Statistical Models for Word Spotting

Gernot A. Fink

TU Dortmund University, Department of Computer Science

Workshop *Machines and Manuscripts 2015*
Karlsruhe, February 20

- Introduction *Why Word Spotting?*
- Evolution of Methods *... towards statistical models*
- BoF-Based Word Spotting
- BoF Hidden-Markov Models *... for Advanced Modeling*
- Summary

Joint work with *Leonard Rothacker*

Introduction: Automatic Reading Systems

State of Automatic Reading:

- One of the earliest application fields studied in computer science
- So-called OCR achieves high-quality results for machine-printed text in well-defined settings.
- Online handwriting recognition again gaining popularity
- Offline handwriting recognition: Remarkable results, but still an open research problem

General Methodology:

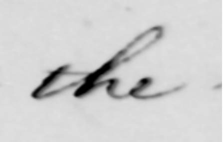
Statistical sequence models (usually Hidden-Markov Models) that are trained from *extensive* amounts of example data

Introduction: Why Word Spotting?

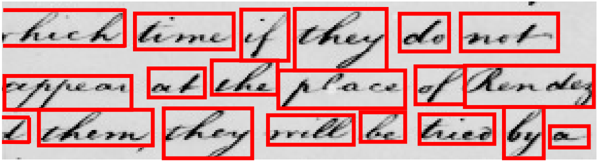
What if automatic transcription of handwriting is no longer feasible?

Alternative: Retrieval of individual words rather than transcription ("query-by-example")

Query word image



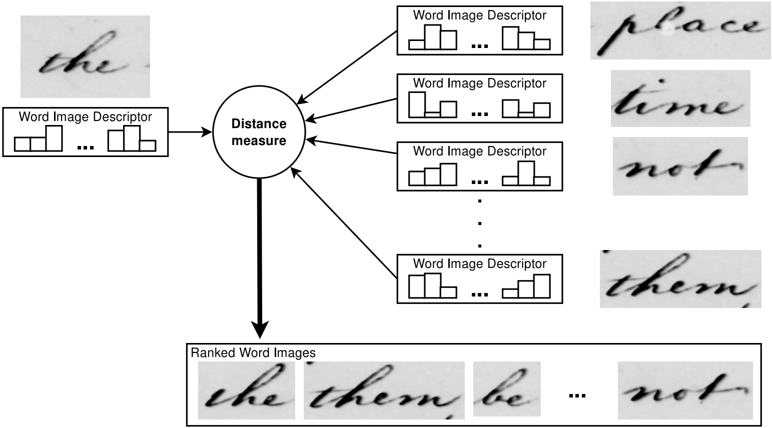
Document image



Images from *The George Washington Papers at the Library of Congress, 1741-1799*

Introduction: Why Word Spotting?

Query-by-example word spotting



Based on [Rath & Manmatha, IJRAR'07]

Word Spotting: Fundamentals

Core Methodology:

- Specialized image retrieval
- Important ingredient: Image matching procedure
- Frequently required: Pre-segmentation (words / lines)

Taxonomy:

- *Segmentation-based*
- *Segmentation-free*, i.e., segmentation problem covered during retrieval
- *Query-by-Example*, i.e., word image directly used as query
- *Query-by-String*, i.e., query model derived from textual query ("string")

Evolution of Methods

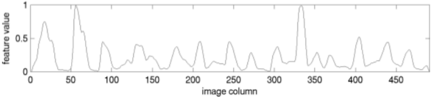
Method: "Indexing Handwriting Using Word Matching"

- Binarization
- Word image alignment
- XOR distance
- Euclidean distance mapping

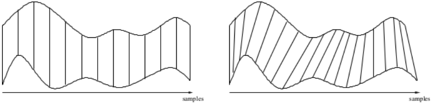
R. Manmatha, Chengfeng Han, E. M. Riseman, and W. B. Croft. 1996. Indexing handwriting using word matching. In Proc. of the first ACM Int. Conf. on Digital libraries (DL '96), Edward A. Fox and Gary Marchionini (Eds.). ACM, New York, NY, USA, 151-159.

Method: Sequence-based Word Image Matching

(a) original image: slant/skew/baseline-normalized, cleaned.



(b) normalized projection profile.



(a) naive alignment after resampling, (b) alignment with DTW.

- Baseline, skew, slant normalization
- Sequence of feature vectors: upper lower word profiles, ink background transitions
- DTW sequence matching

Rath, T. M., and Manmatha, R. Word image matching using dynamic time warping. In Proc. of the Conf. on Computer Vision and Pattern Recognition (Madison, WI, June 18-20 2003), vol. 2, pp. 521-527.

Method: Segmentation-free Word Spotting with Bag-of-Features

- Dense grid of SIFT descriptors [Lowe 2004]
- Bag-of-Features-based Spatial-Pyramid feature representation (originally proposed for scene recognition [Lazebnik *et al.* 2006])
- Encoded in topic space and matched with cosine distance
- Patched-based spotting framework

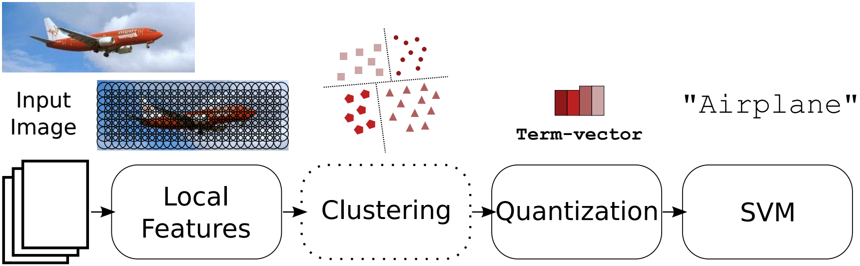
Rusiñol, M., and Aldavert, D., and Toledo, R., and Lladós, J., Browsing heterogeneous document collections by a segmentation-free word spotting method. Proceedings of the International Conference on Document Analysis and Recognition, pages 63-67, 2011.

BoF-Based Word Spotting

Bag-of-Features Models: Fundamentals

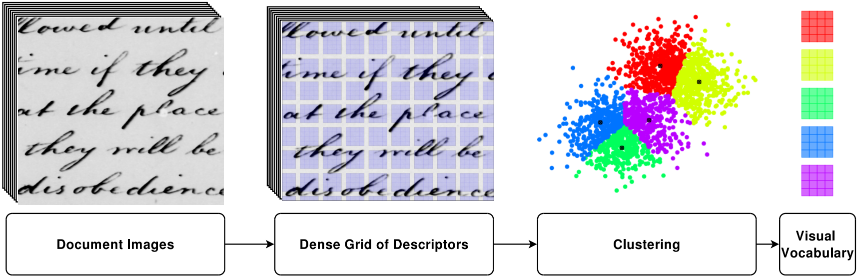
Idea based on Bag-of-Words models proposed for text categorization

BoF-Approach developed for Computer Vision

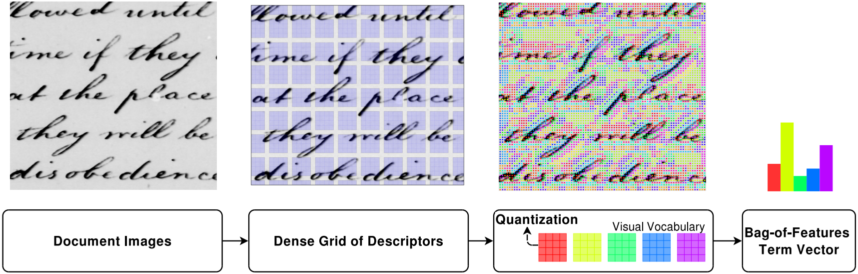


Advantage: Simple but powerful models for visual appearance

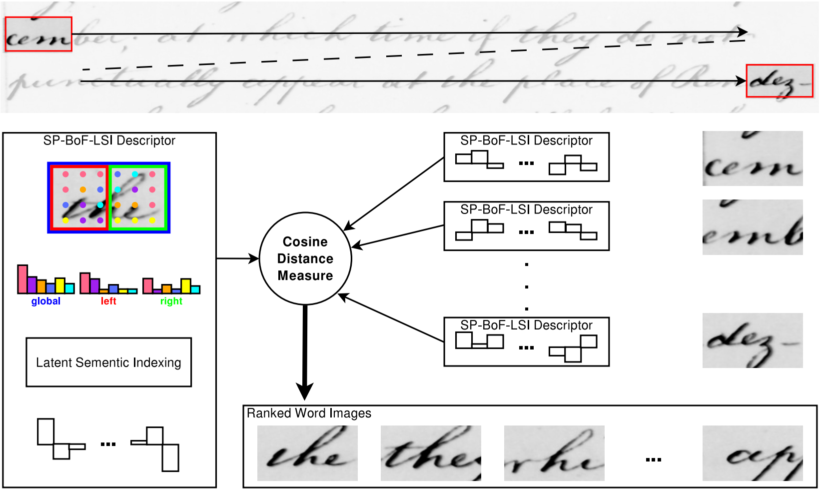
Bag-of-Features for Word Spotting



Bag-of-Features for Word Spotting II



Segmentation-free Word Spotting with Bag-of-Features

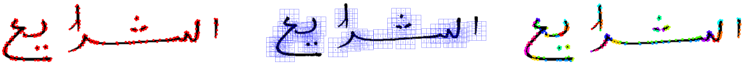


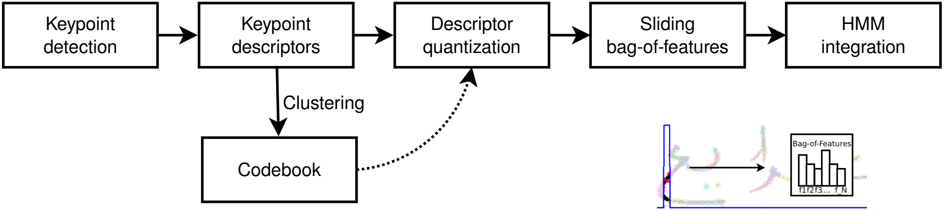
Rusiñol, M., and Aldavert, D., and Toledo, R., and Lladós, J., Browsing heterogeneous document collections by a segmentation-free word spotting method. Proceedings of the International Conference on Document Analysis and Recognition, pages 63-67, 2011.

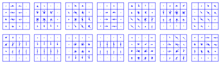
BoF Hidden-Markov Models

Bag-of-Features HMMs

- *Extension* of HMMs towards learned feature representation
- *Extension* of BoF models towards fine-grained script representation



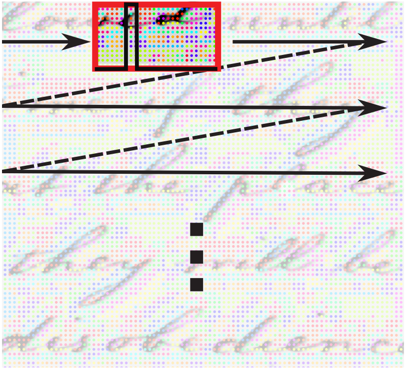




$$b_j(\boldsymbol{f}) = \sum_{k=1}^{|V|} c_{jk} f_k \quad \text{with} \quad \begin{array}{l} \boldsymbol{f} : \text{term vector} \\ V : \text{vis. voc.} \end{array}$$

Rothacker, L., Vajda, S., Fink, G. A.: *Bag-of-Features Representations for Offline Handwriting Recognition Applied to Arabic Script*, In Proc. ICFHR, Bari, 2012.

Bag-of-Features HMMs for Word Spotting



Approach can be sped-up by intelligent patch pre-filtering!

Rothacker, L., Rusinol, M., Fink, G. A.: *Bag-of-Features HMMs for Segmentation-Free Word Spotting in Handwritten Documents*, In Proc. Int. Conf. on Document Analysis and Recognition, Washington DC, USA, 2013.

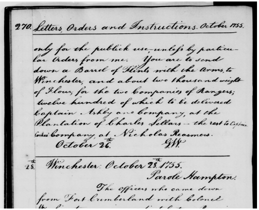
Rothacker, L., Rusinol, M., Llados, J., Fink, G. A.: *A Two-Stage Approach to Segmentation-Free Query-by-Example Word Spotting*, manuscript cultures, 1(7), pages 47-57, 2014.

Segmentation-free QbE Word Spotting: Evaluation

George Washington Benchmark:

Database of handwritten letters

- Limited number of writers
- 20 pages, 4860 words
- Every word acts as query
- Annotated on word level (w/o punctuation / case)


270. Letters Orders and Instructions. October 1755.

only for the publick use,—unless by particu-
lar Orders from me. You are to send
down a Barrel of Flints with the Arms, to
Winchester, and about two thousand weight
of Flour, for the two Companies of Rangers;
twelve hundred of which to be delivered
Captain Ashby and Company, at the
Plantation of Charles Sellars — the rest to Captain
Cocke's Company, at Nicholas Reasmers.
October 26th. GW.

28th. Winchester: October 28th 1755.
Parole Hampton.
The officers who came down
from Fort Cumberland with Colonel

Method	Model	Patch size	mAP	mR
Rusiñol et al. 2011	BoF+LSI	fixed	30.4%	71.1%
Almazán et al. 2014	HoG	query dep.	59.5%	-
Almazán et al. 2014	HoG+QE+RR	query dep.	68.9%	-
Rothacker et al. 2014	IFS+BoF-HMM	query dep.	**71.6%**	**96.3%**

Word Spotting Methods Revisited

Approach considered so far: *Query-by-Example Word Spotting*

Main Advantage / Disadvantage:

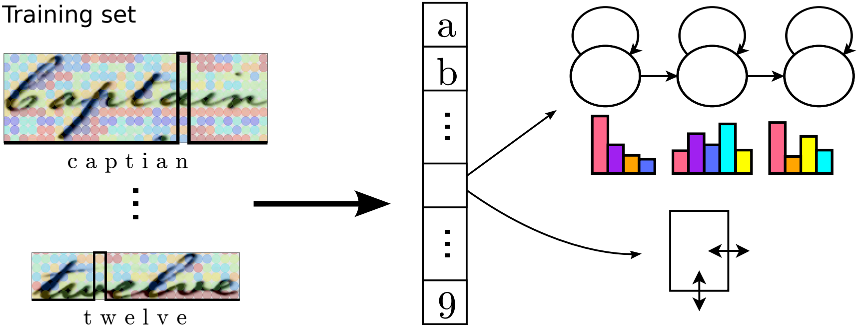
- ✓ No training / annotation required!
- ↯ Queries can only be *selected*.

What about arbitrary queries?

Requirement: Create appearance model of query from *textual* input (= "string")

$\Rightarrow$ *Query-by-String Word Spotting*

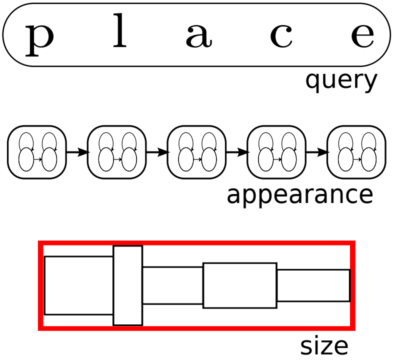
Bag-of-Features HMMs for Query-by-String Word Spotting

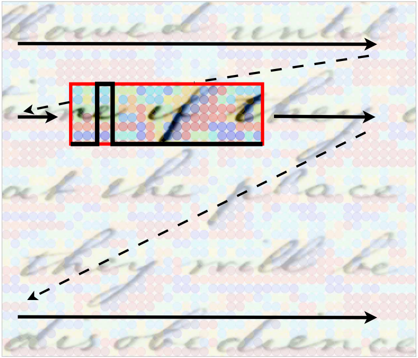


- Appearance model: Bag-of-Features HMMs
- Spatial size model: Width & height estimates

Rothacker, L., Fink, G. A.: *Segmentation-Free Query-by-String Word Spotting with Bag-of-Features HMMs*, Int. Conf. on Document Analysis and Recognition, Tunis, Tunisia, 2015, submitted.

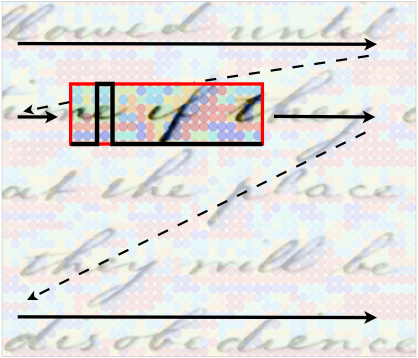
Bag-of-Features HMMs for Query-by-String Word Spotting

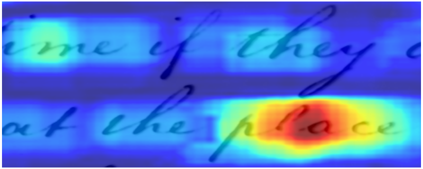


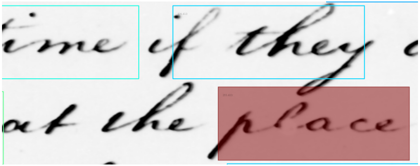


Rothacker, L., Fink, G. A.: *Segmentation-Free Query-by-String Word Spotting with Bag-of-Features HMMs*, Int. Conf. on Document Analysis and Recognition, Tunis, Tunisia, 2015, submitted.

Bag-of-Features HMMs for Query-by-String Word Spotting





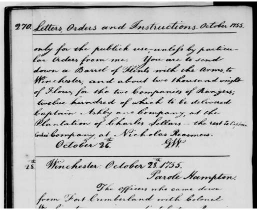


Rothacker, L., Fink, G. A.: *Segmentation-Free Query-by-String Word Spotting with Bag-of-Features HMMs*, Int. Conf. on Document Analysis and Recognition, Tunis, Tunisia, 2015, submitted.

Segmentation-free QbS Word Spotting: Evaluation

George Washington Benchmark:
Database of handwritten letters:

- … as detailed before, now with …
- Cross Validation: 15-5 / 5-15


270. Letters Orders and Instructions. October 1755.

only for the publick use,—unless by particu-
lar Orders from me. You are to send
down a Barrel of Flints with the Arms, to
Winchester, and about two thousand weight
of Flour, for the two Companies of Rangers;
twelve hundred of which to be delivered
Captain Ashby and Company, at the
Plantation of Charles Sellars — the rest to Captain
Cocke's Company, at Nicholas Reasmers.
October 26th. GW.

28th. Winchester: October 28th 1755.
Parole Hampton.
The officers who came down
from Fort Cumberland with Colonel

Method	Evaluation	Segmentation	mAP	mR
Rothacker, Fink 2015	15-5	free (50% ovl.)	77.7 %	92.6 %
Rothacker, Fink 2015	15-5	free (25% ovl.)	**80.3** %	**98.8** %
Rothacker, Fink 2015	5-15	free (25% ovl.)	58.2 %	97.5 %
Aldavert et al. 2013	15-5	word-level	56.5 %	
Almazan et al. 2014	15-5	word-level	91.1 %	
Frinken et al. 2012	10-5-5 / ivw	line-level	71.0 %	
Fischer et al. 2010	10-5-5 / ivw	line-level	60.0 %	

Summary

Method: Segmentation-free *Query-by-Example Word Spotting*

- = Patch-based decoding + stochastic query model (BoF-based)
- ✓ Impressive results for isogeneous document collections
- ✓ No annotation effort required!
- ϟ Poor generalization to different writing styles
- ϟ No truly user-defined queries possible

Alternative: *Query-by-String Word Spotting*

- ✓ Any (textual) query can be used
- STOP Requires general script appearance model (≈ "transcription model")!
- ϟ Consequently requires *at least some* annotated training data!

Further Reading

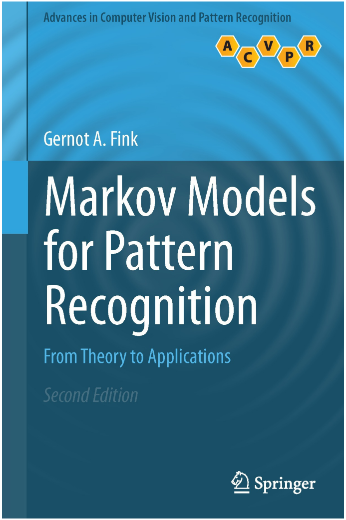
Textbook:

Gernot A. Fink:
Markov Models for Pattern Recognition. Springer, 2014.

Survey Articles:

Josep Lladós *et al.*:
On the Influence of Word Representations for Handwritten Word Spotting in Historical Documents. *IJPRAI*, 26(5), 2012.

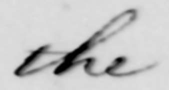
Thomas Plötz & Gernot A. Fink:
Markov Models for Offline Handwriting Recognition: A Survey. *IJDAR*, 12(4), 2009.



Appendix

Segmentation-free Word Spotting: Evaluation

Query word image



Retrieved patches sorted by score

Patch relevance threshold: For example 50% ground truth overlap

- Precision: How relevant is the list?
- Recall: How complete is the list w.r.t. relevant items?
- Average Precision: How well is the retrieval list sorted?

Usually mean values over many queries are reported (mAP and mR).

Attention: Precision and Average Precision only take the list into account!

Segmentation-free Word Spotting: Evaluation II

Average Precision: How well is the retrieval list sorted?

- Let's make the example a little more complex:

$$[1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1]$$

  Total number of relevant items in dataset: 10
- Precision: $\frac{8}{15} \approx 0.53$, Recall: $\frac{8}{10} = 0.8$
  No information about the list's order!
- Average Precision: Precision averaged over all recall levels:

$$\frac{\sum_{k=1}^{n} \mathsf{Precision}_k \times \mathsf{rel}(k)}{\sum_{k=1}^{n} \mathsf{rel}(k)}$$

  $\mathsf{rel}(k)$: Relvance of item k, $\mathsf{Precision}_k$: Precision at cut-off k

  Accumulate Precision whenever the Recall changes (and normalize): $\mathsf{rel}(\cdot)$

Segmentation-free Word Spotting: Evaluation III

$$[1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1]$$

Average Precision:

$$\frac{\sum_{k=1}^{n} \mathrm{Precision}_k \times \mathrm{rel}(k)}{\sum_{k=1}^{n} \mathrm{rel}(k)}$$

$$\frac{\frac{1}{1} + \frac{2}{3} + \frac{3}{4} + \frac{4}{5} + \frac{5}{7} + \frac{6}{11} + \frac{7}{12} + \frac{8}{15}}{1+1+1+1+1+1+1+1}$$

$$\approx 0.7$$