The UOB-Télécom ParisTech Arabic handwriting recognition and translation systems for the OpenHaRT 2013 competition

Olivier Morillot[a], Cristina Oprean[a], Laurence Likforman-Sulem[a], Chafic Mokbel[b], Edgard Chammas[b] and Emmanuèle Grosicki[c]

[a]Télécom ParisTech and CNRS LTCI, Paris, France
[b]University of Balamand,Tripoli, Lebanon
[c]DGA,Paris, France

August 23, 2013

NIST Open Handwriting Recognition and Translation Evaluation & Workshop (OpenHaRT 2013)

Outline

Outline

Team

Two research groups

- Télécom ParisTech (Institut Mines-Télécom): Laurence Likforman-Sulem, Cristina Oprean, Olivier Morillot and Emmanuèle Grosicki
- University of Balamand: Chafic Mokbel and Edgar Chammas

Submitted systems

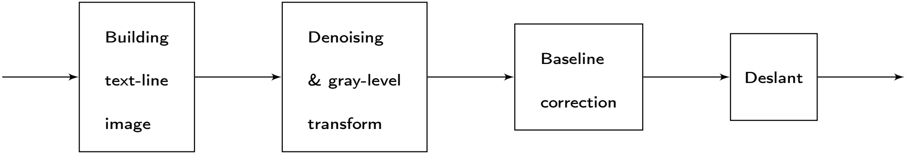
- Document Image Recognition (DIR): 4 systems submitted
  - 1 BLSTM recognizer (primary: p-blstm)
  - 3 HMM recognizers (contrastive: c-baseline, c-contextualhmm, c-hmm)
- Document Text Translation (DTT):
  - 2 systems using MOSES toolkit (constrained: p-baseline_1, unconstrained: p-baseline_2)
- Document Image Translation (DIT):
  - 1 system using HMMs and MOSES toolkit (p-baseline_1)

- Text line approach

Outline

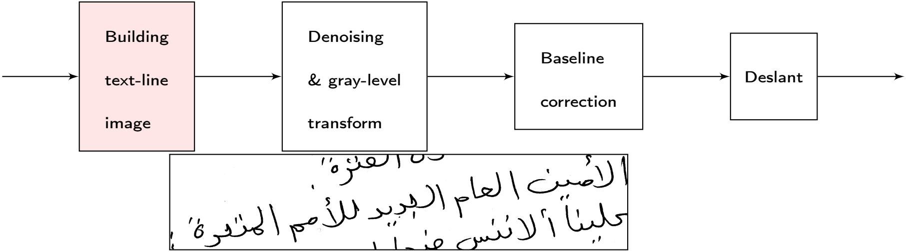
Preprocessing steps

Segmentation-free text-line approach



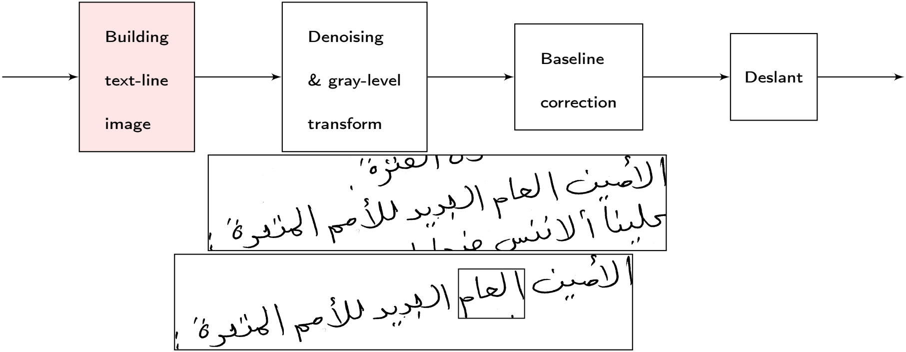
Preprocessing steps

Segmentation-free text-line approach



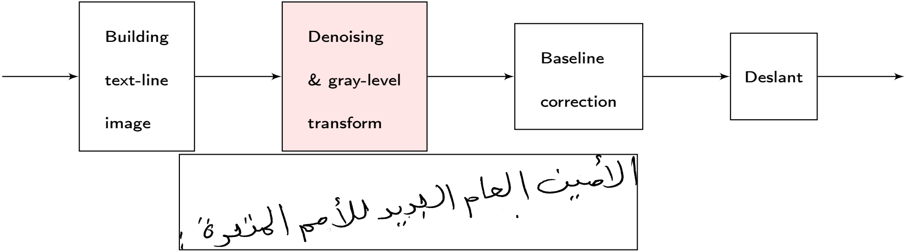
Preprocessing steps

Segmentation-free text-line approach



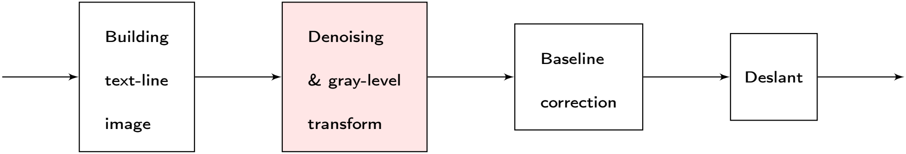
Preprocessing steps

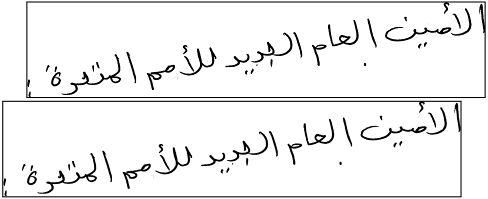
Segmentation-free text-line approach



Preprocessing steps

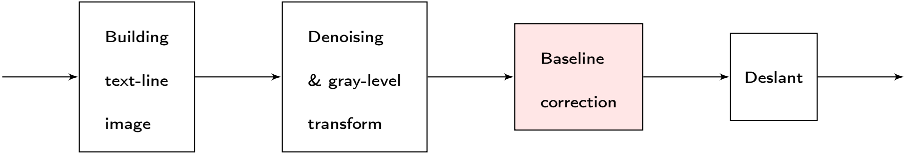
Segmentation-free text-line approach

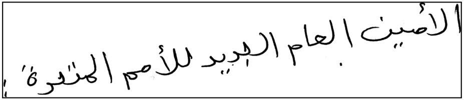




Preprocessing steps

Segmentation-free text-line approach

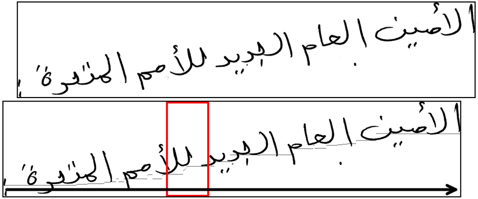




Preprocessing steps

Segmentation-free text-line approach

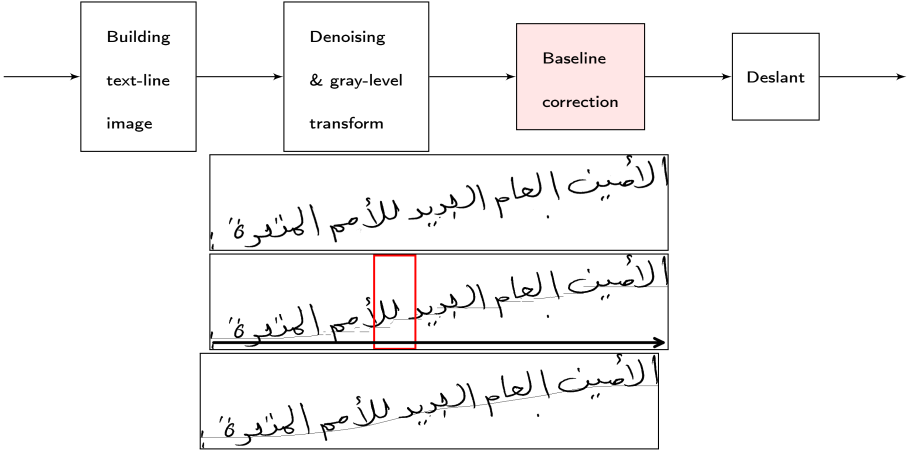
→ Building text-line image → Denoising & gray-level transform → **Baseline correction** → Deslant →



[Morillot et al., A new baseline correction algorithm for text-line recognition with bidirectional recurrent neural networks, Journal of Electronic Imaging, 2013]

Preprocessing steps

Segmentation-free text-line approach

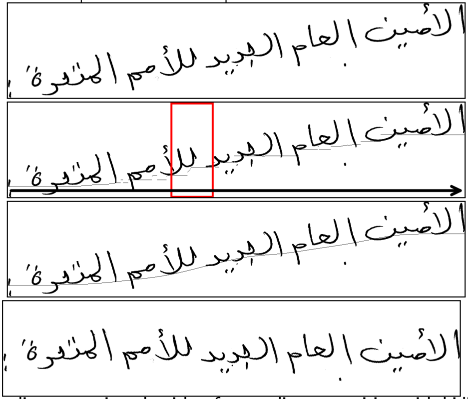


[Morillot et al., A new baseline correction algorithm for text-line recognition with bidirectional recurrent neural networks, Journal of Electronic Imaging, 2013]

Preprocessing steps

Segmentation-free text-line approach

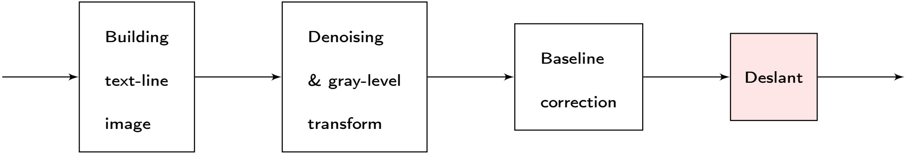
Building text-line image → Denoising & gray-level transform → Baseline correction → Deslant

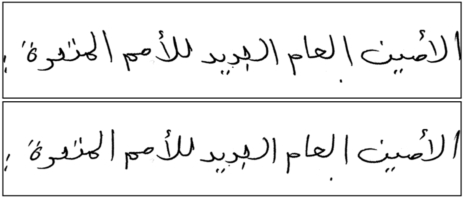


[Morillot et al., A new baseline correction algorithm for text-line recognition with bidirectional recurrent neural networks, Journal of Electronic Imaging, 2013]

Preprocessing steps

Segmentation-free text-line approach





Feature extraction

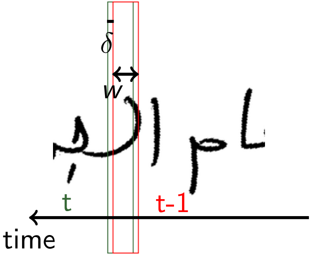


Figure : Overlapping sliding window extraction

- Extraction window width $w = 9$, shift $\delta = 2$
- 37 features: statistical and geometrical

  [Al-Hajj-Mohamad et al., 2009, Oprean et al., 2013]

  - 2 features representing background/foreground transitions,
  - 12 features: concavity configurations,
  - 3 features: gravity center position,
  - 3 features: density of pixels, above and below the baselines,
  - 8 directional features: histogram of gradients for 8 orientations.
- Derivative features -> 74 features

Outline

1 Working team and submitted systems

2 Preprocessing and feature extraction

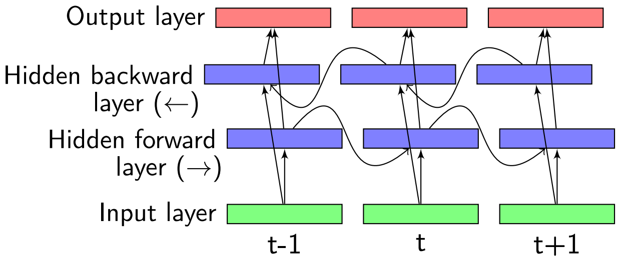
3 Recognition task (DIR)
- BLSTM system
- HMM system
- Language modeling
- Performances

4 Translation tasks (DTT & DIT)

5 Conclusion

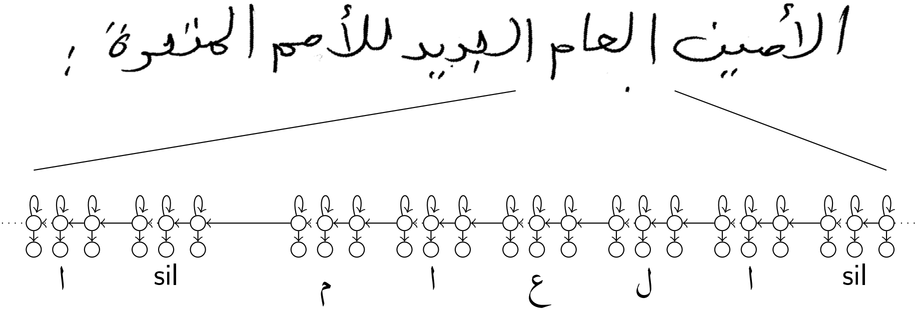
BLSTM system

- Bidirectional recurrent network:



- LSTM blocks [Hochreiter and Schmidhuber, 1997]
- 1 Hidden layer for each direction: 100 blocks per layer
- BLSTM implementation introduced in the works of [Graves et al., 2009]

HMM system



- Bakis topology: no transition from a state to a previous state
- Context-dependent and context-independant modeling of characters
- 5 states for small size letters and punctuations
- 8 states were used for larger letters
- Observation probability density function: mixture of 32 Gaussians
- HTK library

Language modeling

- Bigram model with back-off strategy
- Based on training database text-lines' transcriptions
- SRILM library

DIR experimental conditions

System id	Method	Training data	Dic/LM size
p-blstm	**BLSTM**	**11%**	**22,000 words**
c-hmm	HMM	11%	30,000 words
c-contextualhmm	contextual HMM	11%	30,000 words
c-baseline	contextual HMM	all words and 3.5% of line images	33,000 words

11% of training data: mix between different document categories (AAW, AHR, arb, ...)

DIR results

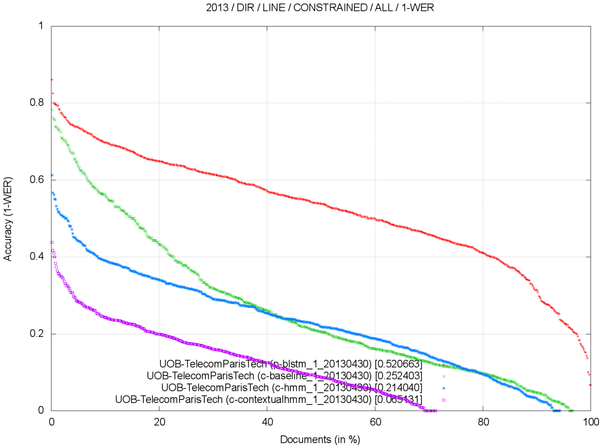


Figure : UOB-TelecomParisTech DIR results

Outline

Machine translation: framework

- f word string in the foreign language (here Arabic)
- e word string in the native language (here English)
- $e*$ set of word strings in the native language
- Goal: Find the most probable native language sentence e provided the observed foreign sentence f
- $\widetilde{e} = \arg\max_{e \in e^*} P(e|f) = \arg\max_{e \in e^*} P(f|e)P(e)$
- $P(e)$ usually known as the statistical language model in the native language.
- Defining $P(f|e)$ not straight-forward

Machine translation: our systems

- Systems built with MOSES toolkit
- Constrained systems using training data provided by the NIST:
  - DTT system trained on actual transcriptions of sentences
  - DIT system trained on actual transcriptions of sentences
- Unconstrained DTT system based on a previous model:
  - Trained on LDC "Arabic News Translation Text Part 1" database (LDC 2004T17)
  - 440,000 arabic words (vocabulary: 25,000 words)
  - Newspaper articles: AFP, Xinhua and An Nahar

DTT & DIT experimental conditions

System id	Training data	Dic/LM size
DTT p-baseline-1	100%	50,000
DIT p-baseline-1	100%	50,000

DTT results

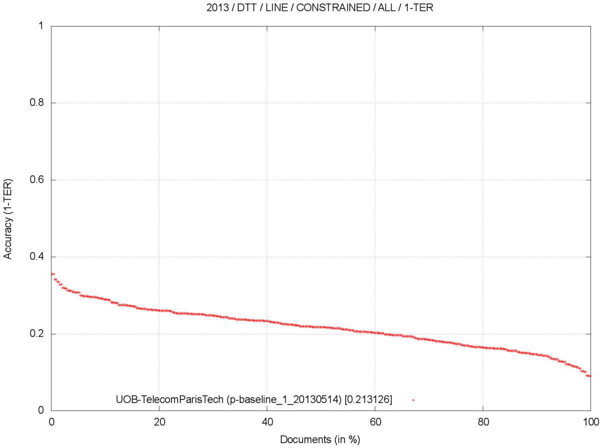


Figure : UOB-TelecomParisTech DTT results

DIT results

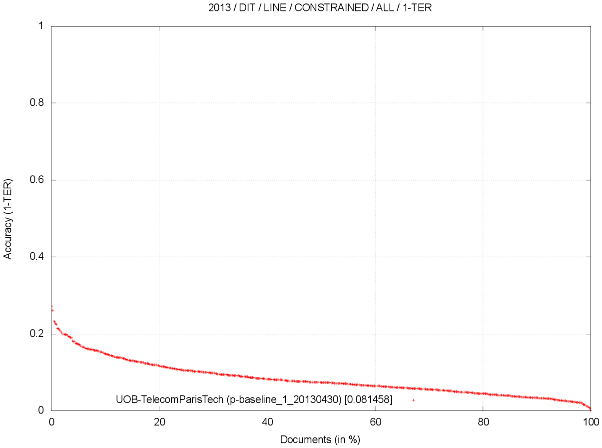


Figure : UOB-TelecomParisTech DIT results

Outline

Conclusion

- Best recognition result with BLSTM: $1 - \mathit{WER} = 52\%$
- Only 11% of the available data used for training (16,000 out of 145,000 text-lines)
- Text-line approach

Perspectives

- Use a larger part of training data
- Combine recognizers
- Using the recognizer's outputs for DIT training
- Factored language models to improve translation

Thank you !

References

Al-Hajj-Mohamad, R., Likforman-Sulem, L., and Mokbel, C. (2009).
Combining slanted-frame classifiers for improved HMM-based arabic handwriting recognition.
Pattern Analysis and Machine Intelligence, IEEE Transactions on, 31(7):1165–1177.

Graves, A., Liwicki, M., Fernandez, S., Bertolami, R., Bunke, H., and Schmidhuber, J. (2009).
A novel connectionist system for unconstrained handwriting recognition.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(5).

Hochreiter, S. and Schmidhuber, J. (1997).
Long short-term memory.
Neural Comput., 9(8):1735–1780.

Morillot, O., Likforman-Sulem, L., and Grosicki, E. (2013).
New baseline correction algorithm for text-line recognition with bidirectional recurrent neural networks.
Journal of Electronic Imaging, 22(2):023028–023028.

Oprean, C., Likforman-Sulem, L., and Mokbel, C. (2013).
Handwritten word preprocessing for database adaptation.
pages 865808–865808–9.