Confidence-Based Discriminative Training for Model Adaptation in Offline Arabic Handwriting Recognition

Philippe Dreuw, Georg Heigold, and Hermann Ney
dreuw@cs.rwth-aachen.de

International Conference on Document Analysis and Recognition – July 2009

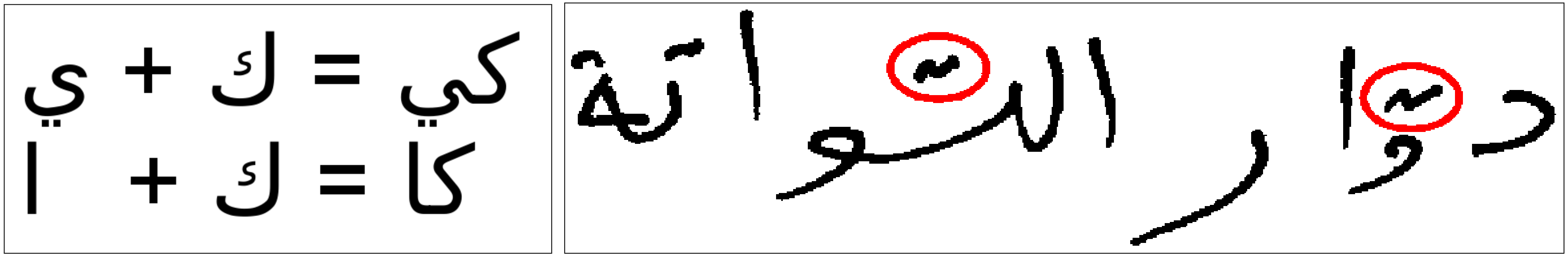
Human Language Technology and Pattern Recognition
Lehrstuhl für Informatik 6
Computer Science Department
RWTH Aachen University, Germany

Outline

1. Introduction
2. Adaptation of the ASR framework for Handwriting Recognition
   - discriminative training using modified MMI criterion
   - unsupervised confidence-based discriminative training during decoding
3. Experimental Results
4. Summary

Introduction

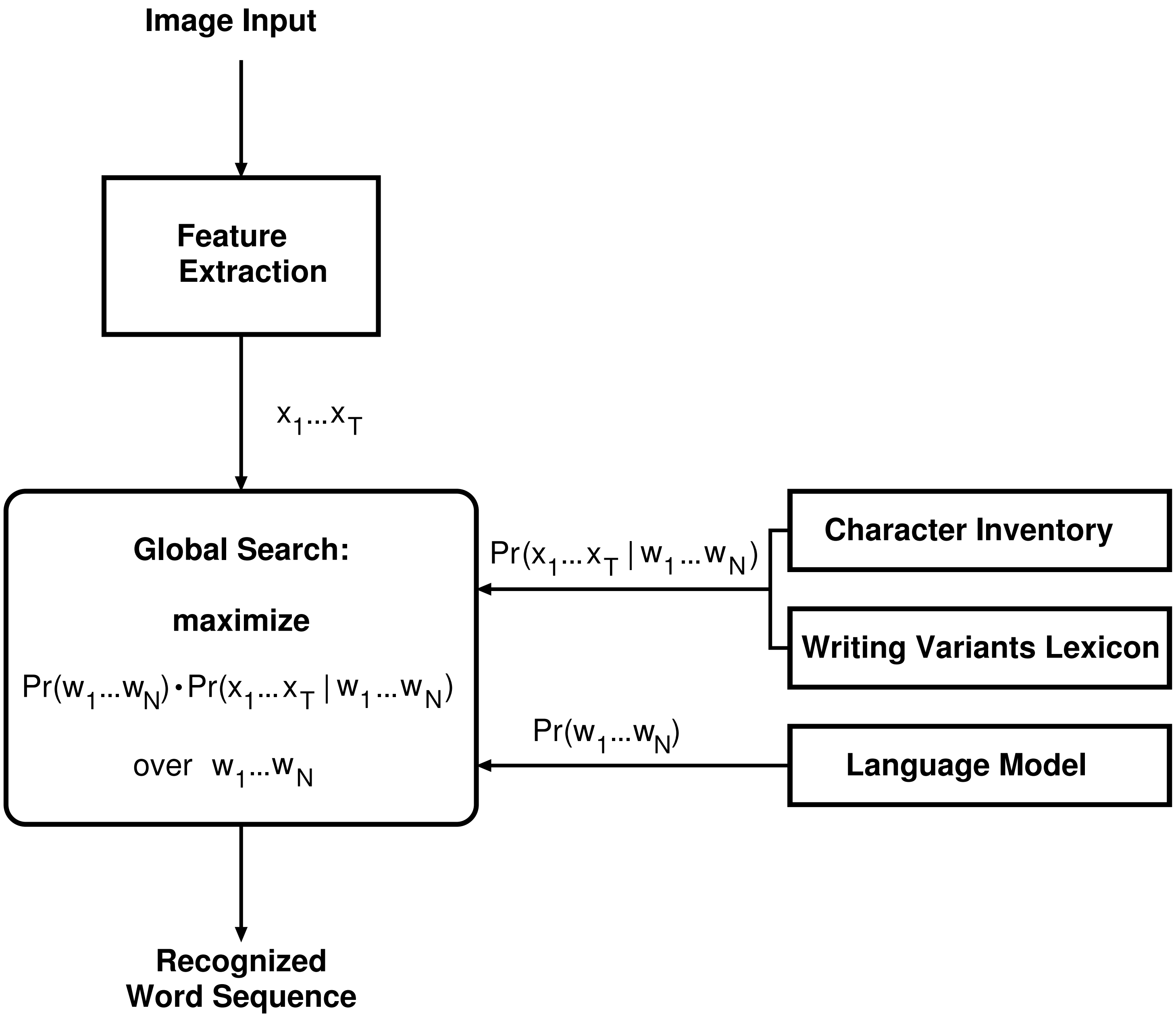
- **Arabic handwriting system**
  - **right-to-left, 28 characters, position-dependent character writing variants**
  - **ligatures and diacritics**
  - **Pieces of Arabic Word (PAWs) as subwords**



(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**

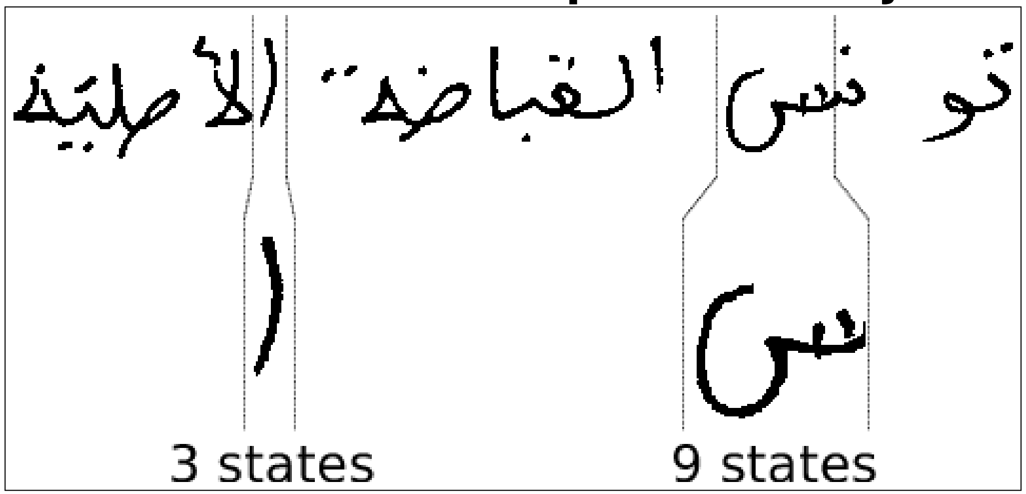
System Overview



Visual Modeling: Feature Extraction and Model Length Estimation

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

- more complex characters should be represented by more HMM states



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**

Training: Modified-MMI Criterion

- 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. ML: Maximum Likelihood

$$\omega_{r,s,t} := 1.0$$

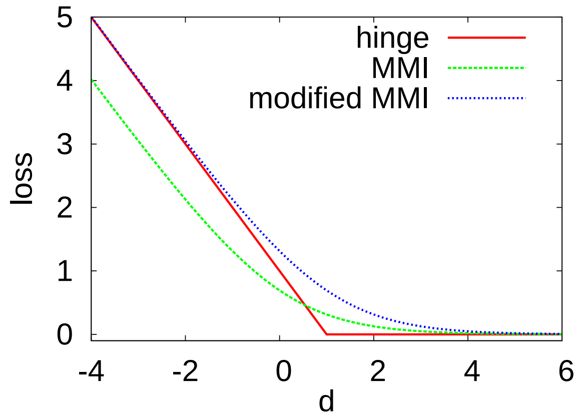
2. MMI: Maximum Mutual Information

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

- $\omega_{r,s,t}$ is the “(true) posterior” weight
- iteratively optimized with Rprop

Training: Modified-MMI Criterion

- **margin-based training for HMMs**
  - **similar to SVM training, but simpler/faster within RWTH-OCR framework?**
  - **M-MMI = differentiable approximation to SVM optimization**



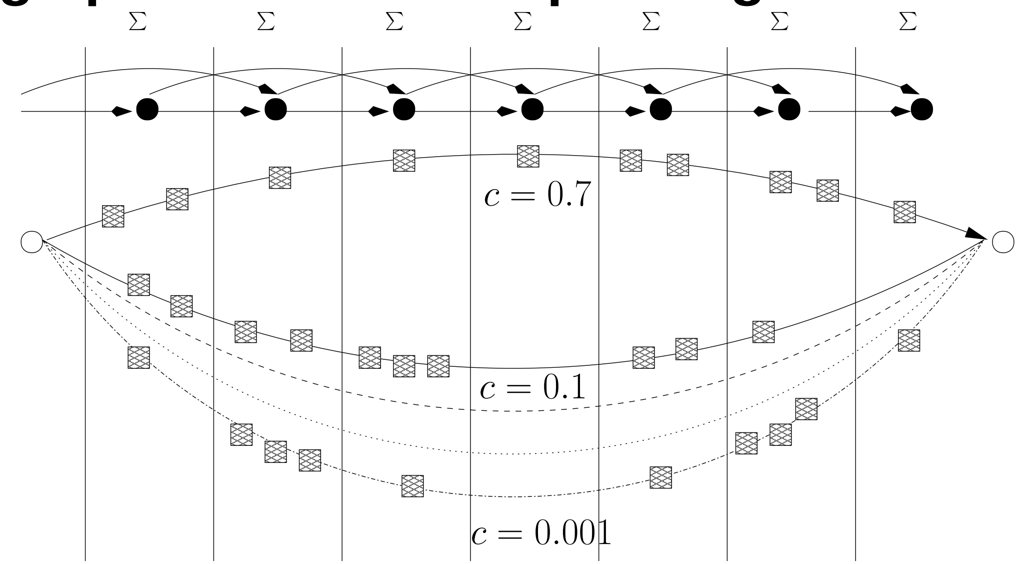
3. M-MMI:

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

- $\omega_{r,s,t}$ **is the "margin posterior" weight**
- $e^{-\rho\delta(W_r,W_r)}$ **corresponds to the margin offset**
- **with $\gamma \rightarrow \infty$ equals to the SVM hinge loss function**
- **iteratively optimized with Rprop**

Decoding: Unsupervised Confidence-Based Discriminative Training

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



- necessary steps for **margin-based 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 models (w/ regularization)
- can be done iteratively with unsupervised corpus update!

Decoding: Modified-MMI Criterion And Confidences

4. 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**

Training Criterions

- **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:**

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

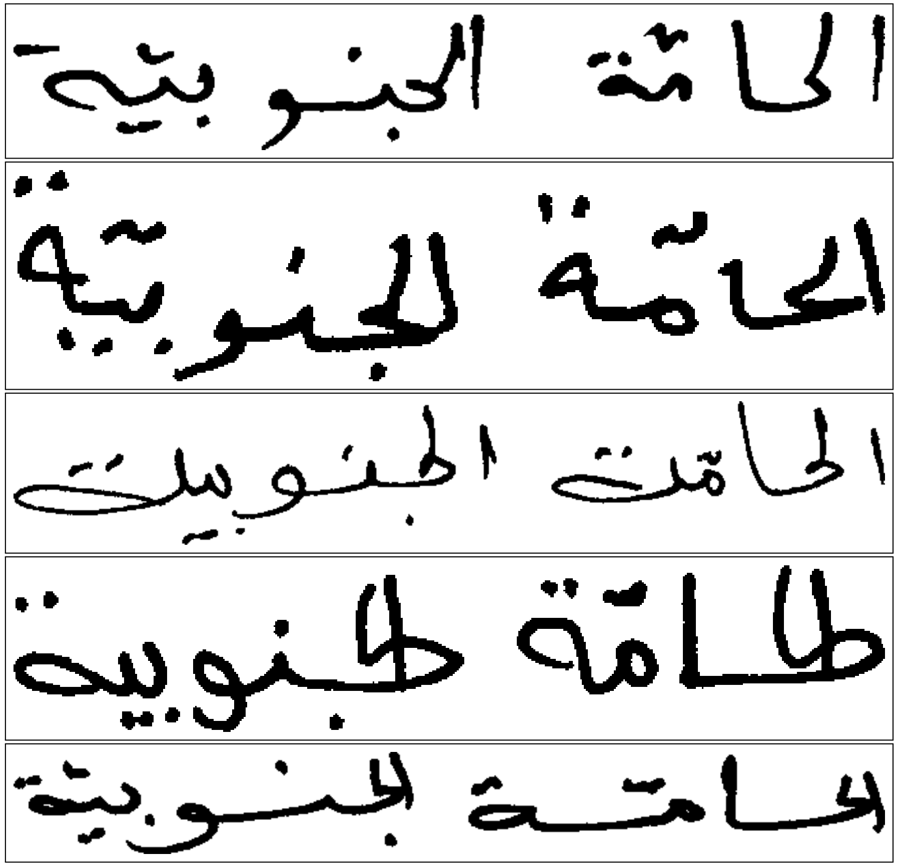
  - **with confidence $c_{r,s,t}$ at sentence-, word, or state-level**

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):



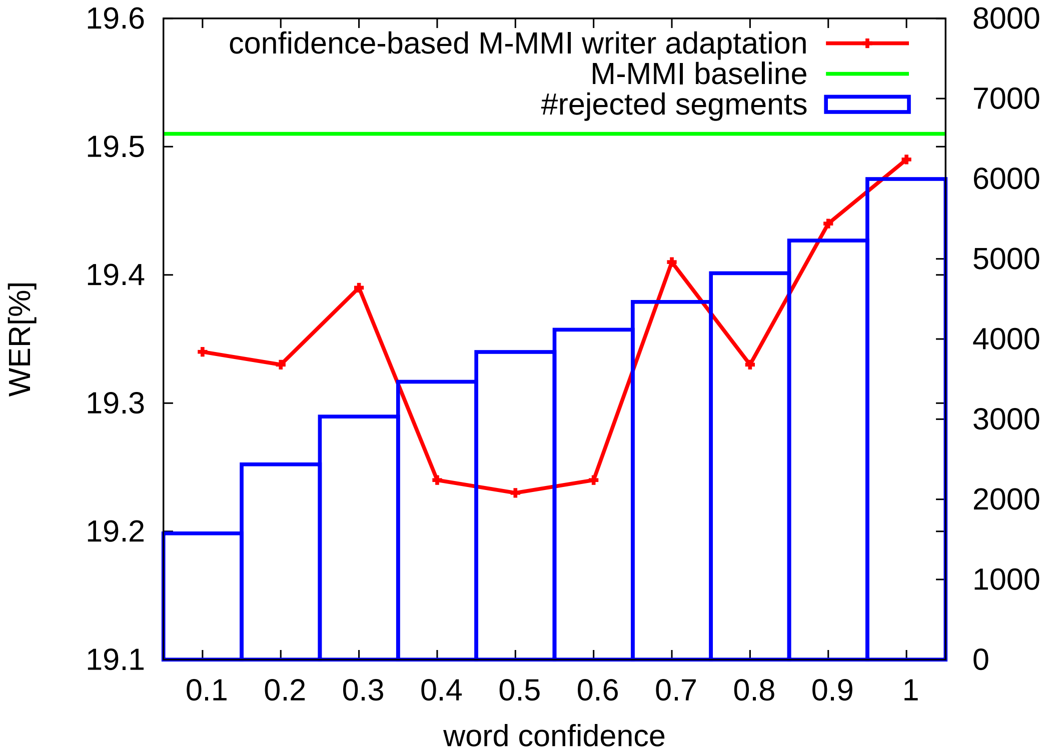
Results - Training: ML vs. MMI vs. Modified-MMI Criterion

- ML = Maximum Likelihood
- MLE = Model Length Estimation
- MMI vs. modified-MMI after 30 Rprop iterations
- ICDAR 2005 Setup [Comparison]

Train	Test	WER [%]			
		ML	+MLE	+MMI	+Modified MMI
abc	d	10.88	7.80	7.44	**6.12**
abd	c	11.50	8.71	8.24	**6.78**
acd	b	10.97	7.84	7.56	**6.08**
bcd	a	12.19	8.66	8.43	**7.02**
abcd	e	21.86	16.82	16.44	**15.35**

Results - Unsupervised Model Adaptation: M-MMI-conf

- M-MMI criterion with posterior confidences (M-MMI-conf)
- unsupervised training for model adaptation during decoding
- word-confidence based M-MMI-conf training and rejections



  - confidence threshold $c = 0.5 \rightarrow$ more than 60% segment rejection rate
  - small amount of adaptation data only

Results - Unsupervised Model Adaptation: M-MMI-conf

- **unsupervised training for model adaptation during decoding**
- **state-confidence based M-MMI-conf training and rejections**
  - **arc posteriors from the lattice output from the decoder**
  - **only word frames aligned with a high confidence in 1st pass**
    → unsupervised model adaptation
  - **only 5% frame rejection rate (20,970 frames of 396,416)**

- **ICDAR 2005 Setup** [Comparison]

Training/Adaptation	WER[%]	CER[%]
ML	21.86	8.11
M-MMI	19.51	7.00
+ unsupervised adaptation	20.11	7.34
+ word-confidences	19.23	7.02
+ state-confidences	**17.75**	**6.49**
+ supervised adaptation	2.06	0.77

Arabic Handwriting - Experimental Results for IFN/ENIT

- **ICDAR 2005 Setup** [Comparison]

Train	Test	WER[%]			
		1st pass			2nd pass
		ML	+MLE	+M-MMI	M-MMI-conf
abc	d	10.88	7.80	**6.12**	**5.95**
abd	c	11.50	8.71	**6.78**	**6.38**
acd	b	10.97	7.84	**6.08**	**5.84**
bcd	a	12.19	8.66	**7.02**	**6.79**
abcd	e	21.86	16.82	**15.35**	**14.55**

Arabic Handwriting - Experimental Results for IFN/ENIT

- evaluation of RWTH-OCR systems at *Arabic HWR Competition*, ICDAR 2009
  - external evaluation at TU Braunschweig, Germany
  - set f and set s are unknown (not available)
  - unsupervised M-MMI-conf model adaptation achieved similar improvements
  - 3rd rank (group)

ID	WRR[%]				
	set f_a	set f_f	set f_g	set f	set s
RWTH-OCR, ID12	86.97	88.08	87.98	85.51	71.33
RWTH-OCR, ID13	87.17	88.63	88.68	85.69	72.54
RWTH-OCR, ID15	86.97	88.08	87.98	83.90	65.99
A2iA, ID8	90.66	91.92	92.31	89.42	76.66
MDLSTM, ID11	94.68	95.65	96.02	93.37	81.06

- **Note:**
  - focus on modeling (ID12 and ID13) and speed (ID15) - no preprocessing

Summary

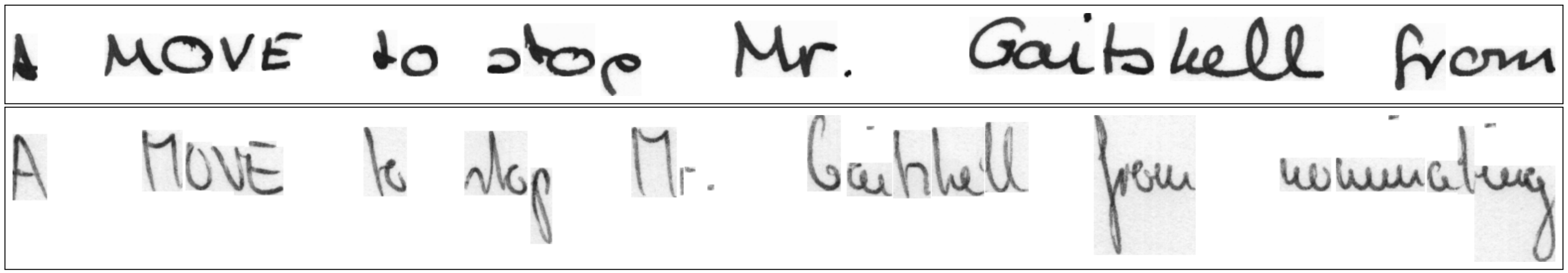
- RWTH-ASR → RWTH-OCR
  - simple feature extraction and preprocessing
  - Arabic: created a SOTA system, ranked 3rd at ICDAR 2009
  - Latin: created a SOTA system, best single system
- discriminative training
  - margin-based HMM training (ML vs. MMI vs. M-MMI)
  - unsupervised confidence-based MMI model adaptation (M-MMI-conf)
  - relative improvements of about 33% w.r.t. ML training
- ongoing work
  - to be evaluated in ASR experiments
  - impact of preprocessing in feature extraction (Arabic vs. Latin)
  - more complex features (e.g. MLP)
  - character context modeling (e.g. CART)
  - further databases/languages

Outlook: Latin Handwriting - IAM Database

- **English handwriting, continuous sentences**

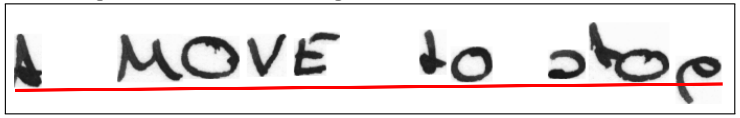
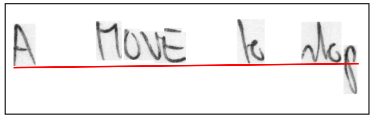
	Train	Devel	Eval 1	Eval 2	Total
Lines	6,161	1,861	900	940	9,862
Running words	53,884	17,720	7,901	8,568	88,073
Vocabulary size	7,754	3,604	2,290	2,290	11,368
Characters	281,744	83,641	41,672	42,990	450,047
Writers	283	128	46	43	500
OOV Rate		≈15%	≈17%	≈15%	

- **Example lines:**

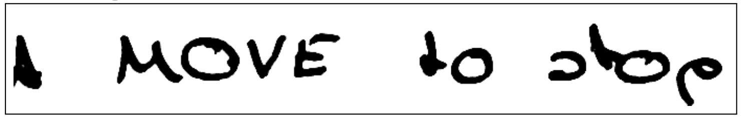
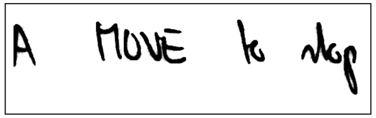


Outlook: Latin Handwriting - UPV Preprocessing

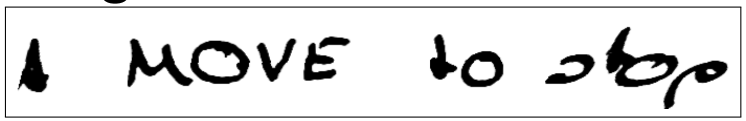
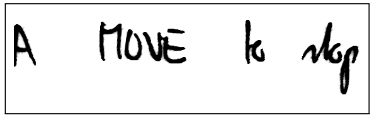
- **Original images**

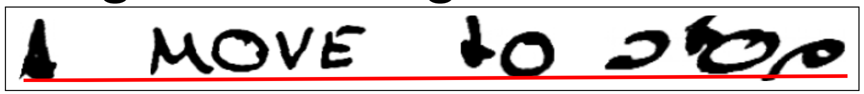
- **Images after color normalisation**

- **Images after slant correction**

- **Images after height normalisation**




Outlook: Latin Handwriting - Experimental Results on IAM Database

Systems	Devel WER [%]	Eval WER [%]
RWTH-OCR*		
Baseline	81.07	83.60
+ UPV Preprocessing	57.59	65.26
+ LBW LM & 20k Lexicon	34.64	41.45
+ discriminative training (M-MMI)	**29.40**	**35.32**
Other Single Systems		
[Bertolami & Bunke 08]	30.98	35.52
[Natarajan & Saleem^{+} 08]	-	*40.01*
[Romero & Alabau^{+} 07]	*30.6*	-
System Combination		
[Bertolami & Bunke 08]	*26.85*	*32.83*

***see [Jonas 09] for details**

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. 21

[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. 33

[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. 21

[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. 21

[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. 21

Appendix: Comparisons for IFN/ENIT

- **ICDAR 2005 Evaluation**

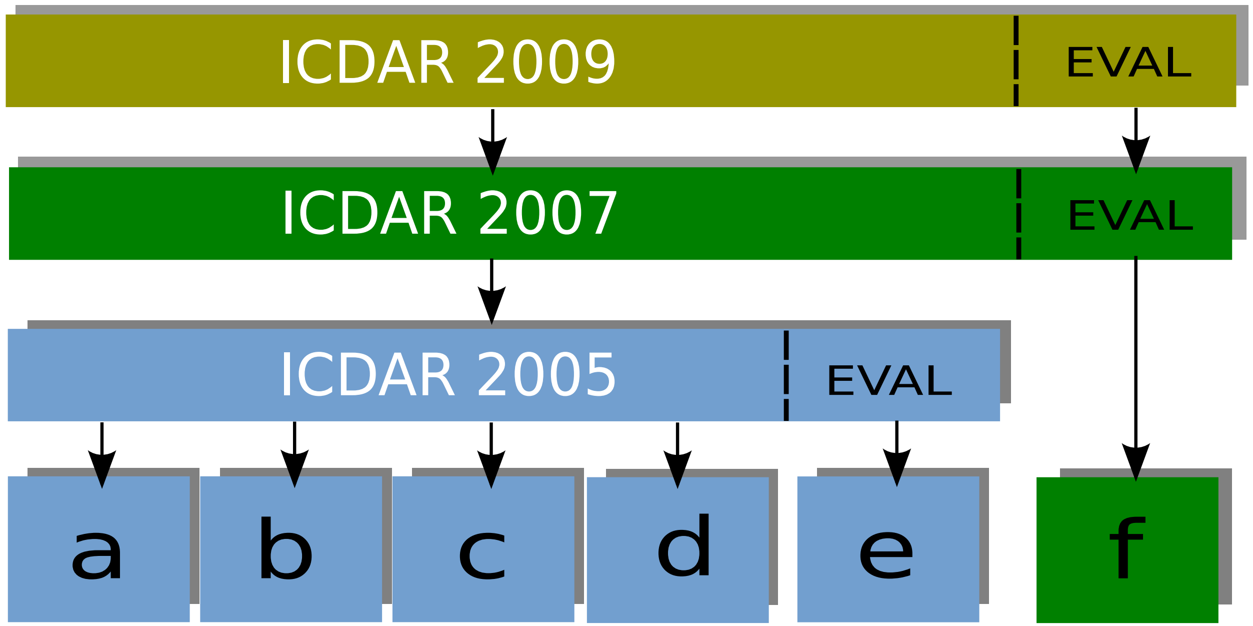
Rank	Group	WRR [%]	
		abc-d	abcd-e
1.	UOB	85.00	75.93
2.	ARAB-IFN	87.94	74.69
3.	ICRA (Microsoft)	88.95	65.74
4.	SHOCRAN	100.00	35.70
5.	TH-OCR	30.13	29.62
	BBN	89.49	N.A.
1*	RWTH	**94.05**	**85.45**

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

Appendix: 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**

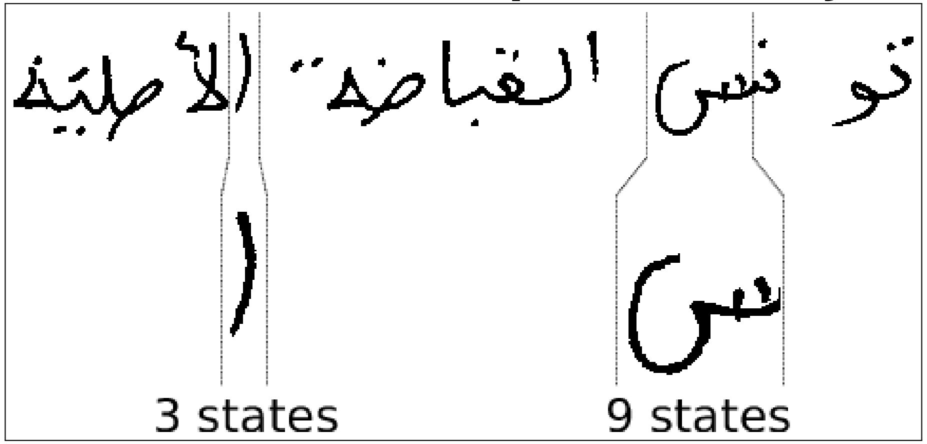


Appendix: Participating Systems at ICDAR 2005 and 2007

- **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

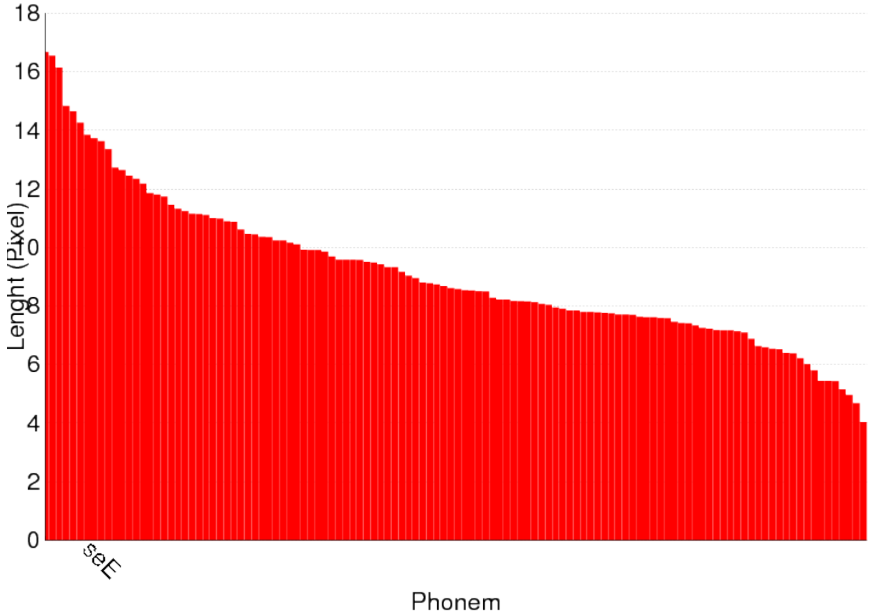
S_c = estimated number states for character c
$N_{x,c}$ = number of observations aligned to character c
N_c = character count of c seen in training
α = character length scaling factor.

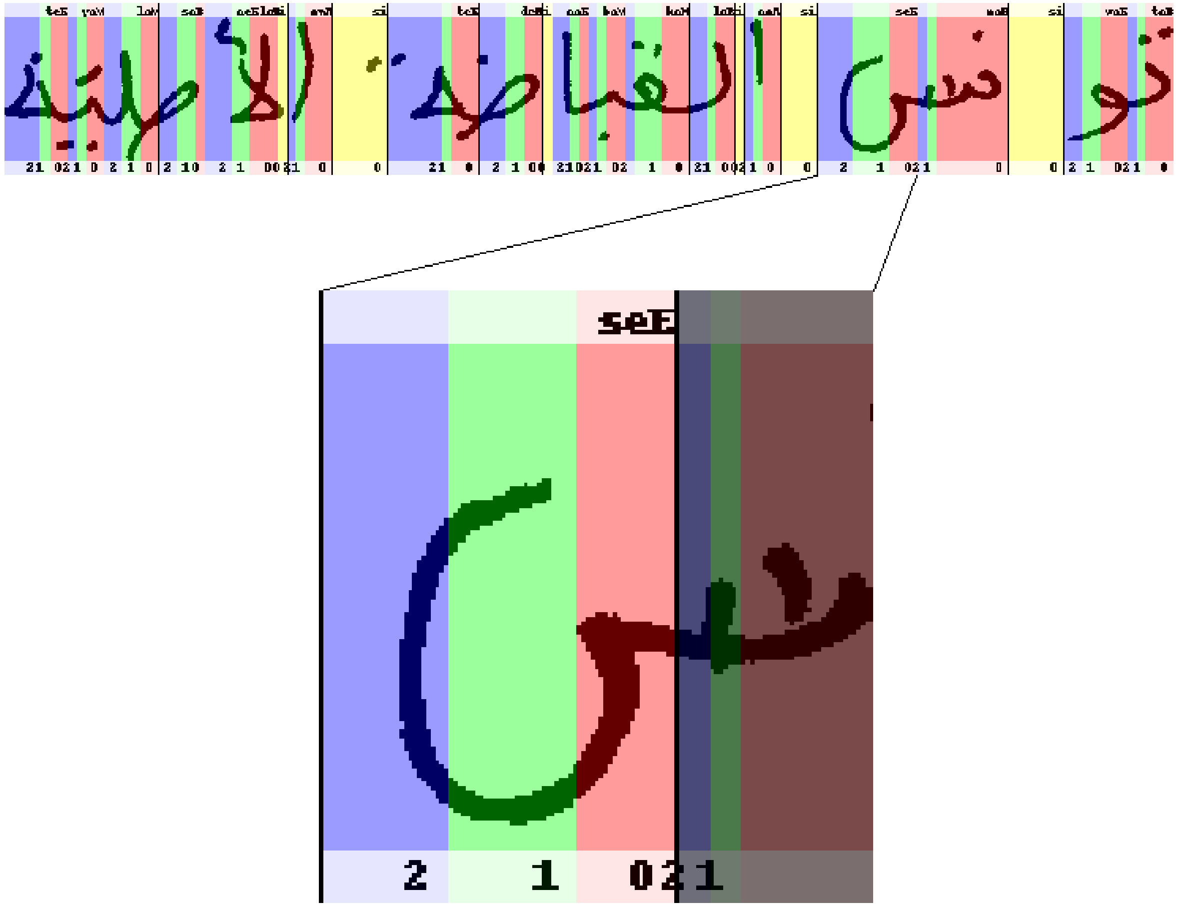
[Visualization]

Appendix: Visual Modeling - Model Length Estimation

Original Length

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

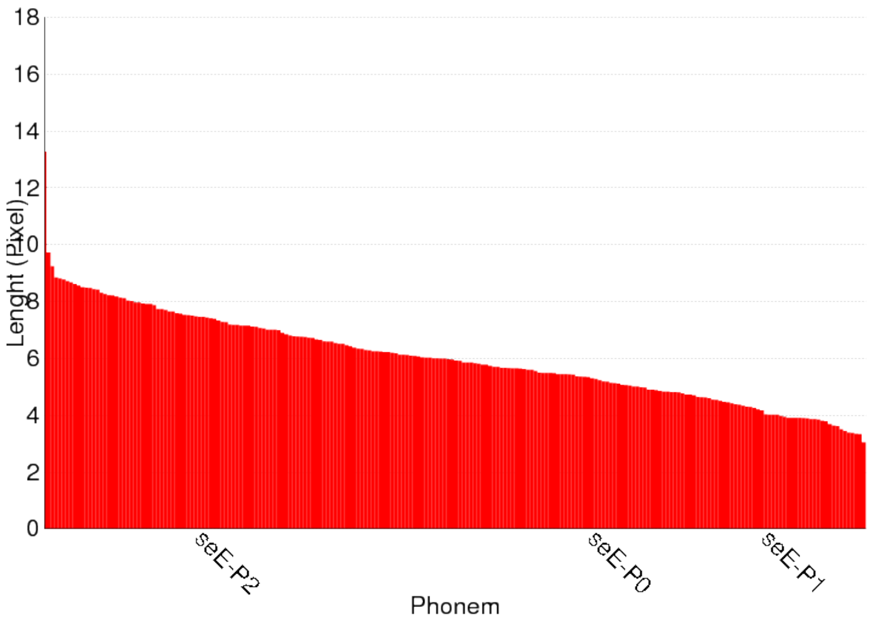


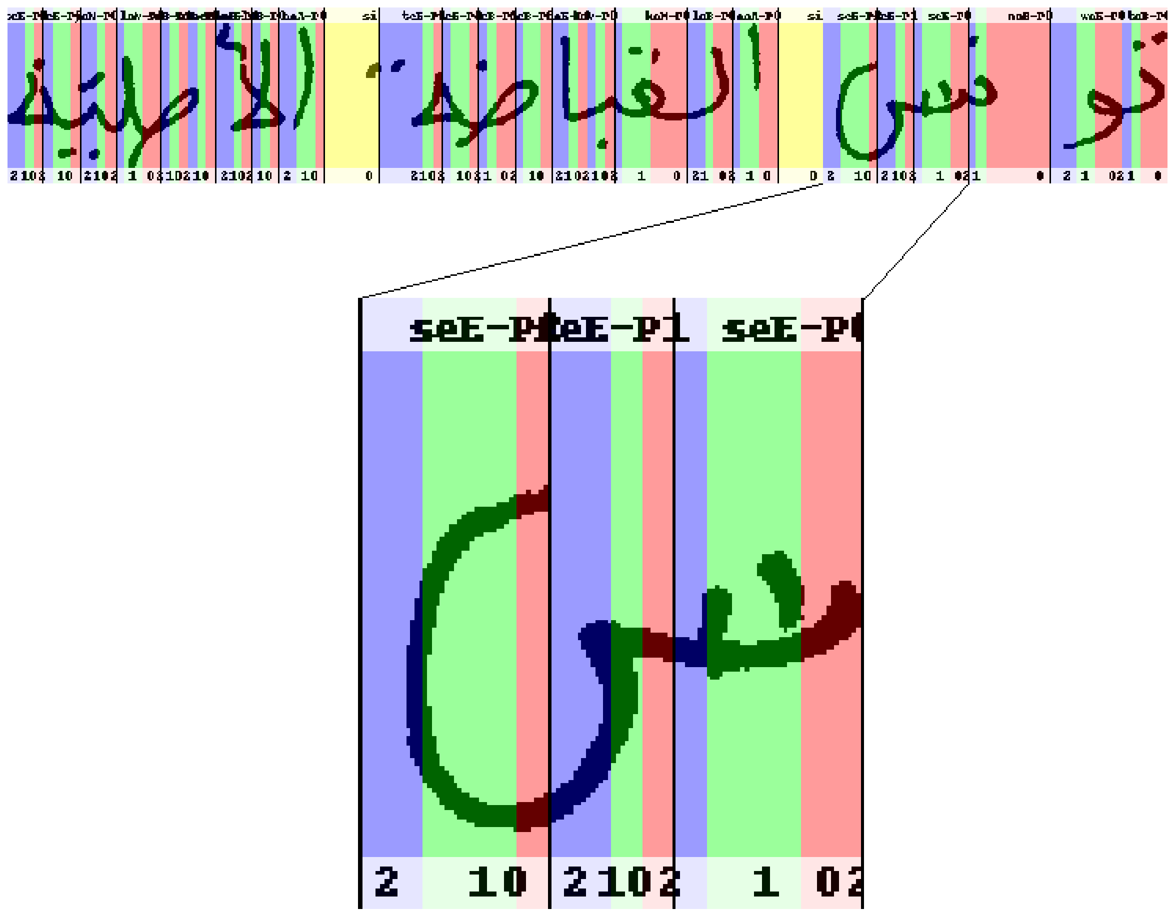


Appendix: Visual Modeling - Model Length Estimation

Estimated Length

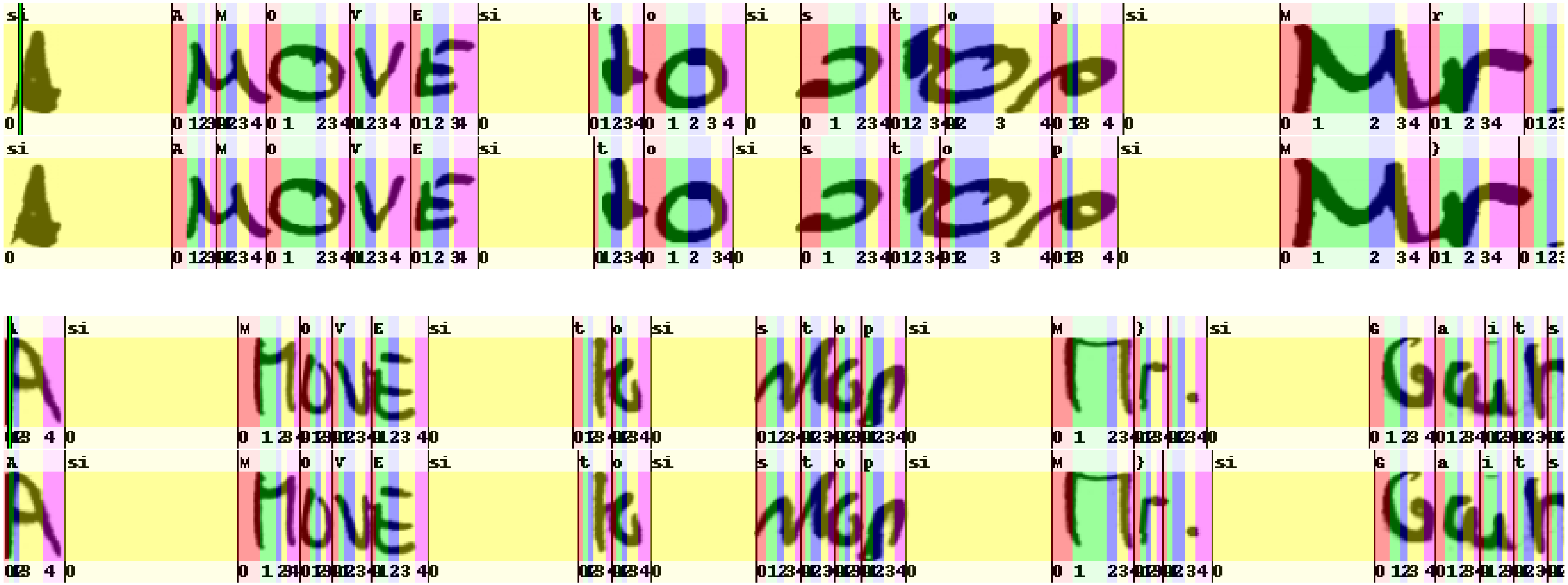
- overall mean of character length = 6.2 pixel (≈ 2.0 pixel/state)
- total #states = 558





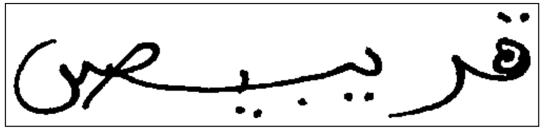
Appendix: Alignment Visualization

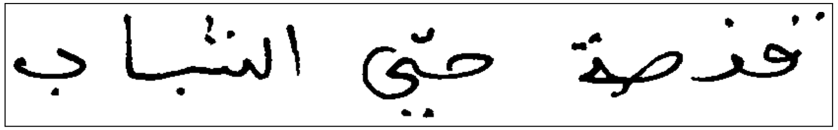
- alignment visualization with and without discriminative training
- upper lines with 5-2 baseline setup, lower lines with additional discriminative training



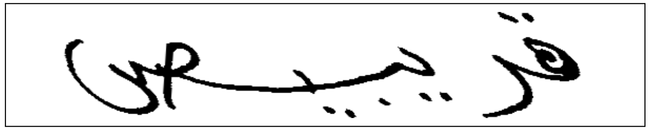
Appendix: Arabic Handwriting - UPV Preprocessing

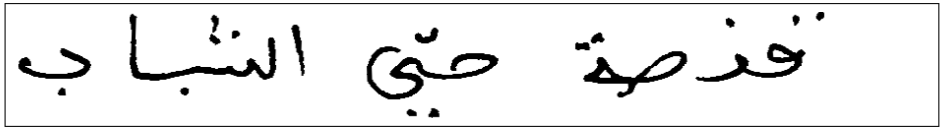
- Original images





- Images after slant correction





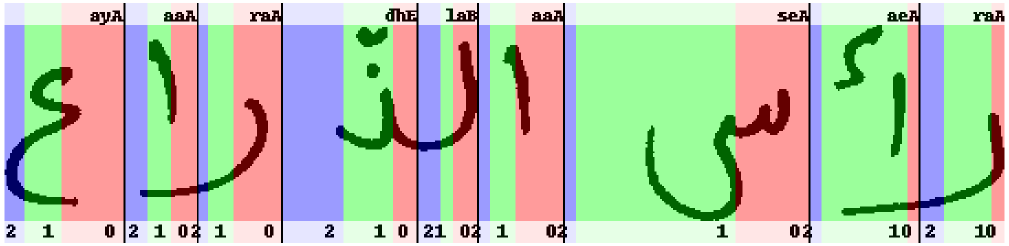
- Images after size normalisation

Experimental Results:

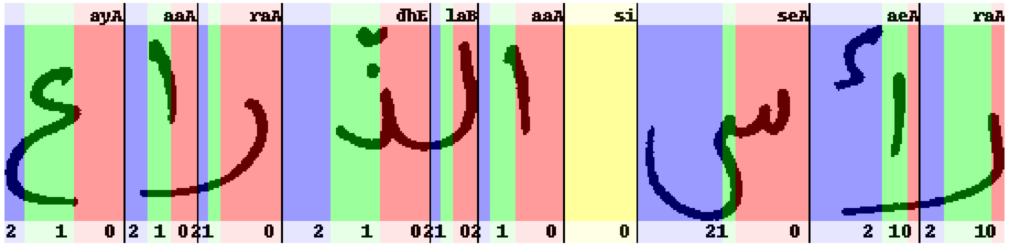
- important informations in ascender and descender areas are lost
- not yet suitable for Arabic HWR

Appendix: Visual Modeling - Writing Variants Lexicon

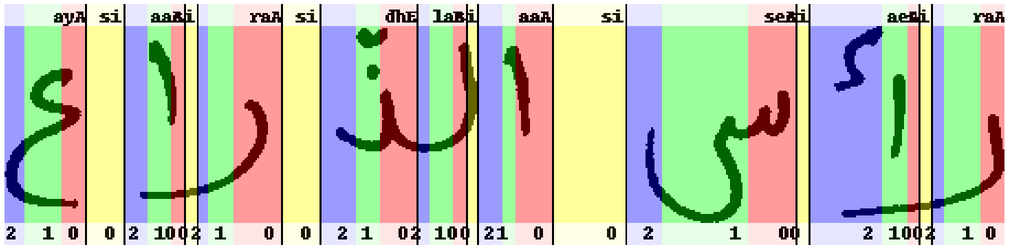
- **most reported error rates are dependent on the number of PAWs**
- **without separate whitespace model**



- **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