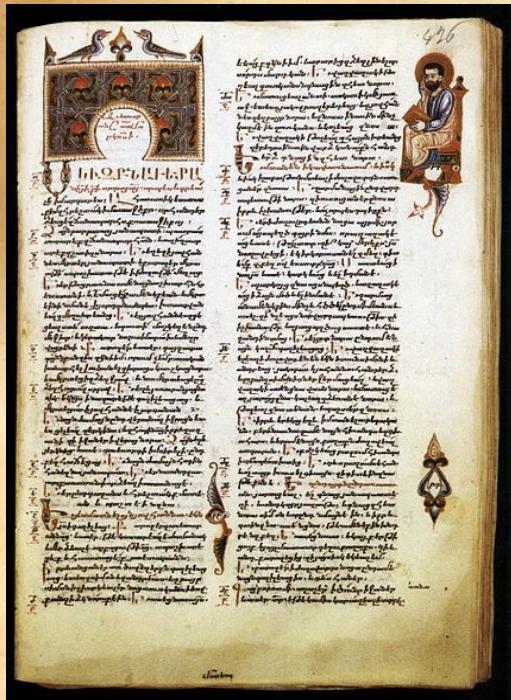
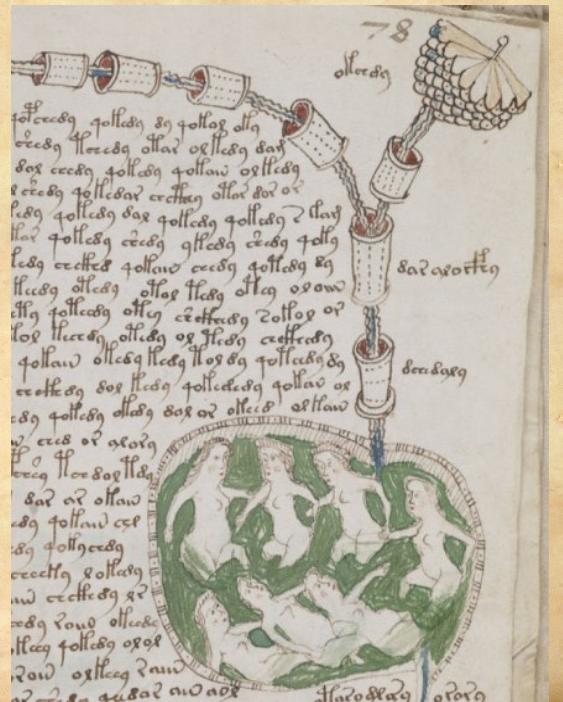


*Query Synthesis  
for Information Retrieval in Machine Printed and  
Latin Script Handwritten Documents*

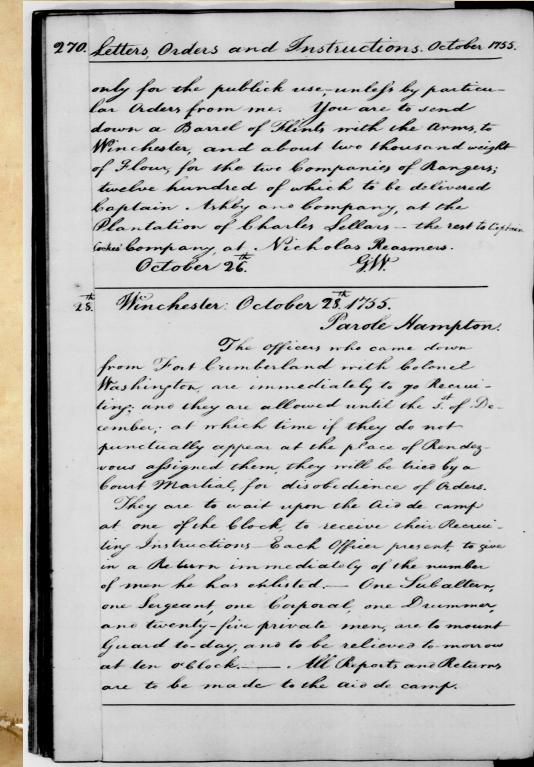
# Information Retrieval



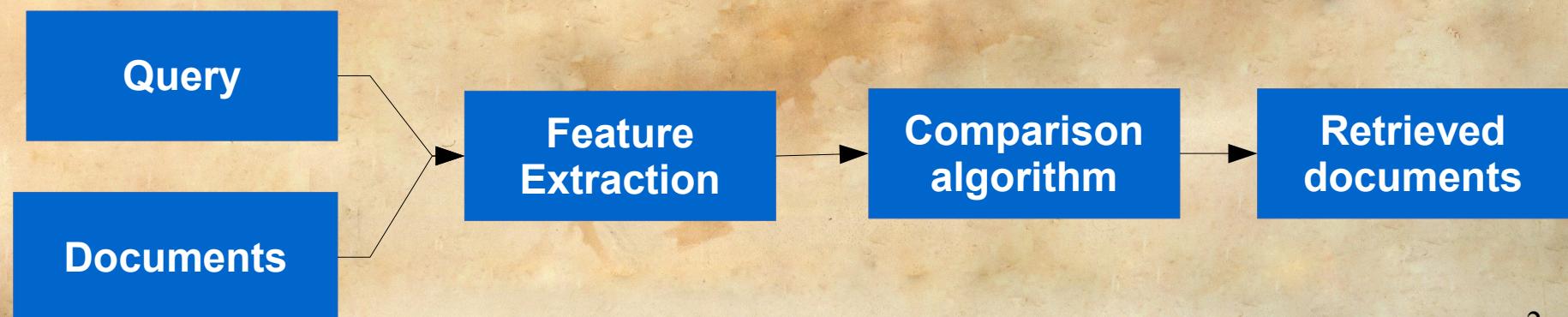
[Latin manuscript , Wikipedia]



[Voynich manuscript, Wikipedia]



George Washington manuscript



# OCR, Word Spotting, Word Retrieval

## Optical Character Recognition

Opticd Charac ter Recognitio

Αναγνώσιση Χαρακτήρων

Word Spotting

Word Spotting

erit signū ihud. fecitq; dñs u  
grauillima in domū pharaois.  
omniē terrā **egypti**. Corrupt  
iuscmodi miseri. vocavit p  
aaron & dixit eis. Ite sacrific  
in terra hac. & ait moyses. N

1. Word-spotting

Α ναγνώρισπ χάθακτήθω

egýptos  
egýpto  
egýpta  
egýptu  
egýpta,  
egýptos

egypt |

Possible representations

egýpt  
egýpt

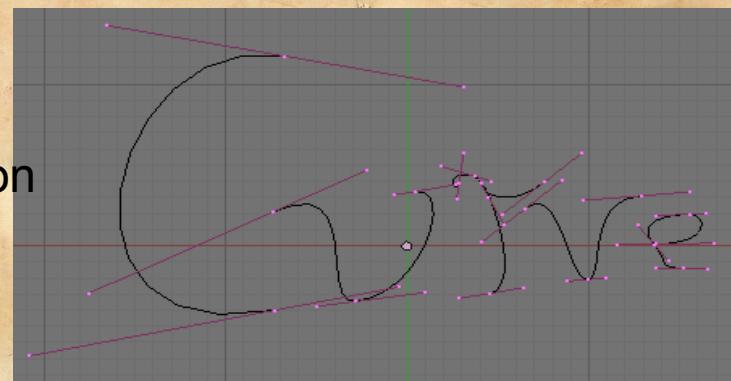
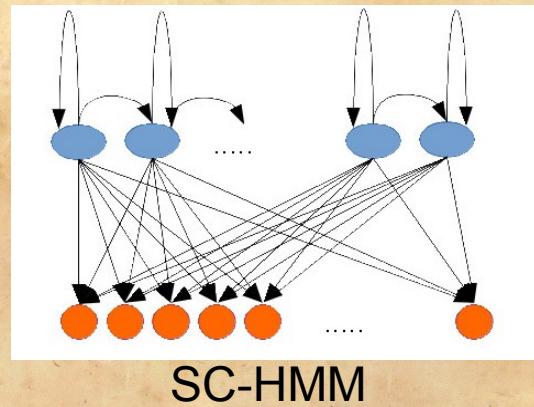
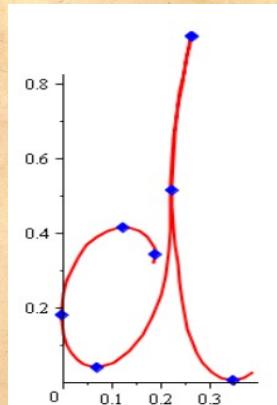
egýptos  
egýpto  
egýpta  
egýptu  
egýpta,  
egýptos

2. Word retrieval

[Towards an omnilingual word retrieval system for ancient manuscripts Y.Leydier , A. Ouji , F. LeBourgeois H. Emptoz Elsevier 2009]

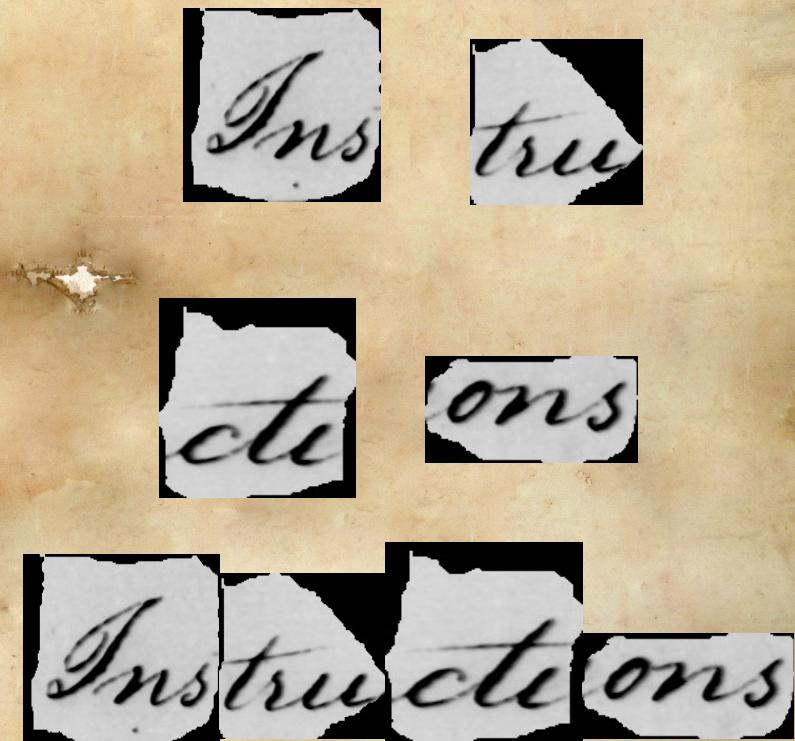
# Synthesis...

- Glyphs
- Bézier curves, bézier splines, polynomials
- Abstract model

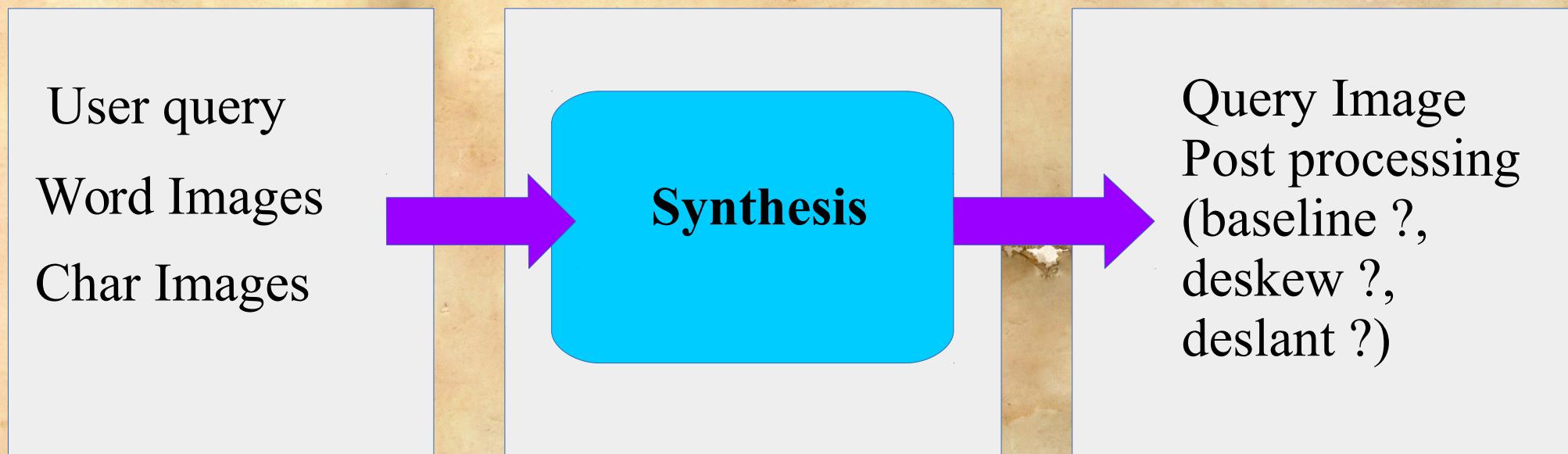


[Polynomial Approximation  
in HR Stephen M.Watt]

Bézier Control points



# *The stages of query synthesis for printed*

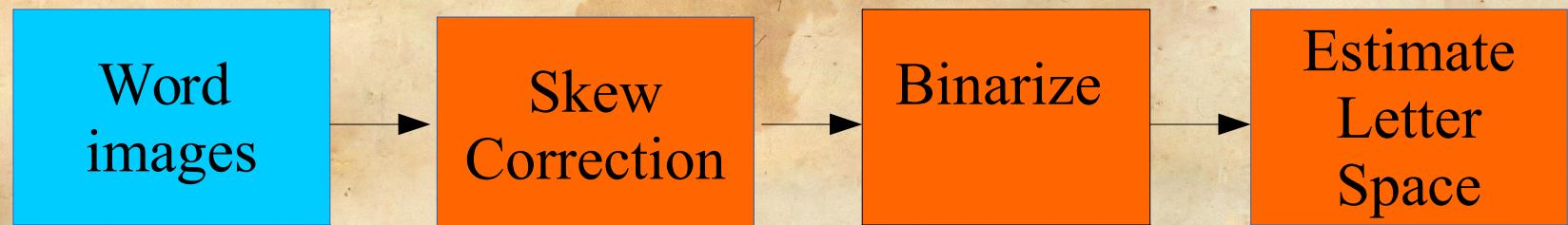
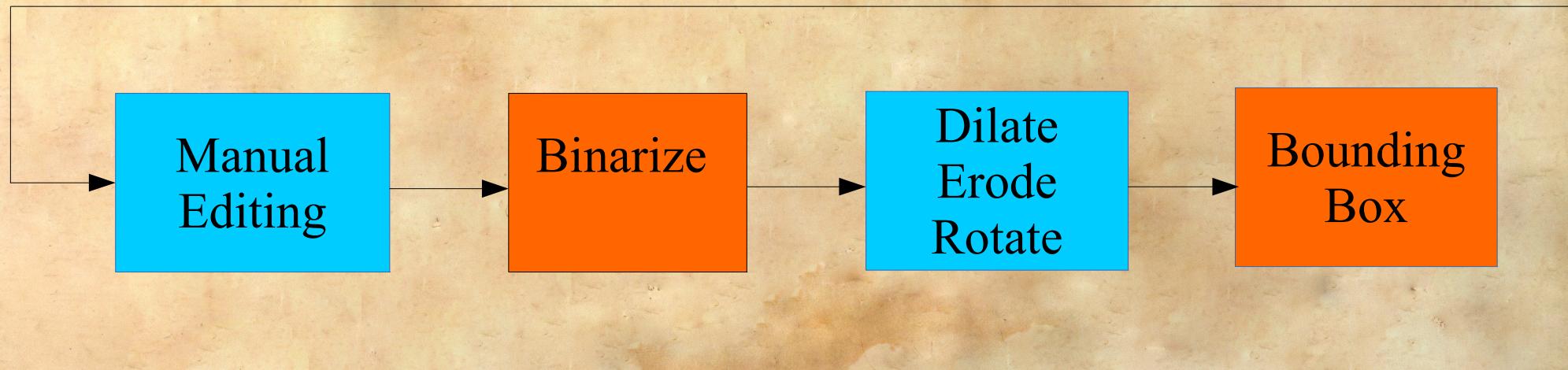
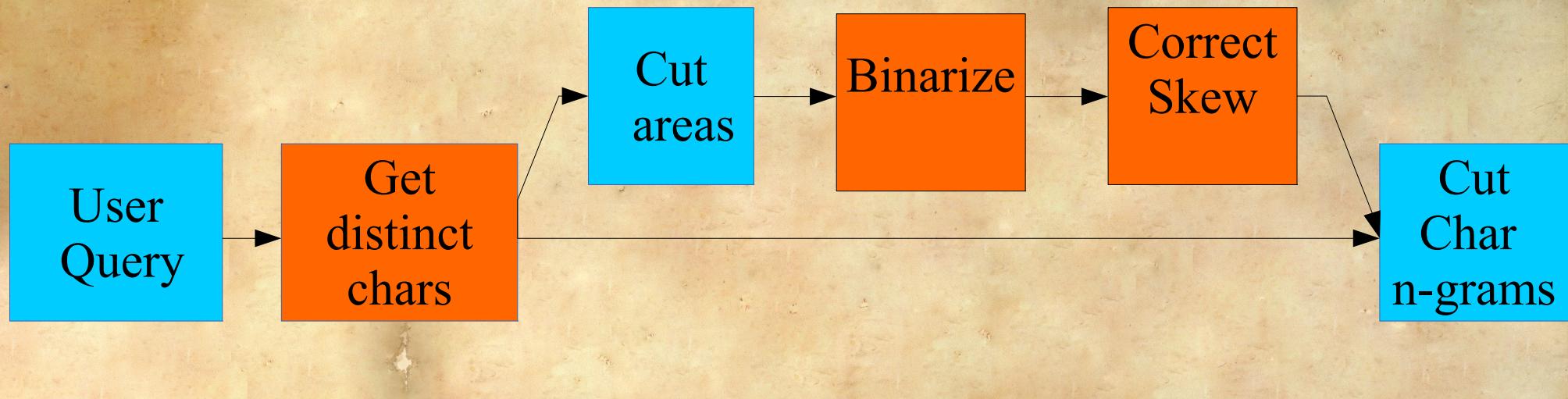


# Old Greek

ἀ	ἄ	ᾶ	ᾶ	ᾶ	ᾶ	ᾶ	ᾶ	Ἄ	Ἄ	Ἄ	Ἄ	Ἄ	Ἄ	Ἄ	Ἄ	Ἄ	Ἄ	Ἄ	Ἄ	Ἄ
έ	έ	ἔ	ἔ	ἔ	ἔ			Ἐ	Ἐ	Ἐ	Ἐ	Ἐ	Ἐ	Ἐ	Ἐ	Ἐ	Ἐ	Ἐ	Ἐ	
ή	ή	ῆ	ῆ	ῆ	ῆ	ῆ	ῆ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ
ἰ	ἴ	ἴ	ἴ	ἴ	ἴ	ἴ	ἴ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ	ἴ
օ	օ	օ	օ	օ	օ			Ὀ	Ὀ	Ὀ	Ὀ	Ὀ	Ὀ	Ὀ	Ὀ	Ὀ	Ὀ	Ὀ	Ὀ	
ύ	ύ	ύ	ύ	ύ	ύ	ύ	ύ		Ὕ		Ὕ		Ὕ		Ὕ		Ὕ		Ὕ	Ὕ
ώ	ώ	ώ	ώ	ώ	ώ	ώ	ώ	Ω	Ω	Ω	Ω	Ω	Ω	Ω	Ω	Ω	Ω	Ω	Ω	Ω
ὰ	ά	ὲ	έ	ή	ή	ἴ	ἴ	օ	ό	ὐ	ὐ	ὐ	ὐ	ὐ	ὐ	ὐ	ὐ	ὐ	ὐ	
ᾳ	ᾳ	ᾳ	ᾳ	ᾳ	ᾳ	ᾳ	ᾳ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ
ῃ	ῃ	ῃ	ῃ	ῃ	ῃ	ῃ	ῃ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ
ῳ	ῳ	ῳ	ῳ	ῳ	ῳ	ῳ	ῳ	Ϙ	Ϙ	Ϙ	Ϙ	Ϙ	Ϙ	Ϙ	Ϙ	Ϙ	Ϙ	Ϙ	Ϙ	Ϙ
ᾳ	ᾳ	ᾳ	ᾳ	ᾳ	ᾳ			Ἄ	Ἄ	Ἄ	Ἄ	Ἄ	Ἄ	Ἄ	Ἄ	Ἄ	Ἄ	Ἄ	Ἄ	
ῃ	ῃ	ῃ	ῃ	ῃ	ῃ			Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	
ῳ	ῳ	ῳ	ῳ	ῳ	ῳ			Ὕ	Ὕ	Ὕ	Ὕ	Ὕ	Ὕ	Ὕ	Ὕ	Ὕ	Ὕ	Ὕ	Ὕ	
ᾳ	ᾳ	ᾳ	ᾳ	ᾳ	ᾳ			Ω	Ω	Ω	Ω	Ω	Ω	Ω	Ω	Ω	Ω	Ω	Ω	
ῃ	ῃ	ῃ	ῃ	ῃ	ῃ			Ῥ	Ῥ	Ῥ	Ῥ	Ῥ	Ῥ	Ῥ	Ῥ	Ῥ	Ῥ	Ῥ	Ῥ	
ῳ	ῳ	ῳ	ῳ	ῳ	ῳ			Ῥ	Ῥ	Ῥ	Ῥ	Ῥ	Ῥ	Ῥ	Ῥ	Ῥ	Ῥ	Ῥ	Ῥ	
	ῳ	ῳ	ῳ	ῳ	ῳ															

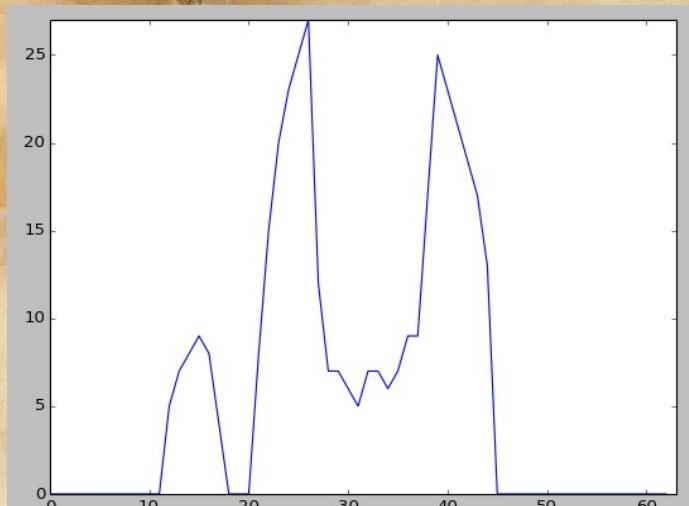
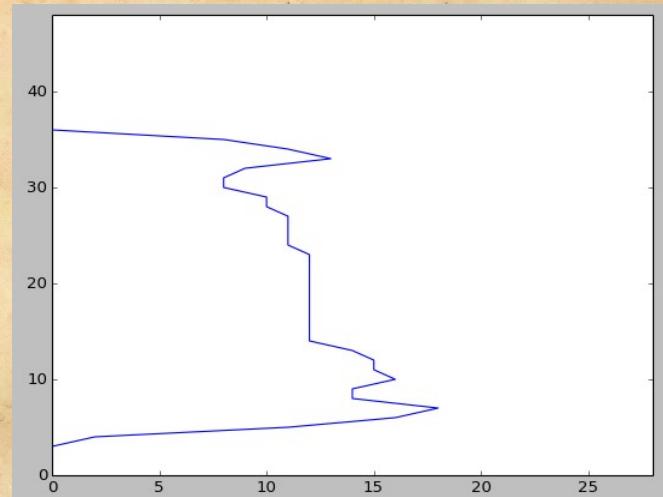
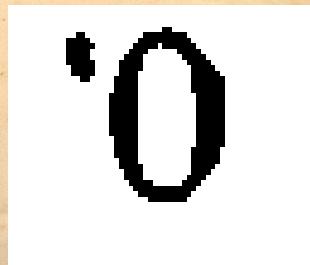
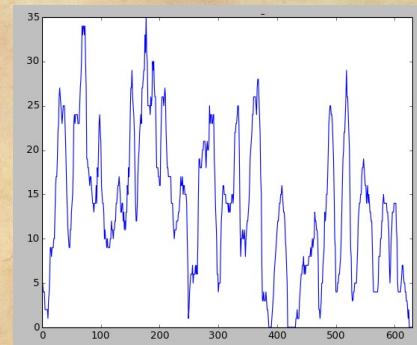
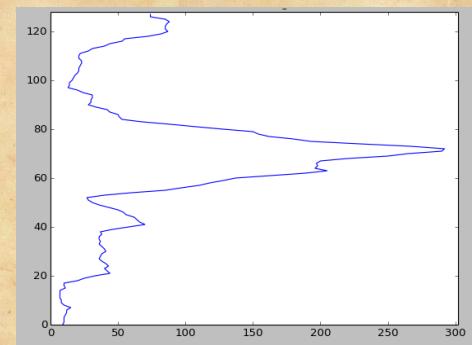
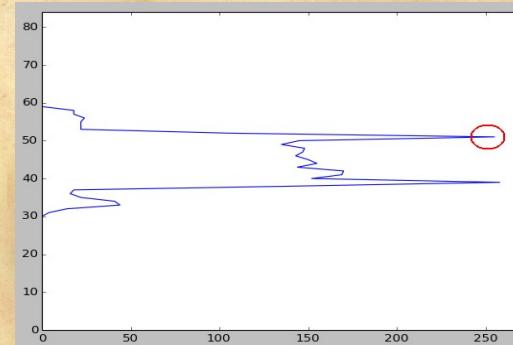
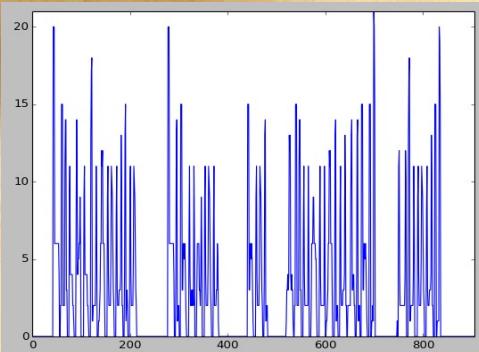
[[http://en.wikipedia.org/wiki/Greek\\_alphabet](http://en.wikipedia.org/wiki/Greek_alphabet)]

## *1st Stage*

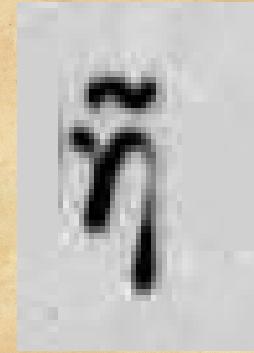
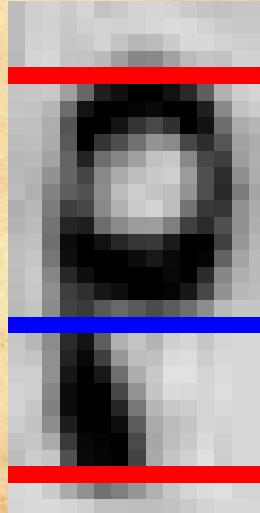


# Histograms

Επεξεργασία Εικόνας και Υπολογιστική Όραση



*Naïve/experimental  $\gamma, \zeta, n, \mu, \xi, \varrho, \varphi, \psi, \varsigma \dots$*



# *Letter /Word Distance Estimation*

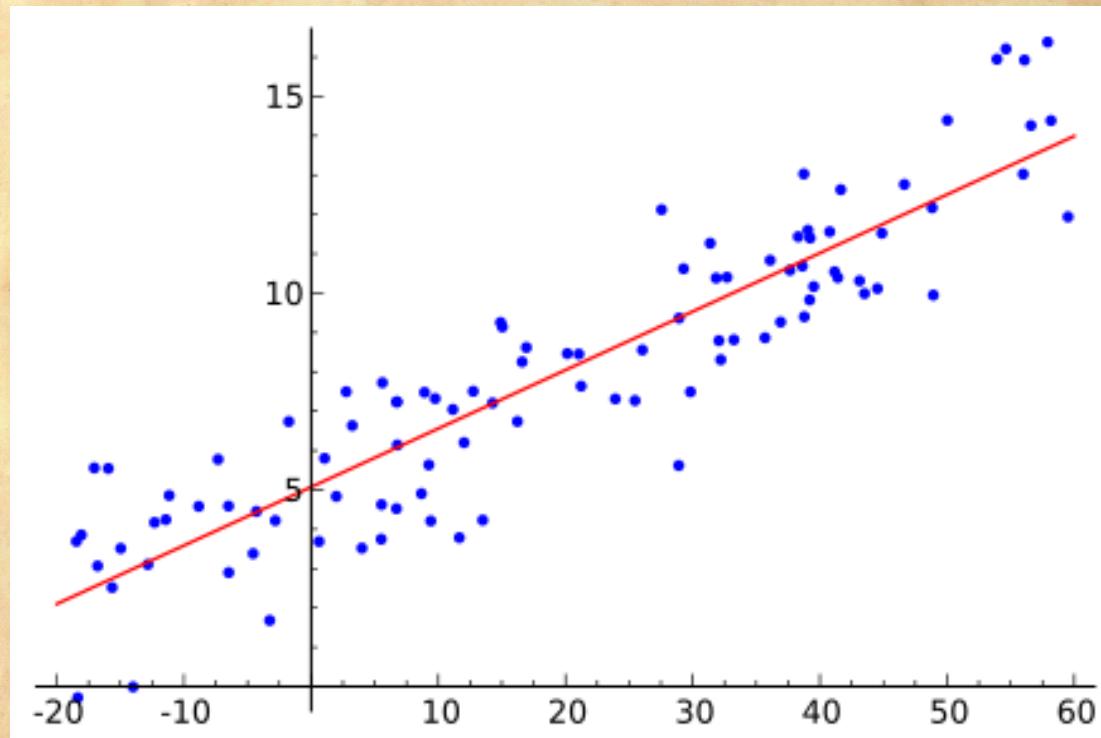
πᾶσαν ἐν τῷ Νοσοχόλειῳ οἰνούχου πεντηκόντην οἰδαπεντακόντην μῆνα  
πάσαν ἐν τῷ Νοσοχόλειῳ οἰνούχου πεντηκόντην οἰδαπεντακόντην μῆνα

Ο παρών νόμος. Σηματίζεις ὑπὸ τῆς Λειχής Συγέλευσεως  
Ο παρών νόμος, Σηματίζεις ὑπὸ τῆς Λειχής Συγέλευσεως

Ψηφίσαμεν οιοσδήποτε μετά τῆς τελευταίας Λειχής τοῦ  
Πηγαδανεύοι οιοσδήποτε μετά τῆς τελευταίας Λειχής τοῦ

τού τὴς Εγκαίσις αὐτῆς, οτε ενεργοῦνται τὰ τοῦ νόμου πεδίῳ μετο  
τού τὴς Εγκαίσις αὐτῆς, οτε ενεργοῦνται τὰ τοῦ νόμου πεδίῳ μετο

# Skew Correction - 1



[Linear Regression Wikipedia]

$$\min \sum_{i=1}^n (y_i - ax_i - b)^2$$

## *Skew Correction- 2*

Χριστόπουλον, δοτις ἔγραψε τὴν ἀνακριτικὴν μου κατάθεσιν ἐν

Χριστόπουλον, δοτις ἔγραψε τὴν ἀνακριτικὴν μου κατάθεσιν ἐν

Χριστόπουλον, δοτις ἔγραψε τὴν ἀνακριτικὴν μου κατάθεσιν ἐν



# 2nd stage Printed

*Synthesize( string query qs, glyph collection G, Average letter space distance d)*

*W=0*

*for each character c in query qs:*

*assign a glyph g to c*

*gbb = BoundingBox(g)*

*W+=gbb.Width+d*

*Create an Empty Image qI [2 \*max(gbb.Height)][ W ][0]*

*qI.BaseLine=qs[0].Glyph.BaseLine*

*col=0*

*for each character c in query qs:*

*Find the vertical shift sh of c.gbb so that gbb.BaseLine=qI.BaseLine*

*Juxtapose gbb to qI at row qI.BaseLine - sh , column col*

*col+=gbb.Width+d*

*Return BoundingBox(qI)*

# *Results*

Διέδοση τριά

Συνέταξη σήμα

Όπεια ή

Επιτροπή

Τρίπλευα

Σάμος

Νοέμβριος

Λιμένι

Νησώ

Βεστήν

Εφορεία

Καρβούνια

Βαθός

Έπος

Σεβτήνος

Επεξεργασία Εκόνας και Έπολογιστική Όραση

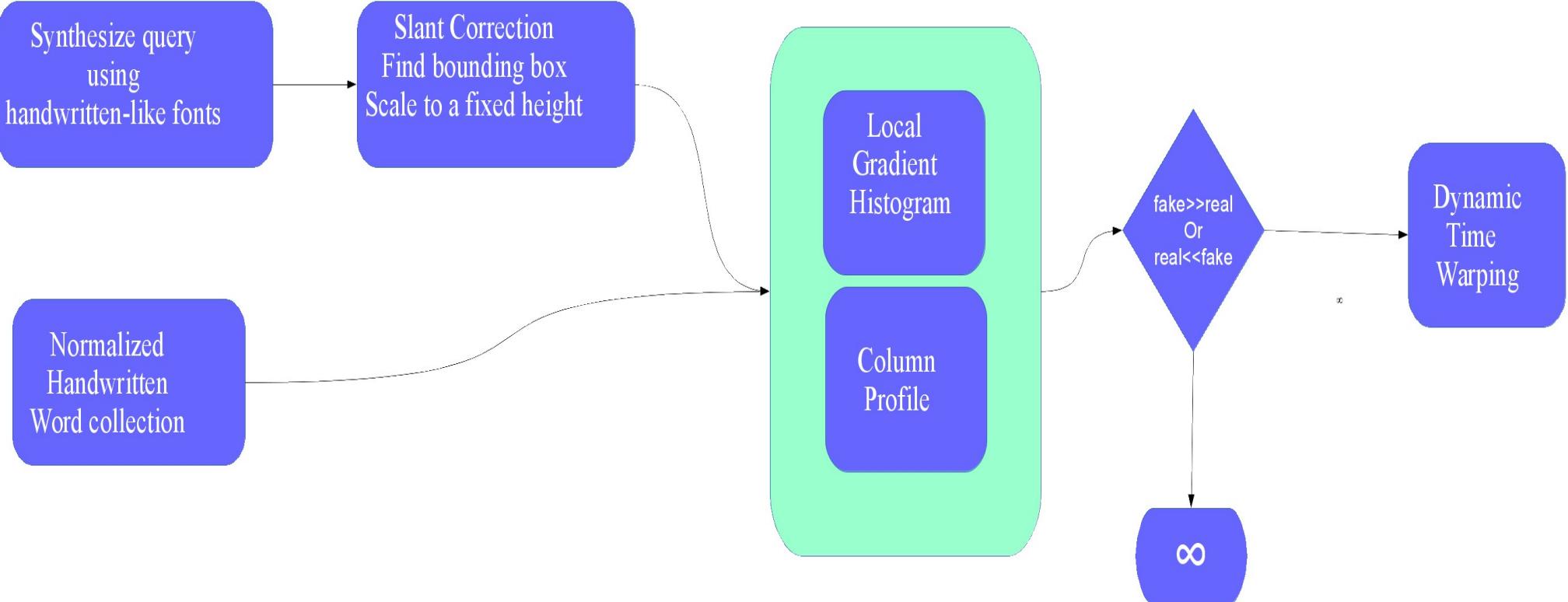
Handwritten...

inclosing  
a

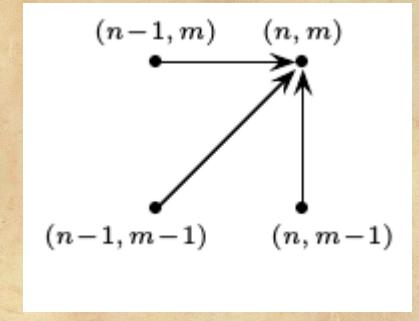
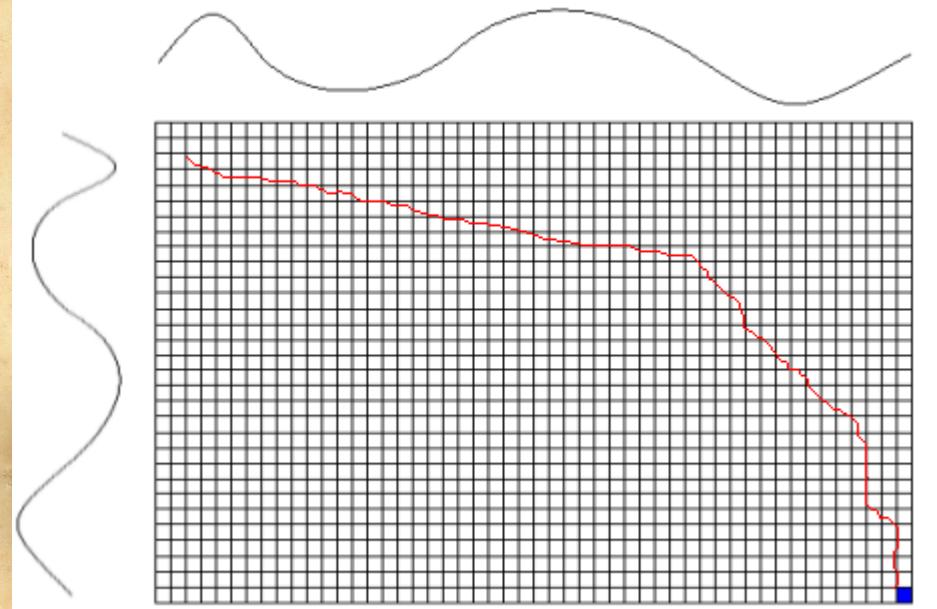
Nicholas

Instructions

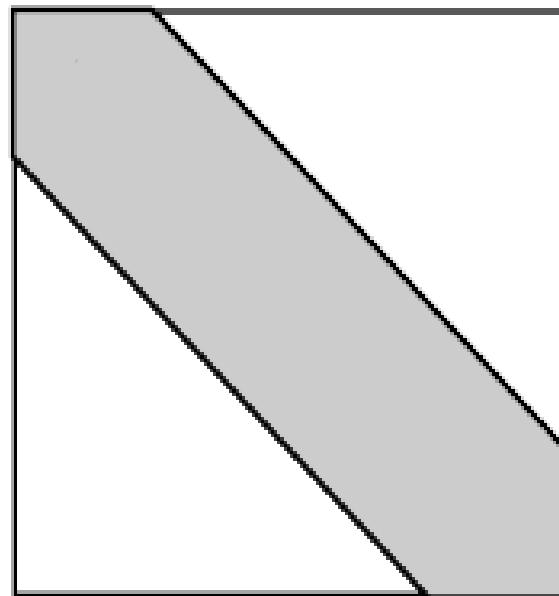
# *Search in handwritten collection*



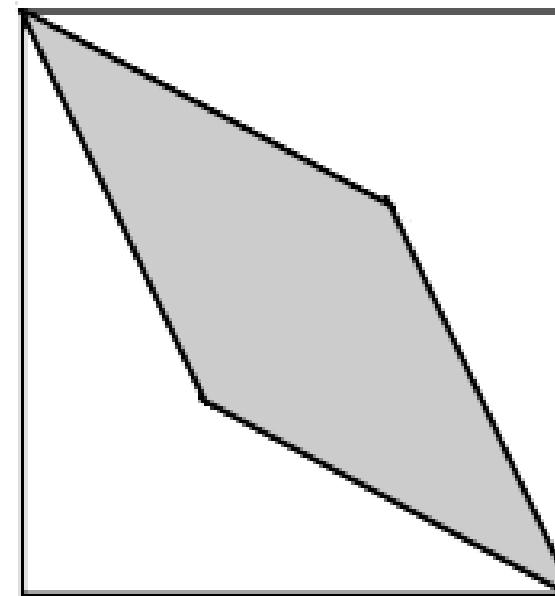
# Dynamic Time Warping



Constraints[Müller]



Sakoe Chiba[Müller]

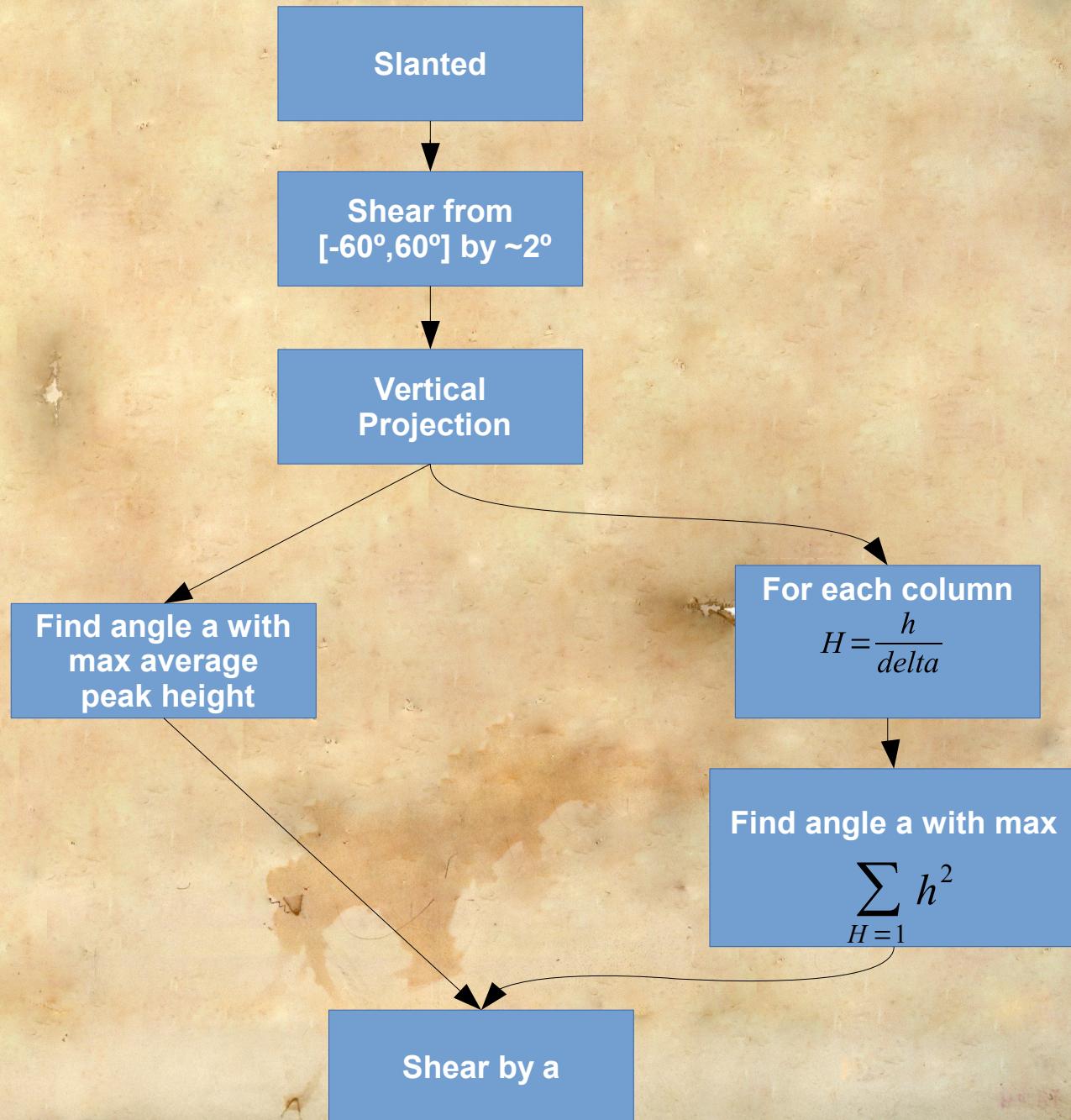


Itakura[Müller]

# Fonts

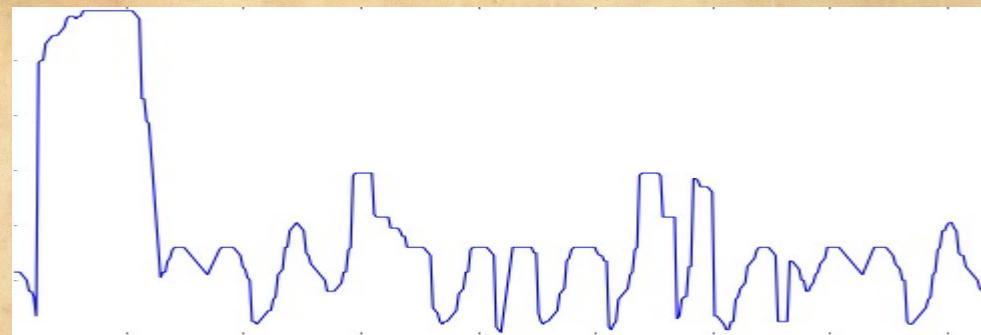
Abecedario	<i>cross the road</i>
<i>Adine Kumboig Script</i>	aquadfont
<i>Ataturk</i>	Breip
<i>Delphine</i>	Digital Handwriting
<i>Elegante</i>	<i>Exmouth</i>
<i>Freebooter</i>	<i>Ginette</i>
<i>Halohand Letter</i>	<i>Kary Morgan Hand</i>
<i>Hoony Script</i>	<i>Jane Austen</i>
<i>Jennifer Lynne</i>	<i>LaurenScript</i>
<i>Lucida Handwriting</i>	<i>Notera</i>
<i>Peciva</i>	<i>Simplesnails</i>
<i>Steve</i>	<i>Taggetes</i>
<i>Tex Gyre Chorus</i>	Times New Roman
<i>URW Chancery L</i>	Vavont
<i>Windsong</i>	

# Slant Correction



# *DTW-Top Profile based*

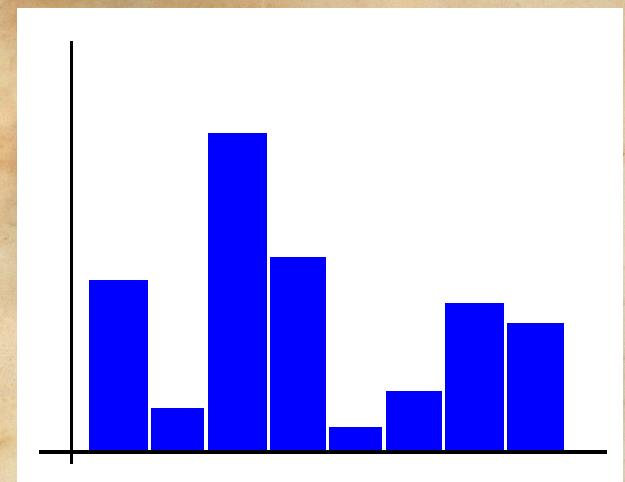
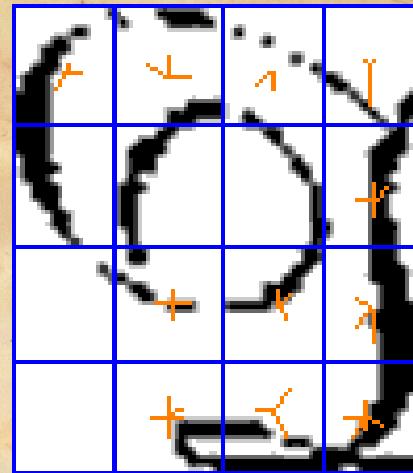
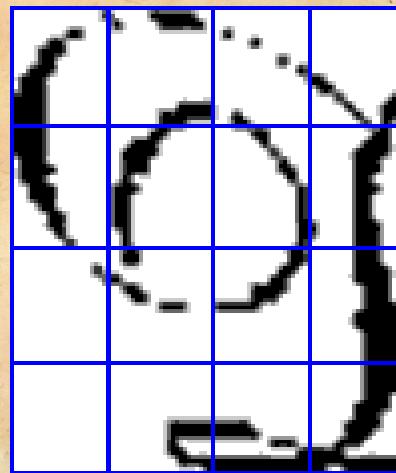
Instructions



# *DTW-LGH based*



December  
De Dec Decem mil nbe ber



Parameters	Score
Prunning:0.8 Band_Width:10	3,51%
Prunning:0.1 Band Type SC Band_Width:80	11,70%
Prunning:0.9 Band Type SC Band_Width:120	11,34%
Prunning:0.5 Band Type SC Band_Width:200	9,52%
Prunning:0.3 Band Type SC Band_Width:8	5,98%
Prunning:0.1 Band Type SC Band_Width:60	7,86%
Prunning:0.65 Band Type SC Band Width 50	4,02%
Prunning:0.65 Band Type ITK Band Width 10	6,71%
Prunning:0.65 Band Type ITK Band Width 100	0,10%
Prunning:0.65 Band Type SC Band Width 100	3,10%

Parameters	Score
Window width=64 Window overlap=0.9 Cells=4 Bins=5 SC Band 28 queries	12.92 %
Window width=90 Window overlap=0.75 Cells=2 Bins=6 SC Band 510 queries	0.8 %
Window width=90 Window overlap=0.75 Cells=2 Bins=6 SC Band 490 queries	0.96 %
Window width=50 Window overlap=0 Cells=2 Bins=6 SC Band 490 queries	0.74 %

## *Conclusions*

- Pleasing to the eye printed queries
- Poor DTW performance