university of
groningen
1614

Advances in writer identification and verification

Lambert Schomaker

icdar-lect-v8-ooffice.ppt

L.R.B. Schomaker (2007) Advances in writer identification and verification [Invited lecture], *Proc. of 9th Int. Conf. on Document Analysis and Recognition (ICDAR 2007)*, IEEE Computer Society, pp. 769-773, vol. II, 23 - 26 September, Curitiba, Brazil.

Researchers

- Marius Bulacu, Axel Brink, Katrin Franke
- Ralph Niels, Louis Vuurpijl
- "The Nijmegen Handwriting Group 1984-1993"
- Netherlands Forensic Institute: Ton Broeders, Wil Fagel, Elisa van den Heuvel
- Isabelle Guyon, Rejean Plamondon
- many others

Overview

1. handwriting-based biometrics
2. the basis of handwriting individuality
3. slant, curvature
4. allography
5. other features
6. outlook

Authentication

- you are x because you **possess** a token: $t = T(x)$
- you are x because you **know** a secret: $p = P(x)$
- you are x because you **behave** as x: $b = B(x)$

- Uniqueness: $B(x) \neq B(y)$ for $x \neq y$
- or: $\forall x\, B(x) \wedge \neg \exists y{\neq}x\, (B(y)=B(x))$

- Permanence: $\forall t_1,t_2\;\; b_{x,t1} = b_{x,t2}$

Authentication

› you are x because you:
- **possess** physiological feature, indissolubly connected to individual x:

- DNA, face, hand, ear, iris, fingerprint

  ➔ General Biometrics!

Behavioral Biometrics

- you are x because you **behave** as x

- speech
- gait
- inter-keystroke delays in typing
- usage of interpunction in text (,.:;)
- word-usage histogram
- **... handwriting biometrics ...**

Behavioral biometrics

Handwriting biometrics

Signature verification
- financial transactions
- legal transactions
- access authorization
- forensics

Writer **1:1** verification
- forensics
- authorization
- humanities

Combined identification & verification
- mail-address scanning for known writers

Writer **1:N** identification
- forensics
- humanities
- writer adaptation in tablet-PCs

Writer identification and verification

1. Interactive, fully manual feature measurement

2. Automatic

3. OCR based

Writer identification and verification

1. Interactive, fully manual

2. Automatic

3. OCR based

1.1 measured Character features

1.2 General geometric script features

Writer identification and verification

1. Interactive, fully manual

2. Automatic

3. OCR based

“Width”

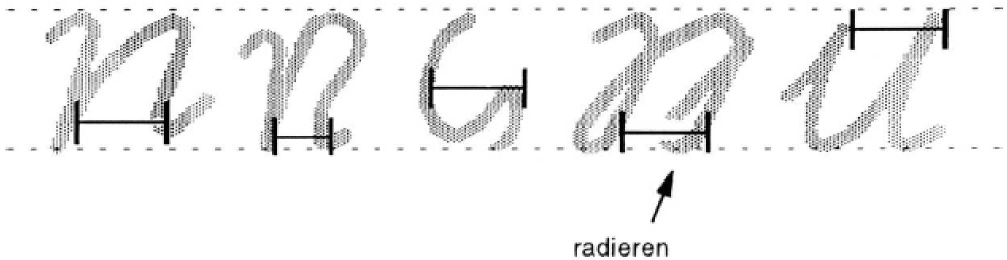


Figure 7.3: Examples of measuring width of handwriting. *Image captured from BKA FISH manual.*

“Corpus”

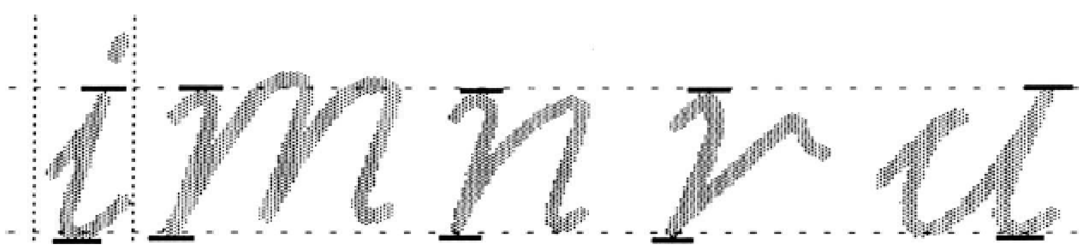


Figure 7.4: Examples of measuring height of handwriting. *Image captured from BKA FISH manual.*

Manual feature measurements in the BKA/Fish system

Writer identification and verification

1. Interactive, fully manual

2. Automatic

3. OCR based

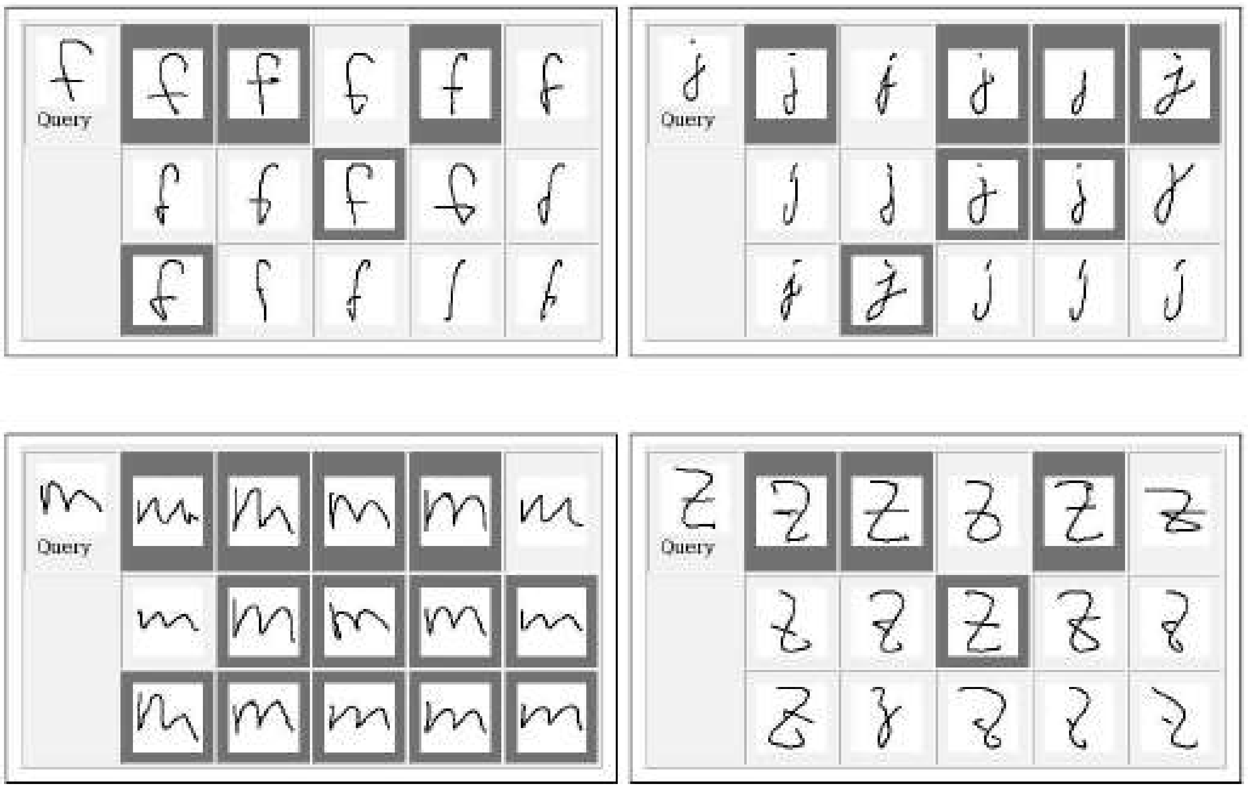
3.1 computed Character features

Writer identification and verification

1. Interactive, fully manual

2. Automatic

3. OCR based



[Niels, Vuurpijl & Schomaker, 2006]

Also cf. Fox system

(Srihari et al. 2006)

"OCR"-based, allographic, semi-automatic

Writer identification and verification

1. Interactive, fully manual
   - 1.1 measured Character features
   - 1.2 General geometric script features

2. Automatic
   - **2.1 Region of Interest (ROI) based**
   - 2.2 Line based HMMs

3. OCR based
   - 3.1 computed Character features

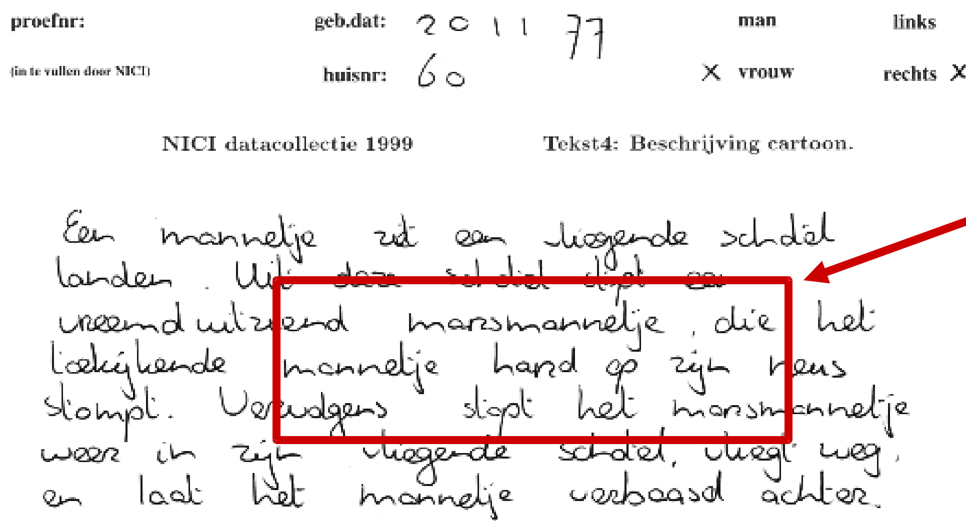
Writer identification and verification

1. Interactive, fully manual

2. Automatic,

ROI based

3. OCR based



Who?

Also cf. Bunke et al:

line HMMs for writers

Writer identification and verification

- 1. Interactive, fully manual
  - 1.1 measured Character features
  - 1.2 General geometric script features
- **2. Automatic**
  - **2.1 Region of Interest (ROI) based**
    - **2.1.1 Textural image features**
    - **2.1.2 Character-fragment features**
    - 2.1.3 Placement statistics
    - 2.1.4 Ink deposition
  - 2.2 Line based HMMs
- 3. OCR based
  - 3.1 computed Character features

Is there a basis for handwriting individuality in behavioral biometrics?

- Handwriting of different persons looks different ...
- But handwriting is learned - thus can be forged -
- Is it stable enough?

- What are the most reliable features?
- How much data are needed in a sample?

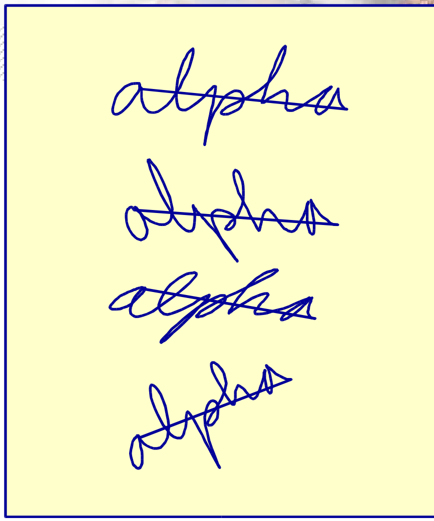
Factors in individuality

- between-writer variation
- within-writer variability

- nature-nurture (genetics-memetics)

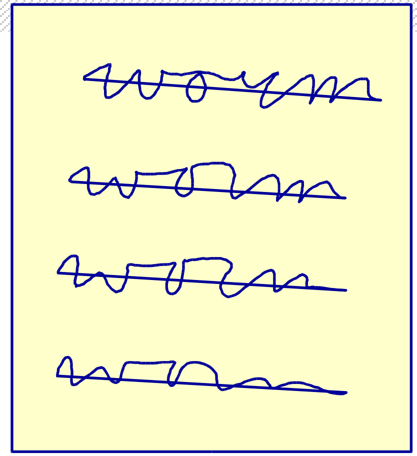
- Four factors in kinematics & 2D
 (that produce problems in the recognition of HWR)

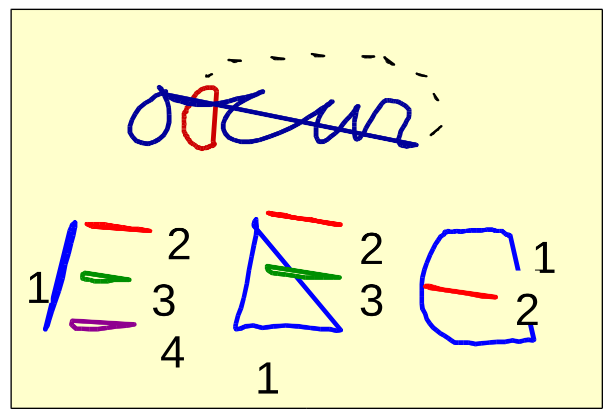
Affine transforms



Allographic variation

Neuro-biomechanical variability



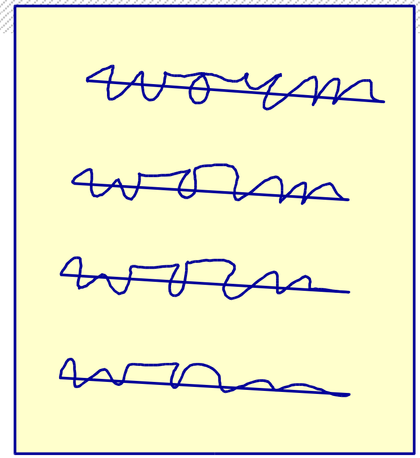


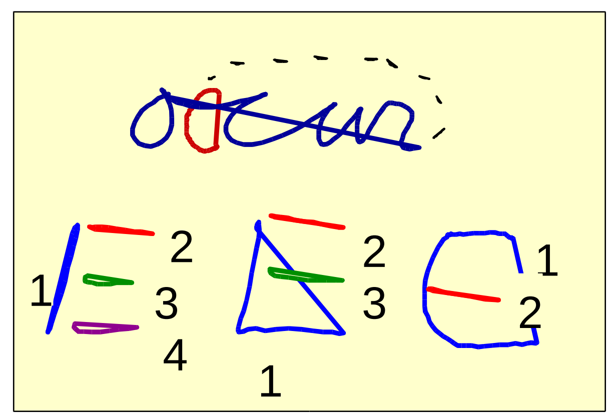
Sequencing variability

System State

- fatigue
- psychofarmaca, alcohol, coffee
- distractions
    - multitasking
    - acoustic noise
- stress

Neuro-biomechanical variability

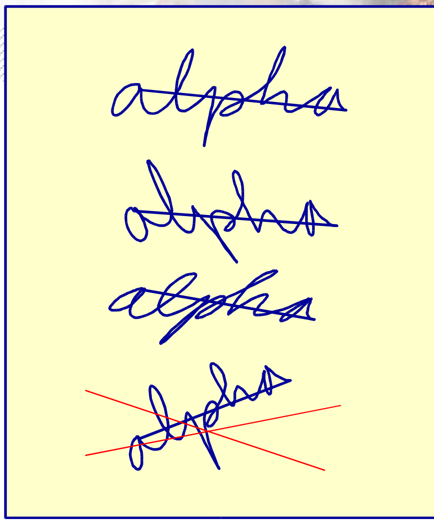




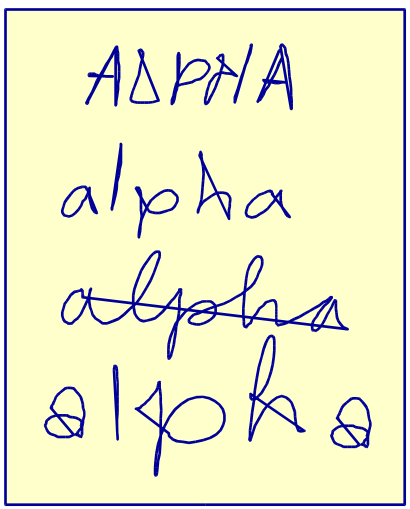
Sequencing variability

Affine transforms

System Identity



Part 1) The habitual pen-grip and biomechanical parameters determine:
size, slant, curvature,
[pen force]

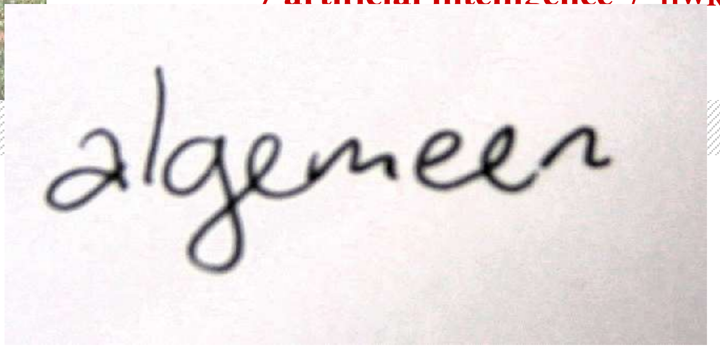


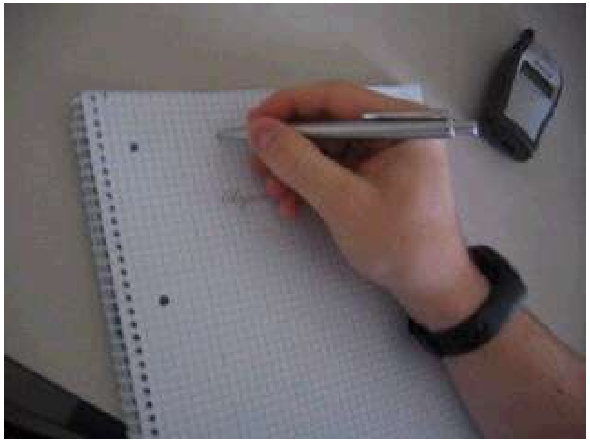
Part 2) The school system and personal preferences determine allography

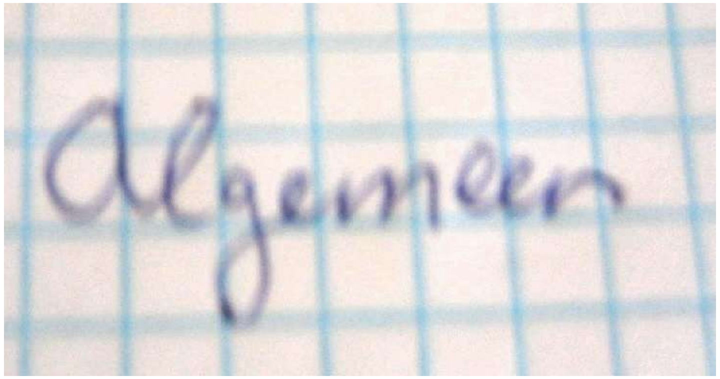
Allographic variation

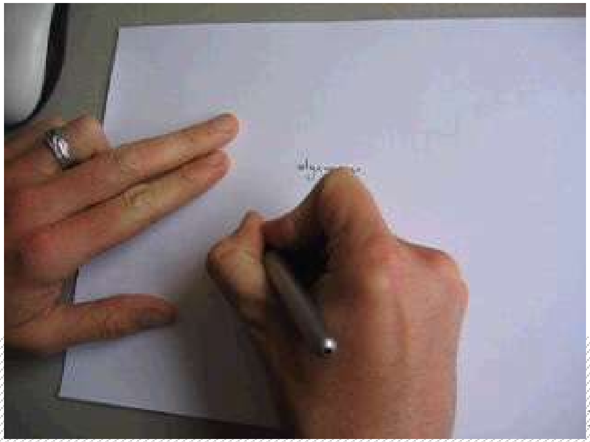
Slant and curvature

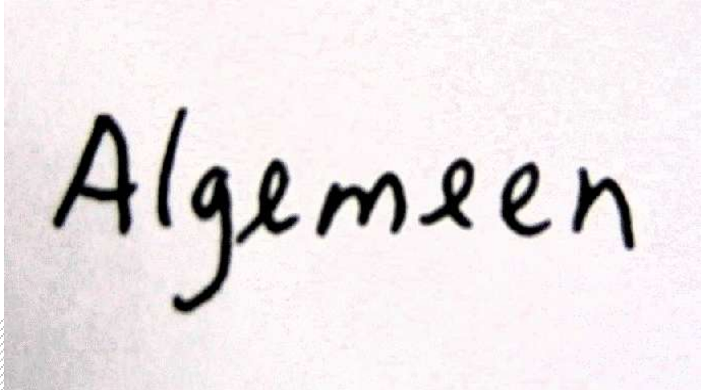
- are under voluntary control by the brain
- but: the learned pen grip provides a
  habitual frame of reference,
  limiting the variation in shape for a given writer
- [videos]

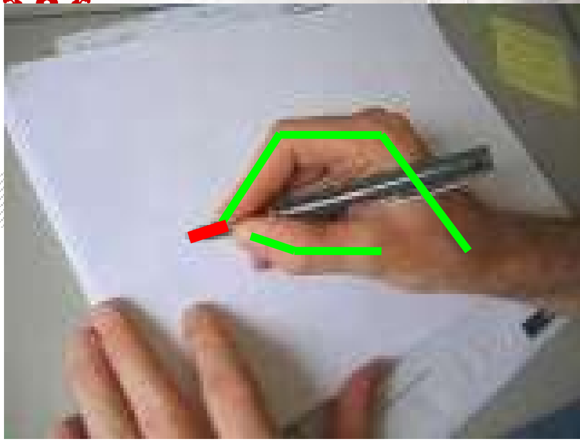

algemeen

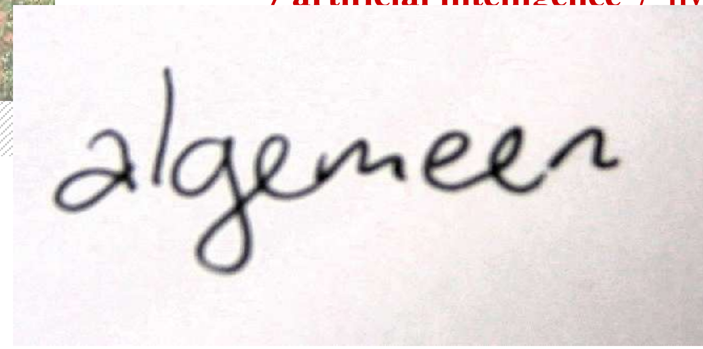


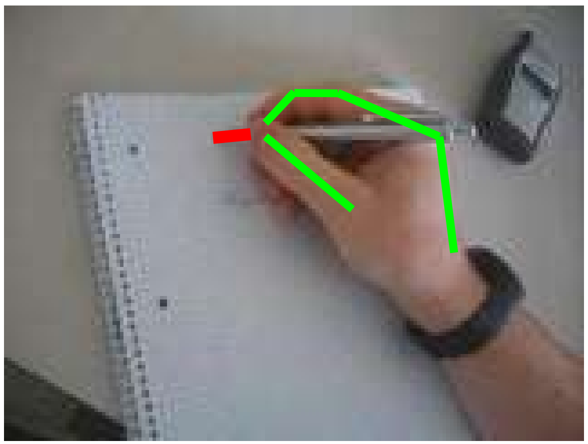

algemeen

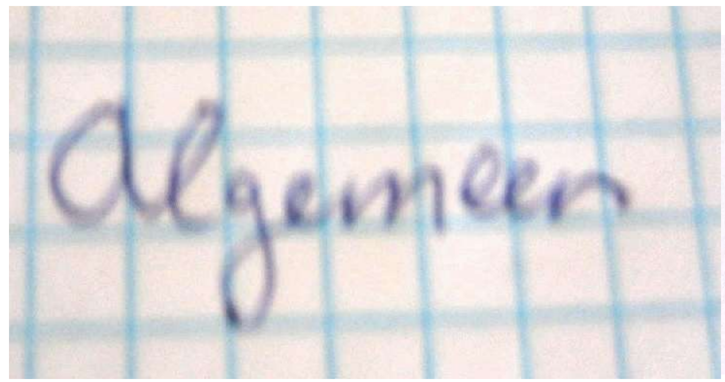


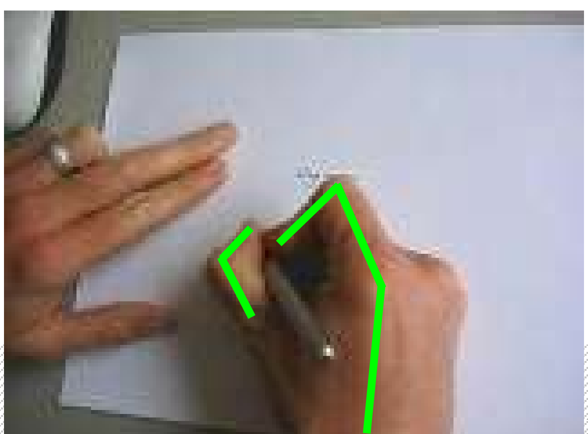

Algemeen

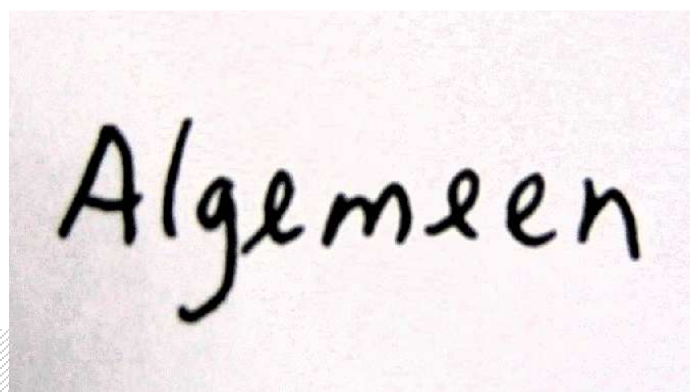


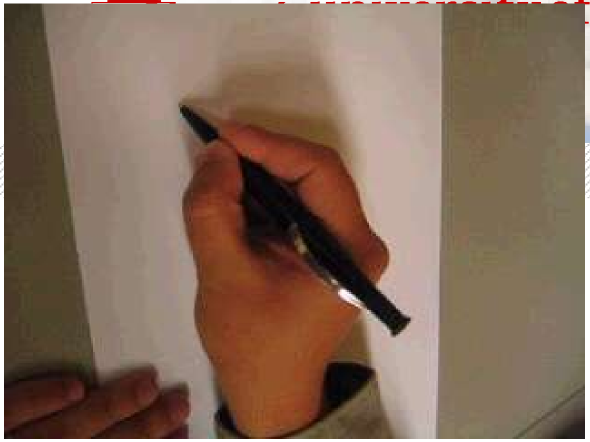

algemeen

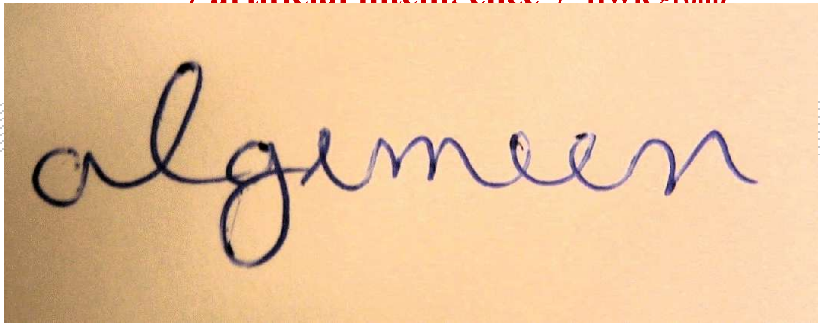


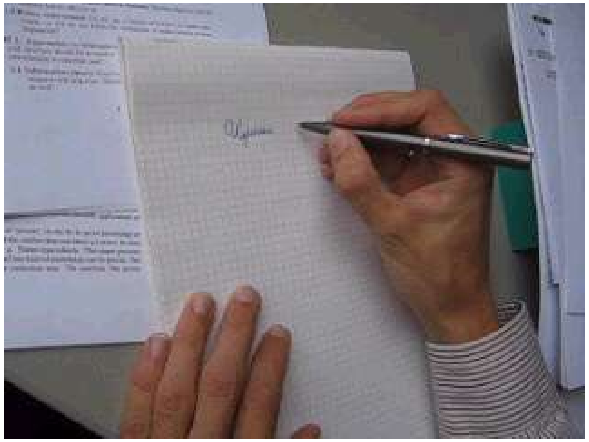

algemeen

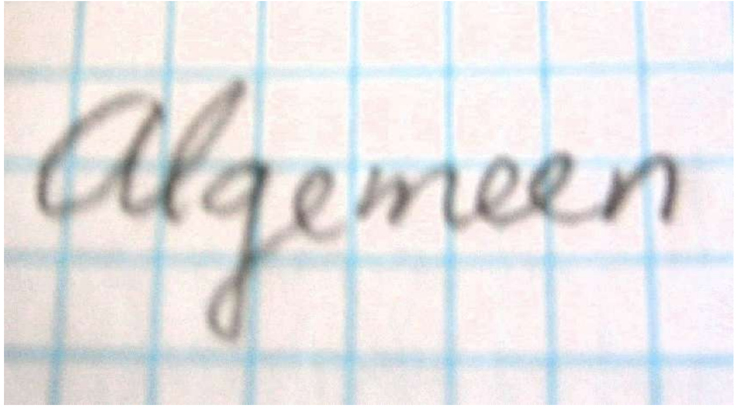


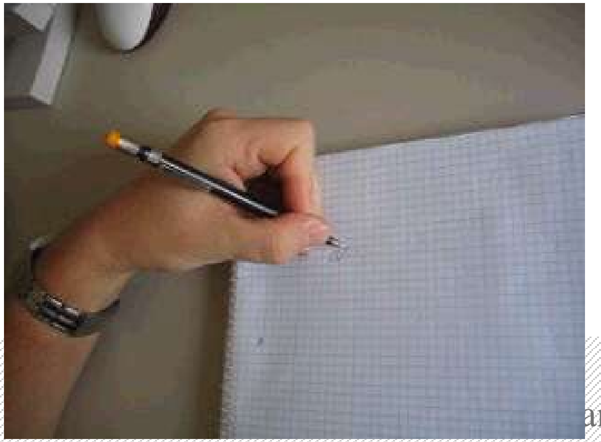

Algemeen

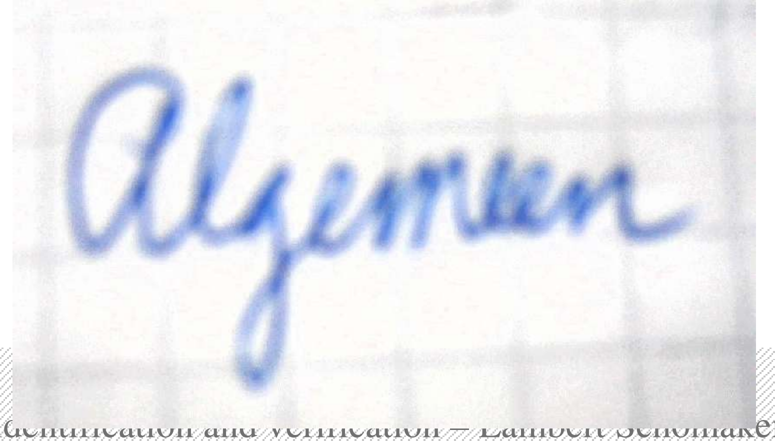


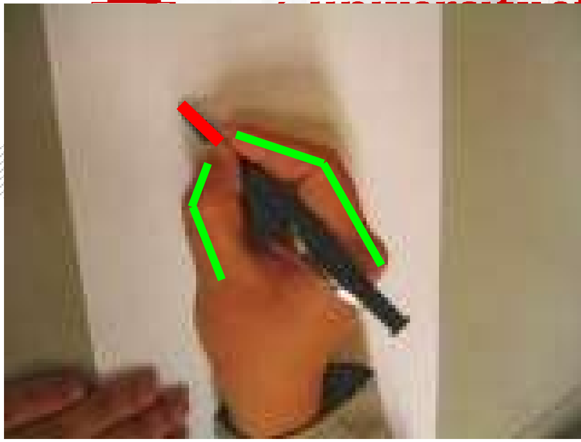

algemeen

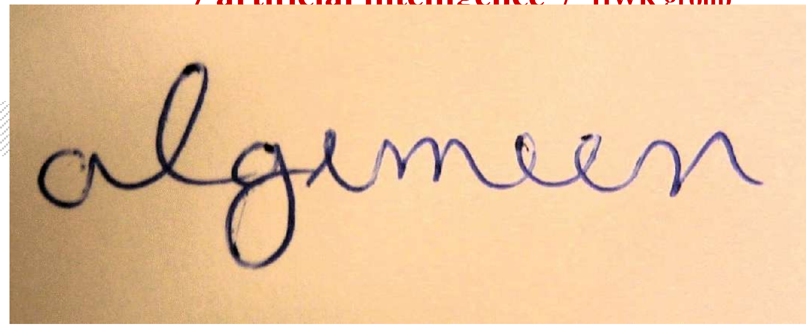


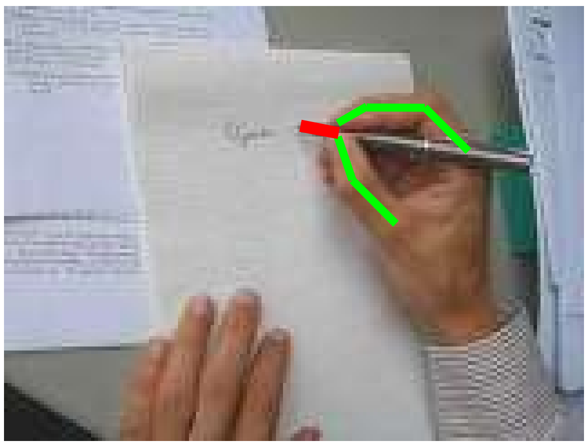

algemeen

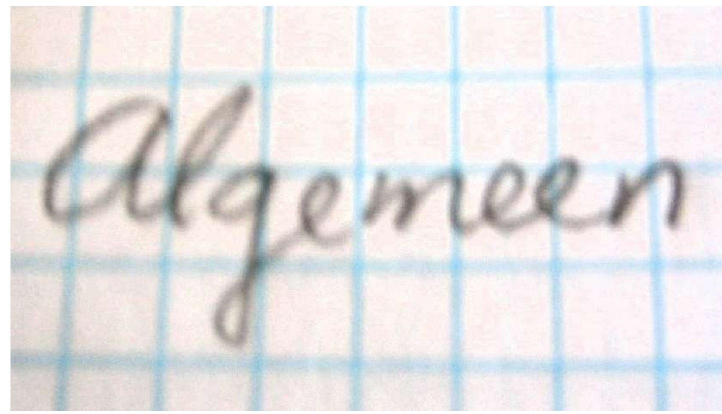


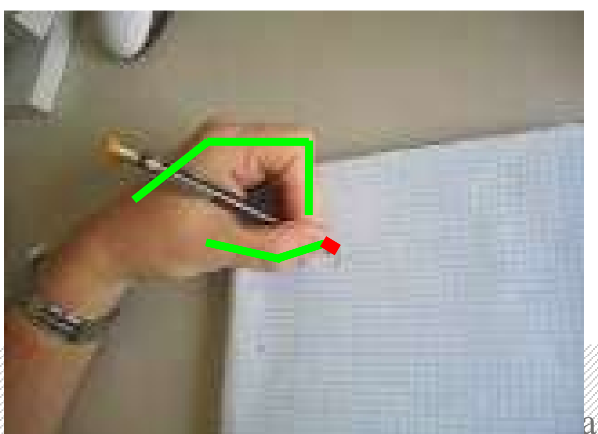

algemeen

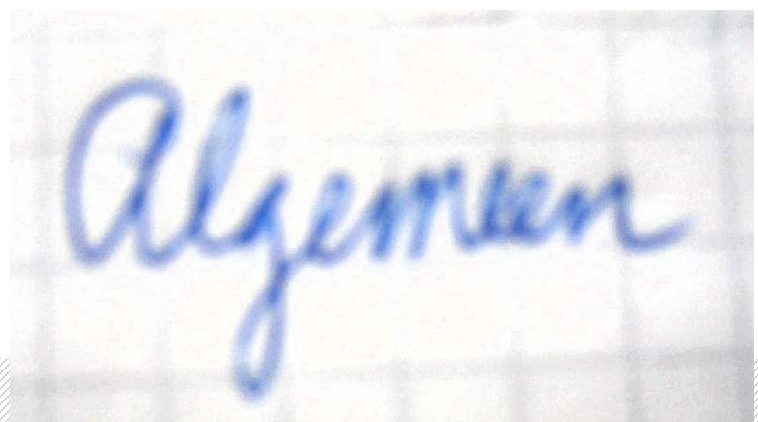



algemeen




algemeen




algemeen

Hypothesis

- if hand biomechanics has a genetic basis,
  i.e., is individual;
- if handwriting shapes are constrained
  by hand biomechanics;
- then basic ink-trace characteristics
  such as slant and curvature
  may be good features in handwriting biometrics

Hand biomechanics & individuality

A.E. Park, J.J. Fernandez, K. Schmedders and M.S. Cohen (2003)
The Fibonacci Sequence: Relationship to the Human Hand
Journal of Hand Surgery, 28, pp. 157-160.

- Park et al (2003) set out to test the urban legend that finger bone lengths obey the Fibonacci sequence

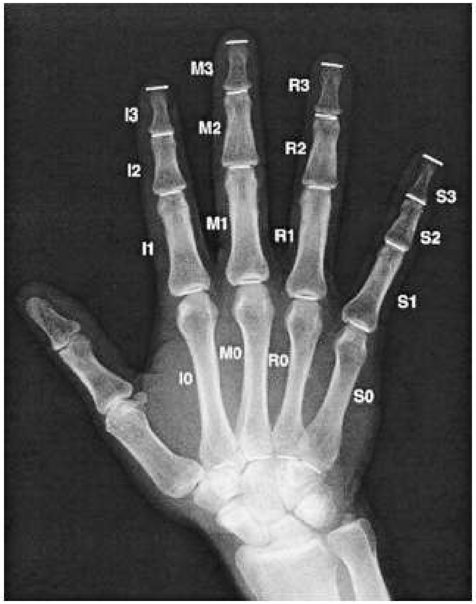


Figure 2. Posteroanterior radiograph of a hand. Bone lengths identified by the name of the field used for statistical analysis. Each data field is composed of a letter to indicate the digit and a number to indicate the bone: I, index finger; M, middle finger; R, ring finger; S, small finger; 0, metacarpal; 1, proximal phalanx; 2, middle phalanx; 3, distal phalanx.

Carpal and phalangeal bone-length ratios

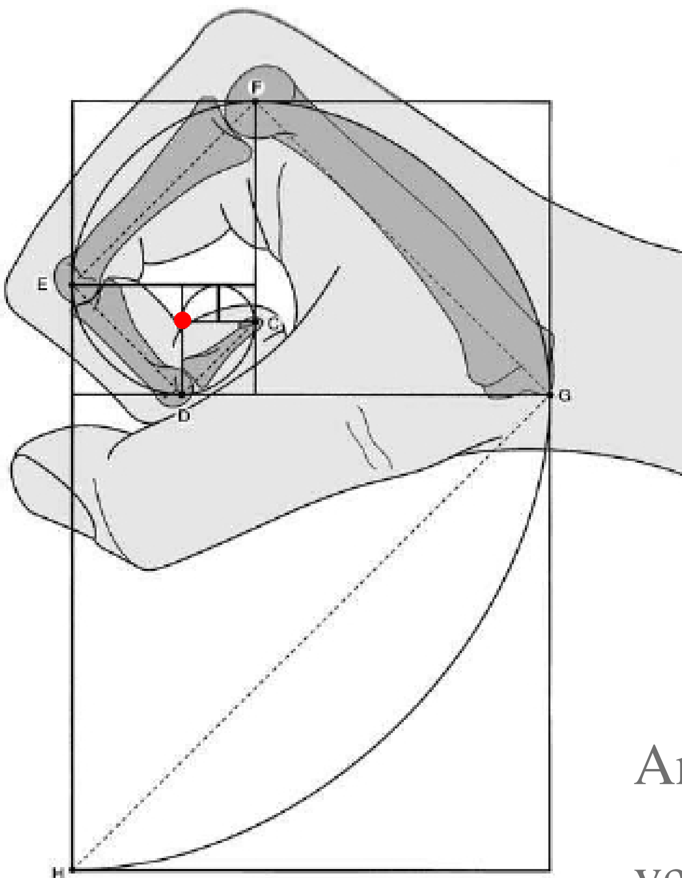


Figure 3. Human hand superimposed on the Fibonacci rectangles and equiangular spiral of Figure 1. This shows the proposed relationship of the Fibonacci sequence to the center of rotation of the joints of the hand. The connection of points

Fibonacci: 0,1,1,2,3,5,8,...

$$r = ae^{b\theta}$$

$$\theta = \frac{1}{b}\ln(r/a),$$

$$\arccos\frac{\langle \mathbf{r}(\theta), \mathbf{r}'(\theta)\rangle}{\|\mathbf{r}(\theta)\|\|\mathbf{r}'(\theta)\|} = \arctan\frac{1}{b} = \phi,$$

Angle φ between tangent and radial vector should be constant (1.618)

Results

- 100 persons, X-rays of hand
- **Park et al. (2003)** measured ratios of all consecutive pairs of bones for each finger
- Fibonacci-compatible value for φ **only** found for **one finger, one pair of bones** (little finger, metacarpal and proximal phalanx lengths)
- Individuals are characterized by their own typical bone-length ratios!

➔ great for biometrics!

Nature/nurture

- Twin studies (Srihari et al, 2007) show that error rate ε in writer verification is ordered:

ε(identical twins) >
ε(fraternal twins) >
ε(arbitrary person pairs)

Part I textural features, rationale

1. brain
2. biomechanics (muscles/bones)
   minimum jerk principle [Flash & Hogan, 1985]
   minimum energy principle [Alexander, 1997]
3. preferred habitual pen grip
4. **slant** and **curvature** distributions

Seminal work on slant

› Dooijes (1984) was the first to use polar histograms of angles in on-line handwriting

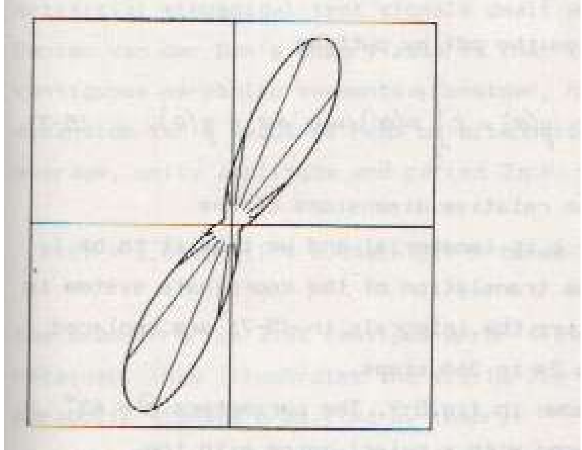


Fig.5-8 (left): distribution of tangent

Seminal work on slant

- Dooijes (1984) made an attempt at decomposing the Cartesian recording of a digitizer tablet in to an oblique system of 2 axes, representing
  the wrist and finger movements:

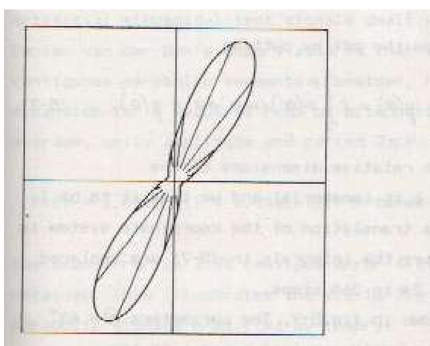
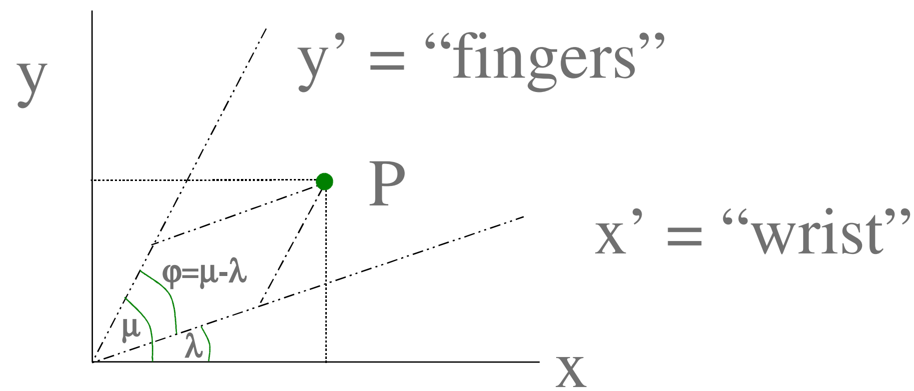
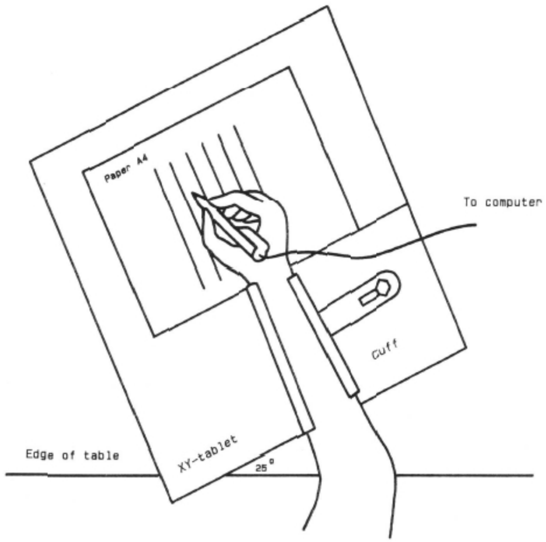


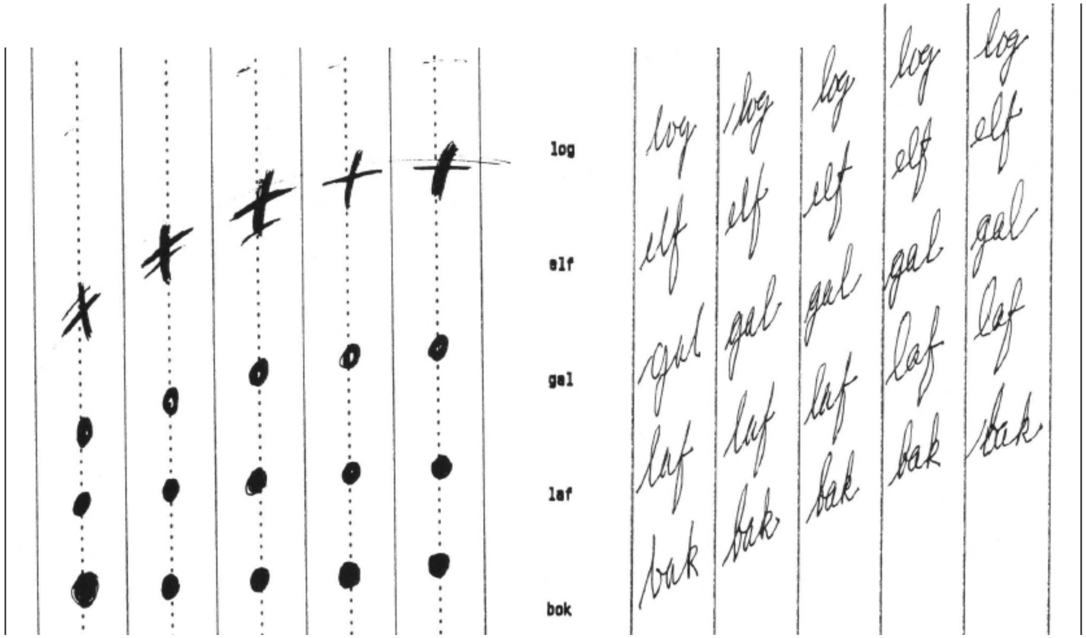
Fig.5-8 (left): distribution of tangent



Slant, continued

- Maarse (1987) investigated constancy of slant under varying wrist orientations, revealing remarkable stability of slant angle, with rotating (x’,y’)





Maarse (1987) and sequel

- The uninstructed writer has a strong tendency to maintain slant, even at the cost of some discomfort in the pen grip
- **Maarse, Schomaker & Teulings (1988).** With 20 writers, and single-line text samples, an identification performance of 79 in 80 (99%) was reached on LDA with five *on-line* features:

1) axial pen force,
2) **slant angle from polar plot,**
3) average velocity while inking,
4) width of rightward strokes,
5) relative pendown-time vs total writing time.

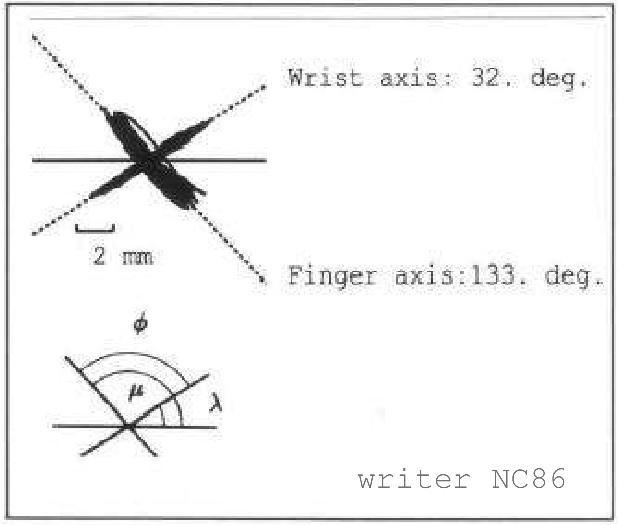
Slant summary

› slant is determined by the biomechanics of the wrist/finger system in interaction with the chosen or trained pen grip

moving fingers, fixed wrist

are more noisy than

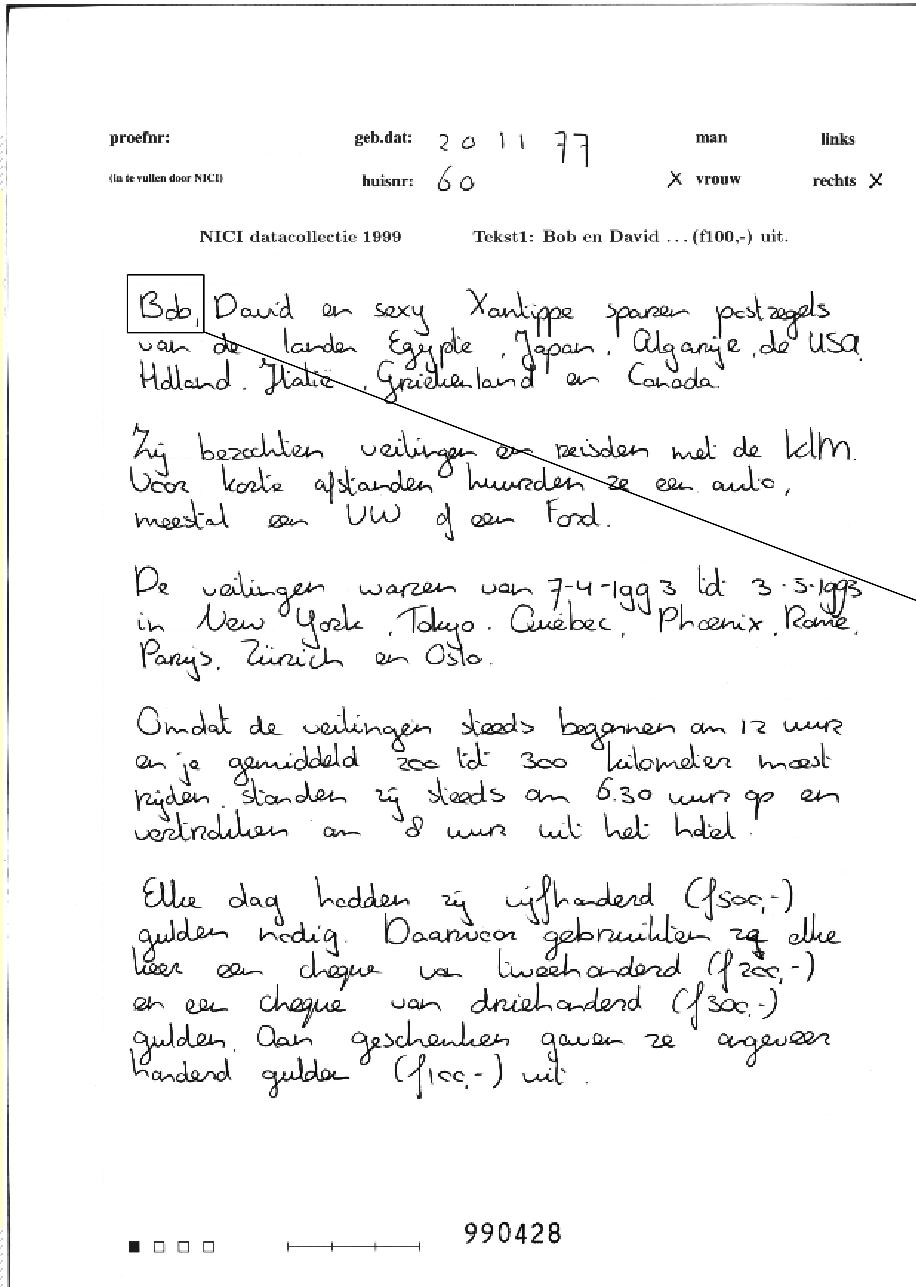
moving wrist, fixed fingers



11 yrs silence, then renewed interest

- NFI: for the Netherlands Forensics Institute, clean handwriting samples, 250 writers, four pages per writer, were collected: **Firemaker** collection [Schomaker & Vuurpijl, 1999]

- Wanda project for BKA: framework for forensic writer identification [Franke,Schomaker,Vuurpijl,Guyon, 2002]


proefnr: geb.dat: 20 11 77 man links

(in te vullen door NICI) huisnr: 60 X vrouw rechts X

NICI datacollectie 1999 Tekst1: Bob en David . . . (f100,-) uit.

Bob, David en sexy Xantippe sparen postzegels van de landen Egypte, Japan, Algarije, de USA Holland. Italie, Griekenland en Canada.

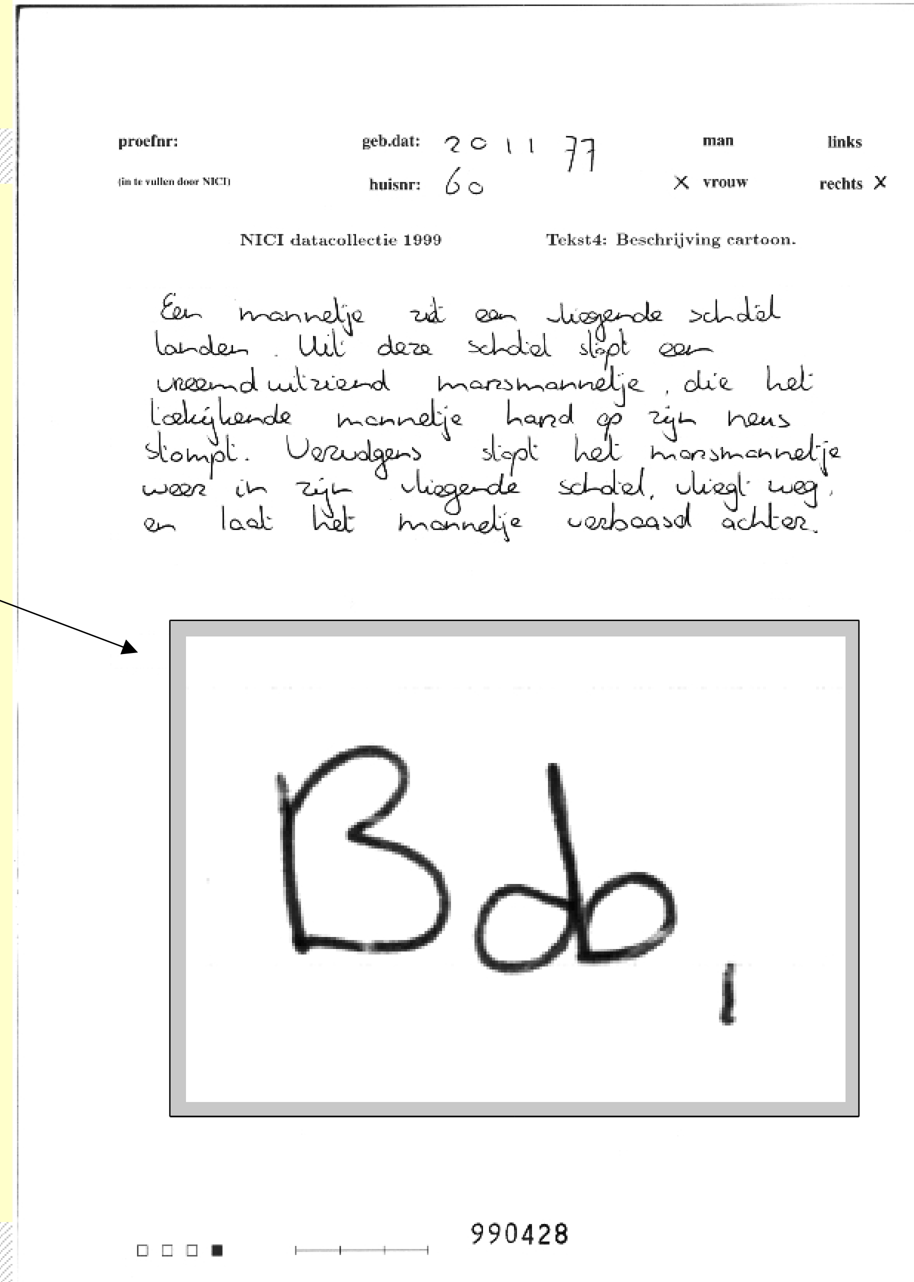
Zij bezochten veilingen en reisden met de KLM. Voor korte afstanden huurden ze een auto, meestal een VW of een Ford.

De veilingen waren van 7-4-1993 tot 3-5-1993 in New York, Tokyo. Québec, Phoenix, Rome, Parijs, Zürich en Oslo.

Omdat de veilingen steeds begonnen om 12 uur en je gemiddeld 200 tot 300 kilometer moest rijden, stonden zij steeds om 6.30 uur op en vertrokken om 8 uur uit het hotel.

Elke dag hadden zij vijfhonderd (f500,-) gulden nodig. Daarvoor gebruikten ze elke keer een cheque van tweehonderd (f200,-) en een cheque van driehonderd (f300,-) gulden. Aan geschenken gaven ze ongeveer honderd gulden (f100,-) uit.

990428


proefnr: geb.dat: 20 11 77 man links

(in te vullen door NICI) huisnr: 60 X vrouw rechts X

NICI datacollectie 1999 Tekst4: Beschrijving cartoon.

Een mannetje ziet een vliegende schotel landen. Uit deze schotel stapt een vreemd uitziend marsmannetje, die het lachende mannetje hard op zijn neus stompt. Vervolgens stapt het marsmannetje weer in zijn vliegende schotel, vliegt weg, en laat het mannetje verbaasd achter.

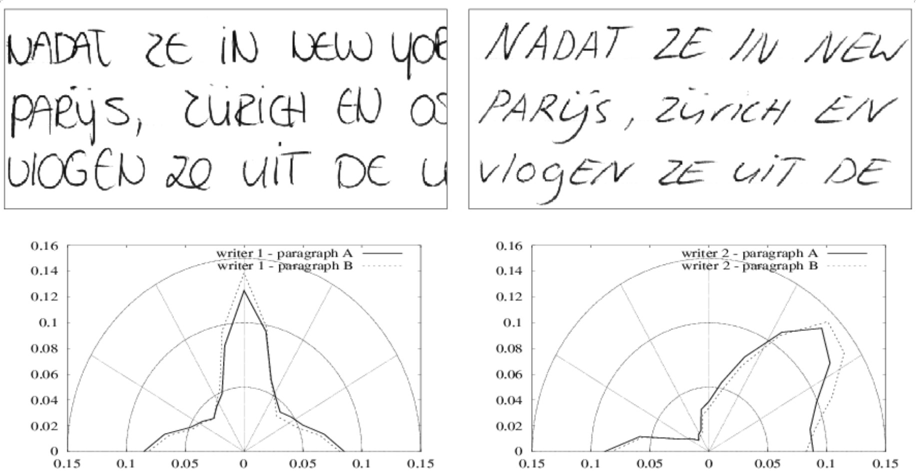
Bob,

990428

p(φ) edge-angles histogram

[Bulacu & Schomaker, CAIP 2003]

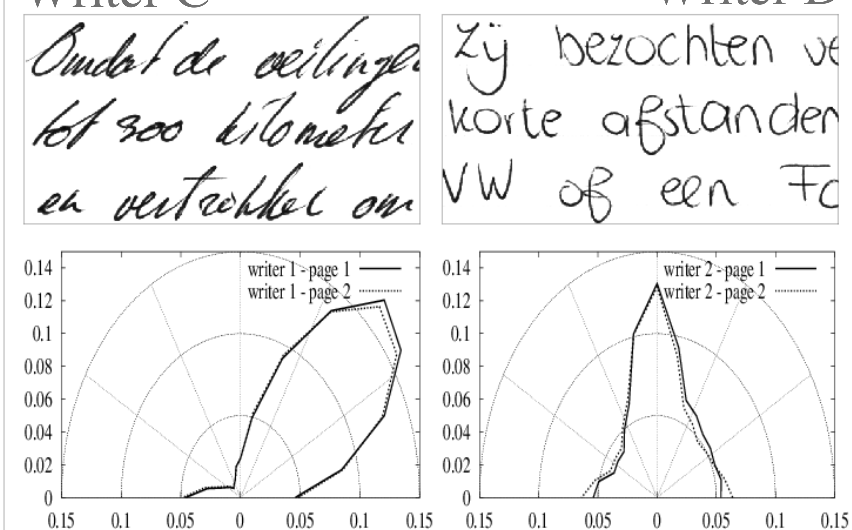
Writer A Writer B



Revived* the polar histogram $r=p(\varphi)$ for off-line

compute angles φ on the ink edges in the image

Writer C Writer D



* [Maarse, 1988; Crettez, 1995]

Distance Measures

- Hamming:

$$d = \sum_i p_i - q_i$$

- Hausdorff

$$d = \max_i \left(p_i - q_i \right)$$

- Euclid

$$d = \sum_i (p_i - q_i)^2$$

- Chi Squared (χ^2)

$$d = \sum_i \frac{(p_i - q_i)^2}{p_i + q_i}$$

- Minkowski

$$d = \left(\sum_i p_i - q_i{}^n \right)^{\frac{1}{n}}$$

- Bhattacharyya

$$d = 1 - \sum_i p_i \cdot q_i$$

Schomaker, Bulacu & Franke (2004) 9th IWFHR

p(φ) edge-angles histogram

[Bulacu & Schomaker, CAIP, 2003]

- Writer identification
- 250 writers, 1 vs 499 samples
- Top-1 = 29 - 34%
- Top10 = 66 - 79%

Are there other factors besides slant?

(Performance confidence margins)

$\alpha = 0.05$	Perf=90%
N:	Conf.band:
100	±6%
150	±5%
250	±4%
900	±2%

Pen-tip movement in 2D

- wrist and finger movements produce horizontally translated Lissajous patterns: pen-tip movement is a phasor signal

  [Hollerbach, 1981; Schomaker et al. 1989]

- roundness, curvature are clear visual characteristics

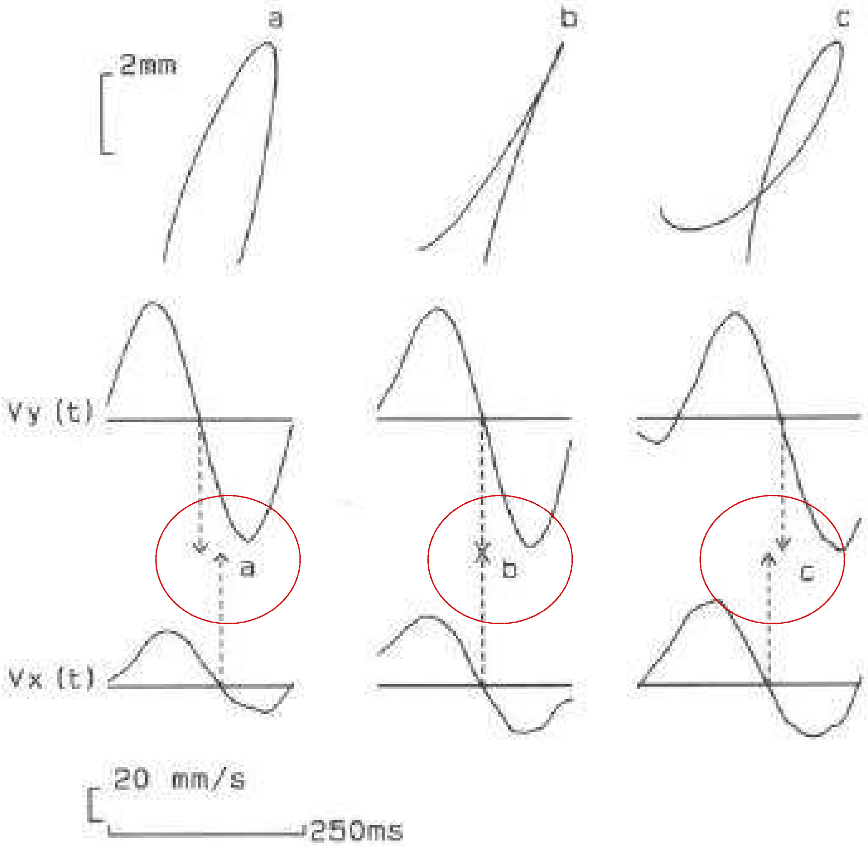


Figure 2. Basic stroke shapes and their relative timing in the velocity domain, a) blunt, clockwise stroke ending, v_x lags v_y, b) sharp stroke ending, no delay, c) blunt, counter-clockwise stroke, looping with next stroke, v_x leads v_y in time. [Schomaker, 1991]

phase v_y, v_x & curvature of handwriting:

$$\Phi = 2\pi \Delta t / T$$

$\Phi \rightarrow 90^0$ "roundish"

$\Phi \rightarrow 0^0$ "sharp"

Roundness of handwriting

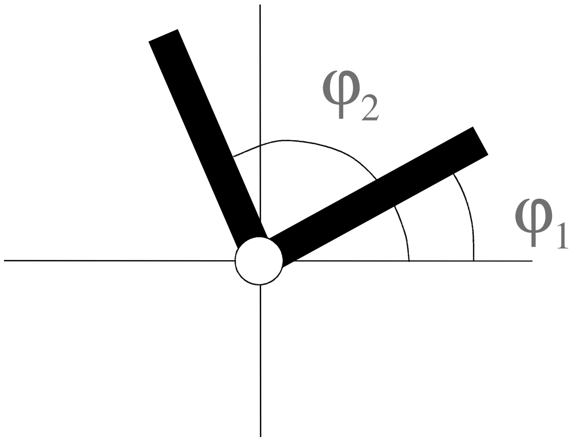
› is caused by the differential phase modulation for the **wrist** and the **finger** system

› Roundness: “transfer function for time signals”
› (Slant: “affine transform”)

Hinge feature [Bulacu & Schomaker, 2003]

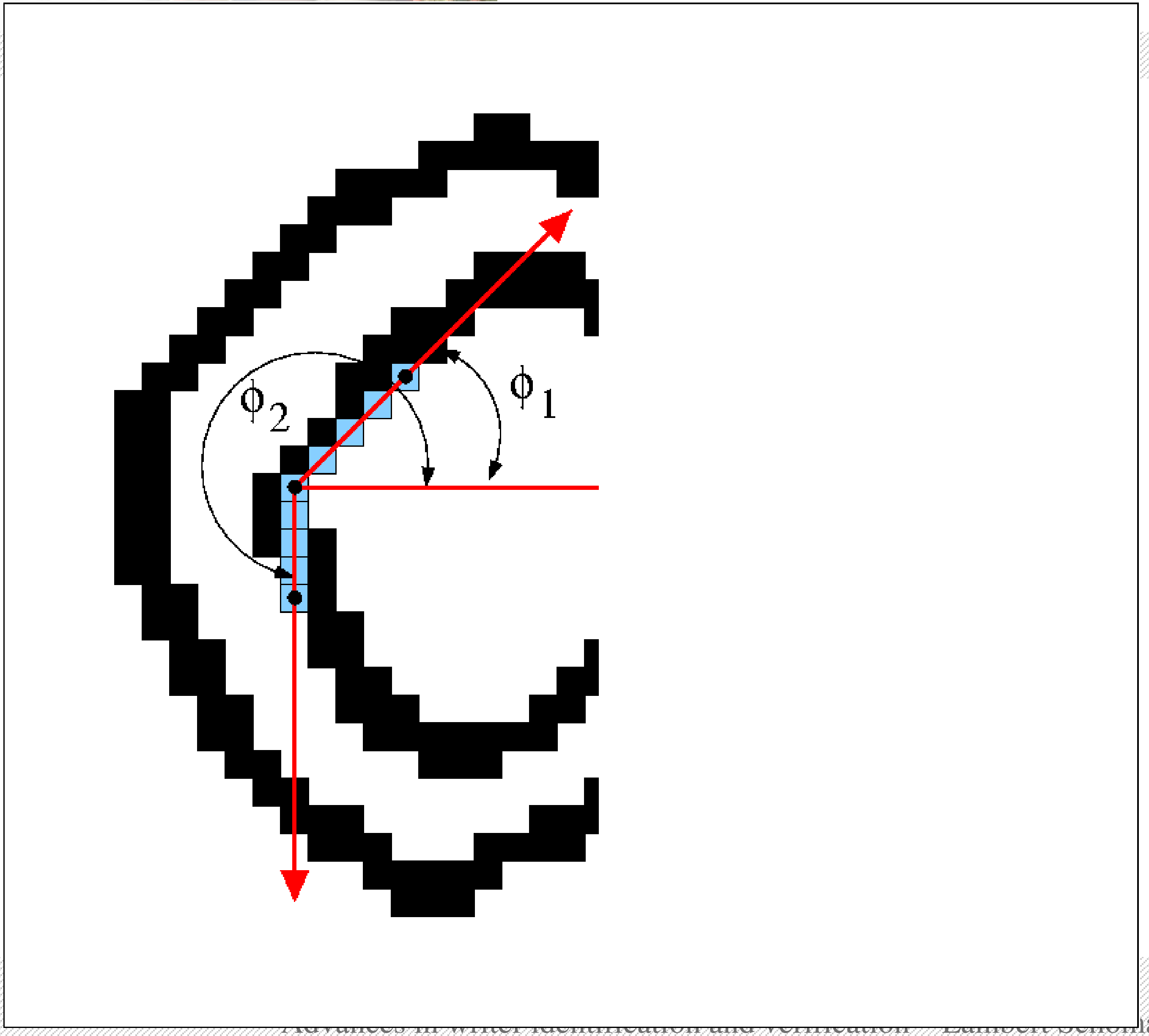
- combines slant information
- and curvature information
- is a transform on images

$$p(\varphi_1,\varphi_2) = H(\varphi_1,\varphi_2,I(x,y))$$



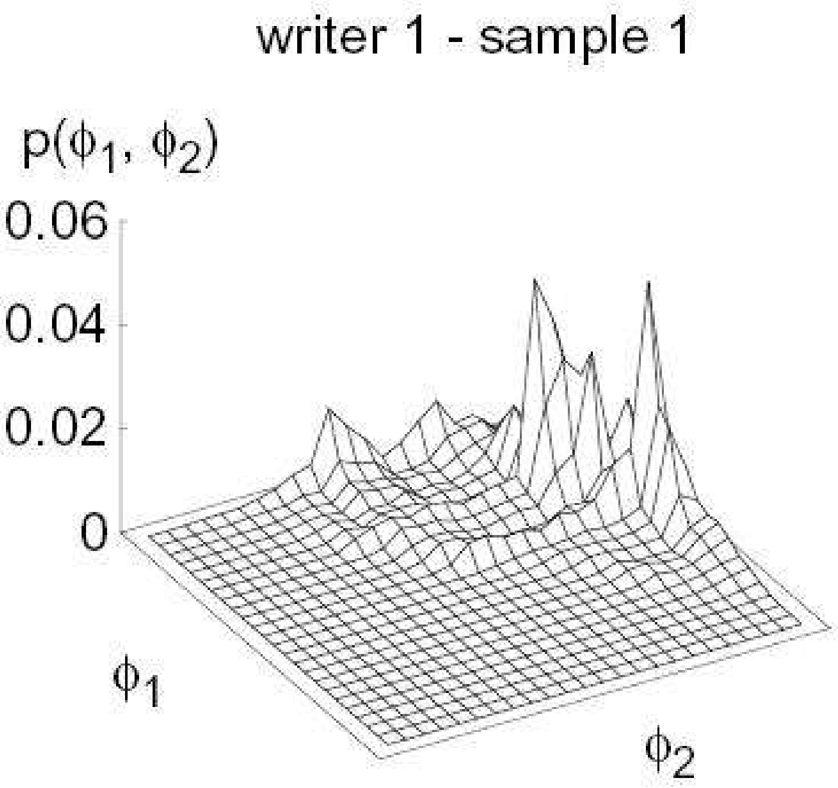
- where the angles φ1,φ2 refer to the legs of a hinge
- probability of angular co-occurrence in hinge patterns along ink contour

Hinge feature



Computing the hinge transform

- Find contours along ink
- Define a set of hinges: φ_1,φ_2 combinations
- And count their occurrence: $p(\varphi_1,\varphi_2)$



[Bulacu & Schomaker, PAMI, 2007]

$p(\varphi_1,\varphi_2)$ hinge histogram [Schomaker & Bulacu, PAMI, 2004]

- Writer identification, Upper case
- nearest neighbour, χ^2 (Chi square) distance measure
- 250 writers, 1 vs 499 samples
- Top-1=83%
- Top10=97%

Part II individual's use of character shapes

- textural features are informative
- what about detailed character shape: **allography**?

Allographs & shape families

How to use allographic information?

- There is no commonly accepted global list of character shapes (allographs) for Western handwriting, nor for other scripts
- Avoid cumbersome human work
  (prefer 'automatic', ROI based analysis)

Allographic writer identification

› Assumption:
- given an exhaustive table of allographs: $L(\lambda, v_\lambda)$ for all letters λ and their variants v_λ
- each writer w is characterized by

$$\vec{w} = p(L)$$

- i.e., a distribution of probabilities of emission of each allograph, by writer w

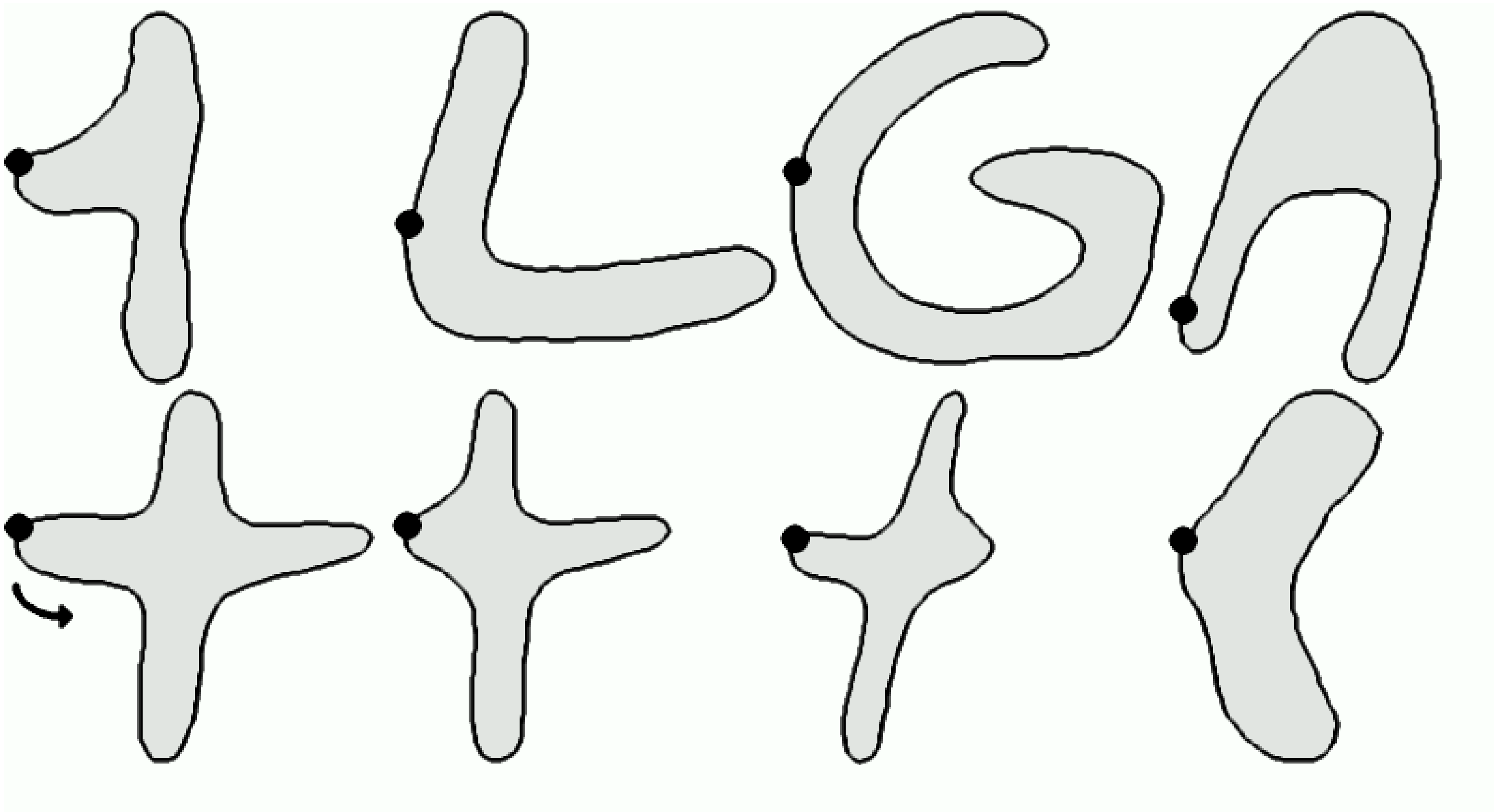
Allographic writer identification

› Hypothesis:
- given an estimated table of character fraglets: *F(v)* with shapes numbered *v*,
- assuming that each fraglet is a *distinct* building block for one or more allographs,
- then each writer *w* can be characterized by

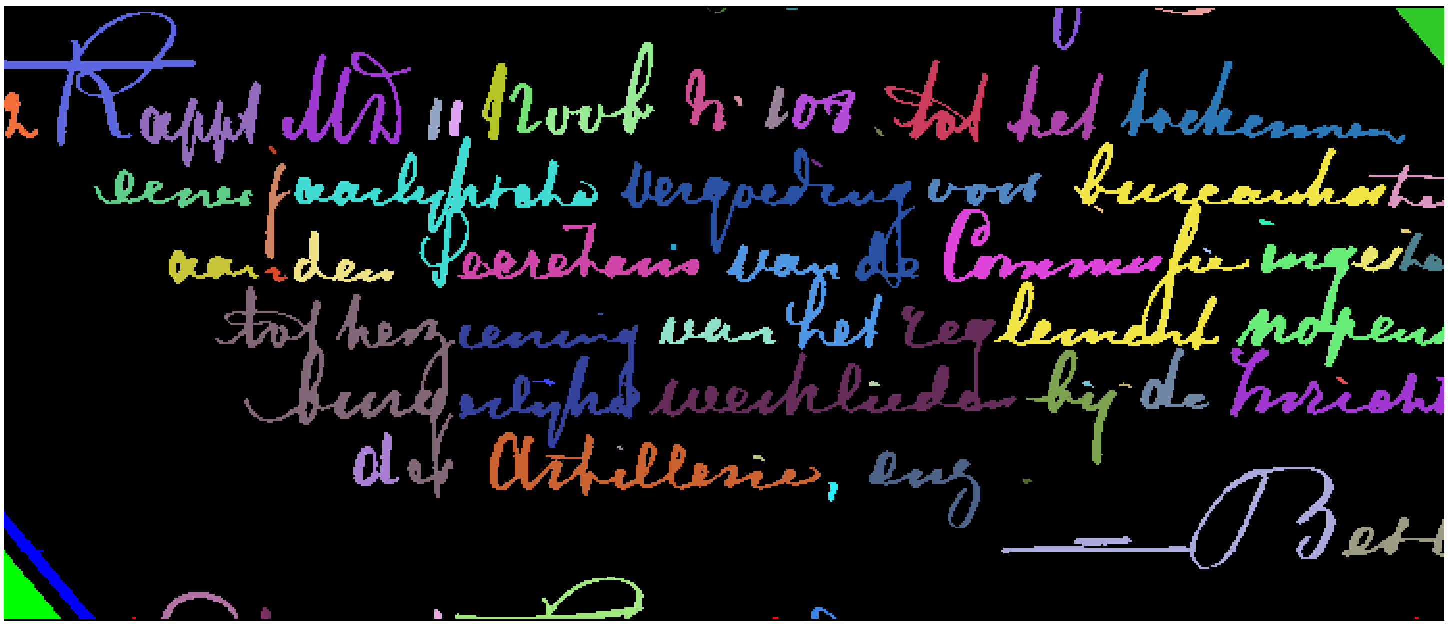
$$w = \vec{p}(F)$$

- i.e., a distribution of probabilities of emission of character fraglets, by writer *w*

Connected-Component Contours (CO^3)

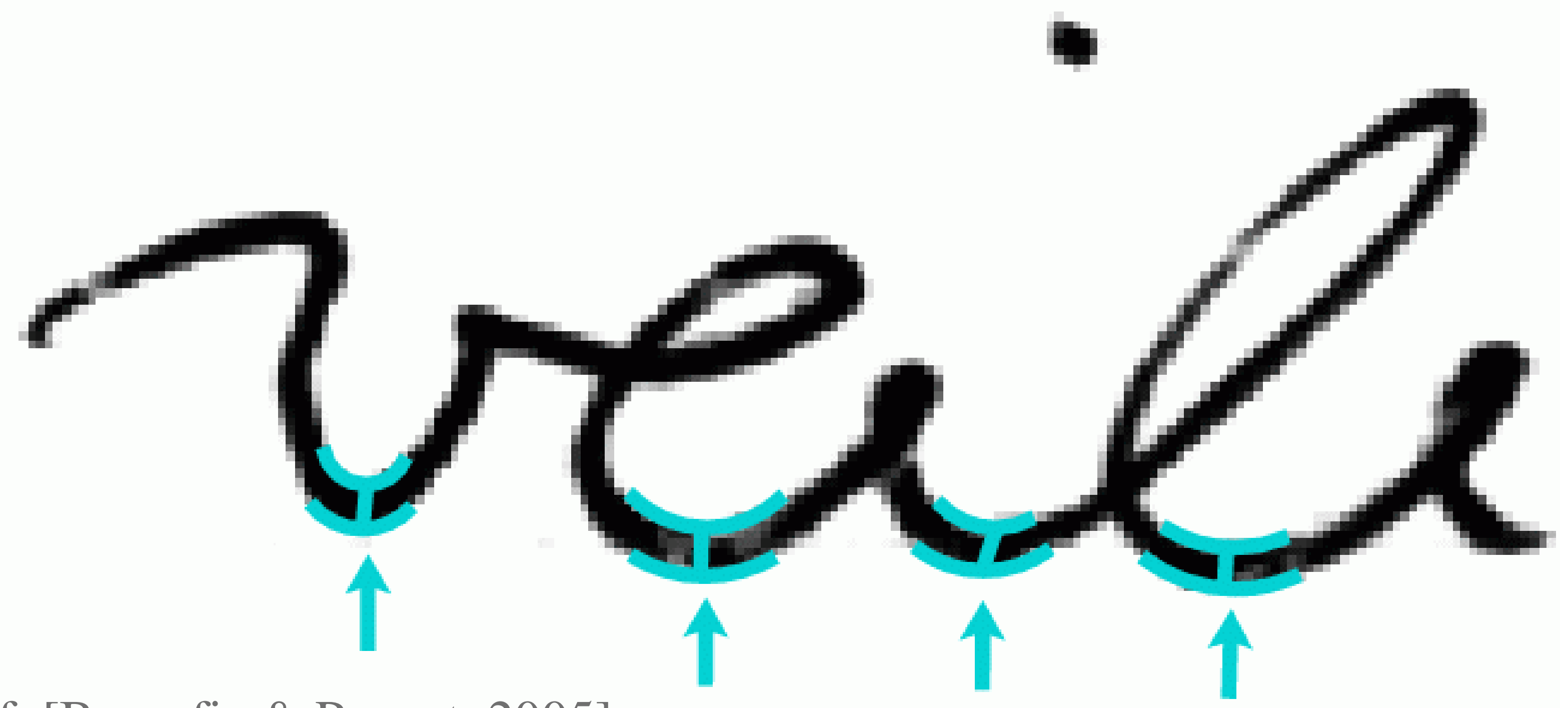


CO^3 works in UPPER case.
How about connected cursive?



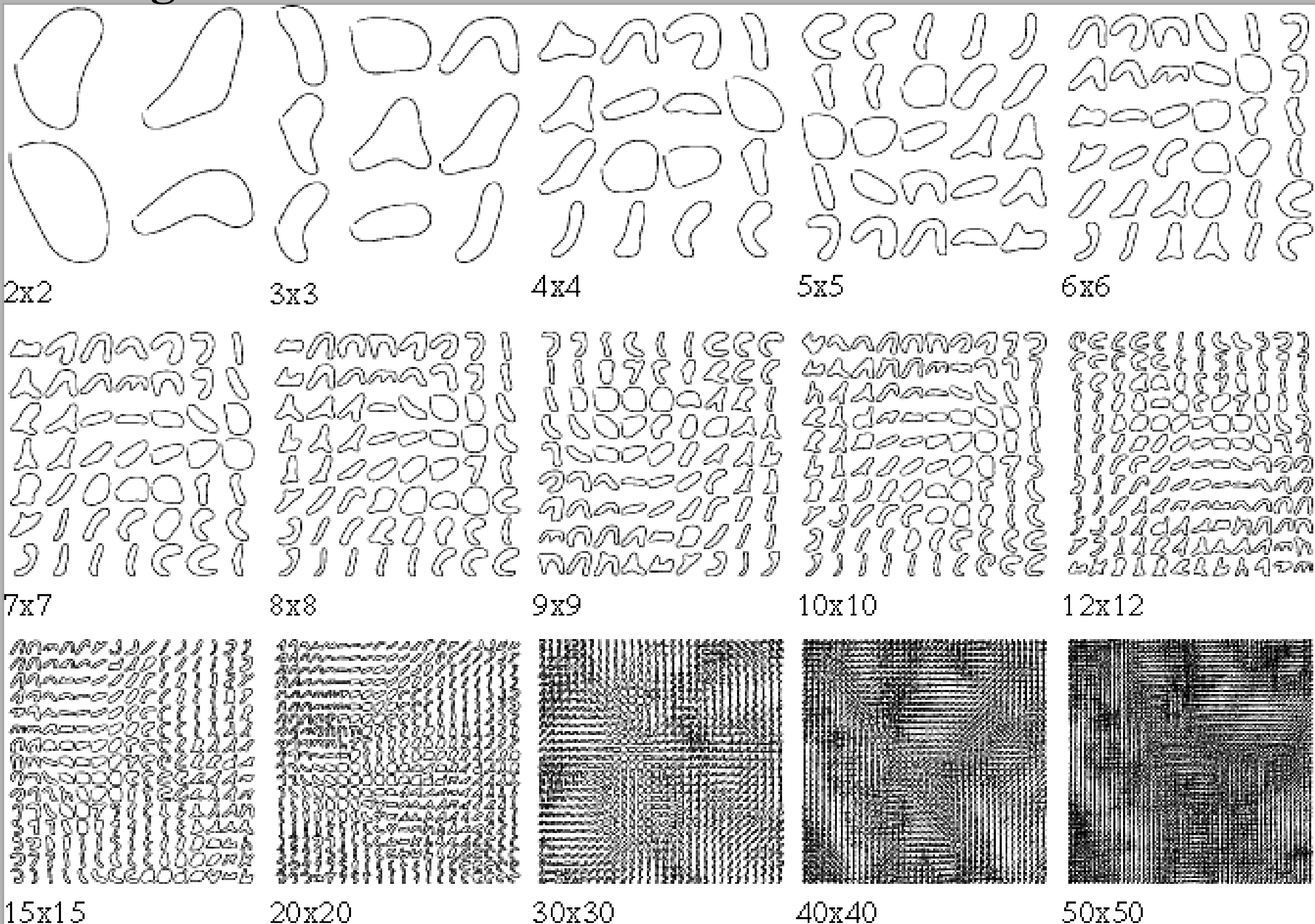
➔Connected components in cursive script are too complex in shape

Segment cursive patterns into fragmented CoCos

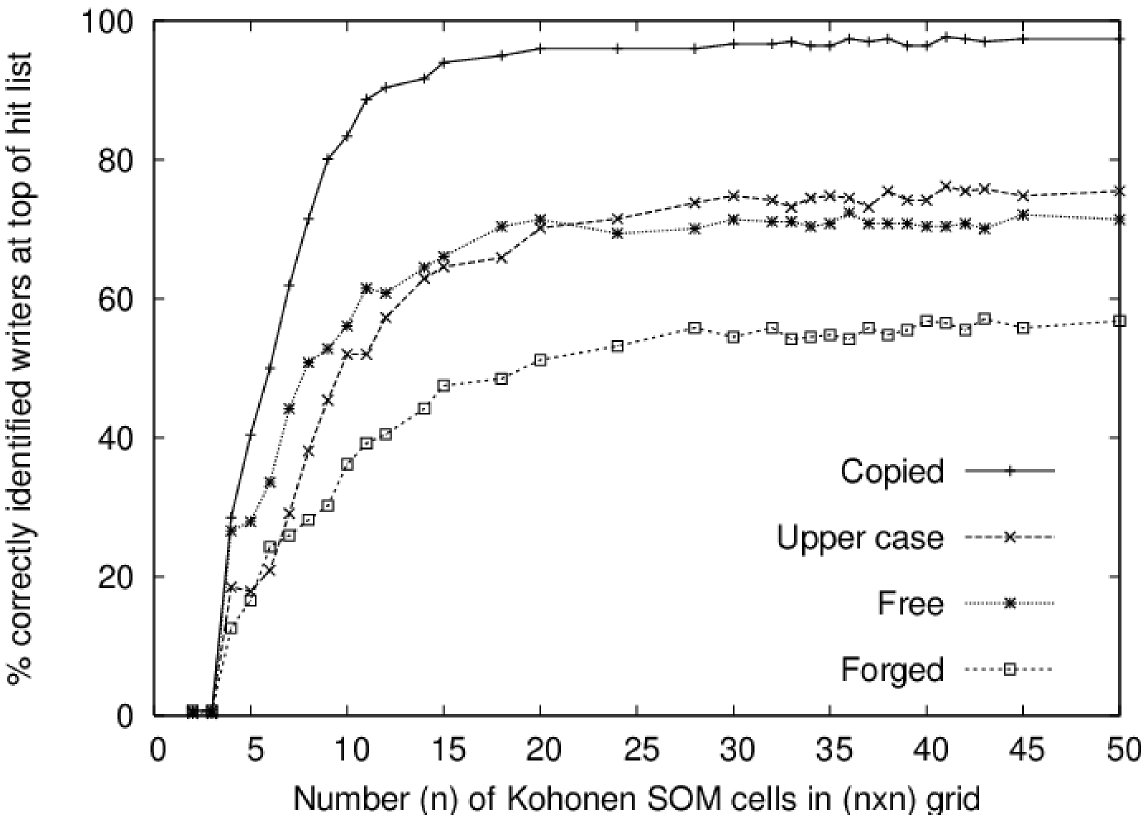


cf. [Bensefia & Paquet, 2005]

Fragmented CO^3 Kohonen SOFMs

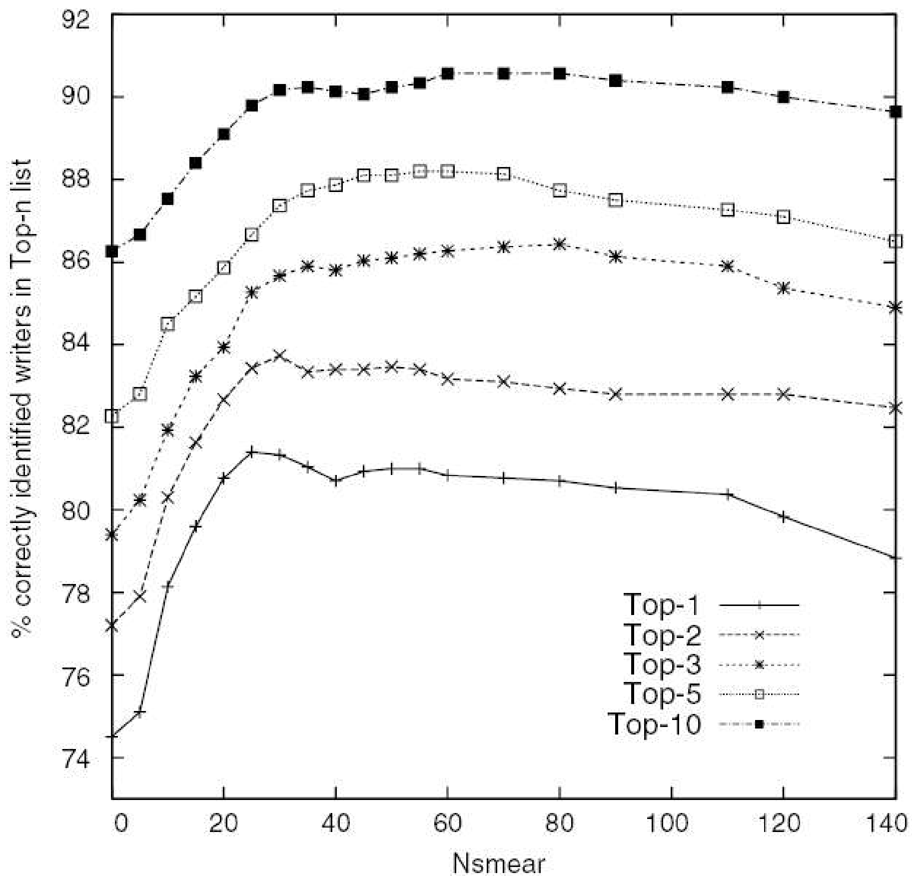


Codebook size and performance



Schomaker, L.R.B. & Bulacu, M. & Franke, K. (2007). Automatic Writer Identification Using Fragmented Connected-Component Contours, In: F. Kimura & H. Fujisawa, Proc. of the 9th IWFHR, IEEE Computer Society, pp. 185-190.

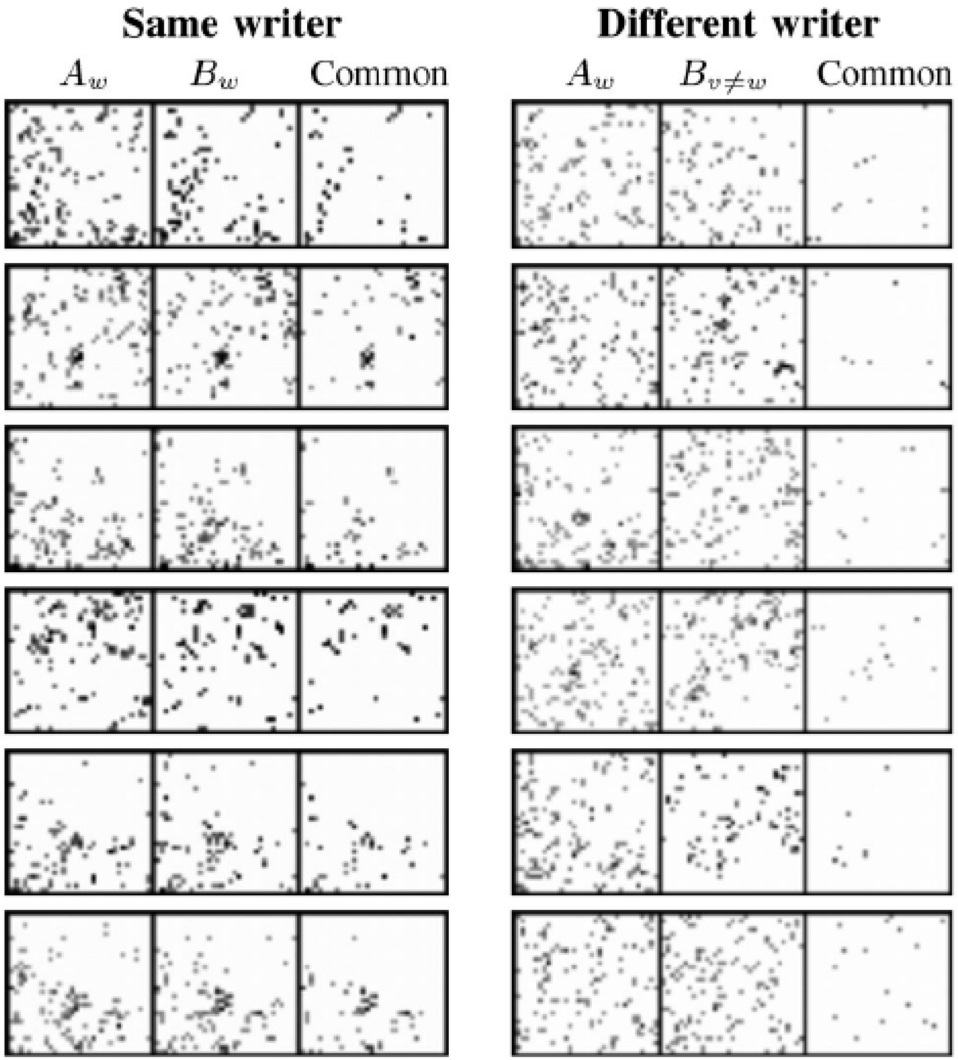
Improved robustness by smearing tallies



Schomaker, L.R.B., Franke, K. & Bulacu, M. (2007). Using codebooks of fragmented connected-component contours in forensic and historic writer identification, *Pattern Recognition Letters, 28(6)*, p. 719-727.

Nsmear: the number of 1NN feature-space neighbours in the codebook to receive a tally, given a connected-component instance

P(CO³) images

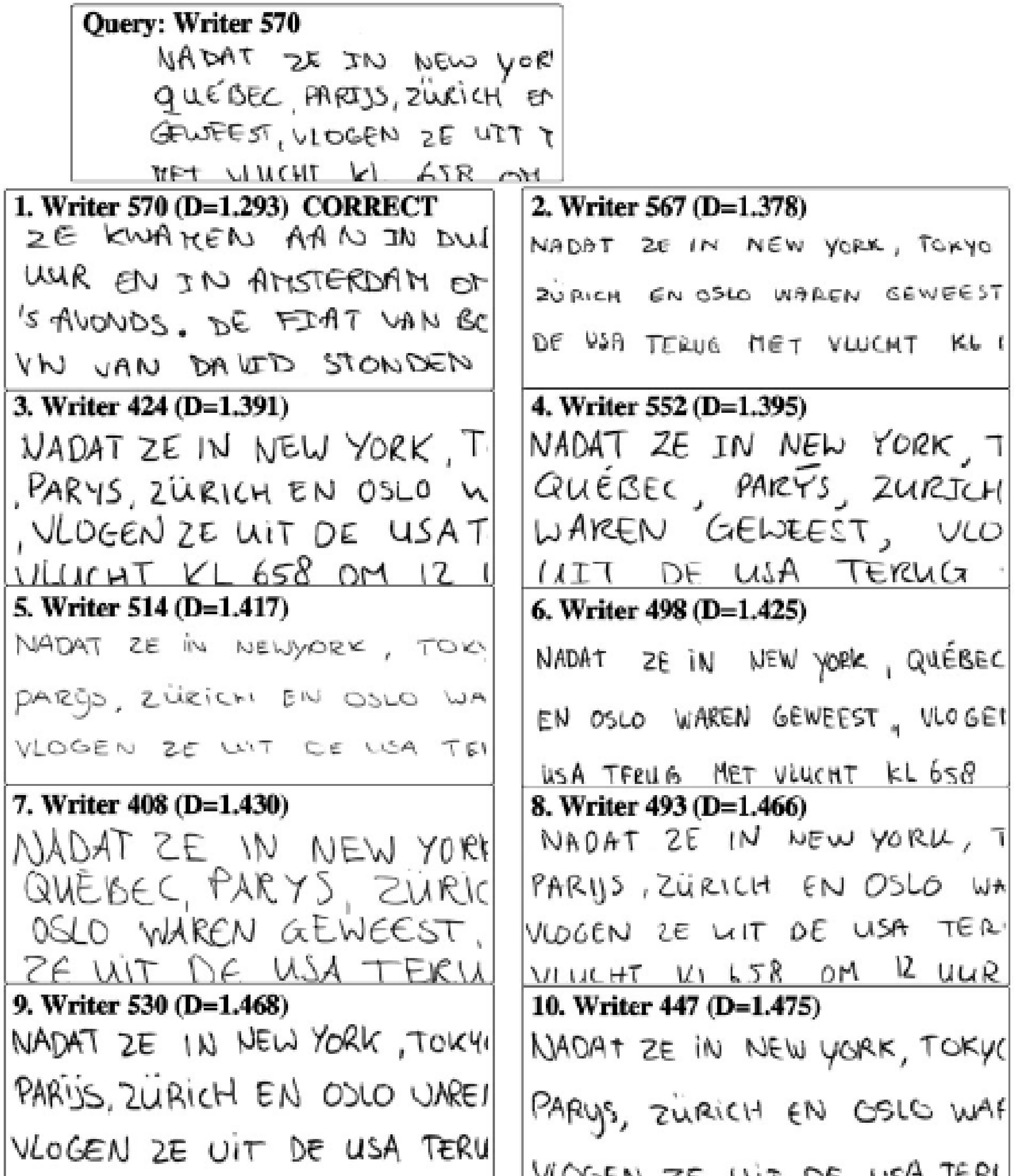


Count fraglet occurrence in input sample

compare to reference vectors

use Hamming or Chi-square distance

Sample hit list



Identification Performance [Schomaker & Bulacu, 2005]

	Feature / Method	N Writers	lowercase		UPPERCASE	
			Top 1	Top 10	Top 1	Top 10
	p(ϕ)	**150**	**53%**	**88%**	**34%**	**79%**
hinge	***p(ϕ₁, ϕ₂)***	**150**	**84%**	**97%**	**84%**	**97%**
	p(ϕ₁, ϕ₃)	**150**	**70%**	**94%**	**68%**	**91%**
	p(CO³)	**150**	**--**	**--**	**72%**	**93%**
fraglets	***p(FCO³)***	**150**	**71 - 97%**	**90 - 100%**	**73%**	**91%**
	system A	**100**	**34%**	**90%**	**--**	**--**
	system B	**100**	**65%**	**90%**	**--**	**--**

Single features, 900 writers [Bulacu & Schomaker, 2006]

		Feature	Identification Top 1	Identification Top 10	Verification EER
	$f1$	$p(\phi)$	43	72	7.1
hinge	$f2$	$p(\phi_1, \phi_2)$	80	91	4.8
	$f3h$	$p(\phi_1, \phi_3)$ h.	65	84	5.9
	$f3v$	$p(\phi_1, \phi_3)$ v.	59	82	9.1
fraglets	$f4$	$p(g)$	76	92	5.8
rl white	$f5h$	$p(rl)$ h.	8	29	16.6
	$f5v$	$p(rl)$ v.	10	34	12.1

Combination performance [Bulacu & Schomaker, 2006]

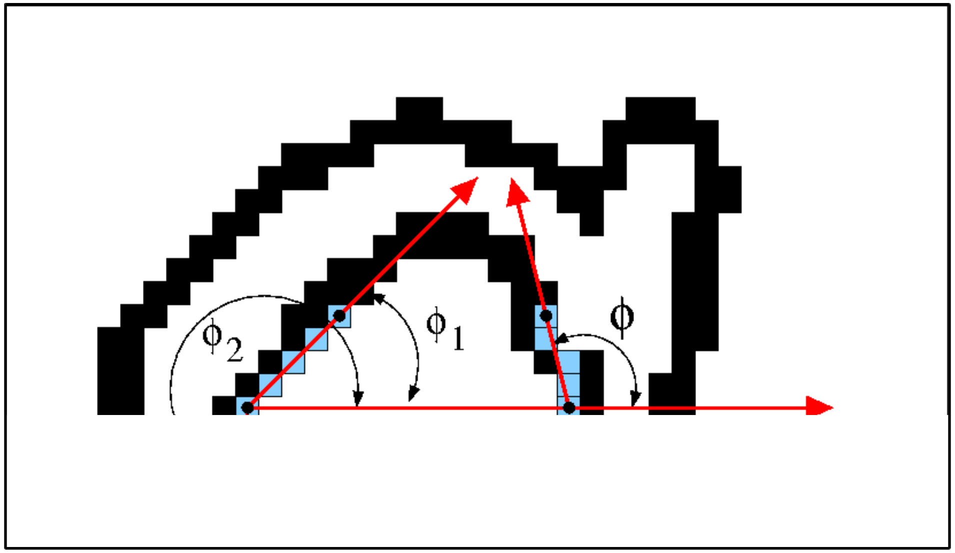
- f1: hinge histogram
- f2: fraglet histogram
- f3: white run-length histogram
- average the distances (Chi square)
- 900 writers
- Top1: 87 %
- Top10: 96%
- EER: 2.6% Using a **single writer-independent** threshold

Other features tested

- Variants on hinge (angle-line-angle, *'ala'*)
- Ink deposition
  - Brush (ink-density related)
- Placement statistics
  - Autocorrelation
  - Horizontal runlength of white
  - Vertical runlength of black
- Misc.
  - #bytes after Lempel-Ziff / #black_pixels
  - wavelets

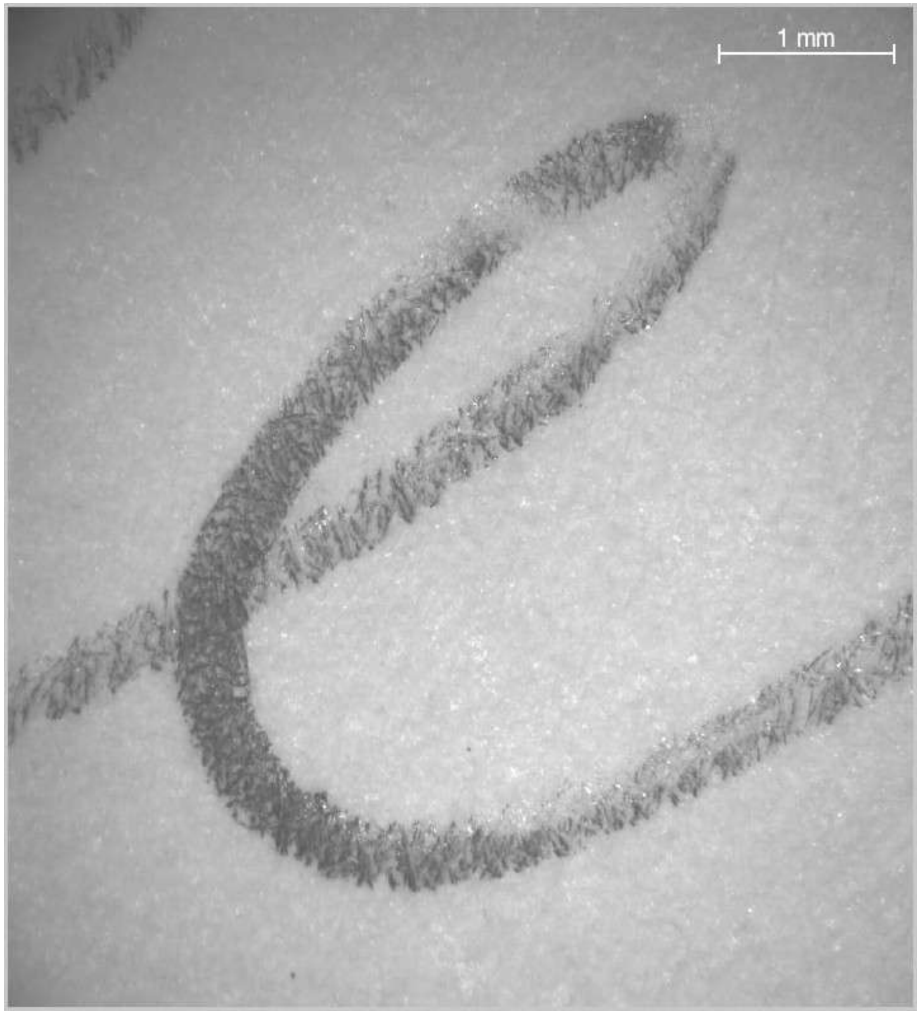
Habitual slant & angle co-occurrence '*ala*'

- P(Phi_1,Phi) where Phi is the angle of the edge to the right of Phi_1 after crossing a non-edge zone [Bulacu & Schomaker, 2003]



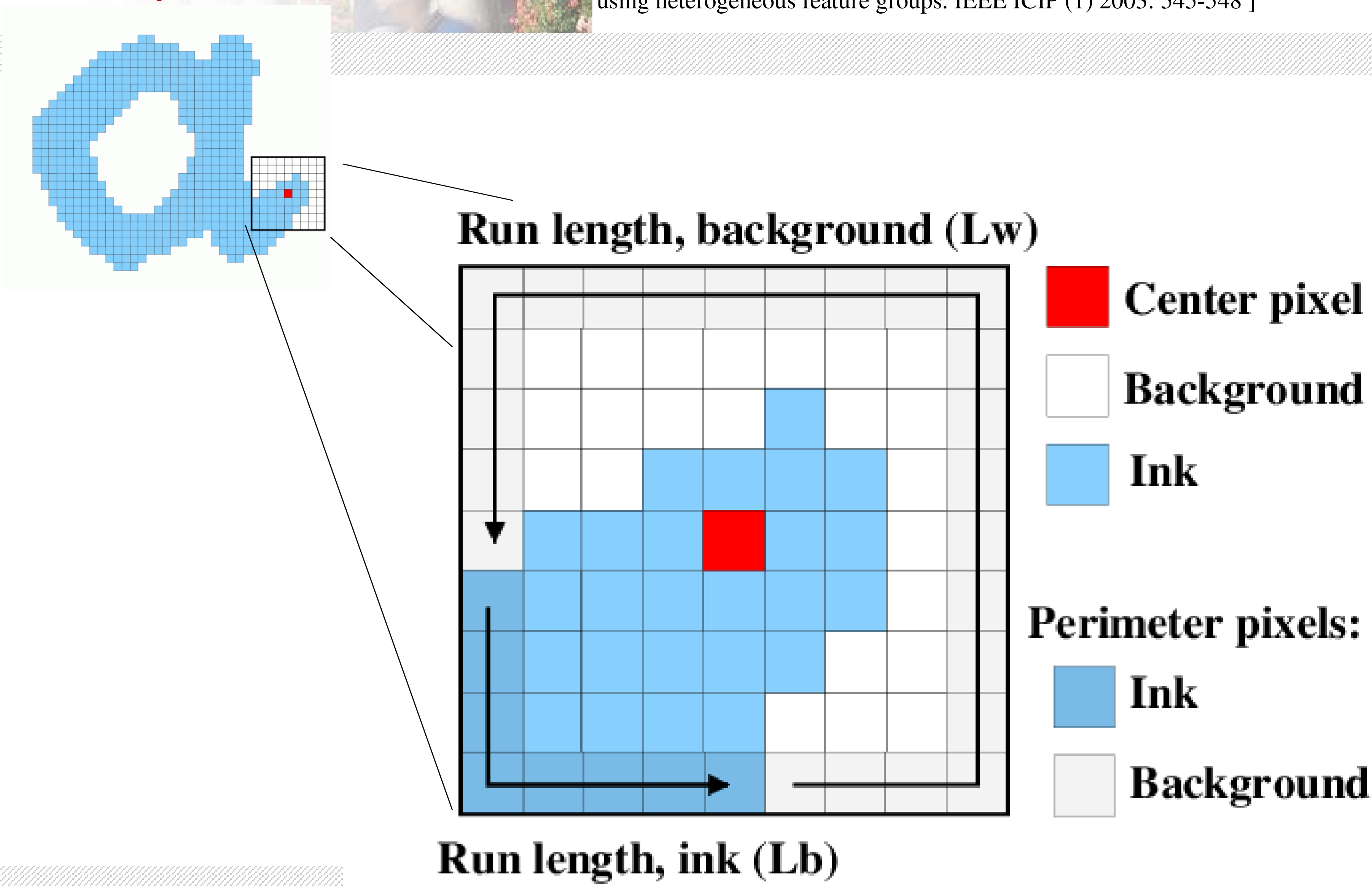
Ink deposition

- Doermann, ICPR 1992
- Franke & Rose (2004); Franke, Schomaker & Köppen (2005): analysis of ink traces by industrial robot

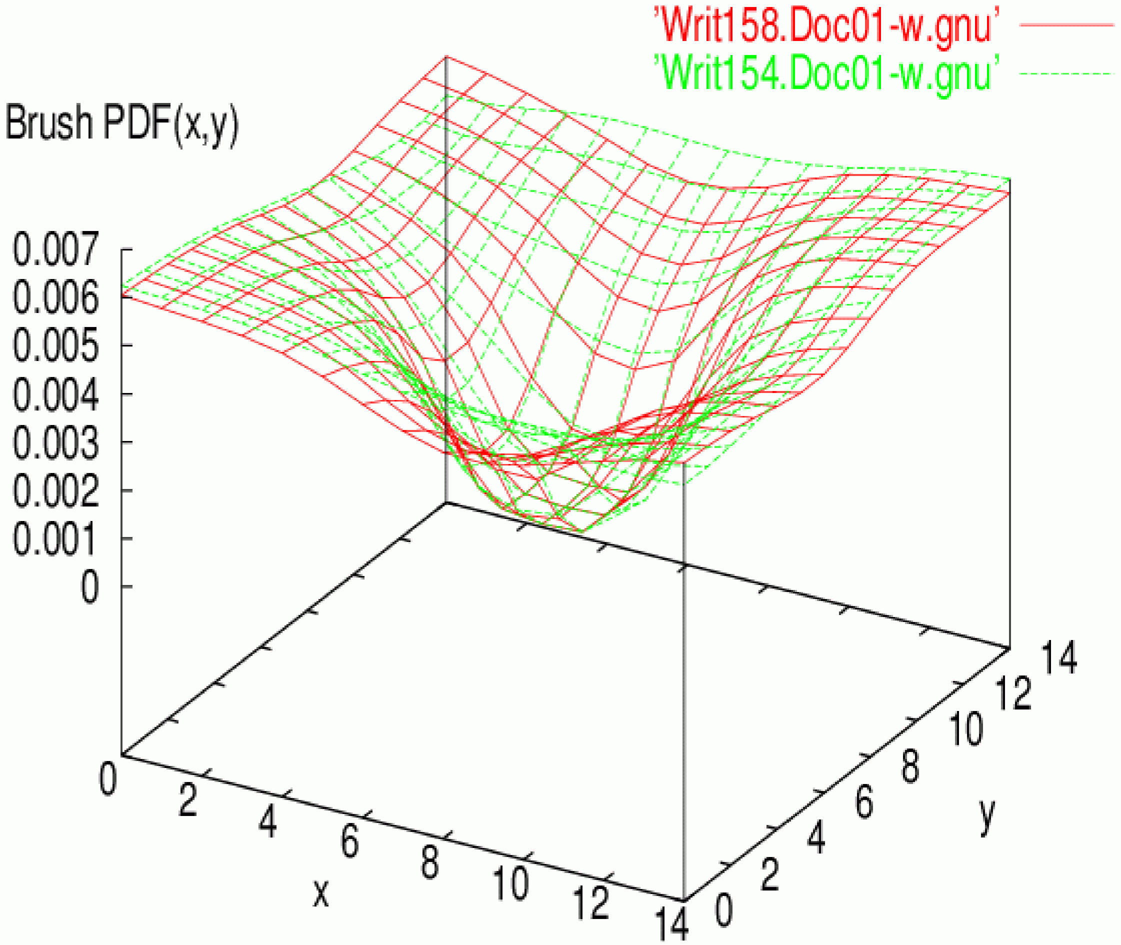


Brush feature

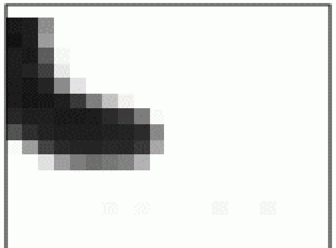
[L.R.B. Schomaker, M. Bulacu, M. van Erp: Sparse-parametric writer identification using heterogeneous feature groups. IEEE ICIP (1) 2003: 545-548]



Brush feature [Schomaker et al., ICIP ’03]




writer A


writer B

Writer identification, 250 writers:

Top1: 53-69%, Top10: 81-93%

Other features UPPER case

Not so good

Nearest-Neighbor Performance of Other Features on Set "ab:" Leave One Out (1 versus 299 Samples), N = 150 Writers, as Before

Feature	Description	$Ndim$	$Top1$ (%)	$Top10$ (%)
e	normalized entropy	1	2	19
w1	wavelets,Haar	99	5	14
w2	wavelets,Odegard	99	14	28
w3	wavelets,Adelson	99	14	29
w4	wavelets,Antonini	99	14	29
w5	wavelets,Brislawn	99	14	29
w6	wavelets,Daubechies 14	99	15	29
w7	wavelets,Villasenor 2	99	15	30
v	vertical run-length PDF	100	21	61
r	horizontal autocorrelation	100	25	61
h	horizontal run-length PDF	100	26	66
f0	edge-angular PDF	16	34	79
b	brush feature, 15x15	225	69	93
f1	CO^3 PDF	1089	72	93
f2	hinge-angular PDF	464	80	97

Given are the dimensionality $Ndim$ of the feature vectors and the $Top1$ and $Top10$ percentages of the correct writer found in a sorted hit list of size 1 and 10, respectively.

“comparison” with other systems

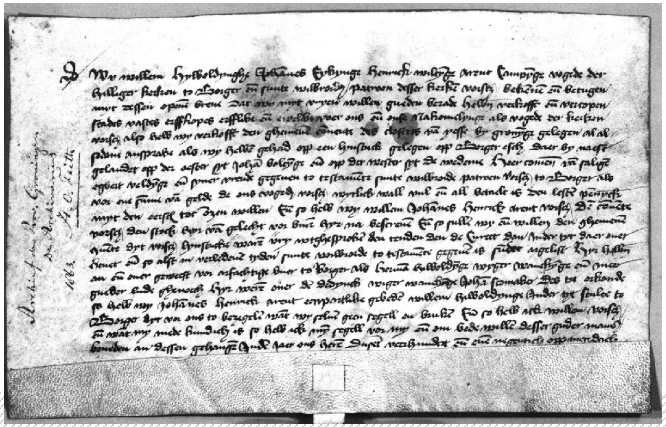
Method/Feature	*N writers*	*Top-1 (%)*	*Top-10 (%)*	*EER (%)*	*Match*	*Notes*	
misc. features	20	91	-	-	MLP	w [17]	Marti et al
'SysA'	100	34	90	-	-	w [28]	
'SysB'	100	65	90	-	LDA	w [28]	
char. models	100	-	-	0.9	HMM	w [25]	Bunke et al
co3	150	72	93	-	1NN	p,U [28]	
brush	250	53	81	-	1NN	n [30]	
splitEdge	250	29	69	-	1NN	n [3]	
splitAla	250	64	86	-	1NN	n [3]	
splitHinge	250	79	96	-	1NN	n [3]	
fco3	900	76	92	5.8	1NN	p [4]	
hinge+fco3+runl	900	79	96	3.3	1NN	p [4]	
misc. features	1500	96	-	3.5	1NN	w [33]	Srihari et al

Wrapping up

- Hinge and allographic fraglet methods are powerful
- Automatic 'ROI' based methods work!
  - Also on Arabic and historical manuscripts!
  - No writer-specific training
- Combining textural and allographic features improves performance
- Human experts are still better at verification!

Future work

- Sensitivity to amount of text in the sample
- Asymptotic performance?
- Explainable decisions?
- Scribe classification of historical charters
  [Bulacu & Schomaker, ICIAP, 2007]
- ... new features [Brink, ...]
- Writer-specific verification



References

- Bulacu, M. & Schomaker, L.R.B. (2007). Text-independent Writer Identification and Verification Using Textural and Allographic Features, *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI),* Special Issue - Biometrics: Progress and Directions, April, 29(4), p. 701-717.
- Niels, R., Vuurpijl, L. & Schomaker, L.R.B. (2007). Automatic allograph matching in forensic writer identification. *International Journal of Pattern Recognition and Artificial Intelligence, 21(1)*, pp. 61-81, ISSN 0218-0014.
- Schomaker, L.R.B., Franke, K. & Bulacu, M. (2007). Using codebooks of fragmented connected-component contours in forensic and historic writer identification, *Pattern Recognition Letters, 28(6)*, p. 719-727.
- L. Schomaker & M. Bulacu (2004). Automatic writer identification using connected-component contours and edge-based features of upper-case Western script. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol 26(6),* June 2004, pp. 787 - 798.
- Marius Bulacu, Lambert Schomaker (2007) "Automatic handwriting identification on medieval documents", *Proc. of 14th Int. Conf. on Image Analysis and Processing (ICIAP 2007)*, IEEE Computer Society, pp. 279-284, 11 - 13 September, Modena, Italy.

http://www.ai.rug.nl/~lambert/publications.html

References

- M. Bulacu, L.R.B. Schomaker, A. Brink (2007) Text-independent writer identification and verification on offline Arabic handwriting", *Proc. of 9th Int. Conf. on Document Analysis and Recognition (ICDAR 2007)*, IEEE Computer Society, pp. 769-773, vol. II, 23 - 26 September, Curitiba, Brazil.
- A. Bensefia, T. Paquet, L. Heutte: Information Retrieval Based Writer Identification. ICDAR 2003: 946-950
- D.S. Doermann and A. Rosenfeld, Temporal clues in handwriting, Proc. ICPR Vol. B, The Hague, 1992, ISBN: 0-8186-2915-0 , pp. 317-320.
- E.H. Dooijes: Decomposition of the pen displacement signal in the analysis of handwriting motorics. In: Signal Processing III: Theories and Applications, I.T. Young et al. (eds), North Holland 1986, pp 1343-1345.
- F. Maarse, L.R.B. Schomaker & H.-L. Teulings, Automatic Identification of Writers, Proc. Conference of the Dutch Psychonomic Society on Human-computer interaction: psychonomic aspects, ISBN:0-387-18901-7 ,New York, Springer 1988, pp. 353-360.