A SIMPLE AND FAST WORD SPOTTING METHOD

Prof. Lior Wolf
The School of Computer Science
Tel Aviv University

Joint work with Alon Kovalchuk and Nachum Dershowitz

Images → Text

OCR is very hard for handwritten material

Instead:

- Approximate match
- **Word spotting**
- Transcript alignment

Word spotting

Given an image of a query word, identify all other occurrences of the same word in a set of input images

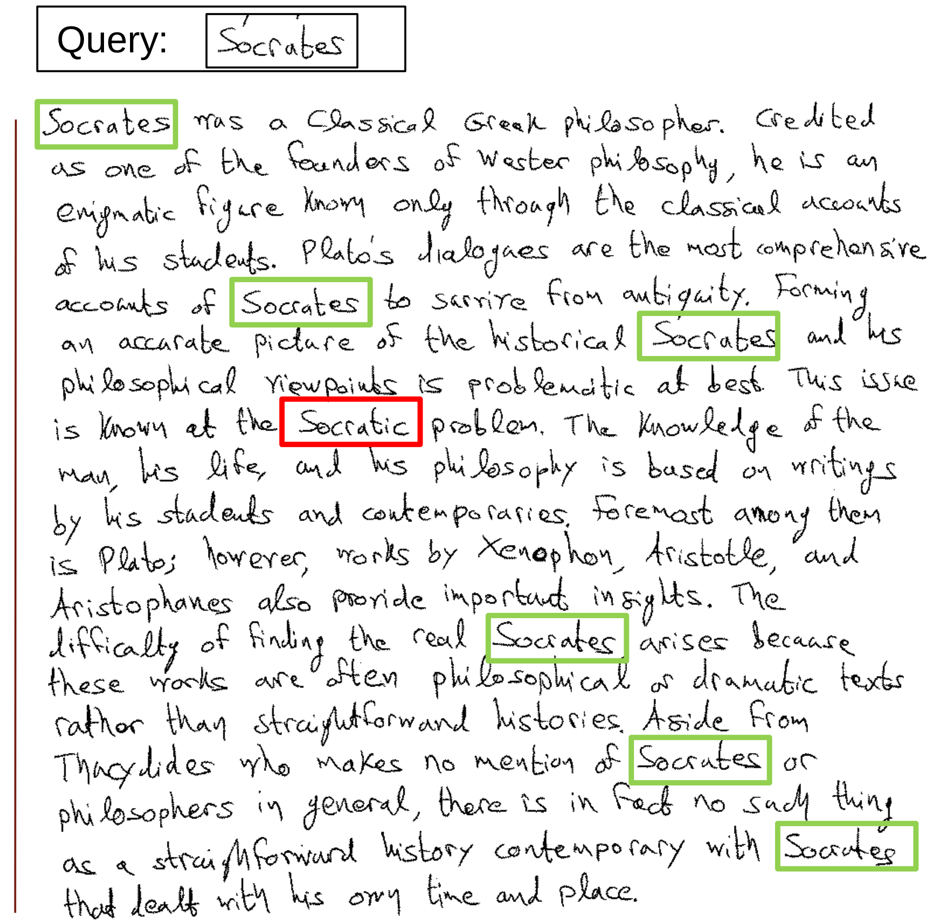
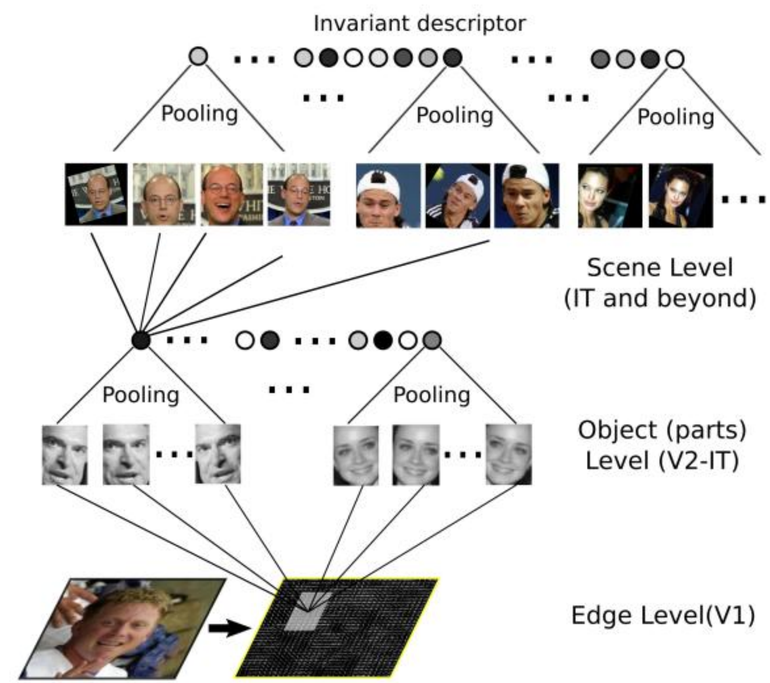


Image source: Handwritten Keyword Spotting Competition (H-KWS 2014)

Magic faces

Our system resembles the biologically plausible face-recognition architecture by Poggio

That system has a very simple architecture; however, it achieves excellent performance



Q. Liao Q, J.Z. Leibo, Y. Mroueh, T. Poggio. **Can a biologically-plausible hierarchy effectively replace face detection, alignment, and recognition pipelines?** arXiv 2014.

Magic faces

Max pooling per identity of stored faces

↑

Vectors are compared to stored vectors

↑

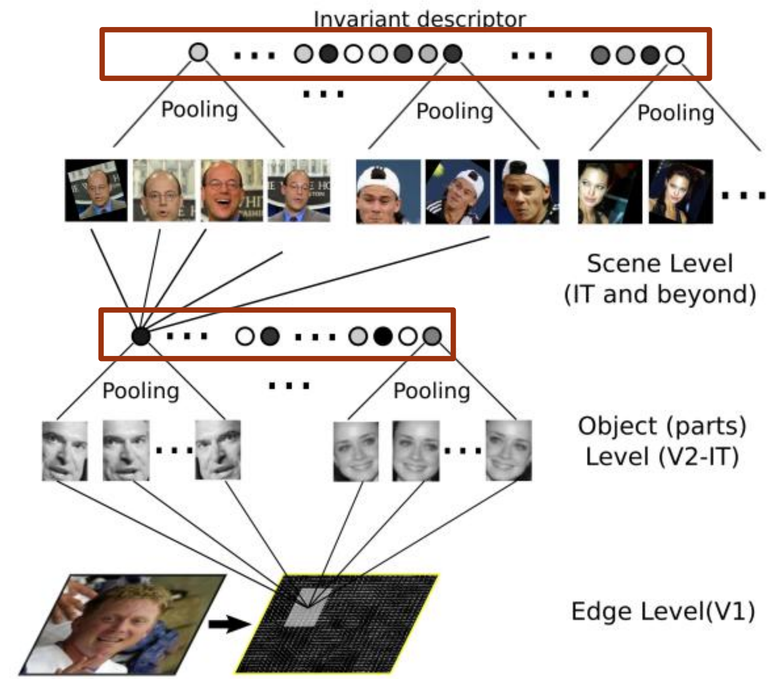
Max pooling per stored image rotation

↑

Patches are compared to stored faces

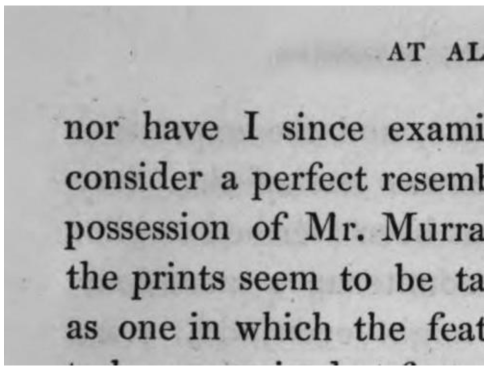
↑

All patches of the image are considered and encoded using HOG

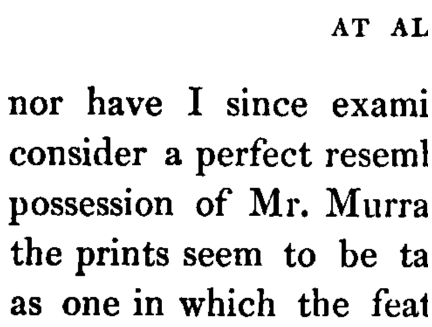


Binarization

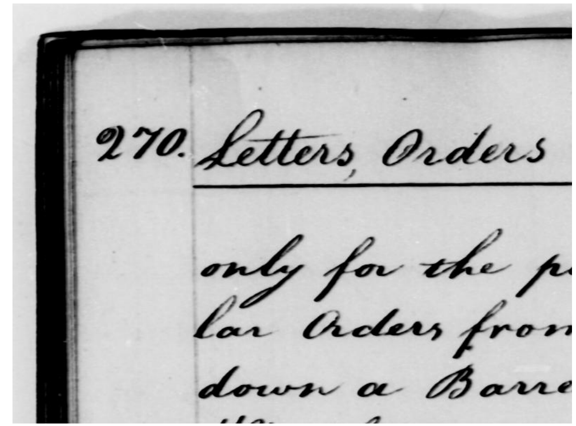
Simple threshold at 85% of mean pixel intensity


AT AL

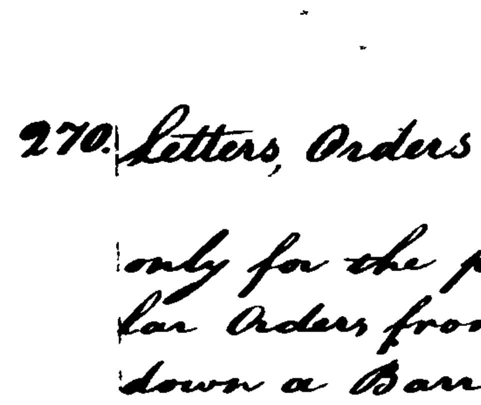
nor have I since exami
consider a perfect reseml
possession of Mr. Murra
the prints seem to be ta
as one in which the feat


AT AL

nor have I since exami
consider a perfect reseml
possession of Mr. Murra
the prints seem to be ta
as one in which the feat


270. Letters, Orders

only for the p
lar Orders from
down a Barre


270. Letters, Orders

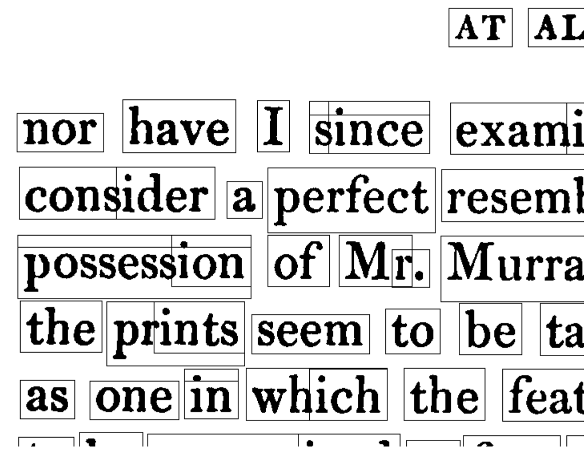
only for the p
lar Orders from
down a Barre

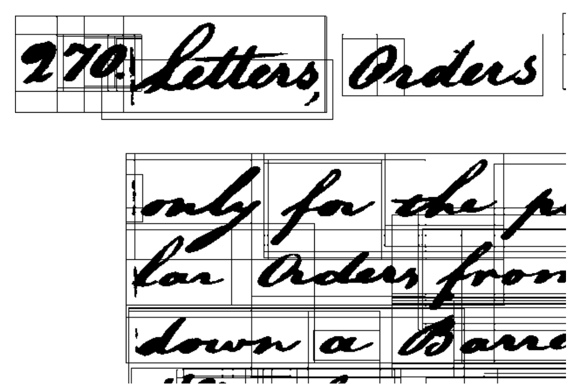
Candidate patch extraction

An elaborate heuristic based on the connected components extracted from the binary image.

Connected components are grouped to form clusters that are:

- “not too big”
- continuous in the x-axis
- components are aligned in y-axis
- contains all components in its bounding box



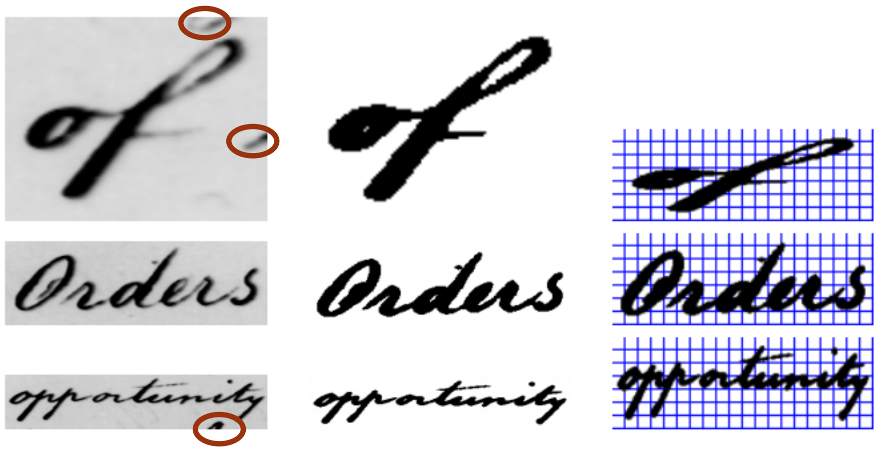
1:2

1:30

Patch normalization

A margin of 8 pixels is placed around the bounding box of each patch

The enlarged patches are then resized to a fixed size of 160 x 56 pixels

Grid of 20 x 7 cells, each 8 x 8 pixels



HOG + LBP encoding

Each cell is represented by:

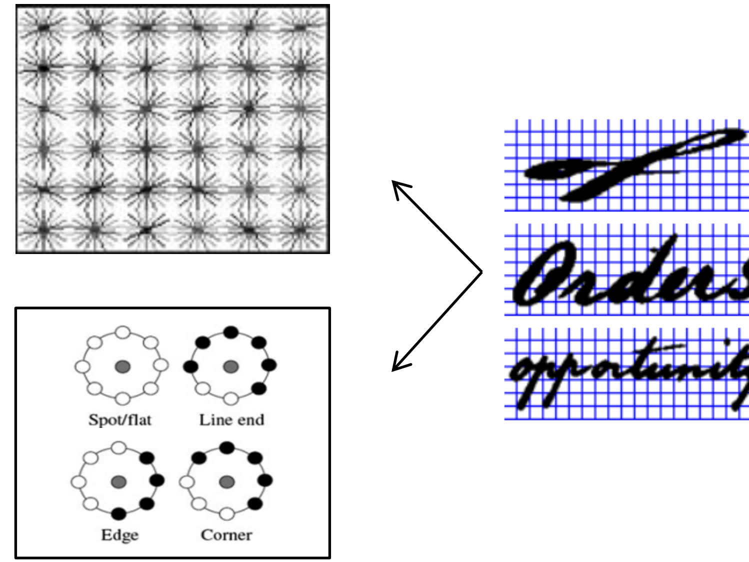
A HOG vector in R^{31}

An LBP vector in R^{58}

All HOG vectors are concatenated to one vector of size 20 x 7 x 31 = 4,340

Same for LBP 20 x 7 x 58 = 8,120

Both vectors are normalized and concatenated to a vector $v \in R^{12460}$



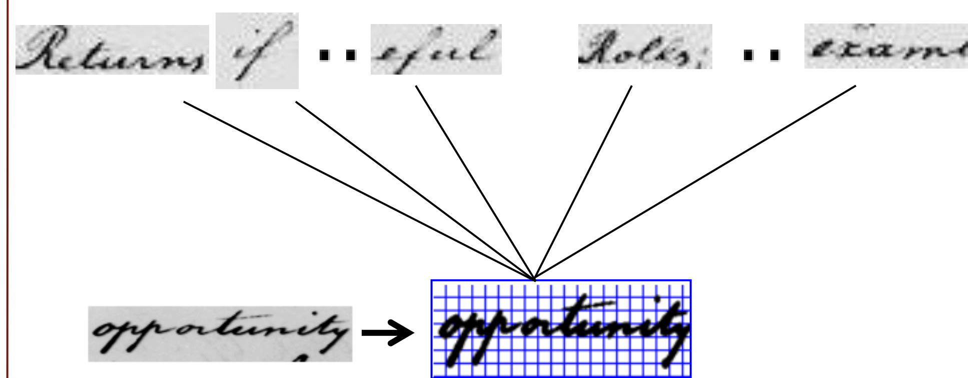
[HOG] N. Dalal and B. Triggs. **Histograms of Oriented Gradients for Human Detection**. CVPR, 2005
[LBP] T. Ojala, M. Pietikäinen, and D. Harwood. **Performance evaluation of texture measures with classification based on Kullback discrimination of distributions**. ICPR 1994.

Similarity to prototypes

3,750 random patches are sampled and split into 250 groups of 15 patches C_i, $i=1..250$

$M \in R^{12460 \times 3750}$ holds their hog+lbp vectors

A patch is first represented by its cosine similarity to the 3,750 prototypes $u=Mv$



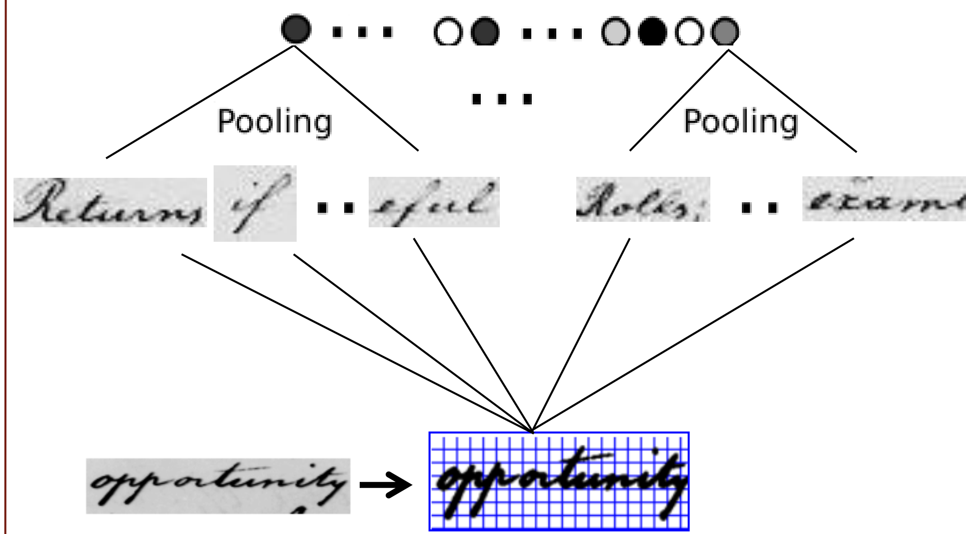
Max pooling

3,750 random patches are sampled and split into 250 groups of 15 patches $C_i, i = 1..250$

Next, max pooling is performed over the groups C_i, resulting in $w \in R^{250}$

$$w_i = \max_{j \in C_i} u_j$$

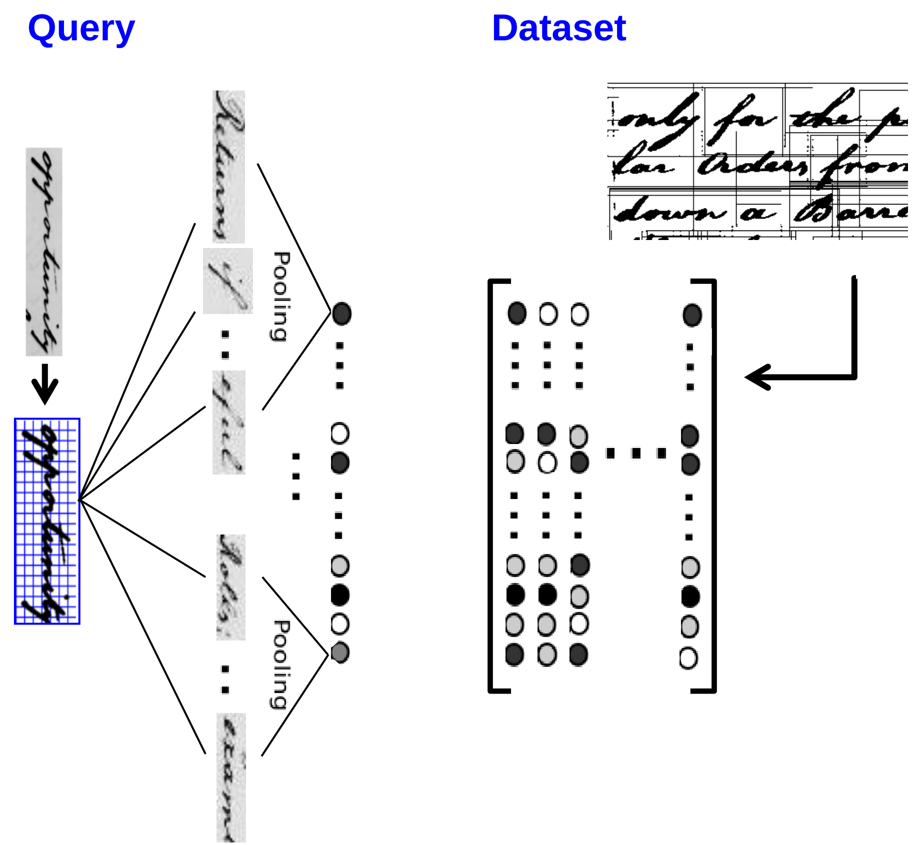
Unlike other deep network architectures pooling here is random!



Query

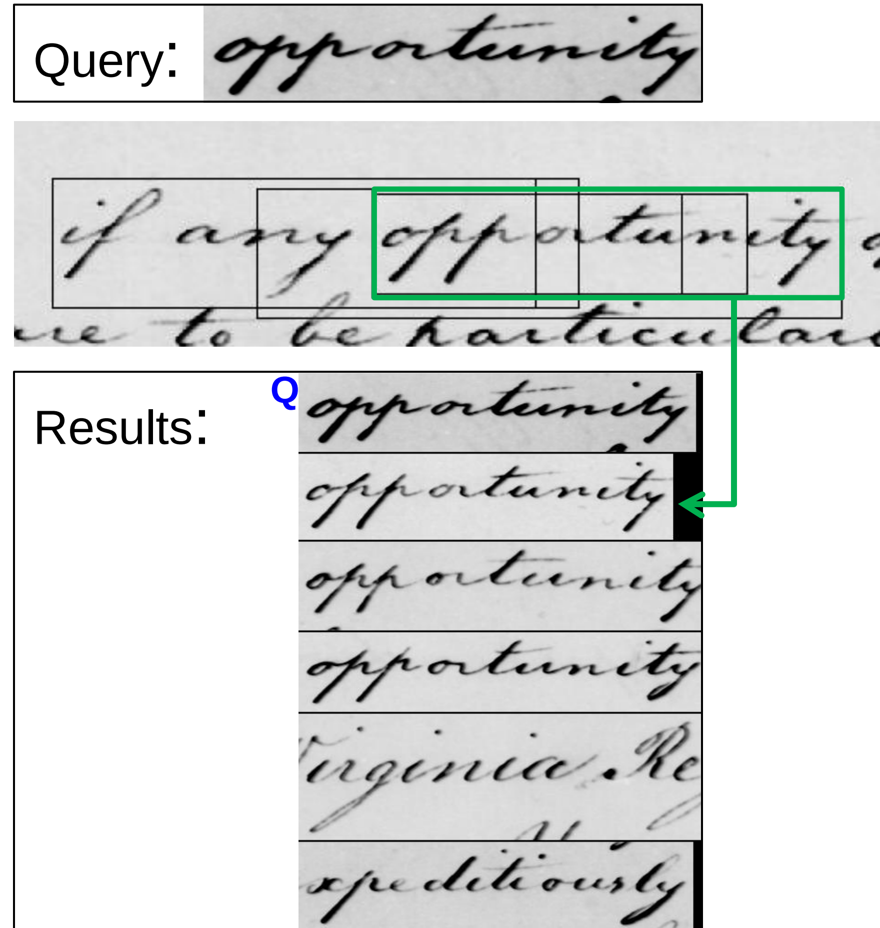
The same process is applied to the query patch and to all image patches

L2 similarity is used to retrieve the most similar patches

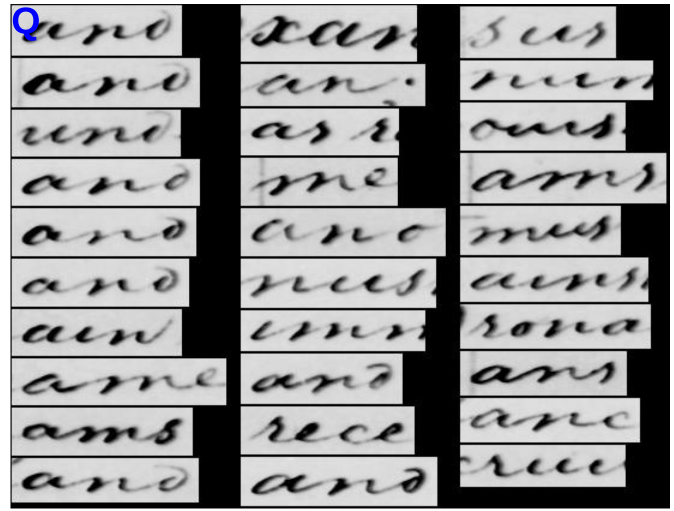


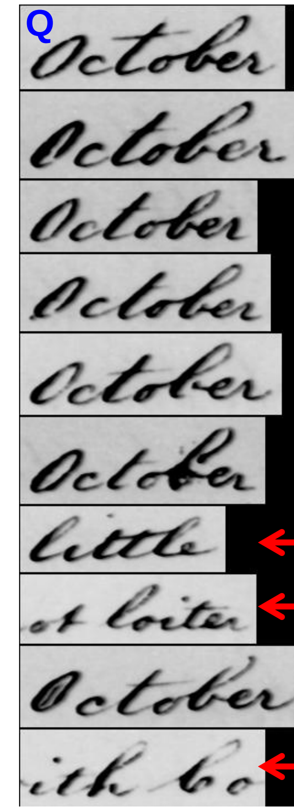
Overlap Suppression

Out of all patches that share a large connected component, only the one most similar to the query is retreived

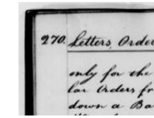


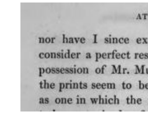
Results – examples





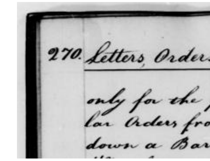
Results – accuracy

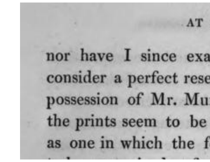




	Method	GW	LB
literature	"Browsing heterogeneous document collections by a segmentation-free word spotting method" M. Rusiñol, D. Aldavert, R. Toledo, J. Lladós, ICDAR, 2011	30.5%	42.8%
	"Efficient exemplar word spotting" J. Almazán, A. Gordo, A. Fornés, E. Valveny, BMVC, 2012	54.5%	85.5%
our system	Complete pipeline	50.1%	90.7%
	Same applied to segmented words	66.3%	92.9%
	Without max-pooling ()	48.8%	90.8%
	Without max-pooling, random subset ()	47.6%	90.7%

Results – run time

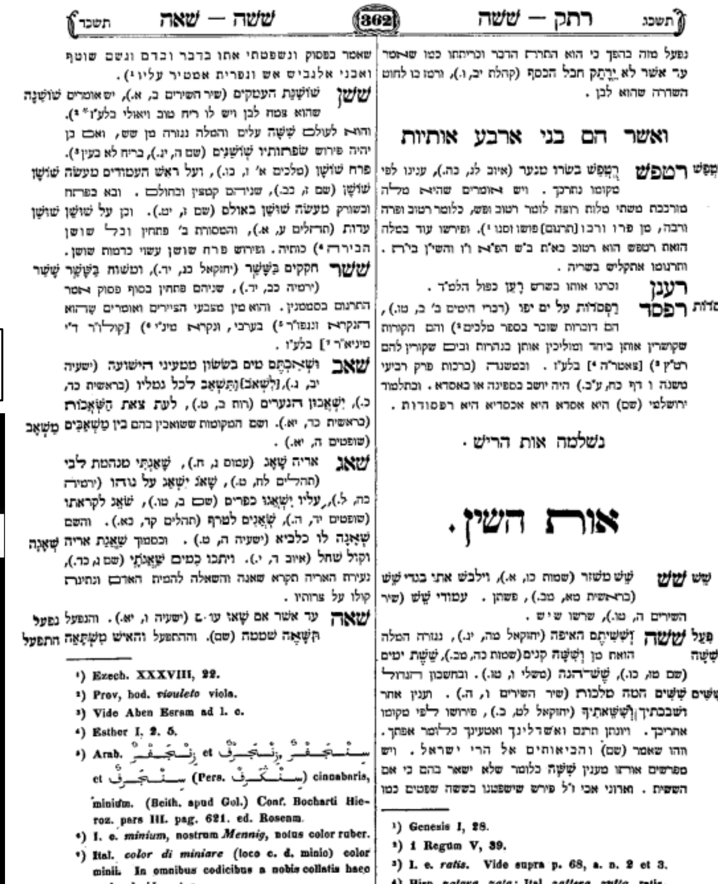

270 Letters Orders
only for the p
lar Orders fron
down a Barre


AT AL
nor have I since exami
consider a perfect resembl
possession of Mr. Murra
the prints seem to be ta
as one in which the feat

	Method	GW	LB
	Number of queries	4,860	4,988
"Efficient exemplar word spotting", Almazán et al. BMVC'12	All queries	5,058sec	4,159sec
	Average per query	1.04sec	0.83sec
Our system	All queries	158sec	46sec
	Average per query	0.03sec	0.01sec
	Single query	0.08sec	0.03sec
	Preprocessing per page	46sec	3sec
	Average memory per page	1,875KB	136KB

Planned code release

We plan to release our code soon to enable efficient off-the-shelf word spotting



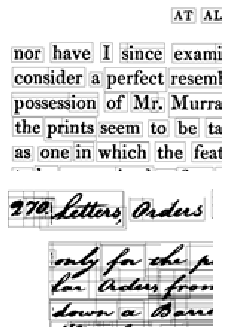
Summary

Candidate patch extraction

An elaborate heuristic based on the connected components extracted from the binary image.

Connected components are grouped to form clusters that are:

- "not too big"
- continious in the x-axis
- components are aligned in y-axis
- contains all components in it's bounding box

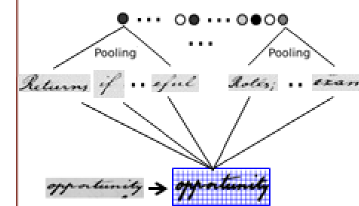


Network based representation

3,750 random patches are sampled and split into 250 groups of 15 patches $C_i, i = 1..250$

Then, the patch is represented by a vector $w \in R^{250}$ by pooling $w_i = \max_{j \in C_i} u_j$

Unlike other deep network architectures pooling is random!



HOG + LBP encoding

Each cell is represented by:

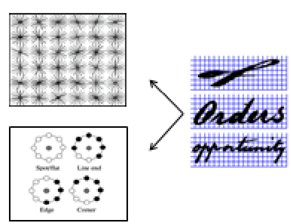
A HOG vector in R^{31}

An LBP vector in R^{58}

All HOG vectors are concatenated to one vector of size $20 \times 7 \times 31 = 4{,}340$

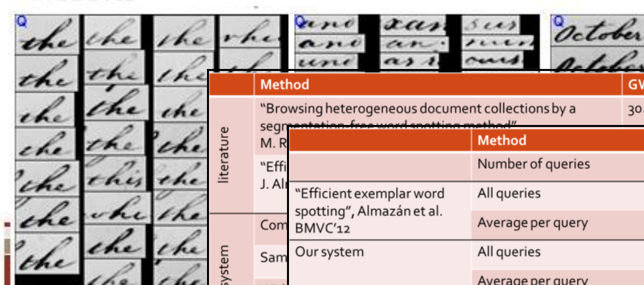
Same for LBP $20 \times 7 \times 58 = 8{,}140$

Both vectors are normalized and concatenated to a vector $v \in R^{12460}$



[HOG] N. Dalal and B. Triggs. **Histograms of Oriented Gradients for Human Detection**. CVPR, 2005
[LBP] T. Ojala, M. Pietikäinen, and D. Harwood. **Performance evaluation of texture measures with classification based on Kullback discrimination of distributions.** ICPR 1994.

Results

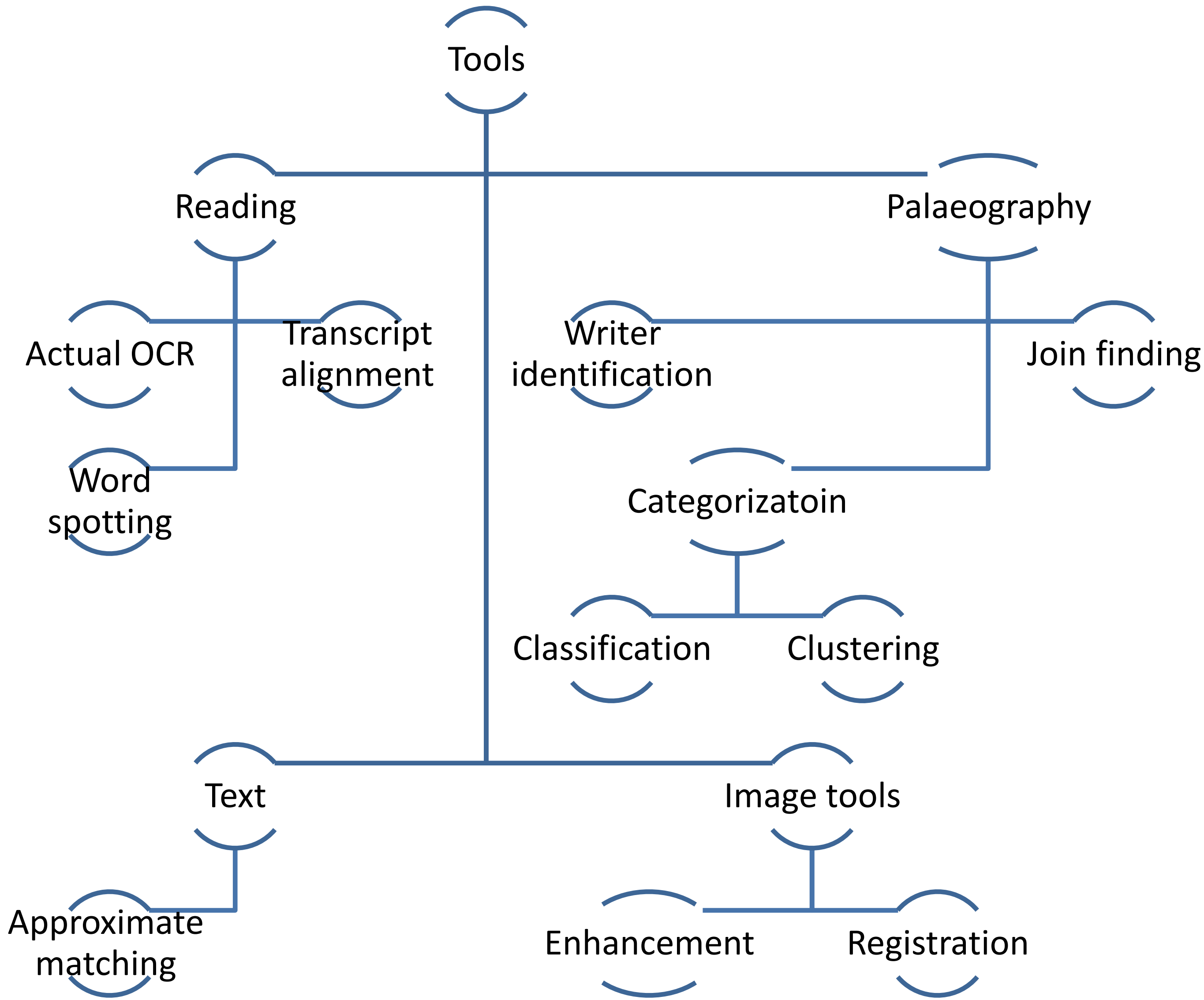


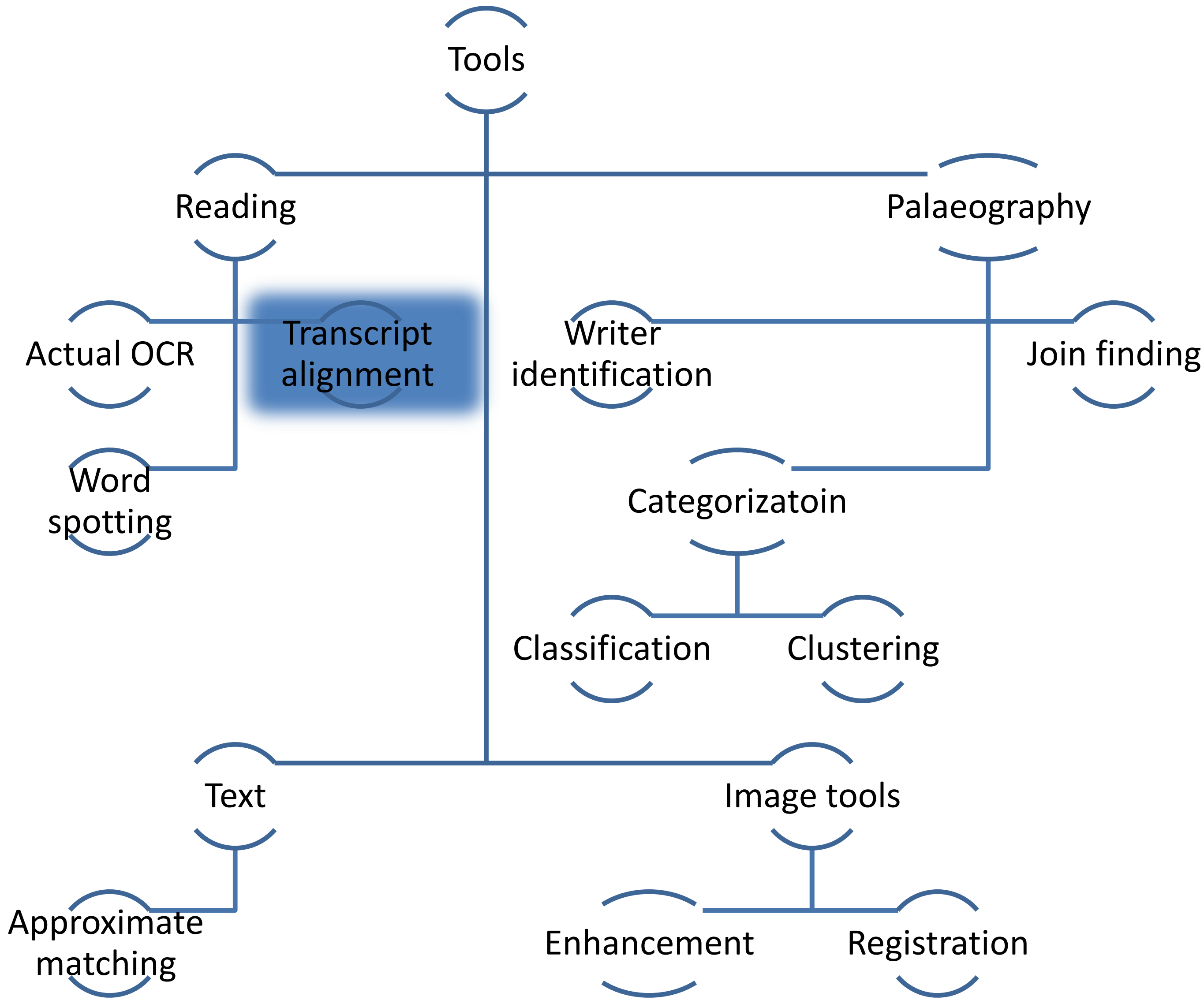
	Method	GW	LB
literature	"Browsing heterogeneous document collections by a segmentation-free word spotting method" M. R	30.5%	42.8%
	"Effi J. Al		
our system	Com		
	Sam		
	With		
	With		

	Method	GW	LB
	Number of queries	4,860	4,988
"Efficient exemplar word spotting", Almazán et al. BMVC'12	All queries	5,058sec	4,159sec
	Average per query	1.04sec	0.83sec
Our system	All queries	158sec	46sec
	Average per query	0.03sec	0.01sec
	Single query	0.08sec	0.03sec
	Preprocessing per page	46sec	3sec
	Average memory per page	1,875KB	136KB

thank

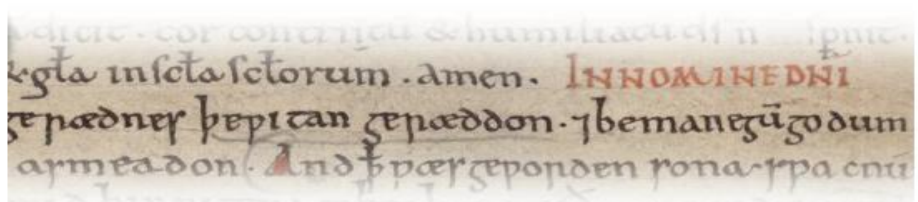
{you}


Tools
Reading
Palaeography
Actual OCR
Transcript alignment
Writer identification
Join finding
Word spotting
Categorizatoin
Classification
Clustering
Text
Image tools
Approximate matching
Enhancement
Registration


Tools
Reading
Palaeography
Actual OCR
Transcript alignment
Writer identification
Join finding
Word spotting
Categorizatoin
Classification
Clustering
Text
Image tools
Approximate matching
Enhancement
Registration

OCR Free Transcript Alignment

Image:



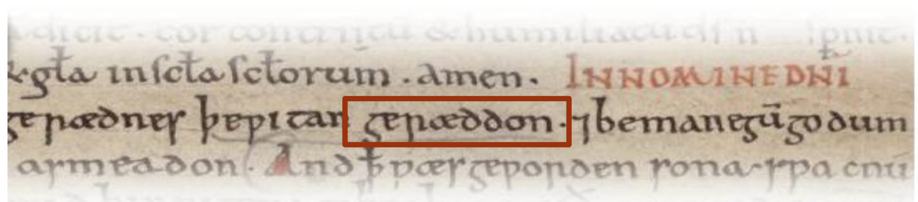
Transcription:

þe witan gerædon. 7 be manegum g

And þæt wæs geworden sona swa cnút

OCR Free Transcript Alignment

Image:

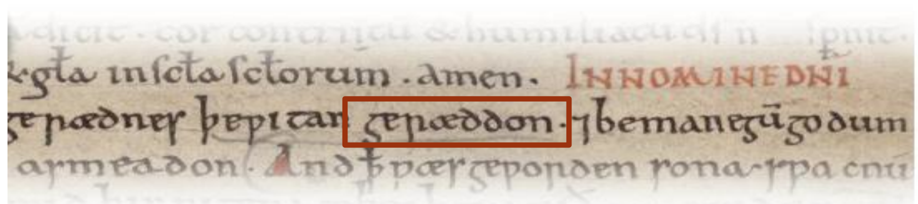


Transcription:

þe witan gerædon. 7 be manegum g
And þæt wæs geworden sona swa cnút

OCR Free Transcript Alignment

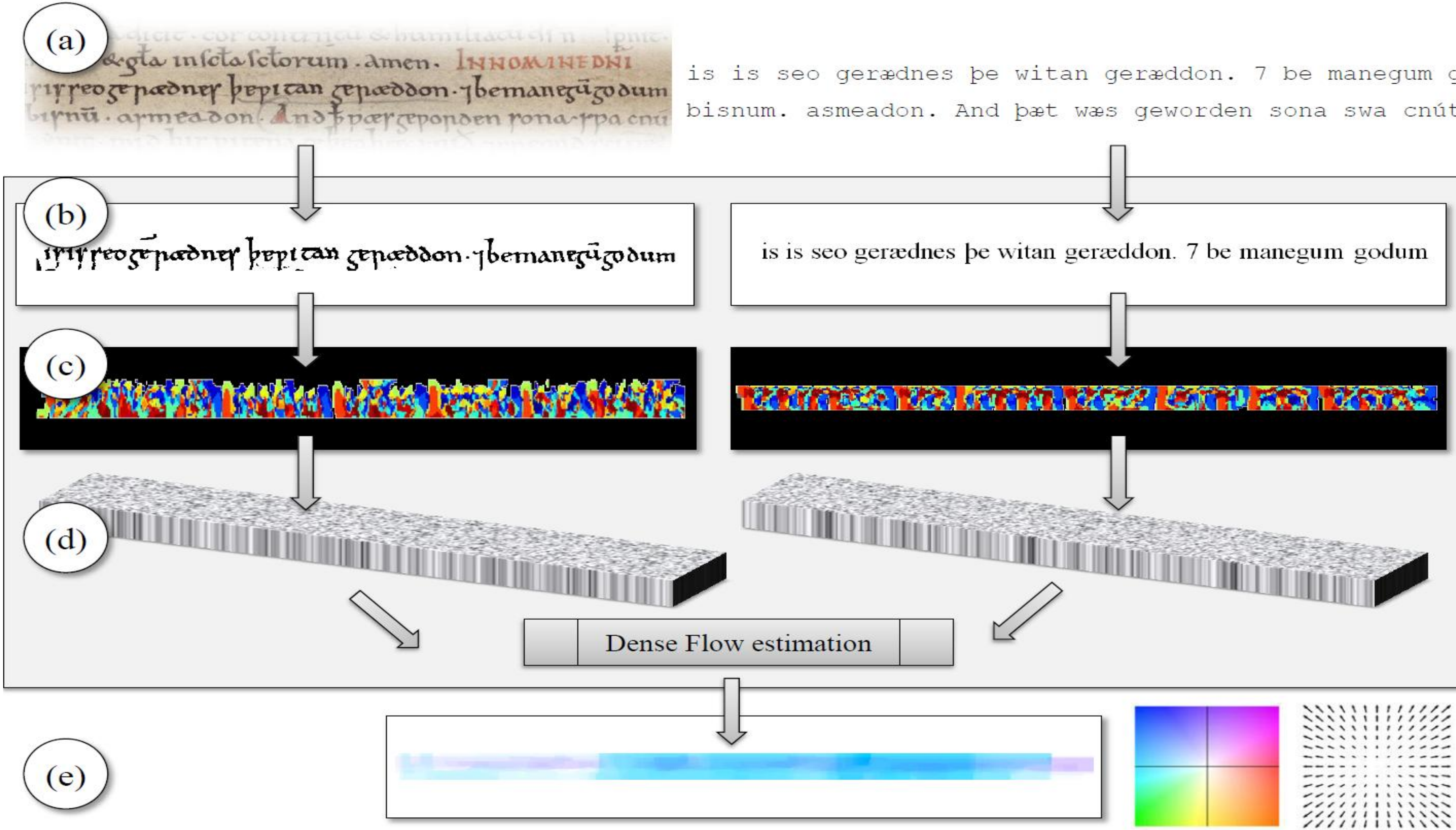
Image:



Transcription:

þe witan geræddon. 7 be manegum g
And þæt wæs geworden sona swa cnút

Motivation for alignment: the human user (UI); the computer user (word spotting)



T. Hassner, L. Wolf, and N. Dershowitz. **OCR-Free Transcript Alignment** *ICDAR,* 2013

Dead Sea Scrolls

ויתרוממו למעלה לכול. אתה הוא יהוה בחרתה באבותינו למקדם.

יתרוממו למעילה לכול. אתה הוא יהוה בחרתה באבותינו למקדם.

Genizah

אלהינו הוא ירחם עלינו וירויח לנו מצרותינו

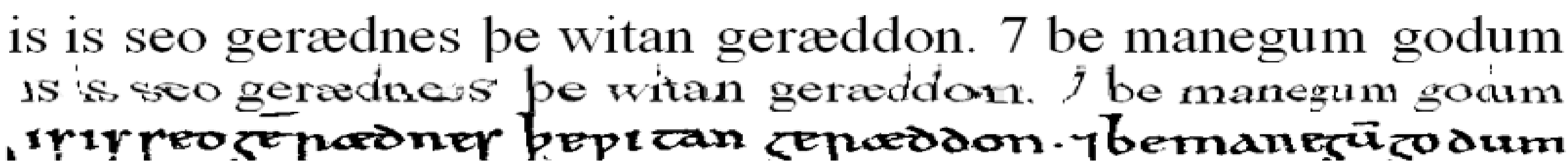
אלהינו הוא ירחם עלינו וירויח לנו מצרותינו
אלהינו הוא ירחם עלינו וירויח לנו מצרותינו

אסתגאת אלעזר הדא אלי וקתנא הדא לם יקדמה אחד אלי בית דין ולא שהד אחד בחצרה בית

אסתגאת אלעזר הדא אלי וקתנא הדא לם יקדמה אחד אלי בית דין ולא שהד אחד בחצרה בית
אסתגאת אלעזר הדא אלי וקתנא הדא לם יקדמה אחד אלי בית דין ולא שהד אחד בחצרה בית

Tibetan bKa gdams gsung bum

Cnut’s Oxford code



Codex Sinaiticus