

Belegarbeit

**Effiziente Einbringung von statistischem
Formwissen in die Segmentierung von
ATPase gefärbten Muskelfaserbildern**

Jens Bornschein

Matrikelnummer: 3017075

<s4170545@inf.tu-dresden.de>

Betreuer und verantwortliche Hochschullehrer:

Dr. rer. nat. habil. Boris Flach

Dr. Thomas Brox

Professur Intelligente Systeme
Institut für künstliche Intelligenz
Fakultät Informatik
Technische Universität Dresden

Zusammenfassung

Diese Arbeit beschäftigt sich mit dem Statistischen Mittel der Hauptkomponentenanalyse zur Nutzung bei der Segmentierung von ATPase eingefärbten Muskelfaserzellen. Vorhandenes a priori Wissen wird hierdurch effizient nutzbar gemacht. Dabei wird die Segmentierung durch einen Energieminimierungsprozess auf Level-Set-Funktionen realisiert und eine Konturevolution auf wenigen formbeschreibenden Parametern ermöglicht.

Inhaltsverzeichnis

1. Einführung	4
2. Medizinische Grundlagen	5
2.1. Muskelfasern	5
2.2. Myofibrilläre ATPase-Färbung	6
3. Hauptkomponentenanalyse	8
3.1. Mathematische Grundlagen und praktische Realisierung	8
3.1.1. Begriffsbestimmung	9
3.1.2. Ermittlung der Hauptkomponenten	11
3.1.3. Approximation und Rücktransformation	12
3.2. Hauptkomponentenanalyse für hochdimensionale Daten	13
3.3. Schwachstellen der PCA	14
3.4. Ein Beispiel	14
3.5. Zusammenfassung	17
4. Segmentierung	18
4.1. Segmentierungsansätze	18
4.2. Formdaten und statistische Auswertung	21
4.2.1. Beispieldaten	21
4.2.2. Formrepräsentation	22
4.3. Umsetzung der Hauptkomponentenanalyse	23
4.4. Energieminimierung	25
4.5. Gradientenabstieg	29
4.5.1. Probleme und deren Lösungsansätze	30
4.6. Seeded-Region-Growing-Ansatz	32
4.7. Zusammenfassung	33
5. Ergebnisse	34
6. Zusammenfassung und Ausblick	37
A. Anhang	40
A.1. Erklärung	40
A.2. Danksagungen	41
A.3. Auszug aus den Beispieldaten	41
A.4. Vergleich über die Anzahl an Datensätzen	42

1. Einführung

Digitale Bilder haben in der heutigen Medizin einen hohen Stellenwert eingenommen. Dabei gewinnt die *Medizinische Bildverarbeitung* ständig an Bedeutung, da sie neue Wege der Diagnostik und Therapie ermöglicht hat. Gekennzeichnet ist sie durch den Einsatz von Techniken und Methoden der Bildanalyse, Mustererkennung und Visualisierung. Im Speziellen können Bildverarbeitungsprozesse bei der Diagnose oder Auswertung eine Entlastung darstellen [Han00]. Die pathologische Bestimmung von Muskelfasertypen in Biopsiprobe¹ ist ein aufwendiger Prozess. Die manuelle Markierung und eventuelle Segmentierung der einzelnen Zellen, sowie die spätere Erhebung der Ergebnisdaten ist oftmals zeit- und kostenintensiv. Diesen Auswertungsprozess zu automatisieren wäre vor allem bei einer größeren Anzahl an auszuwertenden Bildern sinnvoll und komfortabel.

Aus diesem Grund soll die Segmentierung von ATPase-gefärbten Muskelfaserbildern durch einen automatisierten Prozess realisiert werden. Dazu soll ein Energieminimierungsansatz entwickelt werden. Die Aufgabe der Segmentierung zeigt sich in unserem Fall in der Detektierung der bestmöglich passenden Kontur der einzelnen Zellen. Hierzu werden Merkmale über bereits segmentierten Zellen erhoben, welche zur besseren Bestimmung der Zellformen führen sollen. Diese werden durch den Einsatz von statistischen Hilfsmitteln - im Speziellen durch die Hauptkomponentenanalyse - zugänglich und besser nutzbar gemacht. Bei der Realisierung sollen grundlegende Ideen und Techniken zu einem neuen effizienten Ansatz zur lokalem Optimierung kombiniert werden. Hierzu wird der für dieses Problem bereits umgesetzte *Seeded-Region-Growing* Ansatz von A. ASSMUS [Asm08] mit den Arbeiten zur *Formrepräsentation durch Hauptkomponenten* von M. LEVENTON et. al. [LEGF00] und dem Energieminimierungsmodell von M. ROUSSON und D. CREMERS [CR07] genutzt.

Die folgende Arbeit soll in Kapitel 2 durch eine Einführung in die medizinischen Grundlagen begonnen werden. In Kapitel 3 wird das statistische Mittel der Hauptkomponentenanalyse vorgestellt. Dabei werden die mathematischen Grundlagen ebenso wie die praktische Ausführung betrachtet. Der eigentliche Prozess der Segmentierung wird in Kapitel 4 theoretisch entwickelt. Dessen praktische Ergebnisse werden anschließend im Kapitel 5 präsentiert.

¹Biopsie: Medizinischer Begriff für die Entnahme von Gewebematerial aus einem Organismus

2. Medizinische Grundlagen

Um einen groben Überblick über die eigentliche Zielsetzung und den Nutzen dieser Arbeit zu erhalten, soll eine entsprechende Einführung in die medizinischen Grundlagen gegeben werden. Dabei wird am Beispiel der menschlichen Skelettmuskulatur Aufbau und Funktionsweise von Muskeln beschrieben. Zum Verständnis des Färbeverfahrens, welches die Ausgangsdaten für die vorliegende Problemstellung liefert, bedarf es einer Vertiefung, um so eine Interpretation der Resultate zu ermöglichen.

2.1. Muskelfasern

Die Aufgabe der menschlichen Skelettmuskulatur ist die Erzeugung von Kraft. Dabei sind die einzelnen Muskelfasern bis zu 10 cm lang, haben jedoch nur einen Durchmesser von etwa $50 \mu\text{m}$ ¹ und besitzen mehrere Tausend Zellkerne [HH03]. Die einzelnen Fasern bündeln sich dann zu einem Muskelstrang.

Man unterscheidet in erster Linie zwischen zwei Muskelfasertypen: Zum einen den Typ-I, auch als langsame (*slow-twitch*) oder rote Muskelfasern bezeichnet, zum anderen den Typ-II, welche als schnelle (*fast-twitch*) oder weiße Fasern deklariert werden [HH03]. Dabei weisen beim Menschen die Typ-II Fasern eine drei- bis fünffach schnellere Kontraktionsgeschwindigkeit auf. Dies liegt an der Art der Energieumwandlung. Die Kraft, die ein Muskel entwickeln kann, ist einzig durch seinen Durchmesser bestimmt, nicht durch die Verteilung der Zelltypen. Beide Typen können die selbe Kraft erzeugen, wobei Typ-I-Fasern weniger Energie verbrauchen. Bei fast allen Kontraktionen eines Muskels werden zuerst die langsameren Fasern bemüht. Bei zunehmender Kraft werden dann auch die schnelleren Typen hinzugenommen. Anders verhält es sich mit der „Schnellkraft“, also der Anstiegsgeschwindigkeit der Kraft, welche hauptsächlich vom Anteil der Typ-II-Fasern abhängig ist. Die Ausdauer und damit die Zeitspanne bis zur Ermüdung der beiden Typen differiert jedoch zu Gunsten der Typ-I-Fasern. Die Anteile an Fasertypen sind bei den meisten Menschen etwa im gleichen Verhältnis, bei Leistungssportlern kann jedoch eine einseitige Fasertypenverteilung entstehen. So haben Ausdauersportler einen sehr viel höheren Anteil an Typ-I-Fasern [BM70], Sprinter oder Kurzstreckenschwimmer dagegen ein Übermaß an Typ-II-Fasern. Dies ist jedoch auch durch eine sehr hohe genetische Komponente bedingt. Auch gibt es hochspezialisierte Muskelpartien, wie die Augenmuskulatur oder der Herzmuskel, die eine eindeutige Verhältnisverschiebung aufweisen.

¹ $1 \mu\text{m} = \frac{1}{1.000} \text{mm}$

2.2. Myofibrilläre ATPase-Färbung

Zur Unterscheidung der einzelnen Zelltypen gilt die myofibrilläre ATPase-Färbung als das wichtigste und konstanteste histochemische² Verfahren [Sch82]. Grundlage ist dabei die Erkenntnis, dass die Kontraktionsgeschwindigkeit der einzelnen Muskelfasern in direktem Zusammenhang mit der Aktivität ihrer myofibrillären Adenosintriphosphatase (ATPase) steht.

Myosin ist dabei ein Protein, welches in Muskeln Energie in Kraft umsetzt und Bestandteil des kontraktionsfähigen Abschnitts einer *Myofibrille* ist. Diese Myofibrillen sind Zellorganellen³, welche es einer Muskelzelle ermöglichen, sich aktiv zu verkürzen. Bündel dieser auch als Muskelfibrillen bezeichneten Funktionseinheiten bilden eine Muskelfaser. Als Energie für zellinterne Prozesse wird das Molekül Adenosintriphosphat (ATP) benötigt, welches aus drei verschiedenen Phosphaten besteht. Um daraus Energie zu gewinnen, müssen die Phosphate wieder getrennt werden. Dies geschieht durch das Enzym ATPase. All diese Enzyme und Proteine können in unterschiedlichster Form auftreten. Um diesem Umstand Rechnung zu tragen, muss die in Abschnitt 2.1 getroffene Einteilung der Muskelfasertypen etwas verfeinert werden.

Die Typ-II-Fasern werden noch einmal in drei Untergruppen unterteilt, in die schnellen (A) und die langsamen (B) Zuckfasern sowie die Tonusfasern (C). Die Unterscheidung beruht zum einen auf der Art des auftretenden Myosins, zum anderen auf der Anbindung an das Nervensystem [HH03]. So enthalten Typ-I Zellen ausschließlich langsames Myosin, Typ-II-A und Typ-II-B nur rasches und Typ-II-C beide Arten dieses Proteins. Des Weiteren erkannte man, dass die ATPase in schnellen Zellen drei mal so aktiv wie in langsamen ist und das Enzym in den schnellen Zellen alkalistabil und säurelabil, in langsamen genau umgekehrt - also alkalilabil und säurestabil ist. Aus diesem Wissen heraus entwickelten BROOKE und KAISER 1970 [BM70] das Verfahren zur Färbung der ATPase und damit zur Unterscheidung der einzelnen Untertypen.

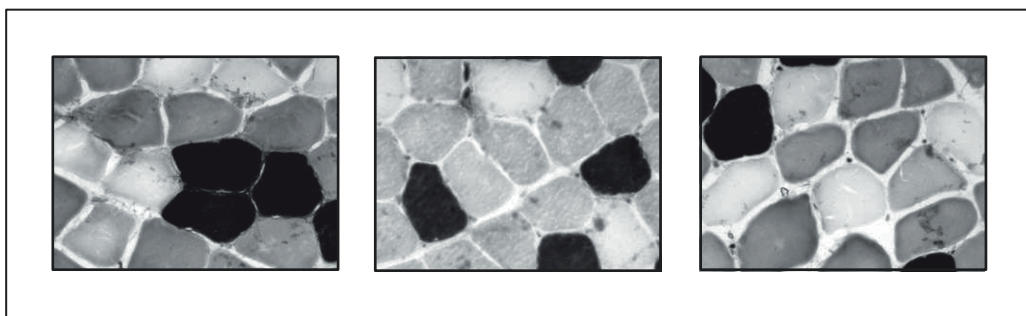


Abbildung 2.1. – Beispiele für eine ATPase Färbung

²Histochemie: Wissenschaft vom chemischen Aufbau und von chemischen Prozessen in Gewebe

³Organellen sind abgegrenzte Funktionseinheiten einer Zelle

Durch die Vorbehandlung der vorher eingefrorenen Muskelschnitte mit verschiedenen sauren oder alkalischen Lösungen⁴ werden die Eigenschaften der Präparate so verändert, dass die ATPase typspezifisch eingefärbt werden kann. Dieser Vorgang wird als *Präinkubation* bezeichnet. Die genauen Zusammensetzungen der einzelnen Lösungen können in [Sch82] auf den Seiten 9 bis 13 eingesehen werden. Aus diesem Färbeverfahren ergeben sich als Endresultat die unterschiedlich stark abgedunkelten Zellen (siehe Abb. 2.1). Eine Übersicht über die verwendete Lösung und die daraus resultierende typspezifische Färbung zeigt Tabelle 2.1. Bei der Zuordnung zu den Fasertypen ist die Färbemethode jedoch nicht immer eindeutig, da es einigen Zellen möglich ist ihren Typ zu ändern.

Muskelfasertyp	I	II-A	II-B	II-C
Standard ATPase bei pH 9,4	○	●	●	●
ATPase nach Präinkubation bei pH 4,6	●	○	●	●
ATPase nach Präinkubation bei pH 4,3	●	○	○	●

Tabelle 2.1. – Muskelfasertypenfärbung nach [Sch82]

Das für die Segmentierung benötigte Verfahren muss somit differenzierende Färbungen und unterschiedlichste Zellformen behandeln können. Zudem sollte es robust gegen Störungen im Bild sein. Schmutzpartikel, schlechte Färbung oder Zellrückstände erschweren den Segmentierungsprozess. Zur finalen Interpretation der einzelnen Bilder muss immer der Kontext über die Erstellung herangezogen werden.

⁴ pH < 7 ⇒ sauer pH = 7 ⇒ neutral (Wasser) pH > 7 ⇒ alkalisch

3. Hauptkomponentenanalyse

Die Hauptkomponentenanalyse (englisch *principal component analysis* - PCA) ist eine multivariate und variablenorientierte Technik. Weiterhin ist sie auch unter dem Namen *Karhunen-Loève-Transformation* bekannt und kann als Merkmalstransformationsverfahren bezeichnet werden [Kar47].

Ziel ist es, aus einer Menge unabhängiger Variablen oder Primärmerkmalen wenige neue, geeignete Variablen oder Sekundärmerkmale zu ermitteln, welche die Ausgangsdaten möglichst genau approximieren [RH95]. Hierzu wird die Redundanz in bzw. die Korrelation (siehe 3.1.1) von zum Teil hochdimensionalen Datensätzen ausgenutzt. Es wird also eine Menge korrelierter Daten durch Informationsverdichtung in eine Menge unkorrelierter Daten umgewandelt, welche ein hohes Maß der Gesamtvarianz abdeckt. Diese neue Menge wird aus *allen* vorhandenen Variablen generiert, sodass eine exakte Repräsentation der Ausgangsdaten möglich ist, sofern man alle neuen Komponenten mit einbezieht [CE36]. Die Eigenschaft, welche sich diese Arbeit hauptsächlich zu Nutzen macht, ist die Reduzierung der Dimension des Variablenraums auf eine im Allgemeinen erheblich kleinere Dimension, ohne dabei allzu viel Informationen zu verlieren.

Es ergeben sich neue statistische Variablen mit auf den Originaldaten gültigen Eigenschaften [Bök05]. Diese sind jedoch als abstrakte Konstrukte zu betrachten, über denen sich *unmittelbare* inhaltliche Interpretationen sehr schwer treffen lassen. Jedoch ist die Entropie - also der mittlere Informationsgehalt einer solchen Variable - meist höher als die der ursprünglichen Messvariablen [Hot33].

Da das gesamte Verfahren auf der Annahme von korrelierten Daten gründet, ist eine vorherige Untersuchung notwendig, ob überhaupt Korrelation besteht. Anderenfalls wäre das Verfahren überflüssig oder würde einen Datensatz liefern, welcher sehr nahe am Original liegt.

3.1. Mathematische Grundlagen und praktische Realisierung

Die Hauptkomponentenanalyse basiert auf verschiedensten mathematischen Techniken und Theoremen. Diese werden im Folgenden erläutert und sollen befähigen, den späteren Aussagen und Annahmen folgen zu können. Zudem soll es eine Anleitung zur Durchführung einer solchen Hauptkomponentenanalyse bieten.

3.1.1. Begriffsbestimmung

Um die Vorgehensweise und Problematiken nachvollziehen zu können, werden zunächst einige mathematische Begrifflichkeiten erläutert.

Eigenwerte und Eigenvektoren

Viele Verfahren auf Matrizen beruhen auf der Lösung des sogenannten *Eigenwertproblems*. Sei also X eine (n,n) -Matrix, so lassen sich die Zahlen λ und die zugehörigen Vektoren v wie folgt ermitteln:

$$Xv = \lambda v \quad (3.1)$$

Dies lässt sich durch Hinzunahme der Einheitsmatrix E zu

$$(X - \lambda E_n)v = 0 \quad (3.2)$$

umstellen.

Alle Vektoren v und deren zugehörige Zahlen λ , die dieser Gleichung genügen, werden Eigenvektoren bzw. Eigenwerte genannt. Da die Lösung $v = 0$ trivial erfüllbar ist, wird die Nebenbedingung $v \neq 0$ eingeführt. Ein Eigenvektor ergibt nach Multiplikation mit der zugehörigen Matrix ein Vielfaches seiner selbst. Dieses Vielfache ist der Eigenwert.

Kovarianz

Die Kovarianz $cov(r, p)$ ist in der Stochastik eine beschreibende Größe für den Zusammenhang zweier Variablen r und p . In unserem speziellen Fall handelt es sich um *Vektoren* der Länge n . Diese sollten jedoch vorher zentriert werden (r^z, p^z), um eine falsche Gewichtung der einzelnen Komponenten auszuschließen. Dies geschieht durch Subtraktion des Mittelvektors. Die Kovarianz kann wie folgt berechnet werden:

$$cov(r, p) = \frac{r^{zT} p^z}{n - 1} \quad (3.3)$$

Dabei zeigt eine positive Kovarianz eine tendenziell gleichartige Entwicklung der Komponenten an. Das heißt, steigt r , so steigt auch p . Eine negative Kovarianz deutet auf eine entgegengesetzte Entwicklung (steigt r , so fällt p und umgekehrt). Ist die Kovarianz Null deutet dies auf das Fehlen eines linearen Zusammenhangs der beiden Vektoren.

Varianz

Die Varianz $var(r)$ ist in der linearen Algebra auch als Streuung bekannt. Sie kann als die Kovarianz eines Vektors r mit sich selbst beschrieben werden. Ganz allgemein beschreibt sie die Abweichung einer Variablen von ihrem Erwartungswert und ist als das Quadrat der Standardabweichung σ , also dem Abstand zum Mittelwert definiert.

$$\sigma(r) = \sqrt{var(r)} = \frac{\|r^z\|}{\sqrt{n-1}} \quad (3.4)$$

$$var(r) = s(r)^2 = cov(r, r) \quad (3.5)$$

Korrelation

Die Korrelation $corr(r, p)$ beschreibt grundsätzlich den selben Sachverhalt wie die Kovarianz. Die Kovarianz hat jedoch den Nachteil, dass sie nicht skaleninvariant ist. Wenn die Werte der Vektoren multipliziert werden (zum Beispiel beim Übergang von Maßzahlen durch eine Multiplikation mit 100), so ändert sich auch die Kovarianz in gleichem Maße. Dem kann durch eine *Standardisierung* (r^s) der Vektoren r entgegengewirkt werden, was zu einem gleich bleibenden Ergebnis führt. Die Korrelation zu berechnen ist entsprechend aufwendiger:

$$r^s = \frac{r^z}{s(r)} \quad (3.6)$$

$$corr(r, p) = \frac{r^{sT} p^s}{n-1} = \frac{cov(r, p)}{s(r)s(p)} \quad (3.7)$$

Hauptkomponenten

In einem dreidimensionalen Raum können alle Vektoren durch drei Basisvektoren vollständig beschrieben werden. Diese werden als Hauptkomponenten bezeichnet. Sie bilden optimale Linearkombinationen, also eine gewichtete Summe der ursprünglichen Variablen [Wil05]. Dabei werden die Hauptkomponenten nicht willkürlich gewählt, sondern folgen zwei Grundprinzipien:

1. Eine Hauptkomponente muss immer orthogonal auf allen bereits gewählten Hauptkomponenten stehen. Dieser Punkt entfällt bei der Wahl der ersten Hauptkomponente.
2. Die verbleibende *Varianz* soll durch die neue Komponente möglichst maximal abgedeckt werden.

3.1.2. Ermittlung der Hauptkomponenten

Die grundlegende Vorgehensweise zur Realisierung einer PCA soll im Folgenden beschrieben werden. Eine PCA realisiert eine lineare Abbildung, bei der die p -dimensionalen Merkmalsvektoren auf eine r -dimensionale Hyperebene projiziert werden. Diese Hyperebene kann auch als Untervektorraum bezeichnet werden, wobei immer $r \leq p$ gilt. Sie wird so gewählt, dass sie einen größtmöglichen Teil der Gesamtvarianz abdeckt. [Bök05]

Zunächst sollten die Rohdaten für die Verwendung vorbereitet werden, d.h. sie sollten in eine Vektorform überführt werden. Ziel dieser Umformung ist es, dass jeder der n Messdatensätze als Vektor x der gleichen Dimension m vorliegt und die Werte an den entsprechenden Stellen von x_i auch gleichbedeutend und untereinander vergleichbar sind. Aus diesen Spaltenvektoren wird der Ausgangsdatensatz $X = \{x_1, x_2, \dots, x_n\}$ erstellt.

Oftmals ist es von Vorteil eine *Datenvorbehandlung* durchzuführen. Hierzu sollten die Originaldaten um ihren Erwartungswert, eine Art Durchschnittsobjekt, korrigiert werden. Eine Mittelwertkorrektur lässt sich einfach durch die Zentrierung jedes Vektors bewerkstelligen. Dabei wird zunächst der Erwartungsvektor μ wie folgt ermittelt:

$$\mu = \frac{1}{n} \sum x_i \quad (3.8)$$

Dieser wird dann von jedem Vektor x_i subtrahiert.

Nun kann die Kovarianzmatrix K berechnet werden:

$$K = \frac{1}{m-1} X X^T \quad (3.9)$$

Man kann die Originaldaten zuvor noch nach Formel (3.6) standardisieren. Damit erhält man aus der Gleichung (3.9) die Korrelationsmatrix. Standardisierung bedeutet, dass die Varianz aller Variablen x gleich 1 ist. Sowohl bei der Kovarianz- als auch bei der Korrelationsmatrix sind die Diagonalelemente gleich der Varianz. Die Eigenwerte und Eigenvektoren der beiden Matrizen sind jedoch *nicht* identisch.

Durch eine Singulärwertzerlegung (englisch *singular value decomposition* - SVD) wird die Kovarianzmatrix bzw. Korrelationsmatrix zerlegt in:

$$K = U \Sigma U^T \quad (3.10)$$

(siehe [LEGF00]). Dabei ist U eine quadratische Matrix, deren Spalten die *Eigenvektoren* von K darstellen. Σ ist eine Diagonalmatrix mit den zu U gehörigen *Eigenwerten* als Diagonalelemente. Dabei sind die Elemente nach der Größe ihrer zugehörigen Eigenwerte in absteigender Reihenfolge geordnet. Eigenvektoren und Eigenwerte können nach Formel (3.1) berechnet werden. Diese Eigenvektoren werden wir im Folgenden als neue Basisvektoren für den Untervektorraum nutzen - es sind die *Hauptkomponenten*.

3.1.3. Approximation und Rücktransformation

Eine zu x gleichartige Variable kann nun durch k Hauptkomponenten, also die ersten k Spalten von U , konstruiert werden. Dabei ist die Wahl der Dimension des Untervektorraums entscheidend. Jede Hauptkomponente deckt einen gewissen Prozentsatz an der Gesamtvarianz ab. Dieser kann mit Hilfe des zugehörigen Eigenwerts bestimmt werden:

$$P(\lambda_i) = 100 \frac{\lambda_i}{\sum_{j=1}^n \lambda_j} \quad (3.11)$$

Damit kann auch der Anteil der k ersten Hauptkomponenten an der Gesamtvarianz errechnet werden und somit als Entscheidungshilfe über die Anzahl genutzter Eigenvektoren dienen:

$$P(\lambda_1, \dots, \lambda_k) = 100 \frac{\sum_{i=1}^k \lambda_i}{\sum_{j=1}^n \lambda_j} \quad (3.12)$$

Aus diesen k Hauptkomponenten kann nun durch Linearkombination jede mögliche neue Variable aus dem aufgespannten Untervektorraum beschrieben werden. Einzig die Multiplikatoren für die einzelnen Basisvektoren müssen noch gespeichert werden. Diese können wieder als k -dimensionaler Vektor α aufgefasst werden. So kann auch jede Variable x oder ein gleichartiger Vektor in einen solchen Formbeschreibungsvektor α überführt werden. [LEGF00]

$$\alpha = U_k^T (x - \mu) \quad (3.13)$$

Im Gegenzug kann aus einem gegebenen α eine repräsentative Variable \tilde{x} errechnet werden:

$$\tilde{x} = U_k \alpha + \mu \quad (3.14)$$

3.2. Hauptkomponentenanalyse für hochdimensionale Daten

Oftmals kommt es vor, dass die Dimension der gesammelten Daten größer als die Anzahl an Datensätzen selbst ist. So möchte man zum Beispiel einen Datensatz von Bildern untersuchen, welche jeweils mehrere Tausend bis Millionen Bildpunkte beinhalten, die ihrerseits aus drei Farbwerten bestehen. Hieraus entsteht schnell ein hochdimensionaler Variablenraum, dessen Analyse nach dem Standardverfahren mit hohem Rechenaufwand und Laufzeitkosten in der Umsetzung verbunden ist. Gerade die Berechnung der Kovarianz- bzw. Korrelationsmatrix wird schnell sehr aufwendig. Hier setzt der folgende Optimierungsansatz an (vgl. [Bis06]).

Die Grundidee des hier vorgestellten Verfahrens beruht auf der Tatsache, dass n Datensätze eines m -dimensionalen Raums höchstens einen Untervektorraum der Dimension $n - 1$ aufspannen können, wenn $n < m$ ist! Führen wir eine Standard PCA durch, so erkennen wir, dass $m - (n + 1)$ Eigenwerte gleich Null sind. Diese gehören zu Eigenvektoren, die keinerlei Varianz repräsentieren.

Ausgangspunkt ist der Datensatz X , dessen Spaltenvektoren die einzelnen Variablen darstellen.

Wir nutzen die oben beschriebene Eigenschaft und berechnen die Matrix \tilde{K} wie folgt:

$$\tilde{K} = \frac{1}{m-1} X^T X \quad (3.15)$$

Hierdurch erhalten wir eine Matrix, welche uns nur die relevanten Varianzen für die Kovarianz- bzw. Korrelationsmatrix liefert. Auch hier haben wir das Eigenwertproblem zu lösen, jedoch auf einer (oftmals stark) reduzierten Dimension. Durch die Lösung dieses Eigenwertproblems erhalten wir exakt die selben *Eigenwerte* wie bei der Standard PCA für die original Matrizen. Die *Eigenvektoren* w_i dagegen, welche zur Rücktransformation benötigt werden, liegen in einem Untervektorraum und müssen in den normalen Raum zurück projiziert werden:

$$\tilde{v}_i = X w_i \quad (3.16)$$

Dabei ist zu beachten, dass die neuen Eigenvektoren \tilde{v}_i normalisiert werden müssen. Auch dies kann effizient durch die Multiplikation mit einer Konstanten u_i geschehen, wobei $\|u_i\| = 1$ gilt. Diese kann wie folgt ermittelt werden:

$$u_i = \frac{1}{\sqrt{n\lambda_i}} X w_i \quad (3.17)$$

Eine ausführliche Herleitung dieses Verfahren findet sich in [Bis06].

3.3. Schwachstellen der PCA

Wie jede Methode hat auch die PCA ihre Grenzen und Nachteile. Einige werden im Folgenden angesprochen. [Wec07]

Ein Nachteil ist, dass sie eher für eine Rekonstruktion und Komprimierung durch Dimensionsreduzierung von Daten ausgelegt und stößt bei der Unterscheidung von verschiedenen Klassen an ihre Grenzen. Zudem unterscheidet sie nicht nach der Mittelwertabweichung, sondern nach dem Kovarianzunterschied.

Der wohl größte Nachteil der PCA ist die Annahme, dass es sich bei den verwendeten Daten um gaußverteilte Daten handelt. Dies kann jedoch auch im hier betrachteten Fall nicht eindeutig bewiesen werden. Sollten die Daten nicht gaußverteilt sein, unterschätzt die PCA die Wahrscheinlichkeiten in den dichten und den spärlichen Bereichen und es entsteht eine unbeabsichtigte Verschiebung. Dies bedeutet auch, dass die PCA nicht in der Lage ist, den Einfluss von Nichtlinearität, also der Wechselbeziehung oder Interaktion der Datensätze untereinander, abzubilden. Dies könnten Nachbarschaftsabhängigkeiten oder der Einfluss von lokalen Strukturen sein, welche auch in den Ausgangsdaten für diese Arbeit vorhanden sein können, da es sich um gewachsene Strukturen handelt, die durch ihre Nachbarn beeinflusst sind. In solchen Fällen könnte man zum Beispiel die nichtlineare Methode der *Kernel-PCA* [Sch01] nutzen. Dabei werden auch hochdimensionale willkürliche Wechselbeziehungen handhabbar.

3.4. Ein Beispiel

Ein kurzes Beispiel soll die beschriebenen Techniken praktisch verdeutlichen. Dabei hält sich das Beispiel nahe an den Gegebenheiten, wie sie für diese Arbeit vorherrschen.

Wir haben einen Satz an Trainingsbildern (siehe Abb. 3.1), welche in die Vektoren \hat{x}_i umgewandelt werden.

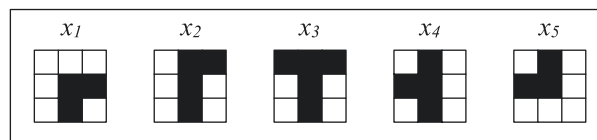


Abbildung 3.1. – Bilderdatensatz

$$\hat{x}_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \hat{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \hat{x}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \quad \hat{x}_4 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \hat{x}_5 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Diese werden durch Subtraktion des Erwartungswertes μ zentriert, welcher durch Formel (3.8) ermittelt wird.

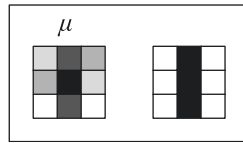


Abbildung 3.2. – Mittelwertbild μ - links: Grauwertumsetzung, rechts: Binärbild mit Entscheidungsschwelle 0,5

$$\mu^T = (0,2 \quad 0,8 \quad 0,4 \quad 0,4 \quad 1 \quad 0,2 \quad 0 \quad 0,8 \quad 0)$$

Wir erhalten den zentrierten Datensatz X :

$$X = \begin{pmatrix} -0,2 & -0,2 & 0,8 & -0,2 & -0,2 \\ -0,8 & 0,2 & 0,2 & 0,2 & 0,2 \\ -0,4 & 0,6 & 0,6 & -0,4 & -0,4 \\ -0,4 & -0,4 & -0,4 & 0,6 & 0,6 \\ 0 & 0 & 0 & 0 & 0 \\ 0,8 & -0,2 & -0,2 & -0,2 & -0,2 \\ 0 & 0 & 0 & 0 & 0 \\ 0,2 & 0,2 & 0,2 & 0,2 & -0,8 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Nun stellen wir nach Formel (3.15) die Matrix \tilde{K} auf:

$$\tilde{K} = \begin{pmatrix} 0,210 & -0,040 & -0,065 & -0,040 & -0,065 \\ -0,040 & 0,085 & 0,060 & -0,040 & -0,065 \\ -0,065 & 0,060 & 0,160 & -0,065 & -0,090 \\ -0,040 & -0,040 & -0,065 & 0,085 & 0,060 \\ -0,065 & -0,065 & -0,090 & 0,060 & 0,160 \end{pmatrix}$$

Daraus ermitteln wir die Eigenwerte λ und Eigenvektoren w (vgl. Formel (3.1)) und ordnen diese nach der Größe der Eigenwerte:

$$\lambda = (0,327 \quad 0,266 \quad 0,059 \quad 0,048 \quad \approx 0,000)$$

$$\tilde{U} = \{w_1, \dots, w_5\} = \begin{pmatrix} 0,000 & -0,886 & 0,122 & 0,000 & -0,447 \\ -0,372 & 0,154 & -0,526 & -0,602 & -0,447 \\ -0,602 & 0,289 & 0,465 & 0,372 & -0,447 \\ 0,372 & 0,154 & -0,526 & 0,602 & -0,477 \\ 0,602 & 0,289 & 0,465 & -0,372 & -0,477 \end{pmatrix}$$

Danach überführen wir die Eigenvektoren w nach Formel (3.16) in den ursprünglichen Raum und normalisieren diese danach zu:

$$U = \{v_1, \dots, v_5\} = \begin{pmatrix} -0,372 & 0,198 & 0,679 & 0,602 & 0,451 \\ 0,000 & 0,607 & -0,177 & 0,000 & -0,090 \\ -0,601 & 0,304 & -0,089 & -0,372 & 0,180 \\ 0,602 & 0,304 & -0,089 & 0,372 & -0,180 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 \\ 0,000 & -0,601 & 0,177 & 0,000 & 0,451 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 \\ -0,372 & -0,198 & -0,679 & 0,602 & -0,721 \\ 0,000 & 0,000 & 0,000 & 0,000 & 0,000 \end{pmatrix}$$

Wir können sehen, dass schon die ersten beiden Eigenvektoren rund 85% der Gesamtvarianz abdecken. Nimmt man noch den dritten hinzu sind es bereits 93% (vgl. Formel (3.12)). Den fünften Eigenvektor könnten wir hingegen vernachlässigen, da er so gut wie keinen Einfluss hat. Wir könnten nun alle Beispielbilder (Abb. 3.1) und alle im aufgespannten Untervektorraum möglichen Variablen durch 2- bzw. 3-dimensionale Vektoren approximieren (vgl. Formel (3.13)).

3.5. Zusammenfassung

Die Hauptkomponentenanalyse ist ein vielseitig einsetzbares Werkzeug, das die wichtigsten Merkmale aus einer großen Menge von Daten extrahieren kann. Dabei gibt es einige Ansätze, das Verfahren in seiner Komplexität zu vereinfachen. Das grundsätzliche Vorgehen bleibt dessen ungeachtet immer gleich:

1. Daten erstellen
2. Eventuell um Mittelwert korrigieren und standardisieren
3. Kovarianzmatrix aufstellen
4. Eigenwerte und Eigenvektoren ermitteln und diese gewichtet ordnen
5. Daten transformieren

4. Segmentierung

Ziel dieser Arbeit ist es, einen Ansatz zur Segmentierung von einzelnen Muskelfasern im Bild eines Zellpreparates zu entwickeln. Dabei wird über zwei Parameter entschieden wie eine Zelle segmentiert wird. Dies ist zum einen die Farbe und zum anderen die Zellform.

Zur Unterscheidung über der Farbe wird die Aufgabenstellung dahingehend modifiziert, dass nur gefärbte Zellen segmentiert werden sollen. Dies ist dem Umstand geschuldet, dass weiße, ungefärbte Zellen sich nur schwer durch ihre Färbung vom Zwischengewebe abheben und zudem durch ihre inneren Zellbestandteile eine starke Strukturierung aufweisen. Eine Berücksichtigung dieser Zellen kann durch ein erneutes Zellpreparat mit entsprechend anderer pH-Wert-Inkubation (siehe Kapitel 2.2) gewährleistet werden. Beide Merkmale, sowohl Farbe als auch Form, sollen in einem Modell miteinander verknüpft werden.

ASSMUS [Asm08] beschreibt in seiner Arbeit einen Seeded-Region-Growing Ansatz (vgl. Kapitel 4.6) auf der Basis von Level-Set-Funktionen (vgl. Kapitel 4.2.2). Dieser Ansatz soll nun dahingehend erweitert werden, dass eine Evolution direkt auf den die Form beschreibenden Parametern vollzogen werden soll.

4.1. Segmentierungsansätze

Die Segmentierung von Bildbereichen ist ein zentrales Thema der Bildanalyse und damit auch in der Medizinischen Bildverarbeitung, da sie für die computergestützte Diagnostik oder Therapie von entscheidender Bedeutung sein kann [Han00]. Sie dient der weitergehenden Analyse oder Vermessung in zwei- oder dreidimensionalen Bilddaten. Es sollen diagnostisch oder therapeutisch relevante Strukturen vom umgebenden Gewebe oder dem Hintergrund abgegrenzt werden. Dabei besteht die Schwierigkeit darin, dass die erzeugten Bilddaten oft von schwankender Qualität sind und die verwendeten Verfahren gegen Bildstörungen oder Bildartefakte unempfindlich sein müssen.

Segmentierungsmethoden können grundsätzlich durch ihre Methodik in *kanten-* und *regionsorientierte* sowie *clusteranalytische* und *klassifikatorbasierte* Ansätze unterschieden werden [Han00]. In der weiteren Betrachtung werden die für diese Arbeit relevanten ersten beiden Ansätze näher beschrieben.

Kantenorientierte Segmentierungsverfahren richten sich vor allem nach starken lokalen Veränderungen in den Bilddaten und detektieren diese. Dies kann jede Art von Bildparameter sein, ob nun Grauwerte oder andere informationstragende Datenelemente wie

Fluss- oder Dichteparameter¹. Eine einfache Methode, die Kanten eines Bildes zu detektieren, ist ein Hochpassfilter. Dieser eliminiert alle niedrigen Frequenzen eines Bildes, sodass nur hohe Frequenzen, also starke Änderungen in der Bildstruktur, erhalten bleiben (siehe Abb. 4.1).

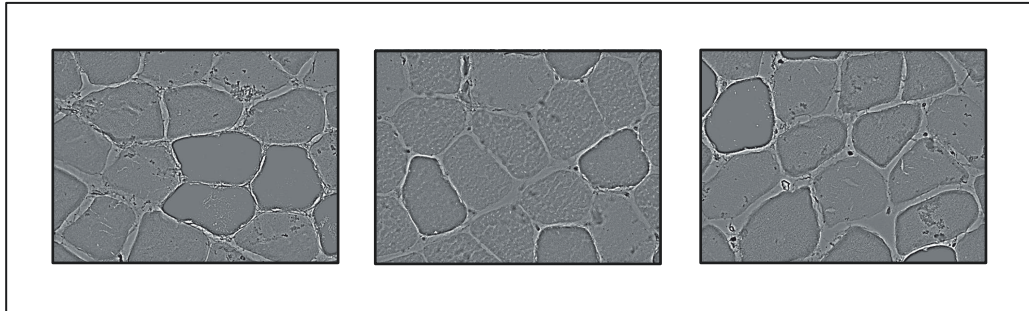


Abbildung 4.1. – Hochpass gefilterte Beispielbilder

Die Methode scheint für diese Arbeit nicht voll geeignet, da sie anfällig gegen Bildstörungen ist. Diese sind oftmals durch starke Änderungen der Bildstruktur und damit durch eine hohe Frequenz gekennzeichnet. Zudem stellt sich die Frage nach der richtigen Wahl der Filterfrequenz und wie die einzelnen Kanten zur Segmentierung genutzt werden können, um zusammenhängende Formen zu trennen.

Regionsorientierte Verfahren extrahieren hingegen Bereiche, welche durch ein Homogenitätskriterium gekennzeichnet sind. Eine der einfachsten Segmentierungsmethoden ist die Binärbildzerlegung, welche auch als Schwellwertverfahren (engl. *thresholding*) bekannt ist [AN07]. Hierbei wird als Homogenitätskriterium ein Farbbereich gewählt, der oftmals global nicht unter- oder überschritten werden darf (siehe Abb. 4.2). Die Festlegung dieses Schwellwertes ist in den meisten Fällen durch den Nutzer selbst notwendig [Asm08, Han00].

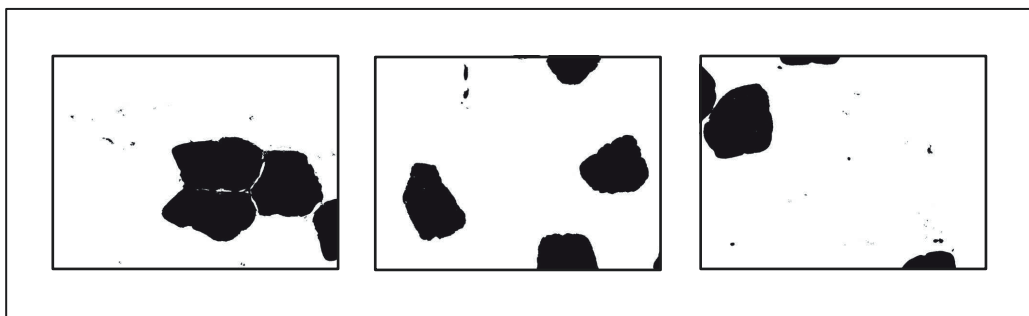


Abbildung 4.2. – Schwellwert gefilterte Beispielbilder (dunkle Zellen)
- Schwellwert 55

¹Diese können in bildgebenden Verfahren wie z.B. der Computertomographie (CT) oder Magnetresonanztomographie (MRT) vorkommen

Da dem Schwellwertverfahren ein globaler Ansatz zu Grunde liegt, wird dieser durch die Verstärkung des Regionsbezuges zum *Bereichs- oder Volumenwachstumsverfahren* erweitert. Das Bereichswachstumsverfahren oder Regionswachstumsverfahren (engl. *region growing*) extrahiert Gebiete von zusammenhängenden Bildsegmenten, welche in ihrem analysierten Entscheidungsparameter homogen sind. Der Basisalgorithmus für diese Art der Segmentierung startet mit einem manuell oder automatisiert gesetzten Startpunkt, welcher auch als Saatpunkt bezeichnet wird (engl. *seed point*). Davon ausgehend wird die direkte Nachbarschaft auf das Homogenitätskriterium geprüft. Alle Punkte, die dieses erfüllen, werden der Region hinzugefügt. Weiterhin wird deren direkte Nachbarschaft überprüft und unter Umständen ebenfalls hinzugenommen. Dies wiederholt sich, bis kein Nachbar mehr die geforderte Eigenschaft erfüllt.

Dieser regionsbasierte Algorithmus allein kann die gegebene Problemstellung nicht lösen. Das größte Problem dieser Ansätze ist das Phänomen des *Überlaufens*. Da das Verfahren einer simplen Markierung von zusammenhängenden Bereichen entspricht, kommt es an Stellen, an denen Bereiche mit sehr ähnlichem Entscheidungskriterium dicht aneinander liegen oder direkt aneinander stoßen, zu Konflikten, falls diese nicht zur gleichen Region gehören sollen. Eben diese Problematik macht das Verfahren für den hier betrachteten Fall nur bedingt geeignet (siehe Abb. 4.3).

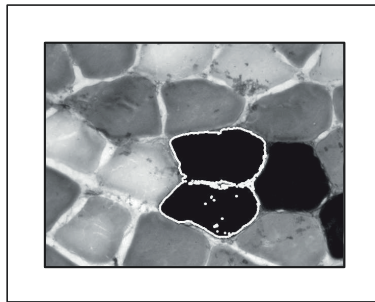


Abbildung 4.3. – Regionswachstumsüberlauf bei Farbdistanz 20 -
weiß: Kontur einer zusammenhängenden Region

Somit muss der Regionswachstumsansatz dahingehend angepasst werden, dass nicht allein der Intensitätswert als Homogenitätskriterium die Regionen definiert. Statistisch gesehen ist jede einigermaßen runde Form als Segmentierungsergebnis zulässig. Dies zeigt auch die Mittelwertform (siehe Abb. 4.4). Man könnte einen Glattheitsterm oder die Minimierung der Konturlänge verwenden, um das Regionenwachstum zu begrenzen. Diese Methoden beschreiben jedoch keinerlei Formcharakteristika, sondern sind nur darauf ausgelegt, eine glatte Konturführung zu gewährleisten. Damit würden eventuell kantige oder mit Auswüchsen versehene Zellen nicht vollständig segmentiert werden. Hier kann statistisch ausgewertetes Vorwissen über die Formausprägungen Abhilfe schaffen - im Folgenden auch als *shape prior* bezeichnet.

4.2. Formdaten und statistische Auswertung

Um das Regionenwachstum auf die zu segmentierenden Bereiche zu begrenzen, sind zusätzliche Informationen über die Regionen notwendig. Das Ausnutzen aller im Bild vorhandenen Daten scheint hier nicht auszureichen. Es müssen zusätzliche Informationen, sogenanntes *a priori*² Wissen, mit eingebracht werden, zum Beispiel Aussagen über die Gestalt der zu segmentierenden Regionen.

4.2.1. Beispieldaten

Um die Segmentierung einzelner Zellen auch in schwierigen Entscheidungsbereichen möglichst genau zu realisieren, werden Merkmale über die Form von Zellen als Entscheidungshilfe benötigt. Es gibt verschiedenste Möglichkeiten dies zu bewerkstelligen. Eine dieser Möglichkeiten wird in [Asm08] vorgestellt. Wie viele andere basiert sie auf einer *Datenbank* von Beispieldaten. Diese sogenannten Trainingssets werden zur Laufzeit dazu benutzt, einen Abgleich der bis dahin segmentierten Form mit den Beispielen zu treffen.

Aus diesem Sachverhalt folgt auch die Tatsache, dass die Qualität des Trainingssets sich unmittelbar auf das Segmentierungsergebnis auswirkt. Dabei ist die Qualität der Trainingsdaten nicht objektiv messbar. Auch steht eine große Anzahl an Trainingsdaten nicht immer in direktem Zusammenhang mit steigender Qualität, jedoch werden häufig auftretende Merkmale dadurch stärker gewichtet. Dennoch ist die Auswahl der einzelnen Beispiele von größter Bedeutung. Die Entscheidung nach der Anzahl und der Art der Beispiele hängt hauptsächlich von den Formen selbst ab. Hat man Objekte, welche sich durch eine große Vielfalt und Varianz auszeichnen, ist ein größerer Datensatz sicherlich sinnvoll. Dieser sollte zudem die verschiedensten Erscheinungsarten einschließen und abbilden können. Ein weiteres Kriterium bei der Auswahl sollte die Leistungsfähigkeit des Segmentierungsalgorithmus selbst darstellen. So sollten Variationen wie Rotation, Spiegelung, Verschränkung oder Verzerrung entweder durch den Segmentierungsprozess oder durch den Datensatz abgedeckt werden.

Im hier vorliegenden Fall sind die zu beschreibenden Daten sehr stark in ihrer Variation. Allerdings sind sie fast symmetrisch, in manchen Ausprägungen jedoch kaum rotationsinvariant. Eine Verschränkung oder Verzerrung der einzelnen Formen führt zu neuen Formen, da keine harte Formcharakteristika eingehalten wird.

Für die Praktische Umsetzung wurde ein Beispieldatensatz von 200 verschiedenen Zellen im Format 100 mal 100 Pixeln als Basis gewählt. Um die vernachlässigte Behandlung der Rotation im Segmentierungsprozess etwas zu kompensieren (siehe Kapitel 4.4), wurden die Trainingsdaten jeweils um 90° gedreht und als weitere Beispiele dem Datensatz hinzugefügt. So ergibt sich ein Trainingsset von 400 vorsegmentierten Bildern, welches ein möglichst gutes Abbild der Zellformen verkörpern soll (siehe Abb. A.1). Vorstellbar wäre

² von vornherein oder vorläufig

zudem eine Spiegelung der Beispieldaten, um das Trainingsset weiter zu vervollständigen. Betrachtungen hierzu finden sich in Anhang A.4.

4.2.2. Formrepräsentation

Hat man eine Auswahl von Beispielformen getroffen, so stellt sich die Frage nach der richtigen Repräsentation der Formen. Es gibt viele verschiedene Arten, Formen zu beschreiben. Dazu zählt die Beschreibung über spezielle Kennzahlen, wie zum Beispiel den Flächeninhalt und Ähnliches. Eine der geeignetsten Merkmale, eine Form zu beschreiben, ist deren *Kontur*. Eine solche Darstellung ist durch unterschiedlichste Verfahren zu bewerkstelligen. Die einfachste Möglichkeit ist die Speicherung der Kontur durch ihre *diskreten* Werte, also eine Abfolge ihrer Standortparameter im 2- oder 3-dimensionalen Gitter.

Eine andere Möglichkeit zur Konturenbeschreibung bieten die *Level-Set-Methoden*. Diese bilden die Form durch eine *Einbettungsfunktion* ϕ (engl. *embedding-function*) in den kontinuierlichen Funktionsraum ab. Dabei ist die Kontur C selbst eine Gruppe von Punkten x mit dem Wert θ in der *Bilddomäne*³ Ω .

$$C_\theta := \{x \in \Omega | I(x) = \theta\} \quad (4.1)$$

Die Kontur bildet somit eine Ebene durch die Funktion. Ein gängiges Beispiel ist die Nutzung des Nullliniendurchgangs als Konturrepräsentation. Dieser wird auch in den Distanzfunktionen (z.B. der klassischen *signed-distance-function*) [Bis06] genutzt, welche einen Wert der Funktion auf die räumliche oder eine entsprechend andere Distanz zur Kontur abbildet. Eine weitere sehr einfache Funktion zur Level-Set-Repräsentation ist die *Einheitssprungfunktion* $H(x)$ (engl. *heaviside-step-function*). Diese nimmt nur zwei bzw. drei Werte an: die Werte a für außerhalb und b für innerhalb der Kontur, sowie den Wert θ für die Kontur selbst.

$$H(x) = \begin{cases} a, & I(x) < \theta \\ b, & I(x) > \theta \end{cases} \quad (4.2)$$

In der konkreten Umsetzung wurde ebenfalls eine Darstellung als Heaviside Level-Set-Funktion implementiert. Dies liegt vor allem an ihrer Einfachheit. Zudem ist die Nutzung einer Distanzfunktion für den Einsatz als Variablendarstellung in der Hauptkomponentenanalyse eher ungeeignet, da die approximierten Variablen selbst keine Distanzfunktion mehr darstellen [RH95]. Die Kontur wird als $\theta = 0,5$ Level festgelegt, die Formfüllung mit $b = 1$ und der Hintergrund mit $a = 0$ definiert. Die hier getroffene Wahl soll numerische Probleme bei der späteren Implementierung minimieren. Aus der gewählten

³Die Bilddomäne ist das Raster des Bildes

Darstellungsform ergeben sich für unsere Beispieldaten einfache Binärbilder, deren Körper mit `high` und deren Hintergrund mit `low` kodiert wurden.

Bei der Formenrepräsentation sollte ein weiterer Punkt angesprochen werden. Da die Beispieldaten gespeichert werden sollten, um eine spätere Erweiterung des Datensatzes zu ermöglichen, ist es sinnvoll diese in ihrem Speicherbedarf zu komprimieren. Hier hat sich die *Laufängenkodierung* als einfache und sehr effektive Lösung erwiesen [Han00], da die einzelnen Datensätze nur aus zwei verschiedenen Zahlen bestehen. Diese Art der Komprimierung beschreibt eine Zeichenkette durch die Zerlegung in Abschnitte gleicher Zeichen. Die Abschnitte werden dann durch das mehrfach auftretende Zeichen und deren folgende Anzahl kodiert⁴. Die Möglichkeit ist bei steigender Dimension der Trainingsdaten von Vorteil, bei sehr kleinen Datensätzen oder solchen mit hoher Änderungsrate ist hingegen eine solche Kodierung eher von Nachteil⁵.

In der vorgenommen Implementierung wurde diese Möglichkeit der Speicheroptimierung für die Beispieldaten genutzt und damit der benötigte Speicherbedarf um mehr als 90% reduziert⁶. Für die Speicherung der Eigenvektoren ist diese Möglichkeit nicht empfehlenswert, da sie aus vielen verschiedenen, stark schwankenden Werten bestehen.

4.3. Umsetzung der Hauptkomponentenanalyse

In Kapitel 3 wurde das Werkzeug der Hauptkomponentenanalyse vorgestellt. Es ermöglicht die effiziente Erzeugung eines endlich großen „Formenraums“, in dem wir agieren können. Zudem sind viele mathematische und statistische Probleme gelöst und müssen nur noch auf den vorliegenden Sachverhalt angepasst oder erweitert werden.

Zur Durchführung werden die Beispielbilder in Level-Set-Repräsentation zeilenweise als Vektoren der Dimension 10.000 aufgetragen, der Erwartungswert (siehe Abb. 4.4) nach Formel (3.8) errechnet und damit der Datensatz zentriert. Eine Standardisierung erfolgt nicht.

Da die Dimension m der Vektoren mit 10.000 deutlich größer als die Anzahl n ⁷ ist, bietet sich die Vorgehensweise für hochdimensionale Datensätze aus Kapitel 3.2 an, um die *Eigenwerte* und *Eigenvektoren* zu ermitteln.

Das Ergebnis dieser Eigenwertzerlegung kann die Güte der ausgewählten Beispieldaten wiedergeben. Der für die Umsetzung herangezogene Beispieldatensatz sollte ein großes Spektrum an Formausprägungen und damit eine große Varianz abbilden, was durch die

⁴Beispiel: Aus (11111000110000) wird durch Laufängenkodierung (51 30 21 40)

⁵siehe Beispiel \hat{x}_4 in Kapitel 3.4 - aus (010110010) wird (10 11 10 21 20 11 10)

⁶normal gespeichert als `double` - kodiert als `String`

400 normal: 31.254 KB kodiert: 474 KB

800 normal: 62.508 KB kodiert: 947 KB

1200 normal: 93.762 KB kodiert: 1.421 KB

⁷In den Experimenten 400, 800 oder 1200 Datensätze

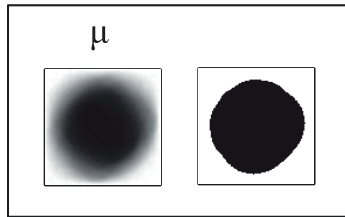


Abbildung 4.4. – Mittelwertbild des Beispieldatensatzes von 400 Bildern - links: Grauwertumsetzung, rechts: Binärbild mit Entscheidungsschwelle 0,5

400 Beispieldaten gut gelungen scheint (siehe Abb. 4.5). Da fast alle Beispieldatensätze zusätzliche Informationen tragen, ergeben sich nur zwei Eigenwerte, die so nah an Null liegen, dass sie vernachlässigbar sind. Dennoch ist ein hoher Grad an Korrelation vorhanden. Dies liegt zum einen an der recht großen Fläche, die alle Beispielformen gemeinsam haben, zum anderen an der stark vergleichbaren Formausrichtung. Wie in Abbildung 4.5 zu sehen ist, beschreiben die ersten sechs Eigenwerte bereits mehr als 50% der Varianz, die ersten 50 schon mehr als 82% und nach 105 Eigenvektoren ist die 90% Grenze überschritten (Berechnung nach Formel (3.12)).

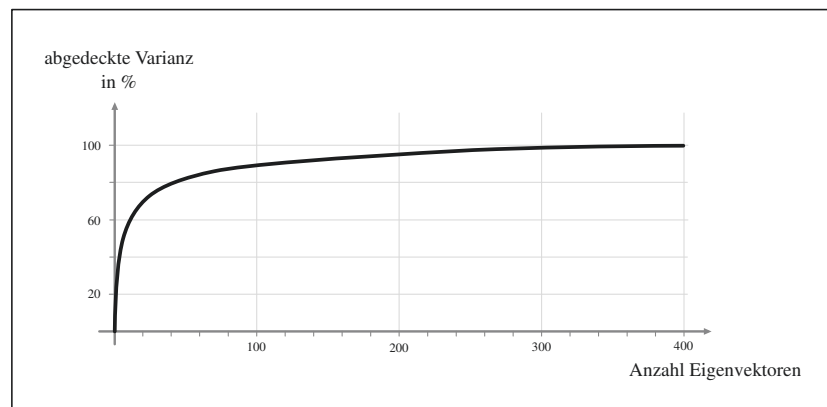


Abbildung 4.5. – Prozentuale Abdeckung der Gesamtvarianz (400 Beispieldaten)

Aus dieser Tatsache lassen sich die fünf hauptsächlichen Formausprägungen identifizieren (siehe Abb. 4.6). Man kann sehen, dass die ersten drei Hauptkomponenten mehr eine Ausrichtung der Form in die drei häufigsten Richtungen modellieren, die beiden darauffolgenden zeigen eine Deformation und erste Detailgestaltung. Dennoch sind auch weniger hoch bewertete Eigenvektoren nicht unbedingt zu vernachlässigen. Es ist ein Kompromiss zwischen Laufzeiteffizienz und Qualität zu finden. Da die Laufzeit in direktem Zusammenhang mit der Anzahl der verwendeten Eigenvektoren steht, ist hier eine geringe Anzahl von Vorteil. Des Weiteren führt eine kleine Anzahl an Eigenvektoren zu einem Regularisierungseffekt. Hierdurch werden besonders häufig auftretende Formcharakteristika hervorgehoben oder der Formenraum gar auf diese beschränkt. Dem entgegen

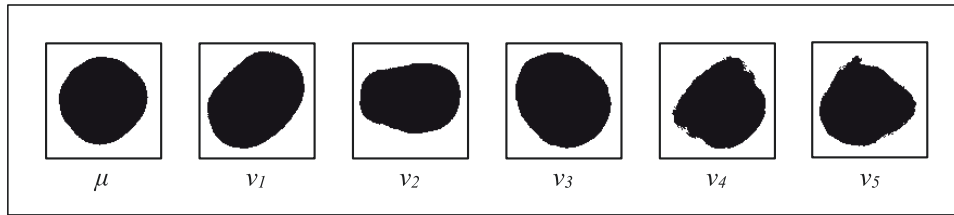


Abbildung 4.6. – Die fünf wichtigsten Formausprägungen für 400 Beispieldaten - errechnet aus Mittelwertform (μ) plus entsprechendem Eigenvektor (v_1 bis v_5)

steht, dass extreme Formausprägungen, welche nur sehr selten vorkommen, durch Eigenvektoren mit niedrigem Eigenwert beschrieben sein können. Diese zu ignorieren kann zu schlechten Segmentierungsergebnissen führen, da solche Formen im Variablenraum nicht beschreibbar sind.

Durch die PCA können wir nun jede zulässige Variable als Formvektor α darstellen oder jede Form durch einen solchen Vektor approximieren, der nach Formel (3.13) erstellt werden kann. Dadurch erhalten wir auch die Möglichkeit festzustellen, ob ein Objekt in unserem Formenraum darstellbar ist. Es kann sogar bestimmt werden, wie wahrscheinlich ein solches Objekt in Bezug auf den aufgespannten Untervektorraum ist.

Die Wahrscheinlichkeit für einen gegebenen Formvektor kann bei der Annahme eines im weitestgehenden Sinne gaußverteilten Trainingsdatensatzes mit Formenvarianz Σ_k nach [LEGF00] wie folgt bestimmt werden:

$$P(\alpha) = \frac{1}{\sqrt{(2\pi)^k |\Sigma_k|}} \exp\left(-\frac{1}{2}\alpha^T \Sigma_k^{-1} \alpha\right) \quad (4.3)$$

4.4. Energieminimierung

Als grundlegendes Modell zur Entscheidung über eine passende Segmentierung gilt die *Maximum-Likelihood Methode*. Dabei handelt es sich um ein statistisches Modell zur Schätzung unbekannter Parameter. Die Grundidee ist relativ simpel, indem man sich fragt, welche Parameterwerte das Zustandekommen eines Funktionswertes oder einer Stichprobe am wahrscheinlichsten macht ohne dabei alle Möglichkeiten auszuprobieren. Methoden, die dieses Modell umsetzen, entscheiden somit über *Wahrscheinlichkeiten*.

Hieraus lässt sich eine weitere gängige Methode in der Bildverarbeitung ableiten - die Nutzung eines pseudo-physikalischen Modells. Die Aufstellung von sogenannten Energiefunktionalen soll die Konturenfindung als *Optimierungsproblem* lösbar machen. Eine *Energie* beschreibt dabei eine Art Maßzahl, welche die Güte eines Objektes widerspiegeln soll [Bis06, CR07]. Sie kann auch als Kostenfaktor bezeichnet werden. Dabei werden

verschiedene Faktoren entweder bestraft, also mit Kosten belegt, oder entsprechend energiereicher modelliert. Je nach Modellierung wird das Energiefunktional maximiert oder minimiert.

Im Folgenden soll ein derartiger Optimierungsansatz entwickelt werden. ROUSSON und CREMERS [CR07] verwenden ein solches Energiefunktional zur Level-Set-Segmentierung. Dieses Modell soll als Grundlage dieser Arbeit dienen, weiterentwickelt und angepasst werden. Dabei stellen ROUSSON und CREMERS ein Modell auf, das auf mehreren Parametern beruht. Dies sind die Form selbst, ihre Position im Bild und ihre Rotation. Bei diesem Ansatz muss man allerdings davon ausgehen, dass ein Vergleich der zu segmentierenden Form und dem Wahrscheinlichkeitsschätzer für die Formen größeninvariant ist. Dies ist im Kontext dieser Arbeit jedoch nicht gegeben, da eine feste Größe für die Beispieldaten gewählt wurde, welche in keinem einheitlichen Zusammenhang mit der Größe der im Bild vorkommenden Zellen steht. Aus diesem Grund muss ein Skalierungsfaktor als weiterer Parameter eingeführt werden. Des Weiteren wird aus Gründen der Komplexität auf den Rotationsparameter verzichtet. Dieser soll durch einen entsprechend gewählten Beispieldatensatz abgedeckt werden (siehe Kapitel 4.2.1).

Zur genauen Darstellung einer Form im Bild müssen die Trainingsdaten unter Berücksichtigung von Verschiebung h und Skalierung s zu einer Parameterdarstellung für die *embedding-function* zur Level-Set-Repräsentation umgestellt werden. Dies geschieht durch das Einsetzen der *embedding-function* in die Gleichung (3.14) und eine entsprechende Umformung zu:

$$\phi_{\alpha,h,s}(x) = \phi_0(sx + h) + \sum_{i=1}^k \alpha_i v_i(sx + h) \quad (4.4)$$

Die Funktion ϕ_0 bezeichnet den Mittelvektor, v_i den i -ten Eigenvektor, die Verschiebung $h \in \mathbb{R}^2$ die Position im zweidimensionalen Bildraster und die Skalierung $s \in \mathbb{R}$ den Größenunterschied zu den Beispieldaten.

Als Faktoren, welche das Energiefunktional bilden, werden wie im *Maximum-Likelihood Ansatz* Wahrscheinlichkeiten genutzt. Dabei kann die Segmentierung durch die Hinzunahme des *shape prior* als Maximierung der *a posteriori*⁸ Verteilung berechnet werden. Dies bedeutet, das Energiefunktional und damit die Wahrscheinlichkeit für die getroffene Segmentierung bzw. gefundene Kontur soll maximal sein. Es lässt sich jede mögliche Segmentierung bewerten und die bestmögliche soll dabei gewählt werden.

$$P(\phi|I) \approx P(I|\phi)P(\phi) \quad (4.5)$$

Dabei bezeichnet der erste Term die Wahrscheinlichkeit für eine getroffene Segmentierung auf Grundlage des vorhandenen Bildes, der zweite Term beschreibt die Wahrscheinlichkeit eines Bildes mit getroffener Segmentierung. Dieser wird als *Datenterm* bezeichnet.

⁸Bezeichnet Wissen, das aus der Erfahrung entstammt oder bereits bewiesen wurde

Der letzte Term beschreibt, wie wahrscheinlich die segmentierte Form nach dem *shape prior* selbst ist. Datenterm und Formenwahrscheinlichkeit können zueinander durch eine Konstante c gewichtet werden.

$$P(\phi|I) \approx P(I|\phi)cP(\phi) \quad (4.6)$$

Die Lösung dieses Problems ist eine unendlichdimensionale Maximierungsaufgabe, da jede dieser Teilterme unendlichdimensional ist. Durch die Parameterdarstellung aus Formel (4.4) können wir das unendlichdimensionale Problem aus Formel (4.6) in ein endlichdimensionales überführen, da die einzelnen Teilterme nun durch die Parameter genauer beschrieben werden und somit nur endlich viele Möglichkeiten zulassen.

$$P(\alpha, h, s|I) \approx P(I|\alpha, h, s)cP(\alpha, h, s) \quad (4.7)$$

Da die Wahrscheinlichkeit $P(\alpha, h, s)$ einzig von α abhängig ist, schreiben wir nur $P(\alpha)$. Dies resultiert aus der Modellannahme, dass jede Verschiebung h und Skalierung s für eine solche Form gleich wahrscheinlich sind.

Aus diesen Vorbetrachtungen lässt sich nun ein finales Energiefunktional definieren, welches maximiert werden soll. Wir formen dieses Funktional zu einer Minimierungsaufgabe um. Hierzu nutzen wir die Eigenschaft des Logarithmus als monotone Funktion. Damit soll nun das Minimum des negativen Logarithmus unseres Energiefunktional gefunden werden (nach [CR07]).

$$E(\alpha, h, s) = -\log P(I|\alpha, h, s) - c \log P(\alpha) \quad (4.8)$$

Durch Hinzunahme der Heaviside Funktion als Darstellung für die Level-Set-Funktion (siehe Kapitel 4.2.2) $H_\phi = H(\phi_{\alpha, h, s}(x))$ und das Einsetzen von Gleichung (4.3) erhalten wir:

$$E(\alpha, h, s) = - \int_{\Omega} H_\phi \log p_{in}(I) + (1 - H_\phi) \log p_{out}(I) dx - c \log \left(\frac{1}{\sqrt{(2\pi)^k |\Sigma_k|}} \exp \left(-\frac{1}{2} \alpha^T \Sigma_k^{-1} \alpha \right) \right) \quad (4.9)$$

Hierzu werden nach [CR07] zwei Gaußwertwahrscheinlichkeitsfunktionen über den Bildinformationen definiert, welche angeben, wie wahrscheinlich die Zugehörigkeit eines Bildintensitätswertes $I(x)$ zum Hintergrund p_{out} oder zur Formfüllung p_{in} ist.

$$p_{out}(I(x)) = \begin{cases} g(x) * \exp\left(-\frac{1}{2}\left(\frac{I(x)-\mu_{out}}{\sigma_{out}}\right)^2\right) & I(x) < \mu_{out} \\ g(x) & I(x) > \mu_{out} \end{cases} \quad (4.10)$$

$$g(x) = \frac{1}{\frac{1}{2} * \int \exp\left(-\frac{1}{2}\left(\frac{I(x)-\mu_{out}}{\sigma_{out}}\right)^2\right) dx + (255 - \mu_{out})} \quad (4.11)$$

$$p_{in}(I(x)) = \frac{1}{\sigma_{in}\sqrt{2\pi}} * \exp\left(-\frac{1}{2}\left(\frac{I(x) - \mu_{in}}{\sigma_{in}}\right)^2\right) \quad (4.12)$$

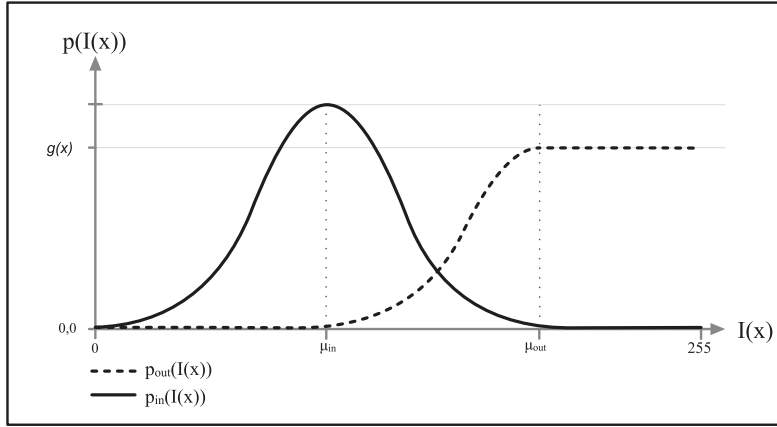


Abbildung 4.7. – Beispiel für Grauwertwahrscheinlichkeitsfunktionen nach Formel (4.10) und (4.12)

Zur Bestimmung der Wahrscheinlichkeiten wird eine Glockenkurve um den jeweiligen Erwartungswert - μ_{in} für den mittleren inneren Grauwert und μ_{out} für den mittleren äußeren Grauwert - mit einer entsprechenden Standardabweichung σ_{in} und σ_{out} für die beiden Regionen aufgespannt. Werte, die oberhalb μ_{out} liegen, sind heller als der Durchschnitt und werden deshalb zum Hintergrund gezählt. Bei der praktischen Umsetzung wurden μ_{in} durch die Mittlung der Grauwerte der bereits segmentierten Form und μ_{out} aus einem umschließenden Bereich ermittelt.

Zudem wurde der dreidimensionale Intensitätsvektor $I(x)$ in einen eindimensionalen Grauwert überführt. Dies dient der Reduzierung der Problematik sowie des Rechenaufwandes und ist durch die recht einseitige und stark ungesättigte Färbung der einzelnen Zellen unserer Ausgangsdaten möglich. In Fällen, in denen Farbe eine größere Rolle spielt, können mehrere Farbwertschätzer hintereinander geschaltet werden, einen für jeden Farbkanal.

$$p_{in|out}(I(x)) = \prod_{channel=1}^n p_{in|out_{channel}}(I(x)) \quad (4.13)$$

Die Daten für die Grauwertschätzer sind nur regional und können deshalb ungenau sein. Schließen die Daten für den mittleren äußeren Grauwert zu viele eigentlich zu einer Form gehörenden Werte ein, wird die Entscheidung über die Wahrscheinlichkeiten sehr ungenau da die Standardabweichung stark zunimmt. Zudem müssen die Farbwerte in jeder Iteration neu bestimmt werden, was wenig lauffzeiteffizient ist. Eine Kombination aus einer geschätzten globalen Komponente und den lokalen Größen könnte hier Abhilfe schaffen.

4.5. Gradientenabstieg

Die Lösung des Minimierungsproblems über dem Energiefunktional aus Formel (4.9) soll ein bestmögliches Segmentierungsergebnis liefern. Funktionen können durch die Verwendung von Ableitungsinformationen minimiert werden. Ein Minimum - ob nun ein globales oder nur ein lokales - zeichnet sich dadurch aus, dass die erste Ableitung Null ist und somit keine Steigung an diesem Punkt vorhanden ist. Hierauf basiert das *Gradientenabstiegsverfahren*, auch als Verfahren des steilsten Abstiegs bekannt [Mei05].

Der Gradient einer Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ bildet einen Vektor, dessen Komponenten durch die einzelnen partiellen Ableitungen bestimmt werden:

$$\nabla f = \text{grad} f = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right) = (f_{x_1}, \dots, f_{x_n}) \quad (4.14)$$

Dies bedeutet, dass ein Extremum oder Sattelpunkt der Funktion f als $\text{grad} f = 0$ definiert ist. Da diese Gleichung oftmals nicht trivial lösbar ist, verwendet man das iterative Verfahren des Gradientenabstiegs. Der Gradient einer Funktion hat die geometrische Eigenschaft, an einem Punkt in Richtung des steilsten Aufstiegs zu zeigen. Diese Eigenschaft kann man sich zu Nutzen machen und ihm in gewünschter Richtung folgen. Im Fall der Minimierung folgt man dem negativen Gradienten bis er in einem Minimum Null wird. Um der Funktion in eine durch den Gradienten bestimmte Richtung zu folgen, addiert oder subtrahiert man die partiellen Ableitungen zu den entsprechenden Funktionsparametern und ermittelt den neuen Funktionswert. Dort bestimmt man erneut den Gradienten und wiederholt den Vorgang bis der Gradient Null ist. Dabei kann die Größe der partiellen Ableitung durch eine Konstante skaliert werden. Dieser Parameter wird *Schrittweite* oder Lernrate ω genannt [Kin92].

$$f\left(x - \omega \frac{\partial f}{\partial x}\right) \leq f(x) \quad (4.15)$$

Dieses Verfahren soll auch zur Lösung des Minimierungsproblems aus Gleichung (4.9) genutzt werden. Hierzu müssen die partiellen Ableitungen für α , h und s bestimmt werden. Dabei ist zu beachten, dass die Heaviside Funktion H_ϕ als erste Ableitung eine *Dirac-Funktion* δ besitzt.

$$\delta(x) = \begin{cases} 0 & x \neq \theta \\ \infty & x = \theta \end{cases} \quad (4.16)$$

Zur Implementierung wurde der Funktionswert von $\delta(\theta)$ auf 1 gesetzt, da dies die Berechnungen wesentlich vereinfacht.

Aus der Ableitung des Energiefunktionals aus Gleichung (4.9) nach ihren Parametern erhält man die Gradientenabstiegsfunktionen (nach [CR07]):

$$\frac{d\alpha}{dt} = \int_{\Omega} \delta(\phi_{\alpha,h,s}(x)) v(sx + h) e(x) dx - c \left(\Sigma_k^{-1} \alpha \right) \quad (4.17)$$

$$\frac{dh}{dt} = \int_{\Omega} \delta(\phi_{\alpha,h,s}(x)) \nabla \phi_{\alpha,h,s}(x) e(x) dx \quad (4.18)$$

$$\frac{ds}{dt} = \int_{\Omega} \delta(\phi_{\alpha,h,s}(x)) \nabla \phi_{\alpha,h,s}(x) x e(x) dx \quad (4.19)$$

wobei gilt:

$$e(x) = \log \frac{p_{out}(I(x))}{p_{in}(I(x))} \quad (4.20)$$

Hierbei ist zu beachten, dass es bei der Implementierung von $e(x)$ zu numerischen Problemen kommen kann, wie der Division durch Null. Auch die Übersteuerung muss bedacht werden, da die Logarithmusfunktion schnell ins Unendliche steigt oder fällt.

Zudem beschreibt $\nabla \phi_{\alpha,h,s}(x)$ den Gradienten der embedding function und wird in der praktischen Umsetzung durch zentrale Differenzen $D^c u(x)$ in x- und y-Richtung geschätzt.

$$\frac{\partial u}{\partial x} \approx D^c u(x) = \frac{u(x+1) - u(x-1)}{2} \quad (4.21)$$

4.5.1. Probleme und deren Lösungsansätze

Der größte Nachteil beim Gradientenverfahren ist die Tatsache, dass ihm nur die lokale Umgebung zu Grunde liegt. Hieraus resultieren die folgenden Probleme [Kin92]:

Lokale Minima Wie bereits erwähnt betrachtet man beim Gradientenverfahren nur einen lokalen Bereich. Dabei kann es passieren, dass das Verfahren in einem Minimum terminiert, welches jedoch nur eine lokale Senke ist. Dieses Problem tritt verstärkt bei steigender Dimension der Hyperebene auf, da sie dadurch stärker zerklüftet wird. Dem kann durch den Vergleich mehrerer Durchläufe von verschiedenen Startpunkten entgegengewirkt werden.

Überspringen von Minima Hierbei kann es passieren, dass Minima einfach übersprungen und nicht detektiert werden. Dies passiert meist, wenn ein solches Minimum nur von geringer räumlicher Ausdehnung ist und die Schrittweite des Verfahrens zu hoch gewählt wurde.

Flache Plateaus Es kommt vor, dass Teile der Hyperebene nicht zerklüftet sind, sondern kaum Änderungen aufweisen. Dadurch kommt es zu einem sehr kleinen Gradienten, der das Verfahren dann nur sehr langsam oder gar nicht terminieren lässt, da er auf einer Stelle stagniert.

Oszillation Das Verfahren springt hierbei von einer Seite eines Tals genau zur gegenüberliegenden Seite. Die beiden Gradienten sind gleich, lediglich ihre Vorzeichen ändern sich. Dem Verfahren gelingt es nicht in die Tiefe abzusteigen. Der Ausgangspunkt einer Oszillation kann auch erst nach einigen Zwischenschritten wieder erreicht werden. Dies wird als indirekte Oszillation bezeichnet.

Um die aufgezeigten Probleme zu lösen oder abzuschwächen, gibt es im Wesentlichen zwei Möglichkeiten [Kin92, DER89]. Zum Ersten können die Initialisierung und die Gewichtungen der Teilterme angepasst werden. In unserem Fall wäre dies die Gewichtung c des *Datenterms* zum Wahrscheinlichkeitsterm der Form. Die Wahl des Startpunktes hat einen ganz entscheidenden Einfluss auf den Verlauf des Verfahrens. Es können also mehrere Startpunkte in ihrem Ergebnis verglichen werden, um das beste Resultat zu wählen.

Die zweite Möglichkeit, die sich bietet, ist die richtige Wahl der *Schrittweite* ω des Verfahrens. Dabei bietet eine große Schrittweite ein schnelleres Voranschreiten, erhöht jedoch die Gefahr des Überspringens von Minima oder von Oszillationen. Eine kleine Schrittweite bietet gute Detektionsmöglichkeiten auch für räumlich kleine Minima. Die Laufzeit kann dabei aber bei flachen Plateaus inakzeptabel lang werden oder lokale Minima können nicht mehr verlassen werden. Die dynamische Anpassung der Schrittweite zur Laufzeit bietet für einige dieser Probleme eine Lösungsmöglichkeit. *Resilient Propagation* ist hierfür ein recht einfacher Ansatz von M. RIEDMILLER und H. BRAUN [MR92]. Dabei wird die Schrittweite um einen frei wählbaren Wert vergrößert, wenn sich das Vorzeichen des Gradienten in zwei aufeinander folgenden Schritten nicht ändert. Ändert sich das Vorzeichen, so wurde folglich ein Minimum übersprungen und die Schrittweite muss reduziert werden.

Diese Art der Schrittweitenanpassung wurde auch in der praktischen Implementierung verwendet. Anwendung findet sie jedoch nur bei der Wahl der Schrittweite für den Gradienten von α , da die Schrittweite für die beiden anderen Parameter unter Umständen unkontrolliert steigen könnte. Dies kann dazu führen, dass Größen- und Positionsveränderungen den Evolutionsbereich über kritische Entscheidungsstellen hinweg bewegen und damit ein Überlaufen der Form begünstigen.

4.6. Seeded-Region-Growing-Ansatz

Der Segmentierungsprozess soll nach dem sogenannten *Seeded-Region-Growing* Ansatz initialisiert werden. Dabei wird eine Startregion nach dem Regionswachstumsverfahren aus Kapitel 4.1 erzeugt. Als Initial wird ein Saatpunkt gesetzt, aus dem sich dann durch Konturevolution mit Hilfe der Energieminimierung die bestmögliche Segmentierung entwickeln soll. Dies kann manuell oder automatisiert geschehen. Dazu muss der Saatpunkt selbst folgenden Ansprüchen genügen:

- Der Saatpunkt s muss unter einer Intensitätsschwelle τ liegen. Dies soll verhindern, dass der Saatpunkt in einer weiß gefärbten Zelle oder im Zellzwischen gewebe liegt.

$$I(s) < \tau \quad \tau \in \mathbb{R}^3 \text{ und } s \in \Omega \quad (4.22)$$

- Eine gewisse Nachbarschaft N um den Saatpunkt sollte im selben Intensitätsbereich liegen. Dabei wird der Intensitätsunterschied durch die Distanz der Farbwerte bestimmt und darf die Schwelle ϵ nicht überschreiten. Dadurch soll verhindert werden, dass ein Saatpunkt zu nah am Rand einer zu segmentierenden Form liegt und hierdurch unnötig viele Iterationsschritte notwendig werden. Zudem soll ausgeschlossen werden, dass der Saatpunkt nur in ein Artefakt⁹ des Bildes gefallen ist.

$$|(I(x) - I(s))| \leq \epsilon \quad \epsilon \in \mathbb{R} \text{ und } x \in N \text{ sowie } N \subset \Omega \quad (4.23)$$

- Die Nachbarschaft um den Saatpunkt darf nicht zu einem bereits segmentierten Bereich S gehören. Auch dies soll die Wahrscheinlichkeit einer Randlage von s minimieren, vor allem in Bezug auf Zellen, die sehr nahe beieinander liegen und den selben Farbwert haben, wo ein Überlaufen in eine andere bereits segmentierte Zelle nicht ausgeschlossen werden kann.

$$x \notin S \quad x \in N \text{ und } S \subset \Omega \quad (4.24)$$

Zur Initialisierung der eigentlichen Formevolution muss eine Initialform gegeben werden, auf der ein erstes statistisches Modell erstellt werden kann, um daraus eine bestmögliche Approximation der vorhandenen Bilddaten in den Formenraum zu schaffen. Dies geschieht durch ein einfaches Füllen der nahen Nachbarschaftsumgebung N , falls sie die Farbraumdistanz ϵ nicht überschreitet. Diese erste Form wird dann nach Formel (3.13) durch einen Formvektor α approximiert und mit den beiden anderen ausgelesenen Parametern s und h in die embedding-function ϕ überführt. Dazu wird die Initialform zentral in eine quadratische *Bounding Box*¹⁰ eingebettet. Die Koordinaten der oberen linken Ecke dieser Box werden als Parameter h herangezogen, die Skalierung der Dimension der Beispieldaten (in unserem Fall 100 x 100 Pixel) zur Ausdehnung der neuen Bounding Box geht als Parameter s ein.

⁹Schmutzpartikel, Bildfehler oder Reste einer zu kleinen Zelle

¹⁰Ein einfacher geometrischer Körper, der ein komplexes dreidimensionales Objekt oder einen komplexen Körper umschließt

4.7. Zusammenfassung

Verschiedene grundlegende Segmentierungsansätze wurden betrachtet. Diese müssen um Komponenten erweitert werden, welche die Form der zu segmentierenden Bereiche näher bestimmt.

Hierzu wurden Beispieldaten nach bestimmten Voraussetzungen ausgewählt und durch die statistische Methode der Hauptkomponentenanalyse aufbereitet. Eine Kosten- oder Energiefunktion über den Wahrscheinlichkeiten der Segmentierungsergebnisse wurde aufgestellt und in eine parametrisierte Darstellung überführt. Zur Lösung des Energieminimierungsproblems wurde das iterative Verfahren des Gradientenabstiegs vorgestellt und auf die aus der Aufgabenstellung hervorgehenden Gegebenheiten angepasst. Durch einen Seeded-Region-Growing-Ansatz soll dann eine Konturevolution initialisiert werden.

5. Ergebnisse

Im Folgenden soll ein Überblick über die Leistungsfähigkeit des entwickelten Verfahrens gegeben werden. Dabei ist festzuhalten, dass in der praktischen Realisierung die Parameter h und s nicht durch die Gradientenfunktionen bestimmt wurden. Sie wurden durch die Zentrierung der Einbettungsfunktion ϕ in ihrer Bounding Box bestimmt, da die Nutzung der Ableitungsfunktionen nicht so einfach zu kontrollieren war und zu schlechteren Ergebnissen führte.

In Abbildung 5.1 ist zu erkennen, dass auch Formausprägungen, welche nicht in den Trainingsdaten vorhanden sind, zufriedenstellend segmentiert werden können. Das hier verwendete vereinfachte Testszenario dient der Evaluation der theoretischen Überlegungen in kontrollierbarer Umgebung.

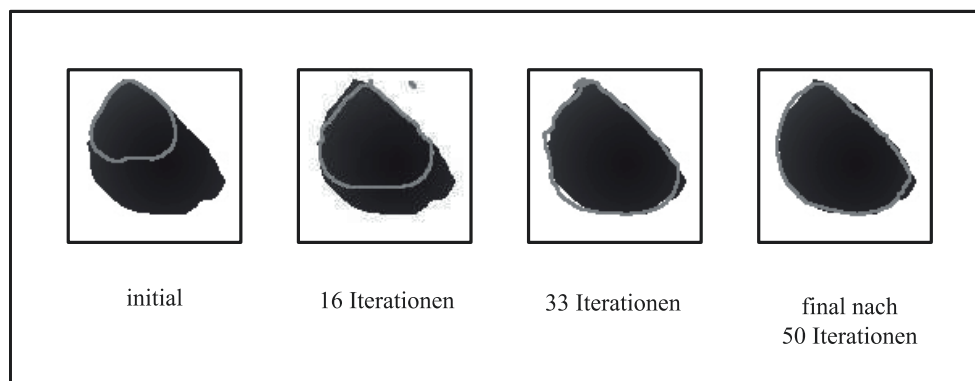


Abbildung 5.1. – Segmentierung auf künstlicher Zellform, welche nicht im Beispielsatz enthalten ist

Auch kritische Entscheidungsstellen werden wie in Abbildung 5.2 gezeigt handhabbar. Dabei ist die Wahl des Startpunktes ein elementarer Faktor für den zufriedenstellenden Erfolg. Zudem wurde die Wahl der Parameter für jede Segmentierung den Bedingungen angepasst. Alle Tests wurden mit dem 1200 Datensätze umfassenden Trainingsset durchgeführt, wobei die ersten zwanzig Hauptkomponenten zur Approximation genutzt wurden, was einer Varianzabdeckung von ca. 69% entspricht.

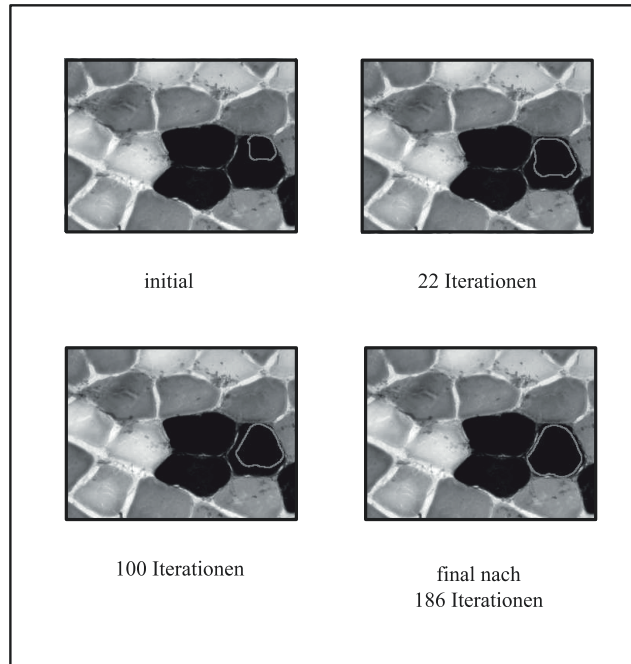


Abbildung 5.2. – Konturevolution an kritischer Stelle

Eine Konturevolution mit kritisch gewähltem Initialpunkt wird in den Abbildungen 5.3 und 5.4 illustriert. Dabei zeigt sich, dass eine zu geringe Gewichtung des Formenfaktors und ungünstige Skalierungs- und Translationsschritte ein Überlaufen in benachbarte Formen ermöglichen.

Eine höhere Gewichtung der Formenwahrscheinlichkeit kann das Überlaufen verhindern. Dies führt jedoch dazu, dass das Segmentierungsergebnis stark geglättet und der Mittelwertform angenähert ist (vgl. Abb. 5.4 und Abb. 4.4).

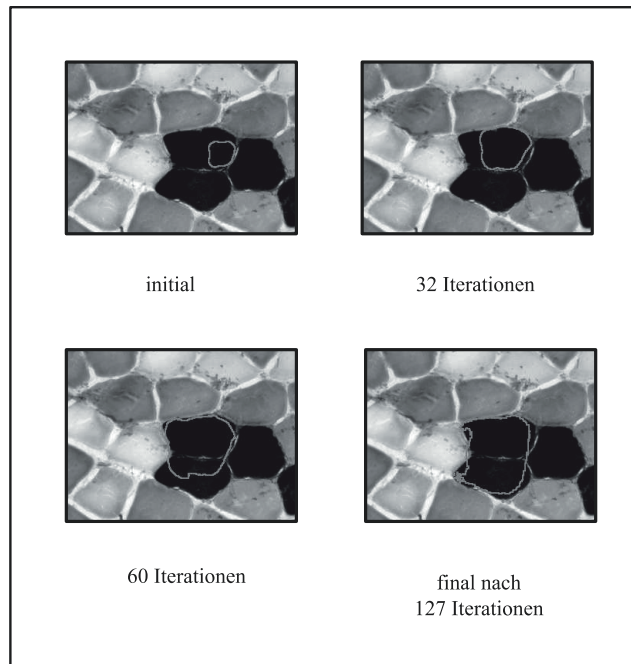


Abbildung 5.3. – Konturevolutionsüberlauf durch zu geringe *shape prior* Gewichtung

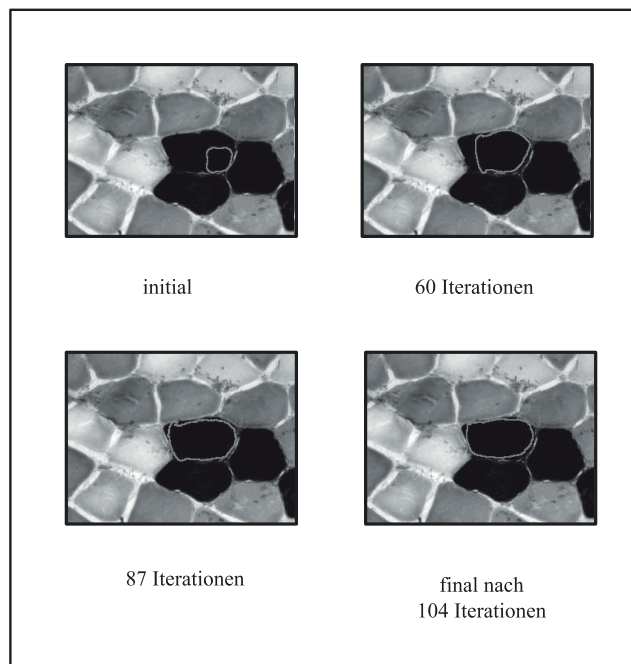


Abbildung 5.4. – Konturevolution mit höher gewichtetem *shape prior*

6. Zusammenfassung und Ausblick

Es konnte gezeigt werden, dass das Mittel der Hauptkomponentenanalyse in diesem Einsatzbereich ein passendes Werkzeug darstellt und dass das Überlaufen von Formen durch den Einsatz des vorgestellten *shape priors* verhindert werden kann.

Das Verfahren liefert in der hier implementierten Version akzeptable, jedoch noch nicht vollständig zufriedenstellende Ergebnisse. Zudem ist es von vielen Variablen und Parametern abhängig, welche das Ergebnis beeinflussen. Diese wurden meist heuristisch ermittelt und machen es wenig nutzerfreundlich.

Für die weitere Entwicklung ist die Einbindung der Ableitungen nach h und s in die praktische Realisierung sowie eine abschließende Auswertung der Ergebnisse wünschenswert. Diese könnte durch eine Clusteranalyse auf den mittleren Grauwerten jeder segmentierten Form realisiert werden. Des Weiteren wären ein Verfahren für die Segmentierung weißer, ungefärbeter Zellen sowie effektive Schätzverfahren für die vielen Parameter weitere Schritte zur Vervollständigung.

Da das Verfahren wenig effizient in der Ressourcennutzung, ist sollte es noch auf gesteigerte Laufzeit weiterentwickelt werden, zumal die genutzte *java* [WB06] Implementierung nicht multi-threading-fähig ist. Die Nutzung paralleler Berechnung könnte einen erheblichen Geschwindigkeitsvorteil mit sich bringen. Auch die Ausnutzung besonderer Hardware ist vorstellbar. Da viele Operationen auf Matrizen und Vektoren basieren, ist eine Ausnutzung von Grafikchips oder des CELL-Prozessors der Spielekonsole PlayStation3 denkbar.

Literaturverzeichnis

- [AN07] ALFRED NISCHWITZ, MAX FISCHER, PETER HABERÄCKER: *Computergrafik und Bildverarbeitung - Alles für Studium und Praxis*. Nummer 978-3-8348-0186-9. Vieweg, 2 Auflage, März 2007.
- [Asm08] ASMUS, ALEXANDER: *Segmentierung von ATPase-gefärbten Muskelfaserschnitten mittels Seeded-Region-Growing*. Großer Beleg, Technische Universität Dresden, Mai 2008.
- [Bis06] BISHOP, CHRISTOPHER M.: *Pattern Recognition and Machine Learning*. Nummer 0-38-731073-8. Springer, 2006.
- [Bök05] BÖKER, PROF. DR. FRED: *Multivariate Verfahren - Hauptkomponentenanalyse*, Kapitel 4, Seiten 32 – 66. 2005.
- [BM70] BROOKE M.H., KAISER K. K.: *Muscle fiber types: How many and what kind?* Arch Neural (Chic), 23:369–379, 1970.
- [CE36] C. ECKART, G. YOUNG: *The Approximation of One Matrix by Another of Lower Rank*. Psychometrika 1, 1936.
- [CR07] CREMERS, D. und M. ROUSSON: *Efficient kernel density estimation of shape and intensity priors for level set segmentation*. In: SURI, J. S. und A. FARAG (Herausgeber): *Parametric and Geometric Deformable Models: An application in Biomaterials and Medical Imagery*. Springer, May 2007.
- [DER89] DAVID E. RUMMELHART, JAMES L. MCCLELLAND: *Parallel Distributed Processing*. MIT Press, 1989.
- [Han00] HANDELS, HEINZ: *Medizinische Bildverarbeitung*. B.G. Teubner, 2000.
- [HH03] H. HOPPELER, R. BILLETTER: *Struktur und Funktion der Skelettmuskulatur*. Therapeutische Umschau - Verlag Hans Huber, Bern, Band 60:363 – 370, 2003.
- [Hot33] HOTELLIG, H.: *Analysis of a complex of statistical variables into principal components*. Journal of Educational Psychology 24, 1933.
- [Kar47] KARHNUEN, K.: *Über lineare Methoden in der Wahrscheinlichkeitsberechnung*. Acad Sci Fennicae A, 137, 1947.
- [Kin92] KINNEBROK, W.: *Neuronale Netze*. R. Oldenbourg Verlag, München/Wien, 1992.

- [LEGF00] LEVENTON, MICHAEL E., W. ERIC, L. GRIMSON und OLIVIER FAUGERAS: *Statistical shape influence in geodesic active contours*. Seiten 316–323. 2000.
- [Mei05] MEISTER, ANDREAS: *Numerik linearer Gleichungssysteme - Eine Einführung in moderne Verfahren*. Nummer 3-528-13135-7. Vieweg, 2 Auflage, Januar 2005.
- [MR92] M. RIEDMILLER, H. BRAUN: *Rprop - A Fast Adaptive Learning Algorithm*. Proceedings of the International Symposium on Computer and Information Science VII, 1992.
- [RH95] RENÉ HENRION, GÜNTER HENRION: *Multivariate Datenanalyse - Methodik und Anwendung in der Chemie und verwandten Gebieten*. Springer, 1995.
- [Sch82] SCHRÖDER, JOHANN MICHAEL: *Spezielle pathologische Anatomie - Band 15 - Pathologie der Muskulatur*. Springer-Verlag, 1982.
- [Sch01] SCHÖLKOPF, SMOLA, MÜLLER: *Nonlinear Component Analysis as a Kernel Eigenvalue Problem*. Ian Fasel, September 25, 2001.
- [WB06] WILHELM BURGER, MARK JAMES BURGE: *Digitale Bildverarbeitung - Eine Einführung mit Java und ImageJ*. Nummer 10 3-540-30940-3. Springer, 2 Auflage, 2006.
- [Wec07] WECHSLER, HARRY: *Reliable face recognition methods: system design, implementation and evaluation*. Springer Verlag, 2007.
- [Wil05] WILM, DR. ANDREAS: *Skript - Psychologische Methodenlehre für Nebenfachstudierende I+II*, 2005. Kapitel 7.3 Hauptkomponentenanalyse.

A. Anhang

A.1. Erklärung

Hiermit erkläre ich, Jens Bornschein, den vorliegenden Großen Beleg zum Thema

Effiziente Einbringung von statistischem Formwissen in die Segmentierung von ATPase gefärbten Muskelfaserbildern

selbstständig und ausschließlich unter Verwendung der im Quellenverzeichnis aufgeführten Literatur- und sonstigen Informationsquellen verfasst zu haben.

Dresden, den 11. September 2009

.....
Unterschrift (Jens Bornschein)

A.2. Danksagungen

Ich möchte meinen beiden Betreuern Dr. rer. nat. habil. Boris Flach und im Besonderen Dr. Thomas Brox danken, die mich in der Umsetzung und Ausarbeitung dieser Arbeit tatkräftig unterstützt haben.

Des Weiteren gilt mein Dank Herrn Dr. Peter Ruck vom Pathologischen Institut am Krankenhaus Leonberg und Frau Prof. Dr. Antje Bornemann vom Institut für Hirnforschung Tübingen für ihre Anregungen.

A.3. Auszug aus den Beispieldaten

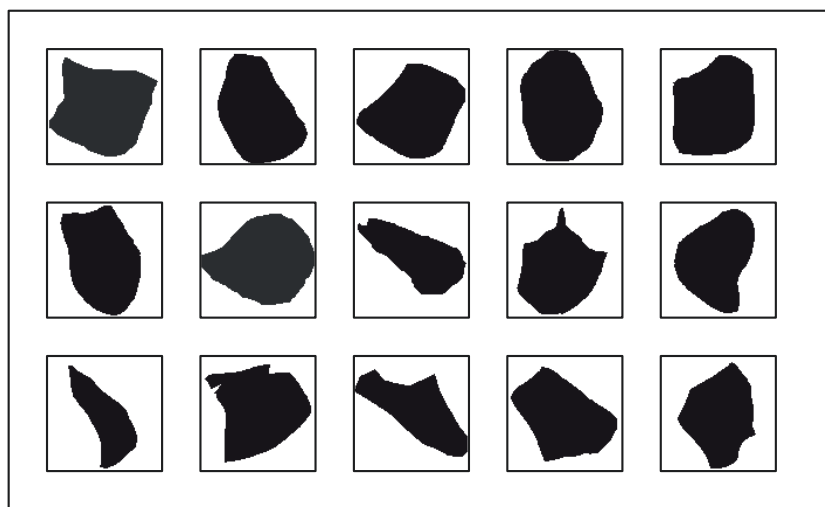


Abbildung A.1. – ausgewählte Beispiele aus dem verwendeten Trainingsset



Abbildung A.2. – zusätzlich generierte Beispieldaten - von links nach rechts: Originalform, horizontal gespiegelt, vertikal gespiegelt, um 90° gedreht, gedrehte Form horizontal gespiegelt, gedrehte Form vertikal gespiegelt

A.4. Vergleich über die Anzahl an Datensätzen

Die Frage nach der Anzahl der zu verwendenden Datensätze wurde in Kapitel 4.2.1 gestellt. Die folgende Tabelle soll die für die praktische Umsetzung herangezogenen Trainingssets näher beschreiben.

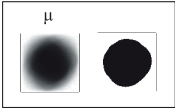
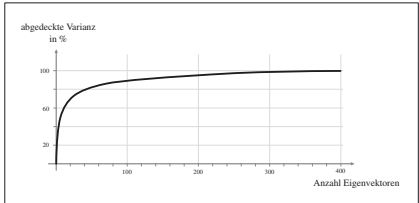
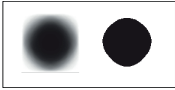
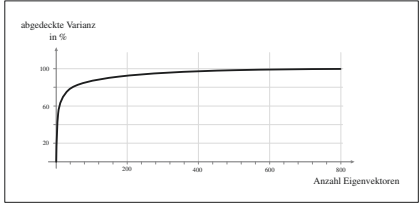

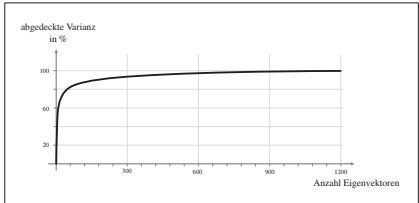
Datensatz	Mittelwertform	Varianzverteilung	Schwellen
400 Beispieldaten: 200 segmentierte Formen +200 um 90° gedrehte			>50% nach 6 >90% nach 105 ≈ 0 sind 2
800 Beispieldaten: 400 Datensätze + 400 horizontal gespiegelte			>50% nach 7 >90% nach 143 ≈ 0 sind 4
1200 Beispieldaten: 400 Datensätze + 400 horizontal gespiegelte + 400 vertikal gespiegelte			>50% nach 7 >90% nach 164 ≈ 0 sind 6

Tabelle A.1. – Vergleich der verwendeten Datensätze nach ihrer Varianz