Writer Adaptive Training and
Writing Variant Model Refinement
for Offline Arabic Handwriting Recognition

Philippe Dreuw, David Rybach,
Christian Gollan, and Hermann Ney
dreuw@cs.rwth-aachen.de

International Conference on Document Analysis and Recognition – July 2009

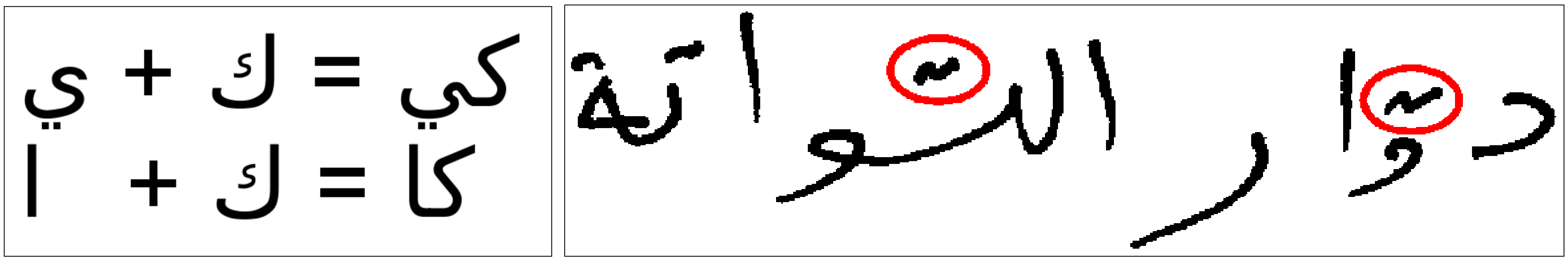
Human Language Technology and Pattern Recognition
Lehrstuhl für Informatik 6
Computer Science Department
RWTH Aachen University, Germany

Outline

1. **Introduction**
2. **Adaptation of an ASR framework for Handwriting Recognition**
   - **Writing Variant Model Refinement**
   - **Writer Adaptation and Writer Adaptive Training**
3. **Experimental Results**
4. **Summary**

Introduction

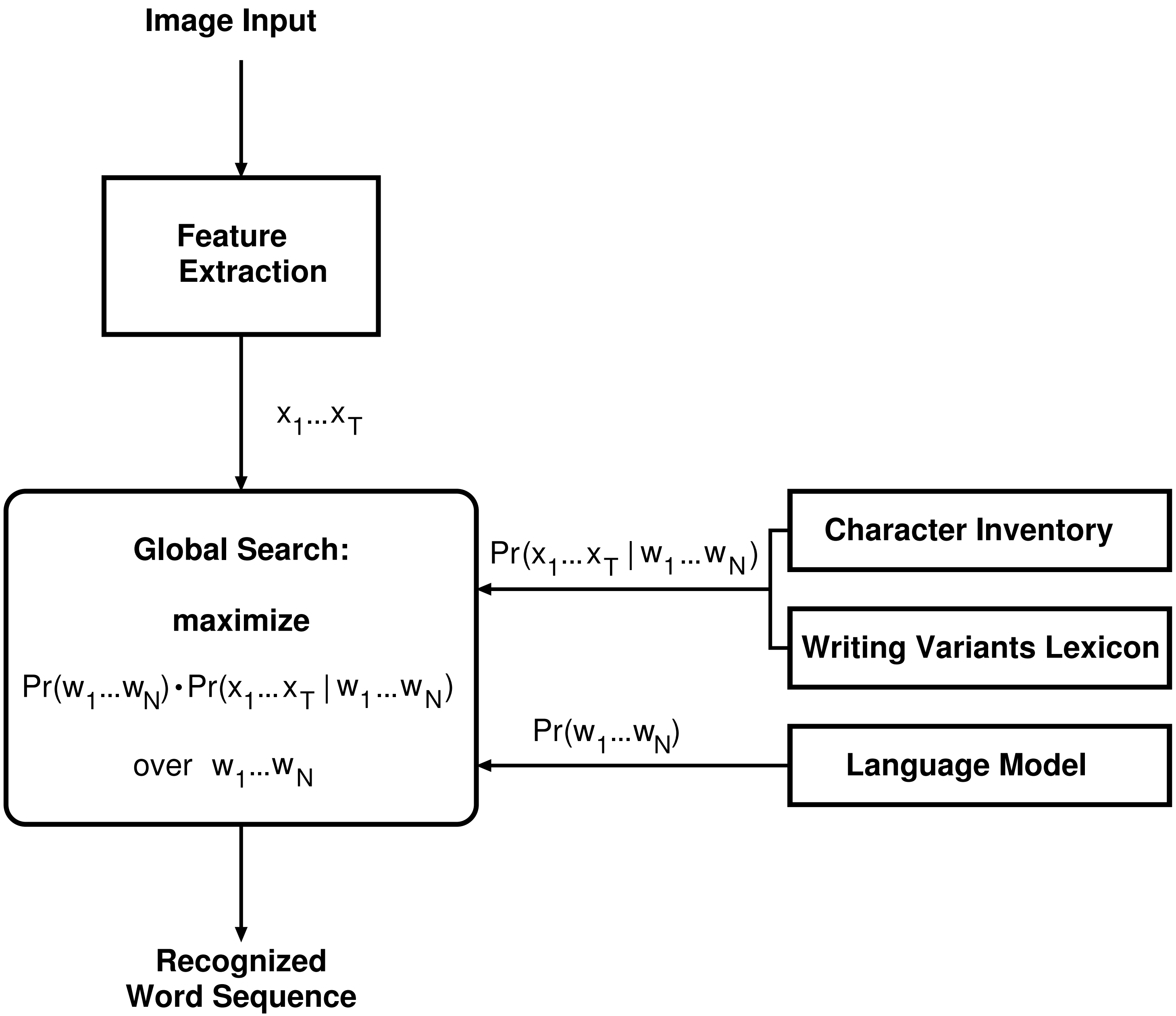
- **Arabic handwriting system**
  - **right-to-left, 28 characters, position-dependent character writing variants**
  - **ligatures and diacritics**
  - **Pieces of Arabic Word (PAWs) as subwords**



(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



Writing Variant Model Refinement

- **HMM baseline system**
  - **searching for an unknown word sequence** $w_1^N := w_1, \ldots, w_N$
  - **unknown number of words** N
  - **maximize the posterior probability** $p(w_1^N | x_1^T)$
  - **described by Bayes' decision rule:**

$$\hat{w}_1^N = \arg\max_{w_1^N} \left\{ p^\gamma(w_1^N) p(x_1^T | w_1^N) \right\}$$

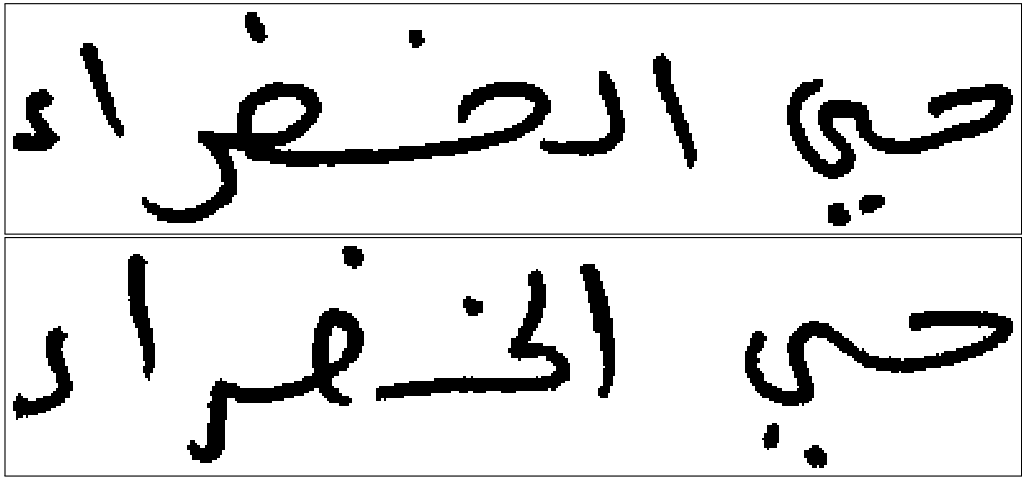
with γ **a scaling exponent of the language model.**

Writing Variant Model Refinement

- ▶ **ligatures and diacritics in Arabic handwriting**
  - ▷ **same Arabic word can be written in several writing variants**

  ⟶ depends on writer's handwriting style

- ▶ **Example: *laB khM* vs. *khMlaB***



- ▶ **lexicon with multiple writing variants** [Details]
  - ▷ **problem: many and rare writing variants**

Writing Variant Model Refinement

- **probability $p(v|w)$ for a variant v of a word w**
  - **usually considered as equally distributed**
  - **here: we use the count statistics as probability:**

$$p(v|w) = \frac{N(v,w)}{N(w)}$$

- **writing variant model refinement:**

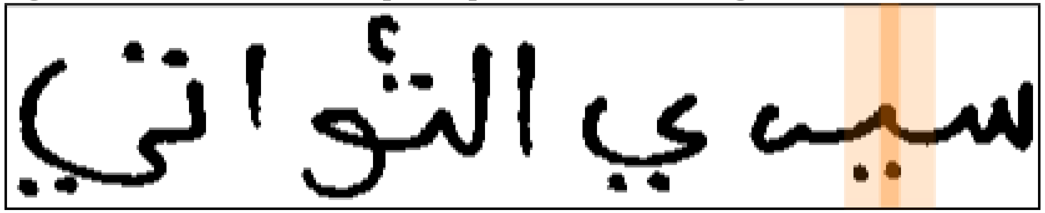
$$p(x_1^T|w_1^N) \approx \max_{v_1^N|w_1^N} \left\{ p^\alpha(v_1^N|w_1^N) p(x_1^T|v_1^N, w_1^N) \right\}$$

with v_1^N a sequence of unknown writing variants
α a scaling exponent of the writing variant probability

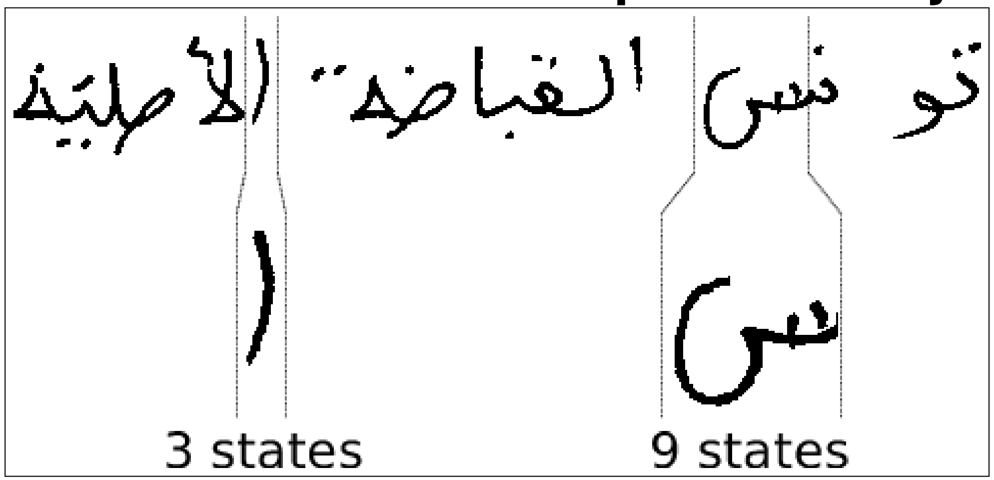
- **training: corpus and lexicon with supervised writing variants possible!**

Visual Modeling

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



- **Model Length Estimation (MLE)**
  - **more complex characters should be represented by more HMM states**



[Dreuw & Jonas+ 08] in ICPR 2008

RWTH-OCR 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**

Constrained Maximum Likelihood Linear Regression (CMLLR)

- **writer adaptation**
  - method for improving visual models in handwriting recognition
  - refine models by adaptation data of particular writers
  - widely used is affine transform based model adaptation

- **CMLLR**
  - **Idea:** normalize writing styles by **adaptation of the features** x_t
  - constrained MLLR feature adaptation technique
  - also known as feature space MLLR (fMLLR) [Details]
  - estimate affine feature transform:

$$x'_t = Ax_t + b$$

  - CMLLR is text dependent
    - requires an (automatic) transcription

Training: CMLLR-based Writer Adaptive Training

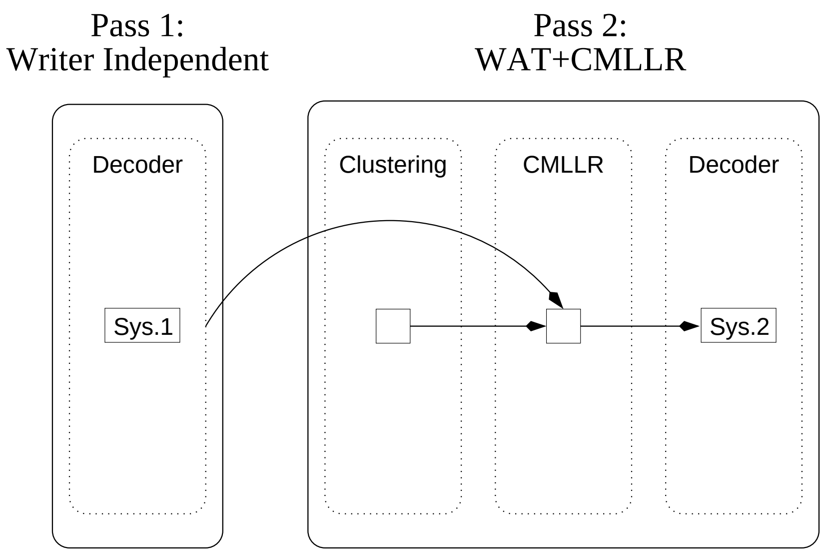
- writer adaptation compensates for writer differences during recognition
  $\longrightarrow$ do the same during visual model training
  $\longrightarrow$ maximize the performance gains from writer adaptation

- writer variations are compensated by writer adaptive training (WAT)
- writer normalization using CMLLR
- necessary steps
  1. train writer independent GMMs model
  2. CMLLR transformations are estimated for each (estimated) writer
     - supervised if writers are known
  3. apply CMLLR transformations on features to train writer dependent GMMs

Decoding: CMLLR-based Writer Adaptation

- writers and writing styles are unknown
- necessary steps

1. estimate writing styles using clustering
   - Bayesian Information Criterion (BIC) based stopping condition
2. estimate CMLLR feature transformations
   for every estimated writing style cluster
3. second pass recognition
   - WAT models + CMLLR transformed features

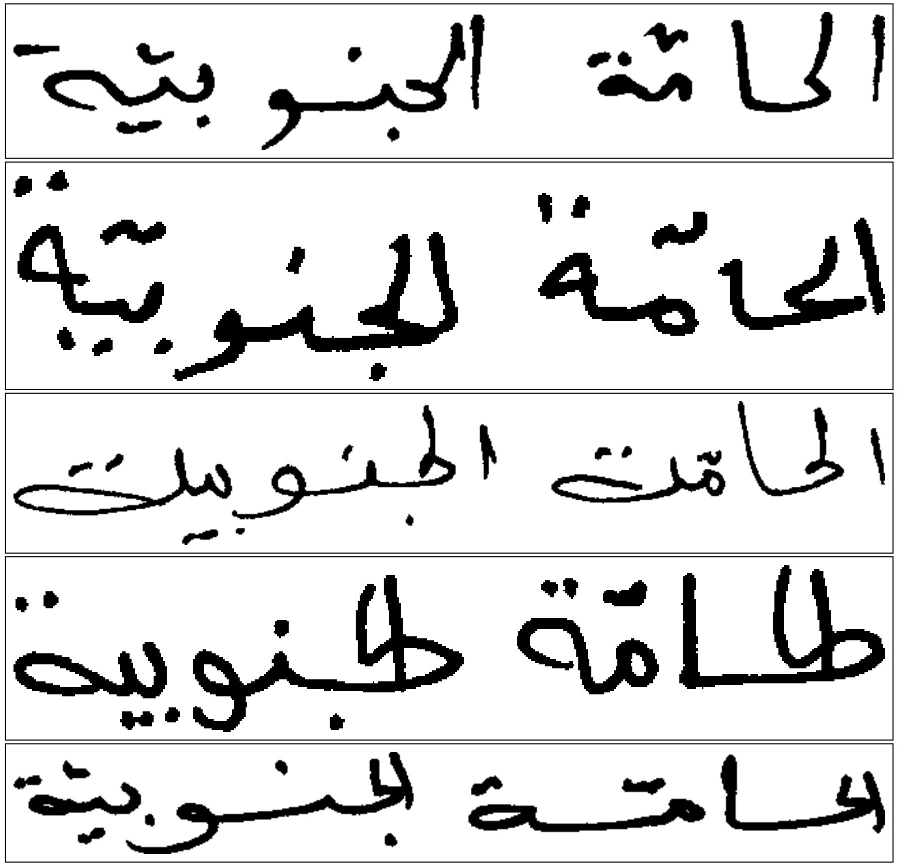


Arabic Handwriting - IFN/ENIT Database

- 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):



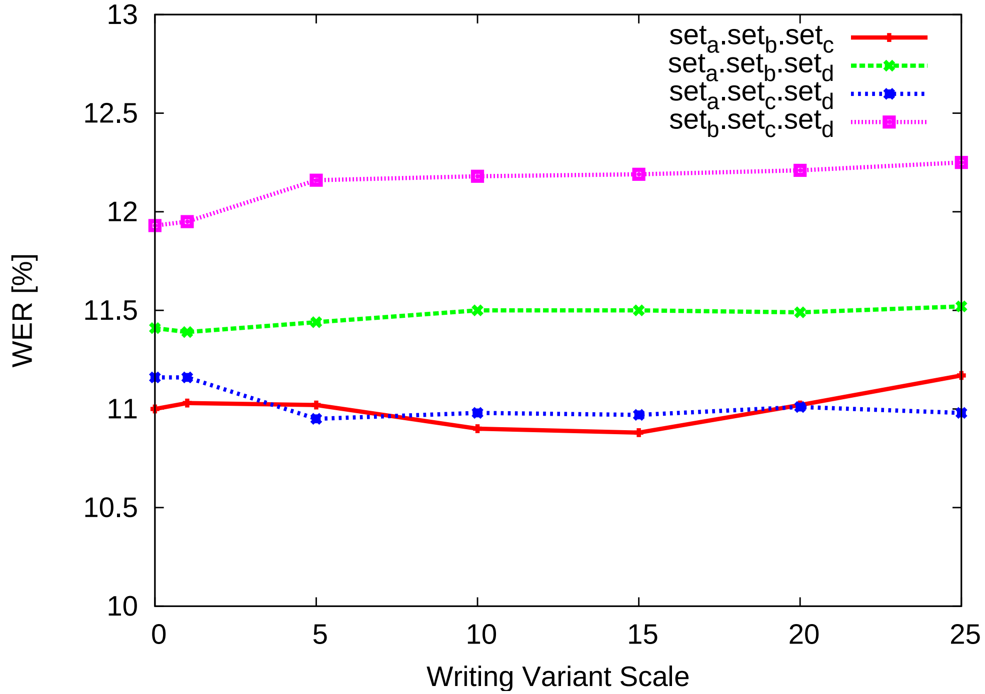
Results - Training: Writing Variant Model Refinement

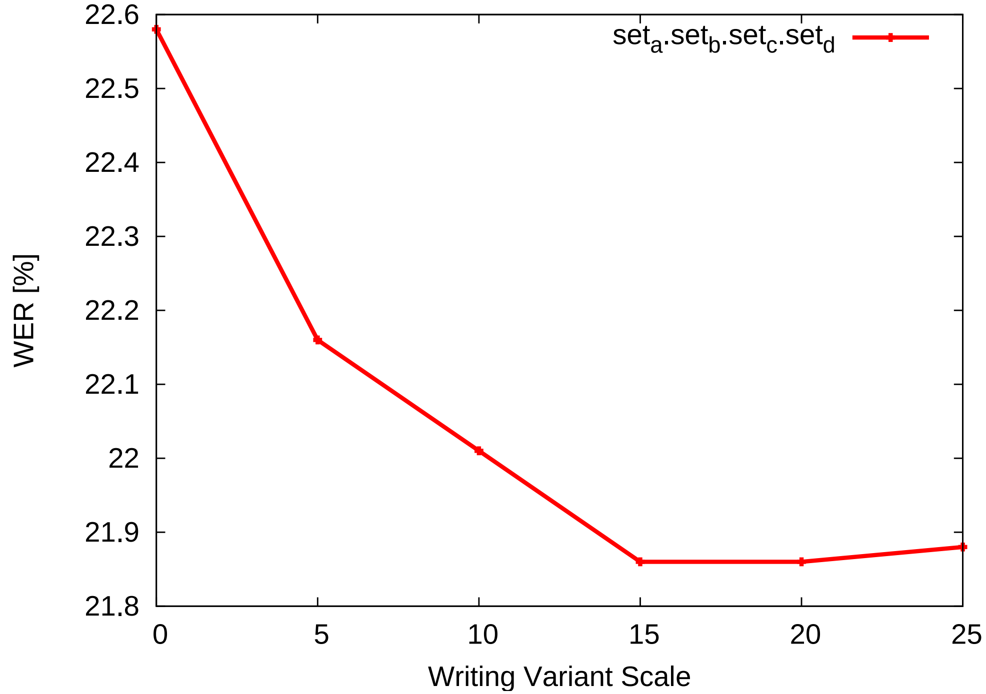
- **comparison of supervised and unsupervised writing variants in training**

Train	Test	unsupervised		supervised	
		WER[%]	CER[%]	WER[%]	CER[%]
abc	d	11.60	3.88	**11.00**	**3.66**
abd	c	12.95	4.60	**11.41**	**3.97**
acd	b	11.98	3.91	**11.16**	**3.65**
bcd	a	12.33	4.26	**11.93**	**4.27**
abcd	e	24.60	9.34	**22.58**	**8.39**

Results - Decoding: Writing Variant Model Refinement

- **empirical optimization of the writing variant scale α on the cross folds**
- **verification on the development set**





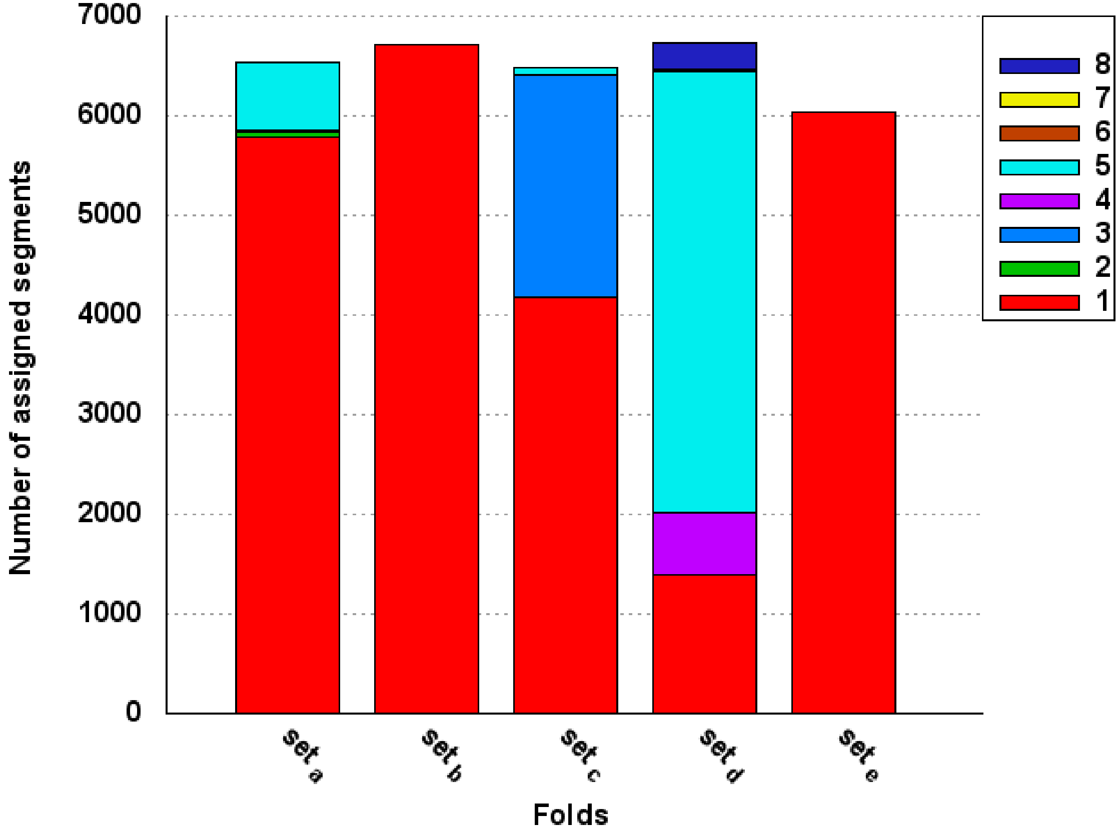
Results - Decoding: Writer Adaptation

- **comparison of MLE, WAT, and CMLLR based feature adaptation**
- **comparison of unsupervised and supervised writer clustering**
  - **decoding always unsupervised**
  - **supervised clustering → only the writer labels!**

Train	Test	WER[%]			
		1st pass		2nd pass	
		SWV	+MLE	WAT+CMLLR	
				unsup.	sup.
abc	d	10.88	**7.83**	**7.72**	5.82
abd	c	11.50	**8.83**	**9.05**	5.96
acd	b	10.97	**7.81**	**7.99**	6.04
bcd	a	12.19	**8.70**	**8.81**	6.49
abcd	e	21.86	**16.82**	**17.12**	11.22

Results - Decoding: Writer Adaptation

- **unsupervised clustering: error analysis**
  - **histograms for segment assignments over the different test folds**
  - **problem: unbalanced segment assignments**



Summary

- RWTH-ASR $\longrightarrow$ RWTH-OCR
  - simple feature extraction and preprocessing
  - writing variants model refinement
  - character model length estimation
- writer adaptive training
  - supervised writer adaptation demonstrated the potential
  - relative improvements of about 33% w.r.t. ML training
- ongoing work
  - improve unsupervised writer clustering
  - discriminative training
    - ranked 3rd at Arabic HWR Competition, ICDAR 2009
    - see second talk (Tuesday, Session 5.2)
  - impact of preprocessing in feature extraction (Arabic vs. Latin)
  - more complex features (e.g. MLP)
  - character context modeling (e.g. CART)
  - further databases/languages

Thank you for your attention

Philippe Dreuw

`dreuw@cs.rwth-aachen.de`

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

References

[Bertolami & Bunke 08] R. Bertolami, H. Bunke: Hidden Markov model-based ensemble methods for offline handwritten text line recognition. *Pattern Recognition*, Vol. 41, No. 11, pp. 3452–3460, Nov 2008.

[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. 8, 28

[Jonas 09] S. Jonas: Improved Modeling in Handwriting Recognition. Master's thesis, Human Language Technology and Pattern Recognition Group, RWTH Aachen University, Aachen, Germany, Jun 2009.

[Natarajan & Saleem+ 08] P. Natarajan, S. Saleem, R. Prasad, E. MacRostie, K. Subramanian: *Arabic and Chinese Handwriting Recognition*, Vol. 4768/2008 of *LNCS*, chapter Multi-lingual Offline Handwriting Recognition Using Hidden Markov Models: A Script-Independent Approach, pp. 231–250. Springer Berlin / Heidelberg, 2008.

[Romero & Alabau+ 07] V. Romero, V. Alabau, J.M. Benedi: Combination of N-Grams and Stochastic Context-Free Grammars in an Offline Handwritten

Recognition System. *Lecture Notes in Computer Science*, Vol. 4477, pp. 467–474, 2007.

Appendix: 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**

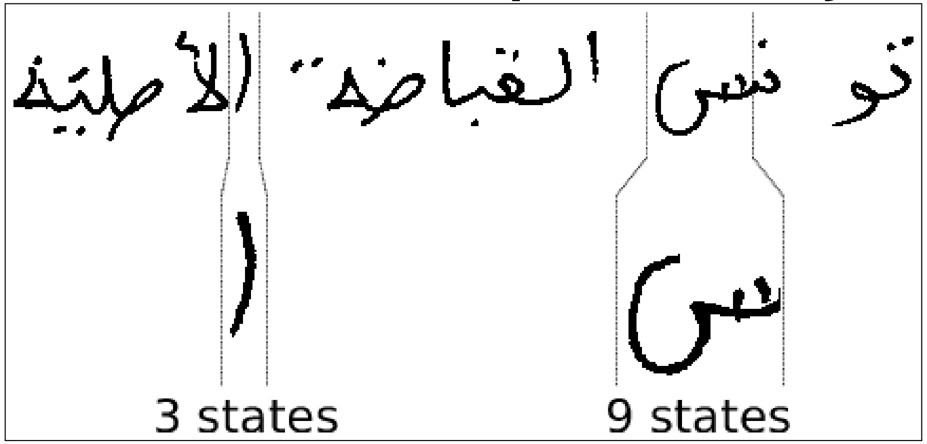
***own evaluation result (no tuning on test data)**

Appendix: 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

Appendix: 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**

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

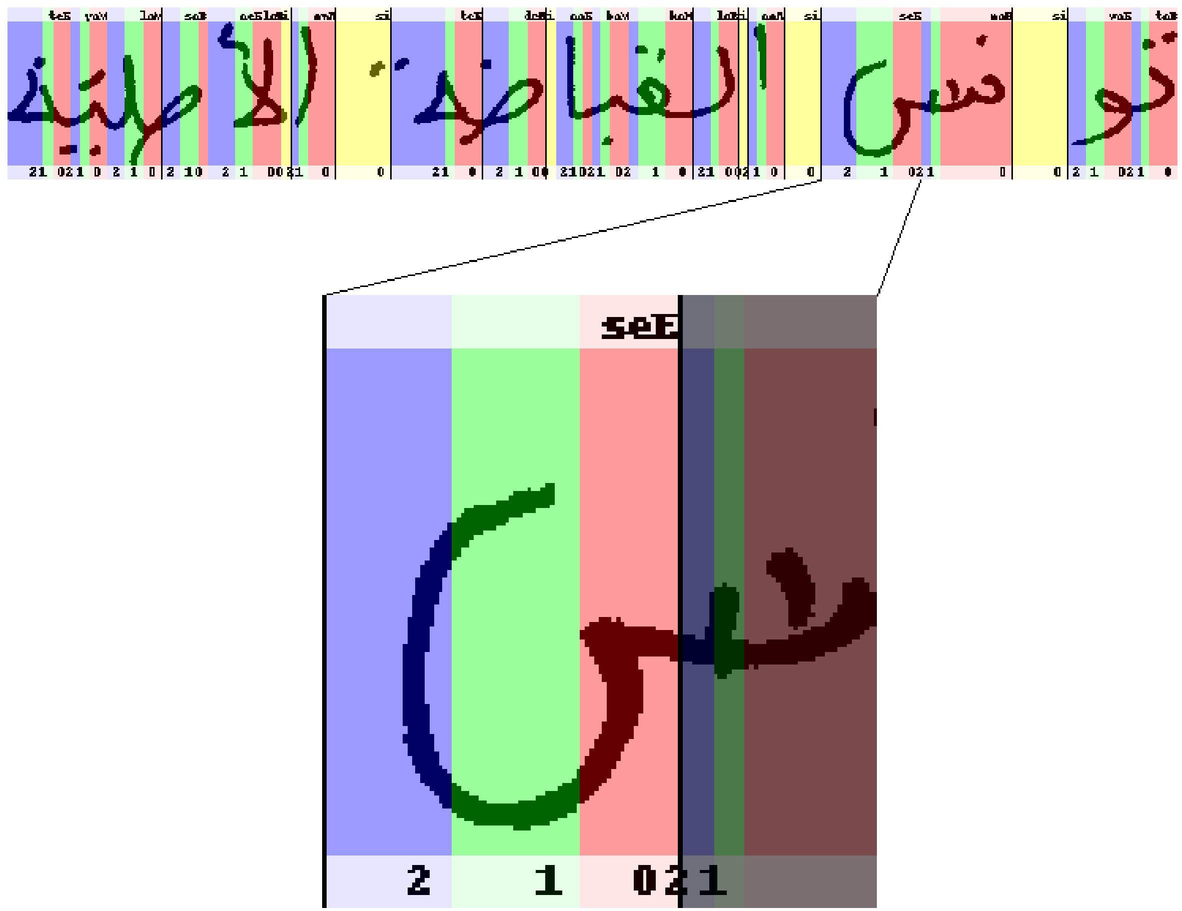
with

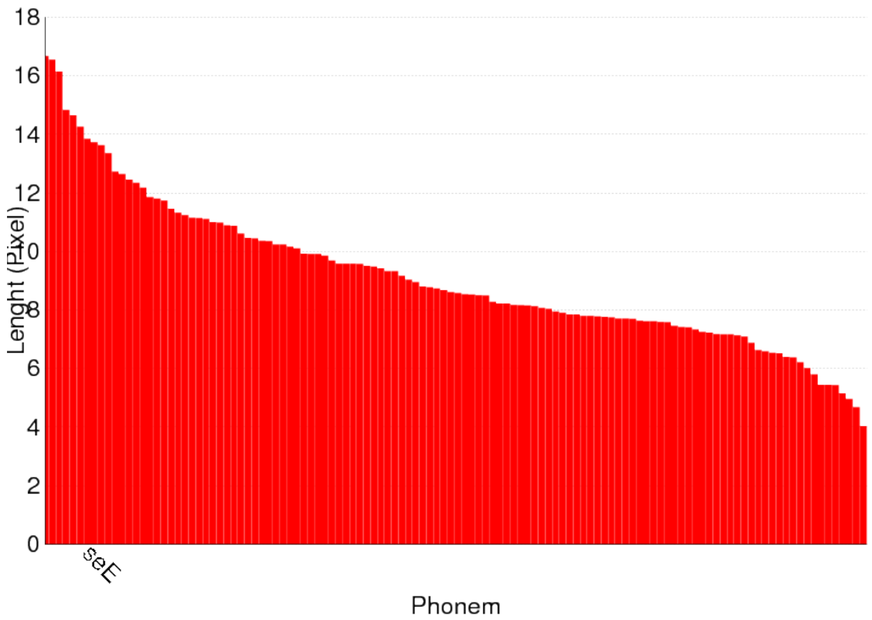
S_c = **estimated number states for character c**
$N_{x,c}$ = **number of observations aligned to character c**
N_c = **character count of c seen in training**
α = **character length scaling factor.**

Appendix: Visual Modeling - Model Length Estimation

Original Length

- overall mean of character length = 7.9 pixel ($\approx$ 2.6 pixel/state)
- total #states = 357

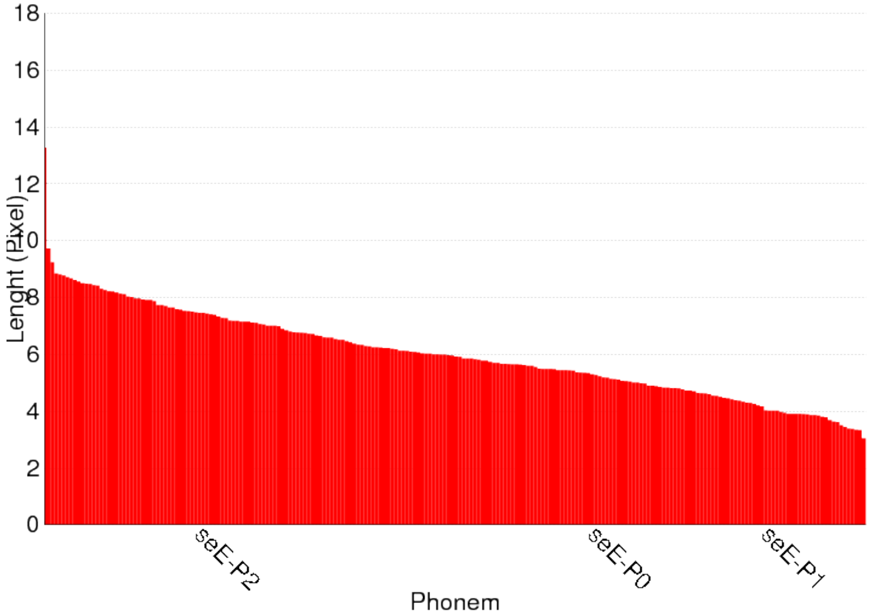


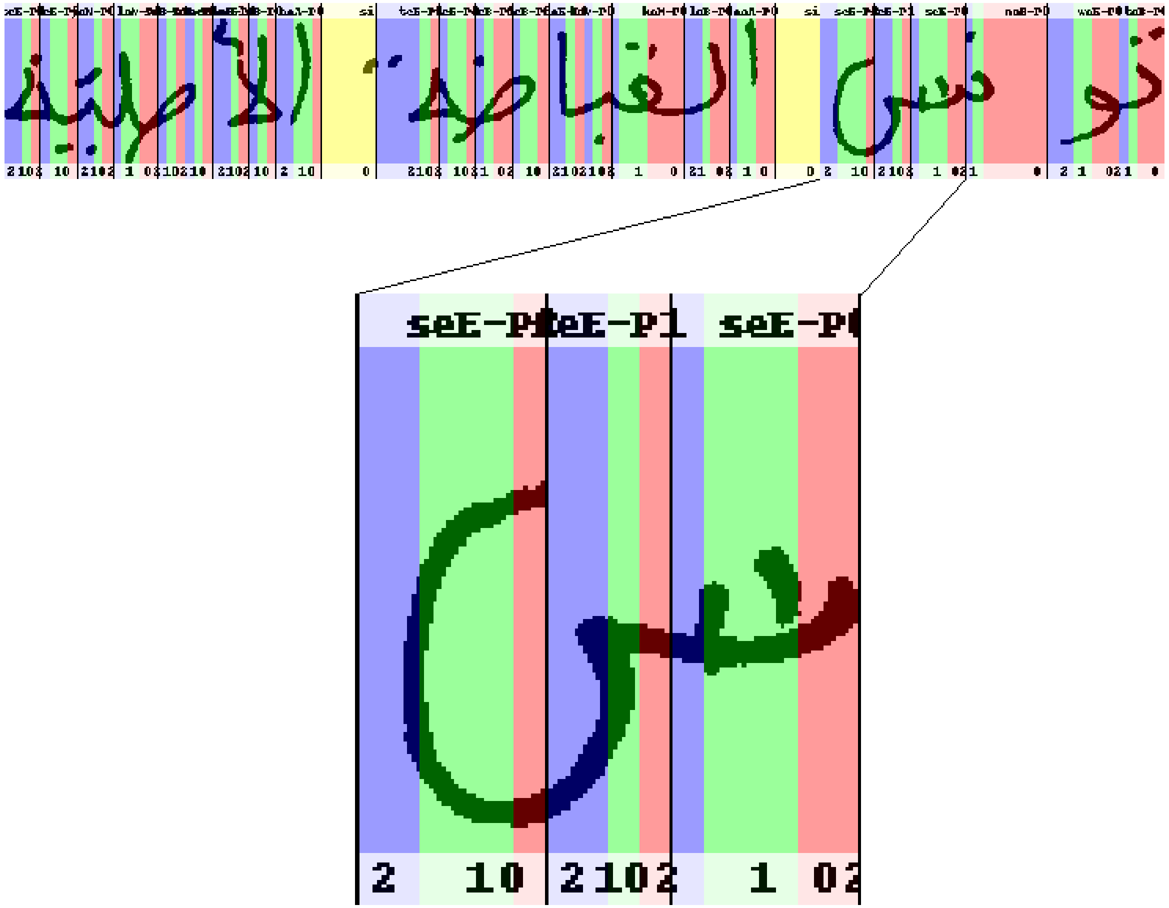


Appendix: Visual Modeling - Model Length Estimation

Estimated Length

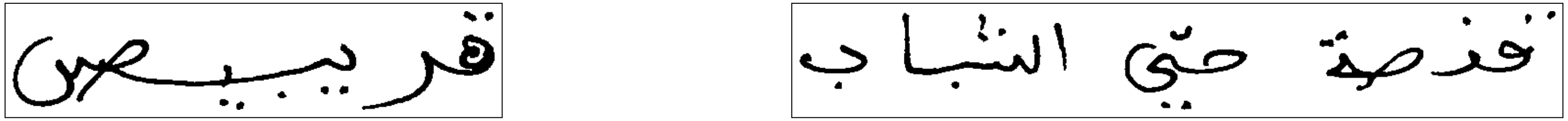
- **overall mean of character length = 6.2 pixel (≈ 2.0 pixel/state)**
- **total #states = 558**



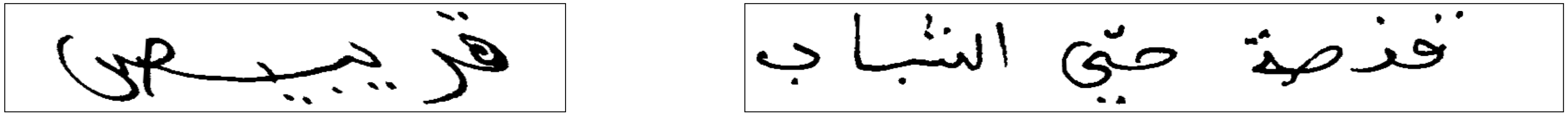


Appendix: Arabic Handwriting - UPV Preprocessing

- **Original images**



- **Images after slant correction**



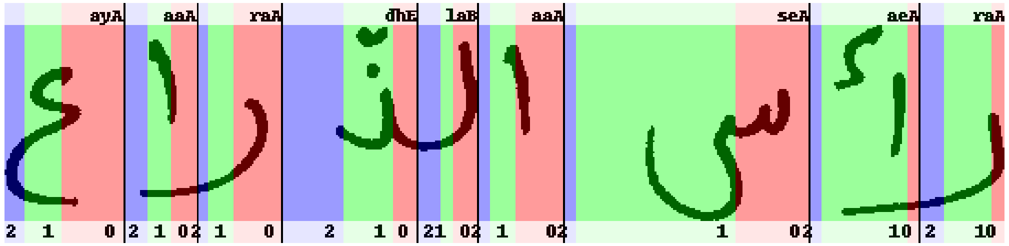
- **Images after size normalisation**

Experimental Results:

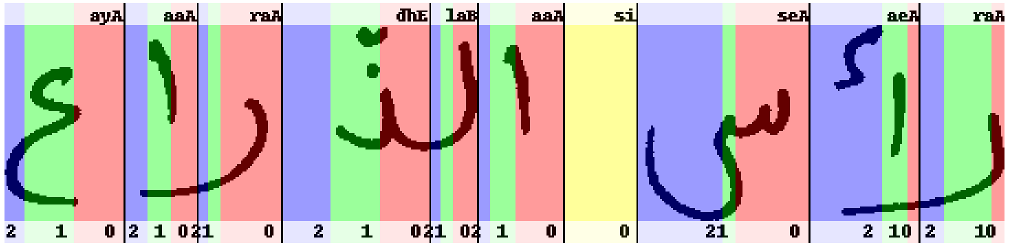
- **important informations in ascender and descender areas are lost**
- **not yet suitable for Arabic HWR**

Appendix: Visual Modeling - Writing Variants Lexicon

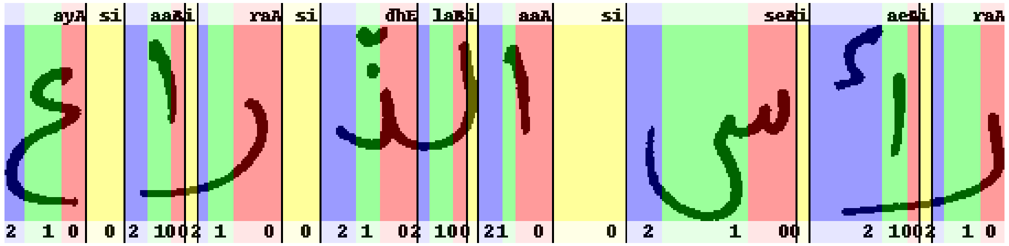
- **most reported error rates are dependent on the number of PAWs**
- **without separate whitespace model**



- **always whitespaces between compound words**



- **whitespaces as writing variants between and within words**

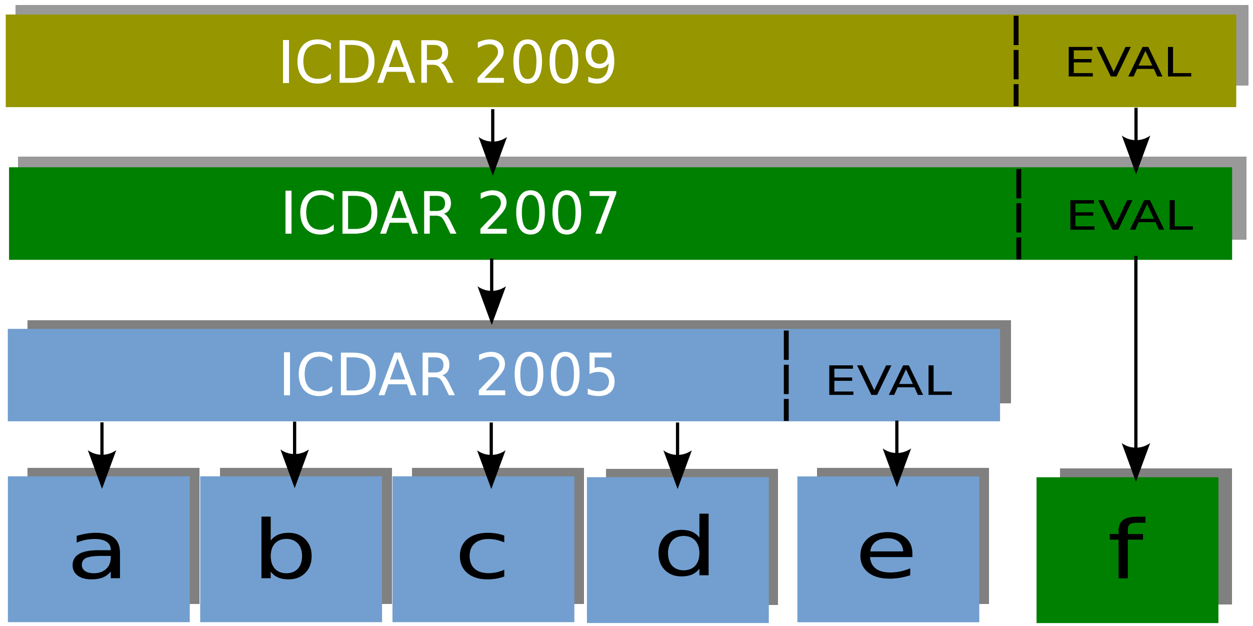


White-Space Models for Pieces of Arabic Words [Dreuw & Jonas^{+} 08] in ICPR 2008

Appendix: Arabic Handwriting - IFN/ENIT Database

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**



Appendix: Constrained Maximum Likelihood Linear Regression

Idea: improve the hypotheses by adaptation of the features x_t

- effective algorithm for adaptation to a new speaker or environment (ASR)
- GMMs are used to estimate the CMLLR transform
- iterative optimization (ML criterion)
  - align each frame x to one HMM state (i.e. GMM)
  - accumulate to estimate the adaptation transform A
  - likelihood function of the adaptation data given the model is to be maximized with respect to the transform parameters A, b
- one CMLLR transformation per (estimated) writer
- **constrained** refers to the use of the same matrix A for the transformation of the mean μ and variance Σ:

$$x'_t = Ax_t + b \longrightarrow N(x|\hat{\mu}, \hat{\Sigma}) \text{ with } \begin{aligned} \hat{\mu} &= A\mu + b \\ \hat{\Sigma} &= A\Sigma A^T \end{aligned}$$