

Handwritten Word Spotting

July 2015

Alicia Fornés
aforres@cvc.uab.es

VISUM Summer School



VISion Understanding and Machine intelligence

Contents

- Introduction
- Segmentation-based / Segmentation-free methods
- Learning-free / Learning-based methods
- Query-by-Example / Query-by-String
- Context aware Word Spotting
- Conclusions

INTRODUCTION

Word Spotting

Spotting is the task of locating a particular element without explicitly recognizing the content

Word spotting is the task of locating a particular keyword without explicitly transcribing the content



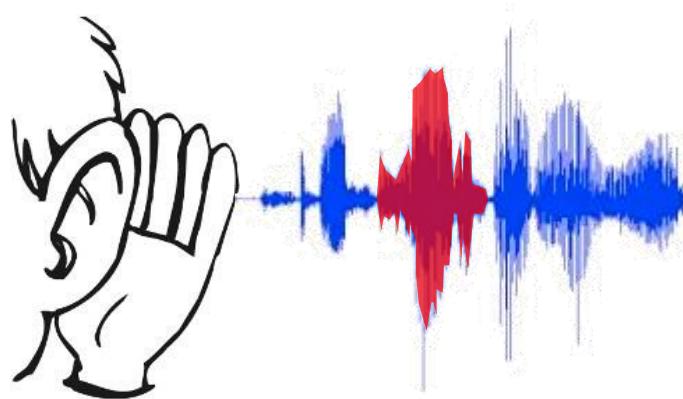
Spoken Word Spotting

Word Spotting was introduced within the speech recognition field in the 70s

Continuous one-dimensional signal

Unconstrained (words are not segmented, no grammar is considered in the sentence,...)

Dynamic Time Warping, Hidden Markov Models, ...



H. Sakoe and S. Chiba. "A Dynamic Programming Approach to Continuous Speech Recognition".
International Congress on Acoustics. 1971.

R.W. Christiansen and C.K. Rushforth. "Word Spotting in Continuous Speech using Linear Predictive Coding".
International Conference on Acoustics, Speech and Signal Processing. 1976

Typewritten Word Spotting

Then applied to typewritten documents in the early 90s

Two-dimensional signal that can be “easily” segmented into words
Character Signatures, Hidden Markov Models,...

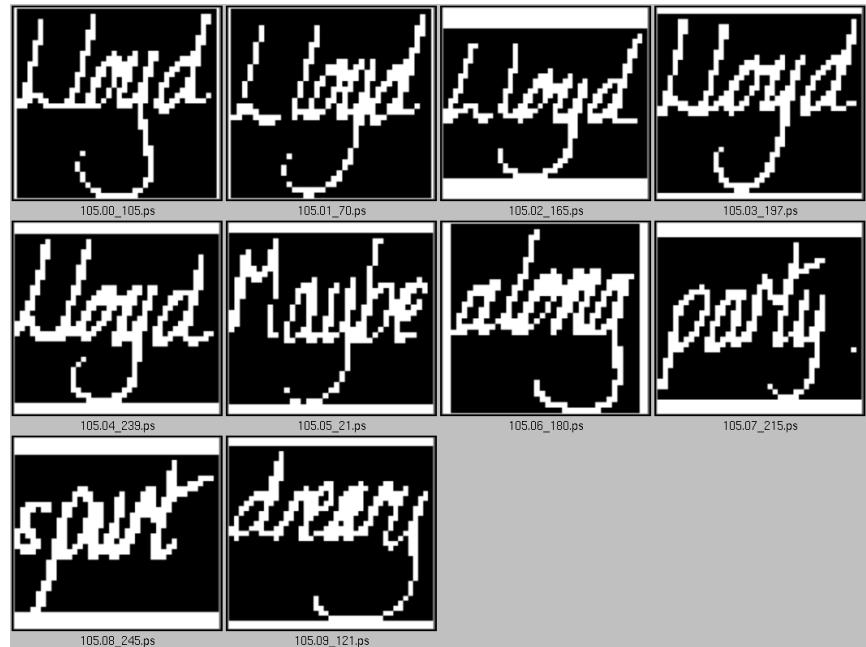


F.R. Chen et al. “Word Spotting in scanned images using hidden Markov models”. International Conference on Acoustics, Speech and Signal Processing. 1993

Handwritten Word Spotting

The first work on Handwritten Word Spotting was published in 1996...

Template matching between query and words in the DB allowing translation and affine transforms



R. Manmatha et al. "Word spotting: a new approach to indexing handwriting". In Proc. of CVPR, 1996.

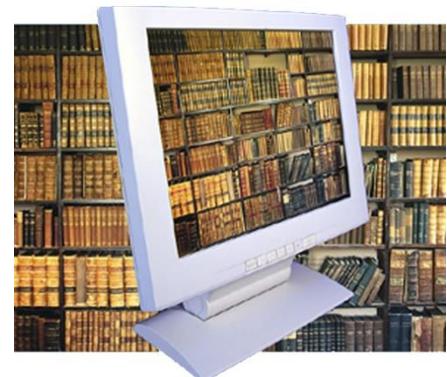
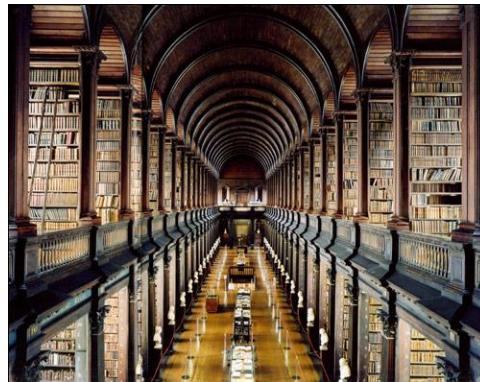
Introduction

Digitization in Archives

Preservation: Face the paper deterioration problem

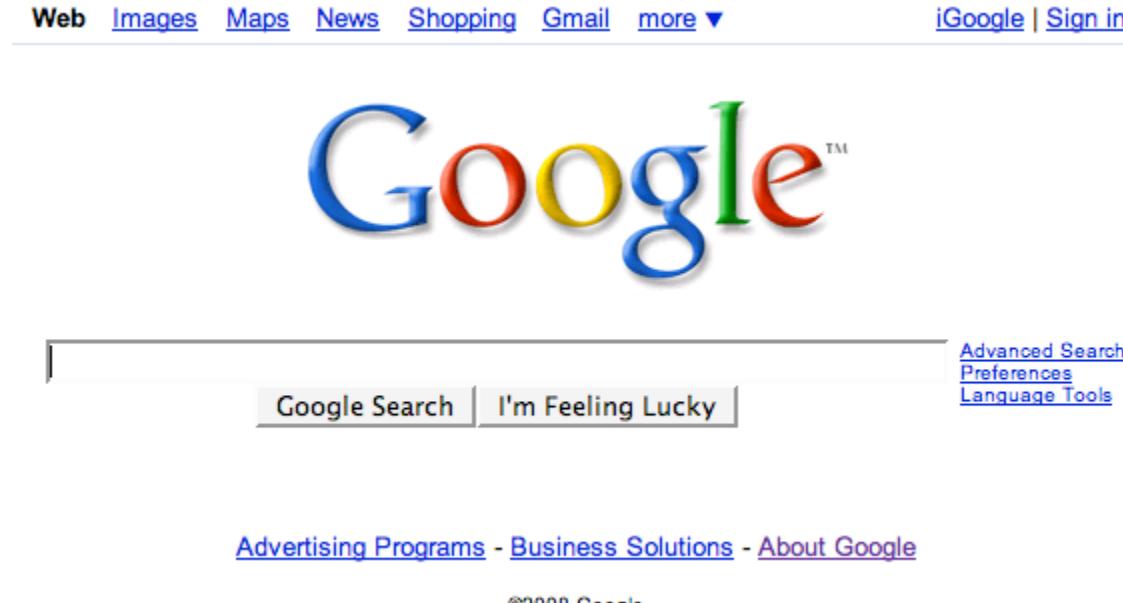
Storage: Avoid having kilometers of shelves

Accessibility: Allow users around the world access the cultural heritage



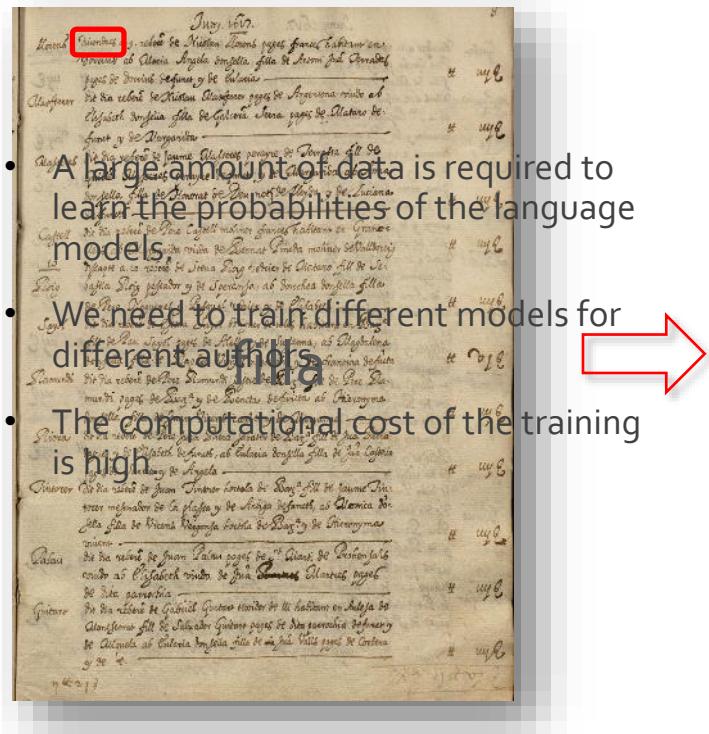
Introduction

Users would like to search information in document images like this!



Word Spotting Systems

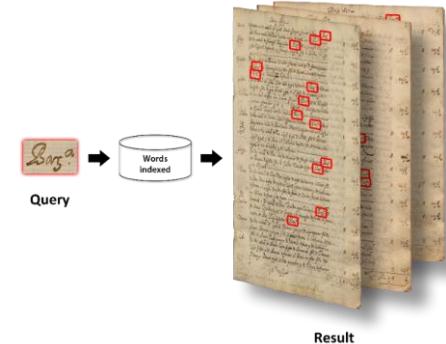
Strategy 1: Full transcription + Search the words in textual transcriptions



- A large amount of data is required to learn the probabilities of the language models
- We need to train different models for different authors
- The computational cost of the training is high

divendres a 9 rebere de Nicolau llorens pages frances habitant en dorrius ab Maria Angela
donsella **filla** de Antoni jua Terrades pages de dorrius defunct y de Eularia
dit dia rebere de Nicolau Masferrer pages de Argentona viudo ab Elisabeth donsell **filla** de
Galcera Serra pages de Mataro de funct y de Margarida
dit dia rebere de jaume Masseres perayre de Terrassa fill de guille Masseres perayre defunct y
de Margarida ab Anna donsell **filla** Honorat de Deu nog de Lleyda y de Luciana defunts
dit dia rebere de Pere castell moliner frances habitant en Grano llers ab Margarida viuda de
Bernat pineda moliner de Valldoreig
dissapte a 10 rebere de Steva Roig vedrier de Mataro fill de Se bastia Roig pescador y de
Speransa ab dorochea donsell **filla** de Pere Nogueres y Pasqual vedrier y de Elisabeth
dit dia rebere de Esteva Sayol botiguer de teles habitant en Barca fill de Pau Sayol pages de
Alella y de Susanna ab Magdalena donsell **filla** de Raphael Vinyes negociant y de francina
defucta
dit dia rebere de Pere Ramundi sastre de Barca fill de Pere ra mundi pages de Barca y de
Beneta defuncta ab Hieronyma donsell **filla** de Juan Vicens carder y de Maryanna
dit dia rebere de Pere juan Riera sabater de Barca fill de jua Riera daguer y de Elisabeth
defunts ab Eularia donsell **filla** de jua Castello pages de Sarria y de Angela
dit dia rebere de juan Tintorer hortola de Barca fill de jaume Tin torer mesurador de la plassa y
de Antiga defunts ab Monica do sella **filla** de Vicens Vergonsa hortola de Barca y de
Hieronyma vivent
dit dia rebere de juan Palau pages de St Marti de Prohensals viudo ab Elisabeth viuda de jua
xxxxx martres pages de dita parrochia
dit dia rebere de Gabriel Guitart texidor de Ili habitant en Aulesa de Montserrat fill de Salvador
Guitart pages de dita parrochia defunct y de Miquela ab Eularia donsell **filla** xxxx jua Valls
pages de Corbera y de t

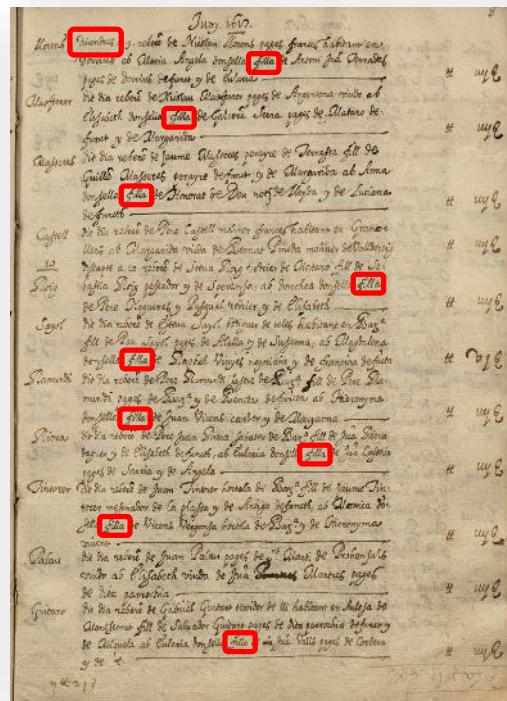
Word Spotting



Strategy 2: handwritten word spotting

- The recognition of words is formulated as a pattern recognition problem
- The transcription is not necessary because the search is done by similarity between the shapes.

silla



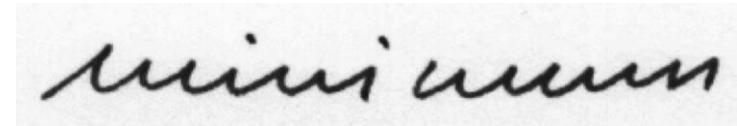
Difficulties

Difficulties:

- Different Handwriting styles



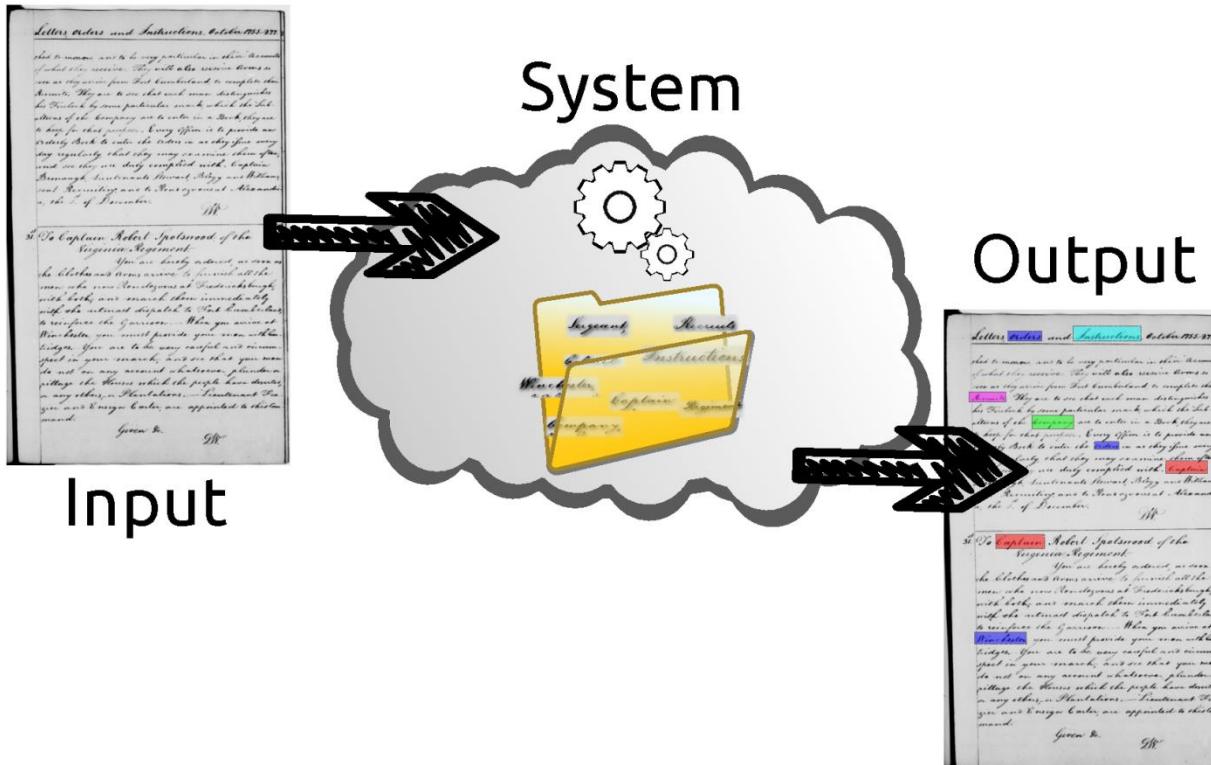
- Different writing instruments
- Large vocabulary
- Segmentation problem
chicken-egg



Handwritten Word Spotting

How to organize different families?

Retrieval application vs. Filtering application

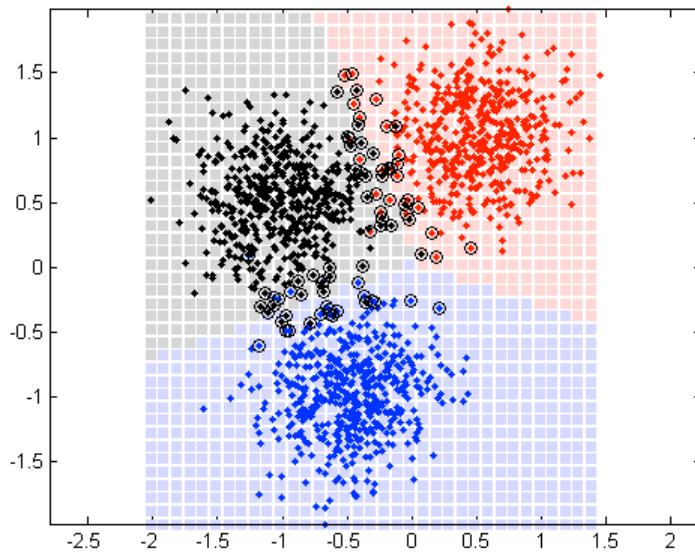


Handwritten Word Spotting

How to organize different families?

Retrieval application vs. **Filtering** application

Learning-free vs. **Learning-based** methods



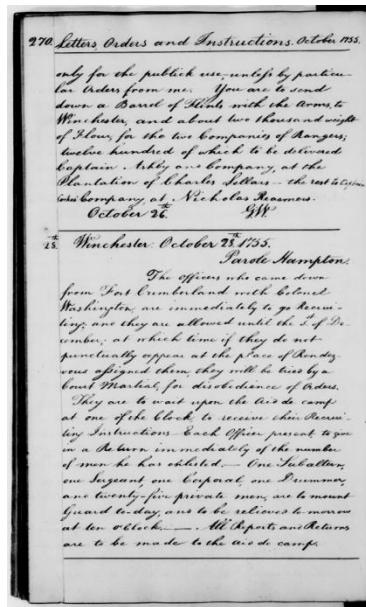
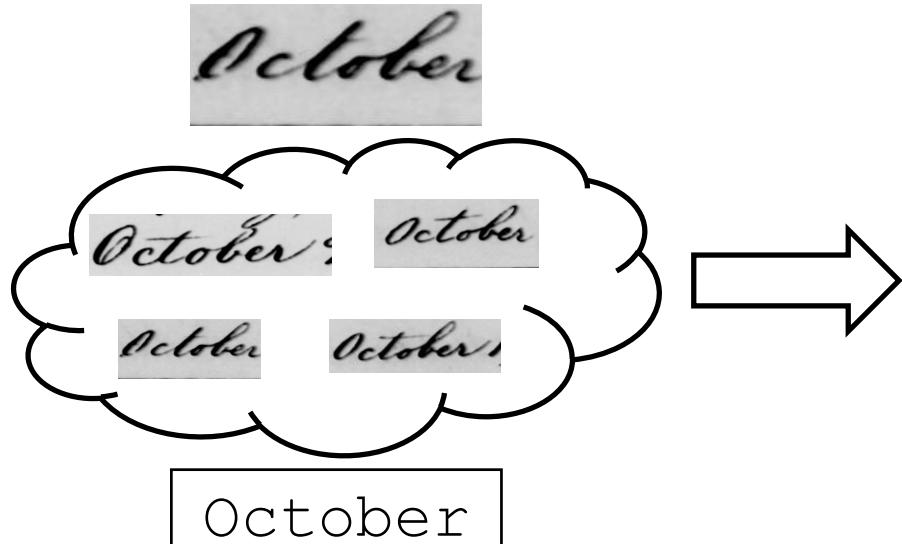
Handwritten Word Spotting

How to organize different families?

Retrieval application vs. Filtering application

Learning-free vs. Learning-based methods

Query-by-example vs. Query-by-class vs. Query-by-string



Handwritten Word Spotting

How to organize different families?

Retrieval application vs. **Filtering** application

Learning-free vs. **Learning-based** methods

Query-by-example vs. **Query-by-class** vs. **Query-by-string**

Segmentation-based vs. **Segmentation-free** methods



Families of HWS methods

- **Segmentation-free vs. Segmentation-based** methods
- **Learning-free vs. Learning-based** methods
- **Query-by-example vs. Query-by-string**

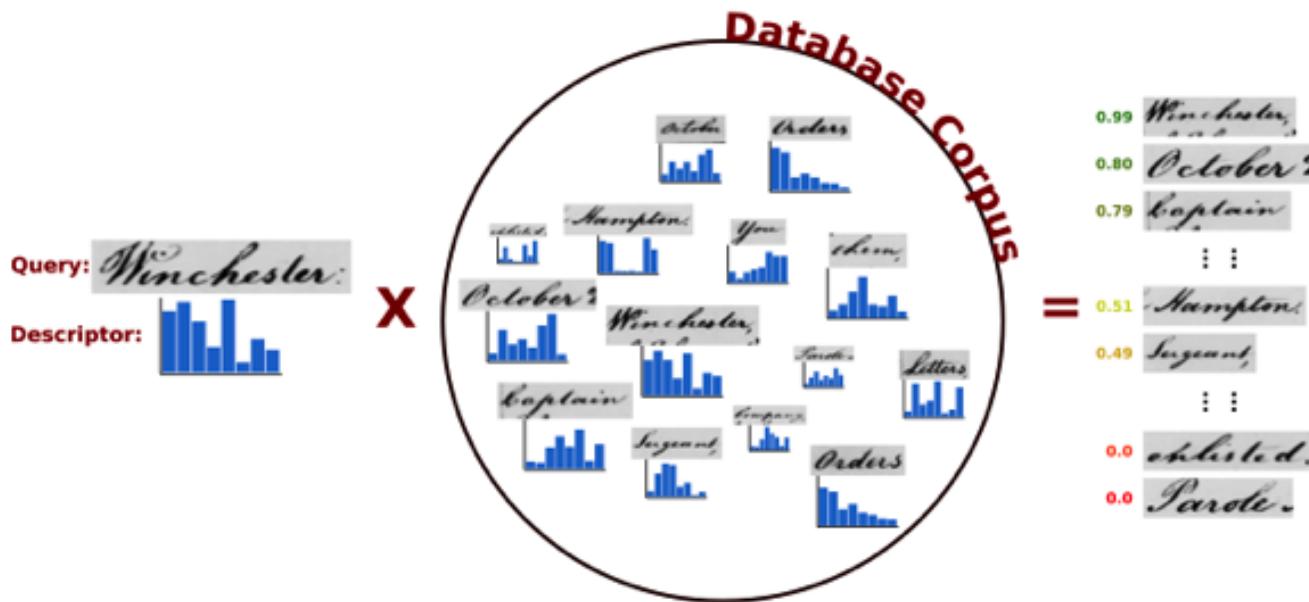
WORD SPOTTING LEARNING-FREE & SEGMENTATION-BASED

Classical Architecture

Layout analysis for word segmentation

Feature extraction: each word has a feature vector

Some distance is used for matching the query with the corpus



Classic Approach: Rath and Manmatha 2003

Break-through reference paper

Word Segmentation

Features: Profiles

Matching: Dynamic Time Warping

T.M. Rath, R. Manmatha: Word Image Matching Using Dynamic Time Warping. CVPR (2), pp. 521-527, 2003.

Classic Approach: Rath and Manmatha 2003

- Feature Extraction

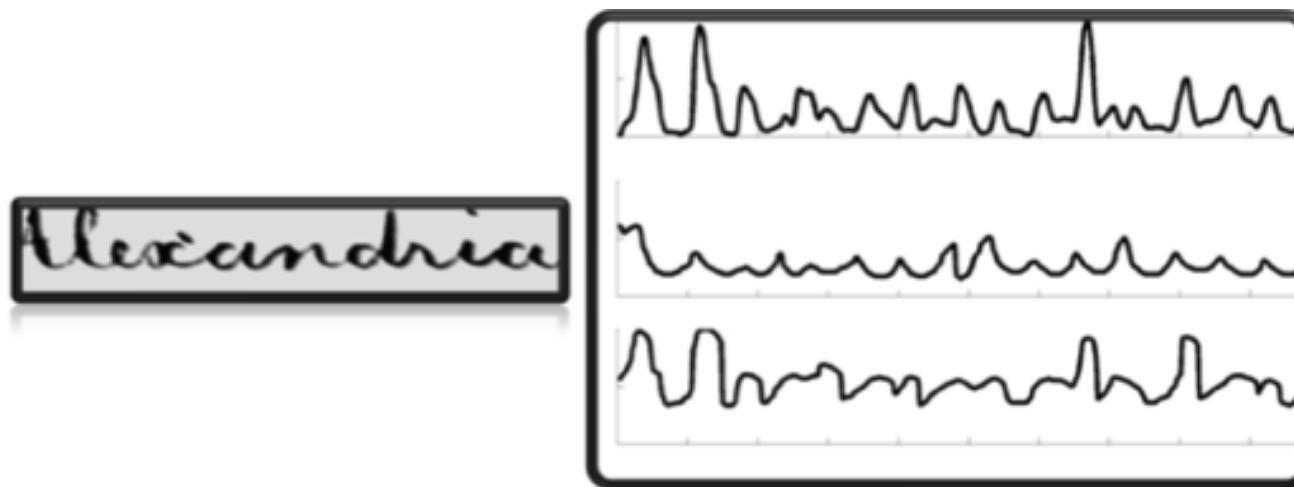
For each column of the word, extract 4 features

f1: upper profile

f2: lower profile

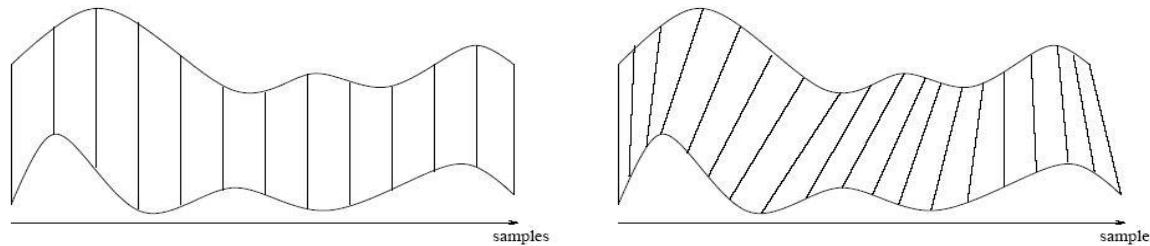
f3: number of foreground pixels

f4: number of transitions (gaps)



Classic Approach: Rath and Manmatha 2003

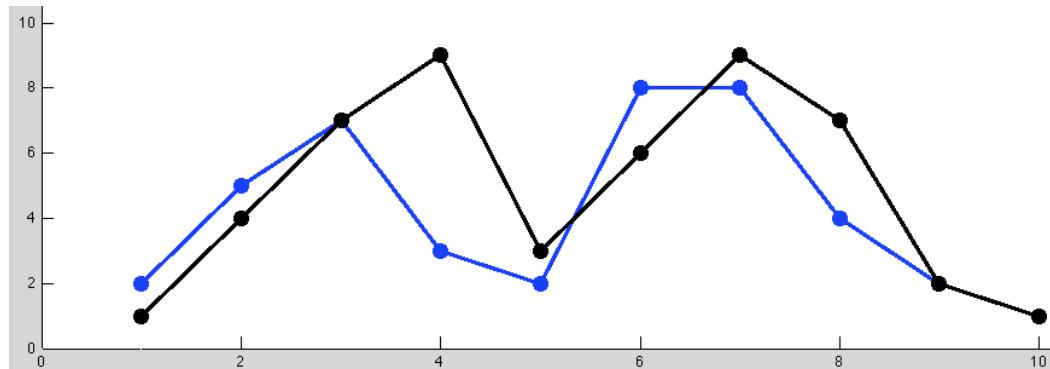
- Matching: Dynamic Time Warping (DTW)
 - DTW computes the distance between two time series optimizing the alignment
 - DTW can distort (warp) the time axis, compressing or expanding when necessary



DTW: Example with 1-D signals

Two 1-Dimensional signals

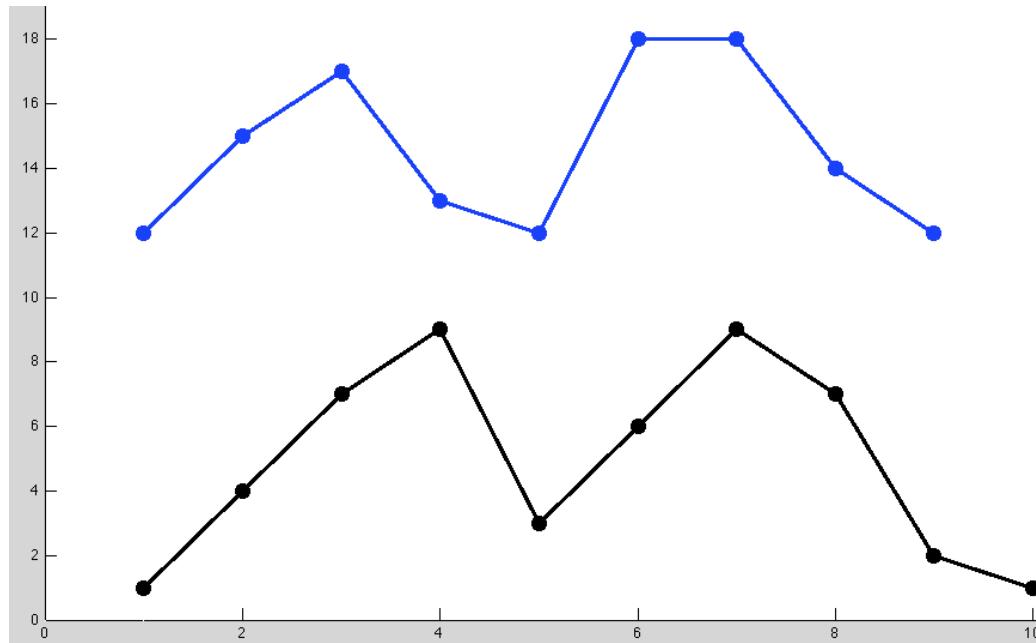
$$X = [2, 5, 7, 3, 2, 8, 8, 4, 2] \leftrightarrow Y = [1, 4, 7, 9, 3, 6, 9, 7, 2, 1]$$



DTW: Example with 1-D signals

Two 1-Dimensional signals

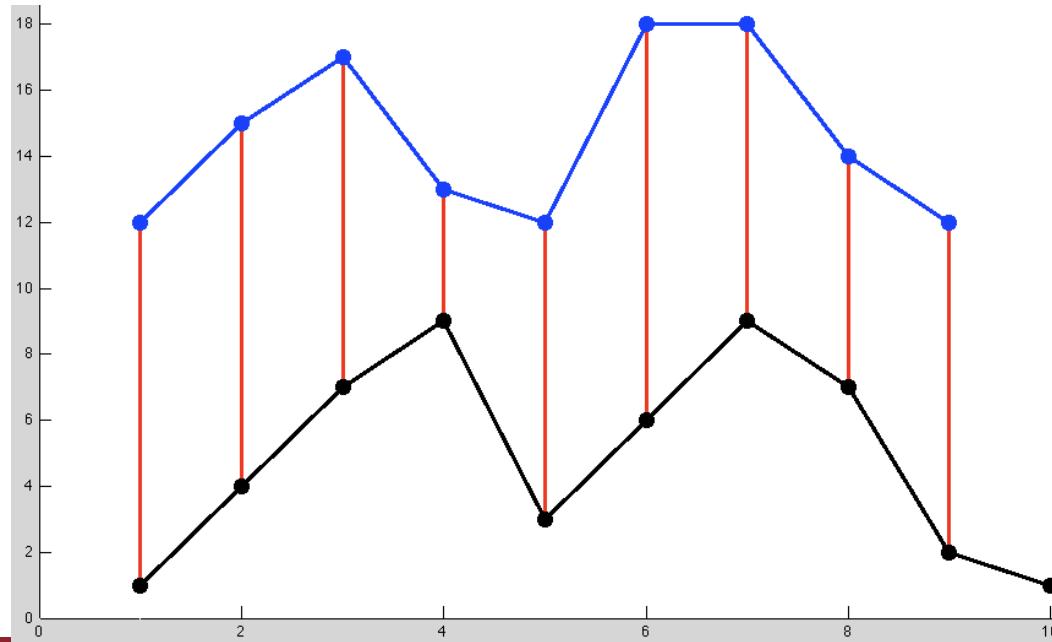
$$X = [2, 5, 7, 3, 2, 8, 8, 4, 2] \leftrightarrow Y = [1, 4, 7, 9, 3, 6, 9, 7, 2, 1]$$



DTW: Example with 1-D signals

Matching:

Euclidean distance → Undesired!

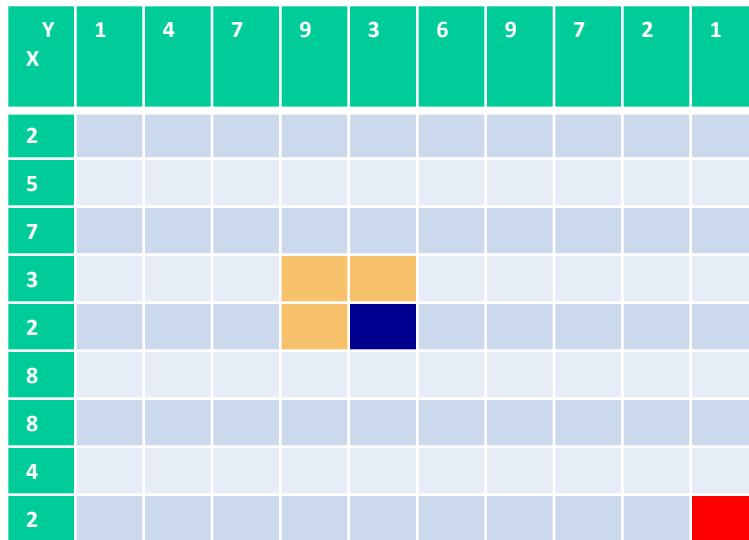


DTW: Example with 1-D signals

DTW is computed using Dynamic Programming

The last cell contains the total cost (distance)

The cost of a cell depends on 3 cells



$$D(i,j) = \min \left\{ \begin{array}{l} D(i,j-1) \\ D(i-1,j) \\ D(i-1,j-1) \end{array} \right\} + d(x_i, y_j)$$

DTW: Example with 1-D signals

| X | Y | 1 | 4 | 7 | 9 | 3 | 6 | 9 | 7 | 2 | 1 |
|---|-----|---|----|----|----|----|-----|-----|-----|-----|---|
| 2 | 1 | 5 | 30 | 79 | 80 | 96 | 145 | 170 | 170 | 171 | |
| 5 | 17 | | | | | | | | | | |
| 7 | 53 | | | | | | | | | | |
| 3 | 57 | | | | | | | | | | |
| 2 | 58 | | | | | | | | | | |
| 8 | 107 | | | | | | | | | | |
| 8 | 156 | | | | | | | | | | |
| 4 | 165 | | | | | | | | | | |
| 2 | 166 | | | | | | | | | | |

$$D(i,j) = \min \left\{ \begin{array}{l} D(i,j-1) \\ D(i-1,j) \\ D(i-1,j-1) \end{array} \right\} + d(x_i, y_j)$$

First row:

$$(1-2)^2 = 1$$

$$(4-2)^2 + 1 = 5$$

$$(7-2)^2 + 5 = 30$$

First column:

$$(5-1)^2 + 1 = 17$$

$$(7-1)^2 + 17 = 53$$

DTW: Example with 1-D signals

| X | Y | 1 | 4 | 7 | 9 | 3 | 6 | 9 | 7 | 2 | 1 |
|---|-----|---|----|----|----|----|-----|-----|-----|-----|---|
| 2 | 1 | 5 | 30 | 79 | 80 | 96 | 145 | 170 | 170 | 171 | |
| 5 | 17 | 2 | 6 | 22 | 26 | 27 | 43 | 47 | 56 | 72 | |
| 7 | 53 | | | | | | | | | | |
| 3 | 57 | | | | | | | | | | |
| 2 | 58 | | | | | | | | | | |
| 8 | 107 | | | | | | | | | | |
| 8 | 156 | | | | | | | | | | |
| 4 | 165 | | | | | | | | | | |
| 2 | 166 | | | | | | | | | | |

$$D(i,j) = \min \left\{ \begin{array}{l} D(i,j-1) \\ D(i-1,j) \\ D(i-1,j-1) \end{array} \right\} + d(x_i, y_j)$$

Second row:

$$(4-5)^2 + 1 = 2$$

$$(7-5)^2 + 2 = 6$$

$$(9-5)^2 + 6 = 22$$

...

DTW: Example with 1-D signals

| X | Y | 1 | 4 | 7 | 9 | 3 | 6 | 9 | 7 | 2 | 1 |
|---|-----|----|----|----|----|----|-----|-----|-----|-----|---|
| 2 | 1 | 5 | 30 | 79 | 80 | 96 | 145 | 170 | 170 | 171 | |
| 5 | 17 | 2 | 6 | 22 | 26 | 27 | 43 | 47 | 56 | 72 | |
| 7 | 53 | 11 | 2 | 6 | 22 | 23 | 27 | 27 | 52 | 88 | |
| 3 | 57 | 12 | 18 | 38 | 6 | 15 | 51 | 43 | 28 | 32 | |
| 2 | 58 | 16 | 37 | 67 | 7 | 22 | 64 | 68 | 28 | 29 | |
| 8 | 107 | 32 | 17 | 18 | 32 | 11 | 12 | 13 | 49 | 77 | |
| 8 | 156 | 48 | 18 | 18 | 43 | 15 | 12 | 13 | 49 | 98 | |
| 4 | 165 | 48 | 27 | 43 | 19 | 19 | 37 | 21 | 17 | 26 | |
| 2 | 166 | 52 | 52 | 76 | 20 | 35 | 68 | 46 | 17 | 18 | |

$$D(i,j) = \min \left\{ \begin{array}{l} D(i,j-1) \\ D(i-1,j) \\ D(i-1,j-1) \end{array} \right\} + d(x_i, y_j)$$

Distance = 18

Normalize by the
length of the path

$$\text{dist}(\mathbf{X}, \mathbf{Y}) = D(M, N) / K$$

DTW: Example with 1-D signals

| X | Y | 1 | 4 | 7 | 9 | 3 | 6 | 9 | 7 | 2 | 1 |
|---|-----|----|----|----|----|----|-----|-----|-----|-----|---|
| 2 | 1 | 5 | 30 | 79 | 80 | 96 | 145 | 170 | 170 | 171 | |
| 5 | 17 | 2 | 6 | 22 | 26 | 27 | 43 | 47 | 56 | 72 | |
| 7 | 53 | 11 | 2 | 6 | 22 | 23 | 27 | 27 | 52 | 88 | |
| 3 | 57 | 12 | 18 | 38 | 6 | 15 | 51 | 43 | 28 | 32 | |
| 2 | 58 | 16 | 37 | 67 | 7 | 22 | 64 | 68 | 28 | 29 | |
| 8 | 107 | 32 | 17 | 18 | 32 | 11 | 12 | 13 | 49 | 77 | |
| 8 | 156 | 48 | 18 | 18 | 43 | 15 | 12 | 13 | 49 | 98 | |
| 4 | 165 | 48 | 27 | 43 | 19 | 19 | 37 | 21 | 17 | 26 | |
| 2 | 166 | 52 | 52 | 76 | 20 | 35 | 68 | 46 | 17 | 18 | |

$$D(i,j) = \min \left\{ \begin{array}{l} D(i,j-1) \\ D(i-1,j) \\ D(i-1,j-1) \end{array} \right\} + d(x_i, y_j)$$

Distance = 18

Normalize by the length
of the path

$$\text{dist}(\mathbf{X}, \mathbf{Y}) = D(M, N) / K$$

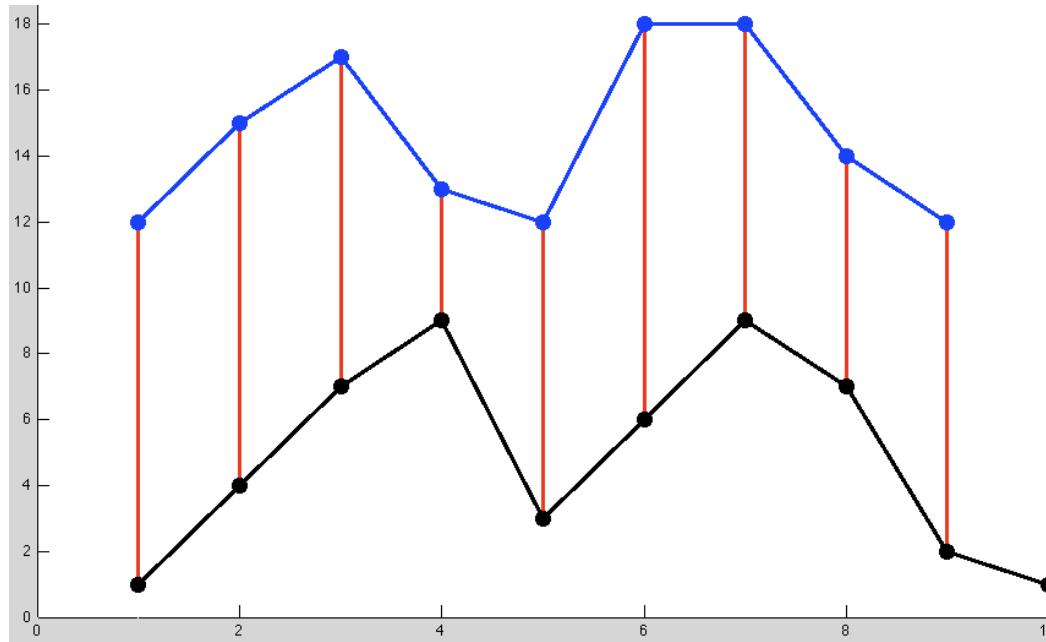
Path? → Backtracking

Final distance = 18/10

DTW: Example with 1-D signals

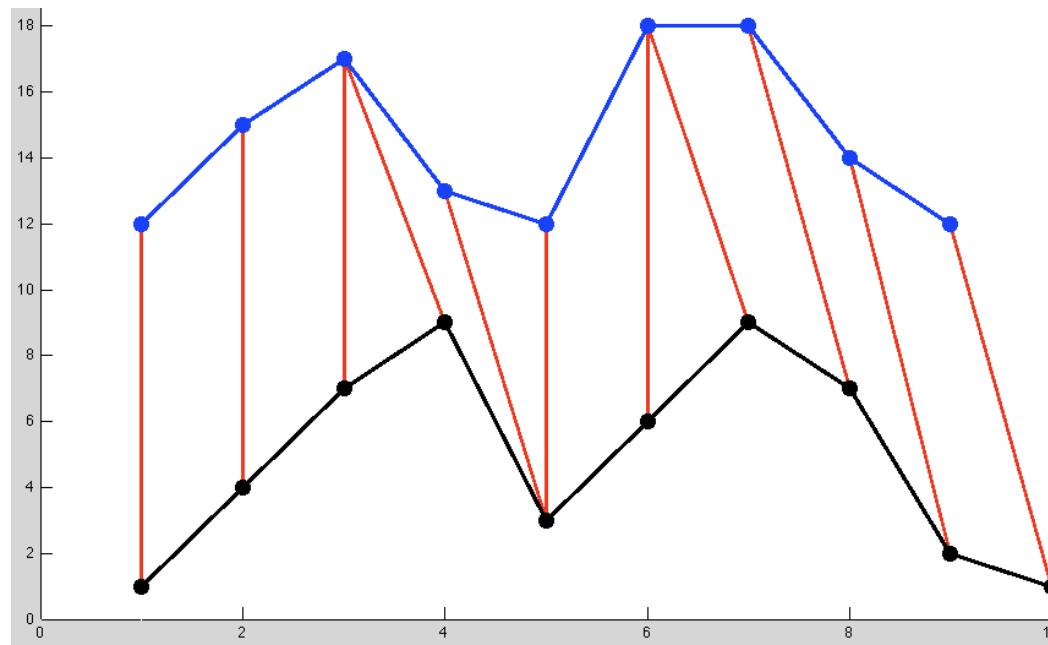
Matching:

Euclidean distance → Undesired!



DTW: Example with 1-D signals

DTW Matching:



Classic Approach: Rath and Manmatha 2003

- Matching: Dynamic Time Warping (DTW)

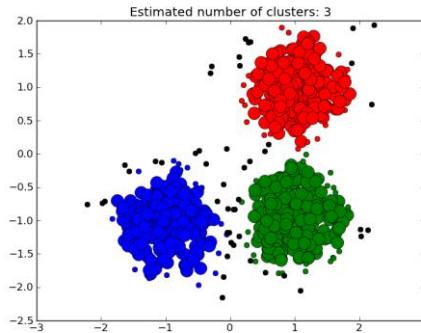
The distance at each point is the square of the euclidean distance between 4-dimensional vector

$$D(i,j) = \min \left\{ \begin{array}{l} D(i,j-1) \\ D(i-1,j) \\ D(i-1,j-1) \end{array} \right\} + d(x_i, y_i)$$

$$d(x_i, y_j) = \sum_{k=1}^d (x_{i,k} - y_{j,k})^2$$

Improvement: Rath and Manmatha 2007

- Same method + Clustering (groups)
 - DFT features from profiles → K-means clustering
 - Matrix with pairwise distances → k-means
 - Average linkage: average distance between all pair items



- Useful for speeding up the transcription of documents
 - The user transcribes only one word in each cluster
 - All words in the cluster are automatically transcribed

T.M. Rath, R. Manmatha: Word spotting for historical documents. IJDAR, 2007.

Improvement: Rath and Manmatha 2007

Advantages

Features and matching robust to deformations

They can handle the variability in the handwriting style

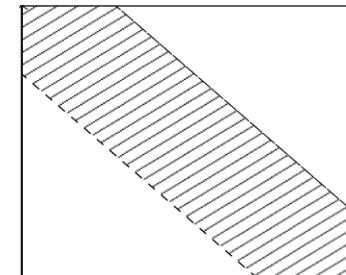
Disadvantages

Depends on a good word segmentation

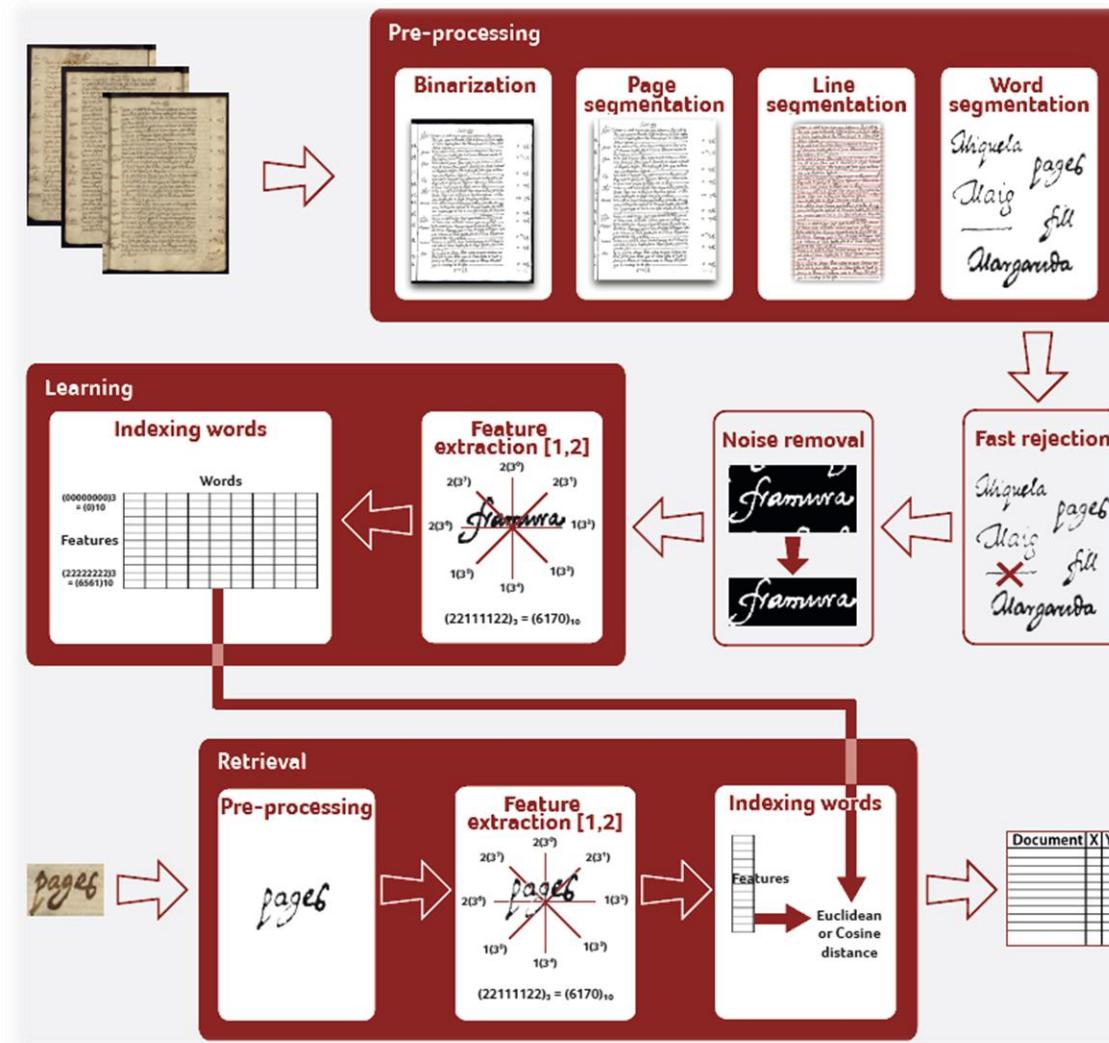
Complexity $O(n^2)$ → Slow method, hardly scalable

All distances between words have to be computed

Optimizations → Sakoe-Chiba band

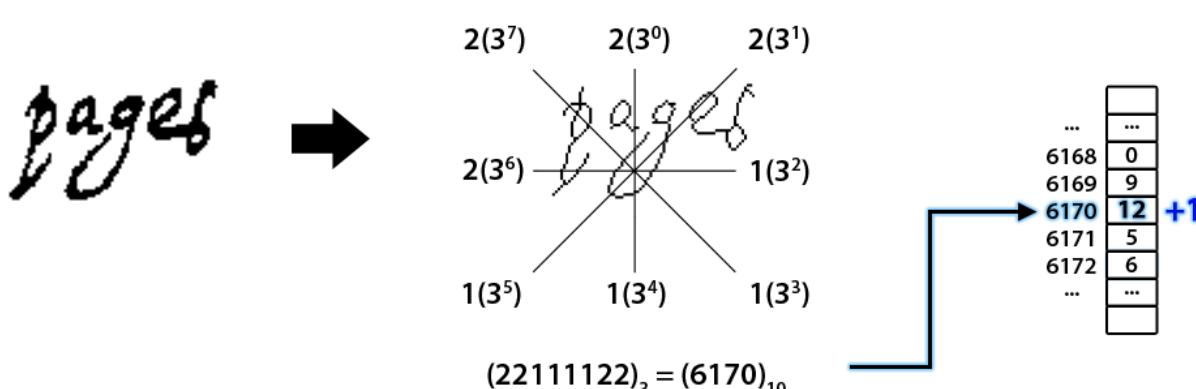


Word Spotting using a pseudo-structural descriptor



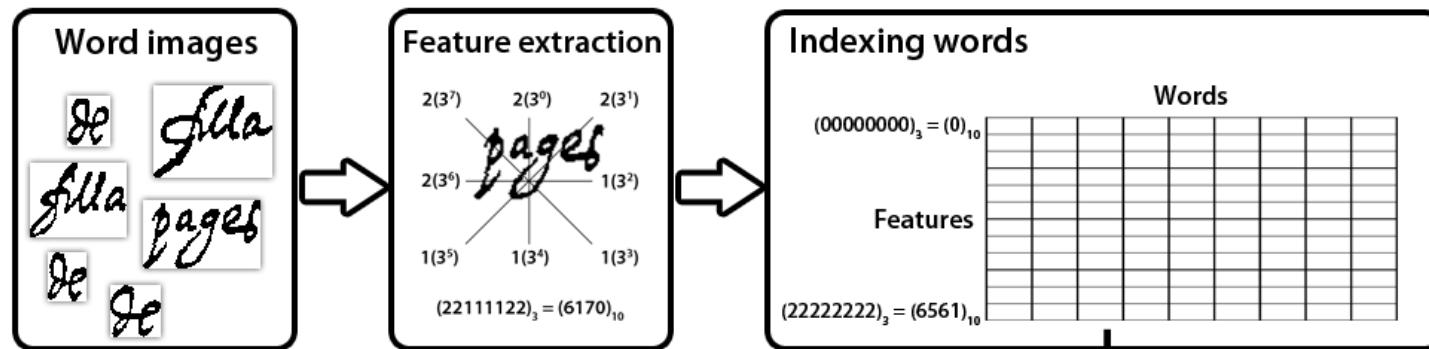
Word Spotting using a pseudo-structural descriptor

Characteristic Loci Features: A characteristic Loci feature is composed by the number of the intersections in four directions (up, down, right and left).

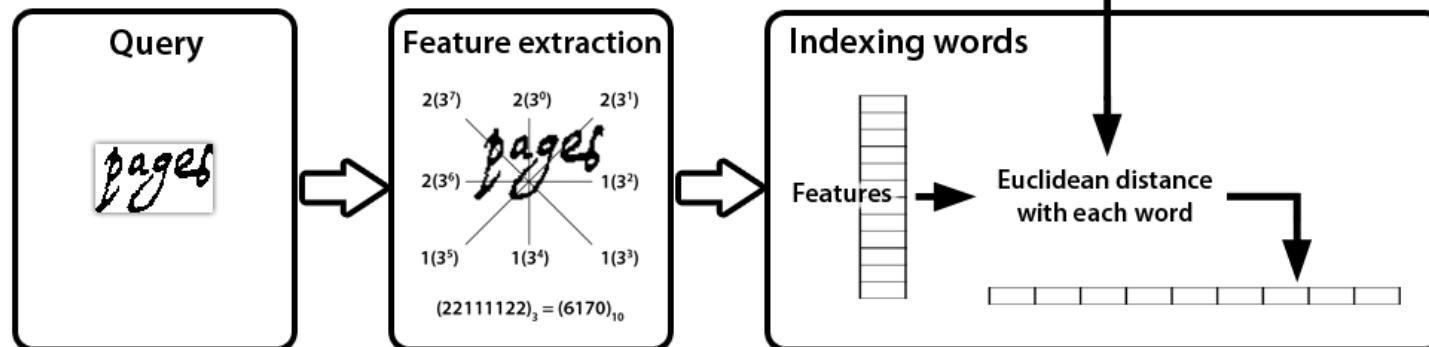


Word Spotting using a pseudo-structural descriptor

Learning



Retrieval

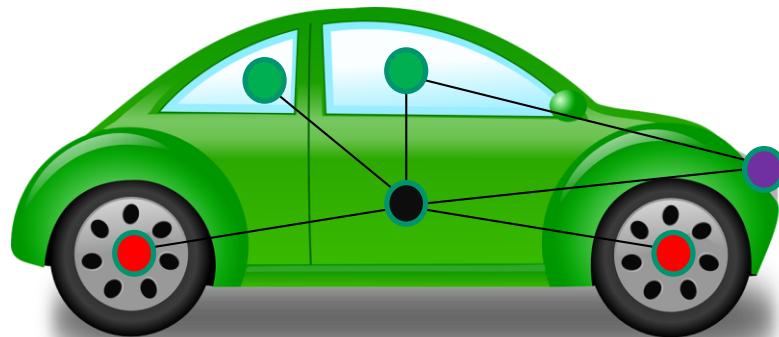


D. Fernandez, J. LLados, and A. Fornes. Handwritten Word Spotting in Old Manuscript Images using a Pseudo-Structural Descriptor Organized in a Hash Structure. Iberian Conference on Pattern Recognition and Image Analysis (IBPRIA), June 2011.

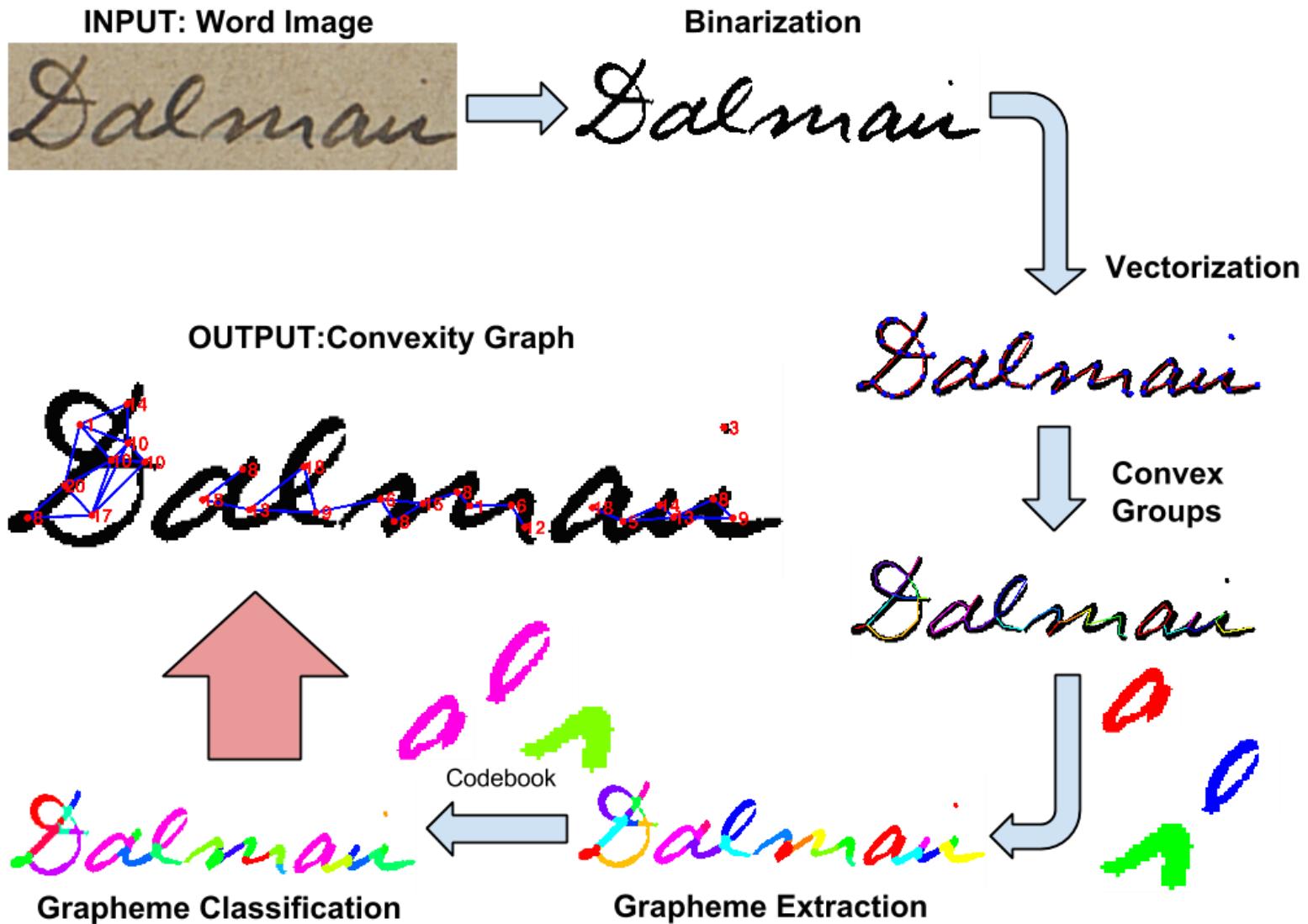
Structural Representations

Graph-based representation.

A **graph** is a set of objects called vertices that can be connected with an edge representing their relation.



Word Spotting using a graph representation



Robust Graph-based Representation

- **Robust graph representation**, tolerant to the inherent deformation of handwritten strokes, and variations among writers.
- Representation based on **convex groups** of the skeleton image (nodes), called **graphemes**, and adjacency relationships (edges).
- **Grapheme** or **Glyph** is the smallest unit used in describing the writing system of a language.



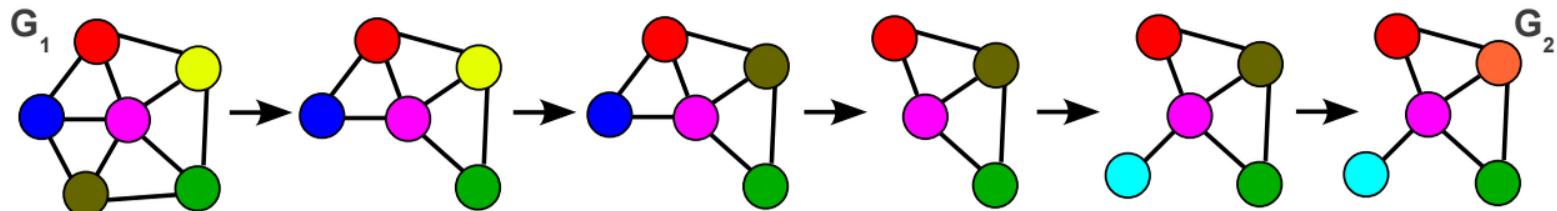
P.Riba, J.Lladós, A.Fornés. Handwritten Word Spotting by Inexact Matching of Grapheme Graphs. International Conference on Document Analysis and Recognition (ICDAR), 2015.

Word Spotting using a graph representation

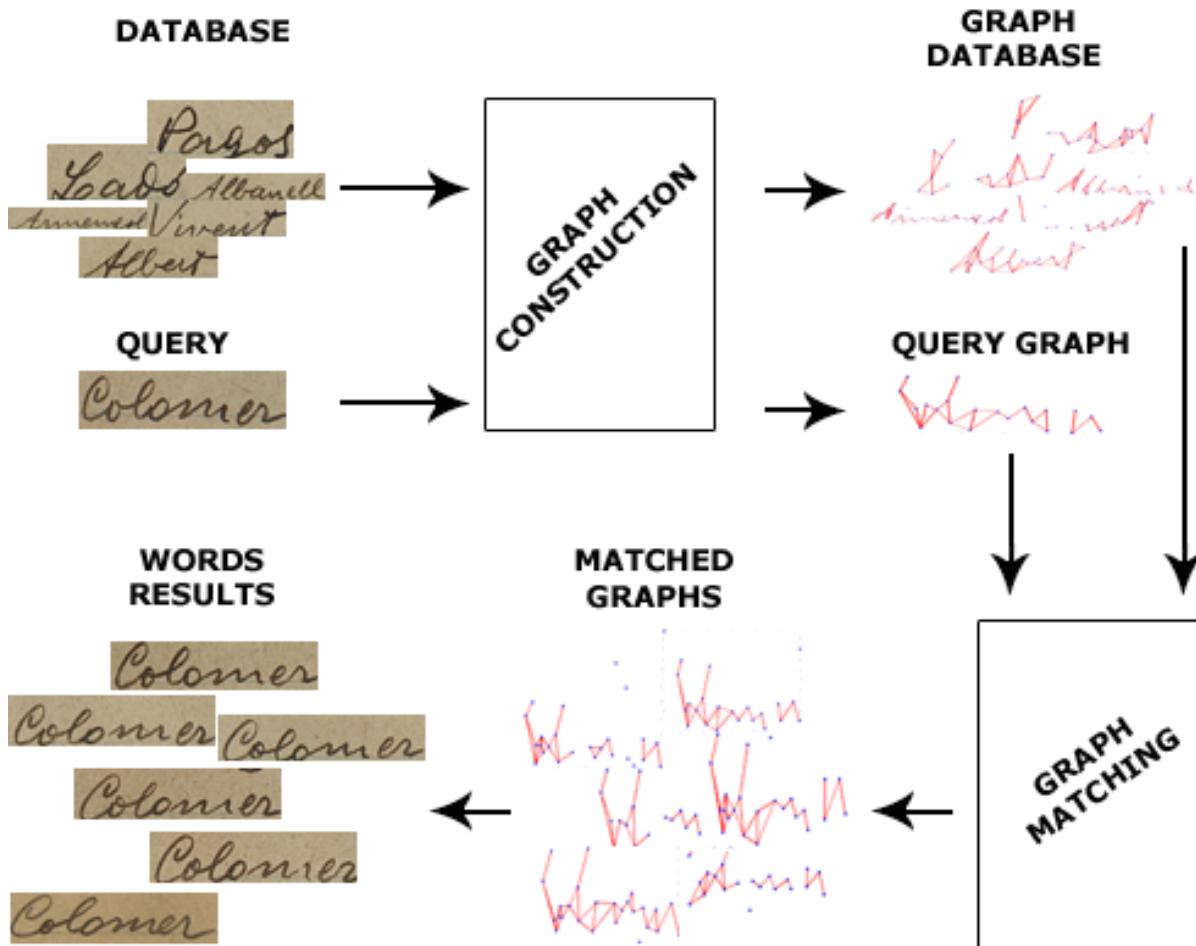


Inexact Graph Matching

- **Graph Edit Distance** follows the idea of string edit distance.
- The minimum cost of operations required to transform one graph into the other. (vertices and edges).
- **Insertion, deletion and substitutions.**
- **Bipartite Graph Matching**, suboptimal approximation to graph edit distance.

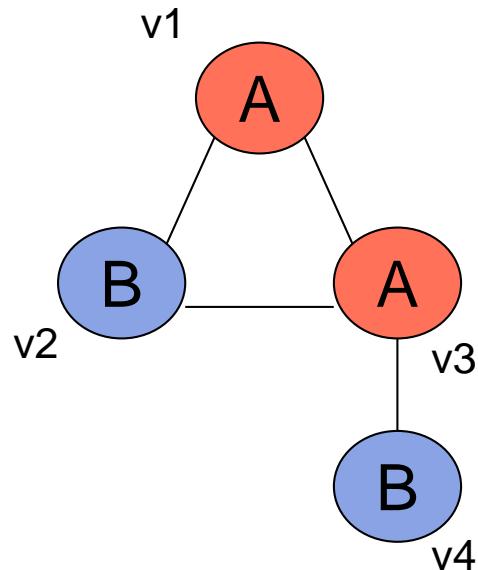


Graph Matching for Word Spotting



Segmentation-free? → Graph Embedding

Binary topological node feature



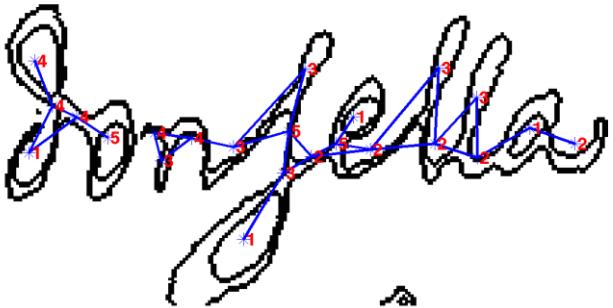
| | Label A | Label B |
|-------------------------------|---------|---------|
| $v_{v1} = [1, 3, 6, 1, 2, 4]$ | | |
| $v_{v2} = [2, 2, 7, 0, 3, 3]$ | | |
| $v_{v3} = [1, 4, 6, 2, 1, 7]$ | | |
| $v_{v4} = [1, 1, 4, 0, 2, 1]$ | | |

| |
|-------------------------------------|
| $\hat{v}_{v1} = [0, 1, 1, 0, 0, 1]$ |
| $\hat{v}_{v2} = [1, 0, 1, 0, 1, 0]$ |
| $\hat{v}_{v3} = [0, 1, 1, 1, 0, 1]$ |
| $\hat{v}_{v4} = [0, 0, 0, 0, 0, 0]$ |

4 paths of length 3 incident in v4 coming from nodes labeled with A. (tottering is allowed)

- [v1, v2, v3, v4]
- [v3, v1, v3, v4]
- [v3, v2, v3, v4]
- [v3, v4, v3, v4]

Experimental Results

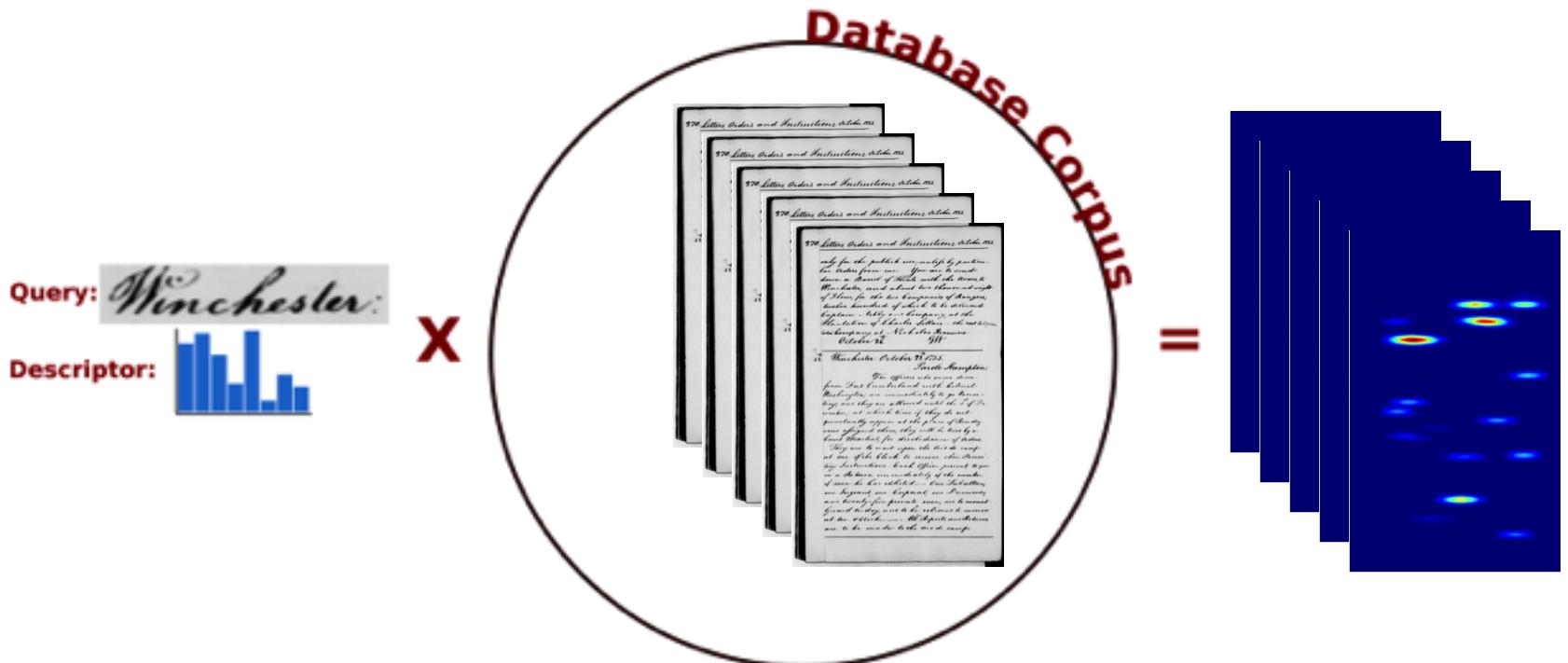


Domingo 6 rebié de D. Pau Colomé y su mujer dona Paula su
hija Pilar pagat y de Elizabeth Margués dona del Villar
de granada sus hijos de Jordal y de Juanita de gonzález —
que ha rebié de Antoni Ramon pagat de su hermano fill de Joan
Carra casas y de t. Agustí al Blavaria S. Angelina
filla de Josep L. matrero de Esteban y Domènec y
de t. Agustí —
Pilar a 2. rebié de Pau Colomé pagat de su hermano
de la Rosa fill de Josep Colomé pagat de su hermano
de Margarida víuda de Miguel iñaki pagat de Josep iñaki
y molt més en tallermanet —
que dia rebié de Juan Joan Francesc habitant en Berga ab Esperanza
iugat filla de Pere Pau i iugat poser real y de Esperanza matrero
de Joan —
Catalina a 3. rebié de Josep Utrera vecindat de Berga filla de
Pere Martí pagat de Beltria Achurri ab Utrera. La viudat
de Josep Utrera que es de Berga —
que dia rebié de Josep Peforante tañedor de aguafrescas
habitante en Berga ab Juan Francesc —
que dia rebié de Juan Santiago frances habitant en Berga ab
Paula Llorente víuda de Antoni que operaria en Berga —
que dia a 11. rebié de Ramona Om víuda de Llana frances habitante
en Terrassa ab Josep Llana — filla de Josep Llana es
viudor de Llana després y de — vivien —
que dia a 12. rebié de Josep frances sepiador de Berga fill de
ab Anna Llana — filla filla de Pere Horroch corredor de vistolas de
Berga y de Catàuria —
que dia rebié de Josep Colomer pagat de Berga fill de Antoni Co
lomer pagat y de Andronica, ab Josep — filla de Josep
Colomer pagat y de Margalida —
que dia rebié de Pere Cufanyol pagat habitante de Berga
de Josep fill de Pere Cufanyol vecindat de Berga y de Elizabeth
que dia rebié de Josep Cufanyol pagat habitante de Berga
y de Elizabeth —

WORD SPOTTING LEARNING-FREE & SEGMENTATION-FREE

Segmentation-free methods

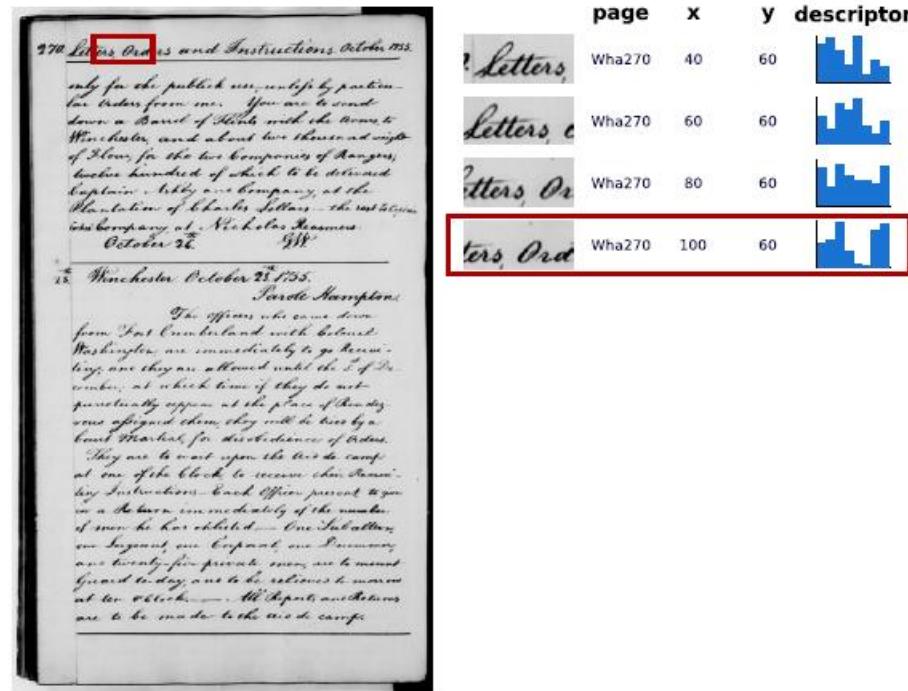
No segmentation of lines nor words



Word Spotting using Bag of Visual Words

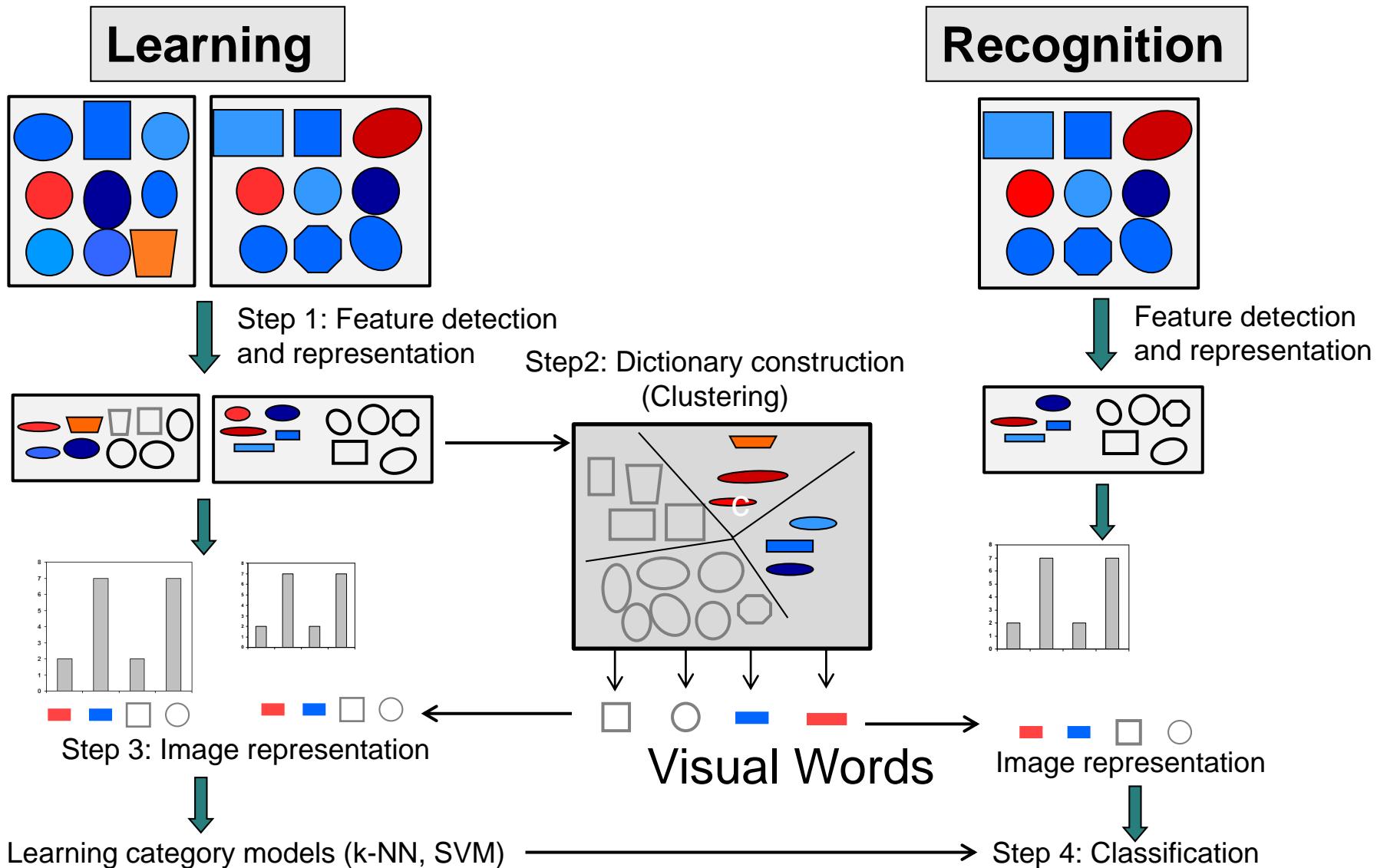
The method does not need any word nor line segmentation

Patch-based framework



M. Rusiñol, D. Aldavert, R. Toledo and J. Lladós. Efficient segmentation-free keyword spotting in historical document collections. Pattern Recognition, 2015

Bag of Words



Problem with Bag-of-Words

integral

Histogram= [a,e,g,i,l,n,r,t]

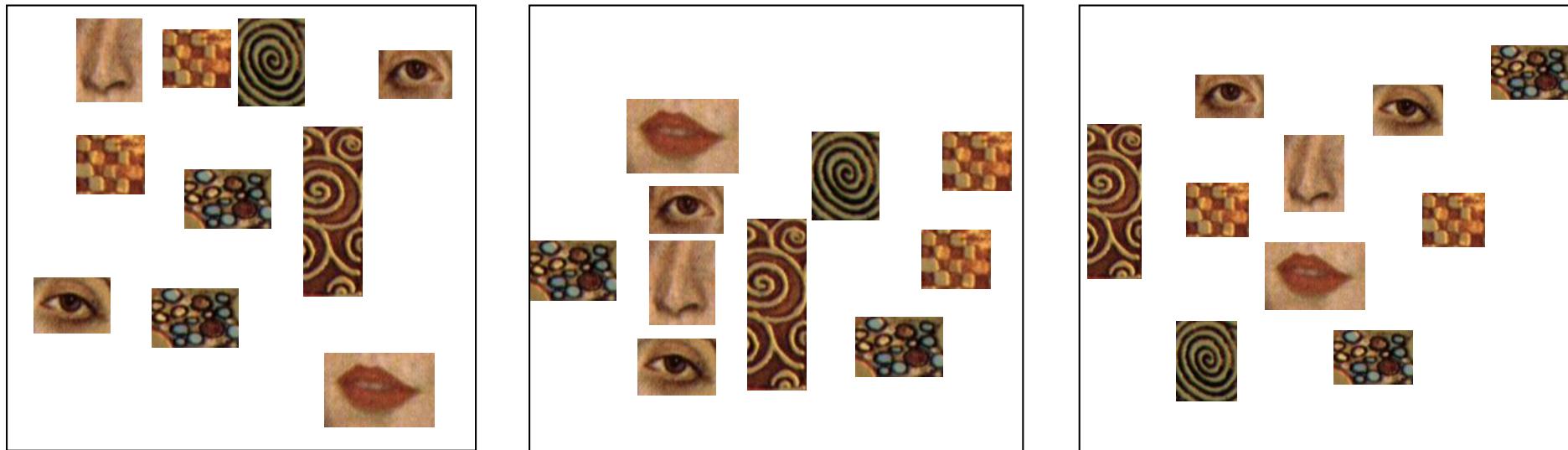
relating

Histogram = [a,e,g,i,l,n,r,t]

triangle

Histogram = [a,e,g,i,l,n,r,t]

Problem with Bag-of-Words



All have equal probability for bag-of-words methods

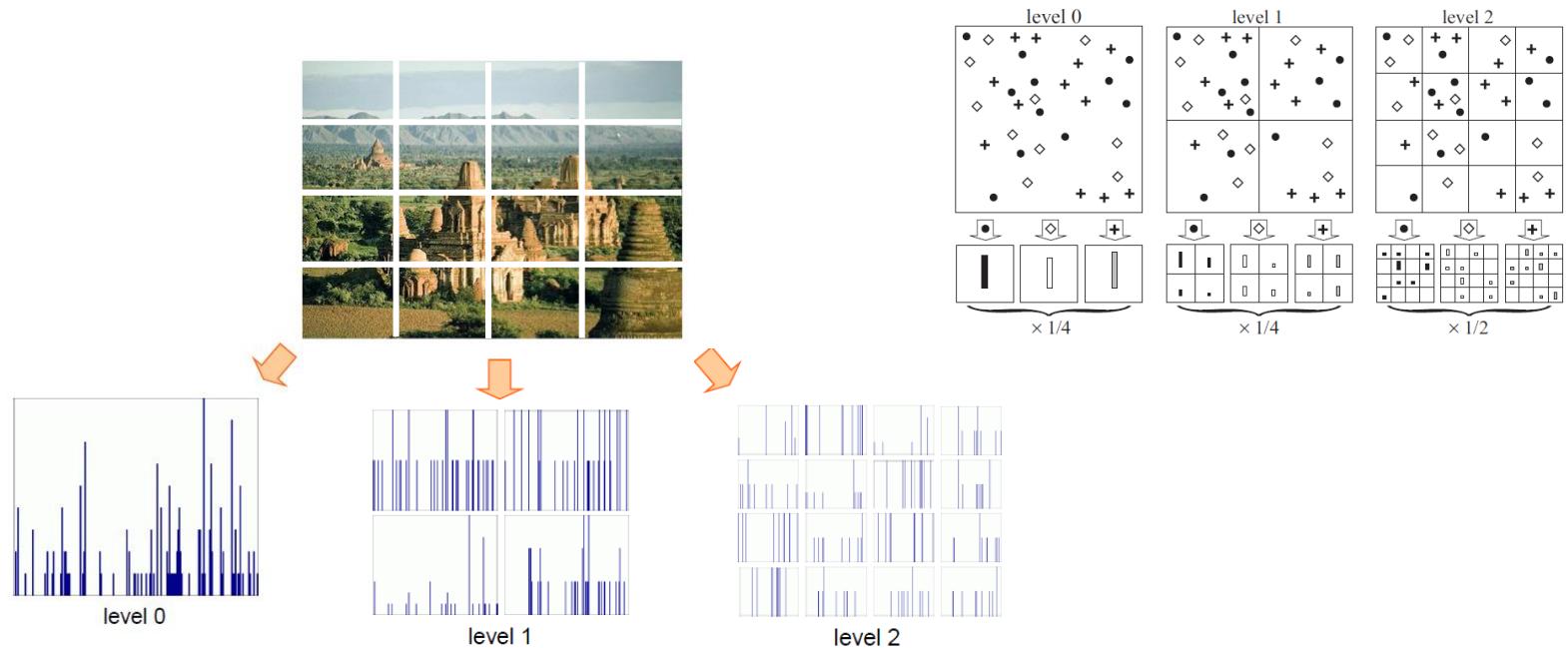
Location information is important

BoW + location still doesn't give correspondence

Improving BoW: Adding spatial information

Spatial Pyramids (Lazebnik, Schmid, Ponce, 2006)

- Hierarchical computation of histograms at different spatial image decompositions
- Combination of histograms at all levels of decomposition
- More weight to finer grids



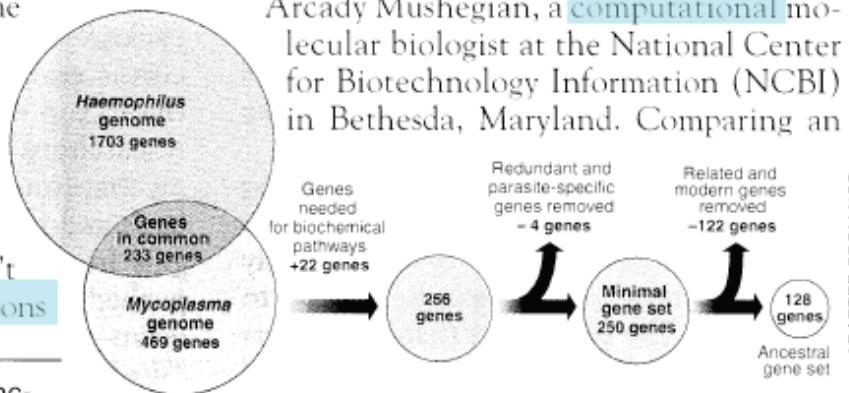
Improving BoW: Topic Models

Seeking Life's Bare (Genetic) Necessities

L COLD SPRING HARBOR, NEW YORK—How many genes does an organism need to survive? Last week at the genome meeting here,* two genome researchers with radically different approaches presented complementary views of the basic genes needed for life. One research team, using computer analyses to compare known genomes, concluded that today's organisms can be sustained with just 250 genes, and that the earliest life forms required a mere 128 genes. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough.

Although the numbers don't match precisely, those predictions

"are not all that far apart," especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson of Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just a genetic numbers game, particularly as more and more genomes are completely mapped and sequenced. "It may be a way of organizing any newly sequenced genome," explains Arcady Mushegian, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an



Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 12.

Improving BoW: Topic Models

Topics

gene 0.04
dna 0.02
genetic 0.01
...

life 0.02
evolve 0.01
organism 0.01
...

brain 0.04
neuron 0.02
nerve 0.01
...

data 0.02
number 0.02
computer 0.01
...

Documents

Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK—How many genes does an organism need to survive? Last week at the genome meeting here,* two genome researchers with radically different approaches presented complementary views of the basic genes needed for life. One research team, using computer analyses to compare known genomes, concluded that today's organisms can be sustained with just 250 genes, and that the earliest life forms required a mere 128 genes. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough.

Although the numbers don't match precisely, those predictions

* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 12.

SCIENCE • VOL. 272 • 24 MAY 1996

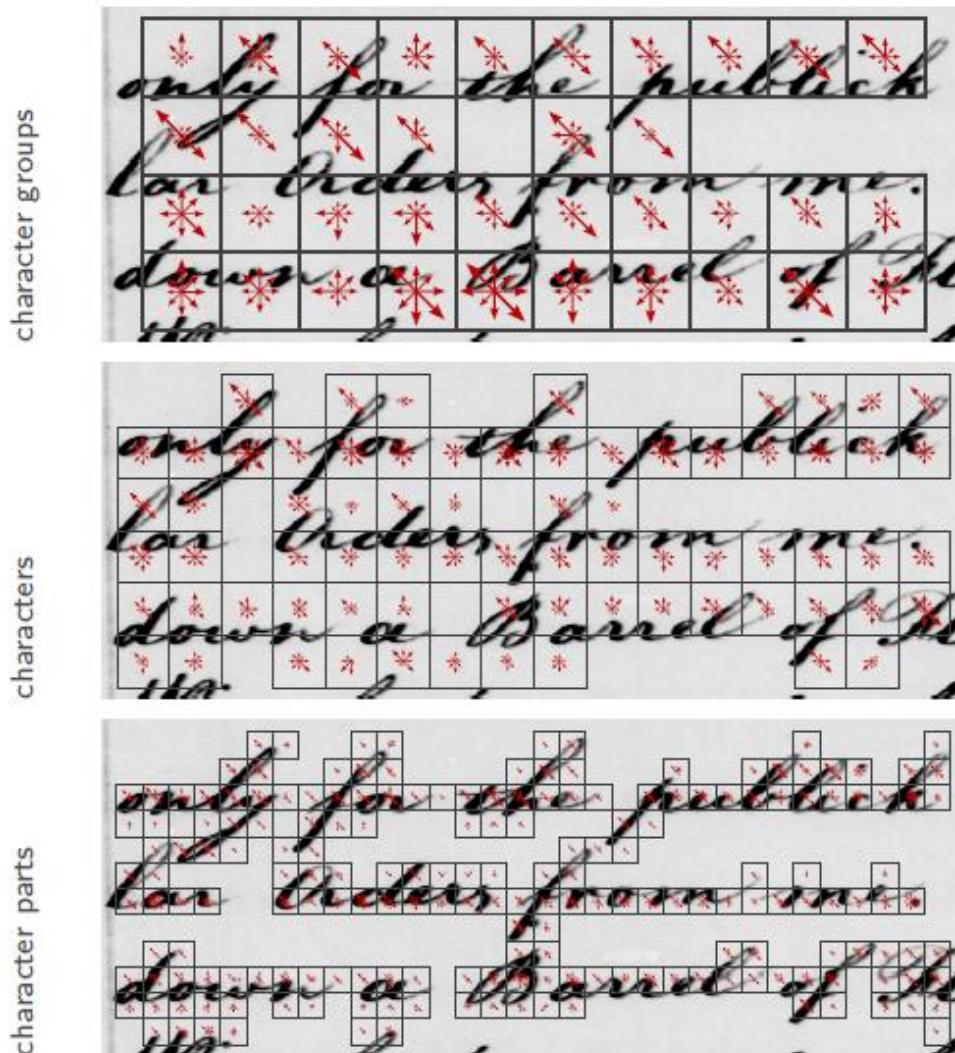
"are not all that far apart," especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson of Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just a matter of numbers alone, particularly as more and more genomes are completely mapped and sequenced. "It may be a way of organizing any newly sequenced genome," explains Arcady Mushegian, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an



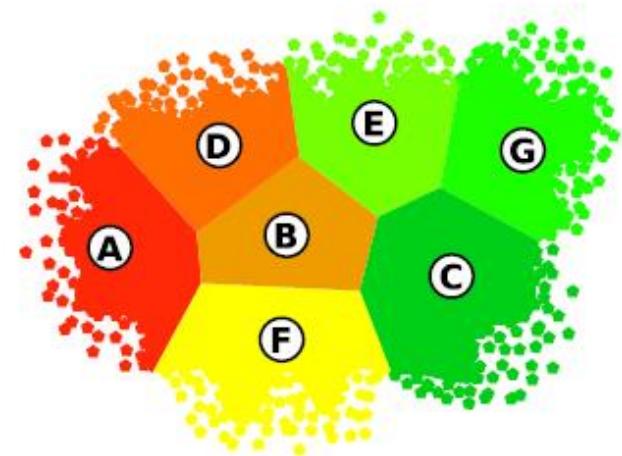
Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

Topic proportions and assignments

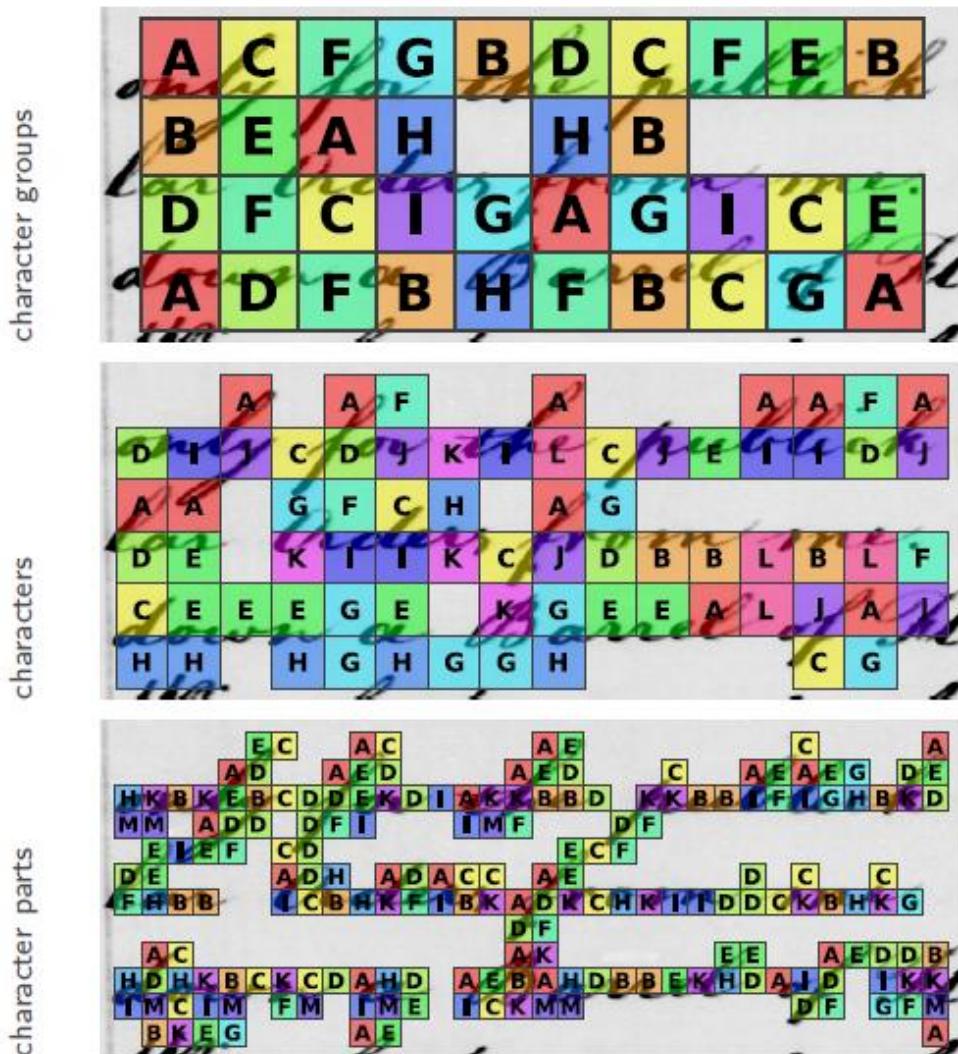
Bag of Visual Words for Word Spotting



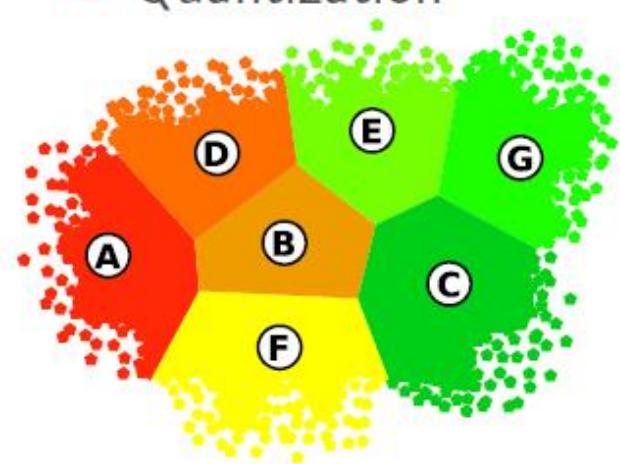
- SIFT features
- Densely extracted at 3 different scales
- Remove low-gradient features
- Build a codebook with k -means



Bag of Visual Words

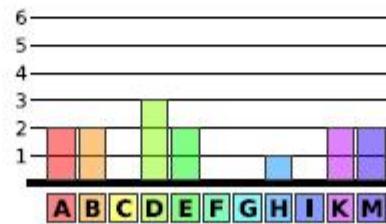
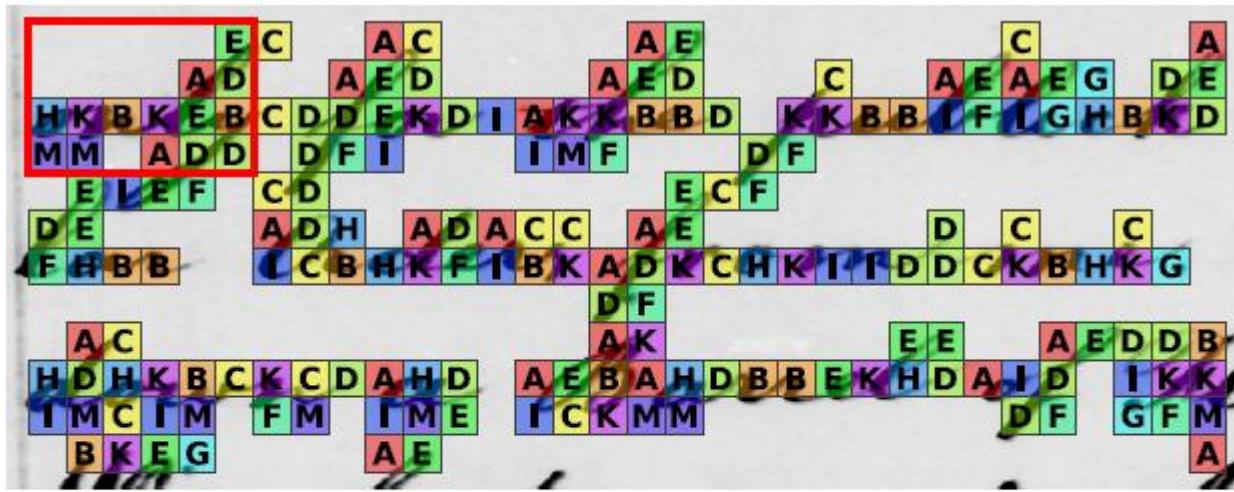


- SIFT features
- Densely extracted at 3 different scales
- Remove low-gradient features
- Build a codebook with k -means
- Quantization



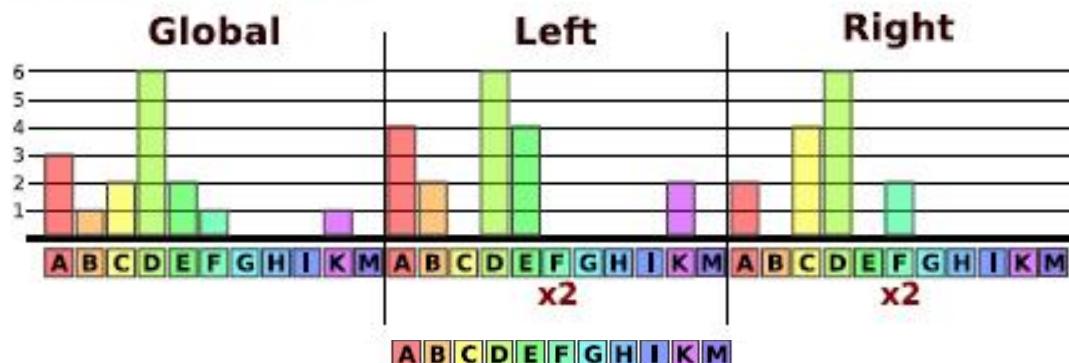
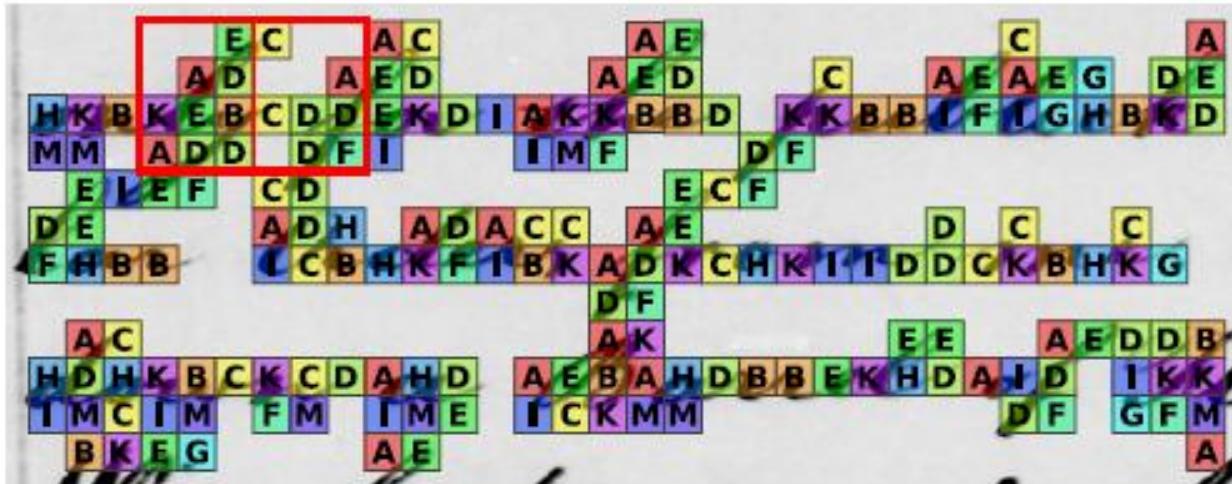
Bag of Visual Words

The document is splitted into overlapping local patches



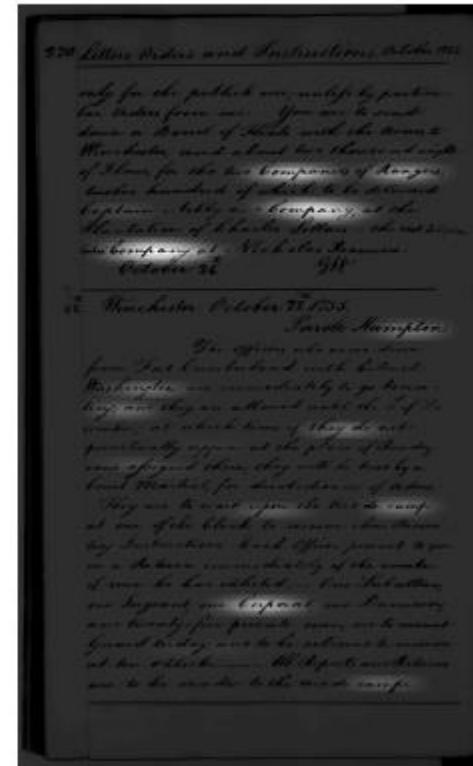
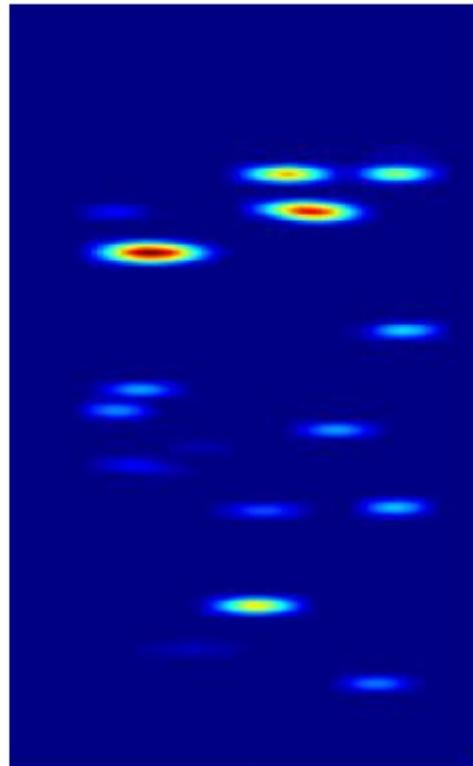
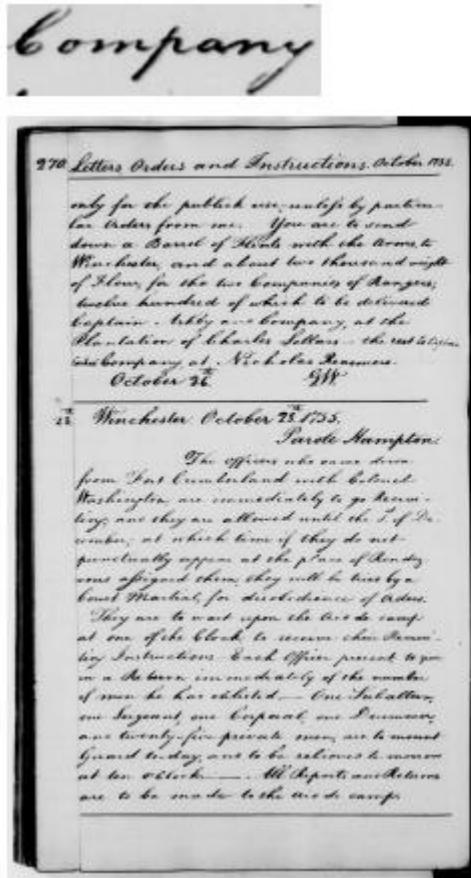
Bag of Visual Words

The document is splitted into overlapping local patches



Bag of Visual Words

- ◆ A voting scheme provides the probability map where to find the queried word



Discussion: Representations

Table 4. Pros and Cons Summary.

| | Size | Time | Indexability | Preprocessing | Performance | Scalability |
|-------------------|------|------|--------------|---------------|-------------|--------------------------|
| BoVW | -- | + | + | + | ++ | — (size) |
| DTW | + | -- | — | — | + | — (time) |
| Pseudo-structural | ++ | ++ | ++ | — | + | — (discriminative power) |
| Structural | ++ | ++ | ++ | --- | — | — (discriminative power) |

J. Lladós, M. Rusiñol, A. Fornés, D. Fernández and A. Dutta. On the Influence of Word Representations for Handwritten Word Spotting in Historical Documents. International Journal of Pattern Recognition and Artificial Intelligence, 2012.

Discussion: Techniques

Techniques

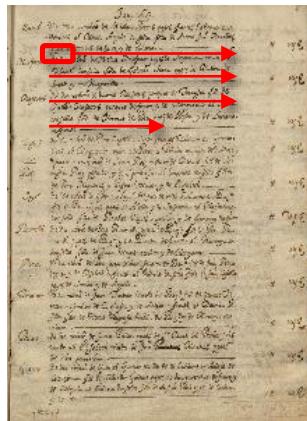
| | Complexity Scalability | Performance | Generalizable Other scenarios | Room for improvements |
|---|---------------------------|-------------|----------------------------------|--------------------------|
| Hashing /Indexing | ++ | - | + | ++ |
| Alignments (DTW / Graph / Trees) | - | ++ | - | - |
| Template matching (Euclidean distance) | + | + | -- | - |

WS - segmentation

Segmentation-free

- ✓ The image is divided into patches (e.g. using a sliding window).

It is computationally costly in terms of time.



Segmentation-based

- ✓ The computational cost is low.
- It requires each document image to be segmented at word level.
- The word classification is strongly influenced by over or under-segmentations



Discussion: Segmentation-free

Segmentation-free methods

Advantages:

Do not depend on any segmentation method

Disadvantages:

Find parts of words (defender = end)

Lord Lord Lord Lord Lord Lord
an an an an an an

Discussion: Learning-free methods

Learning-free methods

Advantages ☺

No training data required

Independent to alphabet and language

They are an image retrieval case

fast with indexing-hashing techniques

Disadvantages ☹

Difficult Multiple writer

How to deal with handwriting variations?

Difficult query-by-string

How to generate a “realistic” word from the ASCII text?

QUERY-BY-EXAMPLE QUERY-BY-STRING

Classification of Word Spotting

Query-by-String

- ✓ Flexibility to search any kind of keyword.

It has to build a writing model and to map this model to the words images.

Labelled datasets are required in order to train the recognition engine.

Query-by-Example

- ✓ It does not require learning.
- ✓ It only requires collecting one or several examples of the keyword.
- ✓ It does not need labeled training data.
- It is sensitive to different writing styles.

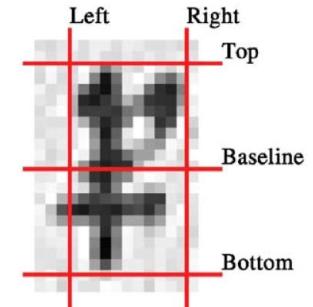
“filla” →



Query-by-String (& Segmentation-free) approach

Writing Model is constructed

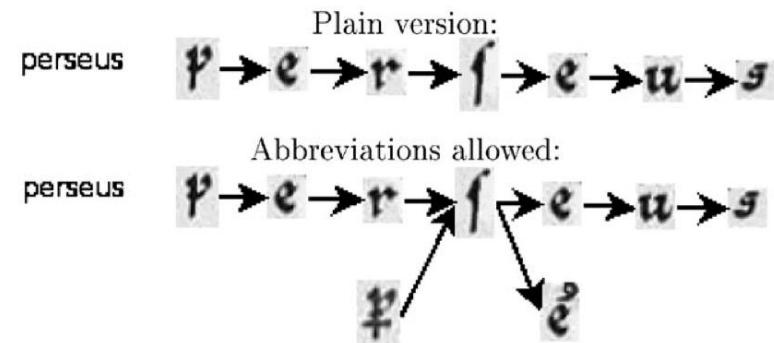
For each letter in the alphabet, extract one image, and provide information about the linking behavior



When the user types the query word

Generates different grammatical variations

Links the letters according to the rules

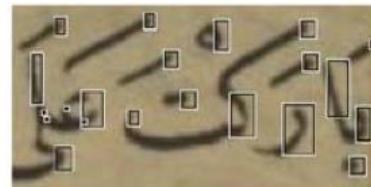
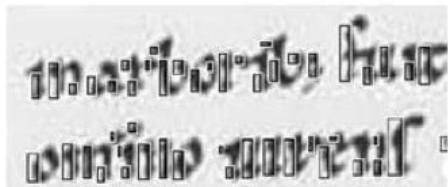


Y.Leydier, A.Ouji, F.LeBourgeois, H.Emptoz; "Towards an omnilingual word retrieval system for ancient manuscripts". PR(42), pp. 2089-2105, 2009.

Query-by-String approach

- Features: Gradients
- Extract guides

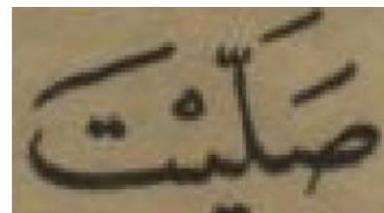
Inverse direction to the writing order



- ZOI location

Expand while variation in
angle gradients

Reduce overlapping regions



Query-by-String approach

- Matching
 - ZOIs are stored in a tree
 - Search the tree through the lowest score path
 - Prune branches with high scores

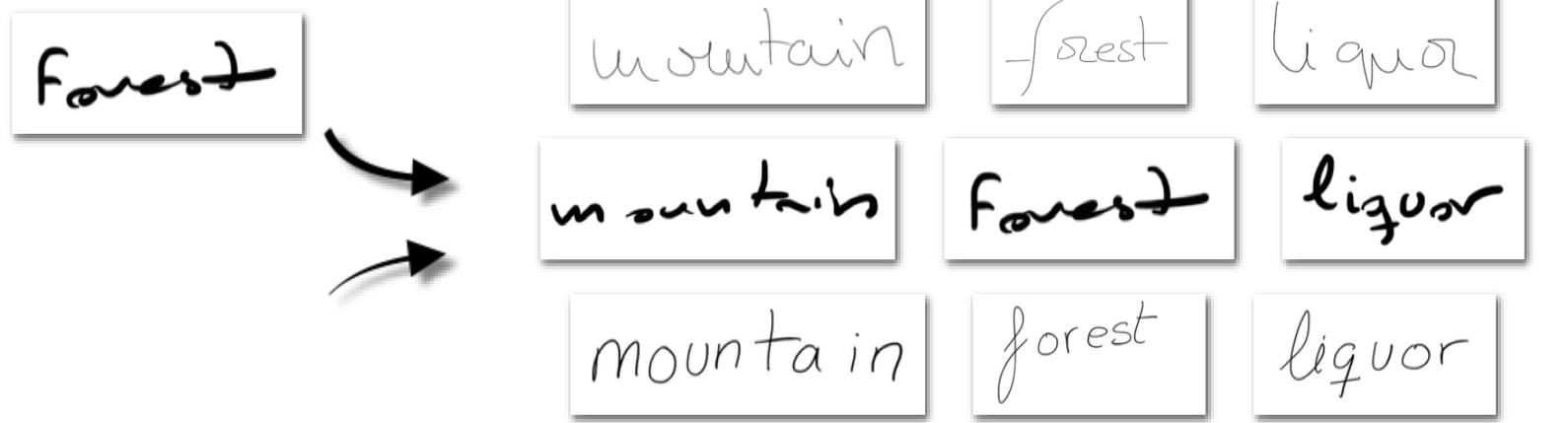
Advantages / Disadvantages

Query-by-string ☺

Model has to be created ☹

LEARNING-BASED WORD SPOTTING

Difficulties



Introduction: word spotting



Learning-based Keyword Spotting

Advantages:

- Higher precisions and recall rates can be expected.
- Flexible systems, can be used for Document filtering and database search.
- More sophisticated system allow the inclusion of more, external data (e.g. language models) to further increase the recognition

Disadvantages:

- A (large) training set is needed
- Fixed costs for training the system have to be taken into account, whenever new writing styles appear.

Approaches

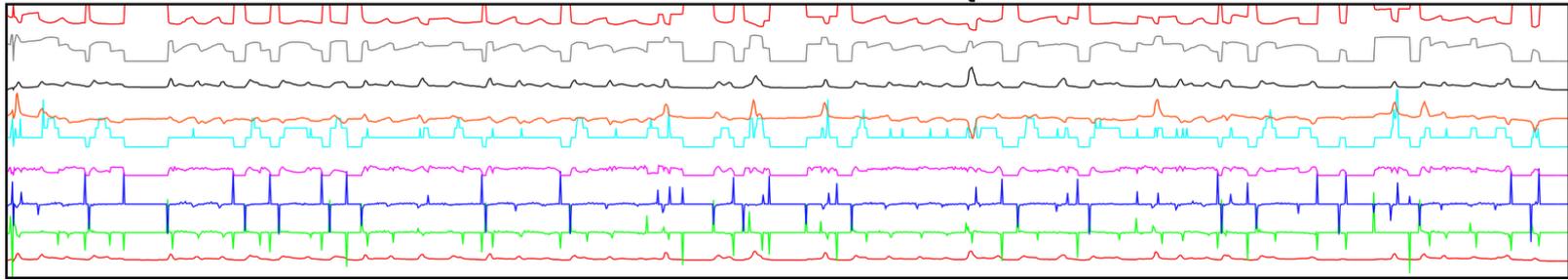
Idea: evaluate efficiently a matching score between the query and the text line

Common Approaches:

- Hidden Markov Models
- Neural Networks: BLSTM NN

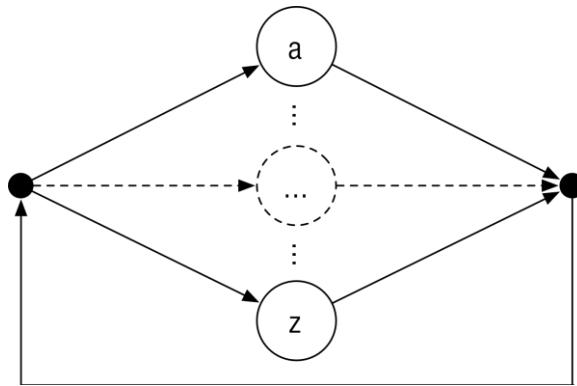
Sequential Representation of a Text Line

the measurement of temperatures. This,

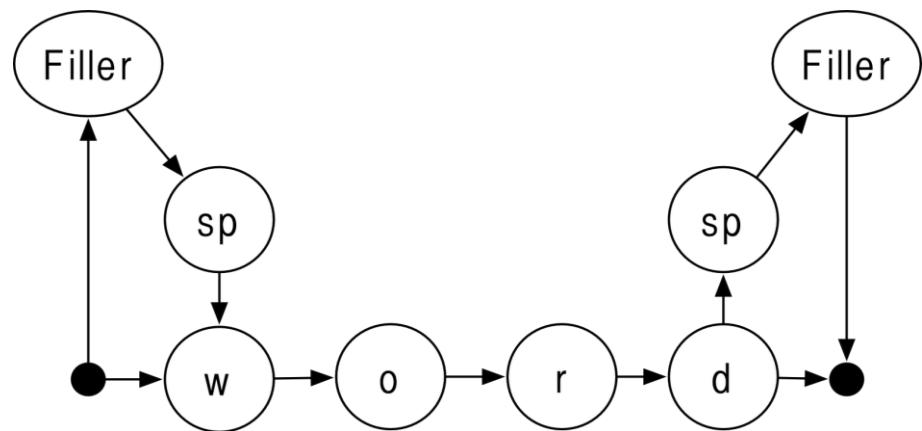


HMM for Word Spotting

Given a Keyword, construct 2 HMMs



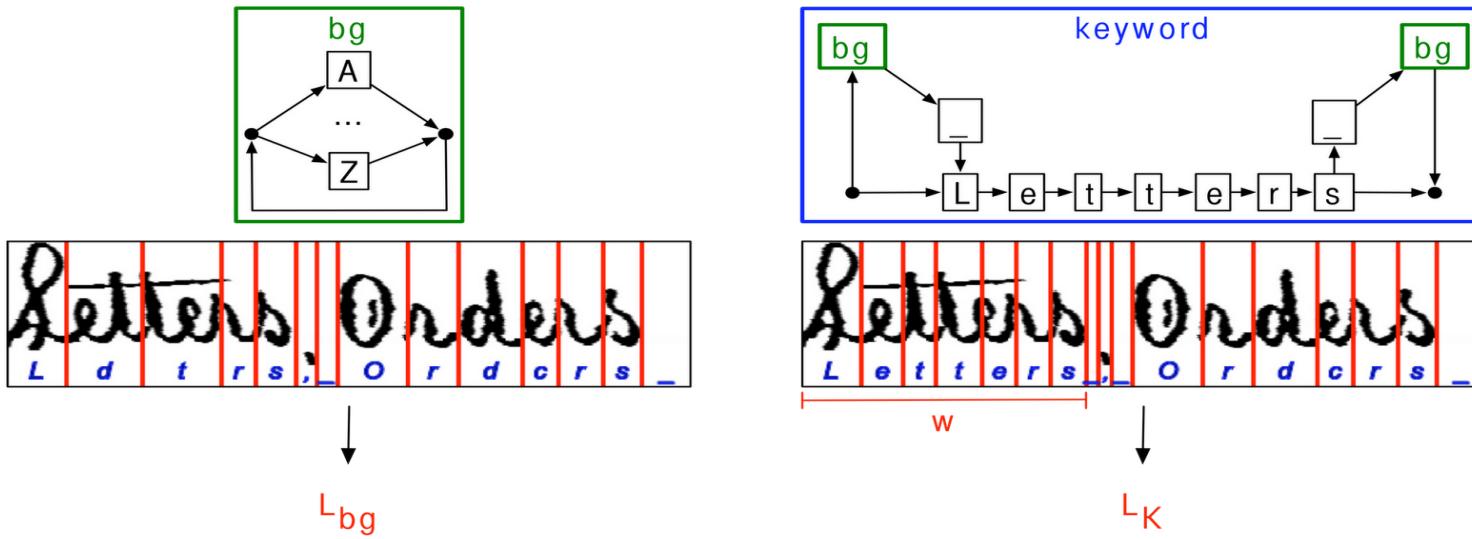
The Background-HMM



The Keyword-HMM

HMM for Word Spotting

Perform Viterbi recognition using both models



Consider likelihood ratio $R = L_K / L_{bg}$ ($0 < R < 1$, since $L_{bg} > L_K$)

Return positive match, iff $R/w > T$, with

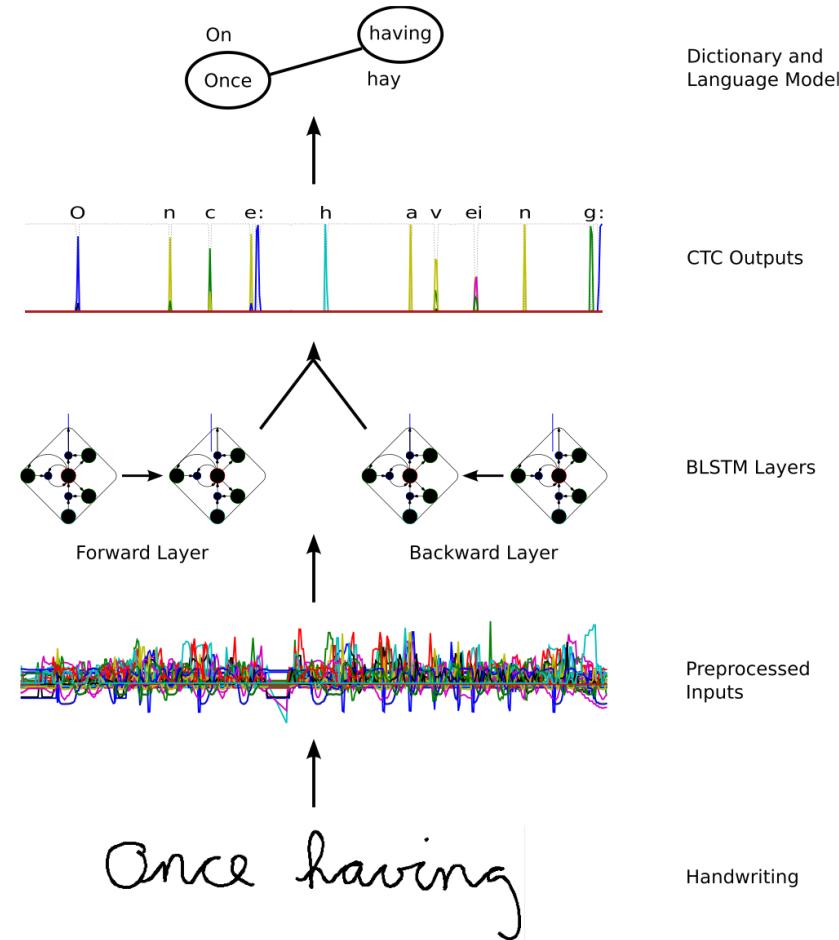
keyword width w and threshold T

Handwriting Recognition using BLSTM NN

Bidirectional Neural Networks + n-grams

Graves et al., "A novel connectionist system for unconstrained handwriting recognition", PAMI 2009

- Recurrent Neural Networks
 - Bidirectional (BLSTM blocks)
 - CTC: Dynamic Programming
 - Word n-grams + dictionary
-
- Best results so far



Bidirectional Long-Short Term Memory Neural Networks

BLSTM NN are bidirectional recurrent NN with a specialized LSTM hidden layer

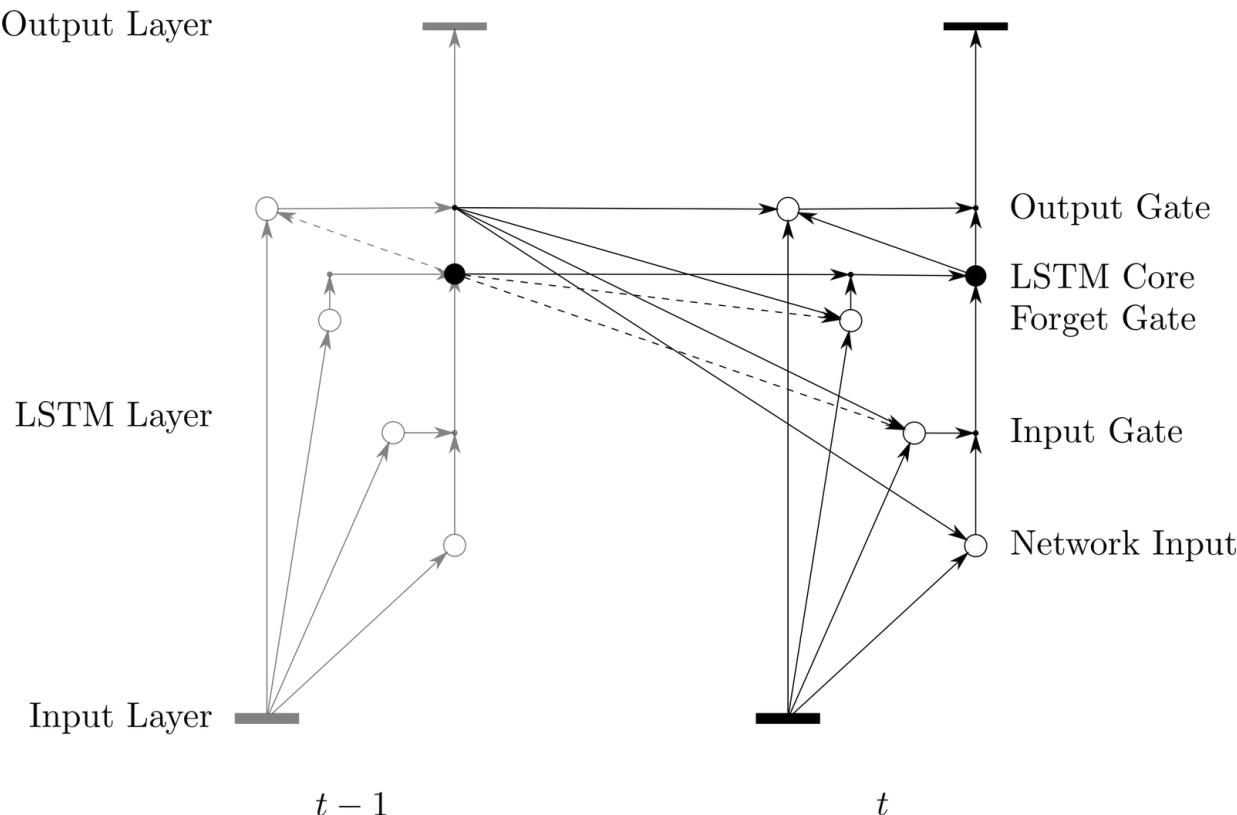
LSTM: Long Short-Term Memory

- A core of the cell stores a value
- Information flow into and out of each cell is controlled by gates
- A forget gate can reset the core's value to 0
- LSTM cells are differentiable memory cells
- LSTM cells overcome the vanishing gradient problem:
 - > Backpropagation for several time-steps is hard
 - > Recurrent NN can therefore not be trained to consider more than just a few time steps

Frinken V, Fischer A, Manmatha R, Bunke H, A Novel Word Spotting Method based on Recurrent Neural Networks, IEEE Trans Pattern Analysis and Machine Intelligence (PAMI), 34(2), pp. 211-24, 2012

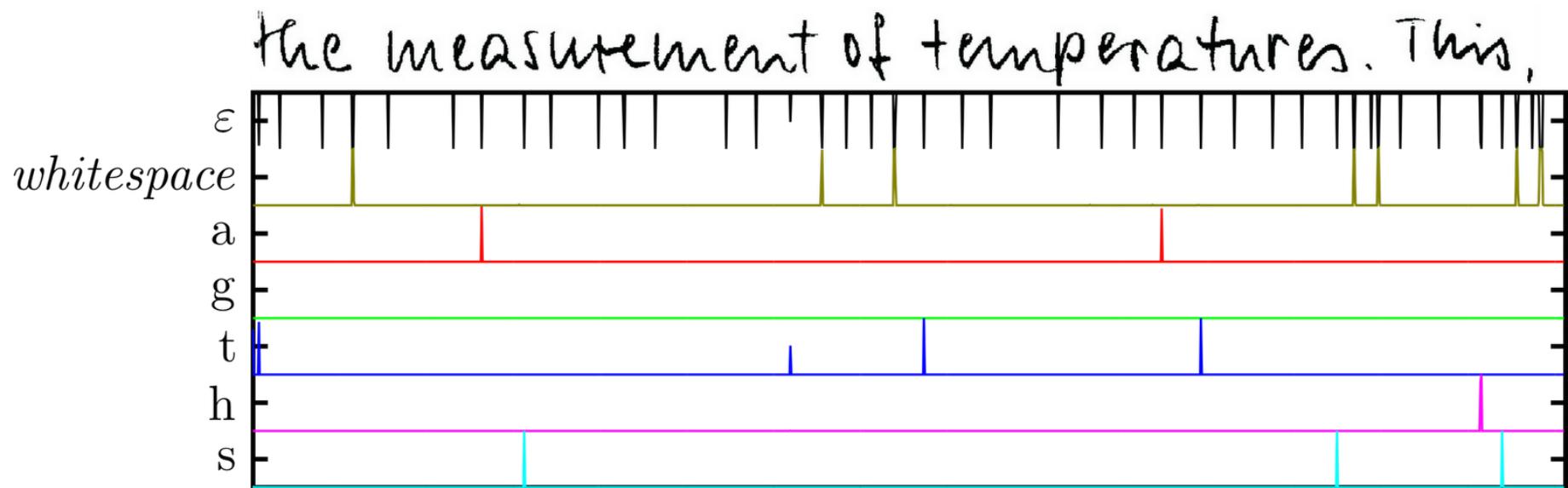
Bidirectional Long-Short Term Memory Neural Networks

The information flow within a LSTM neural network

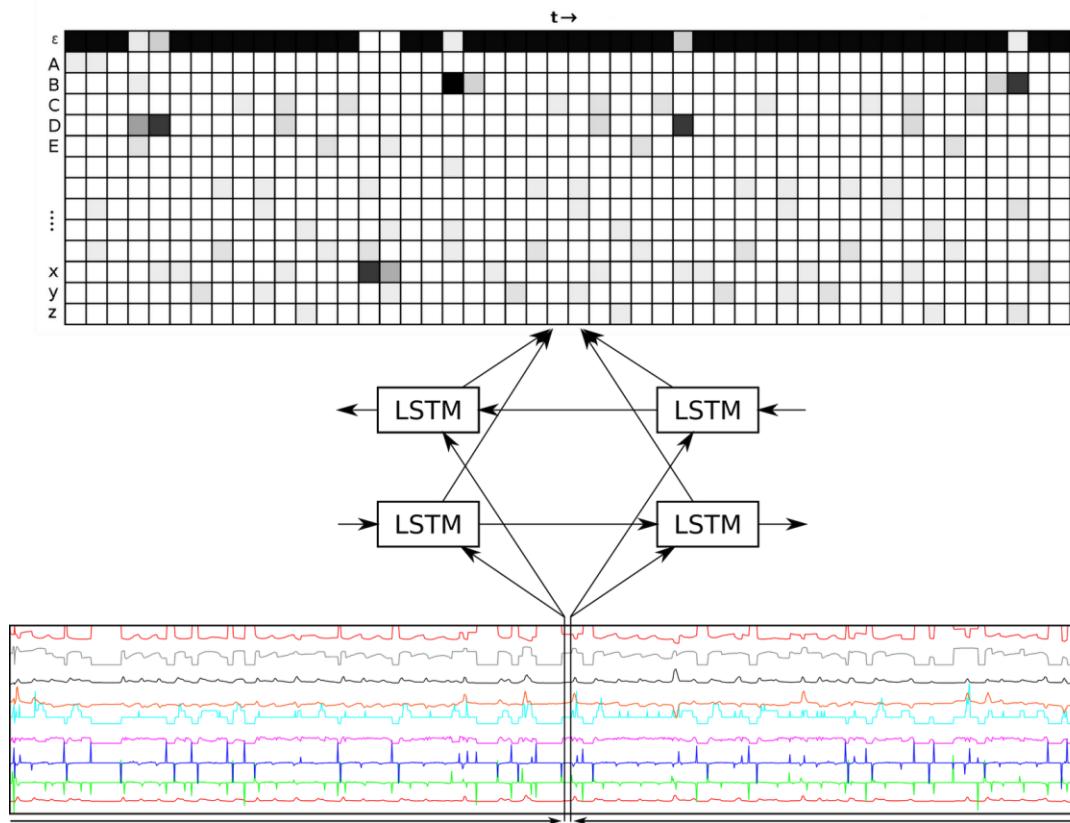


BLSTM NN with CTC Output Layer

The neural network can be trained to detect characters.



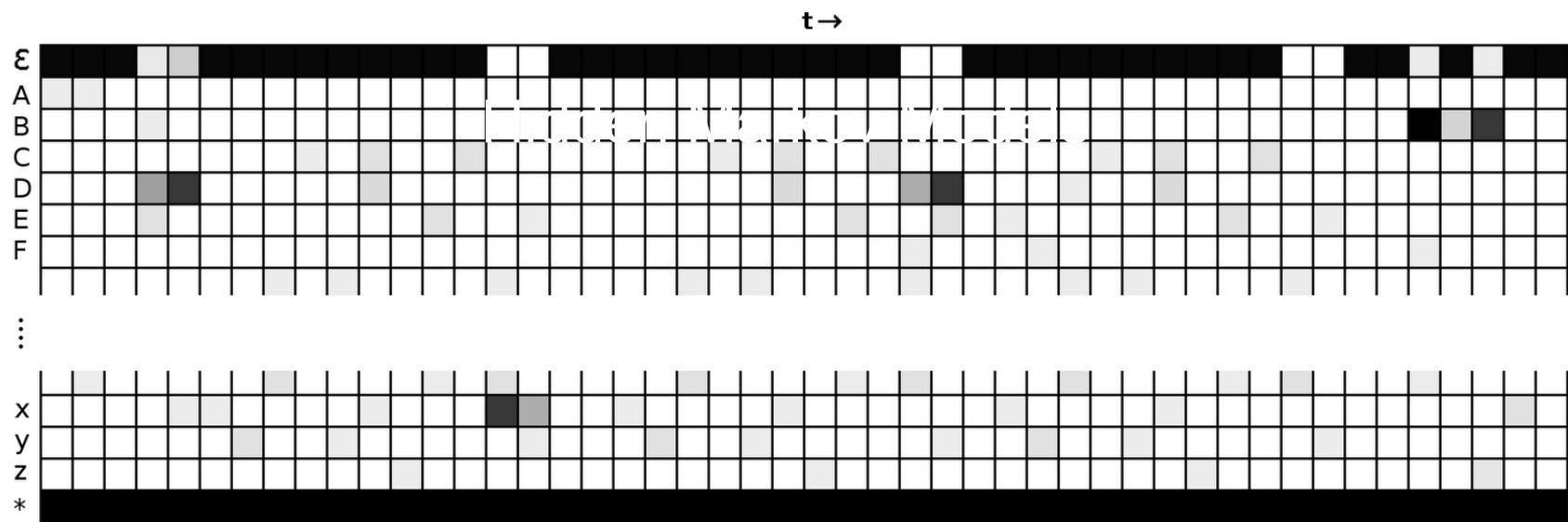
BLSTM NN for Word Spotting



Take home message: For each position, a character probability vector is returned

BLSTM NN for Word Spotting

1. Add a new row to the matrix with a constant activation of 1.0



BLSTM NN for Word Spotting

1. Add a new row to the matrix with a constant activation of 1.0
2. transform the keyword, e.g., **word**:
 2. a) Add the ε -character before and after each character of the search phrase:
 2. b) Add the $*$ -character before and after the new search phrase

word → * ε w ε o ε r ε d ε *

BLSTM NN for Word Spotting

1. Add a new row to the matrix with a constant activation of 1.0
2. transform the keyword, e.g., **word**:
 2. a) Add the ε -character before and after each character of the search phrase:
 2. b) Add the $*$ -character before and after the new search phrase
3. Use dynamic programming to search for the best path in the matrix that follows the characters of the search word
4. The matching score is then the product of the path's nodes

The word is used in the singular

* -
ε -
w -
ε -
o -
ε -
r -
ε -
d -
ε -

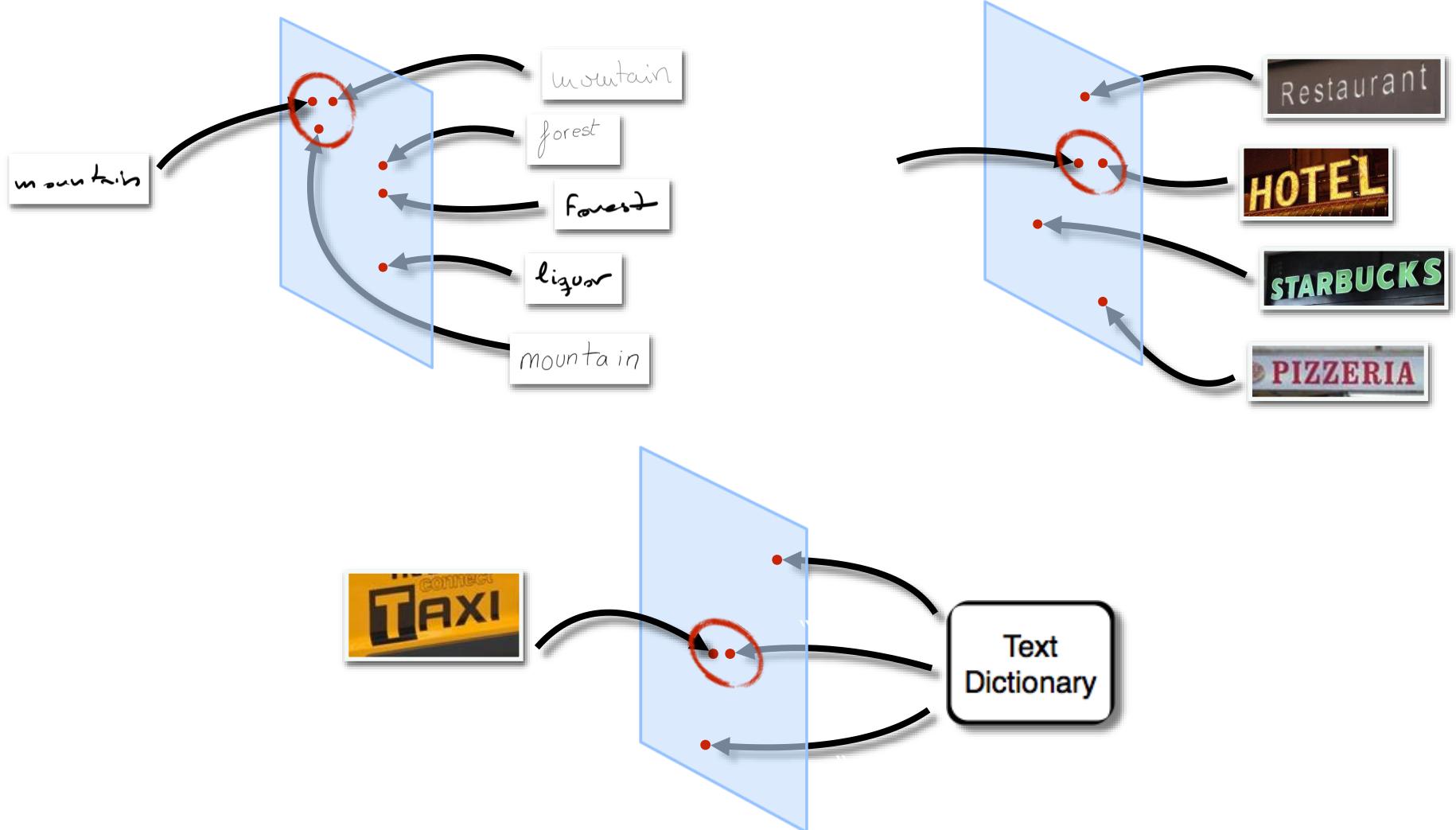
Query-by-Example + Query-by-String approach

Segmentation-based approach

Query-by-string and Query-by-example

J.Almazán, A.Gordo, A.Fornés, E.Valveny. Word Spotting and Recognition with Embedded Attributes. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), volume 36, issue 12, pages 2552-2566, 2014.

Unified Framework

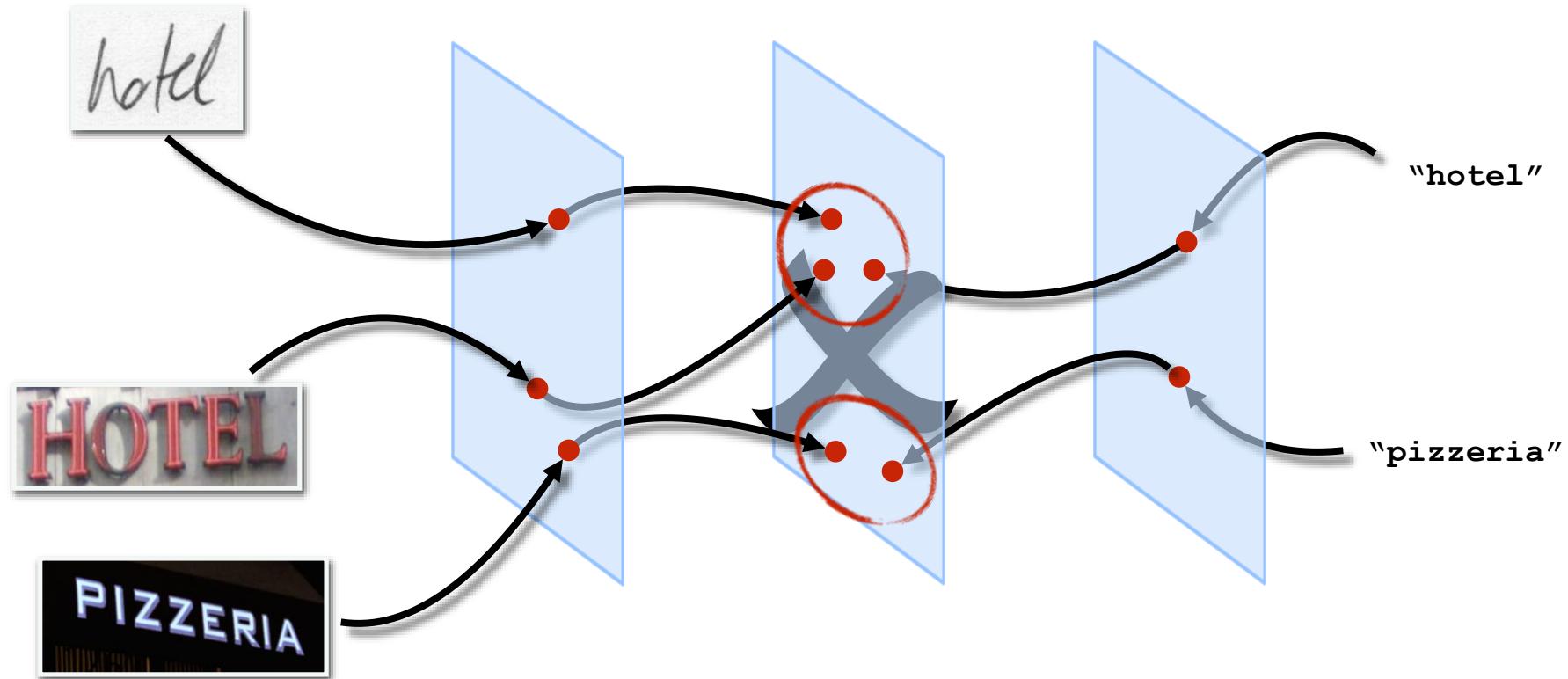


Proposal

Attributes

CCA

PHOC

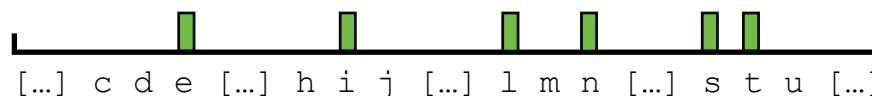


Label Embedding

Proposal: Pyramidal Histogram of Characters (PHOC)

Lvl 1

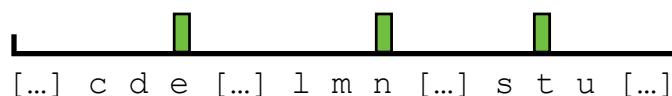
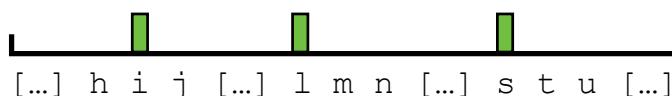
“listen”



Lvl 2

“lis”

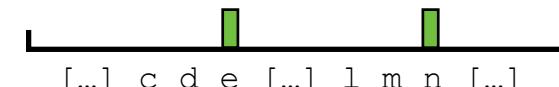
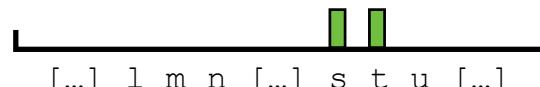
“ten”



Lvl 3
“li”

“st”

“en”



Attribute Embedding

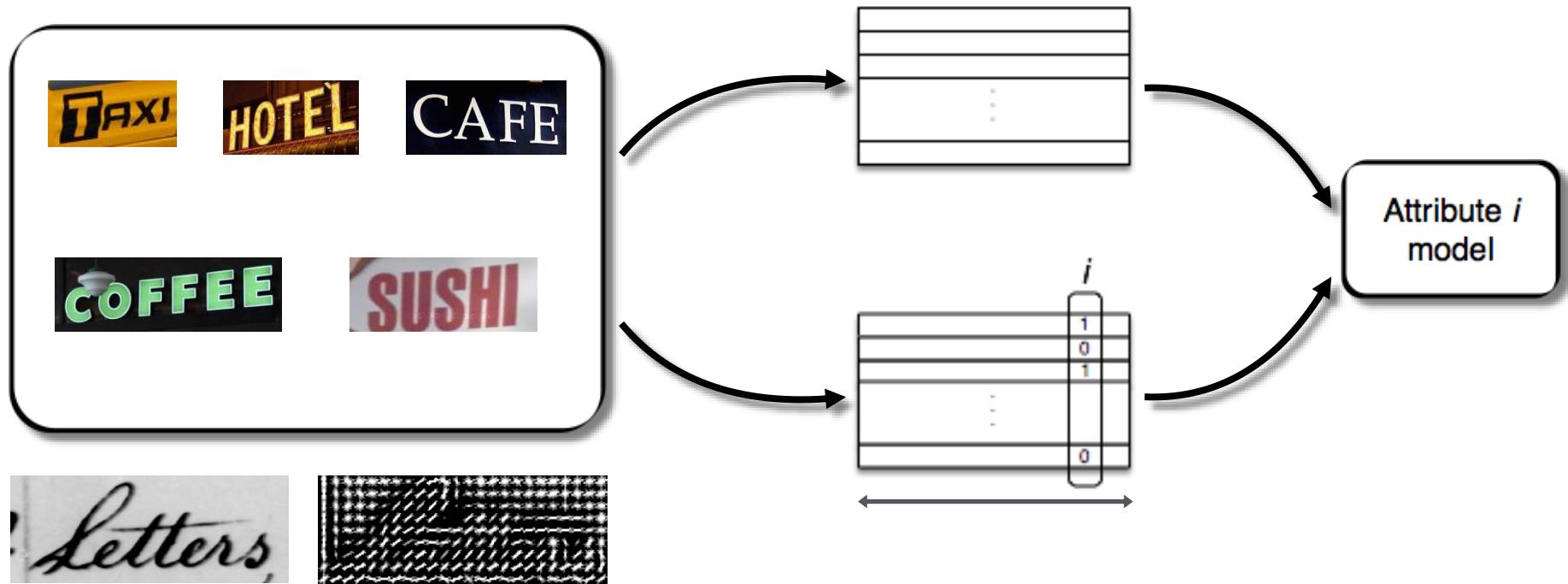


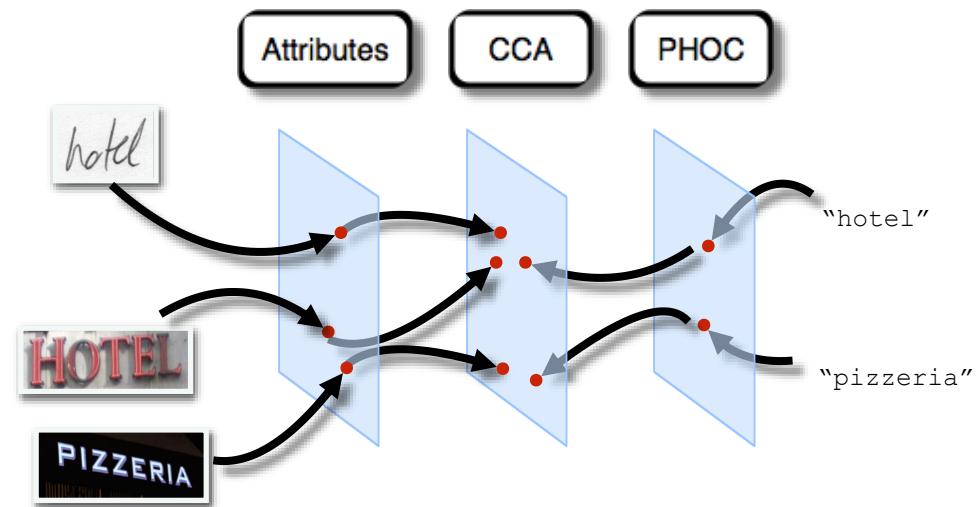
Image representation: Fisher Vector (FV) over SIFT+XY and 2x6 SP
(FV can be seen as a bag of words encoding higher-order statistics)

Attribute model: linear SVM

Common Subspace

Canonical Correlation Analysis (CCA) to learn a common subspace between attribute scores and PHOCs

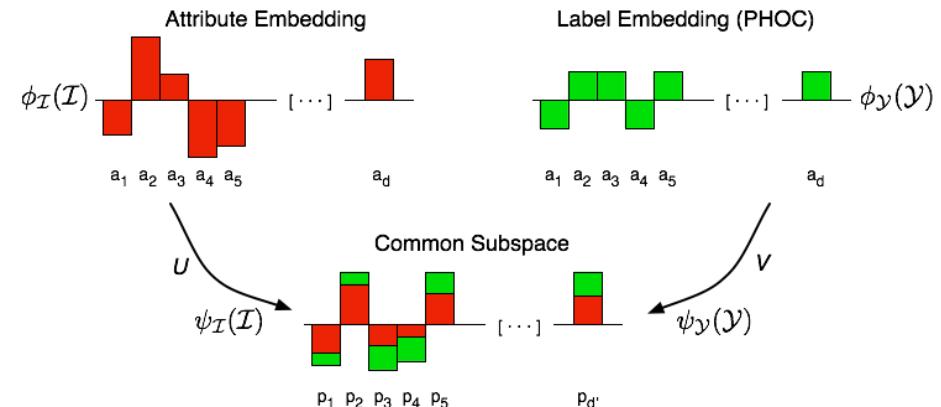
CCA finds a common space where the correlation of the projected views is maximal



Common Subspace

Canonical Correlation Analysis (CCA) to learn a common subspace between attribute scores and PHOCs

CCA finds a common space where the correlation of the projected views is maximal



Three advantages:

Comparison between image and text embeddings is meaningful

Attributes scores of images of the same word are brought together

Dimensionality reduction (**96 dimensions**)

Qualitative Results

| Queries: | Top-5 results: | | | | |
|-------------|----------------|-------------|-------------|-------------|-------------|
| deadly | deadly | dearly | clearly | Deadly | dearly |
| beyond | beyond | beyond | battered | beyond | beyond |
| little | little | little | little | little | little |
| Earth | Earth | earth | earth | earth | Worth |
| towards | towards | towards | towards | towards | towards |
| window | window | written | wild | windows | written |
| attitude | unthinkable | available | attributs | attitude | 'think |
| Advertising | ADVERTISING | Advertising | Advertising | Advertising | advertising |
| WELCOME | WELCOME | WELCOME | WELCOME | Welcome | Welcome |
| Billboards | billboard. | Billboards | BILLBOARD | BILLBOARD | Billboard |
| World | WORLD. | Mobile | WORLD | WORLD | Vozilla |

| | | | | |
|---------------|-----------|----------------|--------------|-----------|
| security | undue | independence | breakfast. | feather |
| security ✓ | undue ✓ | independence ✓ | breakfast ✓ | feather ✗ |
| REGENCY | COTTAGE | SUMMER | VIJAYAWADA | 560 |
| regency ✓ | cottage ✓ | summer ✓ | vijayawada ✓ | 560 ✓ |
| MOTORSPORTS | SUSHI | HOTEL | Man | BURBANK |
| motorsports ✓ | sushi ✓ | hotel ✓ | man ✗ | burbank ✓ |
| STARBUCKS | STUFF | SHOGUN | BOOKSTORE | MEURROS |
| starbucks ✓ | store ✗ | salon ✗ | bookstore ✓ | bimbos ✗ |

Demo application

Segmentation-based WS approach

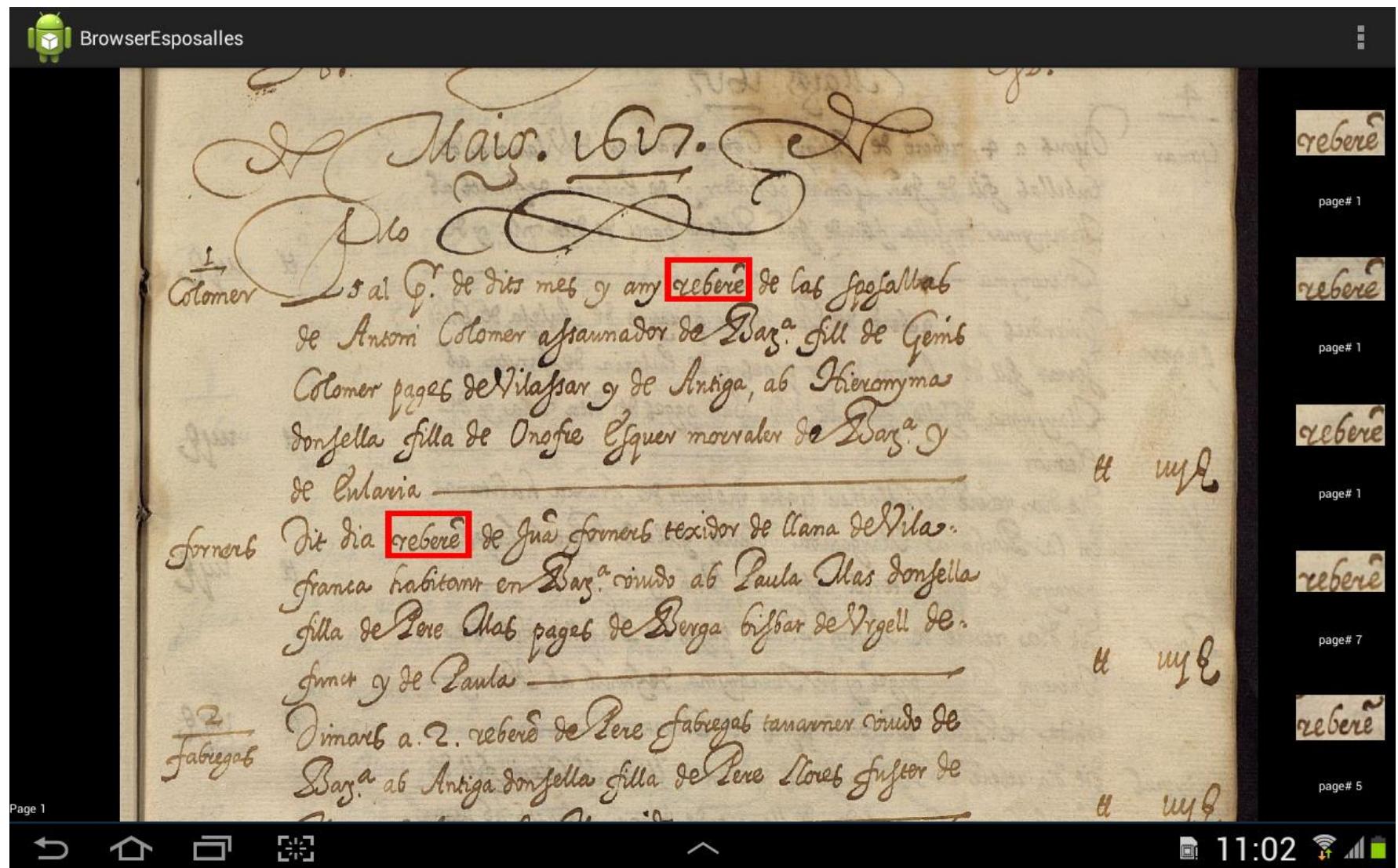
Query-by-string and Query-by-example Combined searches

- AND
- OR

https://www.youtube.com/watch?v=1wUq51t5_tc

P.Riba, J.Almazán, A.Fornés, D.Fernández, E.Valveny, J.Lladós. e-Crowds: a mobile platform for browsing and searching in historical demography-related manuscripts. International Conference on Frontiers in Handwriting Recognition (ICFHR), 2014.

Demo application in Tablet device (Android)



Query-by-String vs Query-by-Example discussion

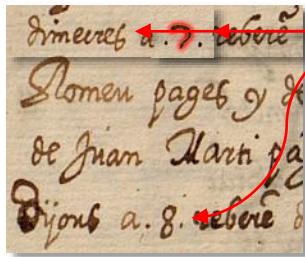
Query-by-string → *USER LOVES IT!* ☺

| Query-by-Example | Query-by-String |
|--|---|
| Need to find one occurrence manually | User types the text query (keyboard) |
| Does not handle different handwriting styles | Can handle different handwriting styles |
| Does not handle grammatical alternatives | Handles grammatical (spelling) alternatives |
| No prior knowledge need | Need to build a writing model |

CONTEXT AWARE WORD SPOTTING

Word Spotting Difficulties

- Efficient strategy in the access by content to historical manuscripts when explicit recognition is not possible.
- **Problem:** Word spotting is usually built based solely on the statistics of local terms which make it **difficult** the recognition in some cases:



If the dash is not a dash, it should be "7".

- **Solution: Contextual information** can provide more relevant information for the recognition of a query word than intrinsic word image information.

The role of the context in visual object recognition

- Classical word spotting is based on the statistics of local terms.
- **Contextual information** can provide more relevant information for the recognition of a query word than intrinsic word image information.
- **Context** of a word in the document can be defined in terms of the other recognized words and their mutual dependences.

Context: The structure of a marriage record

- Date (DI)
- Husband (HI)
- Husband's Parents (HPI)
- Wife (W)
- Wife's Parents (WPI)
- Key-words

Dilluns al p^r de Janer 1601 rebarem delos espous
les de fra^r JULIA pagador dela parroquia de leuane
ras fill de Jo^r JULIA pagos y de ciuicia, ab maria
donqella filla de ramon ferrer pagador q^e y de
Joana

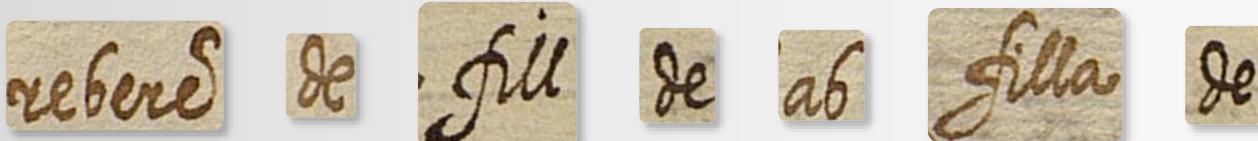
Die d' enero de Poco hica notacion d. ta Salva

Context-aware Word Spotting

Key-words

Word-Spotting

Contextual W-S



Mayo. 1607.
Dijous a. 4. rebere de Miguel Gomar mariner del Vilanova de
Cubellat fill de Juá Gomar pescador y de Enlaria defuntes ab
Hieronymas doncella filla de Juá Dossell pages de dita vila y de
Hieronyma —

dit dia rebere de Sebastia Lloch fuster de La Garriga fill de Garau Lloch
fester y de francesa defuntes ab Elizabeth doncella filla de Bernar
Nadal marxant de Palafrugell del Gavar de Girona y de Manca —
dit dia rebere de Juá Bor pages de Samalús fill de Alanch Bor
pages y de Antonia defuntes ab Antiga doncella filla de Montferran
Busqueta pages de Castellar defuntes y de Marganida —

Mayo. 1607.
Dijous a. 4. rebere de Miguel Gomar mariner del Vilanova de
Cubellat fill de Juá Gomar pescador y de Enlaria defuntes ab
Hieronymas doncella filla de Juá Dossell pages de dita vila y de
Hieronyma —

dit dia rebere de Sebastia Lloch fuster de La Garriga fill de Garau Lloch
fester y de francesa defuntes ab Elizabeth doncella filla de Bernar
Nadal marxant de Palafrugell del Gavar de Girona y de Manca —
dit dia rebere de Juá Bor pages de Samalús fill de Alanch Bor
pages y de Antonia defuntes ab Antiga doncella filla de Montferran
Busqueta pages de Castellar defuntes y de Marganida —

Context-aware Word Spotting

- Applied in documents that presents a **repetitive structure** in the words with text lines along pages.
- Objective: **Word-Spotting using structural information**
 - Discover the words that can be used as **keywords**.
 - Classify the rest of words by **Classes**.
 - Structural **Word Spotting**: find words using the example/string and its class.
 - Speed up the transcription

Keywords

A handwritten document in Spanish with several words highlighted with colored boxes. The highlighted words include "reberé", "fill", "de", "Jennar", "filla", "de", "Antoni", "Janer", "pages", "y", "de", "Eulalia", "de", "Gimuta", "ab", "Maryanna", "do", "Jella", "filla", "de", "Juá", "Sab", "pages", "dita", "vila", "y", "de", "Eleonor". Some words like "filla" and "de" appear multiple times with different bounding boxes.

Classes

The same handwritten document as above, but with words colored according to their class. The colors correspond to the following classes: Day (blue), Joint (pink), Husband (green), Wife (red), and Job (orange). The word "filla" is also highlighted in green.

Words

The handwritten document with words categorized into four main groups: Male name (blue), Surname (orange), Town (red), and Job (green). Below the table, it specifies "Female name" for the word "Eleonor".

| | Male name | Surname | Town |
|----------|-----------|---------|------------------------------------|
| reberé | Pau | Janer | parayre de Antefa de Móz. |
| Jennar | Antoni | Janer | pages y de Eulalia de Gimuta ab |
| Maryanna | do Jella | filla | de Juá Sab pages de dita vila y de |
| Eleonor | | | |

Alignment of records

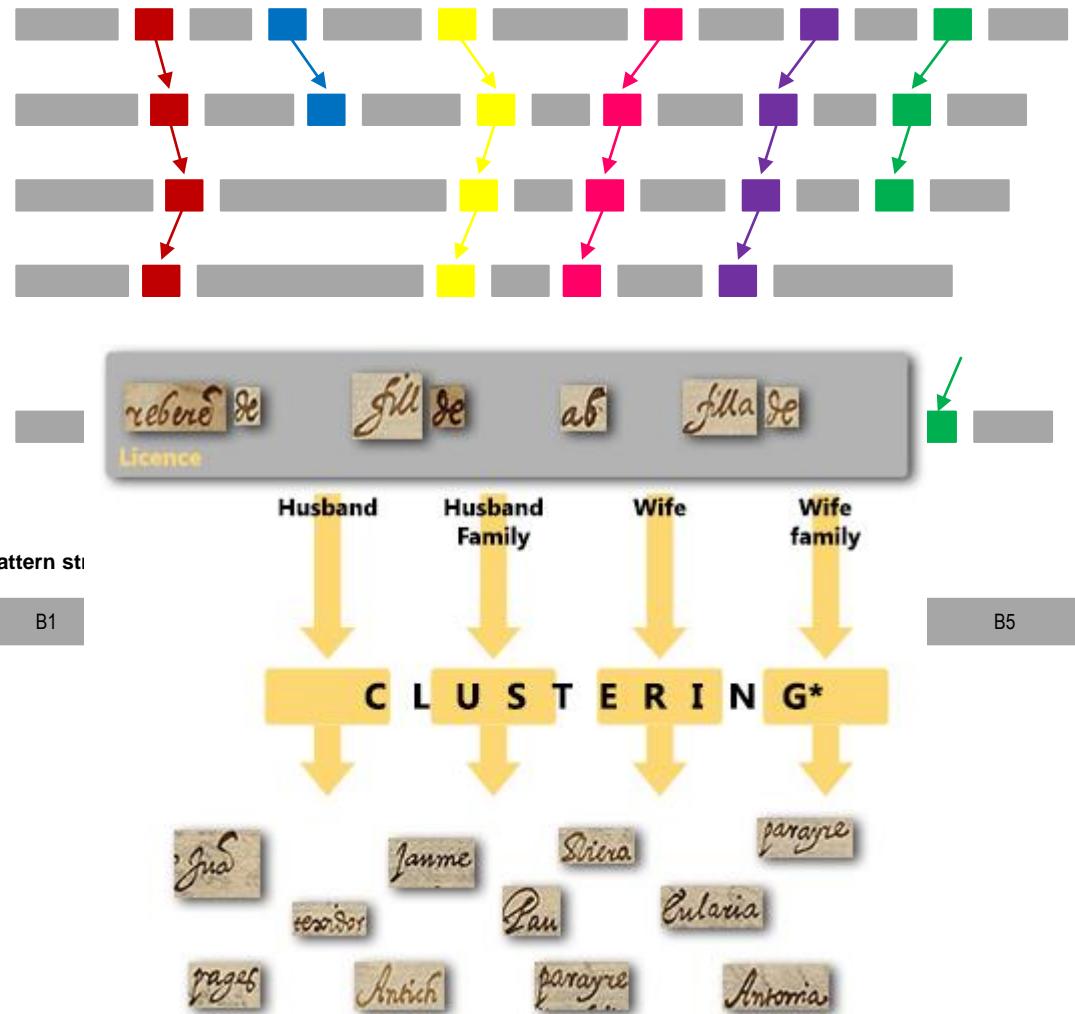
| | Octubre. 1607 — |
|-----------------|---|
| 1 Soler | Al dia rebere de Llorenç Soler labrador habitante en Barc. ^a fill de Llorenç Soler mestre de cases de Granollers y de Maria Juncosa defunta ab Alomia Anna muda de Antoni de Soler oficial real de Barcelona. |
| 2 Callas | dia dia rebere de Juà Llub Callas viudo parayre de Barc. ^a ab Alomia Anna doncella filla de Bernat Torregrosa de Barc. ^a y de Hieronyma defunta. |
| 3 Mauedra | Nt dia rebere de Juà Mauedra parayre de Abriera fill de Juan Mauedra pages y de Francima ab Eulalia doncella filla de Antoni Juà Dafsa pages y de Hieronyma defunta tots de Abriera — |
| 4 Dexon | en la rebere de Juan Dexon pages de franga habitant en Barc. ^a ab Anna Oliva muda de Juan Dexon pages — |
| 5 La torre | de dia rebere de Juan de La torre carmer del regne de Arago habitante en Barc. ^a ab Maria doncella filla de Bernat Sierra pages de frangana de Pinars — y de — |
| 6 Oliveres | de dia rebere de Juà Oliveres pages de Lloja demur viudo ab Alomia doncella filla de Juan Pruna pages del barc. ^a y de Beneta — |
| 7 Mari | Dilhens a. 2. rebere de Llub Mari peixador de Barc. ^a fill de Fran. Mari peixador y de Montferrada ab Solita doncella filla de Pere Colomer treballador de Barc. ^a y de Magdalena defunta — |
| 8 Arguer | de dia rebere de Jaume Arguer pages del regne de franga habitante en Barc. ^a ab Olazgarzón viuda de Pere Cortés pages de Barc. ^a — |
| 9 Clonal | Dijous a. 3. rebere de Juan Clonal pages de Abriera fill de Juan Clonal pages defunto y de Eulalia ab Hieronyma doncella filla de Llorenç Clugments pages de Auleta de Montserrat y de Elizabeth — |
| 10 Glandina | Dilhens a. 6. rebere de Bernat Glandina pages del regne de franga habitante en Gana ab Elizabeth viuda de Quílez Guinó pages — |
| 11 Salamanca | de dia rebere del P. Freror de Salamanca doctor en medicina natural del monestir de Montserrat habitante en Caldes de Montbui viudo ab Esperanza doncella filla de Antich Gibert canstader de cera de Caldes de Montbui y de Angela defunta — |

dit dia rebere de Juà Mauedra parayre de Abriera fill de Juan Mauedra pages y de Francima ab Eulalia doncella filla de Antoni Juà Dafsa pages y de Hieronyma defunta tots de Abriera —

Dilhens a. 2. rebere de Llub Mari peixador de Barc.^a fill de Fran. Mari peixador y de Montferrada ab Solita doncella filla de Pere Colomer treballador de Barc.^a y de Magdalena defunta —

Extraction of key and frequent words

| Octubre. 1607 — | |
|-----------------|---|
| 1 Soler | Al gr ^o de diez reber ^e de Llores. Soler sacerdote librador habitante en Bar ^a fill de Sacerdote. Soler mestre de capes de Granollers y de Mor. Jernada defuntes ab Alomia Anna viuda de Anton de Mor oficial real de Zar ^a . |
| Callas | Die dia rebere de Juá Luis Callas viudo parayre de Bar ^a ab Alomia doncella filla de Bernat Torreg ^a sacerdote de Bar ^a y de Hieronyma defuntes. |
| Mauera | Die dia rebere de Juá Maudra parayre de Arenys fill de Juan Mauera pagés y de Hieronyma ab Eulalia doncella filla de Anton Juá Dacha pagés y de Hieronyma defuntes tots d'Arenys — |
| Dexon | Die dia rebere de Juan Dexon pagés de granja habitante en Sag ^a ab Alomia doncella filla de Bernat Sierra pagés de Vilafamés del Pinardejo y de e ^o . |
| R La torre | Die dia rebere de Juan de La torre carmer del regne de Arago habitante en Sag ^a ab Alomia doncella filla de Bernat Sierra pagés de Vilafamés del Pinardejo y de e ^o . |
| Olivares | Die dia rebere de Juá Olivares pagés de Lloja demur viudo ab Alomia doncella filla de Juan Pruna pagés del fan ^o y de Beneta — |
| 2 Mari | Vilamit a 2. rebere de Luis Alom pagador de Sag ^a fill de Juan Alom pagador y de Alom Jernada ab Juá doncella filla de Pere Colomer treballador de Sag ^a y de Alagadona defuntes — |
| Arguer | Die dia rebere de Juana Arguer pagés del regne de granja habitante en Sag ^a ab Olazgarzón viuda de Pere Cortés pagés de Sag ^a — |
| 5 Coral | Dijous a 3. rebere de Juan Moral pagés de Cervera fill de Juan Moral pagés defuntes y de Culiana, ab Hieronyma doncella filla de Sacerdote — |
| 6 Gandia | Vilamit a 6. rebere de Bernat Gandia pagés del regne de granja habitante en Gana ab Elizabeth viuda de Guillermo Guirao pagés — |
| Salamanca | Die dia rebere del P. Georri de Salamanca doctor en medicina natural del monestir de Olomífera habitante en Caldes de Olomífera viudo ab Speranza doncella filla de Antich Gisbert canstalor de cera de Caldes de Olomífera y de Angela defuntes — |



(*). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise in Second International Conference on Knowledge Discovery and Data Mining (1996), pp. 239-241 by Martin Ester, Hans-P. Kriegel, Jorg Sander, Xiaowei Xu edited by Evangelos Simoudis, Jiawei Han, Usama Fayyad

CONCLUSIONS

Conclusions

- **Segmentation-based vs. Segmentation-free** methods
 - Accuracy of word segmentation methods?
- **Learning-free vs. Learning-based** methods (closer to recognition)
 - Do we have enough transcribed data?
- **Query-by-example vs. Query-by-class vs. Query-by-string**
 - Can we learn the appearance of words?
- **Context aware Word Spotting**
 - Discovers the contextual information (structure).
 - Tool for semi-automatic transcription (most frequent words)
 - Context can improve the methods (controlled conditions)

Achievements

Such scenarios are kind of solved...

Single writer + Query-by-example + presegmented small collections

Multiple writers + large amount of training data

Query-by-string + large amount of training data

Challenges

- Having multiple writers in query-by-example scenarios
- Query-by-string when few or no training data is available
- Segmentation-free methods
- The use of context
- Scalability of the methods
- Non-latin scripts / shapes / signatures?
- Use of relevance feedback ?
 - Role of the User?



Obrigada!

July 2015

Alicia Fornés
aforres@cvc.uab.es

VISUM Summer School



VISion Understanding and Machine intelligence