Computer-based Support for Manuscript Analysis in the Humanities

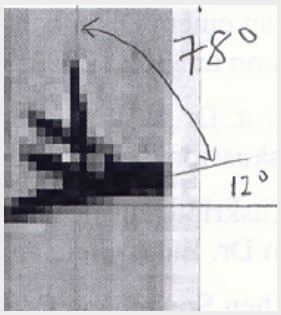
manuSciences '15, Franco-German Summer School

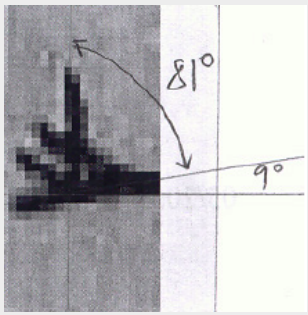
Bernd Neumann
Department of Informatics, University of Hamburg

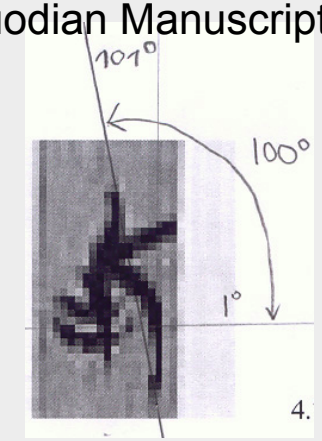
Why Computer-aided Manuscript Analysis?

Richter 2006:
"Tentative Criteria for Discerning Individual Hands in the Guodian Manuscripts"

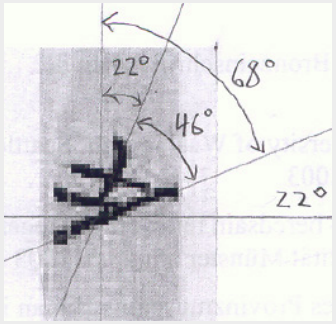
Scribe A:

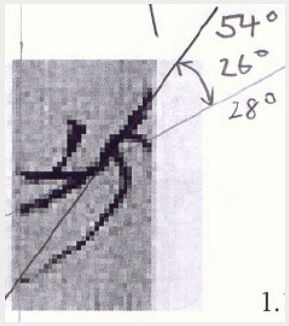


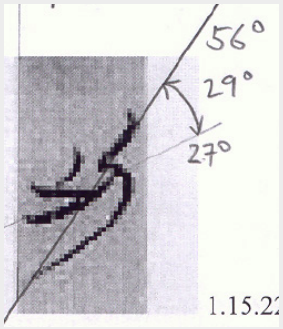




Scribe B:

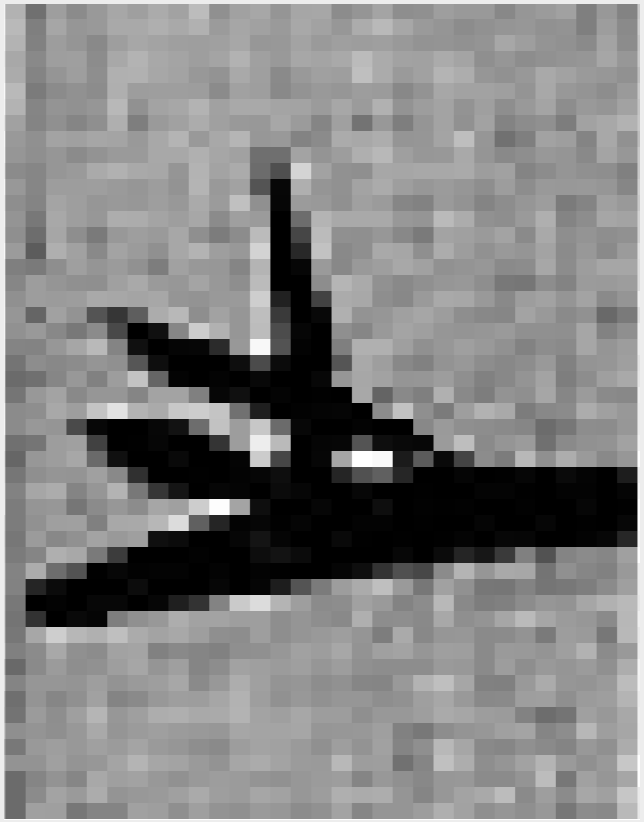






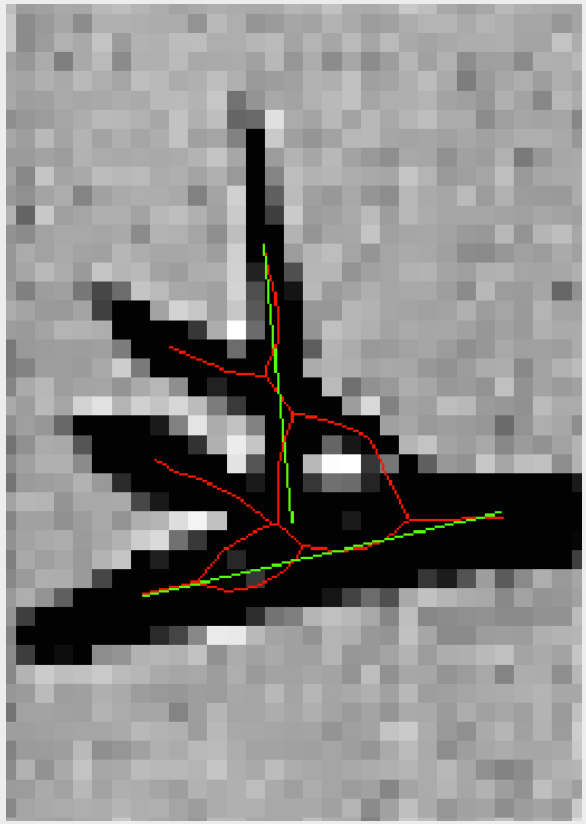
Computer methods can - provide objective and repeatable measurements,
- deal with large data volumes

Example: Computer-based Measurements



coarse image

contours



skeleton and axes

Agenda

- **Centre for the Study of Manuscript Cultures**
  at the University of Hamburg
- **Stroke analysis**
  Determining individual strokes in Chinese characters
- **Layout Analysis**
  Locating text blocks, determining line structure
- **Recognition of composite writing patterns**
  Finding ligature patterns in medieval music notation

Centre for the Study of Manuscripts Cultures

DFG Special Research Unit "Manuscript Cultures in Asia, Africa and Europe"

Manuscripts as material entities, relations to social and cultural context

First phase 2011 – 2015

17 sub-projects in diverse areas of the humanities, organized in 3 project areas

- Paratexts
- Visual Organisation
- Manuscript Collections

3 "Scientific Service Projects"

- Recovering lost writing
- Reconstructing manuscript history with methods of Material Science
- Determining visual manuscript and character features using computer-based image analysis

http://www.manuscript-cultures.uni-hamburg.de

Service Project Image Analysis

Project team in the Department of Informatics:
Rainer Herzog, Arved Solth, Bernd Neumann

Work plan:

A Application of image processing methods for projects of the humanities

B Innovative image processing methods for manuscript analysis

C Prototype of a work place for manuscript analysis

Second Phase 2015 - 2019

Same main topics as in Phase 1:

- Paratexts
- Visual Organisation
- Manuscript Collections

3 Working Groups (cross-section topics)

- Learning
- Ritual
- Agency

Same service projects

- Main goal of image analysis project: Development of Advanced Manuscript Analysis Portal (AMAP)

* Different team: Siegfried Stiehl, Volker Märgner (PIs), N.N., N.N.

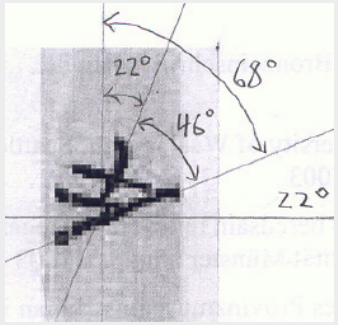
Stroke Analysis

Work by Rainer Herzog and Arved Solth

Motivation for Stroke Analysis

Recovering individual strokes may provide a basis for

- **recognizing characters,**
- **retrieving similar allographs,**
- **comparing the handwriting of different scribes.**



Computer methods are useful

- **for determining statistically significant numbers of stroke features,**
- **for computing features not immediately visible to the eye, e.g. relationship between stroke lengths.**

Basic Stroke Finding Procedure

- **Determine character contours (or silhouette)**
  - Compute gradient magnitude image
  - Sub-pixel watershed segmentation
- **Identify strokes**
  - Constrained Delaunay Triangulation
  - Merge partial strokes at junctions
- **Compute stroke features**
  - Location of medial axis, length, orientation, width
  - Relational properties

Determining Character Contours (1)

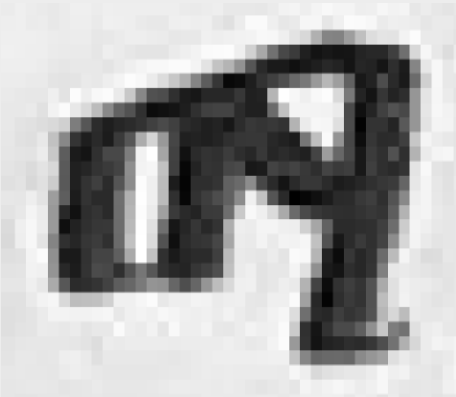
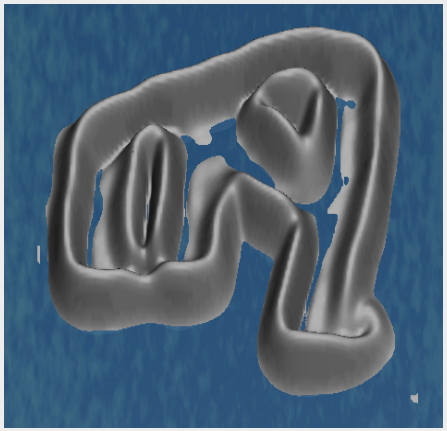
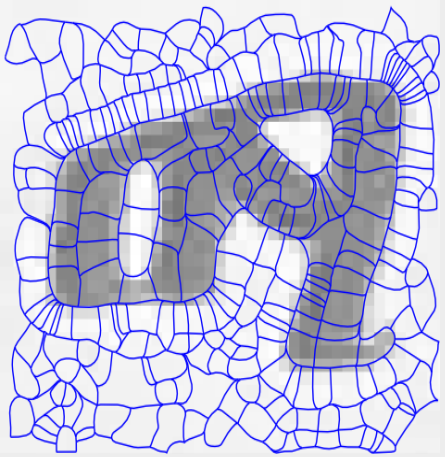


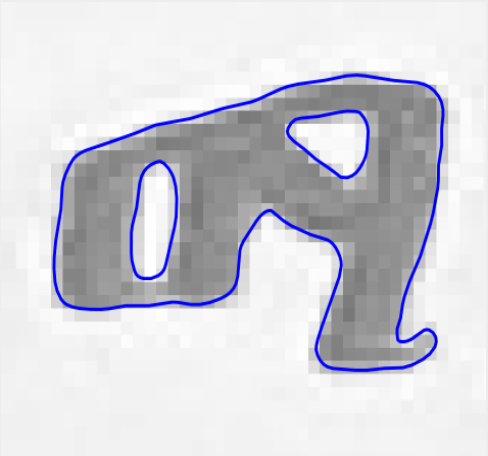
Image of Amharic character



Gradient magnitudes after interpolation



Watershed edges



Final contour after removing insignificant edges

Determining Character Contours (2)

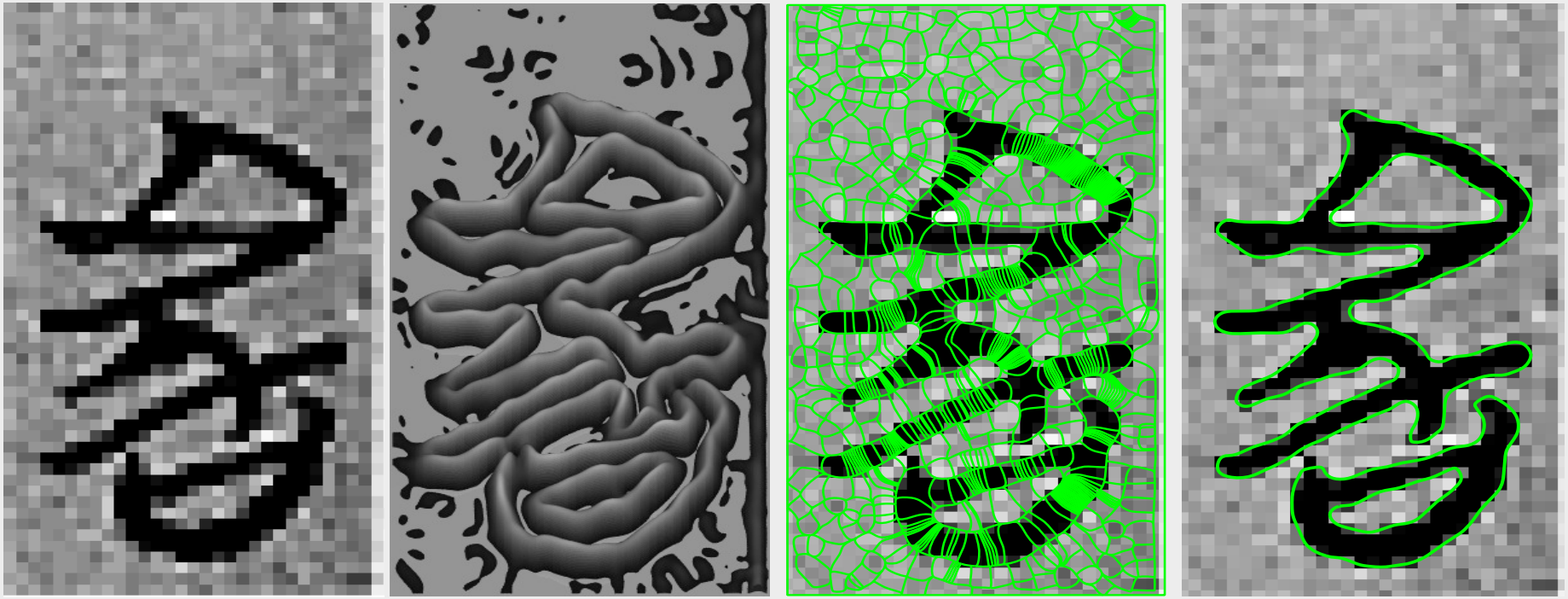


Image of Chinese character

Gradient magnitudes after interpolation

Watershed edges

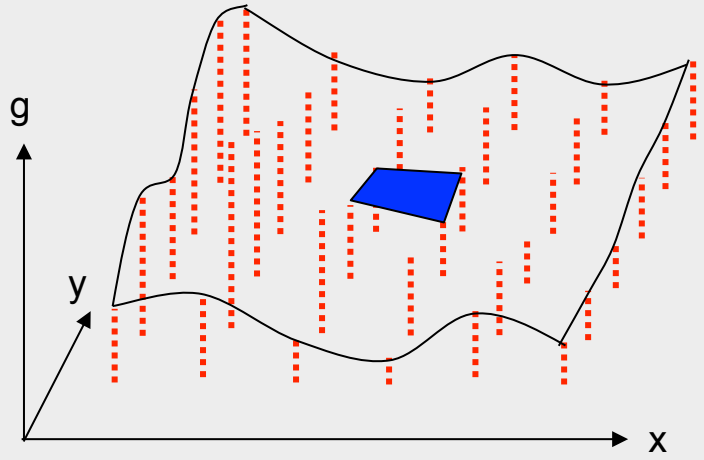
Final contour after removing insignificant edges

Subpixel Watershed Segmentation

A **Determine continuous image by Spline interpolation between pixels of discrete image**

B **Determine watershed lines (mathematically: lines connecting maxima and saddle points)**

C **Remove insignificant lines**



5-th Order Spline interpolation

There exist numerous methods for determining object boundaries.
Nice properties of watershed lines:

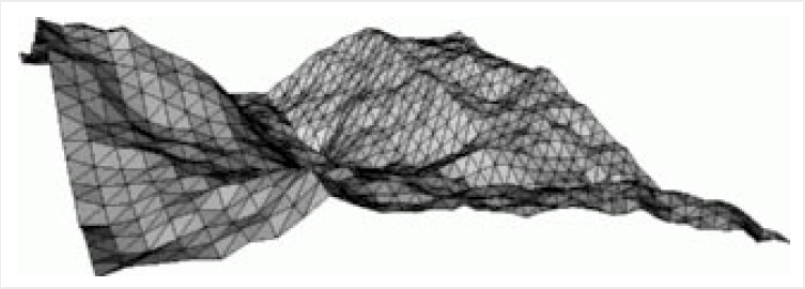
- closed boundaries

- no artefacts at junctions

Constrained Delaunay Triangulation (CDT)

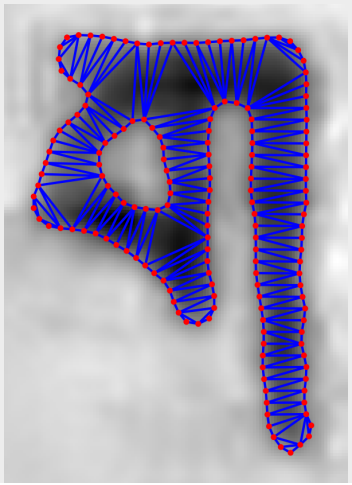
Delaunay Triangulation is used in Computer Graphics and Engineering to obtain a surface representation in terms of triangles.

Example: Obtaining a perspective view of a topological map



CDT is defined for points on a polygonal boundary such that no edge of a triangle crosses the boundary.

The density of boundary points can be chosen as fit for the application.



Why CDT for Stroke Analysis?

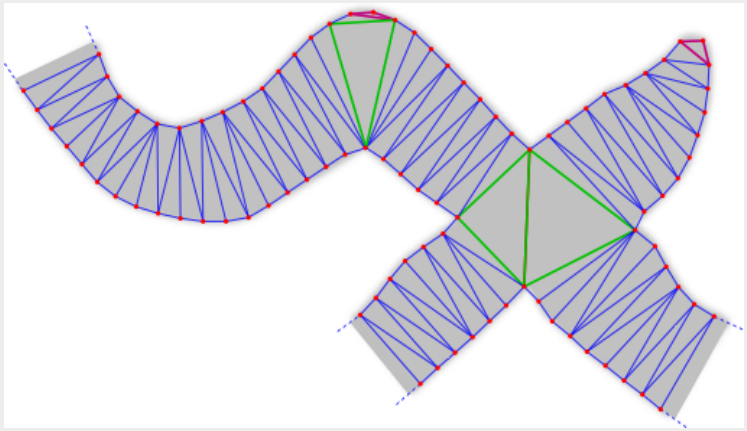
CDT generates triangles which (ideally) allow to recognize (i) stroke junctions, (ii) stroke middle sections and (iii) stroke endings (Solth et al. 2009).

Three types of triangles according to the number of *chords* (edges not coinciding with the contour):

- junction triangles (3 chords, green)

- sleeve triangles (2 chords, blue)

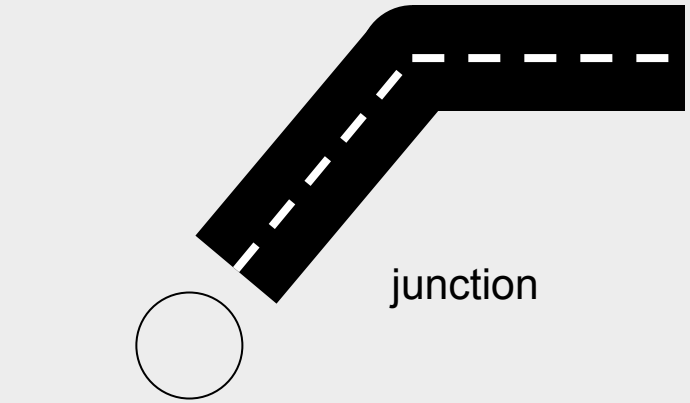
- terminal triangles (1 chord, red)



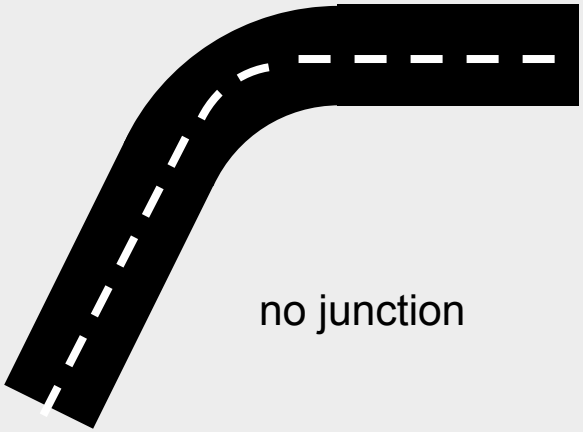
Stroke analysis amounts to collecting the connections from terminal or junction triangles via sleeve triangles to other terminal or junction triangles.

Corners vs. Curves

Junction triangles are generated within a curve, if the stylus has performed a sudden (discontinuous) orientation change.



idealized circular stylus



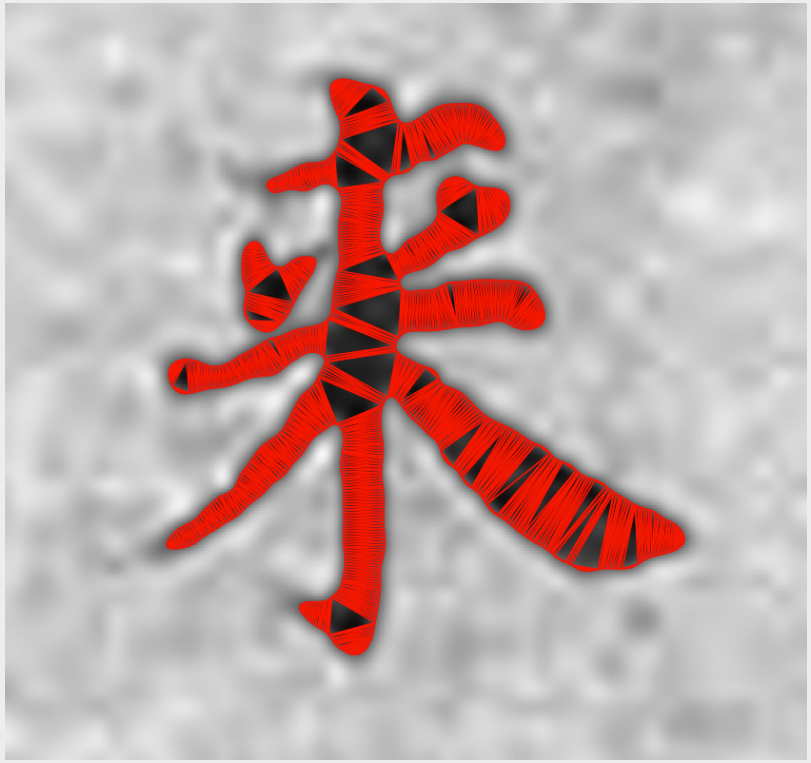
Exact conditions for junction:

- **Dense boundary points**
- **Stylus radius S, center line curve radius R, and angle α meet inequality**

$$R < S \frac{1 - \cos(\alpha/2)}{1 + \cos(\alpha/2)}$$

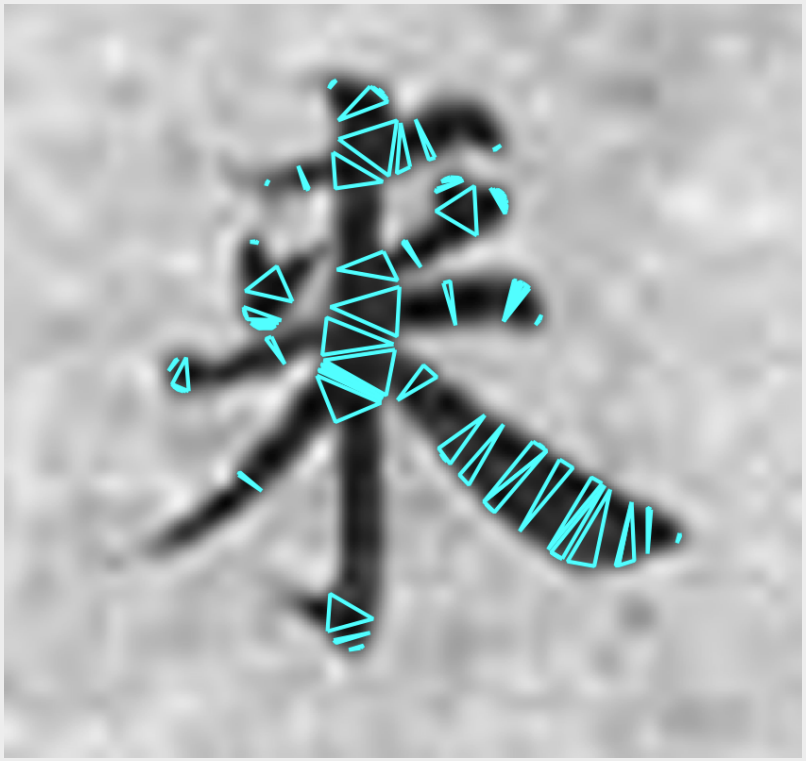
Triangulation Example

Chinese character for "come"
with computed contour

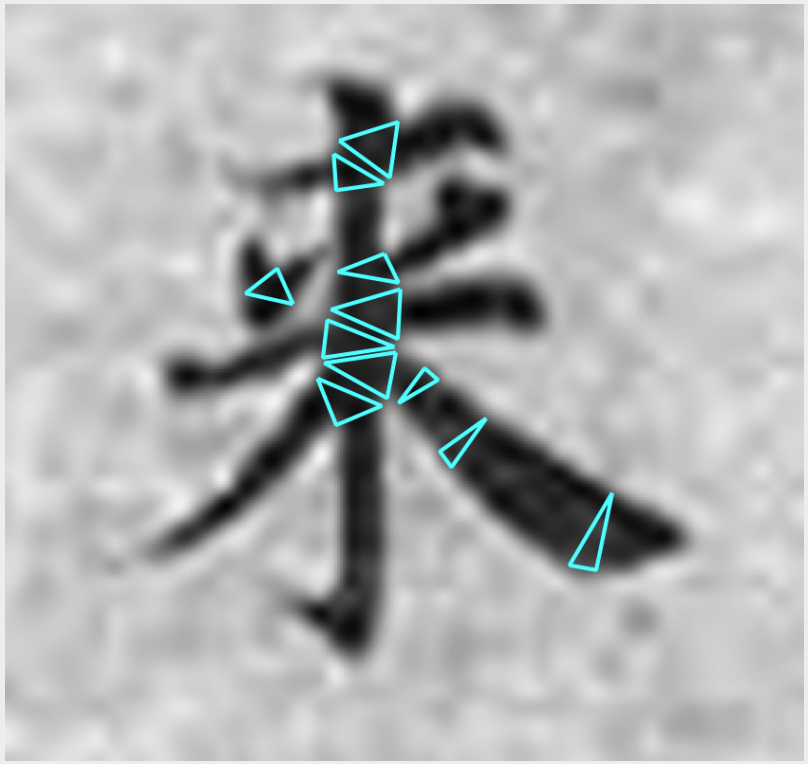


Constrained Delaunay
Triangulation

Junction Triangles



Junction triangles including spurious junctions due to handwriting irregularities

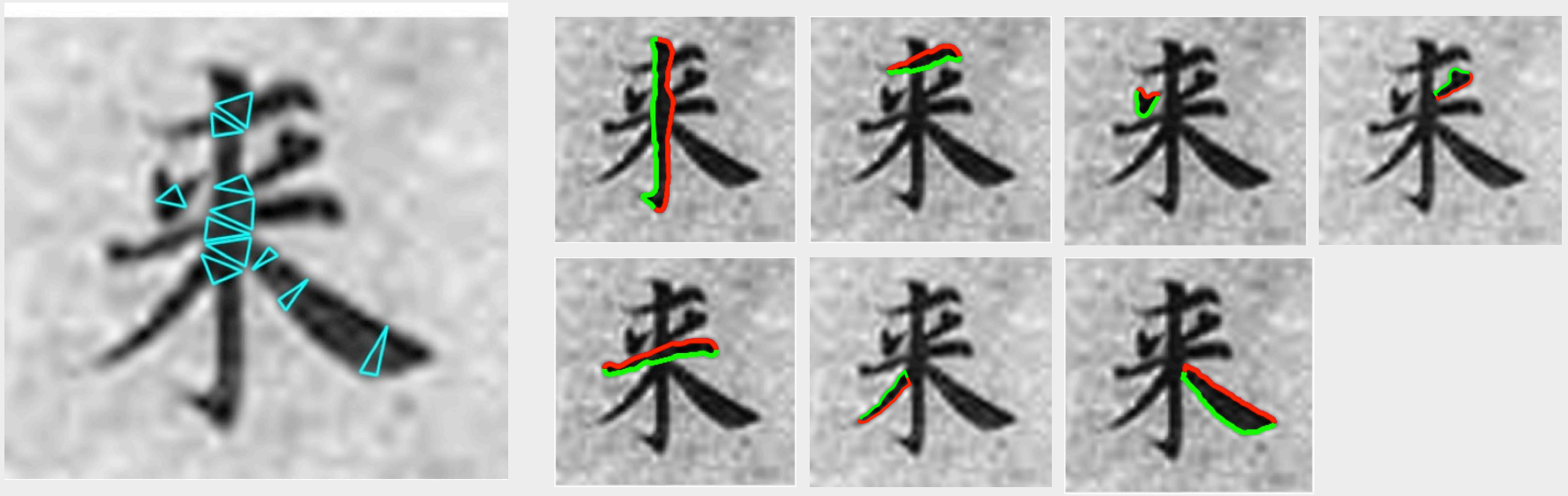


Cleaned up junction triangles

Merging Partial Strokes

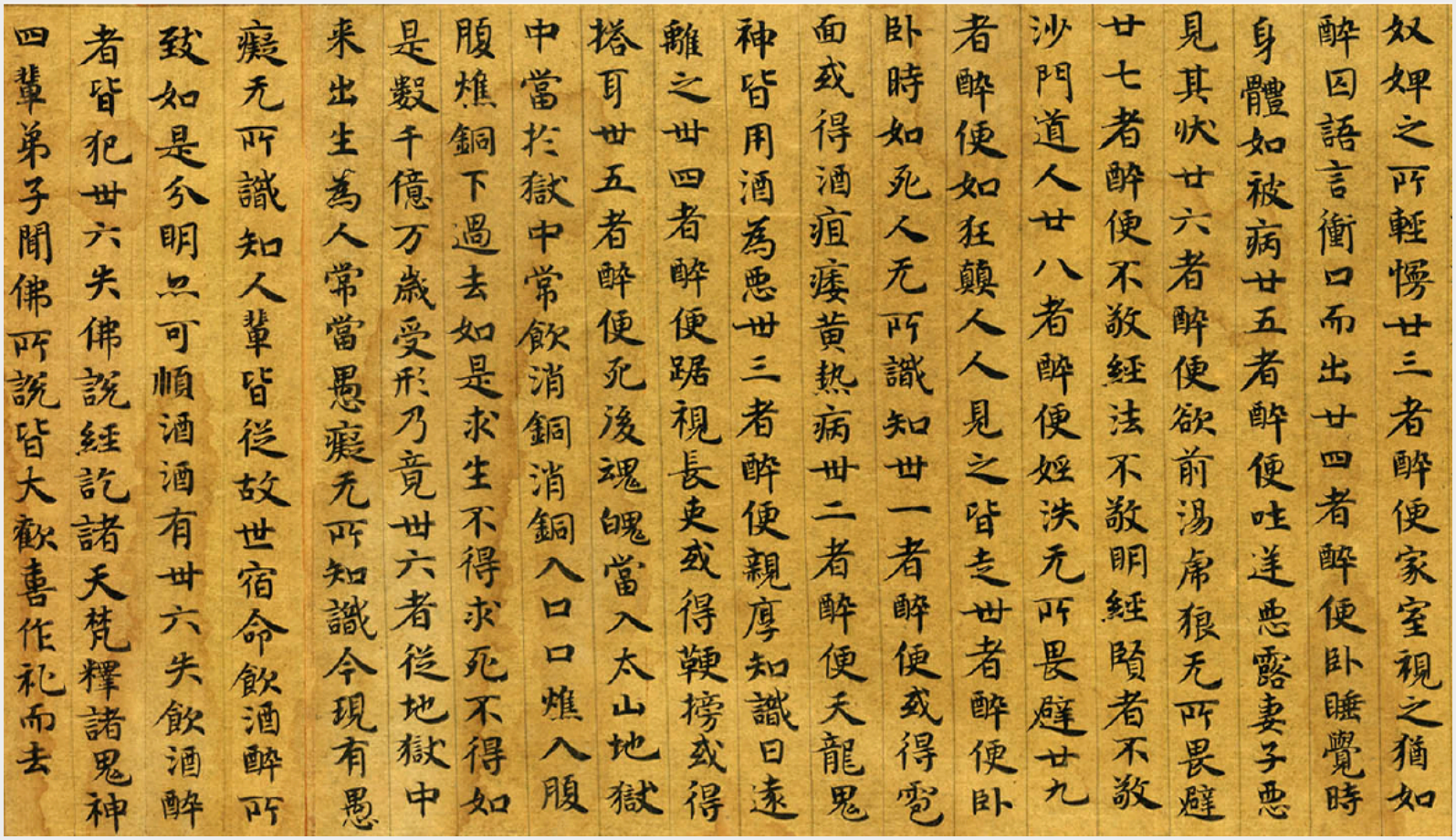
Partial strokes are merged by searching for an optimal stroke configuration:

- **smooth individual continuation**
- **best overall result**



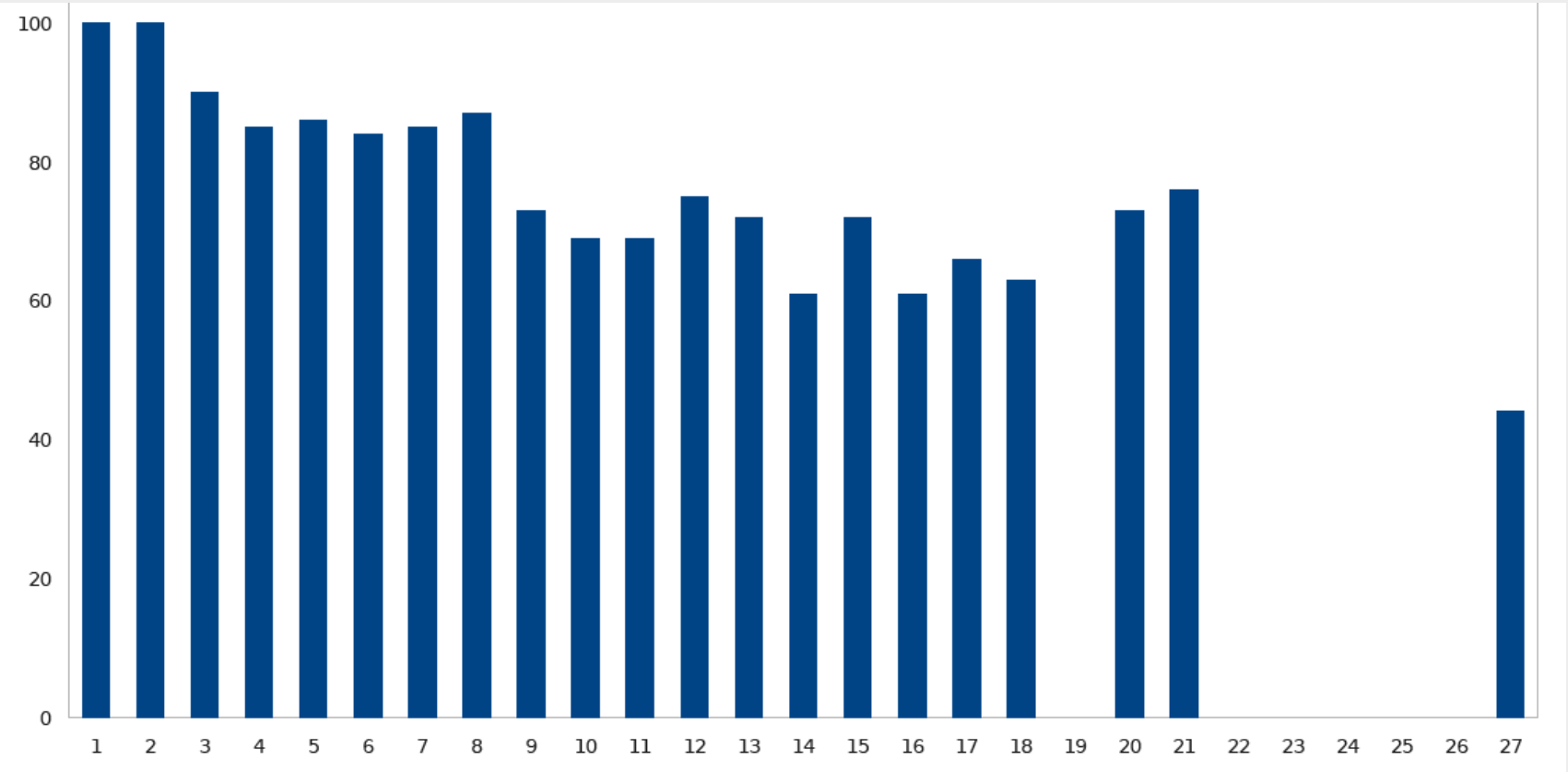
Evaluation (1)

339 Chinese characters, ca. 60 x 60 pixels each



Evaluation (2)

Stroke recognition rate decreases with character complexity



Summary of Stroke Recognition

- Constrained Delaunay Triangulation provides a sound conceptual basis.
- There exist proven and optimized algorithms.
- Merging partial strokes is not straight-forward.
- Results may be improved by knowledge about possible character shapes and writing rules.

Layout Analysis

with work by Rainer Herzog

Sanskrit manuscript on birchbark
(spiral line structure)

What is Layout Analysis?

Determining the locations of

- **text blocks (incl. paratexts)**
- **text lines (columns)**
- **characters**
- **(- strokes)**
- **non-textual elements**

Arabic Manuscript Multaqā al-abḥār (1641)
Ms.or.oct.261, p13v, Staatsbibliothek Berlin

Why Layout Analysis?

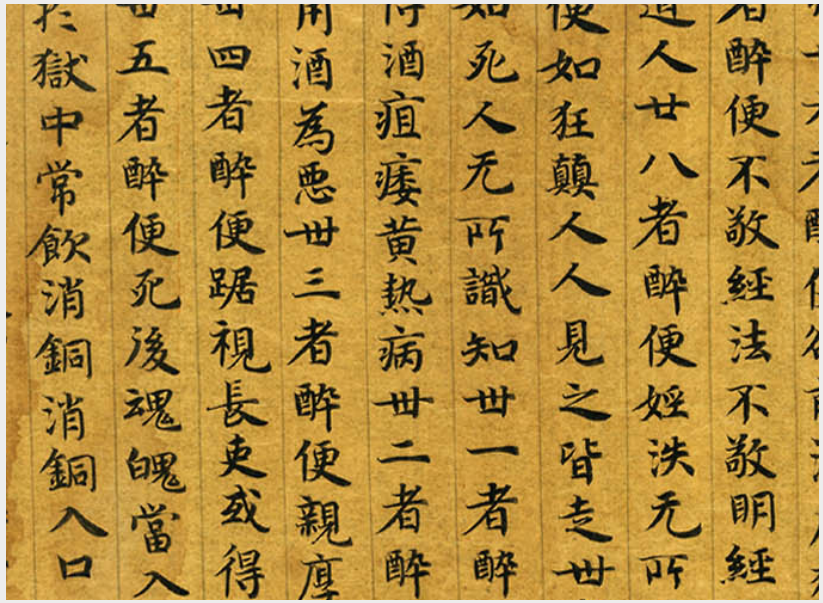
Human eyes are very good at discerning text block and character boundaries.

Main purpose of computer methods is to handle large data volumes.

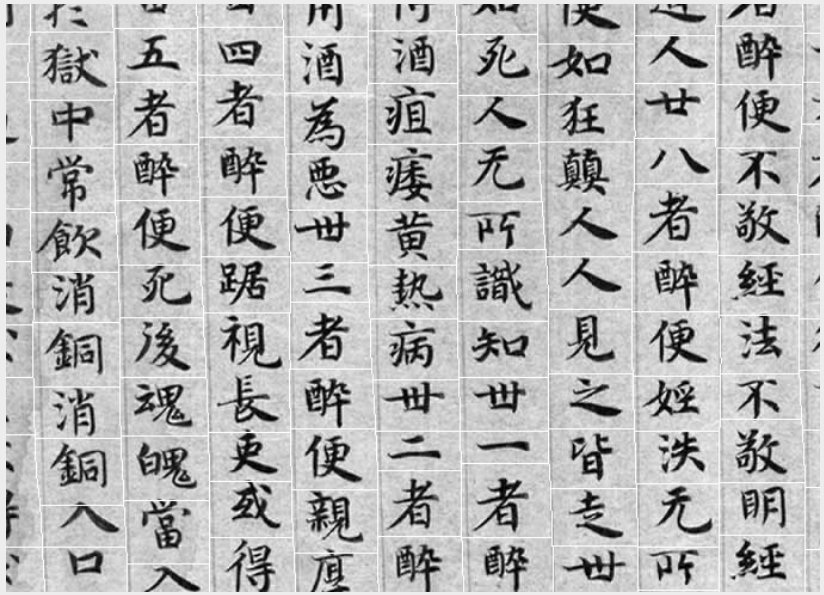
- **Layout analysis provides inventory of main text blocks, paratexts and other layout elements.**
- **Layout analysis delivers useful information (line frequency, orientation) for word and character segmentation.**
- **Layout analysis allows rectification of text blocks and thus application of well-developed methods for analyzing horizontal or vertical lines:**
  - **Line Segmentation**
  - **Word Spotting**
  - **Writer Identification**

Easy Cases

- **Text lines parallel to image rows (columns)**
- **No overlap between characters**



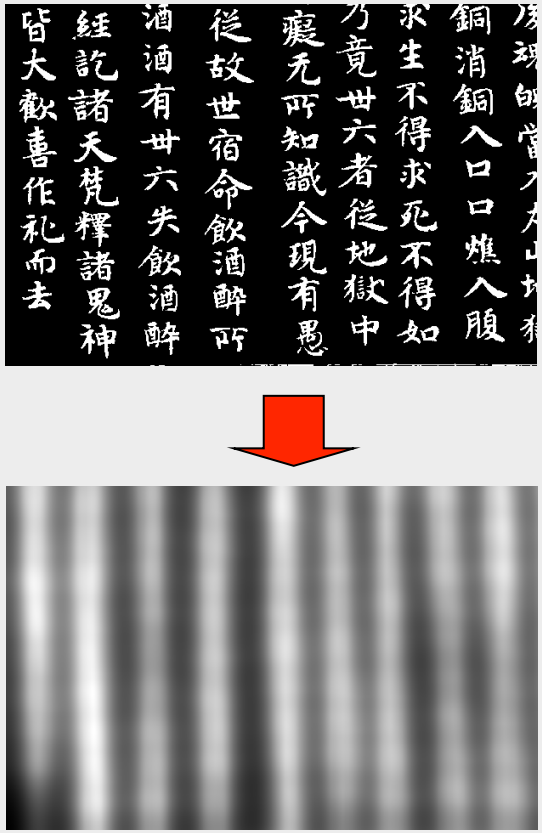
Section of the *Fo shuo Tiwei jing* 佛說提謂經
(British Library Or.8210/S.2051)



Columns and characters automatically isolated (white lines)

Simple Processing Steps

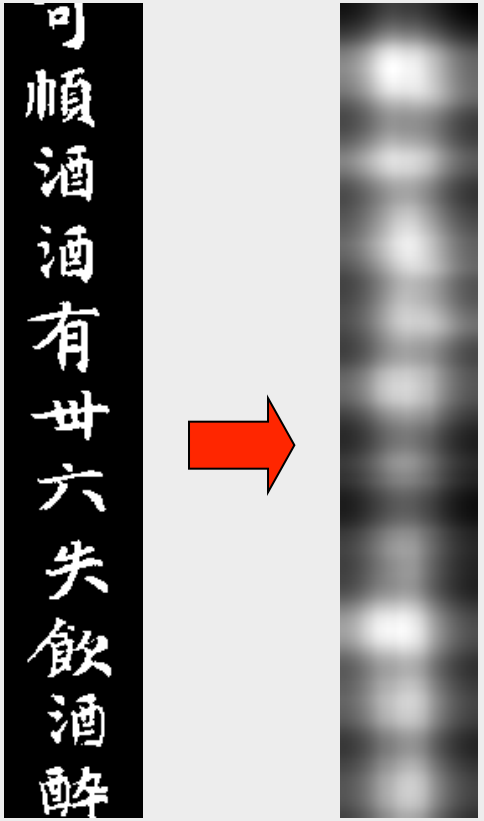
Column structure



horizontal
low-pass
filtering

column boundaries at minima

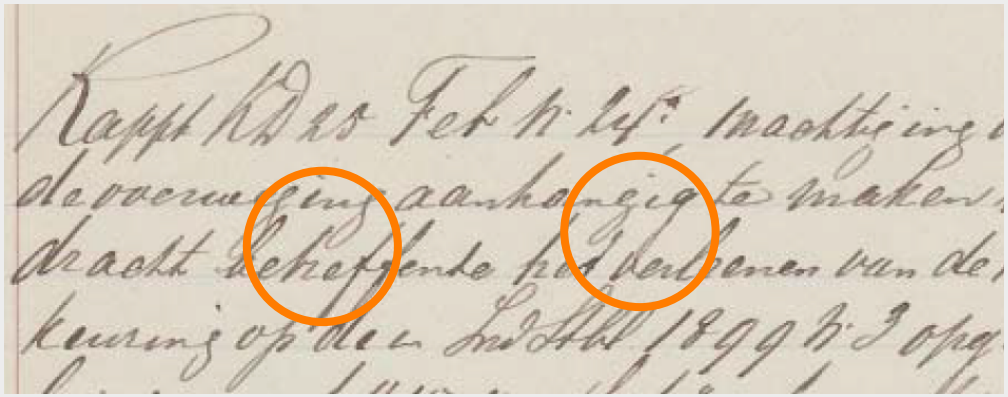
Line structure

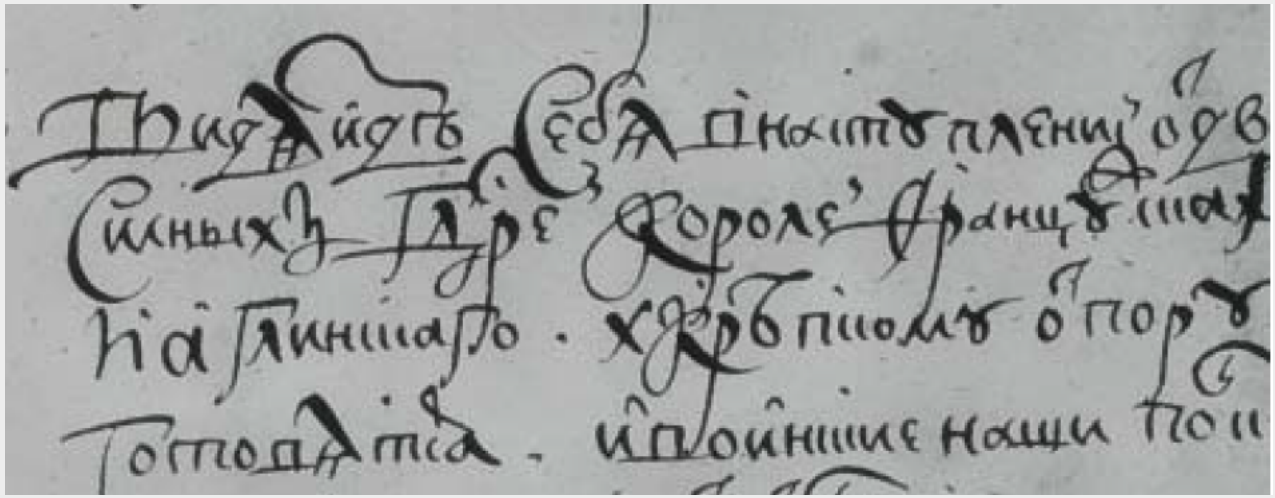


vertical
low-pass
filtering

character boundaries at minima

Localizing Overlapping Text Lines

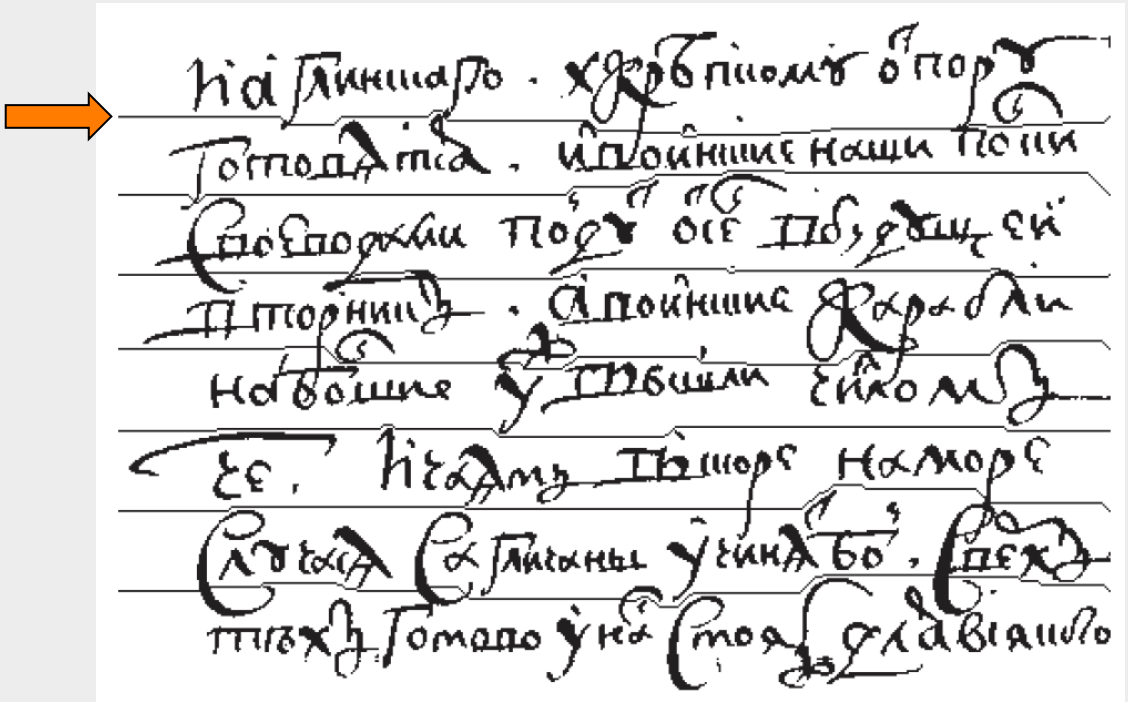




Examples from
Surinta et al. 2014: A* Path Planning for Line Segmentation of Handwritten Documents.

Planning a Path between Lines

A* Algorithm of Artificial Intelligence finds "minimal-cost" paths from initial state (left margin) to final state (right margin).



Judicious cost definition determines performance of algorithm.

Layout Analysis by Gabor Transformation

(Herzog et al. 2014)

- **Location of text blocks in manuscript pages**
- **Line structure**

Main idea:

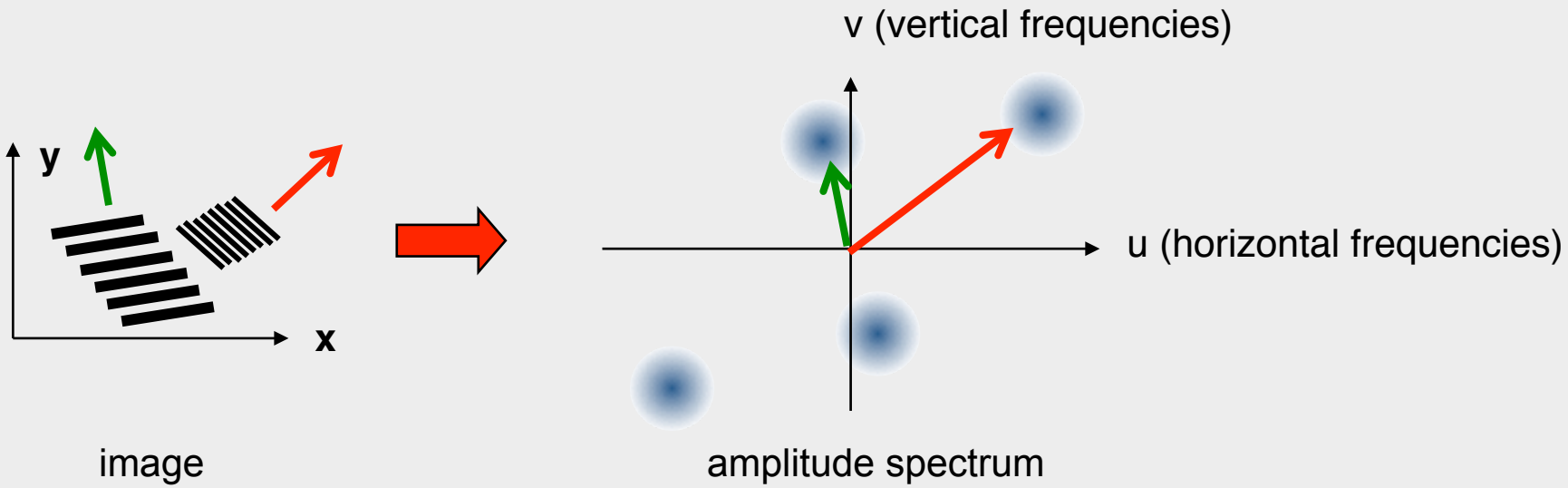
Use local 2D Fourier Transforms (= Gabor Transform) to determine frequency and orientation of text lines.

What is a 2D Fourier Transform?

An image function may be considered a sum of spatial sinusoidal components of different frequencies and directions.

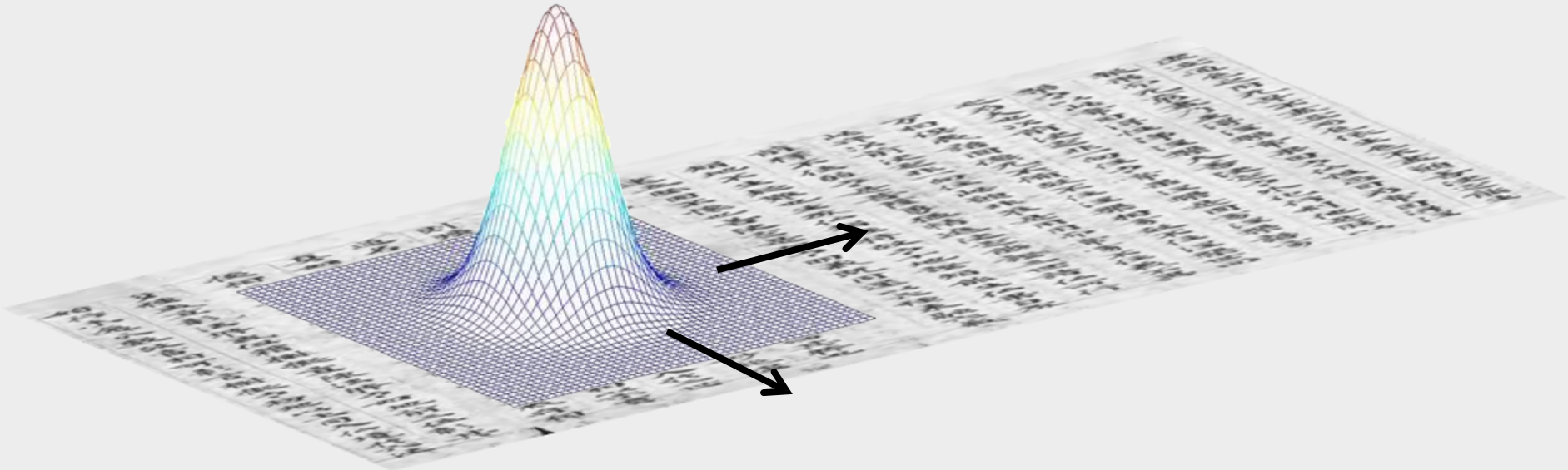
The 2D Fourier Transform computes the *spectrum* of the image function, which indicates the amplitudes, orientations and phases of the spatial sinusoidals contained in an image.

Principle:



What is a Gabor Transform?

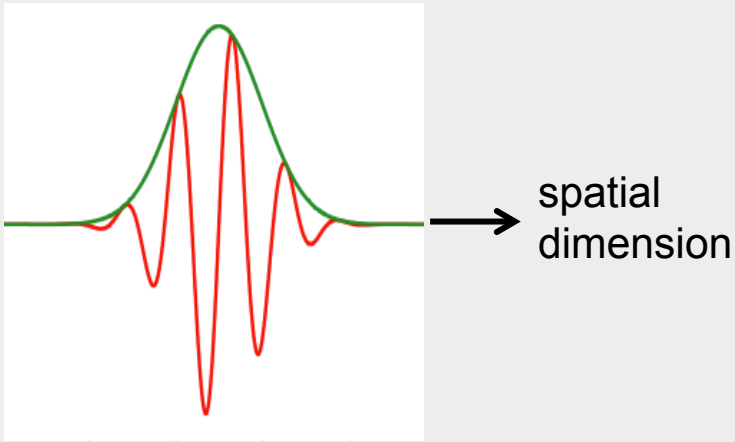
A Gabor Transform is a Fourier Analysis applied to a circular local area, weighted by a Gaussian centered at the circle.

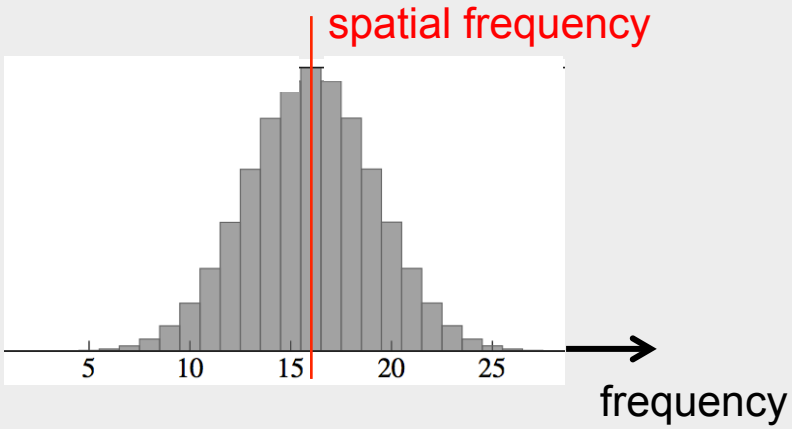


By applying the Gabor Transform at all image locations, the locally dominating line frequencies and orientations can be determined.

Uncertainty Principle

- **Narrow Gabor windows produce spatially fine information, but coarse frequency resolution.**
- **Wide Gabor windows produce spatially coarse information, but fine frequency resolution.**
- **A Gaussian-shaped window offers the best compromise between spatial and frequency resolution.**





Spatial window causes smeared frequency response

Processing Steps

A Determine dominating frequency and orientation ("line signature") at each image location.

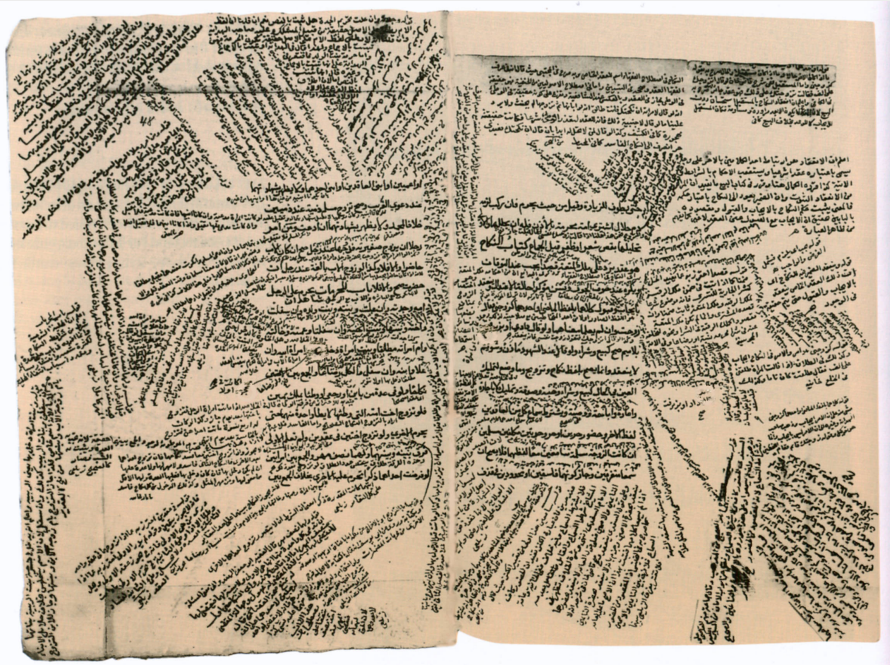
B Determine local inhomogeneity by computing differences (gradient magnitudes) between adjacent line signatures.

C Segment image into text blocks along inhomogeneity maxima.

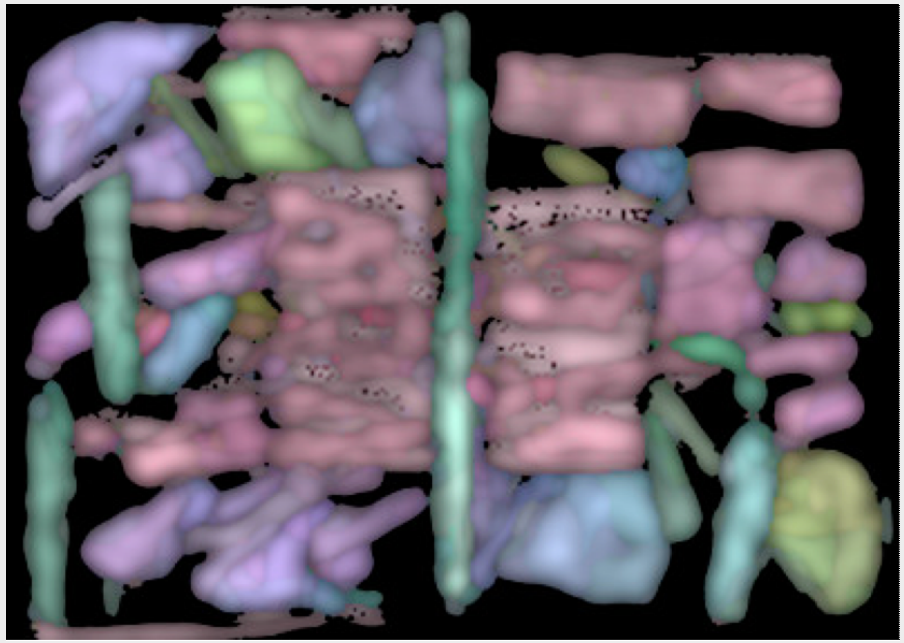
If line distances are unknown or vary strongly, step A must be carried out with several sizes of the Gabor window.

Typical processing time for 2000 x 3000 manuscript page:
10min (in research infrastructure)

Determining Local Line Signatures



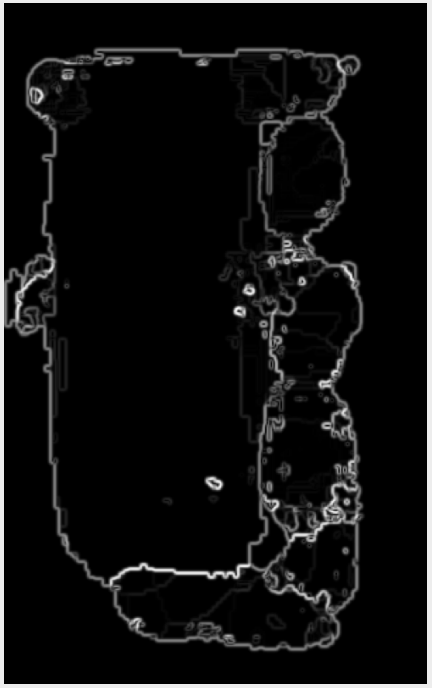
Section of an Arabic manuscript



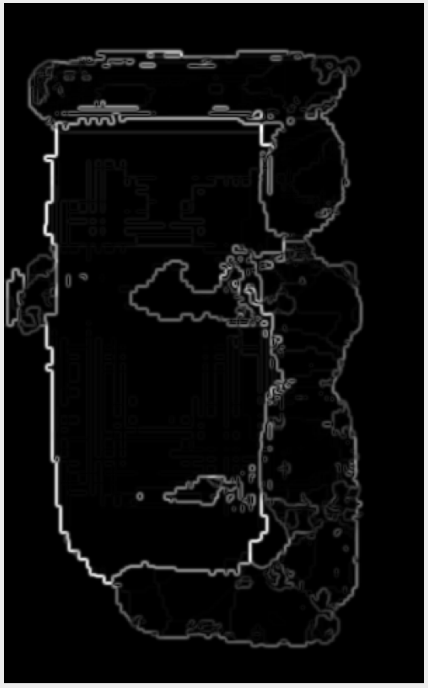
Colour code of line orientations

Note: Orientations ± 180° are not distinguished

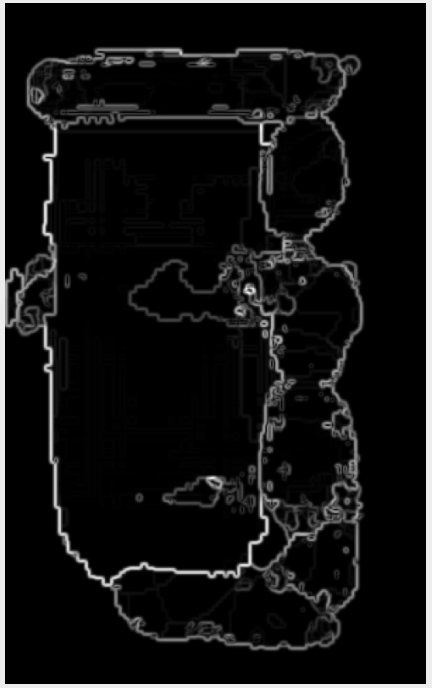
Inhomogeneity Gradients



Orientation gradient magnitudes



Frequency (line distance) gradient magnitudes



Combined inhomogeneity gradient magnitudes

Region Properties

Region boundaries can be derived from the inhomogeneity gradient image with segmentation methods from Computer Vision.

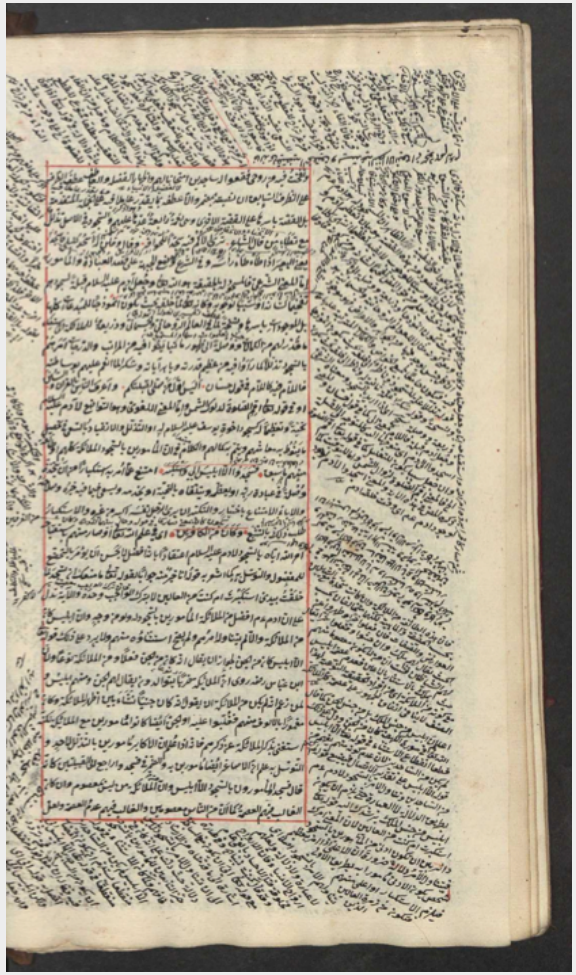
Used here: Watershed segmentation

Several region properties are available for further analysis, e.g.:

	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

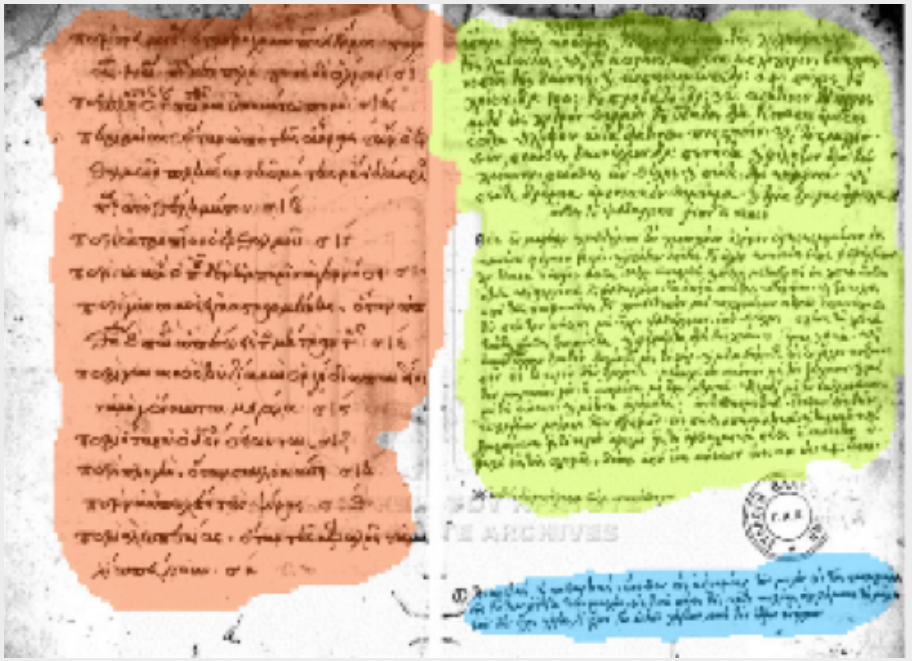
Examples (1)

Examples (2)



Examples (3)

Examples (4)



Examples (5)

Summary of Gabor Transform Layout Analysis

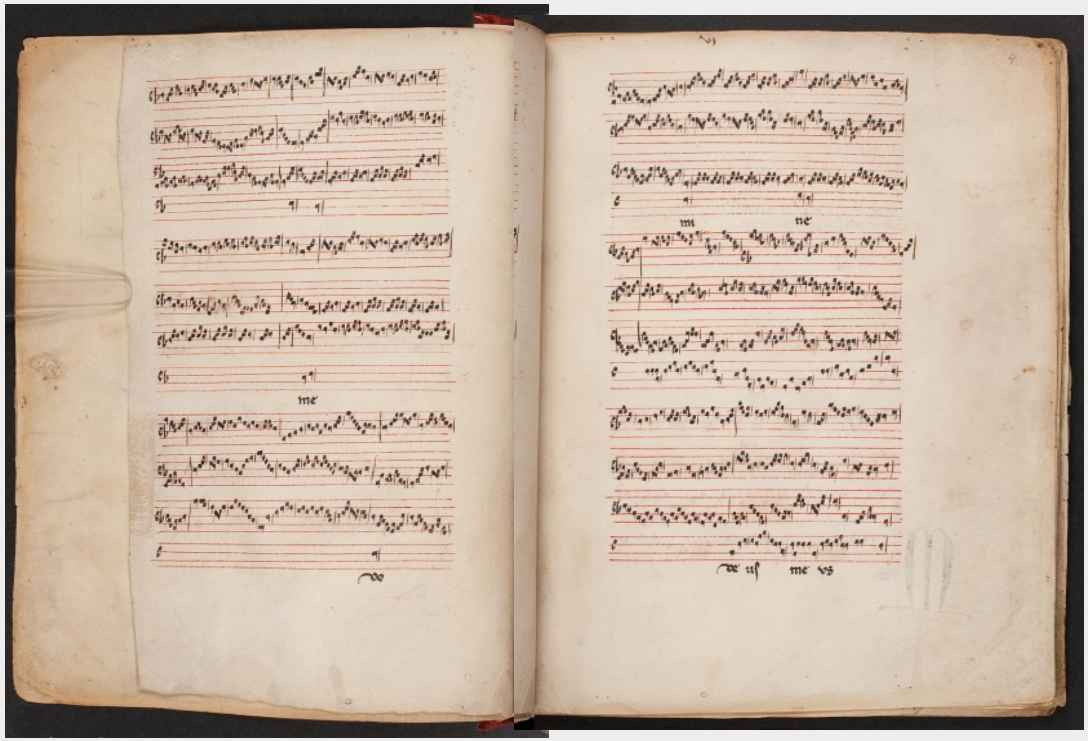
- Text block locations, line distances and orientations can be determined irrespectively of their spatial arrangement.
- No prior binarization required
- Not restricted to any particular kind of handwriting
- Robust against moderate amounts of noise
- Robust against slight irregularities
- Results can be improved further by region analysis methods in the spatial domain.

Limitations and Problem Areas

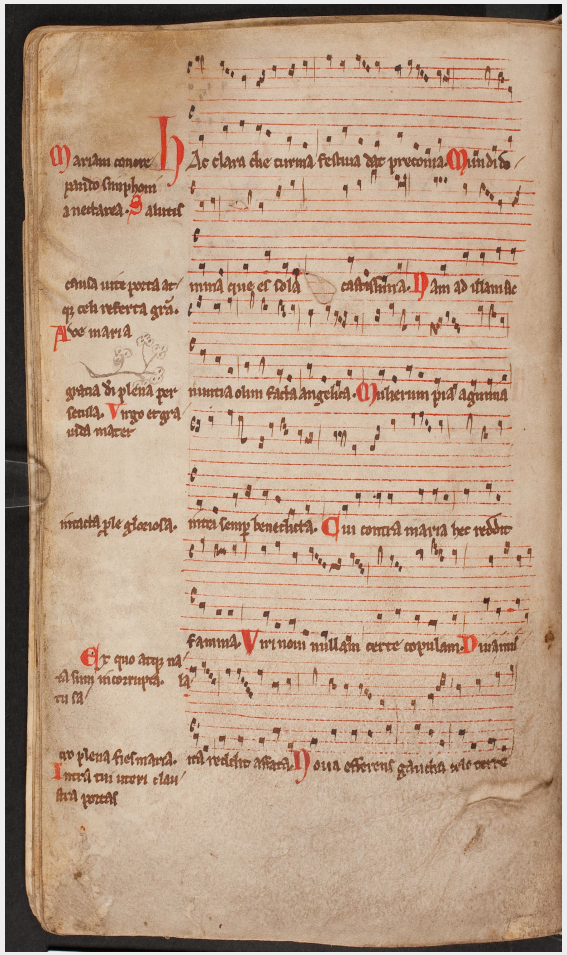
- Uncertain results at boundaries and corners
- Touching text blocks of similar characteristics
- Text blocks with varying characteristics, curved line orientation
- Interlinear text
- Too small text blocks
- Intraline periodicity
- Too much background noise

Pattern Recognition for Medieval Music Notation

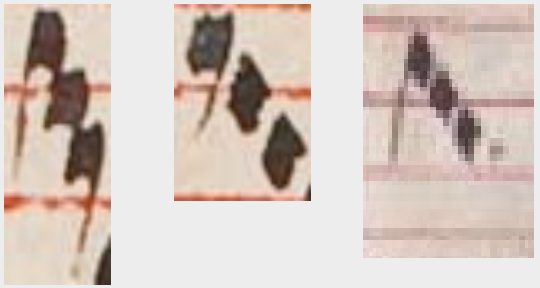
(Solth 2014)



Interesting Patterns of Medieval Music Notation



What is the significance of square and rhomb note ligatures?



Is there any meaning to different stem lengths?

Large data volume of > 1200 pages must be analyzed!

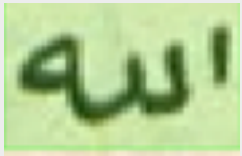
Challenging Pattern Recognition Problem

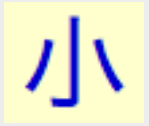
ligatures of same type have significantly different appearances

Search for *compositional structures*: patterns of parts related to each other by certain constraints.

Many examples of compositional structures in manuscripts:

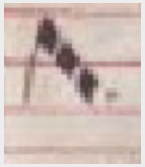
Diacritics 

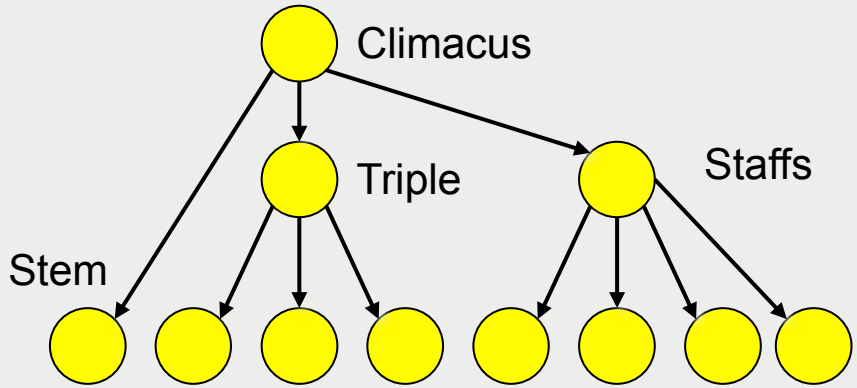
Words 

Stroke radicals 

How does one recognize a compositional structure in a standardized way?

Models for Compositional Structures

A compositional structure  **can be visualized as a graph:**



Nodes represent image parts or aggregates

Edges represent the relation "composed-of" (or "has-part")

Each aggregate node is described by

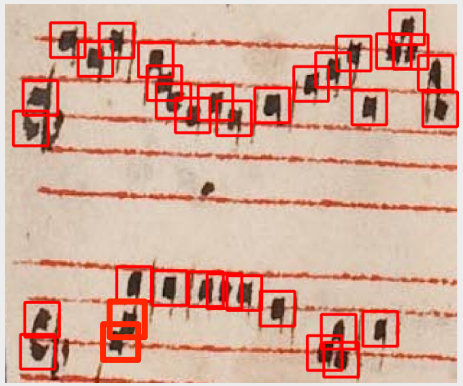
- **aggregate name**
- **parent concepts**
- **aggregate properties**
- **parts**
- **constraints between parts**

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

Recognition of Primitive Parts

"Primitive parts" = elements of a compositional structure which cannot be decomposed further

Here: staff lines, note heads, stems



Standard recognition procedures:

- **Template matching**
- **Normalized Cross Correlation**
- **Feature-based Classification**
- **Specialized methods**

Here: Distinguishing between

- square notes
- rhomb notes
- noise

by template matching.

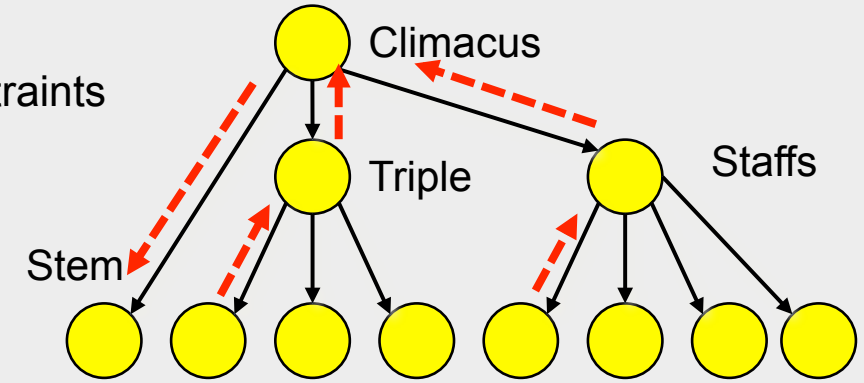
Recognition of Compositional Structures

There exist several standardized algorithms:

- **Constraint Satisfaction**
- **Top-down Search**
- **Bottom-up Search**
- **Mixed Bottom-up Top-down Search**

We used Mixed Bottom-up Top-down Search for ligature recognition:

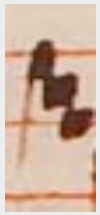
- Find staff lines
- Find staff system, check staff system constraints
- Find note heads
- Find triples, check triple constraints
- Find stem
- Find ligature, check stem constraint

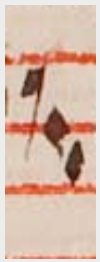


Standard bottom-up steps for aggregate nodes:

- Check constraints on parts
- Compute aggregate properties
- Assign aggregate to parent aggregates

Result Statistics





Manuscript	W1	W2	F
# Pages	379	508	798
# Ligature Type 1	2703	1089	3891
# Ligature Type 2	2	4	6
# Ligature Type 3	1398	549	1364
Ratio Type 1:Type 3	~2:1	~2:1	~3:1
Accuracy Note Classification	89,41%	96,44%	79,11%

W1 = Scottish Manuscript 1 (Herzog August Bibliothek Wolfenbüttel)
W2 = Scottish Manuscript 2 (Herzog August Bibliothek Wolfenbüttel)
F = French Manuscript (Notre Dame)

Experiences with Cooperations

- Interesting for a Computer Scientist to get a glimpse at research topics of the Humanities!
- Precise definitions of required services had to be worked out.
- Target data (e.g. for comparing writing patterns) have often been too sparse for a statistical analysis.
- Development of problem-oriented programs took more time than anticipated.
- Mismatch of language and perception between Computer scientists and humanity scholars.
- Broad support independent of writing systems may be less effective than specialized methods.

Some Useful Links

Our publications:

http://kogs-www.informatik.uni-hamburg.de

Centre for the Study of Manuscript Cultures (CSMC):

http://www.manuscript-cultures.uni-hamburg.de

Collection of generic vision algorithms (VIGRA):

http://ukoethe.github.io/vigra

GeoMap for support of segmentation

https://github.com/hmeine/geomap

Delaunay Triangulation

http://www.cs.cmu.edu/~quake/triangle.html

Our emails:

{herzog, solth, neumann}@informatik.uni-hamburg.de


Thank you for your interest!