Ein generischer Ansatz zur digitalen Layoutanalyse von Manuskripten

An der Universität Hamburg eingereichte

Dissertation

zur Erlangung des akademischen Grades
Dr. rer. nat.

Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Informatik

Rainer Herzog

aus Wedel (Deutschland)

im Juni 2018

Gutachterinnen / Gutachter

Prof. Bernd Neumann, Ph.D.

Prof. Dr.-Ing. H. Siegfried Stiehl

Prof. Dr. Leonie Dreschler-Fischer

Tag der Disputation: 28. November 2018

Danksagung

Die vorliegende Arbeit entstand während meiner Tätigkeit als Doktorand und wissenschaftlicher Mitarbeiter am Arbeitsbereich Kognitive Systeme des Fachbereichs Informatik sowie am Sonderforschungsbereich (SFB) 950 *Manuskriptkulturen in Asien, Afrika und Europa* der Universität Hamburg, unter Förderung der Deutschen Forschungsgemeinschaft (DFG).

Mein Dank gilt zuerst Professor Bernd Neumann für die Betreuung der Arbeit, für seine Unterstützung und Motivation. Diese Arbeit ist deutlich geprägt von seinen vielfältigen Ideen und Hinweisen, die in den Diskussionen angeregt wurden. Auch meiner weiteren Betreuerin Professorin Leonie Dreschler-Fischer gilt mein Dank für ihre Unterstützung.

Meinen Kollegen, Arved Solth und Benjamin Seppke gilt mein herzlicher Dank für die gute Zusammenarbeit und den immer wieder gewinnbringenden Gedankenaustausch aufgrund ihrer vielfältigen Erfahrungen in vielen Bereichen.

Eine besondere Erwähnung muss hier auch Professor Hans Siegfried Stiehl bekommen, der das Teilprojekt Informatik in der zweiten Projektphase des SFB übernommen hat. Seine Ideen, in welchem Rahmen diese Arbeit weitergeführt werden kann, halfen mir besonders, den Kontext der Anwendungsmöglichkeiten dieser Arbeit einschätzen zu können.

Weiterhin gilt mein Dank meinen Kollegen Wilfried Bohlken, Stephanie von Riegen, Lothar Hotz, sowie den Mitarbeiterinnen und Mitarbeitern am Sonderforschungsbereich, durch die mir erst die tieferen Einblicke in die vielfältigen Gesichtspunkte ermöglicht wurden, die bei der Betrachtung und Untersuchung von Manuskripten eine Rolle spielen.

Nicht unerwähnt bleiben darf mein Dank an die unzähligen, in dieser Arbeit genannten und nicht genannten Bibiotheken, die über ihr Webangebot Bilder von Manuskriptseiten zur allgemeinen Verfügung bereitstellen. Insbesondere gilt das für die Franckeschen Stiftungen zu Halle, die sogar Manuskriptfotos nach meinen besonderen Wünschen anfertigten.

Schließlich möchte ich mich noch bei meinen Eltern und meinem Bruder für die Unterstützung bedanken.

Zusammenfassung

In dieser Arbeit wird ein Ansatz zur Layouterkennung von digitalisierten Manuskripten vorgestellt. Die Bestimmung des Layouts liefert einerseits Manuskriptforschern wertvolle Hinweise auf kulturelle Zusammenhänge, kann andererseits auch innerhalb einer Verarbeitungskette nützlich sein, die eine darüber hinausgehende computerbasierte Auswertung von Manuskripten erlaubt. Im Vergleich zu bisherigen Ansätzen liegt der Fokus dieser Arbeit auf der universellen Anwendbarkeit. Es werden keine Vorannahmen bezüglich Schreibsystem, Dokumentenformaten oder Layouteigenschaften, wie etwa der Annahme eines einzigen Haupttextbereichs, gemacht. Vor der Anwendung dieses Verfahrens auf beliebige Manuskriptseiten ist keine Trainingsphase vorgesehen. Dabei soll das Verfahren robust gegenüber Störungen wie Flecken oder Rissen sein, wie sie bei historischen Manuskripten häufig auftreten.

Auf der theoretischen Basis der Gabor-Transformation wird experimentell untersucht, wie die grundlegende Eigenschaft von mehrzeiliger Schrift, die ein abwechselnd helles und dunkles streifenförmiges Muster darstellt, zur Layoutanalyse von Manuskripten geeignet ist. Die Gabor-Transformation lässt durch Anwendung einer Fensterfunktion eine lokal eingeschränkte Merkmalsextraktion einzelner Manuskriptregionen zu. Im hier entwickelten Verfahren wird die Antwort der Gabor-Transformation im Frequenzraum in drei Bänder unterteilt und durch eine Kombination ihrer Antworten werden die jeweiligen Vorteile dieser Bänder bezüglich Orts- oder Frequenzauflösung vereint. Nach systematischer Verringerung der Auflösung der Manuskriptbilder sowie anschließender Analyse unter einheitlicher Fensterfunktion werden die Auswertungen zusammengefasst und lassen so eine Untersuchung von Regionen nahezu beliebiger Textgröße zu. Daraus werden die Abgrenzung von Schrift zum Hintergrund ermittelt, sowie die durchschnittliche Zeilenorientierung und der durchschnittliche Zeilenabstand von Textregionen. Eine Abgrenzung innerhalb der Textregionen hin zu einzelnen Textblöcken erfolgt darauf anhand lokaler Unterschiede der Zeilenstruktur, unter der Annahme, dass diese innerhalb eines Textblocks weitgehend einheitlich ist. Für jeden so ermittelten Textblock können Eigenschaften wie Position, Größe, Zeilenorientierung und Zeilenabstand bestimmt werden.

Um eine Validierung des entwickelten Verfahrens im Sinne eines Vergleichs mit einer Ground Truth zu ermöglichen, wurde ein Testdatensatz von annotierten Manuskriptbildern benötigt. Dieser sollte die Anforderungen an das entwickelte System hinreichend überprüfen können, daher sollte er ein breites Spektrum an unterschiedlichen Herausforderungen enthalten. Da ein derartiger Testdatensatz nicht vorhanden war, wurde für die Validierung ein Datensatz erstellt, der aus 50 Manuskriptseiten besteht. Anhand des Testsatzes wurde das entwickelte Verfahren analysiert. Die Ergebnisse werden detailliert ausgewertet.

Abstract

This thesis is concerned with the layout analysis of digitized manuscripts. By determining layout features, scholars can derive valuable insights regarding the cultural and historic context of manuscripts. The approach developed in this thesis can also be installed within a process chain to support various kinds of advanced computer-assisted manuscript examinations. Compared to other existing approaches, the aim of this work is the versatile application on nearly all kinds of manuscripts. No assumptions are made regarding the writing system, the format of documents or its contents, like the incidence of one single main text area. Moreover, no training phase is considered necessary before applying this approach. The system is designed to be robust against deteriorations like stains or fissures, which are very likely to exist in many historical manuscripts.

Multiple lines of text form a pattern of alternating light and dark stripes. As the Gabor Transform is predestined for a localized analysis of such a pattern, a main part of this work addresses the application of the Gabor Transform on text areas. The frequency spectrum is subdivided into three frequency bands, then the Gabor Transform responses from different bands are combined in a way to profit from their specific advantages regarding their spatial or frequency resolutions. By reducing the resolution of manuscript images in a stepwise way, while applying the Gabor Transform using the same window size for each step, the combination of the responses allow for a systematic examination of manuscript images featuring text areas of nearly arbitrary geometric dimensions. The responses are then used to separate text areas from the background, and to specify the line distance and orientation of text areas, allowing to separate text blocks against each other, while presuming that text blocks are formed by homogeneous text areas and that text blocks are discriminable by their line distance or line orientation. As a result, one can then obtain e.g. the position, size, line distance and orientation for each text block.

Finally, the developed methods should be evaluated on a set of annotated manuscripts. As no adequate set of samples was available featuring the challenges completely different kinds of manuscripts would pose, a test set consisting of 50 annotated manuscripts was created. The techniques developed in this thesis were applied to this test set, followed by an in-depth analysis of the results.

Publikationen im Zusammenhang mit dieser Arbeit

Herzog, R.; Solth, A.; Neumann, B.: Text block recognition in multi-oriented handwritten documents. Technical Report FBI-HH-B-301-14, Department of Informatics, University of Hamburg, 2014.

Solth, A.; Herzog, R.; Neumann, B.: A Modular Workbench for Manuscript Analysis. Manuscript Cultures 2014, University of Hamburg, SFB 950, S.132–137

Weitere Publikationen im Zusammenhang mit dem Thema der computergestützten Manuskriptanalyse

Herzog, Rainer; Solth, Arved; Neumann, Bernd: Using Harris Corners for the Retrieval of Graphs in Historical Manuscripts. 12th International Conference on Document Analysis and Recognition (ICDAR), doi: 10.1109/ICDAR.2013.262, 2013, 1295-1299

Herzog, Rainer; Solth, Arved; Bestmann, Oliver; Scheel, Julian; Neumann, Bernd : Retrieving Writing Patterns From Historical Manuscripts Using Local Descriptors. In: Digital Humanities 2012, Hamburg University Press, 288-291

Inhaltsverzeichnis

1. Einleitung 1
   1.1. Motivation . . . 1
   1.2. Zielsetzung der Arbeit . . . 4
   1.3. Gliederung der Arbeit . . . 5

2. Stand der Forschung 7
   2.1. Layoutanalyse von gedruckten Dokumenten . . . 7
      2.1.1. Frequenzbasierte Ansätze . . . 8
      2.1.2. „Top-Down“ Ansätze . . . 9
      2.1.3. „Bottom-Up“ Ansätze . . . 11
      2.1.4. Hybride und weitere Ansätze . . . 13
      2.1.5. Fazit der Layoutanalyse auf gedruckten Dokumenten . . . 15
   2.2. Layoutanalyse von handgeschriebenen Dokumenten . . . 17
      2.2.1. Binarisierung . . . 18
      2.2.2. Analyse der Dokumentenstruktur . . . 19
      2.2.3. Segmentierung in Textzeilen . . . 23
   2.3. Fazit verschiedener Layoutanalyseverfahren . . . 24

3. Layoutanalyse auf Basis der Fouriertransformation 27
   3.1. Grundlagen der frequenzbasierten Layoutanalyse . . . 28
      3.1.1. Layoutanalyse durch zweidimensionale diskrete Fouriertransformation 29
      3.1.2. Layoutanalyse durch die Gabor-Transformation . . . 31
   3.2. Abgrenzung von Textblöcken zum Hintergrund . . . 41
      3.2.1. Abgrenzung von gleichartigen, benachbarten Textblöcken . . . 44
      3.2.2. Pixelgenaue Untersuchung des Abtastrasters . . . 47
   3.3. Analyse der Frequenzauflösung . . . 50
      3.3.1. Auswertung des Zeilenabstandes auf Manuskriptbeispielen . . . 52
      3.3.2. Auswertung der Zeilenorientierung eines Manuskripts . . . 56
      3.3.3. Kombination der Auswertung des idealen und oberen Frequenzbandes . 61
   3.4. Layoutanalyse auf Dokumenten beliebiger Auflösung . . . 65
      3.4.1. Segmentierung von Textblöcken anhand von lokalen Unterschieden der Zeilenstruktur . . . 70
   3.5. Zusammenfassung . . . 80

4. Anwendung des Verfahrens auf Manuskriptbeispiele 83
   4.1. Auswahlkriterien . . . 84

4.2. Vorbereitungen . . . 91
4.3. Erzeugung der Ground Truth . . . 91
4.3.1. Binarisierung und initiale farbliche Klassifizierung . . . 93
4.3.2. Unterschiedliche Einfärbung der Schrift . . . 93
4.3.3. Bestimmung der Textblockregionen . . . 93
4.4. Auswertung der Ergebnisse . . . 94
4.4.1. Rechnerische Auswertung ohne Segmentierung . . . 95
4.4.2. Rechnerische Auswertung mit Segmentierung . . . 96
4.5. Auswertung des Testdatensatzes . . . 97
4.5.1. Auswertung des Testdatensatzes Berlin . . . 97
4.5.2. Auswertung des Testdatensatzes Leipzig . . . 100
4.5.3. Auswertung des Testdatensatzes Greek . . . 101
4.5.4. Auswertung weiterer Manuskriptseiten . . . 102
4.6. Fazit . . . 105

5. Fazit und Ausblick 107
5.1. Fazit . . . 107
5.2. Ausblick . . . 109

A. Testdatensatz Berlin 113
A.1. Abgleich ohne Segmentierung . . . 114
A.2. Abgleich mit Segmentierung . . . 115
A.3. Darstellung der Manuskriptseiten . . . 117

B. Testdatensatz Leipzig 123
B.1. Abgleich ohne Segmentierung . . . 124
B.2. Abgleich mit Segmentierung . . . 125
B.3. Darstellung der Manuskriptseiten . . . 127

C. Testdatensatz Greek 133
C.1. Abgleich ohne Segmentierung . . . 134
C.2. Abgleich mit Segmentierung . . . 135
C.3. Darstellung der Manuskriptseiten . . . 137

D. Testdatensatz Varia 143
D.1. Abgleich ohne Segmentierung . . . 144
D.2. Abgleich mit Segmentierung . . . 145
D.3. Darstellung der Manuskriptseiten . . . 146

Literaturverzeichnis 149

1. Einleitung

1.1. Motivation

Diese Arbeit entstand im Rahmen des Sonderforschungsbereichs 950 (SFB) „Manuskriptkulturen in Asien, Afrika und Europa“, einer von der Deutschen Forschungsgemeinschaft geförderten Einrichtung, in dem Manuskripte vielfältiger Herkunft und Schreibsysteme wissenschaftlich untersucht werden.

Manuskripte erlauben uns auf vielfältige Weise, Wissen über unsere Geschichte zu erhalten. Während sich die Wissenschaft früher nahezu ausschließlich mit dem Inhalt befasst hat, geht man in letzter Zeit dazu über, zusätzlich die äußeren Eigenschaften der Manuskripte zu untersuchen. Hierbei spielt das Layout der Manuskriptseiten eine besondere Rolle (SFB950, Daub [2013]).

An erster Stelle ist für die Beschreibung des Layouts eine Begriffsbestimmung nötig:

Eine *Textregion* oder ein *Textbereich* beschreibt in dieser Arbeit einen Bereich eines Manuskripts, welcher Text enthält, im Gegensatz zu Ornamenten, Bildelementen oder inhaltslosen Bereichen.

Ein *Textblock* soll dagegen eine in sich homogene Textregion bezeichnen, in der Text bzgl. Zeilenorientierung und Zeilenabstand weitgehend einheitlich ist. Da es bei ein- oder zweizeiligem Text kaum eine Möglichkeit gibt, eine Homogenität des Zeilenabstandes festzustellen, sollte ein Textblock aus mindestens drei Zeilen bestehen.

Im Bereich der Geisteswissenschaften beschränkt sich die Bedeutung des Begriffs *Text* in der Regel auf den Inhalt, der aber in dieser Arbeit nicht untersucht werden soll. Stattdessen werden hier die Begriffe *Text* und *Schrift* synonym verwendet.

Häufig finden sich auf Manuskriptseiten neben den Haupttextblöcken auch Texte an den Rändern, oder auch interlineare Schrift (Abbildung 1.1). In der jüngeren Zeit wurden besonders diese Randtexte, auch als *Glossen* bezeichnet, intensiver untersucht.

Enthält eine Beschreibung des Layouts beispielsweise die Breite der Ränder, die Anzahl von Textblöcken pro Seite, die Breite und Anzahl von Zeilen der Textblöcke und ihre Positionierung, so kann diese Beschreibung direkt verwendet werden und in Form von Metadaten neben einer digitalen Kopie eines Manuskripts als zusätzliche textuelle Beschreibung in Datenbanken für die fachübergreifende Manuskriptforschung dienen.

Abbildung 1.1.: Kano Manuskript aus Borno, Nigeria.
© SOAS, University of London. CC BY-NC-SA

Eine Klassifikation von Manuskriptseiten in Layoutsorten ist schon anhand einfacher Eigenschaften wie Anzahl und Abmessungen der Textregionen möglich.

In einigen Datenbanken (z.B. SfarData) kann über Suchfilter nach Manuskripten mit bestimmten Layoutmerkmalen gesucht werden und so eine Übersicht an Manuskripten mit ähnlichen Merkmalen wie ein gerade untersuchtes Manuskript angezeigt werden. Auf diese Weise können Rückschlüsse auf das Alter oder die Entstehungsregion von undatierten Manuskripten gezogen werden, soweit ein Manuskript nach Einschätzung eines Experten ähnlich genug zu Manuskripten bekannter Herkunft ist. Weiterhin sind Fälle bekannt, in denen einzelne Seiten eines Manuskripts verloren gingen und an unterschiedlichen Orten weltweit wieder auftauchten. Solche Seiten konnten unter anderem anhand ihrer gleichartigen Layoutmerkmale wieder zusammengeführt werden.

Darüber hinaus können durch eine systematische Auswertung Einblicke in kulturelle Gegebenheiten zur Zeit der Entstehung der Manuskripte gewonnen werden.

Vergleichbare Merkmale können auch dazu genutzt werden, um Schreiber zu identifizieren.

Hierzu werden zusätzlich noch weitere Merkmale wie z.B. der Schreibstil ausgewertet, dazu gehören u.a. Regelmäßigkeit und Anzahl von Zeichen innerhalb eines bestimmten Abschnittes, Abstände zwischen Zeilen, Zeichen und Wörtern oder der Schrägheitsgrad der Schrift.

Bezogen auf die Randtextblöcke ist es allerdings erforderlich, die Merkmale einzelner Textblöcke getrennt voneinander zu betrachten, da sie von verschiedenen Schreibern stammen könnten. Unter dem Gesichtspunkt der Schreibererkennung ist es wünschenswert, wenn eine Layoutanalyse größere, in sich homogene Textblöcke findet und gegen andere abtrennt. Es ist dagegen nicht erforderlich, Textblöcke in einzelne Absätze zu trennen, da im Allgemeinen angenommen werden kann, dass sich der Schreiber in diesem Fall nicht ändert. Obwohl auch an der Identifikation von Schreibern kleinerer Textblöcke ein großes Interesse besteht, bieten diese häufig nicht genügend Eigenschaften, anhand derer eine hinreichend genaue Schreiberbestimmung möglich wäre.

Um alle diese Merkmale von Hand zu vermessen, wäre ein beträchtlicher Zeitaufwand nötig. Häufig werden deshalb bei längeren Manuskripten nur einzelne Teile vermessen, die per Augenschein als exemplarisch für das gesamte Manuskript angenommen werden. Manuskriptforscherinnen und -forscher wünschen sich daher, von computerbasierten Werkzeugen unterstützt zu werden, mit denen die meisten der vorangehend beschriebenen Aufgaben semi-interaktiv unterstützt oder in einzelnen Fällen sogar vollautomatisch erfüllt werden können.

Die Vorteile von rechnergestützten Vermessungsverfahren, wie sie unter einer grafischen Benutzungsschnittstelle angeboten werden können, sind:

1. Eine Zeit- und Arbeitserleichterung für den Manuskriptforscher.
2. Eine höhere Präzision der Ergebnisse im Vergleich zu einer stichprobenartigen manuellen Messung. Ein vollautomatischer Prozess kann dagegen eine komplette Analyse über möglicherweise sehr viele Seiten vollziehen, und so Mittelwert und Varianz der Messwerte präziser bestimmen. Einzelne Seiten, die deutlich abweichende Werte aufweisen, könnten einfacher aufgespürt werden.
3. Ein weiterer Vorteil ist die einfache Nachvollziehbarkeit der gemessenen Werte, da alle Vermessungen sofort gespeichert und später wieder abgerufen werden können.
4. Auch eine höhere Objektivität kann als Vorteil für den Fall benannt werden, dass mit Messergebnissen eine bestimmte These, z.B. über den vermuteten Ursprung eines Manuskripts, unterstützt werden soll. Eine automatische Vermessung aller sinnvollen Größen führt zu einer besseren Belegbarkeit, als wenn (aufgrund einer Experteneinschätzung) für den Zweck vermeintlich wenig bedeutsame Vermessungen ausgelassen werden.
5. Möglicherweise können sogar neue Erkenntnisse gewonnen werden, falls eine nicht offensichtliche Kombination automatisch vermessener Werte Auffälligkeiten wie eine besondere Regelmäßigkeit zeigt. Diese könnte beispielsweise Hinweise auf eine bestimmte Herkunft, ausgehend von einer Region über eine spezielle Schreiberschule bis möglicherweise hin zu einem bestimmten Schreiber liefern, der so auch bei weiteren Manuskripten identifizierbar werden könnte.

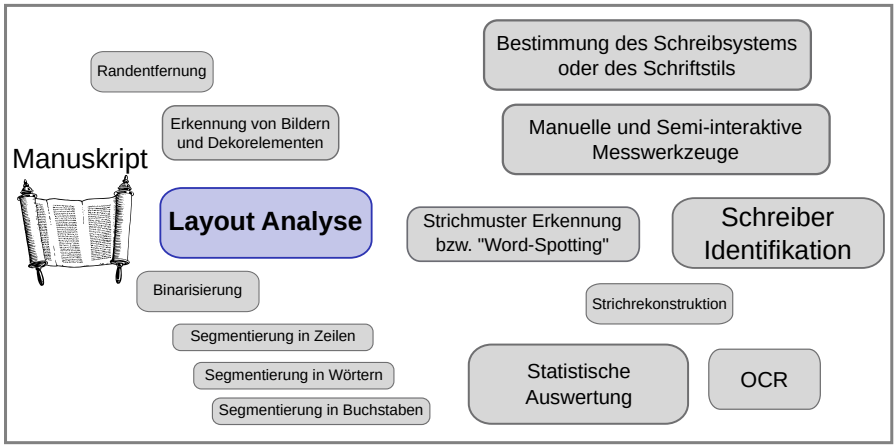


Abbildung 1.2.: Layoutanalyse in einer Verarbeitungskette zur Manuskriptuntersuchung

Eine weitere Anwendungsmöglichkeit eines Ergebnisses der Layouterkennung kann auch eine Zeilensegmentierung sein, welche die Angaben von Zeilenorientierung und Abstand als Grundeinstellung für die nachfolgende Segmentierung nutzt.

Schließlich können selbst segmentierungsfreie Word-Spotting Verfahren von diesen Angaben profitieren, indem sie diese als Grundeinstellung für die erwartete Schrifthöhe und Schriftausrichtung verwenden.

1.2. Zielsetzung der Arbeit

- Die Arbeit beginnt mit einer Analyse der Problemstellung: Es soll eine Layoutanalyse auf vielfältigen Manuskriptarten geleistet werden. Allen zur Verfügung stehenden Manuskripten ist dabei gemein, dass Textzeilen ein periodisches Muster darstellen.

- Da die Gabor-Transformation als theoretisch fundierter Ansatz eine Transformation eines örtlich begrenzten Bereichs eines Manuskripts in den Frequenzraum gestattet, und damit eine lokale Auswertung periodischer Muster erlaubt, soll diese als Basis für die Untersuchungen dienen. Hiermit soll ein Textbereich von anderen, nicht-periodischen Bereichen wie Hintergrund unterschieden werden können.

- Experimentell soll eine Performanzanalyse im Hinblick auf die Erkennung des Zeilenabstandes und der Zeilenorientierung im Wertebereich von 0°-180°, sowie der darauf basierenden Segmentierung von homogenen Textbereichen in Textblöcke durchgeführt werden. Das Verfahren soll dabei robust gegenüber Störungen wie Rissen, Flecken, Verfärbungen oder Löchern durch Wurmfraß sein.

- Das Verfahren soll darauf ausgelegt sein, einen größeren, heterogenen Datenbestand zu analysieren. Hierfür soll weder eine Trainingsphase noch ein Benutzereingriff notwendig sein. Dann können Hunderte unterschiedlicher Manuskriptseiten ohne Unterbrechung analysiert werden, der Nutzer muss keine Veränderung einer Parametrierung vornehmen.
- Das entwickelte System soll auf einem Testdatensatz validiert werden. Der Testdatensatz muss groß genug sein, um die vielfältigen Herausforderungen bezüglich der unterschiedlichen Schreibsysteme, Kontrastverhältnisse, Störungen usw. hinreichend abdecken zu können.

Nicht Teil der Arbeit ist die Entwicklung einer Anwendungsschnittstelle, da die Layoutanalyse als ein Baustein unter der am SFB 950 entwickelten Benutzungsoberfläche AMAP integriert werden soll.

Eine Erweiterung der Erkennung der Zeilenorientierung von 0°-180° auf volle 360° könnte durch eine nachträgliche Benutzereingabe ermöglicht werden. Eine automatische Erkennung würde eine schreibsystemspezifische Anpassung bedingen, auf die hier verzichtet wird, da sie in Anbetracht der kaum überschaubaren Anzahl verschiedener Schreibsysteme und Schreibstilen nicht im Rahmen dieser Arbeit möglich wäre. Ansätze für Erweiterungen der Orientierungserkennung finden sich in [Srihari u. Govindaraju, 1989; Aradhye, 2005; Lu u. Tan, 2006].

1.3. Gliederung der Arbeit

Nach dieser Einleitung wird der Stand der Forschung im 2. Kapitel beleuchtet, angefangen mit der Layoutanalyse auf gedruckten Dokumenten bis zu Ansätzen mit unterschiedlichen Zielsetzungen auf handgeschriebenen Seiten.

Das 3. Kapitel beschreibt das in dieser Arbeit entwickelte Verfahren der Layoutanalyse im Einzelnen, angefangen mit der Anwendung der Gabor-Transformation und der Auswertung des so erhaltenen Frequenzspektrums. Es wird beschrieben, wie anhand des Spektrums eine Auftrennung von Text- zu Hintergrundregionen ermöglicht wird, und die Struktur von Textbereichen bestimmt werden kann. Es folgt die Beschreibung der schrittweisen Auflösungsreduktion des Manuskriptbildes bei gleichzeitiger Anwendung der Gabor-Transformation unter einheitlicher Fenstergröße, durch die eine systematische Erkennung von Textblöcken beliebiger Schriftgrößen ermöglicht wird.

Das 4. Kapitel stellt einen Testsatz von 50 Manuskripten vor und beschreibt, wie dieser Testsatz eigens zur Erstellung einer Ground Truth verarbeitet wurde, da bisher kein Testdatensatz zur Verfügung stand, der die Herausforderungen einer generischen Layoutanalyse hinreichend abbilden konnte. Darauf wird der Abgleich zwischen Ground Truth und den berechneten Ergebnissen beschrieben. Schließlich werden die Ergebnisse der Auswertung vorgestellt und bewertet.

Abschließend erfolgt im 5. Kapitel eine Zusammenfassung der Untersuchungen dieser Arbeit mit einem Ausblick auf mögliche Erweiterungen.

2. Stand der Forschung

In diesem Kapitel wird eine Übersicht über wichtige Arbeiten im Bereich der Layoutanalyse gegeben, angefangen mit gedruckten Dokumenten bis hin zu handgeschriebenen, die den aktuellen Stand der Forschung widerspiegeln.

Die Bildverarbeitung beschäftigt sich schon seit einigen Jahrzehnten mit dem Thema der Layoutanalyse von Dokumenten ([Selfridge u. Neisser, 1960]), anfänglich standen allerdings ausschließlich gedruckte Dokumente im Fokus.

Die physische oder geometrische Layoutanalyse ermittelt häufig zunächst, ob ein eingescanntes Dokument kantenparallel ausgerichtet ist, und korrigiert die Darstellung oder berücksichtigt darauf diesen Winkel vor einer weiteren Bearbeitung, beispielsweise zur Zeilenerkennung. Anschließend werden homogene Bereiche identifiziert, die als Text, Bilder oder Tabellen klassifiziert werden, um z.B. Bilder nicht dem folgenden Prozess der Zeichenerkennung (OCR) zu übergeben.

Die logische Layoutanalyse soll dann die erkannten Texte in unterschiedliche funktionale Entitäten unterteilen, wie Überschrift, Zusammenfassung, Haupttext, Bildunterschrift, Referenzen etc. Bei komplexem Layout soll die richtige Reihenfolge von verstreut angeordneten Textblöcken erhalten bleiben. Da für die logische Layoutanalyse in aller Regel genaue Kentnisse über die Art der vorliegenden Dokumente vorhanden sein müssen, soll hier nicht weiter auf sie eingegangen werden. In [Mao u. a., 2003] findet sich u.a. ein Vergleich vieler Methoden der logischen Layoutanalyse.

Seit einigen Jahren finden auch handschriftliche Dokumente das Interesse der Bildverarbeitung. Bibliotheken benötigen für Annotierungen von Manuskriptseiten Angaben über die Seitengröße, die Größe des Haupttextblockes, die Breite der Seitenränder und ähnliche Maße. Auch eine automatische Bestimmung und Differenzierung von Textbereichen, Bildbereichen und dekorativen Elementen liefert nützliche Informationen für Anwender sowie nachgelagerte automatisierte Verarbeitungsschritte.

2.1. Layoutanalyse von gedruckten Dokumenten

In folgendem Abschnitt wird zunächst ein Überblick über verschiedene Ansätze der physischen Layoutanalyse gegeben, die Textseiten in Spalten und Absätze unterteilt, bishin zu einzelnen Zeilen oder Wörtern.

Die Verfahren der physischen Layoutanalyse können neben frequenzbasierten Ansätzen in „Top-Down“, „Bottom-Up“ und hybride Ansätze unterteilt werden. Bei „Top-Down“ Ansätzen werden zunächst die gröbsten Strukturen erkannt (Seitengröße, Seitenränder), dann einzelne Elemente, die keinen Text enthalten (z.B. Bilder), sowie Textbereiche, welche darauf in Absätze, Zeilen, Wörter usw. zerlegt werden. Häufig werden Backtracking-Verfahren eingesetzt, die einer fehlerhaften Segmentierung entgegenwirken sollen. Auch ein schon frühzeitiger Abgleich mit logischen Mustern ist üblich.

Die „Bottom-Up“ Ansätze setzen dagegen eine erfolgreiche Binarisierung voraus, die den in der Regel hellen Hintergrund von dunklen Vordergrundelementen trennt. Eine Zusammenhangsanalyse ergibt die kleinsten Elemente, also meist einzelne Buchstaben oder Satzzeichen. Durch Auswertung der Abstände dieser Elemente untereinander werden Wörter gefunden, die dann zu Zeilen und daraus zu Absätzen zusammengefasst werden, bis alle Teile der Seite erkannt sind.

Ein Überblick über sehr frühe Ansätze beider Arten wird in [Nadler, 1984] gegeben.

Als *hybrid* bezeichnet man Verfahren, die beide Ansätze kombinieren.

Aktuellere Übersichten über unterschiedliche Herangehensweisen finden sich in [Mao u. a., 2003] und [Jain u. Yu, 1998].

Im folgenden werden einige bedeutende Arbeiten vorgestellt. Obwohl diese häufig schon älter sind, finden sich deren Ansätze, häufig in optimierter und kombinierter Form, bis heute in vielen Systemen wieder. Beispielhaft dafür ist das weiter unten beschriebene Fraunhofer System.

Im Anschluß an die Vorstellung folgt in Abschnitt 2.1.5 eine Bewertung der Ansätze im Hinblick auf diese Arbeit.

2.1.1. Frequenzbasierte Ansätze

Postl [1986] hat bereits 1986 auf der Basis der Fouriertransformation die Zeilenausrichtung eingescannter Dokumente erfasst. Eingescannte Dokumente sind häufig nicht perfekt gerade ausgerichtet, dieses ist jedoch Voraussetzung für viele Methoden, die auf diesen Bildern eine Zeichenerkennung durchführen. Postl nutzt dabei die Tatsache aus, dass gedruckte Textzeilen in einem weitgehend einheitlichen Abstand zueinander stehen und sich aus dem Text und dem Weißraum zwischen den Zeilen ein sehr regelmäßiges Muster aus abwechselnd hellen und dunklen Bereichen ergibt. Die Ausrichtung einer solchen Textur wird dann durch die Fouriertransformation ausgewertet.

Ein weiteres frequenzbasiertes Verfahren haben Jain u. Bhattacharjee [1992] entwickelt. Zielsetzung ist hierbei, zwischen Textblöcken und anderen Elementen wie Bildern, Grafiken oder Weißraum zu unterscheiden, um später reine Texte an OCR-Programme weitergeben zu können.

In diesem Verfahren werden zunächst Fenstergrößen verwendet, die einer Zweierpotenz von 2^n Pixeln entsprechen. Für diese Fenster werden Gabor-Filterbänke von jeweils $n-1$ radial abgestimmten Filtern und für jeden Radius 4 Orientierungen ($0°, 45°, 90°$ und $135°$) verwendet, d.h. $(n-1) \times 4$ Filter insgesamt.

Die daraus resultierenden Bilder werden in einem nächsten Schritt durch ein nichtlineares Filter aufbereitet, das Ergebnis dann als Antwortbild bezeichnet. Innerhalb der Antwortbilder werden zentriert an jedem Pixel kleine, sich überlappende Fenster erzeugt und der Durchschnittswert eines jeden Fensters aus den Beträgen der Werte der einzelnen Bildpixel innerhalb der Fenster berechnet. Die Durchschnittswerte von den sich überlappenden Fenstern werden als „Texturenergie“ beschrieben und in Eigenschaftsvektoren abgespeichert, die die Eigenschaften der Umgebung der zugrundeliegenden Pixel beschreiben.

Schließlich wird ein Clustering-Algorithmus angewandt, der diese Eigenschaftsvektoren auf drei Clusterzentren aufteilt. Ein Clusterzentrum soll dabei den Textblöcken entsprechen, ein weiteres den Nicht-Text Bereichen und das dritte den Grenzabschnitten zwischen den ersten beiden. Die Nicht-Text Bereiche sollen sich dabei durch nur wenige Variationen bzgl. der Intensität auszeichnen. Eine rechtwinklige Maske fasst dann alle Bildpixel im Originalbild zusammen, die dem Cluster entsprechen, welches die Texteigenschaften beschreibt und segmentiert so zuletzt die Textbereiche.

Ein Wavelet-basierter Ansatz wird in [Doermann, 1995] vorgestellt. Die Grundidee ist, aufgrund unterschiedlicher Texturen drei Arten von Inhalten zu unterscheiden: Text, Bilder und Grafiken. Die Merkmalsvektoren einer Wavelet-Packet-Transformation werden von sich überlappenden Fenstern über dem Dokument bestimmt. Eine „Vote Propagation Matrix“ (VPM) nimmt für die einzelnen Bildkoordinaten alle Klassifizierungen der einzelnen Fenster auf, die über diesem Bildbereich liegen. Begonnen wird mit großen Fenstern, deren Klassifizierung allerdings nicht eindeutig sein könnte, da unter einem Fenster Inhalte unterschiedliche Art liegen könnten. Daher werden verschiedene Skalierungen durch sukzessiv kleinere Fenster ausgewertet, deren Klassifizierungen die VPM ergänzen.

Die Entscheidung, welcher Klasse eine bestimmte Bildkoordinate zugeordnet werden soll, wird also innerhalb der VPM einerseits durch die Klassifizierung von Fenstern verschiedener Größe, aber auch durch mehrere Fenster derselben Größe, aber mit unterschiedlichem Mittelpunkt beeinflusst, und wird schließlich durch die gewichtete Kombination aller Klassifizierungen getroffen.

In den Fällen, in denen a-priori Wissen über die Art der Dokumente vorhanden ist, kann die VPM auch vorbelegt werden und so optimierte Ergebnisse liefern.

2.1.2. „Top-Down“ Ansätze

Einer der bekanntesten ortsraumbasierten Ansätze ist XY-Cuts, beschrieben in [Nagy u. a., 1992], frühere Versionen finden sich bereits in [Nagy u. Seth, 1984; Nagy u. a., 1986]. Da das

im Folgenden beschriebene Verfahren besonders störungsanfällig auf schräge Zeilenorientierungen reagiert, werden nur exakt ausgerichtete Dokumente verwendet. Auch Flecken oder anderes Bildrauschen würde auf den binarisierten Dokumenten zu falschen Ergebnissen führen. Daher entfernt das verwendete Entrauschungsverfahren auch größere Elemente wie Satzzeichen, die für die reine Layoutanalyse nicht benötigt werden.

In horizontalen und vertikalen Projektionsprofilen werden nun pixelweise alle Zeilen bzw. Spalten der mit 300dpi eingescannten Dokumente, die keine schwarzen Pixel enthalten, mit Nullen beschrieben (Weißraum), alle anderen dagegen mit Einsen. Im nächsten Schritt werden mittels Lauflängenkodierung Folgen von Nullen und Einsen in Zeilen bzw. Spalten zu Profilabschnitten zusammengefasst. Bereits jetzt wird in dem hier beschriebenen Verfahren versucht, diese Profilbeschreibung logischen Einheiten zuzuordnen; so dürfte eine Überschrift in großer Schrift einen Abschnitt einer Länge von 100-160 aufeinanderfolgenden mit Einsen besetzten Zeilen haben, bei mehrzeiligen Überschriften werden mehrere solcher Abschnitte von jeweils einem von Nullen besetzten Abschnitt der Länge 4-20 unterbrochen. Ähnliche Regeln sind hier für Untertitel definiert, und auch von der weiteren horizontalen und vertikalen Aufteilung wird erwartet, dass sie einer vordefinierten Beschreibung entspricht. Grundsätzlich können Projektionsprofile rekursiv von allen mit Einsen besetzten Abschnitten erstellt werden, womit auch einzelne Spalten des Haupttextes erkannt und aufgetrennt werden können. Diese Rekursion wird solange angewandt, bis ein Textblock einem vorher definierten logischen Typ entspricht.

In [Baird u. a., 1990] wird ein Ansatz beschrieben, der als „global-to-local“-Strategie bezeichnet wird und auf der Auswertung des Weißraumes basiert. Hier werden zunächst auf binarisierten Dokumentbildern die umschließenden Rechtecke aller schwarzen Zusammenhangskomponenten bestimmt. Komponenten, die kein Text sein können, da sie zu groß oder zu klein sind, werden gelöscht. Aus der Anordnung der verbliebenen Komponenten werden Schrägheit und Scherung berechnet und mittels einer affinen Transformation korrigiert.

In einem nächsten Schritt werden alle maximal großen weißen Rechtecke bestimmt. Maximal groß bedeutet, dass alle Seiten des weißen Rechtecks entweder den Rand des Dokuments oder ein schwarzes Bildpixel berühren. Die Vereinigung dieser Rechtecke entspricht dem gesamten Weißraum eines Dokuments.

Anschließend werden die Rechtecke anhand ihrer Relevanz sortiert. Die Relevanz ist das Produkt des Flächeninhalts und des Logarithmus vom Seitenverhältnis. Die Autoren gehen von Dokumenten im Manhattan-Layout aus, in denen sich der Weißraum besonders gut durch entweder hohe, schmale oder durch flache, breite Rechtecke beschreiben lässt. Erstere sollen zwei Textspalten voneinander trennen, die anderen sollen Flächen entsprechen, die bspw. den Weißraum zwischen Überschrift und Haupttext kennzeichnen. Eine obere Grenze verhindert, dass zu flache oder schmale Rechtecke berücksichtigt werden, die im Weißraum zwischen zwei direkt aufeinanderfolgenden Textzeilen entstehen könnten.

Nach Abzug eine Maske, die durch die Vereinigung der relevantesten Rechtecke entsteht, bleiben die Flächen zurück, die den gesuchten einzelnen Textblöcken entsprechen sollten und so die Segmentierung darstellen. In einem Test auf über 100 Dokumenten mit Manhattan-Layout

waren 1% der nach Relevanz sortierten Rechtecke ausreichend, bei besonders komplexem Layout waren mehr als 5% nötig.

Dieser Ansatz wird in [Breuel, 2002] aufgegriffen, allerdings mit einem optimierten Algorithmus für das Aufinden der Menge aller maximal großen Rechtecke, die ausschließlich Weißraum enthalten. Gegeben ist zunächst die Menge der umschließenden Rechtecke aller Vordergrundelemente und ein Rechteck, welches dem äußeren Rand des Dokuments entspricht. Angefangen mit diesem Rechteck wird rekursiv untersucht, ob Vordergrundelemente darin enthalten sind. In dem Fall wird ein möglichst mittig liegendes Vordergrundrechteck gewählt und das aktuell untersuchte Rechteck in die vier Rechtecke unterteilt, die über, unter, links und rechts neben diesem umschließenden Vordergrundrechteck liegen, ansonsten innerhalb des aktuell untersuchten Rechtecks maximal groß sind. Diese vier Rechtecke werden im nächsten Rekursionsschritt zu den aktuell untersuchten Rechtecken. Sobald ein Rechteck keine Vordergrundelemente enthält, ist es Teil der Ergebnismenge.

Auf diese Weise wird sichergestellt, dass die größten Weißraum-Rechtecke zuerst in der Ergebnismenge erscheinen und die Rekursion beim Erreichen eines Qualitätskriteriums vorzeitig abgebrochen werden kann. Überlappungen von Ergebnisrechtecken können vermieden werden, indem diese direkt nach ihrer Bestimmung vom Verfahren als Vordergrundrechteck angesehen werden. Die Qualität des so gefunden Weißraumes hängt allerdings stark von der Formähnlichkeit der Vordergrundelemente mit ihren umschließenden Rechtecken ab.

2.1.3. „Bottom-Up“ Ansätze

In [Wahl u. a., 1982] wird das als „run-length smearing algorithm“ bezeichnete Verfahren vorgestellt. Hier wird ein binarisiertes Dokument in Form einer Bitmap dargestellt, die weißen Hintergrund durch die Ziffern 0 beschreibt, die Ziffern 1 dagegen repräsentieren Pixel von Vordergrundelementen. Aus dieser Bitmap werden nun zwei weitere Bitmaps erzeugt, die die ursprüngliche Bitmap in der Weise verändern, dass sie jeweils spalten- und zeilenweise eine ununterbrochene Folge von höchstens C_1 0-Bits durch 1 ersetzen. Als eine sinnvolle Größe für C_1 wird die in Pixeln gemessene Länge langer Wörter empfohlen. Anschließend wird die spaltenweise erzeugte Bitmap mit der zeilenweise erzeugten durch ein logisches UND verknüpft. Weißraum zwischen Wörtern, der sich vertikal über mehrere Zeilen erstreckt, kann zu kurzen horizontalen Unterbrechungen der kombinierten Bitmap führen. Um diese Störungen zu beseitigen, wird die Bitmap nun nochmals horizontal untersucht und C_2 lange Folgen von 0-Bits werden durch 1 ersetzt, wobei C_2 der in Pixeln gemessenen Breite nur weniger Zeichen entsprechen soll. Die nach einer Zusammenhangsanalyse aufgezeigten Regionen stellen die gewünschte Segmentierung dar, wobei jede einzelne Textzeile als eine eigene Region erfasst wird.

Nachfolgend wird jede Region anhand folgender Kriterien untersucht:

- Größe (Anzahl der Pixel) pro Region,
- Höhe und Breite der Region, und – jeweils bezogen auf die Ursprungsbitmap –

- Anzahl der Vordergrundpixel sowie
- Vordergrund-/Hintergrundtransitionen.

Durch Clustern von Kombinationen dieser Kriterien werden geeignete Schwellenwerte bestimmt, die eine Unterscheidung der Regionen in Text, Grafiken bzw. Bildern, sowie langen horizontalen oder vertikalen Strichen ermöglicht. Ein Zusammenfassen von homogenen Textzeilen zu Absätzen oder eine Unterscheidung zwischen benachbarten Absätzen wird dagegen nicht beschrieben.

In [O'Gorman, 1993] wird das „Docstrum" Verfahren beschrieben, welches ein binarisiertes Dokument erwartet, aus dem bereits Bilder und Diagramme entfernt wurden. In einem ersten Schritt des Verfahrens werden dann weiße oder schwarze Flächen entfernt, deren Größe unterhalb eines Schwellenwertes liegen, da sie häufig durch Rauschen verursacht wurden. Ob die Größe von diakritischen oder Satzzeichen dabei über oder unter dem Schwellenwert liegen, wird als unkritisch erachtet. Nach Bestimmung der Zusammenhangskomponenten werden die k nächsten Nachbarn jeder Komponente ermittelt, wobei für k häufig 5 gewählt wird, um von einem Buchstaben die vier benachbarten Buchstaben sowohl innerhalb der Zeile als auch in den direkt benachbarten Zeilen zu finden. Um Störungen vorzubeugen wird noch eine weitere Komponente mit einbezogen.

Die Abstände und Orientierungen (beschränkt auf den Wertebereich $0° - 180°$) werden ermittelt, und aus diesen Werten werden nach Anwendung einiger Filter lokale Clusterzentren bestimmt, die der groben Orientierung und den durchschnittlichen Abständen zwischen Buchstaben sowohl innerhalb einer Zeile, als auch zwischen benachbarten Zeilen entsprechen. Somit kann zunächst ermittelt werden, welche Buchstabenfolge innerhalb einer Zeile liegt. Die Orientierung einer Geraden, die mit geringstem Fehler durch die Mittelpunkte dieser Buchstaben verläuft, wird als genauere Orientierungsmessung herangezogen, da sie weniger durch nichtsymmetrische Buchstabenformen verfälscht wird.
Benachbarte Zeilen werden daraufhin als zwei Mengen von Buchstaben beschrieben, deren Anordnung annähernd parallel verläuft. Anhand gleichartiger Abstände zwischen Textzeilen werden diese zu Textblöcken zusammengefasst.

Soweit durch dieses Verfahren nicht alle Komponenten eines Dokumentes gleichzeitig ausgewertet werden, sondern z.B. aufgrund der gemessenen Komponentenabstände nur einzelne Regionen, können auch unterschiedliche Orientierungen von Textblöcken auf einer Textseite festgestellt werden, oder es kann auf unterschiedliche Textgrößen, wie in Überschriften, reagiert werden.

Auf Voronoi-Diagrammen basiert das in [Kise u. a., 1998] beschriebene Verfahren. Auf einem binarisierten Dokument werden zunächst Konturpunkte von zusammenhängenden Regionen berechnet; das Voronoi-Diagramm beschreibt nun die Flächen um jeden dieser Zentrumspunkte herum, in denen jede Stelle eine kürzere euklidische Distanz zum jeweiligen Zentrumspunkt

aufweist als zu allen anderen Zentrumspunkten. Kanten beschreiben Pixel, die den gleichen euklidischen Abstand zu zwei Zentrumspunkten haben. Darauf werden alle Kanten zwischen Flächen entfernt, deren Zentrumspunkte auf einer gemeinsamen Kontur liegen; das Resultat wird hier als Voronoi-Regionendiagramm bezeichnet.

Die Zielsetzung ist nun, aus diesem Regionendiagramm nicht-relevante Kanten zu entfernen, so dass die zurückbleibenden Kanten Textblöcke oder Bildelemente abgrenzen. Nicht-relevante Kanten bestehen beispielsweise zwischen Buchstaben, diakritischen Zeichen oder Satzzeichen innerhalb eines Textblockes. Kriterien zur Bestimmung nicht-relevanter Kanten sind neben dem minimalen Abstand zwischen Konturpunkten zweier benachbarter Flächen – der also annähernd dem minimalem Abstand zwischen zusammenhängenden Regionen entspricht – das Verhältnis der Flächengrößen benachbarter Flächen. Hierzu nötige Schwellenwerte werden aus einem Histogramm aller zuvor genannten Abstände ermittelt. Das Größenverhältnis orientiert sich am a-priori Wissen über maximal mögliche Größenunterschiede der im Dokument vorkommenden Zeichen. Experimente zeigen die Unempfindlichkeit auf Rotation und die Anwendbarkeit auf Dokumenten ohne Manhattan-Layout und Textblöcken, deren äußere Grenzen sich nicht durch Rechtecke beschreiben lassen. Selbst Textblöcke, deren Grundlinie nicht gerade ist, können segmentiert werden.

2.1.4. Hybride und weitere Ansätze

Pavlidis hat in [Pavlidis u. Zhou, 1992] eine hybride Methode vorgestellt, die ein Dokument zunächst in Textspalten aufteilt, um danach Elemente in Text, Bilder und Diagramme zu segmentieren. Auch hier wird von rechtwinkligen Textblöcken ausgegangen, die mehrspaltig auf Dokumentseiten auftreten. Ähnlich zu [Baird u. a., 1990] orientiert sich das Verfahren zunächst am Weißraum. Jedes mit 300dpi eingescannte Dokument wird in flache Streifen unterteilt, die eine Höhe von etwa 0,5" haben, auf denen eine vertikale Projektion errechnet wird. Durch vorheriges horizontales Glätten werden kleine Flecken und Rauschen unterdrückt; größere zusammenhängende Weißflächen werden dagegen zuverlässig erkannt und als horizontaler Abstand zwischen zwei Spalten interpretiert. Nach einem Abgleich der Position dieser Weißflächen mit denen der direkt vertikal benachbarten Streifen können die Weißflächen als zusammenhängend interpretiert werden, soweit die Positionen und Größe einem Toleranzmaß entsprechen. Abweichungen in Form breiterer Weißflächen werden als horizontale Abgrenzung zwischen Elementen interpretiert. Durch die Bestimmung der Projektion auf flachen Streifen ist diese Methode auch auf stärker gedrehten Dokumenten anwendbar, und nachdem Geraden an die Weißflächen angepasst worden sind, kann auch effizient der Drehwinkel des Dokuments ermittelt werden.

In einem nachfolgenden Schritt wird zwischen Text, Halbton (Bild) und Diagrammelementen unterschieden, indem die Kreuzkorrelation der Werte von dicht aufeinanderfolgenden horizontalen Pixelzeilen berechnet wird. Text und Diagrammelemente zeigen hier eine hohe Korrelation von direkt aufeinanderfolgenden Pixelzeilen, die bei größerer Entfernung stark abnimmt. Bilder hingegen zeigen eine niedrige Korrelation von direkt aufeinanderfolgenden Zeilen, die sich auch bei größerem Abstand wenig ändert. Schließlich wird der Anteil schwarzer

Pixel innerhalb der Diagramm- oder Textelemente berechnet, anhand dessen zwischen diesen Elementen unterschieden wird.

Als ein Teil eines Dokumentenverarbeitungssystems wird in [Jain u. Yu, 1998] ein weiteres Verfahren beschrieben. Auch hier werden auf einem binarisierten Bild die Zusammenhangskomponenten bestimmt. Mittels Hough-Transformation wird aus den Schwerpunkten der Komponenten, die anhand ihrer Größe etwa einzelnen Buchstaben entsprechen, der Orientierungswinkel des Dokuments bestimmt. Als „Generalized Text Line“ (GTL) wird eine Zusammenfassung der Komponenten bezeichnet, die sich – unter Berücksichtigung des Orientierungswinkels – vertikal überschneiden und deren horizontaler Abstand gering ist (<0,12"). Der Inhalt einer GTL wird als Text gewertet, falls ihre Höhe geringer als ein Schwellenwert ist und die Abweichung der vertikalen Positionen der Unterkanten der enthaltenen Komponenten geringer als ein weiterer Schwellenwert ist, oder aber sowohl die Breite der GTL als auch die durchschnittliche Höhe der enthaltenen Komponenten vorgegebenen Kriterien entspricht. Mehrere untereinanderstehende Textzeilen werden zu Textblöcken zusammengefasst. Darauf werden Nicht-Textregionen wie folgt klassifiziert:

- Komponenten, deren Größe und relativer Anteil an Vordergrundpixeln Schwellenwerte übertreffen, werden als Abbildungen interpretiert.
- Eine Tabelle wird erkannt, sofern nahezu parallele horizontale und vertikale Linien erkannt werden, deren Längen ähnlich sind. Alle weiteren Komponenten der Region müssen eine ähnliche Höhe aufweisen.
- Als Zeichnung wird jede Region gewertet, die nicht als Text, Bild oder Tabelle erkannt wurde.

Ein Ansatz, der viele Ideen vereint, wird am Fraunhofer Institut entwickelt [Konya u. a.]. Er wurde in [Antonacopoulos u. a., 2011] sowie in [Antonacopoulos u. a., 2013] evaluiert, wobei 2013 eine Variante mit verändertem Binarisierungsverfahren vorgestellt wurde.

Die Binarisierung ist der erste Schritt; statt einer Kombination von globalen und adaptiven Binarisierungsschwellenwerten kam in der letzten Fassung neben der globalen Schwellenwertbestimmung ein Schattenentfernungsverfahren zum Einsatz, außerdem wird für die Umwandlung von Farbbildern in Graustufen der L-Kanal des L*a*b*-Farbraumes genutzt.

Nach der Binarisierung folgen einige in [Gatos u. a., 2005] näher beschriebene Schritte: Nach Korrektur verdrehter Dokumentbilder wird durch einen Floodfill-Algorithmus ein äußerer Rand entfernt, der möglicherweise einer dunklen Scanunterlage entspricht. Ein Histogramm über die Höhe aller Zusammenhangskomponenten offenbart als Maximum die durchschnittliche Zeichenhöhe (unter der Annahme, dass einzelne Zeichen die häufigsten Komponenten sind). Die horizontale und vertikale Lauflängenuntersuchung – ähnlich zu der in [Wahl u. a., 1982] vorgestellten – lässt nun von Zeichen und Bildkomponenten horizontale und vertikale schwarze

Striche unterscheiden, die als Trennelemente festgestellt werden. Auch maximal große weiße Rechtecke werden nach der weiter oben beschriebenen Methode von Breuel bestimmt; die durchschnittliche Zeichenhöhe wird als Abbruchkriterium der Rekursion verwendet.

Weiterhin wird der Ansatz von [Jain u. Yu, 1998] integriert, nach dem Textregionen in grober Auflösung erkannt werden und von Nicht-Textregionen unterschieden werden. In einem letzten Schritt werden die zuvor erkannten Textregionen näher untersucht und in einer genauen Zeilenstruktur erfasst. Textzeilen werden anhand von Fontcharakteristiken (Strichdicke, Kursivität etc.) mittels dynamischer Programmierung zu homogenen Regionen gruppiert.

Ein weiteres System wird in [Shafait, 2008] vorgestellt. Die Aspekte der lokal adaptiven Binarisierung, der Erkennung der äußeren Ränder des Dokuments und die Klassifizierung der einzelnen Zonen werden in dieser Arbeit mit optimierten Methoden angegangen. Schließlich wird die Anwendbarkeit der Zeilensegmentierung auf Dokumente im arabischen Schreibsystem gezeigt, die das Urdu-Alphabet verwenden. Dabei werden die Herausforderungen insbesondere bezüglich des Nastaliq-Schriftstils offengelegt. Dieser weist in vielen hier gegebenen Beispielen einen geringen oder gar keinen Durchschuss auf, d.h. die Striche einzelner Zeichen berühren dabei Zeichen der benachbarten Zeilen. In der Arbeit wird auch ein Modell für Textzeilen in Urdu präsentiert, welches zwei Linien unterhalb der Grundlinie nutzt, da der Nastaliq-Schriftstil verschieden lange Unterlängen aufweist.

2.1.5. Fazit der Layoutanalyse auf gedruckten Dokumenten

Auch wenn in [Ramel u. a., 2007] noch Herausforderungen bezüglich historischer Dokumente mit komplexem Layout und in [Moll u. a., 2008] weitere herausfordernde Beispiele gezeigt werden, die mit den aktuellen Verfahren noch nicht zufriedenstellend gelöst werden können, so zeigen doch die Ergebnisse aktueller Wettbewerbe (wie Antonacopoulos u. a. 2013) beeindruckende Erkennungsleistungen auf gedruckten Dokumenten.

Im Einzelnen können „Top-Down“-Verfahren dadurch überzeugen, dass sie sich auf den Weißraum in binarisierten Dokumenten stützen, und anhand dessen schnell eine Grobstruktur der Dokumente erfassen können. Die Unteraufteilung wird beendet, sobald vorgegebene Kriterien erfüllt sind, etwa wenn Abstände zwischen möglichen Regionen kleiner als festgelegte Werte sind. Es wird davon ausgegangen, dass die Konturen der Regionen Rechtecken ähnlich sind und ausreichend Weißraum zwischen Regionen vorhanden ist. Soweit Verfahren mit gedrehten Orientierungen umgehen können, erwarten sie einen einzigen Korrekturwinkel für das gesamte Dokument. Einige Verfahren können Rotationskorrekturen nur bei geringen Abweichungen, etwa Winkel zwischen -20 und +20 Grad, erkennen und korrigieren. Teilweise kommen bei den Verfahren bereits frühzeitig Ontologien zum Einsatz, die die erwartete logische Dokumentenstruktur beschreiben und so beispielsweise die Segmentierung einer Überschrift über einem dreispaltigen Dokument vorgeben.

Verfahren	keine Binarisierung nötig	unterstützt Regionen mit nicht geradlinigen Konturen	bietet Rotationskorrektur	unterstützt unterschiedliche Orientierungen auf gleichem Dokument	unterstützt sich berührende Zeichen	unterstützt beliebige Textgrößen und Abstände auf gleichem Dokument	genauere Klassifikation von nicht-Text Elementen
Frequenzbasierte Verfahren							
Postl 1986	X	X	X		X	X	
Jain u. Bhattacharjee 1992	X	(X)		X	X	X	X
Doermann 1995	X	X		X	X	X	X
Top-Down Verfahren							
Nagy u. a. 1992					X		(X)
Baird u. a. 1990			X		X		
Bottom-Up Verfahren							
O'Gorman 1993		X	X	X		(X)	
Kise u. a. 1998		X		X		(X)	
Wahl u. a. 1982					X		X
Weitere Verfahren							
Pavlidis u. Zhou 1992			(X)		X		X
Jain u. Yu 1998			(X)				X

Tabelle 2.1.: Vergleich prominenter Layoutanalyseverfahren

Auch die „Bottom-Up"-Verfahren sind auf binarisierte Dokumente angewiesen, sie untersuchen allerdings Vordergrundelemente und können flexibler an unterschiedliche Regionen angepasst werden. In [O'Gorman, 1993] wird gezeigt, wie unterschiedlich gedrehte Textblöcke des gleichen Dokuments erkannt werden können, Kise u. a. [1998] zeigen neben der Verarbeitung von Textblöcken mit äußeren Konturen in beliebiger Form sogar eine Möglichkeit, Textblöcke ohne einheitliche Zeilenorientierung verarbeiten zu können. „Bottom-Up"-Verfahren sind allerdings auf eine besonders fehlerarme Binarisierung angewiesen. Häufig wird zunächst nach Zusammenhangskomponenten gesucht, die einzelnen Buchstaben entprechen, um davon ausgehend auf Wörter, Zeilen, Textblöcke sowie andere Elemente wie Bilder oder Diagramme zu schließen. Daher kommt es zu fehlerhaften Segmentierungen in Dokumenten mit Regionen in unterschiedlichen Schriftgrößen oder Zeilenabständen. Auch andere Schreibsysteme, in denen etwa Buchstaben nicht voneinander getrennt werden, oder sich die Zeichen benachbarter Zeilen berühren, werden in der Regel nicht unterstützt.

Die meisten aktuellen Systeme nutzen hybride Ansätze, sie kombinieren also die „Bottom-Up" Idee mit der „Top-Down" Richtung. So vereinen sie die Vorteile beider Systeme und erweitern sie häufig um spezielle Erkennungsverfahren, wie etwa Tabellenerkennung ([Gatos u. a., 2005]). Ein detaillierter Vergleich von sechs Verfahren zur Layoutanalyse findet sich in [Shafait u. a., 2008].

Die frequenzbasierten Ansätze kommen ohne vorherige Binarisierung aus. Jain nutzt dabei Gabor-Filter, deren Ausgaben er mittels Clustering-Verfahren gruppiert, um so Textbereiche und Bildelemente zu unterscheiden. Texte können dabei in beliebigen Orientierungen, unter-

schiedlichen Größen, Wort- und Zeilenabständen vorliegen. Abhängig von der Regelmäßigkeit der Texte sind unterschiedliche Anzahlen von Clusterzentren nötig. Auch Doermann kann durch die Analyse mit Wavelet-Filtern Textbereiche segmentieren, deren Konturen nicht geradlinig sind und deren Hintergrund unterschiedliche Grautöne aufweist, wie etwa bei Text vor einem Bildhintergrund.

Tabelle 2.1 lässt erkennen, dass die frequenzbasierten Ansätze unter den gegebenen Kriterien am flexibelsten einsetzbar sind. Eine holistische Rotationskorrektur ist für die Anwendung der Verfahren nicht nötig oder nicht sinnvoll, falls einzelne Textblöcke in unterschiedliche Richtungen gedreht sind.

2.2. Layoutanalyse von handgeschriebenen Dokumenten

Im folgenden Abschnitt werden Arbeiten vorgestellt, die sich auf handschriftlich verfasste Dokumente (Manuskripte) konzentrieren. Auch bei Manuskripten ergeben sich folgende Problemstellungen:

- Abgrenzen von Textbereichen zu sonstigen Bereichen (Bildern etc.)
- Zeilensegmentierung
- Wortsegmentierung
- Buchstabensegmentierung

Nach vollendeter Layoutanalyse kann neben einer Zeichenerkennung auch versucht werden, Schreiber zu identifizieren oder zu verifizieren, dass zwei Dokumente vom selben Schreiber verfasst wurden. Herausforderungen im Bereich der Layoutanalyse bieten Manuskripte im Vergleich zu gedruckten Dokumenten vor allem in Bezug auf die geringere Regelmäßigkeit der Schrift:

- Textregionen können, durch Nachlässigkeit oder ganz bewusst, stärker von einer rechteckigen Form abweichen.
- Die Abstände zwischen Zeilen sind weniger regelmäßig, ihre Orientierung uneinheitlicher.
- Die Ausprägung der einzelnen Buchstaben (Graphen) sind bei jedem einzelnen Vorkommen unterschiedlich.
- Die Strichdicke kann unterschiedlich sein, bis hin zu sehr dünnen Strichen, und der Tintenverlauf unterschiedlich dunkel

Dazu können weitere Störungen auftreten: Historische Manuskripte können bspw. eingerissen, mit Flecken versehen oder anders beschädigt sein; Tinte kann verschmiert oder ausgeblichen sein; Glossen, Symbole, Verzierungen oder Ergänzungen können sich zwischen Zeilen befinden, einzelne Wörter oder Buchstaben können durchgestrichen sein oder durch Wurmfraß unkenntlich geworden sein.

Im folgenden werden Arbeiten zu den Themen Binarisierung, Segmentierung von Textregionen sowie Zeilensegmentierung vorgestellt.

Auf darauf folgende Verarbeitungsschritte soll hier nicht näher eingegangen werden, in [Arica u. Yarman-Vural, 2001] wird ein Überblick über Fortschritte und Herausforderungen in Bezug auf die handschriftliche Zeichenerkennung gegeben.
Auch Verfahren zur Online-Handschriftenerkennung, wie sie auf modernen Tablet-Computern zum Einsatz kommen, werden hier nicht berücksichtigt.

2.2.1. Binarisierung

Viele Verarbeitungsschritte, die sich an der Prozesskette auf gedruckten Dokumenten orientieren, erwarten binarisierte Vorlagen. Die Binarisierung von historischen, insbesondere handschriftlichen Dokumenten, ist allerdings aus zuvor genannten Gründen ungleich schwieriger als bei modernen Dokumenten, deren Hintergrund im Original bereits nahezu weiß und deren Schrift nahezu schwarz ist. Seit 2010 wird alle zwei Jahre auf der „International Conference on Frontiers in Handwriting Recognition (ICFHR)“ die „Handwritten Document Image Binarization Competition (H-DIBCO)“ veranstaltet, auf der die hierzu angemeldeten Verfahren Binarisierungen zu vorgegebenen herausfordernden Dokumenten liefern (z.B. [Pratikakis u. a.] im Jahr 2016), und so einen vergleichbaren Stand der Kunst präsentieren. Ein ähnlicher Wettbewerb findet mit der DIBCO auf der ICDAR statt; auch deren Testsatz enthält handschriftliche Dokumente.

Das wohl am häufigsten verwendete Verfahren zur Bestimmung eines globalen Schwellenwertes, anhand dessen Dokumentbilder binarisiert werden, wird in [Otsu, 1979] beschrieben. Hierbei wird von einer Bimodalität des Grauwerthistogramms ausgegangen. Für alle möglichen Schwellenwerte, die im Histogramm zwei Klassen bestimmen können, werden die Varianzen der Verteilungen innerhalb der so bestimmten Klassen ermittelt. Ein optimaler Schwellenwert wird durch Minimierung der gewichteten Summe der intra-Klassen Varianzen gefunden. Die Gewichte ergeben sich aus den Anteilen der Bildpixel der Klassen. Laut dem Fazit des H-DIBCO Wettbewerbs von 2016 findet der Otsu-Algorithmus noch immer Anwendung als Teil einiger dort vorgestellter Verfahren und kann auf einigen Beispielen selbst mit den besten aktuellen Verfahren mithalten [Pratikakis u. a., 2016].

In einigen Fällen kann ein einziger Schwellenwert zur Binarisierung nicht ausreichend sein, z.B. bei Schattenwurf in bestimmten Lichtsituationen. Schwellenwerte, die anhand lokal begrenzter Abschnitte bestimmt werden, liefern dann bessere Ergebnisse, soweit die Größe der Abschnitte sinnvoll gewählt ist. Bei zu großen Fenstern könnten Unterschiede innerhalb des Fensters übersehen werden, bei zu kleinen Fenstern könnte eine Region betrachtet werden, in der keine Vordergrundelemente vorhanden sind, und der Schwellenwert aufgrund der fehlenden Information vollkommen falsch bestimmt wird. In [Sauvola u. Pietikäinen, 2000] wird ein fortschrittliches Verfahren vorgestellt, welches Dokumente anhand lokal adaptiver Schwellenwerte binarisiert.

Ein neueres Verfahren zur Binarisierung von Dokumenten wird von Lu u. a. [2010] beschrieben. Zunächst wird der Hintergrund durch ein iteratives polynomiales Glätten modelliert. Durch eine Abschätzung der Varianz über das gesamte Dokument wird zwischen Kontrasten, die durch Verunreinigungen oder durchscheinende Schrift entstehen, und denen der gewünschten Zeichenstriche unterschieden. Unter Zuhilfenahme dieser Informationen werden lokale Schwellenwerte bestimmt, anhand derer das Dokument binarisiert wird.

In [Howe, 2011, 2013] wird ein Verfahren vorgestellt, welches erstens anhand des Laplace-Operators die lokale Wahrscheinlichkeit abschätzt, ob Bildpixel zum Vordergrund oder Hintergrund gehören, und zweitens anhand des Canny-Kantendetektors lokale Übergänge bewerten kann. Beide Komponenten werden zu einer Funktion kombiniert, deren minimale Energie mittels Graph-Cuts bestimmt wird und somit die Binarisierung vorgibt.
Vier der sechs benötigten Parameter werden fest vorgegeben; die zwei weiteren werden als kritisch eingeschätzt. Eine automatische Bestimmung dieser zwei Parameter ist möglich, da durch eine kleine Veränderung der Parameter nur dann sehr ähnliche Resultate erzeugt werden, wenn die Parameterwahl nahezu optimal ist; bei schlechter Parametergröße unterscheiden sich die Ergebnisse durch geringfügige Veränderungen dagegen stark.

Auch wenn aktuelle Verfahren auf den Wettbewerben beeindruckende Resultate liefern können, so findet man hier noch einige Fälle, in denen fehlerhafte Binarisierungen auftreten.

Nachfolgende Verarbeitungsschritte wie eine Zeilensegmentierung werden diese Fehler jedoch nicht erkennen oder korrigieren, sondern reichen den Fehler häufig in verstärktem Umfang weiter, was zu fatalen Endergebnissen führen kann.

2.2.2. Analyse der Dokumentenstruktur

In diesen Bereich fallen alle Verfahren, deren Ziel es ist, auf handgeschriebenen Dokumenten Elemente wie Bilder, Dekorelemente oder Initialen, sowie Haupttext und Nebentextblöcke zu erkennen und zu unterscheiden. Sehr viele Ansätze, handschriftliche Dokumente zu analysieren, sind jedoch beschränkt auf eine Domäne, in der Texte eine zueinander ähnliche Form haben.

So wird z.B. in [Bulacu u. a., 2007] ein Verfahren beschrieben, in dem die handschriftlich verfassten Dokumente des Archivs der niederländischen Königin (Kabinet der Koningin [KdK]) analysiert werden. Bekannt ist das Schreibsystem und die Sprache (niederländisch), die horizontale Zeilenausrichtung sowie eine starre tabellarische Textanordnung in fünf Spalten. In der ersten bis dritten sowie der fünften Spalte befinden sich nur kurze Vermerke, in der vierten breiten Spalte ist dagegen der Haupttext zu finden. Die grobe Layoutanalyse beschränkt sich auf die Bestimmung von Rechtecken, die den einzelnen Tabellenzellen in etwa entsprechen. Diese Tabellenzellen sind in roter Farbe auf dem Papier vorgegeben. Deutlich aufwändiger ist die dann folgende Zeilensegmentierung, die auf Seite 23 beschrieben wird.

Eine Unterscheidung in Bild-, Dekor- und Textbereiche wird in [Grana u. a., 2011] gezeigt. Auf dem Datensatz einer sehr gut erhaltenen illuminierten (d.h. mit gemalten Bildern versehenen) Renaissancebibel wird zunächst der Text extrahiert. Hierzu werden die Manuskriptseiten in quadratische Kacheln aufgeteilt, deren Größe etwa der Höhe von sechs Zeilen Text entspricht. Die Ergebnisse der Autokorrelation lassen in dem Fall, in dem die Kacheln im Bereich der Textblöcke liegen, eine horizontal ausgerichtete Struktur erkennen; hingegen wird keine eindeutige Richtung deutlich, falls Kacheln von Dekor- oder Bildelementen korreliert werden. Hierdurch werden Textblöcke segmentiert und extrahiert, es verbleiben Bild- und Dekorelemente, deren Positionen und Ausmaße durch Binarisierung nach [Otsu, 1979] vom Hintergrund segmentiert werden. Diese so ermittelten Bereiche des Originaldokuments werden wiederum in Kacheln fester Größe unterteilt, und Histogramme ihrer Farb- und Gradientenverteilung werden erstellt. Eine zuvor trainierte Support Vector Machine (SVM) unterscheidet zwischen Bild- und Dekorhistogrammen und ermöglicht so die Segmentierung von Bildern auch innerhalb von Dekorelementen. Dekorelemente unterscheiden sich von Bildelementen in diesem Datensatz durch geringere Farbanzahl und stärkere Symmetrie.

Auch in [Garz u. a., 2011] wird das Ziel verfolgt, den Haupttext von Dekorelementen, aber auch Initialen oder Überschriften zu unterscheiden. Der dazu verwendete Datensatz besteht aus einem Manuskript in glagolitischer Schrift. Digitalisate der Manuskriptseiten zeigen Unregelmäßigkeiten von Hintergrund, Zeilenabstand und Zeilenausrichtung. Unterschiedliche Farbtoninformationen sind nicht vorhanden. Mittels des SIFT-Verfahrens von Lowe [2004] werden Merkmalsvektoren an Schlüsselpositionen bestimmt, die durch eine zuvor trainierte SVM in Haupttext und sonstige Elemente unterschieden werden. Ausgehend von diesen Schlüsselpositionen, ihrer Skalierung und Klassifikation werden mittels einer zweidimensionalen Gaußverteilung jedem einzelnen Bildpixel des Dokuments die jeweils wahrscheinlichste Klasse zugeordnet. Das Ergebnis wird als Flächen repräsentiert, in denen alle enthaltenen Bildpixel genau einer der beiden Klassen zugeordnet sind.

Die zuvor beschriebene Arbeit wurde weitgehend in das HisDoc-Projekt integriert, welches sich seit 2009 mit der Analyse historischer europäischer Handschriften aus dem Mittelalter beschäftigt. In diesem Projekt wurde ein Schwerpunkt auf die Segmentierung der Manuskripte in die vier Klassen Hintergrund, Dekoration, Haupttext und Nebentexte gelegt, wobei Nebentexte sowohl als Randbemerkungen als auch als interlineare Schrift im Haupttextbereich enthalten sein konnten. Die Bedeutung einer derartigen Layoutanalyse wurde durch einen Wettbewerb auf der ICDAR im Jahr 2017 deutlich ([Simistira u. a., 2017]). Hier sollte neben der beschriebenen Layoutanalyse auf Testsätzen aus drei Manuskripten auch eine Zeilensegmentierung stattfinden. Jedem der für diesen Wettbewerb angemeldeten Verfahren wurde ein Teil des Testsatzes von jeweils 30 Seiten pro Manuskript zum Training zur Verfügung gestellt. Nur eines von sechs Systemen, die eine Layoutanalyse ermöglichen, konnte ohne vorherige Trainingsphase auf dem bereitgestellten Datensatz eingesetzt werden. Die übrigen fünf Systeme basierten auf neuronalen Netzen und konnten durch die Trainingsphase ausnahmslos bessere Ergebnisse erzielen. Das beste Ergebnis einer anschließend geforderten Zeilensegmentierung wurde dagegen von einem

System erreicht, welches ohne Verwendung neuronaler Netze entwickelt wurde.

In [Gatos u. a., 2014] wird ein Verfahren zur Bestimmung von Textzonen beschrieben. Der hierzu verwendete Datensatz in Schreibschrift des lateinischen Alphabets enthält ein- und zweispaltigen Text, horizontale und vertikale Linien, Randnotizen oder Stempel. Hier wird zunächst die durchschnittliche Höhe aller Zusammenhangskomponenten auf binarisierten Bildern bestimmt. Da einige Dokumente den Haupttext durch vertikale Linien eingrenzen, wird zunächst nach diesen gesucht. Das zuvor bestimmte Maß der durchschnittlichen Höhe aller Komponenten dient der Unterscheidung dieser Linien von anderen Elementen. Bei Dokumenten ohne vertikale Linien werden vertikal aufeinanderfolgende weiße Pixelreihen bestimmt, die länger als ein Drittel der Dokumentenhöhe sind. Durch Zusammenfassen benachbarter Pixelreihen werden der Weißraum sowie die vertikale Abgrenzung der Textbereiche, und damit die Anzahl der Textspalten, ermittelt. Horizontal werden Textbereiche abgegrenzt, indem die maximalen Grenzen der umschließenden Rechtecke aller Zusammenhangskomponenten bestimmt werden, die weder Teile von Markierungslinien noch horizontal außerhalb der zuvor bestimmten vertikalen Abgrenzungen liegen.

In [Ouwayed u. Belaïd, 2008] wird eine Layouterkennung auf binarisierten Dokumenten vollzogen. Anwendung findet das Verfahren auf arabischen Handschriften, die viele Textbereiche mit unterschiedlicher Zeilenorientierung enthalten, aber keine sonstigen Elemente wie Bilder.

Dazu wird die Manuskriptseite in Kacheln gleicher Größe aufgeteilt. Die Größe der Kacheln wird anhand der durchschnittlichen Größe aller Zusammenhangskomponenten bestimmt und sollte etwa der Höhe dreier Zeilen entsprechen. Darauf erfolgt eine Zuordnung aller Zusammenhangskomponenten anhand der jeweiligen Schwerpunkte zu genau einer Kachel. Kacheln mit nur einer zugeordneten Komponente werden mit einer benachbarten Kachel verschmolzen.

Für alle nicht leeren Kacheln wird anhand der Maxima von Projektionsprofilen auf einem 15°-Raster eine grobe Orientierung bestimmt. Benachbarte Kacheln ähnlicher Orientierung werden darauf auf mögliche Vereinigung überprüft, indem die Projektionsprofile auf den vereinten Flächen bestimmt wird. Durch Auswertung der Wigner-Ville-Verteilungen auf den Projektionsprofilen wird nun das Projektionsprofil bestimmt, welches die höchste Energie ergibt und somit die Zeilenorientierung innerhalb der Fläche festlegt. Nur falls die so ermittelte Zeilenorientierung den Orientierungen der Einzelkacheln ähnlich ist, wird eine Vereinigung vorgenommen.

Sobald erkennbar ist, dass die Orientierung einzelner Kacheln deutlich von den Orientierungen größerer benachbarter Bereiche abweicht, wird angenommen, dass die Kachel gleichzeitig Text mehrerer Orientierungen enthält und weiter unterteilt werden muss.

Im Ergebnis werden zusammenhängende Textbereiche gleichartiger Orientierung segmentiert. Eine anschließende Zeilensegmentierung wird auf Seite 24 beschrieben.

Das Ziel, auf arabischen Dokumenten den Haupttextblock von Randtexten in unterschiedlichen Orientierungen zu trennen, wird in [Bukhari u. a., 2012a] verfolgt. Auch die hierbei verwendeten

Dokumente liegen binarisiert vor, und enthalten keine Dekorelemente oder Bilder. Zunächst werden Zusammenhangskomponenten bestimmt, die in Fenster der Größe 64×64 skaliert werden. Von diesen Komponenten werden ihre Höhe, Orientierung, Position innerhalb des Dokuments und das Verhältnis von schwarzen zu weißen Pixeln ermittelt. Zusätzlich zu diesem Fenster und den Merkmalen wird die Umgebung der Komponente in ein weiteres 64×64 Pixel großes Fenster skaliert. Alle diese Daten werden daraufhin durch ein neuronales Netz (Auto-MLP) klassifiziert. Das Ergebnis wird anschließend geglättet, indem jede Komponente den Durchschnitt der Klassifizierungen der Komponenten in der 150×150 Pixel Nachbarschaft erhält.

Die gleiche Zielsetzung wird in [Asi u. a., 2014] verfolgt. Dabei wird der gleiche binarisierte Datensatz genutzt und anhand der durchschnittlichen Höhe der Zusammenhangskomponenten wird ein Gabor-Filter eingerichtet, der horizontal ausgerichteten Text von Text in anderen Orientierungen und Schriftgrößen unterscheidet. Es wird hierbei davon ausgegangen, dass der Haupttext der überwiegende Bestandteil des Dokuments ist und die Schriftgröße des Haupttexts nahezu einheitlich ist. Da außerdem von rechteckigen Umrissen des Haupttexts ausgegangen wird, wird als nächstes ein Rechteck bestimmt, welches den Unterschied zwischen den vom Gabor-Filter als Hauptbestandteil erkannten Textbereichen innerhalb des Rechtecks und anderen außerhalb maximiert. Ausgehend von dieser Grobklassifizierung wird mittels Graph-Cuts eine Feinklassifizierung vorgenommen, die jede Komponente anhand ihrer umgebenden Klassifizierungen, der Position relativ zum Rechteck der Grobklassifizierung, sowie weiterer Merkmale neu klassifiziert. Diese weiteren Merkmale enthalten die Komponentengröße oder die Kompaktheit, was auf der Annahme beruht, dass der Haupttext tendenziell größer geschrieben ist als die Randtexte. Der Vergleich der endgültigen Segmentierung zeigt bessere Ergebnisse als bei [Bukhari u. a., 2012a].

Zusammenfassend kann festgehalten werden, dass es bisher nur wenige und begrenzt Leistungsfähige Verfahren im Bereich der Layouterkennung auf Manuskripten gibt.

- Es konnte erfolgreich gezeigt werden, wie Dekorelemente oder Bilder von Textbereichen getrennt werden können.
- Einfaches Textlayout, mit der Annahme von rechteckigen äußeren Grenzen, kann auf binarisierten Bildern durch Auswertung der Zusammenhangskomponenten, Markierungslinien und Folgen weißer Pixel erfasst werden.
- Besonders komplexes Layout von Textbereichen mit stark unterschiedlichen Orientierungen findet man häufig bei Manuskripten im arabischen Schreibsystem, welche von drei der beschriebenen Verfahren behandelt werden.

Eines dieser Verfahren unterteilt die Dokumentseiten in Kacheln und bestimmt die Orientierung des Textes innerhalb der einzelnen Kacheln. Durch Zusammenfassen benachbarter Kacheln mit Text gleicher Orientierung werden Textbereiche gleicher Orientierung segmentiert. Kritisch scheint hier die Kachelgröße zu sein, die anhand der durchschnittlichen Textgröße bestimmt

wird. Die verwendeten Dokumente scheinen nur Texte nahezu einheitlicher Größe zu enthalten. Da die Orientierung der Texte anhand von Projektionsprofilen über mehrere Zeilen erkannt wird, dürfte das Verfahren bei einzeiligen Randbemerkungen fehlerhafte Ergebnisse liefern, außerdem scheint es grundsätzlich stark von einer guten Binarisierung abzuhängen.

Auch die beiden weiteren Verfahren ([Bukhari u. a., 2012a], [Asi u. a., 2014]), die sich auf Manuskripte im arabischen Schreibsystem konzentrieren, benötigen eine Binarisierung als Basis. Sie wurden entwickelt, um einen Haupttextblock von Randtexten segmentieren können. Dabei liegen eine Reihe von Vorannahmen zugrunde, wie z.B.

- Es gibt nur einen Haupttextblock,
- dieser ist Hauptbestandteil der Dokumentseite,
- der Haupttextblock ist horizontal ausgerichtet
- und hat rechteckige Umrisse.

2.2.3. Segmentierung in Textzeilen

Eine Segmentierung von handschriftlichen Textbereichen in einzelne Zeilen hat in der letzten Zeit eine starke Aufmerksamkeit gefunden. Allerdings wird erst in den jüngsten Wettbewerben auf die Herausforderungen historischer Manuskripte eingegangen und den Verfahren wird vor der Anwendung ein umfangreicher Trainingsdatensatz von Manuskriptseiten im gleichartigen Stil zur Verfügung gestellt. Noch in den Wettbewerben [Gatos u. a., 2010] oder [Stamatopoulos u. a., 2013] wurden dagegen Dokumente mit nur einem Textbereich von einem Schreiber untersucht. Die dort untersuchten Manuskripte enthalten auch keine Linien oder Bilder, daher ist eine vorherige Bestimmung der Textbereiche unnötig. Die Dokumente sind bereits binarisiert, und enthalten keine Störungen, wie sie durch Flecken oder Risse entstehen könnten. Die Bandbreite geht von sehr gleichmäßiger Schrift mit großem Zeilenabstand bis hin zu ungleichmäßigem Zeilenabstand, schrägem Zeilenverlauf und sich vertikal berührenden Zeichen. In den Wettbewerbsdatensätzen befinden sich hauptsächlich Manuskripte im lateinischen Schreibsystem, aber auch im griechischen oder verschiedenen indischen Schreibsystemen.

Alle Verfahren zur Zeilentrennung konnten auf dem Testdatensatz von 2013 mit über 90-prozentiger Genauigkeit Zeilen trennen, die besten Verfahren erreichten fast 99%. Einige der hier vorgestellten Verfahren zur Zeilensegmentierung beruhen auf Bildglättungen mit unterschiedlicher Stärke, andere reduzieren die Bildauflösung, wieder andere zerteilen das Bild in vertikale Streifen und untersuchen deren Projektionen, oder untersuchen direkt die Zusammenhangskomponenten.

Ein detaillierter Überblick mit einer Auswertung der Stärken und Schwächen verschiedener Ansätze wird in [Likforman-Sulem u. a., 2007] gegeben. Nachfolgend werden exemplarisch drei Verfahren näher vorgestellt.

In [Bulacu u. a., 2007] wird das Droplet-Verfahren vorgestellt. Die sehr gut erhaltenen Dokumente sind bereits liniiert, so dass alle Textzeilen exakt gerade ausgerichtet sind und der Abstand genau gleichmäßig ist. Die Herausforderung liegt darin, dass sich Ober- und Unterlängen der Buchstaben aufeinanderfolgender Textzeilen im Profil überlappen und sich häufig auch berühren. Eine grobe Textzeilenbestimmung wird durch eine horizontale Projektion gegeben. Auf den binarisierten Dokumenten wird nun anhand der Projektionsminima versucht, eine ununterbrochene Folge weißer Bildpixel zu finden. Angefangen von einer Seite werden so lange horizontal benachbarte weiße Pixel markiert, bis auf ein schwarzes Pixel, welches zur Schrift gehört, getroffen wird. Von diesem Punkt ausgehend wird entlang der Kontur in beide Richtungen versucht, eine Ober- oder Unterlänge zu umgehen. In dem Fall, in dem ein Ausweichen entlang der Kontur in beiden Richtungen in die Nähe der Maxima der Zeilenprojektion führt, wird davon ausgegangen, dass sich hier zwei Zeichen berühren, und diese werden getrennt.

In [Ouwayed u. Belaïd, 2008] werden anschließend an das auf Seite 21 beschriebene Verfahren zur Layouterkennung auf arabischen Dokumenten Textzeilen segmentiert, indem an Stellen der bereits vorliegenden Projektionsmaxima einzelner Textregionen gleicher Orientierung Grundlinien angenommen werden. Alle Komponenten, die von diesen Grundlinien geschnitten werden, bilden einzelne Textzeilen. Kleinere Komponenten, wie diakritische Zeichen, werden der am dichtesten benachbarten Grundlinie zugeordnet. Auf die Problemstellung von sich über Zeilen hinweg berührender Zeichen wird nicht eingegangen.

Auf Grauwertbildern von chinesischen Manuskripten mit nur einem Textbereich gleichartiger Größe und Orientierung werden in [Herzog, 2007] Zeilen segmentiert. Durch eine Fouriertransformation wird eine grobe Abschätzung des Zeilenabstandes bestimmt, welche die Parameter für eine Glättung mit einem asymmetrischen Gaußfilter liefert. Auf dem so geglätteten Grauwertbild wird eine Wasserscheidentransformation ausgeführt, die zu einer starken Übersegmentierung führt. Anhand der Kriterien wie Orientierung, erwartete Zeilenhöhe (bzw. Breite bei vertikaler Zeilenorientierung), und den Grauwerten unter den Kanten werden nicht relevante Kanten der Segmentierungsmaske gelöscht, so dass nur die Kanten zurückbleiben, die Zeilen voneinander segmentieren. Dabei kann das Verfahren mit Störungen oder Zeichen umgehen, die sich über Zeilen hinweg berühren.

2.3. Fazit verschiedener Layoutanalyseverfahren

Bereits vor 1990 gab es Ansätze, um auf gedruckten und bereits binarisierten Dokumenten eine Layoutanalyse durchzuführen.

Die meisten Verfahren segmentieren dabei Dokumente, indem sie sich entweder am Hintergrund orientieren, dessen Ausdehnung ausgehend von großen Weißraumflächen hin zu Absatzzwischenräumen erfasst wird („Top-Down“), oder sie starten bei Zusammenhangskomponenten

des Vordergrundes, die i.d.R. einzelnen Buchstaben entsprechen, und fassen später Zeilen und Textblöcke zusammen („Bottom-Up“).

Die „Top-Down“-Verfahren gehen dabei von rechteckigen Umrissen der Textblöcke aus, erwarten Textblöcke mit weitgehend horizontal ausgerichteten Grundlinien und guter Binarisierungsqualität.

Die „Bottom-Up“-Verfahren reagieren noch empfindlicher auf Störungen der Binarisierung. Grundsätzlich sind sie in der Lage, von rechteckigen Umrissen abweichende Textblöcke zu erkennen und Dokumente zu verarbeiten, die von horizontalen Grundlinien stark abweichende Textblöcke aufweisen. Sie werden jedoch durch die Annahme beschränkt, dass sich der überwiegende Anteil kleiner Zusammenhangskomponenten aus einzelnen Buchstaben zusammensetzt, wodurch sie einerseits auf Schreibsysteme mit getrennt geschriebenen Buchstaben optimiert sind, andererseits Dokumente mit nahezu einheitlicher Schriftgröße erwarten. Abweichungen von diesen Annahmen führen bereits frühzeitig zu gravierenden Segmentierungsfehlern.

Die hybriden Ansätze kombinieren beide Techniken und können von den Vorteilen beider Verfahren profitieren.

Frequenzbasierte Verfahren können bereits auf gedruckten Dokumenten ihre Leistungsfähigkeit zeigen. Aufgrund der sehr guten Regelmäßigkeit von Buchstaben- und Zeilenabständen können sie erfolgreich zwischen Text-, Bild und weiteren Elementen unterscheiden. Die Form und Orientierung der Textblöcke ist dabei unerheblich, eine vorherige Binarisierung ist nicht nötig.

Zur Layoutanalyse handschriftlicher Dokumente gibt es bisher nur relativ wenige Arbeiten. Forschungsschwerpunkte waren seit vielen Jahren die Themen Binarisierung sowie Zeilensegmentierung auf binarisierten Dokumenten.

Da sich in der westlichen Welt nur noch sehr wenige moderne handschriftlich verfasste Dokumente von Interesse finden, beziehen sich die meisten aktuellen Arbeiten auf historische Dokumente, bei denen die Binarisierung durch Beschädigungen der Dokumente besonders anspruchsvoll ist. Um direkt anschließend eine Zeilensegmentierung durchführen zu können, müsste jedes Dokument aus lediglich einem Textblock bestehen und keine weiteren Elemente enthalten. Daher ist eine vorherige Layoutanalyse nur selten vermeidbar.

Arbeiten zur Layoutanalyse auf Manuskripten sind bisher in der Lage, zwischen Text-, Bild- und Dekorelementen mittels Farbinformationen oder den Haupttext von Überschrift, Initialen und Dekorelementen zu unterscheiden.

Bestehende Verfahren zur Lokalisierung der Textblöcke treffen in der Regel viele Vorannahmen, so wird meistens von einer sehr guten Binarisierung ausgegangen und davon, dass die Höhe der am häufigsten vorkommenden Zusammenhangskomponenten denen einzelner Wörter entspricht. Auch Position, Anzahl und rechteckige Umrisse von Haupttextblöcken werden meistens antizipiert. Zudem sind die bisherigen Verfahren auf ein einziges Schreibsystem, häufig auf Dokumente einer einzigen Domäne, optimiert und damit bei anderen Schreibsystemen nicht anwendbar.

In den letzten Jahren ist ein Trend festzustellen, der von trainingsfreien hin zu trainingsbasierten Ansätzen auf Basis neuronaler Netze führt. Diese Verfahren scheinen allerdings nur auf größeren, in sich homogenen Datensätzen sinnvoll anwendbar zu sein, und benötigen für die Trainingsphase eine beträchtliche Menge an zuvor fehlerlos klassifizierten Daten der gleichen Art. Unter diesen Vorraussetzungen liefern sie zwar meistens weitaus bessere Ergebnisse als die trainingsfreien Verfahren, allerdings ist es fraglich, inwieweit diese Rahmenbedingungen in der praktischen Anwendung erfüllt werden können.

Für die gewünschte Zielsetzung dieser Arbeit ist kein bisher bestehender Ansatz ausreichend, allerdings können die frequenzbasierten Ansätze schon bei gedruckten Dokumenten durch ihre flexiblen Einsatzmöglichkeiten und ihre Robustheit gegenüber Binarisierungsfehlern überzeugen. Weiterhin werden sie auch mit Erfolg auf handgeschriebenen Dokumenten eingesetzt. Daher wird im folgenden Kapitel ein eigener, frequenzbasierter Ansatz vorgestellt, bei dem auf eine vorherige Binarisierung verzichtet wird.

3. Layoutanalyse auf Basis der Fouriertransformation

Im folgenden Kapitel wird die Methodik der Layoutanalyse dieser Arbeit vorgestellt. Das Ziel dieser Layoutanalyse ist es, auf heterogenen Datensätzen Textblöcke von mindestens drei Zeilen zu erkennen, den Umriss der Blöcke vom Hintergrund zu segmentieren, und benachbarte Textblöcke voneinander zu segmentieren, soweit sie anhand der Zeilenabstände oder Zeilenorientierung unähnlich genug sind. Dabei sollen im Gegensatz zu bisher entwickelten Verfahren keine Bedingungen bezüglich der Form oder Anzahl von Textblöcken gemacht werden, und auch eine Trainingsphase ist nicht vorgesehen, da in der Regel keine bereits annotierten Daten zur Verfügung stehen, die hierfür notwendig wären.

Anschließend kann für jeden Textblock der durchschnittliche Zeilenabstand sowie die Orientierung der Zeilen ausgegeben werden. Die Kombination der Merkmale Zeilenabstand und Zeilenorientierung soll in dieser Arbeit als *Struktur* definiert werden.

In Abschnitt 3.2 wird eine Segmentierung von Textblöcken zum Hintergrund anhand der Stärke der festgestellten vorherrschenden Frequenz vorgestellt. Zueinander benachbarte Textblöcke mit gleichartiger Struktur können mit diesem Verfahren voneinander segmentiert werden, wenn ihr Abstand zueinander mindestens 2,5 Zeilen beträgt.

In Abschnitt 3.3 wird zunächst vorgestellt, mit welcher Auflösung unter welchen Bedingungen die Zeilenstruktur gemessen werden kann. Die Abschnitte 3.3.1 und 3.3.2 zeigen dann mögliche Verteilungen von Zeilenabständen und Orientierungen innerhalb jeweils eines beispielhaften Textblocks.

Da Manuskripte in vollkommen unterschiedlichen Auflösungen vorliegen können, und auch auf einer einzelnen Manuskriptseite Textblöcke von stark unterschiedlicher Schriftgröße auftreten können, scheinen feste Vorgaben von Minimal- und Maximalwerten zu bestimmender Größen für eine universelle Anwendung ungeeignet. Stattdessen sollen Manuskriptbilder in unterschiedliche Auflösungen skaliert werden, um diese auf sich ergänzende festgelegte Frequenzanteile zu untersuchen, und somit ein sehr breites Spektrum von möglichen Textblockstrukturen erfassen zu können. In Abschnitt 3.4 wird dieses Vorgehen beschrieben. Im Anschluß wird eine komplette Segmentierung von Manuskriptseiten in Hintergrund sowie verschiedene Textblöcke mit unterscheidbarer Zeilenstruktur vorgestellt (Abschnitt 3.4.1).

Implementiert wurde das Verfahren in Python, da für Python mit numpy und scipy nicht nur umfangreiche Bibliotheken für wissenschaftliche Zwecke zur Verfügung stehen, sondern auch eine

Schnittstelle zur VIGRA Bildverarbeitungsbibliothek [Köthe, 2000], deren Implementationen in dieser Arbeit hauptsächlich genutzt wurden.

Ergebnisse einer Auswertung des Verfahrens auf 50 Manuskriptbildern folgen in 4. Kapitel.

3.1. Grundlagen der frequenzbasierten Layoutanalyse

Längere Schriftstücke werden in unterschiedlichsten Kulturkreisen durch eine Folge von Zeichen in horizontalen oder vertikalen Zeilen verfasst. Diese Zeilen stehen nebeneinander oder untereinander, in einigermaßen regelmäßigen Abständen, und formen einen Textblock. Dabei ist erkennbar, dass Textblöcke ein ebenso regelmäßiges Streifenmuster bilden, unabhängig davon, ob der Text in dunkler Schrift auf hellem Untergrund oder umgekehrt verfasst ist. Nur im Falle von Inschriften, bei denen das Schreibmaterial in einer dritten Dimension ausgearbeitet wird, wie z.B. im Falle der Keilschrift oder bei Steininschriften, ist ein abwechselnd helles und dunkles Streifenmuster nicht immer erkennbar, sofern keine besondere Beleuchtungssituation vorliegt.

Dokumente, die in Farbdarstellung vorliegen, werden auf Graustufenrepräsentation reduziert. Hierzu wird in fast allen Fällen der Rotkanal genutzt, da bei dem am häufigsten vorkommenden Fall der dunklen Schrift auf hellem Hintergrund der Hintergrund leicht hellrötlich-braun erscheint. In Fällen mit roter Schrift wird stattdessen der Grünkanal benutzt, dadurch ergibt sich der maximale Kontrast zwischen Schrift und Hintergrund.

Die Erfassung dieser Streifenmuster soll als alleiniges Hilfsmittel dienen, um Textblöcke zu erkennen, da es Textblöcke in beliebigen Schreibsystemen und Kulturkreisen charakterisiert. Alle weiteren Annahmen, wie die Anzahl von Textblöcken auf einer Seite, rechteckige Umrisse oder Ausrichtung von Textblöcken, würden die universelle Anwendbarkeit der Methodik einschränken und werden daher vermieden.

Zunächst wird gezeigt, dass die Fouriertransformation in der Lage ist, die Frequenz eines regelmäßigen Streifenmusters zu extrahieren. In Abschnitt 3.1.2 wird die Gabortransformation eingeführt, die als Erweiterung der Fouriertransformation betrachtet werden kann. Sie schränkt den untersuchten Bildbereich durch eine Fensterfunktion ein, so dass neben der Frequenz auch deren Position im Bild ermittelt werden kann.

Die Fenstergröße ist dabei eine wichtige Verfahrensgröße und muss für eine Anwendung optimal bestimmt werden.

3.1.1. Layoutanalyse durch zweidimensionale diskrete Fouriertransformation

Definiert ist die zweidimensionale diskrete Fouriertransformation (2D-DFT) wie folgt ([Gonzalez u. Woods, 2018]):

$$F(u,v) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m,n) \cdot e^{-2\pi i(\frac{mu}{M}+\frac{nv}{N})} \tag{3.1.1}$$

Für ein Grauwertbild der Größe $M \times N$ wird so über alle Pixelwerte ($f(m,n)$) die Stärke aller möglichen diskreten Frequenzen (u,v) auf diesem Bild berechnet, wobei $0 < u < M-1$ und $0 < v < N-1$ gilt.

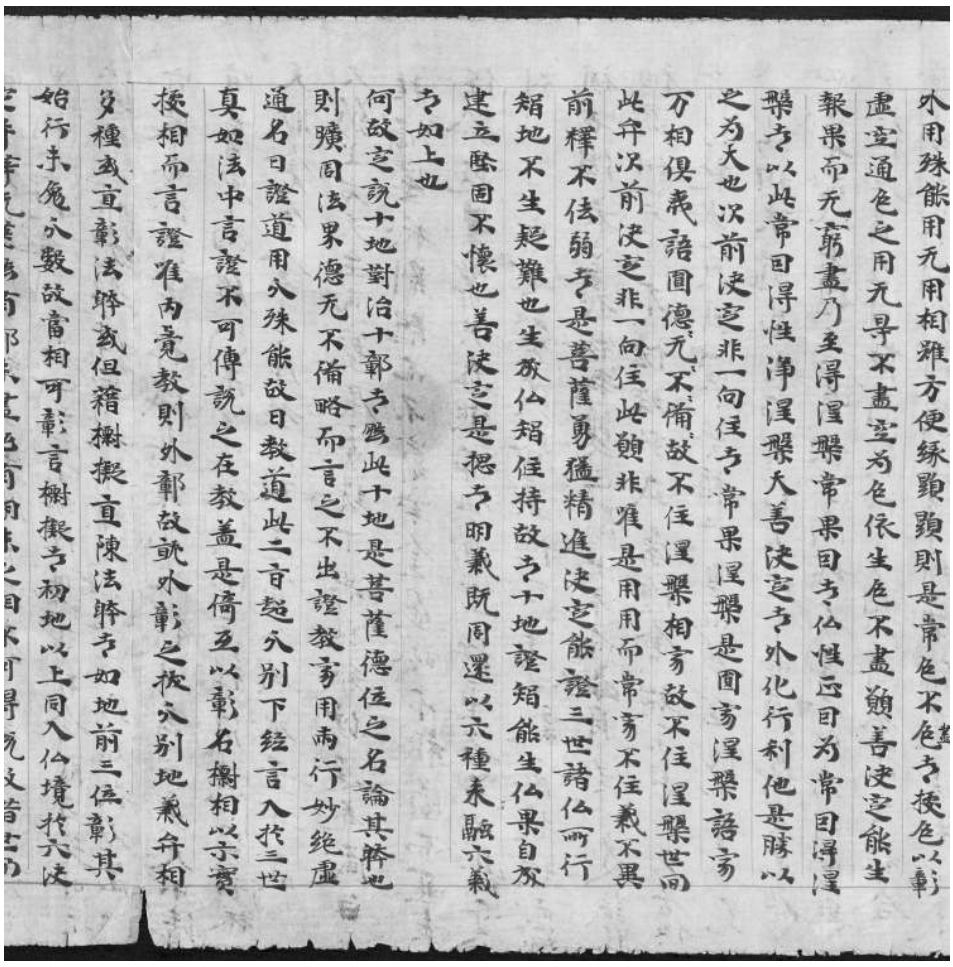
Der Nutzen einer solchen Transformation der Bildwerte aus dem Ortsraum in den Frequenzraum soll zunächst an einem Manuskriptbild demonstriert werden.

Angewendet auf den Manuskriptausschnitt in Abbildung 3.1a ergibt sich das Ergebnis in Abbildung 3.2, welches die Beträge der niedrigen Frequenzen im Wertebereich $-30 < (u,v) < +30$ darstellt, und als Frequenzspektrum bezeichnet wird ([Sonka u. a., 1998]).

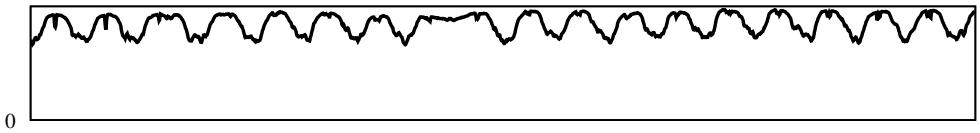
Hier werden lediglich die Grauwerte genutzt, die zwischen 0 und 255 liegen können. Farbbilder werden in dieser Arbeit vor der Frequenzanalyse in geeigneter Weise auf Grauwerte reduziert, so wurde bspw. für Abbildung 3.1 nur der Rotkanal des Originalbildes extrahiert.

Aus Gleichung 3.1.1 ergibt sich für den Koordinatenursprung des Spektrums der Wert, der der Summe aller Bildpixel entspricht. Dieser Wert wird auch als *DC* bezeichnet, für **D**irect **C**urrent. Da diese Summe deutlich größer ist als alle weiteren lokalen Maxima, und zudem keine weitere Relevanz hat, wurde dieser Wert stattdessen aus Darstellungsgründen auf Null gesetzt.

Erkennbar ist ein lokales Maximum an Position [19,0], sowie, aufgrund der Symmetrie, an [-19,0]. Die hier berechnete Frequenz entspricht dem regelmäßigen Zeilenmuster, welches in Abbildung 3.1b als Projektion auf die X-Achse verdeutlicht wird, und einen sinusähnlichen Verlauf hat. Die Position [19,0], exakt horizontal vom Koordinatenursprung gelegen, deutet auf vertikal orientierte Zeilen. Bei einer Bildbreite von M Pixel würde eine Folge von 19 sinusoiden Wellen (mit einer Wellenlänge von $\frac{M}{19}$ Pixel) zu diesem Ergebnis führen. Eine sinusoide Welle entspricht dann der Kombination aus dunkleren Pixeln des Zeichens und helleren des Zeilenzwischenraumes, also entspricht die Wellenlänge genau dem gesuchten Zeilenabstand. Da die diskrete Fouriertransformation von einem einheitlichen, unbegrenzt periodischen Signal ausgeht, bedeutet die Einheit 19 lediglich eine Wellenzahl von 19. Diese Wellenzahl gleicht der Anzahl an Wellen mit einer Wellenlänge von $\frac{M}{19}$Pixel, die in der Summe einer Länge von M Pixeln entsprechen. Dabei darf nicht direkt auf eine Zahl von 19 (vertikal ausgerichteten) Zeilen geschlossen werden. Würden die vier Zeichen in der mittleren Zeile und möglicherweise weitere Zeilen fehlen, würde die Stärke des erkannten Signals bei [19,0] sinken, sich jedoch nicht automatisch die Stärke an anderer Stelle auf einen vergleichbaren Wert erhöhen.



(a)



(b)

Abbildung 3.1.: Ausschnitt aus Or.8210/S.2717R, © The British Library Board.
(a) Manuskript mit einheitlicher Zeilenstruktur
(b) Projektion der Pixelwerte auf die X-Achse.

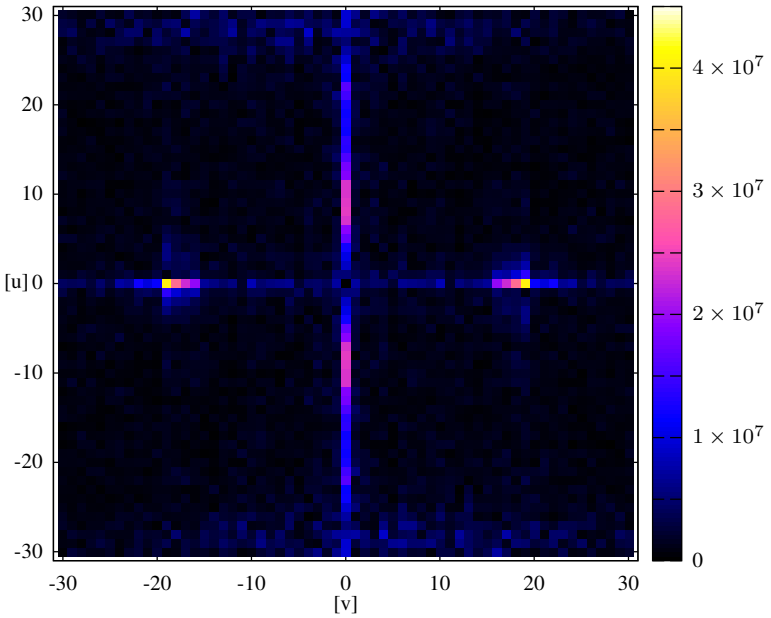


Abbildung 3.2.: Frequenzspektrum des Manuskriptausschnitts aus Abbildung 3.1a

Allgemein gilt:

$$\text{Zeilenabstand [Pixel]} = \frac{\text{Bilddurchmesser [Pixel]}}{\text{Wellenzahl}} \tag{3.1.2}$$

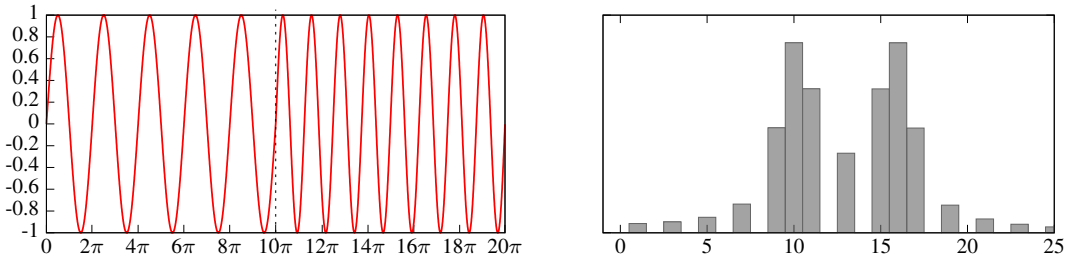
Obwohl der Begriff Wellenzahl häufig nur für ein eindimensionales Frequenzspektrum verwendet wird, findet der Begriff in dieser Arbeit auch für das zweidimensionale Spektrum Anwendung, da die Bedeutung gleich ist. Im zweidimensionmalen Spektrum entspricht die Wellenzahl dem euklidischen Abstand (Radius) vom Koordinatenursprung des Frequenzraumes.

Die weiteren lokalen Maxima im Verlauf der Y-Achse entstehen durch den oberen und unteren hellen Seitenrand. Diese und alle weiteren lokalen Maxima sind jedoch unkritisch, solange die Stärke dieser Frequenzen deutlich niedriger ist, als die durch das Zeilenmuster erzeugte Frequenz.

Solange also nur ein einheitliches, regelmäßiges Zeilenmuster auf einer Seite vorhanden ist, kann mit Hilfe der diskreten Fouriertransformation die durchschnittliche Zeilenorientierung und der durchschnittliche Zeilenabstand bestimmt werden.

3.1.2. Layoutanalyse durch die Gabor-Transformation

Durch die Fouriertransformation wird zwar erkennbar, welche Frequenzen in einem Signal vorhanden sind, nicht jedoch, an welchem Ort. Zur Verdeutlichung werden zunächst Überlegungen an eindimensionalen Signalen gezeigt, die sich direkt auf zweidimensionale Signale übertragen



(a) Signal mit zwei unterschiedlichen Frequenzen

(b) Frequenzspektrum des Signals

Abbildung 3.3.: Signal mit unterschiedlichen Frequenzen und relevanter Ausschnitt des zugehörigen Spektrums

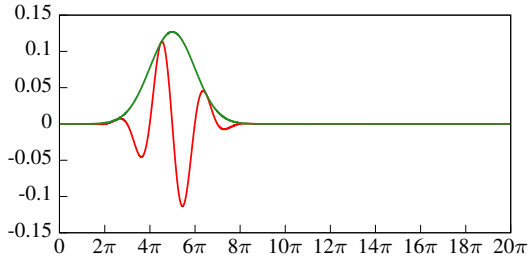
lassen. Abbildung 3.3 zeigt das Ergebnis einer diskreten Fouriertransformation eines eindimensionalen, veränderlichen Signals. Deutlich erkennbar sind die beiden Maxima bei den Wellenzahlen 10 und 16, die den Frequenzen des linken und rechten Teils entsprechen. Somit zeigt das Ergebnis zwar alle enthaltenen Frequenzen (bis zur Wellenzahl 25) auf, nicht jedoch deren Position. (Aus Symmetriegründen wurde in dieser Darstellung auf die Abbildung des negativen Frequenzbereichs verzichtet.)

Wenn ein vorhandenes Signal mit einer Gaußfunktion als Fensterfunktion multipliziert wird, so dass nur ein Teil des Signals ausgewertet wird, erhält man eine räumlich eingeschränkte Fouriertransformation. Die Verwendung der Gaußfunktion als Fensterfunktion wird bei einem diskreten Abtastraster häufig nach Dennis Gabor als Gabor-Transformation bezeichnet ([Gabor, 1946; Feichtinger u. Strohmer, 1998; Feichtinger u. a., 2008]).

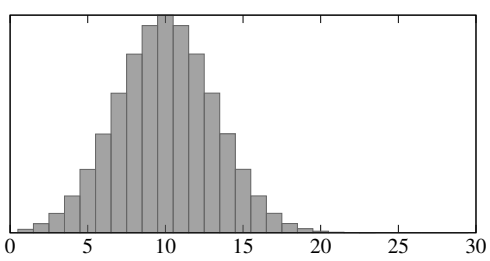
In Abbildung 3.4 wird der Effekt demonstriert: Die grünen Kurven in Abbildung 3.4a und 3.4c repräsentieren dabei die Fensterfunktion, mit der das Signal aus Abbildung 3.3a multipliziert wird. Das resultierende Spektrum charakterisiert den Ort, an dem die Fensterfunktion platziert ist. Da in dem Beispiel die Position des Fensters bekannt ist, kann zu dieser Ortsangabe eine einzelne Frequenz bestimmt werden, indem die Wellenzahl des Maximums bestimmt wird. Grundsätzlich muß dabei die Annahme erfüllt sein, dass das Fenster so schmal ist, dass es nur eine einzelne Frequenz umfasst.

Da mit jeder Fensterfunktion die Pixelwerte im Ortsraum multipliziert werden, wird hierdurch die Stärke der erkannten Frequenzen beeinflusst. Eine Rechteckfunktion als Fensterfunktion führt zu unerwünschten Störeffekten im Frequenzspektrum, sofern sich im Bereich des Fensters keine vollen Perioden eines Signals befinden. Dann führt eine angeschnittene Welle zu einer Diskontinuität, wenn die durch die Rechteckfunktion abgegrenzten Ausschnitte hintereinandergesetzt werden, wie es durch die Fouriertransformation implizit vollzogen wird, und zu einer Streuung über das volle Frequenzspektrum. Dieses als „Leck-Effekt“ bekannte Phänomen wird durch die Wahl einer geeigneten Fensterfunktion, die nahe (oder bei) Null beginnt, und sich über ein mittiges Maximum wieder Null annähert, vermieden.

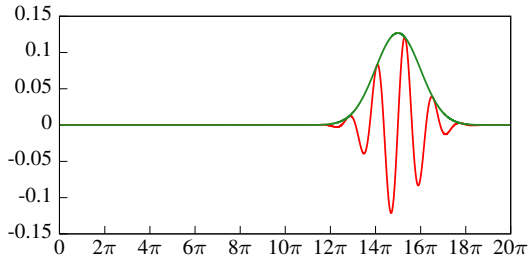
Da die Fouriertransformation einer Gaußfunktion wieder eine Gaußfunktion darstellt, ist die



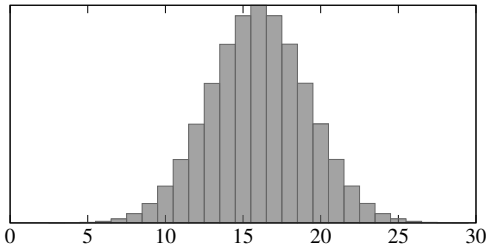
(a) Gauß-Fenster an Position $t = 5\pi$



(b) Frequenzspektrum von 3.4a



(c) Gauß-Fenster an Position $t = 15\pi$



(d) Frequenzspektrum von 3.4c

Abbildung 3.4.: Signal aus Abbildung 3.3 multipliziert mit einer Gaußfunktion, wodurch die Zuordnung der einzelnen Frequenzen zu Positionen ermöglicht wird

Gaußkurve als einzige Fensterfunktion optimal im Sinne der Unschärferelation, da sie sowohl im Orts- als auch im Frequenzraum gleichzeitig die maximale Auflösung erreicht. In dieser Arbeit wird daher die Gabor-Transformation verwendet, die auf der Gauß-Fensterfunktion basiert. Zwar wird auch durch die Gaußfunktion die Stärke aller erkannter Frequenzen beeinflusst, allerdings stützt sich diese Arbeit auf die Position relevanter lokaler Maxima im Spektrum, nicht auf deren absolute Stärke. Da die Gaußfunktion keine Nebenmaxima erzeugt, können Fehler aufgrund von Artefakten ausgeschlossen werden. Als Kompromiss zwischen rechnerischer Effizienz einerseits und andererseits dem Bestreben, Störeinflüsse auf das Spektrum zu minimieren, wird die Gaußfunktion bei $|\sigma| > 3$ mit Null gleichgesetzt.

Eine Übersicht über verschiedene Fensterfunktionen und ihre Eigenschaften finden sich beispielsweise in Harris [1978] und Heinzel u. a. [2002].

Bei der Gabor Transformation werden auch die Stärken der um den DC benachbarten Wellenzahlen beeinflusst, sofern Bildpixelwerte, die im Mittel nicht Null entsprechen, mit der Gaußfunktion multipliziert werden.

In Heinzel u. a. [2002] werden verschiedene Möglichkeiten vorgestellt, den Wert des DC bereits im Ortsraum zu eliminieren. Die Methode, von einem Bildausschnitt den Mittelwert zu berechnen, und diesen Wert von jedem Bildpixel zu subtrahieren (bevor der Bildauschnitt mit der Gaußfunktion multipliziert wird), wird in Abbildung 3.7 genutzt. Daher beschreiben in den links dargestellten Bildausschnitten die Grauwerte an den Bildaußenkanten Werte von nahezu Null,

alle dunkleren Bildteile weisen negative Werte auf und umgekehrt. Trotz dieser Behandlung sind die Werte im Frequenzspektrum um den Koordinatenursprung häufig deutlich verfälscht. Mit zunehmendem Abstand vom Koordinatenursprung wird der Störeinfluss jedoch geringer; selbst ohne DC-Behandlung wird die Stärke der Wellenzahlen von mindestens vier um weniger als ein tausendstel des DC Anteils erhöht. Aus diesem Grund wird in dieser Arbeit der Einfluß des DC vernachlässigt, soweit Wellenzahlen von mindestens vier betrachtet werden.

Die Stärke einer bestimmten Zeilenfrequenz ist von vier Faktoren abhängig:

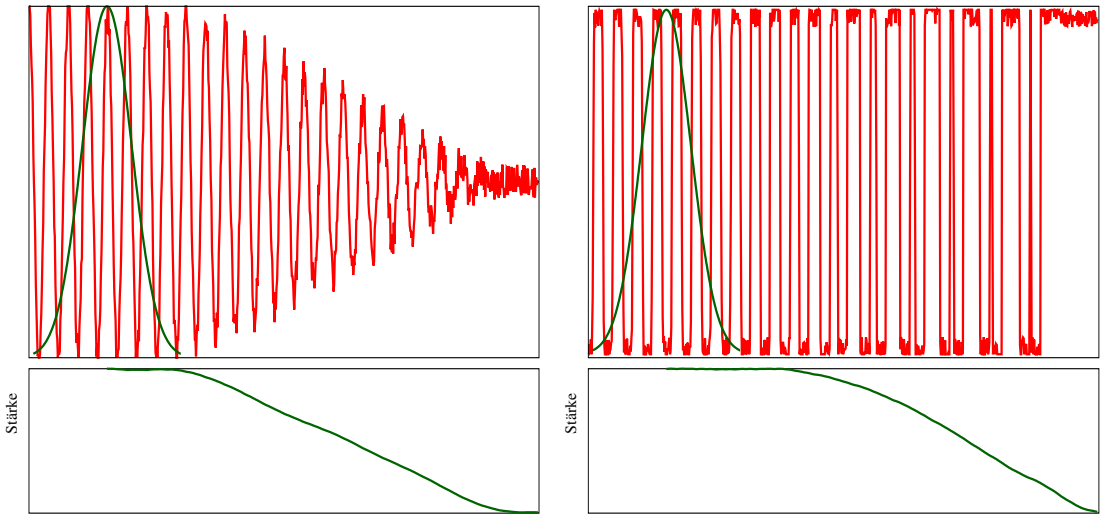
1. Die Amplitude des Signals im Ortsraum. Bezogen auf Textzeilen entspricht das dem Helligkeitsunterschied zwischen Hintergrund und Schrift. Unter Schrift ist hierbei die durchschnittliche Helligkeit einer Textzeile gemeint. Je nach Schriftsystem, Schreibinstrument und Schreibstil kann der Anteil des Hintergrundes innerhalb der Zeile erheblich sein, und so die Amplitude einer Projektion der Bildpixel verringern. In Abbildung 3.1b erreicht der Minimalwert der Summen der aufsummierten Pixelwerte noch immer über 68% des Maximums. In Abbildung 3.5a ist beispielhaft in rot ein leicht verrauschter Sinusoid mit abnehmender Amplitude dargestellt. Die grüne Gaußkurve soll dabei die Fensterfunktion darstellen. Wird dieses Fenster nach rechts verschoben, so nimmt die Stärke der entsprechenden Frequenz ab – wie im unteren Teil in grün dargestellt.

2. Das Verhältnis zwischen hohen und niedrigen Anteilen des Signals. Eine durch Textzeilen verursachte Frequenz ist umso stärker, je ähnlicher eine Projektion dieser Textzeilen (wie in Abbildung 3.1b) einer Sinusform entspricht. Ein sehr enger oder sehr weiter Zeilenabstand im Verhältnis zur Schrifthöhe verringert also die Stärke dieser Frequenz.

3. Die Regelmäßigkeit des Signals. Auch Zeilen mit unregelmäßig geschriebenen Abständen oder ungleicher Schriftgröße verringern die Stärke der korrespondierenden Frequenz. Dieser Effekt soll in Abbildung 3.5b demonstriert werden. In rot wird ein leicht verrauschtes Rechtecksignal dargestellt, dessen hohe und niedrige Anteile im linken Teil der Grafik im gleichen Verhältnis stehen. Dabei könnten die niedrigen Anteile den dunklen Bereichen einer Textzeile entsprechen, die hohen dem Weißraum zwischen zwei Zeilen. Im Verlauf des Signals bleibt die Breite des niedrigen Anteils gleich, während die Breiten zweier aufeinanderfolgender hoher Anteile zunehmend abwechselnd breiter und schmaler werden. Diese Darstellung könnte also einem stark variierendem Zeilenabstand entsprechen, bis sich schließlich zwei aufeinanderfolgende Zeilen berühren und in einer Projektion verschmelzen würden. In dieser Situation nimmt die Stärke der ursprünglichen Frequenz zunehmend ab, während die Stärke einer Frequenz entsprechend doppelter Wellenlänge deutlich zunehmen würde.

4. Die Anzahl der Perioden des Signals im gemessenen Bereich. Ein Textblock mit geringer Anzahl an Zeilen ist daher schwieriger zu bestimmen als ein größerer Textblock. Auch besonders schmale Textblöcke werden nur mit geringer Stärke erkannt.

Soweit das Verschieben eines Abtastfensters bewirkt, dass die Stärke der darin vorherrschenden Frequenz abnimmt, kann daraus nicht entschieden werden, welcher der Faktoren hierfür ausschlaggebend ist.



(a) Signal mit abnehmender Amplitude

(b) Signal bei variierendem Zeilenabstand

Abbildung 3.5.: Auswirkungen von Signaländerungen. Die unteren Bildteile zeigen die Stärke der Frequenz, die im oberen Bildteil im Bereich der Gaußkurve als stärkste erkannt wird. Je weiter dieses Fenster nach rechts bewegt wird, desto schwächer wird diese Frequenz.

3.1.2.1. Bestimmung der optimalen Fenstergröße

Wie bereits in Abbildung 3.2 erkennbar ist, werden lokale Maxima im Frequenzspektrum nicht alleine durch die Zeilenstruktur beeinflusst. Welchen Einfluss die Größe des Fensters auf die Stärke der darin erkannten Frequenzen hat, soll anhand eines griechischen Manuskripts demonstriert werden. Neben einer einigermaßen regelmäßigen Zeilenstruktur enthält es einen Stempelaufdruck, der selbst zu lokalen Maxima im Spektrum führt. Vergleichbare Auswirkungen können durch Spuren von Wurmfraß, Risse, Schattenwürfe, Flecken oder ähnlichen Störungen erzeugt werden.

In Abbildung 3.7 werden die Frequenzspektren von Ausschnitten des Manuskripts aus Abbildung 3.6 an der markierten Stelle in den jeweils angegebenen Größen gezeigt. Auch diese Spektren stellen jeweils nur Ausschnitte des Frequenzbereichs bis hin zur Wellenzahl 16 dar.

Abgesehen von starken Antworten in der Nähe des Koordinatenursprungs lassen die durch $\sigma = 15$ und $\sigma = 25$ bestimmten Ausschnitte ein starkes Ansprechen auf die Artefakte des Stempels erkennen.

Bereits bei einem Ausschnitt der Größe von $\sigma = 40$ Pixel wird der Einfluß des Stempels stark vermindert. An den Koordinaten (0, 5) sowie (0, -5) bilden sich lokale Maxima, die durch die Zeilenstruktur entstanden sind. Aus Gleichung 3.1.2 ergibt sich ein Zeilenabstand von $\frac{40 \cdot 6}{5} = 48$ Pixel, der „Bilddurchmesser“ entspricht hier dem Fensterdurchmesser.

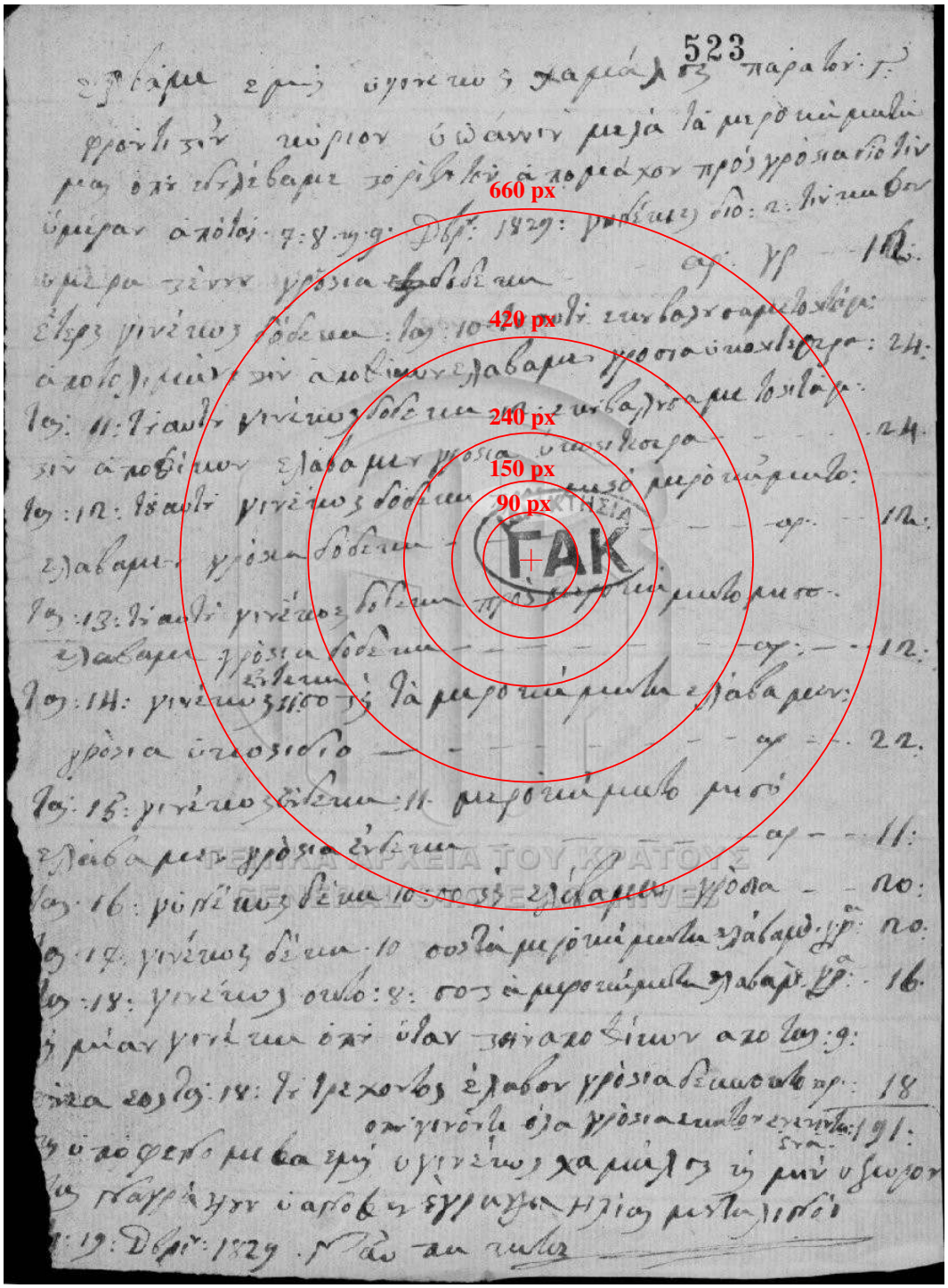
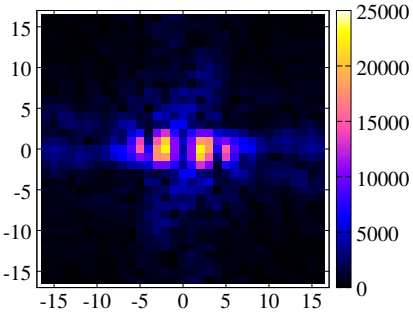
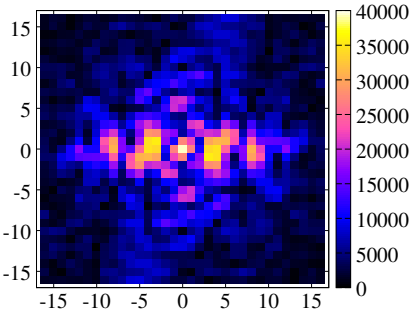


Abbildung 3.6.: Manuskriptseite aus dem Kapodistrias Governing Period Archive, © Greek State Archive GRGSA-CSA_PAK007.00.F000006, S.1032. Die Durchmesser der Abtastfenster entsprechen jeweils 6 σ.

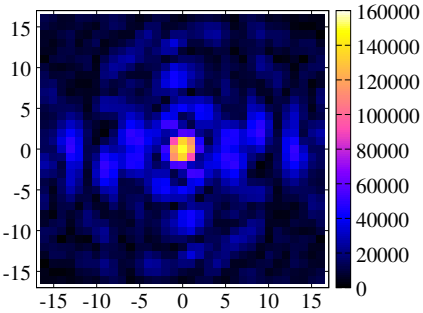
$\sigma = 15$ px



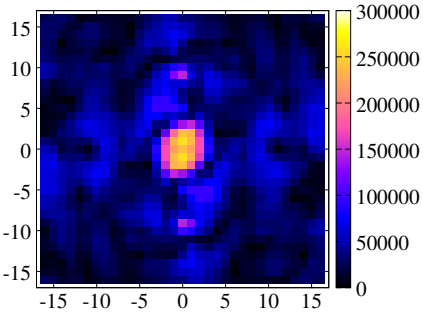
$\sigma = 25$ px



$\sigma = 40$ px



$\sigma = 70$ px



$\sigma = 110$ px

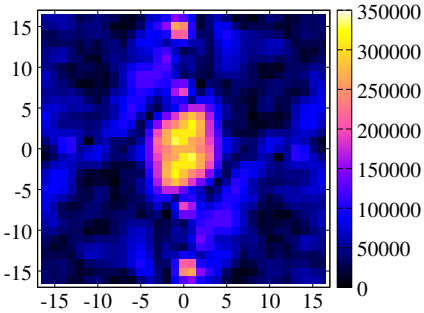


Abbildung 3.7.: Gabor-Transformation der markierten Bereiche in Abbildung 3.6. Rechts jeweils ein Ausschnitt aus dem Spektrum der niedrigsten Frequenzen der auf das Manuskript angewendeten Fensterfunktion (links dargestellt).

Erst bei einer Größe von $\sigma = 70$ Pixel werden die beiden stärksten lokalen Maxima mit einer Wellenzahl von mehr als drei durch die Zeilenstruktur angeregt. Aus den Koordinaten (0, 9) errechnet sich ein Zeilenabstand von 46,67 Pixel.

Durch ein nochmals größeres Sigma von 110 Pixeln sind die durch die Zeilenstruktur verursachten lokalen Maxima nochmals ausgeprägter; die Position von (-1, 15) bedeutet neben einem rechnerischen Zeilenabstand von 43,9 Pixel eine durchschnittliche Neigung der Zeilen um 3,8°.

Durch händisches Vermessen wurde ein durchschnittlicher Zeilenabstand von etwa 45 Pixel festgestellt.

Generell ist es wünschenswert, einen möglichst kleinen Ausschnitt zu erfassen, denn der Ortsbezug ist umso genauer, je kleiner der gewählte Ausschnitt ist. Entsprechend wird die Häufigkeit reduziert, mit der ein Ausschnitt Bereiche mehrerer unterschiedlicher Zeilenstrukturen erfasst, und dadurch kein einheitliches Ergebnis innerhalb des Ausschnitts bestimmt werden kann. Andererseits sind größere Fenster weniger anfällig für lokale Störungen und bieten eine bessere Frequenzauflösung.

Aus den Darstellungen läßt sich herleiten, dass die theoretisch kleinste Ausschnittgröße, die im Durchmesser einer Wellenlänge, d.h. einem Zeilenabstand entspricht, praktisch bei Weitem nicht ausreicht. Die bereits auf Seite 34 beschriebenen Einflussgrößen, wie das ungleiche Verhältnis von dunklen zu hellen Anteilen, dazu ein geringer Kontrast des Zeilenmusters, führen bei den beiden kleinen Ausschnitten der Abbildung 3.7 dazu, dass das deutlich stärker ausgeprägte Muster des Stempels zu einer stärkeren Antwort im Spektrum führt als das Muster der Zeilenfrequenz.

Erst bei einer Ausschnittgröße von $\sigma = 40$, d.h. 240 Pixel, und damit etwa 5,3 Zeilen, bildet die Zeilenstruktur ein lokales Maximum im Spektrum. Die noch bei dem Ausschnitt von $\sigma = 25$ deutlich erkennbaren durch den Stempel verursachten Maxima verlieren dagegen deutlich an Gewicht. Ursächlich dafür sind die deutlich geringere Anzahl an Perioden des Stempelsignals im Vergleich zur Anzahl an Zeilen. Somit ist die Anzahl an Perioden das einzige Kriterium für eine geeignete Fenstergröße, sofern eine hinreichend große Anzahl an Perioden zur Verfügung steht.

Ignoriert man den Bereich mit dem Radius $r < 4$ um den Koordinatenursprung im Spektrum des Ausschnitts mit $\sigma = 70$, so entsprechen die stärksten verbleibenden Maxima denen der Zeilenfrequenz. Um gegen Störungen der dargestellten Größenordnung robust zu sein, sollte die Ausschnittsgröße also mindestens dem Achtfachen des Zeilenabstandes entsprechen. Dann wird die am stärksten ausgeprägte Antwort im Spektrum durch die Zeilenstruktur hervorgerufen, soweit der Bereich der niedrigsten Frequenzen ignoriert wird, und die Zeilenstruktur nicht deutlich unregelmäßiger oder kontrastärmer als im gegebenen Beispiel ist.

Noch stärker ausgeprägt ist das Maximum im Ausschnitt der Größe $\sigma = 110$, welcher fast 15 Zeilen umfasst. Hier wird einerseits deutlich, dass der genaue Ortsbezug bei Ausschnitten dieser Größe verloren geht, andererseits die höhere Frequenzauflösung sogar die leichte Neigung der

Zeilen widerspiegelt. Auch der rechnerische Zeilenabstand von 43,9 Pixeln kommt den tatsächlichen 45 Pixeln am nächsten.

Als Ergebnis der Beispielbetrachtung der Abbildungen 3.6 und 3.7 kann festgehalten werden, dass der Bereich der niedrigsten Frequenzen, die einer Wellenzahl von weniger als vier entsprechen, trotz DC-Behandlung keine verwertbaren Resultate liefert. Auf eine DC-Behandlung wird daher in den folgenden Betrachtungen verzichtet und dieser Bereich ignoriert.

Bei Wahl einer Ausschnittsgröße, die mindestens etwa dem Achtfachen des Zeilenabstandes entspricht, wird davon ausgegangen, dass die stärkste im Frequenzspektrum auftretende Frequenz durch die Zeilenstruktur verursacht wird. Größere Ausschnitte bieten zwar eine höhere Frequenzauflösung, die aber gleichzeitig eine Verminderung der Ortsauflösung bedeutet. Zudem bieten größere Fenster nur bei ausreichender Anzahl an vorhandenen homogenen Zeilen einen unbestreitbaren Vorteil. Somit wird die Betrachtung zunächst auf das Frequenzband beschränkt, welches die Wellenzahlen von 8 bis $8 \times \sqrt{2}$ umfasst, welches im Nachfolgenden als *ideales* Frequenzband bezeichnet wird. Der Bereich im Radius von 4 bis 8 wird dagegen als *unteres*, der Bereich von 12.7 bis 19 als *oberes* Frequenzband definiert. Eine nähere Beschreibung der Frequenzbänder wird in Abschnitt 3.3 gegeben.

3.1.2.2. Beschränkung auf relevante Frequenzen

Für jeden Ausschnitt gilt: Die Wellenlänge der niedrigsten Frequenz entspricht dem Durchmesser des jeweiligen Ausschnittes, die Wellenlänge der höchsten entspricht zwei Pixel. Da hier aber nur niedrige Frequenzen, die der Zeilenstruktur entsprechen, betrachtet werden sollen, kann auf die Berechnung aller deutlich höherer Frequenzen verzichtet werden. Das kann dadurch erreicht werden, indem die Auflösung des zu untersuchenden Bildes reduziert wird.

Abbildung 3.8 veranschaulicht diese Idee anhand des Ausschnitts mit $\sigma = 70$px aus Abbildung 3.7. In der obersten Zeile befindet sich in der mittleren Spalte das komplette berechnete Spektrum bis hin zur Wellenzahl 210. Links ist der Auschnitt in Original-Auflösung abgebildet, rechts der Ausschnitt aus dem Spektrum bis hin zur Wellenzahl 16. Zur besseren Sichtbarkeit werden im Spektrum die niedrigsten Wellenzahlen ausgeblendet.

In den folgenden Zeilen wird nicht nur die Fenstergröße reduziert, sondern gleichzeitig in gleichem Umfang die Auflösung des Originalbildes, wodurch der inhaltliche Umfang des untersuchten Ausschnittes beibehalten wird. Somit wird das Spektrum auf die maximal möglichen Wellenzahlen reduziert, die jeweils dem halben Ausschnittdurchmesser (gemessen in Bildpixeln) je Dimension entsprechen. Bezogen auf die Originalauflösung wirkt die Reduktion der Auflösung also wie ein Tiefpassfilter; bei halber Auflösung je Dimension entspricht nun ein Pixel zwei Pixeln der vorhergehenden Auflösung. Um Aliaseffekte zu unterdrücken, wird das Originalbild vor einer Reduzierung der Auflösung mit einem Gaußfilter von $\sigma = 1.2$ geglättet.

Während sich das Frequenzspektrum im Beispiel der Abbildung 3.8 bei halber und viertel Auflösung ($\sigma = 35$ bzw. 18) im relevanten Ausschnitt der rechten Spalte nicht wesentlich verändert, verschiebt sich das lokale Maximum von der Wellenzahl 9 auf 10 bei $\sigma = 9$. Bei einer weiteren

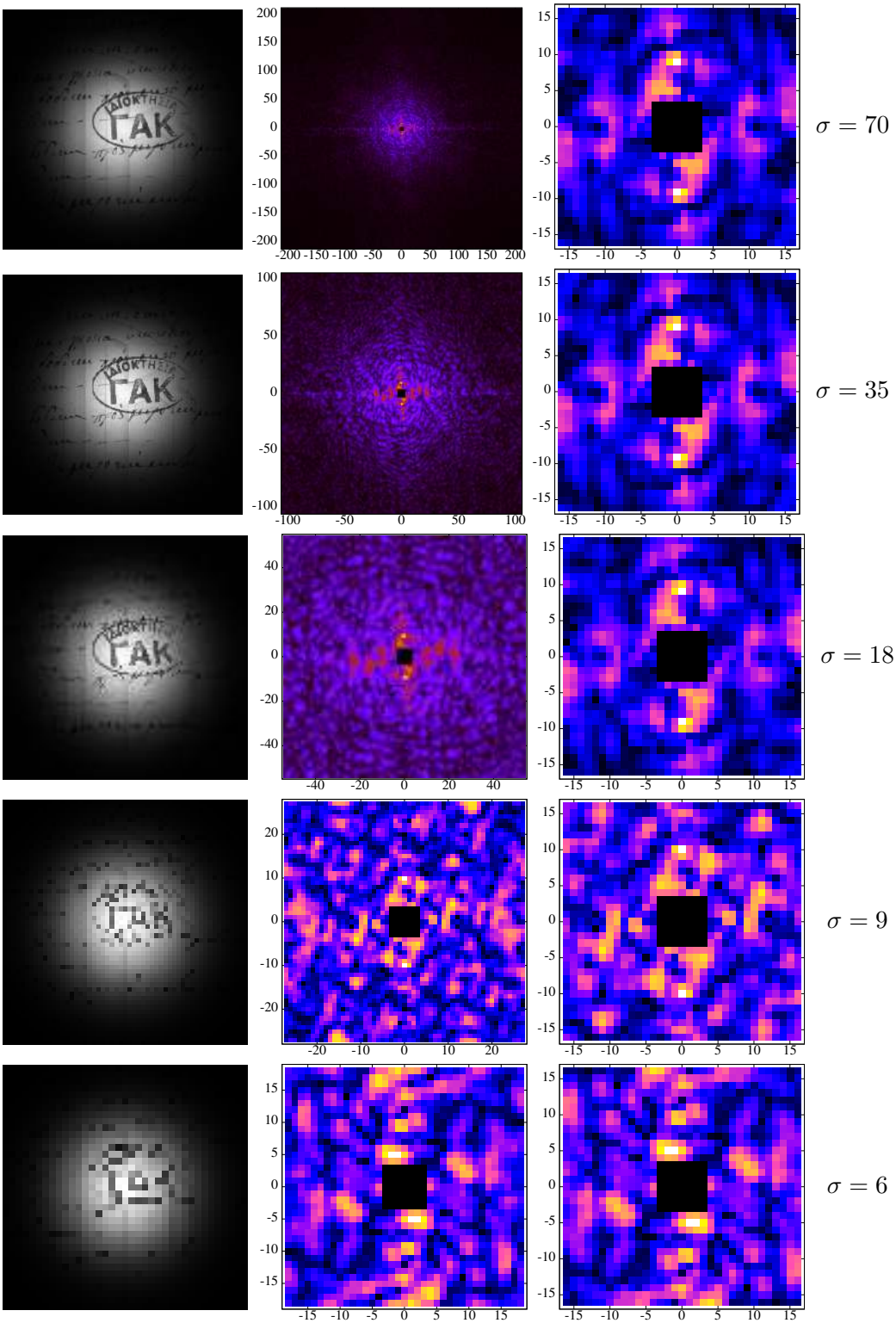


Abbildung 3.8.: Oberste Zeile: Gabor-Transformation des Ausschnitts der Größe 420×420px aus Abbildung 3.7. Darunter jeweils in reduzierter Auflösung bei Beibehaltung des Umfangs. Mittlere Spalte: Volles Spektrum, rechte Spalte: Spektrumausschnitt bis zur Wellenzahl 16.

Reduzierung auf $\sigma = 6$ bildet sich sogar ein lokales Maximum an anderer Stelle. Grund hierfür ist, dass die Projektion des Zeilenmusters keinem idealen Sinusoid entspricht. Ein σ von 18 bedeutet, dass die Frequenz der Wellenzahl 10 einer Wellenlänge von $\frac{6\times18}{10} = 10,8$ Pixeln entspricht. Sollten bei einer derartig verringerten Auflösung von diesen 10,8 Pixeln nur noch zwei Pixel der Projektion eines Zeilenmusters den Text repräsentieren, die restlichen den Zeilenzwischenraum, so wird dieses Muster bei nochmals deutlich stärkerer Verringerung der Auflösung verlorengehen.

Betrachtet man die Berechnungszeit der DFT auf einer Intel „Core i5 4670“ CPU (Singlecore) für die verschiedenen Fenstergrößen, angefangen von 318ms bei $\sigma = 70$, so ergeben sich die folgenden weiteren Werte: 99ms bei $\sigma = 35$, 27,5ms bei $\sigma = 18$, 18,5ms bei $\sigma = 9$ und 14ms bei $\sigma = 6$. Grundsätzlich sinkt also die Berechnungszeit bei kleineren Fenstergrößen beträchtlich. In der für die DFT genutzten VIGRA-Bibliothek wird dabei auf den FFTW-2 Algorithmus ([FFTW]) zurückgegriffen. Die Anwendung des FFTW-Algorithmus ([Frigo u. Johnson, 1998]) ermöglicht dabei aufgrund vielfältiger Optimierungen und der Ausnutzung spezieller CPU-Register für bestimmte Fenstergrößen eine besonders kurze Berechnungszeit.

Aufgrund dieser Betrachtung scheint $\sigma = 20$, entsprechend einem Fenster der Größe von 121×121px, eine gute Wahl bzgl. der Vermeidung unnötiger Berechnungen und genügender Genauigkeit zu sein; hier werden Frequenzen der Wellenzahl 10 durch eine Wellenlänge von 12 Pixeln repräsentiert. Neben einer erträglichen Berechnungszeit von 42,5ms bewirkt die ungerade Kantenlänge eine optimierte Erkennung von genau horizontal oder vertikal verlaufenden Zeilenmustern.

Diese Fenstergröße wird für alle folgenden Untersuchungen beibehalten und die Auflösung der Originalbilder entsprechend der erwarteten Zeilenabstände in die entsprechende Auflösung reduziert. Die universelle Anwendung auf Bilder mit unbekannten Zeilenabständen wird in Abschnitt 3.4 beschrieben.

3.2. Abgrenzung von Textblöcken zum Hintergrund

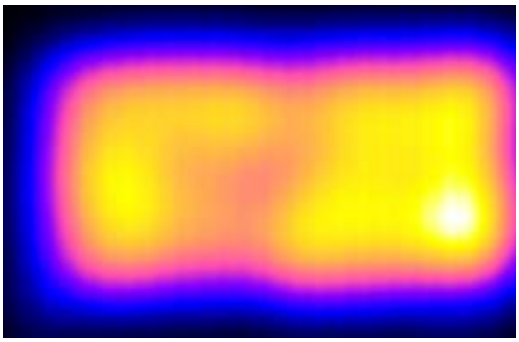
In diesem Abschnitt wird gezeigt, wie genau die äußeren Grenzen von homogenen Textblöcken erkannt werden können. Eine genaue Abgrenzung von Textblöcken ist nicht nur für die Bestimmung der Größe und Form der Textblöcke wichtig, sondern dient auch zur Trennung von eng benachbarten Textblöcken gleichartiger Zeilenstruktur. Die Abgrenzungen werden hier zunächst nur anhand von Abbildungen demonstriert, eine pixelgenaue Auswertung erfolgt im Anschluss. Zur Auftrennung von Textblöcken gleichartiger Struktur oder ihrer Abgrenzung zum Hintergrund wird lediglich die Verteilung der Stärke der erkannten Zeilenfrequenz ausgewertet. Die Bestimmung der Zeilenstruktur wird dann in Abschnitt 3.3 vorgestellt.

Abbildung 3.9a zeigt einen Ausschnitt einer japanischen Schriftrolle. Die Auflösung des verwendeten Ausschnitts beträgt 4636×3012 Pixel, der einfache Zeilenabstand 194 Pixel.

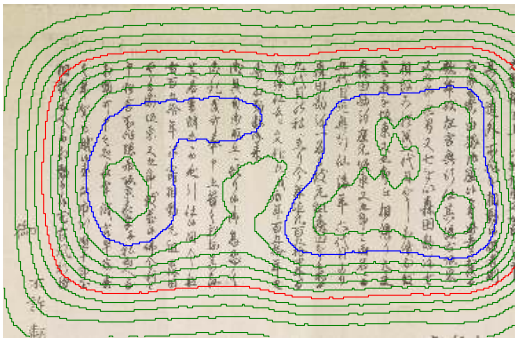
Wenn die Auflösung um das Sechzehnfache in beiden Dimensionen reduziert wird, also auf 289×188 Pixel, beträgt der Zeilenabstand etwa 12 Pixel und kann mit dem zuvor festgelegten Fenster im idealen Band gut erfasst werden. Abbildung 3.9b zeigt die Stärke des lokalen Maximums mit der Wellenzahl 10 für jedes Bildpixel, wobei das Fenster mittig auf jedem Bildpixel platziert wird. Sofern weniger als 60 Pixel Entfernung zu einem Bildrand besteht, wird der durchschnittliche Grauwert des vom Fenster erfassten Manuskriptteils berechnet, und die Fläche des Fensters außerhalb des Bildrandes mit diesem Wert aufgefüllt, um möglichst wenig Artefakte zu erzeugen, die zu ungewollten lokalen Maxima im Frequenzspektrum führen könnten. Trotz des sehr regelmäßigen Zeilenabstandes und homogener Vordergrund- sowie Hintergrundfarbe ist die Stärke der Zeilenfrequenz im rechten unteren Bildbereich am stärksten ausgeprägt, dagegen oben mittig sichtbar schwächer. Nur geringe Auswirkungen hat das Fehlen von Zeichen in nur einer Zeile. Für die Stärke der erkannten Zeilenfrequenz wird ein Maß festgelegt, welches sich an der 75% Quartile F_q aller festgestellten, sortierten Frequenzstärken orientiert, und somit gegenüber lokalen Ausreißern in der Regel unempfindlich ist. Abbildung 3.9c zeigt die Verteilung der Stärke anhand von Isolinien. Für diese Darstellung wurde das Ergebnis wieder auf die Originalgröße heraufskaliert, daher beschreiben die Linien ein gröberes Raster. Die blauen Isolinien beschreiben die Stärke der 75% Quartile F_q, alle weiteren Linien beschreiben Frequenzstärken, die jeweils 10 Prozentpunkte oberhalb oder unterhalb der 75% Quartile liegen. Diese relativ zur 75% Quartile normierten Frequenzstärken werden im folgenden ohne weitere Erwähnung verwendet, ihr relevanter Wertebereich (unterhalb der Quartile) liegt im Intervall $[1...0]$. Die rote Linie beschreibt die Frequenzstärke 0.5, also genau die Hälfte der 75% Quartile. Sie soll in der folgenden Betrachtung als vorläufige Segmentierungslinie dienen. Idealerweise sollte durch ein Fenster, welches genau zur Hälfte auf einem homogenen Textblock, zur Hälfte über einfarbigem Hintergrund liegt, eine halb so starke Zeilenfrequenz ermittelt werden, wie durch ein Fenster, welches komplett über demselben Textblock liegt. Dieser Vorstellung entspricht die vorläufige Segmentierungslinie bereits recht gut, sie kann zur Trennung von Textblock und Hintergrund dienen.

Im folgenden soll nun eine Auswertung des unteren Frequenzbandes untersucht werden, da durch die bessere Ortsauflösung des Abtastfensters bei Verringerung der Bildauflösung Vorteile in Bezug auf die Bestimmung der Textblockgrenzen erwartet werden. Zwar konnte bereits gezeigt werden, dass absolute Maxima im unteren Frequenzband unzuverlässig sein können, dennoch zeigt Abbildung 3.9d die Stärke der Zeilenfrequenz bei einer Wellenzahl von 5. Berechnet wird das, indem das ursprüngliche Manuskriptbild auf eine Auflösung von 579×376 heruntersKaliert wird, also je Dimension ein Achtel der ursprünglichen Auflösung vorweist. Auch in den Frequenzspektren, die hier durch die gleiche Fenstergröße berechnet werden, wird nach lokalen Maxima gesucht. Da das stärkste lokale Maximum durch den Einfluss von Bildstörungen geprägt sein mag, wird stattdessen nach einer Entsprechung des bereits zuvor gefundenen stärksten Maximums im idealen Frequenzband gesucht. Eine Entsprechung ist gefunden, soweit die Orientierungen der Maxima übereinstimmen und ihre Wellenzahl der Hälfte derjenigen beträgt, die bei halber Auflösung im idealen Band gefunden wurde. Grundsätzlich entspricht dieser Suchvorgang der Suche nach relevanten lokalen Maxima im oberen Band, welche im Abschnitt 3.3.3 detailliert beschrieben ist. Da sich bei dieser Berechnung nur die Bildauflösung, nicht jedoch die Fenstergröße geändert hat, ist die absolute Stärke der Frequenzen vergleichbar. So beschreibt die

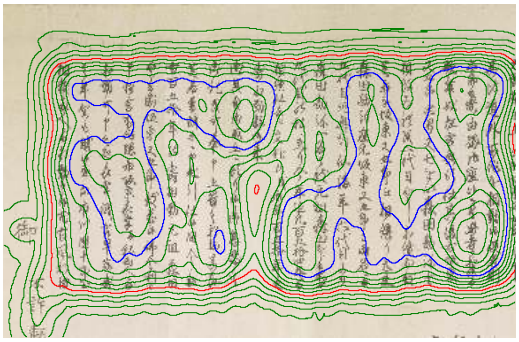
(a) Ausschnitt aus einer japanischen Manuskriptrolle



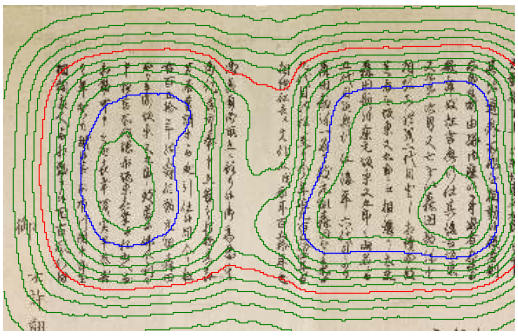
(b) Stärke der erkannten Zeilenfrequenz bei Skalierung auf etwa 10 Zeilen innerhalb 120 Pixel



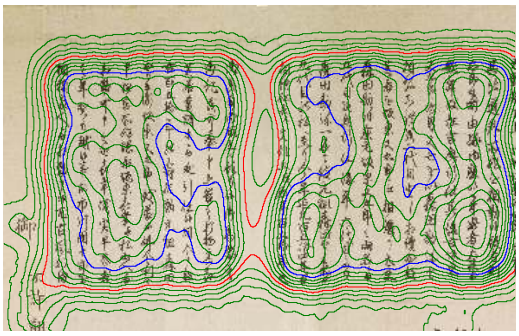
(c) Alternative Darstellung von Abbildung 3.9b durch Isolinien



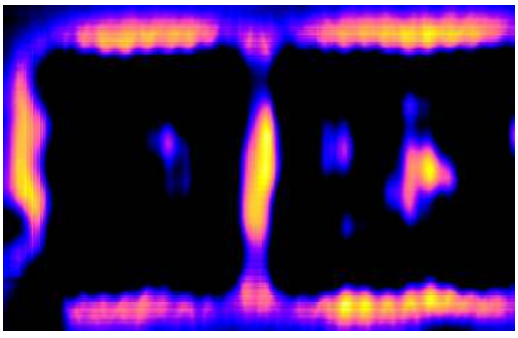
(d) Stärke der erkannten Zeilenfrequenz bei Skalierung auf etwa 5 Zeilen innerhalb 120 Pixel



(e) Zwei Textblöcke mit Abstand von 2,5 Zeilen, Abtastfenster umfasst etwa 10 Zeilen



(f) Zwei Textblöcke mit Abstand von 2,5 Zeilen, Abtastfenster umfasst etwa 5 Zeilen



(g) Differenz der Frequenzstärken aus Abbildungen 3.9e-3.9f, beschränkt auf positive Werte

(h) Abschließende Segmentierung durch Abgleich der Stärken beider Frequenzbereiche

Abbildung 3.9.: Darstellung unterschiedlicher Frequenzstärken auf einem Auschnitt einer japanischen Manuskriptrolle (© Cambridge University Library, FJ.1000.16)

blaue Isolinie in Abbildung 3.9d den absolut gleichen Wert wie in Abbildung 3.9c, gleiches gilt für die rote Linie. An dieser ist besonders die bessere Ortsauflösung erkennbar, insbesondere die Ecken des Textblocks werden in Abbildung 3.9d genauer erkannt.

3.2.1. Abgrenzung von gleichartigen, benachbarten Textblöcken

In Abbildung 3.9a ist nur ein zusammenhängender Textblock zu sehen. Dieser soll trotz des mittigen Absatzes als zusammenhängender Textblock erkannt werden. Eine Segmentierung von zwei gleichartigen Textblöcken bezüglich der Zeilenstruktur, die direkt benachbart sind, ist mit diesem Verfahren möglich, indem eine ausreichend niedrige Stärke der Zeilenfrequenz zwischen den Blöcken gemessen werden kann. Bei zunehmendem Abstand von gleichartigen Textblöcken zueinander wird eine Segmentierung einfacher. Die Abbildungen 3.9c und 3.9d im Bereich des Absatzes zeigen, dass ein zweifacher Zeilenabstand selbst bei einer Fenstergröße, die nur fünf Zeilen erfasst, nicht ausreicht, um mithilfe der vorläufigen Segmentierungslinie diesen Bereich zu trennen. Deshalb soll nun eine Segmentierung bei 2,5-fachem Zeilenabstand untersucht werden. Hierfür wird der gleiche Manuskriptausschnitt genutzt, bei dem durch Bildbearbeitung die Zeichen in der Absatzzeile gelöscht wurden, und die so entstandenen Textblöcke um einen weiteren halben Zeilenabstand, d.h. 97 Pixel – bezogen auf die Originalauflösung – auseinandergezogen wurden. Die gelöschten oder neu hinzugefügten Flächen wurden mit der Hintergrundfarbe aufgefüllt.

Abbildungen 3.9e und 3.9f zeigen, dass bei einer Fenstergröße, die sich über 10 Zeilen erstreckt, keine nennenswerte Änderung feststellbar ist. Selbst ein Fenster über etwa 5 Zeilen kann mithilfe der vorläufigen Segmentierungslinie die Textblöcke noch nicht komplett trennen. Interessant ist nun der Blick auf Abbildung 3.9g. Die farbigen Flächen stellen die Bereiche dar, in denen die Stärke der erkannten Zeilenfrequenz aus Abbildung 3.9f geringer ist als in Abbildung 3.9e. Dieser Abgleich des idealen Frequenzbandes mit dem unteren Frequenzband zeigt deutlich die Bereiche, in denen das relativ größere Fenster des idealen Frequenzbandes an den Rändern der Textblöcke noch einen größeren Bereich des Zeilenmusters erfasst, während das relativ kleinere Fenster einen deutlich kleineren Teil der Zeilen überdeckt und so die Stärke der erkannten Zeilenfrequenz deutlich geringer ausfällt. Insbesondere im Bereich zwischen den Textblöcken wird dieses deutlich.

Um eine komplette Segmentierung zu erreichen, ist die Kombination der in Abbildungen 3.9e und 3.9g dargestellten Werte naheliegend. Jedes Bildpixel wird als zum Textblock gehörend gewertet ($Tb_{x,y}$), wenn folgendes gilt:

$$Tb_{x,y} = IB_{x,y} - (6 \times (1 - IB_{x,y}) \times DB_{x,y}) > 0.45 \tag{3.2.1}$$

Alle übrigen Bildpixel werden dem Hintergrund zugeordnet.

$IB_{x,y}$ bezeichnet dabei die Stärke im idealen Frequenzband an der Bildposition (x, y), entsprechend bezeichnet der Ausdruck $DB_{x,y}$ die Differenz aus idealem und unterem jeweils normiertem Frequenzband, soweit diese Differenz positiv ist, anderenfalls wird 0 angenommen. Durch

$6 \times (1 - IB_{x,y})$ wird die Differenz gewichtet, wobei niedrige Werte des idealen Bandes zu hoher Gewichtung führen, Einbrüche der Werte des unteren Bandes dagegen gering gewichtet werden, soweit das ideale Band ein stabiles Zeilenmuster verspricht.

Abbildung 3.9h zeigt die so entstandene finale Segmentierung, wobei alle Bildpixel, die dem Hintergrund zugerechnet werden, abgedunkelt sind.

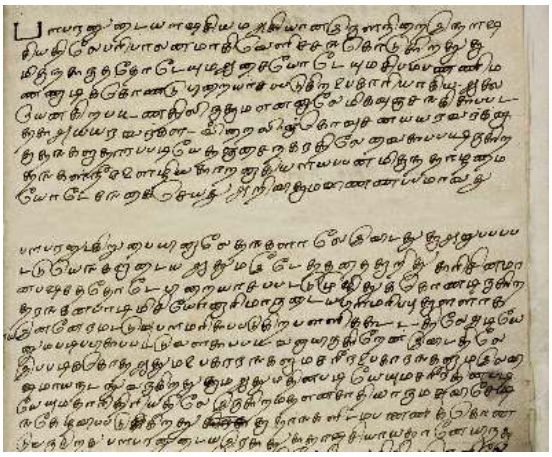
Diese Segmentierungsmethode soll nachfolgend noch an zwei weiteren Manuskripten demonstriert werden.

In Abbildung 3.10a wird ein tamilisches Manuskript gezeigt. Auch dieses wurde verändert; statt des 1,5-zeiligen Abstandes der Absätze wurde der Abstand auf etwa 2,5 Zeilen vergrößert. Im Vergleich zum japanischen Manuskriptausschnitt kann ein deutlich unregelmäßigerer Zeilenabstand festgestellt werden. Die Stärke der erkannten Zeilenfrequenz ist noch recht stabil, soweit das Manuskript auf eine Größe herunterskaliert wird, in der etwa zehn Zeilen in das Abtastfenster passen. Dagegen variiert die erkannte Zeilenfrequenzstärke beträchtlich, soweit nur etwa fünf Zeilen im Abtastfenster liegen. Dieses wird durch die unregelmäßige Schriftgröße mit unregelmäßigem Zeilenabstand und variierende Kontrastverhältnisse verursacht. Trotzdem können die beiden Textblöcke erfolgreich segmentiert werden.

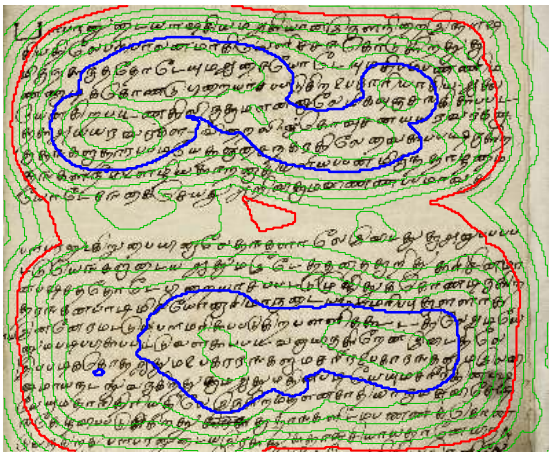
Eine noch stärkere Unregelmäßigkeit von Schriftgröße, Zeilenabstand und Schriftfarbe zeigt Abbildung 3.11. Das tibetische Manuskript weist im Original einen großen Textblock auf. Da das hier vorgestellte Verfahren nicht nur Textblöcke von rechteckiger Form aufteilen kann, wurde der Textblock mit einer gebogenen Linie künstlich auf zwei Blöcke aufgeteilt. Der Abstand entspricht auch hier weitgehend 2,5 Textzeilen. Als zusätzliche Schwierigkeit wurden Störungen in die erzeugte Lücke hinzugefügt, diese Striche weisen eine ähnliche Helligkeit und Farbe wie der übrige Text auf. Bemerkenswert ist nicht nur die Stärke der jeweils erkannten Zeilenfrequenz in Abbildungen 3.11c und 3.11d im Bereich der Lücke, sondern auch die sehr starken Variationen über die komplette Ausdehnung der Textblöcke, die durch die Unregelmäßigkeiten verursacht werden. Bereits mit einem Fenster einer Größe von etwa 5 Textzeilen könnten die Textblöcke trotz der künstlichen Störungen segmentiert werden, das vorgestellte Segmentierungsverfahren liefert allerdings im Bereich der Störungen ein noch besseres Ergebnis. Die Fläche der segmentierten Textblöcke in Abbildung 3.11e weist im unteren Bereich kleine Löcher auf, die sich entweder einfach durch einen morphologischen Operator (Closing) schließen lassen, oder in einer späteren Auswertung verschmelzen lassen (siehe Abschnitt 3.4.1.2).

Die Störungen sind zwar zufällig erzeugt, jedoch wurde darauf geachtet, dass sie kein eigenständiges periodisches Muster darstellen. Auch eine Störung, die ähnlich zum Zeilenmuster ausfällt, kann zu einer fehlerhaften Segmentierung führen. Ein waagerechter Strich in der Dicke einer Textzeile und in Verlängerung von dieser könnte bei diesem bereits sehr unregelmäßigen Manuskript schon die erfolgreiche Segmentierung verhindern.

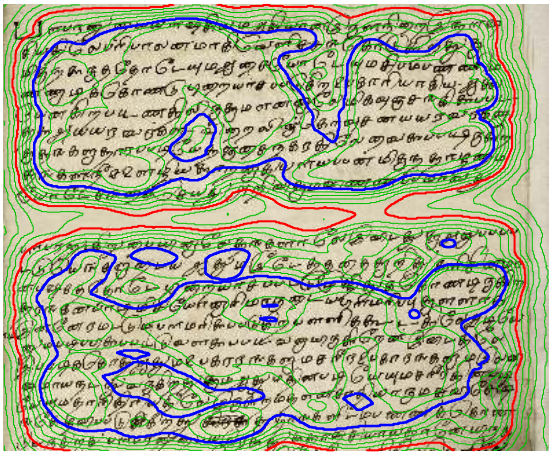
Aus diesen Untersuchungen folgt die Erkenntnis, dass sich die Ausdehnung von Textblöcken aus der Stärke der gemessenen Zeilenfrequenz bestimmen lässt, soweit alle untersuchten Textblöcke im Hinblick auf die Faktoren von Seite 34 eine vergleichbare Frequenzstärke generieren. Eine



(a) Ausschnitt aus einer tamilischen Manuskriptseite



(b) Stärke der erkannten Zeilenfrequenz bei Skalierung auf etwa 10 Zeilen innerhalb des Abtastfensters



(c) Stärke der erkannten Zeilenfrequenz bei Skalierung auf etwa 5 Zeilen innerhalb des Abtastfensters

(d) Abschließende Segmentierung

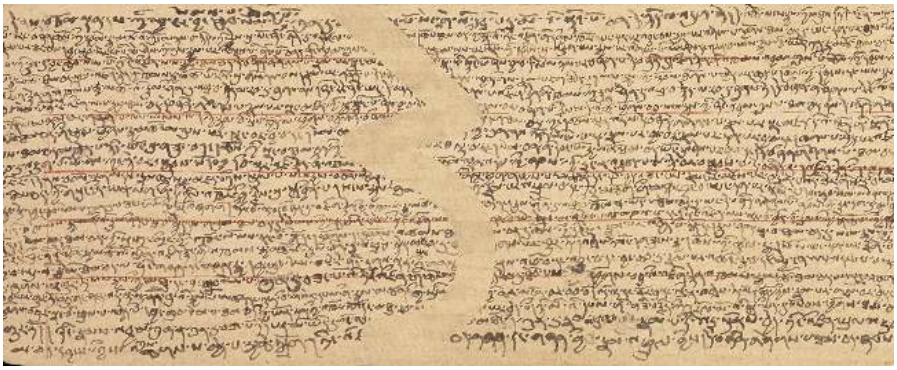
Abbildung 3.10.: Darstellung unterschiedlicher Frequenzstärken auf einem Auschnitt eines tamilischen Manuskripts (© Franckesche Stiftungen, Halle, 1_B_44_22)

präzise Auswertung mithilfe eines pixelgenauen Maßes folgt im nächsten Abschnitt. Gleichartige Textblöcke lassen sich in der Regel voneinander trennen, soweit deren Abstand zueinander mindestens 2,5 Zeilen beträgt. Eine genaue Segmentierung wird mit zunehmender Unregelmäßigkeit des Zeilenmusters schwieriger. Das Verfahren ist stabil gegen Störungen, soweit diese kein eigenständiges periodisches Muster darstellen oder das Zeilenmuster unterstützen. Sollten sich auf einer Manuskriptseite jedoch zusätzlich zu regelmäßigen und stark kontrastierten Textblöcken weitere Blöcke mit weniger als 4 Zeilen, sehr schwachem Kontrast oder sehr unregelmäßigem Zeilenabstand befinden, kann es vorkommen, dass ein Streifen vom Rand dieser Blöcke dem Hintergrund zugeordnet wird. In diesen Fällen ist es ratsam, die Konstante 0,45 in Gleichung 3.2.1 zu verringern. Dadurch werden zusätzliche Bildpixel am Rand aller Textblöcke den Textblöcken statt dem Hintergrund zugeordnet, auf Kosten einer schlechteren Aufteilung zwischen Textblöcken mit starkem Signal.

3.2.2. Pixelgenaue Untersuchung des Abtastrasters

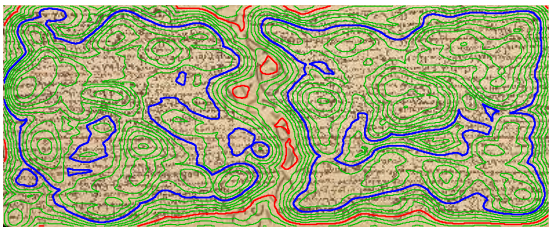
Die in den vorigen Abschnitten gezeigte Verteilung der Frequenzstärken wurde auf den Manuskriptbildern ermittelt, indem das Abtastfenster zentriert über jedem einzelnen Bildpixel positioniert und das Frequenzspektrum ausgewertet wurde. Da sich die Inhalte zweier Fenster, deren Mittelpunkte sich um lediglich einen Pixel unterscheiden, erheblich überschneiden, soll in diesem Abschnitt untersucht werden, welchen Einfluss ein gröberes Abtastraster auf das Segmentierungsergebnis hat. Hierzu soll exemplarisch das bereits in Abbildung 3.11a gezeigte Manuskript dienen, in der Variante ohne Störungen. Das Manuskript weist bei einer Höhe von 624 Pixel Zeilenabstände zwischen etwa 20 und 35 Pixel auf. Für Untersuchungen im idealen und unteren Band wird der Manuskriptausschnitt um den Faktor 2 bzw. 1,4 je Dimension herunterskaliert, um im Fenster zwischen etwa 12 und 6,9 bzw. 8,5 und 4,9 Textzeilen zu erhalten. Das Segmentierungsergebnis soll nun mit einer pixelgenauen Vorgabe verglichen werden, in der jedes Bildpixel einem Textblock zugeordnet wird, welches sich mindestens im Abstand eines halben Zeilenabstandes zu einem Textpixel befindet. In diesem Fall wurden dafür 11 Pixel ab der künstlichen Trennlinie gewählt und der Ausschnitt auf den kritischen Bereich begrenzt. Alle anderen Bildpixel werden in der Ground Truth als Hintergrund gewertet (schwarz dargestellt in Abbildung 3.12a).

Die unterste Zeile in Abbildung 3.12 zeigt in weiß die falsch klassifizierten Pixel im Vergleich zur Ground Truth, bei einem Abtastraster von jeweils 1, 2, 4, 8, 16 und 32 Pixel. In Tabelle 3.1 werden die Ergebnisse zusammengefasst. Da die Abtastfenster immer gleich groß sind, hängt die Berechnungszeit lediglich von der Anzahl der Abtaststellen ab. Bei einem Rasterabstand von 32 Pixel, also etwa der Hälfte des Radius des Abtastfensters, ist die Fehlpixelrate in diesem kritischen Bereich etwa viermal so hoch, als beim Verschieben des Fensters um nur einen Pixel. Zudem werden in einigen Situationen Textblöcke nicht mehr erfolgreich voneinander segmentiert werden können. Auch bei einem Abstand von 16 Pixel ist die Fehlerrate um 57% höher als bei nur einem Pixel. Engere Abtastraster bieten einen Kompromiss aus höherer Präzision zu höherer Berechnungsgeschwindigkeit. Bemerkenswert ist der geringe Anstieg der Fehlpixelrate bei der Vergrößerung des Rastern von einem auf zwei Pixel um nur 2,2%. Dieser Umstand

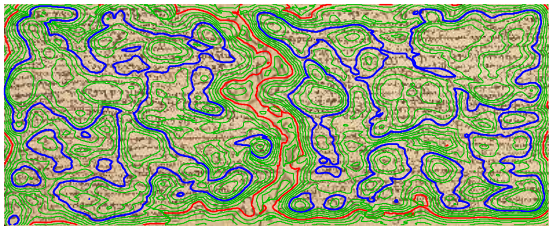


(a) Ausschnitt aus einem tibetischen Manuskript, der Textblock wurde künstlich aufgeteilt

(b) Fläche zwischen Textblöcken wurde mit Störungen verfremdet



(c) Stärke der erkannten Zeilenfrequenz bei Skalierung auf etwa 10 Zeilen innerhalb 120 Pixel



(d) Stärke der erkannten Zeilenfrequenz bei Skalierung auf etwa 5 Zeilen innerhalb 120 Pixel

(e) Abschließende Segmentierung

Abbildung 3.11.: Segmentierung zweier künstlich erzeugter Textblöcke auf einem Ausschnitt eines tibetischen Manuskripts (© The British Library Board, IOL Tib J 131)

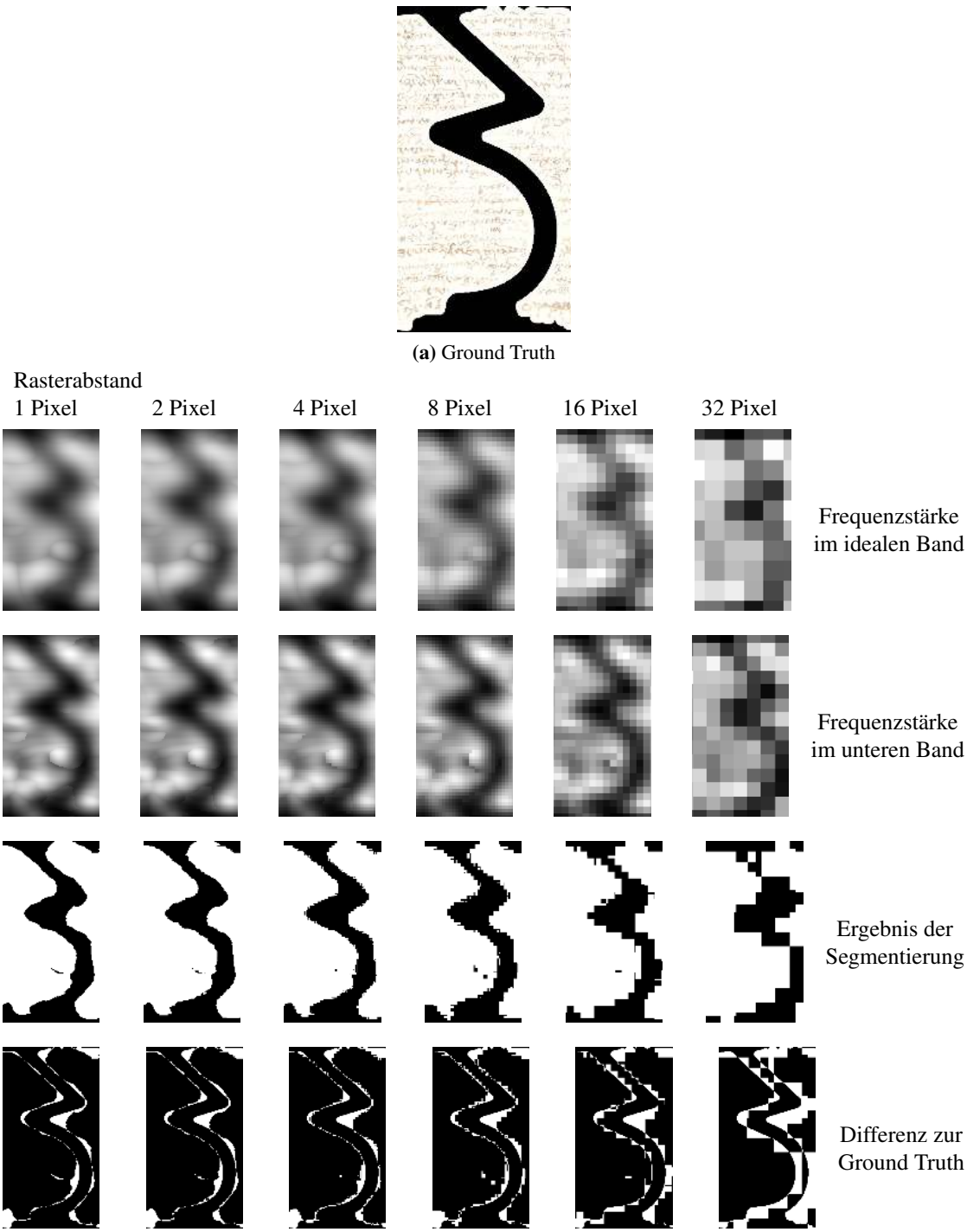


Abbildung 3.12.: Pixelgenaue Segmentierung des kritischen Bereichs aus Abbildung 3.11a

legitimiert die vorherige Reduzierung der Bildauflösung, die bereits implizit eine Vergrößerung des Abtastrasters darstellt. In den folgenden Betrachtungen werden 5 Pixel als Rasterabstand benutzt.

Tabelle 3.1.: Vergleich von Genauigkeit und Berechnungszeit bei verschiedenen Rasterabständen (CPU: Intel Core i5 4670)

Rasterabstand	**1 Pixel**		**2 Pixel**		**4 Pixel**		**8 Pixel**		**16 Pixel**		**32 Pixel**	
Frequenzband	ideales	unteres	ideales	unteres	ideales	unteres	ideales	unteres	ideales	unteres	ideales	unteres
Berechnungszeit [s]	320,15	660,29	79,52	165,67	20,21	42,07	5,10	10,62	1,32	2,69	0,39	0,74
Fehlerrate	8,52%		8,71%		9,15%		10,57%		13,38%		23,58%	

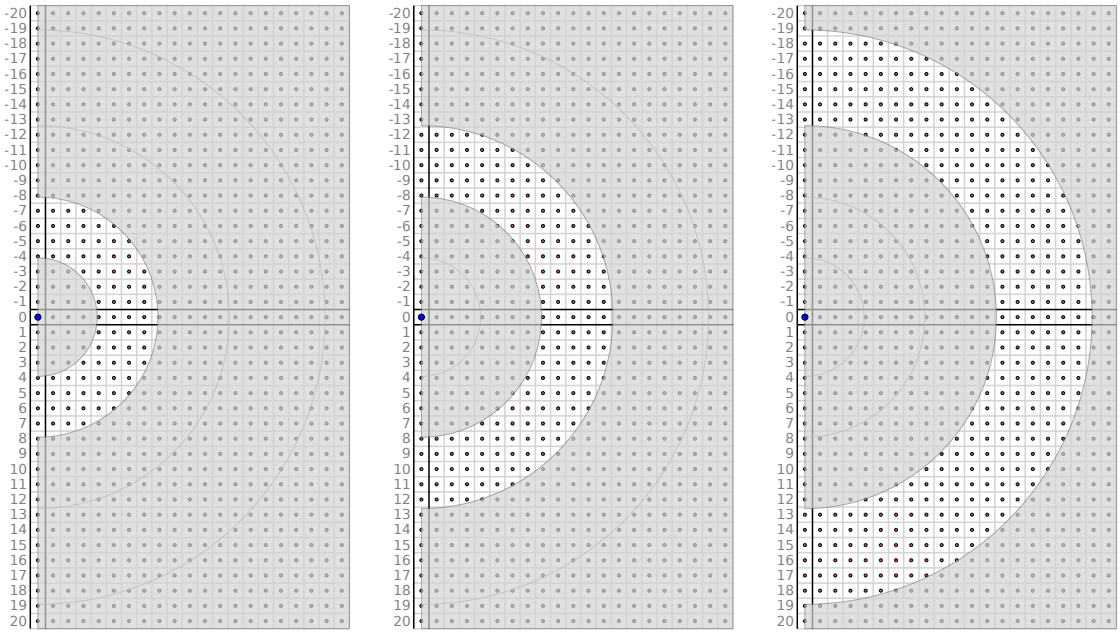
3.3. Analyse der Frequenzauflösung

In diesem Abschnitt soll zunächst ein theoretischer Blick auf mögliche Ergebnisse der Auswertung des diskreten Frequenzspektrums geleistet werden, und diese Betrachtung dann an praktischen Beispielen nachvollzogen werden. Bevor im nächsten Abschnitt eine Segmentierung auf Basis starker lokaler Unterschiede von Wellenzahl oder Orientierung vorgestellt wird, sollen mögliche Varianzen innerhalb von Textblöcken betrachtet werden.

In Abbildung 3.13 werden alle möglichen diskreten Positionen von lokalen Maxima in jedem der drei Frequenzbänder aufgezeigt. Aufgrund der Symmetrie braucht nur eine Hälfte des Spektrums ausgewertet werden, hier wird die rechte Hälfte gewählt. Aus der Position eines lokalen Maximums ergibt sich rechnerisch der Zeilenabstand und dessen Orientierung. Sollte wie in Abbildung 3.14b ein lokales Maximum an Position x,y = (5, -8) auftreten, so entspricht der gemessene Winkel α zur y-Achse von 32° in etwa der Abweichung der Zeilenorientierung von einer horizontalen Ausrichtung. Aus der Wellenzahl, die dem Radius entspricht (hier $d = 9,4$), kann der durchschnittliche Zeilenabstand berechnet werden: $\frac{121\text{px}}{9,4}$ =12,87px. Bei einer gleichmäßigen Verteilung der Textzeilen würden sich 9,4 Zeilen in dem Fenster mit dem Durchmesser von 121 Pixeln befinden. Die Genauigkeit dieser Angaben ist abhängig vom Radius; je größer der Radius, desto präziser die Berechnung:

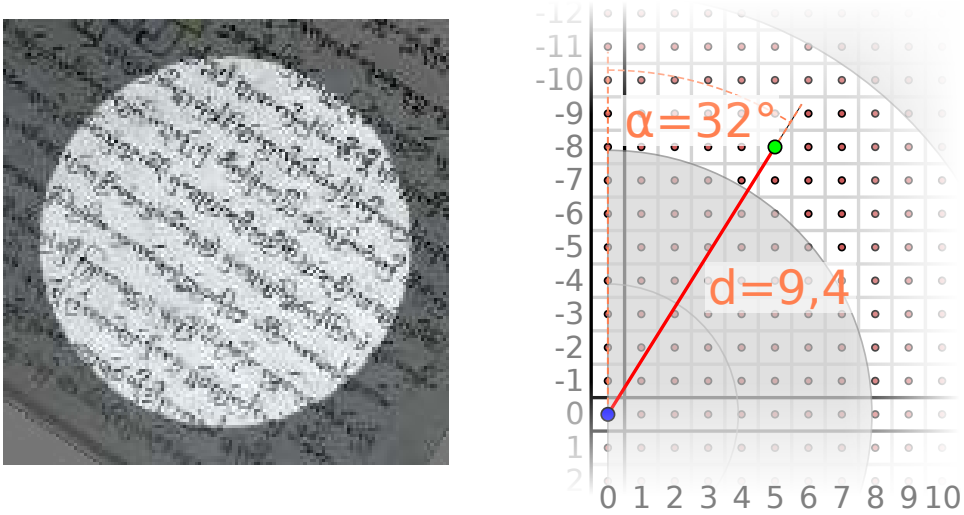
Sollten sich bei gleichmäßiger horizontaler Zeilenorientierung 4,5 Textzeilen innerhalb des Abtastfensters befinden, beträgt der durchschnittliche Zeilenabstand $\frac{121\text{px}}{4,5}$ =26,89px. Ein lokales Maximum kann aufgrund der Diskretisierung allerdings nur an den Stellen x,y = (0,4) oder (0,5) bzw (0,-4) oder (0,-5) entstehen, in der Praxis ist mit alternierendem Auftreten zu rechnen. $d = 4$ würde einem Zeilenabstand von 30,25px, $d = 5$ würde 24,2 Pixel ergeben. Der maximale Fehler liegt im unteren Band also bei 12,5%.

Anders sieht es am oberen Ende des idealen Bandes aus: Bei 11,5 Zeilen im Abtastfenster, also einem Zeilenabstand von 10,52 Pixeln, kann ein Maximum entweder im Radius $d = 11$ oder $d = 12$ liegen, was im ersten Fall einen rechnerischen Zeilenabstand von 11, im zweiten Fall



(a) Unteres Band **(b)** Ideales Band **(c)** Oberes Band

Abbildung 3.13.: Positionen möglicher lokaler Maxima in den verschiedenen Frequenzbändern



(a) Fenster mit einem Durchmesser von 121 Pixeln auf einem Manuskriptausschnitt **(b)** Auswertung eines lokalen Maximums

Abbildung 3.14.: Gabor-Transformation eines Manuskriptausschnitts (links), dabei entsteht an der grün markierten Position ein lokales Maximum (rechts).

von 10,08, und damit einen maximalen Fehler von 4,5% ergibt. An der Grenze von unterem zu idealem Band liegt der maximale Fehler bei 6,25%.

Bezüglich der Orientierung zeigt sich ein ähnliches Bild: Ein lokales Maximum an der Position $(1, -4)$ würde eine Neigung der Zeilen um 14° bedeuten. Bei einer angenommenen Neigung von 7° könnten im unteren Band also nur entweder 0° oder 14° festgestellt werden, sofern der Fensterdurchmesser etwa dem Vierfachen des Zeilenabstandes entspricht. Dadurch entsteht ein Fehler von maximal 3,9%

Im idealen Band könnte ein lokales Maximum an der Position $(1, -12)$ liegen, welches rechnerisch einer Neigung von 4,8° entspricht. Ein Maximum könnte dort auftreten, sofern der Fensterdurchmesser etwa dem Zwölffachen des Zeilenabstandes entspricht und die Neigung bei 2,4° liegt. Diese Abweichung von 2,4° entspricht einem maximalen Fehler von 1,3%. An der Grenze von idealem zu unterem Band liegt der maximale Fehler bei 2,0%.

Bei nahezu diagonal ausgerichteten Textblöcken kann die Abweichung in ungünstigen Fällen um den Faktor $\sqrt{2}$ höher liegen, aufgrund des größeren Pixelabstandes diagonal benachbarter Pixel im Raster.

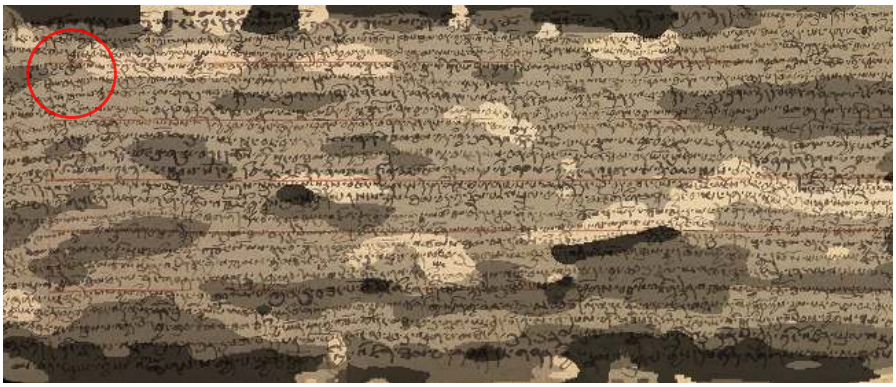
Eine nochmals höhere Frequenzauflösung als das ideale Band bietet das obere Band, welches aber nur dann sinnvoll genutzt werden kann, wenn auf einer ausreichend großen Fläche eine hohe Anzahl an Textzeilen zur Verfügung steht (optimal sind 13-18), die zueinander einen gleichartigen Abstand und durchgehend gleichartige Orientierung aufweisen sollten.

Für die Illustration der Frequenzmessungen an folgenden Beispielen soll die Wellenzahl zunächst getrennt von der Orientierung gezeigt werden.

3.3.1. Auswertung des Zeilenabstandes auf Manuskriptbeispielen

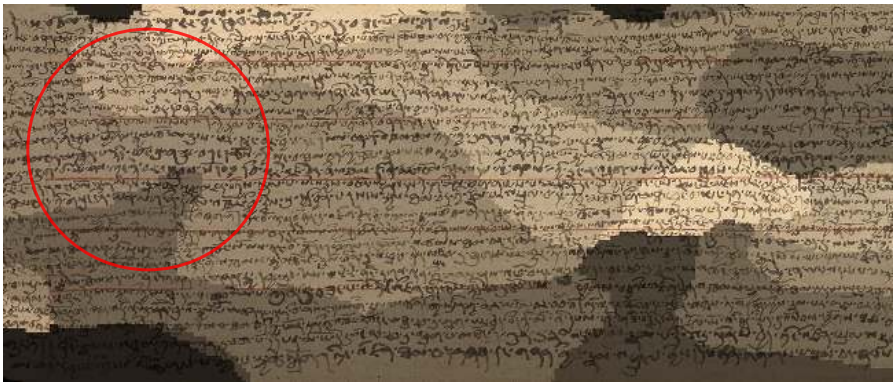
In Abbildung 3.15 wird der bereits bekannte tibetische Manuskriptausschnitt untersucht. Um den Zeilenabstand anhand der Wellenzahl in den jeweiligen Frequenzbändern bestimmen zu können, wird das Bild für das untere, ideale und obere Band um die Faktoren 1,2, 2,2 sowie 3,3 herunterskaliert. Für eine vergleichbare Darstellung des Ergebnisses wurde jeweils wieder die ursprüngliche Auflösung gewählt, in die auch das angedeutete Abtastfenster umgerechnet wurde (roter Kreis jeweils oben links auf den Abbildungen). Die an jeder Stelle ermittelten Zeilenabstände werden durch unterschiedliche Graustufen repräsentiert, die den jeweils unter den Abbildungen angegebenen Werten entsprechen.

Erkennbar ist im unteren Band (Abbildung 3.15a) eine kleinflächige Verteilung unterschiedlicher Ergebnisse. Aufgrund des relativ geringeren Durchmessers des Abtastfensters ist zwar die Lokalität der Ergebnisse präziser, der Messfehler ist aber größer. Zudem könnten einzelne Ergebnisse durch diakritische Zeichen, d.h. Vokalzeichen zwischen den Zeilen beeinflusst worden sein, und somit nicht den tatsächlichen Zeilenabstand beschreiben. Die Zuverlässigkeit der Ergebnisse nimmt zudem mit geringerer gemessener Frequenzstärke ab.



(a) Unteres Band. Rechnerisch ermittelte Zeilenabstände:
36,3px - 29,04px - 24,2px - 20,74px

(b) Ideales Band. Rechnerisch ermittelte Zeilenabstände:
33,28px - 29,58px - 26,62px - 24,2px - 22,18px



(c) Oberes Band. Rechnerisch ermittelte Zeilenabstände:
30,71px - 28,52px - 26,62px - 24,96px - 23,49px - 22,18px

Abbildung 3.15.: Rechnerisch ermittelte Zeilenabstände auf tibetischem Manuskriptausschnitt. Die Graustufen repräsentieren dabei - von dunkel (weit) nach hell (eng) - die unter den Bildern angegebenen Zeilenabstände. Das dabei genutzte Fenster entspricht der relativen Größe des jeweiligen roten Kreises in den Bildern links oben.

Im oberen Band dagegen (Abbildung 3.15c) ist die Auflösung möglicher Zeilenabstände größer, die großflächigere Aufteilung der Ergebnisse lässt auf weniger Fehlmessungen schließen, wobei durch das relativ große Abtastfenster ein Durchschnitt einer größeren Fläche berechnet wird. Hierdurch, und durch eine geringere Anzahl von Abtaststellen, wird eine geringere Ortsauflösung erreicht. Grundsätzlich unterscheidbar sind unterschiedliche Zeilenabstände in der Größenordnung von mindestens 7%, wodurch hier ein durchschnittlicher Fehler von lediglich 3,5% entsteht.

Das ideale Band stellt den Kompromiss zwischen unterem und oberem Band dar. Hier sind grundsätzlich Unterschiede der Zeilenabstände ab einer Differenz von etwa 11% erkennbar.

Insgesamt ist selbst im oberen Band noch eine recht grobe Abstufung der Ergebnisse festzustellen, im Vergleich zu den glatten Verläufen der gemessenen Frequenzstärken in den vorigen Abschnitten. Dieses grobe Raster stellt eine besondere Herausforderung für die Segmentierung in unterschiedliche Textblöcke dar, die im nächsten Abschnitt beschrieben werden soll, da die so bestimmten Unterschiede des Zeilenabstandes von 11% kein hinreichendes Indiz für unterschiedliche Textblöcke darstellen.

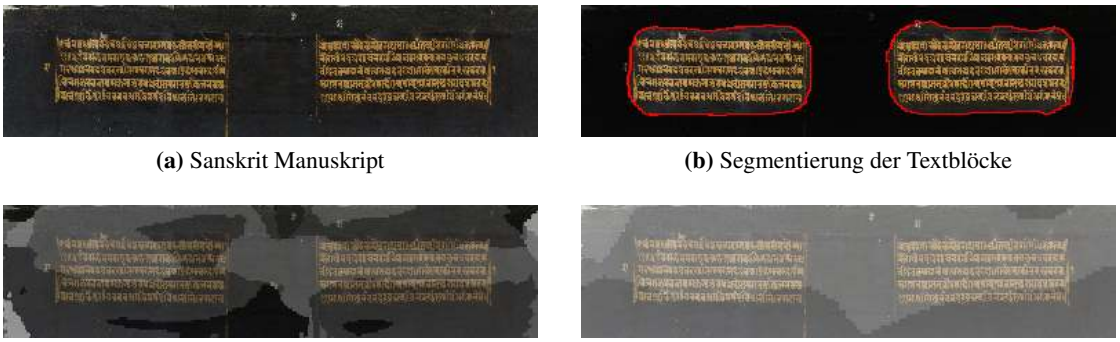
Zuvor soll jedoch ein Ergebnis auf einem sehr gleichmäßig geschriebenen Sanskrit-Manuskript (Abbildung 3.16a) vorgestellt werden. Das Manuskript umfasst 91 Blätter aus schwarzem Papier, die mit jeweils 5 Zeilen in goldfarbener Tinte bechrieben sind. Nur auf der ersten Seite befinden sich zwei Textblöcke (Folio 1v). Die Fotos der Manuskriptseiten enthalten ein Lineal, welches eine einfache Umrechnung von Bildpixeln in cm bzw. mm erlaubt. Von Manuskriptforschern wurden bereits Layoutangaben händisch erfasst, z.B. die Seitenbreite von 33,7cm. Durch Umrechnung der Bildpixel konnten 33,52cm bestimmt werden, womit diese Angabe hinreichend bestätigt wurde. Geringe Abweichungen können durch nicht exakt plan aufgelegtes Papier oder unregelmäßige Papierkanten entstehen.

Da das Verfahren lediglich regelmäßige Muster abwechselnd heller und dunkler Bildbereiche auswertet, kann auch helle Schrift auf dunklem Hintergrund ohne Änderung von Parametern auf gleiche Weise analysiert werden. Abbildung 3.16b zeigt die Segmentierung der Textblöcke. Hiermit kann grob der beschriftete Bereich bestimmt werden. Durch ein Zeilensegmentierungsverfahren im Ortsraum können diese Umrisse präzisiert werden oder die Anzahl an Zeilen ermittelt werden. Durch die von diesem Verfahren ermittelten Angaben wie

- grobe Umrisse der Textblöcke,
- Zeilenabstand sowie
- Zeilenorientierung

können Zeilensegmentierungsverfahren durch a-priori Wissen unterstützt werden.

Interessant ist auch die Bestimmung des Zeilenabstandes. Die vorhandene Beschreibung des Manuskripts enthält lediglich eine einzige Angabe von 5mm Schrifthöhe sowie 2mm Abstand zwischen Zeilen, woraus sich ein Zeilenabstand von 7mm errechnet. Durch Vermessen von Hand



(a) Sanskrit Manuskript

(b) Segmentierung der Textblöcke

(c) Grauwert entspricht Wellenzahl im idealen Band

(d) Grauwert entspricht Wellenzahl im oberen Band

Abbildung 3.16.: Sanskrit Manuskript mit goldfarbener Tinte auf schwarzem Papier aus dem späten 17. Jhd. (© Cambridge University Library, MS Add 875, fol. 1v)

wurden durchschnittlich 88 Pixel, entsprechend 7,9mm Zeilenabstand auf diesem Folium ermittelt.

Für die Bestimmung des Zeilenabstandes im idealen Band wurde das Originalbild um den Faktor 7 je Dimension verkleinert, womit – ausgehend von 88 Pixeln Zeilenabstand – 9,625 Zeilen in das Abtastfenster passen würden. Abbildung 3.16c zeigt eine sehr gleichmäßige Graustufenverteilung im Bereich der Textblöcke, wobei der hellere Grauton einer Wellenzahl von 10 entspricht (umgerechnet einem Zeilenabstand von 7,59mm), der dunklere einer Wellenzahl von 9 (umgerechnet einem Zeilenabstand von 8,45mm). Die Grautöne außerhalb der Textblockbereiche sind irrelevant. Bei genauer Betrachtung des Manuskripts fällt tatsächlich auf, dass der Zeilenabstand im oberen Teil etwas geringer ist als im unteren beider Textblöcke.

Nach Reduzierung der Bildauflösung um den Faktor 13 je Dimension wurde auch im oberen Band das stärkste lokale Maximum ausgewertet. Hier wurde im Bereich der Textblöcke im wesentlichen eine Wellenzahl von 18 ermittelt (entsprechend 7,83mm), an der unteren Kante der Textblöcke dagegen lediglich 17 (8,29mm).

Eine Wellenzahl von 18 bedeutet, dass das Abtastfenster eine Ausdehnung von etwa dem 18-fachen eines einzelnen Zeilenabstandes hat. Die Ergebnisse sind hier selbst bei nur fünf vorhandenen Zeilen stabil und präzise, da in diesem Fall das Zeilenmuster bei jeder Position des Abtastfensters das mit Abstand stärkste Muster innerhalb des Fensters bleibt. Auf der anderen Seite wird deutlich, dass ein so großes Abtastfenster nur dann sinnvolle Ergebnisse liefern kann, wenn die Textblöcke entweder ausreichend groß sind, oder, wie hier, durch breite umgebende Ränder ohne starke regelmässige Muster die vorherrschenden Muster bleiben.

Insgesamt kann festgehalten werden, dass die automatische Vermessung der Zeilenabstände mit diesem Verfahren genauer als die vorhandene Beschreibung ist. Üblicherweise wird bei händischer Vermessung auf den ersten und letzten Seiten eines umfangreichen Korpus vermessen, dazu evtl. noch mehrmals dazwischen, abhängig vom Umfang des Materials, der zur Verfügung

stehenden Zeit, des Engagements der ausführenden Person oder bei offensichtlichen, stichprobenartig festgestellten, deutlichen Abweichungen.

3.3.2. Auswertung der Zeilenorientierung eines Manuskripts

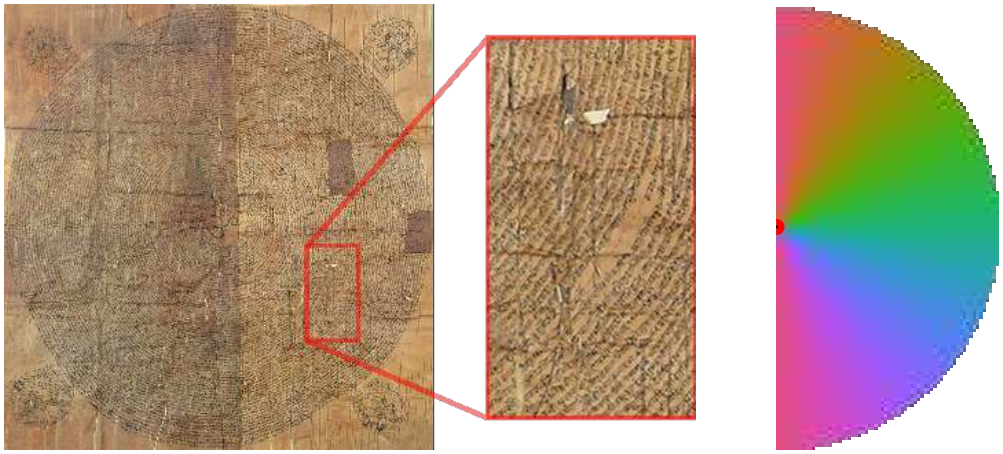
Bei der Auflösung der Orientierung kann von einer feinen Abstufung profitiert werden. Bereits das untere Band lässt dreizehn verschiedene Orientierungswinkel bei einer Wellenzahl zwischen 4 und 5 im Wertebereich von 0°–180° zu. In Abbildung 3.17 wird die Wellenzahl, Orientierung und Frequenzstärke anhand eines spiralförmig geschriebenem Sanskrit-Manuskripts vorgestellt. Die nahezu schwarze Schrift ist auf dunkelbrauner Birkenrinde aufgebracht, wodurch nur ein recht schwacher Kontrast entsteht. Die Bildauflösung des vorliegenden Manuskriptbildes von 1611×1674 Pixel ist kaum ausreichend, um die Schrift lesen zu können. Neben den Maserungen der Rinde werden Störungen auch durch Ausbesserungsstellen, Faltungsknicke, Risse und Löcher erzeugt.

Um die Messergebnisse der Zeilenorientierung nachvollziehbar darstellen zu können, wurde eine Farbkodierung gewählt, die den Wertebereich von 0°-180° als Farbübergang rot-grün-blau-rot repräsentiert (Abbildung 3.17b). Die zyklische Farbanordnung entspricht dem Umstand, dass zwischen 0° und 180° nicht unterschieden werden kann. Der untere Teil der Abbildung 3.17 beschreibt links den ausgewerteten Zeilenabstand (wiederum bedeutet dunkel einen großen, hell einen kleinen Zeilenabstand), mittig die Zeilenorientierung sowie rechts die Stärke des ausgewerteten absoluten Maximums innerhalb des jeweiligen Frequenzbandes. Im oben dargestellten unteren Frequenzband können neben der bekannten, recht groben Auflösung des Zeilenabstandes und einer bereits guten Auflösung der Zeilenorientierung noch viele Fehlstellen erkannt werden, die durch Störungen verursacht wurden. Die Störstellen nehmen über das ideale hin zum oberen Band ab, die Auflösung der Orientierung und des Zeilenabstandes nehmen dagegen zu. Dabei muss im Auge behalten werden, dass die Ergebnisse den mit einer Gaußverteilung gewichteten Durchschnitt eines jeweiligen Abtastfensters repräsentieren, welches im oberen Band einen Durchmesser von etwa 16 Zeilen umfasst.

Die Messergebnisse außerhalb des kreisförmigen Textblocks entstehen durch Auswertung von Bildrauschen oder Störungen. Sie sind vernachlässigbar, da diese Bereiche durch Segmentierung des Textbereiches vom Hintergrund anhand der Frequenzstärke dem Hintergrund zugeordnet werden.

Da in den folgenden Abschnitten die Segmentierung von Textblöcken anhand lokaler Unterschiede der Zeilenstruktur erörtert werden soll, ist die Bestimmung der Messbarkeit von geringen Unterschieden von besonderem Interesse. Folgendes Beispiel zeigt daher detailliert, wann Unterschiede der Zeilenorientierung messbar sind.

Auch hierfür wurde ein tibetisches Manuskript künstlich mittig aufgeteilt (Abbildung 3.18a). Trotz der sehr regelmäßigen Zeilenstruktur sind an einigen Stellen Abweichungen der Zeilenorientierung von bis zu 2 Grad feststellbar, die in der folgenden Auswertung zwar teilweise sichtbar werden, aber hier vernachlässigt werden. Die Fenstergröße für das ideale Band wurde



(a) Sanskrit Manuskript auf Birkenrinde

(b) Farbkodierung der Zeilenorientierung

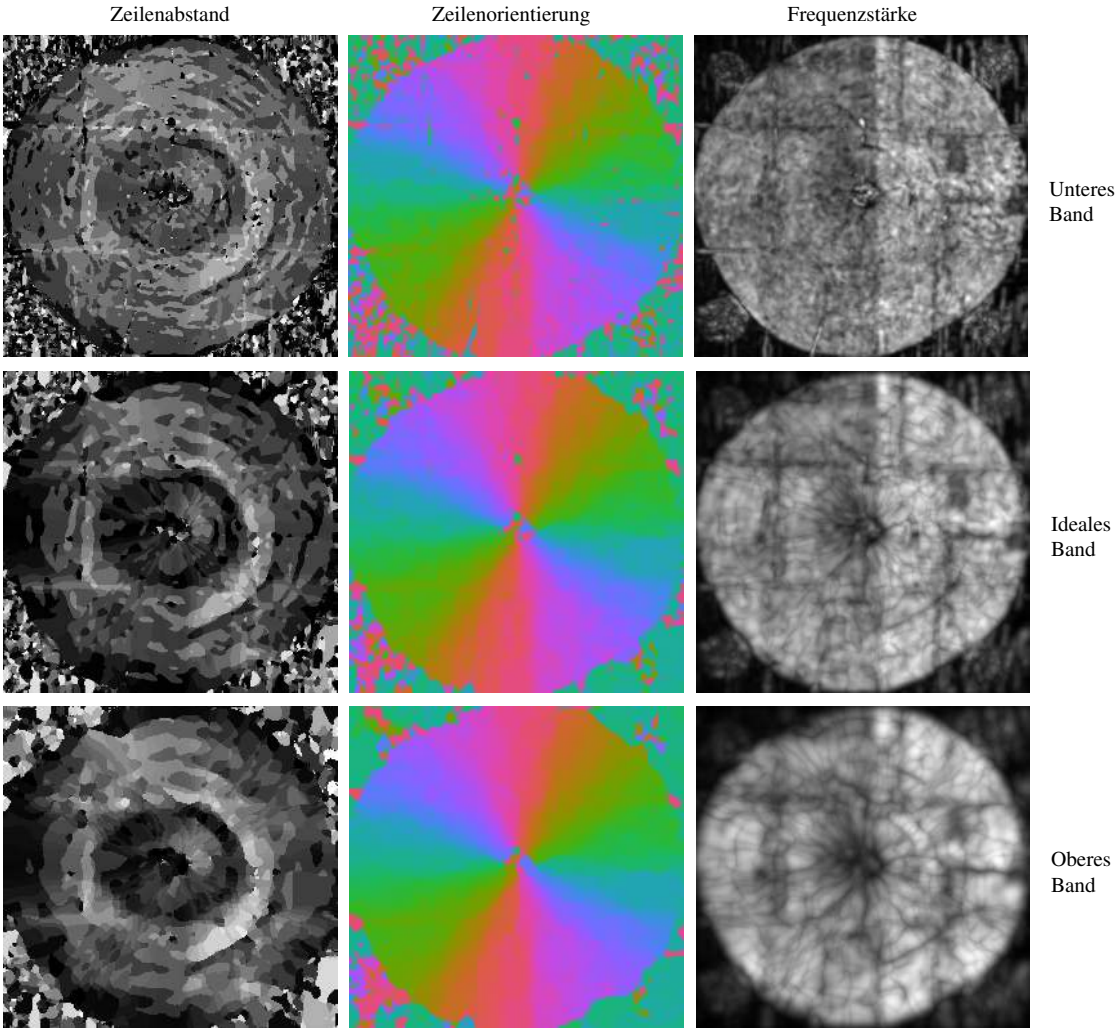
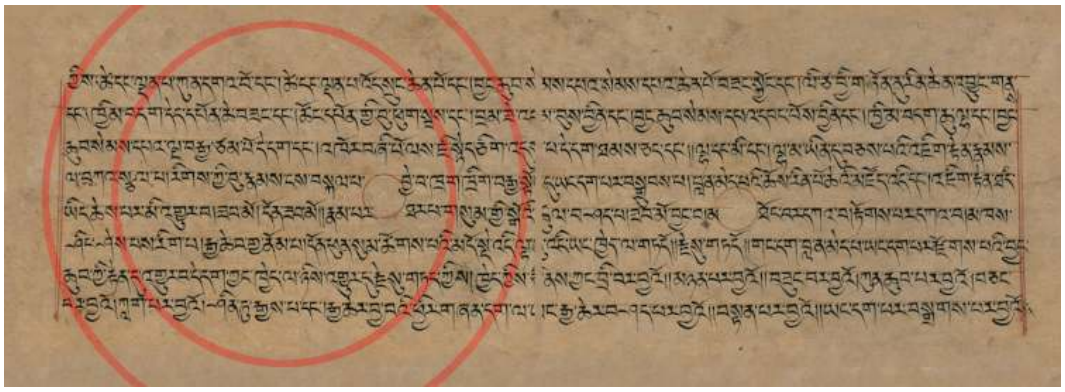
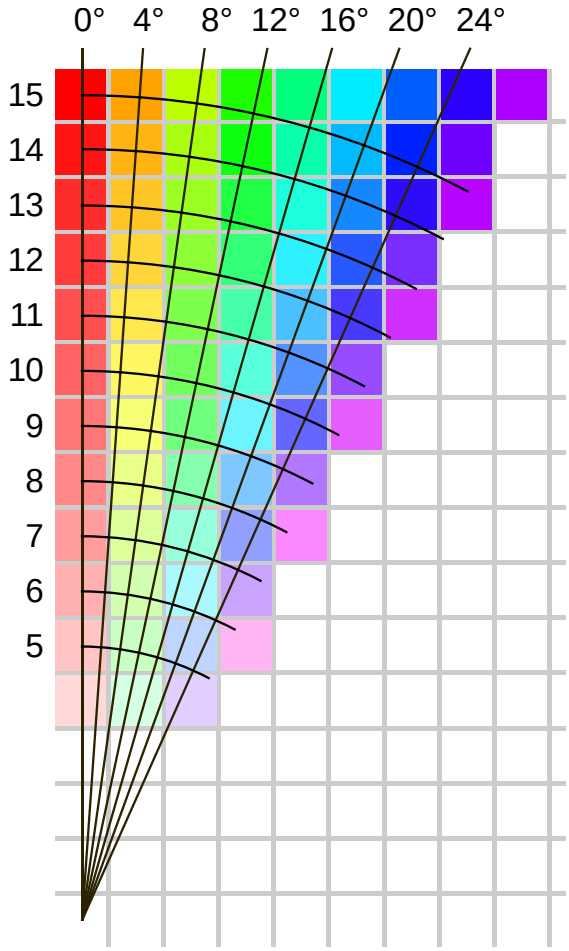


Abbildung 3.17.: Analyseergebnisse eines spiralförmig geschriebenem Sanskritmanuskripts



(a) Tibetisches Manuskript mit angedeuteter relativer Fenstergröße des idealen und oberen Frequenzbandes



(b) Farbkodierung der Orientierungen in Abbildung 3.19.
Die Kreisbögen stellen die Wellenzahl dar.

Abbildung 3.18.: Die tibetische Manuskriptseite wurde künstlich mittig geteilt (© Staatsbibliothek Berlin, MIK I 4565a [PPN 834836645], Folio4v). Die Farbkodierung stellt die ermittelten Orientierungen in Abbildung 3.19 dar, bei gedrehter rechter Manuskripthälfte.

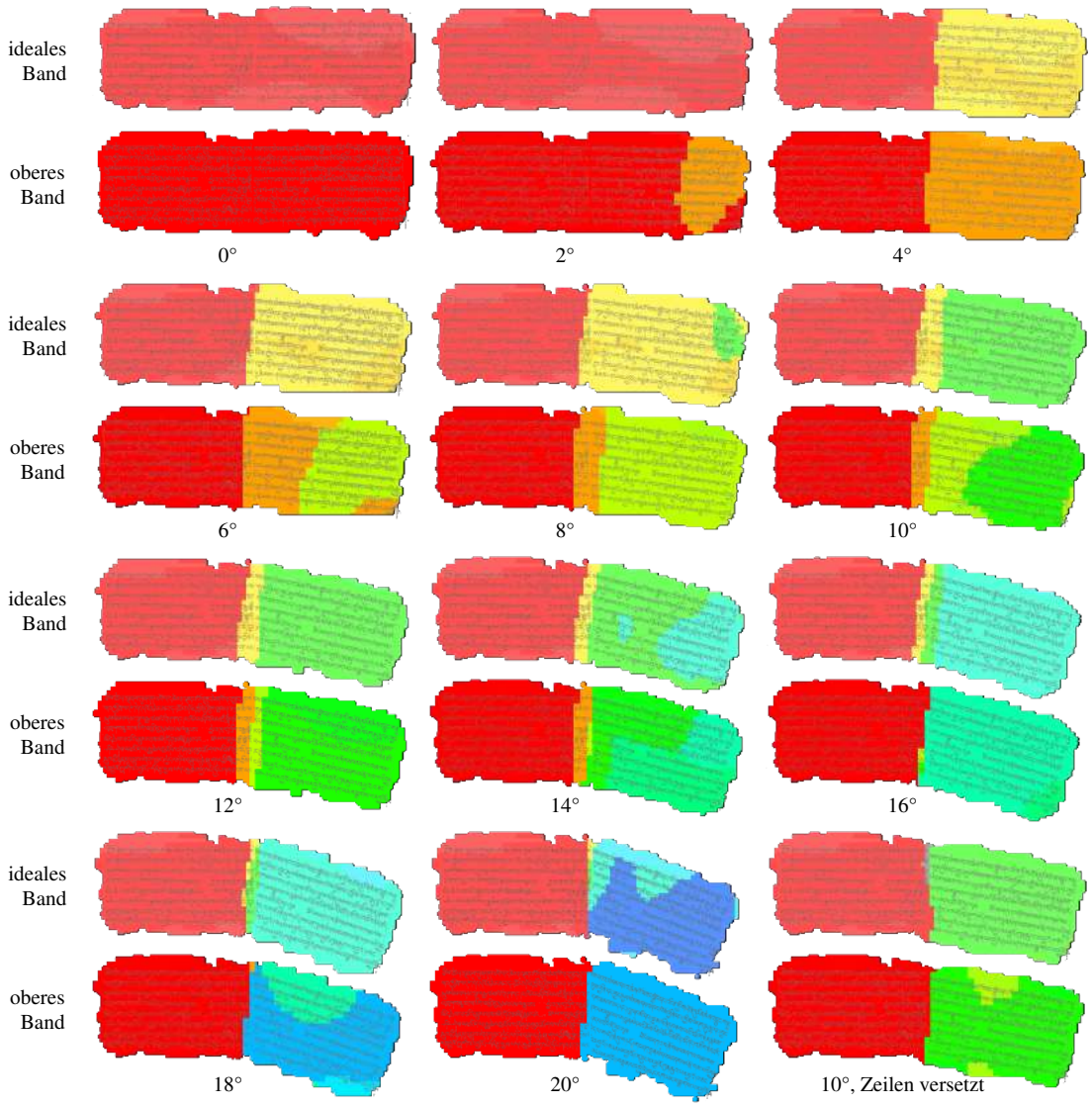


Abbildung 3.19.: Tibetisches Manuskript aus Abbildung 3.18a.
Die rechte Hälfte wurde schrittweise um 2° im Uhrzeigersinn gedreht.

so gewählt, dass es etwa 10,5 Zeilen umfasst, das Fenster des oberen Bandes dagegen knapp 15 Zeilen (Darstellung als rötliche Kreisbögen in Abbildung 3.18a). Auf eine Auswertung des unteren Bandes soll hier verzichtet werden.

Der Abstand zwischen den Texthälften entspricht einem Fünftel des Zeilenabstandes und ist somit zwar enger als häufig in den untersuchten Manuskriptseiten festgestellt werden konnte; ein typischerweise eingehaltener Mindestabstand konnte allerdings auch nicht beobachtet werden, daher wurde dieser geringe Abstand gewählt.

In Abbildung 3.19 wird nun die Auswertung der Messungen dargestellt, bei der die Orientierung der rechten Hälfte schrittweise um jeweils 2 Grad im Uhrzeigersinn gedreht wird, bis die rechte Hälfte eine Orientierung von 20° erreicht hat. Die hier dargestellten Farben entsprechen denen in Abbildung 3.18b, wobei das obere Band Farben nahe der Wellenzahl 15, das ideale Band Farben nahe der Wellenzahlen 10 oder 11 aufweisen. Dabei werden nicht nur die feineren Abstufungsmöglichkeiten des oberen Bandes deutlich, sondern auch, dass aufgrund des diskreten Frequenzspektrums ein Wechsel von einer Farbabstufung zur direkt benachbarten ohne erkennbaren Unterschied der Orientierung erfolgen kann, so z.B. bei 2° oder 6° im oberen Band. Um Textbereiche (und später auch Textblöcke) unterschiedlicher Orientierung unterscheiden zu können, sind deutlichere Unterschiede in kurzer örtlicher Distanz nötig. Da die Textzeilen der gedrehten Hälfte auf gleicher Höhe wie die der horizontal orientierten Hälfte stehen, ergibt sich die stärkste Frequenz im Grenzbereich der Textblockhälften als Interpolation beider Orientierungen, soweit ihre Orientierungen nicht bedeutend voneinander abweichen. Erkennbar ist das bei z.B 8–12° im oberen Band oder 10–16° im idealen Band. Auch hier ermöglicht das obere Band eine bessere Unterscheidung real vorkommender Orientierungsunterschiede, da größere Orientierungsunterschiede in kurzer örtlicher Distanz erkennbar sind. Der Interpolationseffekt tritt besonders bei dieser ungünstigen Zeilenstruktur auf, bei der sich die Zeilenenden beider Textblockhälften genau gegenüberstehen. Wenn die Zeilen verschiedener Textblöcke bei gleichem Zeilenabstand dagegen um eine halbe Zeile versetzt geschrieben sind, werden die Orientierungen bei 10° Drehung im oberen Band korrekt erkannt (Abbildung 3.19, unten rechts); im idealen Band gibt es nur einen schmalen Streifen, in dem eine Orientierung von $-5°$ als stärkste gemessene Frequenz erkannt wurde (grau dargestellt).

Sieht man vom Grenzbereich ab, ist im oberen Band eine Unterscheidung unterschiedlicher Zeilenorientierungen ab einem Unterschied von 6–8° möglich, im idealen Band ab etwa 10°. Dadurch ergibt sich im oberen Band ein durchschnittlicher Fehler von etwa 4°.

Tabelle 3.2.: Fehler zwischen Orientierung der Textzeilen und gemessener Orientierung

	Orientierung (rechte Hälfte)	0°	2°	4°	6°	8°	10°	12°	14°	16°	18°	20°	10° (Zeilen versetzt)
Fehler durchschnitt	ideales Band	0.00	1.01	0.82	0.29	1.40	0.98	0.68	1.63	0.77	0.94	1.16	0.71
	oberes Band	0.00	0.98	0.13	1.04	0.55	1.11	0.60	1.04	0.32	1.19	0.53	0.79
Fehler σ^2	ideales Band	0.00	2.03	1.52	0.61	3.96	2.73	2.24	5.84	3.46	4.30	5.55	1.48
	oberes Band	0.00	1.91	0.16	2.21	1.52	2.77	1.86	3.14	3.79	8.66	7.21	2.02

Die pixelgenaue Auswertung der Fehler dieser Messungen der Orientierungen beider Textblockhälften wird in Tabelle 3.2 gezeigt. Hier fällt auf, dass sich die erwähnten Vorteile des oberen Bandes nur begrenzt widerspiegeln. So kann eine Orientierung von 6° im idealen Band besser repräsentiert werden als im oberen Band. Die auffallend hohe Varianz im oberen Band bei Orientierungen von 18° und 20° entsteht hauptsächlich durch eine schlechte örtliche Auflösung im Grenzbereich: Bildpixel der horizontal orientierten Textblockhälfte werden der um 20° gedrehten Hälfte zugeordnet und umgekehrt. Dabei muss beachtet werden, dass die Auswirkungen dieser im Grenzbereich auftretenden Fehler abhängig sind vom Umfang des Grenzbereichs im Verhältnis zum Umfang des insgesamt gemessenen Bereichs.

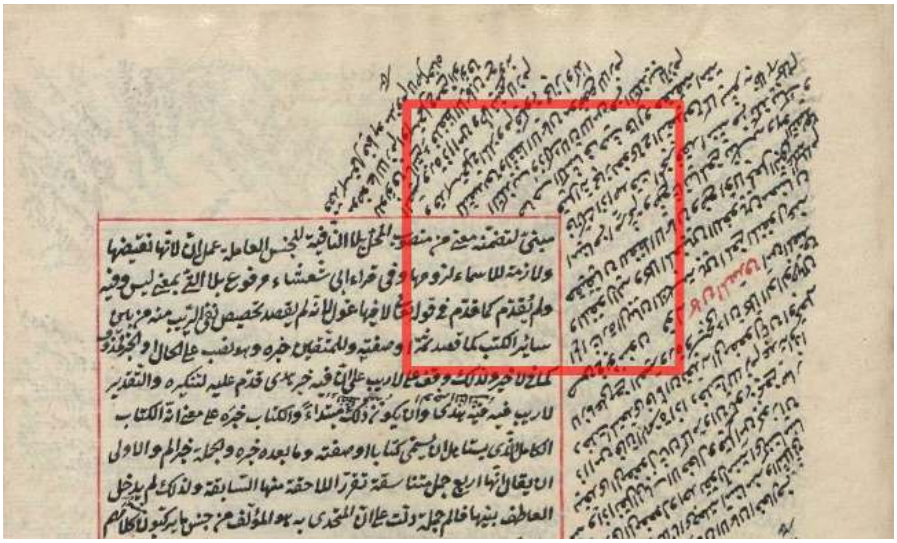
3.3.3. Kombination der Auswertung des idealen und oberen Frequenzbandes

Anhand des nächsten Beispiels soll gezeigt werden, wie sich die Messergebnisse des idealen und oberen Bandes so miteinander verbinden lassen, dass die Vorteile beider Bänder genutzt werden können.

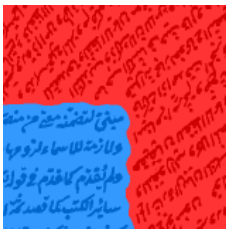
Abbildung 3.20 zeigt die Situation eines arabischen Manuskriptes. Schräg über dem horizontal orientierten Haupttextblock befindet sich ein diagonal geschriebener Randtextblock. Eine Auswertung der Situation im roten Rahmen von Abbildung 3.20a würde idealerweise für die blau eingefärbte Fläche von Abbildung 3.20b eine Zeilenstruktur entsprechend des Haupttextblocks ergeben, die rote Fläche sollte dem Randtextblock entsprechen. Die Auswertung mit einer Fenstergröße, die acht Zeilen des Haupttextblockes umfasst, wird in Abbildung 3.20c dargestellt. Für alle rot eingefärbten Bildpixel entspricht die Position der stärksten Frequenz innerhalb des idealen Bandes der Zeilenstruktur des Haupttextblocks, die der blauen Bildpixel dem Randtextblock. Aufgrund der kreisförmigen Abtastfunktion wird die nahezu rechtwinklige Ecke des Haupttextblockes leicht abgerundet.

Die Untersuchungen auf den vorangehenden Seiten konnten zeigen, dass die Frequenzauflösung im oberen Band die Zeilenstruktur detaillierter beschreiben kann, daher wurde die Bildauflösung um den Faktor $\sqrt{2}$ reduziert, und das Ergebnis ist in Abbildung 3.20d dargestellt. Auch hier stellen die roten Bildpixel das lokale Maximum im oberen Frequenzband dar, welches der Zeilenstruktur des Haupttextblocks entspricht, so wie der blaue Bereich dem Randtextblock entspricht. Erwartungsgemäß ist die örtliche Auflösung schlechter als im idealen Band: Die dunkel- bzw. hellviolett eingefärbten Bereiche in Abbildung 3.20e zeigen die Unterschiede der Ortsauflösung. Naheliegend ist die nochmals stärkere Abrundung der oberen Ecke des Haupttextblockes aufgrund des relativ größeren Fensters (hellviolett dargestellt).

Besonders auffällig ist allerdings die Verschiebung der Grenze zwischen Haupt- und Randtextblock an der oberen und rechten Kante in Richtung des Randtextblocks. Bei genauer Betrachtung fällt eine leichte Verschiebung in diese Richtung bereits im idealen Band auf (im Vergleich zur Ground Truth). Die Ursache dieser Verschiebung ist durch die unterschiedliche Stärke der auf beiden Textblöcken gemessenen Frequenzen gegeben. Diese Eigenschaft ist typisch für nahezu alle untersuchten Manuskripte. Der Haupttextblock wird meistens mit besonderer Sorgfalt



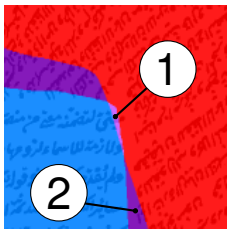
(a) Teil eines arabischen Manuskripts



(b) Ground Truth innerhalb des rot markierten Rahmens

(c) Auswertung im idealen Frequenzband

(d) Auswertung im oberen Frequenzband



(e) Örtliche Unterschiede der Frequenzbänder

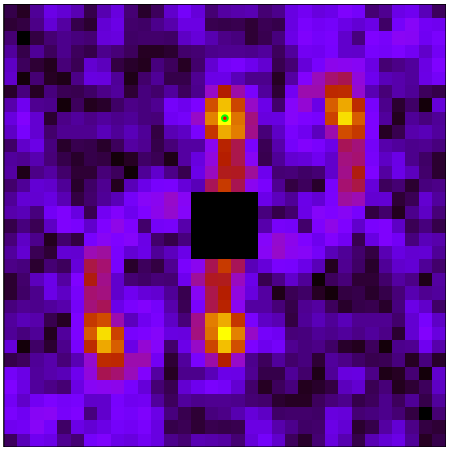
Abbildung 3.20.: Auswertung der jeweils stärksten lokalen Maxima im idealen und oberen Frequenzband bei angepassten Fenstergrößen. Hier gezeigt am Beispiel der rechten oberen Ecke eines Haupttextblocks mit direkt benachbartem diagonal geschriebenen Randtext eines arabischen Manuskripts. Erläuterungen siehe Text.

geschrieben: Die Zeilenorientierung ist meistens nahezu perfekt horizontal, die Zeilenabstände sind sehr gleichmäßig. Dagegen werden Randtextblöcke meistens zu eng geschrieben, als dass sich gleichartige Abstände zwischen hellen und dunklen Bereichen ergeben würden. Alle diese Faktoren führen zusammen zu einer schwächeren Frequenzstärke der Randtextblöcke, und da durch die Größe des Abtastfensters im Grenzbereich zweier Textblöcke auch die Strukturen des benachbarten Textblocks erfasst werden, verschiebt sich mit zunehmender (relativer) Fenstergröße die Erkennung der schwächer ausgeprägten Zeilenstruktur als lokales Maximum innerhalb eines Frequenzbandes stärker in den Bereich des jeweils weniger ideal geschriebenen Textblocks. Dieses Phänomen betrifft die dunkelviolett markierten Bereiche.

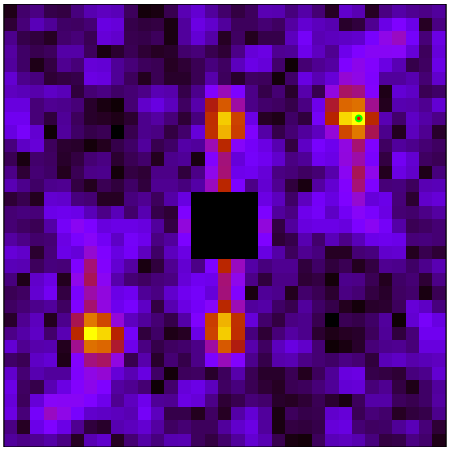
Die grundlegende Idee, diesen Nachteil der Ortsauflösung des oberen Frequenzbandes zu umgehen, und dabei dennoch von der höheren Frequenzauflösung desselben zu profitieren, liegt darin, nicht das stärkste lokale Maximum auszuwählen, sondern jeweils das lokale Maximum, welches der Struktur des stärksten lokalen Maximums des idealen Bandes entspricht (bei gleicher zentraler Position von Abtastfenstern und angepasster Bildauflösung).

Wenn man davon ausgeht, dass das jeweils stärkste Maximum eines jeden Abtastfensters im idealen Frequenzband der Zeilenstruktur entspricht, wie in Abbildung 3.20c dargestellt, so sollte sich die gleiche Zeilenstruktur durch ein lokales Maximum im oberen Band wiederfinden, wenn die Fenstergröße relativ zum Bildinhalt um den Faktor $\sqrt{2}$ vergrößert wird und sich die Zentren der Abtastfenster an gleicher Position befinden. Die Position des Maximums im oberen Frequenzband ergibt sich aus dem Vektor vom Ursprung zum Maximum im idealen Frequenzband, der um den Faktor $\sqrt{2}$ verlängert wird. Durch die Diskretisierung, und dadurch, dass das Abtastfenster mit erweitertem Bildinhalt mehr Informationen auswertet, die auch zu einer geringen Abweichung der Position des Maximums im oberen Band führen können, werden neben der sich rechnerisch ergebenen Position auch die umliegenden 24 Positionen des diskreten Spektrums als Suchraum ausgewertet. Sollte sich in diesem Suchraum kein lokales Maximum befinden, wird die Zeilenstruktur aus dem Maximum des idealen Bandes errechnet, ansonsten aus dem stärksten lokalen Maximum im Suchraum des oberen Frequenzbandes. In Abbildung 3.21 wird dieses Vorgehen verdeutlicht. Exemplarisch werden hier an den beiden in Abbildung 3.20e markierten Positionen die Entsprechungen der Maxima des idealen Bandes im oberen Frequenzband bestimmt. Da davon ausgegangen wird, dass dieses Vorgehen die Information der Zeilenstruktur in nahezu jedem Fall verbessert, wird auf diese Weise die Zeilenstruktur in der folgenden Arbeit an jeder Stelle durchgeführt, eine Fallunterscheidung ist dabei nicht notwendig. Eine Verschlechterung der Information ist nur in den Fällen zu erwarten, in dem das größere Abtastfenster zusätzliche ähnliche Zeilenstrukturen erfasst, die nicht den Zeilenstrukturen des kleineren Abtastfensters entsprechen, aber zu einer Verschiebung der Position des lokalen Maximums führen.

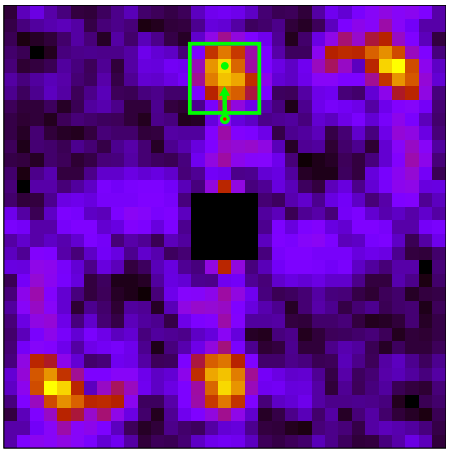
Aus der bisherigen Betrachtung kann festgehalten werden, dass für die Segmentierung zwischen Textblöcken und Hintergrund neben dem idealen Band Informationen des unteren Bandes verwendet werden sollten; eine Auswertung der Stärke der Frequenzen im oberen Band bietet hier keine Vorteile. Umgekehrt sind die ermittelten Werte bzgl. Zeilenabstand oder Orientierung im unteren Band aufgrund der geringen Frequenzauflösung schlechter geeignet als die im idealen



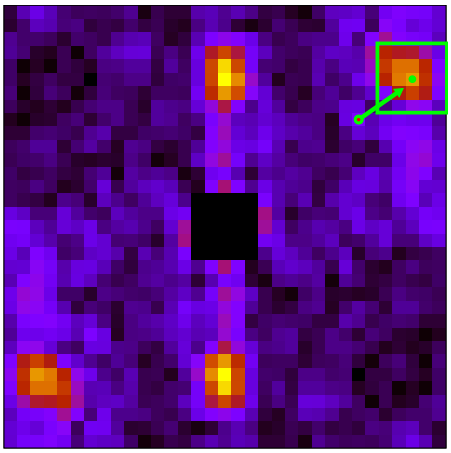
(a) Maximale Frequenzstärke im idealen Band für Situation 1 in Abbildung 3.20e (durch Punkt markiert)



(b) Maximale Frequenzstärke im idealen Band für Situation 2 in Abbildung 3.20e (durch Punkt markiert)



(c) Bestimmung des relevanten lokalen Maximums im oberen Band für Situation 1 in Abbildung 3.20e



(d) Bestimmung des relevanten lokalen Maximums im oberen Band für Situation 2 in Abbildung 3.20e

Abbildung 3.21.: Nach Bestimmung der maximalen Frequenzstärke im idealen Band kann auch bei größerem Abtastfenster das dazu korrespondierende lokale Maximum im hohen Frequenzband bestimmt werden.

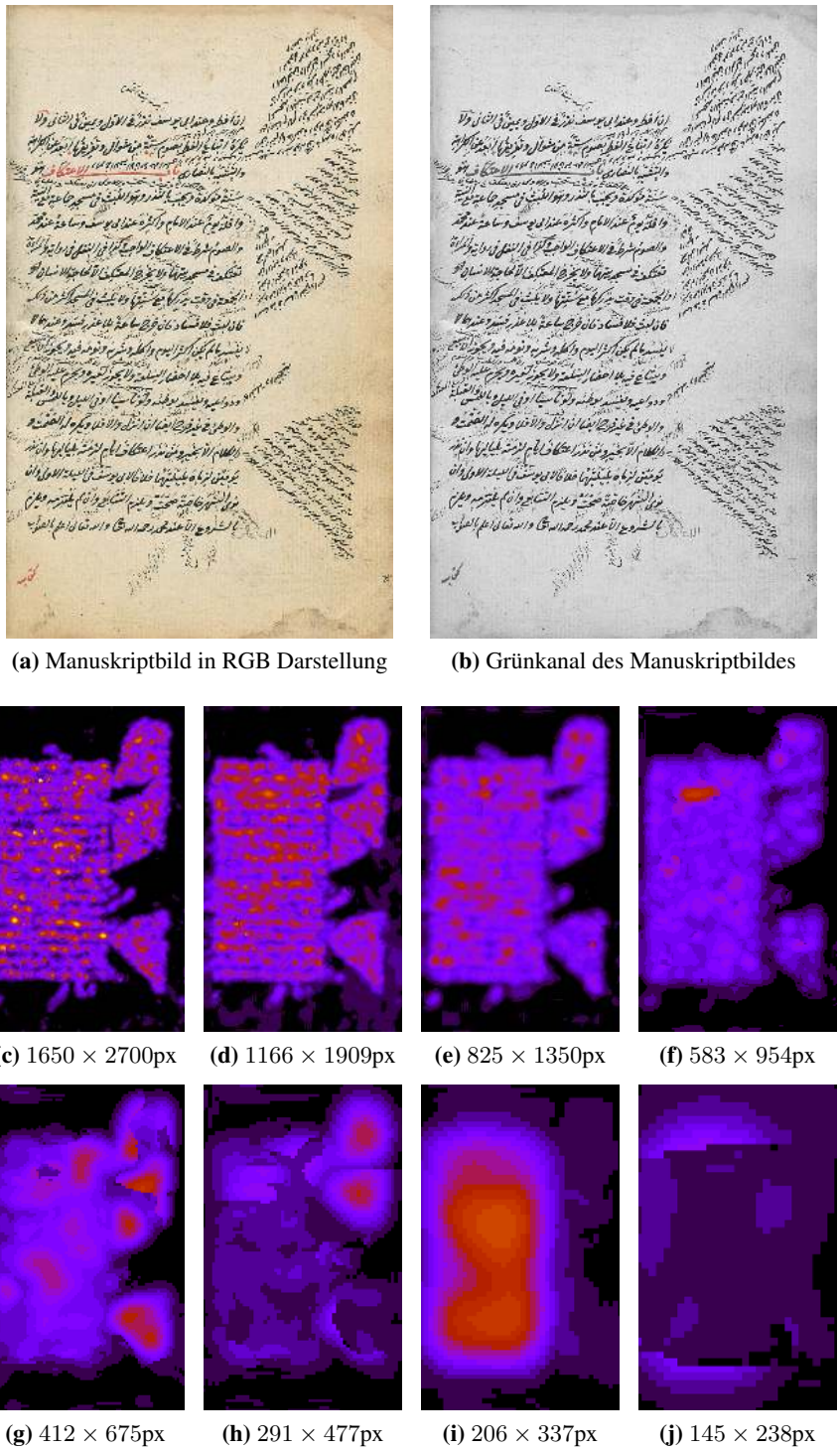
Band und werden daher nicht weiter benötigt. Zusätzlich zu den Werten des idealen Bandes sollte auf die Werte des oberen Bandes zugegriffen werden, soweit Frequenzunterschiede innerhalb der Fenster unerheblich sind. Dann bietet das obere Band aufgrund der besseren Frequenzauflösung genauere Werte.

3.4. Layoutanalyse auf Dokumenten beliebiger Auflösung

In den bisherigen Betrachtungen wurden einzelne Textbereiche mit weitgehend einheitlichem Zeilenabstand betrachtet. Um die Eigenarten der einzelnen Frequenzbänder, die zuvor vorgestellt wurden, optimal zeigen zu können, wurde die Auflösung der untersuchten Manuskriptteile so angepasst, dass die Abtastfenster den jeweiligen Textbereich optimal erfassen können. Da aber diese Anpassungen bei einer Verarbeitung großer, heterogener Datensätze nicht möglich ist, und auf einzelnen Manuskriptseiten gleichzeitig Textblöcke mit vollkommen unterschiedlichen Zeilenabständen vorhanden sein können, soll stattdessen jedes Manuskript systematisch in verschiedene Auflösungen skaliert werden. Theoretisch können digitale Bilder so stark in höhere Auflösungen skaliert werden, bis die Kapazitätsgrenzen des jeweiligen Computers erreicht werden, oder andererseits bis auf eine Größe von nur einem Pixel herabskaliert werden. Innerhalb dieser technischen Grenzen scheint stattdessen eine Eingrenzung auf realistische Maßstäbe sinnvoll, die wie folgt hergeleitet werden.

Anhand vieler Beispiele konnte festgestellt werden, dass Textblöcke aufgrund zu geringer Auflösung tendenziell unleserlich werden, sobald ihr Zeilenabstand weniger als 20 Pixel entspricht. Das Muster eines hinreichend regelmäßigen Textblocks kann dagegen auch bei Zeilenabständen von unter 10 Pixel noch als lokales Maximum im Frequenzspektrum erkennbar sein. Bei Anwendung des Abtastfensters der Größe 121×121 Pixel würde daher eine Zeilenstruktur im idealen Frequenzband ein lokales Maximum bei der Wellenzahl 12 erzeugen, die einen Zeilenabstand von lediglich 10 Pixel hat. Dieses wird als geringstmögliche sinnvolle Textgröße angesehen. Wird nun die Auflösung des Originalbildes sukzessive um den Faktor $\sqrt{2}$ reduziert, und für jede dieser Auflösungen eine Frequenzanalyse durchgeführt, sollte bei passender Auflösung rein rechnerisch jede Zeilenstruktur mit einem größeren Zeilenabstand als 10 Pixel ein lokales Maximum im idealen Frequenzband erzeugen. Die Bildauflösung wird so lange reduziert, bis das Eingabebild in der größeren Dimension eine geringere Auflösung als 121 Pixel aufweist. Das bedeutet, dass der größtmöglich erkannte Textblock einen Zeilenabstand aufweisen darf, der maximal einem Achtel der Breite bzw. Länge der jeweils größeren Dimension des Dokuments entspricht. Derart große Textblöcke im Verhältnis zur Dokumentgröße waren bei den untersuchten Manuskripten nicht vorhanden, und daher wird die obere Grenze auf die Weise festgelegt, dass die schrittweise Reduzierung der Bildauflösung in dem Moment gestoppt wird, sobald beide Kanten des untersuchten Manuskripts eine kürzere Länge als 121 Pixel haben.

Obwohl mit diesen Parametern Textblöcke auch auf untypischen Manuskriptformen erfasst werden können, stellen diese Einstellungen keine Grenzen des Möglichen dar. So ist auch eine Vergrößerung des Manuskripts vor der Verarbeitung denkbar, oder eine weitere Verkleinerung, die



(a) Manuskriptbild in RGB Darstellung **(b)** Grünkanal des Manuskriptbildes

(c) 1650 × 2700px **(d)** 1166 × 1909px **(e)** 825 × 1350px **(f)** 583 × 954px

(g) 412 × 675px **(h)** 291 × 477px **(i)** 206 × 337px **(j)** 145 × 238px

Abbildung 3.22.: Die Bildauflösung einer Manuskriptseite wird schrittweise um den Faktor $\sqrt{2}$ reduziert, womit bei gleichbleibender Fenstergröße des Abtastfensters alle untersuchten Textblockgrößen vermessen werden können. Die Abbildungen 3.22c bis 3.22j zeigen die dabei gemessenen Frequenzstärken im idealen Frequenzband. (© Staatsbibliothek zu Berlin, Hs. or. 5160, 30v)

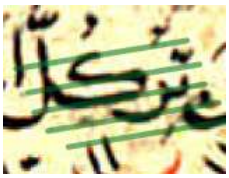


Abbildung 3.23.: Parallel ausgeführte Striche einzelner Zeichen können deutliche Frequenzstärken verursachen.

bei Sonderformaten sinnvoll sein könnte, in denen die Kantenlängen des Dokuments möglicherweise nur dem dreifachen Zeilenabstand eines einzig vorhandenen Textblocks entsprechen.

Sinnvoll kann es dagegen auch sein, diese Grenzen enger zu setzen. In einem Fall, in dem der geringstmögliche Zeilenabstand vor einer Analyse bekannt ist, und dabei deutlich größer als die hier angenommenen zehn Pixel sind, kann durch die Angabe eines Parameters die Bildauflösung bereits vor der Analyse reduziert werden, und so die Berechnung deutlich beschleunigt werden. Genauso könnte die Reduzierung der Bildauflösung früher gestoppt werden, sobald beispielsweise die kürzere Kante des Dokuments eine Länge von weniger als 121 Pixel hat, wie es z.B. bei Schriftrollen oder Palmblattformaten sinnvoll sein kann.

Werden Manuskriptbilder in der zuvor beschriebenen Weise in verschiedene Bildauflösungen transformiert, ist eine leichte Überlappung der Grenzen der Frequenzbänder notwendig. Bedingt wird das durch die Diskretisierung des Frequenzspektrums, und da Abtastfenster bei gleicher Position relativ zur Dokumentengröße bei unterschiedlicher Dokumentenauflösung nicht den identischen Inhalt umfassen. So kann es vorkommen, dass ein Abtastfenster ein lokales Maximum, welches der Zeilenstruktur an dieser Stelle entspricht, mit hoher Wellenzahl im unteren Frequenzband aufweist, bei der nächstgeringeren Auflösungsstufe befindet sich dieses Maximum aber schon im oberen Frequenzband bei hierfür geringer Wellenzahl. Um nun zu vermeiden, dass eine Zeilenstruktur innerhalb der Auflösungsfolge bei keiner Auflösung im idealen Frequenzband erscheint, werden für alle nachfolgenden Betrachtungen die Grenzen der Frequenzbänder leicht abgeändert:

- Das untere Frequenzband wird Frequenzen bis zur Wellenzahl 9 erfassen.
- Das ideale Frequenzband wird sich von Wellenzahlen 8 bis $9 \times \sqrt{2}$ ausdehnen.
- Das obere Frequenzband wird Wellenzahlen ab $8 \times \sqrt{2}$ erfassen.

Zwar kann es durch diese Überlappung nun verstärkt vorkommen, dass eine Zeilenstruktur bei benachbarten Auflösungsstufen wiederholt im idealen Frequenzband erfasst wird, dieser Umstand ist bei der weiteren Analyse jedoch unerheblich.

Die Zielsetzung der folgenden Vorgehensbeschreibung ist, die optimale Bildauflösung für die Bereiche eines jeden Dokuments zu finden, die Textblöcke darstellen, um sie durch das Abtastfenster bestimmen zu können.

In Abbildung 3.22 wird zunächst ein Dokument in Originalauflösung dargestellt. Angefangen bei 1650×2700 Pixel wird dann in sieben Schritten die Auflösung auf bis zu 145×238 reduziert. Für jede Bildauflösung wird das 121×121 Pixel große Abtastfenster in einem 5-Pixel-Raster über das Bild geführt und an jeder dieser Positionen werden Stärke und Positionen der lokalen Maxima in allen Frequenzbändern gespeichert. Für die Originalauflösung bedeutet das 330×540 Abtaststellen. Die Stärke des stärksten lokalen Maximums im idealen Frequenzband kann für jede dieser Abtaststellen wieder in Bildform repräsentiert werden, wie in Abbildung 3.22c dargestellt. Auch für die geringeren Bildauflösungen wird diese Darstellung gezeigt, die – trotz einer geringeren Anzahl an Abtaststellen – ohne Interpolation jeweils auf die gleiche Bildgröße von 330×540 Pixel skaliert wird. Dadurch werden die Frequenzstärken an jeder Bildposition für jede Bildauflösung direkt vergleichbar.

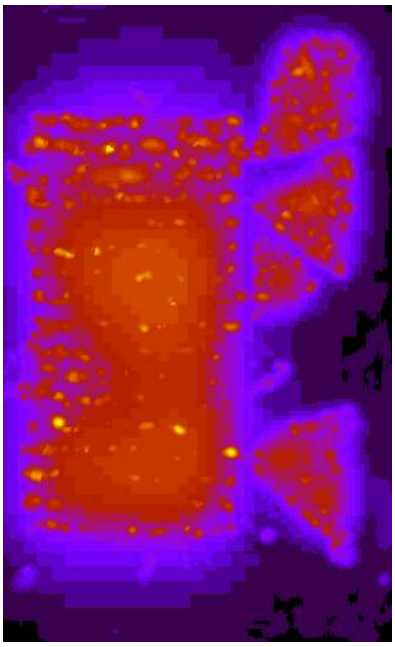
Abtastfensterpositionen, die am Bildrand liegen, können im Umfang über die äußeren Bildränder hinausreichen. Wie bereits zuvor erwähnt, wird für solche Abtastfenster der durchschnittliche Grauwert des Bildbereichs innerhalb der Bildgrenzen und des Fensters errechnet, um das Bild künstlich mit diesem Grauwert zu erweitern. Somit sollte der dadurch entstehende Fehler so klein wie möglich gehalten werden.

Lokale Maxima der Frequenzstärke können durch Bildstörungen verursacht sein, hauptsächlich aber durch die Schrift, und lassen sich dann drei Strukturgruppen zuordnen:

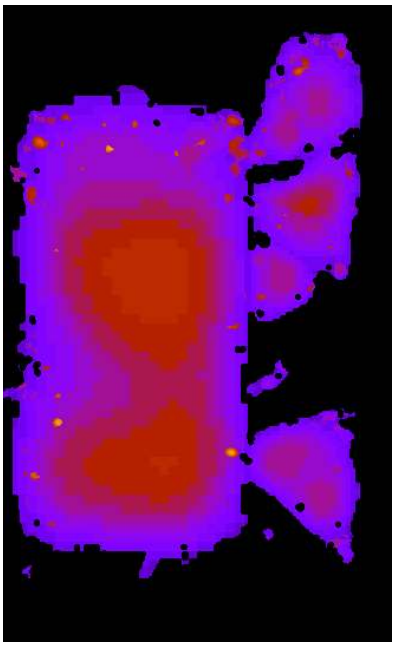
- Intertextblockstruktur: Tritt bei niedrigen Bildauflösungen auf. Häufige Ursache sind Texte in Versform, d.h. z.B. wiederholt drei Zeilen Text, dann ein größerer Weißraum. Meistens ist dieses Auftreten geprägt durch eine geringe Frequenzstärke.
- Textblockstruktur: Tritt in mittlerer bis hoher Frequenzstärke auf.
- Intrazeilenstruktur: Bei hohen Auflösungen können einzelne Zeichen (wie ein 'm') mit parallel ausgeführten Strichen die höchsten Frequenzstärken erreichen, wie in Abbildung 3.22c und Abbildung 3.23 erkennbar.

Untersucht werden nun für jedes Bildpixel der Auflösungsabfolge die so ermittelten Frequenzstärken, angefangen bei der niedrigsten Bildauflösung. In dieser Folge können also bis zu drei lokale Maxima durch die zuvor erwähnten unterschiedlichen Strukturarten entstehen. Wie in Abbildung 3.24a gezeigt, würde die einfache Bestimmung der in der Folge maximalen Frequenzstärke häufig das Maximum der Intrazeilenstruktur ermitteln. Daher verwirft eine Heuristik, angefangen beim zuerst gefundenen lokalen Maximum bei geringer Bildauflösung, weitere lokale Maxima, die bei höherer Bildauflösung entstehen, soweit sie nicht mindestens 20% stärker sind (dieses Maß hat sich in vielen Versuchen als sinnvolle Größe bewährt). Neben weiteren Faktoren wird auch die Stärke des jeweils zweitstärksten lokalen Maximums bei gleicher Bildauflösung mit einbezogen, da die Differenz zwischen stärkstem und zweitstärkstem Maximum beim Erfassen der Intrazeilenstruktur häufig deutlich niedriger ist als die in der Textblockstruktur. Diese Differenz kann als Maß der Zuverlässigkeit der erfassten Maxima interpretiert werden.

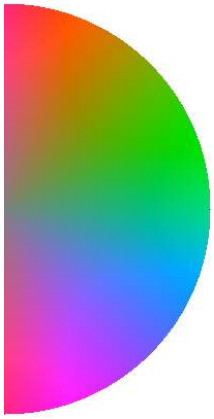
Abbildung 3.24b zeigt das Ergebnis nach Anwendung der Heuristik, bei dem bereits Bereiche als Hintergrund maskiert wurden, die eine insgesamt zu niedrige Frequenzstärke aufweisen. Die



(a) Stärkstes Frequenzmaximum aller Auflösungsstufen für jedes Bildpixel

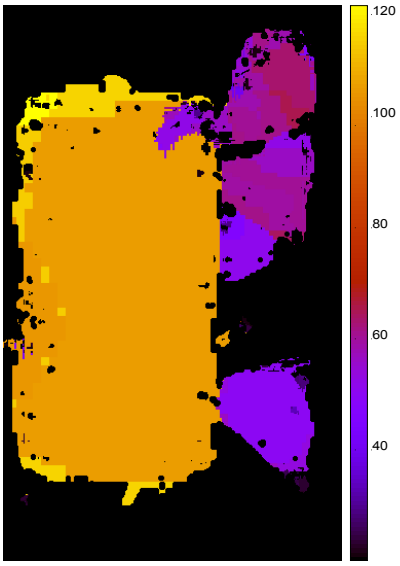


(b) Durch Heuristik ausgewähltes Frequenzmaximum, Hintergrund maskiert



(c) Farbkodierung der Zeilenorientierung

(d) Ermittelte Zeilenorientierung



(e) Ermittelter Zeilenabstand

Abbildung 3.24.: Die Kombination der größten Werte jedes Bildpixels aus Abbildungen 3.22c bis 3.22j ergibt Abbildung 3.24a. Abbildung 3.24b zeigt die durch eine Heuristik bereinigten Werte. Die Farbkodierung der Zeilenorientierung zeigt die der Zeilenfrequenz entsprechenden Werte (orthogonal zur Zeilenorientierung). Der Zeilenabstand ist farblich von dunkelblau (sehr geringer Abstand) bis hellgelb (großer Abstand) kodiert.

nun noch immer falsch interpretierten Maxima können in einem späteren Schritt, bei der Bestimmung und Auswertung homogener Regionen, herausgefiltert werden.

Bereits die Verarbeitung bis zu diesem Schritt lässt aber erste Auswertungen zu, die in Abbildungen 3.24d und 3.24e gezeigt werden. Diese sollen vorweg einen ersten Eindruck von diesem Zwischenergebnis vermitteln. Eine pixelgenaue Auswertung hiervon erfolgt in Abschnitt 4.4.1.

3.4.1. Segmentierung von Textblöcken anhand von lokalen Unterschieden der Zeilenstruktur

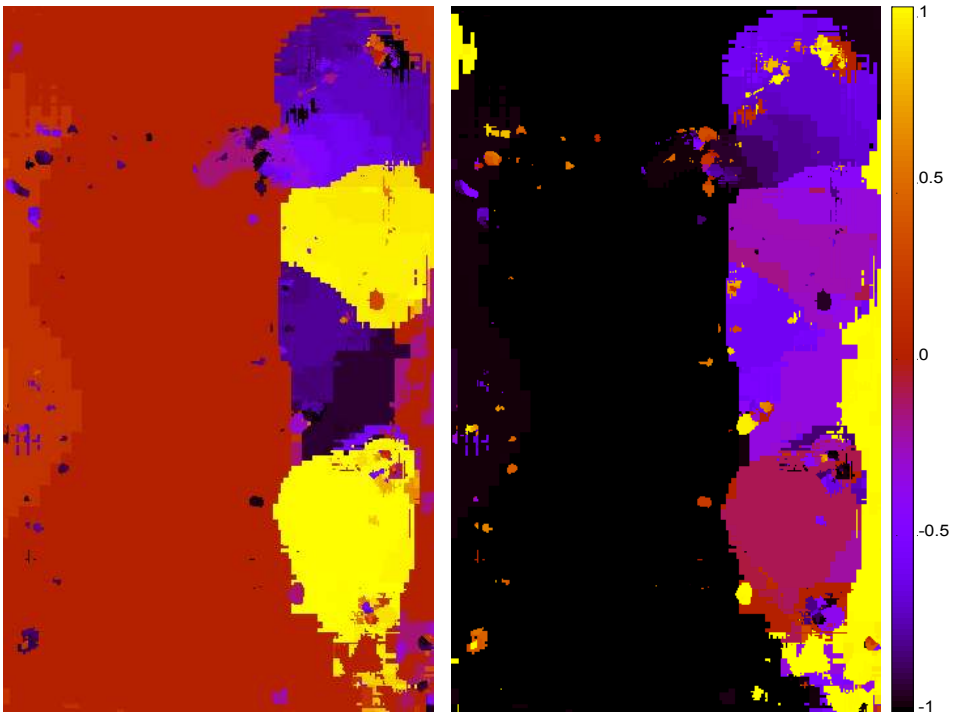
Eine Segmentierung von unterschiedlichen Textblöcken anhand von lokalen Unterschieden bzgl. Zeilenabstand und Orientierung soll im Folgenden gezeigt werden. Wie in den vorherigen Abschnitten gezeigt wurde, können Unterschiede des Zeilenabstandes mit einer Abweichung von etwa 5% oder der Orientierung von etwa 5° erkannt werden. Allerdings können derartige Abweichungen auch innerhalb eines Textblockes mit wenig regelmäßiger Zeilenstruktur vorkommen. Die Herausforderung der im Folgenden beschriebenen Aufgabe liegt nun darin, zu unterscheiden, ob erkannte Unterschiede der Struktur durch unterschiedliche Textblöcke hervorgerufen werden, um diese dann voneinander trennen zu können, oder auf eine Auftrennung zu verzichten, im Falle von in sich heterogenen Textblöcken. Die Schwierigkeit dieser Herausforderung hängt naturgemäß deutlich von den untersuchten Manuskripten ab.

Hierzu wird zunächst ein Gradientenbild erzeugt, welches an den Stellen hohe Werte aufweist, an denen starke lokale Unterschiede auftreten. Anschliessend wird ein Wasserscheidensegmentierungsverfahren auf dieses Gradientenbild angewendet, und die zunächst vorhandene Übersegmentierung sukzessive eliminiert, bis die zurückbleibenden Regionen den ermittelten Textblöcken entsprechen.

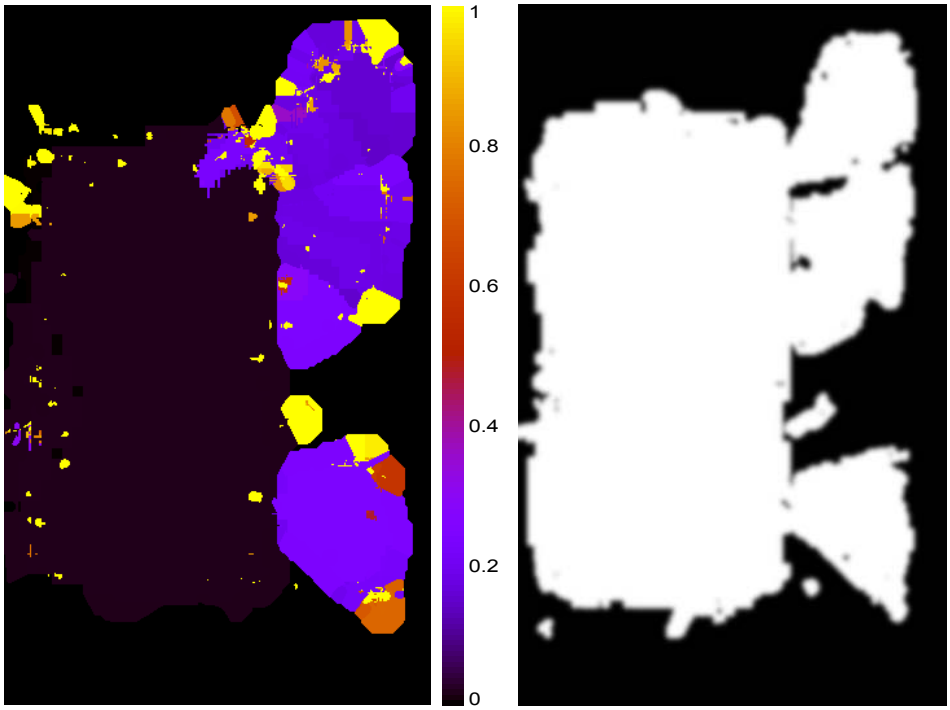
3.4.1.1. Berechnung des Gradientenbildes

Das Ziel ist hier, ein gemeinsames Gradientenbild zu erhalten, indem Gradienten in ausgewogener Gewichtung durch die Zeilenorientierung, den Zeilenabstand sowie durch die Abgrenzung von Text zu Hintergrund entstehen.

Die Zeilenorientierung wird zunächst in Grad errechnet, wobei folgendes gilt: $0° \hat{=} 180°$. Das bedeutet, dass die Differenz von unterschiedlichen Zeilenorientierungen maximal 90° betragen darf. Bei einem vollen Kreisumfang von 0°-360° ist eine übliche Vorgehensweise, Winkeldifferenzen unter Beachtung der Zyklizität zu berechnen, die Umrechnung von Polarkoordinaten in kartesische Koordinaten und die Bestimmung des euklidischen Abstandes zwischen diesen. Diese Methode soll auf den Wertebereich von 0°-180° übertragen werden, indem die errechnete Zeilenorientierung zunächst verdoppelt wird, d.h. 90° wird zu 180° und 180° wird zu 360°. Nun können die Orientierungen als Polarkoordinaten mit dem Radius 1 betrachtet werden, und in kartesische Koordinaten umgerechnet werden. Die maximale Differenz bilden Koordinaten, die



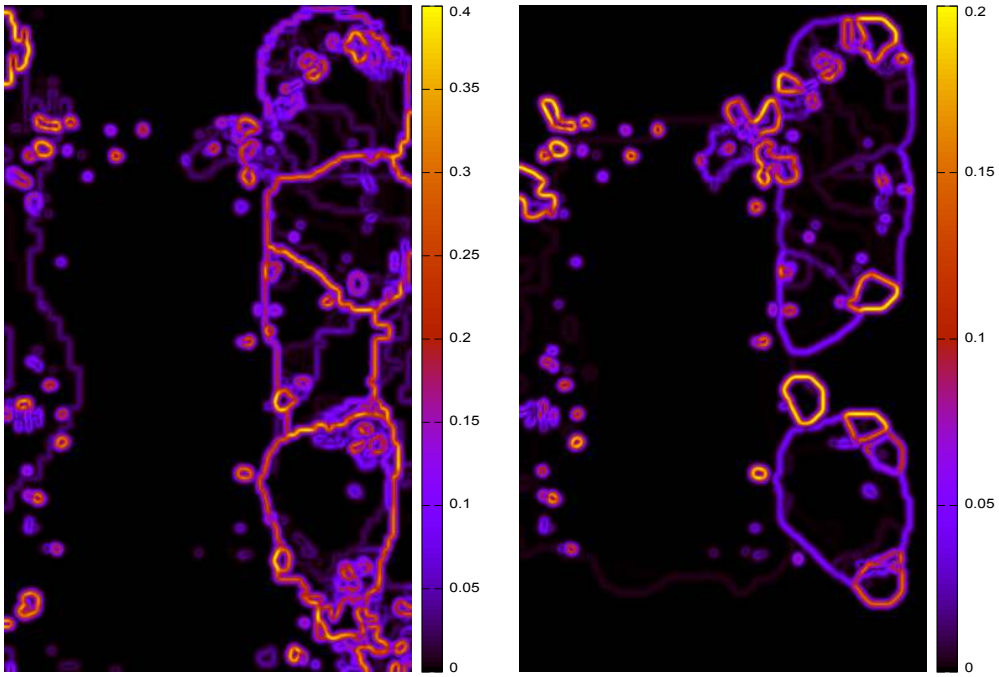
(a) Zeilenorientierung in kartesischen Koordinaten. Links x-Werte, rechts y-Werte



(b) Normierte Zeilendichte

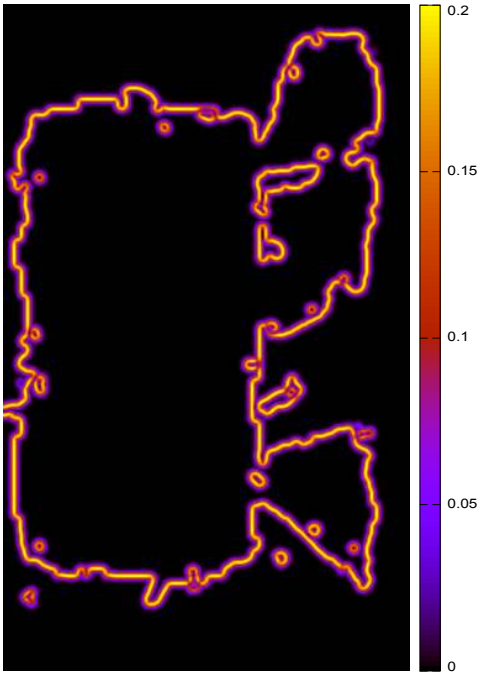
(c) Textmaske

Abbildung 3.25.: Diese normierten Bilder repräsentieren die Zeilenorientierung, die Zeilendichte sowie die Textmaske.

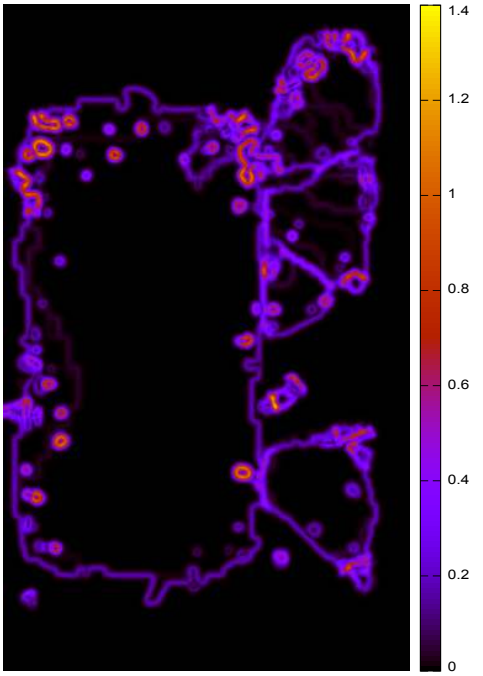


(a) Gradient der Zeilenorientierung

(b) Gradient der Zeilendichte



(c) Gradient der Textmaske



(d) Kombiniertes Bild der drei Gradienten

Abbildung 3.26.: Aus den Bildern von Abbildung 3.25 werden zunächst einzelne Gradientenbilder erzeugt, die dann in ein Bild kombiniert werden, wobei Gradienten außerhalb der Textmaske ignoriert werden.

sich im Einheitskreis gegenüberliegen. Zwischen diesen wird der maximale euklidische Abstand von 2 errechnet, der als maximale Winkeldifferenz von 90° interpretiert werden kann.

Jede errechnete Zeilenorientierung aus Abbildung 3.24d kann in zwei Bildern dargestellt werden, die jeweils die x- und y-Werte repräsentieren, wie in Abbildung 3.25a gezeigt. Um den später beschriebenen Gradienten zwischen Text und Hintergrund nicht zu verfälschen, werden Orientierungen auch außerhalb der Textregionen bestimmt und später verworfen. Beide Bilder sind bereits normiert, da sie nur Werte im Intervall $[-1..1]$ annehmen können.

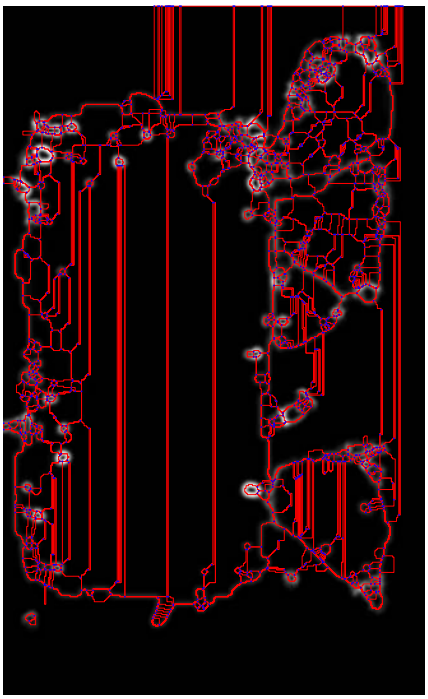
Das Ausgangsbild zur Erstellung des Gradientenbildes zur Repräsentation des Zeilenabstandes ist in Abbildung 3.25b dargestellt. Eine Normierung der Werte fand hier statt, indem alle Werte innerhalb der Textregionen sortiert werden. Um Verfälschungen durch Ausreißer zu vermeiden, werden Minimum und Maximum durch die 2% sowie die 98% Quantile bestimmt. In Abbildung 3.25b wird nun anhand der Extremwerte der niedrigsten Zeilendichte der Wert 0, der höchsten Zeilendichte der Wert 1 zugeordnet.

Die Textregionenmaske dient der Erstellung des Textmaskengradienten. Bildpixel, die Textregionen repräsentieren, erhalten den Wert 1, alle übrigen 0. Das Bild wird durch einen Gaußfilter mit $\sigma = 2$ leicht geglättet (Abbildung 3.25c).

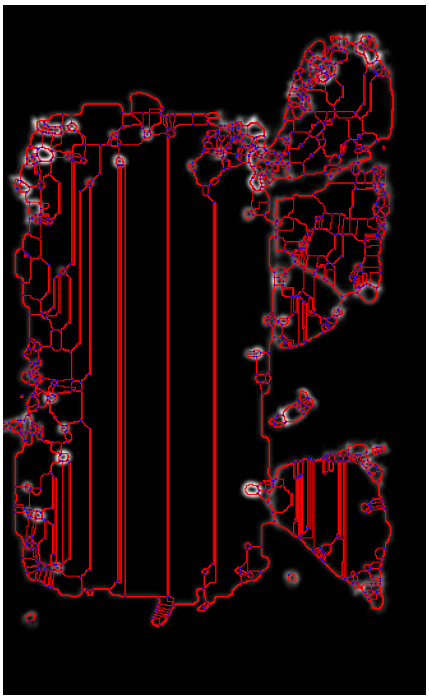
Die Gradientenbilder werden erstellt, indem die vier zuvor beschriebenen Bilder mit der Ableitung der eindimensionalen Gaußfunktion als Kern diskret gefaltet werden, jeweils in x- und y-Richtung, und anschließend der Betrag daraus errechnet wird. Auf diese Weise lassen sich auch die Bilder der Zeilenorientierungen zu einem Gradientenbild kombinieren. Die besten Ergebnisse konnten, unabhängig von der Bildauflösung, durch die Wahl von $\sigma = 2$ erzielt werden. Schließlich werden alle Gradientenbilder (Abbildungen 3.26a bis 3.26c) addiert, wobei das Gradientenbild der Zeilendichte mit vierfacher Gewichtung eingeht. Durch Multiplikation mit der Textregionenmaske werden Gradienten außerhalb der Textregionen eliminiert. Das Ergebnis zeigt Abbildung 3.26d.

3.4.1.2. Segmentierung nach Anwendung des Wasserscheidenverfahrens

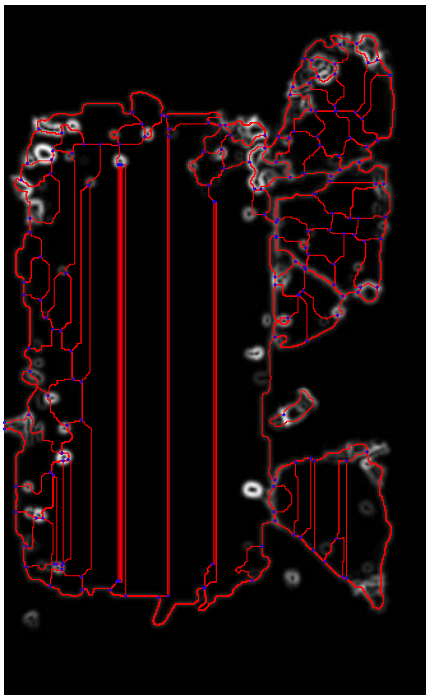
Innerhalb der Textregionen repräsentiert die Stärke der Gradienten die Stärke der lokalen Unterschiede der Zeilencharakteristik. Daher ist eine Auftrennung von Textbereichen durch starke Gradienten zweckmäßig. Eine Herausforderung stellt dagegen die Interpretation schwächerer Gradienten dar. Insbesondere innerhalb der Randtextblöcke kann häufig eine hohe Varianz bzgl. der Zeilenorientierung und des Zeilenabstandes festgestellt werden, wodurch schwache Gradienten entstehen. In diesen Fällen sollen diese Gradienten nicht zu einer Auftrennung führen. Auf der anderen Seite wurde bereits in Abbildung 3.19 dargestellt, dass Unterschiede der Zeilenorientierung von weniger als 16° selbst im oberen Band nur zum kleinstmöglichen Gradienten führen, wie sich anhand der direkt benachbarten Farbdarstellung ableiten lässt. Weisen benachbarte Randtextblöcke einen derartigen Unterschied der Zeilenorientierung auf, wäre eine Auftrennung jedoch erwünscht.



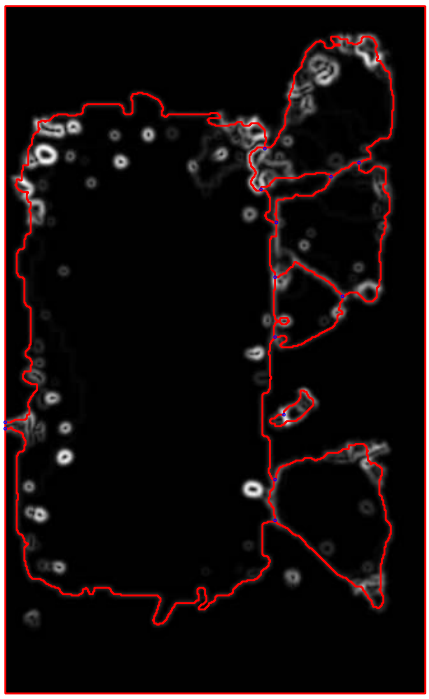
(a) Wasserscheidensegmentierung auf Abbildung 3.26d.



(b) Bereinigte Segmentierung I



(c) Bereinigte Segmentierung II



(d) Abgeschlossene Segmentierung

Abbildung 3.27.: Die zunächst entstandene Übersegmentierung durch das Wasserscheidenverfahren wird in drei Schritten reduziert. Abbildung 3.27d stellt das Ergebnis dar.

Zur Durchführung der Segmentierung wird zunächst das Wasserscheidenverfahren auf das Gradientenbild aus Abbildung 3.26d angewendet. Hierzu wird die „Union-Find“-Variante der VIGRA Bibliothek genutzt, die wiederum als Algorithmus 4.7 in [Roerdink u. Meijster, 2000] beschrieben ist. Dies führt zunächst zu einer Übersegmentierung, wie in Abbildung 3.27a dargestellt wird. Aus dieser Übersegmentierung müssen nun sukzessive Kanten entfernt werden, bis nur noch die Regionen verbleiben, die die gesuchten Textblöcke darstellen.

Auf der Datenbasis des Gradientenbildes wird das GeoMap-Rahmenwerk ([Meine u. Köthe, 2005]) initialisiert, welches die pixelbasierte geometrische und die topologische Repräsentation vereint. Alle Bildregionen sind in dieser Repräsentation von Kanten umgeben, die wiederum jeweils Start- und Endknoten haben. Die GeoMap unterstützt das Adressieren von Knoten, Kanten und Regionen und erlaubt typischerweise nützliche Abfragen wie die Menge aller Kanten, die eine Region umgeben oder die Bestimmung der beiden Flächen beiderseits einer Kante. Auch nach dem Löschen von Kanten verbleibt ein topologisch korrekter Zustand. Die dabei verwendete Operation vereinigt dabei auch zwei Kanten mit einem gemeinsamen Knoten vom Grad 2, und löscht Kanten mit Knoten vom Grad 1. Weiterhin können zusammen mit dem Statistikmodul der VIGRA u.a. Eigenschaften wie die Länge einer Kante, die Größe einer Fläche, oder minimaler, durchschnittlicher und maximaler Pixelwert einer Fläche oder einer Kante ausgegeben werden. Zur Bestimmung dieser statistischen Werte der Kanten werden die Bildwerte unter dem Kantenverlauf aus einem per Splines interpolierten Gradientenbild errechnet, da die Kanten dieser Implementierung der Wasserscheidensegmentierung zwischen den originalen Bildpixeln verlaufen.

Anfänglich sind in diesem Beispiel 766 Flächen entstanden, getrennt von 2209 Kanten. Durch den Gradienten der Textmaske können alle entstandenen Flächen der Segmentierung entweder einer Textregion oder dem Hintergrund zugeordnet werden. In einem ersten Schritt werden nun alle Kanten entfernt, die zwischen zwei Flächen verlaufen, die beide dem Hintergrund zugeordnet werden. Alle Kanten, die auf dem Textmaskengradienten, d.h. zwischen einer Hintergrundfläche und einer Textregionenfläche verlaufen, werden durch ein Flag vor späterem Löschen geschützt. Eine Ausnahme bilden verbleibende Hintergrundflächen, deren Fläche weniger als ein Fünfhundertstel der Fläche des Gesamtdokuments darstellen. Dies ist ein empirisch bestimmter Wert, der häufig auf sehr kleine Einschlüsse von Hintergrundbereichen in der Textmaske zutrifft, die später eliminiert werden können. Die verbleibende Segmentierung zeigt Abbildung 3.27b. Es verbleiben 641 Flächen, getrennt von 1835 Kanten.

In einem zweiten Schritt werden Flächen bestimmt, die relativ zu einem ermittelten Zeilenabstandsmaß unrealistisch klein sind, um sie dann mit ähnlichen, benachbarten Flächen zu vereinigen. Dazu wird von allen Flächen der Median der Zeilenabstände in Pixeln bestimmt, die als Kehrwert der Zeilendichte für alle Bildpixel von Abbildung 3.25b (vor der Normierung) vorliegen. Das Quadrat des Minimums aller Mediane wird als Schwellenwert Θ definiert. Abgeleitet ist dieses Maß von der Vorstellung, dass die Fläche von Textblöcken eine minimale Größe haben sollten, die etwa dem Quadrat des dreifachen Zeilenabstandes entspricht. Da die Gradienten aber nicht immer genau zwischen zwei Textblöcken liegen, sondern, wie in Abbildung 3.20 gezeigt, in Richtung eines schwächer ausgeprägten Textblocks verschoben sein können, werden hier lediglich Flächen behandelt, deren Pixelfläche kleiner als Θ ist. Damit ist auch unter

schwierigen Konstellationen mit keiner unerwünschten Flächenvereinigung zu rechnen. Bei diesen Flächen wird nun die Kante bestimmt, die den kleinsten minimalen Gradienten aufweist, und gelöscht. Diese Prozedur wird solange über alle Flächen angewendet, bis keine unrealistisch kleinen Flächen mehr übrig sind. Eine Ausnahme stellen Textflächen dar, die als Inseln nur von Texthintergrund umgeben sind und deren Außenkante im vorigen Schritt geschützt wurde. Da die Auswertung ergab, dass alleinstehende Textblöcke manchmal nur ein sehr schwaches Frequenzsignal liefern, werden solche Inseltextregionen, deren Größe lediglich $\frac{\Theta}{4}$ entspricht, akzeptiert. Kleinere Regionen werden trotz Kantenschutz mit dem Hintergrund verschmolzen. Übrig bleiben dadurch in dem Beispiel nur noch 95 Flächen, abgegrenzt von 275 Kanten. Das Ergebnis dieses zweiten Schrittes zeigt Abbildung 3.27c.

Mit den beiden zuvor beschriebenen Schritten kann bereits ein großer Teil der Übersegmentierung eliminiert werden. Da lediglich einfache, eindeutige Kriterien verwendet wurden, ist die Ausführung nicht nur sehr schnell, es kann zudem beobachtet werden, dass keine unerwünschte Kantenentfernung auftrat.

Die dritte und letzte Phase der Kantenentfernung soll aufgrund der Komplexität mit Hilfe einer Support Vector Machine (SVM) ausgeführt werden, genutzt wurde dafür die Implementierung aus Chang u. Lin [2011]. Um der SVM eine solide Datenbasis zur Verfügung zu stellen, werden für alle Kanten, und deren direkt benachbarten Flächen, folgende Eigenschaften ermittelt:

Für alle Kanten:

1. Minimumwert des darunterliegenden Gradienten
2. Maximumwert des Gradienten
3. Wert an der 10% Quantile
4. Wert an der 20% Quantile
5. Wert an der 30% Quantile
6. Wert an der 50% Quantile
7. Wert an der 70% Quantile
8. Wert an der 80% Quantile
9. Wert an der 90% Quantile
10. Länge der Kante

Für jede dieser Kanten werden nicht nur die Eigenschaften der zwei direkt angrenzenden Flächen (A und B), sondern vorweg auch die einer dritten Fläche (C) bestimmt, die entstehen würde, wenn die Flächen A und B durch das Löschen der Kante verschmolzen würden.

Für die Flächen:

1. Größe [Pixel] (Fläche A)
2. Größe [Pixel] (Fläche B)

3. Median der Zeilenabstände (Fläche A)
4. Median der Zeilenabstände (Fläche B)
5. Varianz der Zeilenabstände (Fläche A)
6. Varianz der Zeilenabstände (Fläche B)
7. Varianz der Zeilenabstände (Fläche C)
8. Durchschnittliche Zeilenorientierung in Grad (bestimmt aus dem jeweiligen Median der kartesischen Werte, siehe Abbildung 3.25a) (Fläche A)
9. Durchschnittliche Zeilenorientierung in Grad (Fläche B)
10. Differenz dieser Zeilenorientierungen (A/B)
11. Varianz der kartesischen Zeilenorientierungskomponente in x-Richtung (Fläche A)
12. Varianz der kartesischen Zeilenorientierungskomponente in y-Richtung (Fläche A)
13. Varianz der kartesischen Zeilenorientierungskomponente in x-Richtung (Fläche B)
14. Varianz der kartesischen Zeilenorientierungskomponente in y-Richtung (Fläche B)
15. Varianz der kartesischen Zeilenorientierungskomponente in x-Richtung (Fläche C)
16. Varianz der kartesischen Zeilenorientierungskomponente in y-Richtung (Fläche C)
17. Quotient aus Summe der Länge aller Kanten, die direkt den Hintergrund abgrenzen (für Fläche A) und Summe der Länge aller Kanten für Fläche A
18. Quotient aus Summe der Länge aller Kanten, die direkt den Hintergrund abgrenzen (für Fläche B) und Summe der Länge aller Kanten für Fläche B

    und schließlich zur Normierung von Flächengrößen und Kantenlängen:

19. Dokumentbreite in Pixeln
20. Dokumenthöhe in Pixeln

Für das Training der Support Vector Machine wurden nun die mit diesen Kriterien beschriebenen Kanten folgendermaßen klassifiziert: Die Segmentierung von 40 Manuskriptdokumenten wurde bis einschließlich der oben beschriebenen zwei Basisschritte vorbereitet. Von Hand wurden dann bis zum gewünschten Ergebnis Kanten entfernt oder Kanten als notwendig markiert. Angefangen wurde dabei mit besonders kurzen Kanten, für jede so behandelte Kante wurden zusammen mit der Information, ob die Kante gelöscht wurde oder nicht, die oben beschriebenen 30 Merkmale gespeichert. Insgesamt wurden so über 10.000 Kanten klassifiziert.

Die dabei verwendeten 40 Manuskriptseiten überschneiden sich teilweise mit denen aus dem im nächsten Kapitel vorgestellten Testsatz. Für diesen Fall wurde die Datenmenge des Trainings um die jeweilige Manuskriptseite reduziert.

Abbildung 3.28.: Ergebnis der Segmentierung. Unterschiedliche Farben stellen einzeln erkannte Textblöcke dar.

Tabelle 3.3.: Unterschiede zwischen den von Hand vermessenen Textblockstrukturen ("Ground Truth") und den analytisch ermittelten Werten

	Ground Truth		Ermittelt		Differenz	
Textblock	Zeilenabstand	Orientierung	Zeilenabstand	Orientierung	Zeilenabstand	Orientierung
1	62	158	55.4	156.9	10.66%	1.1
2	108	1	103.3	179.8	4.35%	1.2
3	58	38	57.6	35.1	0.69%	2.9
4	53	151	49.0	152.1	7.55%	1.1
5	51	42	49.0	40.8	3.92%	1.2

Eine so trainierte SVM finalisiert nun in einem dritten und letzten Schritt die Segmentierung und bewertet dabei alle Kanten, angefangen bei der kürzesten. Lediglich acht Flächen verbleiben, davon werden zwei dem Hintergrund zugerechnet. Die Flächen werden durch 15 Kanten umgeben. Diese acht Flächen werden in der folgenden Arbeit als berechnete Textblockregionen bezeichnet und sind als Ergebnis in Abbildung 3.27d dargestellt.

3.4.1.3. Auswertung

In Abbildung 3.28 ist dieses Ergebnis übertragen auf das Ausgangsdokument dargestellt. Neben dem Haupttextblock wurden fünf weitere Textblöcke erkannt.

Die äußeren Grenzen zwischen den Textblockregionen und dem Hintergrund kommen einem perfekten Ergebnis schon recht nahe. Bei den oberen Spitzen des hellgrünen und dunkelvioletten Textblocks ist die Frequenzstärke allerdings zu schwach, daher wurde hier Text abgeschnitten. Andererseits führen einzeilige, kurze Randbemerkungen dicht am Haupttextblock zu Verfälschungen der Grenzen, wie etwa mittig oben und unten.

Der (als blauer Textblock dargestellte) falsch erkannte Einzeiler weist Abmessungen auf, die knapp über dem Mindesttextblockmaß Θ liegen. Ein weiterer Fehler ist der nicht erkannte Textblock am linken Seitenrand, der mit seinen drei sehr kurzen Zeilen und kaum lesbarer Schriftfarbe eine zu geringe Frequenzstärke aufweist.

Die Auftrennungen zwischen den Textblockregionen sind grundsätzlich richtig, das Hineinfliessen des oberen Randtextblocks in den Haupttext macht eine perfekt positionierte Auftrennung allerdings besonders schwierig. Falsch zugeordnet ist auch die obere Spitze des dunkelgrünen Textblocks, sowie der kleine Einschluß des Haupttextblocks in den dunkelgrünen Randtextblock. In beiden Fällen war das Zeilensignal des dunkelgrünen Textblocks an den Stellen nicht stark genug.

Tabelle 3.3 vergleicht die von Hand vermessenen Duchschnittswerte mit den errechneten Durchschnittswerten der Textblöcke. Die Nummerierung erfolgt anhand der Position der Textblöcke von oben nach unten, ohne Berücksichtigung der falschen Erkennung. Der Haupttextblock (2) ist

deutlich sorgfältiger geschrieben als die Randtexte und weist dadurch eine sehr geringe Strukturvarianz auf, wodurch die Differenz der Orientierungen besonders klein ist.

3.5. Zusammenfassung

Textblöcke auf Manuskriptbildern weisen ein abwechselnd helles und dunkles Streifenmuster auf, welches mit Hilfe der diskreten Fouriertransformation analysiert werden kann. Eine Spezialisierung der Fouriertransformation ist die Gabor-Transformation, die durch Anwendung einer Gauß-Fensterfunktion eine örtliche Einschränkung des analysierten Bereiches der Manuskriptseite zulässt. Als Antwort werden die Eigenschaften der Signalstärke, der Orientierung und der durchschnittlichen Wellenlänge ausgegeben, die interpretiert eine Abgrenzung von Textbereichen vom Hintergrund zulassen, sowie in der Regel die Orientierung und den Abstand von Textzeilen repräsentieren.

Für die Analyse wird ein Abtastfenster der Größe 121×121 Pixel gewählt, und dieses in einem Raster von 5 Pixeln Abstand über die Manuskriptseiten geführt. Durch eine Veränderung der Auflösung der Manuskriptbilder können Zeilenstrukturen der Textblöcke derartig durch das Abtastfenster erfasst werden, dass dessen Durchmesser entweder dem vier- bis achtfachen des Zeilenabstandes entspricht, oder dem acht- bis etwa 13-fachen, oder dem 12 bis 19-fachen Zeilenabstand. Durch Aufteilung des Frequenzspektrums in drei Bänder erhält man eine dem Zeilenmuster entsprechende Antwort im ersten Fall im unteren Frequenzband, sonst im idealen oder im letzten Fall im oberen Frequenzband.

Eine Kombination der Antworten aus unterem und idealem Frequenzband wird zum Auffinden der Grenzen zwischen Textregion und Hintergrund genutzt, die Kombination der Antworten aus idealem und oberen Frequenzband für eine Interpretation von Zeilenorientierung und Zeilenabstand.

Abgrenzen lassen sich Textblöcke gleichartiger Struktur voneinander, wenn zwischen ihnen ein Mindestabstand von 2,5 Zeilen vorhanden ist. Unterschiedliche Zeilenabstände lassen sich bei Abweichungen von etwa 8% erkennen, unterschiedliche Zeilenorientierungen ab einer Differenz von etwa 6°. Dadurch entstehen durchschnittliche Fehler von etwa 4% bezüglich des Zeilenabstandes sowie etwa 3° hinsichtlich der Orientierung.

Für eine Segmentierung von Textblöcken auf Manuskripten beliebiger Auflösung wird die Originalauflösung eines Manuskriptes sukzessive um den Faktor $\sqrt{2}$ verringert. Die Ergebnisse der Analyse aller Auflösungsstufen lassen sich jeweils wieder in Bildform darstellen, und anschließend in Bildern vereinen, die die Ergebnisse aller Auflösungsstufen kombinieren. Lokale Unterschiede von Bildbereichen lassen sich gut durch Gradientenbilder interpretieren, daher wird jeweils ein Gradientenbild berechnet, das lokale Unterschiede der errechneten Orientierungen, der Zeilenabstände und die Abgrenzungen von Textbereichen zum Bildhintergrund darstellt.

Nach Zusammenfügen der Gradientenbilder wird durch eine Wasserscheidensegmentierung zunächst eine initiale Segmentierung ermöglicht. Die finale Textblocksegmentierung entsteht,

nachdem aus der initialen Segmentierung zunächst nach festgelegten Kriterien, dann mithilfe einer SVM irrelevante Kanten entfernt wurden.

Als Resultat lassen sich neben der Anzahl der gefundenen Textblöcke unter anderem auch ihre Größe, ihre Position, die durchschnittliche Orientierung und der durchschnittliche Zeilenabstand ausgeben.

Auch wenn das Verfahren darauf ausgelegt ist, einen größeren Datenbestand ohne Benutzereingriff, z.B. über Nacht, zu analysieren, so ist das Verfahren dennoch schnell genug, um auch eine einzelne Manuskriptseite ohne unzumutbare Wartezeiten zu verarbeiten. Um das Ergebnis aus Abbildung 3.28 zusammen mit den in Tabelle 3.3 enthaltenen Ergebnissen zu erhalten, vergehen auf einem Computer mit einer älteren Intel Corei5-4670 CPU keine sechs Minuten, wobei der allergrößte Anteil der Zeit für die Berechnung der Gabor-Transformationen benötigt wird. Diese Zeit wird zum einen durch das sehr feine Abtastraster von 5×5 Bildpixeln festgelegt, zum anderen durch die Bildauflösung, die im obigen Beispiel 1650×2700 Pixel beträgt. Für die Analyse einer anderen Manuskriptseite mit 1125×822 Bildpixeln werden bei gleicher Rasterung hingegen nur 54 Sekunden benötigt. Beide Manuskriptseiten sind im Testsatz enthalten, der im nächsten Kapitel vorgestellt wird.

4. Anwendung des Verfahrens auf Manuskriptbeispiele

Zur Überprüfung des Verfahrens wurde ein Testsatz von 50 Manuskriptseiten erstellt, da während der Entwicklung dieser Arbeit kein adäquater Benchmark zur Verfügung stand. Der Testsatz besteht aus 15 islamischen Manuskripten aus der Staatsbibliothek zu Berlin, sowie 15 islamischen Manuskripten aus der Universitätsbibliothek in Leipzig. Weiterhin wurden 15 griechische Manuskripte ausgewählt, die restlichen 5 Manuskripte stellen Sonderfälle dar:

- Zwei Koranmanuskriptseiten aus dem Norden Nigerias, die aufgrund ihrer teilweise ausgeblichenen Schrift und dunkler Ränder die Grenzen eines jeden Textsegmentierungsverfahrens aufzeigen dürften,
- ein Poster mit vier Textblöcken in vier verschiedenen Schreibsystemen,
- eine Sanskrit-Manuskriptseite mit mittig eingeschlossenem Bild, sowie
- ein Ende einer japanischen Manuskriptrolle mit hellem Text auf dunklem Hintergrund und dunklem Text auf hellem Hintergrund.

Der Testdatensatz steht inklusive der (wie nachfolgend beschrieben) erzeugten Ground Truth hier zur öffentlichen Verfügung: [Testset].

Islamische Manuskripte dürften die größten Herausforderungen darstellen, da in diesem Kulturkreis Manuskripte mit vielen verschiedenen Textblöcken auf einer Seite und sehr komplexem Layout zu finden sind. Daher überwiegen sie in diesem Testdatensatz deutlich.

Zunächst werden in den beiden folgenden Abschnitten Anforderungen und Vorbereitungen des Testdatensatzes vorgestellt, dann folgt in Abschnitt 4.3 die detaillierte Beschreibung der Bestimmung einer Ground Truth.

Abschnitt 4.4 erklärt die Berechnung des Abgleichs zwischen vermessener und berechneter Struktur. Schließlich wird in Abschnitt 4.5 der Testdatensatz im Einzelnen beschrieben, die Berechnung der Struktur ausgewertet und das Ergebnis analysiert.

4.1. Auswahlkriterien

Bei der Auswahl der Dokumente wurde darauf geachtet, dass darin Manuskriptdokumente aus verschiedenen Epochen mit unterschiedlichen Schreibsystemen enthalten sind, und dass diese durch eine möglichst hohe Zahl unterschiedlicher Textblöcke genügend Herausforderungen darstellen.

Die Textblöcke des Testdatensatzes sollten nach menschlichem Ermessen möglichst unzweifelhaft erkennbar und aufgeteilt sein. Allein dieser letzte Gesichtspunkt stellt eine größere Herausforderung dar, da selbst in einem einzigen Kulturkreis die Ansichten darüber, welche Eigenschaften einen Textblock ausmachen, häufig unterschiedlich sind, und eine eindeutige Definition über verschiedene Kulturkreise hinweg kaum möglich sein dürfte. Darüber hinaus sollten unterschiedliche Gesichtspunkte nicht miteinander vermengt werden. So ist es beispielsweise in der Paläographie wünschenswert, verschiedene Textstellen, die im Hinblick auf die These des identischen Schreibers untersucht werden sollen, rein nach Kriterien der Schriftmerkmale hin zu untersuchen, und dabei inhaltliche Ähnlichkeiten der Texte zu ignorieren. In gleicher Weise soll auch hier nicht von zusammenhängendem Inhalt auf die Zusammengehörigkeit von Textzeilen zu einem Block entschieden werden, sondern dafür alleine visuelle Kriterien herangezogen werden. Schließlich soll dieses Verfahren auch in einer Prozesskette, die die digitale Paläographie unterstützt, eingesetzt werden können.

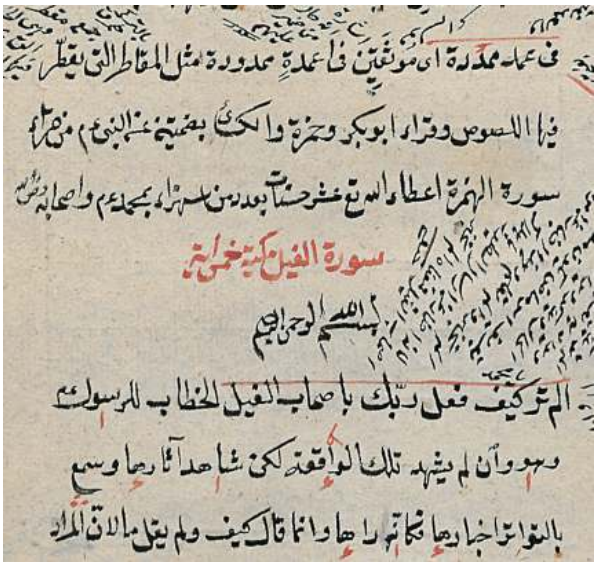
Grundsätzlich soll der Testdatensatz die mit dem beschriebenen Verfahren erzielbaren Ergebnisse nicht nur anschaulich belegen, sondern einen objektiven numerischen Abgleich mit einer pixelgenau erstellten Ground Truth ermöglichen. Für ein objektives, möglichst unzweifelhaftes Erstellen einer Ground Truth sind mindestens folgende drei Anforderungen erforderlich:

1. Jede im Testdatensatz enthaltene Schrift sollte gut erkennbar sein und sich nach menschlichem Ermessen möglichst eindeutig genau einem Textblock zuordnen lassen.
2. Die Varianz der Zeilenstruktur innerhalb eines Textblocks sollte möglichst gering sein.
3. Textblöcke sollten möglichst kompakt sein.

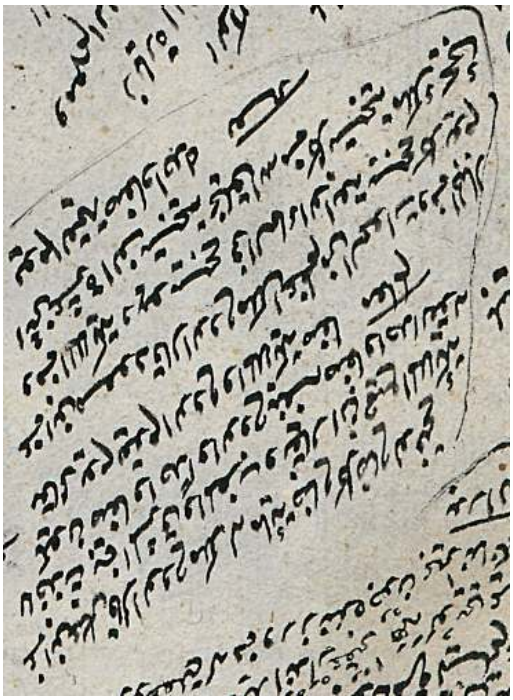
Der erste Punkt bedeutet, dass Textblöcke sich möglichst eindeutig von umstehender Schrift und anderen Textblöcken abgrenzen lassen sollten. Textbereiche werden als ein Textblock gewertet, soweit die Struktur weitgehend gleichmäßig ist. Das bedeutet auch, dass mehrere Absätze, wie es bei Haupttextblöcken häufiger vorkommt, zu einem Textblock zusammengefasst werden. Absätze in Randtexten sind dagegen unüblich. Grenzwertige Beispiele von Textbereichen, die noch als ein Textblock gewertet werden, finden sich in Abbildung 4.1.

Dennoch werden Textbereiche als unterschiedliche Textblöcke gewertet, soweit ihre Struktur zwar ähnlich, aber nach menschlichem Ermessen unterschiedlich ist. Dieser Eindruck kann entstehen, wenn ein Textbereich einen gleichmäßigen Zeilenabstand hat, der sprungweise in einen ähnlichen, aber geringfügig unterschiedlichen, wiederum gleichmäßigen Zeilenabstand übergeht. Visuelle Unterschiede können bis zu einem gewissen Grad auch durch einen unterschiedlichen Schreibstil, engere oder gesperrte Schrift, oder unterschiedlich gesättigte Tinte oder

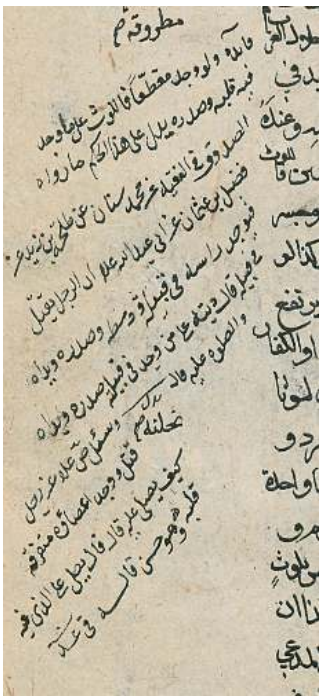
(a) Wertung des Haupttextbereichs als ein Textblock, trotz der optischen Trennung in zwei Spalten in der unteren Hälfte.



(b) Da die Struktur des Haupttextblocks einheitlich bleibt, wird er auch über die zwei sehr kurzen Zeilen hinweg als Einheit gewertet.

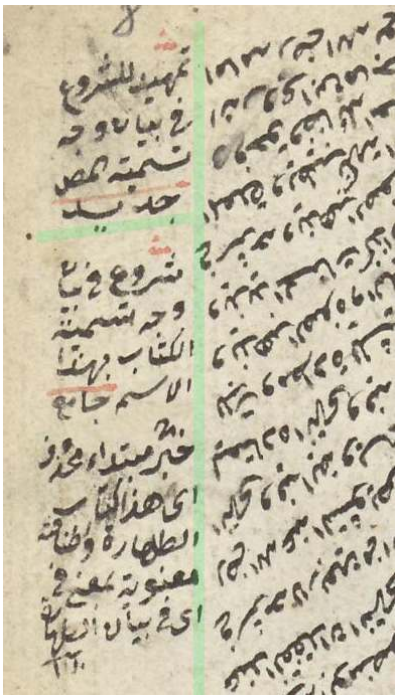


(c) (Die Orientierung entspricht dem Originaldokument)

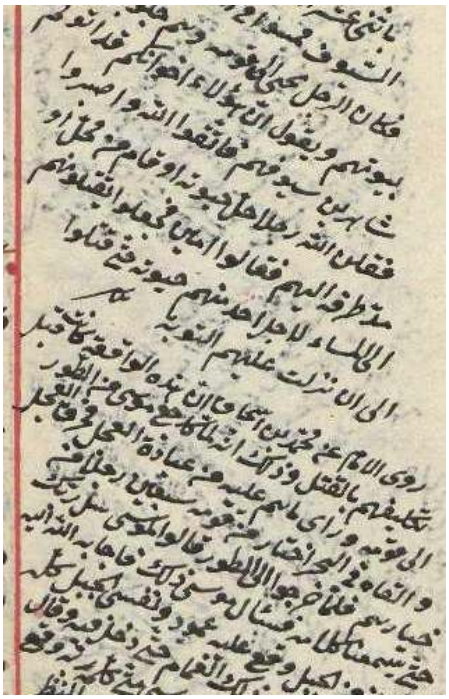


(d)

Abbildung 4.1.: Beispiele für Textbereiche, die aufgrund weitgehend einheitlicher Struktur als ein Textblock gewertet werden.



(a) Links werden zwei Textblöcke vom rechten getrennt (durch eingefügte leicht grünliche Trennlinien verdeutlicht).



(b) Durch den erkennbar unterschiedlichen Zeilenabstand werden hier zwei unterschiedliche Textblöcke angenommen.

(c) Die diagonal geführten Zeilen links werden als ein Textblock gewertet, die gegenläufig diagonalen unten als ein zweiter. Die vertikale Textzeile wird als Einzeiler ignoriert.

Abbildung 4.2.: Beispiele für Textbereiche, die als getrennte Textblöcke gewertet werden.

Schreibwerkzeuge hervorgerufen sein. Grenzwertige Beispiele für Textbereiche mit ähnlicher Struktur, die aber als unterschiedliche Textblöcke aufgefasst werden, zeigt Abbildung 4.2.

In Abbildung 4.3 werden dagegen Textbereiche gezeigt, deren sinnvolle Aufteilung anhand rein optischer Kriterien – teilweise aufgrund mangelnder Erkennbarkeit – nahezu unmöglich ist. Auch sollte es vermieden werden, dass im Testdatensatz unübersichtliche Situationen auftreten, wie ein zweizeiliger Text, gefolgt von einem einzeiligen, mit gleicher Ausrichtung und Schriftgröße. Oft kann dann nicht eindeutig entschieden werden, ob die Situation nicht doch einen dreizeiligen Textblock darstellt, womit die Ground Truth anzweifelbar wäre.

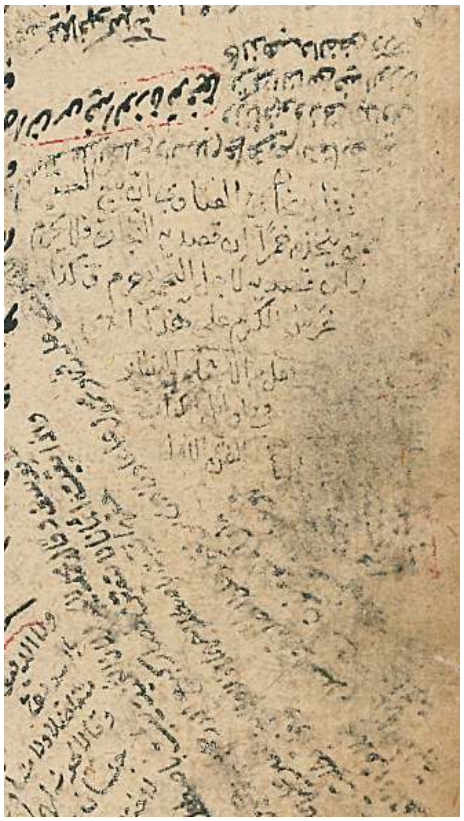
Punkt zwei der obigen Anforderungen (Homogenität) ergibt sich dadurch, dass die Beschreibung der Textstruktur durch die Ground Truth lediglich anhand der beiden Werte

- durchschnittlicher Zeilenabstand und
- durchschnittliche Zeilenorientierung.

bestimmt werden soll. Natürlich könnte die Ground Truth auch deutlich heterogene, aber eindeutig zusammenhängende Textblöcke enthalten. Auch solche könnten in der Regel von dem hier vorgestellten Verfahren erkannt werden. Die Ground Truth soll aber über die Prüfung der Übereinstimmung der Segmentierung hinaus einen Vergleich der innerhalb der Textblöcke vermessenen Werte ermöglichen, und eine manuelle Vermessung der durchschnittlichen Werte wird bei zunehmender Varianz eines Textblocks zunehmend ungenauer, womit die Auswertung verfälscht würde. Daher sollte die Varianz dieser Struktur möglichst gering sein. In Abbildung 3.17 wurde bereits ein extremes Beispiel gezeigt, bei dem man keine sinnvolle durchschnittliche Zeilenorientierung ermitteln kann.

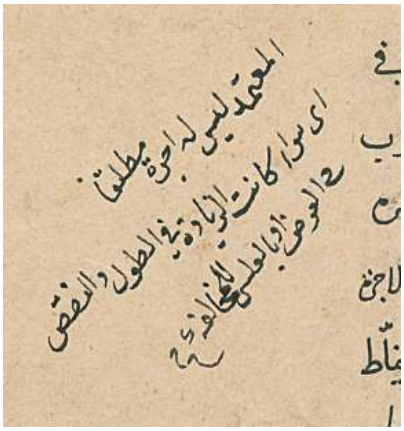
Beispiele für Textblöcke, deren Struktur uneindeutig ist, werden in Abbildung 4.4 dargestellt. Hier kann das in dieser Arbeit beschriebene Verfahren zwar auch lokal eingegrenzt genauere Werte für einzelne Textblockbereiche liefern, aber das Erstellen einer ausreichend genauen Ground Truth mit unterschiedlichen Angaben innerhalb einzelner Textblöcke würde durch eine genaue Berücksichtigung dieser lokalen Informationen unpraktikabel werden. Daher wurde versucht, Manuskriptseiten mit derartigen Textblöcken im Testdatensatz zu vermeiden.

Die Anforderungen von Punkt drei (Kompaktheit) ergeben sich folgendermaßen: Da Randtexte in islamischen Manuskripten häufig ohne jegliche Formatierungsregeln geschrieben werden, kann es vorkommen, dass längere ein- oder zweizeilige Schrift in mehr als dreizeiligen Textblöcken münden. Da aber ein- und zweizeiliger Text nicht als eigenständiger Textblock gewertet werden soll, ist eine Festlegung der äußeren Abgrenzungen von mindestens dreizeiligen Textblöcken problematisch. Daher sollte eine derartige Konstellation möglichst nicht vorkommen. Idealerweise sollten Textblöcke auch eine durchgängige Zeilenbreite aufweisen, die dem dreifachen Zeilenabstand entspricht, da nur so eine zuverlässige Messung der Zeilenstruktur erfolgen kann. Beispiele von Textbereichen mit unzureichender Kompaktheit zeigt Abbildung 4.5.



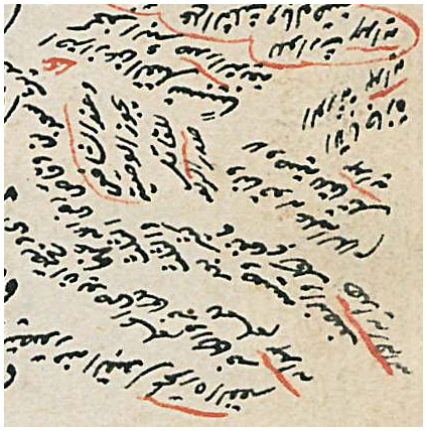
(a) Eine Aufteilung ist bei schlecht erkennbarer Schrift oft nicht möglich.

(b) Wo und in wieviele Textblöcke sollte der linke Textbereich aufgeteilt werden?



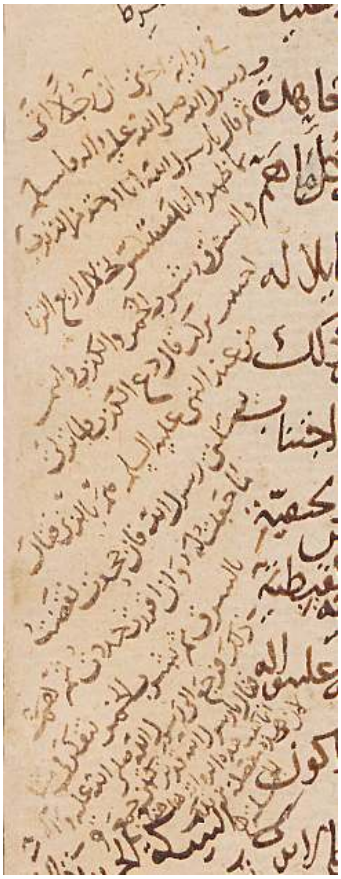
(c) Sollten diese drei Zeilen als ein Textblock gewertet werden?

(d) Ist das ein einzelner Textblock?

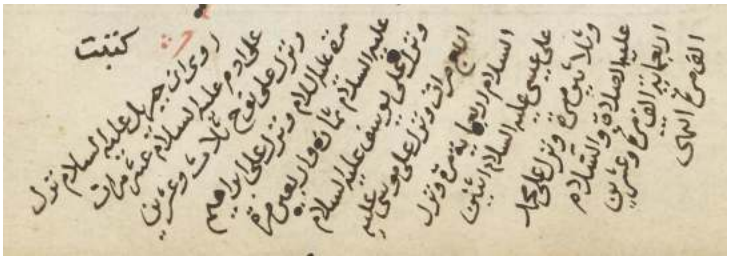


(e) Viele ein- und zweizeilige Textbereiche erschweren eine Aufteilung erheblich.

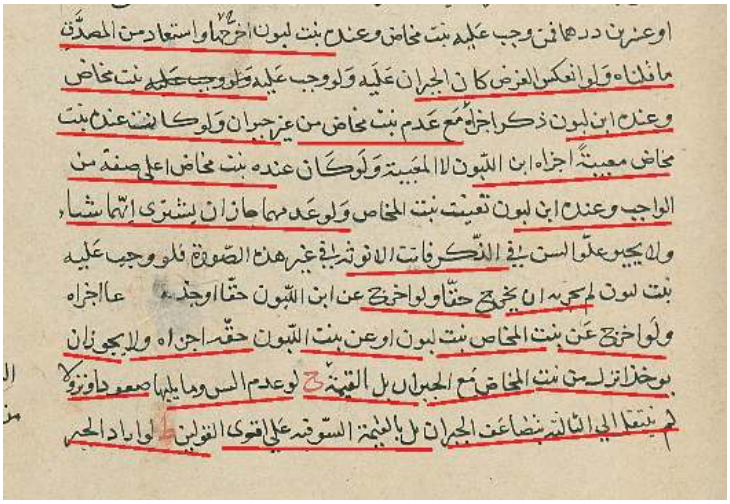
Abbildung 4.3.: Beispiele von Textbereichen, die nach rein optischen Gesichtspunkten nicht eindeutig in einzelne Textblöcke unterteilt werden können. Derartige Situationen sollten daher im Testdatensatz unbedingt vermieden werden.



(a) Die Bestimmung eines durchschnittlichen Zeilenabstandes ist hier sehr schwierig.

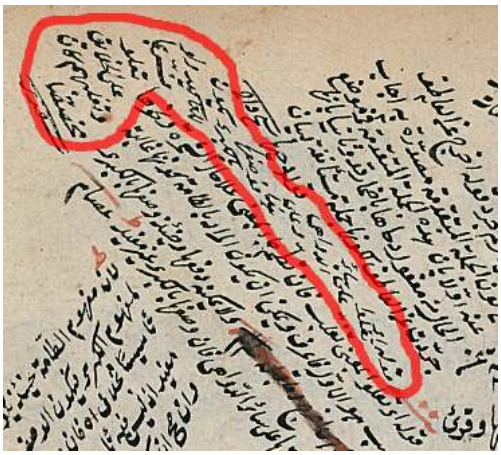


(b) Die Veränderung der Orientierung bedeutet auch, dass der Zeilenabstand oben geringer ist als unten.

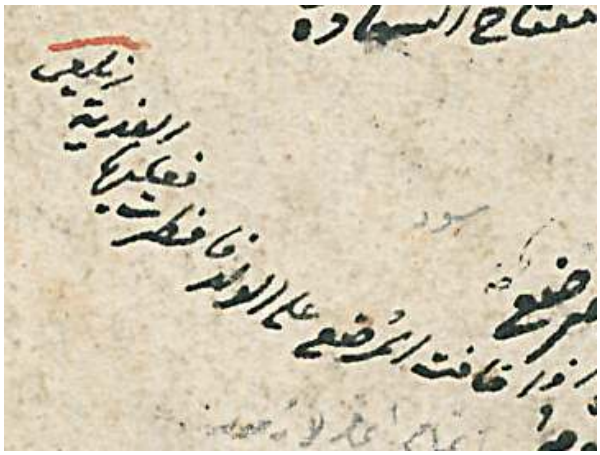


(c) Selbst in einigen Haupttextblöcken variiert die Zeilenorientierung deutlich. (Die roten Unterstriche wurden zur besseren Veranschaulichung eingefügt)

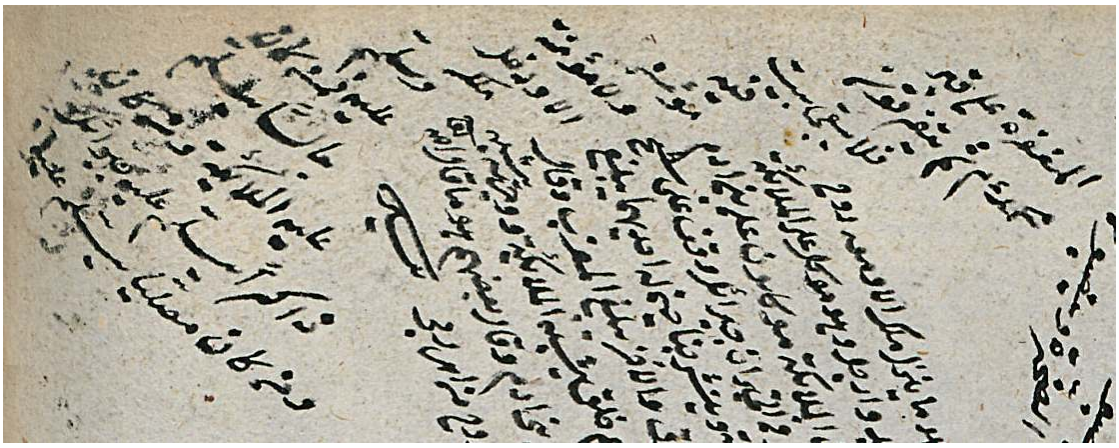
Abbildung 4.4.: Variierende Orientierungen und Zeilenabstände können durch einen einzelnen Durchschnittswert nicht erfasst werden..



(a) Im rot markierten Textbereich finden sich nur links oben mindestens drei untereinanderstehende Zeilen.



(b) Das Verfahren erkennt im oberen Teil einen sehr kleinen, nahezu horizontal ausgerichteten Textblock. Diese Situation soll dagegen als Einzeiler gewertet werden. (Im Testdatensatz enthalten)



(c) Die Grenzen zwischen Textblock und anderen Textbereichen sind häufig fliessend.

Abbildung 4.5.: Bei Textblöcken, die in längere ein- und zweizeilige Textbereiche zerfliessen, sind die äußeren Grenzen schwer zu definieren.

Letztendlich musste allerdings festgestellt werden, dass diese Idealvorstellungen, die oben beschriebenen Situationen zu vermeiden, sich kaum in der Realität der zugänglichen Manuskriptseiten wiederfindet und umsetzen lässt. Also bedeutet die beschriebene Zielsetzung, diese Situationen zu vermeiden, nicht etwa, dass sie sich im Testdatensatz nicht wiederfinden würden. Der Testdatensatz besteht nicht etwa aus synthetisch erzeugten Idealsituationen, sondern er enthält reale Herausforderungen, die dieses Verfahren nicht lösen kann und sicherlich auch für andere Verfahren in naher Zukunft noch ausreichend Herausforderungen bieten wird.

Für die erzeugte Ground Truth bedeutet das daher, dass

1. die erstellten Textblockabgrenzungen nicht immer unbestreitbar perfekt sind, sondern eine möglichst objektive Entscheidung sind, die im Einzelfall auch anders entschieden werden könnte,

2. ein nach menschlichem Ermessen möglichst genauer Durchschnittswert über alle Textblockzeilen – sehr offensichtliche Ausreisser ausgenommen – gemessen wurde, wobei die Genauigkeit der Messungen mit zunehmender Heterogenität deutlich abnimmt, und

3. für die äußeren Grenzen der Textblöcke nur die Bereiche gewertet wurden, in denen sich mindestens drei Zeilen orthogonal zur Zeilenausrichtung befanden. (Nur so scheint ein Abgleich mit der Vorgabe der in dieser Arbeit aufgestellten Textblockbeschreibung in akzeptabler Weise erfüllbar zu sein.)

4.2. Vorbereitungen

Fast alle Beispieldokumente weisen zunächst schwarze Seitenränder oder dort vorhandene Störungen auf, die nicht nur das hier vorgestellte Verfahren negativ beeinflussen könnten. Es wurden bereits automatische Verfahren entwickelt, die Randstörungen dieser Art beseitigen können (z.B. Ávila u. Lins [2004], [Bukhari u. a., 2012b]); solche Verfahren sind allerdings nicht Teil dieser Arbeit. Um zu vermeiden, das die Auswertung durch Fehler verfälscht wird, die bei der Anwendung solcher automatisierten Verfahren entstehen könnten, wurden händisch die Bildränder entfernt, indem entweder der Bildausschnitt auf einen verkleinerten Bereich reduziert wurde, oder durch Bildbearbeitungsmethoden die Störungen entfernt wurden. Die so aufbereiteten Bilder wurden dann einerseits von dem in dieser Arbeit beschriebenen Verfahren analysiert, andererseits wurden sie zur Bestimmung der Ground Truth verwendet. Ein Ergebnis dieser Vorverarbeitung ist in Abbildung 4.6b zu sehen.

4.3. Erzeugung der Ground Truth

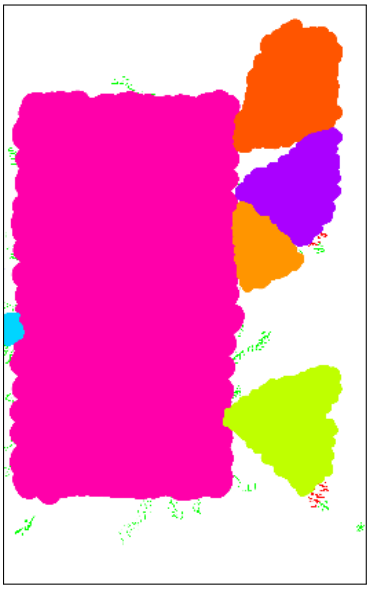
In den folgenden Abschnitten ist die Weiterverarbeitung der Bilder beschrieben, die Abgleiche der Analyseergebnisse auf verschiedenen Untersuchungsebenen ermöglichen soll. In dieser Arbeit werden allerdings lediglich die in Abschnitt 4.3.3 bestimmten Textblockregionen der

(a) In dieser Form sind Manuskriptseitenbilder von Bibliotheken erhältlich.

(b) Das gleiche Bild nach manueller Entfernung der Seitenränder

(c) Farbliche Klassifizierung von Text in Textblock, zwei- und einzeiligen Text, sowie Störungen.

(d) Farbliche Unterscheidung von Schrift unterschiedlicher Textblöcke



(e) Dilatation der Einfärbung zur Bestimmung der Textblockflächen.

(f) Kombination der Abbildungen 4.6b und 4.6e zur Überprüfung des Ergebnisses

Abbildung 4.6.: Die Prozesskette zur Erstellung der Ground Truth

Ground Truth für den Abgleich mit den berechneten Textblockregionen verwendet. Dabei wurde sich an der in Moll u. a. [2008] beschriebenen Methodik orientiert.

4.3.1. Binarisierung und initiale farbliche Klassifizierung

Während für die hier vorgestellte Layoutanalyse keine fehleranfällige Binarisierung nötig ist, sondern in vielen Fällen sogar von unterschiedlichen Helligkeitsverteilungen profitiert wird, ist eine pixelgenaue Auswertungsmöglichkeit für die Überprüfung der Ergebnisse für einen wissenschaftlichen Vergleich von Verfahren üblich. Daher wird jedes Dokument zunächst mit einem manuell optimierten globalen Schwellenwert binarisiert, wodurch sich ein weißer Hintergrund ergibt. Dabei wurden Bildstörungen, die z.B. durch Wurmlöcher oder Flecken entstanden sein könnten, manuell entfernt. Alle verbleibenden schwarzen Pixel wurden nun dunkelblau eingefärbt, soweit sie zu Textblöcken gehören. Interlineare Schriftpixel wurden nach Ermessen eines Schriftunkundigen gelb eingefärbt (daher ohne jeglichen Anspruch auf Perfektion). Zu diesen gehören neben eigenständiger interlinearer Schrift auch diakritische Zeichen, soweit sie deutlich in den Zeilenzwischenraum ragen, Vokalzeichen oder nicht textuelle Markierungen.

Größere Störungen wie Stempel und einzeilige Schrift wurden grün eingefärbt, zweizeilige Schrift rot. Letzteres betrifft auch die Bereiche von Textblöcken, die über den Bereich hinausragen, der von mindestens drei orthogonal zur Zeilenrichtung positionierten Textzeilen umfasst wird.

In der Auswertung dieses Testdatensatzes wird eine Erkennung von nicht blau eingefärbter Schrift als gleichwertiger Fehler gewichtet, durch die unterschiedliche Einfärbung könnte aber später bei anderen Auswertungen eine unterschiedliche Gewichtung von Fehlerursachen eingeführt werden. Eine falsche Erkennung von zweizeiligen Textblöcken könnte so weniger kritisch bewertet werden als eine Erkennung einzeiliger Schrift. Ein Ergebnis dieses Schrittes zeigt Abbildung 4.6c.

4.3.2. Unterschiedliche Einfärbung der Schrift

Während im zuvor beschriebenen Zwischenergebnis noch die Schrift aller gesuchten Textblöcke blau eingefärbt ist, werden nun unterschiedliche Farben für unterschiedliche Textblöcke gewählt. Das hierdurch erreichte Zwischenergebnis wird in Abbildung 4.6d gezeigt. Nun werden für jeden farblich diskriminierten Textblock die durchschnittlichen Zeilenabstände und Zeilenorientierungen von Hand gemessen und im nachfolgenden Schritt gespeichert.

4.3.3. Bestimmung der Textblockregionen

Für alle zuvor ermittelten Bildpixel einer Textblockfarbe wird nun eine morphologische Dilatation angewendet, so dass alle Bildpixel mit der gleichen Farbe einfärbt werden, die im Radius des halben gemessenen durchschnittlichen Zeilenabstandes dieses Textblocks um diese Pixel herum

liegen. In den Bereichen, in denen Textblöcke so eng nebeneinander liegen, dass sich ihre Regionen überschneiden, wurden die Regionsgrenzen mittig von Hand bestimmt. In seltenen Fällen enthalten die so erzeugten Regionen noch Einschlüsse, die abschließend noch aufgefüllt werden. In Abbildung 4.6e ist das Ergebnis dieses Schrittes zu sehen. Alle auf diese Weise gleichartig eingefärbten Bildpixel definieren nun die pixelgenaue Ausdehnung einer Textblockregion. Auf diese Weise kann einerseits ein rechnerisch genauer Abgleich zwischen der Ground Truth und den Ergebnissen des in dieser Arbeit vorgestellten Verfahrens erfolgen, andererseits dürfte eine derart erstellte Textblockregion recht gut den Vorstellungen eines menschlichen Betrachters entsprechen, würde er die äußeren Grenzen eines handgeschriebenen Textblocks markieren.

Eine Sortierung der Textblöcke erfolgt darauf, indem alle Bildpixel zeilenweise und von oben nach unten durchlaufen werden, und durch das erste Auftreffen auf eine Textblockregion die Positionierung in der Sortierung festgelegt wird. Gemäß dieser Sortierung werden dann die im vorherigen Schritt bestimmten Werte der Orientierung und des Zeilenabstandes in einer Datei im maschinenlesbaren Python-numpy-Array Format gespeichert. Diese Regionen können nun mit den Ergebnissen des hier vorgestellten Verfahrens wie im nächsten Abschnitt beschrieben abgeglichen werden.

In Moll u. a. [2008] wurde nahegelegt, die Bildpixel, die eine Region definieren, für eine Auswertung so eng wie möglich um die Schrift zu legen. Das würde also idealerweise nur die zuvor beschriebenen Bildpixel der Schrift selbst umfassen. Durch die Bereitstellung der zuvor beschriebenen Ergebnisse nach Einfärbung der Schrift kann durch den Testdatensatz auch eine derartige Auswertung erfolgen, jedoch sollen für die Ergebnisauswertung in dieser Arbeit zusammenhängende Pixelregionen verglichen werden. Ein Vergleich lediglich der Schriftpixel würde bedeuten, auch die durch das Verfahren berechneten Regionen auf die Schrift zu reduzieren, und damit wiederum eine Binarisierung erfordern, auf die hier ganz bewusst verzichtet werden soll.

4.4. Auswertung der Ergebnisse

Als ein erstes Ergebnis soll ein Vergleich zwischen Textstrukturen an der Position aller Bildpixel jeweils einer Textblockregion der Ground Truth und der an gleicher Position dieser Bildpixel errechneten Strukturwerte dienen. Diese Werte stellen das Teilergebnis vor einer Segmentierung dar.

Als zweite Auswertung erfolgt der Vergleich von Textblockregionen, die einerseites durch die im Verfahren vorgestellte Segmentierung, und andererseits wiederum aus der Ground Truth stammen. Nach der Beschreibung der Berechnungsweise folgt die Anwendung auf den Testdatensatz.

4.4.1. Rechnerische Auswertung ohne Segmentierung

Ein Vergleich der Textblockstruktur ohne Segmentierung bedeutet einen pixelweisen Abgleich der Werte jeder Textblockregion der Ground Truth mit den berechneten Werten an gleicher Position. Diese Auswertung kann sinnvoll sein, da die berechnete Textblocksegmentierung in schwierigen Situationen nicht immer zum gewünschten Ergebnis führt. Weiterhin bedarf eine Fragestellung wie „Gibt es auf einer Manuskriptseite einen Text mit gefragter Struktur?“ nicht unbedingt einer Segmentierung, und daher soll hier eine derartige Auswertung erfolgen, bei der die Korrespondenz über die Pixelposition erfolgt. Diese pixelweise Differenzberechnung kann in angepasster Form auch zum Eingrenzen der Bildpixel mit besonders guter oder besonders schlechter Übereinstimmung, und damit zur Fehleranalyse dienen.

Für die Abweichung zwischen berechneter und manuell gemessener Orientierung einer Textblockregion mit n Pixeln gilt:

$$OrDiffTb = \frac{\sum_{i=0}^{n-1} OrDiff^i}{n} \tag{4.4.1}$$

Dabei beschreibt zunächst der Ausdruck

$$OrDiff^i_{temp} = \left| OrGTTb^i_{(x,y)} - OrBer_{(x,y)} \right| \tag{4.4.2}$$

den Betrag der Differenz aus der Orientierung der Ground Truth jedes Pixels der Textblockregion i ($OrGTTb^i_{(x,y)}$) mit der an gleicher Bildposition berechneten Orientierung $OrBer_{(x,y)}$.

Aufgrund der Zyklizität bedarf diese Differenz noch folgender Normierung:

$$OrDiff^i = \begin{cases} OrDiff^i_{temp} & \text{if} \quad OrDiff^i_{temp} \leq 90 \\ 180 - OrDiff^i_{temp} & \text{else} \end{cases} \tag{4.4.3}$$

Die Orientierungsdifferenz $OrDiffTb$ beschreibt daher den Durchschnittswert zwischen dem Betrag der Differenz der Orientierung eines jeden Bildpixels i aus der als Ground Truth bestimmten Textblockregion ($OrGTTb^i_{(x,y)}$) und der Orientierung, die an gleicher Bildposition berechnet wurde ($OrBer_{(x,y)}$).

Die Abweichung des Zeilenabstandes wird in ähnlicher Weise berechnet:

$$ZADiffTb = \frac{\sum_{i=0}^{n-1} \frac{\left|ZaGTTb^{i}_{(x,y)} - ZaBer_{(x,y)}\right|}{\max\left(ZaGTTb^{i}_{(x,y)}, ZaBer_{(x,y)}\right)}}{n} \tag{4.4.4}$$

$ZADiffTb$ beschreibt damit für jeden Textblock i die prozentuale Abweichung zwischen Ground Truth ($ZaGTTb^{i}_{(x,y)}$) und berechnetem Zeilenabstand ($ZaBer_{(x,y)}$) an gleicher Pixelposition (x, y).

Ab Anhang A.1 werden die ermittelten Ergebnisse für jeden Textblock einzeln aufgelistet, eine Ergebniszusammenfassung wird ab Abschnitt 4.5 dargestellt.

4.4.2. Rechnerische Auswertung mit Segmentierung

Hier wird zunächst eine $(I \times J)$-Matrix erstellt, die die Güte der Überdeckung aller Textblockregionen der Ground Truth ($GTTb$) $i \in I$ mit jeweils n Pixeln zu allen Textblockregionen $j \in J$ der berechneten Segmentierung ($SegTb$) von m Pixeln ermittelt.

Jedes Element der Matrix wird wie folgt berechnet:

$$\ddot{U}berdeckung_{i,j} = \frac{\sum_{i=0}^{n-1} \begin{cases} 1 & \text{if} \quad GTTb^{i}_{(x,y)} = SegTb^{j}_{(x,y)} \\ 0 & \text{else} \end{cases}}{\max(n, m)} \tag{4.4.5}$$

Gebündelt wird das Ergebnis zunächst in vier Überdeckungsklassen ($\ddot{U}K$) von >90%, >80%, >70% bis >50% Überdeckung.

Zusätzlich werden Regionen der Ground Truth als nicht erkannt gewertet, soweit die Summe der Schnittmengenpixel der erkannten Regionen (aufsummiert über alle erkannten Regionen) weniger als 25% der Ground Truth Regionsfläche entspricht.

Als fehlerhafte Erkennung wird umgekehrt bewertet, falls die Schnittmenge zwischen einer oder mehreren Regionen der Ground Truth und einer segmentierten Region weniger als 25% der Pixelfläche derselben entspricht.

Für jede der vier *ÜK*s wird daraufhin wieder die Differenz der Orientierung und des Zeilenabstandes berechnet, nun aber mit den durchschnittlichen Werten der berechneten, segmentierten Textblockregionen, deren Bestimmung bereits auf Seite 70 beschrieben wurde:

$$OrDiff\ddot{U}K = \frac{\sum_{s=0}^{O-1} \begin{cases} OrTbDiff & \text{if} \quad OrTbDiff \leq 90 \\ 180 - OrTbDiff & \text{else} \end{cases}}{O} \tag{4.4.6}$$

wobei

$$OrTbDiff = |OrGTTb_s - OrSegTb_s| \tag{4.4.7}$$

Die Berechnung der durchschnittliche Orientierungsdifferenz (*OrDiffÜK*) erfolgt über alle O Textblockregionspaare, deren Überdeckung der jeweiligen *ÜK* zugeordnet sind. Nach gleichem Schema erfolgt die Berechnung der Zeilenabstandsdifferenzen dieser Textblockregionspaare.

In den tabellarischen Aufstellungen, die ab Anhang A.1 zu finden sind, wird in der Zeile „Flächenanteil“ der Quotient aus der aufsummierten Anzahl der Bildpixel aller Schnittmengen der jeweiligen *ÜK*, zur Pixelsumme aller Textblockregionen der Ground Truth (P_{GT}) angegeben. Hier schließt jede *ÜK* das Ergebnis der jeweils strengeren *ÜK* mit ein. Auch die Summen der Pixel aller nicht erkannten Regionen der Ground Truth, sowie der falsch erkannten Segmentierungen werden durch P_{GT} normiert.

Die Zeile „Korrespondenzen“ beschreibt die Anzahl der Textblockregionspaare, normiert durch die Anzahl der Textblockregionen der Ground Truth.

In den Zeilen „ZA-Differenz“ sowie „O-Differenz“ werden dagegen die oben beschriebenen Zeilenabstandsdifferenzen sowie Orientierungsdifferenzen für jede *ÜK* einzeln, d.h. nicht kumuliert, angegeben.

Ein perfektes Ergebnis würde in der Tabelle daher in der ersten Spalte (Überdeckungsklasse >90%) bereits einen Flächenanteil von 100% zeigen, da 100% aller Pixel aller Regionspaare übereinstimmen würden. Damit würden sich auch alle Regionen der Ground Truth zu genau einer Region der berechneten Textblöcke zuordnen lassen, womit 100% Korrespondenzen erfüllt wären. Weiterhin sind geringstmögliche Differenzen der Zeilenstruktur wünschenswert.

4.5. Auswertung des Testdatensatzes

4.5.1. Auswertung des Testdatensatzes Berlin

Ein Testdatensatz aus 15 islamischen Manuskriptseiten (Quelle: Staatsbibliothek zu Berlin) soll im Folgenden kurz als „Berlin“ bezeichnet werden. Er weist auf jeder Manuskriptseite einen

Tabelle 4.1.: Vergleich zwischen Ground Truth und berechneten Werten ohne Segmentierung der 15 Manuskriptseiten „Berlin“.

Berlin	Differenz von Orientierung [°], Zeilenabstand [%]			
	Haupttextbereich		Randtextbereiche	
Manuskript	Orientierung	Zeilenabstand	Orientierung	Zeilenabstand
1	2,27	2,38%	9,42	12,10%
2	3,76	4,59%	6,63	9,57%
3	0,49	2,37%	10,45	13,04%
4	0,98	1,53%	6,34	12,72%
5	2,45	5,72%	7,57	10,21%
6	1,57	4,65%	10,81	16,01%
7	0,73	2,45%	10,52	13,19%
8	0,75	5,98%	22,07	28,33%
9	0,07	3,53%	20,72	26,20%
10	0,46	3,16%	8,90	7,85%
11	1,48	4,90%	10,73	13,08%
12	2,20	9,30%	5,92	11,60%
13	1,81	6,22%	12,00	11,75%
14	0,78	4,51%	14,53	20,85%
15	0,99	1,56%	12,99	19,56%

Haupttextblock und bis zu 15 Randtextblöcke auf. Diese Texte an den Seitenrändern haben maximal die Schriftgröße des Haupttextes, können aber Zeilenabstände von weniger als einem Drittel des Haupttextzeilenabstandes annehmen. Einige Seiten weisen viele besonders kleine Randtextblöcke auf, deren genaue Erkennung durch die enge Lage zu benachbarten größeren Textblöcken und durch ihre wenig homogene Schreibweise erschwert wird.

Tabelle 4.1 zeigt die Zusammenfassung der Auswertung des Testdatensatzes Berlin (ohne Segmentierung), in der der Haupttextblock jeweils einzeln ausgewertet wurde, das Ergebnis der Randtextblöcke dagegen zusammengefasst wurde. Diese Zusammenfassung stellt den Mittelwert gemittelt über die Anzahl der Randtextblöcke dar, unabhängig von ihrer Größe. Dadurch ist auch die recht große Abweichung der Randtextbereiche von Manuskriptseite 8 zu erklären: Hier gibt es vier besonders kleine Randtextblöcke, deren Struktur nicht hinreichend erkannt wurde, da ihr schwaches Struktursignal von den Werten der größeren benachbarten Textblöcke überlagert wurde. Obwohl auf Manuskriptseite 9 keine derart kleinen Randtextblöcke vorhanden sind, ist auch hier das Signal des Haupttextblocks so stark, dass es deutlich das Signal von zwei Randtextblöcken überlagert, was in hohen Abweichungen resultiert. Insgesamt dürften Abweichungen der Orientierung von unter 5°, sowie des Zeilenabstandes von unter 5% im Rahmen dieser Analyse als geradezu perfektes Ergebnis zu bewerten sein, Abweichungen von unter 10° bzw. 10% Zeilenabstand noch immer als gut bis sehr gut. Da die Haupttextblöcke durch sehr viele Textzeilen charakterisiert werden, und sie sich dazu in der Regel durch eine weitgehend einheitliche Struktur auszeichnen, kommen die Stärken des Verfahrens durch die Nutzung des oberen Frequenzbandes voll zum Tragen, interlineare Störungen spielen erst in den Größenordnungen eine Rolle, wie sie beim folgenden Testdatensatz auf zwei Manuskriptseiten vorkommen.

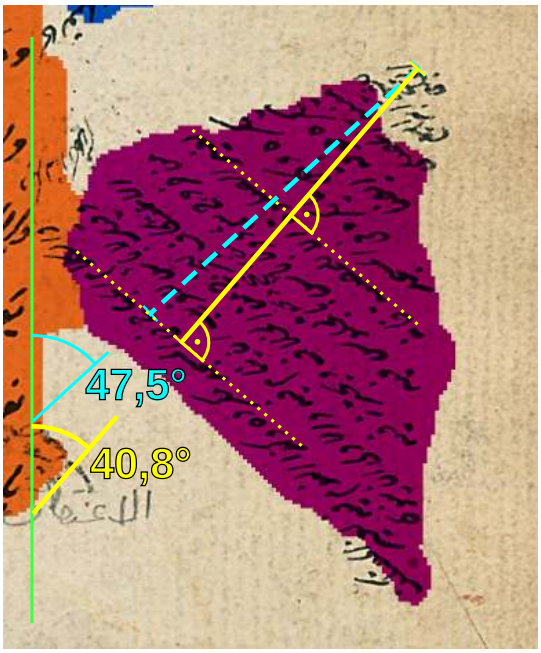


Abbildung 4.7.: Für diesen Textblock wurde ursprünglich händisch eine Zeilenorientierung von 47.5° vermessen (hellblau). Hier fiel eine deutliche Abweichung zur errechneten Zeilenorientierung (gelb) auf.

Die deutlich größeren Abweichungen der Randtextblöcke sind dagegen dadurch zu erklären, dass ihre geringe Signalstärke von umliegenden Textblöcken überlagert wird, und sie so häufig nicht innerhalb ihrer kompletten Fläche durch das von ihrer eigenen Textstruktur verursachte Signal interpretiert werden. Neben einem zu schwachen Signal, verursacht durch zu geringe Textblockgröße und manchmal durch ungünstige Kontrastverhältnisse, spielt auch die Varianz der Struktur eine entscheidende Rolle für die Ausprägung der Abweichungen.

Eine detaillierte Auswertung mit Auflistung aller Textblöcke wird in Anhang A.1 gegeben. Im Anschluss daran wird die Auswertung der Segmentierung vorgestellt.

Die Auswertung der Segmentierung macht den zuvor bemerkten Einfluss der unterschiedlichen Signalstärke benachbarter Textblöcke deutlich. Auf Kosten einer dadurch nicht immer optimalen Segmentierung enthalten die Signale der Pixelschnittmengen zueinander korrespondierender Textblöcke nicht mehr die Störeinflüsse der benachbarten Textblöcke, die Abweichungen der Strukturanalyse werden hier deutlich geringer. Ein recht hoher Unterschied kann auch hier durch eine hohe Varianz der Struktur, durch eine fehlerhafte Segmentierung, oder auch fehlerhafte Ground Truth Daten verursacht werden. Ein Beispiel für letzteres zeigt Abbildung 4.7. Eine unerklärliche Abweichung fiel hier auf, die Ground Truth Daten wurden darauf korrigiert. Noch kleinere Textblöcke bedeuten allerdings tendenziell deutlich weniger eindeutige Messungen.

Abbildungen von Manuskriptseiten, Ground Truth und berechneter Segmentierungsergebnisse finden sich in Anhang A.3.

Tabelle 4.2.: Vergleich zwischen Ground Truth und berechneten Werten ohne Segmentierung der 15 Manuskriptseiten „Leipzig“.

Leipzig	Differenz von Orientierung [°], Zeilenabstand [%]			
	Haupttextbereich		Randtextbereiche	
Manuskript	Orientierung	Zeilenabstand	Orientierung	Zeilenabstand
1	0,25	3,87%	6,98	8,14%
2	0,09	3,71%	10,95	9,58%
3	0,15	3,12%	8,08	9,87%
4	0,15	3,68%	8,24	8,04%
5	0,20	3,65%	10,46	8,61%
6	0,15	3,63%	12,84	12,53%
7	0,55	2,03%	11,22	8,46%
8	1,54	0,66%	12,10	5,68%
9	0,04	2,54%	5,43	6,02%
10	1,52	9,12%	2,71	6,26%
11	8,10	12,56%	9,83	13,08%
12	5,95	17,88%	10,25	15,79%
13	0,23	1,89%	6,38	7,40%
14	0,28	3,30%	18,33	6,89%
15	1,48	4,74%	12,26	7,78%

4.5.2. Auswertung des Testdatensatzes Leipzig

Auch der Testdatensatz „Leipzig“ (Quelle: Orientalisches Institut, Universität Leipzig) besteht aus 15 islamischen Manuskriptseiten mit jeweils einem Haupttextblock und bis zu 15 Randtextblöcken.

Bei einem Manuskript, welches mit mehreren Seiten am Anfang dieses Testdatensatzes vertreten ist, sind fast alle Randtextblöcke besonders regelmäßig geschrieben. Der Zeilenabstand ist bei allen diesen Randtexten sehr ähnlich, der Kontrast zwischen Text und Hintergrund ist hoch und die Schriftfarbe ist über alle Textblöcke gleichmäßig dunkel.

Im hinteren Teil des Testdatensatzes befinden sich zwei Manuskriptseiten, deren Haupttextblock neben den Schriftfarben Schwarz und Rot auch Blau und Gold enthält. Neben der Annahme eines durchgehenden Haupttextblockes, die durch die Umrandung unterstützt wird, wären auch andere Interpretationen von mehreren Haupttextblöcken möglich. Auf weiteren Manuskriptseiten sind größere Randtextblöcke enthalten, die eine beträchtliche Veränderung der Zeilenorientierung innerhalb dieser Textblöcke von über 60° aufweisen.

Tabelle 4.2 fasst die Ergebnisse zusammen, Ergebnisse im Detail finden sich im Anhang B.

Auffällig sind die Abweichungen der Manuskripte 11 und 12 beim Vergleich der Struktur der Haupttextblöcke. Die Ursachen hierfür liegen zum einen in den unterschiedlichen Schriftfarben begründet, da hellere Schrift einen geringeren Kontast zum Hintergrund bewirkt, zum anderen in größeren interlinearen Textregionen, die neben dem Signal der Hauptzeilenstruktur für

Störungen sorgen. Hier treten beide Phänomene an gleicher Stelle auf. Abgesehen von diesen sind die über die Haupttextblöcke gemessenen Ergebnisse ohne Segmentierung, wie auch schon beim Testdatensatz Berlin, als nahezu perfekt zu bewerten. Die zuvor erwähnten besonders homogenen Randtextblöcke machen sich auch in besonders geringen Differenzen des Abgleichs bemerkbar, sowohl vor als auch nach erfolgter Segmentierung. Insgesamt ist der Testdatensatz Leipzig etwas homogener, daher werden bessere Messergebnisse erzielt.

4.5.3. Auswertung des Testdatensatzes Greek

Um die Anwendbarkeit der Layoutanalyse auch auf Manuskripte europäischer Schreibsysteme zu verdeutlichen, wurden auch 15 griechische Manuskripte ausgewählt. Diese Manuskripte aus den Greek State Archives werden im Folgenden kurz als „Greek“ bezeichnet, und liegen oftmals nur in Graustufenrepräsentation vor, teilweise sind sie in einem schlechten Zustand mit dunklen Rändern und Flecken. Nicht alle Textblöcke sind horizontal ausgerichtet, und an einigen Stellen finden sich mehrzeilige Unterschriften, die sich als Textblock interpretieren lassen oder auch nicht. Mit bis zu vier Textblöcken weist dieser Teil des Testdatensatzes deutlich weniger Textblöcke pro Seite als die zuvor vorgestellten Testdatensätze auf. In Tabelle 4.3 sind teilweise mehrere Textblöcke unter der Überschrift „Haupttextbereich“ zusammengefasst, da eine eindeutige Auftrennung, wie in den zuvor behandelten Manuskripten des arabischen Schreibsystems – in einen Haupttextblock und weitere Randtexte – hier nicht möglich ist. Auch für diesen Testdatensatz sind die Einzelergebnisse, die Ergebnisse der Segmentierung und Bilder ab Anhang C zu finden.

Die Auswertung zeigt nicht nur die Anwendbarkeit auf Manuskripte eines weiteren Schreibsystems, sondern auch – ohne Änderung von Parametern – die Übertragbarkeit auf Manuskripte ohne eindeutigen einzelnen Haupttextblock. Die erste Manuskriptseite des Testdatensatzes ist eine Doppelseite, auf der Haupttextbereiche nicht mittig positioniert sind.

Dass einzelne Werte aus der zusammenfassenden Tabelle 4.3 auf ein schlechteres Ergebnis als in den vorigen Testsätzen hindeuten, liegt häufig an der Berechnung des Duchschnittswertes. Da hier insgesamt deutlich weniger Textblöcke pro Seite vorhanden sind, wird der Durchschnittswert stärker von einzelnen Messfehlern beeinflusst. Die Unterschiede des Vergleichs der Textblockstrukturen nach erfolgter Segmentierung zeigen dagegen ein vergleichbar gutes Ergebnis wie bei den zuvor betrachteten Testdatensätzen.

Tabelle 4.3.: Vergleich zwischen Ground Truth und berechneten Werten ohne Segmentierung der 15 Manuskriptseiten „Greek“.

Greek	Differenz von Orientierung [°], Zeilenabstand [%]			
	Haupttextbereich		Randtextbereiche	
Manuskript	Orientierung	Zeilenabstand	Orientierung	Zeilenabstand
1	1,34	7,98%	-	-
2	0,48	2,86%	-	-
3	0,17	3,93%	20,04	28,08%
4	0,99	5,93%	-	-
5	0,45	7,38%	1,79	11,59%
6	2,26	3,58%	3,65	8,14%
7	0,85	4,25%	3,92	25,19%
8	1,02	3,56%	10,23	23,53%
9	1,86	5,43%	5,70	13,08%
10	0,53	8,98%	1,79	2,39%
11	0,16	6,61%	3,82	12,06%
12	2,30	6,66%	1,93	5,91%
13	4,88	18,85%	0,24	4,76%
14	1,22	6,25%	11,82	22,49%
15	4,42	5,43%	5,35	18,52%

4.5.4. Auswertung weiterer Manuskriptseiten

Dem Testdatensatz wurden weitere fünf Manuskriptseiten hinzugefügt, die deutlicher die Grenzen dieses Layout-Analyseverfahrens aufzeigen können (Abbildung 4.8). Dieser Teil des Testdatensatzes erhält den Namen „Varia“.

Bei den ersten beiden Manuskripten handelt sich um Manuskriptseiten aus zwei alten Koranen, die ursprünglich aus der Region Borno stammen, einem Bundesstaat im Nordosten von Nigeria. Der Haupttext sowie einige Kommentare sind auf arabisch geschrieben, weitere Glossen dagegen in altem Kanembu, einer heute nicht mehr üblichen Sprache aus dem Norden Nigerias. Neben dunklen Rändern, Flecken und Löchern stellen vor allem die stark unterschiedlichen Kontraste in Form einer teilweise deutlich ausgeblichenen Schrift eine große Herausforderung dar. Dazu kommen eine Vielzahl an Randtextblöcken in unterschiedlichen Schriftgrößen und benachbarte Textblöcke ähnlicher oder gleichartiger Orientierung.

Ein Sanskritmanuskript enthält mittig ein Bild, davon abgesehen aber lediglich einen sehr regelmäßigen Textblock. Eine weitere Manuskriptseite zeigt ein Poster mit vier Textblöcken in vier verschiedenen Schreibsystemen.

Schließlich gibt es noch den Ausschnitt des Anfangs einer japanischen Schriftrolle. Auf der rechten Seite gibt es in schwacher schwarzer Tinte einen Textblock auf hellem Hintergrund, links ist der Anfang des Haupttextes zu sehen, in heller Schrift auf dunklem Hintergrund, dabei sehr regelmäßig mit sehr gutem Kontrast und zusätzlicher Linierung.

Abbildung 4.8.: Die Manuskripte des Testdatensatzes „Varia“.

Tabelle 4.4.: Vergleich zwischen Ground Truth und berechneten Werten ohne Segmentierung der 5 Manuskriptseiten „Varia“.

Weitere	Differenz von Orientierung [°], Zeilenabstand [%]			
	Haupttextbereich		Randtextbereiche	
Manuskript	Orientierung	Zeilenabstand	Orientierung	Zeilenabstand
1	2,97	9,75%	10,97	22,58%
2	3,35	5,44%	30,96	46,81%
3	0,24	4,09%	-	-
4	1,17	5,45%	38,71	63,86%
5	1,48	1,90%	5,67	18,75%

Auch für diesen Testdatensatz gibt es neben der Zusammenfassung in Tabelle 4.4 in Anhang D die detaillierte Auswertung in Tabellen- und Bildform.

Beeinflusst durch die blasse Schrift der Randtextbereiche beider Borno-Manuskripte entspricht das ausgewertete Signal häufig nur unzureichend dem Textmuster, die Ergebnisse der Auswertung sind daher deutlich schlechter als in den anderen Testsätzen. Gleiches gilt auch für die Segmentierung, hier liegen zudem Textblöcke mit sehr ähnlicher Zeilenstruktur dicht beeinander, eine Auftrennung findet dann nicht statt.

Die Besonderheit des Sanskritmanuskriptes ist das mittig vorhandene Bildelement, welches recht gut vom Text segmentiert wird und dem Hintergrund zugeordnet wird. Da das Verfahren Textbereiche an kontrastreichen zeilenförmigen periodischen Mustern erkennt, sind vorhandene Bildelemente unkritisch, soweit sie keine dem Suchmuster ähnlichen Eigenschaften zeigen. Eine genaue Auswertung, inwieweit das auf in Manuskriptseiten vorhandene Bildelemente zutrifft, fand aber nicht statt. Da sich einige der im Abschnitt 2.2 genannten Arbeiten explizit auf das Entfernen von Bildelementen und Ornamenten konzentrieren, wird der Weg des Entfernens dieser Artefakte im Feld der Vorverarbeitung empfohlen.

Beim vierten Manuskript liegen die speziellen Eigenschaften offensichtlich im Vorkommen von vier verschiedenen Schreibsystemen auf einem Dokument. Besonders ist aber auch das Format, das Original ist etwa 74cm breit und 52cm hoch, wodurch auch die relativ kleine Schrift der linken Spalte lesbar bleibt. In der Zusammenfassung von Tabelle 4.4 wird der Durchschnitt der vier Spalten unter Haupttextbereich geführt, die dreizeilige Überschrift im arabischen Schreibsystem dagegen als Randtext. Die drei Zeilen sind allerdings so eng untereinander geschrieben und ineinander verflochten, dass die Erkennung der Zeilenstruktur hier fehlschlägt.

Keine Herausforderung für dieses Verfahren ist die gleichzeitig auftretende helle Schrift auf dunklem Hintergrund und umgekehrt auf der fünften Manuskriptseite. Auch die zusätzliche Liniierung des linken Textblocks wirkt sich nicht negativ auf das berechnete Ergebnis aus. Einzig die Segmentierung von Textbereichen zu Hintergrund anhand der vorhandenen Signalstärke führt zur Erkennung eines nicht vorhandenen Textblocks in der Mitte des Dokuments. Ursächlich dafür ist der sehr starke Kontrast zwischen der dunklen linken Seite und der hellen rechten.

Die dadurch hervorgerufene Antwort ist selbst im idealen Frequenzband noch ähnlich stark wie die des rechten Textblocks mit wenigen Zeilen bei sehr blasser Schrift.

4.6. Fazit

Zunächst wurde in diesem Kapitel die Erzeugung eines pixelgenauen Ground-Truth-Datensatzes von 50 Manuskriptseiten vorgestellt. Anschließend fand ein Abgleich mit den Ergebnissen des in dieser Arbeit entwickelten Verfahrens statt. Dabei wurde deutlich, dass der Unterschied zwischen den von Hand vermessenen und den rechnerisch ermittelten Textstrukturwerten in den Fällen sehr gering ausfällt, in denen Textblöcke kompakt sind, aus vielen Zeilen bestehen, einen guten Kontrast zwischen Schriftfarbe und Hintergrund bieten und eine homogene Erscheinung haben. Dann wird eine Abweichung zwischen manuell vermessenen Zeilenabständen und deren berechneten Werten von weniger als 5% ermittelt. Dass die Abweichung der Orientierung der Haupttextblöcke häufig weniger als 1° ergibt, liegt an der Art der diskreten Aufteilung des Frequenzspektrums, welches eine Erkennung von exakt horizontalen oder exakt vertikalen Orientierungen stark begünstigt. Für den allgemeinen Fall davon abweichender Orientierung sind aber unter sehr guten Bedingungen Abweichungen von weniger als 4° realistisch.

Je stärker von diesen nahezu idealen Bedingungen abgewichen wurde, umso stärker stieg die Fehlerrate sowohl bezüglich der Textstrukturerkennung als auch der Segmentierung. Dabei sollte beachtet werden, dass auch die händische Bestimmung von Zeilenabstand und Orientierung bei Textblöcken mit wenigen, unregelmäßigen Zeilen eine Herausforderung ist und eine eindeutige und präzise Vermessung oft unmöglich ist.

Soweit ein kontraststarkes homogenes Zeilenmuster vorliegt, ist das Verfahren robust gegenüber unterschiedlichen Schriftfarben, interlinearer Schrift oder Störungen wie Flecken oder Stempeln.

5. Fazit und Ausblick

5.1. Fazit

In dieser Arbeit wurde eine experimentelle Layoutanalyse auf Manuskriptdigitalisaten vorgestellt. Eine Zielsetzung war, dass möglichst alle Projektgruppen des SFB 950 von dem Ergebnis profitieren können. Nach einer Durchsicht des aktuellen Forschungsstands musste festgestellt werden, dass kein bereits existierendes Verfahren alle Vorraussetzungen der Zielvorgaben erfüllt. Nur durch einen generischen Ansatz kann eine breite Anwendbarkeit erreicht werden. Zielsetzungen waren, dass keine Einschränkungen bezüglich

- des Schreibsystems,
- des Dokumentenformats,
- der Farbgestaltung,
- der Auflösung
- sowie keine Vorannahmen bezüglich der Schriftgröße oder des Layouts

getroffen werden. Zudem sollte weder eine Binarisierung noch eine Trainingsphase nötig sein.

Die offensichtliche Gemeinsamkeit zwischen allen untersuchten Manuskripten stellt das periodische Muster von Textzeilen dar. Da die Gabor-Transformation als theoretisch fundierter Ansatz eine Transformation eines örtlich begrenzten Bereichs eines Manuskripts in den Frequenzraum gestattet, und damit eine lokale Auswertung periodischer Muster erlaubt, wurden auf dieser Basis ein Verfahren zur Layoutanalyse entwickelt.

Im Gegensatz zu vielen bisherigen Ansätzen ist dieses Verfahren nicht für ein bestimmtes Schreibsystem entworfen und kann auf Textbereiche beliebiger Größe, äußerer Abmessungen und Farbgebungen angewandt werden, soweit zwischen der Schrift und dem Hintergrund ein ausreichend großer Helligkeitskontrast gegeben ist.

Es konnte festgestellt werden, dass eine Gauß-Fensterfunktion der Größe von 121×121 Pixeln einen guten Kompromiss zwischen Genauigkeit und rechnerischer Effizienz darstellt. Das Frequenzspektrum wurde in drei Bänder aufgeteilt, und der Frequenzbereich zwischen den Wellenzahlen 8 und $9 \times \sqrt{(}2)$ als ideal bestimmt, um einerseits eine hinreichend gute Ortsgenauigkeit zu behalten, andererseits robust gegenüber wenig periodischen Störeinflüssen zu sein. Zur Unterscheidung zwischen Textbereichen und Hintergrund wird zusätzlich ein unteres Frequenzband im Bereich der Wellenzahlen von $4 - 9$ ausgewertet, um von der besseren Ortsgenauigkeit

dieses Frequenzbereichs zu profitieren. Zusätzlich wird von der besseren Frequenzauflösung des Frequenzbereichs ab der Wellenzahl $8 \times \sqrt{2}$ profitiert, soweit ein homogener Textbereich ausreichender Größe zur Verfügung steht.

Durch die Anwendung einer Auflösungspyramide sind keine Voraussetzungen bezüglich des Dokumentenformats oder der Auflösung nötig. Aus rechnerischer Effizienz werden allerdings folgende Grundeinstellungen vorgenommen, um innerhalb der sinnvoll erscheinenden Auflösungstufen zu verbleiben: Es wird davon ausgegangen, dass keine Textbereiche auf dem Dokument vorhanden sind, deren Zeilenabstand weniger als 10 Pixeln entspricht (da ein derartig kleiner handschriftlicher Text kaum lesbar sein dürfte). Die äußeren Abmessungen des Dokuments sollten auch nicht kleiner sein als ein Achtel des maximal vorkommenden Zeilenabstandes. Falls Dokumente verarbeitet werden sollten, auf die diese Bedingungen nicht zutreffen, sind die Parameter der Auflösungsgrenzen dieses Verfahrens entsprechend anzupassen.

Eine Segmentierung in einzelne Textblöcke ist möglich, soweit Textblöcke einen Abstand von mindestens 2,5 Zeilen zueinander haben. Soweit Textblöcke eine zueinander unterschiedliche Zeilenstruktur zeigen, ist eine Segmentierung auch bei dichter benachbarten Textblöcken möglich. Dazu wird zunächst eine Wasserscheidensegmentierung auf einem Gradientenbild durchgeführt, welches die lokalen Unterschiede der Zeilenstruktur von Textbereichen abbildet. Diese Art der Segmentierung stellt zunächst eine Übersegmentierung dar, die sukzessive durch Entfernen von überflüssigen Kanten aufgelöst wird. Zur Entscheidung, welche Kanten gelöscht werden sollen, wurde eine Support Vector Machine auf etwa 10.000 Kanten trainiert. Der Support Vector Machine wird dafür eine Datenbasis mit 30 Merkmalen für jede Kante zur Verfügung gestellt, die neben lokalen Eigenschaften auch die Eigenschaften der Flächen enthält, deren Vereinigung überprüft werden soll. Sobald keine als nicht notwendig erachtete Kante mehr vorhanden ist, sollten alle Flächen jeweils einem Textblock entsprechen, deren Eigenschaften nun ausgegeben werden könnten.

Dieses experimentell erstellte Konzept wurde durch einen Testdatensatz validiert. Da kein Datensatz zur Verfügung stand, der die gegebenen Herausforderungen bezüglich der unterschiedlichen Schreibsysteme, Kontrastverhältnisse, Textblockanordnungen etc. widerspiegeln konnte, wurde eigens ein Testdatensatz anhand von 50 Manuskriptbildern angefertigt. Anhand dieser Ground Truth wurde das Verfahren validiert, und die Ergebnisse wurden detailliert dargestellt.

Hierbei konnte unter anderem festgestellt werden, dass unter idealen Voraussetzungen von Textblöcken mit homogener Zeilenstruktur und mindestens zehn Zeilen der durchschnittliche Erkennungsfehler bei etwa 5% des jeweiligen Zeilenabstandes und der durchschnittliche Fehler der Zeilenorientierung bei etwa 4° liegt. Das Verfahren zeigt sich dabei robust gegenüber Rissen, kleineren Löchern oder Wasserflecken.

Das in dieser Arbeit vorgestellte Verfahren kann sowohl mit dunkler Schrift auf hellem Hintergrund, als auch mit heller Schrift auf dunklem Hintergrund umgehen. Es ist direkt auf heterogene Datenbestände anwendbar, eine Trainingsphase auf zuvor angefertigten Dokumenten ist nicht notwendig, wie sie bei anderen Verfahren zur Anpassung auf einen speziellen Dokumententyp benötigt wird.

Die Segmentierung von Textblöcken kann zu Fehlern führen, wenn Textblöcke nicht im Abstand von mindestens 2,5 Zeilen zueinander liegen und die Zeilenstruktur benachbarter Textblöcke zueinander nicht unähnlich genug ist. Hier sollte eine Differenz der Zeilenabstände von mindestens etwa 20% oder eine Differenz der Zeilenorientierung von mindestens 10° eingehalten werden. In Fällen von besonders kleinen, wenig kompakten, oder bezüglich der Zeilenstruktur besonders inhomogenen Textblöcken können fehlerhafte Segmentierungen nicht ausgeschlossen werden.

Zusammenfassend kann festgehalten werden, dass das in dieser Arbeit entwickelte Verfahren die gestellten Anforderungen weitgehend erfüllt. Der generische Ansatz auf Basis der Gabor-Transformation erlaubt die Erkennung von periodischen Zeilenmustern jeder Art und ist gleichzeitig robust gegen Störungen, soweit diese nicht zu kontraststark oder selbst periodisch sind. Es konnte beispielhaft demonstriert werden, dass die lokale Erkennung der Zeilenstruktur zumindest ähnlich präzise sein kann wie eine Vermessung von Hand. Die Validierung zeigt die Grenzen für eine Segmentierung in einzelne Textblöcke.

5.2. Ausblick

Die folgenden Aspekte sind für die Weiterentwicklung dieses Ansatzes denkbar:

- Präsentation in einer graphischen Nutzungsoberfläche
- Integration der Verfahrens in eine Verarbeitungskette
- Weitere Präzisierung der Ergebnisse

Der nachfolgende Schritt nach Abschluss dieser Arbeit sollte die Integration dieses Ansatzes in die Anwendungsoberfläche des SFB 950 sein, die in der zweiten Projektphase des Teilprojektes Informatik entwickelt wird (Rajan u. Stiehl [2018]). Zur Veranschaulichung der Ergebnisse könnte die Benutzungsoberfläche neben den durchschnittlichen Strukturwerten von Textblöcken auch Teilbereiche innerhalb der Textblöcke hervorheben, die beispielsweise nahe der minimalen oder maximalen Orientierung der Textblöcke liegen. In gleicher Weise könnten Bereiche hervorgehoben werden, die durch eine ungewöhnlich hohe Varianz des Zeilenabstandes auffallen. Über einen Cursor könnte eine Position innerhalb eines Textblocks gewählt werden und die dort lokal bestimmten Werte könnten ausgegeben werden. Auch einzelne Seiten eines Manuskripts könnten von Interesse sein, soweit sie sich durch besondere Regelmäßigkeit oder Unregelmäßigkeit auszeichnen, und diese könnten dann in Form einer geeigneten Statistik erfasst werden. Es gilt mit Manuskriptforscherinnen und -forschern abzuklären, welche von diesen zur Verfügung stehenden Merkmalen in welcher Form und Präzision dabei für ihre Arbeit nützlich sind.

Eine weitere Einsatzmöglichkeit dieses Verfahrens ist die Nutzung der lokal bestimmten Werte der Zeilenstruktur innerhalb einer Verarbeitungskette zur weiteren Manuskriptanalyse. Dass der zuvor bestimmte Zeilenabstand zur Parametrierung eines anisotropen Glättungsfilters genutzt

wird und auf dem Ergebnis eine Zeilensegmentierung stattfinden kann, konnte z.B. in Herzog [2007] gezeigt werden.

Auch für segmentierungsfreie „Spotting“-Verfahren bieten die Ausgaben der Zeilenstrukturwerte eine nützliche Datengrundlage. Das gilt sowohl für den Fall der Strichmustersuche, die die (digitale) Paläographie in dem Bereich unterstützen möchte, in dem eine größere Anzahl an gleichartigen Strichmustern zur späteren Analyse gesammelt werden soll (Herzog u. a. [2013]), als auch für „Keyword-Spotting“-Verfahren. Beide Ansätze müssen häufig auf die Dokumente bzgl. der erwarteten Zeilenstruktur parametriert werden. Ansätze, die ohne Vorgaben bezüglich der erwarteten Worthöhe auskommen, zeigen dagegen häufig eine geringere Effizienz oder Genauigkeit, da ihnen die Dimensionierung der Zeichenhöhe komplett unbekannt ist (Giotis u. a. [2017]). In beiden Fällen hilft die aus der Layoutanalyse erhaltene Zeilenstruktur, die als a-priori-Wissen zum Zeilenabstand, und damit zur in etwa erwarteten Worthöhe genutzt werden kann.

Weiterhin sollte der Ansatz daraufhin überprüft werden, inwieweit mit dem spezielleren Methodenrepertoire der Gabor-Filterbänke oder Gabor-Wavelets eine genauere Segmentierung der äußeren Umrisse von Textblöcken oder eine verbesserte Segmentierung zwischen zueinander strukturell ähnlichen Textblöcken möglich ist.

Neben dem Versuch, die Segmentierung zu verbessern, könnte auch die festgestellte Varianz der Textblöcke daraufhin untersucht werden, inwieweit sich dieser Wert als Konfidenz für eine korrekte Segmentierung eignet.

In den letzten Jahren wurden durch Anwendung neuronaler Netze auch im Hinblick auf die Layoutanalyse von Manuskripten beachtliche Erfolge erzielt (z.B. [Xu u. a., 2017]). Es scheint daher eine Untersuchung sinnvoll, inwieweit diese für die Textblockerkennung und Unterscheidung insbesondere der Randtextblöcke eingesetzt werden können. Der Einsatz künstlicher neuronaler Netze scheint zunächst in den Bereichen sinnvoll, in denen umfangreiche Corpora vorliegen, die innerhalb eines einzelnen Corpus eine homogene Gestaltung zeigen, und von denen mit vertretbarem Aufwand ein Trainingssatz erstellt werden kann. Angesichts der rasanten Entwicklung dieser Verfahren in den letzten Jahren, gilt es im Auge zu behalten, welcher Trainingsaufwand in der Zukunft nötig sein wird, um hinreichend brauchbare Ergebnisse zu erzielen.

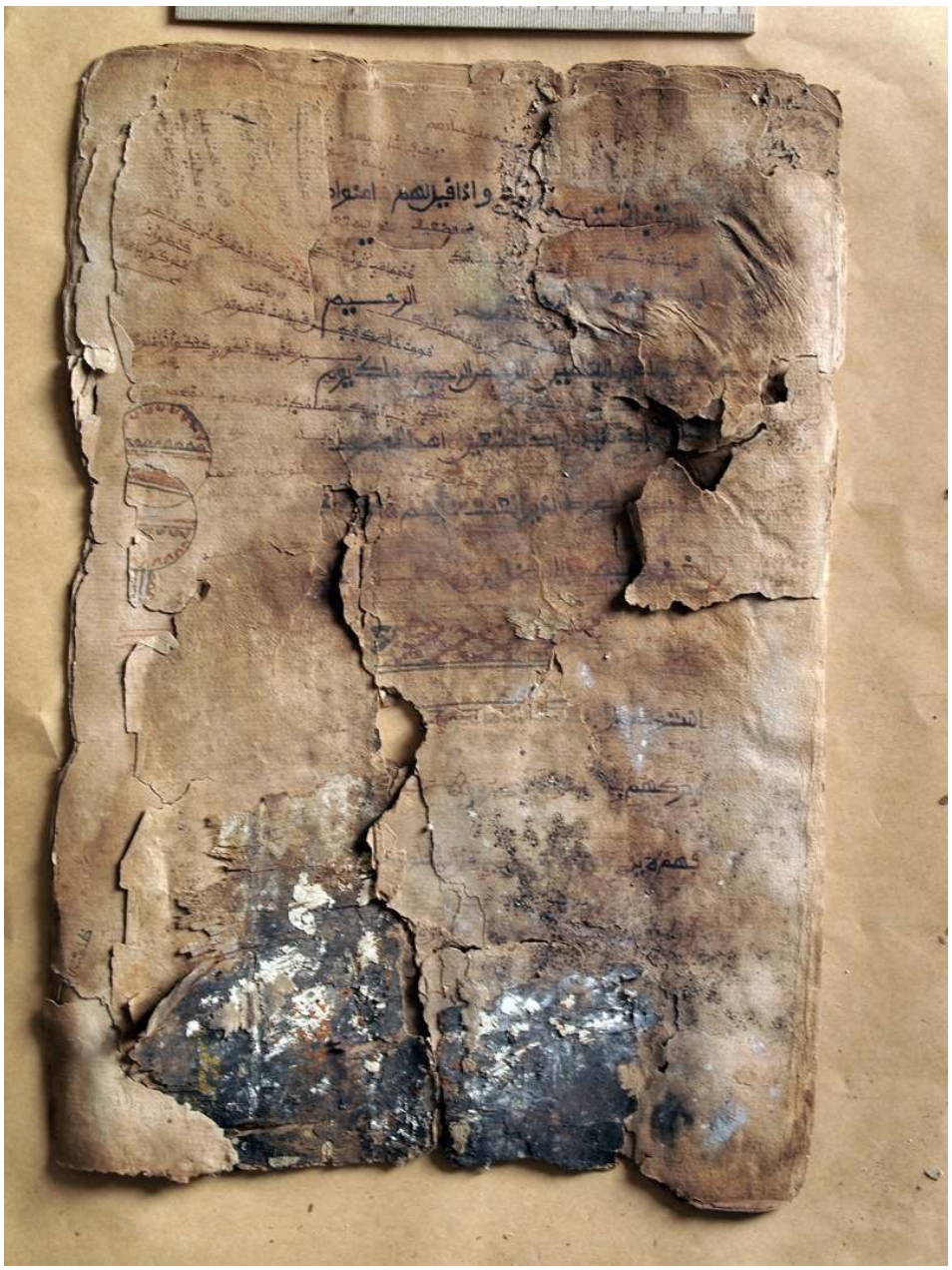


Abbildung 5.1.: Nicht alle Herausforderungen, die sich den Manuskriptforschern stellen, können alleine durch computerbasierte Verfahren gelöst werden.
(Malam Muhammadu Qur'an, © SOAS, University of London. CC BY-NC-SA)

A. Testdatensatz Berlin

Quellen der verwendeten Manuskripte:

	Signatur	Folio	URL
1.	Landberg97	8r	http://resolver.staatsbibliothek-berlin.de/SBB0000AACD00000000
2.	Petermann I 634	206r	http://resolver.staatsbibliothek-berlin.de/SBB00007E2A00000000
3.	Hs. or. 5160	15v	http://resolver.staatsbibliothek-berlin.de/SBB0000AE8B00000000
4.	Hs. or. 5160	30r	http://resolver.staatsbibliothek-berlin.de/SBB0000AE8B00000000
5.	Hs. or. 5160	30v	http://resolver.staatsbibliothek-berlin.de/SBB0000AE8B00000000
6.	Hs. or. 5160	42v	http://resolver.staatsbibliothek-berlin.de/SBB0000AE8B00000000
7.	Hs. or. 5448	45v	http://resolver.staatsbibliothek-berlin.de/SBB0000A63000000000
8.	Hs. or. 5467	6r	http://resolver.staatsbibliothek-berlin.de/SBB0000B30800000000
9.	Hs. or. 5467	186v	http://resolver.staatsbibliothek-berlin.de/SBB0000B30800000000
10.	Hs. or. 5488	8r	http://resolver.staatsbibliothek-berlin.de/SBB0000B30200000000
11.	Hs. or. 14469	5v	http://resolver.staatsbibliothek-berlin.de/SBB000102C300000000
12.	Hs. or. 14469	16r	http://resolver.staatsbibliothek-berlin.de/SBB000102C300000000
13.	Hs. or. 14469	45v	http://resolver.staatsbibliothek-berlin.de/SBB000102C300000000
14.	Petermann II 169	23v	http://resolver.staatsbibliothek-berlin.de/SBB00015BEB00000000
15.	Petermann II 169	30r	http://resolver.staatsbibliothek-berlin.de/SBB00015BEB00000000

Alle verwendeten Manuskripte unterstehen der Creative Commons Lizenz BY-NC-SA.

A.1. Abgleich ohne Segmentierung

Tabelle A.1.: Vergleich zwischen Ground Truth und berechneten Werten ohne Segmentierung. HTB=Haupttextblock, Flächenanteil bezogen auf die Summe aller Pixel der Randtextblockregionen.

Berlin		Differenz		
Seite	Textbereich	Orientierung [°]	Zeilenabstand	Flächenanteil
1	HTB	2,27	2,38%	-
	1	8,31	12,26%	21,52%
	2	11,00	10,14%	17,45%
	3	10,76	14,14%	9,84%
	4	4,81	10,89%	32,13%
	5	38,19	32,78%	5,23%
	6	8,01	7,84%	13,84%
2	HTB	3,76	4,59%	-
	1	6,63	9,57%	100,00%
3	HTB	0,49	2,37%	-
	1	3,34	4,68%	13,27%
	2	7,74	9,27%	27,64%
	3	6,97	17,37%	4,06%
	4	35,14	18,83%	7,82%
	5	31,85	47,16%	8,32%
	6	17,21	33,19%	1,61%
	7	6,64	9,14%	21,44%
	8	3,09	7,99%	15,85%
4	HTB	0,98	1,53%	-
	1	1,68	9,41%	41,70%
	2	9,62	16,48%	29,03%
	3	9,72	13,71%	29,27%
5	HTB	2,45	5,72%	-
	1	8,75	11,57%	32,62%
	2	4,11	6,24%	23,20%
	3	14,73	12,63%	12,72%
	4	42,58	68,14%	1,70%
	5	3,92	7,47%	29,76%
6	HTB	1,57	4,65%	-
	1	5,96	10,57%	14,55%
	2	9,41	14,09%	27,54%
	3	24,71	30,29%	4,14%
	4	23,46	28,55%	3,35%
	5	3,02	7,99%	1,45%
	6	4,60	13,28%	1,75%
	7	9,47	11,09%	3,74%
	8	54,5	48,47%	1,56%
	9	11,26	16,90%	7,62%
	10	15,47	43,14%	1,34%
	11	17,67	33,31%	2,10%
	12	19,63	45,23%	2,40%
	13	10,95	16,18%	14,16%
	14	4,05	4,70%	10,63%
	15	8,84	11,38%	3,65%

Berlin		Differenz		
Seite	Textbereich	Orientierung [°]	Zeilenabstand	Flächenanteil
7	HTB	0,73	2,45%	-
	1	7,04	9,06%	58,49%
	2	14,77	13,68%	34,90%
	3	18,84	47,06%	6,61%
8	HTB	0,75	5,98%	-
	1	23,13	23,12%	20,28%
	2	10,34	17,94%	15,56%
	3	21,20	34,89%	18,56%
	4	29,89	42,66%	4,22%
	5	36,66	64,36%	4,20%
	6	24,73	34,47%	8,56%
	7	47,00	67,06%	3,03%
	8	11,05	12,36%	8,62%
	9	17,16	12,92%	11,39%
	10	46,39	42,10%	5,57%
9	HTB	0,07	3,53%	-
	1	15,65	17,46%	31,21%
	2	2,89	16,69%	12,06%
	3	33,89	30,09%	13,06%
	4	18,37	18,64%	20,33%
	5	31,41	47,22%	23,33%
10	HTB	0,46	3,16%	-
	1	13,56	8,64%	20,31%
	2	6,07	5,91%	44,11%
	3	6,10	11,50%	2,56%
	4	28,33	28,41%	5,86%
	5	12,86	14,24%	4,19%
	6	4,87	4,04%	22,98%
11	HTB	1,48	4,90%	-
	1	20,83	32,45%	2,19%
	2	17,2	19,04%	9,32%
	3	6,34	11,83%	4,79%
	4	6,41	8,90%	7,92%
	5	17,81	31,57%	3,31%
	6	9,65	8,54%	4,25%
	7	3,46	4,84%	7,57%
	8	3,27	3,19%	3,58%
	9	39,83	56,54%	0,78%
	10	8,57	10,78%	13,99%
	11	15,02	14,60%	6,50%
	12	16,27	24,55%	4,59%
	13	34,41	28,38%	3,48%
	14	6,65	7,58%	19,51%
	15	9,09	11,49%	8,23%

Berlin		Differenz		
Seite	Textbereich	Orientierung [°]	Zeilenabstand	Flächenanteil
12	HTB	2,2	9,30%	-
	1	4,60	6,73%	53,94%
	2	14,57	31,90%	5,86%
	3	4,96	9,56%	13,28%
	4	25,93	51,93%	1,72%
	5	2,83	10,47%	7,94%
	6	13,73	40,15%	3,10%
	7	5,85	13,13%	14,17%
13	HTB	1,81	6,22%	-
	1	5,43	12,95%	5,46%
	2	14,17	19,24%	3,65%
	3	8,86	31,40%	1,30%
	4	2,27	5,59%	9,52%
	5	16,28	21,35%	6,11%
	6	29,85	44,30%	5,15%
	7	2,49	4,35%	17,87%
	8	16,90	20,63%	2,51%
	9	5,15	9,16%	13,23%
	10	6,07	5,32%	27,62%
	11	11,44	42,65%	1,68%
	12	9,57	18,93%	5,90%
14	HTB	0,78	4,51%	-
	1	9,74	11,63%	46,80%
	2	8,51	24,56%	6,71%
	3	25,68	43,24%	21,59%
	4	7,89	10,10%	12,05%
	5	22,60	24,97%	12,85%
15	HTB	0,99	1,56%	-
	1	14,05	26,60%	19,07%
	2	18,47	17,87%	17,50%
	3	9,93	16,73%	22,91%
	4	8,53	26,86%	19,60%
	5	14,96	10,82%	20,92%

A.2. Abgleich mit Segmentierung

Tabelle A.2.: Abgleich unter Berücksichtigung der Segmentierung.Vergleich zwischen Ground Truth und berechneten Werten mit Segmentierung.
Zur Erklärung der Berechnung der Werte siehe Abschnitt 4.4.2

Berlin		>90%	>80%	>70%	>50%	False Negative	False Positive
Seite							
1	Flächenanteil	60,30%	85,60%	90,50%	90,50%	0,00%	0,00%
	Korrespondenzen	14,29%	71,43%	85,71%	85,71%	0,00%	0,00%
	ZA-differenz [%]	1,72%	7,63%	4,73%	-		
	O-differenz [°]	1,76	2,02	5,91	-		
2	Flächenanteil	84,41%	93,55%	93,55%	93,55%	0,00%	0,00%
	Korrespondenzen	50,00%	100,00%	100,00%	100,00%	0,00%	0,00%
	ZA-differenz [%]	0,54%	4,69%	-	-		
	O-differenz [°]	2,94	3,10	-	-		
3	Flächenanteil	55,68%	66,16%	81,39%	84,39%	0,00%	0,00%
	Korrespondenzen	11,11%	33,33%	55,56%	77,78%	0,00%	0,00%
	ZA-differenz [%]	1,00%	1,88%	3,38%	10,06%		
	O-differenz [°]	0,52	1,41	0,75	21,15		
4	Flächenanteil	76,15%	91,68%	91,68%	91,68%	0,00%	0,23%
	Korrespondenzen	50,00%	100,00%	100,00%	100,00%	0,00%	25,00%
	ZA-differenz [%]	4,27%	3,42%	-	-		
	O-differenz [°]	0,44	2,44	-	-		
5	Flächenanteil	68,60%	88,10%	88,10%	90,50%	0,45%	0,36%
	Korrespondenzen	16,70%	66,70%	66,70%	83,30%	16,70%	16,70%
	ZA-differenz [%]	4,48%	5,43%	-	-		
	O-differenz [°]	1,16	1,41	-	-		
6	Flächenanteil	51,10%	56,50%	58,00%	68,10%	0,00%	0,13%
	Korrespondenzen	6,20%	12,50%	18,80%	37,50%	0,00%	6,20%
	ZA-differenz [%]	2,60%	2,30%	7,83%	19,86%		
	O-differenz [°]	0,34	1,10	4,89	12,35		
7	Flächenanteil	73,60%	73,60%	81,60%	81,60%	6,56%	0,00%
	Korrespondenzen	25,00%	25,00%	50,00%	50,00%	25,00%	0,00%
	ZA-differenz [%]	1,37%	-	4,20%	-		
	O-differenz [°]	0,51	-	0,25	-		
8	Flächenanteil	0,00%	63,60%	71,40%	78,80%	0,00%	0,00%
	Korrespondenzen	0,00%	9,10%	27,30%	45,50%	0,00%	0,00%
	ZA-differenz [%]	-	5,48%	4,40%	0,41%		
	O-differenz [°]	-	0,12	7,38	2,28		
9	Flächenanteil	0,00%	74,50%	74,50%	82,90%	0,00%	0,00%
	Korrespondenzen	0,00%	33,30%	33,30%	83,30%	0,00%	0,00%
	ZA-differenz [%]	-	7,13%	-	5,25%		
	O-differenz [°]	-	6,36	-	1,23		
10	Flächenanteil	78,20%	88,00%	89,80%	91,50%	0,00%	0,00%
	Korrespondenzen	42,90%	57,10%	71,40%	85,70%	0,00%	0,00%
	ZA-differenz [%]	1,61%	0,28%	9,18%	6,77%		
	O-differenz [°]	2,50	6,89	0,49	5,73		
11	Flächenanteil	0,00%	57,20%	64,10%	72,00%	0,48%	0,00%
	Korrespondenzen	0,00%	18,80%	31,20%	50,00%	6,20%	0,00%
	ZA-differenz [%]	-	3,94%	9,84%	8,78%		
	O-differenz [°]	-	0,66	0,58	1,67		

Tabelle A.3.: Abgleich unter Berücksichtigung der Segmentierung.Vergleich zwischen Ground Truth und berechneten Werten mit Segmentierung (Fortsetzung).
Zur Erklärung der Berechnung der Werte siehe Abschnitt 4.4.2

Berlin		>90%	>80%	>70%	>50%	False Negative	False Positive
Seite							
12	Flächenanteil	70,80%	80,30%	86,20%	88,80%	0,00%	0,61%
	Korrespondenzen	25,00%	50,00%	62,50%	87,50%	0,00%	25,00%
	ZA-differenz [%]	3,27%	7,46%	5,94%	2,58%		
	O-differenz [°]	0,78	1,05	1,95	1,31		
13	Flächenanteil	27,40%	72,90%	72,90%	82,00%	0,91%	0,26%
	Korrespondenzen	23,10%	38,50%	38,50%	84,60%	7,70%	15,40%
	ZA-differenz [%]	2,02%	3,63%	-	7,19%		
	O-differenz [°]	0,91	0,39	-	5,26		
14	Flächenanteil	74,10%	76,10%	76,10%	83,00%	0,00%	0,00%
	Korrespondenzen	16,70%	33,30%	33,30%	66,70%	0,00%	0,00%
	ZA-differenz [%]	3,26%	1,64%	-	3,47%		
	O-differenz [°]	0,01	2,30	-	1,11		
15	Flächenanteil	68,40%	73,80%	77,90%	84,70%	0,00%	0,00%
	Korrespondenzen	16,70%	33,30%	50,00%	83,30%	0,00%	0,00%
	ZA-differenz [%]	0,91%	13,38%	19,69%	5,62%		
	O-differenz [°]	0,04	4,63	0,18	1,10		

A.3. Darstellung der Manuskriptseiten

Die Farben der Textblockregionen wurden zufällig gewählt, unterschiedliche Farben bedeuten lediglich unterschiedliche Textblöcke.

Abbildung A.1.: Manuskript 1-3. Spaltenweise: Original (randbereinigt) (1), Segmentierung: Ground Truth (2), berechnet (3).

Abbildung A.2.: Manuskript 4-6. Spaltenweise: Original (randbereinigt) (1), Segmentierung: Ground Truth (2), berechnet (3).

Abbildung A.3.: Manuskript 7-9. Spaltenweise: Original (randbereinigt) (1), Segmentierung: Ground Truth (2), berechnet (3).

Abbildung A.4.: Manuskript 10-12. Spaltenweise: Original (randbereinigt) (1), Segmentierung: Ground Truth (2), berechnet (3).

Abbildung A.5.: Manuskript 13-15. Spaltenweise: Original (randbereinigt) (1), Segmentierung: Ground Truth (2), berechnet (3).

B. Testdatensatz Leipzig

Quellen der verwendeten Manuskripte:

	Signatur	Seite	URL
1.	B. or. 002	36	https://www.islamic-manuscripts.net/receive/IslamHSBook_islamhs_00000617
2.	B. or. 002	42	https://www.islamic-manuscripts.net/receive/IslamHSBook_islamhs_00000617
3.	B. or. 002	51	https://www.islamic-manuscripts.net/receive/IslamHSBook_islamhs_00000617
4.	B. or. 002	70	https://www.islamic-manuscripts.net/receive/IslamHSBook_islamhs_00000617
5.	B. or. 002	74	https://www.islamic-manuscripts.net/receive/IslamHSBook_islamhs_00000617
6.	B. or. 002	84	https://www.islamic-manuscripts.net/receive/IslamHSBook_islamhs_00000617
7.	B. or. 014	26	https://www.islamic-manuscripts.net/receive/IslamHSBook_islamhs_00000626
8.	B. or. 040-03	4	https://www.islamic-manuscripts.net/receive/IslamHSBook_islamhs_00000768
9.	B. or. 040-02	5	https://www.islamic-manuscripts.net/receive/IslamHSBook_islamhs_00000767
10.	B. or. 040-02	13	https://www.islamic-manuscripts.net/receive/IslamHSBook_islamhs_00000767
11.	B. or. 062-01	12	https://www.islamic-manuscripts.net/receive/IslamHSBook_islamhs_00001046
12.	B. or. 062-01	16	https://www.islamic-manuscripts.net/receive/IslamHSBook_islamhs_00001046
13.	B. or. 141	130	https://www.islamic-manuscripts.net/receive/IslamHSBook_islamhs_00001257
14.	B. or. 235	151	https://www.islamic-manuscripts.net/receive/IslamHSBook_islamhs_00010324
15.	B. or. 235	187	https://www.islamic-manuscripts.net/receive/IslamHSBook_islamhs_00010324

Alle verwendeten Manuskripte unterstehen der Creative Commons Lizenz BY-NC.

B.1. Abgleich ohne Segmentierung

Tabelle B.1.: Vergleich zwischen Ground Truth und berechneten Werten ohne Segmentierung. HTB=Haupttextblock, Flächenanteil bezogen auf die Summe aller Pixel der Randtextblockregionen.

Leipzig		Differenz		Flächen-anteil
Seite	Text-bereich	Orientie-rung [°]	Zeilen-abstand	
1	HTB	0,25	3,87%	-
	1	7,06	15,93%	1,84%
	2	2,75	6,42%	8,10%
	3	4,79	5,67%	4,04%
	4	0,66	20,13%	5,08%
	5	7,65	6,45%	16,83%
	6	4,32	5,36%	11,62%
	7	12,97	16,50%	1,77%
	8	7,38	13,20%	2,11%
	9	10,75	5,78%	6,00%
	10	3,53	5,68%	11,06%
	11	19,77	17,27%	1,79%
	12	5,23	11,29%	2,23%
	13	9,46	9,85%	3,69%
	14	6,76	6,03%	8,20%
	15	11,59	8,99%	15,64%
2	HTB	0,09	3,71%	-
	1	11,72	10,54%	24,99%
	2	21,58	14,50%	7,54%
	3	2,91	9,63%	11,63%
	4	45,85	38,99%	1,98%
	5	9,52	5,18%	30,23%
	6	9,21	6,68%	12,11%
	7	13,42	16,30%	3,12%
	8	8,81	12,81%	8,41%
3	HTB	0,15	3,12%	-
	1	6,01	2,78%	11,02%
	2	7,92	14,33%	5,20%
	3	14,56	16,56%	12,29%
	4	7,6	9,16%	10,84%
	5	5,98	7,08%	36,33%
	6	14,38	20,54%	5,60%
	7	7,85	11,18%	17,10%
	8	4,49	9,75%	1,63%
4	HTB	0,15	3,68%	-
	1	20,69	10,24%	9,33%
	2	5,40	10,61%	12,26%
	3	6,50	6,98%	26,95%
	4	7,83	8,73%	4,30%
	5	10,68	14,03%	1,87%
	6	4,48	6,53%	29,68%
	7	8,67	6,76%	5,88%
	8	15,91	7,70%	6,88%
	9	14,82	14,08%	2,86%

Leipzig		Differenz		Flächen-anteil
Seite	Text-bereich	Orientie-rung [°]	Zeilen-abstand	
5	HTB	0,2	3,65%	-
	1	9,32	7,99%	45,13%
	2	4,79	8,82%	14,60%
	3	19,51	9,78%	8,61%
	4	15,59	7,24%	6,94%
	5	10,59	9,28%	20,14%
	6	14,31	10,93%	4,58%
6	HTB	0,15	3,63%	-
	1	6,21	13,62%	5,23%
	2	11,09	15,40%	22,23%
	3	26,44	24,46%	6,24%
	4	19,96	22,04%	8,89%
	5	6,08	16,66%	12,49%
	6	7,23	8,82%	15,47%
	7	16,12	4,98%	29,45%
7	HTB	0,55	2,03%	-
	1	11,60	8,30%	66,46%
	2	10,48	8,79%	33,54%
8	HTB	1,54	0,66%	-
	1	5,50	4,94%	82,16%
	2	18,69	6,41%	17,84%
9	HTB	0,04	2,54%	-
	1	6,90	8,59%	24,62%
	2	5,77	3,22%	24,33%
	3	5,38	5,46%	32,90%
	4	3,09	7,32%	18,14%
10	HTB	1,52	9,12%	-
	1	7,61	16,02%	3,72%
	2	2,76	4,27%	43,28%
	3	1,50	5,88%	22,51%
	4	2,58	6,22%	24,92%
	5	2,24	11,18%	2,65%
	6	6,54	22,24%	2,92%
11	HTB	8,1	12,56%	-
	1	14,46	18,48%	19,03%
	2	10,50	19,81%	17,31%
	3	10,86	15,62%	24,27%
	4	6,14	10,71%	28,47%
	5	5,78	18,40%	7,51%
	6	13,1	20,28%	3,42%

Leipzig		Differenz		Flächen-anteil
Seite	Text-bereich	Orientie-rung [°]	Zeilen-abstand	
12	HTB	5,95	17,88%	-
	1	3,60	9,13%	14,12%
	2	1,89	6,68%	38,36%
	3	18,15	29,36%	8,89%
	4	28,25	23,40%	12,22%
	5	1,86	28,44%	7,25%
	6	19,9	23,02%	19,16%
13	HTB	0,23	1,89%	-
	1	7,51	7,79%	40,38%
	2	4,30	6,24%	28,83%
	3	2,41	11,07%	8,77%
	4	10,88	6,97%	10,80%
	5	6,41	6,57%	11,21%
14	HTB	0,28	3,30%	-
	1	20,15	4,58%	60,70%
	2	3,92	10,70%	2,86%
	3	14,82	11,82%	19,04%
	4	18,2	8,92%	17,40%
15	HTB	1,48	4,74%	-
	1	15,21	6,64%	59,36%
	2	3,57	22,53%	7,17%
	3	8,38	7,46%	18,59%
	4	9,49	5,61%	14,89%

B.2. Abgleich mit Segmentierung

Tabelle B.2.: Abgleich unter Berücksichtigung der Segmentierung.Vergleich zwischen Ground Truth und berechneten Werten mit Segmentierung.
Zur Erklärung der Berechnung der Werte siehe Abschnitt 4.4.2

Leipzig		>90%	>80%	>70%	>50%	False Negative	False Positive
Seite							
1	Flächenanteil	55,40%	61,40%	67,40%	80,10%	0,00%	0,00%
	Korrespondenzen	12,50%	18,80%	31,20%	62,50%	0,00%	0,00%
	ZA-differenz [%]	3,01%	0,62%	1,97%	1,76%		
	O-differenz [°]	0,13	3,71	1,96	2,87		
2	Flächenanteil	67,30%	84,50%	90,60%	92,70%	0,00%	0,00%
	Korrespondenzen	22,20%	44,40%	77,80%	88,90%	0,00%	0,00%
	ZA-differenz [%]	3,85%	1,63%	6,86%	4,74%		
	O-differenz [°]	0,42	4,63	3,93	0,04		
3	Flächenanteil	75,40%	83,70%	88,10%	91,10%	0,00%	0,00%
	Korrespondenzen	22,20%	44,40%	66,70%	77,80%	0,00%	0,00%
	ZA-differenz [%]	3,20%	2,25%	4,82%	0,01%		
	O-differenz [°]	0,36	0,94	5,34	4,25		
4	Flächenanteil	61,80%	84,70%	89,70%	89,70%	0,00%	0,00%
	Korrespondenzen	20,00%	60,00%	80,00%	80,00%	0,00%	0,00%
	ZA-differenz [%]	3,93%	1,29%	0,60%	-		
	O-differenz [°]	1,01	4,33	9,62	-		
5	Flächenanteil	66,40%	66,40%	85,20%	87,00%	0,00%	0,00%
	Korrespondenzen	14,30%	14,30%	57,10%	71,40%	0,00%	0,00%
	ZA-differenz [%]	1,28%	-	5,05%	5,03%		
	O-differenz [°]	0,34	-	6,50	14,43		
6	Flächenanteil	71,20%	71,20%	81,40%	86,30%	0,00%	0,00%
	Korrespondenzen	12,50%	12,50%	37,50%	75,00%	0,00%	0,00%
	ZA-differenz [%]	2,32%	-	4,84%	6,50%		
	O-differenz [°]	0,01	-	8,77	2,93		
7	Flächenanteil	66,60%	66,60%	66,60%	84,20%	0,00%	0,00%
	Korrespondenzen	33,30%	33,30%	33,30%	100,00%	0,00%	0,00%
	ZA-differenz [%]	0,82%	-	-	0,66%		
	O-differenz [°]	0,33	-	-	3,85		
8	Flächenanteil	41,80%	81,90%	89,40%	89,40%	0,00%	0,00%
	Korrespondenzen	33,30%	66,70%	100,00%	100,00%	0,00%	0,00%
	ZA-differenz [%]	0,09%	0,24%	0,12%	-		
	O-differenz [°]	0,60	0,49	3,73	-		
9	Flächenanteil	0,00%	72,50%	82,20%	90,70%	0,00%	0,00%
	Korrespondenzen	0,00%	60,00%	80,00%	100,00%	0,00%	0,00%
	ZA-differenz [%]	-	0,46%	0,54%	4,70%		
	O-differenz [°]	-	1,81	1,27	2,47		
10	Flächenanteil	36,60%	76,90%	91,10%	92,70%	0,00%	0,00%
	Korrespondenzen	28,60%	42,90%	57,10%	71,40%	0,00%	0,00%
	ZA-differenz [%]	4,19%	3,34%	4,84%	7,15%		
	O-differenz [°]	0,36	0,85	0,05	2,24		
11	Flächenanteil	0,00%	62,60%	71,50%	78,30%	0,00%	0,00%
	Korrespondenzen	0,00%	28,60%	57,10%	85,70%	0,00%	0,00%
	ZA-differenz [%]	-	2,93%	7,52%	5,99%		
	O-differenz [°]	-	0,52	0,99	1,63		

Tabelle B.3.: Abgleich unter Berücksichtigung der Segmentierung.Vergleich zwischen Ground Truth und berechneten Werten mit Segmentierung (Fortsetzung).
Zur Erklärung der Berechnung der Werte siehe Abschnitt 4.4.2

Leipzig		>90%	>80%	>70%	>50%	False Negative	False Positive
Seite							
12	Flächenanteil	0,00%	67,00%	67,00%	81,10%	0,00%	0,00%
	Korrespondenzen	0,00%	14,30%	14,30%	85,70%	0,00%	0,00%
	ZA-differenz [%]	-	3,71%	-	10,07%		
	O-differenz [°]	-	0,23	-	1,51		
13	Flächenanteil	66,20%	73,70%	73,70%	88,70%	0,00%	0,00%
	Korrespondenzen	16,70%	33,30%	33,30%	100,00%	0,00%	0,00%
	ZA-differenz [%]	1,65%	4,80%	-	5,16%		
	O-differenz [°]	0,09	1,47	-	1,36		
14	Flächenanteil	0,00%	75,10%	82,30%	88,60%	0,00%	0,00%
	Korrespondenzen	0,00%	40,00%	60,00%	100,00%	0,00%	0,00%
	ZA-differenz [%]	-	1,77%	5,83%	9,72%		
	O-differenz [°]	-	3,22	6,34	2,19		
15	Flächenanteil	44,10%	54,80%	88,30%	88,30%	0,00%	0,00%
	Korrespondenzen	20,00%	60,00%	100,00%	100,00%	0,00%	0,00%
	ZA-differenz [%]	1,12%	11,90%	3,08%	-		
	O-differenz [°]	0,94	0,74	0,99	-		

B.3. Darstellung der Manuskriptseiten

Die Farben der Textblockregionen wurden zufällig gewählt, unterschiedliche Farben bedeuten lediglich unterschiedliche Textblöcke.

Abbildung B.1.: Manuskript 1-3. Spaltenweise: Original (randbereinigt) (1), Segmentierung: Ground Truth (2), berechnet (3).

Abbildung B.2.: Manuskript 4-6. Spaltenweise: Original (randbereinigt) (1), Segmentierung: Ground Truth (2), berechnet (3).

Abbildung B.3.: Manuskript 7-9. Spaltenweise: Original (randbereinigt) (1), Segmentierung: Ground Truth (2), berechnet (3).

Abbildung B.4.: Manuskript 10-12. Spaltenweise: Original (randbereinigt) (1), Segmentierung: Ground Truth (2), berechnet (3).

Abbildung B.5.: Manuskript 13-15. Spaltenweise: Original (randbereinigt) (1), Segmentierung: Ground Truth (2), berechnet (3).

C. Testdatensatz Greek

Quellen der verwendeten Manuskripte:

	Signatur **URL**	**Seite**
1.	GRGSA-CSA_PCVLA.BMS.F000002 http://arxeiomnimon.gak.gr/browse/resource.html?tab=02&id=482876	Take 008
2.	GRGSA-CSA_PCVLA.BMS.F000019 http://arxeiomnimon.gak.gr/browse/resource.html?tab=02&id=482893	Take 010
3.	GRGSA-CSA_PCMANUSCR.IT000238 http://arxeiomnimon.gak.gr/browse/resource.html?tab=02&id=571357	Page 199r
4.	GRGSA-CSA_PAO001.00.F000001 http://arxeiomnimon.gak.gr/browse/resource.html?tab=02&id=3414	Take 017
5.	GRGSA-CSA_PAO001.00.F000001 http://arxeiomnimon.gak.gr/browse/resource.html?tab=02&id=3414	Take 113
6.	GRGSA-CSA_PAO001.00.F000002 http://arxeiomnimon.gak.gr/browse/resource.html?tab=02&id=3415	Take 201
7.	GRGSA-ACH_JUS027.01.F000020 http://arxeiomnimon.gak.gr/browse/resource.html?tab=02&id=24	Take 036
8.	GRGSA-ACH_JUS027.01.F000020 http://arxeiomnimon.gak.gr/browse/resource.html?tab=02&id=24	Take 204
9.	GRGSA-ACH_JUS027.01.F000020 http://arxeiomnimon.gak.gr/browse/resource.html?tab=02&id=24	Take 288
10.	GRGSA-ACH_JUS027.01.F000020 http://arxeiomnimon.gak.gr/browse/resource.html?tab=02&id=24	Take 291
11.	GRGSA-ACH_JUS027.01.F000020 http://arxeiomnimon.gak.gr/browse/resource.html?tab=02&id=24	Take 1279
12.	GRGSA-ACH_JUS027.01.F000020 http://arxeiomnimon.gak.gr/browse/resource.html?tab=02&id=24	Take 1287
13.	GRGSA-ACH_JUS027.01.F000021 http://arxeiomnimon.gak.gr/browse/resource.html?tab=02&id=25	Take 434
14.	GRGSA-ACH_JUS027.01.F000021 http://arxeiomnimon.gak.gr/browse/resource.html?tab=02&id=25	Take 1221
15.	GRGSA-ACH_JUS027.01.F000008.SF0001 http://arxeiomnimon.gak.gr/browse/resource.html?tab=02&id=11	Take 1019

Alle Manuskripte dieses Testsatzes sind dem Webangebot der Greek State Archives entnommen, und stehen unter folgenden Lizenzbedingungen: http://arxeiomnimon.gak.gr/terms.html

C.1. Abgleich ohne Segmentierung

Tabelle C.1.: Vergleich zwischen Ground Truth und berechneten Werten ohne Segmentierung. HTB=Haupttextblock, Flächenanteil bezogen auf die Summe aller Pixel der Randtextblockregionen.

Greek		Differenz	
Seite	Text-bereich	Orientie-rung [°]	Zeilen-abstand
1	1	1,68	5,94%
	2	0,19	2,30%
	3	1,40	5,95%
	4	2,10	17,73%
2	1	0,34	3,21%
	2	0,61	2,51%
3	HTB	0,17	3,93%
	1	20,04	28,08%
4	1	0,19	6,17%
	2	1,78	5,68%
5	HTB	0,45	7,38%
	1	4,28	22,48%
	2	0,39	6,43%
	3	2,06	11,37%
6	HTB	2,26	3,58%
	1	3,65	8,14%
7	HTB	0,85	4,25%
	1	2,14	11,13%
	2	4,7	31,35%
8	HTB	1,02	3,56%
	1	0,29	34,44%
	2	13,53	19,91%
9	HTB	1,86	5,43%
	1	5,24	14,67%
	2	6,01	12,04%
10	HTB	0,53	8,98%
	1	1,79	2,39%
11	HTB	0,16	6,61%
	1	3,22	9,71%
	2	9,35	33,79%
12	HTB	2,3	6,66%
	1	1,93	5,91%
13	HTB	4,88	18,85%
	1	0,24	4,76%
14	1	1,13	2,65%
	2	11,82	22,49%
	3	1,31	9,84%
15	1	3,17	5,66%
	2	5,86	5,20%
	3	9,56	12,50%
	4	1,14	24,53%

C.2. Abgleich mit Segmentierung

Tabelle C.2.: Abgleich unter Berücksichtigung der Segmentierung.Vergleich zwischen Ground Truth und berechneten Werten mit Segmentierung.
Zur Erklärung der Berechnung der Werte siehe Abschnitt 4.4.2

Seite	Greek	>90%	>80%	>70%	>50%	False Negative	False Positive
1	Flächenanteil	67,10%	85,80%	89,90%	89,90%	0,00%	0,00%
	Korrespondenzen	50,00%	75,00%	100,00%	100,00%	0,00%	0,00%
	ZA-differenz [%]	1,83%	2,86%	9,85%	-		
	O-differenz [°]	0,46	0,05	1,82	-		
2	Flächenanteil	59,00%	92,90%	92,90%	92,90%	0,00%	0,00%
	Korrespondenzen	50,00%	100,00%	100,00%	100,00%	0,00%	0,00%
	ZA-differenz [%]	1,13%	0,06%	-	-		
	O-differenz [°]	0,17	0,00	-	-		
3	Flächenanteil	0,00%	83,90%	83,90%	83,90%	0,00%	0,00%
	Korrespondenzen	0,00%	50,00%	50,00%	50,00%	0,00%	0,00%
	ZA-differenz [%]	-	0,12%	-	-		
	O-differenz [°]	-	0,04	-	-		
4	Flächenanteil	0,00%	93,60%	93,60%	93,60%	0,00%	7,03%
	Korrespondenzen	0,00%	100,00%	100,00%	100,00%	0,00%	100,00%
	ZA-differenz [%]	-	3,22%	-	-		
	O-differenz [°]	-	0,14	-	-		
5	Flächenanteil	0,00%	52,00%	52,00%	78,10%	0,00%	1,08%
	Korrespondenzen	0,00%	25,00%	25,00%	75,00%	0,00%	25,00%
	ZA-differenz [%]	-	0,95%	-	3,78%		
	O-differenz [°]	-	0,0249	-	0,93		
6	Flächenanteil	82,00%	82,00%	91,70%	91,70%	0,00%	0,00%
	Korrespondenzen	50,00%	50,00%	100,00%	100,00%	0,00%	0,00%
	ZA-differenz [%]	1,58%	-	1,70%	-		
	O-differenz [°]	0,37	-	1,59	-		
7	Flächenanteil	84,50%	84,50%	86,70%	90,10%	0,00%	0,00%
	Korrespondenzen	33,00%	33,00%	66,60%	100,00%	0,00%	0,00%
	ZA-differenz [%]	0,68%	-	6,96%	13,87%		
	O-differenz [°]	0,10	-	0,05	1,58		
8	Flächenanteil	0,00%	0,00%	94,60%	94,60%	0,00%	4,52%
	Korrespondenzen	0,00%	0,00%	66,60%	66,60%	0,00%	33,00%
	ZA-differenz [%]	-	-	5,92%	-		
	O-differenz [°]	-	-	6,45	-		
9	Flächenanteil	0,00%	73,00%	90,80%	90,80%	0,00%	2,25%
	Korrespondenzen	0,00%	66,70%	100,00%	100,00%	0,00%	33,30%
	ZA-differenz [%]	-	5,73%	1,27%	-		
	O-differenz [°]	-	0,43	0,11	-		
10	Flächenanteil	71,70%	93,60%	93,60%	93,60%	0,00%	0,00%
	Korrespondenzen	50,00%	100,00%	100,00%	100,00%	0,00%	0,00%
	ZA-differenz [%]	2,21%	0,72%	-	-		
	O-differenz [°]	0,10	0,15	-	-		
11	Flächenanteil	56,00%	89,70%	89,70%	89,70%	0,00%	1,75%
	Korrespondenzen	33,30%	66,70%	66,70%	66,70%	0,00%	33,00%
	ZA-differenz [%]	6,15%	4,14%	-	-		
	O-differenz [°]	0,06	1,40	-	-		

Tabelle C.3.: Abgleich unter Berücksichtigung der Segmentierung.Vergleich zwischen Ground Truth und berechneten Werten mit Segmentierung (Fortsetzung).
Zur Erklärung der Berechnung der Werte siehe Abschnitt 4.4.2

Greek		>90%	>80%	>70%	>50%	False Negative	False Positive
Seite							
12	Flächenanteil	65,80%	65,80%	89,70%	89,70%	0,00%	0,00%
	Korrespondenzen	50,00%	50,00%	100,00%	100,00%	0,00%	0,00%
	ZA-differenz [%]	0,03%	-	7,24%	-		
	O-differenz [°]	1,46	-	0,85	-		
13	Flächenanteil	45,50%	45,50%	83,50%	83,50%	0,00%	0,00%
	Korrespondenzen	50,00%	50,00%	100,00%	100,00%	0,00%	0,00%
	ZA-differenz [%]	0,47%	-	12,95%	-		
	O-differenz [°]	0,15	-	0,04	-		
14	Flächenanteil	0,00%	93,00%	93,00%	93,00%	0,00%	0,00%
	Korrespondenzen	0,00%	66,70%	66,70%	66,70%	0,00%	0,00%
	ZA-differenz [%]	-	1,80%	-	-		
	O-differenz [°]	-	0,89	-	-		
15	Flächenanteil	0,00%	78,60%	78,60%	81,70%	0,00%	0,76%
	Korrespondenzen	0,00%	50,00%	50,00%	75,00%	0,00%	25,00%
	ZA-differenz [%]	-	2,44%	-	5,56%		
	O-differenz [°]	-	1,04	-	0,99601		

C.3. Darstellung der Manuskriptseiten

Die Farben der Textblockregionen wurden zufällig gewählt, unterschiedliche Farben bedeuten lediglich unterschiedliche Textblöcke.

Abbildung C.1.: Manuskript 1-3. Spaltenweise: Original (randbereinigt) (1), Segmentierung: Ground Truth (2), berechnet (3).

Abbildung C.2.: Manuskript 4-6. Spaltenweise: Original (randbereinigt) (1), Segmentierung: Ground Truth (2), berechnet (3).

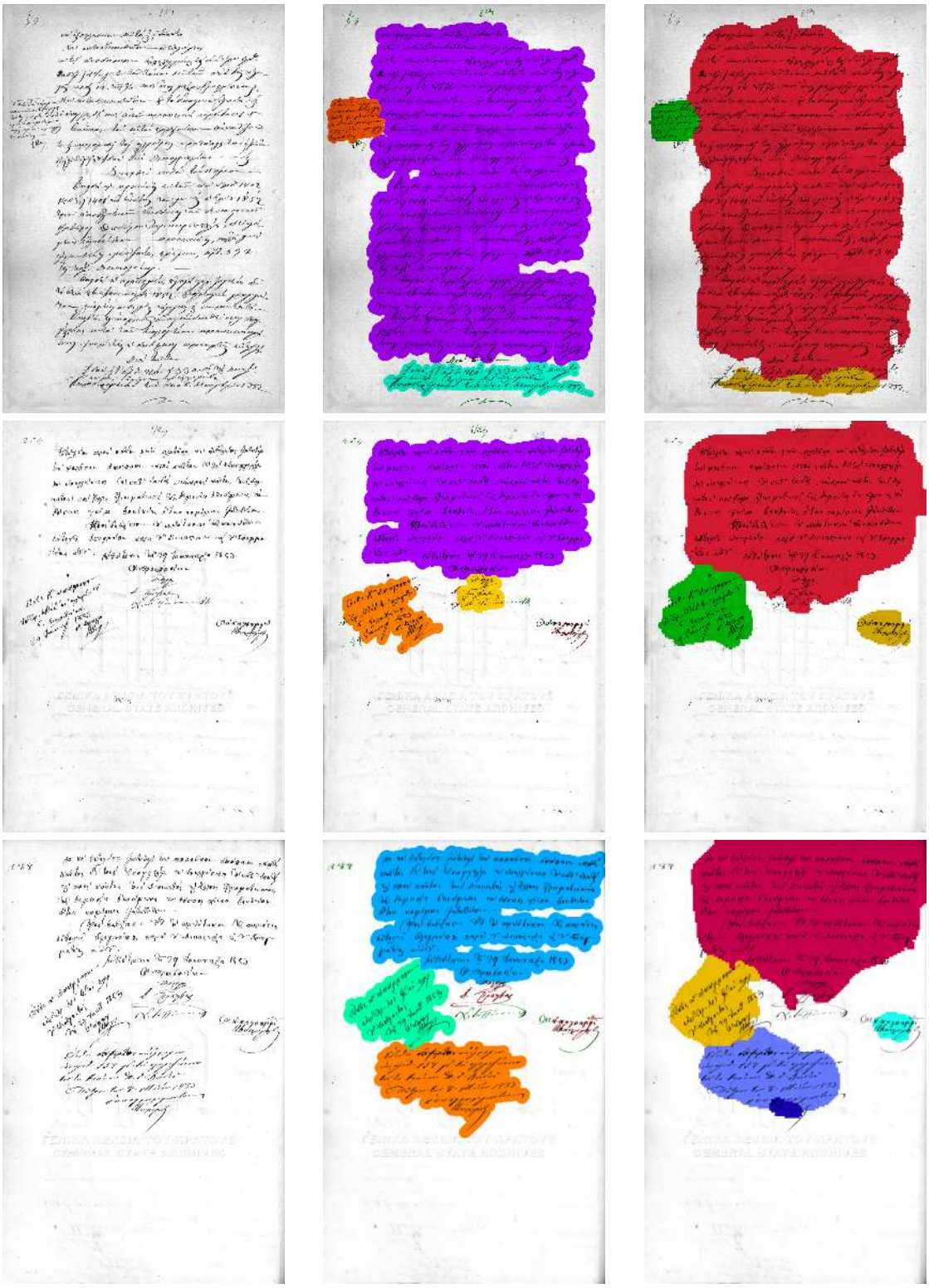


Abbildung C.3.: Manuskript 7-9. Spaltenweise: Original (randbereinigt) (1), Segmentierung: Ground Truth (2), berechnet (3).

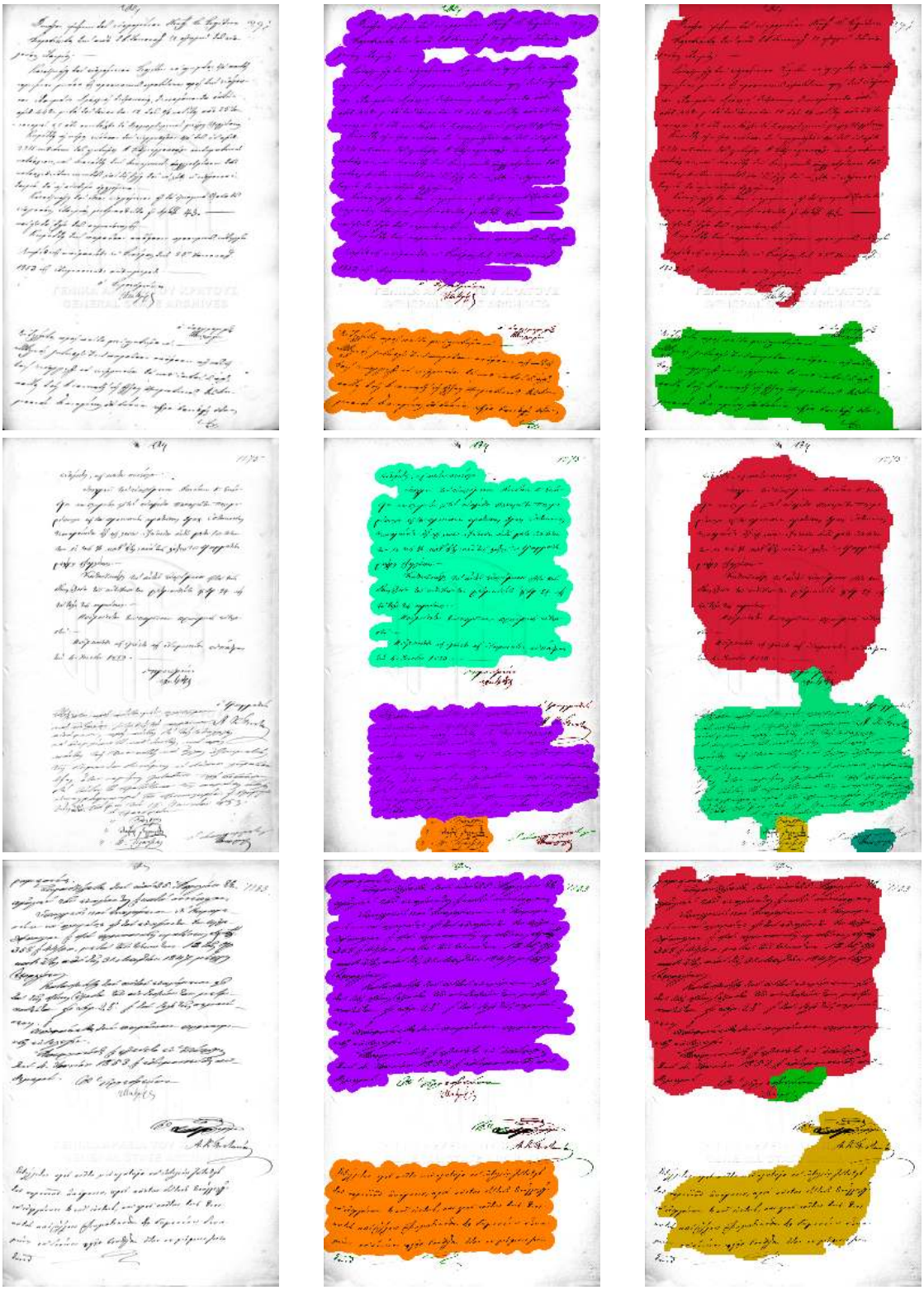


Abbildung C.4.: Manuskript 10-12. Spaltenweise: Original (randbereinigt) (1), Segmentierung: Ground Truth (2), berechnet (3).

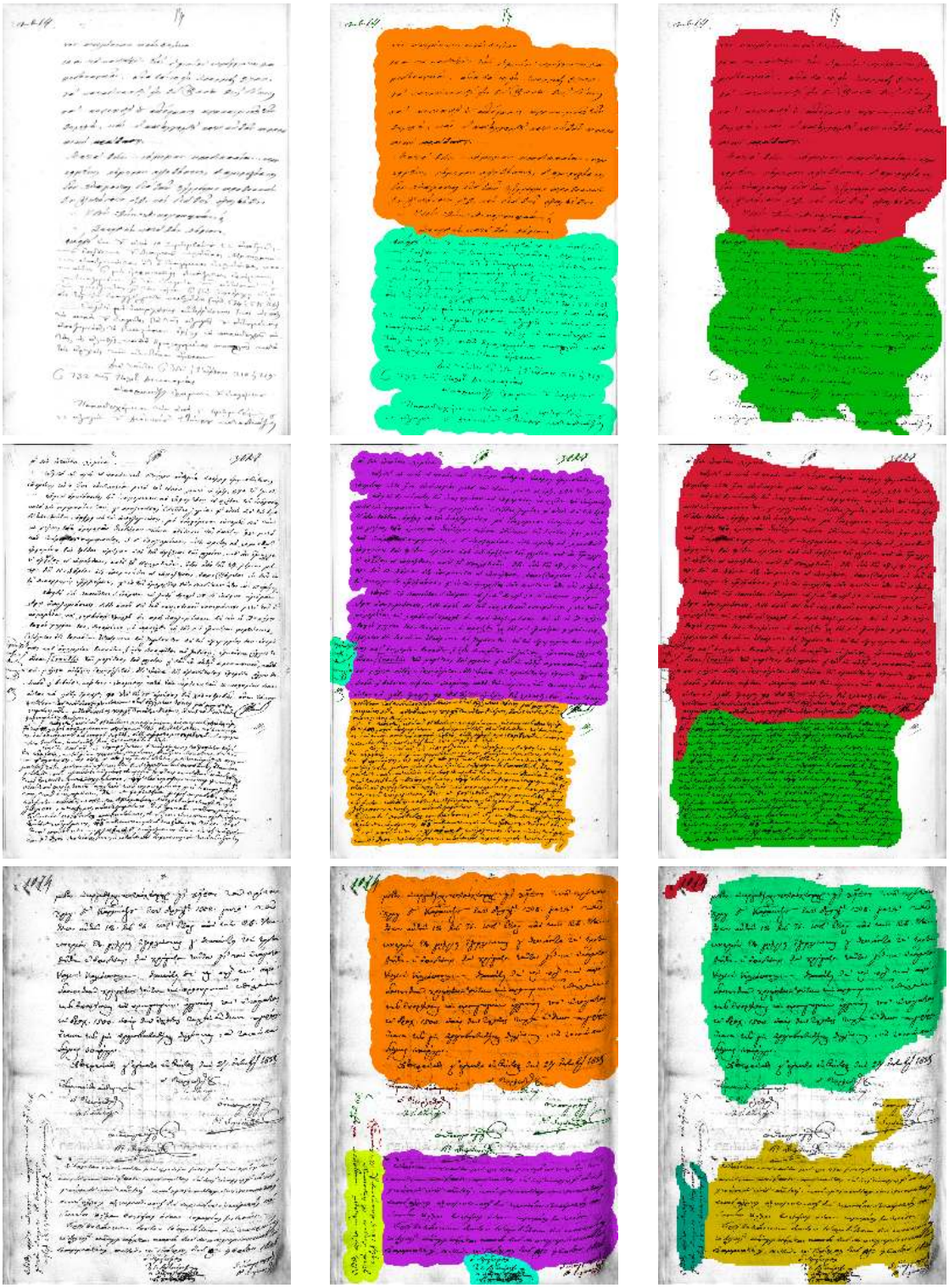


Abbildung C.5.: Manuskript 13-15. Spaltenweise: Original (randbereinigt) (1), Segmentierung: Ground Truth (2), berechnet (3).

D. Testdatensatz Varia

Quellen der verwendeten Manuskripte:

	Signatur	**Seite**
	URL	
1.	MS380808D-2ShK	Q2:91
	http://digital.soas.ac.uk/LOAA003338	
2.	MS380808D-4MM	-
	http://digital.soas.ac.uk/LOAA003340/00001	
3.	MS Or.860	-
	https://cudl.lib.cam.ac.uk/view/MS-OR-00860/1	
4.	Ms. or. fol. 547 b	-
	http://resolver.staatsbibliothek-berlin.de/SBB000083E600000000	
5.	MS FJ.1000.3	74
	https://cudl.lib.cam.ac.uk/view/MS-FJ-01000-00003	

Alle verwendeten Manuskripte unterstehen der Creative Commons Lizenz.
Manuskripte 1+2: © SOAS, University of London. CC BY-NC-SA
Manuskript 3: © Cambridge University Library CC BY-NC
Manuskript 4: © Staatsbibliothek Berlin. CC BY-NC-SA
Manuskript 5: © Cambridge University Library CC BY-NC

D.1. Abgleich ohne Segmentierung

Tabelle D.1.: Vergleich zwischen Ground Truth und berechneten Werten ohne Segmentierung. HTB = Haupttextblock, Flächenanteil bezogen auf die Summe aller Pixel der Randtextblockregionen.

Weitere		Differenz		Flächen-anteil
Seite	Text-bereich	Orientie-rung [°]	Zeilen-abstand	
1	HTB	2,97	9,75%	-
	1	4,82	30,28%	3,84%
	2	7,18	39,36%	4,70%
	3	15,52	33,55%	7,78%
	4	12,28	42,78%	3,61%
	5	11,64	26,58%	3,34%
	6	12,51	25,11%	22,38%
	7	1,07	37,26%	0,51%
	8	20,48	33,68%	8,66%
	9	21,77	36,01%	3,67%
	10	10,34	11,90%	21,73%
	11	3,97	9,57%	19,79%
2	HTB	3,35	5,44%	-
	1	53,00	84,55%	6,12%
	2	17,92	31,36%	4,38%
	3	14,14	84,05%	3,48%
	4	35,67	72,53%	4,43%
	5	16,47	37,77%	6,21%
	6	8,27	50,81%	4,26%
	7	15,95	83,48%	2,98%
	8	29,60	36,45%	43,50%
	9	39,97	41,38%	19,07%
	10	45,19	60,68%	5,56%
3	HTB	0,24	4,09%	-
4	1	38,71	63,86%	4,02%
	2	0,04	4,47%	19,12%
	3	2,59	6,92%	25,54%
	4	1,63	6,39%	27,77%
	5	0,41	3,97%	23,55%
5	1	1,48	1,90%	50,17%
	2	5,67	18,75%	49,83%

D.2. Abgleich mit Segmentierung

Tabelle D.2.: Abgleich unter Berücksichtigung der Segmentierung.Vergleich zwischen Ground Truth und berechneten Werten mit Segmentierung.
Zur Erklärung der Berechnung der Werte siehe Abschnitt 4.4.2

Weitere		>90%	>80%	>70%	>50%	False Negative	False Positive
Seite							
1	Flächenanteil	0,00%	9,10%	58,40%	72,00%	0,00%	0,36%
	Korrespondenzen	0,00%	8,30%	16,70%	41,70%	0,00%	8,30%
	ZA-differenz [%]	-	2,76%	9,26%	24,22%		
	O-differenz [°]	-	0,30	1,00	0,91		
2	Flächenanteil	0,00%	58,50%	58,50%	59,40%	2,12%	0,12%
	Korrespondenzen	0,00%	9,10%	9,10%	18,20%	9,10%	9,10%
	ZA-differenz [%]	-	2,82%	-	17,43%		
	O-differenz [°]	-	0,06	-	6,66		
3	Flächenanteil	97,20%	97,20%	97,20%	97,20%	0,00%	7,03%
	Korrespondenzen	100,00%	100,00%	100,00%	100,00%	0,00%	100,00%
	ZA-differenz [%]	1,16%	-	-	-		
	O-differenz [°]	0,00	-	-	-		
4	Flächenanteil	48,40%	48,40%	84,30%	84,30%	0,00%	1,73%
	Korrespondenzen	40,00%	40,00%	100,00%	100,00%	0,00%	180,00%
	ZA-differenz [%]	1,95%	-	23,66%	-		
	O-differenz [°]	0,61	-	14,06	-		
5	Flächenanteil	46,90%	46,90%	46,90%	46,90%	0,00%	20,13%
	Korrespondenzen	50,00%	50,00%	50,00%	50,00%	0,00%	50,00%
	ZA-differenz [%]	16,75%	-	-	-		
	O-differenz [°]	0,76	-	-	-		

D.3. Darstellung der Manuskriptseiten

Die Farben der Textblockregionen wurden zufällig gewählt, unterschiedliche Farben bedeuten lediglich unterschiedliche Textblöcke.

Abbildung D.1.: Manuskript 1-3. Spaltenweise: Original (randbereinigt) (1), Segmentierung: Ground Truth (2), berechnet (3).

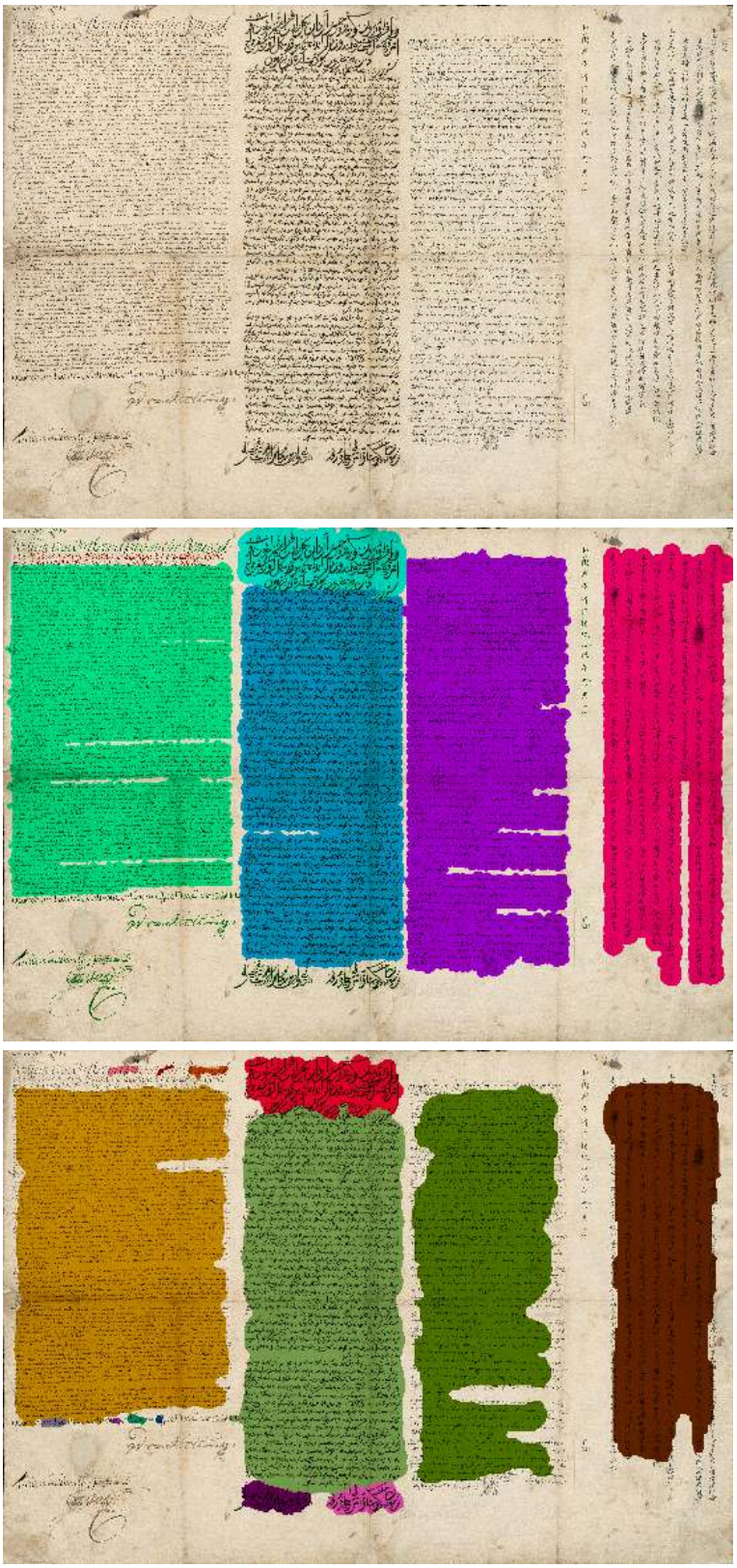


Abbildung D.2.: Manuskript 4. Original (randbereinigt) (oben),
Segmentierung: Ground Truth (mittig), berechnet (unten).

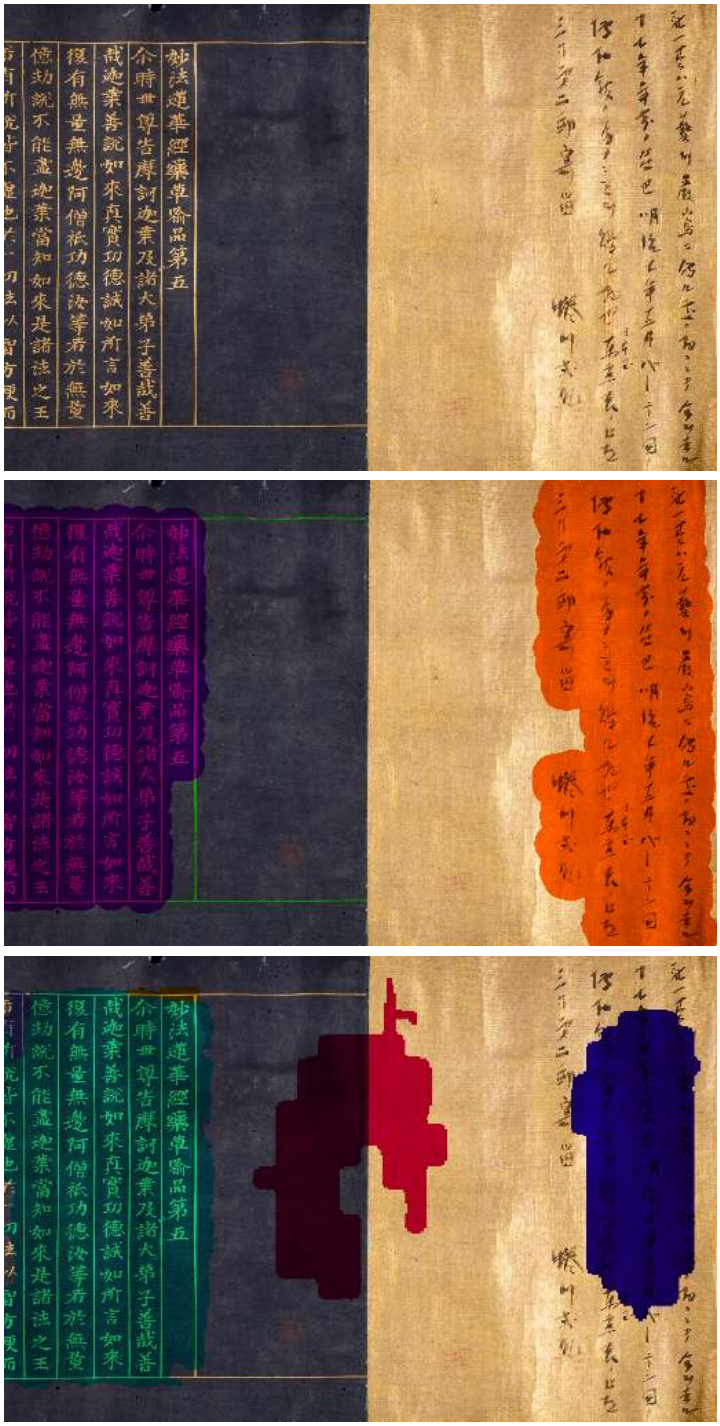


Abbildung D.3.: Manuskript 5. Original (randbereinigt) (oben),
Segmentierung: Ground Truth (mittig), berechnet (unten).

Literaturverzeichnis

[Antonacopoulos u. a. 2011] ANTONACOPOULOS, Apostolos ; CLAUSNER, Christian ; PAPADOPOULOS, Christos ; PLETSCHACHER, Stefan: Historical Document Layout Analysis Competition. In: *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, 2011. DOI: http://dx.doi.org/10.1109/ICDAR.2011.301. – ISSN 1520–5363, S. 1516–1520

[Antonacopoulos u. a. 2013] ANTONACOPOULOS, Apostolos ; CLAUSNER, Christian ; PAPADOPOULOS, Christos ; PLETSCHACHER, Stefan: ICDAR 2013 Competition on Historical Newspaper Layout Analysis (HNLA 2013). In: *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, 2013. DOI: http://dx.doi.org/10.1109/ICDAR.2013.293. – ISSN 1520–5363, S. 1454–1458

[Aradhye 2005] ARADHYE, Hrishikesh B.: A generic method for determining up/down orientation of text in roman and non-roman scripts. In: *Pattern Recognition* 38 (2005), Nr. 11, S. 2114–2131. DOI: http://dx.doi.org/10.1016/j.patcog.2004.12.011. – ISSN 0031–3203

[Arica u. Yarman-Vural 2001] ARICA, Nafiz. ; YARMAN-VURAL, Fatos T.: An Overview Of Character Recognition Focused On Off-line Handwriting. In: *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* 31 (2001), May, Nr. 2, S. 216–233. DOI: http://dx.doi.org/10.1109/5326.941845. – ISSN 1094–6977

[Asi u. a. 2014] ASI, Abedelkadir ; COHEN, Rafi ; KEDEM, Klara ; DINSTEIN, Its'hak ; EL-SANA, Jihad: A Coarse-to-Fine Approach for Layout Analysis of Ancient Manuscripts. In: *Frontiers in Handwriting Recognition (ICFHR), 2014 International Conference on*, 2014. DOI: http://dx.doi.org/10.1109/ICFHR.2014.31. – ISSN 2167–6445, S. 140–145

[Ávila u. Lins 2004] ÁVILA, Bruno T. ; LINS, Rafael D.: Efficient Removal of Noisy Borders from Monochromatic Documents. In: CAMPILHO, Aurélio (Hrsg.) ; KAMEL, Mohamed (Hrsg.): *Image Analysis and Recognition: International Conference, ICIAR 2004, Porto, Portugal, September 29 - October 1, 2004, Proceedings, Part II*, Springer Berlin Heidelberg, 2004. DOI: http://dx.doi.org/10.1007/978-3-540-30126-4_31. – ISBN 978–3–540–30126–4, S. 249–256

[Baird u. a. 1990] BAIRD, Henry S. ; JONES, Susan E. ; FORTUNE, Steven J.: Image Segmentation by Shape-Directed Covers. In: *Pattern Recognition, 1990. Proceedings.,*

10th International Conference on Bd. 1, 1990. DOI: http://dx.doi.org/10.1109/ICPR.1990.118223, S. 820–825

[Breuel 2002] BREUEL, Thomas M.: Two Geometric Algorithms for Layout Analysis. DOI: http://dx.doi.org/10.1007/3-540-45869-7_23. In: LOPRESTI, Daniel (Hrsg.) ; HU, Jianying (Hrsg.) ; KASHI, Ramanujan (Hrsg.): *Document Analysis Systems V, Proceedings of the 5th International Workshop on* Bd. 2423. Springer Berlin Heidelberg, 2002. – ISBN 978–3–540–44068–0, S. 188–199

[Bukhari u. a. 2012a] BUKHARI, Syed S. ; BREUEL, Thomas M. ; ASI, Abedelkadir ; EL-SANA, Jihad: Layout Analysis for Arabic Historical Document Images Using Machine Learning. In: *Frontiers in Handwriting Recognition (ICFHR), 2012 International Conference on*, 2012. DOI: http://dx.doi.org/10.1109/ICFHR.2012.227, S. 639–644

[Bukhari u. a. 2012b] BUKHARI, Syed S. ; SHAFAIT, Faisal ; BREUEL, Thomas M.: Border Noise Removal of Camera-Captured Document Images Using Page Frame Detection. In: IWAMURA, Masakazu (Hrsg.) ; SHAFAIT, Faisal (Hrsg.): *Camera-Based Document Analysis and Recognition*, Springer Berlin Heidelberg, 2012 (Lecture Notes in Computer Science (7139)). DOI: http://dx.doi.org/10.1007/978-3-642-29364-1. – ISBN 978–3–642–29364–1, S. 126–137

[Bulacu u. a. 2007] BULACU, Marius ; KOERT, Rutger van ; SCHOMAKER, Lambert ; ZANT, Tijn van d.: Layout Analysis of Handwritten Historical Documents for Searching the Archive of the Cabinet of the Dutch Queen. In: *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on* Bd. 1, 2007. DOI: http://dx.doi.org/10.1109/ICDAR.2007.4378732. – ISSN 1520–5363, S. 357–361

[Chang u. Lin 2011] CHANG, Chih-Chung ; LIN, Chih-Jen: LIBSVM: A library for support vector machines. In: *ACM Transactions on Intelligent Systems and Technology* 2 (2011), S. 27:1–27:27

[Daub 2013] DAUB, Frederike-Wiebke: Standards and Specifics — the Layout of Arabic Didactic Poems in Manuscripts. In: *manuscript cultures* 5 (2013), S. 52–67. – ISSN 1867–9617

[Doermann 1995] DOERMANN, David: Page Decomposition and Related Research at the University of Maryland. In: *Document Image Understanding Technology (SDIUT), Symposium on*, 1995, S. 39–55

[Feichtinger u. a. 2008] *Kapitel* 4.2. In: FEICHTINGER, Hans G. ; HELFFER, Bernard ; LAMOUREUX, Michael P. ; LERNER, Nicolas ; TOFT, Joachim: *Pseudo-Differential Operators: Quantization and Signals*. Springer Berlin Heidelberg, 2008 (Lecture Notes in Mathematics). – ISBN 9783540682660. DOI: http://dx.doi.org/10.1007/978-3-540-68268-4

[Feichtinger u. Strohmer 1998] FEICHTINGER, Hans G. ; STROHMER, Thomas: *Gabor Analysis and Algorithms: Theory and Applications*. Birkhäuser Boston, 1998 (Applied and Numerical Harmonic Analysis). – S. 6–10. DOI: http://dx.doi.org/10.1007/978-1-4612-2016-9. – ISBN 9781461220169

[FFTW] *The Fastest Fourier Transform in the West*. http://www.fftw.org

[Frigo u. Johnson 1998] FRIGO, Matteo ; JOHNSON, Steven G.: FFTW: An adaptive software architecture for the FFT. In: *Proc. 1998 IEEE Intl. Conf. Acoustics Speech and Signal Processing* Bd. 3, IEEE, 1998. DOI: http://dx.doi.org/10.1109/ICASSP.1998.681704, S. 1381–1384

[Gabor 1946] GABOR, Dennis: Theory of communication. In: *Journal of the Institution of Electrical Engineers* 93 (1946), Nr. 26, S. 429–457

[Garz u. a. 2011] GARZ, Angelika ; SABLATNIG, Robert ; DIEM, Markus: Layout Analysis for Historical Manuscripts Using Sift Features. In: *Document Analysis and Recognition (ICDAR), International Conference on*, 2011. DOI: http://dx.doi.org/10.1109/ICDAR.2011.108. – ISSN 1520–5363, S. 508–512

[Gatos u. a. 2005] GATOS, Basilios ; DANATSAS, Dimitrios ; PRATIKAKIS, Ioannis ; PERANTONIS, StavrosJ.: Automatic Table Detection in Document Images. DOI: http://dx.doi.org/10.1007/11551188_67. In: SINGH, Sameer (Hrsg.) ; SINGH, Maneesha (Hrsg.) ; APTE, Chid (Hrsg.) ; PERNER, Petra (Hrsg.): *Pattern Recognition and Data Mining* Bd. 3686. Springer Berlin Heidelberg, 2005. – ISBN 978–3–540–28757–5, S. 609–618

[Gatos u. a. 2014] GATOS, Basilis ; LOULOUDIS, Georgios ; STAMATOPOULOS, Nikolaos: Segmentation of Historical Handwritten Documents into Text Zones and Text Lines. In: *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, 2014. DOI: http://dx.doi.org/10.1109/ICFHR.2014.84. – ISSN 2167–6445, S. 464–469

[Gatos u. a. 2010] GATOS, Basilis ; STAMATOPOULOS, Nikolaos ; LOULOUDIS, Georgios: ICFHR 2010 Handwriting Segmentation Contest. In: *Frontiers in Handwriting Recognition (ICFHR), 2010 International Conference on*, 2010. DOI: http://dx.doi.org/10.1109/ICFHR.2010.120, S. 737–742

[Giotis u. a. 2017] GIOTIS, Angelos P. ; SFIKAS, Giorgos ; GATOS, Basilis ; NIKOU, Christophoros: A Survey of Document Image Word Spotting Techniques. In: *Pattern Recogn.* 68 (2017), August, Nr. C, S. 310–332. DOI: http://dx.doi.org/10.1016/j.patcog.2017.02.023. – ISSN 0031–3203

[Gonzalez u. Woods 2018] GONZALEZ, Raphael C. ; WOODS, Richard E.: *Digital Image Processing (4th Edition*. Pearson PLC, 2018. – ISBN 1–292–22304–9

[Grana u. a. 2011] GRANA, Costantino ; BORGHESANI, Daniele ; CUCCHIARA, Rita: Automatic segmentation of digitalized historical manuscripts. In: *Multimedia Tools and Applications* 55 (2011), Nr. 3, S. 483–506. DOI: http://dx.doi.org/10.1007/s11042-010-0561-8. – ISSN 1380–7501

[Greek State Archives] GREEK STATE ARCHIVES: *The digital collections of the Greek State Archives*. http://arxeiomnimon.gak.gr/en/index.html

[Harris 1978] HARRIS, Frederic J.: On the use of windows for harmonic analysis with the discrete Fourier transform. In: *Proceedings of the IEEE* 66 (1978), Jan, Nr. 1, S. 51–83. DOI: http://dx.doi.org/10.1109/PROC.1978.10837. – ISSN 0018–9219

[Heinzel u. a. 2002] HEINZEL, Gerhard ; RÜDIGER, Albrecht ; SCHILLING, Roland: Spectrum and spectral density estimation by the Discrete Fourier transform (DFT), including a comprehensive list of window functions and some new at-top windows. (2002), Feb. http://edoc.mpg.de/395068

[Herzog 2007] HERZOG, Rainer: *Analyse historischer chinesischer Manuskripte mit Hilfe des Shape Context Verfahrens*, Universität Hamburg, Diplomarbeit, 12 2007

[Herzog u. a. 2013] HERZOG, Rainer ; SOLTH, Arved ; NEUMANN, Bernd: Using Harris Corners for the Retrieval of Graphs in Historical Manuscripts. In: *Document Analysis and Recognition (ICDAR), 2013 International Conference on*, 2013. DOI: http://dx.doi.org/10.1109/ICDAR.2013.262, S. 1295–1299

[HisDoc] *Historical Document Analysis, Recognition and Retrieval*. http://diuf.unifr.ch/main/hisdoc/

[Howe 2011] HOWE, Nicholas R.: A Laplacian Energy for Document Binarization. In: *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, 2011. DOI: http://dx.doi.org/10.1109/ICDAR.2011.11. – ISSN 1520–5363, S. 6–10

[Howe 2013] HOWE, Nicholas R.: Document binarization with automatic parameter tuning. In: *International Journal on Document Analysis and Recognition (IJDAR)* 16 (2013), Nr. 3, S. 247–258. DOI: http://dx.doi.org/10.1007/s10032-012-0192-x. – ISSN 1433–2833

[Jain u. Bhattacharjee 1992] JAIN, Anil K. ; BHATTACHARJEE, Sushil: Text Segmentation Using Gabor Filters for Automatic Document Processing. In: *Machine Vision and Applications* 5 (1992), Nr. 3, S. 169–184. DOI: http://dx.doi.org/10.1007/BF02626996. – ISSN 0932–8092

[Jain u. Yu 1998] JAIN, Anil K. ; YU, Bin: Document Representation and Its Application to Page Decomposition. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 20 (1998), Mar, Nr. 3, S. 294–308. DOI: http://dx.doi.org/10.1109/34.667886. – ISSN 0162–8828

[Kise u. a. 1998] KISE, Koichi ; SATO, Akinori ; IWATA, Motoi: Segmentation of Page Images Using the Area Voronoi Diagram. In: *Computer Vision and Image Understanding* 70 (1998), Nr. 3, S. 370–382. DOI: http://dx.doi.org/10.1006/cviu.1998.0684. – ISSN 1077–3142

[Konya u. a.] KONYA, Iuliu V. ; EICKELER, Stefan ; SEIBERT, Christoph ; KÖHLER, Joachim: *The Historical Archive/Newspaper Edition*. http://mmprec.iais.fraunhofer.de/document-layout

[Köthe 2000] KÖTHE, Ullrich: *Generische Programmierung für die Bildverarbeitung*. Books on Demand GmbH, 2000. – ISBN 3–8311–0239–2

[Likforman-Sulem u. a. 2007] LIKFORMAN-SULEM, Laurence ; ZAHOUR, Abderrazak ; TACONET, Bruno: Text line segmentation of historical documents: a survey. In: *International Journal of Document Analysis and Recognition (IJDAR)* 9 (2007), Nr. 2-4, S. 123–138. DOI: http://dx.doi.org/10.1007/s10032-006-0023-z. – ISSN 1433–2833

[Lowe 2004] LOWE, David G.: Distinctive Image Features from Scale-Invariant Keypoints. In: *International Journal of Computer Vision* 60 (2004), Nr. 2, S. 91–110. DOI: http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94. – ISSN 0920–5691

[Lu u. a. 2010] LU, Shijian ; SU, Bolan ; TAN, Chew L.: Document image binarization using background estimation and stroke edges. In: *International Journal on Document Analysis and Recognition (IJDAR)* 13 (2010), Nr. 4, S. 303–314. DOI: http://dx.doi.org/10.1007/s10032-010-0130-8. – ISSN 1433–2833

[Lu u. Tan 2006] LU, Shijian ; TAN, Chew L.: Automatic Document Orientation Detection and Categorization Through Document Vectorization. In: *Proceedings of the 14th ACM International Conference on Multimedia*. New York, NY, USA : ACM, 2006 (MM '06). DOI: http://dx.doi.org/10.1145/1180639.1180673. – ISBN 1–59593–447–2, S. 113–116

[Mao u. a. 2003] MAO, Song ; ROSENFELD, Azriel ; KANUNGO, Tapas: Document Structure Analysis Algorithms: A Literature Survey. In: *Proc. SPIE* 5010 (2003), S. 197–207. DOI: http://dx.doi.org/10.1117/12.476326

[Meine u. Köthe 2005] MEINE, Hans ; KÖTHE, Ullrich: The GeoMap: A Unified Representation for Topology and Geometry. In: BRUN, Luc. (Hrsg.) ; VENTO, Mario (Hrsg.): *Graph-Based Representations in Pattern Recognition. GbRPR* Bd. 3434, Springer Berlin Heidelberg, 2005 (Lecture Notes in Computer Science). DOI: http://dx.doi.org/10.1007/978-3-540-31988-7_12. – ISBN 978–3–540–25270–2, S. 132–141

[Moll u. a. 2008] MOLL, Michael A. ; BAIRD, Henry S. ; AN, Chang: Truthing for Pixel-Accurate Segmentation. In: *Document Analysis Systems, 2008. DAS '08. The Eighth IAPR International Workshop on*, 2008. DOI: http://dx.doi.org/10.1109/DAS.2008.47, S. 379–385

[Nadler 1984] NADLER, Morton: Survey Document Segmentation and Coding Techniques. In: *Computer Vision, Graphics, and Image Processing* 28 (1984), Nr. 2, S. 240–262. DOI: http://dx.doi.org/10.1016/S0734-189X(84)80025-0. – ISSN 0734–189X

[Nagy u. Seth 1984] NAGY, George ; SETH, Sharad C.: Hierarchical Image Representation with Application to Optically Scanned Documents. In: *Proceedings of the Seventh International Conference on Pattern Recognition*, 1984, S. 347 – 349

[Nagy u. a. 1986] NAGY, George ; SETH, Sharad C. ; STODDARD, Spotswood D.: Document Analysis with an Expert System. DOI: http://dx.doi.org/10.1016/B978-0-444-87877-9.50016-9. In: *Pattern Recognition in Practice*. Elsevier Amsterdam, 1986. – ISBN 978–0–444–87877–9, S. 149–159

[Nagy u. a. 1992] NAGY, George ; SETH, Sharad C. ; VISWANATHAN, Mahesh: A Prototype Document Image Analysis System for Technical Journals. In: *Computer* 25 (1992), July, Nr. 7, S. 10–22. DOI: http://dx.doi.org/10.1109/2.144436

[O'Gorman 1993] O'GORMAN, Lawrence: The Document Spectrum for Page Layout Analysis. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 15 (1993), Nov, Nr. 11, S. 1162–1173. DOI: http://dx.doi.org/10.1109/34.244677. – ISSN 0162–8828

[Orientalisches Institut, Universität Leipzig] ORIENTALISCHES INSTITUT, UNIVERSITÄT LEIPZIG: *Islamic Manuscripts at the Leipzig University Library*. http://www.islamic-manuscripts.net

[Otsu 1979] OTSU, Nobuyuki: A Threshold Selection Method from Gray-Level Histograms. In: *Systems, Man and Cybernetics, IEEE Transactions on* 9 (1979), Jan, Nr. 1, S. 62–66. DOI: http://dx.doi.org/10.1109/TSMC.1979.4310076. – ISSN 0018–9472

[Ouwayed u. Belaïd 2008] OUWAYED, Nazih ; BELAÏD, Abdel: Multi-oriented Text Line Extraction from Handwritten Arabic Documents. In: *Document Analysis Systems, 2008. DAS '08. The Eighth IAPR International Workshop on*, 2008. DOI: http://dx.doi.org/10.1109/DAS.2008.14, S. 339–346

[Pavlidis u. Zhou 1992] PAVLIDIS, Theo ; ZHOU, Jiangying: Page segmentation and classification. In: *CVGIP: Graphical Models and Image Processing* 54 (1992), Nr. 6, S. 484–496. DOI: http://dx.doi.org/10.1016/1049-9652(92)90068-9. – ISSN 1049–9652

[Postl 1986] POSTL, Wolfgang: Detection of Linear Oblique Structures and Skew Scan in Digitized Documents. In: *Pattern Recognition, International Conference on*, 1986, S. 687–689

[Pratikakis u. a.] PRATIKAKIS, Ioannis ; ZAGORIS, Konstantinos ; BARLAS, George ; GATOS, Basilis: *Dataset for the Competition on Handwritten Document Image Binarization (H-DIBCO 2016)*. http://vc.ee.duth.gr/h-dibco2016/benchmark/

[Pratikakis u. a. 2016] PRATIKAKIS, Ioannis ; ZAGORIS, Konstantinos ; BARLAS, George ; GATOS, Basilis: ICFHR2016 Handwritten Document Image Binarization Contest (H-DIBCO 2016). In: *Frontiers in Handwriting Recognition (ICFHR), 2016 15th International Conference on*, 2016. DOI: http://dx.doi.org/10.1109/ICFHR.2016.0118. – ISSN 2167–6445, S. 619–623

[Rajan u. Stiehl 2018] RAJAN, Vinodh ; STIEHL, H. S.: Bringing Paleography to the Table: Developing an Interactive Manuscript Exploration System for Large Multi-Touch Devices. (2018)

[Ramel u. a. 2007] RAMEL, Jean-Yves ; LERICHE, Stéphane ; DEMONET, Marie-Luce ; BUSSON, Sébastien: User-driven page layout analysis of historical printed books. In: *International Journal of Document Analysis and Recognition (IJDAR)* 9 (2007), Nr. 2-4, S. 243–261. DOI: http://dx.doi.org/10.1007/s10032-007-0040-6. – ISSN 1433–2833

[Roerdink u. Meijster 2000] ROERDINK, Jos B. ; MEIJSTER, Arnold: The Watershed Transform: Definitions, Algorithms and Parallelization Strategies. In: *Fundamenta Informaticae* 41 (2000), Nr. 1,2, S. 187–228. DOI: http://dx.doi.org/10.3233/FI-2000-411207. – ISSN 0169–2968

[Sauvola u. Pietikäinen 2000] SAUVOLA, Jaakko ; PIETIKÄINEN, Matti: Adaptive document image binarization. In: *Pattern Recognition* 33 (2000), Nr. 2, S. 225 – 236. DOI: http://dx.doi.org/10.1016/S0031-3203(99)00055-2. – ISSN 0031–3203

[Selfridge u. Neisser 1960] SELFRIDGE, Oliver G. ; NEISSER, Ulric: Pattern Recognition by Machine. In: *Scientific American* 203 (1960), S. 60–68. DOI: http://dx.doi.org/10.1038/scientificamerican0860-60. – ISSN 0036–8733

[SfarData] *The Codicological Data-Base of the Hebrew Palaeography Project*. http://sfardata.nli.org.il/sfardatanew/home.aspx

[SFB950] *Sonderforschungsbereich Manuskriptkulturen in Asien, Afrika und Europa, Projektbereich B: Visuelle Organisation*. https://www.manuscript-cultures.uni-hamburg.de/Projekte_p2.html#PBB

[Shafait 2008] SHAFAIT, Faisal: *Geometric Layout Analysis of Scanned Documents*, Technische Universiät Kaiserslautern, Diss., 2008

[Shafait u. a. 2008] SHAFAIT, Faisal ; KEYSERS, Daniel ; BREUEL, Thomas M.: Performance Evaluation and Benchmarking of Six-Page Segmentation Algorithms. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30 (2008), Nr. 6, S. 941–954. DOI: http://dx.doi.org/10.1109/TPAMI.2007.70837. – ISSN 0162–8828

[Simistira u. a. 2017] SIMISTIRA, Fotini ; BOUILLON, Manuel ; SEURET, Mathias ; WÜRSCH, Marcel ; ALBERTI, Michele ; INGOLD, Rolf ; LIWICKI, Marcus: ICDAR2017 Competition on Layout Analysis for Challenging Medieval Manuscripts.

In: *Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference on*, 2017. DOI: http://dx.doi.org/10.1109/ICDAR.2017.223. – ISSN 2379–2140, 1361–1370

[Sonka u. a. 1998] SONKA, Milan ; HLAVAC, Vaclav ; BOYLE, Roger: *Image Processing, Analysis, and Machine Vision*. PWS Publishing, 1998. – ISBN 0–534–95393–X

[Srihari u. Govindaraju 1989] SRIHARI, Sargur N. ; GOVINDARAJU, Venugopal: Analysis of textual images using the Hough transform. In: *Machine Vision and Applications* 2 (1989), Jun, Nr. 3, S. 141–153. DOI: http://dx.doi.org/10.1007/BF01212455. – ISSN 1432–1769

[Staatsbibliothek zu Berlin] STAATSBIBLIOTHEK ZU BERLIN: *Datenbank der orientalischen Handschriften der Staatsbibliothek zu Berlin*. http://orient-digital.staatsbibliothek-berlin.de

[Stamatopoulos u. a. 2013] STAMATOPOULOS, Nikolas ; GATOS, Basilis ; LOULOUDIS, Georgios ; PAL, Umapada ; ALAEI, Alireza: ICDAR 2013 Handwriting Segmentation Contest. In: *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, 2013. DOI: http://dx.doi.org/10.1109/ICDAR.2013.283. – ISSN 1520–5363, S. 1402–1406

[Testset] *Test set of 50 manuscript images, including Ground Truth*. http://kogs-www.informatik.uni-hamburg.de/%7Eherzog/Testset

[Wahl u. a. 1982] WAHL, Friedrich M. ; WONG, Kwan Y. ; CASEY, Richard G.: Block Segmentation and Text Extraction in Mixed Text/Image Documents. In: *Computer Graphics and Image Processing* 20 (1982), Nr. 4, S. 375–390. DOI: http://dx.doi.org/10.1016/0146-664X(82)90059-4. – ISSN 0146–664X

[Xu u. a. 2017] XU, Yue ; HE, Wenhao ; YIN, Fei ; LIU, Cheng-Lin: Page Segmentation for Historical Handwritten Documents Using Fully Convolutional Networks. In: *Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference on*, 2017. DOI: http://dx.doi.org/10.1109/ICDAR.2017.94. – ISSN 2379–2140, S. 541–546

Eidesstattliche Versicherung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Hamburg, 30.06.2018