



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Collaborative Research

*Centre for the Study of Manuscript Cultures*

Manuskriptkulturen in Asien, Afrika und Europa

Sonderforschungsbereich 950

# Digitale Manuskriptanalyse

Beitrag zur Konferenz

"AUFBRUCH.hamburg

Digitale Chancen verstehen und nutzen "

10. Oktober 2015, Grand Elysée Hamburg

Bernd Neumann

Fachbereich Informatik, Universität Hamburg



## Historische Manuskripte warten auf Auswertung

- **Ca. 60.000 unbearbeitete Manuskripte in Europa**
- **Bisher vorwiegend inhaltliche Analyse durch Geisteswissenschaftler, materielle Eigenschaften wenig beachtet**
- **Gehütete Manuskriptbestände, kein leichter Zugang**

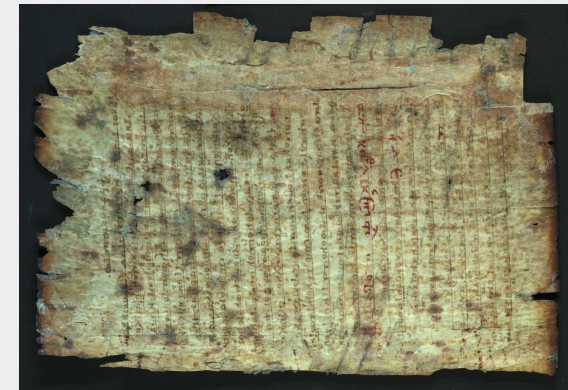
### **Digitalisierung von Manuskripten**

- **fördert die Verbreitung von Kopien**
- **erlaubt rechnerbasierte Analyse großer Datenbestände**
- **ermöglicht neue Analysemethoden**
- **rückt materielle Manuskripteigenschaften in den Fokus**

## Informatik trifft Geisteswissenschaften

Rechnerverfahren ermöglichen fundierte und manchmal unerwartete Erkenntnisse bei der Analyse historischer Manuskripte.

- **Rechnerbasierter Schreibstilvergleich erlaubt**
  - Zuordnung von Manuskriptfragmenten
  - historische Einordnung von Manuskripten
  - Schreiberverifikation
- **Mustererkennungsverfahren unterstützen Textdeutung**
- **Multispektrale Bildanalyse macht ausgelöschte Schrift sichtbar**

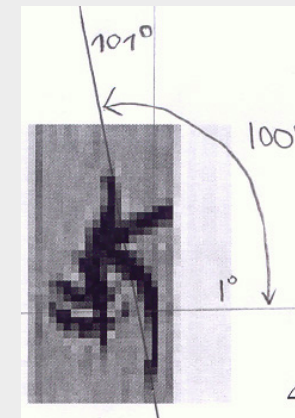
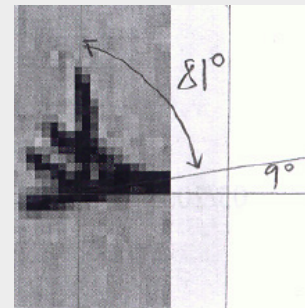
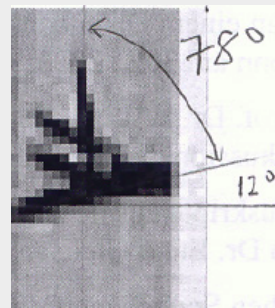


# Anfang der digitalen Manuskriptanalyse an der Universität Hamburg

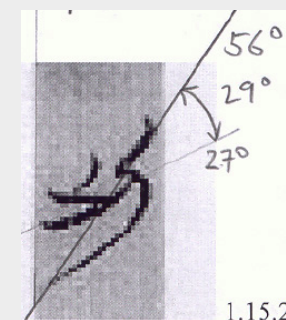
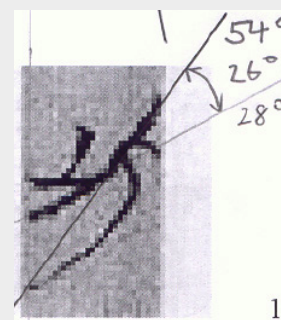
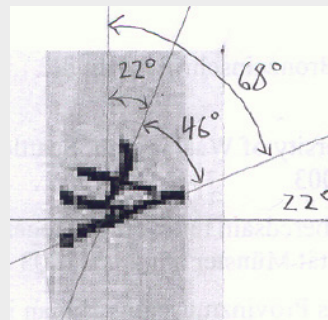
**Kooperation der Sinologie mit der Bildverarbeitung am FB Informatik:**

**Kann man Merkmale chinesischer Schriftzeichen mit Rechnerverfahren bestimmen (Matthias Richter, 2006)?**

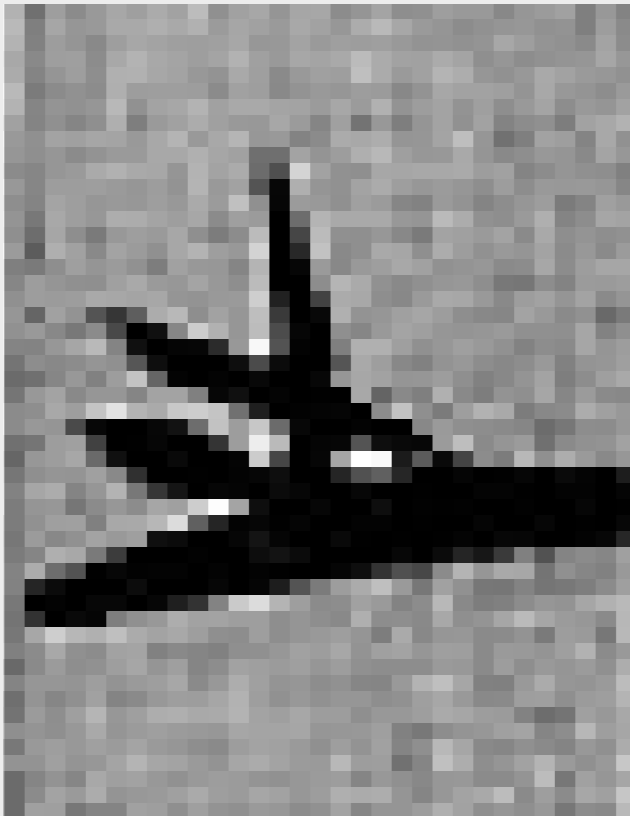
**Schreiber A:**



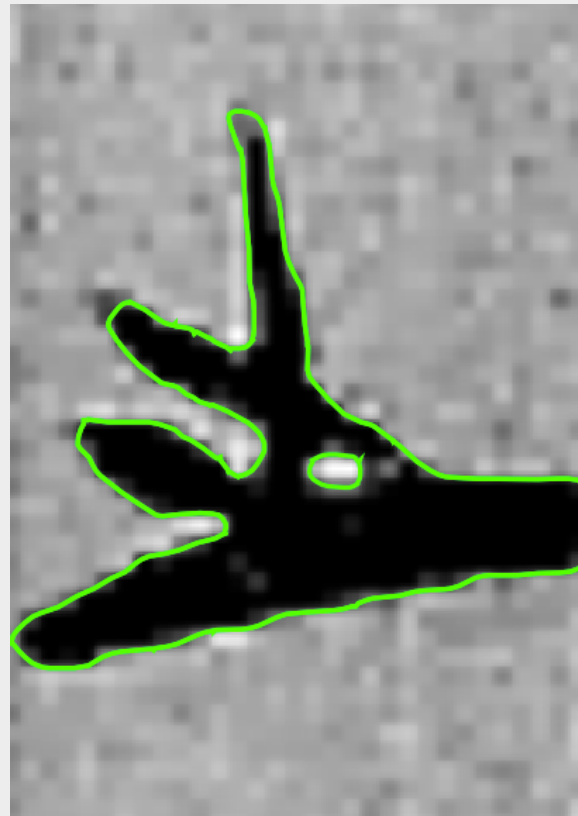
**Schreiber B:**



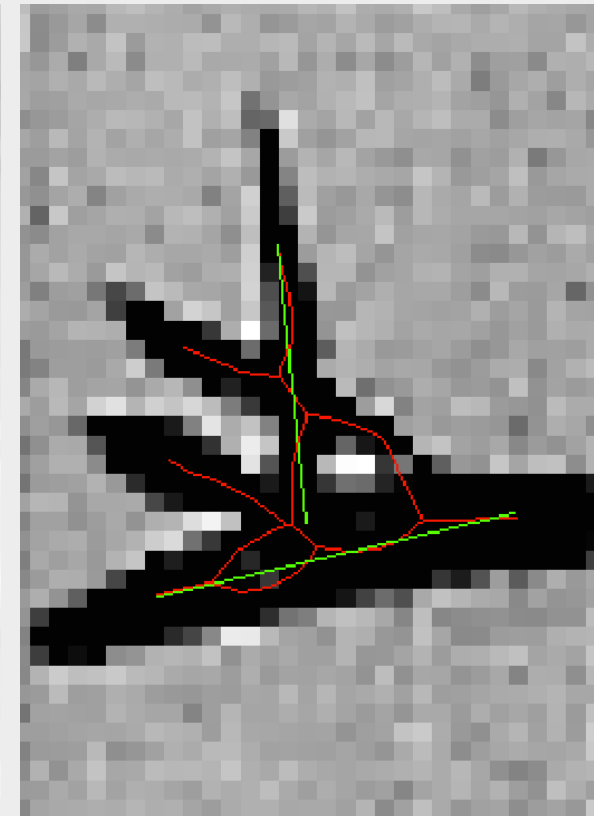
## Erste Ergebnisse 2006



**Grob aufgelöstes Bild  
eines chinesischen  
Zeichens**



**Berechnete Konturen**



**Berechnetes Skelett  
und Medialachsen**

## **Aufbau intensiver Zusammenarbeit bei Digitaler Manuskriptanalyse**

### **DFG-Forschergruppe (2008 – 2011) "Manuskriptkulturen in Asien und Afrika"**

Leitung Prof. Michael Friedrich

**9 Projekte aus der Indologie, Sinologie, Iranistik, Japanologie, Arabistik,  
Sanskritistik, Tamilistik, Tibetologie, Äthiopistik**

**+**

**1 Projekt aus der Informatik**

### **DFG-Sonderforschungsbereich (seit 2011) "Manuskriptkulturen in Asien, Afrika und Europa"**

Leitung Prof. Michael Friedrich

### **Manuskripte als materielle Entitäten in sozialem und kulturellen Kontext**

**z. Zt. 20 Projekte aus den Geisteswissenschaften**

**+**

**1 Projekt aus der Informatik**

**2 Projekte aus Physik und Materialwissenschaften**



**"Wissenschaftliche  
Service-Projekte"**



## **Wissenschaftliches Service-Projekt Bildverarbeitung**

### **Projektgruppe im FB Informatik (2011 – 2015):**

Rainer Herzog, Arved Solth, Bernd Neumann (PI)

### **Arbeitsplan:**

A Anwendung von BV-Methoden in Projekten der Geisteswissenschaftler

B Entwicklung innovativer Methoden für Manuskriptanalyse

C Prototyp eines Arbeitsplatzes zur Manuskriptanalyse

### **Projektgruppe im FB Informatik (2015 – 2019):**

Siegfried Stiehl, Volker Märgner (PIs), N.N., N.N.



## Agenda

### **Anschauliche Beispiele von digitaler Manuskriptanalyse**

- **Rechner als Hilfsmittel: Paläographischer Schreibervergleich**
- **Rechner als Black Box: Forensische Schreiberverifikation**
- **Automatische Layout-Analyse**
- **Mustererkennung in mittelalterlicher Musiknotation**
- **Multispektralanalyse zur Wiedergewinnung ausgelöschter Schrift**





Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Collaborative Research

*Centre for the Study of Manuscript Cultures*

Manuskriptkulturen in Asien, Afrika und Europa

Sonderforschungsbereich 950

# **Rechner als Hilfsmittel: Paläographischer Schreibervergleich**

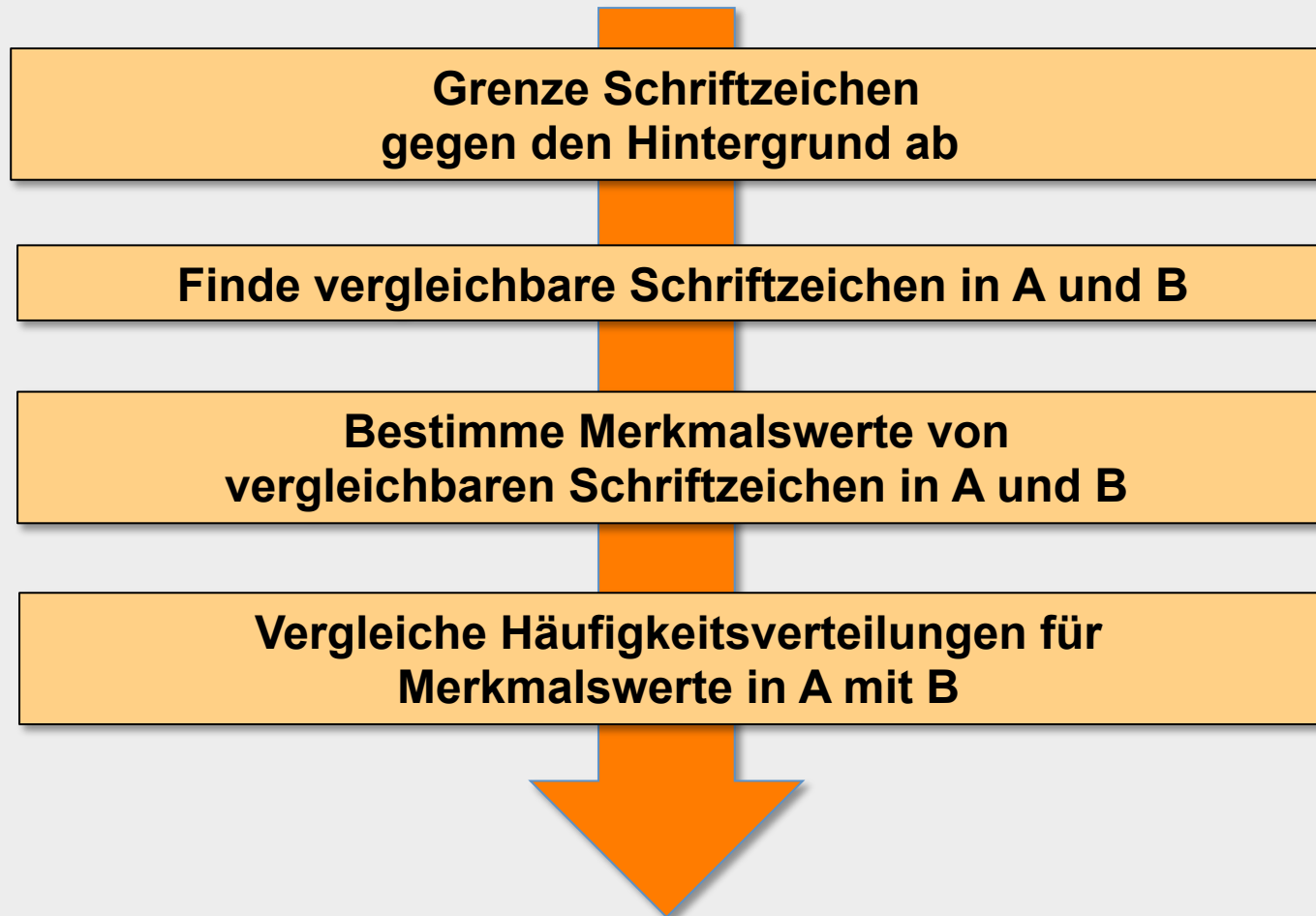
## Schreibervergleich in chinesischen Manuskripten

Zuordnung von ca. 800 Bambusstäben aus einem Grabfund in China  
(Guodian Manuskripte, ca. 300 BC)



## Verarbeitungsschritte

Gehören zwei Manuskripte (Bambusstäbe) A und B zusammen?



## Suche nach vergleichbaren Zeichen

vorgegebenes  
Zeichen

gefundene Zeichen, nach Ähnlichkeit sortiert



Verwendung des Shape-Context-Verfahrens aus der Werkzeugkiste der Bildverarbeitung

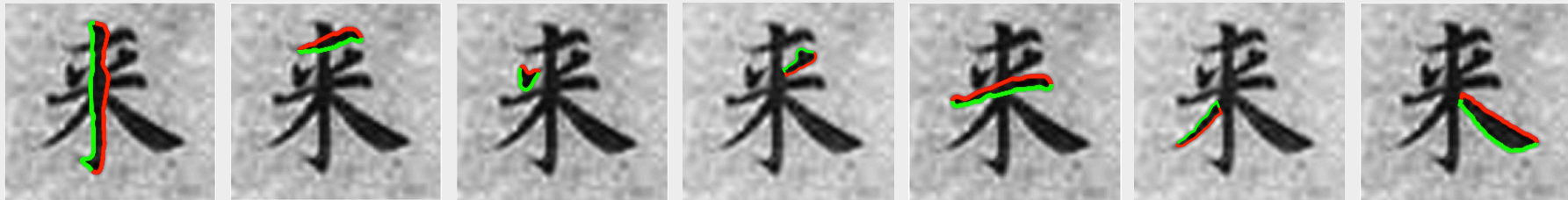
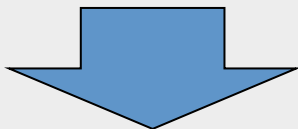
## Bestimmung einzelner Striche



**Innovatives Triangulierungsverfahren liefert**

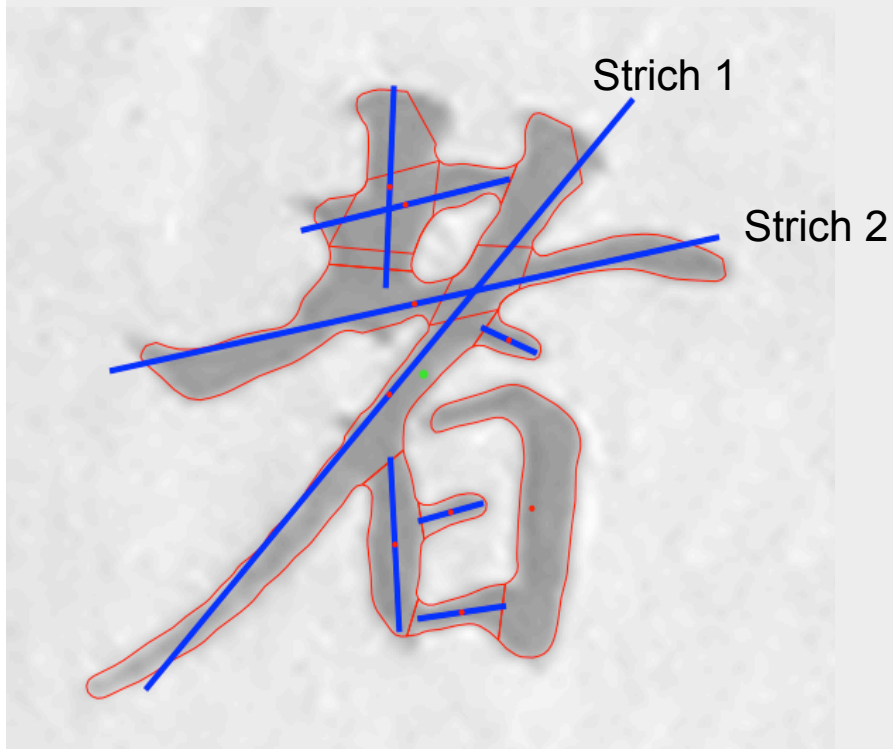
- **Strichverläufe (Medialachsen)**
- **mögliche Strichanfänge und Strichenden**

**Komplexer Verschmelzungsprozess liefert  
endgültige Strichrekonstruktion**



## Vermessen eines Schriftzeichens

Anhand der ermittelten Striche eines Zeichens können zahlreiche Merkmale berechnet werden.



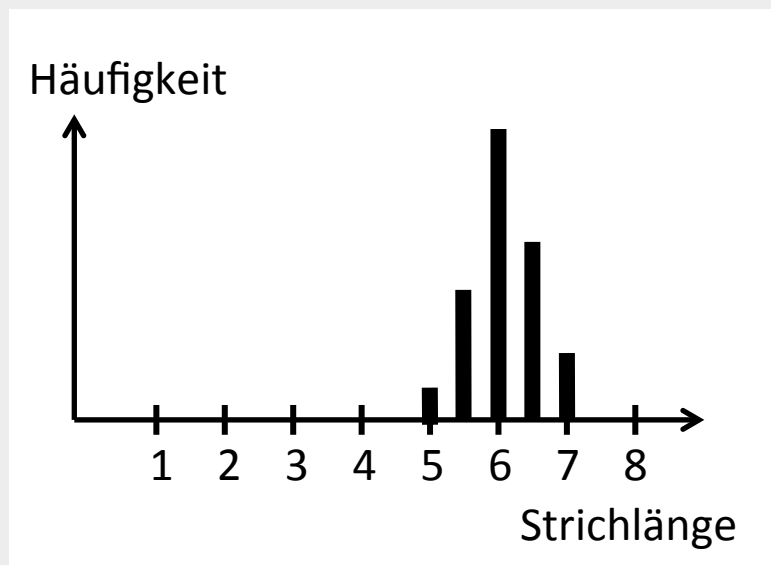
Orientierung Strich 1	49,5°
Orientierung Strich 2	14,3°
Winkel Strich 1 – 2	35,2°
Länge Strich 1	9,2mm
Länge Strich 2	7,5mm
Längenverhältnis Strich 1 – 2	1,23
Seitenverhältnis	1,05
Mittlere Strichdicke	0,52mm
Varianz Strichdicke	0,04mm <sup>2</sup>

## Häufigkeitsverteilungen von Merkmalswerten

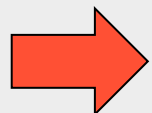
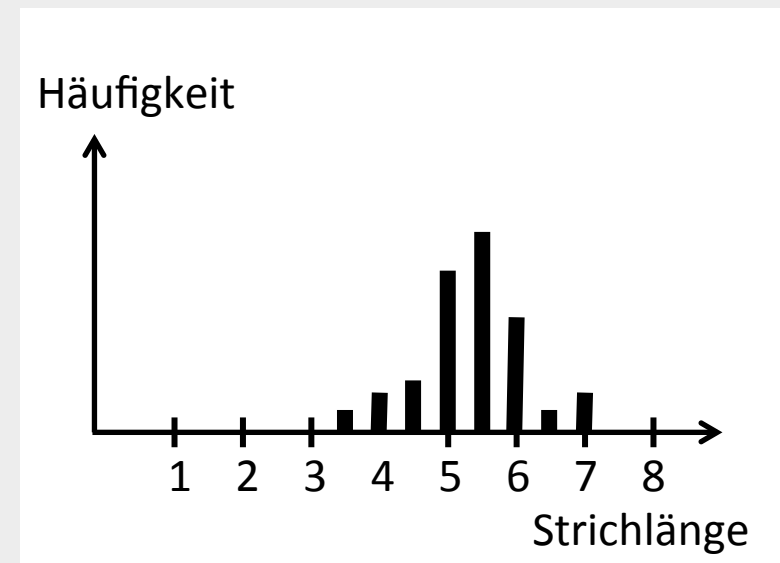
Merkmalswerte verschiedener Schreiber haben meist verschiedene Häufigkeitsverteilungen

Beispiel: Häufigkeitsverteilung der Länge von Diagonalstrichen

Proben Manuskript A



Proben Manuskript B



**"Manuskripte A und B stammen mit 80% Wahrscheinlichkeit von verschiedenen Schreibern"**



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Collaborative Research

*Centre for the Study of Manuscript Cultures*

Manuskriptkulturen in Asien, Afrika und Europa

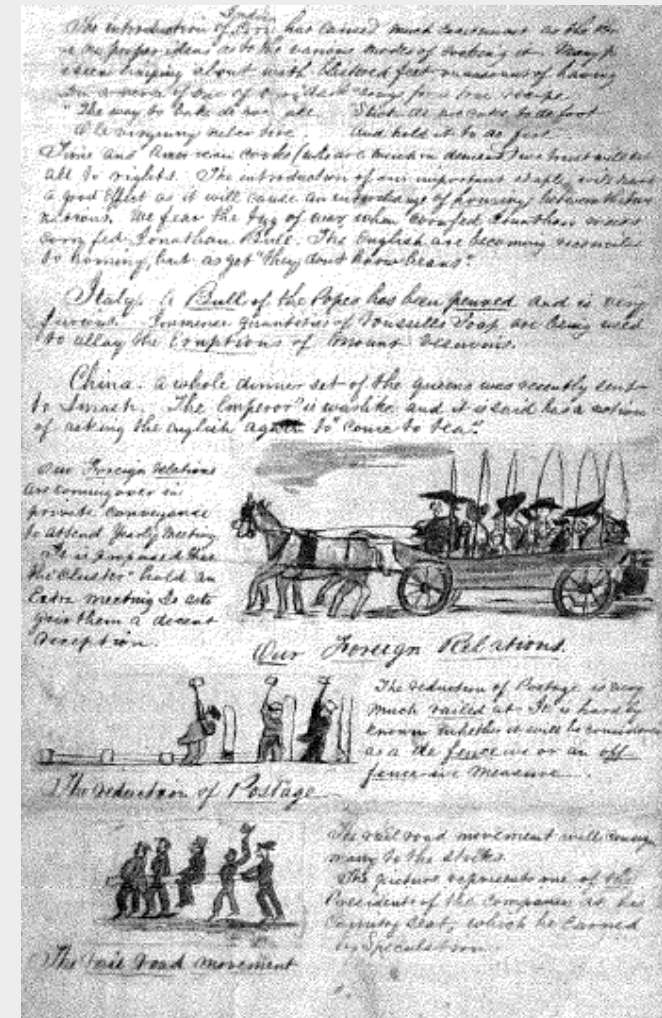
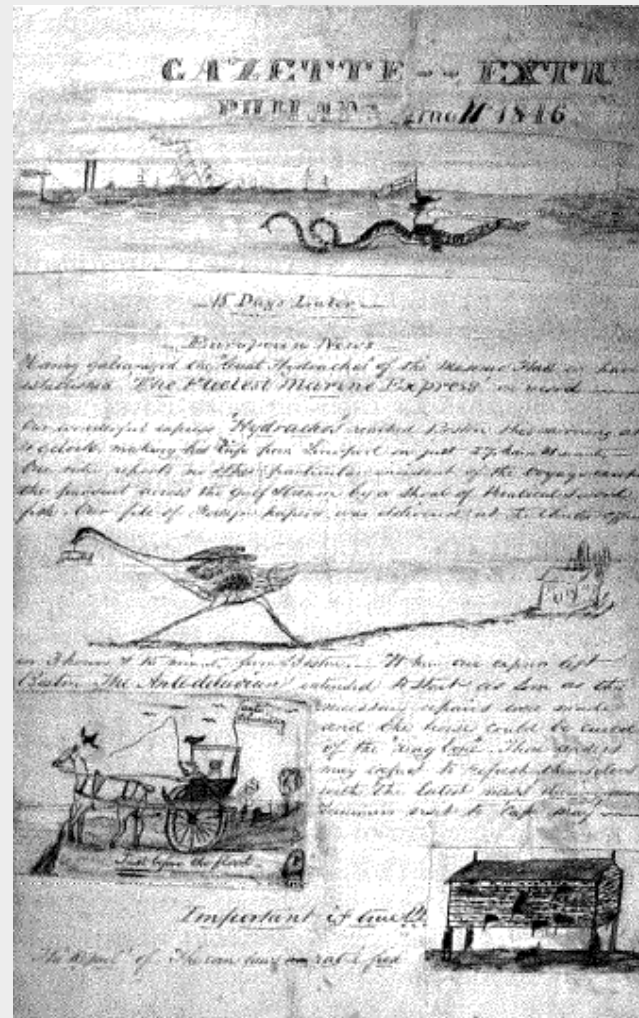
Sonderforschungsbereich 950

# Rechner als Black Box: Forensische Schreiberverifikation



# War Hermann Melville der Schreiber?

Zwei Seiten einer  
 satirischen Zeitung  
 von 1846,  
 Philadelphia, USA



# Von Hermann Melville geschriebene Briefe

My dear Anne  
 Buffalo Nov 20<sup>th</sup> 1857

As my departure from Albany last  
 fall with letters you express a sense that I should write  
 to you when my school should have gone into printing  
 but now a paper wants to appear returning you for  
 report your request. Still however I consider my  
 journal to you - & it is with pleasure that I take  
 pains to inform it.

I should have taken up my pen  
 at an earlier day had not the variety & importance of  
 the scenes incident to my school been so numerous  
 and pressing, that they absorbed a large portion of my  
 time.

But now having become somewhat more  
 quiet into the routine of business - having established  
 a system in my mode of instruction - and being fa-  
 miliar with the character & composition of my scholars  
 in short having got my school more or less upon a  
 regular - a few minutes of time are afforded me which  
 I employ by occasional writing & so forth.

My scholars are about half a number of  
 all ages, sizes, colors, characters, & complexions, some of the  
 white have retained the eyes of light - can see as a man in  
 armor, while others have grasped through the darkness  
 but with the most surprising that they can not recognize objects  
 in the way or a screen forming - & are about an eighth  
 of them as though they had some paper that was before.

They school is situated in a remote & secluded part of the town  
 about five miles from the village, and the house attached to it  
 (over)

New York Jan 22<sup>nd</sup> 1858

My dear Anne  
 What a delicious season is yours  
 the most enjoyable that we can wish for any  
 winter. I don't know how it is possible, but I have  
 always been very partial to the probability of winter &  
 we are never wanting the snow which covers  
 our own fields & meadows - & the snow that  
 falls last winter - which we had a month or two  
 time in - but they had the snow only about to the  
 same that all traces of delightful characteristics  
 - but that issue of general admiration into the  
 air when applied to the other that under the  
 white sign see are a general sense & are making  
 access by general attention - than the white are  
 a general note are decidedly handsome, but of no very  
 great part of them. Some are some of single feet  
 and others they will not be found to be the same  
 good looks, and at every rule they are never particularly  
 ugly. But I have further about some of the  
 birds & "How come it that appearance does" & all  
 that sort of thing - some of the white are only

# Schreiberverifikation mit CEDAR-FOX (1)

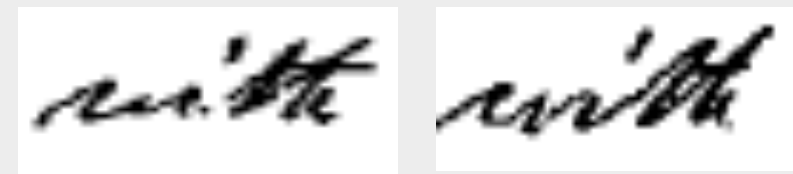
**Center for Excellence in Document Analysis and Recognition**  
**Sargur N. Srihari, University at Buffalo**

## Vorbereitung der Dokumente:

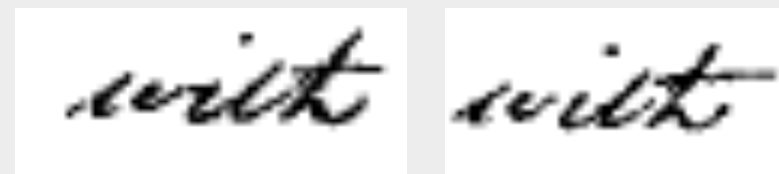
- **Manuelle Beseitigung von Nicht-Text und groben Verunreinigungen**
- **Binarisieren, Zerlegen in Zeilen und Worte, automatische Zeichenerkennung**
- **Manuelle Korrekturen der Wortsegmentierung**
- **Zuordnung gleicher Worte aus bekanntem und unbekanntem Dokument**

## Beispiele:

- bekanntes Dokument

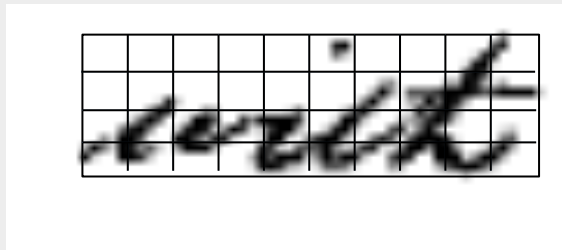


- unbekanntes Dokument



## Schreiberverifikation mit CEDAR-FOX (2)

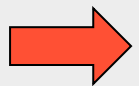
Ein Wort wird in Gitter von 4 x N Zellen unterteilt:



Jede Zelle wird mit lokalen Merkmalen beschrieben:

- überwiegende Strichrichtung
- Präsenz von Ecken, Diagonalen, Horizontalen und Vertikalen

Entsprechende Worte werden anhand einer Kette von ca. 1000 Bits verglichen. Unterschiede werden an einem Schwellenwert gemessen.



**"Ja, Hermann Melville hat die satirische Zeitung geschrieben."**

Forensische Schreiberverifikation ist in den USA gerichtlich zugelassen.



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Collaborative Research

*Centre for the Study of Manuscript Cultures*

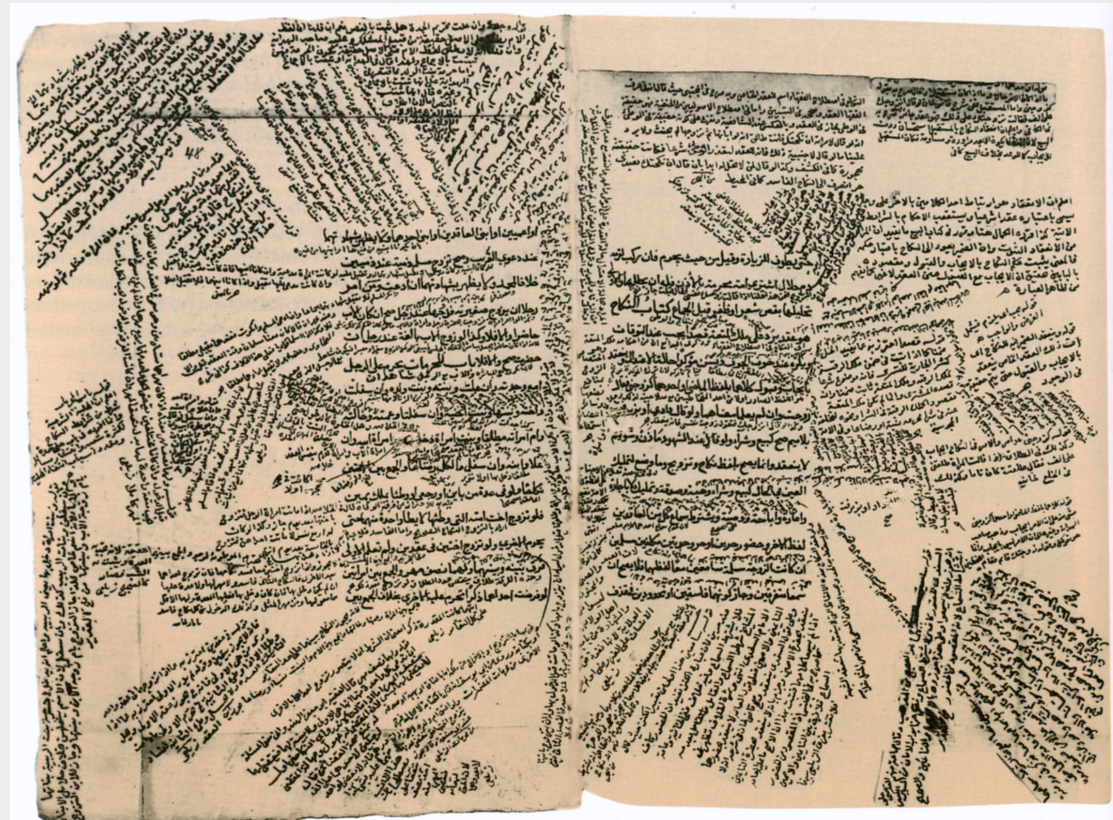
Manuskriptkulturen in Asien, Afrika und Europa

Sonderforschungsbereich 950

# Automatische Layout-Analyse

# Layout-Analyse in historischen Manuskripten

## Bestimmen von Textblöcken als Voraussetzung für automatische Zeichenerkennung (OCR)



# Layout Analyse mit Gabor Transformation

(Herzog et al. 2014)

Lokale Frequenzanalyse - fasst Zeilen mit gleichem Zeilenabstand und gleicher Orientierung als Textblöcke zusammen



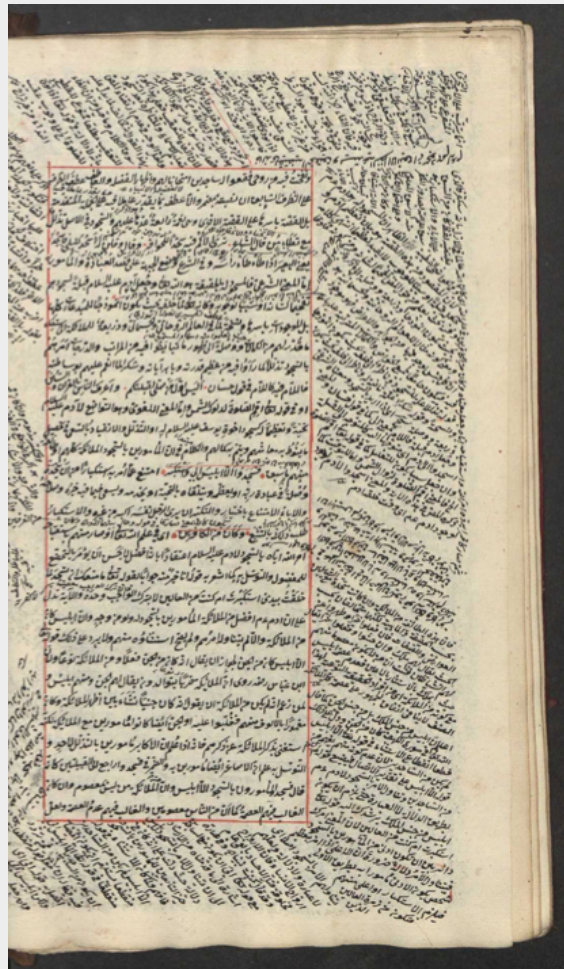
	COG (x,y)	Area [Pixel]	Line distance mean [Pixel]	Line distance variance	Line orientation mean	Line orientation variance
Region 1	(545,1220)	45623	89.93	77.777	90.39	64.39
Region 2	(645,300)	10421	44.95	221.5	93.25	317.08
Region 3	(1050,620)	4412	45.78	85.6	29.58	156.48
Region 4	(775,2150)	7108	35.35	60.97	164.12	134.98
Region 5	(1140,1965)	3247	39.55	69.22	33.62	705.81
Region 6	(1075,1380)	10920	40.67	112.92	114.88	977.64
Region 7	(115,1070)	953	36.65	192.38	98.95	1474.81
Region 8	(655,1045)	2226	49.53	133.38	90.43	245.09

## Beispiele (1)





## Beispiele (2)



## Beispiele (3)





Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Collaborative Research

*Centre for the Study of Manuscript Cultures*

Manuskriptkulturen in Asien, Afrika und Europa

Sonderforschungsbereich 950

# **Mustererkennung in mittelalterlicher Musiknotation**



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Collaborative Research

*Centre for the Study of Manuscript Cultures*

Manuskriptkulturen in Asien, Afrika und Europa

Sonderforschungsbereich 950

# Kooperation zur Analyse von mittelalterlicher Musiknotation

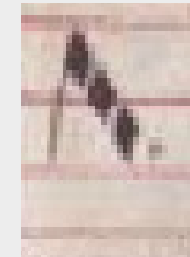
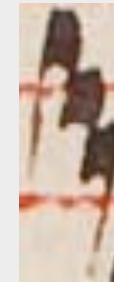
(Solth 2014)



## Musikologische Fragestellungen



**Wie häufig sind spezielle Ligaturen in schottischen bzw. französischen Manuskripten?**



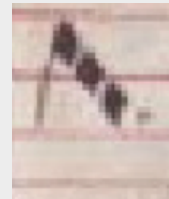
**Haben unterschiedlich lange Notenhälsen eine Bedeutung?**



**Großes Datenvolumen von > 1200 Seiten muss analysiert werden!**

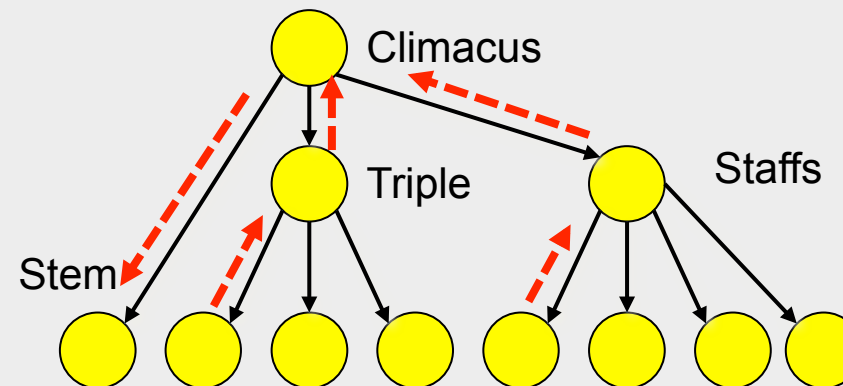
# Formale Modelle für kompositionelle Strukturen

**Kompositionelle Struktur:**



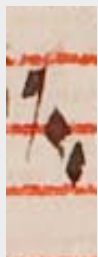
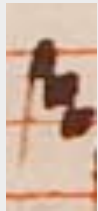
**Computerinterne  
Repräsentation als Graph:**

**Systematische bottom-up  
und top-down Suche  
nach Bildstrukturen, die  
mit dem Modell  
übereinstimmen:**



Name:	Climacus
Parent:	Ligature
Bounding Box:	< 150 x 200
Parts:	Staffs, Triple, Stem
Constraints:	Triple matches Staffs Stem touches upper left of Triple

## Resultierende Statistik



Manuskript	W1	W2	F
# Seitenzahl	379	508	798
# Ligatur Typ 1	2703	1089	3891
# Ligatur Typ 2	2	4	6
# Ligatur Typ 3	1398	549	1364
Verhältnis Typ 1:Typ 3	~2:1	~2:1	~3:1

**W1 = Schottisches Manuskript 1 (Herzog August Bibliothek Wolfenbüttel)**  
**W2 = Schottisches Manuskript 2 (Herzog August Bibliothek Wolfenbüttel)**  
**F = Französisches Manuskript (Notre Dame)**

**Häufigkeit von Ligaturtypen lässt Rückschlüsse auf kulturellen Austausch zu.**



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Collaborative Research

*Centre for the Study of Manuscript Cultures*

Manuskriptkulturen in Asien, Afrika und Europa

Sonderforschungsbereich 950

# Multispektralanalyse zur Wiedergewinnung ausgelöschter Schrift



## Archimedes Palimpsest

**Palimpsest = wieder abgeriebene Schrift, um (teures) Pergament oder Papyrus noch einmal verwenden zu können**

**Kopien von 7 Abhandlungen des Archimedes (287 – 212 v. Chr.) aus dem 10. Jhd. n. Chr.**

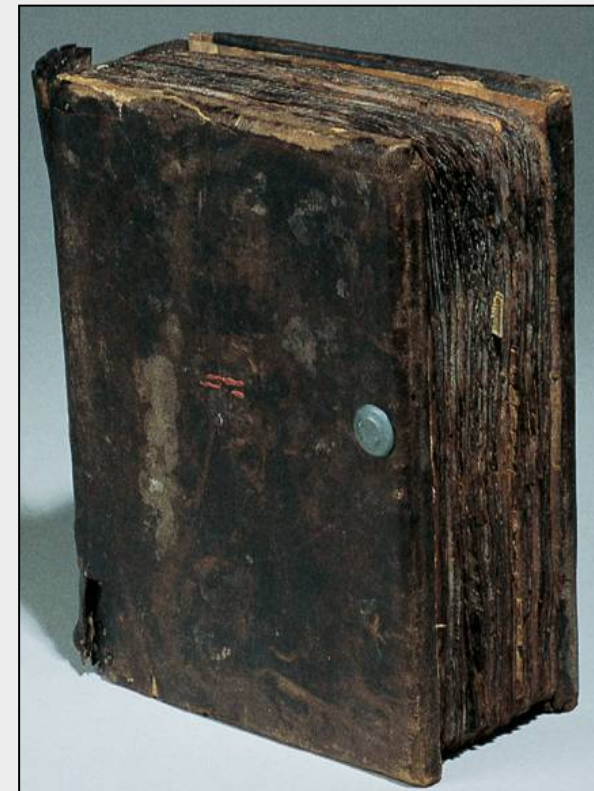
**- einzige Kopie von "Traktat über schwimmende Körper" in original griechischer Sprache**

**- einzige Kopien von Blättern aus "Methode mechanischer Theoreme" und "Stomachion"**

**+**

**weitere Seiten anderer Manuskripte**

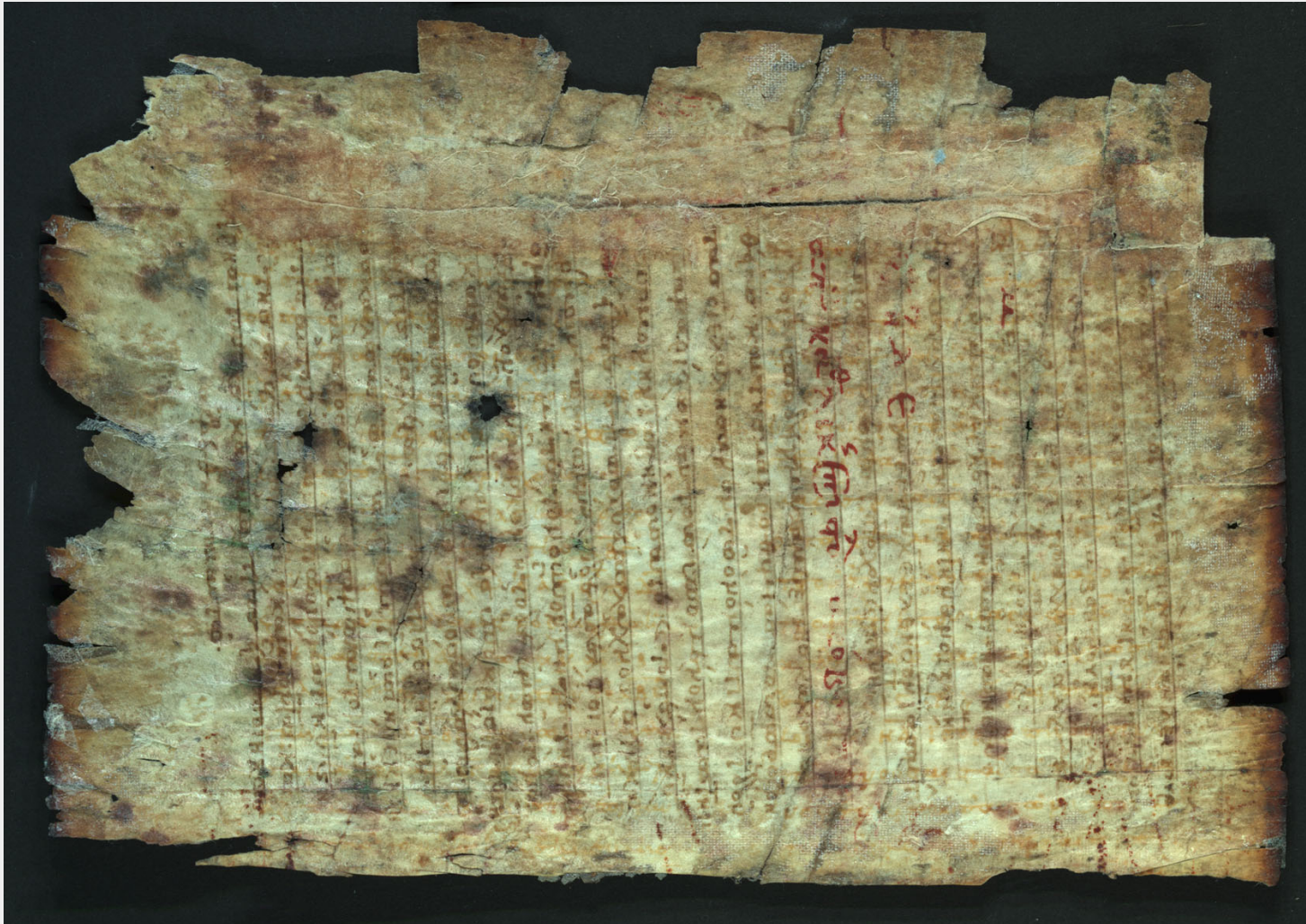
**177 Pergamentseiten, abgerieben und überschrieben mit "Euchologion" (christliches Gebetsbuch) im Jahr 1229 n. Chr.**



Bilder von Roger L. Easton  
Rochester Institute of Technology

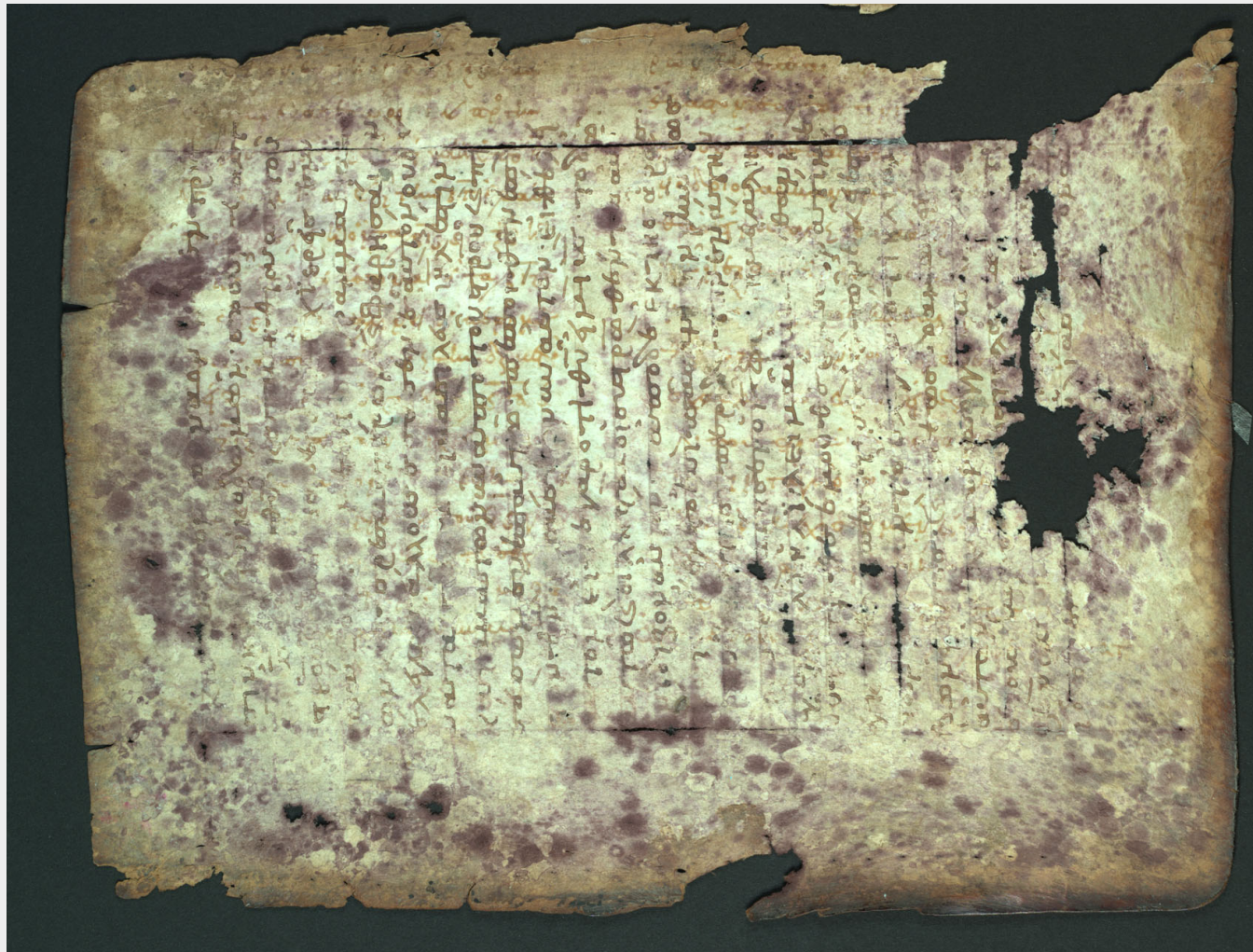


## Seite aus "Methode mechanischer Theoreme"





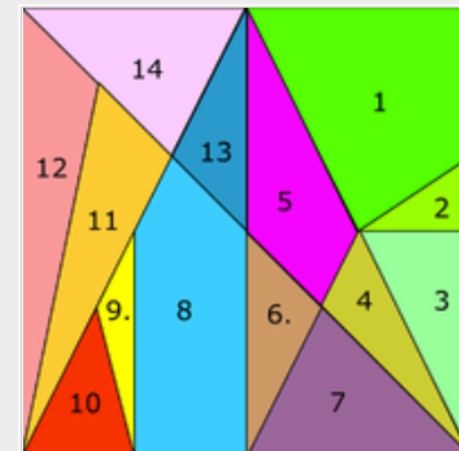
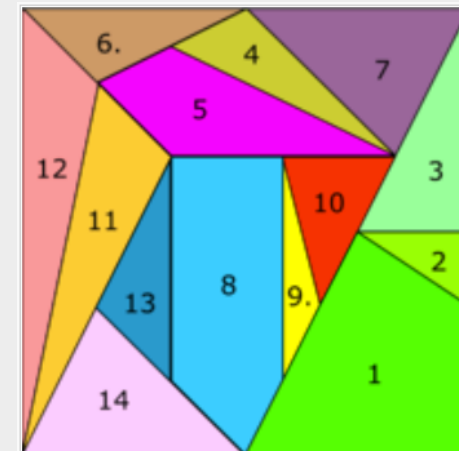
## Seite aus "Stomachion"



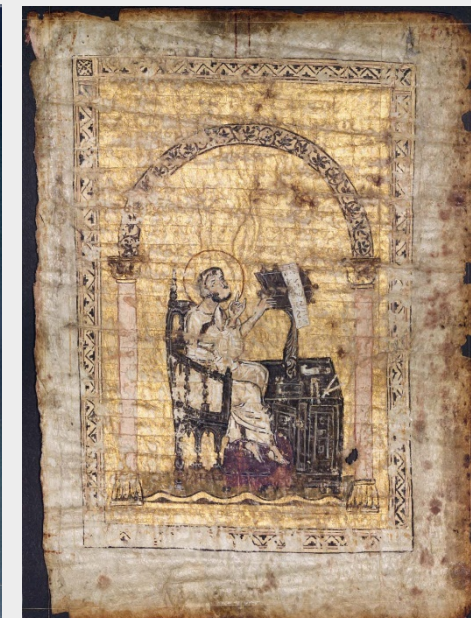
## Das Puzzle "Stomachion" des Archimedes

Traktat untersucht, auf wieviel verschiedene Weisen 14 Puzzleteile zu einem Quadrat zusammgelegt werden können.

(536 ohne Spiegelungen und Drehungen)



## Neuere Übermalungen (1938)



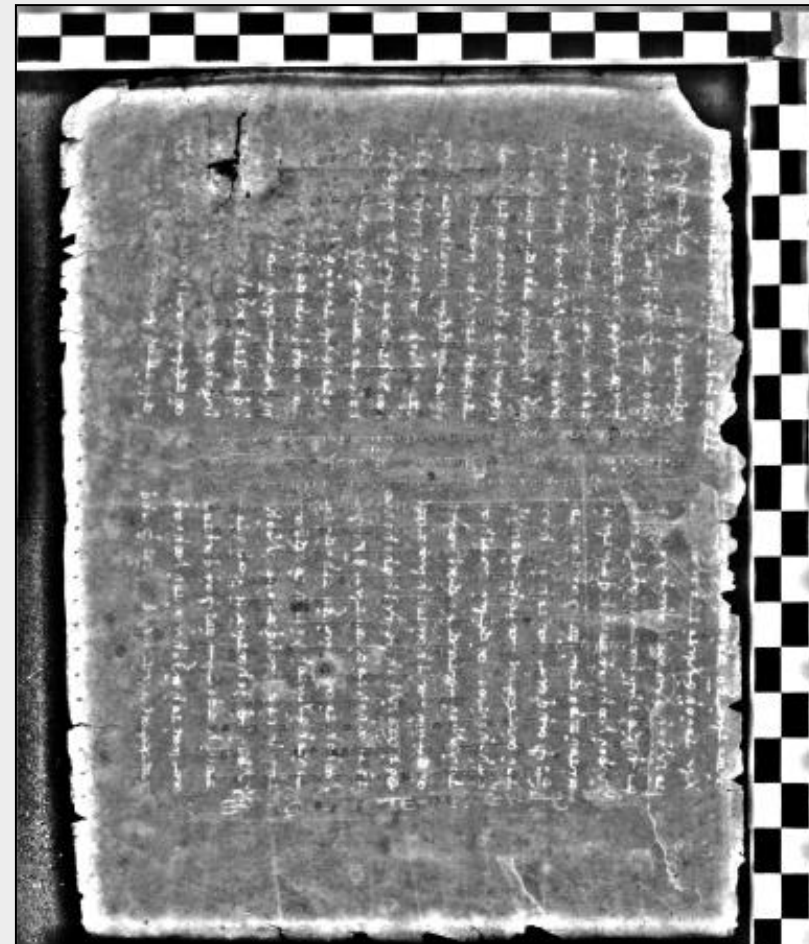
**Mehrere Seiten von "Traktat über schwimmende Körper" und "Methode mechanischer Theoreme" wurden offenbar zur Wertsteigerung mit christlichen Motiven übermalt.**

## Wiederherstellung von abgeriebenem Text

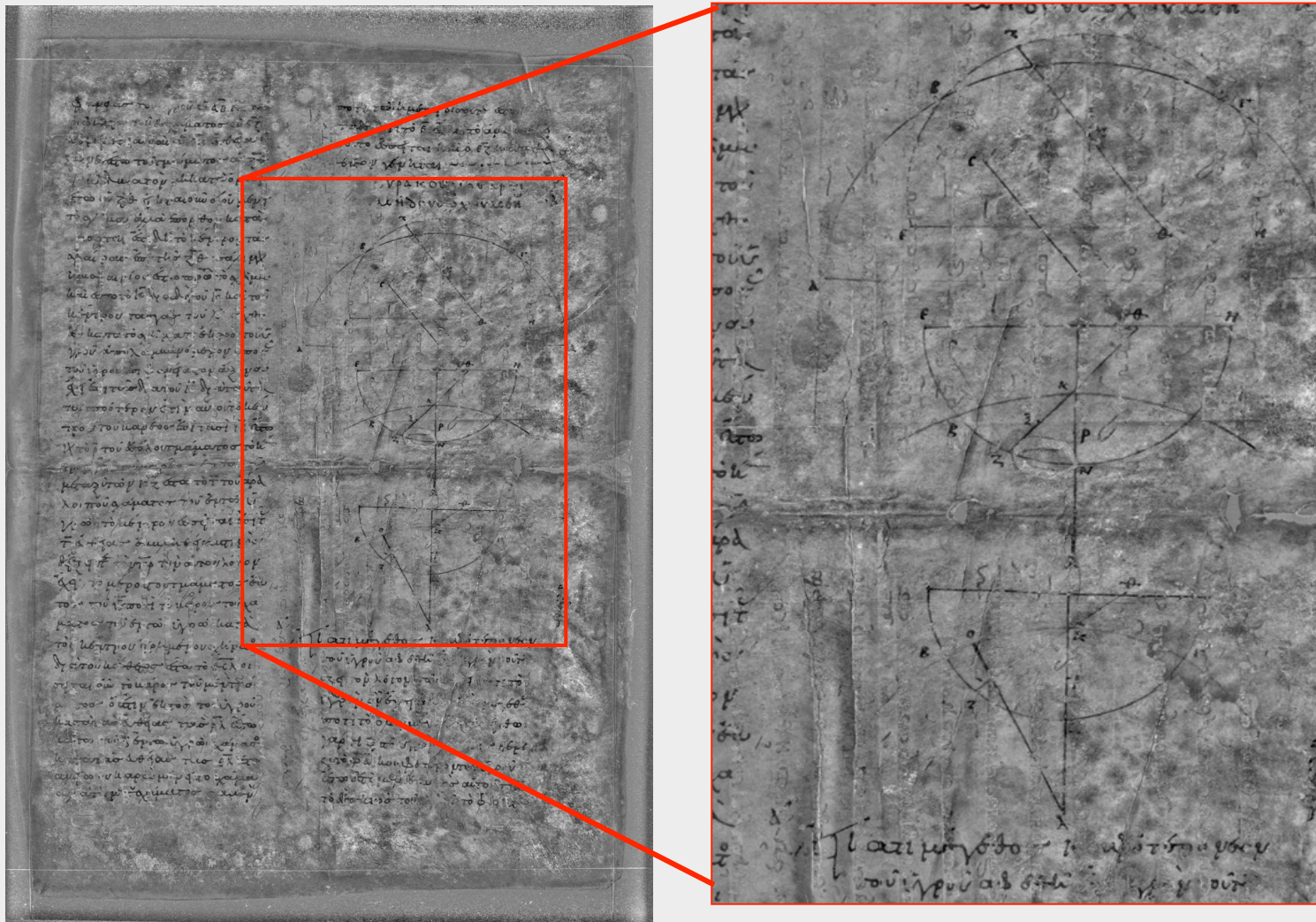
Original



nach Multispektralanalyse



# Wiederherstellung von Zeichnungen





## Digitale Aufbereitung

### Schritt 1:

**Erzeugung multispektraler Bilder mit Spezialkameras**

**Bilder aus dem Infraroten und ultravioletten Bereich enthalten Informationen über unsichtbare Spuren der abgeriebenen Schrift.**

### Schritt 2:

**Bestimmen einer Kombination von Farbauszügen, mit der der Kontrast der abgeriebenen Schrift maximiert wird.**

**Mathematisches Werkzeug aus der Mathematik und Bildverarbeitung:  
PCA (Principle Component Analysis)**

*Es gibt Dinge, die den meisten Menschen unglaublich erscheinen,  
die sich nicht mit Mathematik beschäftigt haben.*

*Archimedes*



## **Rückblick auf sieben Jahre ...**

**... digitaler Unterstützung für Geisteswissenschaftler bei der Manuskriptanalyse**

- **Spannend, wie auch andere interdisziplinäre Forschungsarbeit**
- **Wechselseitige Befruchtung zwischen Bildverarbeitung und Manuskriptforschung**
- **Wesentliche Fortschritte bei der Manuskriptanalyse durch Verarbeitung großer Datenvolumina**
- **Einfache Rechnerhilfe zur Selbstbedienung ist Manuskriptforschern lieber als Zuleistung durch ein komplexes Programm.**
- **Kultureller Abstand zwischen Informatikern und Manuskriptforschern in Sprache und Zielen erschwert die Arbeit.**



**Vielen Dank für Ihr Interesse!**