Improved Modeling for Arabic Handwriting Recognition

Philippe Dreuw, Stephan Jonas, Georg Heigold,
David Rybach, Christian Gollan, Hermann Ney
dreuw@cs.rwth-aachen.de

Quaero CTC Workshop – 12. March 2009

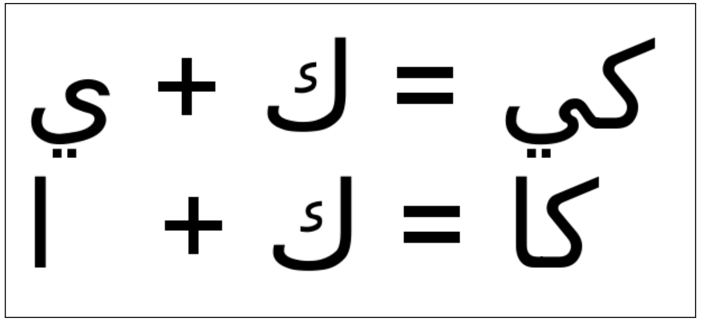
Human Language Technology and Pattern Recognition
Lehrstuhl für Informatik 6
Computer Science Department
RWTH Aachen University, Germany

Outline

1. Introduction
2. Adaptation of the RWTH-ASR framework for Handwriting Recognition
   - Model Length Estimation
   - Unsupervised Confidence-Based Discriminative Training
3. Experimental Results
4. Summary

Introduction

- **Arabic handwriting system**
  - **right-to-left, 28 characters, position-dependent character writing variants**
  - **Pieces of Arabic Word (PAWs) as subwords**
  - **ligatures and diacritics**

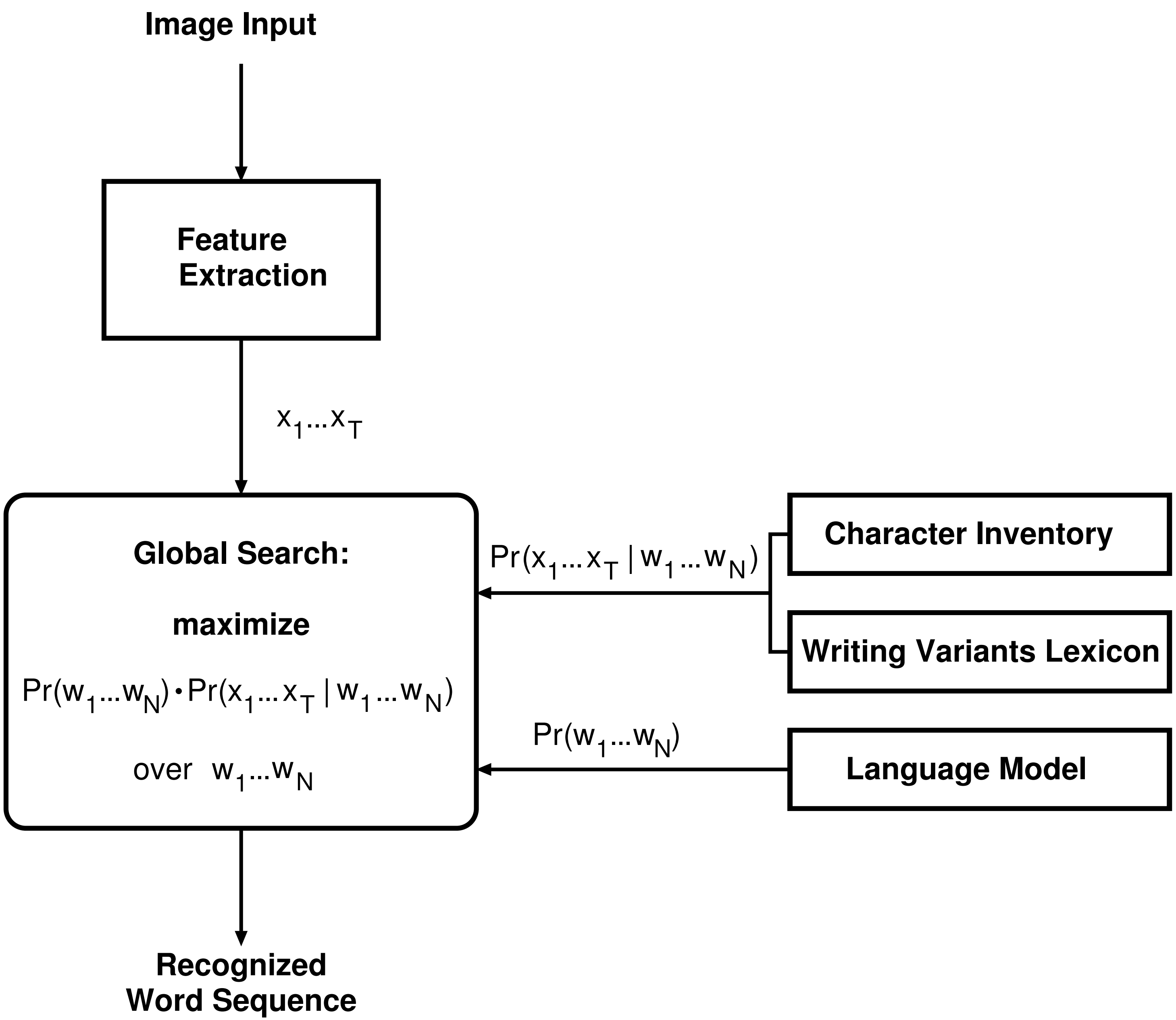


(a) Ligatures

(b) Diacritics

- **state-of-the-art**
  - **preprocessing (normalization, baseline estimation, etc.) + HMMs**
- **our approach:**
  - **adaptation of RWTH-ASR framework for handwriting recognition**
  - **preprocessing-free feature extraction, focus on modeling**

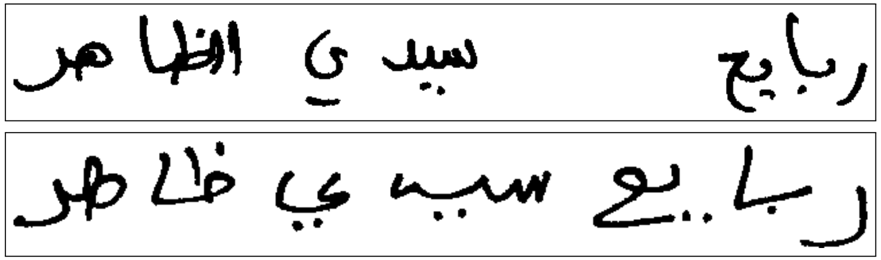
System Overview



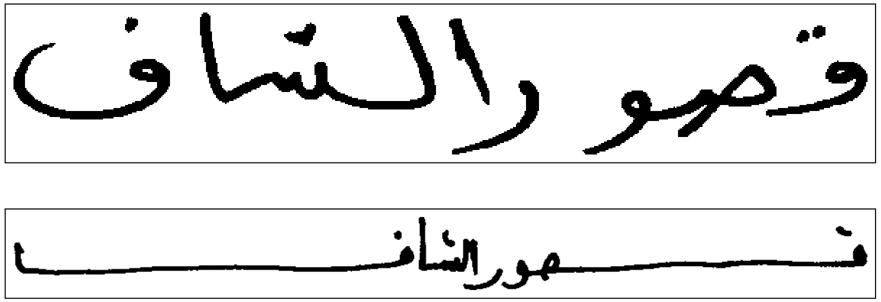
Visual Modeling: Feature Extraction and HMM Transitions

- **recognition of characters within a context, temporal alignment necessary**
- **features: sliding window, no preprocessing, PCA reduction**

- **important: HMM whitespace models (a) and state-transition penalties (b)**



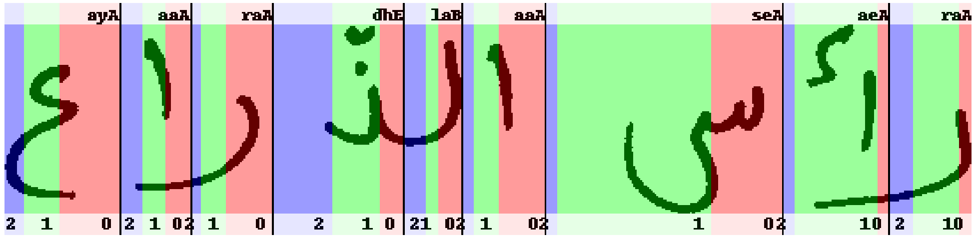
(a)



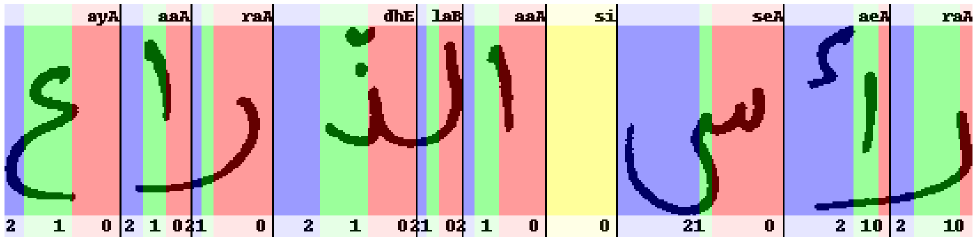
(b)

Visual Modeling: Writing Variants Lexicon

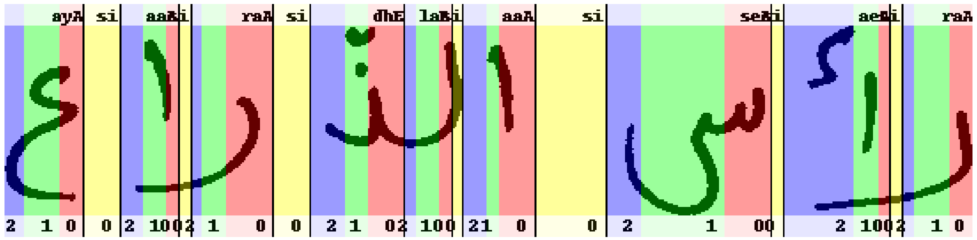
- **White-Space Models for Pieces of Arabic Words** [Dreuw & Jonas^{+} 08] in ICPR 2008
  - most reported error rates are dependent on the number of PAWs
  - without separate white space model (NS)



  - always white spaces between compound words (bws)

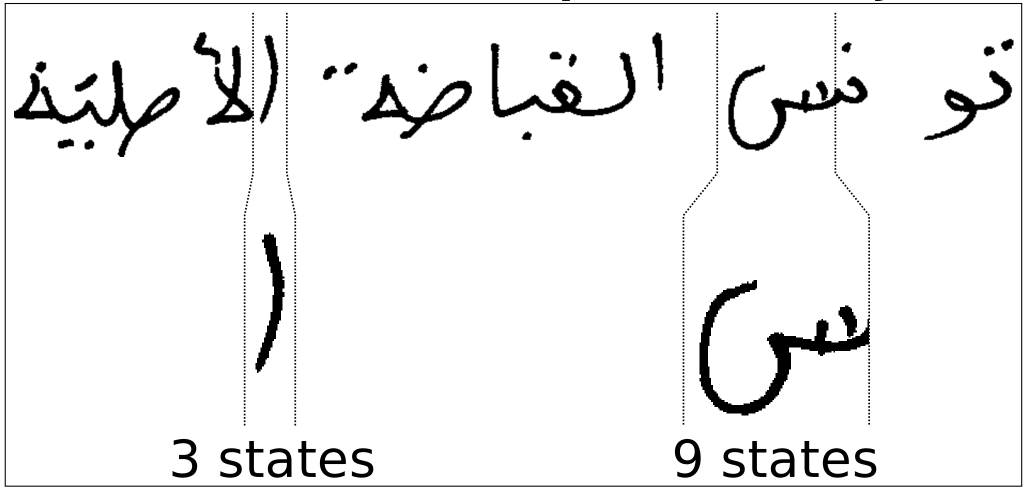


  - white spaces as writing variants between and within words (bwws)



Visual Modeling: Model Length Estimation

- **more complex characters should be represented by more HMM states**



- **the number of states S_c for each character c is updated by**

$$S_c = \frac{N_{x,c}}{N_c} \cdot f_P$$

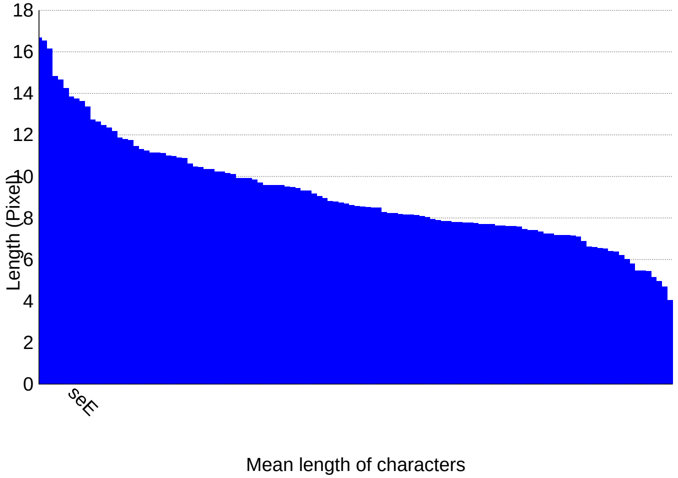
with

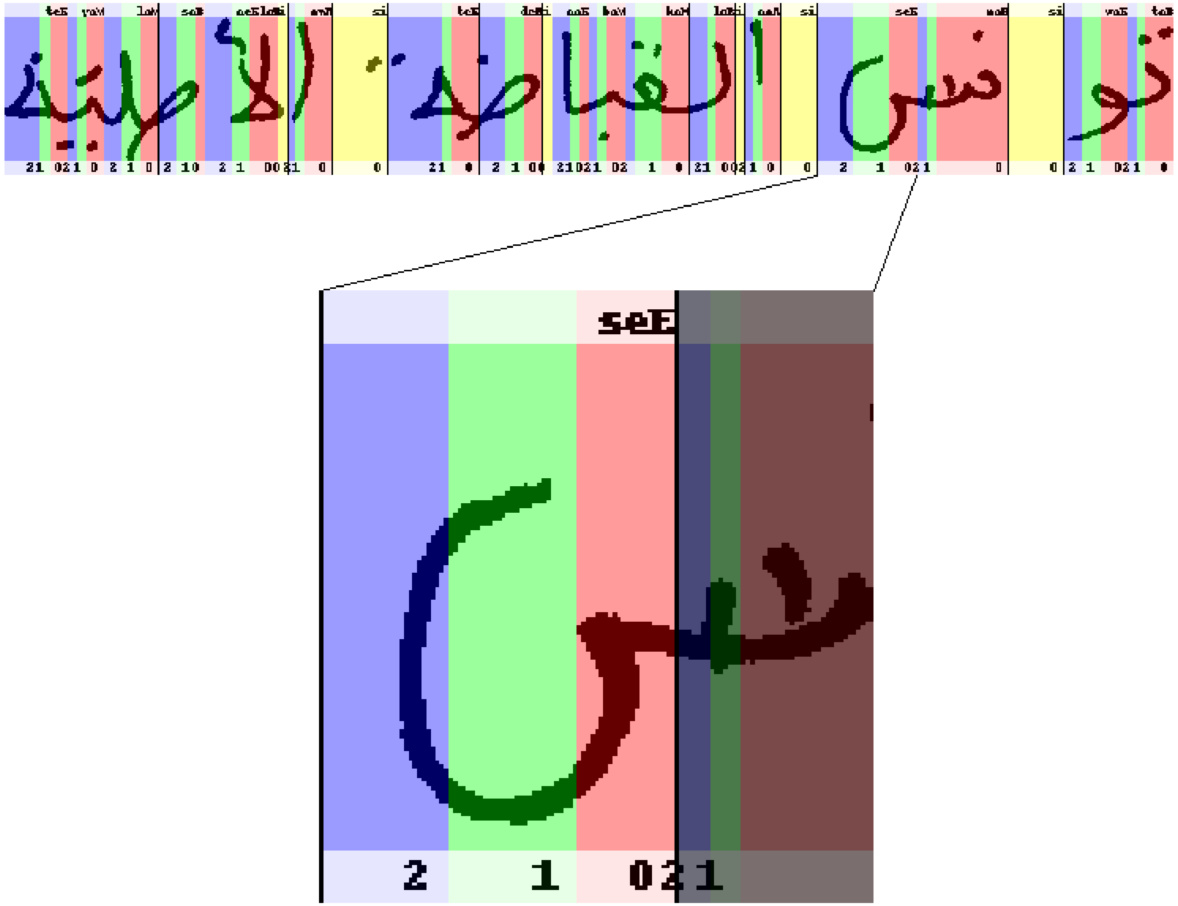
S_c	=	**estimated number states for character c**
$N_{x,c}$	=	**number of observations aligned to character c**
N_c	=	**character count of c seen in training**
f_P	=	**character length scaling factor.**

Visual Modeling: Model Length Estimation

Original Length

- overall mean of character length = 7.9 pixel (≈ 2.7 pixel/state)
- total #states = 357

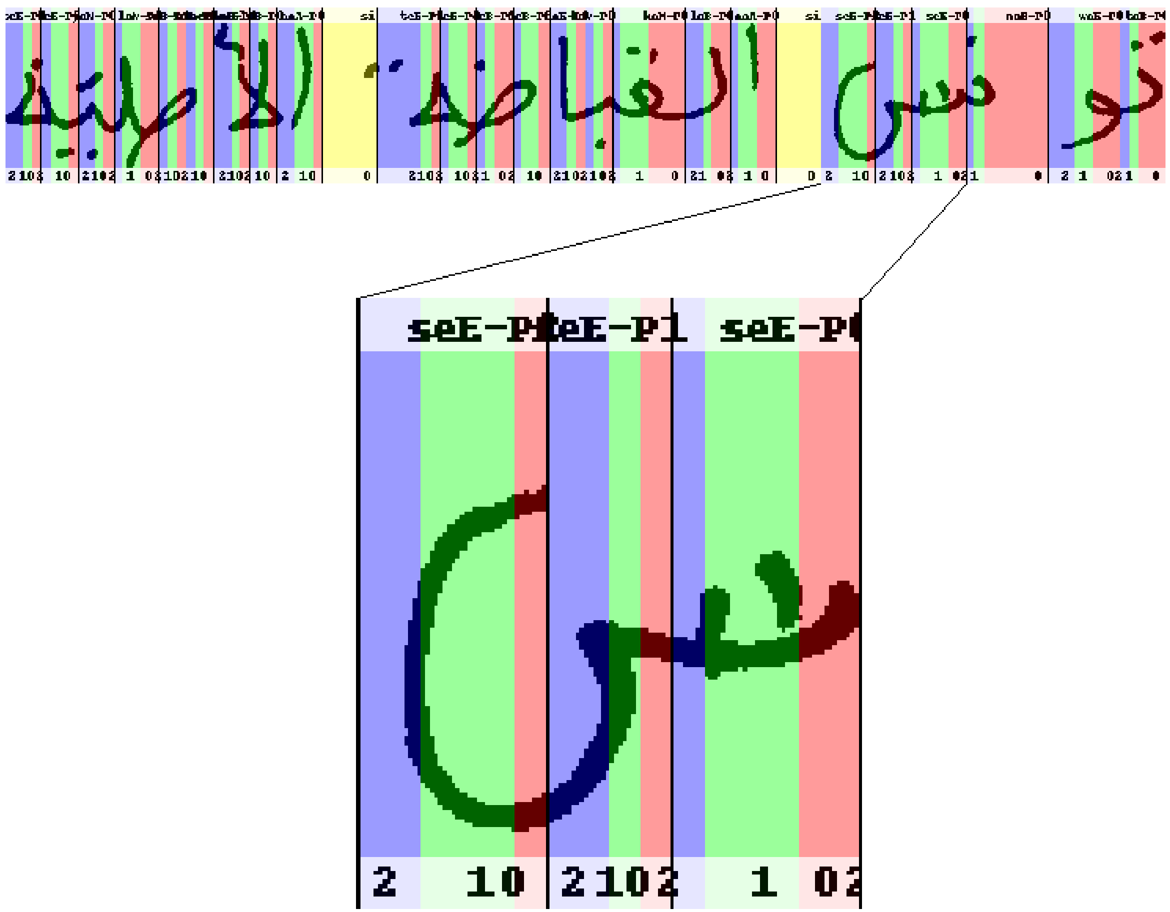


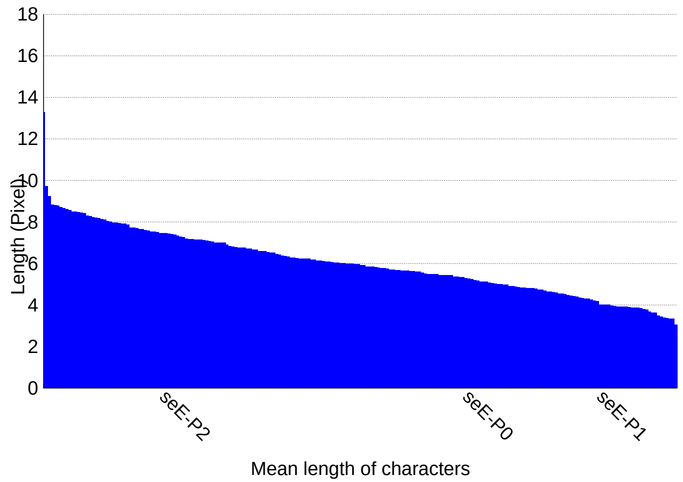


Visual Modeling: Model Length Estimation

Estimated Length

- overall mean of character length = 6.2 pixel ($\approx$ 2.0 pixel/state)
- total #states = 558





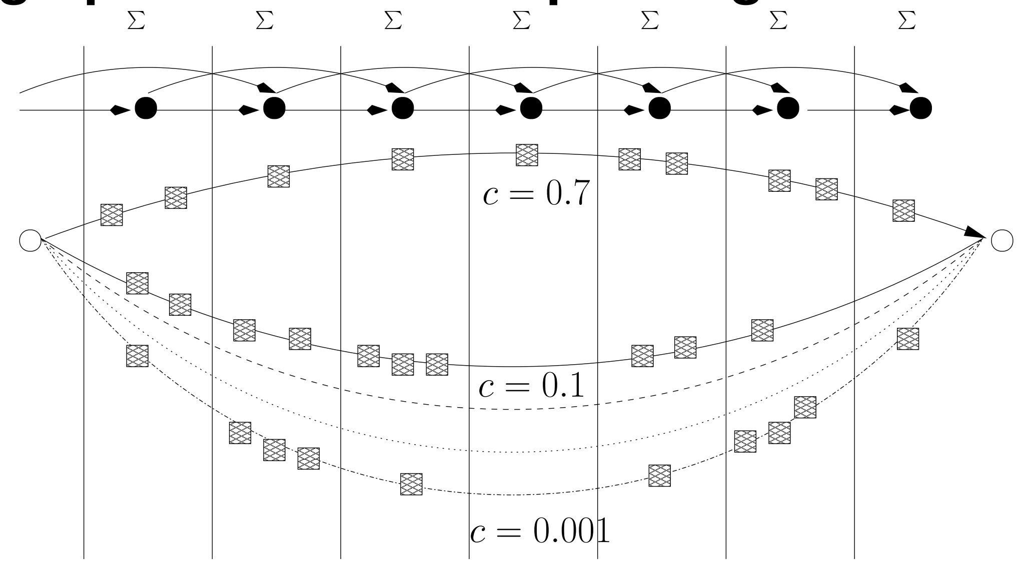
Training and Decoding Architectures

- Training
  - Maximum Likelihood (ML)
  - CMLLR-based Writer Adaptive Training (WAT)
  - discriminative training using modified-MMI criterion (M-MMI)

- Decoding
  - 1-pass
    - ML model
    - M-MMI model
  - 2-pass
    - segment clustering for CMLLR writer adaptation
    - unsupervised confidence-based M-MMI training for model adaptation

Unsupervised Confidence-Based Discriminative Training

- example for a word-graph and the corresponding 1-best state alignment



- necessary steps for **model adaptation during decoding**:
  - 1-pass recognition (unsupervised transcriptions and word-graph)
  - calculation of corresponding confidences (sentence, word, or state-level)
  - unsupervised M-MMI-conf training on test data
    to adapt supervised-trained model (w/ regularization)
- can be done iteratively with unsupervised corpus update!

Unsupervised Confidence-Based Discriminative Training

- **ML training: accumulation of observations x_t:**

$$\textbf{acc}_s = \sum_{r=1}^{R} \sum_{t=1}^{T_r} x_t$$

- **M-MMI training: weighted accumulation of observations x_t:**

$$\textbf{acc}_s = \sum_{r=1}^{R} \sum_{t=1}^{T_r} \omega_{r,s,t} \cdot x_t$$

- **M-MMI-conf training: confidence-weighted accumulation of observations x_t:**

$$\textbf{acc}_s = \sum_{r=1}^{R} \sum_{t=1}^{T_r} \omega_{r,s,t} \cdot c_{r,s,t} \cdot x_t$$

  - **with confidence at sentence-, word, or state-level**

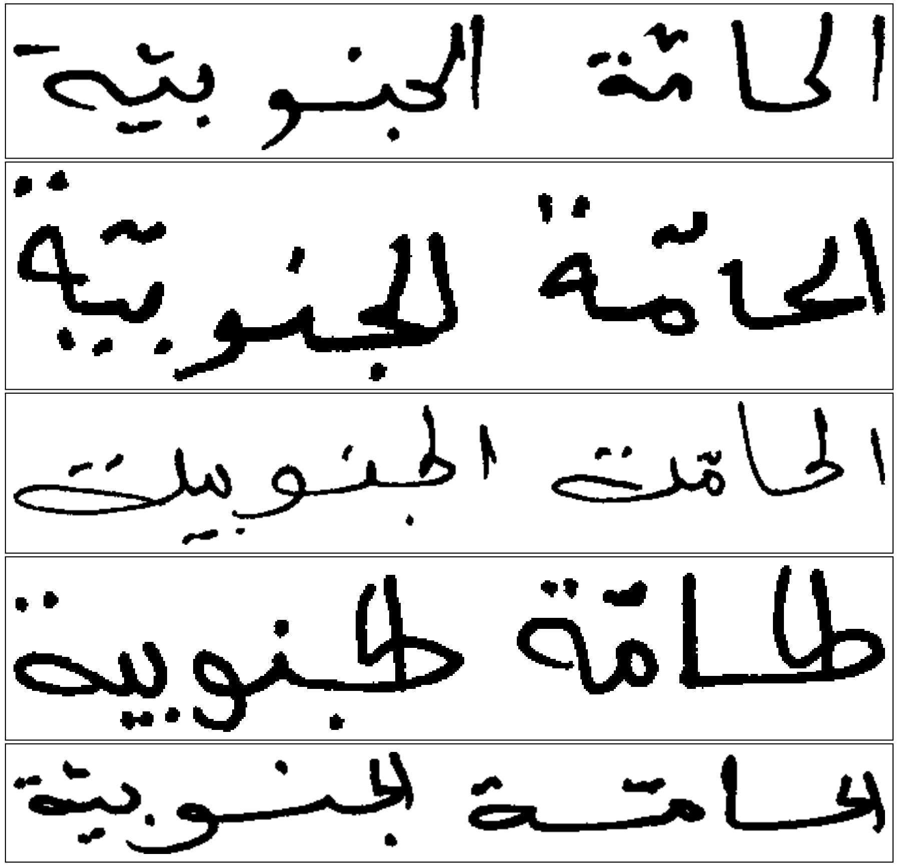
[Details]

Database: IFN/ENIT

- 937 classes
- 32492 handwritten Arabic words (Tunisian city names)
- database is used by more than 60 groups all over the world
- writer statistics

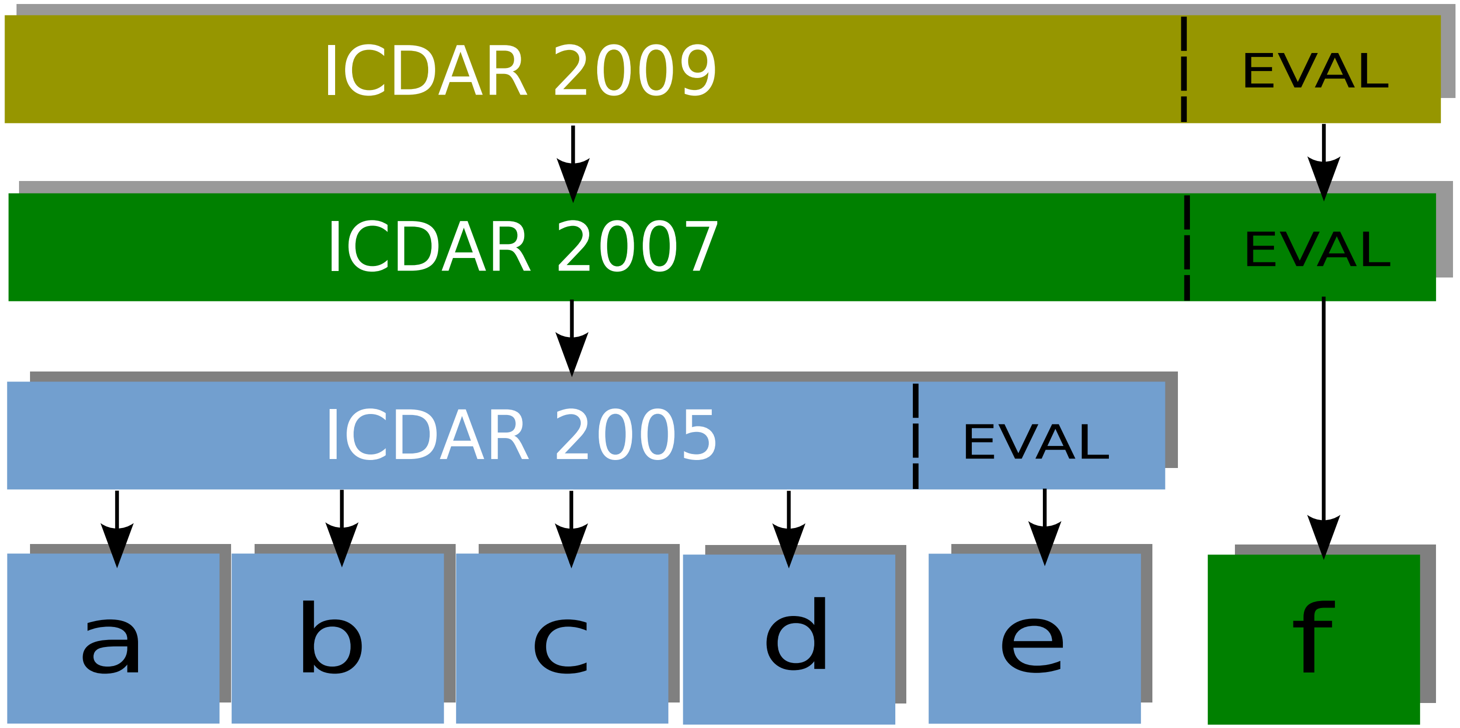
set	#writers	#samples
a	102	6537
b	102	6710
c	103	6477
d	104	6735
e	505	6033
Total	916	32492

- examples (same word):



Corpus development

- ICDAR 2005 Competition: a, b, c, d sets for training, evaluation on set e
- ICDAR 2007 Competition: ICDAR05 + e sets for training, evaluation on set f
- ICDAR 2009 Competition: ICDAR 2007 for training, evaluation on set f



Experimental Results for IFN/ENIT

Train	Test	WER[%]					
		1st pass			2nd pass*		
		ML	+MLE	+M-MMI	WAT+CMLLR		M-MMI-conf
					unsup.	sup.	
abc	d	10.88	7.83	6.12	7.72	5.82	5.95
abd	c	11.50	8.83	6.78	9.05	5.96	6.38
acd	b	10.97	7.81	6.08	7.99	6.04	5.84
bcd	a	12.19	8.70	7.02	8.81	6.49	6.79
abcd	e	21.86	16.82	15.35	17.12	11.22	14.55

- * [Dreuw & Rybach$^+$ 09] and [Dreuw & Heigold$^+$ 09] submitted to ICDAR 2009
- 3 systems are currently evaluated at
  Arabic Handwriting Recognition Competition ICDAR 2009

Summary

- RWTH-ASR $\longrightarrow$ RWTH-OCR
  - simple feature extraction, no preprocessing
  - character model length estimation important
- discriminative training
  - unsupervised confidence-based discriminative training criterion
  - relative improvements of about 14%
  - results outperform all error rates reported in the literature
- ongoing work
  - to be evaluated in ASR experiments
  - impact of preprocessing in feature extraction
  - character context modeling (e.g. CART)
  - further databases

Thank you for your attention

Philippe Dreuw

`dreuw@cs.rwth-aachen.de`

`http://www-i6.informatik.rwth-aachen.de/`

Comparisons for IFN/ENIT

- ICDAR 2005 Evaluation

Rank	Group	WRR [%]	
		abc-d	abcd-e
1.	UOB	85.00	75.93
2.	ARAB-IFN	87.94	74.69
3.	ICRA (Microsoft)	88.95	65.74
4.	SHOCRAN	100.00	35.70
5.	TH-OCR	30.13	29.62
	BBN	89.49	N.A.
1*	RWTH	94.05	85.45

* own evaluation result (no tuning on test data)

Appendix: Modified-MMI Criterion And Confidences

- **Training: weighted accumulation of observations x_t:**

$$\textbf{acc}_s = \sum_{r=1}^{R} \sum_{t=1}^{T_r} \omega_{r,s,t} \cdot x_t$$

1. MMI:

$$\omega_{r,s,t} := \frac{\sum\limits_{s_1^{T_r}: s_t = s} p(x_1^{T_r}|s_1^{T_r}) p(s_1^{T_r}) p(W_r)}{\sum\limits_{V} \sum\limits_{s_1^{T_r}: s_t = s} p(x_1^{T_r}|s_1^{T_r}) p(s_1^{T_r}) p(V)}$$

2. M-MMI:

$$\omega_{r,s,t}(\rho \neq 0) := \frac{\sum\limits_{s_1^{T_r}: s_t = s} p(x_1^{T_r}|s_1^{T_r}) p(s_1^{T_r}) p(W_r) \cdot e^{-\rho\delta(W_r, W_r)}}{\sum\limits_{V} \sum\limits_{s_1^{T_r}: s_t = s} p(x_1^{T_r}|s_1^{T_r}) p(s_1^{T_r}) p(V) \cdot e^{-\rho\delta(W_r, V)}}$$

Appendix: Modified-MMI Criterion And Confidences

3. **M-MMI-conf**:

$$\omega_{r,s,t}(\rho \neq 0) := \underbrace{\frac{\sum_{s_1^{T_r}:s_t=s} p(x_1^{T_r}|s_1^{T_r})p(s_1^{T_r})p(W_r) \cdot e^{-\rho\delta(W_r,W_r)}}{\sum_V \sum_{s_1^{T_r}:s_t=s} p(x_1^{T_r}|s_1^{T_r})p(s_1^{T_r})p(V) \cdot \underbrace{e^{-\rho\delta(W_r,V)}}_{\text{margin}}}}_{\text{posterior}} \cdot \underbrace{\delta(c_{r,s,t} > c_{\text{threshold}})}_{\text{confidence}}$$

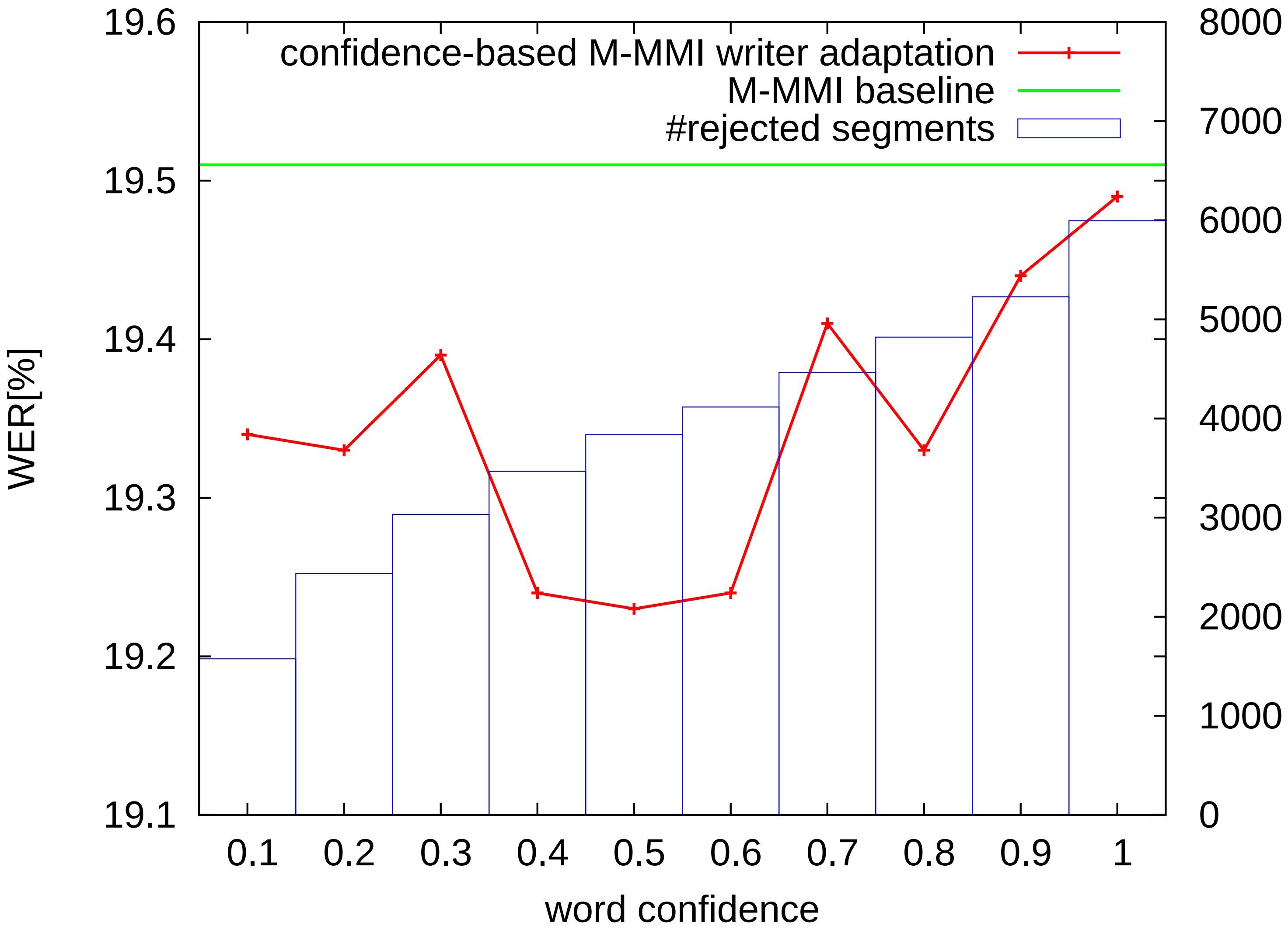
- **weighted accumulation becomes:**

$$\text{acc}_s = \sum_{r=1}^{R} \sum_{t=1}^{T_r} \underbrace{\omega_{r,s,t}(\rho)}_{\text{margin posterior}_{\rho\neq 0}} \cdot \underbrace{c_{r,s,t}}_{\text{confidence}} \cdot x_t$$

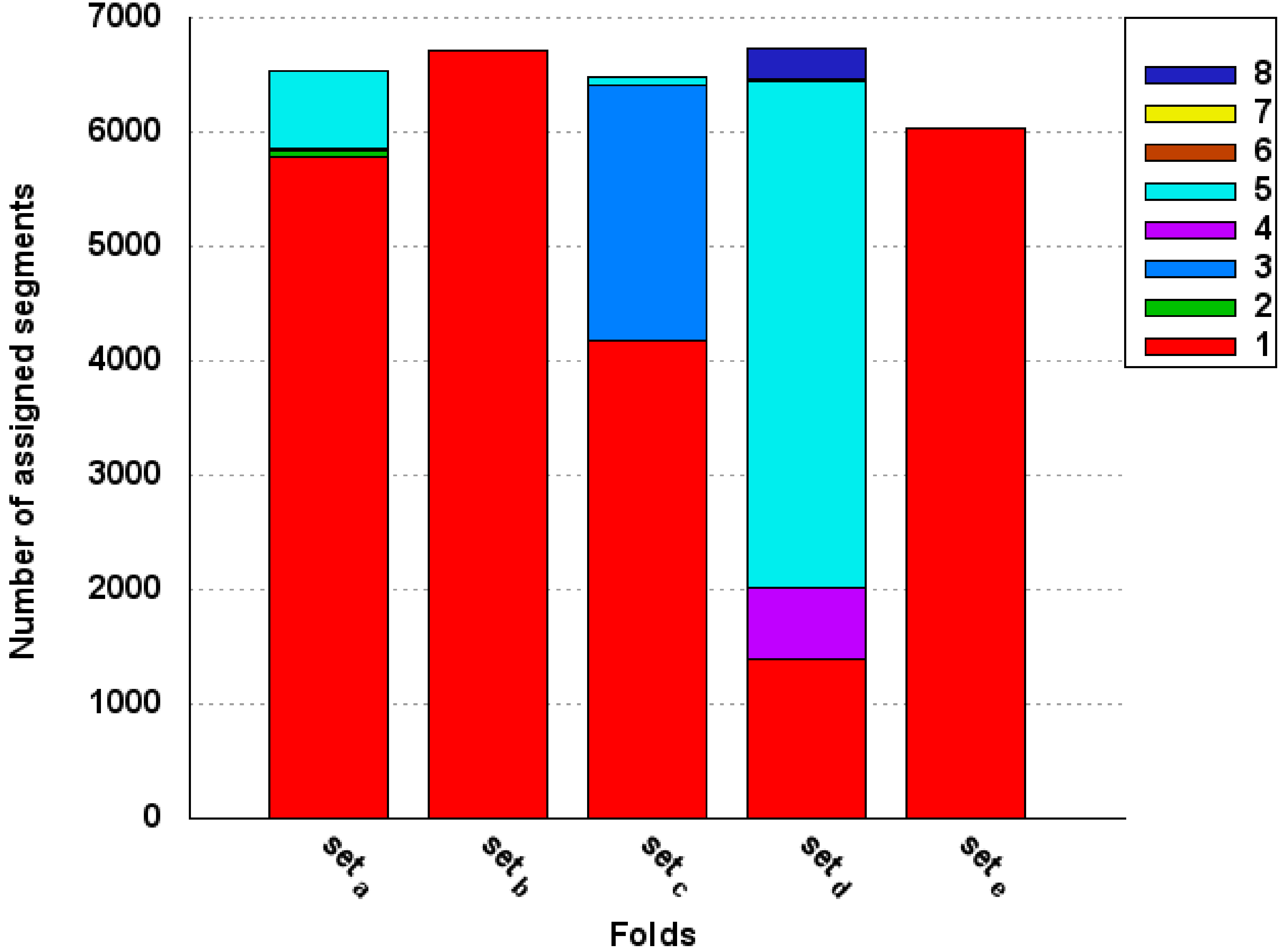
- **confidences at:**
  - **sentence-, word-, or state-level**

Appendix: Modified-MMI Criterion And Confidences

▶ **word-confidence based M-MMI training and rejections**



Appendix: Segment Clustering Histograms



Appendix: Participating Systems at ICDAR 2005 and 2007

- **MITRE: Mitre Cooperation, USA**
  over-segmentation, adaptive lengths, character recognition with post-processing
- **UOB-ENST: University of Balamand (UOB), Lebanon and Ecole Nationale Superieure des Telecommunications (ENST), Paris**
  HMM-based (HTK), slant correction
- **MIE: Mie University, Japan**
  segmentation, adaptive lengths
- **ICRA: Intelligent Character Recognition for Arabic, Microsoft**
  partial word recognizer
- **SHOCRAN: Egypt**
  confidential
- **TH-OCR: Tsinghua Universty, Beijing, China**
  over-segmentation, character recognition with post-processing
- **CACI: Knowledge and Information Management Division, Lanham, USA**
  HMM-based, trajectory features
- **CEDAR: Center of Excellence for Document Analysis and Recognition, Buffalo, USA**
  over-segmentation, HMM-based
- **PARIS V / A2iA: University of Paris 5, and A2iA SA, France**
  hybrid HMM/NN-based, shape-alphabet
- **Siemens: SIEMENS AG Industrial Solutions and Services, Germany**
  HMM-based, adapative lenghths, writing variants
- **ARAB-IFN: TU Braunschweig, Germany**
  HMM-based

References

[Dreuw & Heigold^{+} 09] P. Dreuw, G. Heigold, H. Ney: Confidence-Based Discriminative Training for Writer Adaptation in Offline Arabic Handwriting Recognition. In *International Conference on Document Analysis and Recognition*, Barcelona, Spain, July 2009. 15

[Dreuw & Jonas^{+} 08] P. Dreuw, S. Jonas, H. Ney: White-Space Models for Offline Arabic Handwriting Recognition. In *International Conference on Pattern Recognition*, Tampa, Florida, USA, Dec. 2008. 6

[Dreuw & Rybach^{+} 09] P. Dreuw, D. Rybach, C. Gollan, H. Ney: Writer Adaptive Training and Writing Variant Model Refinement for Offline Arabic Handwriting Recognition. In *International Conference on Document Analysis and Recognition*, Barcelona, Spain, July 2009. 15