Analyse historischer chinesischer Manuskripte mit Hilfe des Shape Context Verfahrens

Diplomarbeit

Rainer Herzog

Betreuer:

Dr. rer. nat. Ullrich Köthe
Prof. Dr. Leonie Dreschler-Fischer

Universität Hamburg, MIN-Fakultät, Department Informatik
Arbeitsbereich Kognitive Systeme
Vogt-Kölln-Straße 30, 22527 Hamburg

Dezember 2007

Zusammenfassung

Diese Arbeit beschäftigt sich mit der Segmentierung und dem Vergleich einzelner chinesischer Zeichen aus historischen Manuskripten. Dazu werden die Manuskripte mittels Wasserscheidentransformation zunächst in Spalten aufgeteilt, dann diese Spalten in Bereiche einzelner Zeichen. Aus Binärbildern dieser Bereiche werden mittels Labeling Strukturen als Zeichen erkannt.

Ein Vergleich verschiedener Zeichen untereinander erfolgt durch das Shape Context Verfahren. Dazu werden auf der Kontur der Zeichen Punkte bestimmt. Zu jedem Punkt wird ein „SHAPE CONTEXT“ erfaßt; dieser beinhaltet den Abstand und die Richtung jeweils aller übrigen Punkte zum zuvor ausgewählten und ermöglicht so eine vergleichbare Beschreibung der Kontur. Die Ähnlichkeit zweier Zeichen läßt sich nun über den Vergleich aller SHAPE CONTEXTe zweier Zeichen beziffern.

Wahlweise können anstatt der Kontur auch die Skelette der Zeichen verglichen werden.

Zur ergonomischen Anwendung wird die gesamte Funktionalität unter einer grafisch bedienbaren Oberfläche präsentiert.

Inhaltsverzeichnis

1 Einleitung 1

2 Grundlagen 5
2.1 Chinesische Handschriften . . . 5
2.2 Optische Zeichenerkennung in der Bildverarbeitung . . . 6
2.2.1 Vorverarbeitung der Bilddokumente . . . 6
2.2.2 Erfassung einzelner Zeichen . . . 7
2.3 Zeichenanalyse in der Paläografie . . . 10

3 Zeichensegmentierung 15
3.1 Segmentierung in Spalten . . . 17
3.1.1 Glättung des Bildes . . . 17
3.1.2 Segmentierungsverfahren . . . 18
3.1.3 Bewertung der Trennkanten mit Hilfe der GeoMap . . . 19
3.1.4 Filtern unerwünschter Kanten . . . 23
3.1.5 Vereinigung von Flächen . . . 23
3.2 Segmentierung der Zeichen . . . 29
3.2.1 Glättung der Spalten . . . 29
3.2.2 Zusammenführen schmaler Flächen . . . 30
3.2.3 Löschen aller verbleibenden Kanten, die ein Zeichen schneiden . . . 32

4 Shape Context 35
4.1 Subpixelgenaue Segmentierung der Zeichen . . . 35
4.2 Zeichenvergleich durch Shape Context . . . 36
4.2.1 Auswahl geeigneter Vergleichskriterien . . . 36
4.2.2 Beschreibung des Shape Context-Verfahrens . . . 38
4.2.3 Eigenschaften des Shape Context . . . 39
4.2.4 Vergleich von Shape Contexten . . . 39
4.2.5 Beispiel eines Zeichenvergleichs . . . 40
4.2.6 Erweiterungen des Vergleichsverfahrens . . . 43

5 Anwendungen 45
5.1 Beispiele für die Segmentierung von Zeichen . . . 45
5.1.1 Vorstellung der grafischen Oberfläche . . . 45
5.1.2 Segmentierung einer einfachen Vorlage . . . 50
5.1.3 Segmentierung problematischer Vorlagen . . . 52

5.2 Zeichenvergleich durch Nutzung der grafischen Oberfläche 58
5.2.1 Segmentierung der Zeichen innerhalb der zuvor festgelegten Flächen . 58
5.2.2 Darstellung von Vergleichsergebnissen 63
5.3 Variationen des SHAPE CONTEXT 63
5.3.1 Variation der Punkteanzahl auf der Kontur 63
5.3.2 Variationen der Zonenmaske 66
5.4 Variation des Kostenmaßes 71
5.5 Transformation von Zeichen 74
5.6 Verwendung der Zeichenskelette 76
5.7 Auffinden gleicher und ähnlicher Zeichen 80
5.8 Auswertung eines Zeichenvergleichs 83

6 Fazit und Ausblick 85
6.1 Zusammenfassung und Bewertung 85
6.2 Ausblick 87
6.2.1 Verbesserung der Segmentieung 87
6.2.2 SHAPE CONTEXT 88
6.2.3 Benutzungsoberfläche 88
6.2.4 Datenbanken 89
6.2.5 Erfassung weiterer Merkmale 90

Anhang 91

Literaturverzeichnis 94

1 Einleitung

In dieser Arbeit werden historische chinesische Manuskripte mit Hilfe computergestützter Analyseverfahren untersucht. Damit sollen neue Möglichkeiten erschlossen werden, die bisher von Sinologen verwendeten Auswertungsmethoden zu unterstützen und zu ergänzen.

In den vergangenen Jahren wurden in China altertümliche Grabstätten entdeckt, welche als Grabbeigaben unter anderem Manuskripte enthielten. Diese Manuskripe sind teilweise nur schlecht erhalten oder liegen in mehreren Teilen vor. Eine Antwort auf die Frage, welche Manuskriptteile tatsächlich zu einem Manuskript gehören oder ob sie unterschiedliche Manuskriptstücke darstellen, könnte ein besseres Verständnis über die historischen chinesischen Anschauungen im Bereich der Philosophie oder der Politik liefern. Die Zuordnung verschiedener Schriftstücke zu verschiedenen Autoren erlaubt eine Einordnung dieser, sowohl in zeitlicher Ordnung, als auch in Hinblick auf deren Bedeutung.

Der Inhalt der Schriftstücke läßt nicht unbedingt auf deren Zusammengehörigkeit schließen; oft werden in Gräbern Schriften entdeckt, die verschiedene Ansichten vertreten. Da es aber prominente Beispiele für historische chinesische Schriftstücke gibt, die sowohl konfuzische als auch daoistische Anschauungen enthalten, die man im allgemeinen als Vertreter vollkommen unterschiedlicher Auffassungen ansieht, kann auf dieser Grundlage nicht ausgeschlossen werden, daß Teilstücke nicht doch zusammengehören.

Es scheint hingegen möglich, anhand der Erkennung einer Handschrift nützliche Schlüsse zur Einordnung von Dokumentstücken ziehen zu können.

Nach [Richter, 2005] lassen sich grob drei paläografische Hierarchieklassen unterteilen.

1. **Der Schrifttyp**: Hiermit wird ein gleichförmiges Aussehen der Buchstaben des kompletten Schriftstücks beschrieben. Oft werden Schrifttypen mit einem Material („Bronzeinschriften“), einer bestimmten zeithistorischen Entwicklung oder einer bestimmten Gegend verbunden, in der dieser Schrifttyp üblich war.

2. **Der Schriftstil**: Dieser beschreibt die Ausführung eines Schrifttyps. Hier mag es Unterschiede in der Gleichförmigkeit von gleichen Buchstabenelementen oder charakteristischer Strichführung geben, auch die Schreibgeschwindigkeit, die Schrägheit der Buchstaben oder deren Abstand.

   Diese Schriftstile können typisch für bestimmte Epochen, Regionen oder bestimmte Lehrrichtungen sein. Ein Schriftstil läßt sich durch Attribute wie *barock, wellig, aufblühend, dynamisch, breit*; ein anderer etwa durch *einfach, geradlinig, statisch, glatt* beschreiben.

3. **Die Handschrift**: Während die Unterscheidung der Schreibweisen verschiedener Autoren deutlich einfacher ist, wird mit der „Handschrift“ eine bestimmte Person eines bestimmten Alters beschrieben; die Schreibweise einer Person kann im Verlauf ihres Lebens durchaus variieren.

   Es ist nicht unüblich, daß Manuskripte ab einer bestimmten Länge von verschiedenen Personen geschrieben worden sind, die den gleichen Schriftstil anwandten.

   Da zusammenhängende Manuskripte für sich innerhalb relativ kurzer Zeit fertiggestellt wurden, können verschiedene Handschriften immer verschiedenen Personen zugeordnet werden; unterschiedliche Manuskripte können aber durchaus unterschiedliche Handschriften derselben Person enthalten, soweit ihre Entstehung zeitlich deutlich auseinanderliegt. Kriterien zur Handschriftbestimmung finden sich in ähnlicher Weise wie zur Unterscheidung der Schriftstile; auch hier können Unterschiede wie der Zeichenabstand, die Zeichengröße, der Schrägheitsgrad, oder die Schreibgeschwindigkeit bewertet werden. In der Regel dienen aber die feineren Details als sicherere Merkmale, wie die Häufigkeit, in der der Pinsel zur Tinte geführt wird, der Anpreßdruck in der Pinselführung oder in bestimmten Zeichen ein typischer Strichabstand oder die Verbindung einzelner Pinselstriche.

Ausgehend von der Zielsetzung, aus von Sinologen bereitgestellten digitalen Bildvorlagen verschiedene Handschriften voneinander unterscheiden zu können, werden in dieser Arbeit die Teilaspekte der Segmentierung einzelner Buchstaben aus diesen Vorlagen sowie die Erstellung eines Vergleichsmaßes zwischen einzelnen Zeichen erarbeitet.

Auch wenn im Rahmen dieser Arbeit eine klare Abgrenzung von ähnlichen Handschriften zueinander nicht erreicht werden kann, so kann – nach erfolgreicher Segmentierung – doch eine Aussage über die Ähnlichkeit verschiedener Buchstaben sowohl gleicher als auch unterschiedlicher Art zueinander getroffen werden. Die Durchführung der Zeichenanalyse wird durch den Benutzer über eine grafische Bedienoberfläche gesteuert, die auch zur Präsentation der Vergleichsergebnisse dient.

Die Beschreibung des Vorgehens erfolgt in den nachstehenden Kapiteln:

Vorgehensweise:

Das **2. Kapitel** beschreibt Grundlagen der chinesischen Zeichen. Mögliche Ansätze zum Vergleich dieser Zeichen werden ebenso vorgestellt wie die damit verbundenen Varianten der Segmentierung einzelner Zeichen aus einer Bildvorlage.

Im **3. Kapitel** wird die durchgeführte Methode der Zeichensegmentierung beschrieben. Um einzelne Zeichen miteinander vergleichen zu können, müssen auf vorhandenen Bildvorlagen, die oft mehrere hundert Zeichen darstellen, einzelne Bildbereiche bestimmt werden, die jeweils genau ein Zeichen enthalten. Alle Bildpixel in diesen Bereichen, die aufgrund ihrer Helligkeit als Pinselstriche erkannt werden, werden darauf als ein Zeichen gewertet.

Das **4. Kapitel** beschreibt das Verfahren zur Gewinnung von SHAPE CONTEXTen, auf deren Basis einzelne, zuvor segmentierte Zeichen verglichen werden können.

Der Zugriff auf die Funktionalität wird im **5. Kapitel** beschrieben, dieser wird mittels einer grafischen Benutzungsoberfläche gewährleistet, die eine einfache Interaktion des Benutzers ermöglicht. Hier werden Ergebnisse der Segmentierung verschiedener Bildvorlagen präsentiert, die Wirkung einzelner über die Benutzungsoberfläche einstellbarer Parameter gezeigt, die daraus gewonnenen Vergleichsergebnisse vorgestellt und aufgetretene Probleme erläutert. Beispielhaft wird eine mögliche Anwendung analysiert.

Abschließend werden im **6. Kapitel** die Ergebnisse dieser Arbeit bewertet, dazu werden einige Erweiterungsmöglichkeiten vorgestellt.

2 Grundlagen

Dieses Kapitel beschreibt einige Eigenschaften chinesischer Schriftzeichen, und die sich daraus ergebenden möglichen Ansätze, diese Zeichen aus einer Bildvorlage zu segmentieren. Weiterhin werden einige Varianten vorgestellt, charakteristische Merkmale einzelner Zeichen zu bestimmen, um diese identifizieren zu können oder sie miteinander zu vergleichen.

2.1 Chinesische Handschriften

Nach [Xigui, 2000] ist Anfang des 17. Jhd. n. Chr. unter Kaiser Kangxi das „*Kāngxī zìdiǎn*“ entstanden, ein chinesisches Wörterbuch, das 47.043 Zeichen umfaßt. Von diesen Zeichen waren aber selbst damals die meisten ungebräuchlich, auch heute rechnet man mit nur rund 5000-6000 Zeichen, die im regulären Gebrauch sind. Mit einer Kenntnis von 3800 Zeichen ist man in der Lage, 99,9% eines durchschnittlichen Buches zu lesen, für 99,99% benötigt man 5200 Zeichen.

In der heute üblichen vereinfachten Schreibweise, die zwischen 1956 und 1964 entwickelt wurde, erfolgt die Schreibrichtung zeilenweise von links nach rechts, die Zeilen darauf von oben nach unten. Laut [Srihari et al., 2007] wurden zudem 2235 Zeichen vereinfacht, indem die Anzahl der Striche, aus denen die Zeichen bestehen, auf durchschnittlich 10,3 Striche pro Zeichen reduziert wurde. In den hier untersuchten historischen Manuskripten wird dagegen die traditionelle chinesische Schreibweise verwendet, in der spaltenweise von oben nach unten geschrieben wird, darauf von rechts nach links. Die Zeichen traditioneller Schreibweise bestehen aus etwa 16,03 Strichen pro Zeichen. Einen Vergleich der Zeichen für „Drache“ zeigt Abbildung 2.1, hier wurde die Strichanzahl von 16 auf 5 reduziert.

Aufgrund der hohen Anzahl von verschiedenen Zeichen bestehen Wörter historischer Schriften in der Regel aus nur 1-2 Zeichen, nach vielen Reformen haben Wörter heute oft 3 oder mehr Zeichen. Handschriftlich erzeugte Zeichen sind nicht miteinander verbunden, wie es bei westlichen Schreibschriften üblich ist, woduch eine Segmentierung einzelner Zeichen zunächst einfach erscheint; da aber die Zeichen selbst meistens aus nicht zusammenhängenden Komponenten bestehen, ist es bei geringem Zeichenabstand nicht immer eindeutig, welche Komponenten welchen Zeichen zuzuordnen sind.

Eine systematische Einteilung in Form von 214 Radikalen fand im Kangxi-Wörterbuch statt. Radikale können einzelne Zeichen oder Teile von Zeichen sein, und werden im Wörterbuch als

Abbildung 2.1: Das chinesische Zeichen für „Drache“ („lóng“).
Links die traditionelle, rechts die vereinfachte Schreibweise.

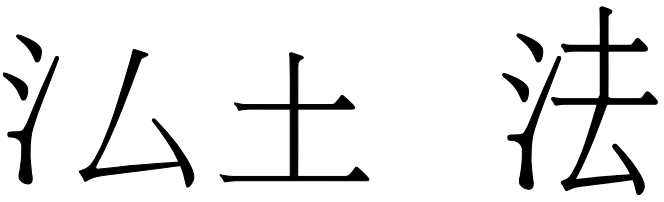


Abbildung 2.2: Die drei Radikale Nr. 85 („shuǐ“), Nr. 28 („sī“) und Nr. 32 („tǔ“)
ergeben zusammengesetzt das Zeichen für „Gesetz“ („fǎ“)

Suchkriterium verwendet. Eingeteilt nach der Anzahl der enthaltenen Striche gibt es Radikale mit nur einem und bis zu 17 Strichen. Das Zeichen für „Gesetz“ besteht so z.B. aus drei Radikalen (Abbildung 2.2).

2.2 Optische Zeichenerkennung in der Bildverarbeitung

2.2.1 Vorverarbeitung der Bilddokumente

Zur Erkennung einzelner Zeichen oder Wörter ist eine Vorverarbeitung des Dokumentes nötig, in der etwa Störpixel entfernt werden oder Linien, die z.B. beim Einscannen gefalzter Papierseiten entstehen können. Hier werden auch Dokumente normalisiert, damit auch unterschiedliche Dokumente vergleichbare Zeichengrößen enthalten oder die Strichdicken der Zeichen aneinander angepaßt werden.

Die Vorverarbeitung umfaßt auch die Segmentierung. Zunächst werden die Dokumente in Spalten respektive Zeilen aufgeteilt, diese werden darauf in einzelne Zeichen unterteilt. Da die Zwischenräume zwischen chinesischen Wörtern nicht größer sind als die zwischen einzelnen Buchstaben, bietet sich die Aufteilung in Wörter nicht an, soweit keine Datenbank vorliegt, mit denen Wörter abgeglichen werden können. Wie bereits erwähnt ist die Trennung einzelner Zeichen oft schwierig, wenn diese aus einzelnen Radikalen bestehen, die für sich selbst auch eigene Zeichen darstellen können. Außerdem kann es vorkommen, daß sich Radikale innerhalb eines Zeichens nicht berühen, sich Striche von Radikalen benachbarter Zeichen dagegen überlappen.

Zur Trennung einzelner Zeichen wird in [Liang u. Shi, 2005], sowie in [Tseng u. Lee, 1999] mittels des Viterbi Algorithmus nach möglichen Pfaden gesucht, an denen eine Trennung möglich scheint. Diese Pfade können sowohl linear als auch nichtlinear sein. Mittels weiterer Regeln werden darauf unwahrscheinliche Pfade verworfen.

In [Tseng u. Chen, 1998] werden dagegen alle Striche der Zeichen durch die sie umschließenden Rechtecke repräsentiert, wissensbasierte Operationen vereinen darauf ausgewählte Rechtecke. Anschließend wird nach passenden Abgrenzungen zwischen einzelnen Zeichen gesucht.

Die Normalisierung einzelner Zeichen beschränkt sich nicht nur auf die lineare Skalierung um alle Zeichen in eine möglichst einheitliche Größe zu transformieren; es werden heute üblicherweise bei der chinesischen Handschrifterkennung nichtlineare Normalisierungen genutzt, die auch die Dicke der einzelnen Striche aneinander angleichen. Cheng-Lin Liu et al. beschreiben darüber hinaus in [Liu u. Marukawa, 2004] und [Liu et al., 2003] weitere Methoden der Normalisierung, die weniger aufwändig sind als die nichtlineare Normalisierung und dennoch zu einer besseren Zeichendarstellung führen als diese. Weitere Bestrebungen richten sich auf die Angleichung der Schrägheit der Buchstaben, da Unterschiede in der Schrägheit große Herausforderungen an die nachfolgenden Methoden zur Zeichenerkennung stellen. In [Kato et al., 1999] wird erwähnt, daß sich die Schrägheit nach deren Erfahrung hauptsächlich auf Striche auswirkt, die horizontal oder vertikal ausgerichtet sein sollten, während diagonale Striche geringere Variationen bezüglich ihrer Schrägstellung aufweisen. Es wird in deren Arbeit versucht, mittels *TPID*, „Transformation based on partial inclination detection“ auch solche Zeichen auszurichten, die nur teilweise schräg ausgerichtet sind, wie z.B. in einem Fall, wo alle Striche, die horizontal verlaufen sollten, um 15° schräg gestellt sind, anders verlaufende Striche dagegen keine Korrektur benötigen.

2.2.2 Erfassung einzelner Zeichen

Auch wenn in dieser Arbeit nicht der Versuch unternommen wird, einzelne Zeichen zu erkennen und zu identifizieren, wie es bei der optischen Zeichenerkennung (OCR) die Zielsetzung ist, so treten bei Einblicken in die Herangehensweisen dieser Aufgaben doch interessante Parallelen hervor.

In dieser Arbeit werden Zeichen untereinander verglichen, wobei dem Vergleich gleichartiger Zeichen besondere Bedeutung zukommt. Dabei sollten individuelle Schreibweisen erkannt werden, die sich an gleichartigen Zeichen einer Handschrift erkennbar wiederholen und sich gegenüber anderen Handschriften abgrenzen sollten. Eine Normalisierung dient also auch hier zur Hilfe der Erkennung gleichartiger Zeichen, weitergehende Untersuchungen sollten dann aber an den Zeichen in ursprünglicher Darstellung vorgenommen werden.

Bei der optischen Zeichenerkennung werden einzelne Zeichen mit einer Normvorlage abgeglichen, um diese identifizieren zu können. Dabei wird durch eine Normalisierung versucht,

individuelle Schreibvariationen herauszufiltern. Nach erfolgter Normalisierung sollten gleichartige Zeichen verschiedener Handschriften idealerweise identisch sein.

Bisherige Ansätze der Zeichenerkennung erfolgen auf drei Arten:

- Vergleich einzelner Striche der Zeichen,
- Vergleich von Radikalen
- oder durch den Vergleich ganzer Zeichen (holistischer Ansatz).

In [Srihari et al., 2007] werden verschiedene Quellen benannt, in denen Vergleiche auf jeweils eine dieser Arten durchgeführt werden. Das auf Strichen basierende Verfahren versucht, Zeichen anhand der Anzahl der Striche, deren Reihenfolge und Position zu erfassen, während die Methode der Radikale den Ansatz verfolgt, Zeichen in deren Radikale aufzuteilen und anhand dieser und deren Anordnung die Zeichen einzuordnen. Holistische Verfahren beruhen dagegen auf der Erfassung eines kompletten Zeichens, wodurch Schwierigkeiten vermieden werden, die entstehen, wenn Radikale sich berühren oder überschneiden und so nicht ohne weiteres einzeln erkannt werden.

Es stellt sich heraus, daß die holistischen Verfahren die besten Erkennungsraten vorweisen. Nachfolgend seien zwei Ansätze dieser Art kurz vorgestellt; zum genauen Verständnis ist ein Einblick in die aufgeführten Quellen ratsam.

1. Ein **holistisch basiertes Verfahren** aus [Liu u. Ding, 2005] kommt zu Erkennungsraten zwischen 98,5% und 99,5% (je nach Vergleichsdatenbank), das für jedes Bildpixel eines Zeichenbildes mittels Sobeloperatoren einen Gradientenvektor berechnet, und diesen Richtungsvektor quantisiert. Zur Quantisierung läßt sich die Umgebung des Pixels z.B. in acht Richtungen von jeweils 45° einteilen; ein Gradient der Stärke 3 in Richtung 75° entspricht somit einer Stärke von 1,1 von 45° und 2,1 von 90°. Diese so quantisierten Eigenschaftsvektoren werden dann positionsbezogen mit passenden Vorlagen abgeglichen. Mittels Fishers linearer Diskriminante läßt sich ein optimaler Klassifikator ermitteln, der einen möglichst großen Umfang an Testdaten von Klassen vorraussetzt, die alle eine multivariate Gaußverteilung aufweisen und eine einheitliche Kovarianzmatrix besitzen. In [Liu, 2006] stellt Cheng-Lin Liu mit der *discriminative feature extraction* eine Variante vor, die keine bestimmte Kovarianzmatrix einer Verteilung von Klassen vorraussetzt, und aus den Eigenschaftsvektoren einen optimalen Klassifikator errechnet; dieser liefert trotz einfacher Berechnung beachtliche Ergebnisse. Für Tests wurden dazu handschriftliche Zeichen aus zwei großen Datenbanken verwendet.

   Nei Kato et al. ([Kato et al., 1999]) stellen ein Verfahren vor, das alle Zeichen auf eine Größe von 64×64 Pixeln skaliert, und dieses Gitter in 49 sich überlappende Zonen von 16×16 Pixeln aufteilt. Statt des Zeichens selbst wird aus der binarisierten Form die Kontur ermittelt, die genau ein Pixel breit dargestellt wird. Für jedes Pixel der 64×64 Zeichenpixeln wird nun die achter-Nachbarschaft betrachtet, und soweit das aktuelle Pixel ein Konturpixel ist, der Verlauf der Kontur registriert. Dafür gibt es genau vier Möglichkeiten, horizontal, vertikal oder jeweils diagonal. Ändert sich der Konturverlauf im betrachteten

Pixel, werden für dieses Pixel zwei der vier Richtungen vorgemerkt. Für jede der 49 zuvor erwähnten Zonen wird der durchschnittliche Gradientenverlauf für die darin liegenden Pixel ermittelt, wobei der Gradient von zentral liegenden Pixeln höher gewichtet wird als der am Rand positionierter Pixel. Es ergibt sich ein Eigenschaftsvektor der Größe 196 für jedes Zeichen, durch 49×4 Dimensionen.

Kato verwendet vor dem genauen Zeichenabgleich eine grobe Klassifikation, um den späteren genauen Abgleich beschleunigen zu können. Dazu werden die zu untersuchenden Zeichen mit 3036 unterschiedlichen Zeichen, die in einer Datenbank vorliegen, mittels spezialisierter City-Block-Distanz abgeglichen, und nur die 30 ähnlichsten feiner klassifiziert, womit er bei Tests Erkennungsraten von 99,4% erreicht.

2. Da **radikalbasierte Systeme** geringere Erkennungsraten aufweisen, wird hier nur kurz ein mögliches Vorgehen verdeutlicht:

   Die Idee der radikalbasierten Systeme ist eine einfachere Erkennung von Radikalen, da die mögliche Anzahl unterschiedlicher Radikale signifikant niedriger ist als die möglicher Zeichen. Ein segmentiertes Zeichen wird dazu zunächst in sein Skelett überführt, in dem nach charakeristischen Merkmalen wie Kreuzungen gesucht wird. Die Hauptachsen der Skelette werden ermittelt und mittels Chamfer Distanz werden im Zeichen befindliche Radikale ermittelt. Ein Viterbi-Algorithmus erkennt darauf das hierarchisch aus einzelnen Radikalen bestehende komplette Zeichen.

3. **Strichbasierte Verfahren** Auch die strichbasierten Verfahren unterliegen der Annahme, daß die Erkennung einzelner Striche aufgrund ihrer geringen Anzahl einfacher sei. So gibt es nach [Schmidt, 1990] im wesentlichen nur 11 Striche, oder 24, wenn zusätzlich auch Erweiterungen dieser 11 Striche gezählt werden. Die vorgestellten Verfahren kategorisieren jeden einzelnen Strich in eine der vier Hauptachsen: Horizontal, vertikal oder jeweils diagonal. Teilweise wird darüber hinaus noch versucht, die Richtung zu bestimmen, etwa „von oben nach unten“. Gabor-Filter, pixel- oder konturbasierte Verfahren werden mitunter angewandt, um die Strichrichtung zu bestimmen. Mit Hilfe statistischer Modelle wird die Position und Anordnung der Striche abgeglichen, so können Wahrscheinlichkeiten für ganze Zeichen errechnet werden.

   Stricherkennung wird außerdem bei der Online-Handschrifterkennung eingesetzt; hier wird die Strichführung während des Schreibens der Zeichen erfaßt. Diese Art der Zeichenerkennung ist wesentlich einfacher, da nach [Schmidt, 1990] nicht nur die Richtung, in der ein Pinselstrich geführt wird, sondern auch die Reihenfolge, in der die Striche eines Zeichens geschrieben werden, per Konvention festgelegt ist. Die Online-Handschrifterkennung stellt somit ein gänzlich anderes Problemfeld dar, und bereits praktikable Lösungen die auch kommerziell eingesezt werden, etwa bei kleinen Mobilrechnern, lassen sich nicht direkt auf die hier beschriebene Offline-Erkennung übertragen.

In [Srihari et al., 2007] wird darüber hinaus auf Arbeiten verwiesen, die verschiedene Klassifikatoren kombinieren und damit die Erkennungsraten im Vergleich zu einfachen Klassifikatoren deutlich verbessern.

Erfolgreich ist darüber hinaus die Vorgehensweise, bei Unsicherheiten bezüglich der Segmentierung zu versuchen, verschiedene (aus Sicht der Segmentierung) mögliche Zeichen zu identifizieren, und Ergebnisse hieraus in den Segmentierungsalgorithmus einfließen zu lassen, um somit das Segmentierungsverfahren zu verbessern.

Erkennungsfehler einzelner Zeichen können aufgelöst werden, wenn stattdessen versucht wird, ganze Wörter zu identifizieren. Da es aber – wie bereits oben erwähnt – neben dem Zeichenabstand keinen expliziten Wortabstand gibt, sind diese Anwendungen bisher auf einzelne Bereiche beschränkt, wie etwa die Adresserkennung auf Briefen oder Worterkennung auf Bankschecks. Hier können Worterkennungsverfahren aufgrund der Vorkenntnisse, d.h. einer dort wesentlich geringeren Wortmenge, überzeugende Ergebnisse liefern.

Im Hinblick auf die Zielsetzung dieser Arbeit ist nicht nur der holistische Ansatz verwendbar, auch die radikalbasierte Methode kann Vorteile bieten. Da die Anzahl verschiedener Zeichen sehr groß ist, die der Radikale dagegen eher klein, ist es leichter, gleichartige Radikale zu finden als Zeichen. Man könnte also auf gleiche Art wie bei ganzen Zeichen einzelne Radikale auf typische Merkmale einer Handschrift untersuchen und dazu evtl. typische Auffälligkeiten der Anordnung dieser innerhalb ganzer Zeichen feststellen. Der strichbasierte Ansatz scheint dagegen weniger geeignet, da einzelne Striche zu wenige charakteristische Merkmale aufweisen.

2.3 Zeichenanalyse in der Paläografie

Stellvertretend sei hier kurz das Vorgehen von Dr. Richter aus [Richter, 2005] erläutert.

Da Originaldokumente zur Untersuchung nicht erreichbar sind, und stattdessen auf Reproduktionen zurückgegriffen werden muß, kann der Anpreßdruck der Schreibinstrumente nicht als Kriterium ermittelt werden, wie es bei forensischen Untersuchungen üblich ist. Dort kann zudem anhand der Schriftprobe die Stiftführung abgeleitet werden, um diese Schriftprobe einem Vergleichsmuster zuzuordnen, soweit so ein Vergleichsmuster vorhanden ist. Im Gegensatz zu Federn oder anderen festen Schreibinstrumenten der westlichen Welt wurden im altertümlichen China in der Regel weiche Pinsel benutzt. Diese lassen es oft zu, die vertikale Handbewegung durch die Strichführung zu ermitteln. Charakteristische Merkmale einer Handschrift finden sich üblicherweise bei der Untersuchung vieler gleichartiger Buchstaben, was sich bei chinesischen Handschriften aufgrund der hohen Anzahl verschiedener Zeichen und der damit verbundenen geringeren Häufung gleichartiger Zeichen als schwieriger gestaltet, als es bei Handschriften des lateinischen Alphabets der Fall ist. Ein Vergleich von Radikalen ist nur begrenzt möglich, da diese in unterschiedlichen Zeichen oft verschiedene Größe und Ausprägung haben.

Beispiele charakteristischer Merkmale werden anhand der *Laozi* Manuskripte gezeigt. Diese auf Bambusstreifen enthaltenen Texte wurden im kleinen Ort Guodian in der Nähe von Jingmen gefunden, etwa 200km nordwestlich der Millionenstadt Wuhan. Dort wurden viele altertümliche Grabstätten entdeckt, aus einem Grab konnten 1993 diese über 700 Bambusstreifen entnommen werden, die auf das späte vierte Jahrhundert v. Chr. datiert werden. Veröffentlicht wurden diese

Bambusstreifen in *Jingmenshi bowuguan: Guodian chumu zhujian - Beijing: Wenwu chubanshe, 1998*. Bezugnehmend auf die dort getroffene Unterteilung in Laozi A, B und C werden in Abbildung 2.3 die ersten zehn Streifen von Laozi A dargestellt.

Bemerkenswert ist hier die Handschrift der Streifen 5 und 6, die sich von der aller übrigen Streifen aus Laozi A unterscheidet. Ausschlaggebend für die Feststellung. daß Streifen 5 und 6 tatsächlich von einer anderen Person geschrieben wurden, ist die Kombination vieler Details, die alleine gesehen auch zufällig auftreten könnten.

Die Beschreibung der Auffälligkeiten im einzelnen:

- **Pinselstriche**: Striche der Zeichen in den Streifen 5 und 6 sind dynamischer ausgeführt als im übrigen Teil A. Kurven sind enger gezogen, die Striche sind breiter als auf den übrigen Streifen, genauer betrachtet scheint der Pinsel fester aufgesetzt, im Verlauf des Strichs aber etwas zurück genommen zu werden. Am Strichende wird der Pinsel flacher und ungenauer weggeführt, wodurch kurze horizontale Strichenden entstehen. In Abbildung 2.4 wird ein Beispiel von Zeichen gezeigt, die Handschrift B, die den Streifen 5 und 6 entspricht, mit der Handschrift A vergleicht, die bei allen übrigen Streifen in Laozi A erkannt wird.
- **Verschiedene Zeichenformen**: Von einzelnen Zeichen gibt es mehrere mögliche Schreibweisen. So wird das Zeichen „*wèi/wéi*“ in Laozi A in Typ 1 und 2 benutzt, in Laozi B in Typ 2 oder 3, in Laozi C ausschließlich in Typ 2. Das 11. Zeichen auf dem 6. Bambusstreifen ist dagegen vom Typ 4, der sonst nirgendwo in Laozi A-C nochmals gebraucht wird.
- **Schrägstellung der Pinselstriche**: Auffällig sind Schrägstellungen von Buchstaben im Verhältnis zur Grundlinie; einen Indikator für eine bestimmte Handschrift bieten diese Hinweise allerdings kaum, da sie durch die Position der Hand oder der Pinselhaltung zufällig beeinflußt werden können. Weniger auffällig sind Unterschiede der Strichanordnung innerhalb eines Zeichens; stellt man diese aber häufiger in gleicher Form fest, und kann diese klar von einer anderen Strichführung unterscheiden, läßt sich hiervon ein recht guter Hinweis auf unterschiedliche Handschriften herleiten. Beispiele dazu für Zeichen „*zhī*“ in Abbildung 2.5 oder für „*zhě*“ in Abbildung 2.6. Andere Zeichen haben dagegen durchgängig eine höhere Variabilität, so daß Abweichungen keine guten Hinweise liefern.

Soweit sich Veränderungen der Schreibweise nicht anderweitig begründen lassen, kann nur die Kombination verschiedener solcher Merkmale eine sichere Aussage über unterschiedliche Schreiber zulassen, wobei eher unbewußt entstandene Änderungen, wie ein unterschiedlicher Winkel von Strichen zueinander, höhere Bedeutung haben, als bewußt getroffene Entscheidungen, wie die über die Zeichenform.

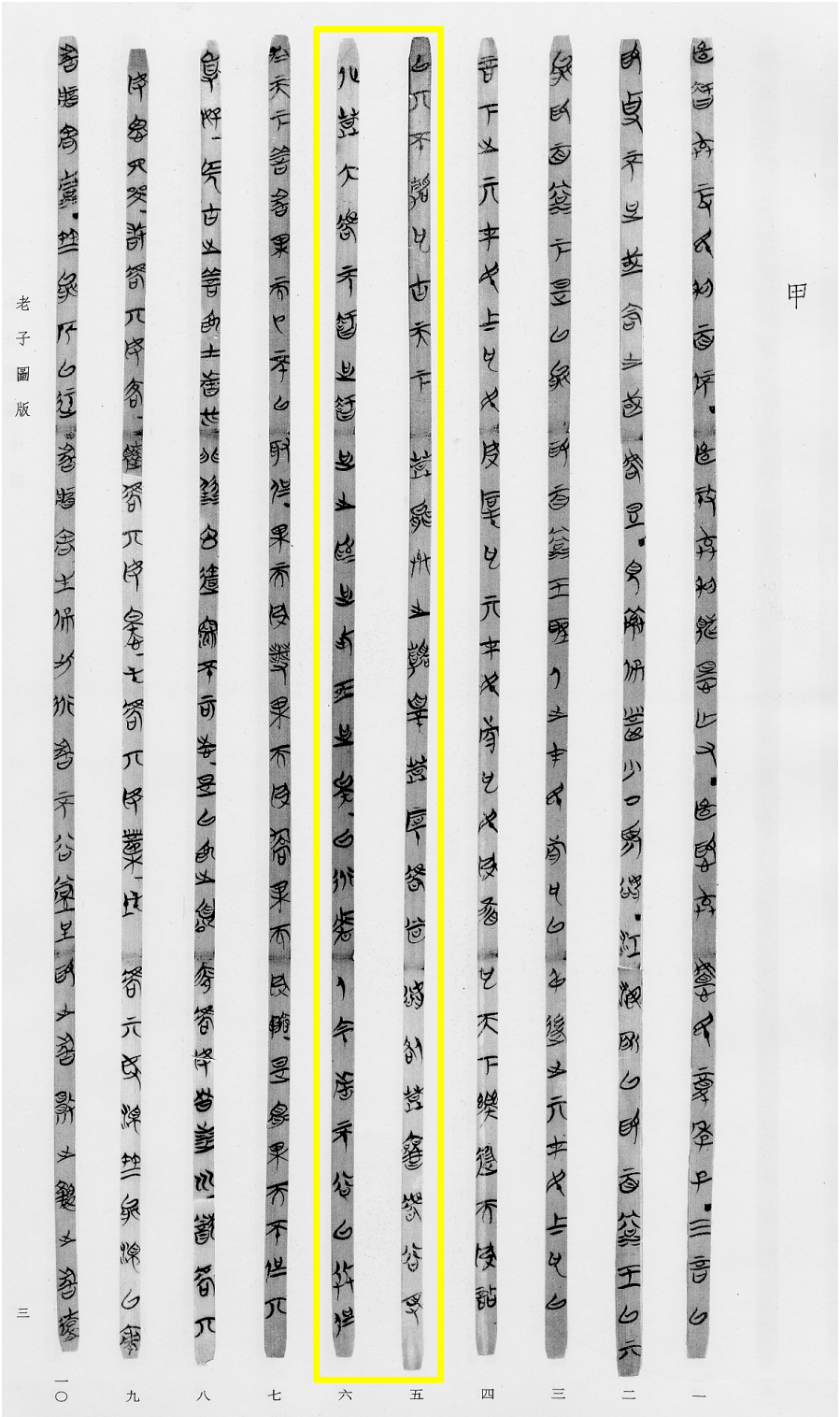
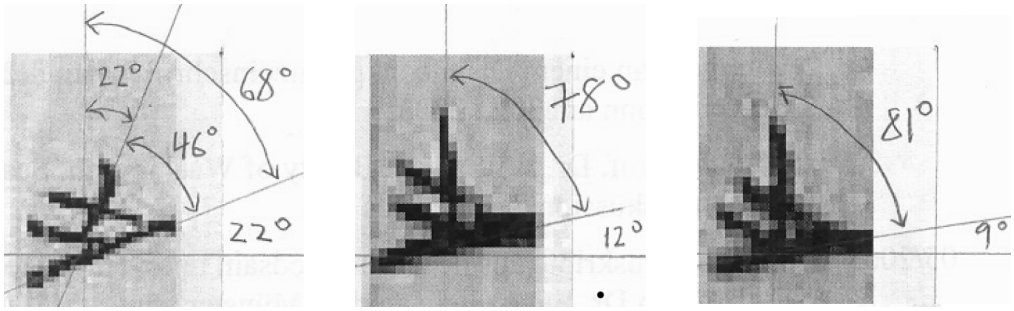


Abbildung 2.3: Die ersten zehn Bambusstreifen des Laozi Manuskripts, Teil A. Bei der Zählweise von rechts nach links wurde beim 5. und 6. Streifen eine andere Handschrift erkannt. Aus: [Richter, 2005]

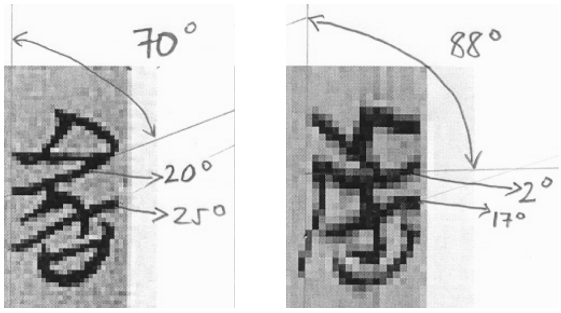
(a) Handschrift A **(b)** Handschrift B

Abbildung 2.4: Unterschiede der Pinselführung. Aus: [Richter, 2005]



(a) Streifen 27, Zeichen 16 **(b)** Streifen 5, Zeichen 12 **(c)** Streifen 6, Zeichen 10

Abbildung 2.5: Vergleich der Striche im Zeichen „*zhī*“.
(a): Handschrift A, (b) und (c): Handschrift B. Aus: [Richter, 2005]



(a) Streifen 7, Zeichen 5 **(b)** Streifen 6, Zeichen 22

Abbildung 2.6: Vergleich der Striche im Zeichen „*zhě*“.
(a): Handschrift A, (b): Handschrift B. Aus: [Richter, 2005]

3 Zeichensegmentierung

In diesem Kapitel wird die Segmentierung der Bilddokumente chinesischer Manuskripte in Bereiche, die einzelne Zeichen enthalten, beschrieben. Hierzu werden die zuvor geglätteten Bilder zunächst mittels Wasserscheidentransformation in Spalten aufgeteilt, welche darauf in Bereiche einzelner Zeichen unterteilt werden.

Für fast alle der hier und in den folgenden Kapiteln durchgeführten Bildoperationen wurde auf die Hilfe der VIGRA *([Köthe, 2000]) zurückgegriffen.* VIGRA *ist eine von Dr. Köthe in der Programmiersprache C++ entwickelte Bibliothek generischer Algorithmen, mit denen z.B. einzelne Pixeloperationen oder spezielle Filter auf Grauwert- oder Farbbilder angewendet werden können; dazu liefert die* VIGRA *die passenden Datentypen, die einen komfortablen Zugriff auf ausgewählte Bilddaten gewährleisten.*

Da sowohl die VIGRA *als auch das weiterhin verwendete* GEOMAP*-Rahmenwerk in den Programmiersprachen C++ und Python verfaßt sind, werden auch die Programmstrukturen dieser Arbeit in diesen Sprachen verfaßt. Dabei werden aufgrund des Geschwindigkeitsvorteils rechenintensive oder zeitkritische Bereiche in C++ geschrieben; die Ansteuerung bzw. die Hauptroutine ist dagegen in Python gehalten, da Python eine deutliche weniger aufwändige und interaktive Programmierung erlaubt.*

Als Grundlage für den Entwurf einer computergestützten Zeichensegmentierung dienten digitale Bildvorlagen, die vom Asien-Afrika Institut der Universität Hamburg sowie von Dr. Matthias Richter überreicht wurden. Ein Ausschnitt eines solchen Dokumentes zeigt Abbildung 3.1. Eine Segmentierung, die sich am Farbton orientiert, scheint hier nicht angebracht zu sein, da die Bildpixel, die den Vordergrund darstellen, also die Zeichen, nahezu schwarz sind. Der Hintergrund ist dagegen weitgehend einfarbig, in mittlerem bis hellem Braun gehalten. Farbbilder werden auf Computern üblicherweise in einem RGB-Farbraum repräsentiert, in dem jedes Bildpixel drei diskrete Werte zwischen 0 und 255 annehmen kann, mit denen Rot- Grün- und Blaukanal jeweils einzeln adressiert werden. Wie in Abbildung 3.1 zu erkennen ist, hat der Rotkanal die höchsten Werte bei den Hintergrundpixeln, die Werte der Vordergrundpixel liegen in allen Kanälen sehr nahe bei Null. Da alle vorliegenden Farbbildvorlagen diese Form haben, wird für die weitere Bearbeitung nur der Rotkanal verwendet, dieser bietet neben dem größten Abstand zwischen Vorder- und Hintergrund auch nur wenige Pixel mittlerer Helligkeit, die sich nicht eindeutig zuordnen lassen. Somit verspricht der Rotkanal, die Trennung am genauesten zu ermöglichen. Gegenüber den RGB-Bildern wird eine Datenreduktion erreicht, ohne wertvolle Informationen zu verlieren. Neben Farbdokumenten liegen auch Manuskriptausschnitte als Grauwertbilder vor, deren weitere Verarbeitung dann identisch zu der Bearbeitung der extrahierten Rotkanäle erfolgt.

(a) Rotkanal

(b) Grünkanal

(c) Blaukanal

Abbildung 3.1: Ausschnitt einer Bildvorlage eines chinesischen Manuskripts.

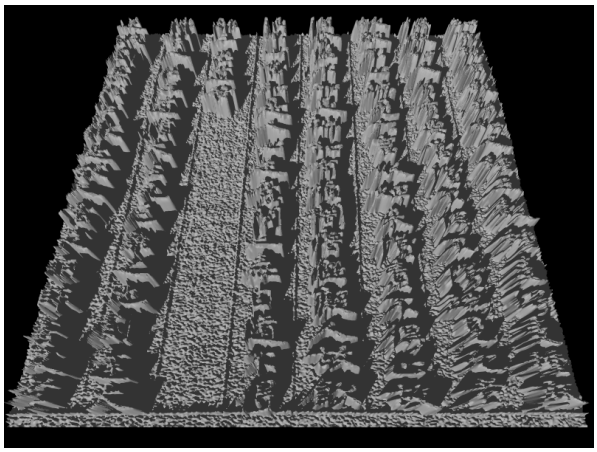


Abbildung 3.2:

Unterschiedliche Grauwerte der Abbildung 3.1a werden im Höhenprofil dargestellt.

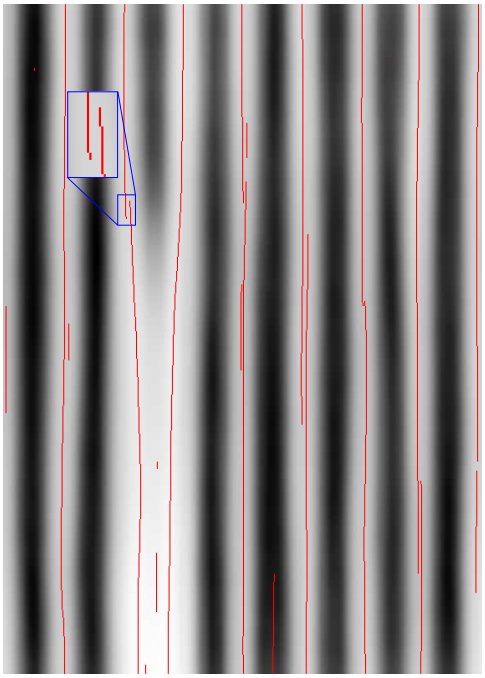
3.1 Segmentierung in Spalten

Alle vorliegenden Manuskripte wurden in traditioneller chinesischer Schreibweise verfaßt, was eine spaltenweise Anordnung aufeinanderfolgender Zeichen bedeutet. Man erkennt auf Abbildung 3.1 die Textspalten als annähernd vertikal verlaufende Streifen mit vielen dunklen Bildpixeln, die fast ausschließlich mit Pixeln der Zeichen übereinstimmen, dazwischen liegen Bildbereiche mit fast ausschließlich relativ hellen Hintergrundpixeln. Eine geringe Anzahl dunkler Pixel durch Verunreinigungen oder Striche scheint unerheblich. Abbildung 3.2 zeigt die unterschiedliche Grauwerte in diesem Bild als Höhenprofil, wobei Pixel mit niedrigeren Grauwerten höher dargestellt werden.

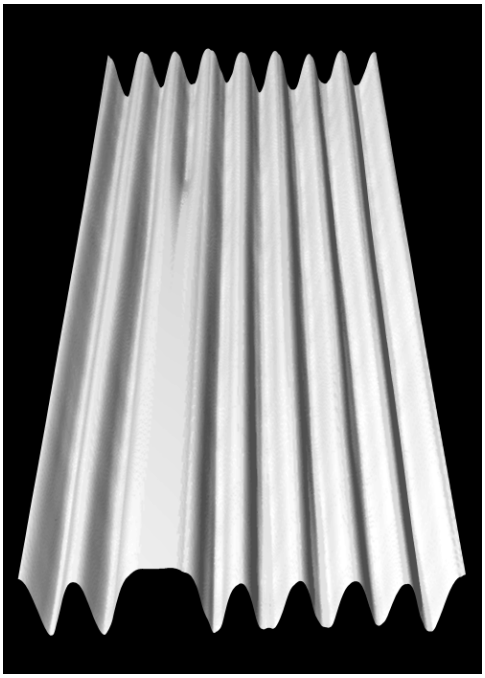
3.1.1 Glättung des Bildes

Wird das Bild tiefpaßgefiltert, wobei ein exponentieller, asymmetrischer Kern verwendet wird, der horizontal einem Zwölftel der Spaltenbreite und der einfachen Spaltenbreite vertikal entspricht, erhält man das Bild in Abbildung 3.3a. Das Grauwertgebirge hierzu zeigt Abbildung 3.3b, hellere Bildpixel werden hier höher dargestellt. Somit wird eine Glättung des Bildes bewirkt, die in Richtung der Textspalte stärker ausgeprägt ist als quer zur Spalte. Die Stärke der Glättung in Abhängigkeit der Spaltenbreite ermöglicht die Anwendung dieses Filters auf Bildvorlagen mit unterschiedlichen Spaltenbreiten, gemessen in Bildpixeln. Unterschiedliche Spaltenbreiten – gemessen in Bildpixeln – können selbst aus einem identischem Original bei Erfassung mit unterschiedlicher Bildauflösung entstehen. Die Spaltenbreite wird dabei berechnet aus dem Quotienten aus Bildbreite und der ermittelten Anzahl der Textspalten.

Geht man davon aus, daß die Spaltenbreite innerhalb einer Spalte nahezu konstant ist, und die Breite aller Spalten sowie aller Spaltenzwischenräume innerhalb eines Bildes annähernd gleich bleibt, kann man die Textspalten als periodisches Signal auffassen. Mittels einer zweidimensionalen diskreten Fouriertransformation kann das Amplitudenspektrum eines Bildes errechnet



(a) Tiefpaßgefiltertes Bild aus Abbildung 3.1a



(b) Grauwerte im Höhenprofil zu (a)

Abbildung 3.3: Anwendung eines asymmetrischen exponential-Tiefpaßfilters auf Abbildung 3.1a.

werden, in dem sich die Stärke dieser Frequenzen erkennen läßt. Bei dieser Bildvorlage befinden sich die beiden Maxima aller so ermittelten Frequenzen>0 vom Ursprung aus gesehen auf einer horizontalen Gerade im Radius 8, womit die ermittelte Spaltenanzahl dieses Ausschnittes 8 beträgt.

3.1.2 Segmentierungsverfahren

Die Segmentierung des Bildes in einzelne Spalten scheint durch die Glättung eindeutig zu sein, doch das einfache Vorgehen, im Bild zeilenweise lokale Maxima zu bestimmen, die jeweils der Mitte der Spaltenzwischenräume entsprechen sollten, liefert Bildpixel, die den roten Markierungen in Abbildung 3.3a entsprechen. Selbst bei dieser recht starken Glättung liegen die ermittelten Maxima eines Spaltenzwischenraumes in zwei aufeinanderfolgenden Zeilen teilweise einige Pixel auseinander; an anderen Stellen scheint die Glättung noch immer nicht stark genug zu sein, um zu verhindern, daß durch die Spaltentrennmarkierungen im Originalbild zwei Maxima in einem Zwischenraum erkannt werden.

Die oben erwähnten empirisch ermittelten Werte für den Filterkern stellen einen Kompromiß zwischen zu schwacher und zu starker Glättung dar. Eine zu schwache Glättung führt dazu,

daß viele unbedeutende lokale Maxima verbleiben, während eine zu starke horizontale Glättung Spalten verschmelzen könnte, insbesondere dann, wenn deren Abstand relativ gering ist. Eine zu starke vertikale Glättung kann zu den gleichen Problemen führen, da die Textspalten nicht immer genau vertikal ausgerichtet sind.

Um einzelne Spalten segmentieren zu können, sind diese ohne Unterbrechung voneinander zu trennen, was auf Grundlage dieses Verfahrens nur mit zusätzlicher, fehleranfälliger Heuristik möglich wäre, mit der nahe untereinander oder nebeneinander liegende Trennlinien miteinander verbunden werden könnten.

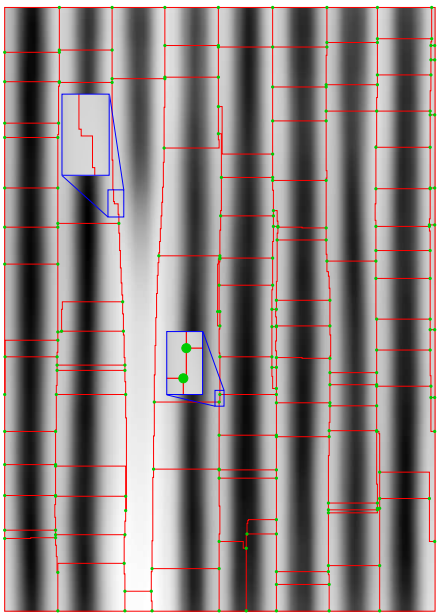
Eine bessere Möglichkeit zur Segmentierung wird durch das Wasserscheidenverfahren ermöglicht. In der Natur wird Wasser nach Regenfällen einen Weg in tiefergelegene Teiche oder Seen finden. Diese stellen in der Landschaft regionale Minima dar. Hügel oder Berge, die zwischen Seen liegen, bilden Wasserscheiden; Regenwasser wird an ihnen aufgeteilt und fließt entweder auf der einen oder anderen Seite in verschiedene Seen. An dieser Vorstellung ist die Wasserscheidensegmentierung ausgerichtet; Grauwertbilder können wie in Abbildung 3.3b dargestellt als Höhenprofil verstanden werden. Durch Auffüllen mit Wasser werden sich an den lokalen Minima Auffangbecken bilden, analog zu den Seen in der Natur. Lokale Maxima im Bild stellen Wasserscheiden dar, diese befinden sich genau an den Stellen im Bild, an denen sich jeweils zwei oder mehr Auffangbecken berühren, wenn entsprechend viel Wasser aufgefüllt wird, bis sie sich vereinen. Abbildung 3.4b zeigt einen so erhaltenen Zwischenstand.

Bemerkenswert ist das Ergebnis der Wasserscheidensegmentierung, die immer ununterbrochene Segmentierungslinien liefert, was einen klaren Vorteil bietet gegenüber der oben beschriebenen einfachen Methode. Nachteilig ist dagegen die oft recht starke Übersegmentierung, d.h. auch in sehr unbedeutenden lokalen Minima entstehen Auffangbecken, die zur Trennung von Flächen führen, deren Trennung nicht beabsichtigt ist. Das Ergebnis der Wasserscheidentransformation auf das geglättete Bild angewandt zeigt Abbildung 3.4a. Die Erkennung der lokalen Maxima erfolgt zweidimensional, so daß auch vertikale Trennlinien entstehen.

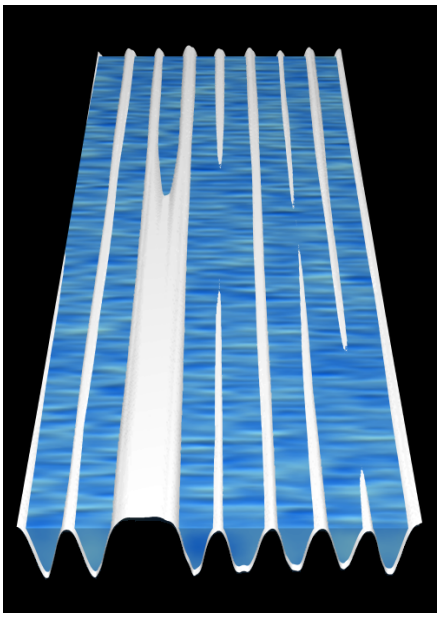
Bei dieser Segmentierung werden die Trennlinien nicht auf den Pixeln geführt, womit nicht festgelegt wäre, zu welcher Seite der Trennung die unter der Trennlinie befindlichen Pixel gehören, sondern zwischen den Pixeln. Dieses *crack edge* Verfahren wird in Abbildung 3.5 verdeutlicht. Durch Splineinterpolation ist es möglich, Bildwerte an beliebiger Position zwischen einzelnen Bildpixeln zu erhalten. Somit kann eindeutig bestimmt werden, auf welcher Seite der Pixel die Trennung erfolgen soll.

3.1.3 Bewertung der Trennkanten mit Hilfe der GeoMap

Die GEOMAP (siehe [Meine, 2003] und [Meine, 2007]) ist ein Rahmenwerk, mittels dessen geometrische und topologische Informationen vereint werden. Abbildung 3.4a zeigt eine Darstellung der GEOMAP, die auf das segmentierte Bild übertragen wurde. Durch den äußeren Rahmen werden alle Bildbereiche in endliche Flächen aufgeteilt. Kanten (rot dargestellt) berühren sich in Knoten (grün dargestellt). Eine Kante ist dabei die Verbindung von immer genau zwei



(a) Anwendung der Wasserscheiden-segmentierung auf Abbildung 3.1a



(b) Prinzip der Wasserscheidensegmentierung

Abbildung 3.4: Wasserscheidensegmentierung nach Anwendung eines asymmetrischen exponential-Tiefpaßfilters auf Abbildung 3.1a.

Knoten, die sich nur am Anfang und Ende der Kante befinden. Die GEOMAP ermöglicht es z.B. durch Datenstrukturen für Knoten, Kanten und Flächen sowohl Eigenschaften von diesen wie die Position, Größe oder durchschnittliche Farbe, als auch Nachbarschaftsbeziehungen zugreifbar zu machen, wie die Auflistung der beiden Flächen, die von einer Kante getrennt werden, oder alle eine Fläche umgebenden Kanten. Zudem werden Modifikationen wie die Vereinigung zweier Flächen in der Weise unterstützt, daß immer ein topologisch und geometrisch konsistenter Zustand gewährleistet wird.

Betrachtet man Abbildung 3.4a, so fällt auf, daß unerwünschte Kanten überwiegend horizontal und durch dunkle Bildbereiche verlaufen.

Normalisierung der Grauwerte

Wünschenswert ist eine Angleichung aller Pixelwerte, die Vordergrundbereiche darstellen, ebenso wie aller Werte der Hintergrundbildpixel zueinander, um möglichst enge und einheitliche Wertebereiche für Vorder- und Hintergrundpixel über der gesamten Bildbereich zu erhalten.

Um eine solche Normalisierung zu erreichen, wird eine Schwellwertberechnung mit dem Quad-Tree Verfahren kombiniert. Geht man davon aus, daß im Originalbild alle Bildpixel in zwei Klassen aufgeteilt werden können, d.h. aufgrund der Helligkeit der Pixel lassen sich die Klasse „Vordergrundpixel“ und „Hintergrundpixel“ trennen, dann ist ein optimaler Schwellwert

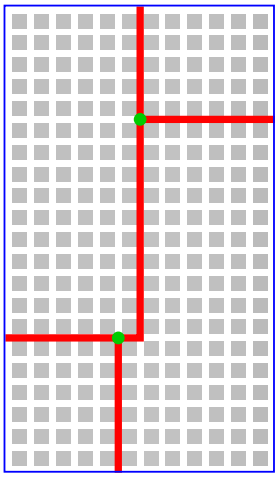


Abbildung 3.5:

Vergrößerung des Ausschnitts aus Abbildung 3.4a: Mit „crack edges“ werden Trennlinien zwischen Bildpixeln angeordnet.

der Wert, der alle Pixel mit (in diesem Fall) höherem Wert als Hintergrund bewertet, und alle übrigen der Vordergrundklasse zuordnet, so daß die Anzahl der dann falsch klassifizierten Pixel minimal ist. Ein in [Duda et al., 2001] beschriebenes Verfahren auf der Grundlage der Annahme, daß die beiden Klassen je eine Normalverteilung mit gleicher Varianz aufweisen, ermittelt als Schwellwert die Mitte zwischen den beiden Klassenmittelwerten. Da aber die Annahme gleicher Varianz der beiden Klassen nicht immer gegeben ist, wird in dieser Arbeit das Verfahren aus [Sonka et al., 1999, Algorithmus 5.2] verwendet, das für zwei Klassen auch dann einen akzeptablen Schwellwert ermittelt, wenn das Grauwerthistogramm keine bimodale Verteilung aufweist (Algorithmus 3.1). In der Regel terminiert der Algorithmus nach vier bis zehn Iterationen.

Algorithmus 3.1: Algorithmus zur Schwellwertermittlung
(aus [Sonka et al., 1999, Algorithmus 5.2])

Sei $f(x,y)$ eine Funktion, die zu jedem Pixel an der Stelle (x,y) den zugehörigen Pixelwert liefert.

Initialisiere im Schritt t_0 den Durchschnittswert der Hintergrundpixel $\mu_H^{t_0}$ mit dem Durchschnittswert der Pixel an den vier Bildecken, und den Durchschnittswert der Vordergrundpixel $\mu_V^{t_0}$ mit dem Durchschnittswert aller übrigen Pixel.

Initialisiere den Schwellwert $T^{t_0} = \frac{\mu_H^{t_0} + \mu_V^{t_0}}{2}$.

Wiederhole

$$\mu_V^t = \frac{\sum f(x,y) \mid f(x,y) < T^t}{\text{Anzahl Vordergrundpixel}}$$

$$\mu_H^t = \frac{\sum f(x,y) \mid f(x,y) >= T^t}{\text{Anzahl Hintergrundpixel}}$$

$$T^{t+1} = \frac{\mu_H^t + \mu_V^t}{2}$$

$$t = t + 1$$

Solange bis $(T^t = T^{t-1})$

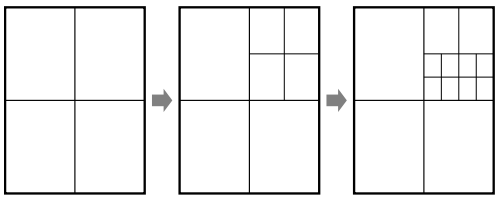


Abbildung 3.6: Prinzip des Quadtree-Algorithmus: Eine Fläche wird so lange rekursiv in vier Quadranten unterteilt, bis diese eine hinreichende Ähnlichkeit aufweisen.

Wurde ein solcher Schwellwert für das ganze Bild berechnet, wird das Bild darauf horizontal und vertikal jeweils mittig geteilt und auf gleiche Art Schwellwerte für die so entstandenen vier Quadranten ermittelt. Sollte das Quadrat der Differenz aus dem Schwellwert des gesamten Bildes und dem eines oder mehrerer Quadranten größer als 10 sein, werden alle Quadranten wiederum in vier Teile geteilt und deren Schwellwerte berechnet. Sollten die Abweichungen aller Quadranten dagegen innerhalb des Toleranzbereiches liegen, werden die hier ermittelten Schwellwerte verworfen und stattdessen der zuvor ermittelte Schwellwert des Gesamtbereichs verwendet. Dieser Vorgang wird rekursiv angewendet, bis entweder in allen Unterbereichen keine zu große Abweichung mehr erkannt wird oder eine Seitenlänge eines Quadranten kleiner ist als das 1,5-fache der Spaltenbreite. Es sollte somit in jedem Ausschnitt ein Teil eines Zeichens vorhanden sein, also die Bereiche immer Vorder- und Hintergrundpixel enthalten. Das Prinzip des Quadtree Algorithmus verdeutlicht Abbildung 3.6.

Würde man nun mit den ermittelten Schwellwerten eine Binarisierung durchführen, d.h. Pixel mit Grauwerten unterhalb des Schwellwertes schwarz und Pixel oberhalb weiß färben, gingen Bildinformationen verloren. Man kann in der Regel davon ausgehen, daß Pixel, die falsch klassifiziert werden, überwiegend Werte aufweisen, deren Abstand zum Schwellwert gering ist, während die Güte der Klassifizierung von Pixeln mit zum Schwellwert deutlich unterschiedlichen Werten wesentlich besser ist.

Um diesen Informationsverlust zu vermeiden wird auf das Bild eine Kontrastverstärkung angewendet mit folgender Funktion:

$$P_n = tanh(\phi \cdot \frac{P_o - T}{127.5} + 1) \cdot 127.5$$

P_n bezeichnet dabei den neu berechneten Pixelwert, P_o den Pixelwert im Ursprungsbild, T ist der für diesen Bereich ermittelte Schwellwert und ϕ eine Konstante, mit der die Stärke der Kontrastanpassung gesteuert werden kann. Genau wie das Intervall der Pixelwerte im Ursprungsbild werden Werte im Intervall $[0 - 255]$ ausgegeben. Für einen ermittelten Schwellwert von 138 und $\phi = 2$ ist diese Funktion in grüner Farbe in Abbildung 3.7b geplottet.

Da alle vorhandenen Vorlagen nur geringe Abweichungen der Vorder- und Hintergrundhelligkeit innerhalb eines Dokuments aufweisen und die Steigung der Anpassungsfunktion im gesamten Bereich endlich ist, konnten keine negativen Auswirkungen des Quadtree-Verfahrens, wie Helligkeitssprünge an den Seitenrändern einzelner Quadranten, festgestellt werden.

Abbildung 3.7 zeigt die Auswirkungen dieser Kontrastanpassung an einem Ausschnitt. Hier wurde ein Schwellwert von 138 erkannt. Aufgrund der Gleichmäßigkeit von Vorder- und Hintergrund wurde mittels Quadtree keine Unterteilung vorgenommen. Da in dieser Darstellung keine absolute Klassifizierung stattfindet, sind die Auswirkungen der Transformation für Pixel mit Werten in Schwellwertnähe gering.

3.1.4 Filtern unerwünschter Kanten

Nach diesen Grauwertanpassungen wird für jede Kante der GEOMAP das Minimum bestimmt, d.h. den dunkelsten Bildwert, den die Kante durchläuft. Auch über die Menge der Minima aller Kanten könnte mit Algorithmus 3.1 ein Schwellwert berechnet werden, der idealerweise alle Kanten, die zwischen zwei Textspalten verlaufen, von denen unterscheiden läßt, die sich über eine Textspalte erstrecken. In den meisten Bildvorlagen ist ein Löschen solcher Kanten problemlos möglich. Da es aber Manuskripte gibt, bei denen sich einzelne Zeichen zwischen zwei Spalten befinden, und dadurch – oder durch andere Störeinflüsse wie Flecken oder Risse auf der Schreibunterlage – auch die Kanten der Spaltentrennung durch dunklere Bereiche verlaufen, und diese durch den berechneten Schwellwert als Vordergrund klassifiziert würden, wird auf das Löschen von Kanten nur anhand eines globalen Helligkeitsschwellwertes verzichtet.

Die Orientierung von Kanten kann als weiteres Kriterium dienen. Die X-Komponente des Einheitsvektors aus Start- und Endpunkt einer Kante kann als Maß für die Horizontalität, die Y-Komponente für die Vertikalität dienen. Hierbei interessiert nur der absolute Betrag, der immer zwischen 0 und 1 liegt.

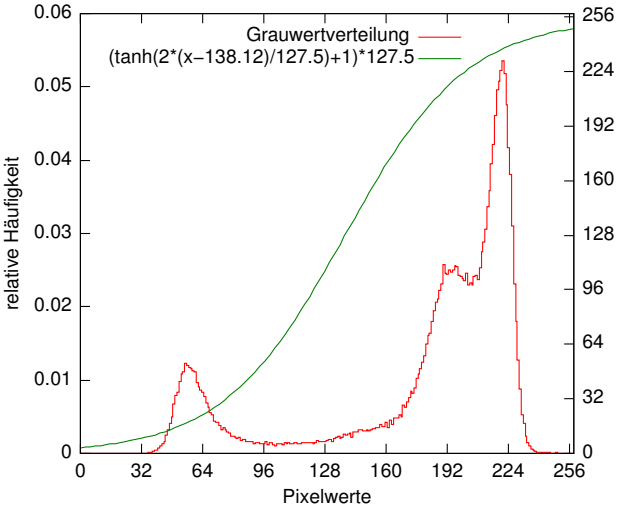
Abbildung 3.8 zeigt allerdings, daß die Orientierung alleine auch kein gutes Kriterium liefert, da durch das Entfernen der kurzen roten Kanten, die genau horizontal verlaufen, die notwendige Verbindung zweier Kantenstücke zur Spaltenbegrenzung verloren ginge.

Empirisch konnte festgestellt werden, daß in allen vorhandenen Bildvorlagen alle Kanten gelöscht werden konnten, deren horizontale Orientierung mindestens 0.9 beträgt und deren Länge gleichzeitig mindestens einem Drittel der durchschnittlichen Spaltenbreite entspricht, ohne dadurch eine zur Spaltentrennung notwendige Kante zu entfernen.

3.1.5 Vereinigung von Flächen

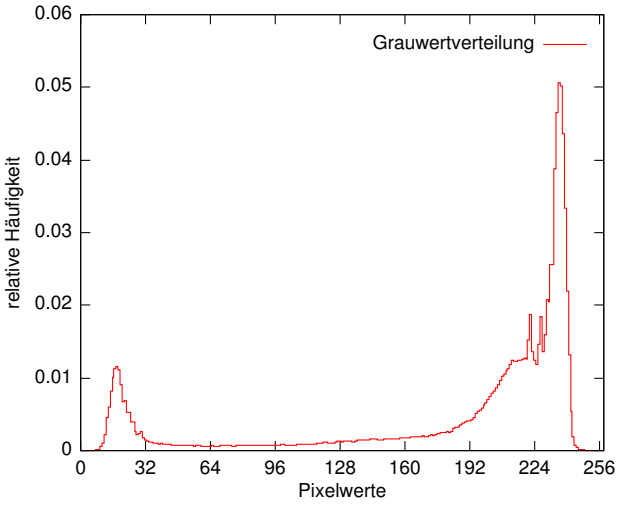
Nachdem einige Kanten wie beschrieben mit einfachen Mitteln entfernt werden konnten, werden als nächstes die Flächen betrachtet. Hier wird davon ausgegangen, daß Flächen, deren Größe höchstens ein Viertel des Produkts der Bildhöhe mit der durchschnittlichen Spaltenbreite entspricht, keine ausreichend großen Flächen darstellen, um eine Textspalte zu repräsentieren. Alle diese Flächen werden nach ihrer Größe sortiert, und ausgehend von der kleinsten Fläche wird für jede Fläche eine Kante gesucht, durch deren Entfernung zwei Flächen zu einer größeren vereint werden.

(a) Manuskriptausschnitt mit Original Grauwertverteilung



(b) Grauwerthistogramm (linke Ordinate) mit Anpassungskurve (rechte Ordinate)

(c) Manuskriptausschnitt nach veränderter Grauwertkurve



(d) Grauwerthistogramm nach Anpassung

Abbildung 3.7: (b) zeigt das Grauwerthistogramm des Auschnitts in (a). Mit dem errechneten Schwellwert von 138 werden die Grauwerte mit der grün dargestellten Funktion angepaßt. Es resultiert die Abbildung (c) mit zugehörigem Histogramm (d)

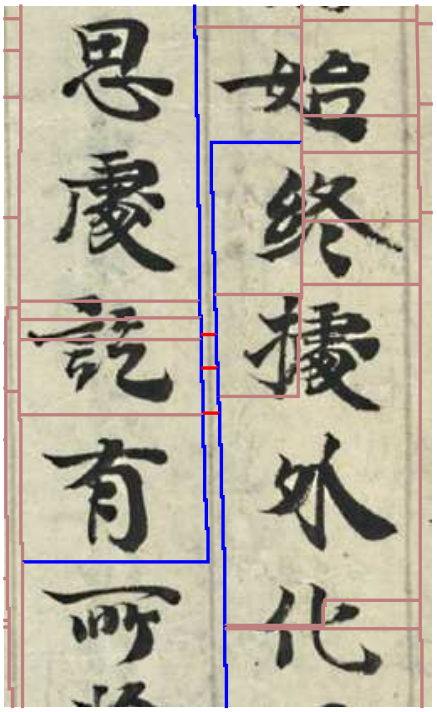


Abbildung 3.8:
Horizontal verlaufende Kanten sind bei der Segmentierung in Textspalten i.d.R. zu verwerfen. Die kurzen roten Kanten in der Bildmitte stellen allerdings das notwendige Bindeglied zwischen zwei Spaltentrennkanten dar; eine dieser Kanten muß erhalten bleiben.

Die GEOMAP ermöglicht es, Kanten mit Markierungen zu versehen. Alle Kanten der äußeren Bildbegrenzung werden als diese markiert und somit von einer Entfernung ausgeschlossen. Für jede andere Kante wird nun folgender Wert berechnet:

$$K = \frac{P_{\text{min}}}{100} + \frac{|L_X|}{H} - \frac{|L_Y|}{Sb} \cdot \tau$$

P_{min} beschreibt dabei den niedrigsten Pixelwert, der von der Kante durchlaufen wird. Gemessen wird dieser auf der geglätteten Vorlage, um Ausreisser abzuschwächen. L bezeichnet den Vektor vom Start- zum Endpunkt der Kante, L_X entspricht dabei der X-Komponente, normalisiert an der Bildhöhe H; L_Y der Y-Komponente dieses Vektors, normalisiert an der durchschnittlichen Spaltenbreite Sb. τ ist ein zusätzlicher konstanter Faktor, ein Wert von 0,75 hat sich hier bewährt.

Hohe Werte erhalten Kanten, die nur durch helle Bildbereiche verlaufen und vertikal ausgerichtet sind, niedrige Werte beschreiben Kanten, die dunkle Bildbereiche durchlaufen und horizontal verlaufen. Sehr kurze oder diagonal verlaufende Kanten werden überwiegend durch die Helligkeitswerte beschrieben, die sie durchlaufen, bei langen Kanten ist der Einfluß der Ausrichtung besonders stark, soweit sie horizontal oder vertikal verlaufen.

Von jeder Fläche, die nach obigen Kriterien als zu klein bewertet wurde, wird die Kante mit dem kleinsten Wert K entfernt, und somit zwei Flächen zu einer größeren vereinigt. Dieser Vorgang wird so lange wiederholt, bis keine Fläche mehr die Mindestgröße unterschreitet.

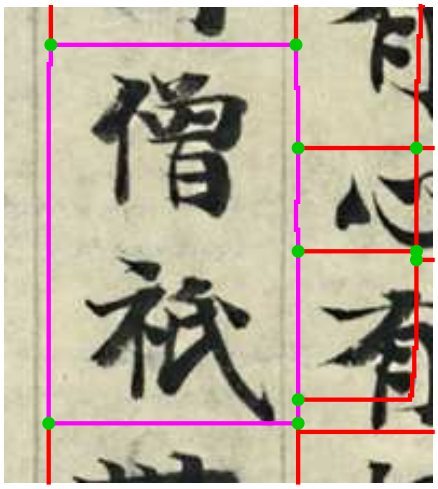


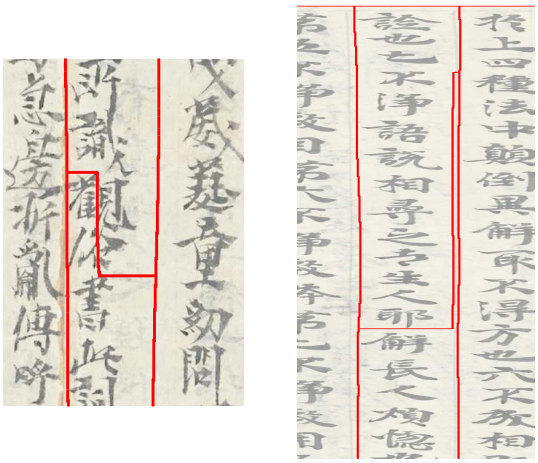
Abbildung 3.9:
Die linke Seite der in Magenta umrandeten Fläche hat nur eine Kante, während die rechte Seite von vier Kanten repräsentiert wird.

Betrachtet man also alle Kanten, die eine zu kleine Fläche begrenzen, so kann sichergestellt werden, daß eine dieser Kanten gelöscht werden muß. Würden stattdessen unter einzelnen Kanten, die auch für sich nach Kriterien wie Helligkeit der Bildbereiche in deren Verlauf, Orientierung oder Länge klassifiziert werden könnten, Kandidaten zur Löschung gesucht, wäre nicht nur die Bestimmung eines Schwellwertes notwendig, der festlegt, welche Kanten erhalten bleiben sollen; es wäre zudem wahrscheinlicher, daß Kanten entfernt werden würden, die erhalten bleiben müssen.

Das Löschen einer Kante bewirkt zudem eine Veränderung anderer Kanten. Ein Knoten, in dem drei Kanten enden, wird gelöscht, wenn eine der dort endenden Kanten entfernt wird, und die beiden dort verbleibenden Kanten werden vereint, was eine Änderung der Länge und oft auch der Ausrichtung der verbleibenden Kante bewirkt. Große Flächen werden häufig von vielen Kanten umgeben, die direkt hintereinander liegen und die gleiche Ausrichtung haben (Abbildung 3.9). Bei der Bewertung nach obigem Verfahren werden so andere Ergebnisse erzielt, als wenn diese Kanten als eine einzige lange Kante betrachtet würden. Durch die Vorsortierung und die Bewertung der Kanten der jeweils kleinsten Fläche wird dieser Effekt vermindert, da die jeweils relativ kleinsten Flächen von deutlich weniger Kanten umschlossen werden, und dort nur selten zwei oder mehr hintereinander liegende Kanten die gleiche Ausrichtung haben.

Die meisten unerwünschten Kanten können so entfernt werden, einige Kanten verbleiben allerdings, die mit obigem Verfahren nicht erfasst werden.

Da es durchaus möglich ist, daß eine Spalte deutlich schmaler als der Durchschnitt ist, oder eine Spalte nicht auf der gesamten Bildhöhe Zeichen enthält und die erkannte Spalte sich dadurch evtl. nicht über die gesamte Bildhöhe erstreckt, dürfen Kanten anhand dieser Merkmale nicht entfernt werden.



(a) „geteilte Spalte“ **(b)** „Doppelkante mit geteilter Spalte“

Abbildung 3.10:
Beispiele von Kantenverläufen, die erst vom Kriterium der vertikalen Korrelation erfaßt werden.

Vertikale Korrelation

Die Abbildung 3.10 stellt Probleme dar, die sich mit der Betrachtung einzelner Flächen oder Kanten nicht behandeln lassen. In einigen Dokumenten werden Spalten so in einen oberen und einen unteren Bereich geteilt, daß trennende Kanten durch helle Bildbereiche verlaufen und ihre Ausrichtung nicht überwiegend horizontal ist. Zudem ist keine Fläche unterhalb der Mindestgröße.

Da die Spalten nahezu vertikal verlaufen, kann als weiteres Kriterium die vertikale Übereinstimmung von Bildflächen darauf hindeuten, daß Flächen mit starker Übereinstimmung zu einer Spalte zusammengefaßt werden sollten. Dazu wird für jede Bildpixelspalte festgehalten, wie viele Pixel welcher Fläche diese enthält. Für jede Bildpixelspalte, deren Projektion auf eine horizontale Achse Bildpixel von mehr als einer Fläche enthält werden die Anzahl der Pixel aller darin enthaltenen Flächen miteinander multipliziert und für alle Pixel-Spalten werden diese Produkte für jede Fläche einzeln aufsummiert. Teilt man nun für jede Fläche die Summe der Produkte durch die Anzahl von Bildpixeln, die die jeweilige Fläche enthält, und normiert das Ergebnis über die Gesamtmenge von Bildpixeln einer Pixelspalte, erhält man die (asymmetrische) vertikale Korrelation einer Fläche mit einer anderen.

Dieses Vorgehen soll an folgendem Beispiel verdeutlicht werden: Die Abbildung 3.11 zeigt vier Flächen. Fläche 1 besteht aus 20 Pixeln, Fläche 2 aus 37, Fläche 3 aus 35 und Fläche 4 aus 28 Pixeln. Eine Pixelspalte enthält 12 Bildpixel. Die vertikale Korrelation der Flächen zueinander läßt sich wie folgt berechnen: Die Fläche 1 überschneidet sich nur in der 2. Pixelspalte mit Fläche 2 und 4. Sie enthält 8 Pixel der Fläche 1 und 3 der Fläche 2. Die Korrelation von Fläche 1 zu 2 entspricht daher $\frac{8 \cdot 3}{\text{(Anzahl Flächenpixel)} \cdot \text{(Pixel pro Spalte)}} = \frac{8 \cdot 3}{20 \cdot 12} = \frac{1}{10}$ oder 10%.

Die Korrelation von Fläche 1 zu 4 ergibt in gleicher Weise $\frac{8 \cdot 1}{20 \cdot 12} = \frac{1}{30}$ oder 3,33%.

Die Korrelation von Fläche 2 zu 1 ergibt dagegen $\frac{3 \cdot 8}{37 \cdot 12} = \frac{2}{37}$ oder 5,4%.

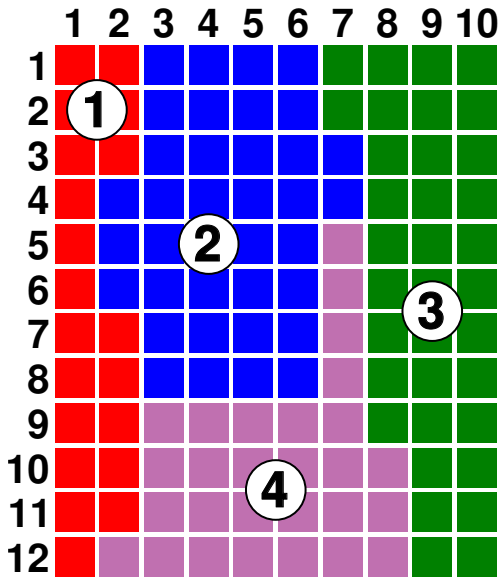


Abbildung 3.11:
Beispielgrafik zur vertikalen Korrelation von Flächen. Erläuterungen siehe Text.

Eine deutliche Korrelation berechnet sich aus dem Vergleich von Fläche 2 zu 4. Hier überschneiden sich 6 Pixelspalten, deren Werte aufsummiert werden: $\frac{(3 \cdot 1) + 4 \cdot (8 \cdot 4) + (2 \cdot 8)}{37 \cdot 12} = \frac{147}{444}$ oder 33,11%.
Umgekehrt ist das Ergebnis von Fläche 4 zu 2: $\frac{147}{28 \cdot 12}$ oder 43,75%.

Die übrigen Korrelationen können folgender Übersicht entnommen werden:

Korrelationen	Fläche 1	Fläche 2	Fläche 3	Fläche 4
zu Fläche 1	–	5,41%	0,00%	2,38%
zu Fläche 2	10,00%	–	0,95%	**43,75%**
zu Fläche 3	0,00%	0,90%	–	12,80%
zu Fläche 4	3,33%	**33,11%**	10,24%	–

Als Kriterium für die Vereinigung von Flächen wird nun die Summe der Korrelationen zueinander korrespondierender Flächen benutzt. Beträgt diese Größe mehr als 50, werden Flächen verschmolzen, was im Beispiel auf die Fläche 2 und 4 (33,11+43,75=76,86) zutreffen würde.

Mit diesem letzten, recht aufwändigen Verfahren werden nur noch wenige Kanten entfernt, die mit den vorher durchgeführten Verfahren nicht erfaßt werden konnten. Selbst bei Vorlagen in schlechter Qualität konnten Spalten eindeutig segmentiert werden, soweit durch die Fouriertransformation die korrekte Spaltenanzahl erkannt wurde. In Kapitel 5 werden dazu einige Beispiele vorgestellt.

3.2 Segmentierung der Zeichen

Nach der Auftrennung in Spalten ist die Auftrennung jeder einzelnen Spalte in Bereiche, die einzelne Zeichen enthalten, das nächste Ziel. Das Vorgehen ist ähnlich wie bei der Spaltenerkennung. Als Basis dient auch hier die Übersegmentierung der Wasserscheidentransformation auf einer geglätteten Vorlage; mit geeigneten Kriterien werden darauf unerwünschte Kanten gelöscht. Dieses Vorgehen hat sich als leichter erwiesen als in einer nichtsegmentierten Basis nach Segmentierungspfaden an geeigneter Position zu suchen.

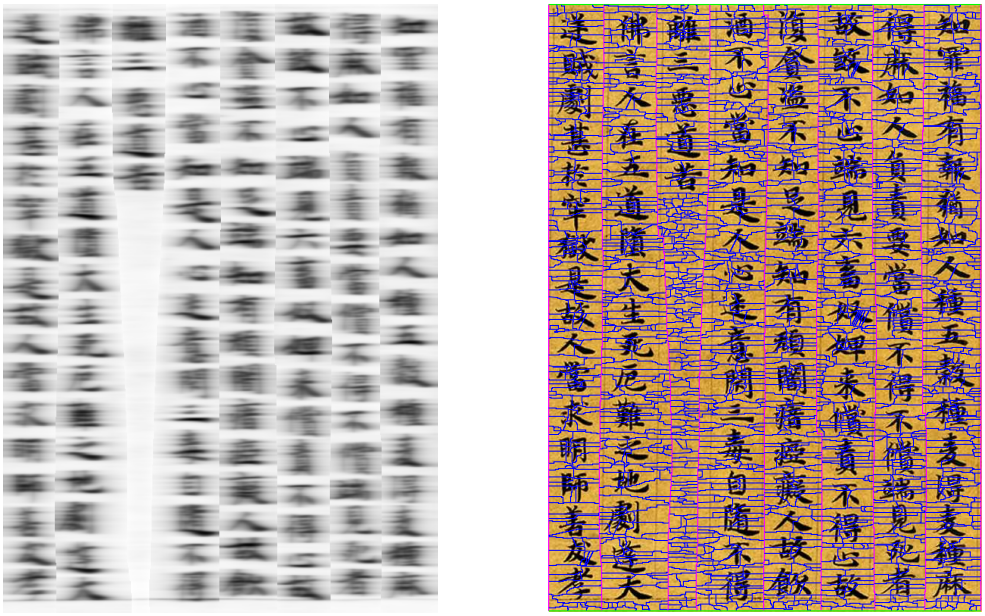
3.2.1 Glättung der Spalten

Die Stärke der Glättung wird von der durchschnittlichen Zeichenhöhe abhängig gemacht, wobei davon ausgegangen wird, daß es keine Zeichen gibt, die innerhalb einer Spalte nebeneinanderstehen. Somit bildet wiederum die mittels Fouriertransformation berechnete Zeichenfrequenz dividiert durch die Spaltenhöhe die durchschnittliche Zeichenhöhe. Aufgrund von unterschiedlichen und teilweise sehr geringen Zeichenabständen ergibt sich im Powerspektrum teilweise ein absolutes Maximum für alle Frequenzen>1, welches nicht der Zeichenfrequenz entspricht. In Schmidt [1990] wird erwähnt, daß als generelle Raumvorgabe für chinesische Zeichen ein Quadrat vorgesehen ist, das möglichst gleichmäßig und symmetrisch ausgefüllt werden sollte. Aus diesem Zusammenhang aus Zeichenbreite und Zeichenhöhe wird im Powerspektrum der Bereich, in dem ein lokales Maximum bestimmt wird, so eingeschränkt, daß die sich daraus ergebende durchschnittliche Zeichenhöhe nicht größer ist als die zuvor bestimmte durchschnittliche Spaltenbreite (die der Zeichenbreite inklusive dem Spaltenzwischenraum entspricht), und nicht kleiner als einem Viertel der durchschnittlichen Spaltenbreite sein kann. Diese Einschränkung und Wahl der Grenzen hat sich bei allen vorliegenden Manuskriptproben bewährt.

Die Bildglättung wird darauf durch eine Faltung mit exponentiellem Kern erreicht, der in der horizontalen einem Viertel und in der vertikalen einem Fünfzigstel der durchschnittlichen Zeichenhöhe entspricht. Diese Glättung wird auf jede Spalte einzeln angewendet, anschließend werden die geglätteten Spalten wieder zu einem Bild zusammengesetzt. Somit können benachbarte Spalten sich nicht gegenseitig beeinflussen. Das Ergebnis dieser Glättung zeigt Abbildung 3.12a, eine darauf angewendete Wasserscheidentransformation zeigt die Darstellung der GEOMAP mit hinterlegtem Originalbild in Abbildung 3.12b.

Das Ziel für nachfolgendes Vorgehen ist, die Fläche der Spalte durch Kanten so aufzuteilen, daß sich in jeder Fläche ein ganzes Zeichen befindet. Größere Zeichenzwischenräume dürfen dagegen auch durch mehrere Kanten getrennt werden; Flächen, die dann kein Zeichen enthalten, werden durch die darauf folgenden Verfahren ignoriert.

Durch die vertikale Glättung kann zwischen größeren und kleineren Zwischenräumen zwischen Zeichen oder Teilen von Zeichen unterschieden werden, die im Originalbild gleiche Helligkeit aufweisen, da kleinere Zwischenräume in der geglätteten Darstellung dunkler erscheinen als größere, und somit darin verlaufende Kanten entsprechend bewertet werden. Zwischenräume



(a) Zusammengesetztes Bild aus einzeln geglätteten Spalten.

(b) GEOMAP-Darstellung nach Wasserscheidentransformation auf (a)

Abbildung 3.12: Nach Auftrennung in Spalten werden diese einzeln geglättet und anschließend wieder zusammengesetzt. Darauf wird eine Wasserscheidentransformation angesetzt.

zwischen zwei Zeichen sollten dabei immer größer sein als helle Flächen innerhalb eines Zeichens. Eine zu starke Glättung läßt Zeichen mit sehr geringem Abstand ineinander verschmelzen.

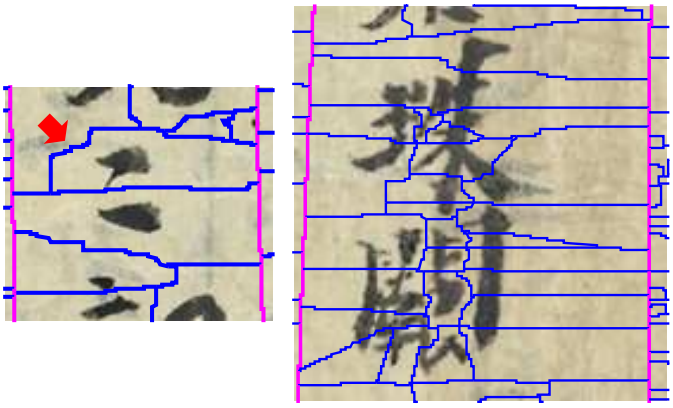
Auch die Stärke der horizontalen Glättung darf nicht zu groß sein, denn die genaue Trennung von Zeichen, von denen sich Teile auf gleicher Pixelzeilenhöhe befinden, wird dann deutlich schwieriger oder unmöglich; eine zu schwache horizontale Glättung kann dagegen besonders bei geringer Zeichenbreite im Verhältnis zur Spaltenbreite dazu führen, daß zeichentrennende Kanten sich nicht über die gesamte Spaltenbreite erstrecken, eine Trennung auf Basis des entwickelten Verfahrens wird so erschwert (Abbildung 3.13).

3.2.2 Zusammenführen von Flächen, die sich nicht über die Spaltenbreite erstrecken

An dieser Stelle wird davon ausgegangen, daß durch die vorhergehende Spaltensegmentierung die Vorlage so aufgeteilt wurde, daß die Zeichen in einer Spalte strikt untereinander angeordnet sind, und sich durch überwiegend horizontal verlaufende Kanten trennen lassen. Das bedeutet, daß Flächen, die einzelne Zeichen enthalten, beidseitig die Spaltenbegrenzungskanten berühren

Abbildung 3.13:

Während in (a) eine stärkere horizontale Glättung die durch den Pfeil markierte Kante horizontal an die Spaltentrennkante geführt hätte, würde eine stärkere horizontale oder vertikale Glättung keine Kante zwischen den Zeichen in (b) ergeben, wodurch keine Trennung der Zeichen möglich wäre.



(a) Eine stärkere Glättung wäre hier vorteilhaft

(b) Eine stärkere Glättung hätte ein Verschmelzen der Zeichen ohne Trennkante zur Folge.

sollten. In Abbildung 3.12b werden die Spaltenbegrenzungskanten magentafarben dargestellt und in der GEOMAP als solche markiert. Damit sind sie bei den folgenden Verfahren vor Löschvorgängen geschützt.

Initial wird als **1. Kriterium** der Ansatz verfolgt, alle Flächen, die nicht beidseitig die Spaltenbegrenzungen berühren, so mit benachbarten Flächen zu vereinen, bis alle Flächen diese Anforderung erfüllen.

Dazu wird für Kanten von allen Flächen, die nicht beidseitig an die Spaltenbegrenzungen heranreichen, folgender Wert berechnet:

$$K \;=\; P_{\min} - \frac{|L_Y|}{Zh} \cdot \rho$$

$P_{\min}$ entspricht dem niedrigstem Helligkeitswert im Verlauf dieser Kante, L_Y der Y-Komponente des Vektors vom Start- zum Endpunkt der Kante, normiert an der durchschnittlichen Zeichenhöhe Zh, die wie zuvor unter Abschnitt 3.2.1 beschrieben berechnet wurde. ρ schließlich ist ein Faktor, der einem Zehntel des Helligkeitsminimums aller Kanten entspricht, die die jeweilige Fläche begrenzen.

In der Praxis bedeutet das, daß Flächen, deren Kanten ein Zeichen schneiden, in diesem Bereich dunkle Bildbereiche durchlaufen[1]. Bei diesen Flächen werden solche Kanten gelöscht, die das Zeichen durchlaufen. Bei anderen Flächen in ausschließlich hellen Bildbereichen ist

[1] Durch Anwendung des CrackEdge-Verfahrens entsprechen die Helligkeitswerte bestimmter Kantenpositionen keinen direkten Pixelwerten. Statt dessen werden die Helligkeitswerte durch Interpolation der Werte benachbarter Bildpixel berechnet.

das Kriterium der Helligkeit weniger relevant, stattdessen soll hier die am stärksten vertikal verlaufende Kante entfernt werden.

Alle ausgewählten Flächen werden nun nach dem Wert K der Kante mit dem jeweils niedrigstem K sortiert, und angefangen bei der Fläche mit der Kante mit niedrigstem K wird für jede Fläche diese eine Kante entfernt.

Da auch hier das Löschen einer Kante immer zwei Flächen beeinflußt, muß diese Sortierung ständig aktualisiert gehalten werden, da nach dem Vereinen zweier Flächen in der GEOMAP eine dieser Flächen nicht mehr addressierbar ist. Dieses kann beim Löschen einer Kante von Fläche (a) die Fläche (b) auf der anderen Seite der Kante sein, falls deren Inhalt zur soeben behandelten Fläche zugewertet wird. Es könnte aber auch die Adresse der soeben behandelten Fläche (a) gelöscht worden sein, und die Fläche (b) könnte so erweitert worden sein, daß sie nun beidseitig die Spaltenbegrenzungen berührt, ohne daß die Fläche (b) explizit behandelt wurde. In beiden Fällen darf nicht zusätzlich eine Kante der Fläche (b) entfernt werden.

Sind keine Flächen mehr vorhanden, die sich nicht über die gesamte Spaltenbreite erstrecken, haben alle verbleibenden Flächen neben den Kanten, die der Spaltenbegrenzung entsprechen, jeweils nur noch eine obere und eine untere Kante. Nur diese beiden Kanten können durch weitere Löschoperationen entfernt werden, ausgenommen davon sind die oberste und unterste Kante jeder Spalte, da diese als Vorlagenbegrenzung in der GEOMAP als solche markiert wurden und damit von Löschvorgängen geschützt sind.

3.2.3 Löschen aller verbleibenden Kanten, die ein Zeichen schneiden

Das **2. Kriterium** versucht nun die verbleibenden Flächen so zu vereinen, daß keine Kante mehr ein Zeichen schneidet. In Experimenten konnte festgestellt werden, daß kein Zeichen von einer Fläche umschlossen werden kann, die niedriger als ein Drittel der durchschnittlichen Zeichenhöhe ist. Von allen diesen Flächen wird jeweils die Kante gelöscht, die die dunkleren Bildbereiche durchläuft.

Alle übrigen Flächen werden nach ihrer durchschnittlichen Höhe sortiert, und für die beiden horizontal begrenzenden Kanten wird das Minimum der durchlaufenden Helligkeitswerte festgestellt. Über diese Minima wird ein Schwellwert Sw mit Hilfe von Algorithmus 3.1 berechnet. (Im folgenden wird die Kante von den beiden eine Fläche horizontal begrenzenden Kanten, die den niedrigsten Helligkeitswert durchläuft, als Kante (a) bezeichnet.)

Für jede Fläche wird nun wieder folgender Wert berechnet:

$$K = -\frac{P_{\min} - Sw}{100} + \eta - \frac{Fh_u}{Zh} - \frac{Fh_u + Fh_v}{2 \cdot Zh} \cdot \theta$$

$P_{\min}$ ist dabei das Helligkeitsminimum der Kante (a). η ist eine Konstante, die bezüglich des Kriteriums die Helligkeit im Kantenverlauf zur Höhe der Fläche gewichtet, Fh_u bezeichnet

die durchschnittliche Höhe der Fläche, diese wird mit der durchschnittlichen Zeichenhöhe Zh normiert, Fh_v beschreibt die durchschnittliche Höhe der Fläche auf der gegenüberliegenden Seite der Kante (a), θ ist schließlich eine Konstante zur Gewichtung der Werte von einfacher Flächenhöhe zur Summe der Flächenhöhen beider Flächen an Kante (a).

Experimentell haben sich für η ein Wert von 1,35 und für θ ein Wert von 0,57 als geeignet erwiesen.

Sollte der Wert K größer als 0 sein, wird die Kante (a) gelöscht.

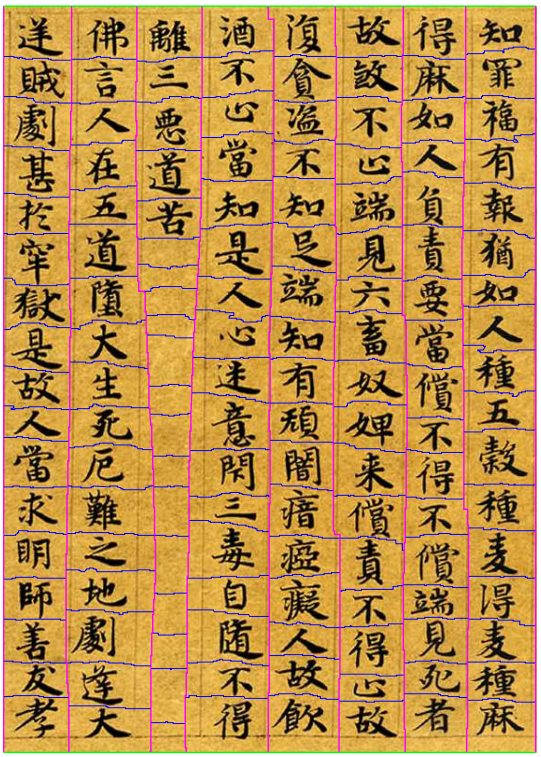
Ein Löschen dieser Kante wird also wahrscheinlicher, umso dunkler der durchlaufene Bildbereich ist, und umso kleiner die Höhe der untersuchten Fläche ist. Außerdem wird die Höhe der Fläche miteinbezogen, die entsteht, wenn die Kante (a) entfernt würde; diese entspricht der Summe der durchschnittlichen Höhen der Flächen u und v beiderseits von Kante (a).

So werden von allen Flächen entsprechende Kanten entfernt, bis keine Kante mehr einen Wert kleiner oder gleich Null aufweist.

Abbildung 3.14 stellt die Ergebnisse nach Anwendung von Kriterien 1 und 2 gegenüber, Ergebnisse weiterer Vorlagen werden in Kapitel 5 präsentiert.

(a) Initiale Wasserscheidensegmentierung

(b) (a) nach Anwendung von Kriterium 1



(c) (b) nach Anwendung von Kriterium 2

Abbildung 3.14: Während die initiale Wasserscheidensegmentierung eine deutliche Übersegmentierung zeigt, kann nach Anwendung beider Kriterien die Segmentierung so verbessert werden, daß jedes Zeichen genau einer Fläche zugeordnet werden kann.

4 Shape Context

Dieses Kapitel beschreibt den Vergleich einzelner Zeichen untereinander. Dazu wird jede Fläche, die durch die im Kapitel 3 beschriebene Segmentierung definiert wurde auf ein darin enthaltenes Zeichen untersucht, und – soweit vorhanden – dieses Zeichen segmentiert. Die Kontur jedes Zeichens wird durch darauf angeordnete Punkte beschrieben, deren relative Positionierung zueinander in einem Satz von SHAPE CONTEXTen abgelegt wird. Vergleiche verschiedener Zeichen können nun über den Vergleich dieser SHAPE CONTEXTe erfolgen.

Erweiterungen der Vergleichskriterien, wie die Berücksichtigung des Konturverlaufs in jedem Punkt oder eine affine Transformation der Zeichen aufeinander, bieten Möglichkeiten, den Zeichenvergleich noch zu optimieren. Sie werden in diesem Kapitel kurz vorgestellt, im Kapitel 5 dann näher untersucht.

Bevor in den durch die in Kapitel 3 beschriebene Spalten- und Zeichensegmentierung erstellten Flächen Zeichen segmentiert werden können, wird zunächst festgestellt, ob diese Zeichen enthalten. Dazu wird auf die Schwellwerte zurückgegriffen, die bereits (wie unter Abschnitt 3.1.3 beschrieben) berechnet wurden. Während eine Binarisierung dort keine optimale Grundlage bildete, wird die Bildvorlage nun anhand dieser Schwellwerte binarisiert, so daß Bildelemente von Schriftzeichen weiß, andere Bildelemente schwarz dargestellt werden. Sollte eine Fläche weniger weiße Pixel enthalten als der Wert, der drei Viertel der durchschnittlichen Zeichenhöhe entspricht, wird davon ausgegangen, daß die Fläche kein Zeichen enthält und sie wird nicht weiter verwendet. Experimentell konnten mit dieser Größe zuverlässige Ergebnisse ermittelt werden.

Die durchschnittliche Zeichenhöhe, die wie unter Abschnitt 3.2.1 beschrieben berechnet wurde, entspricht dabei der Höhe der Spalte in Pixeln geteilt durch die Anzahl von Zeichen in dieser Spalte, und schließt somit den Zeichenzwischenraum mit ein. Auf die Hinzunahme eines Maßes für die Breite der Fläche wurde dabei bewußt verzichtet, da die Spalte in Einzelfällen wesentlich breiter sein kann als das Zeichen.

4.1 Subpixelgenaue Segmentierung der Zeichen

Zur subpixelgenauen Segmentierung der Zeichen innerhalb jeder verbleibenden Fläche wird diese auf das Neunfache ihrer Originalgröße vergrößert. Dabei werden Werte für die zusätzlich vorhandenen Pixel durch die Splineinterpolationsfunktion der VIGRA berechnet. Wird über diese

Fläche durch Algorithmus 3.1 ein Schwellwert berechnet und dieser Ausschnitt binarisiert, können wesentlich genauere Ergebnisse erzielt werden, als auf einem Ausschnitt in Originalgröße (Abbildung 4.1, Abbildung 4.2).

Die binarisierte Darstellung erlaubt ein Labeling der Flächen, d.h. benachbarte Pixel, die im Binärbild als Vordergrundpixel dargestellt werden, werden zu einer Fläche zusammengefaßt. Als Nachbarschaftsbeziehung wird dabei für die in Weiß dargestellten Vordergrundpixel die Achternachbarschaft gewählt, in der neben horizontal und vertikal auch diagonal angrenzende Pixel zugerechnet werden; für den Hintergrund gilt die Vierernachbarschaft.

Für alle so entstandenen Vordergrundflächen wird die Anzahl der zugehörigen Pixel festgestellt. Um Störpixel oder Bildrauschen auszufiltern, die Flächen mit sehr geringen Ausmaßen ergeben, werden Flächen gelöscht, die weniger Pixel beinhalten, als es der quadrierten Anzahl eines Drittels der durchschnittlichen Zeichenhöhe entspricht.

4.2 Zeichenvergleich durch Shape Context

4.2.1 Auswahl geeigneter Vergleichskriterien

Um Zeichen miteinander vergleichen zu können, ist es nötig, die Bildinformationen auf ein geeignetes Vergleichsmaß zu reduzieren. Dazu müssen Informationen erkannt werden, die invariant in Bezug auf nicht relevante Eigenschaften oder Einflüsse sein sollten. So ist die Größe einer Zeichenfläche, gemessen in Bildpixeln, abhängig von der Auflösung, mit der die Vorlage eingescannt wird und bietet in direkter Form keine wertvolle Information; auch die Lage eines Zeichens innerhalb eines Dokuments erscheint zunächst unerheblich.

Während ein Vergleich verschiedener Zeichen aufgrund von Farb- oder Helligkeitsinformationen wenig sinnvoll erscheint, bietet die Form gute Anhaltspunkte. Interessant zur Unterscheidung sind Maße wie die Exzentrizität, d.h. das Verhältnis von der Höhe zur Breite der Zeichen. Gegen dieses Maß spricht allerdings der Umstand, daß die Vorgabe für chinesische Schriftzeichen möglichst gleiche Höhe und Breite vorsieht. Auch der Umfang im Verhältnis zur Größe kann eine geeignete Vergleichsbasis darstellen, oder zentrale Momente zweiter oder dritter Ordnung. Der Informationsgehalt wird allerdings für zu gering erachtet, da es viele ähnlich aussehende Zeichen gibt. Die Anzahl der Löcher, die Hintergrundpixel innerhalb eines Zeichens einschließen, kann bei ungenauer Schreibweise bei gleichen Zeichen variieren; Striche, die sich bei gleichen Zeichen je nach Schreibgenauigkeit berühren oder nicht, liefern eine unterschiedliche Anzahl an Regionen, so daß auch dieser Vergleich schlechte Ergebnisse liefert..

Geeignet zur Charakterisierung scheint das Skelett eines Zeichens, d.h. die Rückführung der einzelnen Pinselstriche auf einfache Linien, die zentriert im Pinselstrich geführt werden. Vorstellbar wäre hier, in ähnlicher Weise wie von Dr. Richter bereits von Hand ausgeführt und in Abbildungen 2.5 und 2.6 dargestellt, die Winkel einzelner Striche zu messen, sie untereinander zu vergleichen, die Anzahl und Position von Knotenpunkten festzustellen oder die Länge der Striche zwischen den Knoten.

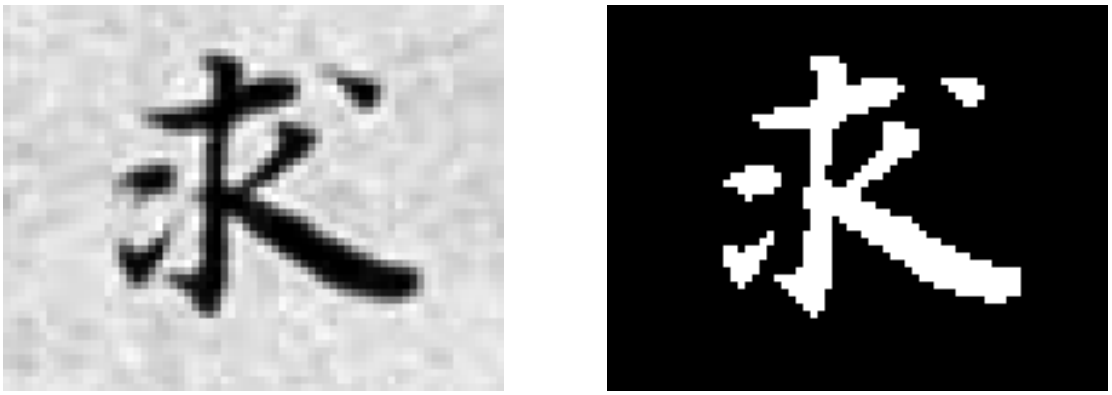
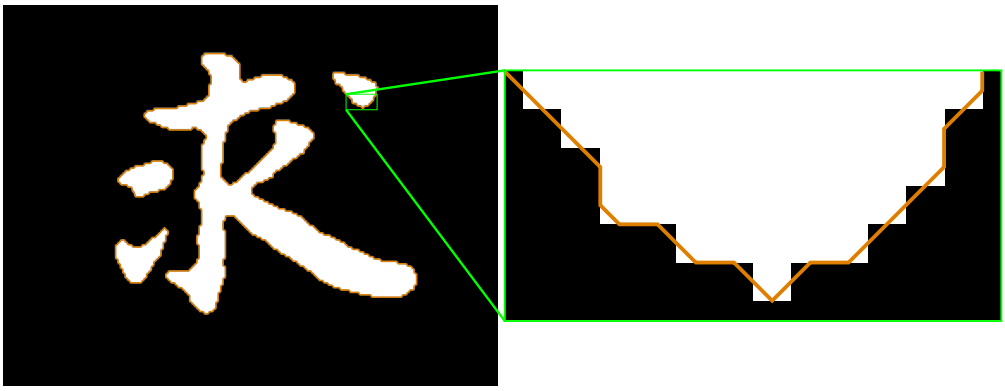


Abbildung 4.1: In der binarisierten Darstellung der Originalvorlage gehen viele Details verloren.

Abbildung 4.2:

Rechts die Vergrößerung des Originals auf die neunfache Anzahl von Bildpixeln; mittels Splines werden fehlende Pixelwerte interpoliert.

Unten dargestellt ist das Ergebnis der Binarisierung dieser interpolierten Vorlage; VIGRA-Funktionen erlauben das Erfassen der Kontur als detailreiches Polygon.



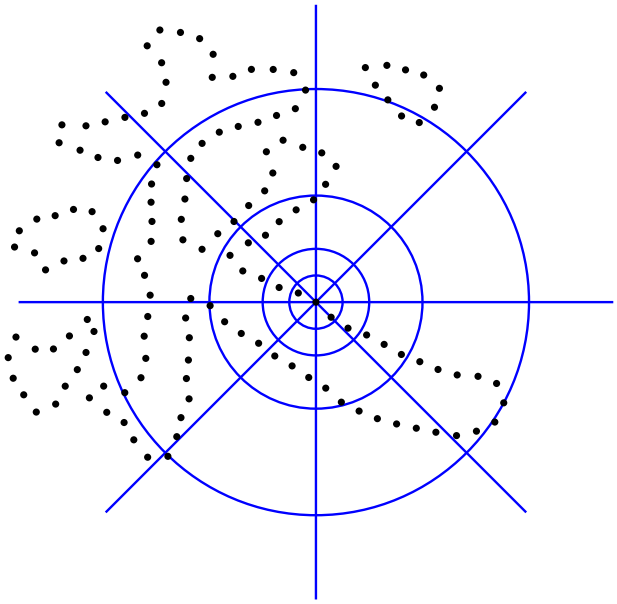


Abbildung 4.3:
Über jeden Konturpunkt wird eine Maske gelegt; die Anzahl der übrigen Konturpunkte in jeder Zone ergibt den SHAPE CONTEXT.

Auch der Umriß bietet eine gute Grundlage, hier könnte man die Kurvenscheitelpunkte und Wendepunkte festhalten. In dieser Arbeit wird das Verfahren auf der Grundlage von SHAPE CONTEXTen angewendet, dessen Beschreibung folgt.

4.2.2 Beschreibung des Shape Context-Verfahrens

Das SHAPE CONTEXT Verfahren, beschrieben in Belongie u. Malik [2000]; Belongie et al. [2000, 2001b,a], nutzt die Kontur als Vergleichskriterium. Bemerkenswert dabei ist, daß nicht nur der äußere Umriß bewertet wird, sondern auch innere Konturen miteinbezogen werden. Stellt man die Kontur als Menge aller Punkte hierauf dar, so kann die Menge von Vektoren von einem Punkt zu allen diesen Konturpunkten die Konturinformationen perfekt abbilden. Um eine angemessene Datenmenge zu erhalten und um geringfügige Unterschiede auszugleichen, wie sie beispielsweise bei mehrfacher digitaler Erfassung identischer Objekte möglich sind, werden einzelne Punkte in groben, annähernd gleichen Abständen auf der Kontur verteilt. Von jedem Konturpunkt aus wird die Umgebung in ein grobes Koordinatensystem unterteilt, wobei sich ein logarithmisch-polares System als besonders nützlich erweist. Die Einteilung erfolgt hier in gleichmäßig große Winkelzonen, während die Einteilung in Entfernungszonen auf logarithmischer Basis eine stärkere Gewichtung kurzer Abstände gegenüber längeren verspricht. Die Histogramme, die für jede Zone eines solchen Koordinatensystems die Anzahl der darin enthaltenen Punkte der übrigen Kontur darstellen, werden als SHAPE CONTEXT bezeichnet. Die Histogramme für alle Punkte der Kontur zusammen bilden einen Satz von SHAPE CONTEXTen, anhand dessen ein Vergleich von Objekten ermöglicht wird (Abbildung 4.3).

4.2.3 Eigenschaften des Shape Context

- Invarianz bezüglich der Zeichengröße:

  Der Median aller Abstände aller festgelegten Konturpunkte zueinander stellt ein stabiles Maß für die Zeichengröße dar. Sofern die Radien der Zoneneinteilung in Abhängigkeit dieser Größe gewählt werden, wird der Einfluß unterschiedlicher Scanauflösungen stark vermindert.

- Rotationsinvarianz:

  Werden die Achsen der Zonenmaske für jeden Punkt nicht an den Grundachsen ausgerichtet, sondern an der Konturtangente, wird eine Rotationsinvarianz erreicht.
  Auf diese Möglichkeit wird in dieser Arbeit verzichtet, da es auch chinesische Schriftzeichen gibt, die bei Rotationsinvarianz wie „6“ und „9“ oder „n“ und „u“ nicht mehr unterscheidbar wären.

- Translationsinvarianz:

  Da alle Abstände und Richtungen nur zwischen Konturpunkten untereinander gemessen werden, können unterschiedliche Zeichenpositionen keinen Einfluß auf den SHAPE CONTEXT nehmen.

4.2.4 Vergleich von Shape Contexten

Die Autoren geben folgende Funktion vor, die die Kosten K_{SC} für den Vergleich zweier SHAPE CONTEXTe g und h ergibt:

$$K_{SC_{g,h}} = \frac{1}{2}\sum_{z=1}^{Z}\frac{\Big(g(z)-h(z)\Big)^2}{g(z)+h(z)} \tag{4.1}$$

Z beschreibt dabei die Anzahl der Zonen, also der Quantisierungen im Histogramm und $g(z)$ bzw. $h(z)$ bezeichnet die relative Anzahl der Punkte, die der entsprechenden Histogrammzone zugeordneten wurden.

Diese Kosten ergeben 0 für perfekte Übereinstimmung und 1 für totale Ungleichheit. Für den Vergleich von zwei Zeichen G und H benötigt man eine Kostenmatrix mit den Vergleichskosten von jedem SHAPE CONTEXT für jeden Punkt von Zeichen G für jeden SHAPE CONTEXT für jeden Punkt von Zeichen H.

Das Ziel ist nun, eine optimale bijektive Punktezuordnung zu finden, so daß die Gesamtkosten möglichst niedrig sind, die sich aus den für jedes Punktepaar aufsummierten Kosten ergeben.

Während die Entwickler des SHAPE CONTEXT-Verfahrens zunächst durch die 1955 von Harold Kuhn bzw. 1957 von James Munkes entwickelte *Ungarische Methode*, später durch eine spezielle Implementierung der Kürzesten-Wege-Methode von Dijkstra (Jonker u. Volgenant [1987]) versuchen, eine optimale Lösung zu finden, wird in dieser Arbeit das von Gale und Shapley entwickelte *Stable Marriage Matching* Verfahren verwendet, wie es in Sedgewick [1991] beschrieben ist, da dieses bereits in der VIGRA-Bibliothek implementiert ist.

Die gefundene Lösung beschreibt neben einer nach gegebenen Kriterien optimalen Punktezuordnung die Höhe der Gesamtkosten, die die Ähnlichkeit der Zeichen zueinander beschreibt.

4.2.5 Beispiel eines Zeichenvergleichs

Das beschriebene Verfahren soll an folgendem einfachen Beispiel demonstriert werden:

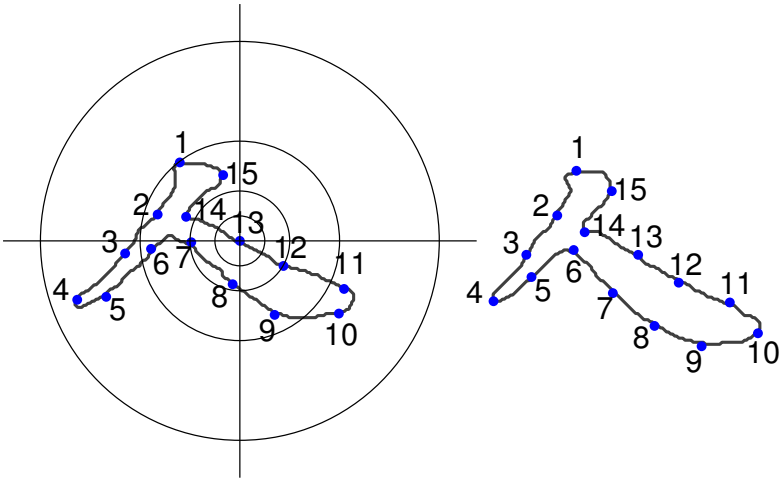


Abbildung 4.4: Auf den Konturen beider Zeichens wurden jeweils 15 Punkte verteilt. Eine Maske teilt die Umgebung eines Punktes in 20 Zonen auf.

Abbildung 4.4 zeigt zwei einfache gleichartige chinesische Zeichen. Auf der Kontur jedes Zeichens wurden jeweils 15 Punkte bestimmt. Eine einfache Maske[1] wird über jeden Punkt gelegt. Im SHAPE CONTEXT wird die Aufteilung der übrigen 14 Punkte auf die einzelnen Zonen der Maske bestimmt. Tabellen 4.1 und 4.2 listen die einzelnen SHAPE CONTEXTe aller Punkte beider Zeichen auf. Der 1. Quadrant der Maske befindet sich dabei rechts oben, die weitere Einteilung erfolgt entgegen dem Uhrzeigersinn. Die Radien werden von innen nach außen gezählt, Punkte außerhalb des äußersten Zonenkreises werden in Zone 5 gewertet.

[1] Näheres zu Masken im Abschnitt 5.3.2

SHAPE CONTEXT	1. Quadrant					2. Quadrant					3. Quadrant					4. Quadrant				
Radien:	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Punkt 1:	0	0	0	0	0	0	0	0	0	0	0	0	2	3	0	0	1	3	3	2
Punkt 2:	0	0	2	0	0	0	0	0	0	0	0	1	2	1	0	0	2	1	4	1
Punkt 3:	1	0	3	3	0	0	0	0	0	0	0	1	1	0	0	0	0	0	3	2
Punkt 4:	0	1	2	7	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Punkt 5:	0	1	2	7	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1
Punkt 6:	0	3	2	1	0	0	0	0	0	0	1	0	2	0	0	0	0	1	3	1
Punkt 7:	0	1	1	0	0	0	2	1	0	0	0	1	1	2	0	0	0	2	3	0
Punkt 8:	0	1	1	1	0	0	0	3	4	0	0	0	0	2	0	0	0	1	1	0
Punkt 9:	0	1	1	0	0	0	0	2	9	0	0	0	0	0	0	0	0	1	0	0
Punkt 10:	0	1	0	0	0	0	0	2	5	6	0	0	0	0	0	0	0	0	0	0
Punkt 11:	0	0	0	0	0	0	0	1	6	2	0	1	1	1	2	0	0	0	0	0
Punkt 12:	0	0	0	0	0	0	0	2	6	0	0	1	1	1	1	0	0	2	0	0
Punkt 13:	0	0	0	0	0	0	0	4	0	0	0	2	1	3	0	0	0	2	2	0
Punkt 14:	0	0	1	0	0	0	1	1	0	0	0	1	1	2	0	0	1	2	4	0
Punkt 15:	0	0	0	0	0	0	1	0	0	0	0	0	3	4	0	0	0	1	5	0

Tabelle 4.1: SHAPE CONTEXTe von Zeichen 1 aus Abbildung 4.4.

SHAPE CONTEXT	1. Quadrant					2. Quadrant					3. Quadrant					4. Quadrant				
Radien:	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Punkt 1:	0	0	0	0	0	0	0	0	0	0	0	0	2	3	0	0	1	1	4	3
Punkt 2:	0	0	2	0	0	0	0	0	0	0	0	0	2	1	0	0	2	2	4	1
Punkt 3:	0	1	2	2	0	0	0	0	0	0	0	0	1	0	0	1	0	1	4	2
Punkt 4:	0	1	1	8	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
Punkt 5:	0	0	3	3	0	1	0	0	0	0	0	1	0	0	0	0	0	1	3	2
Punkt 6:	1	0	2	0	0	0	1	0	0	0	0	1	1	1	0	0	0	2	4	1
Punkt 7:	0	1	1	0	0	0	0	5	2	0	0	0	0	1	0	0	0	1	3	0
Punkt 8:	0	0	2	0	0	0	0	2	8	0	0	0	0	0	0	0	0	1	1	0
Punkt 9:	0	0	2	0	0	0	0	2	8	2	0	0	0	0	0	0	0	0	0	0
Punkt 10:	0	0	0	0	0	0	1	1	3	8	0	0	1	0	0	0	0	0	0	0
Punkt 11:	0	0	0	0	0	0	0	1	6	3	0	0	2	0	1	0	1	0	0	0
Punkt 12:	0	0	0	0	0	0	1	0	7	0	0	0	2	1	0	0	0	3	0	0
Punkt 13:	0	0	0	0	0	0	0	4	2	0	0	1	0	2	0	0	1	1	3	0
Punkt 14:	0	1	0	0	0	0	1	1	0	0	1	0	2	1	0	0	0	2	4	1
Punkt 15:	0	0	0	0	0	0	1	0	0	0	0	1	2	3	0	0	0	1	5	1

Tabelle 4.2: SHAPE CONTEXTe von Zeichen 2 aus Abbildung 4.4.

Ein Vergleich der SHAPE CONTEXTe wird durch Gleichung 4.1 errechnet. Vergleicht man etwa die Punkte 1 und 1 beider Zeichen, die sich nur in den Zonen 3, 4 und 5 des 4. Quadranten unterscheiden, so erhält man Kosten von

$$\frac{1}{2}\left(\frac{\left(\frac{3}{14}-\frac{1}{14}\right)^2}{\frac{4}{14}}+\frac{\left(\frac{3}{14}-\frac{4}{14}\right)^2}{\frac{7}{14}}+\frac{\left(\frac{2}{14}-\frac{3}{14}\right)^2}{\frac{5}{14}}\right)=\frac{1}{2}\left(\frac{1}{14}+\frac{1}{98}+\frac{1}{70}\right)=0.047959.$$

Berechnet man so die Vergleichskosten aller Punkte zueinander, erhält man Tabelle 4.3.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	**5**	14	41	75	54	28	56	79	100	90	79	54	44	24	17
2	19	**5**	31	70	35	17	52	67	86	90	76	68	42	33	22
3	49	39	20	33	**14**	27	68	72	83	93	90	90	71	56	47
4	78	67	35	**4**	33	67	73	79	86	100	100	100	89	74	81
5	78	67	35	6	25	60	73	79	**86**	100	100	100	82	74	74
6	43	30	**15**	47	34	35	51	68	86	90	86	75	71	27	45
7	39	33	42	75	51	18	33	61	81	76	83	54	36	**18**	20
8	64	60	55	66	61	60	**13**	21	35	65	55	43	24	52	64
9	93	81	76	86	82	81	35	**9**	16	57	39	33	50	74	93
10	100	100	93	93	100	100	52	42	20	**15**	23	57	61	83	100
11	80	83	93	100	93	79	60	41	27	34	**14**	37	50	76	73
12	70	69	83	100	83	64	41	27	37	55	31	**20**	33	60	63
13	40	46	64	90	64	39	31	62	81	81	79	63	**18**	35	30
14	28	19	45	81	48	**17**	37	60	81	79	76	56	26	24	18
15	19	30	50	88	66	28	55	81	100	82	83	54	47	23	**8**

Tabelle 4.3: Prozentuale Vergleichskosten aller SHAPE CONTEXTe beider Zeichen aus Abbildung 4.4. Die Zeilen entsprechen den Punkten von Zeichen 1, die Spalten denen von Zeichen 2.

Das Stable-Marriage-Matching Verfahren bestimmt nun Punktepaare, deren Summe der Vergleichskosten so gering wie möglich ist. Dabei muß jedem Punkt eines Zeichens genau ein Punkt des anderen zugeordnet werden. In der Tabelle hervorgehoben sind die ausgewiesenen Punktepaare; die Gesamtkosten für den Zeichenvergleich entsprechen der Summe der Kosten aller Punktepaare, gewichtet mit der Anzahl dieser Paare: $\frac{2.61}{15} = 0.174$.

Abbildung 4.5 zeigt die Zuordnung dieser Punktepaare.

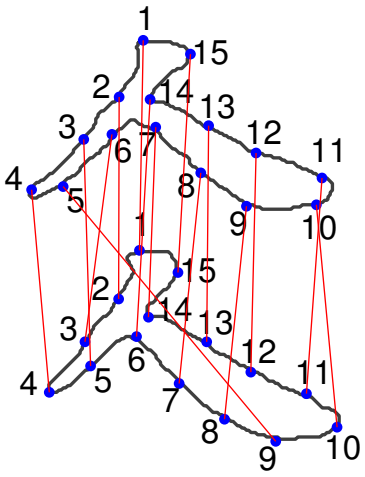


Abbildung 4.5:

Punkte, deren SHAPE CONTEXTe ähnlich sind, werden einander zugeordnet.

4.2.6 Erweiterungen des Vergleichsverfahrens

Verwendung von Zeichen-Skeletten

Ein am Arbeitsbereich von Hans Meine implementiertes Skelettierungsverfahren, das über die Konturpolygonpunkte eine bedingte Delaunay-Triangulation durchführt, die wiederum als Grundlage für eine Chordal-Axis-Transformation dient, liefert die Skelette der einzelnen Zeichen. Für eine genaue Beschreibung des Verfahrens siehe Prasad [1997]. Anstatt auf der Kontur Punkte festzulegen, von denen SHAPE CONTEXTe erzeugt werden, können diese Punkte auch auf dem Skelett definiert werden.

Hierbei wird die Idee verfolgt, daß das Zeichenskelett die Form des Zeichens deutlicher beschreibt als die Kontur, da die SHAPE CONTEXTe von Konturpunkten auf beiden Seiten eines Pinselstriches sehr ähnlich ausfallen und sich beim Abgleich von Konturpunkten zweier Zeichen so falsche Zuordnungen ergeben können. Skelette haben allerdings den Nachteil, daß die Dicke der Pinselstriche nicht erfaßt wird, obwohl diese – unter Berücksichtigung der Zielsetzung dieser Arbeit – eine wertvolle Information darstellen kann.

Einbinden lokaler Informationen

Ein SHAPE CONTEXT beschreibt für jeden Punkt nur dessen Umgebung, nicht aber Informationen über den Punkt selbst. Die Autoren des SHAPE CONTEXT-Verfahrens schlagen deshalb u.a. vor, den Winkel der Konturnormalen an der Position des ausgewählten Konturpunktes zu erfassen (siehe Abbildung 4.6) und diese Werte in eine Kostenfunktion zu integrieren. Auch in dieser Arbeit kann diese Information miteinbezogen werden.

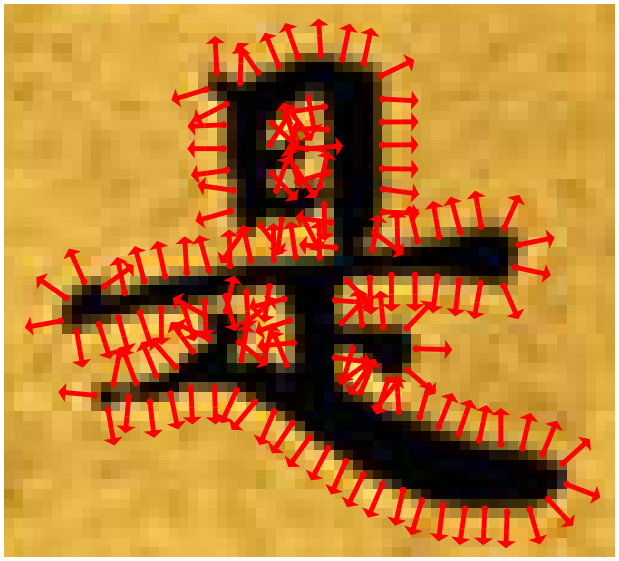


Abbildung 4.6:

Die Normale zur Kontur an den einzelnen Punkten bietet eine weitere Vergleichsmöglichkeit. Auch die Stärke des Gradienten könnte ausgewertet werden, der Informationsgehalt wird aber in dieser Anwendung als nicht maßgebend erachtet.

Angleichen zweier Konturen durch affine Transformation

Durch die Bildung von Punktepaaren können Korrespondenzvektoren zwischen Koordinaten erstellt werden, die es ermöglichen, diese mit möglichst kleinem Fehler aufeinander abzubilden. Die VIGRA-Bibliothek bietet eine Funktion, die aus diesen Vektoren eine affine Abbildungsmatrix erstellt.

Die durch den Vergleich einzelner SHAPE CONTEXTe gegebenen Kosten repräsentieren die Übereinstimmung der Punkte anhand der festgelegten Kriterien und lassen sich somit zur Gewichtung der einzelnen Vektoren verwenden, so daß Vektoren von Punkten mit guter Übereinstimmung stärkeren Einfluß auf die Berechnung der Abbildungsmatrix nehmen als andere.

Werden alle Punkte des einen Zeichens darauf anhand der durch die Matrix beschriebenen affinen Transformation auf die des anderen Zeichens angepaßt, können von diesem Zeichen neue SHAPE CONTEXTe gebildet werden. Dieser Vorgang aus Errechnung der SHAPE CONTEXTe und anschließender Anpassung der Zeichen kann beliebig oft wiederholt werden.

Darstellungen, Beispiele und weitere Details zum SHAPE CONTEXT*-Verfahren und zu diesen Erweiterungen finden sich in Kapitel 5.*

5 Anwendungen

In diesem Kapitel werden Anwendungsbeispiele der in den vorangegangenen Kapiteln beschriebenen Verfahren vorgestellt. Dazu gehört die Vorstellung der grafischen Oberfläche, mit der eine Segmentierung sowie ein anschließender Zeichenvergleich schrittweise durchgeführt und dargestellt werden kann und die zudem eine strukturierte Eingabe der zu den Verfahren entsprechenden Parametern ermöglicht. Neben Ergebnissen des Zeichenvergleichs alleine durch das SHAPE CONTEXT-Verfahren werden Auswirkungen verschiedener Erweiterungen zu diesem präsentiert. Abschließend wird eine Fragestellung des Sinologen Dr. Matthias Richter untersucht und dargestellt.

5.1 Beispiele für die Segmentierung von Zeichen

5.1.1 Vorstellung der grafischen Oberfläche

Um Zeichen vergleichen zu können, müssen diese zunächst in einer Bildvorlage erkannt und aus dieser segmentiert werden. Dazu dient das in Kapitel 3 beschriebene Verfahren. Um diese Segmentierung und auch die Verfahren des Zeichenvergleichs übersichtlich und ergonomisch bedienen zu können, wird eine grafische Oberfläche verwendet. Da in dieser Arbeit sowohl Teile der VIGRA-Bibliothek als auch der GEOMAP verwendet werden, und diese für die Darstellung einzelner Teile das Qt-Rahmenwerk benutzen, wurde auch in dieser Arbeit davon Gebrauch gemacht, so daß sich die benutzten Teile der GEOMAP und der VIGRA nahtlos in die Oberfläche integrieren lassen. So wird zur Darstellung der Bildvorlagen der „Imageviewer“ der VIGRA verwendet, die Darstellung der Kanten der Flächensegmentierung ist der GEOMAP entnommen. Die Steuerung der Qt-Elemente erfolgt über Python bzw. PyQt.

Abbildung 5.1 zeigt diese grafische Oberfläche. Der oberste Teil ist symmetrisch angelegt; mittels der Eingabefelder an oberster Position kann jeweils eine Bildvorlage in das rechte und linke Vorlagenfenster geladen werden. Im linken Vorlagenfenster wird ein einzelnes Zeichen ausgesucht, das mit allen Zeichen im rechten Vorlagenfenster verglichen werden soll. Falls ein Zeichen mit allen anderen innerhalb derselben Vorlage verglichen werden soll, genügt es, diese Vorlage in nur eines der beiden Fenster zu laden, durch Betätigung des Pfeils wird die Vorlage in das jeweils andere Fenster kopiert.

In Abbildung 5.1 ist bereits eine der zur Verfügung stehenden Bildvorlagen in das rechte Vorlagenfenster geladen worden.

Abbildung 5.1: Nach Programmstart sind beide Vorlagenfenster uninitialisiert. Hier wurde dagegen bereits eine Vorlage ins rechte Fenster geladen; diese kann entweder auf einen Ausschnitt beschränkt werden oder direkt duch den Segmentierungsablauf behandelt werden.

Mit dem Werkzeug „Schere“ kann im Bild ein Ausschnitt gewählt werden, auf den der Bildinhalt daraufhin beschränkt wird. Das ist besonders nützlich, wenn – wie im gezeigten Vorlagenfenster – das Manuskript selbst breite Ränder aufweist oder die Bildvorlage neben dem Manuskript breite Ränder zeigt. Somit wird vermieden, daß z.B. die Farbtafel am unteren Bildrand Einfluß auf die Segmentierung hat; außerdem ist die Verarbeitungszeit abhängig von der Bildgröße und kann durch Begrenzung auf relevante Bildteile positiv beeinflußt werden.

Von diesem Zustand aus kann die Segmentierung der rechten Bildvorlage gestartet werden. Diese umfaßt die Segmentierung der Spalten und die darauf folgende Wasserscheidensegmentierung der einzelnen Spaltenbereiche, ohne auf diese eine Heuristik anzuwenden.

Zur Benutzerführung werden alle Werkzeugelemente, deren Anwendungen im jeweiligen Zustand nicht sinnvoll sind, deaktiviert.

Abbildung 5.2 zeigt die Bearbeitung der rechten Bildvorlage mit der in Abschnitt 3.2.2 beschriebenen Methode. Eine Statusanzeige hilft dem Benutzer, die für die Abarbeitung benötigte Zeit einzuschätzen, welche vor allem durch die Größe der Vorlage bestimmt wird und somit recht unterschiedlich sein kann.

Während die Ergebnisse der Spaltensegmentierung bei den zur Verfügung stehenden Vorlagen sehr zuverlässig sind, kommt es bei der Segmentierung der Flächen zu einzelnen Zeichen häufiger zu falschen Ergebnissen. Zu beachten ist, daß zur Segmentierung nur die Kanten genutzt werden können, die durch die grundlegende Wasserscheidensegmentierung entstanden sind. Es ist also nicht vorgesehen, zusätzliche Kanten hinzuzufügen, oder einmal gelöschte Kanten wiederherzustellen. Sowohl die Parameter des in Abschnitt 3.2.2 beschriebenen Verfahrens als auch die in Abschnitt 3.2.3 getroffenen Kriterien sind daher so ausgelegt, daß im Zweifel eine Kante nicht gelöscht wird. Sollten also nach Anwendung beider Verfahren Kanten verbleiben, die nicht erwünscht sind, können diese einfach mit dem Werkzeug „Remove edges“ durch Benutzung der Maus an entsprechender Stelle in der Vorlage entfernt werden.

Das Werkzeug „Protect edges“ stellt dazu das Gegenstück dar, hierdurch können durch die Maus in der Vorlage markierte Kanten vor automatischen Löschvorgängen geschützt werden. In der Praxis wird dieses Werkzeug eingesetzt, nachdem zunächst mindestens eine der beiden Heuristiken zur Anwendung gekommen ist und darauf vom Anwender festgestellt wurde, daß eine notwendige Kante gelöscht wurde.

Da es keine Möglichkeit gibt, das Löschen von einzelnen Kanten rückgängig zu machen, wird vor Anwendung des 1. Kriteriums automatisch ein Sicherungspunkt gesetzt, zu dem später zurückgekehrt werden kann. Dieses erfolgt durch Anwahl von „Revert to last saved state“, worauf alle Veränderungen seit dem letzten Sicherungspunkt verworfen werden. Nun können Kanten vor Löschvorgängen geschützt werden, die so von später durchgeführten heuristischen Verfahren als nicht löschbar erkannt werden. Man kann hier nicht von einem Lernprozeß sprechen; die Heuristik wird für diesen bestimmten Fall nur indirekt angepaßt, da die Kriterien zur Kantenauswertung verändert werden. So sucht z.B. das 2. Kriterium bei einer als zu klein erkannten Fläche die 1.) nicht geschützte Kante Kante 2.) die Kante mit dem niedrigeren minimalen Helligkeitswert, der von dieser durchlaufen wird, und überprüft, ob diese gelöscht werden sollte. Geschützt

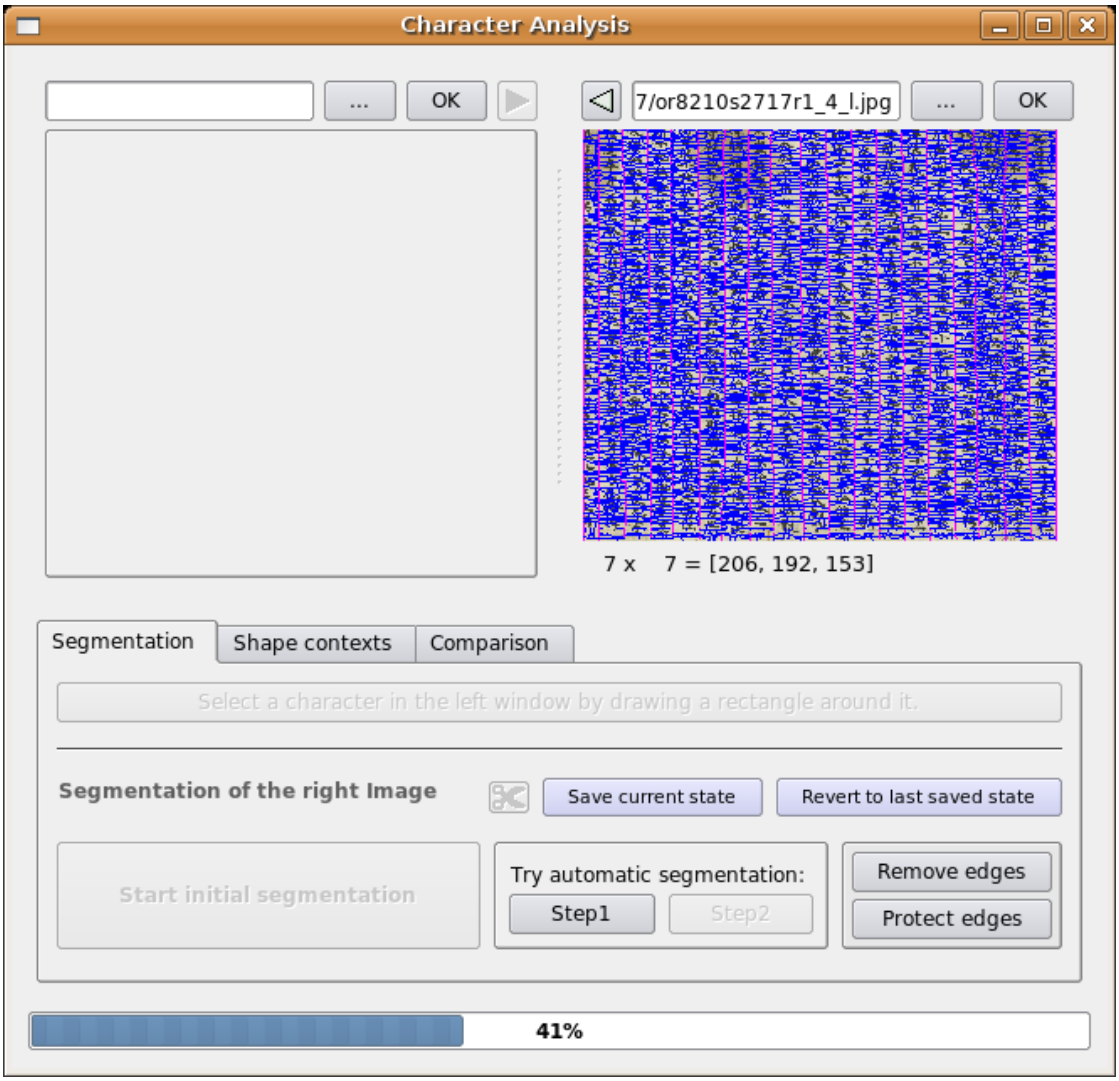


Abbildung 5.2: Nach Segmentierung der Spalten können nacheinander die in Abschnitt 3.2.2 und Abschnitt 3.2.3 beschriebenen Verfahren zur Zeichensegmentierung angewendet werden.

Abbildung 5.3: Im rechten Fenster eine voll segmentierte Vorlage, in der drei Kanten noch von Hand entfernt werden müssen. Eine vergrößerte Darstellung dieser Vorlage wird in Abbildung 5.5 dargestellt.
Im linken Fenster wurde durch einen mit der Maus gezeicheten Rahmen ein Zeichen ausgewählt, mit dem alle im rechten Fenster verglichen werden sollen.
Als nächster Schritt ist darauf „*Shape contexts*“ anzuwählen, um diese zu erstellen.

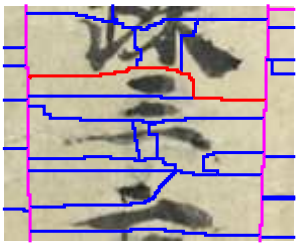


Abbildung 5.4:

Alle vier Kanten, die von Hand mit dem Werkzeug „Protect edges“ markiert wurden und somit von automatischen Löschvorgängen unberührt bleiben, sind rot dargestellt.

sind - wie in Abschnitt 3.2.3 erwähnt - die Kanten der Spaltenmarkierungen; ein von Hand gesetzter Schutz ist damit gleich zu setzen. Sollte also bei einer zu kleinen Fläche die Kante, die durch niedrigere Helligkeitswerte verläuft, von Hand geschützt sein, so wird die verbleibende, der geschützten gegenüberliegende Kante gelöscht, soweit deren Kriterien dem entsprechen. In diesem Fall werden nun andere Flächen entstehen, deren Auswertung im weiteren Verlauf zu veränderten Ergebnissen führen wird.

Oft ist es einfacher, einen Kantenschutz erst nach Anwendung des ersten Kriteriums zu setzen, da dann schon viele unerwünschte Kanten entfernt wurden und der Kantenschutz über die komplette Spaltenbreite gesetzt wird. Vor Anwendung des 1. Kriteriums erstrecken sich Kanten dagegen normalerweise nicht über die komplette Spaltenbreite. Wird dann ein Knoten entfernt, an dem eine geschützte Kante endet, und diese mit einer ungeschützten Kante verschmolzen, kann der Verschmelzungsprozeß der GEOMAP entweder die ungeschützte Kante entfernen und die geschützte über die gesamte Länge erstrecken lassen, oder es wird der Verlauf der geschützten Kante durch die ungeschützte ersetzt. In dem Fall geht der Kantenschutz verloren; die „neue“ Kante könnte von einem folgenden Löschvorgang erfaßt werden. Abbildung 5.4 zeigt die in diesem Fall empfohlene Vorgehensweise, alle Kanten zu schützen, die im Verlauf einer gedachten Kante liegen, die erhalten bleiben soll. Alle von Hand geschützt markierten Kanten werden rot dargestellt.

Mit „Save current state“ kann der Sicherungspunkt zur Wiederherstellung eines Zustandes auch von Hand gesetzt werden.

5.1.2 Segmentierung einer einfachen Vorlage

Abbildung 5.5 zeigt eine für die Segmentierung relativ einfache Bildvorlage:

- Der Kontrast zwischen den Zeichen und dem Hintergrund ist hoch
- Es gibt keine Löcher, Risse oder sonstige schwarze Flecken in der Schreibunterlage
- Der Spaltenabstand und
- Der Zeichenabstand innerhalb einer Spalte ist ausreichend groß und ziemlich gleichmäßig.
- Alle Zeichen haben eine relativ einheitliche Größe.

Abbildung 5.5: Bei der Segmentierung dieser Vorlage wurden 3 Zeichen von jeweils einer – hier orange gefärbten – Kante durchtrennt (zwei Kanten befinden sich nahe der linken unteren Ecke und eine in der rechten oberen Ecke).

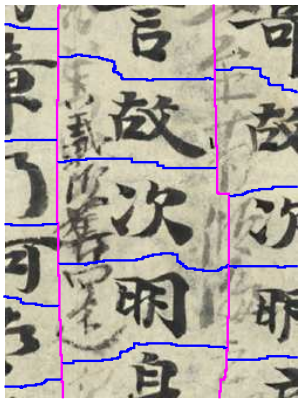


Abbildung 5.6:
Ausschnitt aus Abbildung 5.5: Zwischen den Spalten sind weitere Fragmente einzelner Zeichen zu erkennen, diese erschweren die Segmentierung.

Von den 496 Zeichen werden nur drei Zeichen von jeweils einer Kante durchtrennt. Diese 3 Kanten müßten von Hand gelöscht werden. Es wurde keine notwendige Kante gelöscht.

Da die Helligkeit der von Kanten durchlaufenen Bereiche immer nur in Abhängigkeit zu Helligkeitswerten der übrigen Kanten jeweils einer Fläche verglichen wird, es somit also keinen für die gesamte Bildvorlage geltenden Schwellwert gibt, anhand dessen Kanten gelöscht werden, stellen die drei großen Flecken kein Hindernis dar, das eine korrekte Segmentierung verhindern würde. Auch das Quadtree-Verfahren gleicht diese Helligkeitsschwankungen aus.

Nicht ohne Probleme ist dagegen die Segmentierung der Spalten am rechten Vorlagenrand. Zwischen der zweiten und dritten sowie zwischen der dritten und vierten Spalte von rechts sind Schatten von Zeichenfragmenten zu sehen, die wahrscheinlich von der Rückseite der Schreibunterlage stammen (Vergrößerung in Abbildung 5.6). Hier sind die Grundannahmen verletzt, daß alle Spalten etwa gleiche Abstände zueinander haben oder in einer Spalte keine zwei Zeichen nebeneinander stehen dürfen.

Auch wenn das Segmentierungsergebnis durchaus so interpretiert werden kann, daß in jeder Fläche nur ein Zeichen von der Vorderseite der Schreibunterlage enthalten ist, so wird das Problem lediglich auf das nachfolgende Verfahren, das in jeder Fläche ein Zeichen zu segmentieren versucht, verschoben.

5.1.3 Segmentierung problematischer Vorlagen

Abbildung 5.7 zeigt eine Vorlage, dessen Segmentierung deutlich schwieriger ist:

- Der Zeichenabstand innerhalb einer Spalte ist recht unterschiedlich und oft sehr klein, teilweise berühren sich einzelne Zeichen sogar.
- Die Abstände innerhalb einzelner Zeichen sind größer als der Abstand zwischen diesen und benachbarten Zeichen.
- Der Spaltenabstand ist besonders in der rechten Bildhälfte gering, auch hier berühren sich nebeneinanderstehende Zeichen.

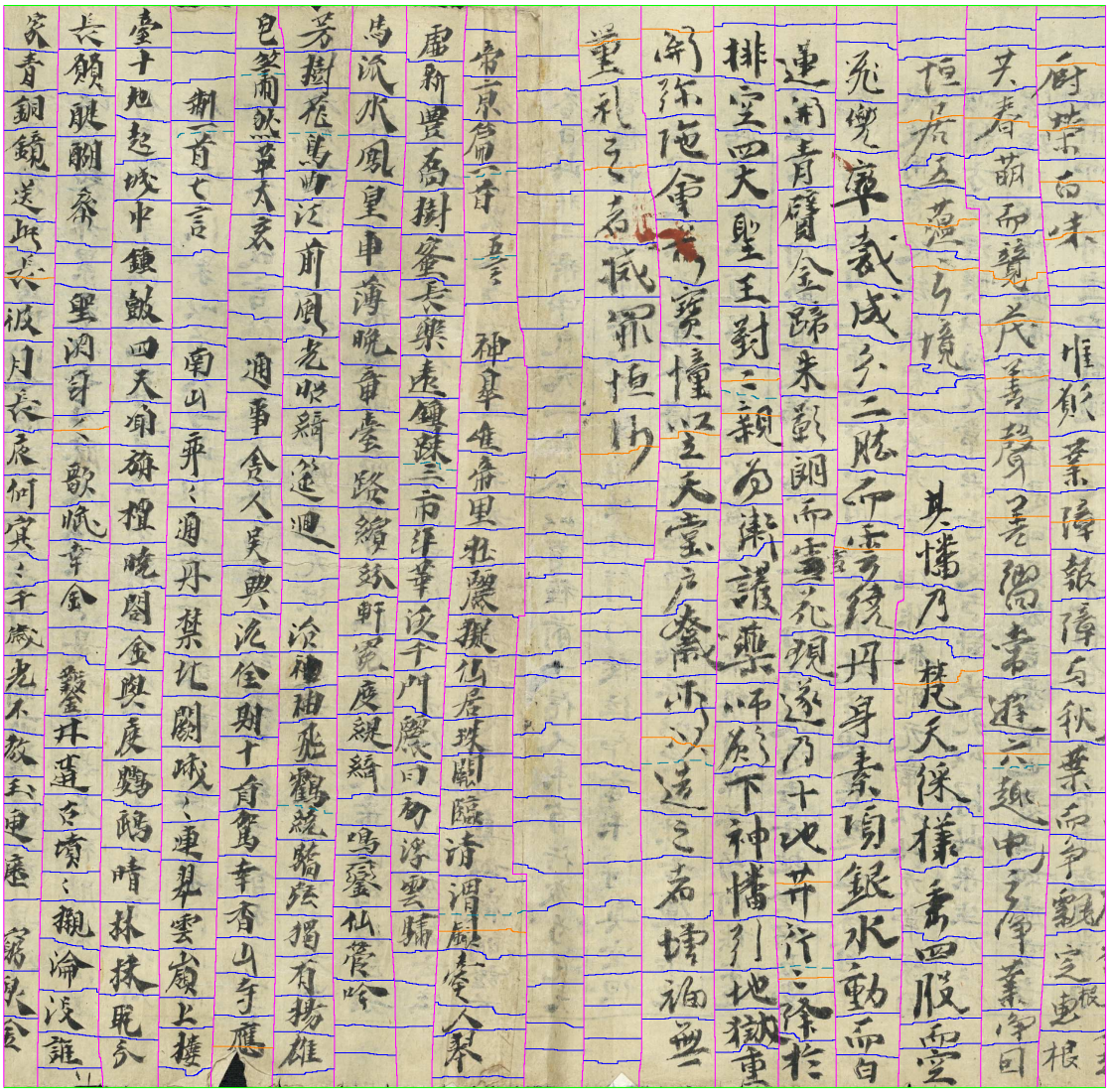


Abbildung 5.7:

- Auch in dieser Bildvorlage sind blau gefärbte Kanten vom Segmentieralgorithmus richtig erkannt worden,
- weitere 32 vorhandene Kanten in diesem Orange müssen noch von Hand entfernt werden.
- In Cyan und unterbrochen dargestellt sind dagegen zwölf Kanten, die nach Anwendung des 1. Kriteriums noch erhalten blieben, durch die Anwendung des 2. Kriteriums jedoch fälschlicherweise entfernt wurden.

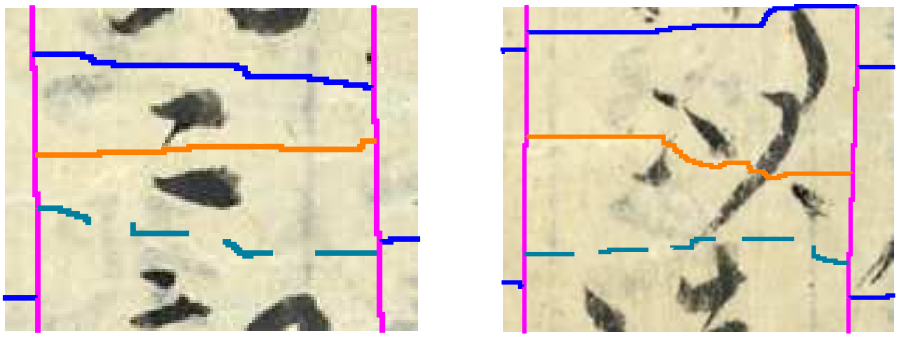


Abbildung 5.8: Während die orange dargestellte Kante nach der Segmentierung noch vorhanden ist, und von Hand gelöscht werden muß, wurde die cyan dargestellte Kante durch das Segmentierverfahren gelöscht; diese hätte erhalten bleiben müssen.

- Große Bereiche, die kein Zeichen enthalten, erschweren die Segmentierung der Spalten; alleine aufgrund der durch die Schreibunterlage durchscheinenden Zeichen und anderer – eher zufälliger – Helligkeitsunterschiede in der Schreibunterlage wurde die mittige Spalte ohne Zeichen segmentiert, ein Verschmelzen dieser mit der rechts oder links benachbarten Spalte wäre hier allerdings unkritisch.
- Die Zeichen in der rechten Bildhälfte sind deutlich größer als auf der linken Seite.
- Am rechten Bildrand ist der Kontrast zwischen den Zeichen und Störeinflüssen der Schreibunterlage niedriger.

Von 366 Zeichen werden 32 durch jeweils eine Kante durchtrennt. Auch der Umstand, daß 12 weitere Kanten automatisch entfernt wurden, die zwischen Zeichen liegen, ist nicht richtig.

An einigen Stellen wurde eine unerwünschte Kante beibehalten, eine benachbarte notwendige Kante dagegen entfernt. Abbildung 5.8 verdeutlicht Probleme an zwei Beispielen. In beiden Fällen stößt das 2. Kriterium auf eine Fläche, die sich zwischen den orange und cyanfarbenen Kanten befindet. Die Helligkeit der Bildbereiche unter beiden Kanten wird in der geglätteten Vorlage überprüft und aufgrund der geringen Höhe der Fläche und des minimalen Helligkeitswertes der Kante, die durch die dunkleren Bildbereiche verläuft, wird entschieden, diese zu löschen. In beiden Fällen in Abbildung 5.8 ist der Unterschied der Minima der von den Kanten durchlaufenen Bildbereiche sehr gering, die cyanfarbene Kante weist aber den geringfügig dunkleren Bildbereich auf.

Während in Abbildung 5.8a der Abstand innerhalb des Zeichens zu groß ist, um diesen Bildbereich durch die vertikale Glättung und die damit erwünschte Zusammenführung der Zeichenelemente dunkler ausfallen zu lassen als die Bereiche über und unter dem Zeichen, wird in Abbildung 5.8b durch die Glättung ein dunklerer Bildbereich zwischen den Zeichen generiert als im Verlauf der orangenen Kante.

Abbildung 5.9:

- Kanten in diesem Blau werden durch die Segmentierung richtig erkannt.
- 64 Kanten in Orange wurden nach der Segmentierung beibehalten und müssen von Hand gelöscht werden.
- 30 Kanten in diesem Cyan wurden durch Anwendung des 2. Kriteriums gelöscht; diese müssen vorher von Hand geschützt werden.
- 39 Kanten in diesem Grün wurden bereits durch Anwendung des 1. Kriteriums gelöscht; diese müssen bereits direkt nach der initialen Segmentierung von Hand geschützt werden.

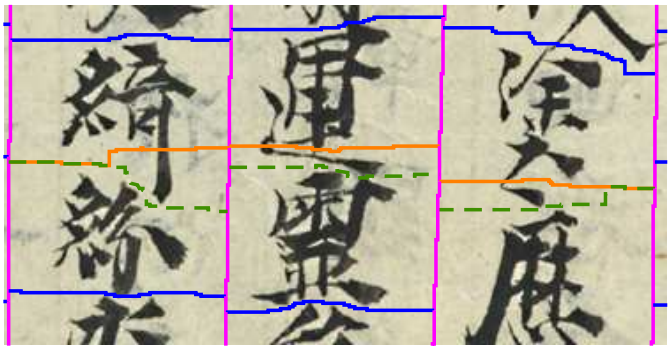


Abbildung 5.10:
Ausschnitt aus Abbildung 5.9: Die orangefarbenen Kanten wurden automatisch erkannt, der Verlauf der grünen Kanten entspricht dagegen der richtigen Segmentierung.

Abbildung 5.9 enthält 436 Zeichen. Auch hier

- ist der Zeichenabstand innerhalb einer Spalte fast ausnahmslos sehr klein, einzelne Zeichen berühren oder überlappen sich häufig.

- Der Spaltenabstand ist auch hier häufig gering, Berührungen nebeneinanderstehender Zeichen sind deutlich erkennbar.

64 Kanten müssen von Hand entfernt werden, weil sie Zeichen durchtrennen. Einige dieser Kanten durchtrennen Zeichen mittig, so daß die Segmentierungsergebnisse wertlos sind, andere Trennungen betreffen nur Randbereiche der Zeichen (Abbildung 5.10), sie stellen somit keine gravierende Beeinflussung der Segmentierung dar. Von letzteren Kanten wurde nur eine Auswahl korrigiert, abhängig von gewünschter Segmentierungsgenauigkeit müßten noch einige weitere Kanten „verschoben“ werden.

30 Kanten wurden durch Anwendung des 2. Kriteriums, 39 weitere Kanten bereits durch Anwendung des 1. Kriteriums entfernt, obwohl diese zur Segmentierung notwendig sind. Diese Kanten müssen durch Markierungen von Hand geschützt werden.

An einigen Stellen scheint die korrekte Segmentierung ohne Kenntnisse der chinesischen Schriftzeichen nicht möglich zu sein, hier können auch mit diesen Segmentierungsverfahren ohne Wissensbasis keine fehlerfreien Ergebnisse erzielt werden. Aber selbst mit solchen Kenntnissen könnte die richtige Segmentierung teilweise nur im Zusammenhang mit umliegenden Zeichen erkannt werden (Abbildung 5.12).

In Abbildung 5.11 sind etwa 327 Zeichen erkennbar, 16 davon werden durch eine Kante geteilt, die jeweils noch von Hand zu löschen ist. Recht häufig sind in diesem Bildausschnitt zwei kurze dicht aufeinanderfolgende horizontale Striche zu sehen, diese dienen als Wiederholungszeichen oder als Ligatur. Mit drei Kanten, die bereits durch die Anwendung des ersten Kriteriums gelöscht wurden, sollten Zeichen von diesen Doppelstrichen getrennt werden.

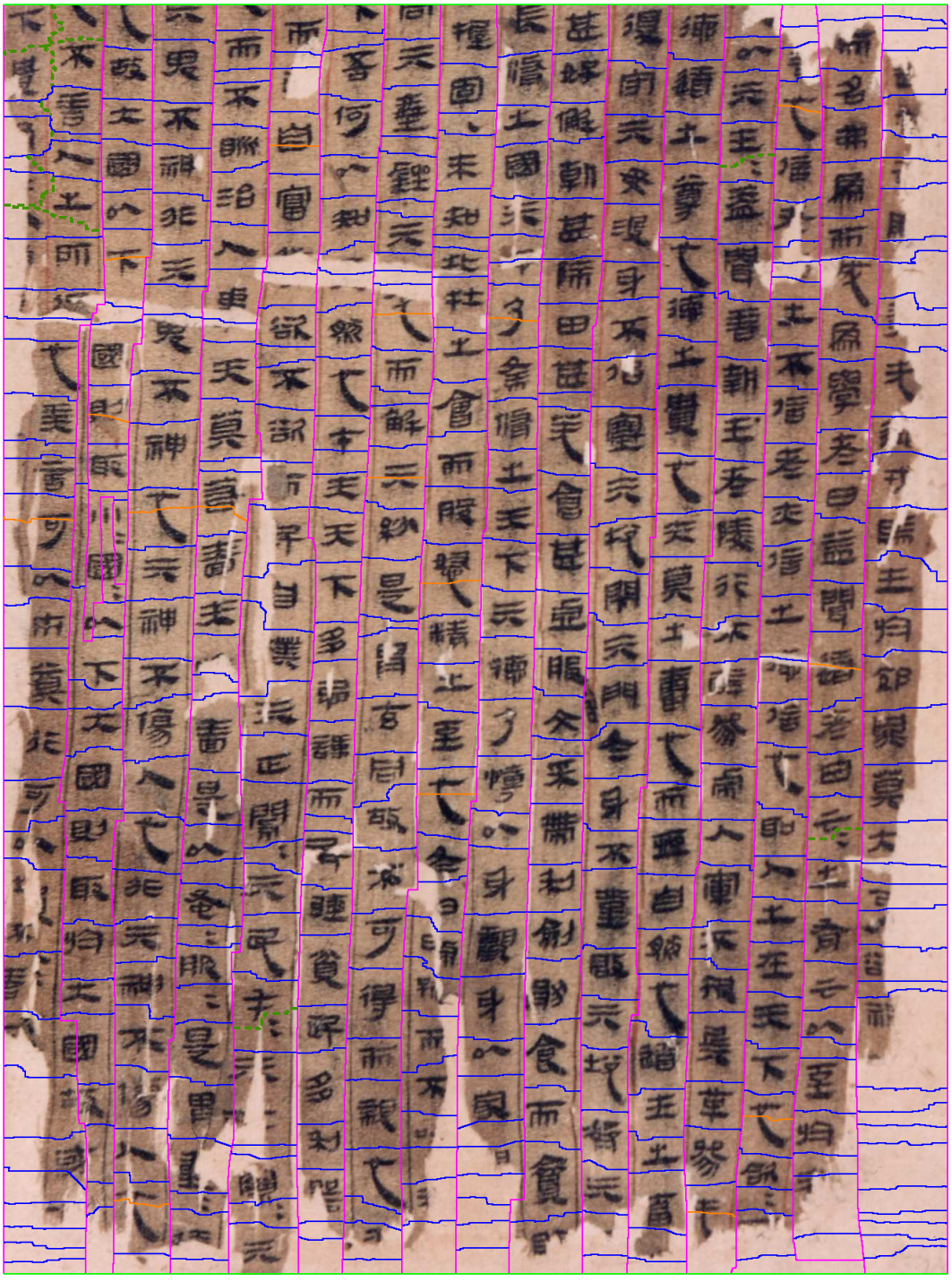


Abbildung 5.11:

- Kanten in diesem Blau werden durch die Segmentierung richtig erkannt.
- 16 Kanten in diesem Orange wurden nach der Segmentierung beibehalten und müssen von Hand gelöscht werden.
- Kanten in diesem Grün wurden bereits durch Anwendung des 1. Kriteriums gelöscht; diese müssen bereits direkt nach der initialen Segmentierung von Hand geschützt werden.

Abbildung 5.12:

Ob es sich hierbei um ein oder zwei Zeichen handelt, kann nur durch den Kontext der benachbarten Zeichen bestimmt werden.

Auch wenn die Abstände zwischen den Zeichen fast überall relativ groß sind, bestehen die Schwierigkeiten bei dieser Segmentierung vor allem in diesen Punkten:

- Die Aussrichtung der Spalten ist nicht streng vertikal, so daß eine Kante der Spaltensegmentierung, die die zweite Spalte von links mittig teilt, durch das Kriterium der vertikalen Korrelation nicht erfaßt wird.
- Die Schreibunterlage hat viele Risse und offene Stellen, die darunterliegende Scanunterlage ist deutlich heller als das Seidenpapier, auf dem geschrieben wurde. Ist die Scanunterlage direkt neben dunklen Stellen auf der Schreibunterlage sichtbar, so heben durch die Glättung die hellen Bereiche der Scanunterlage die dunklen teilweise auf, dadurch wird z.B. die Kante der zweiten Spalte von links im unteren Bildbereich verschoben.
- Durch den gleichen Effekt werden einige Zeichen von Kanten durchtrennt oder nicht als Zeichen erkannt.
- Auch hier variiert die Zeichenhöhe recht stark.

In der linken oberen Ecke ist der Ansatz einer weiteren Spalte erkennbar; die Kante, die diese von der rechts daneben liegenden Spalte trennen würde, trennt aber einen zu kleinen Bereich ab, so daß dieser nicht als eigene Spalte bewertet wird. Somit wird diese Kante automatisch gelöscht; einzelne Zeichen müssen dann von Hand durch Kanten getrennt werden, wie es in der Abbildung 5.11 zu erkennen ist.

5.2 Zeichenvergleich durch Nutzung der grafischen Oberfläche

5.2.1 Segmentierung der Zeichen innerhalb der zuvor festgelegten Flächen

Wie bereits in Kapitel 4 beschrieben, wird nach der Einteilung des Bildes in Flächen versucht, in diesen jeweils ein Zeichen zu segmentieren. Benachbarte Pixel, deren Helligkeitswerte unterhalb des innerhalb der jeweiligen Fläche festgestellten Schwellwertes liegen, werden zu Regionen zusammengefaßt („Labeling“). Sollte die maximale Ausdehnung eines Labels weniger als ein

Abbildung 5.13: Die Darstellung der Zeichen, die aus zusammenhängenden Bildpixeln mit Helligkeitswerten unterhalb des gefundenen Schwellwertes bestehen, erfolgt durch Polygone, die äußere und innere Konturen der Zeichen repräsentieren.

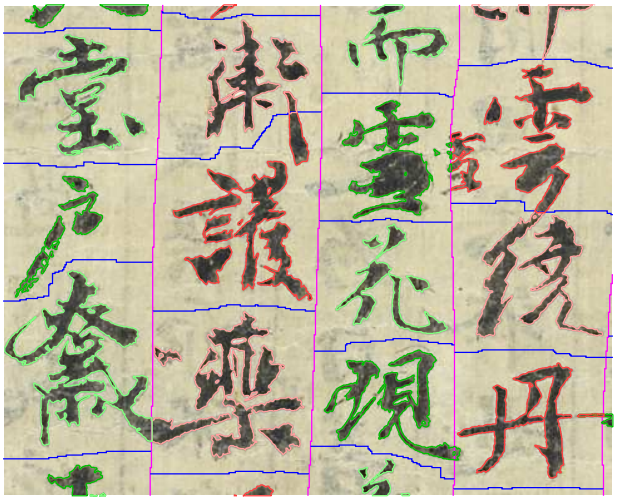
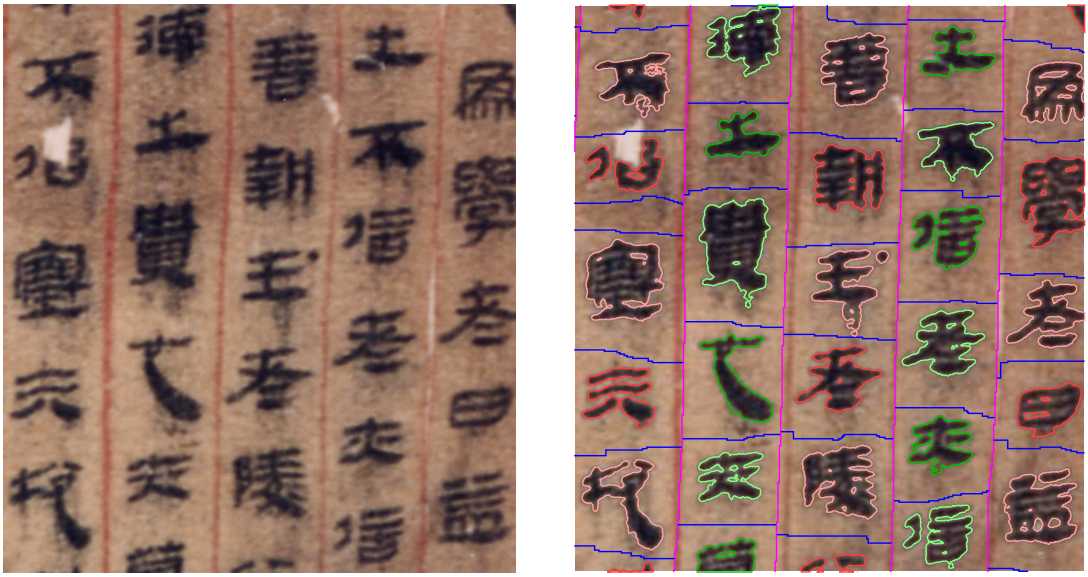


Abbildung 5.14:
Fehler der Flächenbegrenzungen wirken sich direkt auf die erkannten Zeichenumrisse aus.

Fünftel der durchschnittlichen Zeichenhöhe betragen, so wird angenommen, daß es sich bei diesem Label um einen Störeinfluß handelt, wie er etwa durch Bildrauschen entstehen kann, und dieses Label wird verworfen.

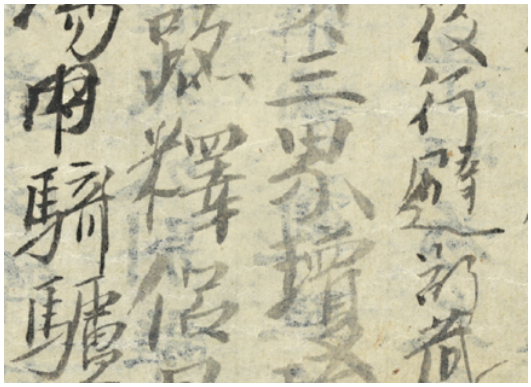
Auch der Durchschnittswert aller Pixel, die dunkler als der errechnete Schwellwert sind, wird durch Anwendung des Algorithmus 3.1 berechnet. Um die Segmentierung von Störeinflüssen möglichst auszuschließen, werden alle Umrisse ignoriert, bei denen die dunkelsten Werte der jeweils umschlossenen Flächen nicht mindestens diesem Durchschnittswert entsprechen.



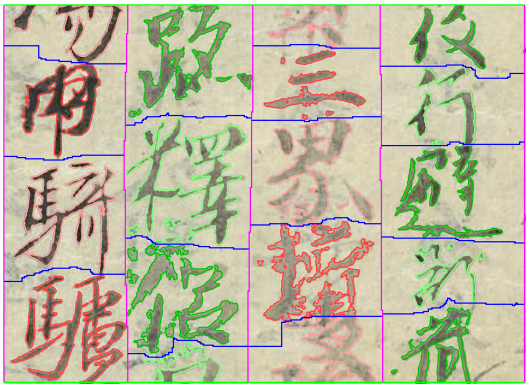
(a) Ausschnitt einer Bildvorlage

(b) Erkannte Zeichen innerhalb dieses Ausschnittes

Abbildung 5.15: Nicht immer führt der automatisch bestimmte Schwellwert zu optimalen Zeichensegmentierungen



(a) Ausschnitt einer weiteren Bildvorlage



(b) Erkannte Zeichen innerhalb dieses Ausschnittes

Abbildung 5.16: Auch hier sind die Kontrastverhältnisse nicht ausreichend; die Erkennung der Zeichen ist ungenügend.

Alle übrigen Label werden mit dem *CrackContourCirculator*, der in der VIGRA enthalten ist, umlaufen, und die Kontur wird in einem Polygon festgehalten.

Nach Betätigung der Schaltfläche „Show shapes“ erfolgt nach dieser Berechnung die Anzeige aller verbleibender Polygone, wobei Polygone benachbarter Flächen farblich unterschiedlich dargestellt werden, was das Auffinden von fehlerhaften Segmentierungen erleichtert und eine Erklärung für im weiteren Verlauf unerwartete Vergleichsergebnisse darstellt. Abbildung 5.13 zeigt die Darstellung dieser Polygone innerhalb der grafischen Oberfläche an einem Beispiel. Die Erklärungen und Möglichkeiten der weiteren Schaltflächen folgen in späteren Abschnitten. Man erkennt in Abbildung 5.13 aber auch, daß bei ausreichenden Kontrastverhältnissen und ohne Bildstörungen eine nahezu perfekte Segmentierung möglich ist. Abbildung 5.14 zeigt dagegen die Probleme der Segmentierung bei sich berührenden Zeichen. Die Trennung von Pinselstrichen direkt an den Linien, die die Flächen begrenzen, führt meistens zu zwei falsch erkannten Polygonen. Auch größere Flecken werden segmentiert.

In Abbildung 5.15 werden dagegen Probleme durch schlechte Kontrastverhältnisse deutlich. Die Vorsegmentierung der Bildvorlage in einzelne Flächen liefert hier zwar stabile Ergebnisse, die Umrisse dieser Zeichen können dagegen nicht zuverlässig bestimmt werden, da die einzelnen Striche eines Zeichens sich nicht deutlich genug vom Hintergrund abheben und so Teile des Hintergrunds als Zeichen erkannt werden.

Genau entgegengesetzt dazu werden einzelne Striche in Abbildung 5.16 nicht erkannt, da diese nur wenig dunkler als der Hintergrund sind. Zudem werden auch Teile von Strichen segmentiert, die durch die scheinbar beidseitige Beschriftung der Schreibunterlage von der Unterseite hindurchscheinen.

Abbildung 5.17: Im rechten Fenster wird die Ähnlichkeit aller Zeichen im Verhältnis zum ausgewählten Zeichen im linken Fenster dargestellt. Die Helligkeit der Zeichen entspricht dabei der Ähnlichkeit; eine Mausbewegung auf einzelne Zeichen gibt die genauen Vergleichskosten aus.

5.2.2 Darstellung von Vergleichsergebnissen

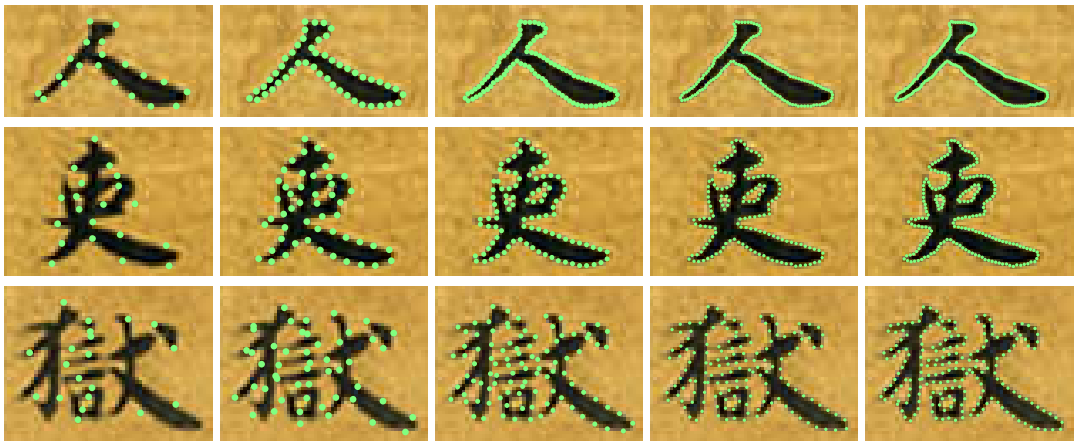
Die Ergebnisse des Zeichenvergleichs werden innerhalb der grafischen Oberfläche wie in Abbildung 5.17 angezeigt.

Im linken Teilfenster kann noch das Zeichen betrachtet werden, welches mit allen Zeichen im rechten Teilfenster verglichen wird. Alle Zeichen im rechten Teilfenster werden farblos dargestellt, die Helligkeit entspricht dabei der errechneten Ähnlichkeit, die anhand der eingestellten[1] Vergleichskriterien ermittelt wurde. Auf schwarzem Hintergrund heben sich die hellsten Zeichen optisch am stärksten hervor, diese entsprechen den als ähnlichste bewerteten Zeichen; unähnlichere Zeichen werden entsprechend dunkler dargesellt.

Die genauen Vergleichskosten können durch Positionierung des Mauszeigers auf die jeweiligen Zeichen abgerufen werden, und erscheinen in roter Schrift unterhalb des rechten Teilfensters. So wird sowohl ein schneller grober Überblick über die gemessene Ähnlichkeit als auch eine genaue Unterscheidung von Zeichen mit sehr ähnlichen Helligkeitswerten ermöglicht.

5.3 Variationen des Shape Context

5.3.1 Variation der Punkteanzahl auf der Kontur



(a) 20 Punkte **(b)** 50 Punkte **(c)** 100 Punkte **(d)** 150 Punkte **(e)** 200 Punkte

Abbildung 5.18: Bestimmung der Kontur unterschiedlich komplexer Zeichen bei unterschiedlicher Anzahl von Punkten.

[1] Die Auswirkungen der möglichen Einstellungen *Anteil der Winkel der Konturnormalen an den Vergleichskosten* und *Zeichentransformaton*, die mittels der grafischen Oberfläche getroffen werden können, werden in den folgenden Abschnitten beschrieben.

Die Autoren des SHAPE CONTEXT-Verfahrens beschreiben lediglich, daß Punkte auf der Kontur in annähernd gleichförmigem Abstand verteilt werden sollen. Es findet sich jedoch kein Hinweis darauf, in welchem Abstand dieses sinnvoll ist. Zur Anschauung werden in Abbildung 5.18 je ein einfaches, ein durchschnittliches und ein komplexes Zeichen gezeigt, dazu eine Verteilung von 20 bis 200 Punkten auf deren Kontur.

Lassen sich einfache Zeichen schon mit etwa 20 Punkten beschreiben, so sind bei komplexeren 150 oder mehr Punkte nötig, um deren Strukturen zu erfassen.

Die Überlegung, einen festen Punktabstand zu bestimmen und so auf komplexeren Zeichen mehr Punkte zu verteilen als auf einfachen, liegt nahe. Probleme ergeben sich daraus allerdings bei der Bildung von Punktepaaren. Auch wenn der Stable-Marriage-Matching Algorithmus durchaus aus $n \times m$ Matrizen Punktepaare bilden kann, so daß bei einem Zeichen N mit n Punkten und einem Zeichen M mit m Punkten bei $(n < m)$ n Punktepaare entstehen, so ist diese Lösung dennoch nicht optimal. Erstens werden bei $(n < m)$ nur mit den Punkten von M Paare gebildet, die am besten zu Punkten aus N passen, was bedeutet, daß genau die am stärksten abweichenden Punkte aus M ignoriert werden. Zweitens fließt auch der Unterschied der Punkteanzahl nicht in die Bewertung ein.

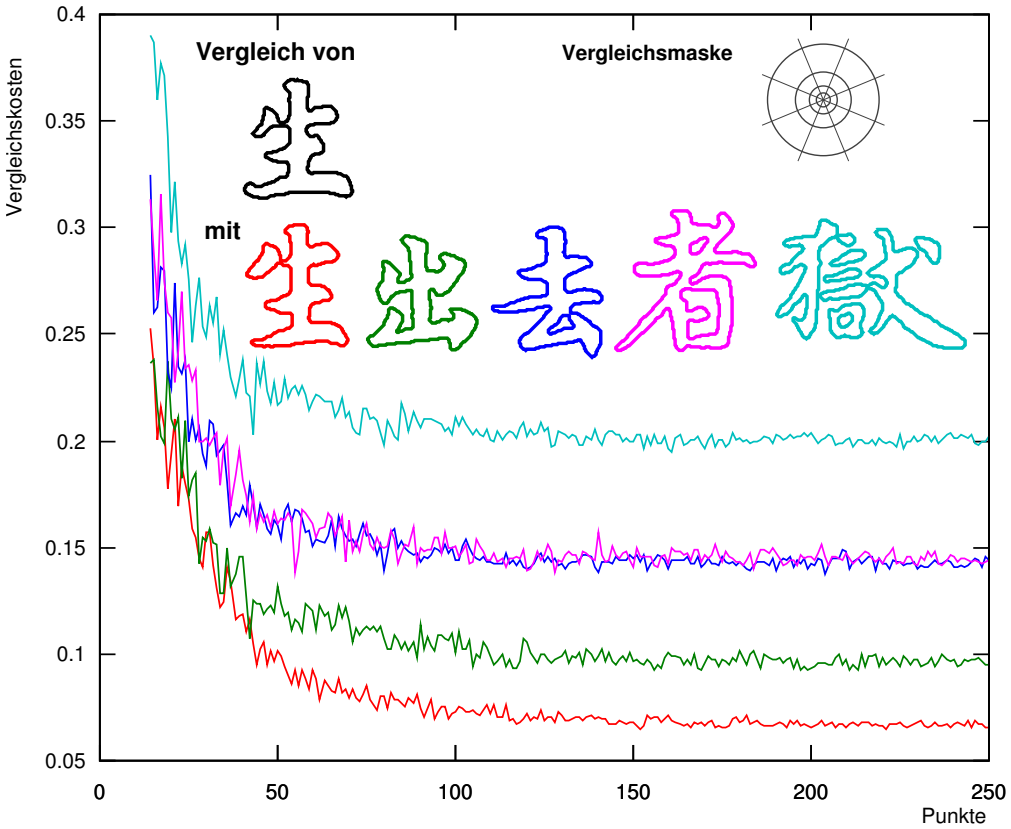


Abbildung 5.19: Vergleichsergebnisse eines einfachen Zeichens mit gleichem (nicht identischem) Zeichen, ähnlichen und unähnlichen Zeichen.

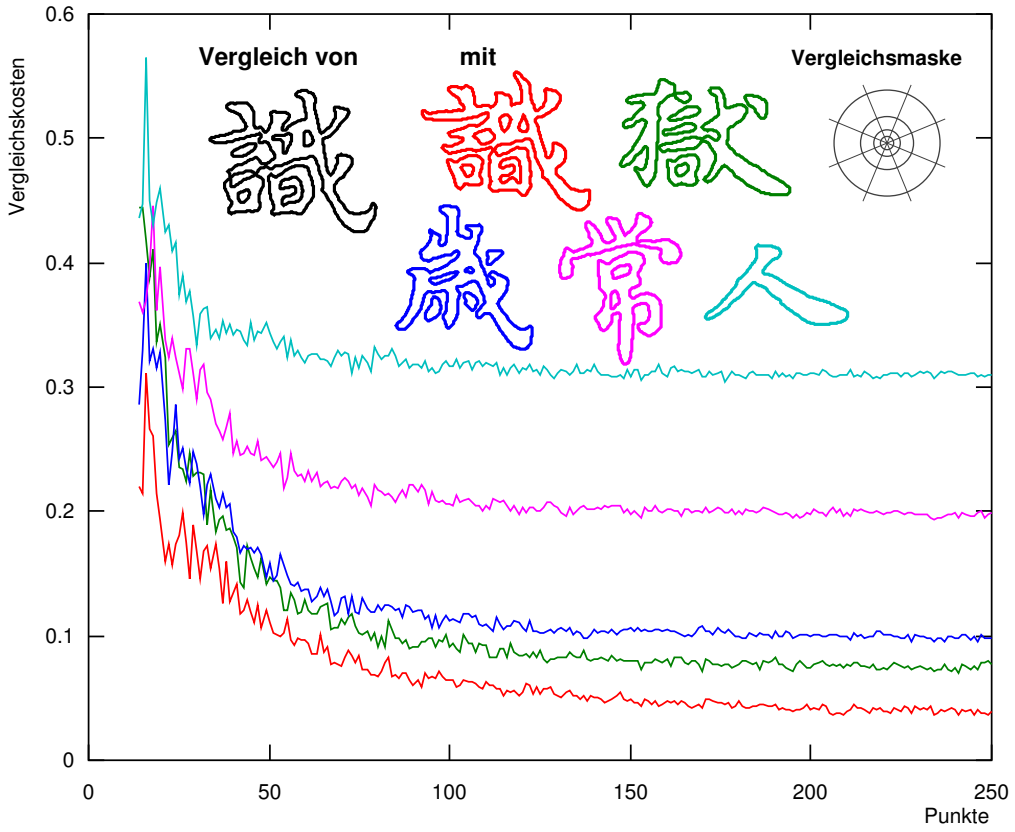


Abbildung 5.20: Vergleichsergebnisse eines komplexen Zeichens mit gleichem (nicht identischem) Zeichen, ähnlichen und unähnlichen Zeichen.

Zwar läßt sich die unterschiedliche Punkteanzahl in das Vergleichsergebnis durch einen festen Kostenbetrag, multipliziert mit dem Betrag der Punktedifferenz, miteinbeziehen, doch ist es schwer festzulegen, wie hoch diese Kosten sein sollten; zudem kann dieses statische Maß nicht die tatsächliche Ähnlichkeit dieser Punkte zu anderen beschreiben.

Aus diesen Gründen wird auf die Anwendung eines festen Punktabstandes verzichtet, stattdessen werden alle Zeichen mit gleicher Punkteanzahl versehen, so daß alle Punkte beider Zeichen paarweise einander zugeordnet werden und somit zum Vergleichskostenmaß beitragen.

Wie in Abbildung 5.13 zu erkennen ist, läßt sich die Anzahl der zu verwendenden Punkte über die grafische Oberfläche vorgeben und die genaue Anordnung der Punkte in ähnlicher Weise wie in Abbildung 5.18 darstellen.

Abbildungen 5.19 und 5.20 sollen Antwort auf die Frage geben, wie viele Punkte auf der Kontur benötigt werden, um sinnvolle Vergleichsergebnisse zu erhalten. Daneben stellen sie Ergebnisse von Vergleichen gleicher, ähnlicher und unähnlicher Zeichen dar. Zur Auswertung läßt sich festhalten:

- Bei Verwendung von wenigen Konturpunkten ergeben sich starke Schwankungen in Abhängigkeit der verwendeten Punktmenge.
- Vergleichsergebnissse mit weniger als 50 Punkten auf der Kontur dieser Zeichen sind unbrauchbar.
- Mit zunehmender Anzahl von Konturpunkten stabilisieren sich die Vergleichsergebnisse, sie tendieren gegen einen festen Wert. Die Verwendung einer deutlich größeren Punktmenge verändert die Ergebnisse also nicht. Der Einfluß einer kleinen Abweichung der Punktmenge ist deutlich geringer als bei Verwendung nur weniger Punkte.
- Vergleiche von oder mit Zeichen mit einfacher Kontur konvergieren bei zunehmender Punkteanzahl schneller gegen einen stabilen Wert als beim Vergleich von Zeichen mit komplexer Struktur.
- Die Verwendung von rund 150 Punkten scheint in allen Fällen ausreichend zu sein; eine Mittelung über den Verlauf bei Verwendung einer unterschiedlichen Punkteanzahl führt zu stabileren Ergebnissen.
- In beiden Grafiken sind die Vergleichskosten zwischen gleichen Zeichen am niedrigsten. Sehr unähnliche Zeichen ergeben hohe Vergleichskosten. In einigen Fällen sind Vergleichskosten beim Vergleich eines Zeichens mit unterschiedlichen, nicht ähnlichen Zeichen nach menschlichem Empfinden nicht immer sofort nachvollziehbar.

5.3.2 Variationen der Zonenmaske

Einteilung der Radien

Auch die Anzahl der Zonen der Maske wurde von den Urhebern des SHAPE CONTEXT-Verfahrens nicht exakt festgelegt. Es wird lediglich vorgeschlagen, die Radien der Ringeinteilung in logarithmischen Abständen anzuordnen, so daß Punkte in näherer Umgebung stärkeren Einfluß auf den SHAPE CONTEXT nehmen als weiter entfernte Punkte. Darüber hinaus wird vorgeschlagen, die Radien in Abhängigkeit der Zeichengröße zu verteilen, um damit eine Größeninvarianz zu erreichen.

Als Maß M für die Zeichengröße wird der Median der Abstände von allen Punkten eines Zeichens zu jeweils allen anderen Punkten desselben Zeichens vorgeschlagen.

Wenn die äußerste Zone der Maske alle Punkte enthält, die weiter entfernt sind als $2 \times M$, die darauf folgende Zone alle Punkte enthält, deren Entfernung größer als M ist, die darauf folgende Zone alle Punkte mit Entfernung größer $\frac{M}{2}$ enthält, eine weitere Zone alle Punkte mit größerer Entfernung als $\frac{M}{4}$ und die innerste Zone alle übrigen Punkte aufnimmt, so erhält man 5 Zonen.

Die gemittelte Aufteilung der Punkte in diese Zonen beträgt:

Zone:	Äußerste	2	3	4	Innerste
Anteil:	4,49%	45,66%	33,11%	11,93%	4,81%

Diese Aufteilung liefert unter der Berücksichtigung der Abstände mit logarithmischer Abstufung und der damit gewollten höheren Wertung der Punkte in näherer Umgebung eine sinnvolle Aufteilung. Abbildung 5.25 zeigt Masken mit Radien dieser Einteilung.

Es wurden keine Versuche mit hiervon abweichenden Radien unternommen, wie es z.B. durch vier Zonen bei größerer Logarithmusbasis oder bei kleinerer Basis mit sechs Zonen denkbar wäre.

Einteilung der Winkel

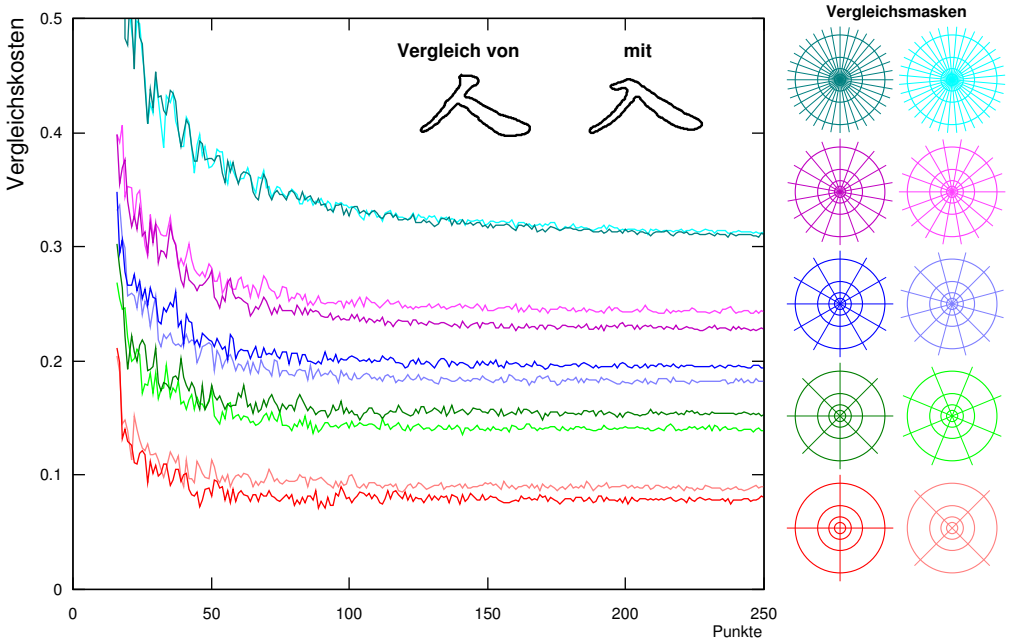


Abbildung 5.21: Vergleichsergebnisse von zwei ähnlichen einfachen Zeichen bei Einteilung der Maske in 4, 8, 12, 18 und 36 Winkelzonen.

Die Auswirkungen eines Vergleichs von zwei einfachen, gleichen Zeichen bei Verwendung von Masken mit unterschiedlicher Winkeleinteilung zeigt Abbildung 5.21.

Auch hieraus können folgende Punkte festgehalten werden:

- Bei Einteilung der Maske in wenige Zonen ergeben sich niedrigere Vergleichskosten, diese steigen mit zunehmender Anzahl an Zonen.
- Die Konvergenz gegen einen stabilen Kostenwert erfolgt bei weniger Zonen schon bei geringerer Punkteanzahl als bei Einteilung in viele Zonen.
- Nicht nur die Anzahl an Zonen beeinflußt das Vergleichsergebnis, auch deren Anordnung ist entscheidend. Die Vergleichskosten werden paarweise mit jeweils zwei Masken gleicher Zonenanzahl dargestellt. Von den Masken im gleichen Farbton teilt die Maske geringerer Helligkeit mit einer vertikal verlaufenden Achse jeweils vier Zonen auf, während die heller dargestellte Maske genau zwei vertikal zentrierte Zonen aufweist.

Die Ursache der im letzten Punkt festgestellten unterschiedlichen Ergebnisse kann Abbildung 5.22 näher erläutern. Bei chinesischen Zeichen gibt es vorwiegend die horizontale, vertikale oder jeweils diagonale Strichführung. Bei ungünstiger Zoneneinteilung trifft der Konturverlauf in einer dieser Richtungen genau auf eine Grenze zwischen Winkelzonen, so daß Punkte der Kontur trotz überwiegend geradlinigem Verlauf in mehrere Zonen fallen.

Natürlich wird dieser Effekt besonders deutlich bei einfachen Zeichen, die nur Striche dieser Hauptrichtungen aufweisen, wie bei den beiden in Abbildung 5.21 dargestellten Zeichen mit fast ausschließlich diagonal verlaufender Kontur. Bei besonders komplexen Zeichen oder einer Einteilung der Maske in besonders viele Zonen gleichen sich die Ergebnisse dagegen einander an. (Siehe dazu auch Abbildung 5.25).

Abbildung 5.23 zeigt den Vergleich der gleichen Zeichen wie in Abbildung 5.19 bei Verwendung unterschiedlicher Masken. In allen zehn Fällen wird das Zeichen mit schwarz gefärbter Kontur mit den fünf übrigen Zeichen mit jeweils einer Maske identischer Zonenanzahl und Anordnung verglichen.

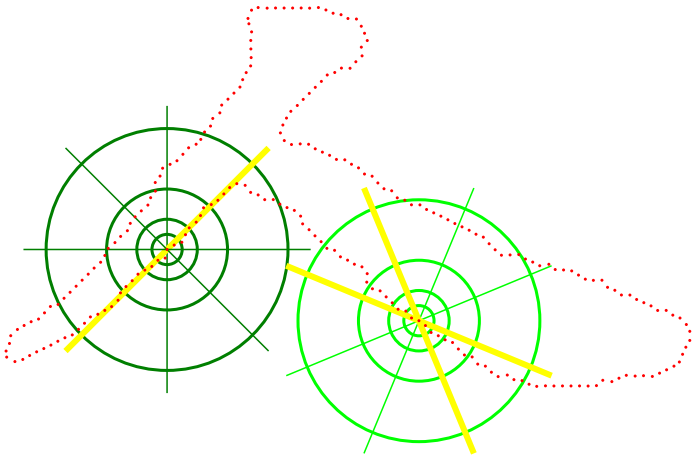


Abbildung 5.22:
Während der diagonale Konturverlauf bei Einteilung der dunkelgrünen Maske genau auf die Grenze der Winkelzonen trifft, wird die Kontur durch die hellgrüne Maske zonenmittig erfaßt. (Die Radien der Masken sind verkleinert dargestellt.)

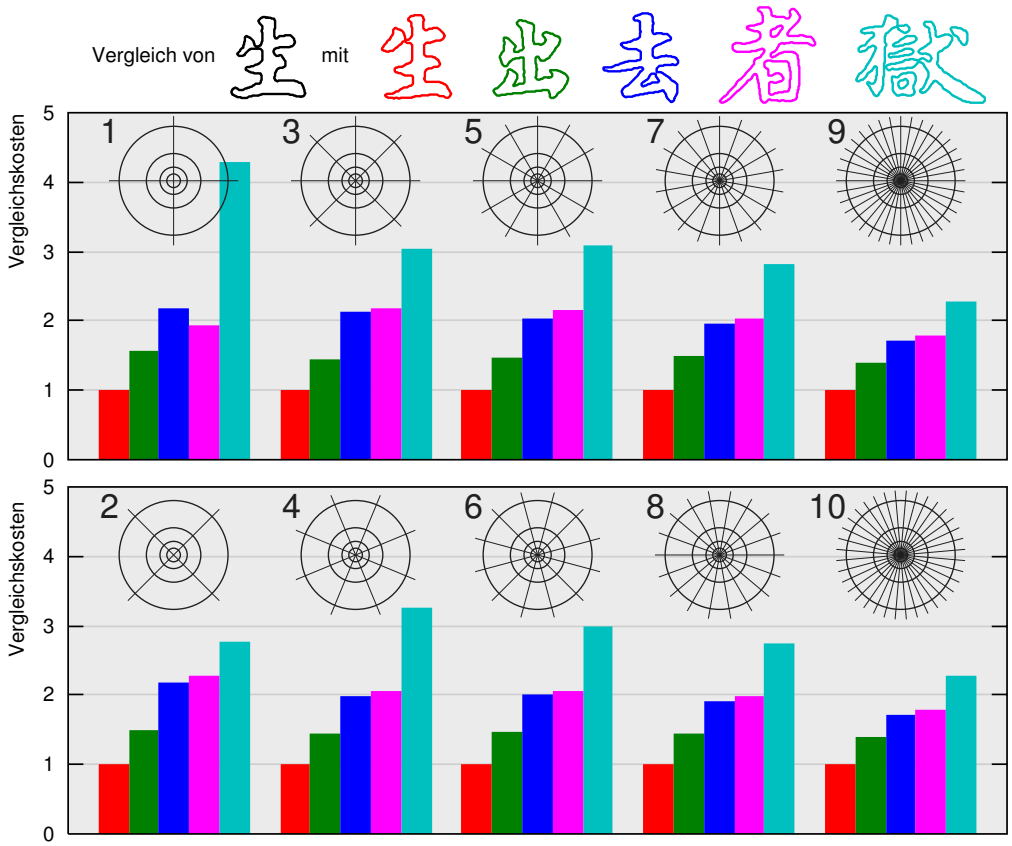


Abbildung 5.23: Vergleichsergebnisse eines einfachen Zeichens mit gleichem (nicht identischem) Zeichen, ähnlichen und unähnlichen Zeichen bei Einteilung der Maske in 4, 8, 12, 18 und 36 Winkelzonen in normalisierter Darstellung.

Da die Konvergenz gegen einen Vergleichswert sowohl bei Verwendung komplexer Zeichen als auch bei Verwendung komplexer Masken eine größere Anzahl an Konturpunkten vorraussetzt, wurden die Ergebnisse dieses Vergleichs genau wie in der folgenden Abbildung 5.24 ermittelt bei Verwendung des Durchschnitts der Vergleichskosten aus 290-310 Konturpunkten.

Die ideale Maske würde sehr geringe Vergleichskosten bei Verwendung gleicher Zeichen ergeben, während sehr unähnliche Zeichen sehr hohe Kosten verursachen sollten.

Wie in Abbildung 5.21 zu erkennen ist, steigen die Vergleichskosten auch gleicher Zeichen bei Verwendung von Masken mit höherer Zonenanzahl an. Um diesen Effekt zu kompensieren, wurden die Vergleichsergebnisse normalisiert, d.h. die Kosten der Zeichenvergleiche werden bei allen Masken im Verhältnis zu den Kosten des Vergleichs gleicher Zeichen bei Verwendung derselben Maske dargestellt, um den Kostenunterschied herausstellen zu können.

Bei Verwendung der Maske (1) ergeben sich hohe Kosten beim Vergleich mit einem kom-

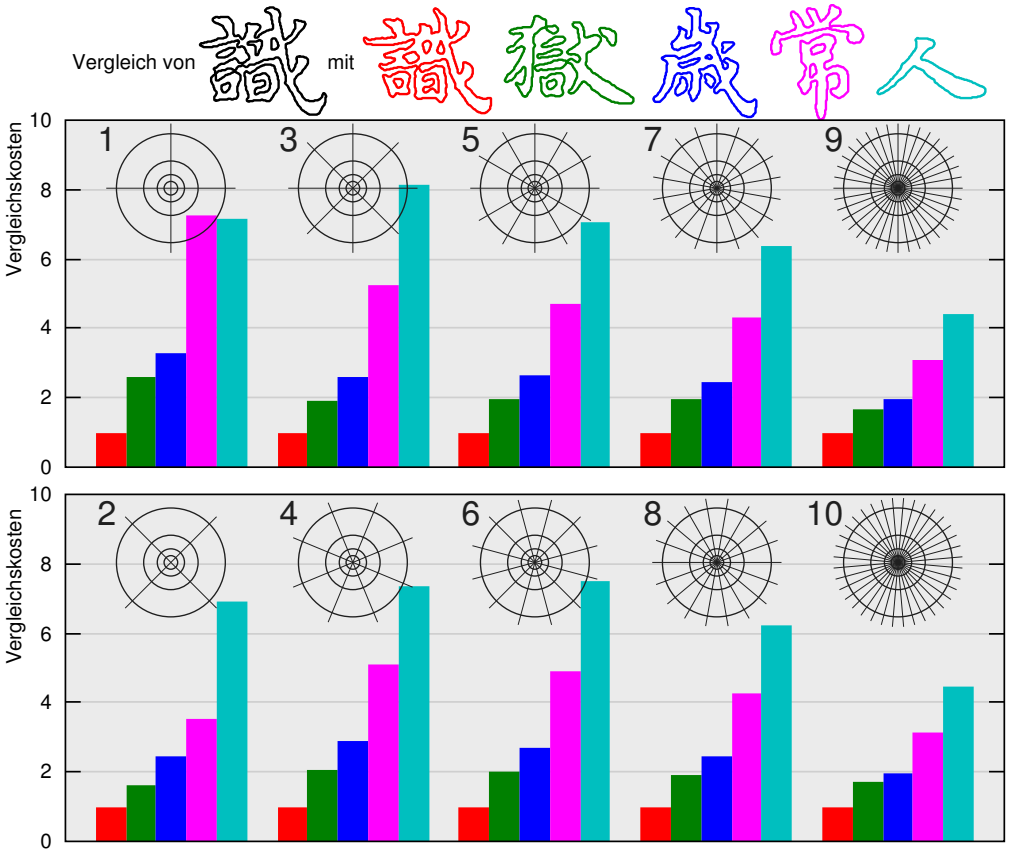


Abbildung 5.24: Vergleichsergebnisse eines komplexen Zeichens mit gleichem (nicht identischem) Zeichen, ähnlichen und unähnlichen Zeichen bei Einteilung der Maske in 4, 8, 12, 18 und 36 Winkelzonen in normalisierter Darstellung.

plexen Zeichen, die Stärke der Ähnlichkeit anderer Zeichen wird vertauscht im Vergleich zur Verwendung aller übrigen Masken. Die Auswertung von Maske (2) ergibt geringe Kostenabstände der drei unähnlichsten Zeichen.

Recht gleichwertige Ergebnisse liefern die Masken (3)-(8); die Masken (7) und (8) mit 18-facher Winkelunterteilung weisen dabei die geringsten Abstände auf.

Der in Abbildung 5.22 veranschaulichte Effekt bei Verwendung der Maske (4) zeigt nur wenig Auswirkungen, nur beim Vergleich des komplexen Zeichens sind die Kosten am höchsten von allen sechs Masken.

Die Masken (9) und (10) mit Einteilung in 10° Zonen liefern die schlechtesten Ergebnisse, die Vergleichsunterschiede sind am geringsten.

Sehr ähnliche Ergebnisse liefert die Gegenüberstellung der Zeichen aus Abbildung 5.20,

Abbildung 5.24 veranschaulicht diese Ergebnisse.

Auch hier weichen die Ergebnisse der Masken mit Einteilung in 90° Zonen teilweise deutlich von denen mit anderen Masken ab. Maske (3) scheint beim Vergleich mit dem einfachen Zeichen das beste Ergebnis zu liefern, Maske (4) errechnet höhere Unterschiede der übrigen Zeichen. Die übrigen Masken ergeben mit höherer Anzahl an Winkelzonen geringere Vergleichsunterschiede; auch hier schneiden die Masken (9) und (10) mit 10°-Einteilung am schlechtesten ab.

Somit sollte die Maske (4) trotz geringer Komplexität und damit niedrigem Rechenaufwand beim Vergleich der Anzahl der in den einzelnen Zonen enthaltenen Punkte stabile Ergebnisse liefern, welche nur in Einzelfällen von Masken mit anderer Zonenanzahl oder Anordnung übertroffen werden können. Für alle folgenden Untersuchungen wird daher diese Maske gewählt.

5.4 Variation des Kostenmaßes

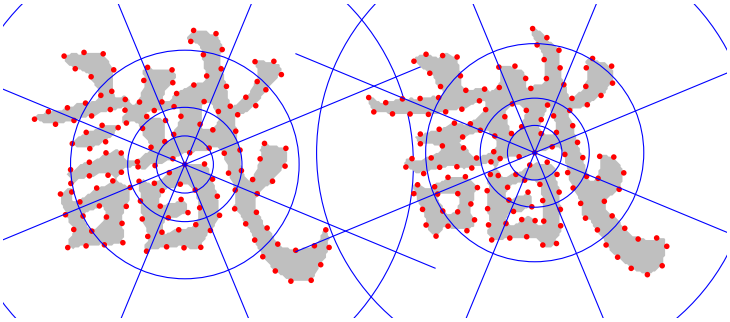


Abbildung 5.25:
Die SHAPE CONTEXTe der beiden Masken sind einander recht ähnlich, obwohl die erste Maske am oberen Rand einer Kontur liegt, die zweite am unteren.

Wie in Abbildung 5.25 zu erkennen ist, weisen zwei SHAPE CONTEXTe eine hohe Ähnlichkeit auf, wenn die übrigen Konturpunkte ähnlich um sie verteilt sind. Da bei diesem Vergleich lediglich die Umgebung des aktuellen Punktes maßgeblich ist, könnten die beiden hervorgehobenen Punkte bei einem Vergleich einander zugeordnet werden, obwohl der Verlauf der Kontur an diesen deutlich unterschiedlich ist. Wie schon in Abschnitt 4.2.6 angeführt, sollte es durch die Einbindung der Konturnormalen an der Position des jeweiligen Punktes möglich sein, die Vergleichsergebnisse durch diese Zusatzinformation zu verbessern.

Auch wenn die Größe der Winkel zweier Normalen einen Unterschied von bis zu 360° aufweisen könnte, kann deren Differenz aufgrund der Zyklizität höchstens 180° betragen. Für diese „entgegengesetzte“ Richtung werden Kosten von 1 angenommen, Kosten von 0 bedeuten vollkommene Übereinstimmung. Mit einem Faktor f können diese Kosten der Ausrichtung der Normalen K_N mit den Kosten K_S des SHAPE CONTEXT-Vergleichs wie folgt kombiniert werden:

$$K_{komb} \;=\; f \cdot K_N + (1 - f) \cdot K_S$$

Somit ergeben sich wieder Gesamtkosten von maximal 1.

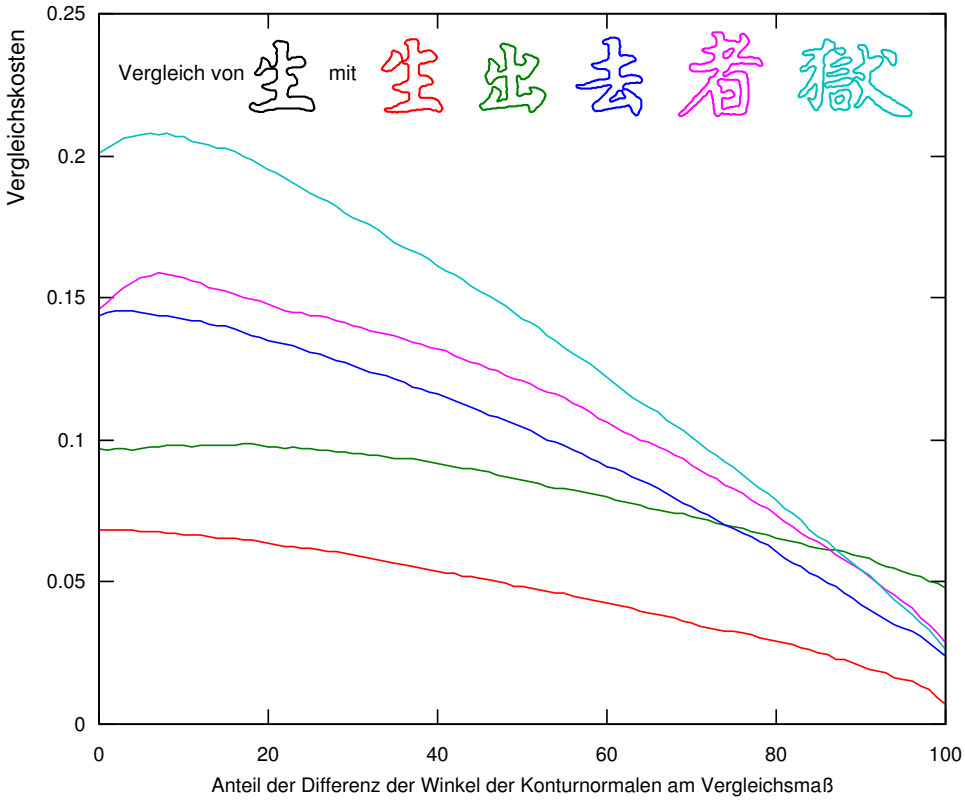


Abbildung 5.26: Gegenüberstellung der Vergleichskosten bei Verwendung unterschiedlicher Anteile von Kostenmaßen an den gesamten Vergleichskosten.

Abbildung 5.26 stellt die Vergleichskosten der bekannten Zeichen unter Verwendung aller Kombinationen dieser Kosten gegenüber. Dazu wurde auf den Durchschnitt der Kosten bei Verwendung zwischen 140 und 160 Konturpunkten zurückgegriffen. Testweise wurde auch eine Versuchsreihe mit durchschnittlichen Kosten zwischen 290 und 310 Konturpunkten angelegt; die Ergebnisse dieser waren mit obigen nahezu deckungsgleich.

Auffällig ist ein Anstieg der Kosten bei der magenta- und cyanfarbenen Kurve mit einem Maximum bei einem Anteil von 7-8% des Konturnormalenvergleichs an den Gesamtkosten. Daraufhin fallen die Kurven wieder ab. Die blaue Kurve steigt nur sehr gering an, die grüne und rote fallen sofort ab. Allen Vergleichskosten gemeinsam sind die deutlich geringeren Kosten bei reiner Verwendung der Winkel der Konturnormalen als Vergleichsmaß. Da an dieser Stelle Punkte an beliebiger Position der Konturen einander zugeordnet werden können, soweit deren Konturnormale ausreichende Übereinstimmung aufweist, scheint die Findung korrespondierender Punktepaare deutlich einfacher zu werden. Es ergaben sich bei einem untersuchten Beispiel beim Vergleich von 150 Konturpunkten immerhin 67 Paare mit Winkeldifferenzen von $<1°$, weitere 45 mit $<10°$, und nur 14 mit $>30°$, wovon nur 1 Paar einen Unterschied von mehr als 90° aufwies.

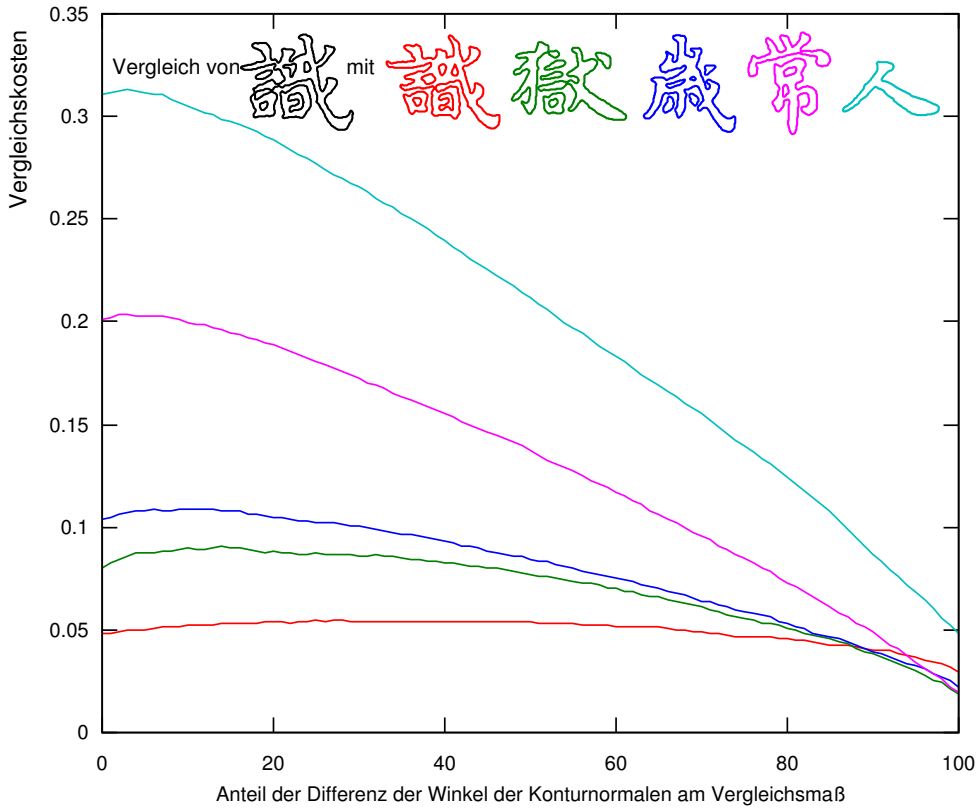


Abbildung 5.27: Gegenüberstellung der Vergleichskosten bei Verwendung unterschiedlicher Anteile von Kostenmaßen an den gesamten Vergleichskosten.

Ein ähnliches Bild ergibt der Vergleich des komplexen Zeichens mit den bekannten übrigen Zeichen (Abbildung 5.27); auch hier wurde der Mittelwert der Kosten zwischen 140 und 160 Konturpunkten bei gleicher Maske verwendet.

Der Kostenanstieg ist bei zunehmendem Einfluß des Konturnormalenvergleichs in die Gesamtkosten bei allen Kurven relativ gering. Wie auch schon bei Abbildung 5.26 festzustellen war, verlaufen die Kurven beim Vergleich ähnlicher Zeichen deutlich flacher; der Abstand aller Kurven ist bei reinem Vergleich der Konturnormalen der Punkte am geringsten.

Die Ursachen für diese Kurvenverläufe sind darin zu suchen, daß die Findung korrespondierender Punktepaare bei hohem Anteil der Vergleiche von SHAPE CONTEXTen an den Kosten eben anhand dieser SHAPE CONTEXT-Paare gesucht werden; die Vergleichskosten werden durch anhand der Konturnormalen schlecht korrespondierender Paare lediglich erhöht. Umso geringer der Anteil der SHAPE CONTEXT Vergleiche an den Kosten wird, umso mehr Punkte werden allein durch Übereinstimmung ihrer Konturwinkel einander zugeordnet, es finden nun also neue Punktezuordnungen statt, deren Auffindung deutlich einfacher ist und insgesamt geringere Kosten verursacht.

Aus diesen Versuchen kann man festhalten, daß die Miteinbeziehung der Konturnormalen in die Kosten mit rund 10% die Ergebnisse durchaus verbessern kann (so werden in Abbildung 5.26 das blaue und magentafarbene Zeichen besser unterschieden), ein höherer Einfluß sich aber auf die Unterscheidung eher nachteilig auswirkt.

5.5 Transformation von Zeichen

Die Zuordnung von Punkten zweier Zeichen zueinander liefert Vektoren, aus denen sich über eine VIGRA-Funktion eine Transformationsmatrix für eine affine Transformation bilden läßt. Vektoren, deren Vergleichskosten der zugehörigen Punktepaare höher sind lassen sich dabei geringer bewerten als Vektoren, deren Punktepaare anhand deren Vergleichskosten eine bessere Übereinstimmung versprechen.

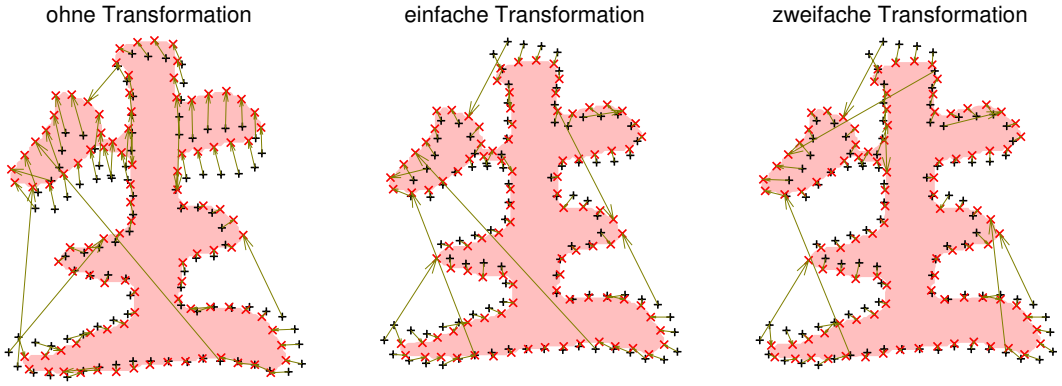


Abbildung 5.28: Abgleich der Punktepaare zweier gleichartiger einfacher Zeichen.

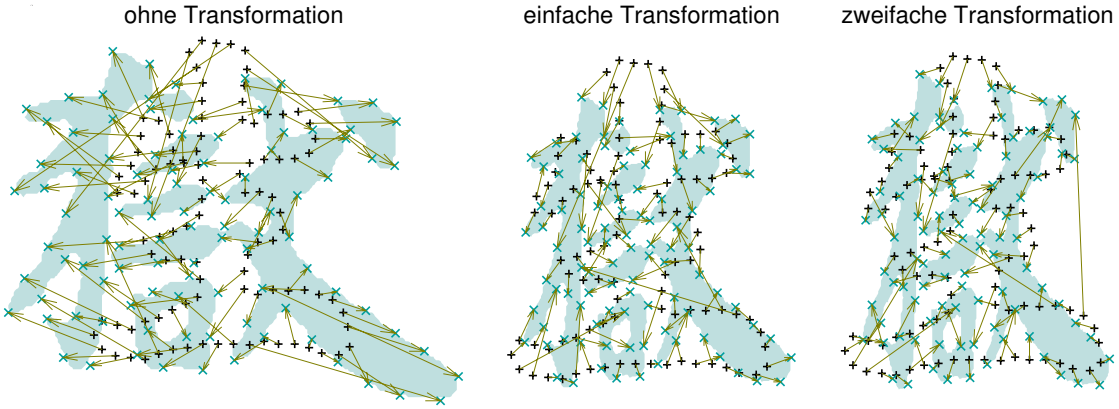


Abbildung 5.29: Abgleich der Punktepaare eines einfachen Zeichens mit einem komplexen.

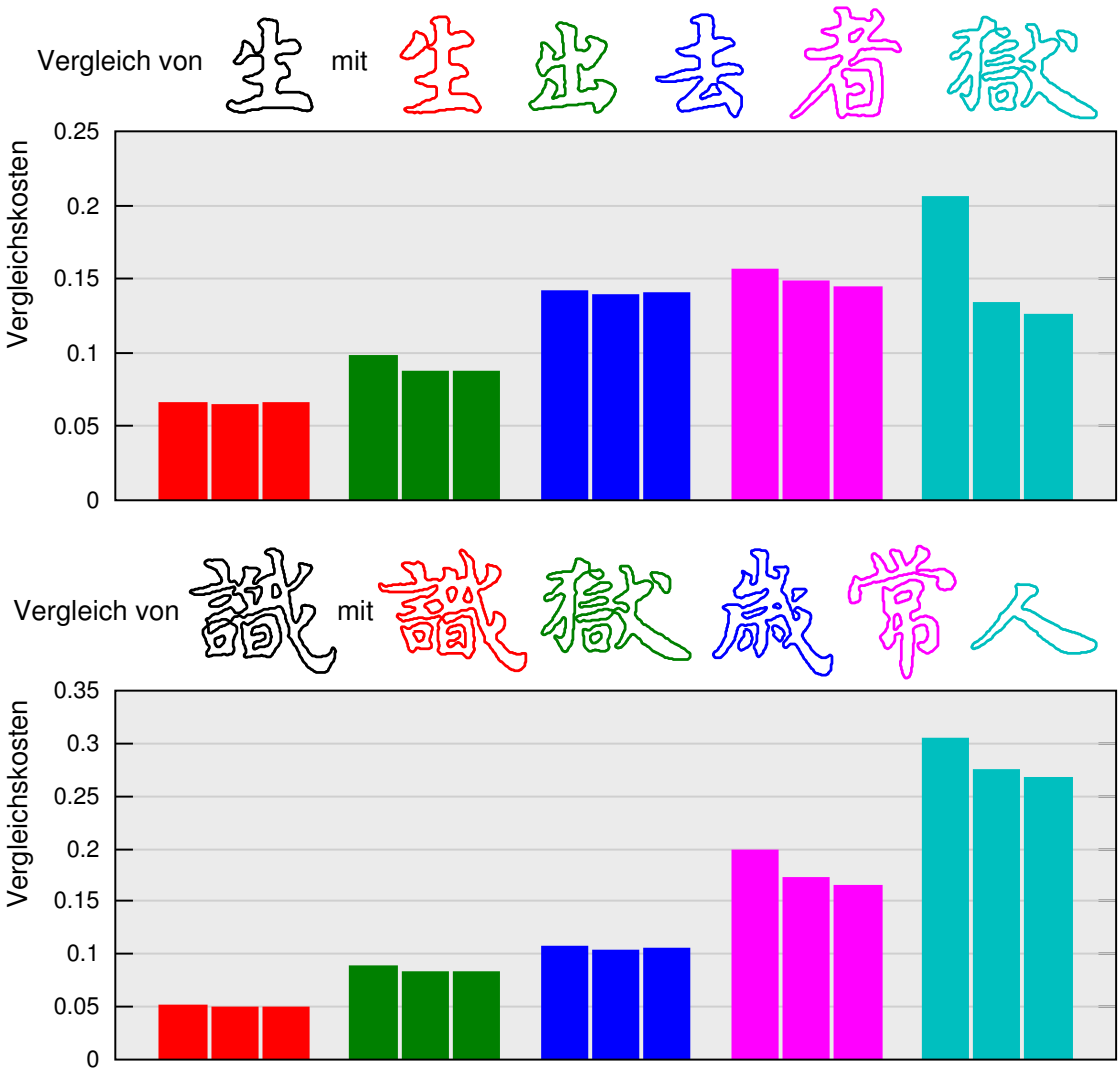


Abbildung 5.30: Vergleiche von Zeichen vor und nach affinen Transformationen.

Werden zwei gleichartige Zeichen durch eine affine Transformation aufeinander abgebildet, so sollte die Übereinstimmung deutlich höher werden, als sie bei ungleichen Zeichen möglich ist. Auch dieses Vorgehen verfolgt das Ziel, ähnliche Zeichen besser von unähnlichen unterscheiden zu können. Auch wenn bei solchen Transformationen evtl. Eigenarten eines Schreibers verloren gehen, so sollte es doch möglich sein, zunächst gleichartige Zeichen zu erkennen und festzuhalten. Zur Bewertung der Unterschiede können weiterhin die nicht-transformierten Zeichen verwendet werden, allerdings ist die Untersuchung von Unterschieden gleichartiger Zeichen wertvoller als die der Unterschiede verschiedenartiger.

In Abbildung 5.28 wird ein einfaches Zeichen auf ein gleichartiges transformiert, in Abbildung 5.29 wird die Transformation eines komplexen Zeichens auf dasselbe einfache versucht.

Jeweils links dargestellt werden die Zeichen vor deren Anpassung; gelbe Pfeile stellen die Punktezuordnungen dar. Anhand dieser Zuordnungen wird darauf die Transformationsmatrix erstellt, die das verglichene Zeichen wie in den Abbildungen mittig dargestellt transformieren läßt. Aus dieser transformierten Repräsentation lassen sich nun erneut SHAPE CONTEXTe bilden, anhand derer sich neue Punktepaare bilden lassen. Diese neuen Zuordnungen können als Vektoren erneut eine Transformationsmatrix bestimmen, die eine weitere Transformation zuläßt, wie sie jeweils rechts in den Abbildungen gezeigt wird. Theoretisch lassen sich diese Vorgänge beliebig oft wiederholen, jedoch bewirkt nur die erste Transformation nennenswerte Unterschiede.

Die durch die Transformationen veränderten Vergleichskosten werden in Abbildung 5.30 veranschaulicht. Die Kosten wurden auch hier gemittelt aus 140-160 Konturpunkten, und das Vergleichsmaß wurde zu 10% vom Abgleich der jeweiligen Konturnormalen beeinflußt.

Die in der Abbildung dargestellten Balken entsprechen dem Vergleich mit den farblich korrespondierenden Zeichen, die sich jeweils über der Grafik befinden. Der linke Balken stellt jeweils die Vergleichskosten ohne Transformation dar, der mittlere bei einfacher Transformation, der rechte bei zweifacher. Die Kosten werden in allen Fällen als absolute Beträge angezeigt.

Wie zu erwarten war, sinken die Kosten in allen Fällen nach einmaliger Transformation. Die Änderung der Kosten zwischen der ersten und zweiten Transformation ist in allen Fällen recht gering, daher wurde auf weitere Transformationen verzichtet.

Deutlich wird bei diesen Vergleichen allerdings, daß besonders die Zeichen, die vor der Transformation sehr hohe Unähnlichkeiten aufwiesen, durch die Transformation am meisten profitieren, d.h. ihre Vergleichskosten reduzieren sich am stärksten, während Vergleichskosten von zueinander gleichartigen Zeichen sich nur in sehr geringem Umfang ändern. Letzteres kann als Hinweis auf die Stabilität des Verfahrens gegenüber geringen Veränderungen interpretiert werden. Nur wenig abweichende Ergebnisse wurden bei Verwendung einer deutlich höheren Punkteanzahl festgestellt.

Somit wird scheinbar durch die affine Transformation ein Ergebnis erzielt, das im genauen Gegensatz zum angestrebten Ziel steht. Ohne die Stärke einer Transformation in die resultierenden Vergleichskosten miteinzubeziehen, sollte daher von der Anwendung affiner Transformationen Abstand genommen werden, um optimale Ergebnisse zu erhalten.

5.6 Verwendung der Zeichenskelette

Wie bereits in Abschnitt 4.2.6 beschrieben, kann jedes Zeichen auf sein Skelett reduziert werden. Werden dann die Punkte statt auf der Kontur auf dem Skelett angeordnet, können die Zeichen auf gleiche Weise miteinander verglichen werden.

Die Skelette stellen dabei nur die grundlegenden Eigenschaften der Zeichen dar; gleichartige Zeichen sollten so eine noch höhere Ähnlichkeit aufweisen, wenn sie sich hauptsächlich durch die Dicke ihrer Pinselstriche unterscheiden, da diese Information nicht berücksichtigt wird. Trotz dieser Reduzierung sollte das Verfahren tolerant gegenüber Bildfehlern sein; auch eine über eine

Abbildung 5.31:
Beispiele für die Skelettierung von Zeichen.

kurze Distanz unterbrochene Linie sollte sehr ähnlich zu einer ununterbrochenen Linie sein, da nur die Punkte auf dieser Linie – unabhängig von der Reihenfolge der Anordnung der Punkte – miteinander verglichen werden.

Eine Auswertung eines Zeichenvergleichs anhand der Zeichenskelette zeigt Abbildung 5.32. Auch hier erfolgt ein Vergleich der bekannten Zeichen, deren Farbe mit den jeweiligen Kostenbalken übereinstimmen. Für alle Balken werden die Kosten gemittelt aus den Vergleichskosten von 140 bis 160 Punkten errechnet. Zur Bedeutung der Balken im einzelnen:

- Der rechte Balken wiederholt zum besseren Vergleich noch einmal die Kosten der Punkteverteilung auf der Kontur unter Miteinbeziehung eines Kostenanteils des Konturnormalenabgleichs von 10%. Hier erfolgte keine affine Transformation.

- Der mittlere Balken repräsentiert die Vergleichskosten der Punkteverteilung auf den Skeletten, die den in der Abbildung dunkel markierten Linien entsprechen.

- Der rechte Balken entspricht dem mittleren, im Gegensatz zu diesem wurde hier eine affine Transformation der Skelette durchgeführt, die in gleicher Weise wie zuvor beschrieben auch hier anwendbar ist.

Ein Vergleich der nicht-transformierten Skelette mit dem einfachen Zeichen, wie es in der oberen Hälfte der Abbildung dargestellt ist, zeigt eine veränderte Reihenfolge der Ähnlichkeit. Das in blau eingefärbte Zeichen ist nun zum schwarzen das ähnlichste nicht gleichartige Zeichen, die grün und magenta gefärbten Zeichen folgen mit geringem Unterschied zueinander. Der Vergleich mit dem cyanfarbenen komplexen Zeichen zeigt eine deutlich größere Unähnlichkeit als beim Konturvergleich. Die Auswirkungen der affinen Transformation sind durchweg gering, allerdings wird die ursprüngliche Reihenfolge der Ähnlichkeiten wiederhergestellt. Nur das komplexe Zeichen wird, ähnlich wie auch bei der Anordnung der Punkte auf der Kontur, durch die Transformation deutlich ähnlicher; der Abstand zu den anderen Zeichen wird sehr gering.

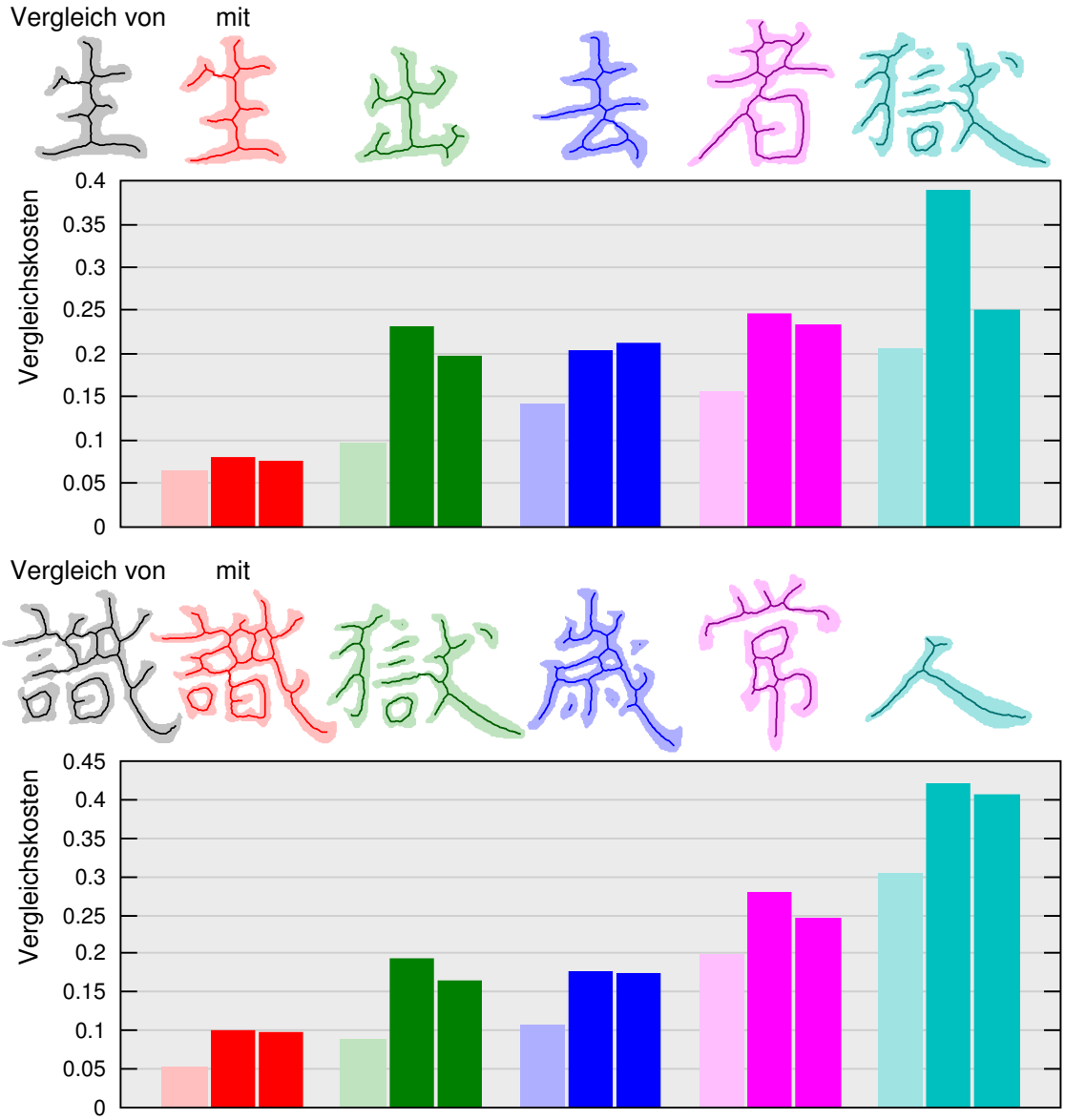


Abbildung 5.32: Vergleiche von skelettierten Zeichen. Erläuterungen im Text.

Der untere Teil von Abbildung 5.32 vergleicht das komplexe Zeichen mit den bekannten übrigen Zeichen. Zunächt muß festgestellt werden, daß die Vergleichskosten mit dem gleichartigen roten Zeichen nahezu doppelt so hoch ausfallen, hingegen steigen die Vergleichskosten bei den übrigen Zeichen, abgesehen vom grünen, nicht im gleichen Verhältnis. Somit wird der relative Unterschied des Vergleichs zweier gleichartiger Zeichen geringer zum Vergleich von unterschiedlichen Zeichen, was bei dem rechts dargestellten, einfachen Zeichen, besonders deutlich wird.

Auch hier ändert sich die Reichenfolge der Ähnlichkeit, das blau gefärbte Zeichen ist ähnlicher als das grüne.

Die Auswirkungen der affinen Transformation sind bei allen Zeichen gering, allerdings wird auch hier die Reihenfolge der Ähnlichkeit wieder auf die bekannte Anordnung verändert.

Bei genauer Betrachtung der erstellten Skelette fällt auf, das einzelne Zeichenkomponenten, die teilweise zufällig oder mit sehr dünnen Pinselstrichen miteinander verbunden sind, einen durchgehenden Skelettstrang aufweisen, wohingegen andere Komponenten ohne direkte Verbindung, aber mit äußerst geringem Abstand zueinander, in der Skelettierung deutlich getrennt werden. Zwar wird dieser Unterschied dadurch abgeschwächt, daß auf diesen Linien nur Punkte verteilt werden, die dann einzeln miteinander verglichen werden, trotzdem wird das Vergleichsergebnis von dieser Eigenart der Skelette beeinflußt.

Etwas heikel ist die Skelettierung der Anfangs- und Endbereiche der Pinselstriche, besonders in breit zulaufenen Enden wird durch das gewählte Skelettierungsverfahren der dort zulaufende Skelettstrang in einzelne kurze Stränge aufgeteilt, die in unterschiedliche Richtungen verlaufen.

Zu jedem Skelettstrang kann bestimmt werden, durch welches Konturteilstück dieser bestimmt wurde. Mit einem Parameter, dessen Größe sich in dieser Arbeit an der jeweiligen Zeichengröße orientiert, werden nun einzelne, kurze Skelettstränge verworfen, die durch einen kürzeren Konturverlauf bestimmt wurden, als durch den Parameter festgelegt wurde. Detailiertere Informationen zur Vereinfachung der Skelette liefert Prasad [1997].

Durch dieses Vorgehen lassen sich deutlich bessere Skelettierungsergebnisse erzielen, auch wenn sie noch nicht in allen Fällen zufriedenstellend sind. So weist das in der Abbildung 5.32 untere cyanfarbene Zeichen am oberen Ende zwei Skelettstränge auf, die beide etwas länger als der eingestellte Parameter sind und somit beide erhalten bleiben.

Bei den oben dargestellten schwarzen Zeichen entscheidet der Algorithmus für ein nach links gerichtetes oberes Skelettende, beim roten Zeichen dagegen ist es nach rechts gerichtet. Schon sehr geringe Änderungen der Kontur können also deutliche Änderungen des Skeletts hervorrufen.

Als Resultat aus Abbildung 5.32 kann festgehalten werden, daß die Beschränkung auf die Zeichenskelette durchaus als Alternative zur Kontur der Zeichen gesehen werden kann, um darauf die Vergleichspunkte zu positionieren. Vergleichsergebnisse zeigen sich hier allerdings weniger tolerant bei leichten Veränderungen der Zeichen. Berühren sich Teile von Zeichen, beeinflussen diese stärker das Vergleichsergebnis als bei den Konturpunkten.

Veränderungen durch eine affine Transformation können bei der Anpassung der Skelette noch stärker ausfallen, als es bei der Konturanpassung möglich war, wie in Abbildung 5.33 gezeigt wird: Das Skelett des chinesischen Zeichens für „1“ besteht nur aus einem einzelnen geradlinigen Strich. Die Transformation bewirkt bei beliebig komplexen Zeichen, daß auch diese nahezu eindimensional angeordnet werden. Dadurch sinken die Vergleichskosten von 0.62 vor

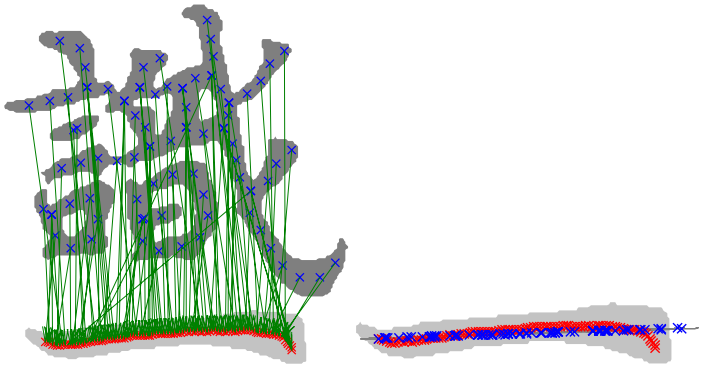


Abbildung 5.33:
Das chinesische Zeichen für „1“ ist ein waagerechter Pinselstrich. Durch eine affine Transformation werden selbst komplexe Zeichen auf einen Strich transformiert (rechtes Bild).

der Transformation auf 0.08 nach der Transformation, was eine sehr hohe Ähnlichkeit bedeutet. Ergebnisse auf dieser Grundlage scheinen für diese Arbeit wertlos zu sein.

5.7 Auffinden gleicher und ähnlicher Zeichen

In folgendem Beispiel soll ein kurzer Überblick über die Leistungsfähigkeit des entworfenen Systems gegeben werden.

Abbildung 5.34 zeigt einen Ausschnitt von 380 Zeichen eines längeren Manuskripts. Mit 150 Punkten werden auf der Kontur dieser Zeichen SHAPE CONTEXTe gebildet, durch die ein Vergleich mit anderen Zeichen die nachfolgend aufgeführten Ergebnisse liefert. Dabei werden die Zeichen nicht transformiert, und es erfolgt eine Miteinbeziehung der Kosten des Vergleichs der Konturnormalen von 10% in die Gesamtkosten.

Vergleicht man nun ein Zeichen mit allen 380 Zeichen dieses Ausschnittes, sollten die gleichartigen Zeichen die niedrigsten Vergleichskosten aufweisen. Die nach den Vergeichskosten sortierte Reihenfolge der Zeichen mit den niedrigsten Kosten wird bei den folgenden Vergleichen ausgegeben; die Reihenfolge von Zeichen mit durchschnittlichen oder hohen Vergleichskosten ist für die folgenden Beispiele nicht relevant.

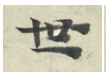
Vergleich von	Position 1	2	3	4	5	6	7	8	9	10
										
Kosten	0.0567	0.0631	0.0643	0.0801	0.0913	0.0976	0.1040	0.1040	0.1043	0.1044

Tabelle 5.1: Sortierte Ausgabe der zehn ähnlichsten Zeichen des Vergleichs des linken Zeichens mit allen aus Abbildung 5.34.

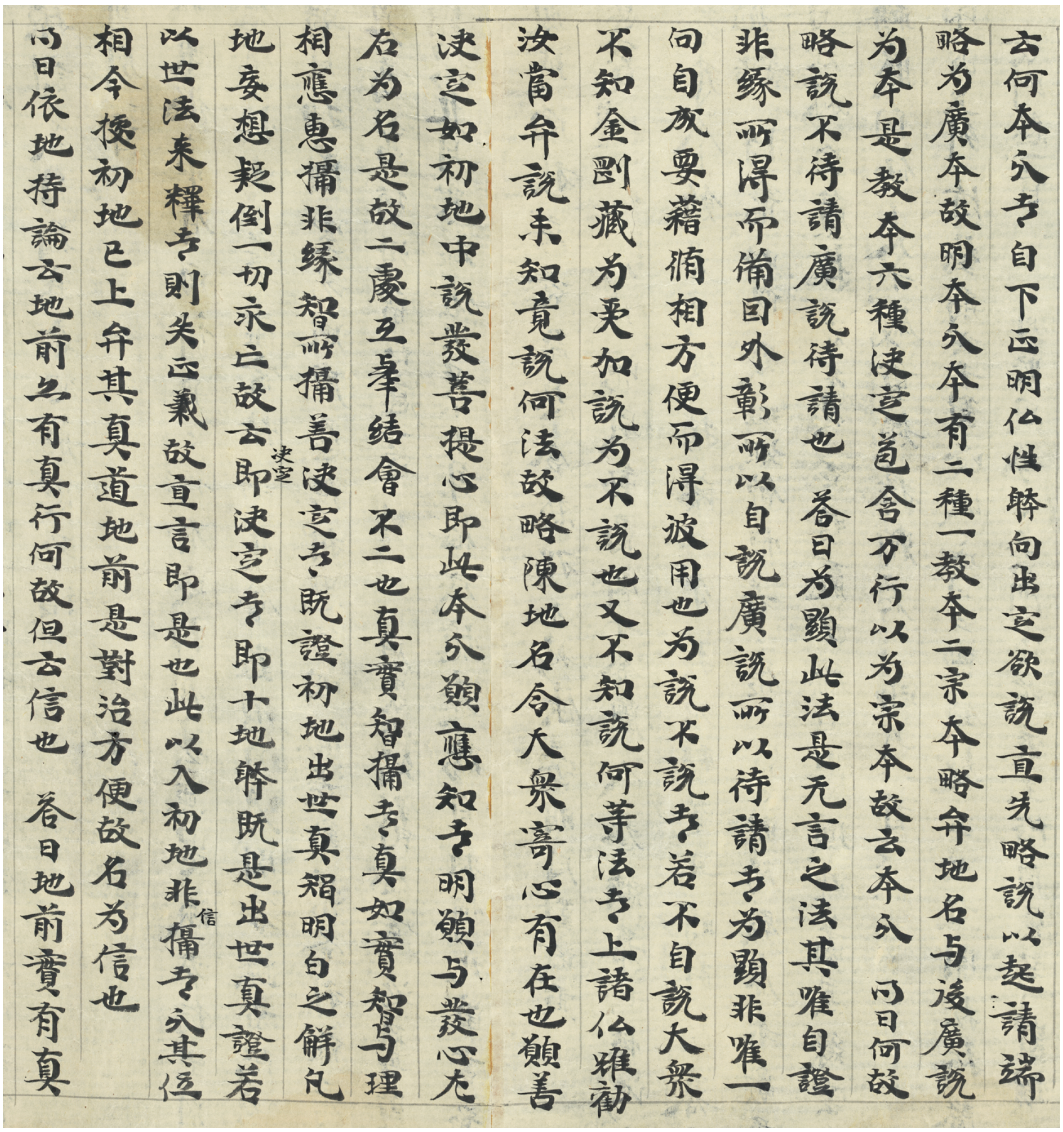


Abbildung 5.34: Manuskriptausschnitt mit 380 Zeichen.

Abbildung 5.35:
Während die Kontur des linken Zeichens durch zwei Polygone repräsentiert wird, werden beim rechten Zeichen acht Polygone erfaßt.

Die Ausgabe der Tabelle 5.1 zeigt als beste Vergleichsergebnisse alle drei gleichartigen Zeichen. Besonders bemerkenswert ist dabei die topologische Unähnlichkeit der Konturen dieser Zeichen. Besteht die erkannte Kontur des Zeichens an Position (2) aus nur zwei Polygonen, so werden an Position (3) acht Polygone erfaßt (Abbildung 5.35). Damit werden die Vorteile des in Kapitel 2 vorgestellten holistischen Ansatzes deutlich – oder die Schwierigkeiten, mit denen zu rechnen ist, wenn ein Zeichen in seine Radikale aufgetrennt werden sollte, um diese einzeln zu bewerten.

Vergleich von	Position 1	4	7	10	13	15	16	17	18	31
Kosten	0.0463	0.0541	0.0575	0.0588	0.0664	0.0727	0.0902	0.1001	0.1005	0.1222

Tabelle 5.2: Sortierte Ausgabe gleicher und ähnlicher Zeichen des Vergleichs des linken Zeichens mit allen aus Abbildung 5.34.

Von den 16 gleichartigen Zeichen zeigt Tabelle 5.2 15 auf den ersten 15 Positionen. Beim Zeichen an Position 31 werden rechts über dem Zeichen zwei Flecken erfaßt, die dunkler und größer als die entsprechenden Schwellwerte sind. Es scheint angebracht zu sein, zusätzlich zur Größe einzelner Polygone auch deren Abstand zu anderen oder zum Schwerpunkt aller Polygone als Kriterium miteinzubeziehen, und daran die Bewertung vorzunehmen, ob einzelne Bildbereiche verworfen werden sollten.

Vergleich von	Position 1	2	3	4	5	6	7	8	9	10
Kosten	0.0440	0.0480	0.0489	0.0536	0.0538	0.0539	0.1133	0.1171	0.1191	0.1192

Tabelle 5.3: Sortierte Ausgabe der ähnlichsten zehn Zeichen des Vergleichs des linken Zeichens mit allen aus Abbildung 5.34.

Vergleich von	Position 1	2	3	4	5	6	7	8	9	10
Kosten	0.0466	0.0557	0.0600	0.0618	0.0656	0.0692	0.0700	0.0704	0.0837	0.0938

Tabelle 5.4: Sortierte Ausgabe der ähnlichsten zehn Zeichen des Vergleichs des linken Zeichens mit allen aus Abbildung 5.34.

In Tabellen 5.3 und 5.4 werden alle im Vergleich gleichartigen Zeichen als die sechs bzw. zehn ähnlichsten erkannt.

Interessant sind die Unterschiede der Kosten beider Tabellen. Können alle gleichartigen Zeichen in Tabelle 5.3 mit recht niedrigen und konstanten Kosten gefunden werden, so werden in Tabelle 5.4 für das unähnlichste gleichartige Zeichen doppelt so hohe Kosten als für das ähnlichste gleichartige Zeichen errechnet. An Position (11) befindet sich dann ein deutlich unähnliches Zeichen mit Kosten von 0.1377, was zwar auch noch einen erkennbaren Kostensprung darstellt, dieser aufgrund der hohen Varianz aber deutlich geringer in Erscheinung tritt.

5.8 Auswertung eines Zeichenvergleichs

Abschließend soll eine Fragestellung des Sinologen Dr. Matthias Richter mit den in dieser Arbeit entwickelten Funktionen untersucht werden. Hierbei soll eine Antwort auf die Frage gefunden werden, ob die Abweichungen der Schreibweise gleichartiger Zeichen innerhalb eines Manuskriptes bei komplexen Zeichen höher sind als bei einfachen.

Exemplarisch werden dafür sechs unterschiedliche Zeichen von elf Seiten eines Dokuments mit jeweils etwa 370 Zeichen pro Seite ausgewählt. Es handelt sich um zwei einfache und vier durchschnittliche bis besonders komplexe Zeichen. Jedes Zeichen wird zehnfach zufällig und möglichst gleichverteilt aus den Seiten ausgewählt, so daß jeweils zehn Exemplare eines Zeichens untereinander verglichen werden.

Im Anhang werden zu allen Zeichen die ausführlichen Vergleichstabellen wiedergegeben, auch diese wurden aus den durchschnittlichen Kosten von 140-160 Konturpunkten bei zehnprozentigem Anteil der Winkeldifferenz der Konturnormalen in jedem Punkt errechnet. Es wurde keine Transformation durchgeführt. Der Mittelwert μ und die Varianz σ für jede Zeichenklasse läßt sich aus diesen Tabellen A.1–A.6 wie folgt errechnen:

$$\mu = \frac{1}{45} \sum_{y=\text{Zeile 1}}^{\text{Zeile 10}} \sum_{x=\text{Spalte } (y+1)}^{\text{Spalte 10}} i_{x,y} \qquad \sigma = \frac{1}{45} \sum_{y=\text{Zeile 1}}^{\text{Zeile 10}} \sum_{x=\text{Spalte } y+1}^{\text{Spalte 10}} (i_{x,y} - \mu)^2$$

In Tabelle 5.5 werden die so errechneten Werte zusammengefaßt. Diese können nicht als absolut objektiv angesehen werden, sondern nur objektiv in Verbindung mit dem gewählten Meßverfahren. So können z.B. die deutlich höheren Kosten von Zeichen 6 in Tabelle A.3 zu den übrigen Zeichen damit erklärt werden, daß der mittlere Teil des oberen Pinselstriches keinen deutlichen Kontrast zum Hintergrund bietet und das Schwellwertverfahren daher nur die linke und rechte Seite des Striches als Zeichenbestandteil erkennt. Schon ein anderer Schwellwert dürfte also unterschiedliche Ergebnisse liefern.

	1	2	3	4	5	6	
durchschnittliche Kosten	7,44678	6,93179	7,16248	5,69824	6,97278	6,80637	$\times 10^{-2}$
Varianz	2,24017	2,12090	3,22100	0,44270	1,11443	0,34424	$\times 10^{-4}$

Tabelle 5.5: Mittelwert und Varianz der Zeichenklassen aus Tabellen A.1–A.6.

Auch wenn der Vergleich von nur zehn Zeichen einer Art keine repräsentativen Rückschlüsse zuläßt, so kann man doch davon ausgehen, daß bestimmte Zeichen mit größerer Konstanz als andere geschrieben werden. Die höhere Varianz bei den einfacheren Zeichen könnte darauf hindeuten, daß sich der Schreiber bei komplexeren Zeichen stärker darum bemüht hat, diese gleichmäßig zu schreiben. Es sind allerdings weitere Untersuchungen mit anderen Zeichen nötig, um diese Vermutung zu unterstützen.

6 Fazit und Ausblick

6.1 Zusammenfassung und Bewertung

In dieser Arbeit wurden Verfahren vorgestellt, die es erlauben, digitale Bildvorlagen historischer chinesischer Manuskripte in einzelne Zeichen zu segmentieren. Die Zeichen können auf unterschiedliche Weise miteinander verglichen werden. Eine vollständige Bedienung der Funktionen wird durch eine grafische Oberfläche ermöglicht.

Die Ergebnisse der Vergleiche bieten eine mögliche Grundlage, auf der das Bestreben der Sinologen, Manuskripte gleicher Urheber einander zuordnen zu können, vereinfacht werden kann.

Zur Segmentierung der Vorlagen wird das Wasserscheidenverfahren benutzt, das zunächst auf die gesamte tiefpaßgefilterte Bildvorlage angewendet wird; nach Einteilung der Vorlagen in Spalten werden diese dann mit entsprechenden Parametern einzeln geglättet. Diese Tiefpaßfilterung bewirkt dabei sowohl eine Reduzierung kleinerer Bilddetails als auch ein Zusammenführen von einzelnen Zeichenelementen, da sich benachbarte helle und dunkle Bildpixel einander angleichen.

Die Stärke der Tiefpaßfilterung wird dabei in Abhängigkeit der Spaltenbreite und Zeichenhöhe angepaßt, die berechnet werden aus der Spaltenanzahl und Zeichenanzahl je Spalte, welche – als periodisches Signal interpretiert – durch eine Fouriertransformation bestimmt werden.

Durch das GEOMAP-Rahmenwerk wird das Ergebnis einer Wasserscheidentransformation über Flächen, Kanten und Knoten zugreifbar. Mit den dabei bereitgestellten Informationen wie Flächengröße, Kantenlänge und Orientierung, oder der Helligkeit der unter den Kanten liegenden Bildbereichen, wird entschieden, wie die anfängliche Übersegmentierung durch Löschen von Kanten und Zusammenführen von Flächen so reduziert wird, daß eine möglichst optimale Segmentierung erreicht wird.

Mit Funktionen der VIGRA-Bibliothek werden Konturen der einzelnen Zeichen subpixelgenau in Polygonen erfaßt. Da die Konturen eine sehr gute Charakterisierung der Zeichen darstellen, werden diese genutzt, um Zeichen miteinander zu vergleichen. Um die Informationen der Konturen auf ein handhabbares Maß zu verringern und vergleichbar zu machen, werden äquidistant Punkte auf der Kontur festgelegt, und Abstand und Winkel der Vektoren dieser Punkte zueinander erfaßt. Quantisierte Abstände und Winkel von einem Punkt zu allen anderen werden als SHAPE CONTEXT bezeichnet, die Menge aller SHAPE CONTEXTe bietet eine komplette Beschreibung eines Zeichens.

SHAPE CONTEXTe werden mittels der Differenz der Punkteanzahl in jeder Zone der Quantisierungen verglichen, normalisierte Differenzen bieten ein bezifferbares Vergleichsmaß, das als Kosten verstanden wird. Eine Zuordnung aller Punkte eines Zeichens zu allen eines anderen Zeichens mit gleicher Punkteanzahl auf die Weise, daß die Summe der Differenzen aller SHAPE CONTEXTe möglichst niedrig ist, liefert die Vergleichskosten zweier Zeichen, die der Summe der Kosten aller einander zugeordneten Punktepaare entspricht.

In Kapitel 5 wurden dann Beispiele gezeigt, die erkennen lassen, daß die Güte der Ergebnisse der Segmentierung stark von der Art der Bildvorlage abhängen. Während Bildvorlagen mit deutlichem Zeichenabstand und gutem Kontrast perfekte Segmentierungsergebnisse liefern, ist bei Vorlagen mit sich berührenden Zeichen oder starken Bildstörungen ein Eingriff von Hand nötig, durch den über die grafische Oberfläche mit der Maus Kanten geschützt oder gelöscht werden können, die von den automatischen Verfahren nicht als notwendige oder falsch positionierte Kanten erkannt wurden.

Tests mit unterschiedlicher Punkteanzahl auf den Zeichenkonturen ergaben, daß Zeichen mit mindestens 50 Punkten versehen werden sollten und mehr als 300 Punkte keine nennenswerten Verbesserungen ergeben. Eine Begrenzung auf 150 Punkte liefert bei geringerem Rechenaufwand ein brauchbares Ergebnis.

Die als Maske bezeichnete Einteilung der Umgebung einzelner Punkte sollte weder Zonen aufweisen, die 90° oder mehr abdecken, noch solche, die nur eine 10° Einteilung besitzen. Besonders bei wenigen, großen Zonen ist der Einfluß der Positionierung der Zonengrenzen für das Ergebnis ausschlaggebend. Eine Maske mit acht Winkelzonen, die jeweils mittig zu den Hauptachsen ausgerichtet sind, stellt auch hier einen guten Kompromiß zwischen Rechenaufwand und einem stabilen Ergebnis dar, zudem muß hier ergänzt werden, daß eine höhere Zonenanzahl nicht notwendigerweise bessere Ergebnisse liefert. Es wurde bei keinem Versuch von der Maskeneinteilung in fünf radial angeordnete Zonen mit logarithmisch von innen nach außen ansteigender Größe abgewichen.

Während die Abbildung gleichartiger Zeichen aufeinander mittels affiner Transformation keine bemerkenswert höhere Ähnlichkeit liefert, werden unähnliche Zeichen von deutlich unterschiedlichen Abmessungen durch die Transformation laut den Meßergebnissen deutlich ähnlicher zueinander, was dem entgegengesetzten Ergebnis entspricht, das erreicht werden sollte. Auf dieser Grundlage stellen affine Transformationen keine wertvolle Ergänzung dar. Eine geringe Einbeziehung der Normalen zur Kontur an der Position der einzelnen Punkte kann das Vergleichsergebnis dagegen verbessern.

Wahlweise kann statt der Kontur auch das Skelett der Zeichen zur Positionierung der Punkte genutzt werden. Die Form der Skelette hängt stark vom Verlauf der Kontur und der Topologie der Zeichen ab. Da aber der Abgleich der Punkte nur mittels ihrer SHAPE CONTEXTe vollzogen wird, deren Lage auf gleichen oder unterschiedlichen Zeichenteilstücken also nicht in die Bewertung einbezogen wird, werden diese Besonderheiten der Skelette teilweise wieder ausgeglichen. Je nachdem, welche Wertigkeit der Topologie zugesprochen wird, liefern Vergleiche auf Grundlage der Skelette bessere Ergebnisse als bei der Punkteanordnung auf der Kontur, hier kann keine allgemeingültige Empfehlung gegeben werden.

Beispiele zeigen, daß bei richtiger Segmentierung und Zeichenerfassung mit dem SHAPE CONTEXT-Vergleich gleichartige von ungleichen Zeichen unterschieden werden können. Da hier aber Handschriften untersucht werden, deren Ausprägungen vom jeweiligen Schreiber abhängen und sehr unterschiedlich sein können, ist auch diese Aussage abhängig von der Qualität der Vorlagen.

Es konnte weiterhin gezeigt werden, daß durch diese Arbeit eine mögliche Antwort auf eine Fragestellung eines Sinologen gefunden werden konnte; das erhaltene Ergebnis konnte alleine durch Eingaben über die grafische Bedienoberfläche gefunden werden. Somit sollte es auch ungeübten Nutzern möglich sein, die hier geleistete Arbeit zu evaluieren.

6.2 Ausblick

Dieser Abschnitt dient der Zusammenfassung der in den vorangegangenen Abschnitten bereits erwähnten Möglichkeiten, die Resultate dieser Arbeit zu verbessern und kann darüber hinaus noch einige weitere Ansätze aufzeigen, welche Erweiterungen zur bisherigen Vorgehensweise sinnvoll erscheinen.

6.2.1 Verbesserung der Segmentieung

In Kapitel 5 wurden einige Beispiele von typischen Segmentierungsfehlern gezeigt. Es sollte aber möglich sein, einige dieser Fehler beheben zu können:

- Eine stärkere Berücksichtigung der durchschnittlichen Zeichenhöhe sollte schon vor der Anwendung des ersten Kriteriums zum Löschen der Kanten stattfinden. Bisher werden erst bei der Anwendung des zweiten Kriteriums nur jeweils zwei Flächen auf ihre Höhe überprüft, und in Abhängigkeit der durchschnittlichen Zeichenhöhe entschieden, ob die zwischen diesen Flächen liegende Kante entfernt werden kann. Die Höhe der umliegenden Flächen wird dagegen nicht berücksichtigt. So kommt es besonders bei Zeichen mit größeren freien Zwischenräumen wie der Zahlen 2 und 3 öfters zu Fehlern, die mit einer zusätzlichen Gewichtung aller Kanten, die sich bei grober Spaltenaufteilung in durchschnittliche Zeichenhöhen etwa an der Stelle einer möglichen Trennkante befinden, umgangen werden könnte.
- Nach Segmentierung und Aufteilung des Bildokuments in einzelne Flächen der GEOMAP kann bei diesen festgestellt werden, ob Bereiche möglicher Zeichen die Begrenzung dieser Flächen berühren. In diesen Fällen könnte zunächst von der Flächensegmentierung abgesehen werden und dieser Bereich im unsegmentierten Bild mit dem Labeling-Verfahren erkannt werden. Schließlich kann anhand des Schwerpunkts des Labels entschieden werden, welcher Fläche – und somit welchem Zeichen – dieses zugeordnet werden sollte. So sollten auch Zeichen, von denen äußere Pinselstriche durch die Flächensegmentierung abgetrennt wurden, komplett erfaßt werden können

- Die Erkennung möglicher Bildstörungen könnte erreicht werden, wenn nicht nur die Helligkeit und Größe von Bildbereichen alleine ausgewertet wird, sondern die Größe auch in Abhängigkeit zur Position gewichtet wird. So ist die Wahrscheinlichkeit größer, das kleine Flächen, die sich dichter an übrigen Bereichen des Zeichens befinden, auch Bestandteile desselben sind, als bei weiter entfernten. Als Abstandsmaß könnte dabei entweder der Schwerpunkt aller in einer Fläche erkannten Vordergrundpixel dienen, oder der minimale Abstand zum jeweils nächstliegenden Vordergrundbereich (Label). Evtl. kann hier auch die Einbeziehung von Farben oder Farbmustern Labels identifizieren, die Bildstörungen repräsentieren.
- Neben der Erfassung der Vordergrundbereiche mittels Schwellwertverfahren sollten auch gradientenbasierte Kantenfindungsoperationen getestet werden, die bei kontrastschwachen Vorlagen bessere Ergebnisse liefern könnten.

6.2.2 Shape Context

Auch dieser Bereich kann durch folgende Berücksichtigungen optimiert werden:

- Auch wenn bisherige Versuche mit affinen Transformationen nicht die gewünschten Ergebnisse erzielten, so ist noch abzuklären, ob eine Erweiterung auf nichtaffine Transformationen, etwa die von den Autoren durchgeführten Abbildungen auf Basis von Thin-Plate-Splines, bessere Resultate ergeben kann.
- Es wäre hilfreich, die Stärke einer Transformation zu erfassen, etwa die Bezugspunkte mit den längsten Verschiebungsvektoren festzuhalten. Mit einer geeigneten Darstellung könnte so nicht nur sichtbar gemacht werden, mit welchem Betrag Zeichen unähnlich zueinander sind, sondern auch an welchen Stellen.
- Auch bei gleichartigen Zeichen konnte in fast allen Fällen festgestellt werden, daß einzelne Punktezuordnungen hohe Kosten verursachen. Die bisherige Praxis, als Vergleichskosten zweier Zeichen den Mittelwert der Vergleichskosten aller Punktezuordnungen zu verwenden, wird hiervon beeinflußt. Alternativ bietet sich der Gebrauch z.B. des Medians an, der nicht von abweichenden Einzelwerten beeinflußt wird. Untersuchungen müssen zeigen, ob das Ziel, beim Vergleich gleichartiger Zeichen deutlich niedrigere Vergleichskosten als bei unterschiedlichen Zeichen zu erhalten, durch das Weglassen der Extremwerte besser erreicht werden kann.

6.2.3 Benutzungsoberfläche

Es gibt sicherlich einige Erweiterungsmöglichkeiten, die sich erst nach häufigerer praktischer Benutzung herausstellen werden; aus bisheriger Sicht können zumindest folgende Abwandlungen sinnvoll erscheinen:

- Sollen Zeichen aus mehr als zwei Bilddokumenten verglichen werden, muß bisher ein Fenster der Oberfläche durch das neue Bild ersetzt werden; Ergebnisse dieser Auswertung gehen dann verloren. So könnte eine Erweiterung auf mehrere Bildfenster, die evtl. mit Karteireitern hintereinander angeordnet werden und nicht unbedingt gleichzeitig sichtbar sein müssen, das Problem lösen.

- Wie bereits oben erwähnt fehlt bisher noch eine konkrete Darstellung der Unterschiede zweier Zeichen. Denkbar wäre es, daß nach vollendetem Vergleich ein Mausklick auf ein Zeichen im rechten Fenster ein weiteres Fenster öffnet, das die beiden vergleichenden Zeichen übereinanderlegt, wie es in Abbildung 5.28 gezeigt wurde. Es bleibt noch zu testen, ob und auf welche Art unterschiedliche Bereiche innerhalb der Zeichen hervorgehoben werden könnten.

- Eine sortierte Darstellung der Vergleichsergebnisse wie in Tabelle 5.1 ist bisher nicht möglich. Diese Art der Ausgabe scheint aber eine durchaus nützliche Erweiterung zu sein.

6.2.4 Datenbanken

Unter diesem Punkt sollen zwei unterschiedliche Absichten zusammengefaßt werden:

- Anbindung an bestehende Datenbanken
  Auch wenn bisher nicht geklärt werden konnte, inwieweit es bereits öffentlich zugängliche Datenbanken gibt, die Zeichen enthalten, die den Zeichen der historischen Manuskripte entsprechen, die untersucht werden sollen, so scheint doch ein Abgleich mit etwa der in [Srihari et al., 2007] erwähnten ITRI Datenbank sinnvoll. So könnte ein Vergleich von mehreren möglichen Segmentierungsergebnissen mit Zeichen in dieser Datenbank dazu führen, daß richtig segmentierte Zeichen in der Datenbank gefunden werden, Segmentierungen, die keine vollständigen Zeichen oder mehr als ein ganzes Zeichen enthalten, dagegen verworfen werden.

- Aufbau einer eigenen Datenbank
  Vorstellbar wäre die Anlage einer eigenen Zeichendatenbank, in der Zeichen, die nachweislich vom gleichen Autor stammen, festgehalten werden. Auch für dieses Anliegen wäre die Anbindung an die oben erwähnte Datenbank hilfreich, um ein systematisches Einsortieren der Zeichen nach gleichartigen Zeichen zu ermöglichen.

  So sollte es möglich sein, alle Zeichen von allen vorliegenden Bilddokumenten zu erfassen und nach Urheber getrennt zu archivieren. Dazu sollten wie in Abschnitt 5.8 gezeigt die Abweichungen der Schreibweisen aller Zeichen eines Autors festgehalten werden. Nach einer Feststellung, welche Zeichen welchen Autors die geringsten Abweichungen aufweisen, sollten diese ein gutes Wiedererkennungsmerkmal bieten. So sollte es möglich sein, neue, unbekannte Dokumente mit allen bisher untersuchten Zeichenschreibweisen abzugleichen und daraus Wahrscheinlichkeiten abzuleiten, inwieweit dieses neue Dokument den bekannten Autoren zuzuordnen ist.

6.2.5 Erfassung weiterer Merkmale

Zur weiteren Bestimmung des Autors sollten aus den Dokumenten weitere Informationen extrahiert werden. Dazu zählen u.a.

- Die äußeren Abmessungen sowie das Material[1] des Manuskripts
- Die obere, untere und seitlichen Randbreiten des Manuskripts
- Der typische Zeichen-, Satz- und Spaltenabstand sowie deren Varianz
- Die durchschnittliche Häufigkeit, mit der der Pinsel neu eingefärbt wurde und auch deren Varianz.

[1] Schriftstücke können z.B. aus Bambusstreifen oder Seide bestehen. Diese Eingabe kann von Hand erfolgen.

Anhang

	Position[2]									
	1 (1-9-21)	**2** (2-8-7)	**3** (3-11-3)	**4** (4-6-8)	**5** (5-6-4)	**6** (6-8-11)	**7** (7-8-7)	**8** (8-7-2)	**9** (9-12-8)	**10** (10-11-3)
	0.0	0.0733	0.0875	0.0832	0.1096	0.0875	0.0968	0.0951	0.0868	0.0979
	0.0733	0.0	0.0722	0.0638	0.1083	0.0739	0.0872	0.0807	0.0621	0.0784
	0.0875	0.0722	0.0	0.0492	0.0746	0.0610	0.0761	0.0546	0.0780	0.0689
	0.0832	0.0638	0.0492	0.0	0.0675	0.0488	0.0696	0.0521	0.0671	0.0669
	0.1096	0.1083	0.0746	0.0675	0.0	0.0794	0.0694	0.0550	0.1000	0.0597
	0.0875	0.0739	0.0610	0.0488	0.0794	0.0	0.0643	0.0595	0.0731	0.0676
	0.0968	0.0872	0.0761	0.0696	0.0694	0.0643	0.0	0.0763	0.0901	0.0698
	0.0951	0.0807	0.0546	0.0521	0.0550	0.0595	0.0763	0.0	0.0766	0.0504
	0.0868	0.0621	0.0780	0.0671	0.1000	0.0731	0.0901	0.0766	0.0	0.0813
	0.0979	0.0784	0.0689	0.0669	0.0597	0.0676	0.0698	0.0504	0.0813	0.0

Tabelle A.1: Vergleichskosten zehn gleichartiger Zeichen aus einem Manuskript.

[2] In Klammern angegeben sind die Positionen der Zeichen im Manuskript: (Seite-Spalte-„Zeile“).

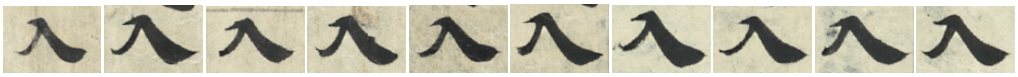
	Position									
	1 (1-1-23)	**2** (1-2-5)	**3** (1-3-1)	**4** (1-3-5)	**5** (1-5-19)	**6** (1-10-13)	**7** (5-12-9)	**8** (5-12-23)	**9** (10-12-11)	**10** (10-16-16)
										
	0.0	0.0814	0.0893	0.0918	0.0607	0.0749	0.0775	0.0902	0.0967	0.0700
	0.0814	0.0	0.0518	0.0553	0.0487	0.0678	0.0603	0.0585	0.0549	0.0585
	0.0893	0.0518	0.0	0.0531	0.0543	0.0797	0.0795	0.0818	0.0447	0.0524
	0.0918	0.0553	0.0531	0.0	0.0540	0.0830	0.0718	0.0638	0.0688	0.0489
	0.0607	0.0487	0.0543	0.0540	0.0	0.0798	0.0714	0.0558	0.0698	0.0386
	0.0749	0.0678	0.0797	0.0830	0.0798	0.0	0.0785	0.1043	0.0666	0.0699
	0.0775	0.0603	0.0795	0.0718	0.0714	0.0785	0.0	0.0775	0.0753	0.0770
	0.0902	0.0585	0.0818	0.0638	0.0558	0.1043	0.0775	0.0	0.0868	0.0746
	0.0967	0.0549	0.0447	0.0688	0.0698	0.0666	0.0753	0.0868	0.0	0.0693
	0.0700	0.0585	0.0524	0.0489	0.0386	0.0699	0.0770	0.0746	0.0693	0.0

Tabelle A.2

	Position									
	1 (1-12-6)	**2** (1-12-10)	**3** (3-13-9)	**4** (4-7-4)	**5** (5-4-12)	**6** (6-6-23)	**7** (8-3-1)	**8** (8-11-7)	**9** (9-18-2)	**10** (10-14-21)
	0.0	0.0599	0.0597	0.0581	0.0477	0.0889	0.0613	0.0632	0.0592	0.0678
	0.0599	0.0	0.0766	0.0640	0.0666	0.1186	0.0883	0.0823	0.0838	0.0810
	0.0597	0.0766	0.0	0.0528	0.0533	0.1067	0.0684	0.0520	0.0708	0.0508
	0.0581	0.0640	0.0528	0.0	0.0520	0.1117	0.0759	0.0612	0.0744	0.0634
	0.0477	0.0666	0.0533	0.0520	0.0	0.0952	0.0627	0.0635	0.0578	0.0557
	0.0889	0.1186	0.1067	0.1117	0.0952	0.0	0.0934	0.1057	0.0784	0.1086
	0.0613	0.0883	0.0684	0.0759	0.0627	0.0934	0.0	0.0635	0.0544	0.0702
	0.0632	0.0823	0.0520	0.0612	0.0635	0.1057	0.0635	0.0	0.0670	0.0531
	0.0592	0.0838	0.0708	0.0744	0.0578	0.0784	0.0544	0.0670	0.0	0.0733
	0.0678	0.0810	0.0508	0.0634	0.0557	0.1086	0.0702	0.0531	0.0733	0.0

Tabelle A.3

	Position									
	1 (1-9-1)	**2** (2-12-22)	**3** (3-9-2)	**4** (4-9-21)	**5** (5-7-19)	**6** (6-12-8)	**7** (7-8-8)	**8** (8-6-1)	**9** (9-5-24)	**10** (10-7-3)
	0.0	0.0597	0.0536	0.0582	0.0600	0.0574	0.0531	0.0487	0.0707	0.0525
	0.0597	0.0	0.0520	0.0513	0.0642	0.0612	0.0559	0.0508	0.0495	0.0562
	0.0536	0.0520	0.0	0.0471	0.0528	0.0540	0.0576	0.0447	0.0603	0.0537
	0.0582	0.0513	0.0471	0.0	0.0480	0.0597	0.0584	0.0496	0.0575	0.0515
	0.0600	0.0642	0.0528	0.0480	0.0	0.0659	0.0666	0.0631	0.0758	0.0596
	0.0574	0.0612	0.0540	0.0597	0.0659	0.0	0.0634	0.0568	0.0714	0.0605
	0.0531	0.0559	0.0576	0.0584	0.0666	0.0634	0.0	0.0506	0.0590	0.0561
	0.0487	0.0508	0.0447	0.0496	0.0631	0.0568	0.0506	0.0	0.0584	0.0474
	0.0707	0.0495	0.0603	0.0575	0.0758	0.0714	0.0590	0.0584	0.0	0.0597
	0.0525	0.0562	0.0537	0.0515	0.0596	0.0605	0.0561	0.0474	0.0597	0.0

Tabelle A.4

	Position									
	1 (1-11-18)	**2** (2-14-12)	**3** (3-13-18)	**4** (4-6-11)	**5** (6-6-24)	**6** (6-11-9)	**7** (7-4-24)	**8** (7-15-16)	**9** (9-15-1)	**10** (10-10-21)
	0.0	0.0771	0.0616	0.0624	0.077	0.0682	0.0733	0.0639	0.0754	0.0672
	0.0771	0.0	0.0816	0.0627	0.0794	0.0842	0.0749	0.0667	0.0927	0.0728
	0.0616	0.0816	0.0	0.0685	0.0777	0.0861	0.0807	0.0673	0.0955	0.0705
	0.0624	0.0627	0.0685	0.0	0.0665	0.0562	0.0698	0.0556	0.0712	0.0513
	0.077	0.0794	0.0777	0.0665	0.0	0.0656	0.0635	0.0621	0.0823	0.0558
	0.0682	0.0842	0.0861	0.0562	0.0656	0.0	0.0727	0.0633	0.0646	0.0602
	0.0733	0.0749	0.0807	0.0698	0.0635	0.0727	0.0	0.0569	0.0887	0.0519
	0.0639	0.0667	0.0673	0.0556	0.0621	0.0633	0.0569	0.0	0.0685	0.0525
	0.0754	0.0927	0.0955	0.0712	0.0823	0.0646	0.0887	0.0685	0.0	0.0711
	0.0672	0.0728	0.0705	0.0513	0.0558	0.0602	0.0519	0.0525	0.0711	0.0

Tabelle A.5

	Position									
	1 (1-8-5)	**2** (1-9-26)	**3** (4-7-5)	**4** (4-18-1)	**5** (6-10-14)	**6** (6-12-23)	**7** (6-15-11)	**8** (10-6-16)	**9** 11-7-25)	**10** (11-14-15)
	0.0	0.0694	0.0753	0.0705	0.0816	0.0728	0.0658	0.0657	0.0737	0.0720
	0.0694	0.0	0.0678	0.0793	0.0689	0.0692	0.0640	0.0686	0.0690	0.0768
	0.0753	0.0678	0.0	0.0670	0.0633	0.0714	0.0670	0.0687	0.0599	0.0648
	0.0705	0.0793	0.0670	0.0	0.0725	0.0728	0.0683	0.0576	0.0612	0.0620
	0.0816	0.0689	0.0633	0.0725	0.0	0.0638	0.0601	0.0761	0.0664	0.0674
	0.0728	0.0692	0.0714	0.0728	0.0638	0.0	0.0667	0.0797	0.0699	0.0731
	0.0658	0.0640	0.0670	0.0683	0.0601	0.0667	0.0	0.0611	0.0610	0.0654
	0.0657	0.0686	0.0687	0.0576	0.0761	0.0797	0.0611	0.0	0.0567	0.0702
	0.0737	0.0690	0.0599	0.0612	0.0664	0.0699	0.0610	0.0567	0.0	0.0581
	0.0720	0.0768	0.0648	0.0620	0.0674	0.0731	0.0654	0.0702	0.0581	0.0

Tabelle A.6

Literaturverzeichnis

[Belongie u. Malik 2000] BELONGIE, Serge; MALIK, Jitendra: Matching with Shape Contexts. In: *IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL'00)* (2000), S. 20–26. http://dx.doi.org/10.1109/IVL.2000.853834. ISBN 0–7695–0695–X

[Belongie et al. 2001a] BELONGIE, Serge; MALIK, Jitendra; MORI, Greg: Shape contexts enable efficient retrieval of similar shapes. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01)* 1 (2001), S. 723–730. http://dx.doi.org/10.1109/CVPR.2001.990547. – ISSN 1063–6919

[Belongie et al. 2000] BELONGIE, Serge; MALIK, Jitendra; PUZICHA, Jan: Shape Context: A new descriptor for shape matching and object recognition. In: *Advances in Neural Information Processing Systems (NIPS'00)* 13 (2000). ISBN 978–0–262–12241–2

[Belongie et al. 2001b] BELONGIE, Serge; MALIK, Jitendra; PUZICHA, Jan: Matching shapes. In: *Eighth IEEE International Conference on Computer Vision (ICCV'01)* 1 (2001), S. 454–461. http://dx.doi.org/10.1109/ICCV.2001.937552. ISBN 0–7695–1143–0

[Duda et al. 2001] DUDA, Richard O.; HART, Peter E.; STORK, David G.: *Pattern Classification*. 2nd edition. New York, NY, USA : John Wiley & Sons, 2001. – ISBN 0–471–05669–3

[Jonker u. Volgenant 1987] JONKER, R.; VOLGENANT, A.: A shortest augmenting path algorithm for dense and sparse linear assignment problems. In: *Computing* 38 (1987), Nr. 4, S. 325–340. http://dx.doi.org/10.1007/BF02278710. – ISSN 0010–485X

[Kato et al. 1999] KATO, Nei; SUZUKI, Masato; OMACHI, Shin'ichiro; ASO, Hirotomo; NEMOTO, Yoshiaki: A Handwritten Character Recognition System Using Directional Element Feature and Asymmetric Mahalanobis Distance. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (1999), Nr. 3, S. 258–262. http://dx.doi.org/10.1109/34.754617. – ISSN 0162–8828

[Köthe] KÖTHE, Ullrich: V*IGRA* - *Vision with Generic Algorithms*. http://kogs-www.informatik.uni-hamburg.de/~koethe/vigra/, Abruf: Dezember 2007. – Software

[Köthe 2000] KÖTHE, Ullrich: *Generische Programmierung für die Bildverarbeitung*, Cognitive Systems Group, University of Hamburg, Germany, Dissertation, Februar 2000. – Veröffentlicht über Libri Books on Demand, – ISBN 3-8311-0239-2

[Liang u. Shi 2005] LIANG, Zhizheng; SHI, Pengfei: A metasynthetic approach for segmenting handwritten chinese character strings. In: *Pattern Recognition Letters* 26 (2005), Nr. 10, S. 1498–1511. http://dx.doi.org/10.1016/j.patrec.2004.12.001. – ISSN 0167–8655

[Liu 2006] LIU, Cheng-Lin: High Accuracy Handwritten Chinese Character Recognition Using Quadratic Classifiers with Discriminative Feature Extraction. In: *18th International Conference on Pattern Recognition (ICPR'06)* 2 (2006), S. 942–945. http://dx.doi.org/10.1109/ICPR.2006.624. – ISSN 1051–4651

[Liu u. Marukawa 2004] LIU, Cheng-Lin; MARUKAWA, Katsumi: Global Shape Normalization for Handwritten Chinese Character Recognition: A New Method. In: *Ninth International Workshop on Frontiers in Handwriting Recognition (IWFHR'04)* 00 (2004), S. 300–305. http://dx.doi.org/10.1109/IWFHR.2004.47. – ISSN 1550–5235

[Liu et al. 2003] LIU, Cheng-Lin; SAKO, Hiroshi; FUJISAWA, Hiromichi: Handwritten Chinese Character Recognition: Alternatives to Nonlinear Normalization. In: *Seventh International Conference on Document Analysis and Recognition (ICDAR'03)* 01 (2003), S. 524. http://dx.doi.org/10.1109/ICDAR.2003.1227720. ISBN 0–7695–1960–1

[Liu u. Ding 2005] LIU, Hailong; DING, Xiaoqing: Handwritten Character Recognition Using Gradient Feature and Quadratic Classifier with Multiple Discrimination Schemes. In: *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)* 0 (2005), S. 19–25. http://dx.doi.org/10.1109/ICDAR.2005.123. – ISSN 1520–5263

[Meine 2003] MEINE, Hans: *XPMap-Based Irregular Pyramids for Image Segmentation*, Cognitive Systems Group, University of Hamburg, Germany, Diplomarbeit, August 2003

[Meine 2007] MEINE, Hans: *Image Analysis within the GeoMap Framework*, Cognitive Systems Group, University of Hamburg, Germany, Dissertation, Oktober 2007

[Prasad 1997] PRASAD, Lakshman: Morphological Analysis of Shapes. In: *CNLS Newsletter* 139 (1997)

[Richter 2005] RICHTER, Matthias: *Tentative Criteria for Discerning Individual Hands in the Guodian Manuscripts*. in the proceedings of the conference: *Confucianism Resurrected: The Third International Conference on Excavated Chinese Manuscripts*, 2005. – at Mount Holyoke College, South Hadley, Mass., April 2004. ed. Xing Wen

[Schmidt 1990] SCHMIDT, Wolfgang G. A.: *Einführung in die chinesische Schrift- und Zeichenkunde*. Hamburg : Helmut Buske, 1990. – ISBN 3–87118–953–7

[Sedgewick 1991] SEDGEWICK, Robert: *Algorithmen*. Bonn [u.a.] : Addison-Wesley, 1991. – ISBN 3–89319–402–9

[Sonka et al. 1999] SONKA, Milan; HLAVÁČ, Vaclav; BOYLE, Roger: *Image Processing, Analysis and Machine Vision*. 2nd edition. Brooks/Cole Publishing Company, 1999

[Srihari et al. 2007] SRIHARI, Sargur N.; YANG, Xuanshen; BALL, Gregory R.: Offline Chinese handwriting recognition: an assessment of current technology. In: *Frontiers of Computer Science in China* 1 (2007), Nr. 2, S. 137–155. http://dx.doi.org/10.1007/s11704-007-0015-2. – ISSN 1673–7350, 1673–7466. – http://www.cedar.buffalo.edu/~srihari/papers/chinese-survey.pdf, Abruf: September 2007.

[Tseng u. Chen 1998] TSENG, Lin Y.; CHEN, Rung C.: Segmenting handwritten chinese characters based on heuristic merging of stroke bounding boxes and dynamic programming. In: *Pattern Recognition Letters* 19 (1998), Nr. 10, S. 963–973. http://dx.doi.org/10.1016/S0167-8655(98)00073-7. – ISSN 0167–8655

[Tseng u. Lee 1999] TSENG, Yi-Hong; LEE, Hsi-Jian: Recognition-based handwritten Chinese character segmentation using a probabilistic Viterbi algorithm. In: *Pattern Recognition Letters* 20 (1999), Nr. 8, S. 791–806. http://dx.doi.org/10.1016/S0167-8655(99)00043-4. – ISSN 0167–8655

[Xigui 2000] XIGUI, Qiu: *Chinese Writing*. Chicago, USA : The Society for the Study of Early China, 2000. – ISBN 1–55729–071–7. – translated by Gilbert L. Mattos and Jerry Norman

Ich möchte mich bei Frau Dreschler-Fischer und Ullrich Köthe für die Übernahme der Betreuung und die Durchsicht der Arbeit bedanken. Mit vielen wertvollen Tips und Hinweisen konnte ich gezielt die Problemstellungen angehen und ein „Abdriften in Sackgassen“ weitgehend vermeiden.

Bedanken möchte ich mich insbesondere auch bei Hans Meine, dem „inoffiziellen“ Betreuer meiner Arbeit. Seine Tür stand immer offen und trotz seiner eigenen Zeitknappheit erhielt ich von Ihm immer den Rat – teilweise recht umfangreich und immer verständlich – der mir bei Schwierigkeiten ein sofortiges Weiterarbeiten ermöglichte.

Ich versichere, daß ich die vorstehende Arbeit selbständig und ohne fremde Hilfe angefertigt und mich anderer als der im beigefügten Verzeichnis angegebenen Hilfsmittel nicht bedient habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht.

Ich bin mit einer Einstellung in den Bestand der Bibliothek des Departments einverstanden.

_____

Rainer Herzog