beschreiben.

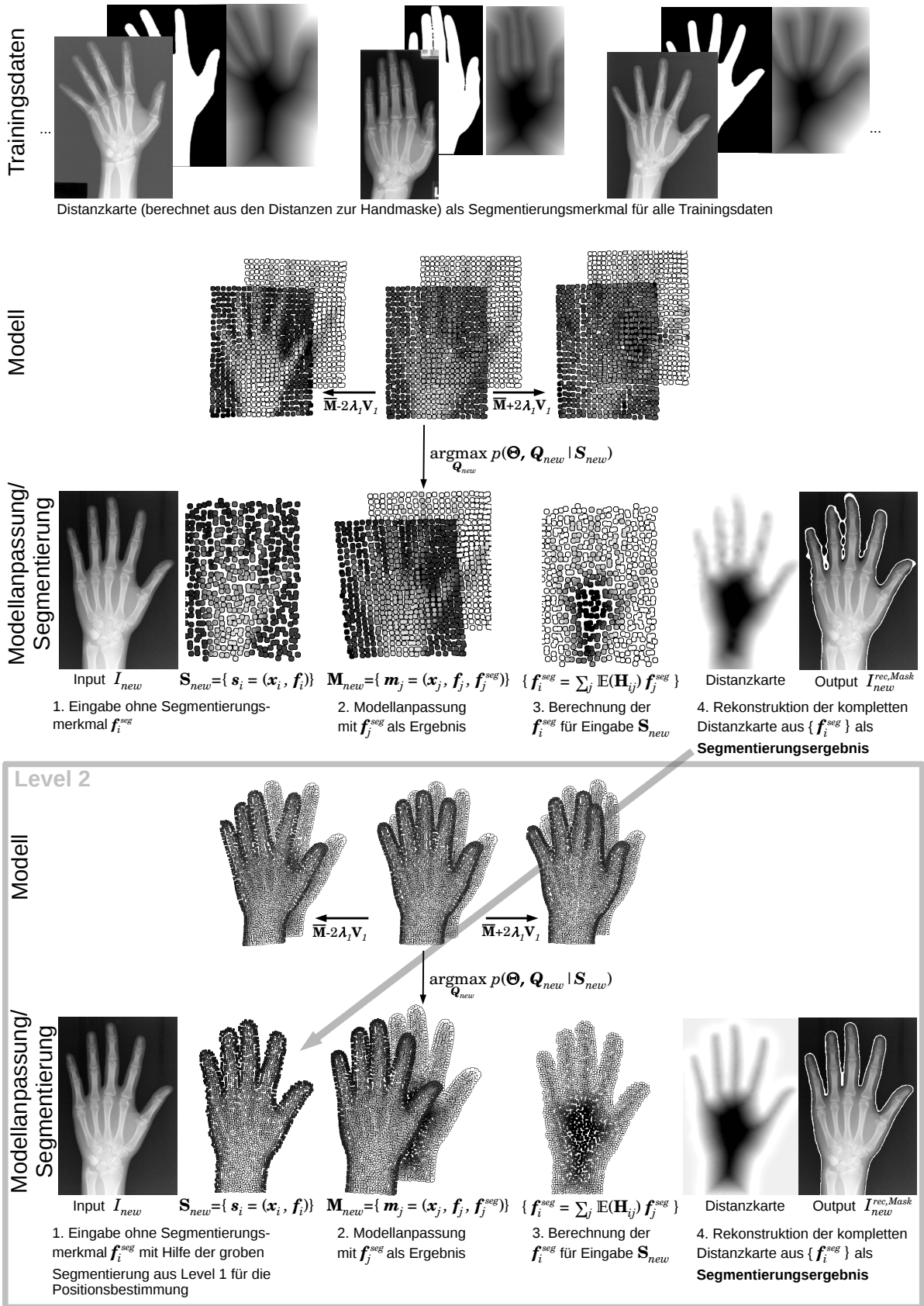


Abbildung 34: Segmentierungsprozess mit zwei Leveln: das grobe Segmentierungsergebnis des ersten Modells wird genutzt, um die Merkmalsvektoren der Eingabe S_{new} im zweiten Level/Modell auf das Innere der Handfläche zu beschränken (Abschnitt 6.4).

6.3.1. Rekonstruktion von Landmarken

Distanzkarten haben zwar den Vorteil, dass die Distanz zur gesuchten Struktur von jedem Merkmalsvektor aus bekannt ist. Die Richtung, in welcher diese Struktur liegt, ist jedoch unbekannt. Bei der Modellierung von Landmarken gibt es keine negativen Distanzwerte, sodass der Nulldurchgang nicht als Zielstruktur definiert werden kann, wie im vorangegangenen Abschnitt 6.2.1. In diesem Fall kann eine naive Rekonstruktion nur exakt sein, wenn sich Merkmalsvektoren direkt auf den gesuchten Positionen (Landmarken) befinden. Um dieses Problem zu umgehen, muss Vorwissen über die zu interpolierenden Werte in die Rekonstruktion einfließen. Da es sich um einfache euklidische Distanzen zur gesuchten Struktur handelt, kann eine Hough-Transformation für Kreise eingesetzt werden: Die Position eines Merkmalsvektors beschreibt dabei das Zentrum und der Distanzwert beschreibt den Radius des Kreises, auf dem die gesuchte Landmarke liegen muss. Wie bei der klassischen Hough-Transformation befindet sich die gesuchte Struktur (gesuchte Landmarke) an den Positionen im Bild, in denen sich die meisten Kreise schneiden.

Beispielbox 12: Rekonstruktion von Kontur- und Landmarkendaten

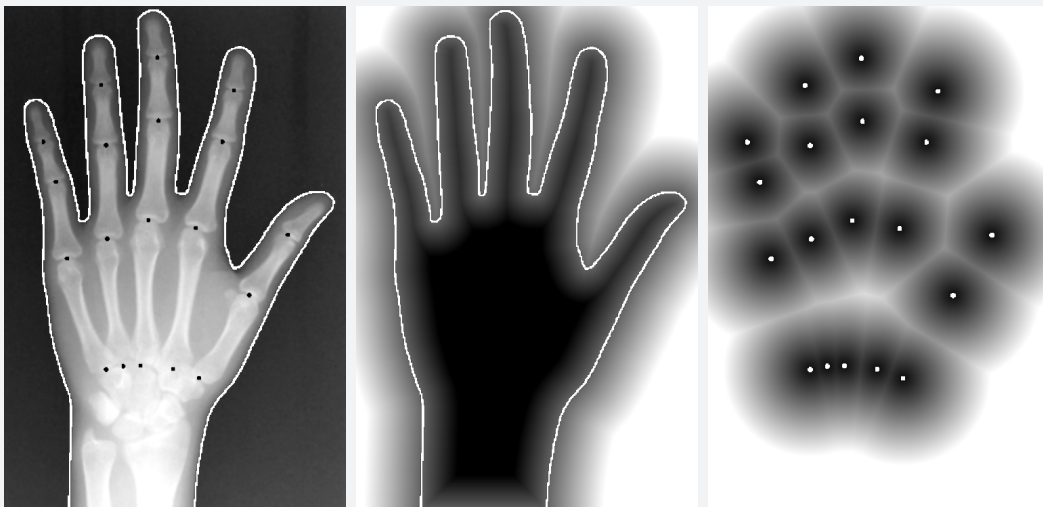


Abbildung 35: Distanzkarten für die Handkontur (Mitte) und für 19 Landmarken zwischen den Fingerknochen (rechts).

Für das vorgestellte Handmodell in vorangegangenen Beispielboxen wurden bereits die Distanzwerte zur Handkontur als Appearance-Merkmal genutzt. Im Folgenden werden diese als “Segmentierungsmerkmale” betrachtet. Außerdem wird ein weiteres Segmentierungsmerkmal, welches 19 Landmarken auf den Fingerknochenzwischenräumen beschreibt, hinzugefügt (Punkte im oberen Bild), um das Modellieren von Landmarkenpositionen zu demonstrieren.

Die vier folgenden Bilder zeigen beispielhaft die beiden “Segmentierungsmerkmale” ((a-b): Handkontur, (c-d): Landmarke)) als Eingabedaten (a,c). Die weißen Punkte visualisieren dabei die Positionen der Merkmalsvektoren von $\mathbf{S}_k$. Die beiden Bilder b) und d) zeigen die aus den Merkmalsvektoren rekonstruierte Kontur/Landmarke (rot). Um die Rekonstruktionsgenauigkeit zu untersuchen, wurden für alle 60 Trainingsdaten die mittleren Kontur-/Landmarkendistanzen zwischen Eingabe und rekonstruierter Kontur/Landmarken berechnet. Die nachfolgende Tabelle zeigt, dass bei einem Merkmalsvektorabstand sp von 3 mm, woraus sich 1744 Merkmalsvektoren ergeben (1.70% der Pixelanzahl der Handfläche), eine Rekonstruktionsgenauigkeit von 0.045 mm für die Landmarken und 0.09 mm für die Handkontur ergibt.

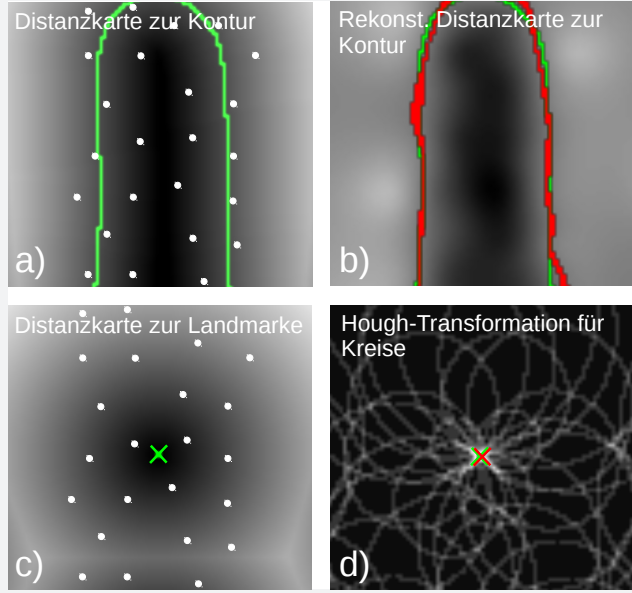


Abbildung 36: Kontur-/Landmarkenrekonstruktion

sp [mm]	$N_s(k)$	$N_s(k)/$ Handfläche [%]	Landmarken- Distanz [mm]	Kontur- Distanz [mm]
6	450 (± 57)	0.44 %	0.060 (± 0.02)	0.28 (± 0.11)
5	651 (± 84)	0.63 %	0.054 (± 0.01)	0.18 (± 0.06)
4	1041 (± 133)	1.01 %	0.046 (± 0.01)	0.12 (± 0.04)
3	1744 (± 230)	1.70 %	0.045 (± 0.01)	0.09 (± 0.01)

Tabelle 1: Quantitative Auswertung der Distanzen zwischen den Groundtruth-Konturen/Landmarken und den rekonstruierten Konturen/Landmarken. Siehe Abschnitt 8.3 für die Berechnung der Werte.

Damit wird deutlich, dass ein Merkmalsvektorabstand von 3 bis 6 mm ausreicht, um z. B. Segmentierungsmasken mit einer Genauigkeit von weit unter 1 mm zu beschreiben, ohne die Merkmalsvektoren direkt auf die entsprechenden Strukturkonturen zu setzen.

Parameterbelegung	
Input-/Trainingsdaten	
N	$= 60$
D	$= 5$ ($D_f = 1$, $D_f^{seg} = 2$, $D_x = 2$)
f_i	$: (1) \text{ Gau\ss-gegl\ddot{a}tete Signalwerte: } (G_{\sigma=6mm} * I_k)(x_i)$
f_i^{seg}	$: (1) \text{ Distanzwerte zur Handmaske: } d(x_i, I_k^{Mask}) \text{ (2) Distanzwerte zur Fingerknochen: } d(x_i, I_k^{Landmarken})$
x_i	$: \text{ kantenmerkmalsabh\ddot{a}ngig mit Mindestabstand } 6 \text{ mm, innerhalb der Hand } (\forall i f_{i,2} < 5mm)$
sp	$= 6, 5, 4, 3 \text{ mm}$
$N_s(k)$	$= 450 \pm 57 \text{ bis } 1744 \pm 230$

6.4. Initialisierung von S_{new}

Bei klassischen modellbasierten Ans\atzen ist der erste Schritt der Anpassung des Modells an neue Daten die Initialisierung der Position des Modells im Bild. Anschließend wird in der N\ahе der Punkte des Modells im Bildraum nach Punktkandidaten gesucht, an welche das Modell angepasst werden soll. In weiteren Iterationen werden diese Kandidaten mit der aktuellen Modellposition/-form aktualisiert. Dieser klassische Ansatz verlangt f\ur die “Bestimmung der Punktkandidaten” die Entwicklung eines weiteren Algorithmus. Daher ist die Qualit\at der klassischen modellbasierten Segmentierung oft von diesem Algorithmus und nicht vom Modell abh\angig. Da die optimalen Positionen der Punktkandidaten meist in der lokalen aktuellen Umgebung der Modellpunkte ermittelt werden, sind klassische modellbasierte Segmentierungen von der initialen Platzierung des Modells im Bild abh\angig (siehe Abbildung 37, links).

Im vorgestellten probabilistischen Modellansatz hingegen gestaltet sich das “Initialisierungsproblem” anders. Die Korrespondenzwerte werden zwischen allen m\oglichen Vektorpaaren zwischen S_{new} und M_{new} bestimmt und nicht nur in deren lokaler Umgebung (mit ausreichend gro\en Merkmalsvarianzen Σ – siehe Abschnitt 4.2 und 5.3). Daher ist die Positionierung des initialen Modells nicht mehr ausschlaggebend f\ur die Qualit\at der Ergebnisse. Au\erdem werden die Merkmalsvektoren in S_{new} nicht neu bestimmt – im Gegensatz zu den “Punktkandidaten” bei der klassischen Modellsegmentierung.

Wird also das gesamte Bild durch das gegebene Modell modelliert, indem z.B. die Merkmalsvektoren als Grit oder randomisiert \uber die Bilddaten verteilt werden, ist die Modellanpassung komplett unabh\angig von jeglicher Initialisierung. Um die Anzahl der Vektoren allerdings sinnvoll zu reduzieren, k\onnen die Positionen der Vektoren ausschlie\lich auf das betrachtete Objekt bzw. dessen Umgebung eingeschr\ankt werden. Au\erdem kann eine solche Reduktion der Merkmalsvektoren auch die Stabilit\at des Modells erh\ohen: F\ur das Beispiel der Handdaten besteht die Umgebung des betrachteten Objekts nur aus “Hintergrund”, der f\ur die Modellierung der Hand nicht von Interesse ist. Vor allem bei stark unterschiedlicher Handstellung/Fingerspreizung w\urden f\ur eine robuste Modellierung die Merkmalsvektoren zwischen den Fingern zu stark variieren. Darum k\onnen f\ur die Modellgenerierung die Merkmalsvektoren des Trainingsdatensatz-

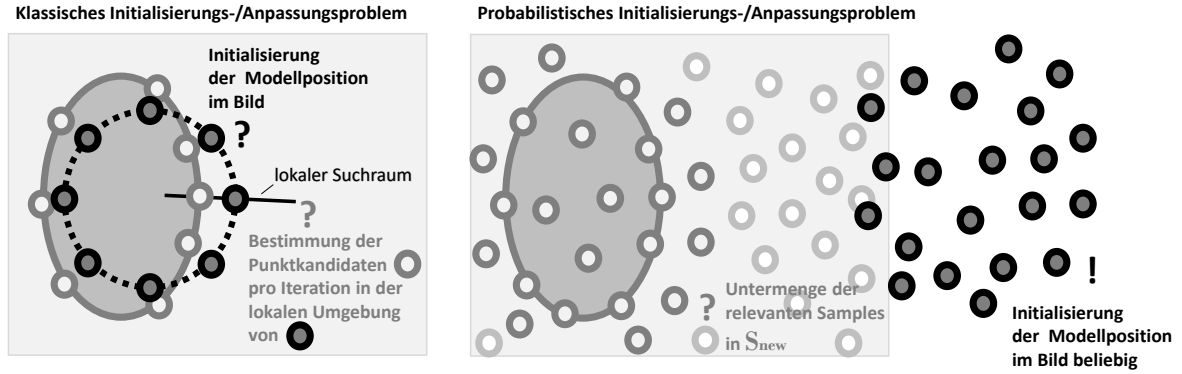


Abbildung 37: Initialisierungsproblem bei der Modellanpassung: Bei der klassischen modellbasierten Segmentierung (links) ist die Qualität des Ergebnisses abhängig von der Initialisierung der Modellposition im Bild und von der Art der Neubestimmung der Punktkandidaten (graue Punkte) in der lokalen Umgebung der Modelllandmarken in jeder Iteration. Beim vorgestellten probabilistischen Modellansatz hingegen ist die Initialisierung der Modellposition nicht von Interesse, da die Korrespondenzwerte zwischen allen möglichen Vektorpaaren und nicht nur in der lokalen Umgebung der Merkmalsvektoren des Modells berechnet werden (mit ausreichend großen Merkmalsvarianzen Σ – siehe Abschnitt 4.2 und 5.3). Außerdem werden die Merkmalsvektoren in S_{new} (analog zu “Punktkandidaten” im klassischen Ansatz) nicht neu bestimmt. Modelliert das gegebene Modell jedoch nur eine Untermenge der Bildpunkte (Merkmalsvektoren ausschließlich innerhalb/in der Umgebung des Objekts), muss diese Untermenge in S_{new} ebenfalls bestimmt/approximiert werden.

zes z. B. durch die gegebenen Distanzwerte zur Objektkontur beschränkt werden mit $S_k = \{s_{k,i} = (x_{k,i}, f_{k,i}, f_{k,i}^{seg} | f_{k,i}^{seg} < 5mm)\}$.

Modelliert das gegebene Modell jetzt also ausschließlich die Merkmalsvektoren innerhalb bzw. in der Umgebung des Objekts, muss diese Untermenge in neuen S_{new} für die Modellanpassung ebenfalls bestimmt werden. Die Werte $f_{new,i}^{seg}$ sind für S_{new} gesucht und daher für eine Initialisierung der Merkmalsvektorpositionen ($S_{new} = \{s_{new,i} = (x_{new,i}, f_{new,i}, \cancel{f_{new,i}^{seg}}, \cancel{f_{new,i}^{seg}} < 5mm)\}$) nicht gegeben. Demnach muss die betrachtete Untermenge an Vektoren für S_{new} auf eine andere Weise initialisiert werden. Eine Möglichkeit dies zu erzielen, ist eine Aneinanderreihung mehrerer Modelle. Abbildung 34 umschreibt eine derart gestaffelte Modellnutzung: Ein erstes gröberes Modell basiert auf komplett objektunabhängigen Merkmalsvektoren und liefert so ein erstes approximiertes Segmentierungsergebnis, welches für die Initialisierung eines feineren Modells bzw. eines S_{new} ohne Merkmalsvektoren im Objekthintergrund genutzt werden kann.

Eine weitere Möglichkeit, das “Problem” des fehlenden Vorwissens über die Lage des Objekts zu lösen, ist für S_{new} das gesamte Bild mit Merkmalsvektoren zu repräsentieren. Die Abbildung 38 visualisiert die Modellanpassung an eine Bildrepräsentation S_{new} mit kompletter Abtastung. Die Merkmalsvektoren in S_{new} können nach jeder Iteration der Modellanpassung reduziert werden, indem alle Vektoren $s_{new,i}$ ohne Korrespondenz zum

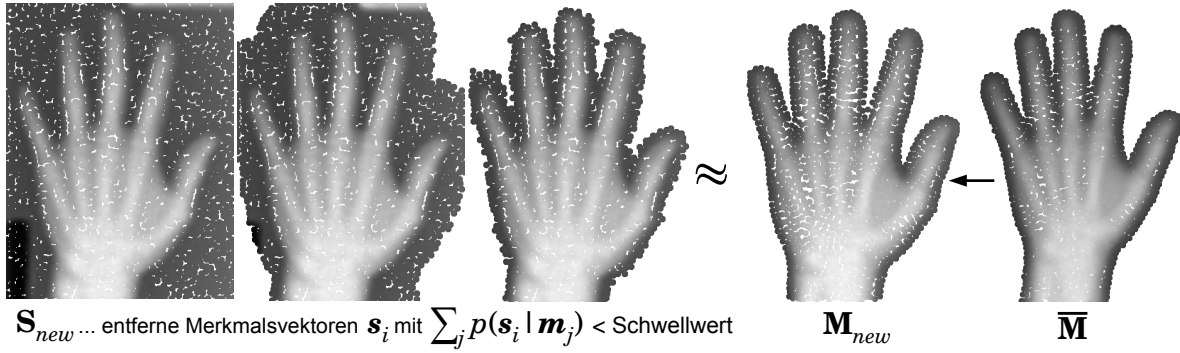


Abbildung 38: Anpassung eines Modells mit positionsabhängigen Merkmalsvektoren (ausschließlich innerhalb der Hand) an eine Bildrepräsentation $\mathbf{S}_{new}$ mit unbekannter Handkontur. Es wird initial das gesamte Bild abgetastet und nach jeder Iteration werden die nicht-korrespondierenden Merkmalsvektoren gelöscht.

Modell ($\sum_{j=1}^{N_m} p(\mathbf{s}_{new,i} | \mathbf{m}_{new,j}) < \text{Schwellwert}$) gelöscht werden. Da zu Beginn der Modellanpassung die Merkmalsvarianzen in Σ für die Berechnung von $p(\mathbf{s}_{new,i}, \mathbf{m}_{new,j})$ relativ groß sind, wird zunächst nur eine geringe Anzahl von “weit entfernten” Merkmalsvektoren gelöscht. Mit konvergierender Adaption werden die Werte in Σ kleiner (Abschnitt 4.2) und weitere Merkmalsvektoren werden als “nichtkorrespondierend” entfernt.

6.5. Rekonstruktion von fehlenden Bildkanälen in multispektralen Bilddaten

Anstelle von Objektmasken oder Landmarken wie beim Segmentierungsproblem kann jedes beliebige Appearance-Merkmal während der Modellanpassung für ein neues Bild rekonstruiert werden. Bei MRT-Aufnahmen kann z. B. der Fall eintreten, dass für einzelne Patienten bestimmte MRT-Sequenzen nicht akquiriert wurden. Manche Algorithmen sind jedoch auf alle Sequenzen angewiesen. Werden Convolutional Neural Networks (CNNs) beispielsweise auf einem vollständigen Satz an Merkmalen/Bildkanälen trainiert, müssen diese ebenfalls komplett für neuen Daten vorhanden sein, welche mittels des CNNs klassifiziert werden sollen. Liegt jedoch bei manchen Patienten ausschließlich eine Untermenge an z. B. MRT-Sequenzen vor, muss ein CNN für jede mögliche Kombination an vorhandenen Kanälen trainiert werden. In diesem Fall kann das probabilistische Modell zur Rekonstruktion der fehlenden Merkmale genutzt werden. Dies kann zum einen den Einsatzbereich von trainierten CNNs verbreitern (ein CNN für alle Merkmalskombinationen) und zum anderen kann dadurch die Anzahl der Trainingsdaten für solche maschinellen Lernverfahren erhöht werden. Die Herangehensweise der Merkmals-/Kanalrekonstruktion entspricht dabei genau dem Segmentierungsproblem (Abschnitt 6.2) mit dem Unterschied, dass kein Segmentierungsmerkmal im neuen Bild, sondern ein beliebiges Appearance-Merkmal aus $\mathbf{f}_i$ fehlt. Die Modellanpassung und die

Rekonstruktion des gesuchten Merkmals

$$f_i = \sum_{j=1}^{N_m} \mathbb{E}(H_{ij}) f_j \quad (85)$$

ist identisch mit der in Abschnitt 6.2 behandelten Problemstellung.

6.6. Klassifikation gesunder und pathologischer Strukturen

Eine pixelweise Klassifikation von medizinischen Strukturen kann als Multi-Objektsegmentierung betrachtet werden, bei der pro Klasse ein Label genutzt wird. Dies kann wie in Abschnitt 6.2 umgesetzt werden. Dabei können die einzelnen Klassen als Labels oder Distanzkarten modelliert werden.

Ist jedoch nach einer Klassifikation von pathologischen Strukturen gefragt, sind diese auf Grund ihrer zu hohen Heterogenität oft nicht modellierbar. Da Form- bzw. Appearance-Modelle die Appearance-Merkmale bzw. deren Variationen mit einer bestimmten Position im Modell/Organ assoziieren, wäre ein Trainingsdatensatz nötig, der alle möglichen Manifestationen und Lokalisationen der Pathologie enthält. Dies ist für solch heterogene Pathologien wie z. B. Tumore oder Schlaganfallläsionen nicht praktikabel.

Wird allerdings nach einer Klassifikation in physiologisch und pathologisch bzw. “der Norm entsprechend” und “abweichend der Norm” gefragt, kann dieses Problem mittels statistischer Modellierung gelöst werden. Dazu wird ein Modell der gesunden/normalen Struktur generiert und anschließend die Abweichungen vom Modell als Pathologie angenommen. Die Nutzung der Korrespondenzwahrscheinlichkeiten liefert für ein solches Szenario den Vorteil, dass das Modell auf Bilder mit Pathologien oder Artefakten adaptiert werden kann, da nicht von Eins-zu-Eins-Korrespondenzen zwischen Modell und neuem Bild ausgegangen wird.

6.6.1. Korrespondenzwahrscheinlichkeiten als Grundlage für die Klassifikation

Kleine Werte der resultierenden Korrespondenzwahrscheinlichkeiten nach Modellanpassung weisen auf Regionen im Bild hin, die nicht durch das “gesunde” Modell dargestellt werden konnten und somit pathologisch/anormal sind.

Im Gegensatz zur modellbasierten Segmentierung werden hier die gesuchten Informationen nicht modelliert (als zusätzliche Merkmale). Stattdessen sind sie automatisch durch die Korrespondenzwahrscheinlichkeiten nach beendeter Modellanpassung zwischen gesunder Modellinstanz und (potentiell pathologischem) neuem Bild gegeben. Dabei ist für jeden Merkmalsvektor der Eingabe $\mathbf{S}_{new}$ der Korrespondenzwert $p(\mathbf{s}_{new,i}|\mathbf{M}_{new})$ ge-

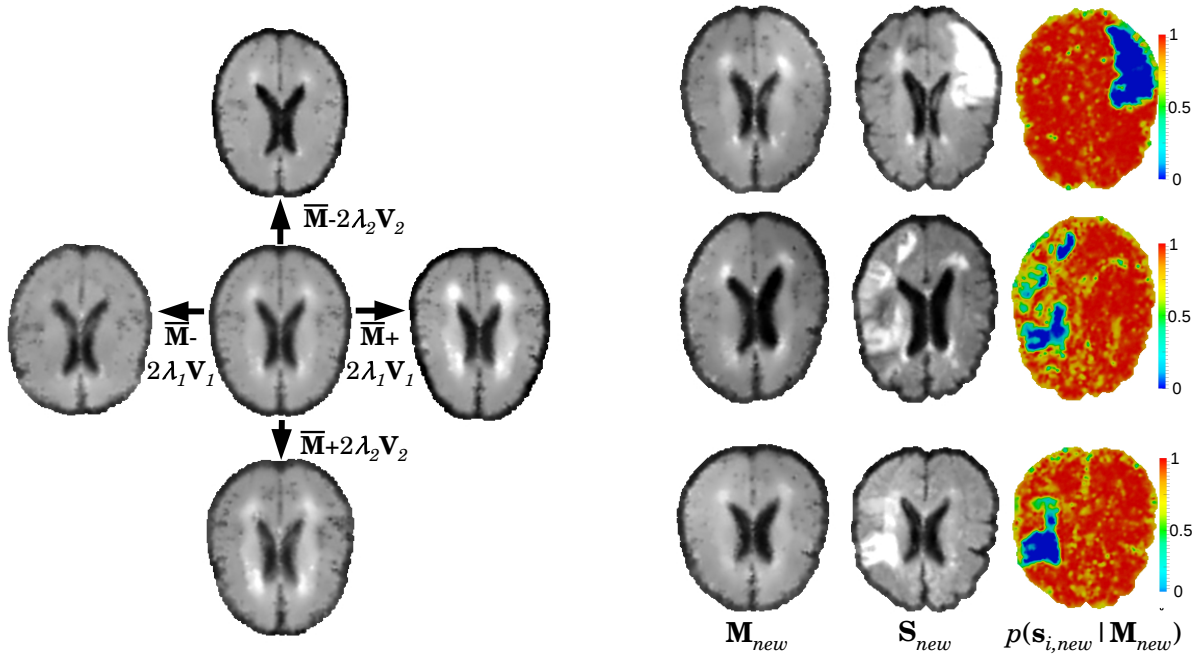


Abbildung 39: Klassifikation in physiologisch und pathologisch mithilfe von Korrespondenzwahrscheinlichkeiten: Links ist das gesunde Modell des Gehirns visualisiert, welches rechts auf neue Bildinstanzen ($\mathbf{S}_{new}$) mit Pathologie (Schlaganfallläsionen) adaptiert wird ($\mathbf{M}_{new}$). Die resultierenden Korrespondenzwerte ($p(\mathbf{s}_{new,i} | \mathbf{M}_{new})$) zwischen $\mathbf{S}_{new}$ und $\mathbf{M}_{new}$ indizieren, welche Regionen (dunkelblau bis hellblau) im Bild nicht durch das Modell dargestellt werden können und daher als pathologisch angenommen werden.

geben durch:

$$p(\mathbf{s}_{new,i} | \mathbf{M}_{new}) = \sum_{j=1}^{N_m} p(\mathbf{s}_{new,i} | \mathbf{m}_{new,j}). \quad (86)$$

Abbildung 39 visualisiert beispielhaft ein gesundes Hirnmodell, das auf Bilddaten mit Schlaganfallläsionen adaptiert wird. Die einzelnen Korrespondenzwerte können ebenfalls mittels Scattered Data B-Spline Interpolation zur einer kompletten Wahrscheinlichkeitskarte ($p(\mathbf{x} | \mathbf{M}_{new}), \forall \mathbf{x} \in \Omega$) rekonstruiert werden (Abschnitt 6.1). Die dunkel bis hellblauen Bereiche in Abbildung 39 indizieren dabei Regionen mit sehr geringer Korrespondenz und können daher als pathologisch interpretiert werden.

Die einfachste Methode, die Bilder in gesunde und pathologische Gebiete zu klassifizieren, ist die Anwendung eines Schwellwertes auf den rekonstruierten Korrespondenzkarten. Die Korrespondenzwerte können darüber hinaus als Grundlage beliebig komplexer Klassifikatoren oder maschineller Lernverfahren eingesetzt werden. Im Abschnitt 8.5 wird die Klassifikation von 3D MRT-Hirndaten in gesund und pathologisch evaluiert.

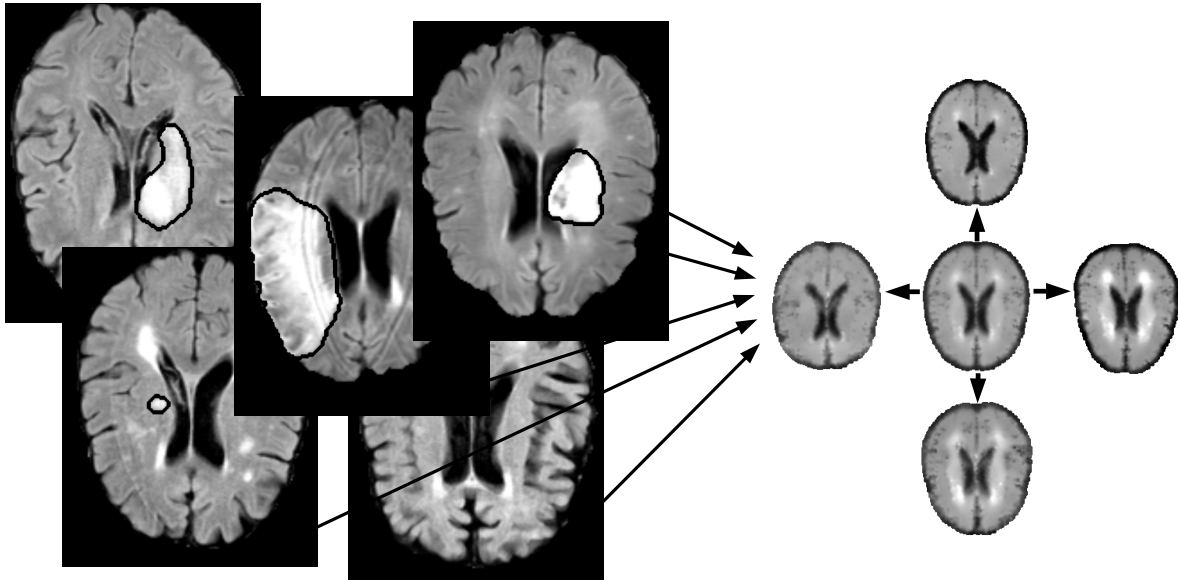


Abbildung 40: Generierung eines kompletten Modells mit partiellen/unvollständigen Trainingsdaten: Links sind Trainingsbilder gegeben mit bekannten Regionen (umrandet), die nicht in das Modell mit einfließen sollen. Während der Modellgenerierung werden die Korrespondenzwahrscheinlichkeiten für die Merkmalsvektoren der Trainingsdaten, die in diesen Regionen liegen, auf Null gesetzt.

6.7. Modellgenerierung auf der Grundlage von partiellen Trainingsdaten

Eine weitere Herausforderung bei der Betrachtung pathologischer Bilddaten ist der mögliche Mangel an gesunden Trainingsdaten. Im Abschnitt 6.6 wird beispielhaft das Problem der Klassifikation von Bilddaten in gesunde und pathologische Gebiete angesprochen. Die Grundlage solcher Klassifikationen bilden oft Bilddaten aus dem klinischen Alltag. Bei diesen Daten handelt es sich zum Großteil um Aufnahmen von Patienten mit vorliegender Pathologie – je nach Fragestellung für die Klassifikation. Für die Generierung des “gesunden Modells” aus pathologischen Bilddaten ist auch hierfür eine einfache Lösung im vorgestellten Ansatz enthalten: Dafür muss lediglich bekannt sein, welche Regionen im Bild nicht für die Modellgenerierung genutzt werden sollen. Dabei liegt die Annahme zugrunde, dass die Pathologien (z. B. Läsionen) lediglich einen lokal beschränkten Einfluss auf das umliegende “gesunde” Gewebe haben. In Abbildung 40 wird ein Modell des gesunden Gehirns aus Bilddaten mit Schlaganfallläsionen generiert. Es liegen die Masken der Läsionen (umrandet) vor. Während der Modellgenerierung werden die Korrespondenzwahrscheinlichkeiten für die Merkmalsvektoren der Trainingsdaten, die innerhalb einer Läsion liegen, auf Null gesetzt, sodass diese nicht in den Modellgenerierungsprozess einfließen. Es sei darauf hingewiesen, dass das resultierende Modell (Θ) keinerlei Informationen über die ignorierten Regionen in den Trainingsdaten beinhaltet. Im Abschnitt 8.5 wird die Klassifikation von 3D MRT-Hirndaten in gesund

und pathologisch evaluiert, wobei das genutzte “gesunde” Modell auf diese Weise aus “partiellen”/pathologischen Trainingsdaten generiert wurde.

6.8. Bildrekonstruktion bei artefaktbehafteten Daten

Die Korrespondenzwahrscheinlichkeiten nach Modellanpassung liefern Informationen über Bereiche im Bild, die nicht durch das Modell darstellbar sind. Unter der Annahme, dass das Modell vollständig ist, können diese Bereiche als pathologisch (siehe Abschnitt 6.6) oder als Bildartefakt interpretiert werden. Diese Information und die lokal korrespondierenden Appearance-Merkmale der entsprechenden Modellinstanz können anschließend zur Rekonstruktion der fehlerhaften Bereiche genutzt werden:

$$\text{if}(p(\mathbf{s}_{new,i}|\mathbf{M}_{new}) < \text{threshold}) : \mathbf{f}_{new,i} = \frac{\sum_{j=1}^{N_m} p(\mathbf{x}_{new,i}|\mathbf{x}_{new,j}) \mathbf{f}_{new,j}}{\sum_{j=1}^{N_m} p(\mathbf{x}_{new,i}|\mathbf{x}_{new,j})}. \quad (87)$$

Hier wird die Wahrscheinlichkeit der Positionswerte der Merkmalsvektoren $p(\mathbf{x}_{new,i}|\mathbf{x}_{new,j})$ anstelle der gesamten Vektoren $p(\mathbf{s}_{new,i}|\mathbf{m}_{new,j})$ für die Korrespondenzbestimmung genutzt – unter der Annahme, dass der Merkmalsvektor $\mathbf{s}_{new,i}$ in einem Bildbereich mit stark abweichenden (fehlerhaften) Appearance-Merkmalen ($p(\mathbf{s}_{new,i}|\mathbf{M}_{new}) < \text{threshold}$) liegt und $p(\mathbf{s}_{new,i}|\mathbf{m}_{new,j})$ daher nicht aussagekräftig ist, um die Korrespondenz innerhalb dieser Regionen zu beschreiben.

7. Probabilistische Appearance-adaptierende Registrierung beliebigdimensionaler Bilder

Wird ein gegebenes Form- oder Appearance-Modell an ein neues Bild angepasst, kann dies auch als “Registrierung” des Modells auf ein neues Bild betrachtet werden. Die Freiheitsgrade der Deformation sind dabei durch die Anzahl der Modengewichte des gegebenen Modells und eine affine Transformation eingeschränkt, was zu einer wesentlich höheren Robustheit aber auch Beschränkung der resultierenden Transformation führt. Das klassische nichtrigide Registrierungsproblem – die Anpassung eines Bildes auf ein zweites – hat maximal so viele Freiheitsgrade wie Pixel im Bild. Die Korrespondenzen zwischen beiden Bildern sind unbekannt und werden über die Optimierung der Deformation (während der Registrierung) gesucht. Dies wiederum ist dem Modellgenerierungsprozess des vorgestellten Ansatzes ähnlich, da im Gegensatz zu klassischen Modellansätzen hier die Korrespondenzen ebenfalls während der Optimierung gesucht werden. Im Folgenden wird das betrachtete Modell mit dem beschriebenen Modellgenerierungsverfahren (Abschnitt 4) zum Registrierungsverfahren zweier Bilder vereinfacht.

Für die Registrierung eines Bildes $I_{\mathbf{M}} : \Omega \rightarrow \mathbb{R}^{D_f}, \Omega \subset \mathbb{R}^{D_x}$ auf ein zweites Bild $I_{\mathbf{S}} : \Omega \rightarrow \mathbb{R}^{D_f}, \Omega \subset \mathbb{R}^{D_x}$ wird in klassischen Registrierungsansätzen die Transformation/Deformation $\phi : \Omega \rightarrow \Omega$ gesucht, welche $I_{\mathbf{M}}$ deformiert. Für die nichtrigide Deformation kann ϕ über ein Verschiebungsfeld $\mathbf{u} : \mathbb{R}^{D_x} \rightarrow \mathbb{R}^{D_x}$ definiert werden, welches für jede Pixelposition im Bild einen Offset beinhaltet:

$$(I_{\mathbf{M}} \circ \phi)(\mathbf{x}) = I_{\mathbf{M}}(\mathbf{x} + \mathbf{u}(\mathbf{x})). \quad (88)$$

Optimiert wird die Deformation ϕ durch

$$\mathcal{D}(I_{\mathbf{S}}, I_{\mathbf{M}} \circ \phi) + \alpha \mathcal{R}(\phi) \rightarrow_{\phi} \min, \quad (89)$$

wobei $\mathcal{D}(\cdot, \cdot)$ ein Distanzmaß beschreibt und $\mathcal{R}(\phi)$ ein mit α gewichteter Regularisierungsterm ist, um unplausible Transformationen zu bestrafen.

Der vorgestellte Modellansatz kann zu einem nichtrigiden Registrierungsansatz für zwei Bilder vereinfacht werden. Beide zu registrierende Bilder werden, wie in Abschnitt 2 dargelegt, als Merkmalsvektormengen/Bildrepräsentationen $\mathbf{S} = \{\mathbf{s}_i | i = 1, \dots, N_s\}$ und $\mathbf{M} = \{\mathbf{m}_j | j = 1, \dots, N_m\}$ betrachtet. Anstatt das mittlere Modell $\bar{\mathbf{M}}$ mit den n gewichteten vorher bestimmten Moden zu variieren, wird in diesem Fall die Merkmalsvektormenge $\mathbf{M}$ mit nur einer Mode an $\mathbf{S}$ angepasst:

$$\mathbf{S} \approx T^{-1}(\mathbf{M} + \mathbf{V}). \quad (90)$$

Mit $n = 1$ werden die Modell-/Observationsparameter λ_p und ω_p ebenfalls konstant Eins und entfallen. Der räumliche Anteil (für die Positionsmerkmale) der eine Mode $\mathbf{V}$ entspricht dem nichtrigiden Verschiebungsfeld $\mathbf{u}$ und T beinhaltet den affinen Anteil der

gesuchten Transformation. Durch die vektorielle Repräsentation der Merkmale, bei der Koordinaten und beliebig viele Appearance-Merkmale kombiniert werden in $\mathbf{s}_i$, $\mathbf{m}_j$ und $\mathbf{v}_j$ (für $\mathbf{S}$, $\mathbf{M}$ und $\mathbf{V}$), stellt der so definierte Ansatz eine “beliebigdimensionale” Registrierung mit $\mathbf{V} : \mathbb{R}^{D_x+D_f} \rightarrow \mathbb{R}^{D_x+D_f}$ und $T : \mathbb{R}^{D_x} \rightarrow \mathbb{R}^{D_x}$ für “beliebigdimensionale” Bilder $I_{\mathbf{S},\mathbf{M}} : \Omega \rightarrow \mathbb{R}^{D_f}$, $\Omega \subset \mathbb{R}^{D_x}$ dar. Bei der klassischen Registrierung werden ausschließlich Koordinatenverschiebungen berechnet und das klassische Verschiebungsfeld entspricht demnach dem Koordinatenanteil der “Mode” ($\mathbf{V}^{\mathbf{x}} = \mathbf{u}$), wohingegen beim vorgestellten Ansatz ebenfalls eine Transformation der Signalwerte/Appearance-Werte vorgenommen werden kann.

Zur Bestimmung der gesuchten Transformation wird das bekannte globale Optimierungskriterium – siehe Gleichung 37 – vereinfacht zu:

$$C(T, \mathbf{V}) \rightarrow_{T, \mathbf{V}} \min = \underbrace{-\log p(\mathbf{S}|T, \mathbf{V}, \mathbf{M})}_{\approx \mathcal{D}(I_{\mathbf{S}}, \phi \circ I_{\mathbf{M}})} - \underbrace{\log p(\mathbf{V})}_{\approx \alpha \mathcal{R}(\phi)} \rightarrow_{T, \mathbf{V}} \min. \quad (91)$$

Dabei entspricht der erste Teil dem Distanzmaß und der zweite Teil dem Regularisierungsterm der klassischen Registrierungsformulierung, wobei der affine Anteil keiner Beschränkung unterliegt. Die Werte in $\Sigma_{\mathcal{N}}$ für $p(\mathbf{V})$ bestimmen dabei analog zu α die Stärke der Regularisierung – allerdings nicht nur für die topologischen Änderungen, sondern ebenfalls für die Anpassungen der Appearance-Werte. Auf $\Sigma_{\mathcal{N}}$ wird genauer in Abschnitt 7.1 eingegangen. Das gesamte Kriterium lässt sich zusammenfassen als:

$$\begin{aligned} C(T, \mathbf{V}) &= -\log p(T, \mathbf{V}, \mathbf{M}|\mathbf{S}) = -\log \left(\frac{p(\mathbf{S}|T, \mathbf{V}, \mathbf{M})p(T|\mathbf{V}, \mathbf{M})p(\mathbf{M})p(\mathbf{V})}{p(\mathbf{S})} \right), \quad (92) \\ &\quad p(T|\mathbf{V}, \mathbf{M}), p(\mathbf{M}) \text{ und } p(\mathbf{S}) \text{ const.} \\ &= -\log(p(\mathbf{S}|T, \mathbf{V}, \mathbf{M})p(\mathbf{V})) \\ &= -\log p(\mathbf{S}|T, \mathbf{V}, \mathbf{M}) - \log p(\mathbf{V}) \\ &= -\sum_{i=1}^{N_s} \log \left(\sum_{j=1}^{N_m} \exp \left(-\frac{1}{2} (T \star \mathbf{s}_i - (\mathbf{m}_j + \mathbf{v}_j))^T \Sigma_j^{-1} (T \star \mathbf{s}_i - (\mathbf{m}_j + \mathbf{v}_j)) \right) \right) \\ &\quad + \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \left(\frac{1}{2} \left(\frac{((\mathbf{m}_j \pm 2 \cdot \mathbf{v}_j) - (\mathbf{m}_\eta \pm 2 \cdot \mathbf{v}_\eta)) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \right. \\ &\quad \left. \Sigma_{\mathcal{N}}^{-1} \left(\frac{((\mathbf{m}_j \pm 2 \cdot \mathbf{v}_j) - (\mathbf{m}_\eta \pm 2 \cdot \mathbf{v}_\eta)) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right) \end{aligned}$$

Die Optimierung des Kriteriums nach $\mathbf{v}_j$ und T erfolgt analog zu Abschnitt 4.1.4 und 4.1.7 (mit $\omega_x = 1$, $\lambda_p = 1$ und konstant $\bar{\mathbf{M}} = \mathbf{M}$). Die Merkmalsvarianzen Σ werden ebenfalls automatisch bestimmt (Abschnitt 4.2.1). Im Vergleich zur klassischen Registrierung kann der Appearance-Anteil von Σ , als Regularisierung/Gewichtung für das Distanzmaß interpretiert werden. Der räumliche Anteil von Σ mit der Interpretation: “Wie weit dürfen Koordinaten auseinanderliegen, um noch als korrespondierend zu gelten?” wäre

```

IN   Bilder  $I_S$  und  $I_M$  (mit  $D_f$  Appearance-Merkmalen) samt Positionen  $\mathbf{x}_i$  (für alle Level);
        Anzahl der Level  $N_l$ 
FOR  $l = 1 : N_l$ 
    INIT  $\mathbf{S}^l = \{\mathbf{s}_i^l = (\mathbf{x}_i^l, \mathbf{f}_i^l) | i = 1, \dots, N_s^l\}$  und  $\mathbf{M}^l = \{\mathbf{m}_j^l = (\mathbf{x}_j^l, \mathbf{f}_j^l) | j = 1, \dots, N_m^l\}$ 
        IF  $l = 1$ : Abs. 4.3
            Setze  $\mathbf{V}^l = 0, T^l = id$ , Initialisiere  $\Sigma^l$  Abs. 4.3.1
        IF  $l > 1$ : Abs. 5.1.1
            Initialisiere Parameter ( $\mathbf{V}^l$ ) und Korrespondenzbeziehungen  $\mathbb{E}(H_{ij}^l)$ 
            für Level  $l$  mit Parametern von Level  $l - 1$ 
            Initialisiere Nachbarschaftsbeziehungen  $\mathcal{N}(j) \forall j$  in  $\mathbf{M}^l$  Abs. 4.3.3
        DO
            0. Update  $\Sigma, \Sigma_{\mathcal{N}}$  (und berechne  $\mathbb{E}(H_{ij})$ ) Abs. 4.2.1/2
            1. Berechne  $C(T^l, \mathbf{V}^l)_{start}$  Abs. 4.1
            2.   Optimierte  $\mathbf{V}^l$  (und berechne  $\mathbb{E}(H_{ij})$ )
            3.   Optimierte  $T^l$  (und berechne  $\mathbb{E}(H_{ij})$ )
            4. Berechne  $C(T^l, \mathbf{V}^l)_{curr}$ 
            WHILE  $(C(T^l, \mathbf{V}^l)_{start} - C(T^l, \mathbf{V}^l)_{curr}) > \epsilon$ 
        OUT Deformationsfeld  $DF(\mathbf{V}^{N_l}, T^{N_l})$  Abs. 7.2

```

Abbildung 41: Multi-Level-Registrierung: Zusammenfassung des Registrierungsprozesses mit mehreren Leveln $l = 1, \dots, N_l$.

vergleichbar mit der Schrittweite bzw. Größe des lokalen Suchraums pro Pixel bzw. in diesem Fall pro Merkmalsvektor.

Multi-Level-Registrierung: Durch die hohe Anzahl von Freiheitsgraden und die zusätzlich wegfallende “Mittelung” der Parameter durch den Trainingsdatensatz, wird die Nutzung eines Multi-Level-Ansatzes bei der Registrierung weit relevanter als bei der Modellgenerierung. Dabei werden die Level, ebenso wie in Abschnitt 5.1 beschrieben, initialisiert. Abbildung 41 fasst den Registrierungsansatz mit dem bekannten globalen Kriterium zusammen.

7.1. Appearance-adaptierende Registrierung

Bei der klassischen Registrierung wird ausschließlich nach einer räumlichen Deformation gesucht. Eine Änderung der Signalwerte der Bilder wird nicht vorgenommen. Durch die beliebigdimensionale Formulierung der Bildrepräsentationen und des globalen Kriteri-

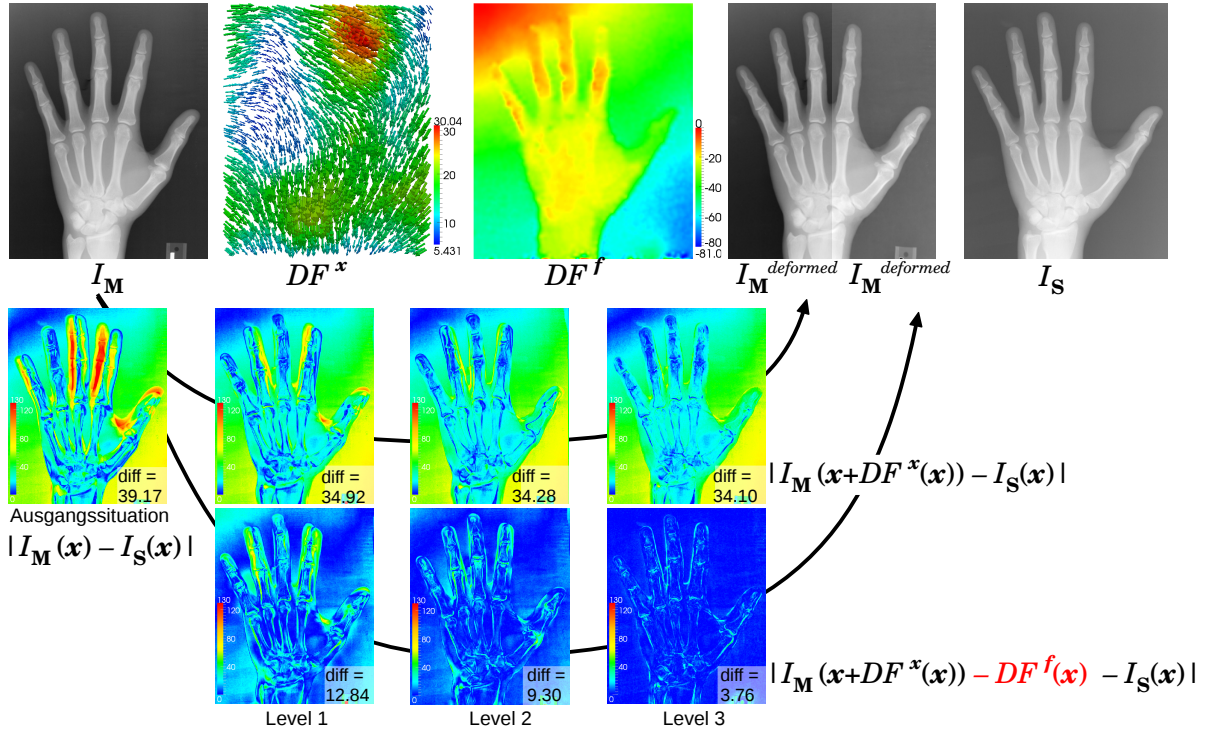


Abbildung 42: Multi-Level-Registrierung zwischen zwei Bildern I_M und I_S : DF^x visualisiert das resultierende Deformationsfeld für die Koordinatenverschiebung und DF^f zeigt die berechnete Signalwertänderung zwischen I_M und I_S . Die zweite Zeile stellt die Signalwertdifferenzen zwischen deformierten I_S und I_M dar: vor der Registrierung und nach erstem/zweiten/dritten Level, wobei wie bei der klassischen Registrierung lediglich DF^x genutzt wird. Die letzte Zeile zeigt dementsprechend die Differenzbilder mit Anwendung von DF^f . Dabei wird deutlich, dass DF^f vor allem starke Signalwertunterschiede in den Bildhintergründen ausgleicht und "topologische Fehler", wie z. B. nicht übereinstimmende Fingerknochen, nicht durch Signalwertänderungen ausgeglichen werden. Des Weiteren zeigen die mittleren Signalwertdifferenzen ("diff"-Wert in der jeweils rechten Ecke des Bildes), dass sich die Werte ohne Anpassung der Signalwerte zwischen den Levels nur sehr geringfügig unterscheiden. Nach der Signalwertanpassung sind die Differenzwerte wesentlich aussagekräftiger, da fast ausschließlich topologische Fehler übrigbleiben.

ums, bei welcher Positionsmerkmale und Appearance-Merkmale gleichbehandelt werden, ist die entsprechende Registrierung ebenfalls nicht beschränkt auf Positionsänderungen. Das Framework erlaubt automatisch eine beliebigdimensionale Transformation eines Bildes. Das in Abschnitt 3.2.2 diskutierte Problem der Positionsänderungen gegenüber der Appearance-Änderungen spielt bei der Registrierung eine noch wesentlichere Rolle als bei der Modellerstellung mit einer großen Anzahl an Trainingsdaten. Eine Positionsanpassung der Merkmalsvektoren der Bildrepräsentationen sollte bei der Registrierung immer eine höhere Priorität haben als eine Appearance-Anpassung, da die Registrierung in den meisten Fällen dem Zweck der topologischen Transformation dient bzw.

lokal korrespondierende Merkmalsvektoren in beiden Bildern gefunden werden sollen. Selbst wenn die Appearance-Anpassung bei den meisten Registrierungsanwendungen nicht von Interesse ist, kann sie während des Optimierungsprozesses zu einer genaueren Bestimmung der Korrespondenzwahrscheinlichkeiten zwischen $\mathbf{S}$ und $\mathbf{M}$ führen. Werden Signalwertunterschiede während der Optimierung ausgeglichen, werden die Varianzwerte in Σ für die Berechnung von $\mathbb{E}(H_{ij})$ bzw. $p(\mathbf{s}_i|\mathbf{m}_j)$ kleiner und es kommt zu einer genaueren Zuordnung korrespondierender Merkmalsvektoren.

Um zu gewährleisten, dass die Appearance-Änderung keine topologische Änderung übernimmt, müssen die Werte in $\Sigma_{\mathcal{N}}$ für die Appearance-Merkmale entsprechend gewichtet werden. In Abschnitt 4.2.2 wurde die Bestimmung und Gewichtung von $\Sigma_{\mathcal{N}}$ diskutiert. Für die Modellgenerierung hat sich herausgestellt, dass vor allem eine Regularisierung der Positionswerte von Bedeutung ist und die Werte in $\Sigma_{\mathcal{N}}$ der Appearance-Änderungen daher hoch gewichtet werden können (hohe Änderungsunterschiede zwischen benachbarten Merkmalsvektoren seien erlaubt). Dies gilt allerdings nur, da die Modellmoden für einen ganzen Trainingsdatensatz optimiert werden und daher automatisch eine gewisse Mittelung der Werte stattfindet. Da bei der Registrierung das “Modell” ($\mathbf{M}$) an nur eine Observation ($\mathbf{S}$) angepasst wird, muss wesentlich stärker regularisiert werden, indem man beispielsweise die Werte in $\Sigma_{\mathcal{N}}$ mit Eins gewichtet:

$$\Sigma_{\mathcal{N}} = \min(\text{diag}(0.1, \dots, 0.1), \text{diag}(\text{Cov}(\mathbf{M}, \mathbf{M} \pm 2 \cdot \mathbf{V}))). \quad (93)$$

Abbildung 42 zeigt ein Beispiel, bei welchem durch eine Anpassung der Signalwerte starke Signalwertunterschiede im Bildhintergrund oder der Handfläche ausgeglichen werden, “topologische Fehler” (unterschiedliche Position von Fingern/Knochen) jedoch ausschließlich durch Positionsänderungen minimiert werden. In Abschnitt 8.6 wird die probabilistische Registrierung an 2D MRT Gehirn-Daten evaluiert, wobei gezeigt wird, dass die Appearance-adaptierende Registrierung für diese Daten bessere Ergebnisse liefert als die probabilistische Registrierung ohne Appearance-Anpassung.

7.2. Rekonstruktion eines Deformationsfeldes

Durch die Betrachtung der spärlichen merkmalsvektorbasierten Bildrepräsentationen $\mathbf{S}$ und $\mathbf{M}$ sind die resultierenden Deformationswerte zunächst ebenfalls nur an den Positionen der jeweiligen Merkmalsvektoren in $\mathbf{M}$ vorhanden. Die Deformationswerte werden pro Vektor aus der affinen Transformation und der nichttrigiden Deformation durch $\mathbf{V}$ zusammengesetzt:

$$df_j^{\mathbf{x}}(\mathbf{x}_j) = ((T^{-1} \star (\mathbf{m}_j + \mathbf{v}_j)) - \mathbf{m}_j)^{\mathbf{x}} := T^{-1}(\mathbf{x}_j + \mathbf{v}_j^{\mathbf{x}}) - \mathbf{x}_j \quad (94)$$

$$df_j^{\mathbf{f}}(\mathbf{x}_j) = \mathbf{v}_j^{\mathbf{f}} \quad (95)$$

Dabei beschreibt $\cdot^{\mathbf{x}}$ den Positionsanteil des Deformations- bzw. des Verschiebungsvektors und $\cdot^{\mathbf{f}}$ den Appearance-Anteil. Ein gesamtes Deformationsfeld $DF : \Omega \rightarrow \mathbb{R}^{D_x + D_f}$ für das

vollständige Eingabebild kann mit der Scattered Data B-Spline Interpolation berechnet werden (Abschnitt 6.1). Abbildung 42 zeigt ein Deformationsfeld. Das DF beschreibt nach dem so vorgestellten Registrierungsansatz die Deformation von I_M nach I_S (mit dem Ziel $I_S \approx I_M^{deformed}$):

$$I_M^{deformed}(\mathbf{x} + DF^{\mathbf{x}}(\mathbf{x})) = I_M(\mathbf{x}) + DF^{\mathbf{f}}(\mathbf{x}) \quad (96)$$

Da die neue Bildkoordinate $\mathbf{x} + DF^{\mathbf{x}}(\mathbf{x})$ gegebenenfalls nicht auf eine ganzzahlige Pixelkoordinate fällt, erschwert sich die Interpolation der neuen Signalwerte. Des Weiteren ist nicht gegeben, dass das Deformationsfeld mit $\mathbf{x} + DF^{\mathbf{x}}(\mathbf{x})$ das gesamte Bild abdeckt, wodurch Lücken im deformierten Bild entstehen würden und ebenfalls eine weitere Interpolation nötig wäre. Aus diesem Grund wird bei der Registrierung zweier Bilder die Inverse der gesuchten Deformation (DF^{-1}) optimiert, für:

$$I_M^{deformed}(\mathbf{x}) = I_M(\mathbf{x} + (DF^{-1})^{\mathbf{x}}(\mathbf{x})) - (DF^{-1})^{\mathbf{f}}(\mathbf{x}). \quad (97)$$

Dadurch ist gegeben, dass für ganzzahlige lückenlose Pixelkoordinaten ($\mathbf{x}$) nur die Signalwerte an den Koordinaten $\mathbf{x} + (DF^{-1})^{\mathbf{x}}(\mathbf{x})$ in I_M interpoliert abgetastet werden müssen.

7.2.1. Inverses Deformationsfeld

Um die Inverse DF^{-1} zu bestimmen, gibt es zwei Möglichkeiten.

Variante 1: Da die inverse Deformation von M auf S der Deformation von S auf M entsprechen sollte ($DF_{M \rightarrow S}^{-1} \approx DF_{S \rightarrow M}$), muss lediglich M und S bei der vorgeschlagenen Registrierung vertauscht werden ($T^{-1}(\mathbf{S} + \mathbf{V})$). Das rekonstruierte Feld, kann anschließend einfach als die inverse Deformation für M interpretiert werden

$$df_i^{-1} \mathbf{x}(\mathbf{x}_i) = ((T^{-1} \star (\mathbf{s}_i + \mathbf{v}_i)) - \mathbf{s}_i)^{\mathbf{x}} := T^{-1}(\mathbf{x}_i + \mathbf{v}_i^{\mathbf{x}}) - \mathbf{x}_i \quad (98)$$

$$df_i^{-1} \mathbf{f}(\mathbf{x}_i) = \mathbf{v}_i^{\mathbf{f}} \quad (99)$$

und Formel 97 genutzt werden.

Variante 2: Um das gesamte Deformationsfeld aus $\{df_j | j = 1, \dots, N_m\}$ zu berechnen, wird die Scattered-Data-B-Spline-Interpolation eingesetzt (Abschnitt 6.1), da die Koordinaten $\mathbf{x}_j$ der Bildrepräsentationen weder gleichmäßig verteilt noch auf ganzzahligen Pixelkoordinaten liegen müssen. Daher kann auch – im Gegensatz zum klassischen Verschiebungsfeld, bei dem die Verschiebungsvektoren meist für gleichmäßig verteilte Pixelkoordinaten berechnet werden – die Berechnung der Deformationswerte einfach “umgedreht” werden:

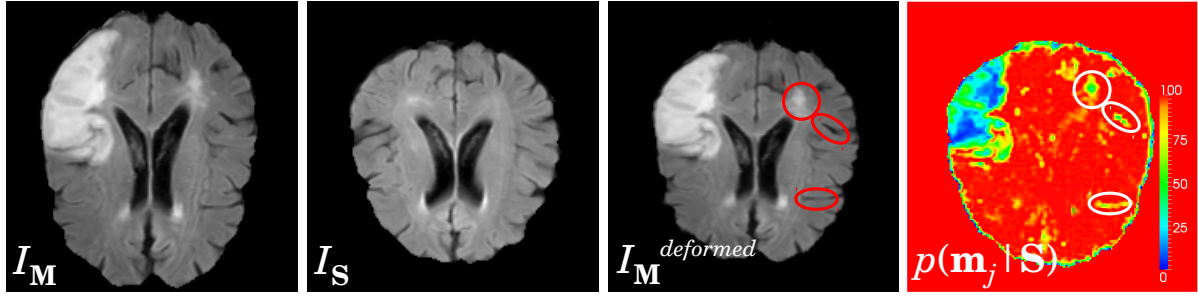


Abbildung 43: Lokale Unsicherheiten: Wird ein Bild I_M mit Schlaganfallläsion auf ein Bild ohne Läsion registriert, indizieren die Korrespondenzwerte (für die Berechnung von $p(\mathbf{m}_j|\mathbf{S})$ siehe Abschnitt 6.6.1) eindeutig die pathologische Region. Außerdem weisen Strukturen (rot und weiß markierte Bereiche), die in einem der Bilder nicht vorhanden sind, geringere Korrespondenzwahrscheinlichkeiten auf.

$$df_j^{-1} \mathbf{x}(\mathbf{x}_j + \mathbf{v}_j^{\mathbf{x}}) = (\mathbf{m}_j - (T^{-1} \star (\mathbf{m}_j + \mathbf{v}_j)))^{\mathbf{x}} := \mathbf{x}_j - T^{-1}(\mathbf{x}_j + \mathbf{v}_j^{\mathbf{x}}) \quad (100)$$

$$df_j^{-1} \mathbf{f}(\mathbf{x}_j + \mathbf{v}_j^{\mathbf{x}}) = -\mathbf{v}_j^{\mathbf{f}}. \quad (101)$$

Dabei sei zu beachten, dass die inversen Deformationsfelder beider Varianten nicht identische Ergebnisse liefern, da die Anzahl der Merkmalsvektoren bzw. der Werte in $\mathbf{V}$, aus denen DF jeweils interpoliert wird, unterschiedlich sind. Bei Variante 1 enthält $\mathbf{V}$ so viele Vektoren wie $\mathbf{S}$ (da $T^{-1}(\mathbf{S} + \mathbf{V})$, $df_i^{-1}(\mathbf{x}_i) = \dots, i = 1, \dots, N_s$) und in Variante 2 entspricht die Vektoranzahl für $\mathbf{V}$ der von $\mathbf{M}$ (da $T^{-1}(\mathbf{M} + \mathbf{V})$, $df_j^{-1}(\mathbf{x}_j + \mathbf{v}_j^{\mathbf{x}}) = \dots, j = 1, \dots, N_m$).

7.3. Registrierung von Bilddaten mit fehlenden Korrespondenzen

Eine große Herausforderung bei der Registrierung zweier Bilder ist die Annahme, dass alle Strukturen des einen Bildes ebenfalls im anderen Bild vorhanden sind. Soll beispielsweise ein Bild mit einer Pathologie (Tumor/Läsion) auf ein gesundes Bild registriert werden, kommt es bei den klassischen Registrierungsverfahren zum fehlerhaften Verhalten auf Grund der falschen Korrespondenzannahme. Eine einfache Möglichkeit, das Problem der fehlenden Strukturen zu umgehen, ist die Maskierung der pathologischen Bildregion. Jedoch muss in diesem Fall der Registrierung ein Segmentierungsschritt vorgangestellt werden.

Auch bei der Herausforderung von pathologischen Bilddaten/fehlenden Bereichen liefern die genutzten Korrespondenzwahrscheinlichkeiten eine Lösung des Problems. Zum einen indizieren sie, wie bereits in Abschnitt 6.6 genutzt, Bereiche geringer Korrespondenz zwischen den Bildern, zum anderen wird durch diese geringe Korrespondenz si-

chergestellt, dass an den entsprechenden Bereichen keine (unschlüssige) Deformation stattfindet. Es kommt also zu einem automatischen Maskierungseffekt. Durch die automatische Anpassung der Merkmalsvarianzen (Σ , Abschnitt 4.2.1) ist auch in diesem Fall keine manuelle Anpassung von Parametern notwendig. Abbildung 43 visualisiert beispielhaft die lokalen “Unsicherheiten” bei der Registrierung von einem pathologischen und einem gesunden Bild.

7.4. Probabilistische Registrierung als Nachverarbeitungsschritt für die Modellanpassung

Ein wie in Abschnitt 3 bis 6 beschriebener modellbasierter Ansatz bietet den Vorteil der Robustheit durch die generierten Moden. Die Deformation/Anpassung eines generierten Modells ist durch die bekannten n Variationen eingeschränkt. Allerdings stellt dies auch eine Beschränkung der Freiheitsgrade dar. Daher ist oft nach der Anpassung eines Modells auf das entsprechende Bild ein Nachbearbeitungsschritt sinnvoll, um die restlichen Fehler zwischen $\mathbf{S}_{new}$ und $\mathbf{M}_{new}$ mit einer Operation mit einer größeren Anzahl an Freiheitsgraden auszugleichen. Da die vorgestellte Registrierung genau dieselbe Parameteranzahl wie das bekannte Modell aufweist, bietet sich ein anschließender Registrierungsschritt als Nachbearbeitung an. Dabei erweitert sich die Modellanpassung um eine weitere Mode $\mathbf{V}^{Reg}$:

$$\mathbf{M}_{new} = T_{new}^{-1}(\bar{\mathbf{M}} + \omega_{new} \mathbf{V} + \mathbf{V}^{Reg}). \quad (102)$$

Dies bedeutet, dass eine Registrierung zwischen $\mathbf{S}_{new}$ und $\bar{\mathbf{M}} + \omega_{new} \mathbf{V}$ durchgeführt wird, wobei T_{new} bereits durch die Modellanpassung (Optimierung von T_{new} und ω_{new}) initialisiert worden ist. In Abschnitt 8.4 wird die probabilistische Registrierung als Nachverarbeitung nach der Modellanpassung des generierten Handmodells an Bilddaten angewendet und evaluiert.

8. Experimente

Dieser Abschnitt der Arbeit stellt verschiedene Experimente vor, in denen das beschriebene Framework angewendet wurde. Die Methode ist von der Wahl der Appearance-Merkmale und von der Art der Positionierung der Merkmalsvektoren unabhängig. Beides kann anwendungsabhängig gewählt werden. In den Abschnitten 8.1 und 8.2 werden zuerst mögliche Appearance-Merkmale und die genutzte Positionierung vorgestellt. Abschnitt 8.3 fasst die angewendeten Metriken für die Evaluierung zusammen und die Abschnitte 8.4 bis 8.6 beschreiben Anwendungsfälle und deren Ergebnisse. Dabei wurden Röntgen-Bilder der Hand und 2D bzw. 3D MRT-Bilder des Gehirns genutzt.

8.1. Merkmalsauswahl

Die Auswahl der Appearance-Merkmale ist wie bei den meisten Bildanalysemethoden ausschlaggebend für eine gute Modellierung der Daten. Die Merkmale sollten zum einen die individuellen Daten aussagekräftig beschreiben, zum anderen Gemeinsamkeiten eines gesamten Trainingsdatensatzes hervorheben.

8.1.1. Signalwerte

Die direkt akquirierten Signalwerte der medizinischen Bilder sind die einfachsten Merkmale. Durch die spärliche Abtastung der Bilder ist in jedem Falle eine Gauß-Glättung der Daten sinnvoll. Diese kann abhängig vom Merkmalsvektorabstand sp erfolgen:

$$f_{i,.} = (G_{\sigma=sp} * I)(\mathbf{x}_i). \quad (103)$$

Für die vorgestellten Modelle wurde keine Anpassung (wie Histogrammangleichung) der Daten eines Trainingsdatensatzes vorgenommen.

8.1.2. Gabor-Wavelets

Gradienten-/Kantenfilter stellen eine einfache und effektive Möglichkeit dar, bestimmte Strukturinformationen im Bild hervorzuheben. Jedoch gehen dabei Informationen verloren. Die Gabor-Wavelet Filter bieten die Möglichkeit, ein Bild in lokale Frequenzbereiche zu unterteilen. Die sogenannte Motherwavelet [Lee 1996] für 2D Bilder

$$\Psi_{2D}(\mathbf{x}) = \Psi_{2D}(x, y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x^2 + y^2)\right) (\exp(i\kappa x) - \exp(-\kappa^2/2)) \quad (104)$$

wird dabei in rotierter ($or = 0 \dots N_{or} - 1$) und skaliertem ($sc = 0 \dots N_{sc} - 1$) Form als Filter genutzt:

$$f_{i,.} = (I * \Psi_{2D}^{or,sc})(\mathbf{x}_i) \quad (105)$$

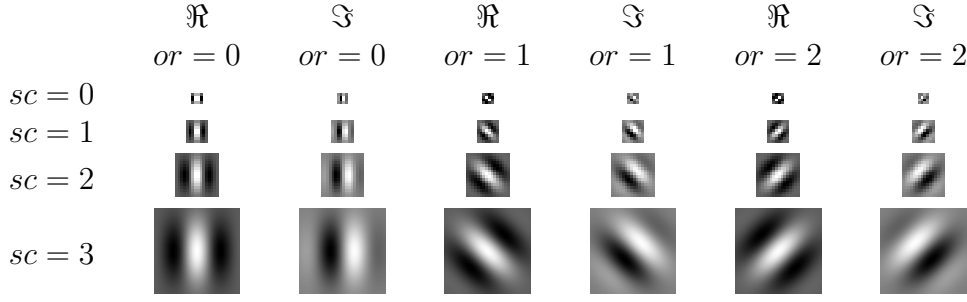


Abbildung 44: Beispiele für Gabor-Wavelet-Filter mit drei Rotationswinkeln und vier Skalierungsstufen mit jeweils Real- und Imaginäranteil

mit den Rotationswinkeln $\theta = or \cdot (\pi/N_{or})$ und den Skalierungsstufen $a = a_0^{sc}$ ($a_0 > 1$):

$$\Psi_{2D}^{\theta,a}(x, y) = \Psi_{2D} \left(\frac{x \cos(\theta) + y \sin(\theta)}{a}, \frac{-x \sin(\theta) + y \cos(\theta)}{a} \right). \quad (106)$$

Da Ψ_{2D} komplex ist, werden in der Anwendung zwei Filter pro Skalierung und Rotation genutzt – Imaginär- und Realanteil:

$$\Im(\Psi_{2D})(x, y) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2}(x^2 + y^2) \right) \sin(\kappa x) \quad (107)$$

$$\Re(\Psi_{2D})(x, y) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2}(x^2 + y^2) \right) (\cos(\kappa x) - \exp(-\kappa^2/2)) \quad (108)$$

Abbildung 44 zeigt Beispielfilter für drei Rotationswinkel und vier Skalierungen. Die Anzahlen der Skalen und Winkel werden abhängig von den Bildern gewählt. Für eine komplette Beschreibung der Bilddaten wird a_0 meist auf 2 gesetzt, was eine Verdopplung der Filtergröße mit jeder Skalierungsstufe bedeutet ($a = a_0^{sc}$). Die letzte Stufe (größter Filter) kann so gewählt werden, dass die resultierende Filtergröße die Hälfte des Bildes ausmacht. Die Anzahl der Winkel kann zum Beispiel auf 8 gesetzt werden, um eine sehr gute Beschreibung der Daten zu gewährleisten (siehe [Lee 1996]).

Von Interesse bei den Gabor-Filtern ist die Möglichkeit einer Rekonstruktion der originalen Bilddaten aus den gefilterten Bildmerkmalen. Da die Fouriertransformierte der Gabor-Wavelets ebenfalls eine Gabor-Wavelet ist, wird das gefilterte Bild ($I * \Psi_{2D}^{or,sc}$) für die Rekonstruktion erneut gefiltert:

$$\begin{aligned} \hat{I} &= \sum_{or} \sum_{sc} (I * \Psi_{2D}^{or,sc}) * \Psi_{2D}^{or,sc} \\ &= \sum_{or} \sum_{sc} (I * \Im(\Psi_{2D}^{or,sc})) * \Im(\Psi_{2D}^{or,sc}) + (I * \Re(\Psi_{2D}^{or,sc})) * \Re(\Psi_{2D}^{or,sc}). \end{aligned} \quad (109)$$

Bei ausreichend großer Filteranzahl (große Frequenzabdeckung) gehen so im Gegensatz

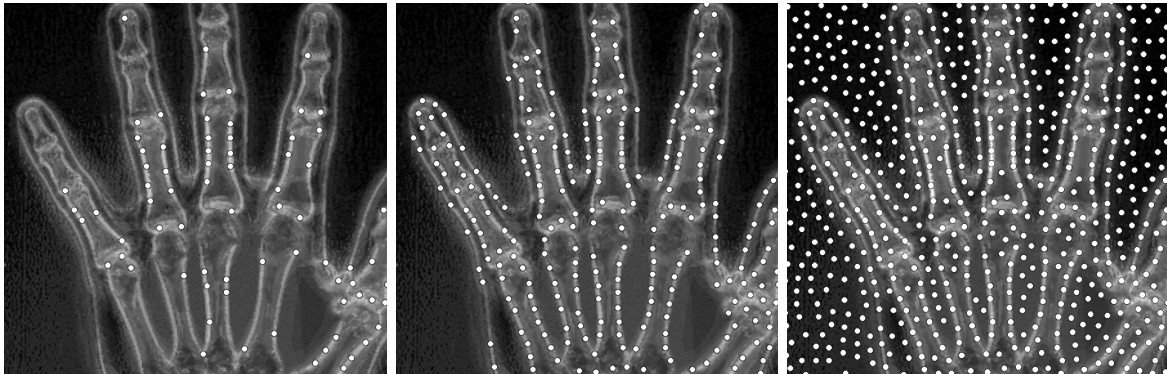


Abbildung 45: Abtastung eines Bildes nach absteigender Distinctiveness: Zuerst (links) werden die Merkmalsvektoren mit höchster Aussagekraft (hier: Summe aller Gabor-Filterantworten) gewählt. Mit absteigender Distinctiveness wird das ganze Bild abgetastet – unter der Bedingung, dass ein Mindestabstand zwischen dem neuen Merkmalsvektor und allen bereits gewählten Vektoren eingehalten wird.

zu einfachen Gradienten-/Kantenfiltern kaum Informationen verloren.

Es sei angemerkt, dass bei der klassischen Anwendung der Gabor-Wavelets die Filter nicht überlappend angewendet werden, das heißt, je größer der Filter, desto weniger Merkmale werden extrahiert. Da bei der vorgestellten Bildrepräsentation jedoch jeder Merkmalsvektor dieselbe Anzahl an Merkmalen enthält – unabhängig von dessen Position – wird hier jede Bildposition mit allen Filtermasken (überlappend) gefiltert.

8.1.3. Segmentierungsmerkmale: Distanzkarten

Wie in Abschnitt 6.2 beschrieben, werden Distanzkarten genutzt, um Objektkonturen oder Landmarken zu repräsentieren. Dies hat den Vorteil, dass auch Merkmalsvektoren, die nicht auf bzw. in der Nähe der gesuchten Kontur liegen, Informationen über deren Position beinhalten. Die Voraussetzung für die Berechnung einer Distanzkarte ist eine vorhandene Objektsegmentierung/-maske für die Daten. Ungeachtet des Zwecks der Segmentierung können solche Merkmale ebenfalls als reguläre Appearance-Merkmale eingesetzt werden. Diese führen während der Modellierung zu einer Stabilisierung: Sind die Signalwertbereiche der Bilder des Trainingsdatensatzes beispielsweise sehr unterschiedlich, können rein strukturbasierte Merkmale die Korrespondenzfindung robuster gestalten – besonders in den ersten Iterationen der Modellgenerierung.

8.2. Positionen der Merkmalsvektoren

Der vorgestellte Algorithmus ist von der Wahl der Positionen der Merkmalsvektoren unabhängig. Wie bereits ausgeführt, ist die Anzahl der Merkmalsvektoren ebenfalls für jedes Bild des Trainingsdatensatzes beliebig wählbar. Eine Motivation der Nutzung von Korrespondenzwahrscheinlichkeiten im Gegensatz zu Eins-zu-Eins-Korrespondenzen ist

das Wegfallen einer aufwendigen Landmarkenbestimmung in den Bildern. Eine komplett bildunabhängige Wahl der Merkmalsvektoren wäre ein regelmäßiges Grit oder randomisiertes Abtasten der Bilddaten. Um dennoch zu gewährleisten, dass wichtige Strukturen in den Bildern durch die Merkmalsvektoren repräsentiert werden, ist hier eine Abtastung der Merkmale in Abhängigkeit der Bilder gewählt worden. Dabei soll erreicht werden, dass das interessierende Objekt im Bild gleichmäßig abgetastet ist und außerdem wichtige und repräsentative Strukturen ohne manuelle Landmarkenpositionierung abgedeckt sind. Dafür wird eine Methode von Murphy et al. [Murphy *et al.* 2008] genutzt, bei der das Bild nach “Distinctiveness” (“Aussagekraft”) sortiert wird. Die Positionen der Merkmalsvektoren werden anschließend mit absteigender Distinctiveness gewählt. Als Bedingung gilt, dass zwischen den gewählten Merkmalsvektoren ein Mindestabstand (sp) eingehalten wird:

IN “Distinctivenessbild” $I^{dst} : \Omega \rightarrow \mathbb{R}$, minimale Distanz zwischen Vektoren sp

DO

sortiere Pixelkoordinaten absteigend mit $I^{dst}(\mathbf{x}_a) > I^{dst}(\mathbf{x}_{a+1}), \forall \mathbf{x}_a \in \Omega :$

$$\mathbf{X}^{sort} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_a, \dots, \mathbf{x}_{|\Omega|}]$$

initialisiere: $\mathbf{X}^{out} = \{ \}$

for $a = 1 : |\Omega|$

$a_is_featurevector = \mathbf{true}$

$i = 1$

while $a_is_featurevector == \mathbf{true} \ \& \ i < |\mathbf{X}^{out}|$

if $\|\mathbf{x}_a - \mathbf{x}_i\|_2 < sp$ mit $\mathbf{x}_i \in \mathbf{X}^{out}, \mathbf{x}_a \in \mathbf{X}^{sort}$

$a_is_featurevector = \mathbf{false}$

$i++$

if $a_is_featurevector == \mathbf{true}$

$$\mathbf{X}^{out} = \mathbf{X}^{out} \cup \mathbf{x}_a$$

OUT $\mathbf{X}^{out} = \{\mathbf{x}_i | i = 1, \dots, N_s\}$

Welche Merkmale als aussagekräftig gelten, muss pro Anwendungsfall geklärt werden. Eine einfache Wahl der Distinctiveness-Werte wäre die Nutzung des Gradientenbetrages der Bilddaten. Werden Gabor-Wavelets als Appearance-Merkmale genutzt, bietet es sich an, die Summe über alle Gabor-Filterantworten (alle Orientierungen und Skalierungen) zu wählen. Da die stärksten Filterantworten auf den Kanten im Bild liegen, entsteht ebenfalls ein “Kantenbild”. Abbildung 45 zeigt ein solches Distinctiveness-Bild. Die weiß

markierten Punkte zeigen, dass zuerst Merkmalsvektoren an Stellen mit hoher Aussagekraft gesetzt werden (Kanten der Fingerknochen oder der Hand). Dies gewährleistet, dass bei allen Bildern des Trainingsdatensatzes bestimmte Strukturen (in diesem Fall die Knochen oder Handkontur) auf ähnliche Weise repräsentiert werden. Die genauen Positionen stimmen jedoch zwischen den Bildern des Trainingsdatensatzes nicht exakt überein. Für klassische Form- oder Appearance-Modellansätze, die auf Eins-zu-Eins-korrespondierenden Landmarken basieren, muss in allen Bildern ein Merkmalsvektor genau auf z. B. der Fingerspitze liegen. Durch die Nutzung der sortierten Distinctiveness ist ausschließlich gegeben, dass die “Kontur der Fingerspitze” (mit einer Genauigkeit von sp) mit gleicher Priorität in allen Bildern repräsentiert ist, da die Gradientenbeträge dort z. B. höher sind als im Bildhintergrund. Durch die probabilistischen Korrespondenzen ist dies dennoch vollkommen ausreichend. Sind außerdem möglichst viele unterschiedliche Varianten an Positionierungen der Merkmalsvektoren auf der z. B. Handoberfläche durch die Trainingsmenge abgedeckt, wird die Anpassung des Modells an Bilddaten mit “ungenau” gesetzten Merkmalsvektoren robuster.

Die Wahl von sp ist abhängig von den Bilddaten, dem betrachteten Objekt und der Anwendung:

- Wie groß ist das Objekt?
- Wie “genau” sollen die Signalwerte rekonstruierbar sein? (Abschnitt 6.1)
- Für Segmentierungen: Wie “fein” sind die zu segmentierenden Objekte deren jeweilige Kontur durch die Distanzkarte beschrieben wird/rekonstruiert werden soll? (Abschnitt 6.2.1)
- Für Landmarkendetektion: Wie weit liegen die Landmarken, die durch jeweils eine Distanzkarte beschrieben werden, auseinander? (Abschnitt 6.2 und 6.3.1)

Neben dem Merkmalsvektorabstand kann ebenfalls der Bereich, der mit Merkmalsvektoren abgedeckt werden soll, angepasst bzw. eingeschränkt werden. In Abschnitt 6.4 wurde bereits diskutiert, dass für die Modellierung von Objekten die Ausschließung von Merkmalsvektoren im Objekthintergrund das Modell stabiler machen kann. Eine andere Möglichkeit, die Merkmalsvektoren auf “interessante” Bereiche einzuschränken, ist die Nutzung eines Schwellwerts für die Distinctiveness-Werte:

```

IN ...
  for  $a = 1 : |\Omega|$ 
    if  $I^{dst}(\mathbf{x}_a) > threshold$ 
      ...

```

Dies führt dazu, dass Bereiche mit geringer Distinctiveness (z. B. geringen Gradientenbeträge) nicht abgetastet werden. Darüber hinaus kann der Merkmalsvektorabstand sp

gestaffelt angepasst werden: je geringer die Werte in I^{dst} , desto größer sp . Dadurch werden Bereiche, die wenig strukturelle Informationen (Gradienteninformationen) beinhalten, weniger detailliert abgetastet.

8.3. Metriken

Um die Performanz der Modelle quantitativ zu evaluieren, wurden die im Folgenden vorgestellten Methoden und Metriken genutzt. Diese zeigen, wie gut das Modell die Trainingsdaten bzw. neue Observationen beschreibt.

8.3.1. Konturdistanzen

Wie in Abschnitt 6.2.1 erläutert, können aus den Segmentierungsmerkmalen komplette Distanzkarten (I_f^{rec}) rekonstruiert werden, bei denen der Nulldurchgang die Kontur des gesuchten Objekts beschreibt. Um die “mittlere symmetrische Konturdistanz” ($MSKD$) zwischen rekonstruierter Distanzkarte (I_f^{rec} , Modellanpassung) und gegebener Distanzkarte (I_f , Groundtruth, Trainingsdaten,...) zu berechnen, werden die entsprechenden Distanzen aufsummiert:

$$\begin{aligned} MSKD(I_f, I_f^{rec}) &= \frac{1}{|\mathbf{x}|} \sum_{\mathbf{x}} |I_f(\mathbf{x})| \quad \forall \mathbf{x} : I_f^{rec}(\mathbf{x}) = 0 \\ &+ \frac{1}{|\mathbf{x}_2|} \sum_{\mathbf{x}_2} |I_f^{rec}(\mathbf{x}_2)| \quad \forall \mathbf{x}_2 : I_f(\mathbf{x}_2) = 0 \end{aligned} \quad (110)$$

8.3.2. Punktdistanzen

Abschnitt 6.3.1 beschreibt, wie Landmarkenpositionen mittels Hough-Transform aus den Distanzwerten rekonstruiert werden. Anschließend kann für diese Landmarken erneut eine Distanzkarte (I_f^{rec}) berechnet werden, die ebenso wie im vorangegangenen Paragraphen mit den gegebenen Landmarken (durch die Distanzkarte I_f beschrieben) mit der “mittleren symmetrischen Punktdistanz” ($MSPD$) verglichen wird:

$$\begin{aligned} MSPD(I_f, I_f^{rec}) &= \frac{1}{|\mathbf{x}|} \sum_{\mathbf{x}} |I_f(\mathbf{x})| \quad \forall \mathbf{x} : I_f^{rec}(\mathbf{x}) = 0 \\ &+ \frac{1}{|\mathbf{x}_2|} \sum_{\mathbf{x}_2} |I_f^{rec}(\mathbf{x}_2)| \quad \forall \mathbf{x}_2 : I_f(\mathbf{x}_2) = 0 \end{aligned} \quad (111)$$

8.3.3. Dice-/Jaccard-Koeffizienten

Die beiden vorangegangenen Metriken sind sinnvoll, wenn die genaue Kontur/Landmarken-Position des gesuchten Objekts bekannt ist. Wird jedoch nach einem komplett unbekannten Objekt gesucht, wie z. B. bei der Klassifikationsproblematik von gesundem

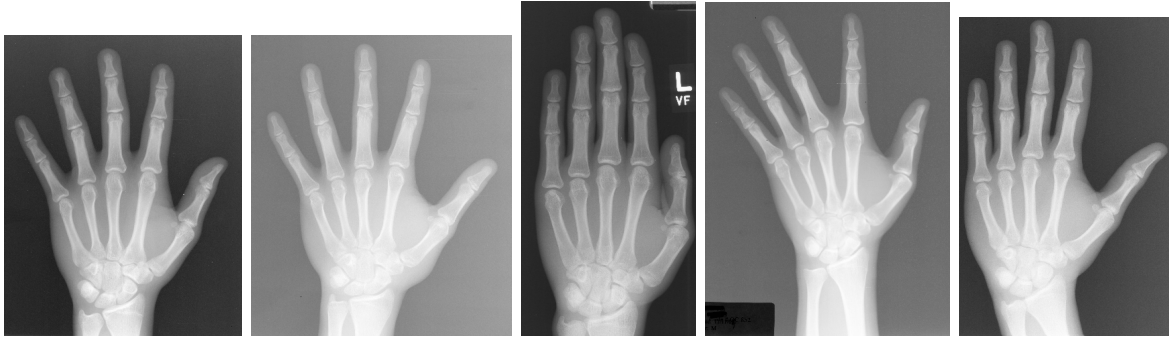


Abbildung 46: Fünf Beispiele der verwendeten 2D Handdaten: Die Bildgröße und Signalwertverteilung variieren in den Daten.

und pathologischem Gewebe, ist nicht vorrangig von Interesse, wie genau die Kontur der ermittelten Segmentierung zur Groundtruth passt. Von größerer Bedeutung ist die Frage, wie viel vom gesuchten Gewebe auch korrekt gefunden wurde. Daher sind hier Metriken, die Mengen beschreiben wie der Dice- (DC) oder Jaccard-Koeffizient (JC), interessanter. Die zu vergleichenden Bilder/Merkmale (I_f, I_f^{rec}) beschreiben in dem Fall ein Labelbild mit einem Labelwert (LV) pro gesuchtem Objekt:

$$JC(I_f, I_f^{rec})^{LV} = \frac{|\{x | I_f(x) = LV\} \cap \{x | I_f^{rec}(x) = LV\}|}{|\{x | I_f(x) = LV\} \cup \{x | I_f^{rec}(x) = LV\}|} \quad (112)$$

$$DC(I_f, I_f^{rec})^{LV} = \frac{2|\{x | I_f(x) = LV\} \cap \{x | I_f^{rec}(x) = LV\}|}{|\{x | I_f(x) = LV\}| + |\{x | I_f^{rec}(x) = LV\}|} \quad (113)$$

8.4. Experiment 1: Anwendung zur Handsegmentierung und Landmarkendetektion

Um die Performanz des vorgestellten Modellansatzes zur Segmentierung und Landmarkendetektion zu evaluieren, wurden 120 2D Hand Röntgenbilder der *Digital Hand Atlas Database*¹ genutzt (siehe Beispielbox 1). Es wurden Bilder von 16- bis 18-jährigen männlichen und weiblichen Jugendlichen ausgewählt. Für alle 120 Bilder wurden die Hände manuell segmentiert und 19 Landmarken zwischen die Fingerknochen gesetzt. 60 der Bilder wurden mit den Handmasken und Landmarken für das Training eines Modells eingesetzt und die weiteren 60 Bilder wurden zur Evaluierung genutzt, wobei die gegebenen Handmasken und Landmarken als Groundtruth für die Evaluierung eingesetzt wurden. Die Analyse der Knochen und der Knochenzwischenräume in Röntgenbildern der

¹<http://www.ipilab.org/BAAWeb>

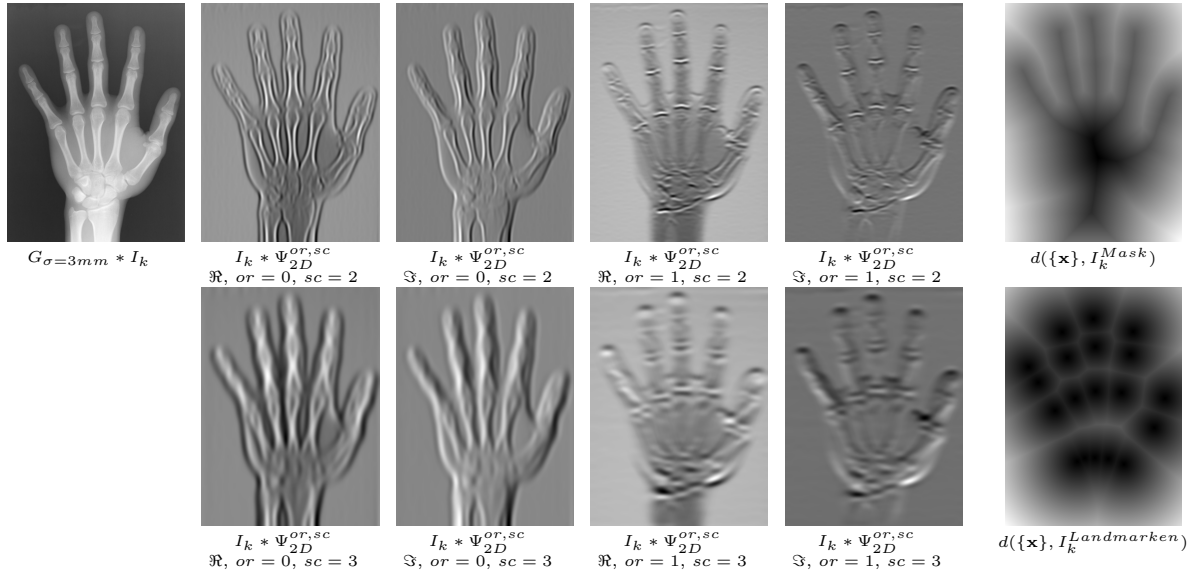


Abbildung 47: Elf genutzte Appearance- bzw. Segmentierungsmerkmale für die modellbasierte Handsegmentierung: (1) Gauß-gefiltertes Röntgenbild (2-9) Gabor-gefilterte Bilder (zwei Skalierungen, zwei Rotationen, Real-/Imaginär-Teil) und (10-11) zwei Segmentierungsmerkmale (rechts) für die Handkontur und 19 Landmarken auf den Fingerknochen.

Hand kann für die Altersbestimmung bei Kindern und Jugendlichen eingesetzt werden. Für eine solche Anwendung ist eine automatische Detektion der Fingerknochenverbindungen von Interesse. Die Segmentierung der gesamten Hand kann als Vorverarbeitungsschritt der Knochenanalyse genutzt werden. Insbesondere wird die Handsegmentierung im folgenden Experiment dafür genutzt, die Qualität der Anpassung des generierten Modells an ein Bild zu bewerten.

Die größte Motivation für die vorliegende Arbeit die Handdaten für die Evaluierung der Modellgenerierung und Modellanpassung zu nutzen, ist die große Variabilität der Daten: Die Hände weisen keine gleichmäßige Handstellung (siehe Abbildung 46) auf und die Signalwerte der Bilder sind nicht aneinander angeglichen. Außerdem haben die Bilder alle unterschiedliche Größen. Um den Rechenaufwand zu verringern, wurden die Bilder auf einen Pixelabstand von $0.5 \times 0.5 \text{ mm}$ downgesampelt. Die durchschnittliche Bildgröße liegt anschließend bei: $332.90 (\pm 43.83) \times 460.13 (\pm 35.84)$ Pixel. Artefakte (‘‘schwarze Boxen’’ oder Beschriftungen im Bild) wurden nicht entfernt.

Modell: Aus 60 Handdaten wurde ein Modell generiert. Da die Bilder vor allem durch die Knochen definiert sind und die Signalwerte selbst wenig Aussagekraft haben, wurden die Gabor-Filter zusätzlich zu den Signalwerten als Appearance-Merkmale genutzt. Abbildung 47 zeigt elf genutzte Appearance-Merkmale: (1) Gauß-gefiltertes Röntgenbild (2-9) Gabor-gefilterte Bilder (zwei Skalierungen, zwei Rotationen, Real-/Imaginär-Teil) und (10-11) zwei Segmentierungsmerkmale (rechts), um die Handkontur und 19 Land-

marken, die zwischen den Fingerknochen liegen, zu beschreiben. Die Handmaske und die Landmarken wurden für alle Bilder manuell erstellt. Um die Positionen der Merkmalsvektoren zu bestimmen, wurde die Summe über alle Gabor-gefilterten Bilder als Distinctiveness gewählt (siehe Abschnitt 8.2). Außerdem sind die Merkmalsvektoren für die Trainingsdaten auf das Innere und die direkte Umgebung der Hand eingeschränkt ($\forall i : f_{i,1}^{seg} < 5mm$). Da das Modell zur Segmentierung eingesetzt wird und nicht zur Rekonstruktion der gesamten Bilddaten, wurde hier nur eine relativ geringe Anzahl an Gabor-Wavelet-Filtern ($sc=2, 3$) genutzt. Die feinsten Skalierungslevel ($a = 2^{sc=0}$ und $a = 2^{sc=1}$) werden nicht verwendet, da sie vor allem Rauschen in den Bildern beschreiben. Außerdem repräsentieren die Gabor-Wavelet-Filter die Daten erst vollständig, wenn die Filtermaske größer ist als der Merkmalsvektorabstand. Größere Skalierungslevel beschreiben vor allem grobe Bildstrukturen, wohingegen $sc=2$ und 3 die Knochen gut repräsentieren (Abbildung 47). Um die Anzahl der Merkmale gering zu halten, wurden zwei Rotationswinkel gewählt.

Die folgende Tabelle fasst alle genutzten Parameter für die Modellgenerierung zusammen:

Input-/Trainingsdaten	
I_k	: Röntgen-Bilder mit Pixelabstand $0.5\text{ mm} \times 0.5\text{ mm}$
N	= 60
D	= 13 ($D_f = 9, D_f^{seg} = 2, D_x = 2$)
$\mathbf{f}_i$	: (1) Gauß-geglättete Signalwerte: $(G_{\sigma=3mm} * I_k)(\mathbf{x}_i)$ Abs. 8.1
	: (2-9) Gabor-Wavelet-gefilterte Bilder ($sc=2, 3, or=0, 1$)
$\mathbf{f}_i^{seg}$	: (1) Distanzwerte zur Handmaske: $d(\mathbf{x}_i, I_k^{Mask})$ Abs. 6.2
	: (2) Distanzwerte zu Fingerknochen: $d(\mathbf{x}_i, I_k^{Landmarken})$
$\mathbf{x}_i$	: Distinctiveness-abhängig mit Mindestabstand 3 mm , Abs. 8.2
	innerhalb der Hand ($\forall i f_{i,1}^{seg} < 5mm$)
sp	= (9, 7, 5) 3 mm (letztes Level)
$N_s(k)$	= 1744 ± 230 (letztes Level)
N_m	= 1837 (letztes Level)
Modellparameter	
n	= 10-30 Abs. 5.2
$T_k \star \mathbf{s}_{k,i}$	= $(T_k(\mathbf{x}_i), \mathbf{f}_i)$ mit $(\mathbf{S}_k - T_k^{-1}(\bar{\mathbf{M}})) \rightarrow_{T_k} \min$ Abs. 4.1.7
Σ	= $\frac{1}{N} \sum_{k=1}^N \text{diag}(\text{Cov}(\mathbf{S}_k, \mathbf{M}_k) \cdot 0.5 + \text{Cov}(\mathbf{S}_k, \bar{\mathbf{M}}) \cdot 0.5)$ Abs. 4.2.1
$\Sigma_{\mathcal{N}}$	= $(1, 1, 10, 10, \dots, 10) \text{diag}(\text{Cov}(\bar{\mathbf{M}}, \bar{\mathbf{M}} \pm 2 \cdot \lambda \mathbf{V}))$ Abs. 4.2.2
$ \mathcal{N}(j) $	≈ 4 Abs. 4.3.3
N_l	= 4 Abs. 5.1

Experimente: Das generierte Modell wurde anschließend auf die 60 bekannten Trainingsbilder und auf 60 unbekannte Bilder angewendet. Dabei sind die beiden Segmentierungsmerkmale während der Modellanpassung unbekannt (siehe Abschnitt 6.2). Stattdessen werden die Distanzkarten aller 120 Daten als Groundtruth für die Evaluierung nach der Modellanpassung genutzt.

E1.1: Um die Modellgenerierung zu evaluieren und zu zeigen, dass das Modell an die Trainingsdaten, aus denen es generiert wurde, wieder angepasst werden kann, wurden zuerst die bekannten 60 Trainingsdaten ($I_k, k = 1, \dots, N$) zur Evaluierung der Segmentierung/Landmarkendetektion genutzt. Die Segmentierungsmerkmale werden dabei während der Modellanpassung nicht verwendet (Abschnitt 6.2). Wie in Abbildung 38 beschrieben, werden die Merkmalsvektoren im ganzen Bild gesetzt und dann iterativ gelöscht, wenn sie zu weit von der adaptierten Modellinstanz entfernt liegen (Abschnitt 6.4).

E1.2: Um zu evaluieren, ob diese iterative Bestimmung/Löschung der Merkmalsvektoren in E1.1 das Ergebnis beeinflusst, wurden die Trainingsbilder (bzw. $\mathbf{S}_k$) ein weiteres Mal mit der Untermenge an Vektoren innerhalb der Hand, die bereits für die Modellgenerierung genutzt wurden, segmentiert.

E1.3: Die Evaluierung der Segmentierung/Landmarkendetektion von neuen Daten wurde an 60 unbekannten Bilddaten durchgeführt. Für die Modellanpassung wurden wie in E1.1 die Merkmalsvektoren für $\mathbf{S}_{new}$ im ganzen Bild gesetzt und dann iterativ gelöscht (Abschnitt 6.4).

E1.4: Zusätzlich wurde nach Modellanpassung an $\mathbf{S}_{new}$ ein probabilistischer registrierungsbasierter Nachverarbeitungsschritt zwischen $\mathbf{S}_{new}$ und $\mathbf{M}_{new}$ durchgeführt (siehe Abschnitt 7.4).

Ergebnisse: Alle Ergebnisse (zusammengefasst in Tabelle 2) wurden mit der Groundtruth-Handmaske bzw. -Landmarken (GT) verglichen und über die jeweiligen 60 Trainingsdaten (E1.1, E1.2) und die 60 unbekannten Testdaten (E1.3, E1.4) gemittelt. Für beide Datensätze (bekannt und unbekannt) wurden die Werte für die Ausgangssituation vor der Modellanpassung berechnet (GT vs. $\bar{\mathbf{M}}$). Nach Modellanpassung wurden einmal die Zwischenergebnisse mit angewandter affiner Transformation (GT vs. $T_{k/new}^{-1}(\bar{\mathbf{M}})$) und die Endergebnisse mit zusätzlich angepassten Moden bestimmt. Dabei beschreibt “ GT vs. $\mathbf{M}_{k/new}$ ” die Gegenüberstellung der aus $\{\mathbf{f}_j^{seg}\}$ rekonstruierten Distanzkarte und “ GT vs. $\mathbf{S}_{k/new}$ ” die aus $\{\mathbf{f}_i^{seg} | \mathbf{f}_i^{seg} = \sum_{j=1}^{N_m} \mathbb{E}(H_{ij}) \mathbf{f}_j^{seg}\}$ rekonstruierte Distanzkarte zur Groundtruth (siehe Abschnitt 6.2 und Abbildung 33 für die Gegenüberstellung von $\mathbf{f}_j^{seg}$ und $\mathbf{f}_i^{seg}$).

Dabei wird deutlich, dass jeweils eine eindeutige Verbesserung zwischen allen vier Werten (vier Zeilen für jeweils E1.1 und E1.3) stattfindet. Die Verbesserung zwischen den jeweils letzten beiden Werten (GT vs. $\mathbf{M}_{k/new}$ und GT vs. $\mathbf{S}_{k/new}$) zeigt, dass die Mittelung der Distanzwerte über die probabilistischen Korrespondenzen eine weitere Verbesserung darstellt (Abb. 33).

E1.1 vs. E1.2: Die Qualität der erzielten Ergebnissen für den Trainingsdatensatz ist unabhängig davon, ob die Merkmalsvektoren vor der Modellanpassung für $\mathbf{S}_k$ “korrekt” initialisiert werden (E1.2) oder erst iterativ reduziert werden (E1.1). Dies zeigt, dass die Initialisierung der Merkmalsvektoren für die Modellanpassung robuste Ergebnisse liefert (Abschnitt 6.4).

	MSKD (mm)	MSPD (mm)
E1.1: Trainingset I_k mit allen Merkmalsvektoren		
GT vs. $\bar{\mathbf{M}}$	3.38 (± 0.74)	8.49 (± 2.70)
GT vs. $T_k^{-1}(\bar{\mathbf{M}})$	1.69 (± 0.46)	2.58 (± 0.89)
GT vs. $\mathbf{M}_k$	0.58 (± 0.17)	0.93 (± 0.23)
GT vs. $\mathbf{S}_k$	0.28 (± 0.07)	0.67 (± 0.18)
E1.2: Trainingset I_k mit Untermenge der Merkmalsvektoren		
GT vs. $\mathbf{M}_k$	0.58 (± 0.17)	0.93 (± 0.24)
GT vs. $\mathbf{S}_k$	0.29 (± 0.08)	0.69 (± 0.19)
E1.3: unbekannte Bilder I_{new} mit allen Merkmalsvektoren		
GT vs. $\bar{\mathbf{M}}$	3.11 (± 0.85)	10.49 (± 3.05)
GT vs. $T_{new}^{-1}(\bar{\mathbf{M}})$	1.80 (± 0.38)	2.80 (± 0.77)
GT vs. $\mathbf{M}_{new}$	0.93 (± 0.25)	1.46 (± 0.30)
GT vs. $\mathbf{S}_{new}$	0.40 (± 0.12)	1.00 (± 0.18)
E1.4: zusätzliche Registrierung		
GT vs. $\mathbf{M}_{new}$	0.49 (± 0.15)	0.99 (± 0.20)
GT vs. $\mathbf{S}_{new}$	0.38 (± 0.10)	0.98 (± 0.18)

Tabelle 2: Ergebnisse der Segmentierung und der Landmarkendetektion für die Handdaten: für die Segmentierung der Handkontur ist jeweils die mittlere symmetrische Konturdistanz (MSKD) und für die Detektion der Landmarken auf den Fingerknochenzwischenräumen ist die mittlere symmetrische Punktdistanz (MSPD) angegeben.

E1.1 vs. E1.3: Die Ergebnisse für beide Datensätze sind deutlich geringer als die Distanzen der Ausgangssituation. Für bekannte Bilder (E1.1) sind die Distanz-Werte am kleinsten (0.29 mm), wobei für unbekannte Bilder die Konturdistanzen mit 0.40 mm ebenfalls deutlich sowohl unter dem Merkmalsvektorabstand (3 mm) des Modells als auch unter dem Pixelabstand (0.5 mm) der Bilder liegen. Das Modell ist demnach in der Lage, unbekannte Bilder zu segmentieren.

Die Werte für die Detektion der Fingerknochen liegen mit $\approx 1\text{ mm}$ über den Werten der Konturdistanzen. Dies liegt an der Art der Rekonstruktion der Landmarkenpositionen (Abschnitt 6.3.1) und der Tatsache, dass die Distanzwerte, welche die Landmarken beschreiben, lediglich positive Werte aufweisen: Weichen einzelne Merkmalsvektoren der angepassten Modellinstanz innerhalb und außerhalb der Hand vom “Ideal-” Wert ab, so kann der Nulldurchgang für die rekonstruierte/interpolierte Distanzkarte für die Objektkontur trotzdem auf der gesuchten Handkontur liegen. Für die Landmarkenbestimmen bedeutet eine Abweichung jedoch, dass der mit der Hough-Transformation gesuchte Punkt “breiter” wird.

E1.3 vs. E1.4: Die zusätzliche Registrierung verbessert die Ergebnisse – vor allem die Gegenüberstellung von GT und $\mathbf{M}_{new}$ ist nach der Registrierung von deutlich höherer

Qualität. GT vs. S_{new} zeigt mit der zusätzlichen Registrierung nur minimal bessere Ergebnisse, da die Mittelung über die Korrespondenzwahrscheinlichkeiten bereits einen ähnlichen Effekt wie die Registrierung hat (siehe Abschnitt 6.2 und Abbildung 33). Im Fall des Handmodells mit 30 Moden ist also nur eine relativ kleine “Korrektur” nach der Modellanpassung nötig, welche nicht größer ist als der Abstand der Merkmalsvektoren.

Die Werte der Landmarkendetektion sind kaum mit Werten in der Literatur zu vergleichen, da die meisten Algorithmen als Ziel die Altersbestimmung haben und die Detektion von z. B. Fingerknochen ausschließlich ein Zwischenschritt in den beschriebenen Algorithmen darstellt und daher die Genauigkeit dieser Detektion meist nicht umfassend evaluiert wird. In [Harmsen *et al.* 2013] basiert die Altersbestimmung z. B. auf den Regionen um die Fingerzwischenräume. In [Cootes *et al.* 2012] wird ein Appearance-Modell genutzt, welches die Fingerknochenzwischenräume als Landmarken für das zugrundeliegende Formmodell einsetzt. Die Genauigkeit wird ausschließlich mit “unter 1 mm für 90% der Daten” angegeben. In [Adeshina *et al.* 2017] wird ein aufwendiges mehrstufiges modellbasiertes Verfahren vorgestellt. Für die Detektion von Landmarken zwischen den Knochen wird eine Kombination aus zwei Appearance-Modellen und Random Forest Regression (um die korrekten Landmarkenpositionen zu lernen) angewendet. Die Genauigkeit für die Landmarkendetektion wird mit 0.87 mm angegeben. Für eine anschließende Segmentierung aller Handknochen wird eine Genauigkeit von “ungefähr 1 mm ” genannt. Selbst wenn die Resultate nicht direkt zu vergleichen sind, liegt die in der Arbeit erreichte Genauigkeit von ca. 1 mm für die Fingerknochendetektion und <0.5 mm für die Handkontur mit einer vergleichbar einfachen Methode (“nur” eine Modellanpassung ohne problemspezifische Nachverarbeitung) im Bereich der Ergebnisse der aktuellen Literatur.

8.5. Experiment 2: Detektion von Schlaganfallgewebe

Um die Klassifikation von gesundem und pathologischem Gewebe zu evaluieren, wurde ein Modell von 3D Gehirn MRT-Daten gebaut. Dafür wurden die Daten der ISLE Challenge² von 2015 genutzt [Maier *et al.* 2015]. Die Challenge stellt die Segmentierung von akut ischämischem Schlaganfallgewebe zur Aufgabe. 28 Daten stehen mit Groundtruth-Segmentierungen für das Training von Modellen bzw. von maschinellen Lernalgorithmen zur Verfügung. Weitere 36 Daten sind ohne bekannte Groundtruth-Segmentierungen gegeben. 8 dieser Daten wurden jedoch mit zum Trainingsdatensatz sehr unterschiedlichen MRT-Protokollen aufgenommen, so dass hier ausschließlich 28 der 36 Testdaten für die Evaluierung nutzbar sind. Es liegen die MRT-Sequenzen: FLAIR, DWI, T1 und T2 vor. Die Daten (Abbildung 48) haben teilweise sehr unterschiedliche Signalwertverteilungen. Die Vorverarbeitung der gegebenen Daten wird in [Maier *et al.* 2015] beschrieben.

²<http://www.isles-challenge.org/>

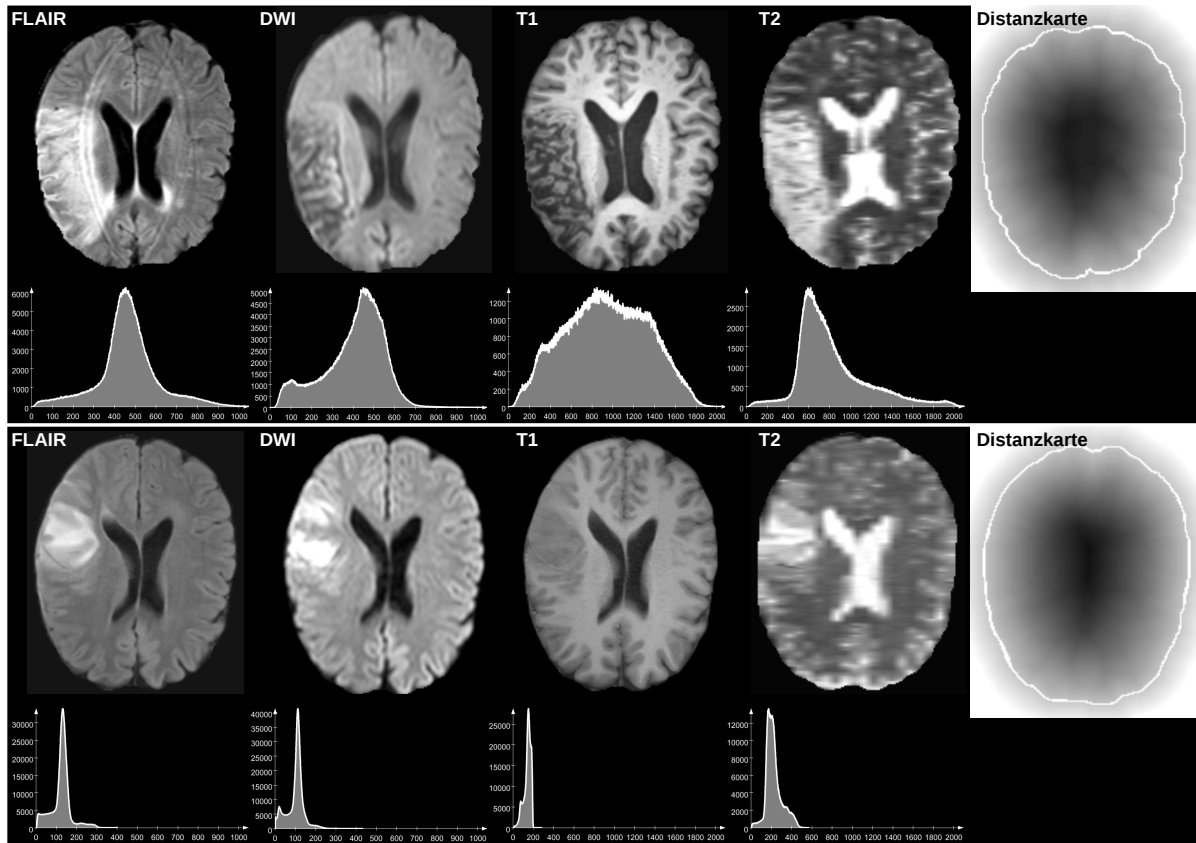


Abbildung 48: Appearance-Merkmale des 3D-Gehirnmodells: die beiden Zeilen visualisieren zwei Beispiele der gegebenen Gehirndaten. Die jeweils ersten vier Bilder zeigen die Signalwerte für die FLAIR, DWI, T1 und T2 Sequenzen, die als Appearance-Merkmale genutzt werden. Darunter sind jeweils die Histogramme der Daten abgebildet. Dabei wird deutlich, dass die Bilder stark unterschiedliche Signalwertverteilungen aufweisen. Die jeweils letzten Bilder der Zeilen zeigen das fünfte Appearance-Merkmal (die Distanzkarte zur Gehirnoberfläche), das zur Verbesserung der Stabilität genutzt wird.

Modell: Da Schlaganfallgewebe ein sehr heterogenes Erscheinungsbild und heterogene Lokalität aufweist, kann das Schlaganfallgewebe selbst nicht modelliert werden. Wie in Abschnitt 6.6 vorgestellt, wird daher ein Modell des gesunden Gehirns generiert. Da die 28 Trainingsdaten ausschließlich pathologische Daten enthalten, wird der in Abschnitt 6.7 vorgestellte Ansatz, bei dem partielle Daten zur Modellgenerierung genutzt werden, angewendet. Die für die Challenge vorliegenden Masken für die Schlaganfälle werden genutzt, um automatisch die Korrespondenzen für die nicht modellierten Gebiete auf Null zu setzen.

Das Modell wurde aus den 28 Daten mit den vier MRT-Sequenzen als Appearance-Merkmale gebaut. Als fünftes Merkmal wurde ebenfalls eine Distanzkarte zum Gehirnrand eingesetzt. Die Gehirnmasken können dafür automatisch generiert werden, da alle Bilder skull-stripped sind und der Hintergrund einheitlich den Signalwert Null hat [Maier

et al. 2015]. Dieses Merkmal wird nicht zur Segmentierung genutzt, sondern zur Stabilisierung, da die Signalwerte der Daten sehr unterschiedliche Verteilungen haben und in den ersten Iterationen dieser Unterschied erst durch die Moden ausgeglichen werden muss. Diese starken Signalwertunterschiede können zu fehlerhaften Korrespondenzwahrscheinlichkeiten führen, wohingegen die Distanzwerte für alle Daten des Trainingssatzes ähnlich sind. Für das Modell werden keine Gradienten- oder Gabor-Wavelet-Werte eingesetzt, da die gesuchten Strukturen (Schlaganfallläsionen) vor allem durch die abweichenden Signalwerte (z. B. Hyperintensitäten im FLAIR-Bild) gekennzeichnet sind. Um die Positionen der Merkmalsvektoren zu bestimmen, wurde die FLAIR-Sequenz als Distinctiveness gewählt. Dabei entsteht der Effekt, dass die Merkmalsvektoren zuerst auf die Bildpositionen mit hellen bzw. hyperintensiven Werten gesetzt werden. Die folgende Tabelle fasst alle genutzten Parameter für die Modellgenerierung zusammen:

Input-/Trainingsdaten

I_k	: MRT-Bilder (FLAIR, T1, T2, DWI) mit Pixelabstand $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$	
N	= 28	
D	= 8 ($D_f = 5$, $D_x = 3$)	
$\mathbf{f}_i$	: (1-4) Gauß-geglättete Signalwerte: $(G_{\sigma=3\text{mm}} * I_k)(\mathbf{x}_i)$ Abs. 8.1 (5) Distanzkarte zum Gehirnrand	
$\mathbf{x}_i$	: Distinctiveness-abhängig mit Mindestabstand 5 mm , Abs. 8.2 innerhalb des Gehirns ($\forall i f_{i,1}(FLAIR) > 0$)	
sp	= (20, 15, 10) 5 mm (letztes Level)	
$N_s(k)$	= 4811 ± 388 (letztes Level)	
N_m	= 4993 (letztes Level)	

Modellparameter

n	= 5-15	Abs. 5.2
$T_k \star \mathbf{s}_{k,i}$	= $(T_k(\mathbf{x}_i), \mathbf{f}_i)$ mit $(\mathbf{S}_k - T_k^{-1}(\bar{\mathbf{M}})) \rightarrow_{T_k} \min$	Abs. 4.1.7
Σ	= $\frac{1}{N} \sum_{k=1}^N \text{diag}(\text{Cov}(\mathbf{S}_k, \mathbf{M}_k) \cdot 0.5 + \text{Cov}(\mathbf{S}_k, \bar{\mathbf{M}}) \cdot 0.5)$	Abs. 4.2.1
$\Sigma_{\mathcal{N}}$	= $(1, 1, 1, 10, 10, \dots, 10) \text{diag}(\text{Cov}(\bar{\mathbf{M}}, \bar{\mathbf{M}} \pm 2 \cdot \lambda \mathbf{V}))$	Abs. 4.2.2
$ \mathcal{N}(j) $	≈ 6	Abs. 4.3.3
N_l	= 5	Abs. 5.1

Experimente: Wie in Abschnitt 6.6 angemerkt, wird die gesuchte Struktur (Läsion) in diesem Fall nicht als Merkmal modelliert, sondern aus den Korrespondenzwahrscheinlichkeiten zwischen Eingabebild und adaptierter Modellinstanz rekonstruiert. Es stellt sich die Frage nach dem Schwellwert, um die Korrespondenzwerte in gesund und pathologisch zu unterteilen. Hier wurde ein automatisches Regiongrowing-Verfahren eingesetzt: als Saatpunkte werden alle Merkmalsvektoren mit $p(\mathbf{s}_{\text{new},i} | \mathbf{M}_{\text{new}}) < \text{Quantil}_{0.1}$ gewählt. Der Schwellwert für das Regiongrowing-Verfahren wird ebenfalls automatisch bestimmt und auf den Quantilwert (Quantil_q) mit höchsten Anstieg zum nächsten Quantilen gesetzt ($\max_q(\text{Quantil}_{q+0.05} - \text{Quantil}_q)$), um den Korrespondenzschwellwert zu finden, der

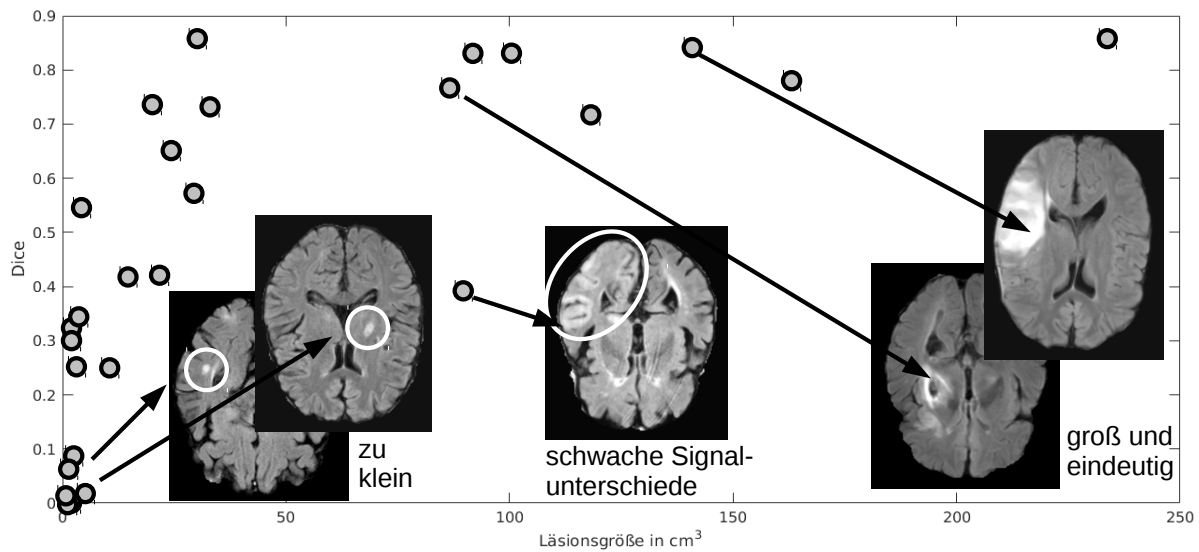


Abbildung 49: Ergebnisse der Gehirnklassifikation in Abhängigkeit von der Läsionsgröße in mm^3 (x-Achse): für größere Läsionen ($> 10 \text{ cm}^3$) kann die gesuchte Pathologie relativ gut durch das generierte Modell identifiziert werden. Für Läsionen ($< 10 \text{ cm}^3$), deren Volumen unter dem Abstand der Merkmalsvektoren liegt, ist das Modell zu ungenau. Außerdem können Läsionen, dessen Signalwerte lediglich gering von gesundem Gewebe abweichen, nicht ausreichend durch die Korrespondenzwerte abgegrenzt werden.

die Bereiche geringer Korrespondenz deutlich von den höheren Korrespondenzen abtrennt.

E2.1: Das Modell wird einmal auf die 28 Trainingsdaten angewendet. Dabei sei darauf hingewiesen, dass das Modell keinerlei Informationen über die Läsionen der Trainingsdaten enthält.

E2.2: Da eine einfache Abgrenzung mittels automatischem Regiongrowing-Verfahren zwischen gesunden und pathologischen Korrespondenzwerten gewählt wurde, werden zusätzlich manuelle Schwellwerte für die Segmentierung der Korrespondenzkarte gesetzt. Diese Werte werden lediglich nach optischer Auswertung der resultierenden Segmentierung der Schlaganfallregion gewählt.

E2.3: Darüber hinaus wurden die 28 von der Challenge bereitgestellten Testdaten evaluiert. Für diese liegen jedoch die Groundtruth-Masken nicht vor. Hier wurden ausschließlich die manuellen Regiongrowing-Schwellwerte genutzt.

Ergebnisse: Abbildung 49 visualisiert die Ergebnisse (Dice-Koeffizienten) für E2.1 in Abhängigkeit von der Läsionsgröße. Dabei wird deutlich, dass große Läsionen (Volumen $> 10 \text{ cm}^3$) durch das Modell gut identifiziert werden können. Für kleine Läsionen (Volumen $< 10 \text{ cm}^3$) ist das Modell jedoch eindeutig mit einem Abstand zwischen den Merkmalsvektoren von 5 mm zu grob.

E2.1: Die Ergebnisse für die 28 Trainingsdaten wurden anhand der gegebenen Maske

für die Schlaganfallsegmentierung in kleine Läsionen (13 Bilder) und große Läsionen (15 Bilder) unterteilt:

	alle (28)	klein $< 10 \text{ cm}^3$ (13)	groß $> 10 \text{ cm}^3$ (15)
Dice	0.45 (± 0.31)	0.17 (± 0.17)	0.69 (± 0.16)

Da das Modell mit einem minimalen Merkmalsvektorabstand von 5 mm ($\approx 12.5 \text{ cm}^3$) zu grob für die Detektion kleiner Läsionen ($< 10 \text{ cm}^3$) ist, sind die Dice-Koeffizienten für diese 13 Fälle entsprechend klein. Für große Läsionen hingegen liegen die Dice-Werte in einem für das Problem angemessenen Bereich.

E2.2: Durch eine manuelle Bestimmung der Schwellwerte können die Ergebnisse von E2.1 weiter verbessert werden:

	alle (28)	klein $< 10 \text{ cm}^3$ (13)	groß $> 10 \text{ cm}^3$ (15)
Dice	0.50 (± 0.30)	0.24 (± 0.25)	0.73 (± 0.14)

Dies zeigt, dass eine ausführlichere Betrachtung der Korrespondenzwerte als Indikatoren für pathologische Bereiche sinnvoll ist. Anstelle eines einfachen Schwellwertverfahrens könnten diese Werte z. B. als zusätzliches Merkmal für maschinelle Lernverfahren genutzt werden.

E2.3: Für die Testdaten liegen keine Groundtruth-Masken vor. Daher konnten die Daten nicht automatisch in kleine/große Läsionen unterteilt werden. Hier wurde eine manuelle grobe Unterteilung vorgenommen. Die Daten wurden über das von der Challenge bereitgestellte Evaluierungsportal ausgewertet.

	alle (28)	klein (12)	groß (16)
Dice	0.45 (± 0.33)	0.13 (± 0.14)	0.72 (± 0.13)

Die Ergebnisse zeigen ähnliche Werte wie bei den Trainingsdaten aus E2.2.

Für eine genauere Segmentierung ist die Generierung eines feineren Modells notwendig. Aufgrund der hohen Variabilität der Signalwerte in den Daten wäre außerdem eine größere Trainingsmenge von Vorteil. Die Ergebnisse zeigen dennoch, dass der Ansatz, pathologische Strukturen durch ein "gesundes Modell" zu identifizieren, funktioniert.

Die Challenge war vor allem an maschinelle Lernverfahren gerichtet. Bei diesen Verfahren wird in den meisten Fällen eine pixelweise Klassifikation der Daten vorgenommen. Das heißt, zum einen wird meist nicht die gesamte globale Gehirnform/-Erscheinung gelernt, sondern ausschließlich lokale Merkmale, wodurch ein Trainingsdatensatz von 28 Daten ausreicht, zum anderen ist die resultierende Klassifikation höher aufgelöst bzw. unabhängig von der Läsionsgröße, da jedes Pixel im Bild klassifiziert wird. Die besten drei Ergebnisse der gesamten Challenge weisen einen Dice-Wert von 0.47 bis 0.59 auf, wohingegen der Großteil der teilgenommenen Verfahren bei einem Dice-Wert von 0.35 bis 0.44 liegt³. Dies zeigt, dass es sich um ein sehr schweres Problem handelt. Des Weiteren sei darauf hingewiesen, dass hier ein genereller Ansatz und keine auf das Problem spezifisch angepasste Methode angewendet wurde.

³<http://www.isles-challenge.org/ISLES2015/>

8.6. Experiment 3: Probabilistische Registrierung von 2D MRT-Gehirnschichten

Die probabilistische Registrierung zweier Bilder (Abschnitt 7) wurde an 40 2D MRT-Schichten des “LPBA40”-Datensatzes [Shattuck *et al.* 2008] getestet. Dafür wurden korrespondierende traversale Schichten der affin vorregistrierten Volumen extrahiert. Für alle Bilder liegen Labelbilder mit Labelwerten für 20 segmentierte Strukturen vor (Abbildung 50).

Registrierung: Alle 40 Bilder wurden paarweise miteinander registriert, woraus sich 1560 Registrierungen ergeben. Als Appearance-Merkmal wurden die MRT-Signalwerte genutzt und die Merkmalsvektoren wurden ebenfalls mit den Signalwerten als Distinctiveness innerhalb des Gehirns gesetzt. Nach der Optimierung der Parameter wurden die Deformationsfelder aus der Mode und der affinen Transformation berechnet (Abschnitt 7.2) und auf die jeweiligen Labelbilder angewendet.

Input-/Trainingsdaten

I_k	: MRT-Bilder mit Pixelabstand $1\text{ mm} \times 1\text{ mm}$	
N	= 40 (pro “Modell” $N=1$)	
D	= 3 ($D_f = 1$, $D_x = 2$)	
$\mathbf{f}_i$	: (1) Signalwerte	
$\mathbf{x}_i$	: Distinctiveness-abhängig mit Mindestabstand 2 mm , Abs. 8.2 innerhalb des Gehirns ($\forall i f_{i,1} > 0$)	
sp	= (8, 7, 6, 5, 4, 3) 2 mm (letztes Level)	
N_s/N_m	= 3814 ± 216 (letztes Level)	

Modellparameter

n	= 1	Abs. 5.2
$T_k \star \mathbf{s}_{k,i}$	= $(T_k(\mathbf{x}_i), \mathbf{f}_i)$ mit $(\mathbf{S}_k - T_k^{-1}(\bar{\mathbf{M}})) \rightarrow_{T_k} \min$	Abs. 4.1.7
Σ	= $\frac{1}{N} \sum_{k=1}^N \text{diag}(\text{Cov}(\mathbf{S}_k, \mathbf{M}_k) \cdot 0.5 + \text{Cov}(\mathbf{S}_k, \bar{\mathbf{M}}) \cdot 0.5)$	Abs. 4.2.1
$\Sigma_{\mathcal{N}}$	= $(1, 1, 1) \text{diag}(\text{Cov}(\bar{\mathbf{M}}, \bar{\mathbf{M}} \pm 2 \cdot \lambda \mathbf{V}))$	Abs. 4.2.2
$ \mathcal{N}(j) $	≈ 4	Abs. 4.3.3
N_l	= 7	Abs. 5.1

Experimente: Für den Vergleich mit einem klassischen Registrierungsverfahren wurde die diffeomorphische variationelle nichtlineare Registrierung von Werner *et al.* gewählt [Werner *et al.* 2014]. Dieses Registrierungsverfahren befindet sich unter den effektivsten Methoden für den LPBA40-Datensatz. Dies wurde in [Ehrhardt *et al.* 2015] gezeigt, wobei der Registrierungsalgorithmus mit 14 bereits in [Klein *et al.* 2009] evaluierten Registrierungsmethoden verglichen wurde. Da diese Evaluierung auf den 3D Daten erfolgte, sind die Werte aus [Klein *et al.* 2009] nicht direkt vergleichbar.

E3.1: Daher werden in dieser Auswertung die Ergebnisse der probabilistischen Registrierung mit den Ergebnissen der variationellen Registrierung auf denselben 2D Schichten

				Dice		Jaccard	
	S	vs.	M	0.6097	(± 0.0671)	0.4611	(± 0.0646)
E3.1: varReg	S^{deformed}	vs.	M	0.7012	(± 0.0511)	0.5626	(± 0.0525)
E3.2: probReg	S^{deformed}	vs.	M	0.7096	(± 0.0433)	0.5710	(± 0.0444)
E3.3: probAAREg	S^{deformed}	vs.	M	0.7180	(± 0.0417)	0.5813	(± 0.0430)

Tabelle 3: Ergebnisse der Registrierung. Die Ergebnisse für beide probabilistische Registrierungen (prob(AA)Reg) sind signifikant besser als für die variationelle Registrierung (varReg) mit einem p -Wert $\ll 0.01$. Die Appearance-adaptierenden Registrierung ist ebenfalls signifikant besser als die probabilistische Registrierung ohne Appearance-Anpassung (p -Wert $\ll 0.01$).

des LPBA40-Datensatzes verglichen. Die Parameterwahl für die Anwendung der variationellen Registrierung auf 2D Schichten wird in [Uzunova *et al.* 2017] beschrieben.

Die vorgestellte probabilistische Registrierung wurde zweimal auf die 40 2D Daten angewendet:

E3.2: Bei der ersten probabilistischen Registrierung (probReg) wurden ausschließlich die Positionsmerkmale bei der Parameteroptimierung angepasst. Das heißt, $\mathbf{V}^f$ bleibt Null während der Optimierung von $C(T, \mathbf{V}) \rightarrow \min_{T, \mathbf{V}}$ (Abschnitt 7).

E3.3: Bei der zweiten Anwendung der probabilistischen “Appearance adaptierenden” Registrierung (probAAREg) wurden darüber hinaus die Appearance-Merkmale (MRT-Signalwerte) zwischen den zu registrierenden Bildern ausgeglichen ($\mathbf{V}^f$ wird mit optimiert).

Bei der Berechnung der Deformationsfelder für die Labels (Abschnitt 7.2) beider Fälle (probReg, probAAREg) spielt die Appearance-Anpassung durch $\mathbf{V}$ keine Rolle, da sich eine Änderung der MRT-Signalwerte nicht auf die Labelwerte (Änderung der Zahlenwerte der Labels) auswirken sollte.

Ergebnisse: Tabelle 3 fasst die Ergebnisse zusammen. Dabei wurden die Dice- und Jaccard-Werte über alle 20 Label und anschließend über alle 1540 Registrierungen gemittelt. Zum einen wurden die Werte vor der Registrierung (**S** vs. **M**) berechnet und zum anderen nach der Registrierung. Die beiden letzten Zeilen zeigen die Ergebnisse des vorgestellten probabilistischen (Appearance-adaptierenden) Verfahrens.

E3.1 vs. E3.2/E3.3: Die Ergebnisse machen deutlich, dass das probabilistische Registrierungsverfahren mit einem etablierten Ansatz wie der variationellen Registrierung verglichen werden kann und im Fall der Gehirndaten sogar signifikant bessere Ergebnisse liefert (p -Wert $\ll 0.01$). Abbildung 51 stellt zwei Beispiele dar, bei denen die klassische Registrierung im Gegensatz zur probabilistischen Probleme aufweist.

E3.2 vs. E3.3: Außerdem zeigt sich, dass die Adaption der Appearance-Werte während der Optimierung zu einer besseren Zuordnung der Korrespondenzwerte führt und die Ergebnisse für probAAREg signifikant (p -Wert $\ll 0.01$) besser sind als für probReg. Abbildung 51 visualisiert zwei Beispiele, bei denen die Anpassung der Signalwerte die Registrierung deutlich verbessert.

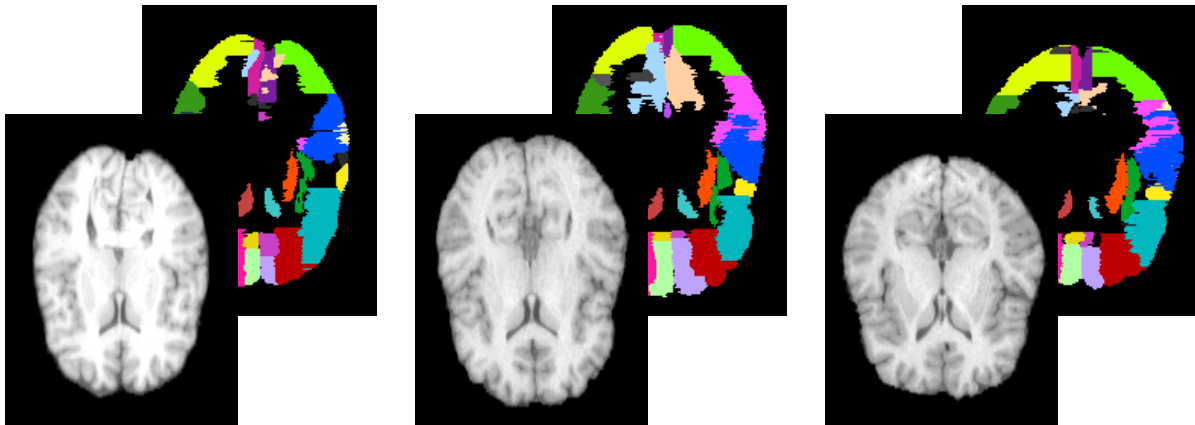


Abbildung 50: Drei Beispiele für die zu registrierenden Gehirndaten mit den Groundtruth-Labeldaten.

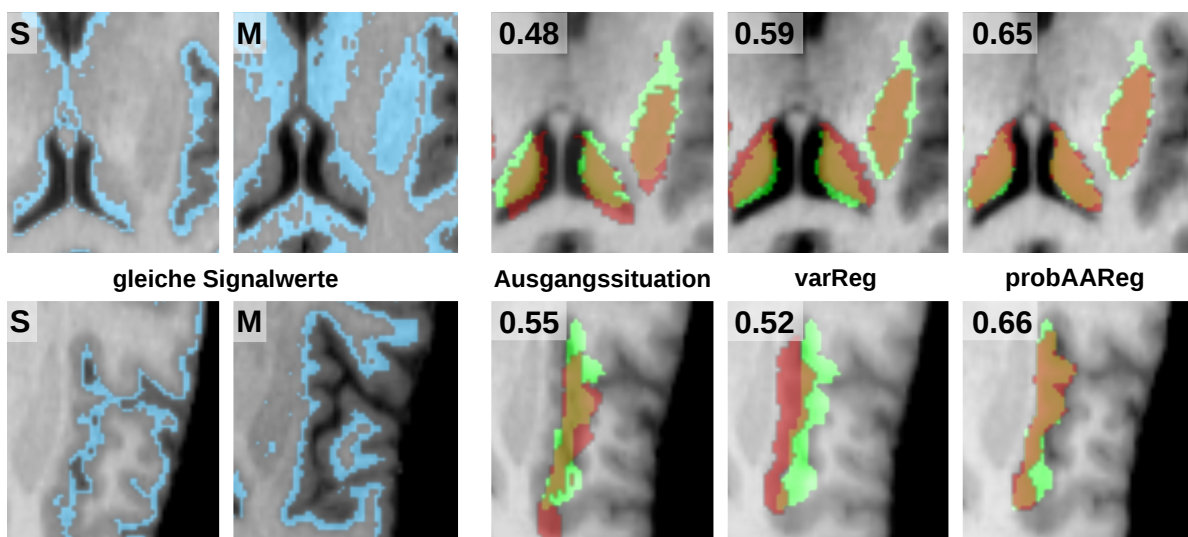


Abbildung 51: Zwei Beispiele für Registrierungsergebnisse, bei denen die variationelle Registrierung Probleme aufweist: Links sind für beide Eingabebilder (**S** und **M**) die gleichen Signalwertebereiche hellblau markiert. Rechts sind die Überlagerungen der *GT*-Labelmaske (grün) und der registrierten Labelmaske (rot) vor und nach der jeweiligen Registrierung mit dem jeweiligen Jaccard-Koeffizienten dargestellt. Für das erste Beispiel (obere Zeile) sind die Signalwerte des äußeren Bereichs des “Nucleus Caudatus” in **M** ähnlich mit dem inneren Bereich derselben Region in **S**. Das hat zur Folge, dass die variationelle Registrierung dieser Bereiche (innen und außen) aufeinander “zieht”. Im zweiten Fall (untere Zeile) passiert dies analog mit dem Bereich der “Cortex Insularis”, wobei die Jaccard-Werte sogar durch die variationelle Registrierung kleiner werden. Die vorgestellte probabilistische Registrierung (rechts) gleicht diese leichten Signalwertunterschiede aus und kann daher diese Bereiche besser aufeinander registrieren.

9. Diskussion

Die vorliegende Arbeit beschreibt einen probabilistischen Maximum-A-posteriori-Ansatz für statistische Appearance-Modelle. Die Grundlage der vorangestellten Problemformulierung bildet eine merkmalsvektorbasierte Repräsentation von Bildern durch eine Menge unabhängiger D -dimensionaler Merkmalsvektoren, welche Positions- und Appearance-Informationen kombinieren. Im Kontrast zu klassischen statistischen Form- oder Appearance-Modellen nutzt das probabilistische Modell Korrespondenzwahrscheinlichkeiten, wodurch eine vorangehende, potentiell aufwendige Definition von Eins-zu-Eins-Korrespondenzen unnötig wird. Dies eliminiert den Bedarf einer elaborierten Vorverarbeitung der Trainingsdaten und unterstützt außerdem ein arbiträres Abtasten der Test- und Trainingsdaten. Ein abgeleitetes globales Optimierungskriterium wird sowohl für die Modellgenerierung als auch -anpassung genutzt. Für die Optimierung der gesuchten Parameter über das globale Kriterium wird ein iterativer Expectation-Maximization-Algorithmus angewendet, um abwechselnd Modellparameter und probabilistische Korrespondenzen zwischen den Trainingsinstanzen und dem Modell zu bestimmen. Für die Steigerung der Robustheit während der Modelloptimierung wurden verschiedene Multi-Level-Erweiterungen vorgestellt.

9.1. Zusammenfassung der Vorteile des vorgestellten Frameworks

- Die Nutzung der **probabilistischen Korrespondenzen** liefert eine Vielzahl an Vorteilen im Gegensatz zu vorher aufwendig bestimmten Eins-zu-Eins-Korrespondenzen:
 - Eine kostenaufwendige Bestimmung von Landmarken und Eins-zu-Eins-Korrespondenzen vor der Modellgenerierung ist nicht notwendig. Außerdem ist die Qualität des resultierenden Modells nicht von potentiell falsch gesetzten Landmarken abhängig.
 - Die probabilistischen Korrespondenzen führen zu einer gesteigerten Robustheit gegenüber nicht vorhandener korrespondierender Strukturen in den Bilddaten.
 - Außerdem sind zusätzliche Informationen über lokale Unsicherheiten des Modells durch probabilistische Korrespondenzen nach abgeschlossener Modellanpassung gegeben. Es werden Bereiche (mit geringen Korrespondenzwerten) in neuen Bilddaten indiziert, die nicht oder “schlecht” durch das Modell abgebildet werden konnten. Diese können beispielsweise automatisch auf pathologische Bereiche in den Bilddaten hinweisen.
- Im Gegensatz zu klassischen Verfahren, die zu meist auf einer Vielzahl an verschiedenen Algorithmen bestehen, basiert die vorgestellte Methode auf einer **geschlos-**

senen mathematischen Formulierung eines globalen Optimierungskriteriums:

- Die Qualität des Verfahrens ist nicht abhängig von Ergebnissen von einzelnen unabhängigen aneinandergereihten Verfahren.
 - Sowohl die Modellgenerierung und als auch die Modellanpassung werden mit demselben Optimierungsverfahren durchgeführt. Im Gegensatz dazu stellt bei klassischen Modellansätzen die Modellanpassung oftmals ein komplett neues und sogar schwereres Problem als die Modellgenerierung dar.
 - Alle Parameter werden entweder über das Kriterium optimiert oder datengetrieben an die aktuelle Situation während der Optimierung angepasst. Es findet kein manuelles Tuning von Parametern statt.
- Die Grundlage der Modellformulierung bildet eine Datenrepräsentation über eine **gemeinsame Definition von Positions- und Appearance-Merkmalen** in einer Merkmalsvektormenge:
 - Durch die Repräsentation der Merkmalsvektoren als beliebig lange Vektoren, bei denen Positionsmerkmale und Appearance- oder Segmentierungsmerkmale gleich behandelt werden, gibt es keinerlei Einschränkungen die Art oder Anzahl der Merkmale betreffend.
 - Es können automatisch “beliebigdimensionale” Bilder verarbeitet werden: Ob 2D, 3D oder 4D (3D+Zeit), ob mehrkanal Bilder (verschiedene MRT-Sequenzen) oder multimodale Bilddaten (MRT und CT Daten), ob eine Segmentierungsmaske oder Multi-Organsegmentierung – es muss keinerlei Anpassung an die vorgestellte Methode erfolgen.
 - Durch die zusätzliche Anpassung der Appearance-Werte während der Modellgenerierung ist ebenfalls keine Vorverarbeitung der Signalwerte (Histogrammangleichung, etc.) der Trainingsdaten notwendig.
 - Außerdem ist die Anwendung des Ansatzes nicht auf die Segmentierungsproblematik beschränkt.
 - Der Modellgenerierungsansatz wurde zu einer **probabilistischen Registrierung zweier Bilder** vereinfacht, wobei eine solche Registrierung folgende Vorteile liefert:
 - Im Gegensatz zur klassischen Registrierung weist die probabilistische Registrierung durch die Nutzung der Korrespondenzwahrscheinlichkeiten eine hohe Robustheit gegenüber pathologischen Bereichen, die nicht in beiden Bildern vorhanden sind, auf.
 - Solche Bereiche werden darüber hinaus automatisch durch geringe Korrespondenzwerte nach der Registrierung indiziert und können daher zusätzlich segmentiert werden.

- Durch die “beliebigdimensionale Formulierung” der Bildrepräsentationen und der Deformation können im Gegensatz zur klassischen Registrierung nicht nur die Verschiebungsfelder der Bildpositionen bestimmt werden, sondern ebenfalls die Signalwertunterschiede beider Bilddaten ausgeglichen werden.

Im Gegensatz zum klassischen Form- oder Appearance-Modell, das vor allem zur Segmentierung des modellierten Objekts eingesetzt wird, liefert der vorgestellte Modellansatz eine Vielzahl an Anwendungsmöglichkeiten – zum einen durch die flexible Formulierung der Bildrepräsentation mit beliebigdimensionalen Merkmalsvektoren und zum anderen durch die Nutzung der probabilistischen Korrespondenzen. Anwendungsszenarien umfassen unter anderem:

- Segmentierung von Objektkonturen,
- Detektion von Landmarken,
- Multi-Objektsegmentierung bzw. Übertragung von beliebig vielen Labeln auf ein neues Bild,
- Rekonstruktion von fehlenden Merkmalen (z.B. fehlende Sequenz in MRT-Bildern).
- robuste Modellanpassung bei fehlenden Korrespondenzen bzw. “fehlerhaften Bereichen” im neuen Bild,
- Identifikation von lokalen Unsicherheiten des Modells,
- Klassifikation von pathologischen Regionen oder
- Rekonstruktion von “fehlerhaften Bereichen” in Bildern.

Neben solchen Modellanwendungen, welche Vorwissen über das betrachtete Objekt in einem Modell speichern und dieses Vorwissen als Grundlage für die Analyse neuer Bilddaten nutzen, wurde außerdem die Anwendung des beschriebenen mathematischen Frameworks zur probabilistischen Registrierung zweier Bilder vorgestellt.

9.2. Diskussion der Parameteroptimierung und der Parameterwahl

Für die Modellierung des betrachteten Objekts wird eine Vielzahl an Parametern optimiert. Diese Optimierung findet über die Minimierung des globalen Kriteriums statt. Da es möglich ist, die auf Null gesetzte Ableitung des Kriteriums geschlossen nach allen Parametern umzustellen, kann pro Iteration das aktuelle Optimum (abhängig von den aktuellen Korrespondenzwahrscheinlichkeiten) geschlossen bestimmt werden. Bei der Berechnung der optimalen Variationsmoden kommt lediglich ein iteratives Verfahren zum Einsatz, um die Orthogonalität zwischen den Moden zu gewährleisten.

Da das aktuelle Optimum der Parameter vollständig von den probabilistischen Korrespondenzen pro Iteration abhängt, können die Merkmalsvarianzen Σ als wichtigste Parameter des vorgestellten Frameworks betrachtet werden. Diese bestimmen, welche Merkmalsvektoren wie stark miteinander korrespondieren – abhängig von ihrem Abstand (pro Merkmal) zueinander.

Die vorliegende Arbeit zeigt, dass die Merkmalsvarianzen sehr gut automatisch an die aktuellen Differenzen zwischen den Observationen des Trainingsdatensatzes und den entsprechenden Modellinstanzen angepasst werden können. Dies macht ein manuelles Parameter-Tuning überflüssig.

Um die Robustheit der Korrespondenzbestimmung weiterhin zu steigern, wurden verschiedene Erweiterungen des Ansatzes zur Regularisierung vorgestellt:

- Eine nicht konstante A-priori-Wahrscheinlichkeit der Modellparameter wurde für die lokale Regularisierung der Merkmalsvektoren im Modell über deren jeweilige Nachbarn eingesetzt. Die Varianzen, welche bestimmen, wie stark die Distanzen zwischen benachbarten Merkmalsvektoren variieren dürfen, werden ebenfalls – analog zur den Merkmalsvarianzen – datengetrieben bestimmt, sodass keine manuelle Definition erforderlich ist.
- Ein Multi-Level-Ansatz, bei dem die Anzahl der Merkmalsvektoren in Modell und Trainingsdaten sukzessive inkrementiert wird, wurde angewendet, um in den ersten Leveln sicherzustellen, dass grobe Unterschiede (korrespondierende Merkmalsvektoren mit großem Abstand zueinander) zwischen den Daten ausgeglichen werden und in weiteren feineren Leveln immer kleinere Unterschiede (kleine Distanzen zwischen korrespondierenden Merkmalsvektoren) eliminiert werden. Die Initialisierung eines neuen Levels erfolgt wieder über die Betrachtung probabilistischer Korrespondenzen zwischen den Modellen unterschiedlicher Level.
- Ein sukzessives Inkrementieren der Anzahl der Moden kann außerdem genutzt werden, um die Anzahl der zu optimierenden Parameter zu Beginn der Optimierung gering zu halten und diese erst mit fortschreitender Berechnung (mit “besseren” Korrespondenzen) zu erhöhen.
- Für die Modellanpassung können darüber hinaus die Merkmalsvarianzen zwischen optimiertem resultierendem Modell und neuer Bildrepräsentation zu Beginn groß gehalten werden (durch einen Faktor), um sicherzustellen, dass große Unterschiede zwischen Modell und neuem Bild erkannt werden. Anschließend erfolgt die Berechnung der Varianzen erneut komplett datengetrieben.

Alle Erweiterungs- und Regularisierungsmaßnahmen erfolgen komplett automatisch und lassen sich entweder mit einer Erweiterung des globalen Kriteriums in die geschlossene mathematische Formulierung einbeziehen (A-priori-Wahrscheinlichkeit des Modells)

oder können durch eine weitere Nutzung probabilistischer Korrespondenzen (Initialisierung eines neuen Levels) umgesetzt werden, sodass keine “Extra”-Verfahren eingesetzt werden müssen.

9.3. Diskussion der Ergebnisse

Die Evaluierung des probabilistischen Modellansatzes und der probabilistischen Registrierung in der vorliegenden Arbeit hat vor allem die Intention, die Möglichkeiten der Methode aufzuzeigen und nicht einzelne Problemstellungen so genau wie möglich zu lösen. Aufgrund dessen wurden die genutzten Daten nicht vorverarbeitet – eine Histogrammangleichung würde z. B. die Variabilität der Appearance-Werte in jedem Trainingsdatensatz verringern.

Segmentierung der Hand: Die Segmentierung der 2D Röntgenbilder der Handdaten demonstriert, dass der Modellansatz in der Lage ist, relativ flexible und stark variierende Handstellungen zu modellieren. Außerdem wird zum einen die Segmentierung einer Objektkontur (Handkontur) und zum anderen eine Landmarkenbestimmung (der Fingerknochen) demonstriert. Da die Genauigkeit der Ergebnisse ($\leq 1mm$) wesentlich kleiner ist als der Abstand zwischen den Merkmalsvektoren ($3mm$), kann geschlussfolgert werden, dass die Modellierung der Kontur bzw. der Landmarken über die Werte einer Distanzkarte eine ausreichend große Genauigkeit bereitstellt. Die Segmentierung von unbekannten Bildern zeigt, dass das generierte Modell erfolgreich an unbekannte Daten angepasst und dass fehlende Merkmale (Distanzkartenwerte) für die Bilder rekonstruiert werden können. Außerdem wurde gezeigt, dass die Initialisierung der Merkmalsvektoren für ein neues Bild über die Reduktion der Vektoren pro Iteration stabile Ergebnisse liefert. Eine weitere nachgeschaltete probabilistische Registrierung zwischen Bild und angepasstem Modell als Nachbearbeitungsschritt liefert die Möglichkeit, die Ergebnisse zu verbessern.

Klassifikation von 3D Gehirndaten: Das 3D Gehirnmodell über die 3D MRT-Daten wurde genutzt, um die Modellgenerierung mit partiellen Trainingsdaten zu demonstrieren. Darüber hinaus wurde das generierte “gesunde” Modell eingesetzt, um pathologische (Schlaganfall-) Regionen in neuen Bildern mittels resultierenden Korrespondenzwerten zu identifizieren. Durch die 3D Daten erhöht sich die Anzahl der Merkmalsvektoren stark und durch die Berechnung der Korrespondenzwahrscheinlichkeiten (potentiell zwischen allen möglichen Merkmalsvektorpaaren), welche in jeder Iteration durchgeführt wird, steigt der Rechenaufwand quadratisch mit der Anzahl der Merkmalsvektoren. Das heißt, um den Rechen- und Speicherplatzaufwand für die Evaluierung gering zu halten, wurde ein relativ grober Abstand zwischen den Merkmalsvektoren von $5mm$ gewählt. Dies spiegelt sich dementsprechend in den Ergebnissen wider: große Läsionen ($> 10cm^3$) können gut durch die vorgestellte Methode segmentiert werden,

wohingegen kleine Läsionen nicht gefunden werden, da das Modell zu grob ist. Außerdem wurde ein relativ einfaches automatisches Schwellwertverfahren zur Aufteilung in “gute” und “schlechte” Korrespondenzwerte genutzt. Der Vergleich mit den manuell gewählten Schwellwerten zeigt, dass an dieser Stelle eine Verbesserung durch aufwendigere Verfahren möglich ist. Die Evaluierung zeigt jedoch, dass aus den resultierenden Korrespondenzwerten eindeutig Informationen über “nichtmodellierbare” – in diesem Fall pathologische – Regionen extrahiert werden können. Für genauere Ergebnisse (vor allem für kleine Läsionen) muss ein feineres Modell generiert werden. Dies erfordert ebenfalls einen größeren Trainingsdatensatz.

Registrierung von 2D Gehirndaten: Das probabilistische Appearance-adaptierende Registrierungsverfahren wurde an 2D-MRT-Bildern des Gehirns getestet. Verglichen wurden die Ergebnisse mit einem gut etablierten nichtrigiden Registrierungsverfahren. Die Ergebnisse zeigen, dass die probabilistische Registrierung ohne Appearance-Adaption bereits eine signifikant bessere Performanz zeigt als der klassische Ansatz. Die zusätzliche Anpassung der Appearance-Werte während der Deformationsoptimierung stellt eine weitere eindeutige Verbesserung da. Daraus lässt sich überdies verallgemeinern, dass das vorgestellte Verfahren der iterativen Korrespondenzwahrscheinlichkeitsbestimmung im Vergleich zur registrierungsbasierten Korrespondenz-/Landmarkenbestimmung gute Ergebnisse liefert. Wie beim Modellansatz können hier ebenfalls beliebig viele Merkmale in der Registrierung genutzt werden. Die Anpassung der Appearance-Werte wird angewendet, um die Korrespondenzen genauer bestimmen zu können, da so “störende” Appearance-Unterschiede ausgeglichen werden. Außerdem kann diese Anpassung eingesetzt werden, um die Signalwerte der Bilder einander anzugleichen (mit oder ohne Positionsdeformation). Bei der evaluierten Anwendung war dies nicht von Bedeutung, da die anhand der Signalwerte optimierte Positionsdeformation auf Labelbilder übertragen wurde.

9.4. Einschränkungen des Ansatzes

In klassischen modellbasierten Ansätzen ist der kostenintensivste Schritt die Bestimmung der korrekten Landmarken bzw. Korrespondenzen in den Bildern des Trainingsdatensatzes. Die anschließende Generierung des mittleren Modells und der Moden wird in einem Schritt z. B. mittels PCA durchgeführt. Im Gegensatz dazu ist die Idee des vorliegenden Ansatzes, so wenig wie möglich vordefinierte Informationen zu nutzen und alle Parameter – mittleres Modell, Moden und Korrespondenzen – “gleichzeitig” in einer iterativen Optimierung eines globalen Kriteriums zu bestimmen. Dies erhöht die Kosten für die Modellgenerierung enorm, minimiert bzw. eliminiert allerdings die Kosten für die Vorverarbeitung. Des Weiteren hat die spärliche Datenrepräsentation zur Folge, dass das resultierende Modell die Daten nicht vollständig abbildet. Das heißt, der vorgestellte Ansatz eignet sich vor allem für Problemstellungen, bei denen die exakten Korrespondenzen nicht gut definiert sind. Außerdem ist es von Vorteil, wenn

sich die Fragestellung der Anwendung gut durch eine spärliche merkmalsvektorbasierte Bildrepräsentation beantworten lässt. Ist die gesuchte Struktur zu fein (z. B. kleine Läsionen im Gehirn), steigt die Anzahl der benötigten Merkmalsvektoren und somit die Anzahl der Parameter, wodurch sich der Rechenaufwand stark erhöht: Durch die iterativ wiederholte Berechnung der Korrespondenzwahrscheinlichkeiten zwischen allen möglichen Merkmalsvektorkpaaren zwischen Modell und Trainingsinstanz steigt der Aufwand des Algorithmus annähernd quadratisch mit der Anzahl der Merkmalsvektoren. Eine Möglichkeit den Aufwand für die Korrespondenzschätzung zu verringern, ist die Reduktion der pro Iteration betrachteten Merkmalsvektoren. Das bedeutet, für jeden Vektor in einer Bildrepräsentation wird lediglich eine lokale Untermenge statt der gesamten Menge der Merkmalsvektoren der zweiten Repräsentation betrachtet.

9.5. Ausblick

Um das Problem der quadratischen Komplexität mit steigender Merkmalsvektoranzahl zu lösen, muss künftig eine weitere Optimierung des Ansatzes in Hinblick auf den Rechenaufwand erfolgen.

Darüber hinaus wäre es von Interesse, die Nutzung von lokal bestimmten Merkmalsvarianzen (Σ) im Gegensatz zu aktuell global identischen Varianzen zu betrachten. Speziell für Objekte, die unterschiedlich stark an unterschiedlichen Regionen innerhalb des Objekts variieren, wäre es interessant, lokal unterschiedliche Werte für die Merkmalsvarianzen zu untersuchen.

Die Dimension der Positionswerte der Merkmalsvektoren ist beliebig – in dieser Arbeit wurden 2D- und 3D-Bilddaten betrachtet. In Zukunft wäre eine Modellierung von 4D Daten (3D + Zeit) interessant, da dies unkompliziert durch den Ansatz ohne weitere Extraschritten umsetzbar ist. Dafür müsste die Rechenkomplexität des Ansatzes reduziert werden, da die vierte Dimension eine weitere Erhöhung der Anzahl der Merkmalsvektoren bedeuten würde.

Wie dargestellt, wurde für die Evaluierung keine Vorverarbeitung der Daten durchgeführt. Um die bestmöglichen Ergebnisse für ein bestimmtes Problem zu erhalten, ist jedoch eine Vorverarbeitung oder Vorausrichtung der Daten immer sinnvoll. Auch wenn eine genaue bzw. reproduzierbare Determinierung der Merkmalsvektoren unter der Nutzung der Korrespondenzwahrscheinlichkeiten nicht nötig ist, wird eine genauere Bestimmung der Merkmalsvektoren immer zu besseren Ergebnissen führen. Allerdings ausschließlich unter der Voraussetzung, dass ähnliche Merkmalsvektoren ebenfalls in neuen unbekannten Bildern definiert werden können. Das heißt, die Betrachtung unterschiedlicher Positionierungen der Merkmalsvektoren wäre ebenfalls von Interesse. Außerdem könnte eine initiale nichtlineare Registrierung der Trainingsbilder zum mittleren Modell für eine initiale Bestimmung der Korrespondenzwahrscheinlichkeiten genutzt werden, welche die Anzahl der notwendigen Iterationen der Modellgenerierung verringern könnte. Dies wäre allerdings nur sinnvoll, wenn die Registrierung “schneller” und robuster ist als die entsprechende Anzahl an Iterationen der Modellgenerierung.

Ein Nachteil des vorgestellten Modellansatzes ist die spärliche Datenrepräsentation der resultierenden Modellinstanzen. Dies kann umgangen werden, indem für das mittlere Modell ein korrespondierendes vollständig definiertes Bild gegeben wäre. Aus der affinen Transformation und den durch die Modellanpassung gewichteten Moden kann anschließend ein komplettes Deformationsfeld rekonstruiert werden (analog zur vorgestellten probabilistischen Registrierung), mit dem das “mittlere Bild” deformiert wird. Dabei können sowohl die Koordinaten als auch die Appearance-Merkmale des Bildes angepasst werden. Dieses deformierte Bild entspricht demnach dem vollständigen “Modellinstanzbild”, das zum gegebenen (Trainings-)Bild korrespondiert. Da das mittlere Modell im vorgestellten Ansatz durch ein beliebiges Trainingsbild initialisiert wird, kann dieses Trainingsbild als “mittleres Bild” genutzt werden. Dabei ist allerdings zu beachten, dass im Verlauf der Modellgenerierung das mittlere Modell nicht mit optimiert werden darf (sonst korrespondiert das mittlere Modell nicht mehr zum mittleren Bild) und daher vorher eine “sinnvolle” Wahl dieser Trainingsinstanz getroffen werden muss.

Für das Registrierungsproblem wurden ausschließlich gut definierte und vollständige 2D Bilddaten betrachtet, um die Genauigkeit der Registrierung zu untersuchen. Ein großer Vorteil des Registrierungsverfahrens liegt allerdings in der Nutzung der probabilistischen Korrespondenzen, was zu einer größeren Robustheit bei korrupten oder pathologiebelasteten Daten führen sollte. Daher ist in Zukunft eine ausführliche Evaluation auf Daten mit Pathologien notwendig. Die resultierenden Korrespondenzen können in diesem Rahmen für eine mögliche Maskierung von Regionen, die nicht in beiden Bilddaten vorhanden sind, genutzt werden.

A. Ableitung des globalen Kriteriums

A.1. Ableitung nach beliebigem Parameter x

Das Globale Kriterium (aufgestellt in Abschnitt 3.2)

$$\begin{aligned}
& \operatorname{argmin}_{Q, \Theta} C(Q, \Theta) \\
&= \operatorname{argmin}_{Q, \Theta} - \sum_{k=1}^N \left(\sum_{i=1}^{N_s(k)} \log \left(\sum_{j=1}^{N_m} \exp \left(-\frac{1}{2} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \Sigma_j^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \right) \right) \right) \\
&+ \left(\left(\sum_{k=1}^N \sum_{p=1}^n \log(\lambda_p) + \left(\frac{\omega_{k,p}^2}{2\lambda_p^2} \right) \right) + \left(\frac{\|\frac{1}{N} \sum_{k=1}^N \boldsymbol{\omega}_k\|^2}{2\sigma_{\bar{\omega}}^2} \right) \right. \\
&+ \left. \left(\frac{\|(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}\|_F^2}{2\sigma_{\bar{T}}^2} \right) \right) \\
&+ \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n \left(\frac{1}{2} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \right. \\
&\left. \Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right)
\end{aligned}$$

kann abgeleitet werden nach einem beliebigen Parameter x :

$$\begin{aligned}
& \frac{\partial}{\partial x} C(Q, \Theta) \\
&= \sum_{k=1}^N \left(\sum_{p=1}^n \frac{\partial}{\partial x} \left(\log(\lambda_p) + \left(\frac{\omega_{k,p}^2}{2\lambda_p^2} \right) \right) \right. \\
&- \sum_{i=1}^{N_s(k)} \frac{\partial}{\partial x} \log \left(\sum_{j=1}^{N_m} \exp \left(-\frac{1}{2} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \Sigma_j^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \right) \right) \Bigg) \\
&- \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n \frac{\partial}{\partial x} \left(-\frac{1}{2} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \right. \\
&\left. \Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right) \\
&- \frac{\partial}{\partial x} \left(\log(\sigma_{\bar{\omega}}) - \left(\frac{\|\frac{1}{N} \sum_{k=1}^N \boldsymbol{\omega}_k\|^2}{2\sigma_{\bar{\omega}}^2} \right) + \log(\sigma_{\bar{T}}) - \left(\frac{\|(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}\|_F^2}{2\sigma_{\bar{T}}^2} \right) \right)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{k=1}^N \left(\sum_{p=1}^n \frac{\partial}{\partial x} C_1(Q, \Theta) - \sum_{i=1}^{N_s(k)} \frac{\partial}{\partial x} C_2(Q, \Theta) \right) \\
&\quad - \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n \frac{\partial}{\partial x} C_3(Q, \Theta) - \frac{\partial}{\partial x} C_4(Q, \Theta)
\end{aligned}$$

$\frac{\partial}{\partial x} C_1$, $\frac{\partial}{\partial x} C_2$, $\frac{\partial}{\partial x} C_3$ und $\frac{\partial}{\partial x} C_4$ sind gegeben durch:

$$\frac{\partial}{\partial x} C_1(Q, \Theta) = \frac{\partial}{\partial x} \log(\lambda_p) + \left(\frac{\omega_{k,p}^2}{2\lambda_p^2} \right)$$

Mit $\frac{\partial}{\partial x} \log(u(x)) = \frac{\frac{\partial}{\partial x} u(x)}{u(x)}$ und $\frac{\partial}{\partial x} \exp(f(x)) = \exp(f(x)) \frac{\partial}{\partial x} f(x)$:

$$\begin{aligned}
&\frac{\partial}{\partial x} C_2(Q, \Theta) \\
&= \frac{\partial}{\partial x} \log \left(\sum_{j=1}^{N_m} \exp \left(-\frac{1}{2} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \Sigma_j^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \right) \right) \\
&= \frac{\frac{\partial}{\partial x} \left(\sum_{j=1}^{N_m} \exp \left(-\frac{1}{2} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \Sigma_j^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \right) \right)}{\left(\sum_{j=1}^{N_m} \exp \left(-\frac{1}{2} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \Sigma_j^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \right) \right)} \\
&= \frac{\sum_{j=1}^{N_m} \exp \left(-\frac{1}{2} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \Sigma_j^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \right)}{\underbrace{\left(\sum_{j=1}^{N_m} \exp \left(-\frac{1}{2} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \Sigma_j^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \right) \right)}} \\
&\quad \mathbb{E}(H_{ij})_k = \frac{p(\mathbf{s}_i | \mathbf{m}_j)}{\sum_{j2}^{N_m} p(\mathbf{s}_i | \mathbf{m}_{j2})} = \frac{\left(\exp \left(-\frac{1}{2} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \Sigma_j^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \right) \right)}{\sum_{j=1}^{N_m} \exp \left(-\frac{1}{2} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \Sigma_j^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \right)} \\
&\quad \frac{\partial}{\partial x} \left(-\frac{1}{2} \underbrace{(\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \Sigma_j^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})}_{\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{B} \mathbf{x} = (\mathbf{B} + \mathbf{B}^T) \mathbf{x}} \right) \\
&= - \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \frac{\partial}{\partial x} \left(\frac{1}{2} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \Sigma_j^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \right) \\
&= - \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \frac{1}{2} \underbrace{(\Sigma_j^{-1} + \Sigma_j^{-1T})}_{=2\Sigma_j^{-1}} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \frac{\partial}{\partial x} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})
\end{aligned}$$

$$= -\sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma_j^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \frac{\partial}{\partial x} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})$$

$$\begin{aligned} & \frac{\partial}{\partial x} C_3(Q, \Theta) \\ &= \frac{\partial}{\partial x} \left(-\frac{1}{2} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_\eta \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \right. \\ & \quad \left. \Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_\eta \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right) \\ &= \left(-\frac{1}{2} (\Sigma_{\mathcal{N}}^{-1} + \Sigma_{\mathcal{N}}^{-1T}) \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_\eta \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \right. \\ & \quad \left. \frac{\partial}{\partial x} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_\eta \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right) \\ &= -(\Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_\eta \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \\ & \quad \frac{\partial}{\partial x} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_\eta \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right) \\ &= -(\Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_\eta \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \\ & \quad \frac{\partial}{\partial x} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_\eta \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right) \end{aligned}$$

$$\begin{aligned} & \frac{\partial}{\partial x} C_4(Q, \Theta) \\ &= \frac{\partial}{\partial x} \left(\log(\sigma_{\bar{\omega}}) - \left(\frac{\|\frac{1}{N} \sum_{k=1}^N \boldsymbol{\omega}_k\|^2}{2\sigma_{\bar{\omega}}^2} \right) + \log(\sigma_{\bar{T}}) - \left(\frac{\|(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}\|_F^2}{2\sigma_{\bar{T}}^2} \right) \right) \\ &= \frac{\partial}{\partial x} \log(\sigma_{\bar{\omega}}) - \left(\frac{\frac{1}{N} \sum_{k=1}^N \boldsymbol{\omega}_k}{\sigma_{\bar{\omega}}} \right) \frac{\partial}{\partial x} \left(\frac{\frac{1}{N} \sum_{k=1}^N \boldsymbol{\omega}_k}{\sigma_{\bar{\omega}}} \right) \\ & \quad + \frac{\partial}{\partial x} \log(\sigma_{\bar{T}}) - \left(\frac{(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}}{\sigma_{\bar{T}}} \right) \frac{\partial}{\partial x} \left(\frac{(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}}{\sigma_{\bar{T}}} \right) \end{aligned}$$

$$\begin{aligned}
&= - \left(\frac{\frac{1}{N} \sum_{k=1}^N \boldsymbol{\omega}_k}{\sigma_{\bar{\omega}}^2} \right)^T \frac{\partial}{\partial x} \left(\frac{1}{N} \sum_{k=1}^N \boldsymbol{\omega}_k \right) \\
&\quad - \left(\frac{(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}}{\sigma_{\bar{T}}^2} \right) \frac{\partial}{\partial x} \left(\left(\frac{1}{N} \sum_{k=1}^N T_k \right) - I_{D+1} \right)
\end{aligned}$$

Es kann zusammengefasst werden:

$$\begin{aligned}
&\frac{\partial}{\partial x} C(Q, \Theta) \\
&= \sum_{k=1}^N \left(\sum_{p=1}^n \frac{\partial}{\partial x} \left(\log(\lambda_p) + \left(\frac{\omega_{k,p}^2}{2\lambda_p^2} \right) \right) \right. \\
&\quad + \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \frac{\partial}{\partial x} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \\
&\quad - \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n - (\Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \\
&\quad \left. \frac{\partial}{\partial x} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta}))}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right) \\
&\quad - \left(\frac{\frac{1}{N} \sum_{k=1}^N \boldsymbol{\omega}_k}{\sigma_{\bar{\omega}}^2} \right)^T \frac{\partial}{\partial x} \left(\frac{1}{N} \sum_{k=1}^N \boldsymbol{\omega}_k \right) \\
&\quad - \left(\frac{(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}}{\sigma_{\bar{T}}^2} \right) \frac{\partial}{\partial x} \left(\left(\frac{1}{N} \sum_{k=1}^N T_k \right) - I_{D+1} \right)
\end{aligned}$$

A.2. Ableitung nach den Merkmalsvektoren des mittleren Modells $\bar{\mathbf{m}}_j$

$$\begin{aligned}
& \frac{\partial}{\partial \bar{\mathbf{m}}_j} C(Q, \Theta) \\
&= \sum_{k=1}^N \left(\sum_{p=1}^n \frac{\partial}{\partial \bar{\mathbf{m}}_j} \log(\lambda_p) + \left(\frac{\omega_{k,p}^2}{2\lambda_p^2} \right) \right. \\
&\quad + \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \frac{\partial}{\partial \bar{\mathbf{m}}_j} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})) \\
&\quad - \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n - (\Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \\
&\quad \left. \frac{\partial}{\partial \bar{\mathbf{m}}_j} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta}))}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right) \\
&= \sum_{k=1}^N \left(\sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k - \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - \bar{\mathbf{m}}_j - \sum_{p=1}^n \omega_{k,p} \mathbf{v}_p)^T \right) \\
&\quad - \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n - (\Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \\
&\quad \left(\frac{1}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right) \\
&= - \sum_{k=1}^N \left(\sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - \sum_{p=1}^n \omega_{k,p} \mathbf{v}_p)^T + \sum_{k=1}^N \left(\sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \Sigma^{-1} \bar{\mathbf{m}}_j^T \right) \right. \\
&\quad + \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n \Sigma_{\mathcal{N}}^{-1} \left(\frac{((\pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|^2} \right)^T \\
&\quad + \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n \Sigma_{\mathcal{N}}^{-1} \left(\frac{\bar{\mathbf{m}}_j^T}{\|\mathcal{N}_{j,\eta}^0\|^2} \right) \\
&\quad \left. \frac{\sum_{k=1}^N \left(\sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j})^T \right)}{\sum_{k=1}^N \sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \Sigma^{-1} + \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n \Sigma_{\mathcal{N}}^{-1} \left(\frac{1}{\|\mathcal{N}_{j,\eta}^0\|^2} \right)} \right. \\
&\quad \left. + \frac{\sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n \Sigma_{\mathcal{N}}^{-1} \left(\frac{(-(\pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) + (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) + \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|^2} \right)^T}{\sum_{k=1}^N \sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \Sigma^{-1} + \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n \Sigma_{\mathcal{N}}^{-1} \left(\frac{1}{\|\mathcal{N}_{j,\eta}^0\|^2} \right)} \right)
\end{aligned}$$

A.3. Ableitung nach den Variationsmoden V

$$\begin{aligned}
& \frac{\partial}{\partial \mathbf{v}_{p,j}} C(Q, \Theta) \\
&= \sum_{k=1}^N \left(\sum_{p=1}^n \frac{\partial}{\partial \mathbf{v}_{p,j}} \log(\lambda_p) + \left(\frac{\omega_{k,p}^2}{2\lambda_p^2} \right) \right. \\
&\quad + \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \frac{\partial}{\partial \mathbf{v}_{p,j}} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \\
&\quad - \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n - (\Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \\
&\quad \left. \frac{\partial}{\partial \mathbf{v}_{p,j}} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta}))}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right) \\
&= \sum_{k=1}^N \left(\sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - \bar{\mathbf{m}}_j - \sum_{q=1}^n \omega_{k,q} \mathbf{v}_{q,j})^T \right. \\
&\quad \frac{\partial}{\partial \mathbf{v}_{p,j}} (T_k \star \mathbf{s}_{k,i} - \bar{\mathbf{m}}_j - \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}) \\
&\quad - \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n - (\Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \\
&\quad \left. \frac{\partial}{\partial \mathbf{v}_{p,j}} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta}))}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right) \\
&= \sum_{k=1}^N \left(\sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - \bar{\mathbf{m}}_j - \sum_{q=1}^n \omega_{k,q} \mathbf{v}_{q,j})^T (-\omega_{k,p}) \right. \\
&\quad - \sum_{\eta \in \mathcal{N}(j)} - (\Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \\
&\quad \left. \left(\frac{(\pm 2 \cdot \lambda_p)}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right) \\
&= - \sum_{k=1}^N \left(\sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - \bar{\mathbf{m}}_j)^T \omega_{k,p} + \sum_{k=1}^N \sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \Sigma^{-1} \left(\sum_{q=1}^n \omega_{k,q} \mathbf{v}_{q,j} \right)^T \omega_{k,p} \right. \\
&\quad + \sum_{\eta \in \mathcal{N}(j)} \Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0)}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \left(\frac{((\pm 2 \cdot \lambda_p))}{\|\mathcal{N}_{j,\eta}^0\|} \right) \\
&\quad \left. + \sum_{\eta \in \mathcal{N}(j)} \left(\frac{(\pm 2 \cdot \lambda_p \Sigma_{\mathcal{N}}^{-1} \mathbf{v}_{p,j})}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \left(\frac{(\pm 2 \cdot \lambda_p)}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right)
\end{aligned}$$

$$\begin{aligned}
0 &= \frac{\partial}{\partial \mathbf{v}_{p,j}} C(Q, \Theta) \\
0 &= -\mathbf{a}_{p,j} + \sum_{q=1}^n (\mathbf{b}_{p,q,j} \Sigma^{-1} \mathbf{v}_{q,j})^T + \sum_{q=1}^n (\mathbf{b}'_{p,q,j} \Sigma_{\mathcal{N}}^{-1} \mathbf{v}_{q,j})^T \\
\mathbf{a}_{p,j} &= (a_{p,j}^1, \dots, a_{p,j}^D) \\
&= \sum_{k=1}^N \sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - \bar{\mathbf{m}}_j)^T \omega_{k,p} \\
&\quad + \sum_{\eta \in \mathcal{N}(j)} \Sigma_{\mathcal{N}}^{-1} \left(\frac{(\bar{\mathbf{m}}_j - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \left(\frac{((\pm 2 \cdot \lambda_p))}{\|\mathcal{N}_{j,\eta}^0\|} \right) \\
\mathbf{b}_{p,q,j} &= \sum_{k=1}^N \left(\sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \omega_{k,q} \omega_{k,p} \right) \\
\mathbf{b}'_{p,q,j} &= \sum_{\eta \in \mathcal{N}(j)} \left(\frac{\pm 2 \cdot \lambda_p}{\|\mathcal{N}_{j,\eta}^0\|} \right)^2, \text{ for } p = q, \text{ else } b'_{p,q,j} = 0 \\
\frac{\partial}{\partial \mathbf{V}_j} C(Q, \Theta) &= -\mathbf{A}_j + \mathbf{B}_j \mathbf{V}_j \\
\mathbf{A}_j \in \mathbb{R}^{D \cdot n} &= (a_{1,j}^1, \dots, a_{1,j}^D, a_{n,j}^1, \dots, a_{n,j}^D)^T \\
\mathbf{V}_j \in \mathbb{R}^{D \cdot \text{cot } n} &= (v_{1,j}^1, \dots, v_{1,j}^D, v_{n,j}^1, \dots, v_{n,j}^D)^T \\
\mathbf{B}_j \in \mathbb{R}^{D \cdot n \times D \cdot n} &= \begin{pmatrix} \mathbf{b}_{1,1,j} \Sigma^{-1} + \mathbf{b}'_{1,1,j} \Sigma_{\mathcal{N}}^{-1} & \dots & \mathbf{b}_{1,n,j} \Sigma^{-1} + \mathbf{b}'_{1,n,j} \Sigma_{\mathcal{N}}^{-1} \\ \dots & \mathbf{b}_{p,q,j} \Sigma^{-1} + \mathbf{b}'_{p,q,j} \Sigma_{\mathcal{N}}^{-1} & \dots \\ \mathbf{b}_{n,1,j} \Sigma^{-1} + \mathbf{b}'_{n,1,j} \Sigma_{\mathcal{N}}^{-1} & \dots & \mathbf{b}_{n,n,j} \Sigma^{-1} + \mathbf{b}'_{n,n,j} \Sigma_{\mathcal{N}}^{-1} \end{pmatrix}
\end{aligned}$$

A.4. Ableitung nach der Standardabweichung λ_p der Moden

$$\begin{aligned}
\frac{\partial}{\partial \lambda_p} C(Q, \Theta) &= \sum_{k=1}^N \left(\sum_{p=1}^n \frac{\partial}{\partial \lambda_p} \log(\lambda_p) + \left(\frac{\omega_{k,p}^2}{2\lambda_p^2} \right) \right. \\
&\quad + \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1}(\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \frac{\partial}{\partial \lambda_p} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j})) \\
&\quad - \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n - (\Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \\
&\quad \left. \frac{\partial}{\partial \lambda_p} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta}))}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right) \\
&= \sum_{k=1}^N \left(\frac{1}{\lambda_p} - \frac{\omega_{k,p}^2}{\lambda_p^3} \right) \\
&\quad + \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \left(\Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \right. \\
&\quad \left. \left(\frac{(\pm 2 \cdot \mathbf{v}_{p,j}) - (\pm 2 \cdot \mathbf{v}_{p,\eta})}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right) \\
&= \sum_{k=1}^N \left(\frac{1}{\lambda_p} - \frac{\omega_{k,p}^2}{\lambda_p^3} \right) + \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \left(\Sigma_{\mathcal{N}}^{-1} \left(\frac{(\bar{\mathbf{m}}_j - \bar{\mathbf{m}}_{\eta}) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \right. \\
&\quad \left. \left(\frac{(\pm 2 \cdot \mathbf{v}_{p,j}) - (\pm 2 \cdot \mathbf{v}_{p,\eta})}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right) \\
&\quad + \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \left(\Sigma_{\mathcal{N}}^{-1} \left(\frac{(\pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \left(\frac{(\pm 2 \cdot \mathbf{v}_{p,j}) - (\pm 2 \cdot \mathbf{v}_{p,\eta})}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right) \\
&= \lambda_p^2 \sum_{k=1}^N (1) + (-\omega_{k,p}^2) \\
&\quad + \lambda_p^3 \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \left(\Sigma_{\mathcal{N}}^{-1} \left(\frac{(\bar{\mathbf{m}}_j - \bar{\mathbf{m}}_{\eta}) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \left(\frac{(\pm 2 \cdot \mathbf{v}_{p,j}) - (\pm 2 \cdot \mathbf{v}_{p,\eta})}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right) \\
&\quad + \lambda_p^4 \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \left(\Sigma_{\mathcal{N}}^{-1} \left(\frac{(\pm 2 \cdot \mathbf{v}_{p,j}) - (\pm 2 \cdot \mathbf{v}_{p,\eta})}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \left(\frac{(\pm 2 \cdot \mathbf{v}_{p,j}) - (\pm 2 \cdot \mathbf{v}_{p,\eta})}{\|\mathcal{N}_{j,\eta}^0\|} \right) \right) \\
&= \lambda_p^2 a_p + b_p + \lambda_p^3 c_p + \lambda_p^4 d_p = 0
\end{aligned}$$

A.5. Ableitung nach den Gewichten der Variationsmoden ω_k

$$\begin{aligned}
& \frac{\partial}{\partial \omega_{k,p}} C(Q, \Theta) \\
&= \sum_{k=1}^N \left(\sum_{p=1}^n \frac{\partial}{\partial \omega_{k,p}} \log(\lambda_p) + \left(\frac{\omega_{k,p}^2}{2\lambda_p^2} \right) \right. \\
&\quad + \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1}(\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \frac{\partial}{\partial \omega_{k,p}} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \\
&\quad - \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n - (\Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_\eta \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \\
&\quad \frac{\partial}{\partial \omega_{k,p}} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_\eta \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta}))}{\|\mathcal{N}_{j,\eta}^0\|} \right) \\
&\quad - \left(\frac{\frac{1}{N} \sum_{k=1}^N \omega_k}{\sigma_{\bar{\omega}}^2} \right)^T \frac{\partial}{\partial \omega_{k,p}} \left(\frac{1}{N} \sum_{k=1}^N \omega_k \right) \\
&\quad - \left(\frac{(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}}{\sigma_T^2} \right) \frac{\partial}{\partial \omega_{k,p}} \left(\frac{1}{N} \sum_{k=1}^N T_k - I_{D+1} \right) \\
&= \sum_{k=1}^N \left(\sum_{p=1}^n \frac{\partial}{\partial \omega_{k,p}} \log(\lambda_p) + \left(\frac{\omega_{k,p}^2}{2\lambda_p^2} \right) \right. \\
&\quad + \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1}(\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \frac{\partial}{\partial \omega_{k,p}} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) - \left(\frac{\sum_{k=1}^N \omega_{k,p}}{N^2 \sigma_{\bar{\omega}}^2} \right) \\
&= \sum_{k=1}^N \left(\sum_{p=1}^n \frac{\partial}{\partial \omega_{k,p}} \log(\lambda_p) + \left(\frac{\omega_{k,p}^2}{2\lambda_p^2} \right) \right. \\
&\quad + \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - \bar{\mathbf{m}}_j - \sum_{q=1}^n \omega_{k,q} \mathbf{v}_{q,j})^T \\
&\quad \frac{\partial}{\partial \omega_{k,p}} (T_k \star \mathbf{s}_{k,i} - \bar{\mathbf{m}}_j - \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}) - \left(\frac{\sum_{k=1}^N \omega_{k,p}}{N^2 \sigma_{\bar{\omega}}^2} \right) \\
&= \frac{\omega_{k,p}}{\lambda_p^2} + \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - \bar{\mathbf{m}}_j - \sum_{q=1}^n \omega_{k,q} \mathbf{v}_{q,j})^T (-\mathbf{v}_{p,j}) - \left(\frac{\sum_{k=1}^N \omega_{k,p}}{N^2 \sigma_{\bar{\omega}}^2} \right)
\end{aligned}$$

$$\begin{aligned}
&= \frac{\omega_{k,p}}{\lambda_p^2} - \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - \bar{\mathbf{m}}_j)^T \mathbf{v}_{p,j} \\
&\quad + \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \sum_{q=1}^n \omega_{k,q} (\Sigma^{-1} \mathbf{v}_{q,j})^T \mathbf{v}_{p,j} - \frac{\omega_{k,p}}{N^2 \sigma_{\bar{\omega}}^2} - \left(\frac{\sum_{l=1, l \neq k}^N \omega_{l,p}}{N^2 \sigma_{\bar{\omega}}^2} \right) \\
&= \frac{\omega_{k,p}}{\lambda_p^2} - \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - \bar{\mathbf{m}}_j)^T \mathbf{v}_{p,j} \\
&\quad + \sum_{q=1}^n \omega_{k,q} \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \mathbf{v}_{q,j}^T \Sigma^{-1} \mathbf{v}_{p,j} - \frac{\omega_{k,p}}{N^2 \sigma_{\bar{\omega}}^2} - \left(\frac{\sum_{l=1, l \neq k}^N \omega_{l,p}}{N^2 \sigma_{\bar{\omega}}^2} \right) \\
&= \frac{\omega_{k,p}}{\lambda_p^2 - N^2 \sigma_{\bar{\omega}}^2} - d_{k,p} + \sum_{q=1}^n \omega_{k,q} g_{k,p,q} \\
\mathbf{G}_k \in \mathbb{R}^{n \times n} \quad \dots \quad g_{k,p,q} &= \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \mathbf{v}_{q,j}^T \Sigma^{-1} \mathbf{v}_{p,j} \\
\mathbf{d}_k \in \mathbb{R}^{1 \times n} \quad \dots \quad d_{k,p} &= \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - \bar{\mathbf{m}}_j)^T \mathbf{v}_{p,j} + \left(\frac{\sum_{l=1, l \neq k}^N \omega_{l,p}}{N^2 \sigma_{\bar{\omega}}^2} \right) \\
\mathbf{d}_k &= \left(\left(\begin{pmatrix} \frac{1}{\lambda_1^2 - N^2 \sigma_{\bar{\omega}}^2} & & \\ & \ddots & \\ & & \frac{1}{\lambda_n^2 - N^2 \sigma_{\bar{\omega}}^2} \end{pmatrix} + \mathbf{G}_k \right) \boldsymbol{\omega}_k \right)
\end{aligned}$$

A.6. Ableitung nach der Transformation $T_k = \{A_k, t_k\}$

$$\begin{aligned}
& \frac{\partial}{\partial T_k} C(Q, \Theta) \\
&= \sum_{k=1}^N \left(\sum_{p=1}^n \frac{\partial}{\partial T_k} \log(\lambda_p) + \left(\frac{\omega_{k,p}^2}{2\lambda_p^2} \right) \right. \\
&\quad + \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1}(\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \frac{\partial}{\partial T_k} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \\
&\quad - \sum_{j=1}^{N_m} \sum_{\eta \in \mathcal{N}(j)} \sum_{p=1}^n - (\Sigma_{\mathcal{N}}^{-1} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta})) - \mathcal{N}_{j,\eta}^0}{\|\mathcal{N}_{j,\eta}^0\|} \right)^T \\
&\quad \frac{\partial}{\partial T_k} \left(\frac{((\bar{\mathbf{m}}_j \pm 2 \cdot \lambda_p \mathbf{v}_{p,j}) - (\bar{\mathbf{m}}_{\eta} \pm 2 \cdot \lambda_p \mathbf{v}_{p,\eta}))}{\|\mathcal{N}_{j,\eta}^0\|} \right) \\
&\quad - \left(\frac{\frac{1}{N} \sum_{k=1}^N \boldsymbol{\omega}_k}{\sigma_{\bar{\omega}}^2} \right)^T \frac{\partial}{\partial T_k} \left(\frac{1}{N} \sum_{k=1}^N \boldsymbol{\omega}_k \right) \\
&\quad - \left(\frac{(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}}{\sigma_{\bar{T}}^2} \right) \frac{\partial}{\partial T_k} \left(\left(\frac{1}{N} \sum_{k=1}^N T_k \right) - I_{D+1} \right) \\
&= + \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1}(\mathbf{s}_{k,i} - \mathbf{m}_{k,j})^T \frac{\partial}{\partial T_k} (\mathbf{s}_{k,i} - \mathbf{m}_{k,j}) \\
&\quad - \left(\frac{(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}}{\sigma_{\bar{T}}^2} \right) \frac{\partial}{\partial T_k} \left(\left(\frac{1}{N} \sum_{k=1}^N T_k \right) - I_{D+1} \right) \\
&= + \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1}(\mathbf{s}_{k,i} - (T_k^{-1} \star (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j})))^T \\
&\quad \frac{\partial}{\partial T_k} (\mathbf{s}_{k,i} - (T_k^{-1} \star (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}))) \\
&\quad - \left(\frac{(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}}{\sigma_{\bar{T}}^2} \right) \frac{\partial}{\partial T_k} \left(\left(\frac{1}{N} \sum_{k=1}^N T_k \right) - I_{D+1} \right) \\
&= + \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1}(T_k \star \mathbf{s}_{k,i} - (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}))^T \\
&\quad \frac{\partial}{\partial T_k} (T_k \star \mathbf{s}_{k,i} - (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j})) \\
&\quad - \left(\frac{(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}}{\sigma_{\bar{T}}^2} \right) \frac{\partial}{\partial T_k} \left(\left(\frac{1}{N} \sum_{k=1}^N T_k \right) - I_{D+1} \right)
\end{aligned}$$

nach fixPos-Variante:

$$\begin{aligned}
\frac{\partial}{\partial T_k} C(Q, \Theta) &= \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}))^T \\
&\quad \frac{\partial}{\partial T_k} (T_k \star \mathbf{s}_{k,i} - (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j})) \\
&= \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}))^T \\
&\quad \frac{\partial}{\partial T_k} (I(T_k^{-1}(\mathbf{x}_{k,i})) - (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j})) \\
&= \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}))^T \\
&\quad \frac{\partial}{\partial T_k} (I(T_k^{-1}(\mathbf{x}_{k,i}))) \\
&= \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (T_k \star \mathbf{s}_{k,i} - (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}))^T \\
&\quad \nabla I(T_k^{-1}(\mathbf{x}_{k,i})) \frac{\partial}{\partial T_k} (T_k^{-1}(\mathbf{x}_{k,i}))
\end{aligned}$$

nach fixApp-Variante:

$$\begin{aligned}
\frac{\partial}{\partial T_k} C(Q, \Theta) &= \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (A_k \mathbf{s}_{k,i} + t_k - (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}))^T \\
&\quad \frac{\partial}{\partial T_k} (A_k \mathbf{s}_{k,i} + t_k - (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j})) \\
&\quad - \left(\frac{(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}}{\sigma_{\bar{T}}^2} \right) \frac{\partial}{\partial T_k} \left(\left(\frac{1}{N} \sum_{k=1}^N T_k \right) - I_{D+1} \right)
\end{aligned}$$

Aufteilung in Translation t und affinem Anteil A :

$$\begin{aligned}
& \frac{\partial}{\partial t_k} C(Q, \Theta) \\
&= \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (A_k \mathbf{s}_{k,i} + t_k - (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}))^T \\
&\quad - \left(\frac{(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}}{\sigma_{\bar{T}}^2} \right) \frac{\partial}{\partial t_k} \left(\left(\frac{1}{N} \sum_{k=1}^N T_k \right) - I_{D+1} \right) \\
&= \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (A_k \mathbf{s}_{k,i})^T + \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (t_k) \\
&\quad - \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} ((\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j})^T) \\
&\quad - \frac{t_k}{N^2 \sigma_{\bar{T}}^2} - \left(\frac{\sum_{l=1, l \neq k}^N t_l}{N^2 \sigma_{\bar{T}}^2} \right) \\
&= \sum_{i=1}^{N_s(k)} \Sigma^{-1} (A_k \mathbf{s}_{k,i})^T + N_s(k) \Sigma^{-1} (t_k) - \sum_{j=1}^{N_m} \Sigma^{-1} ((\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j})^T) \sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k \\
&\quad - \frac{t_k}{N^2 \sigma_{\bar{T}}^2} - \left(\frac{\sum_{l=1, l \neq k}^N t_l}{N^2 \sigma_{\bar{T}}^2} \right) \\
t_k &= \frac{-\Sigma^{-1} A_k \sum_{i=1}^{N_s(k)} \mathbf{s}_{k,i}^T + \Sigma^{-1} \sum_{j=1}^{N_m} (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j})^T \sum_{i=1}^{N_s(k)} \mathbb{E}(H_{ij})_k}{\Sigma^{-1} N_s(k) - \frac{1}{N^2 \sigma_{\bar{T}}^2}} \dots \\
&\quad + \left(\frac{\sum_{l=1, l \neq k}^N t_l}{N^2 \sigma_{\bar{T}}^2} \right) \\
&\quad \frac{1}{\Sigma^{-1} N_s(k) - \frac{1}{N^2 \sigma_{\bar{T}}^2}}
\end{aligned}$$

$$\begin{aligned}
& \frac{\partial}{\partial A_k} C(Q, \Theta) \\
&= \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k \Sigma^{-1} (A_k \mathbf{s}_{k,i} + t_k - (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}))^T \mathbf{s}_{k,i} \\
&\quad - \left(\frac{(\frac{1}{N} \sum_{k=1}^N T_k) - I_{D+1}}{\sigma_{\bar{T}}^2} \right) \frac{\partial}{\partial A_k} \left(\left(\frac{1}{N} \sum_{k=1}^N T_k \right) - I_{D+1} \right) \\
&= \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k ((A_k \mathbf{s}_{k,i} + t_k) \mathbf{s}_{k,i}^T) \\
&\quad - \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k ((\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}) \mathbf{s}_{k,i}^T) - \left(\frac{(\frac{1}{N} \sum_{k=1}^N A_k) - I_D}{N \sigma_{\bar{T}}^2} \right) \\
&= \sum_{i=1}^{N_s(k)} (A_k \mathbf{s}_{k,i} \mathbf{s}_{k,i}^T) - \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k ((\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}) \mathbf{s}_{k,i}^T - t_k \mathbf{s}_{k,i}^T) \\
&\quad - \left(\frac{A_k}{N^2 \sigma_{\bar{T}}^2} \right) - \left(\frac{(\frac{1}{N} \sum_{l=1, l \neq k}^N A_l) - I_D}{N \sigma_{\bar{T}}^2} \right) \\
&= A_k \sum_{i=1}^{N_s(k)} (\mathbf{s}_{k,i} \mathbf{s}_{k,i}^T) - \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k ((\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}) \mathbf{s}_{k,i}^T - t_k \mathbf{s}_{k,i}^T) \\
&\quad - \left(\frac{A_k}{N^2 \sigma_{\bar{T}}^2} \right) - \left(\frac{(\frac{1}{N} \sum_{l=1, l \neq k}^N A_l) - I_D}{N \sigma_{\bar{T}}^2} \right) \\
E &= \sum_{i=1}^{N_s(k)} (\mathbf{s}_{k,i} \mathbf{s}_{k,i}^T) - \left(\frac{1}{N^2 \sigma_{\bar{T}}^2} \right) \\
F &= \sum_{i=1}^{N_s(k)} \sum_{j=1}^{N_m} \mathbb{E}(H_{ij})_k ((\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}) \mathbf{s}_{k,i}^T - t_k \mathbf{s}_{k,i}^T) \\
&\quad + \left(\frac{\frac{1}{N} (\sum_{l=1, l \neq k}^N A_l) - I_D}{N \sigma_{\bar{T}}^2} \right) \\
0 &= A_k E - F \\
A_k &= F E^{-1}
\end{aligned}$$

Literatur

- Adeshina, Steve A., Lindner, Claudia, & Cootes, Timothy F. 2017. Automatic Annotation of Radiographs using Random Forest Regression Voting for Building Statistical Models for Skeletal Maturity. *International Journal of Computer Techniques*, **4**.
- Allasonnière, S., Amit, Y., & Trouvé, A. 2007. Towards a Coherent Statistical Framework for Dense Deformable Template Estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **69**(1), 3–29.
- Andreopoulos, Alexander, & Tsotsos, John K. 2008. Efficient and Generalizable Statistical Models of Shape and Appearance for Analysis of Cardiac MRI. *Medical Image Analysis*, **12**(3), 335 – 357.
- Behiels, Gert, Maes, Frederik, Vandermeulen, Dirk, & Suetens, Paul. 2002. Evaluation of Image Features and Search Strategies for Segmentation of Bone Structures in Radiographs using Active Shape Models. *Medical Image Analysis*, **6**(1), 47 – 62.
- Brett, A D, & Taylor, C J. 2000a. Automated Construction of 3D Shape Models using Harmonic Maps. *In: Medical Image Understanding and Analysis*. BMVA.
- Brett, A.D, & Taylor, C.J. 2000b. A Method of Automated Landmark Generation for Automated 3D PDM Construction. *Image and Vision Computing*, **18**(9), 739 – 748.
- Chitphakdithai, Nicha, & Duncan, James S. 2010. *Non-rigid Registration with Missing Correspondences in Preoperative and Postresection Brain Images*. Berlin, Heidelberg: Springer Berlin Heidelberg. Pages 367–374.
- Chui, Haili, & Rangarajan, Anand. 2003. A New Point Matching Algorithm for Non-rigid Registration. *Computer Vision and Image Understanding*, **89**(2-3), 114 – 141.
- Cootes, T. F., & Taylor, C. J. 1994 (Oct). Using Grey-level Models to Improve Active Shape Model Search. *Pages 63–67 vol.1 of: Proceedings of 12th International Conference on Pattern Recognition*, vol. 1.
- Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. 1992. Training Models of Shape from Sets of Examples. *Pages 2.1–2.10 of: Proceedings of the British Machine Vision Conference*. BMVA Press. doi:10.5244/C.6.2.
- Cootes, T. F., Edwards, G. J., & Taylor, C. J. 2001. Active Appearance Models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **23**(6), 681–685.
- Cootes, T. F., Twining, C. J., Babalola, K. O., & Taylor, C. J. 2008. Diffeomorphic Statistical Shape Models. *Image Vision Comput.*, **26**(3), 326–332.

- Cootes, T.F., Taylor, C.J., Cooper, D.H., & Graham, J. 1995. Active Shape Models – Their Training and Application. *Computer Vision and Image Understanding*, **61**(1), 38 – 59.
- Cootes, Tim F., Ionita, Mircea C., Lindner, Claudia, & Sauer, Patrick. 2012. *Robust and Accurate Shape Model Fitting Using Random Forest Regression Voting*. Berlin, Heidelberg: Springer Berlin Heidelberg. Pages 278–291.
- Cristinacce, David, & Cootes, Tim. 2006. Feature Detection and Tracking with Constrained Local Models. *Pages 95.1–95.10 of: Proceedings of the British Machine Vision Conference*. BMVA Press.
- Cristinacce, David, & Cootes, Tim. 2008. Automatic Feature Localisation with Constrained Local Models. *Pattern Recognition*, **41**(10), 3054 – 3067.
- Dam, Erik B., Fletcher, P. Thomas, & Pizer, Stephen M. 2008. Automatic Shape Model Building Based on Principal Geodesic Analysis Bootstrapping. *Medical Image Analysis*, **12**(2), 136 – 151.
- Davatzikos, C., Tao, Xiaodong, & Shen, Dinggang. 2003. Hierarchical Active Shape Models, Using the Wavelet Transform. *IEEE Transactions on Medical Imaging*, **22**(3), 414–423.
- Davies, Rhodri H., Twining, Carole J., Cootes, Tim. F., Waterton, John C., & Taylor, Chris. J. 2001. A Minimum Description Length Approach to Statistical Shape Modelling. *IEEE Transactions on Medical Imaging*, **21**, 525–537.
- Davies, Rhodri H., Twining, Carole J., Cootes, Tim F., Waterton, John C., & Taylor, Chris J. 2002. *3D Statistical Shape Models Using Direct Optimisation of Description Length*. Berlin, Heidelberg: Springer Berlin Heidelberg. Pages 3–20.
- de Bruijne, Marleen, van Ginneken, Bram, Viergever, Max A., & Niessen, Wiro J. 2003. Adapting Active Shape Models for 3D Segmentation of Tubular Structures in Medical Images. *Pages 136–147 of: Taylor, Christopher J., & Noble, J. Alison (eds), IPMI. Lecture Notes in Computer Science*, vol. 2732. Springer.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**(1), 1–38.
- Donner, Rene, Langs, Georg, & Bischof, Horst. 2007. Sparse MRF Appearance Models for Fast Anatomical Structure Localisation. *Page 111 of: In British Machine Vision Conference*.

- Ecabert, O., Peters, J., Schramm, H., Lorenz, C., von Berg, J., Walker, M. J., Vembar, M., Olszewski, M. E., Subramanyan, K., Lavi, G., & Weese, J. 2008. Automatic Model-Based Segmentation of the Heart in CT Images. *IEEE Transactions on Medical Imaging*, **27**(9), 1189–1201.
- Ehrhardt, Jan, Krüger, Julia, & Handels, Heinz. 2014. Statistical Shape and Appearance Models without One-to-One Correspondences. *Pages 90340U,1–7 of: Ourselin, S., & Styner, M.A. (eds), SPIE Medical Imaging 2014, Image Processing*, vol. 9034.
- Ehrhardt, Jan, Schmidt-Richberg, Alexander, Werner, René, & Handels, Heinz. 2015. *Variational Registration*. Berlin, Heidelberg: Springer Berlin Heidelberg. Pages 209–214.
- Ferrarini, Luca, Olofsen, Hans, Palm, Walter M., van Buchem, Mark A., Reiber, Johan H. C., & Admiraal-Behloul, Faiza. 2007. GAMEs: Growing and Adaptive Meshes for Fully Automatic Shape Modeling and Analysis. *Medical Image Analysis*, **11**(3), 302–314.
- Fletcher, P. Thomas, Joshi, Sarang C., Lu, Conglin, & Pizer, Stephen M. 2003. Gaussian Distributions on Lie Groups and Their Application to Statistical Shape Analysis. *Pages 450–462 of: Taylor, Christopher J., & Noble, J. Alison (eds), IPMI. Lecture Notes in Computer Science*, vol. 2732. Springer.
- Floater, Michael S., & Hormann, Kai. 2005. Surface Parameterization: a Tutorial and Survey. *Pages 157–186 of: Advances in Multiresolution for Geometric Modelling*.
- Frangi, Alejandro F., Niessen, Wiro J., Rueckert, Daniel, & Schnabel, Julia A. 2001. *Automatic 3D ASM Construction via Atlas-Based Landmarking and Volumetric Elastic Registration*. Berlin, Heidelberg: Springer Berlin Heidelberg. Pages 78–91.
- Fripp, J., Crozier, S., Warfield, S., & Ourselin, S. 2005 (Dec). Automatic Initialization of 3D Deformable Models for Cartilage Segmentation. *Pages 74–74 of: Digital Image Computing: Techniques and Applications (DICTA '05)*.
- Gold, Steven, Rangarajan, Anand, ping Lu, Chien, & Mjolsness, Eric. 1997. New Algorithms for 2D and 3D Point Matching: Pose Estimation and Correspondence. *Pattern Recognition*, **31**, 957–964.
- Gollmer, Sebastian T., & Buzug, Thorsten M. 2013. Relaxed Statistical Shape Models for 3D Image Segmentation – Application to Mandible Bone in Cone-beam CT Data. *Current Medical Imaging Reviews*, **9**(2), 129–137.
- Granger, Sébastien, & Pennec, Xavier. 2002. Multi-scale EM-ICP: A Fast and Robust Approach for Surface Registration. *Pages 418–432 of: Computer Vision - ECCV 2002. Lecture Notes in Computer Science*, vol. 2353. Springer.

- Guimond, A., Meunier, J., & Thirion, J.-P. 2000. Average Brain Models: A Convergence Study. *Computer Vision and Image Understanding*, **77**(2), 192–210.
- Harmsen, M., Fischer, B., Schramm, H., Seidl, T., & Deserno, T. M. 2013. Support Vector Machine Classification Based on Correlation Prototypes Applied to Bone Age Assessment. *IEEE Journal of Biomedical and Health Informatics*, **17**(1), 190–197.
- Hartkens, Thomas, Hill, Derek L. G., Castellano-Smith, Andy D., Hawkes, David J., Maurer, Jr., Calvin R., Martin, A. J., Hall, Walter A., Liu, H., & Truwit, Charles L. 2002. Using Points and Surfaces to Improve Voxel-Based Non-rigid Registration. *Pages 565–572 of: Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention-Part II. MICCAI '02*. London, UK, UK: Springer-Verlag.
- Heimann, Tobias, & Meinzer, Hans-Peter. 2009. Statistical Shape Models for 3D Medical Image Segmentation: A Review. *Medical Image Analysis*, **13**(4), 543 – 563.
- Heitz, Jeremy, Rohlfing, Torsten, & Jr., Calvin R. Maurer. 2004. Automatic Generation of Shape Models Using Nonrigid Registration with a Single Segmented Template Mesh. *Pages 73–80 of: Proceedings of the Vision, Modeling, and Visualization Conference 2004 (VMV 2004), Stanford, California, USA, November 16–18, 2004*.
- Heitz, Jeremy, Rohlfing, Torsten, & Maurer, Jr., Calvin R. 2005. *Statistical Shape Model Generation using Nonrigid Deformation of a Template Mesh*.
- Hufnagel, Heike. 2010 (July). *A Probabilistic Framework for Point-based Shape Modeling in Medical Image Analysis*. PhD Thesis, University of Lübeck.
- Hufnagel, Heike, Pennec, Xavier, Ehrhardt, Jan, Handels, Heinz, & Ayache, Nicholas. 2007. Shape Analysis Using a Point-Based Statistical Shape Model Build on Correspondence Probabilities. *Pages 959 –967 of: 10th International Conference on Medical Image Computing and Computer-Assisted Intervention 2007, MICCAI 2007*, vol. Part I. Springer Berlin Heidelberg.
- Hufnagel, Heike, Pennec, Xavier, Ehrhardt, Jan, Ayache, Nicholas, & Handels, Heinz. 2008. Generation of a Statistical Shape Model with Probabilistic Point Correspondences and the Expectation Maximization-Iterative Closest Point Algorithm. *International Journal of Computer Assisted Radiology and Surgery*, **2**(3/2008), 265 – 273.
- Hufnagel, Heike, Pennec, X., Ehrhardt, Jan, Ayache, Nicholas, & Handels, Heinz. 2009. Computation of a Probabilistic Statistical Shape Model in a Maximum-a-Posteriori Framework. *Methods of Information in Medicine*, **48**(2009), 314 – 319.
- Hug, Johannes, Brechbühler, Christian, & Székely, Gábor. 2000. *Model-Based Initialisation for Segmentation*. Berlin, Heidelberg: Springer Berlin Heidelberg. Pages 290–306.

- Jian, Bing, & Vemuri, Baba C. 2011. Robust Point Set Registration Using Gaussian Mixture Models. *IEEE Trans. Pattern Anal. Mach. Intell.*, **33**(8), 1633–1645.
- Jiao, Feng, Li, Stan, Shum, Heung-Yeung, & Schuurmans, D. 2003 (June). Face Alignment Using Statistical Models and Wavelet Features. *Pages I-321–I-327 vol.1 of: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 1.
- Josephson, Klas, Ericsson, Anders, & Karlsson, Johan. 2005. Segmentation of Medical Images Using Three-dimensional Active Shape Models. *Pages 719–728 of: Proceedings of the 14th Scandinavian Conference on Image Analysis. SCIA'05.* Berlin, Heidelberg: Springer-Verlag.
- Kelemen, A., Szekely, G., & Gerig, G. 1999. Elastic Model-based Segmentation of 3-D Neuroradiological Data Sets. *IEEE Transactions on Medical Imaging*, **18**(10), 828–839.
- Klein, Arno, Andersson, Jesper, Ardekani, Babak A., Ashburner, John, Avants, Brian, Chiang, Ming-Chang, Christensen, Gary E., Collins, D. Louis, Gee, James, Hellier, Pierre, Song, Joo Hyun, Jenkinson, Mark, Lepage, Claude, Rueckert, Daniel, Thompson, Paul, Vercauteren, Tom, Woods, Roger P., Mann, J. John, & Parsey, Ramin V. 2009. Evaluation of 14 Nonlinear Deformation Algorithms Applied to Human Brain MRI Registration. *NeuroImage*, **46**(3), 786 – 802.
- Krüger, Julia, Ehrhardt, Jan, Bischof, Arpad, & Handels, Heinz. 2014. Simulation of Mammographic Breast Compression in 3D MR Images Using ICP-Based B-Spline Deformation for Multimodality Breast Cancer Diagnosis. *International Journal of Computer Assisted Radiology and Surgery*, **9**, 367–377.
- Krüger, Julia, Ehrhardt, Jan, & Handels, Heinz. 2015a. A Maximum-A-Posteriori Framework for Statistical Appearance Models with Probabilistic Correspondences. *In: Bayesian and graphical Models for Biomedical Imaging 2015, MICCAI 2015.*
- Krüger, Julia, Ehrhardt, Jan, & Handels, Heinz. 2015b (Dez). Probabilistic Appearance Models for Segmentation and Classification. *Pages 1698–1706 of: The IEEE International Conference on Computer Vision (ICCV).*
- Krüger, Julia, Ehrhardt, Jan, & Handels, Heinz. 2016. Multi-level Approach for Statistical Appearance Models with Probabilistic Correspondences. *Pages 9784,1–6 of: Ourselin, S., & Styner, M.A. (eds), SPIE Medical Imaging 2016, Image Processing*, vol. 9784.
- Krüger, Julia, Ehrhardt, Jan, & Handels, Heinz. 2017. Statistical Appearance Models based on Probabilistic Correspondences. *Medical Image Analysis*, **37**, 146 – 159.

- Kwon, D., Niethammer, M., Akbari, H., Bilello, M., Davatzikos, C., & Pohl, K. M. 2014. PORTR: Pre-Operative and Post-Recurrence Brain Tumor Registration. *IEEE Transactions on Medical Imaging*, **33**(3), 651–667.
- Lai, Rongjie, & Osher, Stanley. 2014. A Splitting Method for Orthogonality Constrained Problems. *Journal of Scientific Computing*, **58**(2), 431–449.
- Larsen, Rasmus, Stegmann, Mikkel B., Darkner, Sune, Forchhammer, Søren, Cootes, Timothy F., & Ersbøll, Bjarne Kær. 2007. Texture Enhanced Appearance Models. *Computer Vision and Image Understanding*, **106**(1), 20 – 30. Special issue on Generative Model Based Vision.
- Lee, Seungyong, Wolberg, George, & Shin, Sung Yong. 1997. Scattered Data Interpolation with Multilevel B-Splines. *IEEE Transactions on Visualization and Computer Graphics*, **3**, 228–244.
- Lee, Tai Sing. 1996. Image Representation Using 2D Gabor Wavelets. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **18**, 959–971.
- Lindner, C., Bromiley, P. A., Ionita, M. C., & Cootes, T. F. 2015. Robust and Accurate Shape Model Matching Using Random Forest Regression-Voting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **37**(9), 1862–1874.
- Maier, Oskar, Wilms, Matthias, von der Gablentz, Janina, Krämer, Ulrike M., Münte, Thomas F., & Handels, Heinz. 2015. Extra Tree Forests for Sub-acute Ischemic Stroke Lesion Segmentation in MR Sequences. *Journal of Neuroscience Methods*, **240**(0), 89 – 100.
- Malladi, R., Sethian, J. A., & Vemuri, B. C. 1995. Shape Modeling with Front Propagation: a Level Set Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**(2), 158–175.
- Mang, A, Toma, A, ans S Becker, T A Schuetz, & Buzug, T M. 2012. A Generic Framework for Modeling Brain Deformation as a Constrained Parametric Optimization Problem to Aid Non-diffeomorphic Image Registration in Brain Tumor Imaging. *Methods Inf Med*, **51**(5), 429–440.
- Murphy, Keelin, van Ginneken, Bram, Pluim, Josien P.W., Klein, Stefan, & Staring, Marius. 2008. Semi-automatic Reference Standard Construction for Quantitative Evaluation of Lung CT Registration. *Pages 1006–1013 of: Metaxas, Dimitris, Axel, Leon, Fichtinger, Gabor, & Szekely, Gabor (eds), Medical Image Computing and Computer-Assisted Intervention - MICCAI 2008. Lecture Notes in Computer Science*, vol. 5242. Springer Berlin Heidelberg.

- Myronenko, A., & Song, Xubo. 2010. Point Set Registration: Coherent Point Drift. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **32**(12), 2262–2275.
- Parisot, Sarah, Wells, William, Chemouny, Stéphane, Duffau, Hugues, & Paragios, Nikos. 2014. Concurrent Tumor Segmentation and Registration with Uncertainty-based Sparse Non-uniform Graphs. *Medical image analysis*, **18**(4), 647–659.
- Press, William H., Teukolsky, Saul A., Vetterling, William T., & Flannery, Brian P. 1992. *Numerical Recipes in C (2Nd Ed.): The Art of Scientific Computing*. New York, NY, USA: Cambridge University Press.
- Rohlfing, Torsten, Jr., Calvin R. Maurer, Bluemke, David A., & Jacobs, Michael A. 2003. Volume-Preserving Nonrigid Registration of MR Breast Images Using Free-Form Deformation with an Incompressibility Constraint. *IEEE Transactions on Medical Imaging*, **22**, 730–741.
- Rueckert, D., Frangi, A. F., & Schnabel, J. A. 2003. Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. *IEEE Transactions on Medical Imaging*, **22**(8), 1014–1025.
- Shang, Y., & Dossel, O. 2004 (Sept). Statistical 3D Shape-model Guided Segmentation of Cardiac Images. *Pages 553–556 of: Computers in Cardiology, 2004*.
- Shattuck, David W., Mirza, Mubeena, Adisetiyo, Vitria, Hojatkashani, Cornelius, Salamon, Georges, Narr, Katherine L., Poldrack, Russell A., Bilder, Robert M., & Toga, Arthur W. 2008. Construction of a 3D Probabilistic Atlas of Human Cortical Structures. *NeuroImage*, **39**(3), 1064 – 1080.
- Shen, Dinggang, Zhan, Yiqiang, & Davatzikos, Christos. 2003. Segmentation of Prostate Boundaries from Ultrasound Images using Statistical Shape Model. *IEEE Transactions on Medical Imaging*, **22**(4), 539–551.
- Staib, L. H., & Duncan, J. S. 1996. Model-based Deformable Surface Finding for Medical Images. *IEEE Transactions on Medical Imaging*, **15**(5), 720–731.
- Staring, Marius, Klein, Stefan, & Pluim, Josien P W. 2007. Nonrigid Registration with Tissue-dependent Filtering of the Deformation Field. *Physics in Medicine & Biology*, **52**(23), 6879.
- Styzt, Martin R., & Parrott, Rob W. 1993. Using Kriging for 3D Medical Imaging. *Computerized Medical Imaging and Graphics*, **17**(6), 421 – 442.
- Subsol, Gérard, Thirion, Jean-Philippe, & Ayache, Nicholas. 1998. A Scheme for Automatically Building Three-dimensional Morphometric Anatomical Atlases: Application to a Skull Atlas. *Medical Image Analysis*, **2**(1), 37 – 60.

- Sundberg, Rolf. 1974. Maximum Likelihood Theory for Incomplete Data from an Exponential Family. *Scandinavian Journal of Statistics*, **1**(2), 49–58.
- Tsagaan, Baigalmaa, Shimizu, Akinobu, Kobatake, Hidefumi, & Miyakawa, Kunihisa. 2002. An Automated Segmentation Method of Kidney Using Statistical Information. *Pages 556–563 of: Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention-Part I. MICCAI '02*. London, UK, UK: Springer-Verlag.
- Turk, Matthew, & Pentland, Alex. 1991. Eigenfaces for Recognition. *J. Cognitive Neuroscience*, **3**(1), 71–86.
- Tzimiropoulos, Georgios, & Pantic, Maja. 2014. *Gauss-Newton Deformable Part Models for Face Alignment In-the-Wild*. IEEE Computer Society. eemcs-eprint-25815. Pages 1851–1858.
- Tzimiropoulos, Georgios, & Pantic, Maja. 2016. Fast Algorithms for Fitting Active Appearance Models to Unconstrained Images. **122**(1), 17–33.
- Uzunova, Hristina, Wilms, Matthias, Handels, Heinz, & Ehrhardt, Jan. 2017. Training CNNs for Image Registration from Few Samples with Model-based Data Augmentation. *In: 20th International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI 2017*.
- Vos, F. M., de Bruin, P. W., Aubel, J. G. M., Streekstra, G. J., Maas, M., van Vliet, L. J., & Vossepoel, A. M. 2004 (Aug). A Statistical Shape Model without using Landmarks. *Pages 714–717 Vol.3 of: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 3.
- Werner, René, Schmidt-Richberg, Alexander, Handels, Heinz, & Ehrhardt, Jan. 2014. Estimation of Lung Motion Fields in 4D CT Data by Variational Non-Linear Intensity-Based Registration: A Comparison and Evaluation Study. *Physics in Medicine and Biology*, **59**, 4247–4260.
- Wu, C. F. Jeff. 1983. On the Convergence Properties of the EM Algorithm. *Ann. Statist.*, **11**(1), 95–103.
- Xu, Hao, Thirion, Bertrand, & Allasonnière, Stéphanie. 2014. Probabilistic Atlas and Geometric Variability Estimation to Drive Tissue Segmentation. *Statistics in Medicine*, **33**(20), 3576–3599.
- Yu, P., Grant, P. E., Qi, Y., Han, X., Segonne, F., Pienaar, R., Busa, E., Pacheco, J., Makris, N., Buckner, R. L., Golland, P., & Fischl, B. 2007. Cortical Surface Shape Analysis Based on Spherical Wavelets. *IEEE Transactions on Medical Imaging*, **26**(4), 582–597.

- Yu, Yen-YunY., Fletcher, PThomas, & Awate, SuyashP. 2014. Hierarchical Bayesian Modeling, Estimation, and Sampling for Multigroup Shape Analysis. *Med Image Comput Comput Assist Interv*, **17**(Pt 3), 9–16.
- Zacharaki, E. I., Hoge, C. S., Biros, G., & Davatzikos, C. 2008. A Comparative Study of Biomechanical Simulators in Deformable Registration of Brain Tumor Images. *IEEE Transactions on Biomedical Engineering*, **55**(3), 1233–1236.
- Zhang, Pei, & Cootes, Timothy F. 2012. Automatic Construction of Parts+Geometry Models for Initializing Groupwise Registration. *IEEE Trans. Med. Imaging*, **31**(2), 341–358.
- Zhao, Zheen, & Teoh, Eam Khwang. 2005. A Novel Framework for Automated 3D PDM Construction using Deformable Models. vol. 5747.