technische universität
dortmund

Statistical Models for Handwriting Recognition and Retrieval

— ICFHR 2014 Tutorial, Crete —

Gernot A. Fink & Leonard Rothacker

TU Dortmund University, Dortmund, Germany

September 1, 2014

- Introduction
- Markov Model-Based Handwriting Recognition ... *Fundamentals*
- Word Spotting for Handwriting Retrieval ... *Models & Methods*
- Summary ... *and Further Reading*

Machine Reading

- One of the earliest applications considered in Computer Science (since 1950s)
- Addresses mainly document images (offline) with special case of online documents
- Machine printed vs. handwritten script
- Mature machine-reading techiques exist, e.g.:
  - For reading machine-printed documents
  - For special application areas (e.g. postal automation)

⇒ Most challenging: Handwriting!

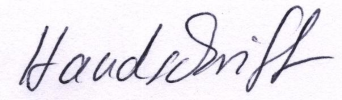
Challenges in Handwriting Recognition

- Considerable freedom in script appearance

  Typical handwriting $\hat{=}$ cursive writing

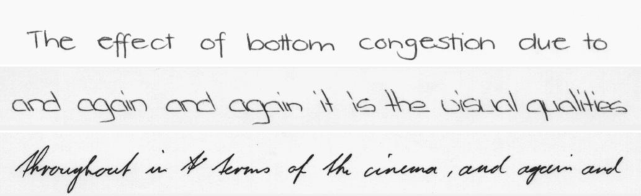
  Also: "hand printed" characters

  Mostly: Combination $\hat{=}$ *unconstrained* ...



- Large Variability of individual symbols
  - Writing style

[Image source: IAM-DB]



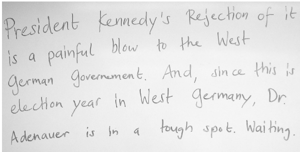
  - Stroke width and quality
  - Considerable variations even for the same writer!

- Segmentiation problematic (especially for cursive writing)

  "Merging" of neighboring symbols

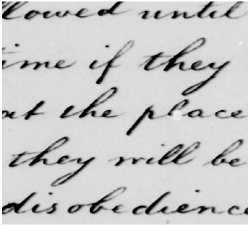
Tasks in Handwriting Recognition

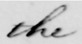
Document Transcription (= “classical” recognition)

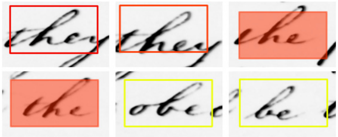


President Kennedy is Rejection of it
is a painful blow to the West
German Government And since this is
selection year in West Germany of
Adenauer is in a tough spot waiting

Document Retrieval (aka “Word Spotting”)







Focus of this Tutorial

Processing type: Offline (documents captured by scanner or camera)

Script type & Writing style:

- Alphabetic scripts, especially Roman script
- *No* restriction w.r.t. writing style, size etc.
  ⇒ Unconstrained handwriting!

Methods: Statistical Recognition Paradigm

- Markov Models for segmentation-free recognition
- Bag-of-Features Models for segmentation-free retrieval

Goal: Understand ...

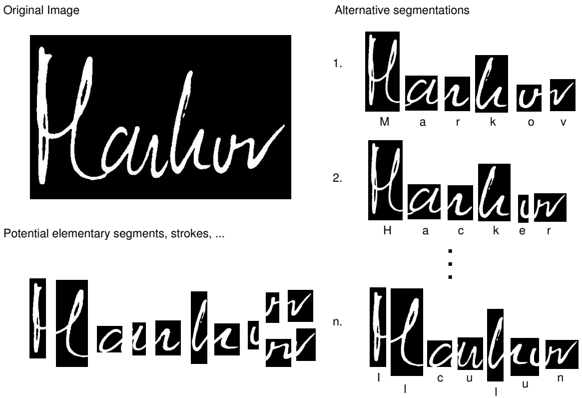
- ... concepts and methods behind Hidden Markov Models and Bag-of-Features models *and* ...
- ... how these are applied in handwriting recognition and retrieval.

⇒ *Recognize advantages of statistical methods in machine reading!*

Overview

"Traditional" Recognition Paradigm

Segmentation + Classification:



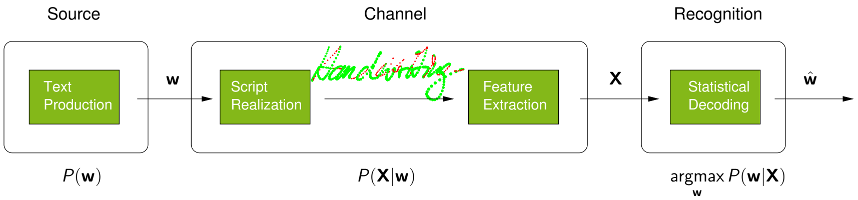
✓ Segment-wise classification possible using various standard techniques

↯ Segmentation is
- costly,
- heuristic, and
- needs to be optimized manually

↯ *Segmentation is especially problematic for unconstrained handwriting!*

Statistical Recognition Paradigm: The Channel Model

(Model originally proposed for automatic speech recognition)



Wanted: Sequence of words/characters $\hat{\mathbf{w}}$, which is most probable for given signal/features $\mathbf{X}$

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmax}}\, P(\mathbf{w}|\mathbf{X}) = \underset{\mathbf{w}}{\operatorname{argmax}} \frac{P(\mathbf{w})P(\mathbf{X}|\mathbf{w})}{P(\mathbf{X})} = \underset{\mathbf{w}}{\operatorname{argmax}}\, P(\mathbf{w})P(\mathbf{X}|\mathbf{w})$$

The Channel Model II

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmax}} P(\mathbf{w}|\mathbf{X}) = \underset{\mathbf{w}}{\operatorname{argmax}} \frac{P(\mathbf{w})P(\mathbf{X}|\mathbf{w})}{P(\mathbf{X})} = \underset{\mathbf{w}}{\operatorname{argmax}} P(\mathbf{w})P(\mathbf{X}|\mathbf{w})$$

Two aspects of modeling:

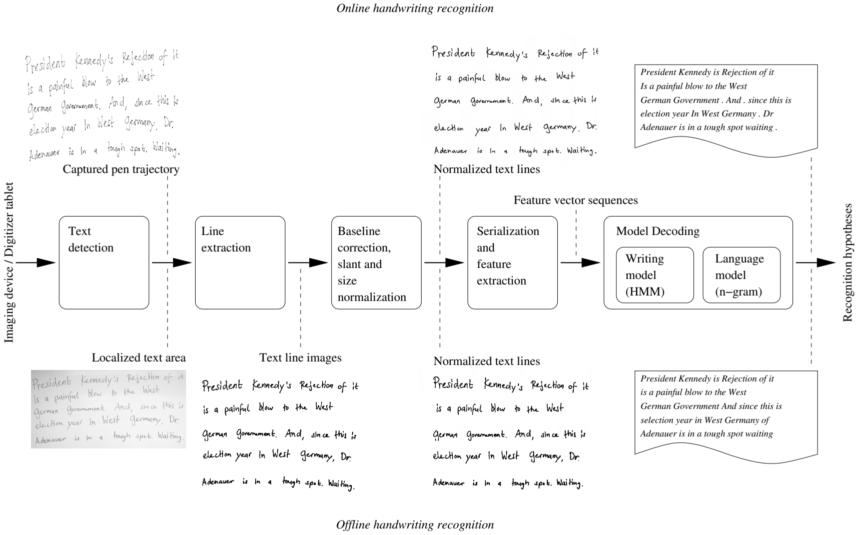
- Script (appearance) model: $P(\mathbf{X}|\mathbf{w})$ $\Rightarrow$ Representation of words/characters
  Hidden-Markov-Models
- Language model: $P(\mathbf{w})$ $\Rightarrow$ Restrictions for sequences of words/characters
  Markov Chain Models / n-Gram-Models

Specialty: Script or trajectories of the pen (or features, respectively) interpreted as *temporal* data

✓ Segmentation performed implicitly! $\Rightarrow$ "segmentation free" approach

STOP Script or pen movements, respectively, must be serialized!

General Architecture

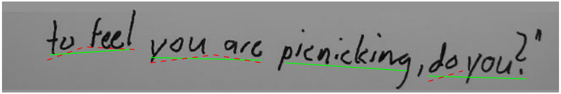


Overview

Preprocessing I

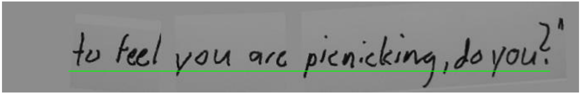
Assumption: Documents are already segmented into text lines
(Text detection and line extraction *highly* application specific!)

Baseline Estimation:



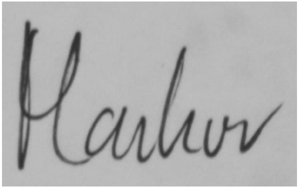
Potential method:
- Initial estimate based on horiz. projection histogram
- Iterative refinement and outlier removal (cf. [4, 25])

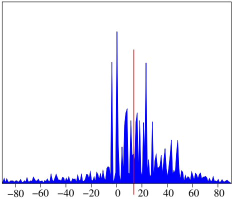
Skew and Displacement Correction:



Preprocessing II

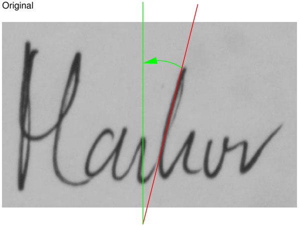
Slant estimation: E.g. via mean orientation of edges obtained by Canny operator (cf. e.g. [28])



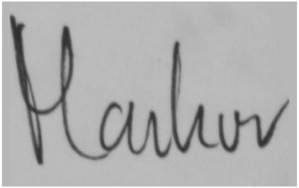


Slant normalization (by applying a shear transform)

Original



Corrected Slant



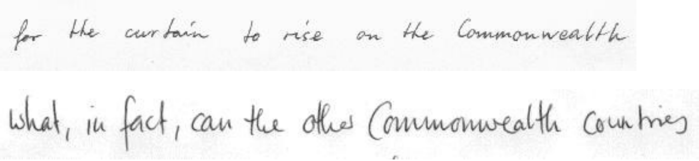
Preprocessing III

Note: Depending on writer and context script might largely vary in size!

Size normalization methods:

- "manually", heuristically, to predefined width/height???
- depending on estimated core size ($\leftarrow$ estimation crucial!)
- depending on estimated character width [15]

Original text lines (from IAM–DB)



Results of size normalization (avg. distance of contour minima)

Serialization: The Sliding Window Method

Problem: Data is two-dimensional, images of writing!

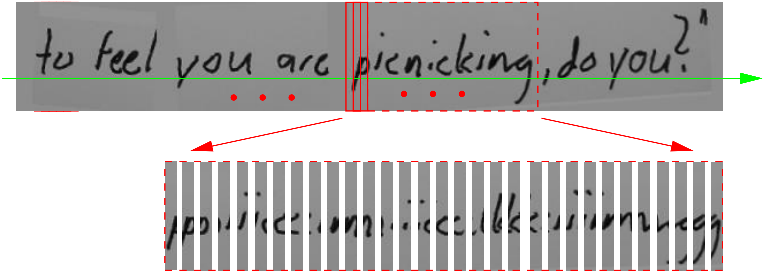
↯ No chronological structure inherently defined!

Exception: Logical sequence of characters within texts

Solution: Sliding-window approach

First proposed by researchers at Daimler-Benz Research Center, Ulm [5], pioneered by researchers at BBN [26]

- Time axis runs in writing direction / along baseline
- Extract small overlapping analysis windows



[Frames shown are for illustration only but actually too large!]

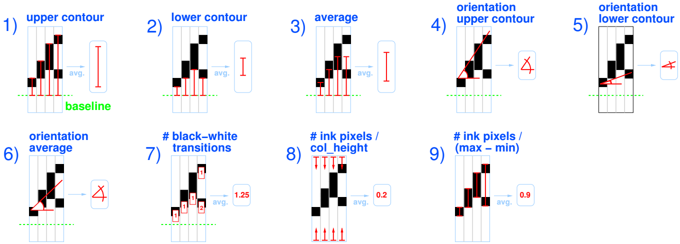
Feature Extraction

Basic Idea: Describe appearance of writing within analysis window

- No "standard" approaches or feature sets
- No holistic features used in HMM-based systems

Potential Methods:

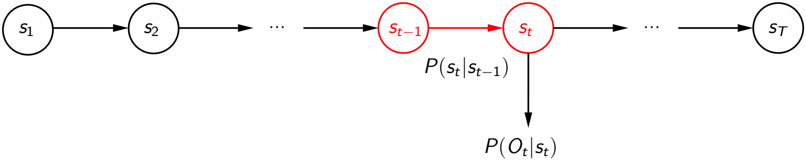
- (For OCR) Local analysis of gray-value distributions (cf. e.g. [3])
- Salient elementary geometric shapes (e.g. vertices, cusps)
- Heuristic geometric properties (cf. e.g. [29])



Additionally: Compute dynamic features
(i.e. discrete approximations of temporal derivatives, cf. e.g. [7])

Overview

Hidden Markov Models: Two-Stage Stochastic Processes



1. Stage: discrete stochastic process ≈ "probabilistic" finite state automaton

   stationary: Process independent of absolute time t

   causal: Distribution s_t only dependent on previous states

   simple: *particularly* dependent only on *immediate* predecessor state ($\hat{=}$ first order)

   $$\Rightarrow P(s_t|s_1, s_2, \ldots s_{t-1}) = P(s_t|s_{t-1})$$

2. Stage: Output O_t generated for every time t depending on current state s_t

   $$\Rightarrow P(O_t|O_1 \ldots O_{t-1}, s_1 \ldots s_t) = P(O_t|s_t)$$

   Note: Only outputs can be observed ⇒ hidden Markov model

Hidden-Markov-Models: Formal Definition

A Hidden-Markov-Model λ of *first order* is defined as:

- a finite set of states:

$$\{s | 1 \leq s \leq N\}$$

- a matrix of state transition probabilities:

$$\mathbf{A} = \{a_{ij} | a_{ij} = P(s_t = j | s_{t-1} = i)\}$$

- a vector of start probabilities:

$$\boldsymbol{\pi} = \{\pi_i | \pi_i = P(s_1 = i)\}$$

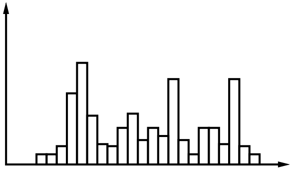
- state specific output probability distributions:

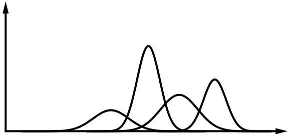
$$\mathbf{B} = \{b_{jk} | b_{jk} = P(O_t = o_k | s_t = j)\} \text{ (discrete case)}$$

or

$$\{b_j(O_t) | b_j(O_t) = p(O_t | s_t = j)\} \text{ (continuous case)}$$

Modeling of Outputs





Discrete inventory of symbols: ✓ Suited for discrete data only (e.g. DNA)

↯ Problematic for non-discrete data – use of vector quantizer required!

Continuous modeling: ✓ Treatment of real-valued vector data (i.e. vast majority of "real-world" data)

✓ Defines distributions over $\mathbb{R}^n$

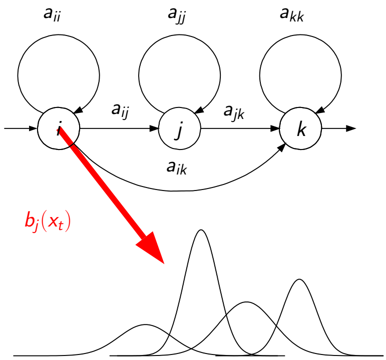
Problem: No general parametric description

Procedure: Approximation using mixture densities

$$b_j(\mathbf{x}) \approx \sum_{k=1}^{M_j} c_{jk} \underbrace{\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{jk}\mathbf{C}_{jk})}_{\text{state specific}} \approx \sum_{k=1}^{M} c_{jk} \underbrace{\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{k}\mathbf{C}_{k})}_{\text{shared}}$$

Popular Simplification: *semi-continuous* models with shared "codebook"

Modeling of Outputs – II

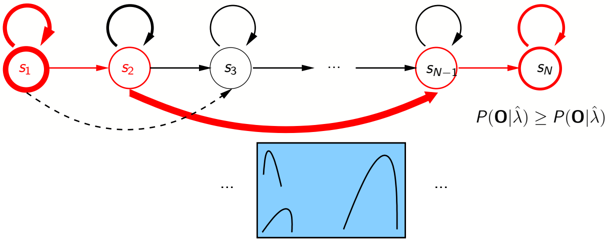


Mixture density modeling:

- Base Distribution?
  ⇒ Gaussian Normal densities
- Shape of Distributions (full / diagonal covariances)?
  ⇒ Depends on pre-processing of the data (e.g. redundancy reduction)
- Number of mixtures?
  ⇒ Clustering (. . . and heuristics)
- Estimation of mixtures?
  ⇒ e.g. Expectation-Maximization

Note: In HMMs integrated with general parameter estimation

Usage Concepts for Hidden-Markov-Models



Assumption: Patterns observed are generated by stochastic models which are comparable *in principle*

Scoring: *How well does the model describe some pattern?*
$\rightarrow$ Computation of the output probability $P(\mathbf{O}|\lambda)$

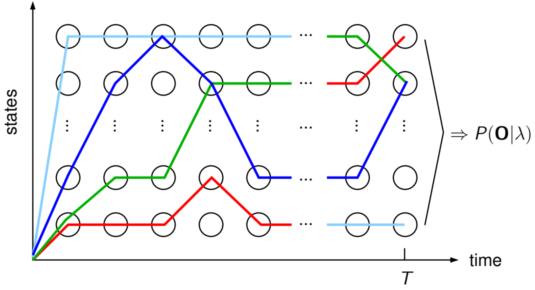
Decoding: *What is the "internal structure" of the model?* ($\hat{=}$ "Recognition")
$\rightarrow$ Computation of the optimal state sequence
$\mathbf{s}^* = \underset{\mathbf{s}}{\operatorname{argmax}} P(\mathbf{O}, \mathbf{s}|\lambda)$

Training: *How to determine the "optimal" model?*
$\rightsquigarrow$ Improvement of a given model λ with $P(\mathbf{O}|\hat{\lambda}) \geq P(\mathbf{O}|\lambda)$

The Output Probability

Wanted: Assessment of HMMs' quality for describing statistical properties of data

Widely used measure: *Output probability* $P(\mathbf{O}|\lambda)$ that observation sequence $\mathbf{O}$ was generated by model λ – along an arbitrary state sequence

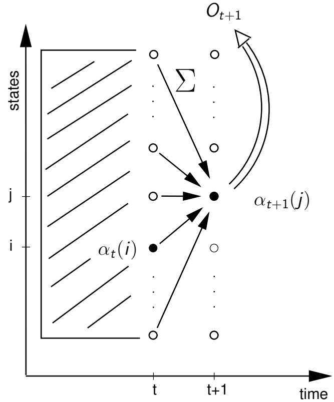


Naive computation infeasible: Exponential complexity $O(TN^T)$

The Output Probability: The Forward-Algorithm

More efficient: Exploitation of the Markov-property, i.e. the "finite memory"
$\Rightarrow$ "Decisions" only dependent on immediate predecessor state

Let:
$\alpha_t(i) = P(O_1, O_2, \ldots O_t, s_t = i|\lambda)$
(*forward variable*)



1. $\alpha_1(i) := \pi_i b_i(O_1)$
2. $\alpha_{t+1}(j) := \left\{ \sum_{i=1}^{N} \alpha_t(i) a_{ij} \right\} b_j(O_{t+1})$
3. $P(\mathbf{O}|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$

✓ Complexity: $O(TN^2)$!
(vs. $O(TN^T)$ from naive computation)

Note: There exists an analogous *Backward-Algorithm* required for parameter estimation.

Decoding

Problem: Global output probability $P(\mathbf{O}|\lambda)$ not sufficient for analysis if individual states are associated to meaningful segments of data

$\Rightarrow$ (Probabilistic) Inference of optimal state sequence $\mathbf{s}^*$ necessary

Maximization of posterior probability:

$$\mathbf{s}^* = \underset{\mathbf{s}}{\operatorname{argmax}}\, P(\mathbf{s}|\mathbf{O}, \lambda)$$

Bayes' rule:

$$P(\mathbf{s}|\mathbf{O}, \lambda) = \frac{P(\mathbf{O}, \mathbf{s}|\lambda)}{P(\mathbf{O}|\lambda)}$$

$P(\mathbf{O}|\lambda)$ irrelevant (constant for fixed $\mathbf{O}$ and given λ), thus:

$$\mathbf{s}^* = \underset{\mathbf{s}}{\operatorname{argmax}}\, P(\mathbf{s}|\mathbf{O}, \lambda) = \underset{\mathbf{s}}{\operatorname{argmax}}\, P(\mathbf{O}, \mathbf{s}|\lambda)$$

Computation of $\mathbf{s}^*$: *Viterbi-Algorithm*

The Viterbi Algorithm

…inductive procedure for efficient computation of $\mathbf{s}^*$ exploiting Markov property

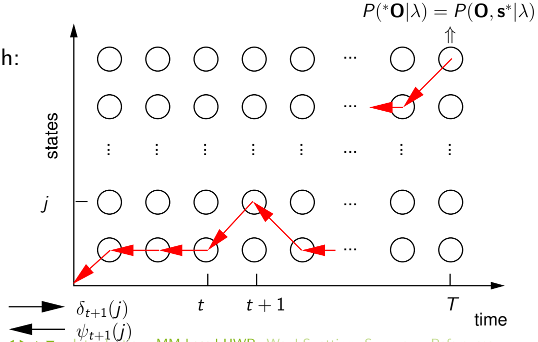
Let: $\delta_t(i) = \max\limits_{s_1, s_2, \ldots s_{t-1}} P(O_1, O_2, \ldots O_t, s_t = i|\lambda)$

1. $\delta_1(i) := \pi_i b_i(O_1)$ $\qquad \psi_1(i) := 0$
2. $\delta_{t+1}(j) := \max\limits_i(\delta_t(i) a_{ij}) b_j(O_{t+1})$ $\qquad \psi_{t+1}(j) := \operatorname*{argmax}\limits_i \ldots$
3. $P^*(\mathbf{O}|\lambda) = P(\mathbf{O}, \mathbf{s}^*|\lambda) = \max\limits_i \delta_T(i)$
   $s_T^* := \operatorname*{argmax}\limits_j \delta_T(j)$
4. Back-tracking of optimal path:
   $s_t^* = \psi_{t+1}(s_{t+1}^*)$

✓ Implicit *segmentation*

✓ Linear complexity in time

STOP Quadratic complexity w.r.t. #states



Parameter Estimation – Fundamentals

Goal: Derive optimal (for some purpose) statistical model from sample data

Problem: No suitable analytical method / algorithm known

"Work-Around": Iteratively improve existing model λ
$\Rightarrow$ Optimized model $\hat{\lambda}$ better suited for given sample data

General procedure: Parameters of λ subject to growth transformation such that

$$P(\mathbf{O}|\hat{\lambda}) \geq P(\mathbf{O}|\lambda)$$

1. "Observe" model's actions during generation of an observation sequence
2. Original parameters are replaced by relative frequencies of respective events

$$\hat{a}_{ij} = \frac{\text{expected number of transitions from } i \text{ to } j}{\text{expected number of transitions out of state } i}$$

$$\hat{b}_i(o_k) = \frac{\text{expected number of outputs of } o_k \text{ in state } i}{\text{total number of outputs in state } i}$$

Limitation: Initial model required!

Parameter Estimation: How to Get Started?

Problem: Parameter training only defined on the basis of *initial* parameters!

Possible Solutions:

- Random / Uniform initialization
  ↯ Only possible for discrete models
- (Fully) Supervised:
  ↯ *Detailed* annotation of training data required
- (Partly) Supervised: Compute annotation automatically with existing model

Pragmatic Solution: Use *semi-continuous* models, i.e, with shared mixture-component densities

$\Rightarrow$ Initialization as combination of:

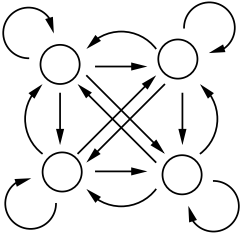
1. *Unsupervised* estimation of initial codebook
2. *Uniform* initialization of remaining parameters (i.e. transition probabilities and mixture weights)

Configuration of HMMs: Topologies

Generally: Transitions between arbitrary states possible within HMMs …
potentially with arbitrarily low probability

Topology of an HMM: Explicit representation of allowed transitions
(drawn as edges between nodes/states)

Any transition possible
$\Rightarrow$ *ergodic* HMM



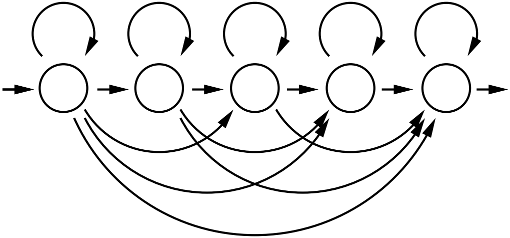
Observation: Fully connected HMM does usually not make sense for describing chronologically organized data

↯ "backward" transitions would allow arbitrary repetitions within the data

Configuration of HMMs: Topologies II

Idea: Restrict potential transition to *relevant* ones!
... by omitting irrelevant edges / setting respective transition probabilities to "hard" zeros (i.e. never modified!)

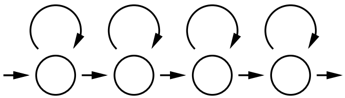
Structures/Requirements for modeling chronologically organized data:



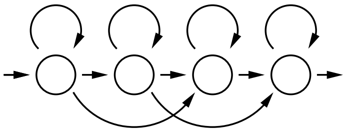
- "Forward" transitions (i.e. progress in time)
- "Loops" for modeling variable durations of segments
- "Skips" allow for optional/missing parts of the data
- Skipping of one or multiple states forward

Configuration of HMMs: Topologies III

Overview: The two most common topologies for handwriting (and speech) recognition:



linear HMM



Bakis-type HMM

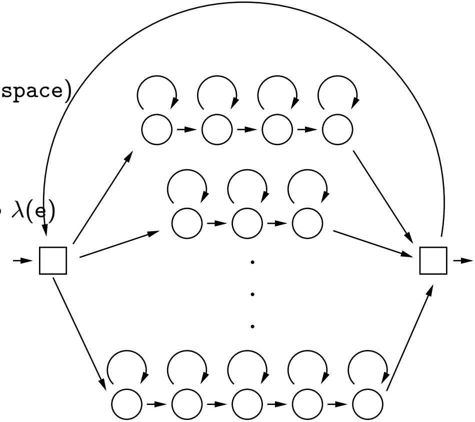
Note: General left-to-right models (allowing to skip any number of states forward) are not used in practice!

“Segmentation-free” approach: Features represent text lines!
⇒ Model for representation of word / character sequences

Configuration of HMMs: Compound Models

Goal: Segmentation

- Basic units: Characters
  $\lambda(\texttt{a}), \ldots \lambda(\texttt{Z}), \lambda(\texttt{0}), \ldots \lambda(\texttt{9}), \lambda(\texttt{@space})$
- Words formed by concatenation
  $\lambda(\texttt{Crete}) =$
  $\lambda(\texttt{C}) \circ \lambda(\texttt{r}) \circ \lambda(\texttt{e}) \circ \lambda(\texttt{t}) \circ \lambda(\texttt{e})$
- Lexicon = parallel connection
  [Non-emitting states merge edges]
- Model for arbitrary text
  by adding loop



$\Rightarrow$ Decoding the model produces segmentation
(i.e. determining the optimal state/model sequence)

Overview

Handwriting Recognition: Evaluation

Principle: Measure transcription performance on independent test set

Performance Measure: Word Error Rate (WER)

Errors determined w.r.t. reference transcription

I.e. substitutions/insertions/deletions

$\Rightarrow$ Levenshtein distance [computed via DTW]

$$\text{WER} = \frac{\#\text{Sub} + \#\text{Del} + \#\text{Ins}}{\#Ref.Words}$$

Alternative performance measure: Word Accracy (WA = 100% - WER)

HWR Evaluation: Example

Ref.	handwriting	recognition	is	difficult	WER
Hypotheses	handwriting	recognition	is	different	25%
				Sub = 1	
	hand / writing	recognitions		difficult	100%
	Sub = 1 / Ins = 1	Sub = 1 / Del = 1			
	handwriting	recognition	is	difficult	0%

HWR Evaluation: Performance Examples

Performance Examples

- Roman script / sencences

  E.g. IAM-DB ($\approx$ 20% WER), RIMES ($\approx$ 15% WER)

- Arabic / isolated words

  E.g. IFN/ENIT ($\approx$ 10% WER [f], 20% WER [s])

  $\Rightarrow$ Ahmad *et al.*, ICFHR 2014: 8% WER [f], **15% WER [s!]**

Caveats

- WER figures alone are useless!
  $\Rightarrow$ Only comparable for identical tasks!

- Preprocessing and feature extraction play a crucial role!

Overview

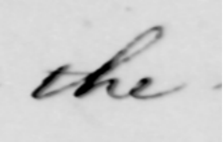
- Introduction
- Markov Model-Based Handwriting Recognition ... *Fundamentals*
- Word Spotting for Handwriting Retrieval ... *Models & Methods*
  - Motivation ... *Why Wordspotting?*
  - Evolution of Methods ... *What has been done before*
  - Word Spotting: Bag-of-Features Models
  - Evaluation ...... *of HW Retrieval*
- Summary ... *and Further Reading*

Handwriting Retrieval: Motivation

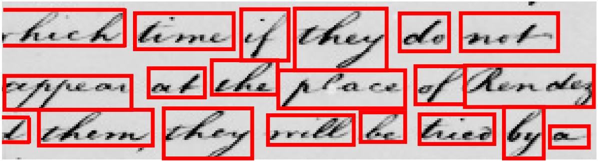
What if automatic transcription of handwriting is no longer feasible?

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

Query word image



Document image

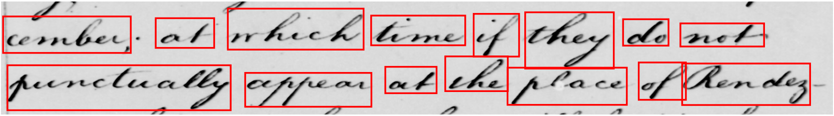


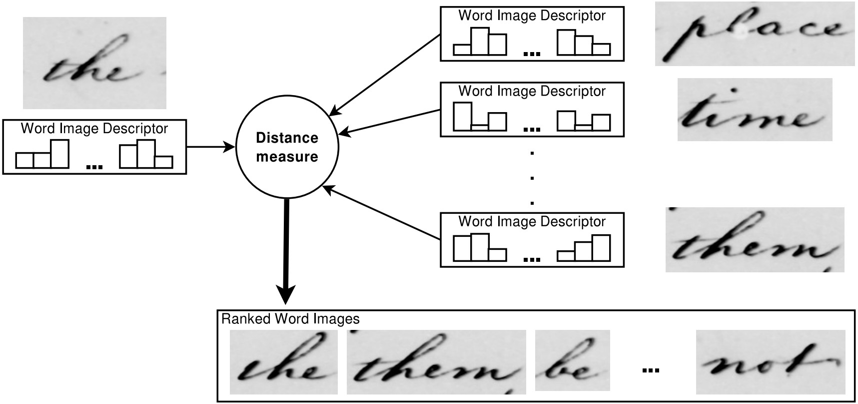
Query-by-example word spotting

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

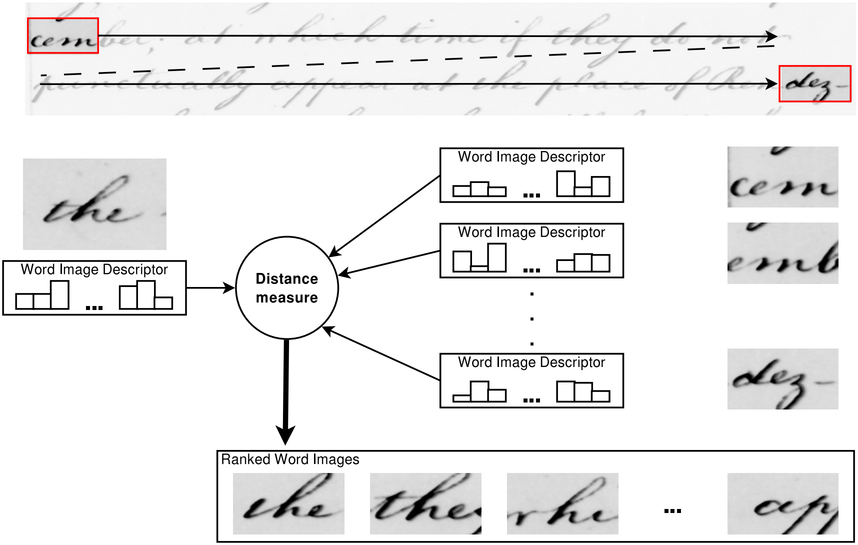
Overview

Word Spotting: Segmentation-based Methods





Word Spotting: Segmentation-free Methods

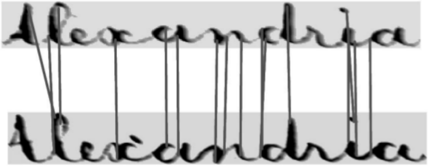


Indexing Handwriting Using Word Matching

- 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.

Corner Feature Correspondences for Word Image Ranking

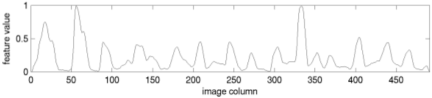


- Word image size normalization (according to query)
- Harris Corner keypoint detection
- Intensity value based keypoint description
- SSD based correspondences
- Euclidean distances of correspondeces

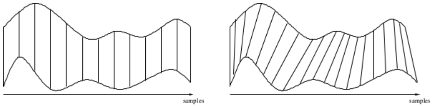
Rothfeder, J. L., Feng, S., and Rath, T. M. Using corner feature correspondences to rank word images by similarity. In Proc. of the Workshop on Document Image Analysis and Retrieval (electronically published) (Madison, WI, June 20 2003).

Sequence-based Word Image Matching

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



(b) normalized projection profile.

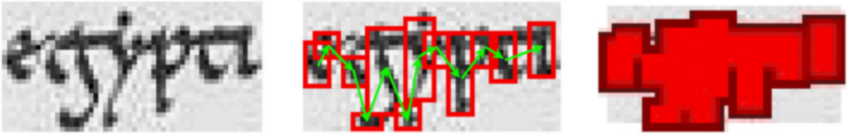


(a) naive alignment after resampling, (b) alignment with DTW.

- 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.

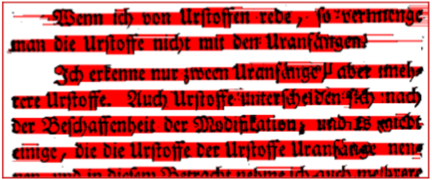
Segmentation-free Word Spotting with Zones-of-Interest

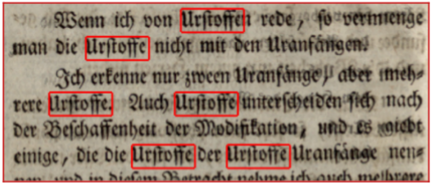


- Zones-of-Interest
- Gradient-based features
- Cohesive elastic ZOI matching

Y. Leydier, F. Lebourgeois and H. Emptoz, "Text search for medieval manuscript images", Pattern Recognition, vol. 40, 2007, pp. 3552-3567.

Segmentation-free Patch-based Word Spotting





- Dense patches in salient document regions
- To some extent invariant to rotation and scale
- Pixel density based feature representation
- Euclidean distance based feature matching

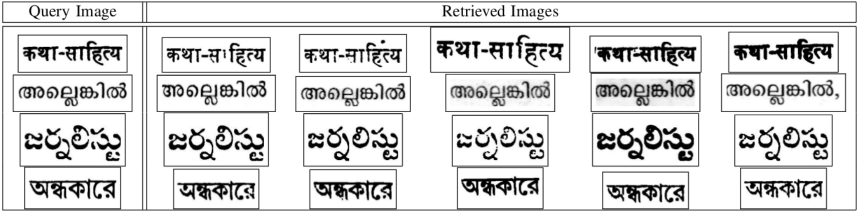
B. Gatos and I. Pratikakis, "Segmentation-free word spotting in historical printed documents," in Proc. of the Int. Conf. on Document Analysis and Recognition, 2009, pp. 271-275.

Segmentation-free Word Spotting with Bag-of-Features

- Dense grid of SIFT descriptors [14]
- Bag-of-Features-based Spatial-Pyramid feature representation
  (Spatial Pyramids originally proposed for scene recognition [11].)
- 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.

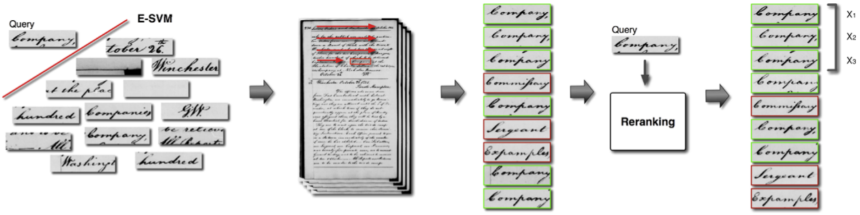
Segmentation-based Word Spotting with Bag-of-Features



- Harris Corners keypoint detection
- SIFT-based Bag-of-Features representations
- Word image indexing with inverted file structure
- Word image reranking by spatial keypoint consistency criterion

Shekhar, R., and Jawahar, C., Word image retrieval using bag of visual words. International Workshop on Document Analysis Systems, pages 297-301, 2012.

Segmentation-free Word Spotting with Exemplar SVM

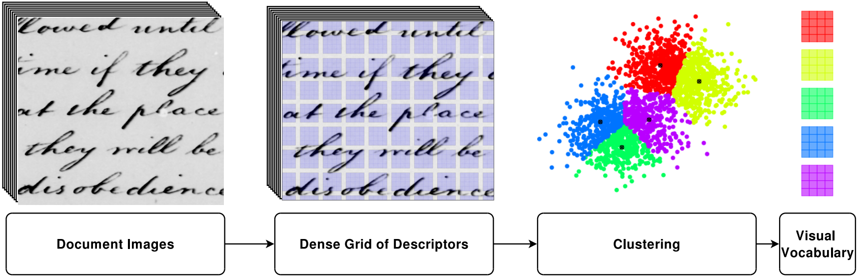


- Features based on Histograms of oriented Gradients (HoG) [6]
- Patched-based framework
- Matching with Examplar SVM
- Reranking and Query expansion

Jon Almazán, Albert Gordo, Alicia Fornés, Ernest Valveny, Segmentation-free word spotting with exemplar SVMs, Pattern Recognition, Available online 17 June 2014,

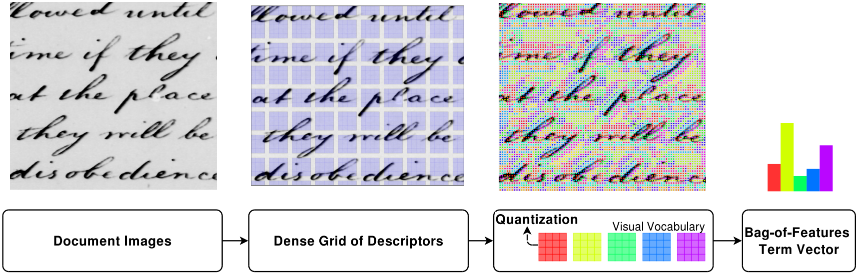
Overview

Bag-of-Features Models: Visual Vocabulary



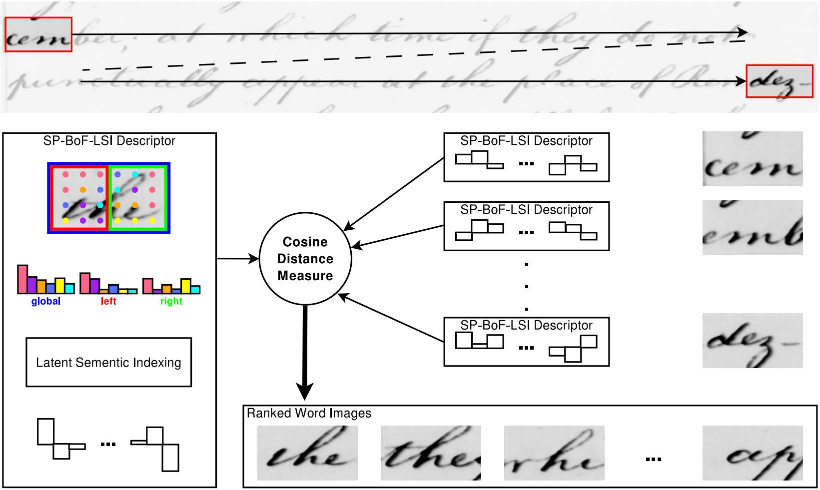
(For an introduction to BoF models see [17].)

Bag-of-Features Models: Term Vector



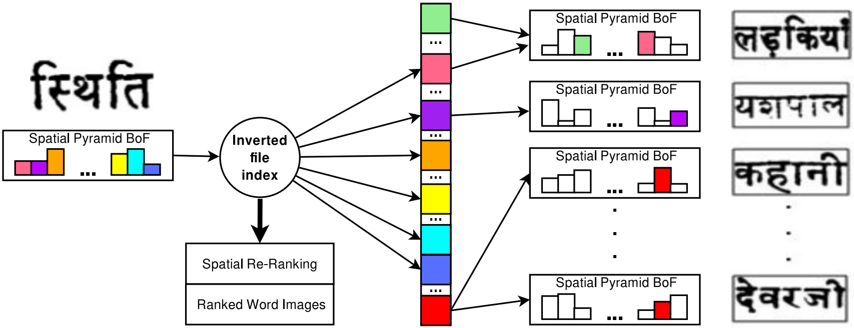
(For an introduction to BoF models see [17].)

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.

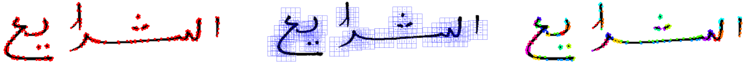
Segmentation-based Word Spotting with Bag-of-Features

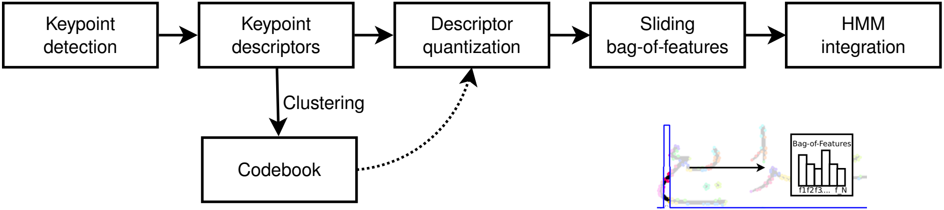


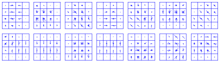
Shekhar, R., and Jawahar, C., Word image retrieval using bag of visual words. International Workshop on Document Analysis Systems, pages 297-301, 2012.

Bag-of-Features HMMs

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



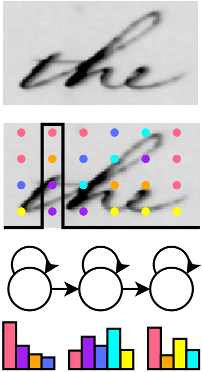




$$b_j(\mathbf{f}) = \sum_{k=1}^{|V|} c_{jk} f_k \quad \text{with } \mathbf{f} : \text{term vector}, V : \text{vis. voc.}$$

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

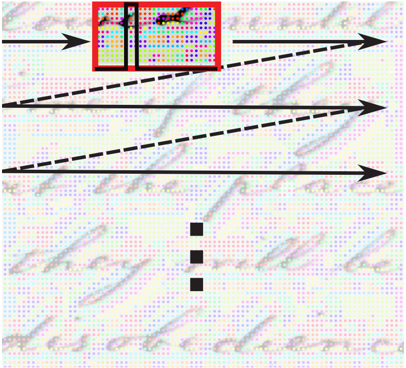
Bag-of-Features HMMs for Word Spotting



Query Word

Bag-of-Features Sequence
(sliding window)

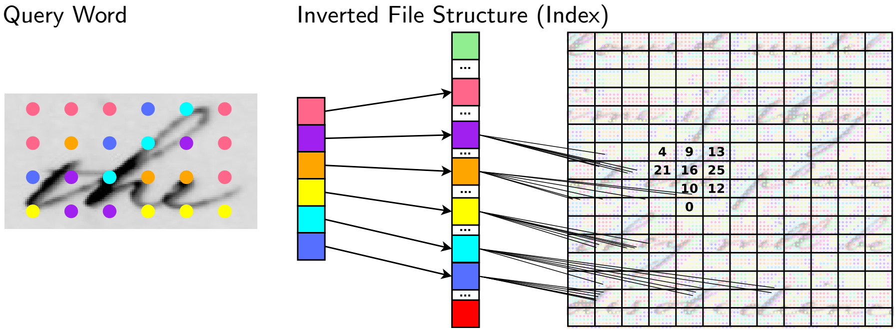
Bag-of-Features
Hidden Markov Model



- Brute-force approach is accurate but slow!

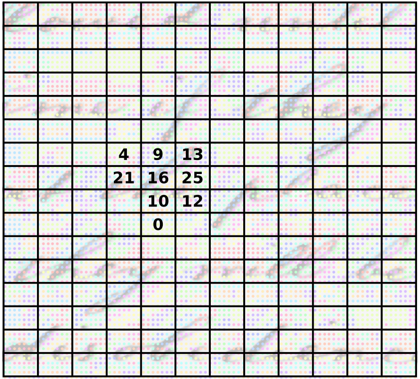
Rothacker, L., Rusinol, M., Fink, G. A.: *Bag-of-Features HMMs for Segmentation-Free Word Spotting in Handwritten Documents*, In Proc. Int. Conf. on Document Analysis and Recognition, Washington DC, USA, 2013.

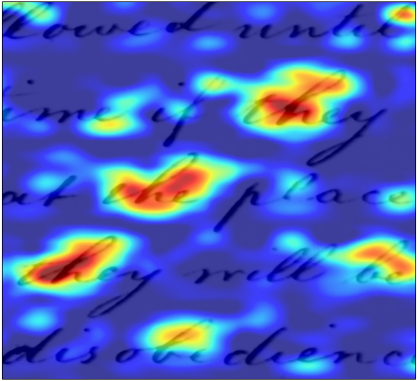
Bag-of-Features HMMs: Speed-Up with Inverted File Structure



Rothacker, L., Rusinol, M., Llados, J., Fink, G. A.: *A Two-Stage Approach to Segmentation-Free Query-by-Example Word Spotting*, Manuscript Cultures, submitted 2014.

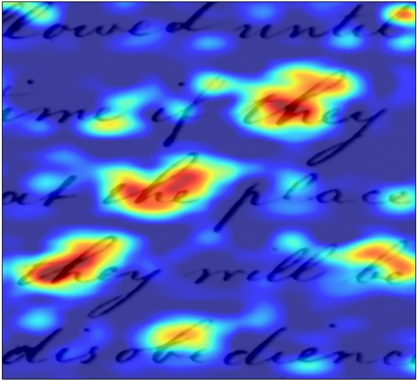
Bag-of-Features HMMs for Word Spotting





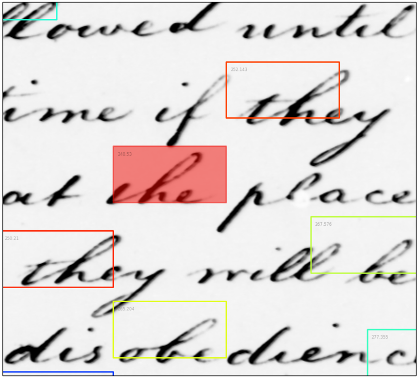
Rothacker, L., Rusinol, M., Llados, J., Fink, G. A.: *A Two-Stage Approach to Segmentation-Free Query-by-Example Word Spotting*, Manuscript Cultures, submitted 2014.

Bag-of-Features HMMs for Word Spotting



Rothacker, L., Rusinol, M., Llados, J., Fink, G. A.: *A Two-Stage Approach to Segmentation-Free Query-by-Example Word Spotting*, Manuscript Cultures, submitted 2014.

Bag-of-Features HMMs for Word Spotting



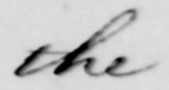
Rothacker, L., Rusinol, M., Llados, J., Fink, G. A.: *A Two-Stage Approach to Segmentation-Free Query-by-Example Word Spotting*, Manuscript Cultures, submitted 2014.

Overview

- Introduction
- Markov Model-Based Handwriting Recognition ... *Fundamentals*
- **Word Spotting for Handwriting Retrieval** ... *Models & Methods*
  - Motivation ... *Why Wordspotting?*
  - Evolution of Methods ... *What has been done before*
  - Word Spotting: Bag-of-Features Models
  - **Evaluation** ...... *of HW Retrieval*
- Summary ... *and Further Reading*

Segmentation-free Word Spotting: Evaluation

Query word image



Retrieved patches sorted by score

Patch relevance threshold: For example 50% ground truth overlap

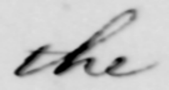
- Precision: How relevant is the list?
- Recall: How complete is the list w.r.t. relevant items?
- Average Precision: How well is the retrieval list sorted?

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

Average Precision: Single value summary averaging precision values at different recall levels. Recall levels are obtained at each new relevant item in retrieval list.

Segmentation-free Word Spotting: Evaluation

Query word image



Retrieved patches sorted by score

- Transform result into relevant / non-relevant list (overlap threshold):

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

- Precision: How relevant is the list?

$$\frac{\#\text{Relevant Retrieved Items}}{\#\text{Retrieved Items}} = \frac{2}{6}$$

- Recall: How complete is the list w.r.t. relevant items?

$$\frac{\#\text{Relevant Retrieved Items}}{\#\text{Relevant Items in Dataset (let's say 10)}} = \frac{2}{10}$$

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 at different recall levels (cf. e.g. [13]):

$$\frac{\sum_{k=1}^{n} \text{Precision}_k \times \text{rel}(k)}{\#\text{Relevant Items in Dataset}}$$

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

  Accumulate Precision whenever the Recall changes and normalize.

Segmentation-free Word Spotting: Evaluation III

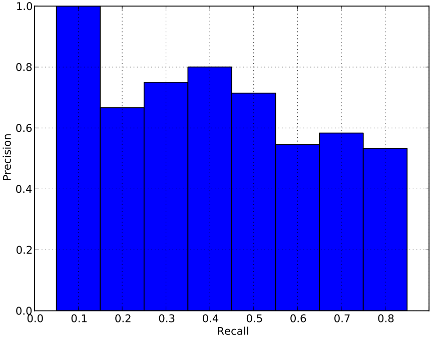
$$[1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1]$$

Average Precision:

$$\frac{\sum_{k=1}^{n} \text{Precision}_k \times \text{rel}(k)}{\#\text{Relevant Items in Dataset}}$$

$$\frac{\frac{1}{1} + \frac{2}{3} + \frac{3}{4} + \frac{4}{5} + \frac{5}{7} + \frac{6}{11} + \frac{7}{12} + \frac{8}{15}}{10}$$

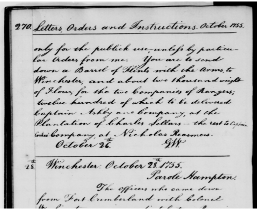
$$\approx 0.56$$



Segmentation-free Word Spotting: Evaluation IV

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. [24]	BoF+LSI	fixed	30.4%	71.1%
Almazán et al. [2]	HoG	query dep.	59.5%	-
Almazán et al. [2]	HoG+QE+RR	query dep.	68.9%	-
Rothacker et al. [21]	IFS+BoF-HMM	query dep.	**69.9**%	**96.5**%

For a more challenging task see Fink *et al.* ICFHR'14

Overview

- Introduction
- Markov Model-Based Handwriting Recognition ... *Fundamentals*
- Word Spotting for Handwriting Retrieval ... *Models & Methods*
- Summary ... *and Further Reading*

Summary: Handwriting Recognition

- ✓ Statistical approach yields powerful appearance models for handwritten script ($\hat{=}$ Hidden Markov Models)
- ✓ Achieves integrated segmentation and classification: Segmentation-free recognition
- ✓ Combination of appearance model and language model possible
- ↯ Model structure needs to be pre-defined
- ↯ Considerable amounts of training data necessary (as for *all* stochastic models)

"There is no data like more data!"

[Robert L. Mercer, IBM]

Summary: Handwriting Retrieval

- ✓ Patch-based decoding + stochastic query model (BoF-based) = Segmentation-free query-by-example word spotting
- ✓ Impressive results for isogeneous document collections
- ✓ No annotation effort required!
- ⚡ Poor generalization (of QbE word spotting) to different writing styles
- ⚡ 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")!
- ⚡ Consequently requires *large amounts* of annotated training data!

Further Reading

Survey Articles:

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

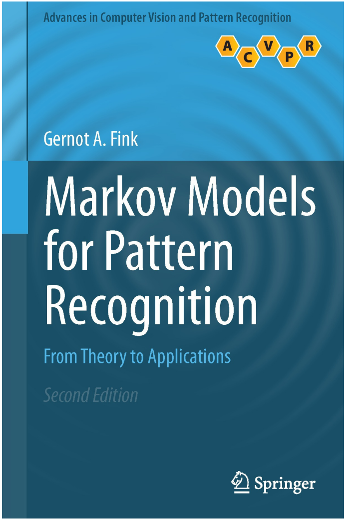
Thomas Plötz & Gernot A. Fink:
Markov Models for Offline Handwriting Recognition: A Survey.
IJDAR, 12(4):269–298, 2009.

- ✓ Open access publication!

Textbook:

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

- ✓ Inspection copy available!
- ✓ Conference discount: 20%!



References I

[1] Irfan Ahmad, Gernot A. Fink, and Sabri Mahmoud.
Improvements in sub-character HMM model based Arabic text recognition.
In *Proc. Int. Conf. on Frontiers in Handwriting Recognition*, Crete, Greece, 2014.

[2] Jon Almázan, Albert Gordo, Alicia Fornés, and Ernest Valveny.
Segmentation-free word spotting with exemplar SVMs.
Pattern Recognition, (0):–, 2014.

[3] Issam Bazzi, Richard Schwartz, and John Makhoul.
An omnifont open-vocabulary OCR system for English and Arabic.
IEEE Trans. on Pattern Analysis and Machine Intelligence, 21(6):495–504, 1999.

[4] Radmilo M. Bozinovic and Sargur N. Srihari.
Off-line cursive script word recognition.
IEEE Trans. on Pattern Analysis and Machine Intelligence, 11(1):69–83, 1989.

References II

[5] T. Caesar, J. M. Gloger, and E. Mandler.
Preprocessing and feature extraction for a handwriting recognition system.
In *Proc. Int. Conf. on Document Analysis and Recognition*, pages 408–411, Tsukuba Science City, Japan, 1993.

[6] Navneet Dalal and Bill Triggs.
Histograms of oriented gradients for human detection.
In Cordelia Schmid, Stefano Soatto, and Carlo Tomasi, editors, *Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 886–893, 2005.

[7] J. G. A. Dolfing and R. Haeb-Umbach.
Signal representations for Hidden Markov Model based on-line handwriting recognition.
In *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, volume IV, pages 3385–3388, München, 1997.

References III

[8] Gernot A. Fink.
Markov Models for Pattern Recognition, From Theory to Applications.
Advances in Computer Vision and Pattern Recognition. Springer, London, 2 edition, 2014.

[9] Gernot A. Fink, Leonard Rothacker, and Rene Grzeszick.
Grouping historical postcards using query-by-example word spotting.
In *Proc. Int. Conf. on Frontiers in Handwriting Recognition*, Crete, Greece, 2014.

[10] B. Gatos and I Pratikakis.
Segmentation-free word spotting in historical printed documents.
In *Document Analysis and Recognition, 2009. ICDAR '09. 10th International Conference on*, pages 271–275, July 2009.

[11] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce.
Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories.
In *Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 2169–2178, 2006.

References IV

[12] Yann Leydier, Frank Lebourgeois, and Hubert Emptoz.
Text search for medieval manuscript images.
Pattern Recognition, 40(12):3552 – 3567, 2007.

[13] Josep Lladós, Marçal Rusiñol, Alicia Fornés, David Fernández, and Anjan Dutta.
On the influence of word representations for handwritten word spotting in historical documents.
Int. J. Pattern Recognition and Artificial Intelligence, 26(5), 2012.

[14] David Lowe.
Distinctive image features from scale-invariant keypoints.
Int. J. of Computer Vision, 60(2):91–110, 2004.

[15] S. Madhvanath, G. Kim, and V. Govindaraju.
Chaincode contour processing for handwritten word recognition.
IEEE Trans. on Pattern Analysis and Machine Intelligence, 21(9):928–932, September 1999.

References V

[16] R. Manmatha, Chengfeng Han, E. M. Riseman, and W. B. Croft.
Indexing handwriting using word matching.
In *Proc. of the First ACM Int. Conf. on Digital Libraries*, DL '96, pages 151–159, New York, NY, USA, 1996. ACM.

[17] Stephen O'Hara and Bruce A. Draper.
Introduction to the bag of features paradigm for image classification and retrieval.
Computing Research Repository, arXiv:1101.3354v1, 2011.

[18] Thomas Plötz and Gernot A. Fink.
Markov Models for Offline Handwriting Recognition: A Survey.
Int. Journal on Document Analysis and Recognition, 12(4):269–298, 2009.

[19] Toni M. Rath and R. Manmatha.
Word image matching using dynamic time warping.
In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–521–II–527 vol.2, June 2003.

References VI

[20] Leonard Rothacker, Marcal Rusinol, and Gernot A. Fink.
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.

[21] Leonard Rothacker, Marçal Rusiñol, Josep Lladós, and Gernot A. Fink.
A two-stage approach to segmentation-free query-by-example word spotting.
Manuscript Cultures, 2014.
(submitted).

[22] Leonard Rothacker, Szilard Vajda, and Gernot A. Fink.
Bag-of-features representations for offline handwriting recognition applied to Arabic script.
In *Proc. Int. Conf. on Frontiers in Handwriting Recognition*, Bari, Italy, 2012.

References VII

[23] Jamie L. Rothfeder, S. Feng, and Toni M. Rath.
Using corner feature correspondences to rank word images by similarity.
In *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW '03. Conference on*, volume 3, pages 30–30, June 2003.

[24] M. Rusiñol, D. Aldavert, R. Toledo, and J. Llados.
Browsing heterogeneous document collections by a segmentation-free word spotting method.
In *Proc. Int. Conf. on Document Analysis and Recognition*, pages 63 –67, Bejing, China, 2011.

[25] M. Schenkel, I. Guyon, and D. Henderson.
On-line cursive script recognition using time delay neural networks and hidden Markov models.
In *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, volume 2, pages 637–640, Adelaide, Australia, April 1994.

References VIII

[26] Richard Schwartz, Christopher LaPre, John Makhoul, Christopher Raphael, and Ying Zhao.
Language-independent OCR using a continuous speech recognition system.
In *Proc. Int. Conf. on Pattern Recognition*, volume 3, pages 99–103, Vienna, Austria, 1996.

[27] R. Shekhar and C.V. Jawahar.
Word image retrieval using bag of visual words.
In *Document Analysis Systems (DAS), 2012 10th IAPR International Workshop on*, pages 297–301, March 2012.

[28] M. Wienecke, G. A. Fink, and G. Sagerer.
Experiments in unconstrained offline handwritten text recognition.
In *Proc. 8th Int. Workshop on Frontiers in Handwriting Recognition*, Ontario, Canada, August 2002. IEEE.

[29] M. Wienecke, G. A. Fink, and G. Sagerer.
Toward automatic video-based whiteboard reading.
Int. Journal on Document Analysis and Recognition, 7(2–3):188–200, 2005.