

# Zeitschrift für digitale Geisteswissenschaften

---

Artikel aus:

Sonderband 1 der ZfdG: Grenzen und Möglichkeiten der Digital Humanities. Hg. von Constanze Baum und Thomas Stäcker. 2015. DOI: [10.17175/sb01](https://doi.org/10.17175/sb01)

Titel:

Vom Zeichen zur Schrift: Mit Mustererkennung zur automatisierten Schreiberhanderkennung in mittelalterlichen und frühneuzeitlichen Handschriften

---

Autor/in:

Daniel Fecker

Kontakt: [fecker@ifn.ing.tu-bs.de](mailto:fecker@ifn.ing.tu-bs.de)

Institution: Technische Universität Braunschweig

GND: [1084240270](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-63862-p0011-9) ORCID:

---

Autor/in:

Volker Märgner

Kontakt: [maergner@ifn.ing.tu-bs.de](mailto:maergner@ifn.ing.tu-bs.de)

Institution: Technische Universität Braunschweig

GND: [1084238829](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-63862-p0011-9) ORCID:

---

Autor/in:

Torsten Schaßan

Kontakt: [schassan@hab.de](mailto:schassan@hab.de)

Institution: Herzog August Bibliothek Wolfenbüttel

GND: [1027794041](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-63862-p0011-9) ORCID:

---

DOI des Artikels:

[10.17175/sb001\\_008](https://doi.org/10.17175/sb001_008)

Nachweis im OPAC der Herzog August Bibliothek:

[830167242](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-63862-p0011-9)

Erstveröffentlichung:

19.02.2015

Lizenz:

Sofern nicht anders angegeben 

Medienlizenzen:

Medienrechte liegen bei den Autoren

Letzte Überprüfung aller Verweise:

24.05.2016

GND-Verschlagwortung:

[Mustererkennung](#) | [automatische Klassifikation](#) | [Handschrift \(Grafologie\)](#) |

Zitierweise:

Daniel Fecker, Volker Märgner, Torsten Schaßan: Vom Zeichen zur Schrift: Mit Mustererkennung zur automatisierten Schreiberhanderkennung in mittelalterlichen und frühneuzeitlichen Handschriften. In: Grenzen und Möglichkeiten der Digital Humanities. Hg. von Constanze Baum / Thomas Stäcker. 2015 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 1). PDF Format ohne Paginierung. Als text/html abrufbar unter DOI: [10.17175/sb001\\_008](https://doi.org/10.17175/sb001_008).

Daniel Fecker, Volker Märgner, Torsten Schaßan

## **Vom Zeichen zur Schrift: Mit Mustererkennung zur automatisierten Schreiberhanderkennung in mittelalterlichen und frühneuzeitlichen Handschriften**

---

### Abstracts

Für die Digital Humanities im Bereich Mediävistik und Frühnezeitforschung stellt die Digitalisierung von Handschriften ein zentrales Feld dar. Da jede Handschrift eigene Charakteristika aufweist, führt die automatische Erstellung eines maschinenlesbaren Textes durch Optical Character Recognition (OCR) anhand von Digitalisaten in den allermeisten Fällen zu fehlerhaften Ergebnissen. Andererseits können Charakteristika dieser Schrift wie Buchstabengröße und -abstand, Dichte des Schriftbildes, Neigung u.a. genutzt werden, um die Identifikation der schreibenden Hand bzw. Hände zu ermöglichen. In dem Beitrag wird gezeigt, wie die Analyse von Handschriftenabbildungen zur Identifikation der schreibenden Hand bzw. Hände genutzt werden kann. Ein Algorithmus soll sonstige paläographische oder kodikologische Befunde unterstützen und Argumente zur Veri- oder Falsifikation von unsicheren Zuschreibungen liefern.

For Digital Humanities in medieval studies and early modern studies, the digitization of manuscripts is a central field. Since each manuscript displays its own unique characteristics, the automatic generation of a machine-readable text using Optical Character Recognition (OCR) as applied to digital images leads, in most cases, to error-prone results. However, characteristics of handwriting such as the size of letters and spacing, slope, and so on can be used to identify the scribe or scribes. This paper demonstrates how the analysis of manuscript images can be used to identify the scribe or scribes. An algorithm will support additional paleographic and codicological findings and provide evidence for the verification or falsification of uncertain attributions.

## **1. Einleitung**

»Deutschland befindet sich in einer Phase intensiv betriebener und mit einem hohen finanziellen Aufwand verbundener Digitalisierung seiner historischen Bestände. Für die Mediävistik und Frühnezeitforschung stellt hierbei die Digitalisierung der dem Mittelalter und der Renaissance entstammenden Handschriften ein zentrales Feld dar.«<sup>1</sup>

Die Nutzung der Digitalisate allein als digitale Lesekopie durch den betrachtenden Forscher oder die Forscherin würde das Erkenntnispotential, das dem Digitalisat selbst innewohnt, schlechterdings vergeuden. Daher ist eine der ersten Fragen von ›traditionell‹ mit Handschriften arbeitenden Geisteswissenschaftlern in der Regel die, ob oder wie man aus den Bildern der Handschriften einen Text gewinnen kann.

---

<sup>1</sup> Vgl. Hays / Müller 2011, S. 416.

Die zentrale Anwendung wäre die Optical Character Recognition (OCR automatisierte Texterkennung), die der Herstellung eines maschinenlesbaren Textes aus bildhaft vorliegender Information dient. Die Güte der aus OCR-Algorithmen gewonnenen Texte ist von mehreren Faktoren abhängig, darunter auch von der Qualität der Vorlage sowie von der Qualität des Scanprozesses bzw. der daraus resultierenden Digitalisate.<sup>2</sup> Doch selbst wenn diese Faktoren theoretisch eine optimale Weiterverarbeitung möglich machen würden, böten historische und insbesondere handschriftliche Materialien weitere Herausforderungen: Schon bei Drucken mit gebrochenen, gotischen Schriften mit nur geringen Unterschieden zwischen einzelnen Buchstabenformen wie dem langen s -f- und dem normalen f produziert gängige OCR-Software in der Regel ein für eine wissenschaftliche Nutzung unbrauchbares, da zu fehlerbehaftetes Ergebnis. Vor diesem Hintergrund erscheint die Aufgabe bei Handschriften noch schwieriger. Dass OCR bei der Verarbeitung von Handschriftendigitalisaten an ihre Grenzen stoßen muss, lässt sich anhand des formalen Ablaufs einer Texterkennung veranschaulichen:

- Anfertigen eines Digitalisates in geeigneter Qualität gegebenenfalls Bildkorrekturen, wie das Geraderücken schräg aufgenommener Seiten, Glättung von Rundungen aufgrund von Materialbiegung, etc.
- Binarisierung der Farbwerte zur deutlichen Trennung von Schrift und Hintergrund
- Segmentierung der Schrift, z.B. in Linien und Einzelworte
- Mustererkennung, bei der zu erkennende Formen mit einem vorhandenen Zeichenvorrat verglichen werden
- Im Falle der Übereinstimmung Zuweisung eines Zeichencodes nach üblicher Textkodierung (=UTF-8)

Die eigentliche Herausforderung an die OCR für Handschriften liegt in Arbeitsschritt 4, da für die Zuweisung eines zu speichernden Codes in Schritt 5 hier zunächst die Zuweisung eines Codes zu einer großen Zahl von Digitalisaten manuell durchgeführt werden muss (man spricht davon, die sogenannte ›ground truth‹ für die Digitalisate anzulegen). Diese Aufgabe ist sehr zeitaufwendig aber notwendig, um das sogenannte ›Training‹ eines Klassifikators zu ermöglichen, wobei der Klassifikator an die jeweilige Schriftart angepasst wird. Der Klassifikator ist Kern des OCR-Systems und ermöglicht es, Muster (pattern) in der Vorlage Zeichen aus dem Zeichenvorrat zuzuordnen. Diese auf einer festgesetzten Wahrheit beruhende Beziehung zwischen den Bildmustern und den kodierten Zeichen ist nur für die Typen von Mustern gültig, die zum Training verwendet wurden, also z. B. für eine Schrifttype oder eine bestimmte Schriftform. Da die Handschrift jedes Schreibers<sup>3</sup> eigene Charakteristika aufweist, welche sie zumindest von anderen Händen unterscheidbar macht, können die in den meisten Handschriftendigitalisaten aufgefundenen Muster nicht ohne weiteres eindeutig codierten Zeichen zugewiesen werden.

---

<sup>2</sup> Stäcker 2013, S. 123–135.

<sup>3</sup> Den Autoren ist bewusst, dass es im frühen Mittelalter Schreiberinnen gegeben hat. Die in diesem Beitrag behandelten Beispiele stammen jedoch durchweg aus dem Kloster Weissenburg, das ein Männer-Konvent war. Dort wird es nur Schreiber gegeben haben, weshalb im Folgenden durchgehend die männliche Form gewählt wird.

Die Varianz zwischen einzelnen Handschriften, vor allem von Buchschriften wie der karolingischen Minuskel, ist als vergleichsweise klein einzuschätzen, so dass die Anwendung von OCR auf diese Schriften demnach recht gute Ergebnisse erwarten lassen dürfte, allerdings ist die Aufgabe der Textgewinnung gerade für die in dieser Schrift geschriebenen Werke relativ uninteressant. Es handelt sich dabei um zumeist wohlbekannte und gut erforschte, in der Regel auch schon kritisch edierte Texte<sup>4</sup> oder um Texte, deren Varianz in den Handschriften als äußerst gering anzusetzen ist, wie z.B. Missale.

Aufgrund dieser Annahmen stand am Anfang der Kooperation zwischen dem Institut für Nachrichtentechnik an der TU Braunschweig (IFN) und der Herzog August Bibliothek Wolfenbüttel (HAB) die Überlegung, das Untersuchungsziel kurzerhand umzukehren und als Ergebnis der Analyse von Handschriftenabbildungen nicht einen durch OCR zu gewinnenden elektronischen Text anzustreben, sondern in der Schrift Merkmale zur Identifikation der schreibenden Hand zu herauszufiltern. Sollte es nämlich gelingen, anhand spezifischer Merkmale eine Schreiberhand (unter Abstraktion von nicht mehr quantifizierbaren Abweichungsquellen wie Lebensalter und Tagesform des Amanuensis oder dem Zustand der Schreibmaterialien und -utensilien) von anderen Schreiberhänden abzugrenzen, so könnte damit die wichtige Fragestellung nach dem Schreiber automatisiert werden.

Als Charakteristika dieser Schrift sollten Buchstabengröße und –abstand, Dichte des Schriftbildes, Neigung u. a. untersucht werden, aber nicht notwendig, wie in der klassischen, vom forschenden menschlichen Auge ausgehenden Paläographie, einzelne Buchstabenformen. Dabei werden in dieser Arbeit verschiedene Merkmale vorgestellt, die auf unterschiedlichen Grundprinzipien, wie der Schriftkontur, der durch die Periodizität des Schriftbildes zugrunde liegenden Textur und speziellen Schlüsselpunkten in der Schrift, beruhen.

Diese Überlegung lässt sich nun in zweierlei Richtung ausdeuten:

1. Zum einen könnte versucht werden, die Unterschiede einer Schrift in Relation zu anderen Schriften zu bringen und so eine der Grundfragen der Kodikologie und Paläographie beantworten zu helfen, die Frage nämlich, wieviele Schreiber bei der Entstehung der Handschrift mitgewirkt haben.
2. Zum anderen können mit vorhandenem Wissen über die Merkmale einer Schrift oder einer Schreiberhand andere Handschriftendigitalisate auf mögliche Identität oder Differenz hin untersucht werden. Es müsste so möglich sein, diese Schrift oder Hand in anderen Handschriften nachzuweisen bzw. die Differenz festzustellen, also einen anderen Schreiber anzunehmen. Wenn man die Schwellwerte (thresholds) entsprechend anpassen würde, sollte es auch möglich sein, Schriftfamilien voneinander zu trennen und bei einer entsprechenden Materialbasis Hinweise auf Datierung und Lokalisierung zu sammeln.

---

<sup>4</sup> Vgl. die Textausgaben im Rahmen der [Patrologia Latina](#) (PL), eine der ersten Werke, die als elektronischer Volltext verfügbar waren, oder der [Library of Latin Texts](#) (CLCLT).

Für die Beantwortung beider Fragen ist sowohl die Festsetzung der Schwellwerte als auch die Abschätzung zentral, welche räumliche Ausdehnung der Untersuchung zugrunde gelegt wird: Die Schwellwerte müssen genutzt werden, um kleinere Abweichungen zu tolerieren. Kein Schreiber ist in der Lage, in einem handschriftlichen Text zwei völlig identische Buchstabenformen zu erzeugen. Zu genaues Betrachten des Untersuchungsgegenstandes würde somit nur noch Unterschiede erkennen lassen und keinerlei Übereinstimmung mehr zu Tage fördern.<sup>5</sup> Die räumliche Ausdehnung ist insofern wichtig, da die in dieser Arbeit verwendeten Algorithmen immer ganze Seiten betrachten und daraus die Merkmalsvektoren berechnen. Unterschiede, die sich auf einer Seite befinden, beispielsweise durch Handwechsel mitten auf der Seite, können dabei allerdings nicht aufgefunden werden.

Die Erkennungsgenauigkeit muss dazu aufgrund des Trainings mit einer Handschrift in einem anderen Codex über einem zu definierenden Schwellwert (threshold) liegen, um als Indiz gewertet zu werden, dass derselbe Schreiber die Handschrift geschrieben haben könnte. Der angestrebte Algorithmus würde damit sonstige paläographische oder kodikologische Befunde unterstützende bzw. ergänzende Argumente zur Verifikation von unsicheren Zuschreibungen liefern. Im Gegenzug müsste das Unterschreiten dieses Schwellwertes Argumente für Fälschungen solcher Zuschreibungen ermöglichen.

Der typische Aufbau einer Prozesskette für die Identifikation eines Schreibers ist die einer OCR nicht unähnlich.<sup>6</sup> In einem ersten Schritt werden die gescannten Bilder eines Manuskriptes mit einer Vorverarbeitung für die Extraktion von geeigneten Merkmalen aufbereitet. Anschließend werden in einem weiteren Schritt geeignete Merkmale gewonnen, die in dem finalen Schritt für eine Klassifikation verwendet werden. Abhängig von der Art der Klassifikation stehen dafür Referenzdokumente mit bekannten Schreibern zur Verfügung.

Dieser Beitrag ist wie folgt gegliedert: nach dieser Einleitung wird in **Abschnitt 2** der Schritt der Vorverarbeitung erläutert. Anschließend werden in **Abschnitt 3** unterschiedliche Methoden der Merkmalsextraktion vorgestellt. In **Abschnitt 4** werden unterschiedliche Aufgaben für die Klassifikation von Schreibern in historischen Dokumenten vorgestellt, zu denen in **Abschnitt 5** exemplarische Ergebnisse von verschiedenen Experimenten gezeigt werden.

## 2. Vorverarbeitung

In der Vorverarbeitungsstufe werden die gescannten Bilder der historischen Dokumente für die Schreibererkennung aufbereitet. Idealerweise sollten nach dieser Stufe die Bilder nur die reine Schrift enthalten, beispielsweise in Form von Binärbildern (Schrift schwarz, Beschreibstoff weiß). Aufgrund der Alterung der Dokumente enthalten die Seitenbilder allerdings verschiedenste Arten von Verunreinigungen, wie z.B. Wasserflecken (vgl. **Abbildung 2**) oder Texte, die von der Rückseite durchscheinen, die die Vorverarbeitung erschweren.

---

<sup>5</sup> Vgl. Robinson 1996, S. 106, der dieses Problem im Zusammenhang mit textueller Überlieferung diskutiert und konstatiert: »In fact, the closer we look at a text, the more variation we see. It is not only that there is more variation: the text itself changes depending on how closely we are looking at it.«

<sup>6</sup> Vgl. M / Idicula 2011, S. 23–33.

Des Weiteren finden sich in solchen Dokumenten oft Zusätze anderer Schreiber an den Seitenrändern. Ornamente unterschiedlichster Ausprägung können zusätzlich die automatische Merkmalsgewinnung erschweren. Um die Zuweisung von Schriftmerkmalen des Haupttextes zu verbessern, wird in einem ersten Schritt die Region des Haupttextes ermittelt. Abbildung 1 (links) zeigt ein solches Beispiel einer Seite eines arabischen, historischen Dokumentes mit vielen Kommentaren.<sup>7</sup>

Im nächsten Schritt wird die Handschrift des Haupttextes vom Hintergrund getrennt (segmentiert), um eine reine, möglichst ungestörte Version der Schrift zu erhalten. Hierfür wird eine Binarisierung angewendet, die ein farbiges Eingabebild in ein binäres Bild transformiert. Abbildung 1 (Mitte und rechts) zeigt ein Beispiel einer solchen Segmentierung mit einem Binarisierungsverfahren aus dem aktuellen Stand der Technik.<sup>8</sup> Trotz der Wasserflecken kann die Schrift erfolgreich segmentiert werden. Einzig die rote Schrift wird nicht ausreichend gut segmentiert. Um auch in solchen Fällen die Schrift erfolgreich zu segmentieren, können spezielle Farbsegmentierungsalgorithmen verwendet werden.<sup>9</sup>



Abb. 1: Seite eines arabischen, historischen Dokumentes mit Haupttext und vielen Kommentaren (links); Beispiel einer Segmentierung der Handschrift mit Hilfe einer Binarisierung: Teil eines Ausgangsbildes (Mitte) und binäres Ergebnisbild (rechts) (Quelle: Autoren).

### 3. Merkmalsextraktion

Nach der Binarisierung erfolgt die Merkmalsextraktion. Das Ziel dieser Stufe ist es, diskriminative Merkmale aus der Handschrift für die Identifikation von Schreibern zu generieren. Allgemein wird zwischen Mikro- und Makromerkmalen unterschieden. Mikromerkmale spiegeln die feine Charakteristik der Handschrift wider. Als Beispiel hierfür gibt es z.B. Grapheme-basierte Ansätze.<sup>10</sup> Im Gegensatz hierzu versuchen Makromerkmale die globale Charakteristik der Schrift eines Schreibers zu ermitteln. In dieser Arbeit wird der Fokus auf den zweiten Typ von Merkmalen gelegt. Dabei wird ein Merkmalsvektor für ganze Textblöcke erzeugt, der die globale Information über den Schreibstil eines Schreibers wie z.B. die Neigung oder die Krümmung der Schrift enthält. Im Folgenden werden verschiedene Arten von Merkmalen, basierend auf den unterschiedlichen Grundprinzipien Kontur, Textur und der Extraktion von Schlüsselpunkten vorgestellt.

<sup>7</sup> In Fecker et al. 2014, S. 1551, wird hierfür ein Verfahren zur automatischen Detektion vorgeschlagen.

<sup>8</sup> Vgl. Bar-Yosef et al. 2007, S. 89–99.

<sup>9</sup> Pantke et al. 2014, S. 151–156.

<sup>10</sup> Bulacu et al. 2007, S. 771.

### 3.1 Konturbasierte Merkmale

Die konturbasierten Merkmale verwenden eine Statistik über die Verteilung der Winkel, die in der Kontur einer Handschrift eines Schreibers enthalten sind.<sup>11</sup> In den binären Bildern wird diese Verteilung mit Hilfe einer Konturverfolgung ermittelt. Die Winkel werden zwischen der Verbindungsgeraden zweier Konturpunkten mit vorgegebenem Abstand und der horizontalen Achse gemessen. Sämtliche ermittelten Winkel einer Schrift werden in einem Winkelhistogramm gesammelt. Anschließend wird dieses zu einer Wahrscheinlichkeitsdichteverteilung normalisiert, welche letztendlich als Merkmalsvektor verwendet wird. Es gibt auch Beispiele für die ermittelten Wahrscheinlichkeitsdichteverteilungen zweier unterschiedlicher Schreiber. Dabei wird eine Log-Polar-Darstellung dieser Verteilungen verwendet. In Diagrammen werden jeweils die Ergebnisse zweier unterschiedlicher Texte eines Schreibers eingetragen. Trotz der unterschiedlichen Texte ändern sich die Verteilungen nur geringfügig. Ein großer Unterschied manifestiert sich dagegen zwischen den Verteilungen unterschiedlicher Schreiber.

In *Text-Independent Writer Identification and Verification on Offline Arabic Handwriting* wurde ein fester Abstand zwischen zwei Konturpunkten für die Winkelmessung verwendet.<sup>12</sup> Als Modifikation wurde in *Writer Identification for Historical Arabic Documents* eine variable Distanz für die Winkelmessung vorgeschlagen, welche mit Hilfe eines Fehlerkriteriums automatisch bestimmt werden kann.<sup>13</sup> Diese Anpassung ermöglichte eine genauere Approximation der Kontur.

### 3.2 Texturbasierte Merkmale

In *Natural Image Character Recognition Using Oriented Basic Image Features* werden die Merkmale »Oriented Basic Image Features« vorgeschlagen, die auf Basis der Symmetrie und Orientierung in der lokalen Nachbarschaft von Pixeln eine Textur beschreiben.<sup>14</sup> Hierbei handelt es sich um einen Multi-Skalen-Ansatz, basierend auf Differenzbildern von Gauß-gefilterten Bildern. Dafür werden Filter unterschiedlicher Ordnungen und Orientierungen verwendet. Basierend auf diesen Differenzbildern werden sieben Symmetrie-Merkmale konstruiert. Einigen dieser Symmetriemerkmale kann eine Orientierung zugewiesen werden. Die übrigen Merkmale sind rotationsinvariant. Sämtliche vorkommende Kombinationen von Symmetrie-Merkmal, Orientierung und Skalierungsstufe der Pixel eines Eingabebildes werden wieder zu einem Histogramm zusammengefügt. Nach einer Normalisierung zu einer Wahrscheinlichkeitsdichtefunktion ergibt sich daraus ebenfalls ein Merkmalsvektor.<sup>15</sup>

---

<sup>11</sup> Bulacu et al. 2007, S. 770.

<sup>12</sup> Bulacu et al. 2007, S. 770.

<sup>13</sup> Fecker et al. 2014, S. 1551.

<sup>14</sup> Newell / Griffin 2011, S. 191–196.

<sup>15</sup> Vgl. Hassane / Al-Madeed 2012, S. 835–840. Dieser Ansatz erreichte in einem Wettbewerb zur arabischen Schreibererkennung in zeitgenössischen Schriften bei der Konferenz International Conference on Frontiers of Handwriting Recognition (ICFHR) im Jahr 2012 den ersten Platz.

### 3.3 Schlüsselpunktbasierte Merkmale

Diese Merkmale basieren auf den Scale Invariant Feature Transform (SIFT)-Deskriptoren.<sup>16</sup> Diese basieren ebenfalls auf Differenzen von Gauß-gefilterten Bildern unterschiedlicher Skalierungen. Anders als bei den Oriented Basic Image Features werden hier über lokale Extremwerte der ermittelten Differenzbilder Schlüsselpunkte detektiert. Über die Gradienten in der Umgebung wird diesen Schlüsselpunkten eine Orientierung zugewiesen (Abbildung 2, links). Anschließend werden ebenfalls auf Basis der lokalen Orientierung und Beträgen der Gradienten in einer begrenzten Region um einen Schlüsselpunkt die Deskriptoren ermittelt (Abbildung 2, rechts).

Für Handschriften werden die Schlüsselpunkte auf Kreuzungen, Krümmungen und Spitzen detektiert. Basierend auf den Distanzen aller Deskriptoren in einem Textblock wird ein Merkmalsvektor abgeleitet.<sup>17</sup> Die Distanzen liefern Informationen über das Verhältnis der Orientierungen und der Beträge der Gradienten um die Schlüsselpunkte, und man erhält somit Informationen über den Schreibstil eines Schreibers.

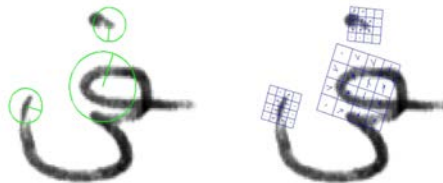


Abb. 2: Exemplare für Schlüsselpunkte in einer Handschrift. Die Mittelpunkte der Kreise deuten auf den Ort des Schlüsselpunktes, der Durchmesser auf die Skalierung, und die Orientierung ist durch die Linie gegeben (links). Deskriptoren der ermittelten Schlüsselpunkte (rechts) (Quelle: Autoren).

## 4. Klassifikation

Es lassen sich verschiedene Klassifikationsaufgaben im Bereich der Schreibererkennung von historischen Dokumenten identifizieren. In dieser Arbeit wird der Fokus auf die drei Klassifikationsaufgaben Intra-Manuskript-Schreiberanalyse, der direkte Vergleich zweier Manuskripte und die One-to-Many-Klassifikation eines Manuskriptes mit unbekanntem Schreiber gelegt. Diese Verfahren werden im Folgenden erläutert, wobei hier speziell die Klassifikation von mehrseitigen Dokumenten behandelt wird. Allgemein wird dabei für jede Seite eines Dokumentes ein Merkmalsvektor extrahiert. Für den Vergleich zweier Merkmalsvektoren wird in der Regel eine Distanzmetrik eingesetzt.

### 4.1 Intra-Manuskript-Schreiberanalyse

---

<sup>16</sup> Lowe 2004, S. 91–110.

<sup>17</sup> Fecker et al. 2014, S. 1552.



Bei der Intra-Manuskript-Schreiberanalyse wird innerhalb eines Dokumentes untersucht, ob an einem Dokument ein oder mehrere unterschiedliche Schreiber beteiligt waren. Zusätzlich können explizit die Seiten ermittelt werden, die von den einzelnen Schreibern geschrieben wurden. Ermöglicht wird dies durch die Berechnung der Distanzen sämtlicher Merkmalsvektoren eines Dokumentes zueinander. Hierdurch ergibt sich eine Distanzmatrix. In *Abbildung 5* ist schematisch eine solche Distanzmatrix eines fünfseitigen Dokumentes dargestellt. Sämtliche Elemente der Hauptdiagonale enthalten den Wert Null, da dies der Distanz von einer Seite zu sich selbst entspricht. Die Matrix ist entlang dieser Hauptdiagonalen gespiegelt und enthält somit redundante Informationen. Die Darstellung in Matrixform birgt beispielsweise bei der Visualisierung Vorteile (vgl. auch [Abschnitt 5.1](#)).

Die Distanzmatrizen lassen sich dazu verwenden, unterschiedliche Schreiber anhand der Unterschiede in den Beträgen der Differenzen zu erkennen. Um die Beträge der Differenzen einordnen zu können, können optional bereits analysierte Referenzdokumente mit bekannter Anzahl von Schreibern verwendet werden, um Schwellwerte abzuleiten.

## 4.2 Direkter Vergleich zweier Manuskripte

Anders als bei der Intra-Manuskript-Schreiberanalyse werden hier die Schreiber von zwei Dokumenten miteinander verglichen. Hierfür werden ebenfalls Distanzmatrizen eingesetzt. Allerdings werden hier die Distanzen der Merkmalsvektoren eines Dokumentes zu denen des anderen Dokumentes ermittelt. Um die Beträge der Distanzen einschätzen zu können, können zusätzlich Referenzdokumente für die Ermittlung von Schwellwerten verwendet werden (*Tabelle 1, Abbildung 3*).

0	D(S1,S2)	D(S1,S3)	D(S1,S4)	D(S1,S5)
D(S2,S1)	0	D(S2,S3)	D(S2,S4)	D(S2,S5)
D(S3,S1)	D(S3,S2)	0	D(S3,S4)	D(S3,S5)
D(S4,S1)	D(S4,S2)	D(S4,S3)	0	D(S4,S5)
D(S5,S1)	D(S5,S2)	D(S5,S3)	D(S5,S4)	0

Abb. 3: Tabelle 1: Schematische Darstellung einer Distanzmatrix. Zur Erläuterung: D(S1,S2) entspricht der Distanz der Merkmalsvektoren der ersten und zweiten Seite eines Manuskripts (Quelle: Autoren).

## 4.3 One-to-Many- Klassifikation

Bei der One-to-Many-Klassifikation wird versucht, einen unbekanntem Schreiber eines Manuskriptes mit Hilfe eines Datenbestandes von bekannten Schreibern zu identifizieren. Dazu wird in diesem Datenbestand dasjenige Manuskript ermittelt, bei dem die Handschrift die größte Ähnlichkeit mit der des unbekanntem Schreibers aufweist. Hier wird dafür die Nächster-Nachbar-Klassifikation eingesetzt. Da es sich dabei um mehrseitige Manuskripte handelt, für die ebenfalls mehrere Merkmalsvektoren extrahiert werden, existieren unterschiedliche

Strategien, diese für die Klassifikation zu kombinieren.<sup>18</sup> Bei der Mittelung werden sämtliche Merkmalsvektoren für die untersuchten Seiten eines Manuskriptes in einen einzelnen Merkmalsvektor zusammengefasst. Dabei wird ein Schreiber dem gesamten Manuskript zugeordnet (Abbildung 4, links). Beim Voting dagegen wird jeder Seite des zu untersuchenden Manuskriptes ein Schreiber zugeordnet. Der Kandidat mit den meisten zugeordneten Seiten wird als Schreiber des Manuskriptes erkannt (Abbildung 4, rechts). Beim Voting ist zusätzlich eine Gewichtung durch die ermittelte Distanz der Nächster-Nachbar-Klassifikation möglich. Je kleiner die Distanz ist, desto höher ist die Konfidenz für die Erkennung eines Schreibers.



Abb. 4: One-to-Many-Klassifikation eines mehrseitigen Manuskriptes durch Mittelung der Merkmalsvektoren (links). One-to-Many-Klassifikation eines mehrseitigen Manuskriptes durch Voting (rechts) (Quelle: Autoren).

## 5. Experimente

In diesem Kapitel werden exemplarische Ergebnisse für Experimente bei der Schreibererkennung präsentiert. Die durchgeführten Experimente sind die Intra-Manuskript-Schreiberanalyse, der direkte Vergleich zweier Manuskripte und die One-to-Many-Klassifikation. Für die Experimente wurden lateinische und arabische Manuskripte verwendet.

### 5.1 Experimente zur Intra-Manuskript-Schreiberanalyse und direktem Vergleich zweier Manuskripte

Um beide Fragestellungen adressieren zu können, wurde folgende Arbeitsweise festgelegt und darauf basierend Testmaterial ausgewählt:

- Es soll eine Handschrift (1) für das Training ausgewählt werden, deren Schreiber bekannt ist. Die ganze Handschrift sollte erwiesenermaßen vom gleichen Schreiber geschrieben worden sein, um dem Algorithmus genügend Trainingsmaterial anbieten zu können. Diese Handschrift dient als Referenz für Hand 1.
- Der Schreiber von Handschrift 1 sollte zugleich in einer anderen Handschrift (2) nachweisbar sein, in der er allerdings nicht alleine schreibt, sondern wiederum erwiesenermaßen weitere Hände identifiziert wurden. Die anderen Schreiberhände müssten dann durch eigene Merkmalsvektoren von Hand 1 zu unterscheiden sein.

Bei der Materialauswahl wurden zwei Handschriften ausgemacht, welche die vorgenannten Bedingungen erfüllen: Cod. Guelf. 62 Weissenburg und Cod. Guelf. 63 Weissenburg. Beide Handschriften sind in der wichtigsten frühmittelalterliche Buchschrift, der karolingischen Minuskel, geschrieben, die sich durch einheitliche, relativ stark standardisierte Formen

---

<sup>18</sup> Fecker et al. 2014, S. 1553.

und eine meist geringe individuelle Varianz auszeichnet. Diese allgemeinen Spezifika unterstützen die Brauchbarkeit von Digitalisaten karolingischer Handschriften ebenso wie deren meist hohes kodikologisches Niveau, das bei entsprechender Fotoaustattung nur wenig Nachbearbeitung erforderlich macht. Beide Handschriften entstammen zudem der Sammlung der im Rahmen des **Europeana Regia-Projekts** digitalisierten Codices Weissenburgenses, die weitere Vorteile aufweist: Die weitaus meisten Codices stammen aus dem Skriptorium des Klosters Weißenburg im Elsass, sind also regional und zeitlich gut einzuordnen. Mit dem den DFG-Richtlinien entsprechenden Katalog von Hans Butzmann sind diese Handschriften außerdem kodikologisch gut erschlossen. Die Beschreibungen liefern die Vorlagen der Schreiberidentifikation, die es zu verifizieren (oder falsifizieren) gilt. **Cod. Guelf. 62 Weiss.** ist dem Kolophon zufolge zwischen 819 und 826 im elsässischen Kloster Weissenburg während des Abbatats von Gerhoh von dem Mönch Waldmann geschrieben.<sup>19</sup> In **Cod. Guelf. 63 Weiss.**, geschrieben ebenfalls in der ersten Hälfte des 9. Jahrhunderts, sind durch die klassische Paläographie drei Hände identifiziert, von denen eine die Hand Waldmanns ist. Waldmann werden die Seiten 82r bis 107r zugeschrieben.<sup>20</sup>

### 5.1.1 Intra-Manuskript-Schreiberanalyse mit 62 Weiss.

Abbildung 5 zeigt die ermittelte Distanzmatrix in einer dreidimensionalen Darstellung. Hier zeigt sich, dass das Buch größtenteils von einem Schreiber (im Folgenden: Schreiber A) geschrieben wurde. Nur zum Ende des Manuskriptes treten variierende Schreiber (im Folgenden zusammengefasst als Schreiber B) auf, was sich in den höheren Distanzen ausdrückt. **Abbildung 6 (links)** zeigt eine Seite aus der Mitte des Manuskriptes, die von dem Hauptschreiber verfasst wurde. In **Abbildung 6 (rechts)** ist dagegen eine Seite vom Ende des Manuskriptes dargestellt, die sich im Schreibstil eindeutig von dem der Seiten in der Mitte des Buches unterscheidet.

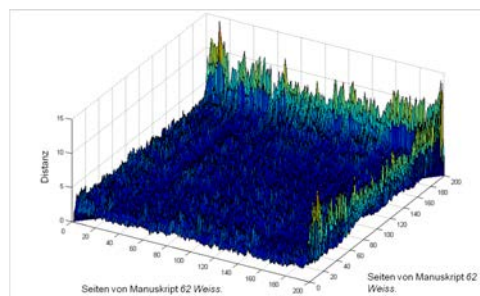


Abb. 5: Distanzmatrix für die Intra-Manuskript Schreiberanalyse von Manuskript 62 Weiss. (Quelle: Autoren).

---

<sup>19</sup> Butzmann 1964, S. 202.

<sup>20</sup> Butzmann 1964, S. 203.

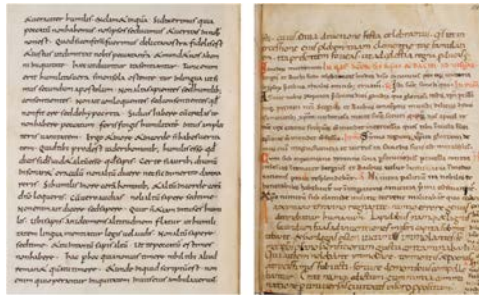


Abb. 6: Seite 20 aus dem Manuskript 62 Weiss, des Hauptschreibers (links). Seite 209 aus dem Manuskript 62 Weiss, mit variierendem Schreiber aus den letzten Seiten des Manuskripts (rechts) (Quelle: Autoren).

### 5.1.2 Intra-Manuskript-Schreiberanalyse mit 63 Weiss.

Ein etwas anderes Ergebnis zeigt die Distanzmatrix des Cod. Guelf. 63 Weiss. in Abbildung 7. Hier lassen sich aufgrund der Distanzunterschiede drei unterschiedliche Schreiber A, B und C identifizieren. Auffällig ist hier, dass die Unterschiede zwischen den Schreibern im Gegensatz zu den Unterschieden in Cod. Guelf. 62 Weiss. nur sehr gering sind und zu deutlich geringeren Distanzen führt. Grund hierfür dürfte die Tätigkeit im gleichen Skriptorium bzw. die Zugehörigkeit zu derselben Schule sein. Abbildung 8 (links, Mitte, rechts) zeigt von jedem Schreiber jeweils eine Seite aus dem Manuskript. Auf diesen ist die Ähnlichkeit sichtbar, im Gegenteil wird es schwierig, Differenzen zu bemerken. Die Unterschiede in der Schrift liegen jetzt nur in einzelnen Buchstabenformen wie z.B. den g-Formen, der Verwendung von Groß- und Kleinbuchstaben sowie im Schriftduktus.

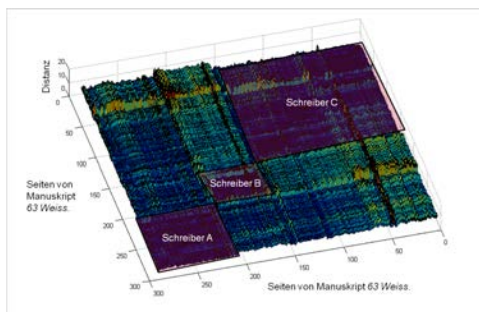


Abb. 7: Distanzmatrix für die Intra-Manuskript-Schreiberanalyse von Manuskript 63 Weiss. (Quelle: Autoren).

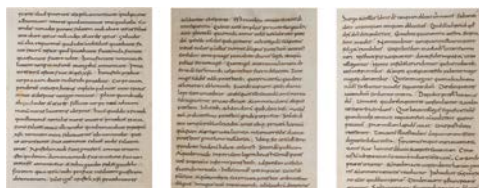


Abb. 8: Seite 94 aus Manuskript 63 Weiss, des Schreibers A (links). Seite 187 aus Manuskript 63 Weiss, des Schreibers B (Mitte). Seite 248 aus Manuskript 63 Weiss, des Schreibers C (rechts) (Quelle: Autoren).

### 5.1.3 Direkter Vergleich von Cod. Guelf. 62 Weiss. und 63 Weiss.

In diesem Experiment werden die Schreiber der Cod. Guelf. 62 Weiss. und 63 Weiss. miteinander verglichen. In *Die Weissenburger Handschriften* ist beschrieben, dass Teile von Cod. Guelf. 63 Weiss. von demselben Schreiber verfasst wurden, der auch 62 Weiss. verfasst hat.<sup>21</sup> In der zwischen den beiden Manuskripten ermittelten Distanzmatrix, zu sehen in *Abbildung 9*, ist tatsächlich zu erkennen, dass die von dem Schreiber B geschriebenen Seiten von 63 Weiss. zu denen von 62 Weiss. eine hohe Ähnlichkeit aufweisen. Tatsächlich stimmen die mit diesem Verfahren ermittelten Seitennummern auch mit den Angaben von Butzmann überein.<sup>22</sup>

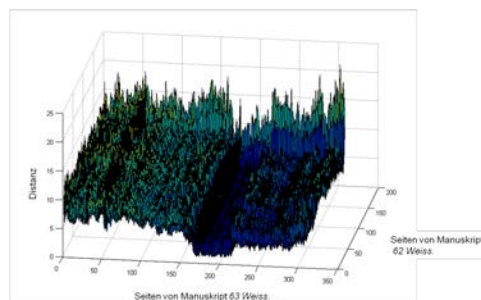


Abb. 9: Distanzmatrix für den direkten Vergleich der Manuskripte 62 Weiss. und 63 Weiss. (Quelle: Autoren).

## 5.2 Experimente zur One-to-Many-Klassifikation

In *Writer Identification for Historical Arabic Documents* wurden Untersuchungen zur One-to-Many-Klassifikation mit arabischen historischen Dokumenten im Rahmen des DFG-geförderten HADARA-Projektes<sup>23</sup> durchgeführt. Hierfür wurde eine frei-verfügbare Sammlung von arabischen, historischen Manuskripten verwendet, die Harvard Islamic Heritage Database.<sup>24</sup> Aus dieser Datenbank wurde eine Teilmenge von 60 Manuskripten ausgewählt. 43 dieser Manuskripte mit insgesamt 2313 Seiten stammen dabei von elf Schreibern, von denen jeder mindestens zwei Manuskripte geschrieben hat. Die restlichen 17 dieser Manuskripte mit insgesamt 2282 Seiten stammen von jeweils einem Schreiber. *Abbildung 10* zeigt Seiten aus drei Manuskripten dieser Datenmenge. Diese Manuskripte beinhalten die typischen Degradierungserscheinungen von historischen Dokumenten.

Für die Evaluation wird die Leave-one-out-Kreuzvalidierung eingesetzt. Bei dieser wird ein Manuskript als Testobjekt und die restlichen werden als Trainingsmenge verwendet. Hier werden nur diejenigen 43 Manuskripte, von denen ein Schreiber jeweils mindestens zwei Manuskripte verfasst hat, als Testobjekte verwendet. Die restlichen 17 Manuskripte werden

---

<sup>21</sup> Butzmann 1964, S. 203.

<sup>22</sup> Butzmann 1964, S. 203.

<sup>23</sup> Pantke et al. 2013, S. 151–156.

<sup>24</sup> Harvard University. [Islamic Heritage Project \(IHP\)](#).

in die Trainingsmenge mit aufgenommen. Für die Experimente bedeutet das, dass jedes der 43 Manuskripte einmal als Testobjekt verwendet wird. Die restlichen 59 Manuskripte dienen als Trainingsmenge mit bekannten Schreibern, von denen dem Testobjekt ein Schreiber zugeordnet werden soll.

In den Experimenten wurden jeweils die in **Abschnitt 3** und **4** vorgestellten Merkmalsextraktions- und Klassifikationsverfahren eingesetzt. Sämtliche Ergebnisse sind in Tabelle 1 (**Abbildung 3**) dargestellt. Neben der Genauigkeit der Klassifikation auf Manuskriptebene, bei der einem Manuskript genau ein Schreiber mit Hilfe der Verfahren Mittelung, Voting oder gewichtetem Voting zugeordnet wird, ist zusätzlich die Genauigkeit auf Seitenebene ermittelt worden. Dazu wird jeder Seite eines Manuskripts genau einem Schreiber zugeordnet. Am schlechtesten schneiden hier die konturbasierten Merkmale ab. Wie in *Writer Identification for Historical Arabic Documents* gezeigt, ergibt sich durch die Modifikation der Winkelmessung eine Verbesserung gegenüber dem originalen Ansatz,<sup>25</sup> jedoch reicht die Performanz nicht an die der anderen Verfahren heran.<sup>26</sup> Am besten schneiden die ebenfalls hier vorgestellten schlüsselpunktbasierten Merkmale ab. Diese sind mit der Voting-Klassifikationsstrategie besser als die texturbasierten Merkmale, sowohl auf Seitenebene als auch auf Manuskriptebene. Mit dem gewichteten Voting wird darüber hinaus auch eine perfekte Klassifikation erreicht, d.h. für jedes Manuskript wird der korrekte Schreiber erkannt. Einzig bei der Mittelung der Merkmalsvektoren schneiden die schlüsselpunktbasierten Merkmale schlecht ab. Es ist zu vermuten, dass durch die Mittelung bei diesen Merkmalen zu viel Information über einen Schreiber verloren geht.

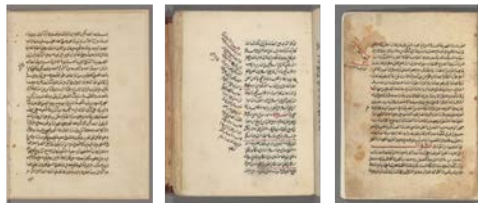


Abb. 10: Seiten von drei Manuskripten aus der [Harvard Islamic Heritage Database](#)(Quelle: Autoren).

Merkmale	Genauigkeit Seitenebene (in %)	Genauigkeit Manuskriptebene (in %)		
		Mittelung	Voting	Gewichtetes Voting
Konturbasiert	68,6	81,0	88,1	92,9
Texturbasiert	87,6	92,9	92,9	92,9
Schlüsselpunktbasier	92,5	59,5	97,6	100,0

Abb. 11: Tabelle 2: Ergebnisse der One-to-Many Klassifikation bei Verwendung der verschiedenen Merkmale und Klassifikationsstrategien (Quelle: Autoren).

## 6. Zusammenfassung

<sup>25</sup> Bulacu et al. 2007, S. 770.

<sup>26</sup> Fecker et al. 2014, S. 3054.

In diesem Beitrag wurde gezeigt, dass eine computergestützte Identifikation der schreibenden Hand bzw. Hände in historischen Dokumenten möglich ist. Hierfür wird eine Prozesskette aus Vorverarbeitung, Merkmalsextraktion und Klassifikation für die automatische Analyse gescannter Dokumente eingesetzt. Neben einer allgemeinen Vorgehensweise für die Vorverarbeitung dieser Dokumente wurden Merkmale vorgestellt, die effektiv die Charakteristika eines Schreibers beschreiben. Diese Merkmale basieren auf unterschiedlichen Charakteristika wie Kontur, Textur und Schlüsselpunkte der Handschrift. Darüber hinaus wurden diverse Einsatzgebiete für die Klassifikation von Schreibern vorgestellt, wie die Intra-Manuskript-Schreiberanalyse, der direkte Vergleich zweier Manuskripte und die One-to-Many-Klassifikation. Experimente mit diesen Verfahren zeigen vielversprechende Ergebnisse, die eine sinnvolle Unterstützung von Arbeiten im Bereich der Kodikologie und Paläographie erwarten lassen.

In der konkreten Anwendung sind vor allem zwei Szenarien denkbar:

- In einer Handschrift werden die Unterschiede der Schriften extrahiert, um die Anzahl schreibender Hände zu ermitteln;
- Im Vergleich zweier oder mehrerer Handschriften wird ermittelt, ob identifizierte Schreiber aus einer Handschrift auch in anderen Codizes tätig waren. Die Merkmale identifizierter Schriften könnten in einer Datenbank gesammelt werden und als Ausgangsmaterial für die Anwendung weiterer Algorithmen genutzt werden.

Die in den Szenarien 1 und 2 ermittelten Merkmalsvektoren könnten zur Identifikation von Schrifttypen generalisiert werden. Damit wäre es denkbar, regionale (insulare vs. kontinentale, länderspezifische) oder temporale (karolingische Minuskeln des 9. bzw. 11. Jhs.) Unterschiede zu identifizieren.<sup>27</sup>

---

<sup>27</sup> Hiermit wird der der Deutschen Forschungsgemeinschaft (Projekt: FI 1494-3-2) für ihre finanzielle Unterstützung gedankt.

## Bibliographische Angaben

Itay Bar-Yosef / Issac Beckman / Klara Kedem / Itshak Dinstein: Binarization, character extraction, and writer identification of historical hebrew calligraphy documents. In: International Journal on Document Analysis and Recognition 9 (2007), 2, S. 89–99. [\[Nachweis im GBV\]](#)

Marius Bulacu / Lambert Schomaker / Axel Brink: Text-Independent Writer Identification and Verification on Offline Arabic Handwriting. In: Proceedings of the International Conference of Document Analysis and Recognition (ICDAR). Hg. von IEEE. Parana 2007, S. 769–773. DOI 10.1109/ICDAR.2007.4377019. [\[Nachweis im GBV\]](#)

Hans Butzmann: Die Weissenburger Handschriften. Frankfurt/Main 1964 (= Kataloge der Herzog August Bibliothek Wolfenbüttel: Neue Reihe, Bd. 10). [\[Nachweis im OPAC\]](#)

Daniel Fecker / Abed Asi / Volker Märgner / Jihad El-Sana / Tim Fingscheidt: Writer Identification for Historical Arabic Documents. In: Proceedings of the International Conference on Pattern Recognition (ICPR). Stockholm 2014, S. 3050–3055. [\[Nachweis im GBV\]](#)

Abdelaali Hassane / Somaya Al-Madeed: ICFHR 2012 Competition on Writer Identification Challenge 2: Arabic Scripts. In: Proceedings of the International Conference on Frontiers in Handwriting Recognition (ICFHR). Hg. von IEEE. Bari 2012, S. 835–840. DOI 10.1109/ICFHR.2012.218. [\[Nachweis im GBV\]](#)

Thomas Haye / Stephan Müller: Mittelalter-Philologie im Internet. Digitalisierung mittelalterlicher Handschriften aus Sicht der Forschung. In: Zeitschrift für deutsches Altertum und deutsche Literatur 140 (2011), S. 416–420. [\[Nachweis im OPAC\]](#)

David G. Lowe: Distinctive Image Features from Scale-Invariant Keypoints. In: International Journal of Computer Vision 2004 (60), S. 91–110. [\[Nachweis im GBV\]](#)

Sreeraj M / Sumam Mary Idicula: A Survey on Writer Identification Schemes. In: International Journal of Computer Applications 26 (2011), Nr. 2, S. 23–33. DOI 10.5120/3075-4205. [\[online\]](#)

Andrew J. Newell / Lewis D. Griffin: Natural Image Character Recognition Using Oriented Basic Image Features. In: Proceedings of the International Conference Digital Image Computing Techniques and Applications (DICTA). Hg. von IEEE. Noosa 2011, S. 191–196. DOI 10.1109/DICTA.2011.39. [\[Nachweis im GBV\]](#)

Werner Pantke / Arne Haak / Volker Märgner: Color Segmentation for Historical Documents Using Markov Random Fields. In: Proceedings of the 6th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2014). Tunis 2014, S. 151–156. DOI 10.1109/SOCPaR.2014.7007997.

Werner Pantke / Volker Märgner / Daniel Fecker / Tim Fingscheidt / Abed Asi / Ofer Biller / Jihad El-Sana / Raid Saabni / Mohammed Yehia: HADARA – A Software System for Semi-Automatic Processing of Historical Handwritten Arabic Documents. In: Proceedings of IS&T Archiving 2013, Washington D.C. 2013, S. 161–166 [\[online\]](#). [\[Nachweis im GBV\]](#)

Peter M.W. Robinson: Is There a Text in These Variants? In: The Literary Text in the Digital Age. Hg. von Richard J. Finneran. Ann Arbor (Mi) 1996, S. 99–115. [\[Nachweis im OPAC\]](#)

Thomas Stäcker: Erfahrungsbericht Helmstedter Drucke Online an der Herzog August Bibliothek Wolfenbüttel. In: Volltext via OCR. Möglichkeiten und Grenzen. Testsznarien zu den Funeralschriften der Staatsbibliothek zu Berlin – Preußischer Kulturbesitz. Hg. von Maria Federbusch / Christian Polzin. Berlin 2013 (= Beiträge aus der Staatsbibliothek zu Berlin – Preußischer Kulturbesitz Band 43), S.123–135. [\[online\]](#)

## Abbildungslegenden und -nachweise

Abb. 1: Seite eines arabischen, historischen Dokumentes mit Haupttext und vielen Kommentaren (links); Beispiel einer Segmentierung der Handschrift mit Hilfe einer Binarisierung: Teil eines Ausgangsbildes (Mitte) und binäres Ergebnisbild (rechts) (Quelle: Autoren).

Abb. 2: Exemplare für Schlüsselpunkte in einer Handschrift. Die Mittelpunkte der Kreise deuten auf den Ort des Schlüsselpunktes, der Durchmesser auf die Skalierung, und die Orientierung ist durch die Linie gegeben (links). Deskriptoren der ermittelten Schlüsselpunkte (rechts) (Quelle: Autoren).

Abb. 3: Tabelle 1: Schematische Darstellung einer Distanzmatrix. Zur Erläuterung: D (S1,S2) entspricht der Distanz der Merkmalsvektoren der ersten und zweiten Seite eines Manuskripts (Quelle: Autoren).

Abb. 4: One-to-Many-Klassifikation eines mehrseitigen Manuskriptes durch Mittelung der Merkmalsvektoren (links). One-to-Many-Klassifikation eines mehrseitigen Manuskriptes durch Voting (rechts) (Quelle: Autoren).

Abb. 5: Distanzmatrix für die Intra-Manuskript Schreiberanalyse von Manuskript 62 Weiss. (Quelle: Autoren).

Abb. 6: Seite 20 aus dem Manuskript 62 Weiss. des Hauptschreibers (links). Seite 209 aus dem Manuskript 62 Weiss. mit variierendem Schreiber aus den letzten Seiten des Manuskripts (rechts) (Quelle: Autoren).

Abb. 7: Distanzmatrix für die Intra-Manuskript-Schreiberanalyse von Manuskript 63 Weiss. (Quelle: Autoren).



Abb. 8: Seite 94 aus Manuskript 63 Weiss. des Schreibers A (links). Seite 187 aus Manuskript 63 Weiss. des Schreibers B (Mitte). Seite 248 aus Manuskript 63 Weiss. des Schreibers C (rechts) (Quelle: Autoren).

Abb. 9: Distanzmatrix für den direkten Vergleich der Manuskripte 62 Weiss. und 63 Weiss. (Quelle: Autoren).

Abb. 10: Seiten von drei Manuskripten aus der [Harvard Islamic Heritage Database](#) (Quelle: Autoren).

Abb. 11: Tabelle 2: Ergebnisse der One-to-Many Klassifikation bei Verwendung der verschiedenen Merkmale und Klassifikationsstrategien (Quelle: Autoren).