Handwritten Text Recognition

M.J. Castro-Bleda, S. España-Boquera, F. Zamora-Martínez

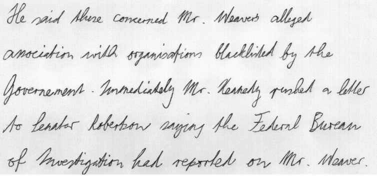
Universidad Politécnica de Valencia
Spain

Avignon, 9 December 2010

The problem: Handwriting recognition

- Handwriting recognition: offline and online handwriting recognition.

An offline handwriting recognition system extracts the information from previously scanned text images



Offline systems are applicable to a wider range of tasks, given that online recognition require the data acquisition to be made with specific equipment.

... whereas online systems receive information captured while the text is being written (stylus and sensitive tablets).

Online systems are more reliable due to the additional information available, such as the order, direction and velocity of the strokes.

The problem: Handwriting recognition

▷ Recognition performance of current automatic offline handwriting transcription systems: far from being perfect.
→ Growing interest in assisted transcription systems, which are more efficient than correcting by hand an automatic transcription.

▷ A recent approach to interactive transcription involves multi-modal recognition, where the user can supply an online transcription of some of the words: STATE system.

▷ Bimodal recognition.

Offline Handwritten Recognition

▷ A preprocessed text line image can be considered a sequence of feature vectors to be generated by a statistical model, as is done in Speech Recognition:

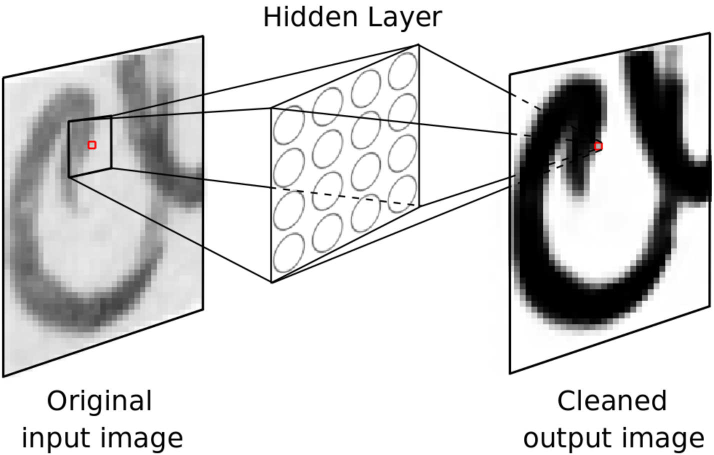
$$\hat{S} = \operatorname*{argmax}_{S \in \Omega^\star} p(S|X) = \operatorname*{argmax}_{S \in \Omega^\star} p(X|S)p(S)\,.$$

▷ **This work** proposes a handwriting recognition system based on

- **MLPs** for preprocessing
- **hybrid HMM/ANN models**, to perform optical character modeling
- **statistical or connectionist** *n*-gram language models: words or characters

Preprocessing

▷ **MLP to enhance and clean images**



Preprocessing

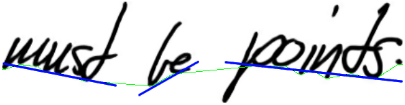
▷ **Slope and slant removal, and size normalization**

Original

Cleaned

Contour

Lower baseline

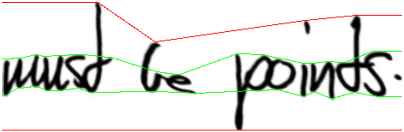


Preprocessing

Desloped

Desloped and deslanted

Reference lines

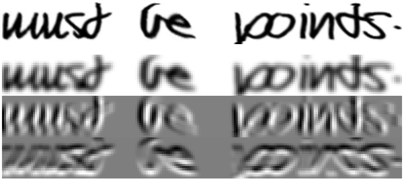


Size normalization

Preprocessing

- **Feature extraction**

Final image



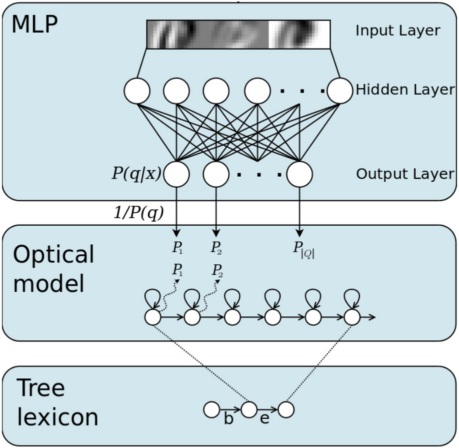
Feature extraction

Frames with 60 features

- grid of 20 square cells
- horizontal and vertical derivatives

Optical models

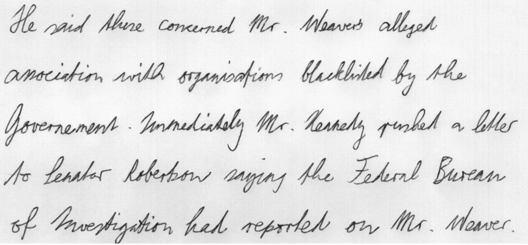
▷ **Hybrid HMM/ANN models:** emission probabilities estimated by ANNs



- A MLP estimates $p(q|x)$ for every state q given the frame x. Emission probability $p(x|q)$ computed with Bayes' theorem.
- Trained with EM algorithm: MLP backpropagation and forced Viterbi alignment of lines are alternated.
- Advantages:
  - each class trained with all training samples
  - not necessary to assume an a priori distribution for the data
  - lower computational cost compared to Gaussian mixtures
- 7-state HMM/ANN using a MLP with two hidden layers of sizes 192 and 128

Corpora for optical modeling

▷ Lines from the IAM Handwriting Database version 3.0



- 657 different writers
- a subset of 6,161 training, 920 validation and 2,781 test lines
- 87,967 instances of 11,320 distinct words (training, validation, and test sets)

Corpora for language modeling

- Three different text corpora: LOB, Brown and Wellington

Corpora	Lines	Words	Chars
LOB + IAM Training	174K	2.3M	11M
Brown	114K	1.1M	12M
Wellington	114K	1.1M	11M
Total	402K	4.5M	34M

Testing the system

▷ Error Rate of the HMMs and the hybrid HMM/ANN models on the test set. Language models estimated with the three corpora and an open dictionary are used.

	Results of Test (%)			
Best model	WER		CER	
8-state HMMs	38.8	±1.0	18.6	±0.6
7-state HMMs, MLP 192-128	22.4	±0.8	9.8	±0.4

Comparing the system

- Comparing is always difficult!!!
- Same conditions (we have contacted the authors).
- Error Rate of the hybrid HMM/ANN models and recurrent networks [Graves et al, 2010] on the test set.

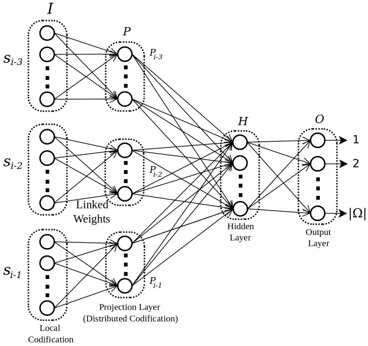
Model	Results of Test (%) WER
7-state HMMs, MLP 192-128	25.9
Recurrent NN (BLSMT)	25.9

The best published performance!!!

Connectionist Language modeling

▷ **SRI language models** smoothed using the modified Kesner-Ney discount.

▷ **Neural Network Language Models**



- linearly combined with standard *n*-grams
- **trained** with stochastic Backpropagation
  - learning rate 0.002, momentum term 0.001, weight decay 10^{-9}
  - cross-entropy error function
  - hidden units ⇨ hyperbolic tangent
  - output layer ⇨ softmax
- fast **evaluation** memoizing softmax normalization constants

Testing the system with NNLM

▷ Error Rate of the hybrid HMM/ANN models on the test set. Language models estimated for a 105 K vocabulary and bigrams (SRI and NNLMs).

	Results of Test (%)	
Language model	WER	CER
SRI bigrams	23.3	9.3
NNLMs	22.6	9.0

Character-based language modeling

- Character-based language models:
  - high order *n*-grams of characters (upt to 8-grams)
  - the language model is able to learn words and sequence of words appearing in the training corpus but also to model words not belonging to the vocabulary,
  - no explicit lexicon is used during recognition: the recognizer is thus able to recognizer out-of-vocabular y words.
- Graphemes for the IAM corpus:

Lower case letters	a b c d e f g h i j k l m n o p q r s t u v w x y z
Upper case letters	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Digits	0 1 2 3 4 5 6 7 8 9
Punctuation marks	<space> - , ; : ! ? / . ' ( ) * & # +

Testing the system with character-based LMs

- **Final results on Test:**

Model	WER (%)	CER (%)
SRI	30.9	13.8
NN LM	24.2	10.1

- **Test OOV word accuracy. 554 OOV words in the test partition:**

Model	# OOV recognized words	% accuracy
SRI	162	29.8
NN LM	184	33.8

Conclusions

▷ HMM/ANN: Performance competitive with state-of-the-art systems.

Improving Offline Handwritten Text Recognition with Hybrid HMM/ANN Models (2010), in: IEEE Trans. PAMI

▷ NN LMs advantages:

- they are very scalable with respect to the corpus size: the size of the trained language model grows with the vocabulary size but not with the number of training samples,
- NN LM represents the tokens in a continuous space, thus allowing a better smoothing as can be observed when comparing SRI and NN LM *n*-grams models using the same optical models.

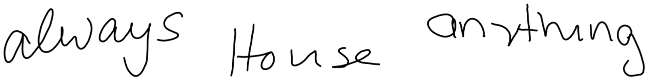
Fast Evaluation of Connectionist Language Models, in: 10th IWANN, p. 33-40, Springer, 2009.

▷ Character language models can alleviate the problem of OOV words.

Unconstrained Offline Handwriting Recognition using Connectionist Character N-grams, in: IEEE IJCNN, p. 4136–4142, 2010.

Online and Bimodal Handwritten Recognition

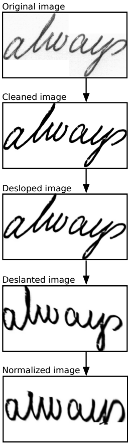
▷ Online samples are sequences of coordinates describing the trajectory of an electronic pen (more information than the offline case).

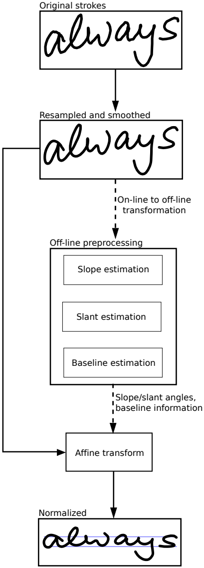


▷ Hybrid HMM/ANN optical models for online and offline recognition.

▷ Isolated word recognition.

▷ Bimodal recognition. Core idea: N -best word hypothesis scores for both the offline and the online samples are combined using a log-linear combination, achieving very satisfying results.





Offline preprocessing Online preprocessing

Optical models: HMM/ANN

▷ **On-line** HMM/ANN configuration:

- Same HMMs topologies and MLP, but
- MLP input wider context: 12 feature frames at both sides
- Models trained with the training partition of the IAM-online DB

Bimodal system

1. Scores of the 100 most probable word hypothesis for the offline sample using the offline preprocessing and HMM/ANN optical models.
2. Same process applied to the online sample.
3. The final score for each bimodal sample is computed from these lists by means of a log-linear combination of the scores computed by both the offline and online HMM/ANN classifiers:

$$\hat{c} = \underset{1 \leq c \leq C}{\operatorname{argmax}} ((1-\alpha) \log P(x_{\text{off-line}}|c) + \alpha \log P(x_{\text{on-line}}|c))$$

4. Combination coefficient estimated over the validation set.

Experimental results

- Word Error Rate:

		Unimodal		Bimodal	
System		Off.	On.	Combination	Relative improv.
Validation	Baseline	27.6	6.6	4.0	39%
	HMM/ANN	12.7	2.9	1.9	34%
(Hidden) Test	HMM/ANN	12.7	3.7	1.5	59%

- Performance of the bimodal recognition engine: close to 60% of improvement is achieved with the bimodal system when compared to using only the online system for the test set.

Conclusions

▷ Perfect transcription for most handwriting tasks cannot be achieved: human intervention needed to correct it $\rightarrow$ Assisted transcription systems aim to minimize human correction effort.

▷ Integration of online input into the offline transcription system can help in this process (STATE system).

▷ Hybrid HMM/ANN optical models perform very well for both offline and online data, and their naive combination is able to greatly outperform each system.

▷ More exhaustive experimentation is needed, with a larger corpus, in order to obtain more representative conclusions.

Hybrid HMM/ANN models for bimodal online and offline cursive word recognition, in: ICPR 2010, IEEE, 2010.