cse@buffalo

# Phrase based direct model for improving handwriting recognition accuracies

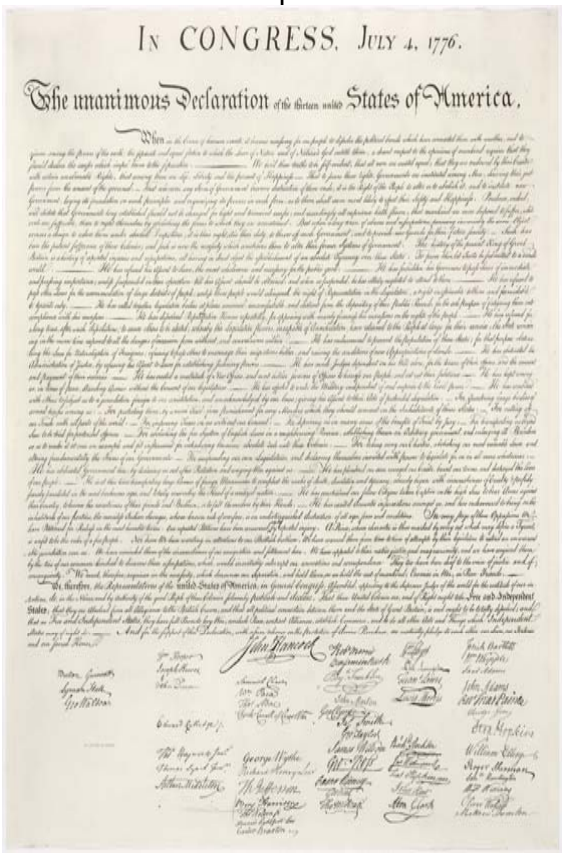
Damien Jose

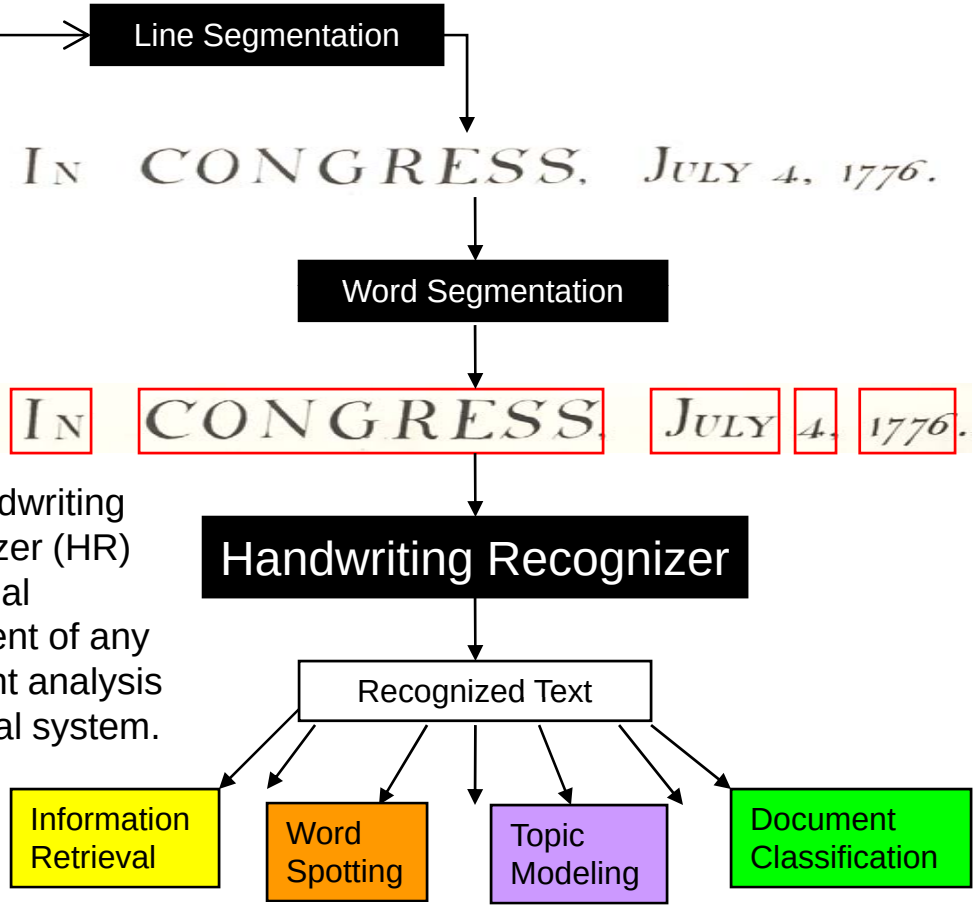
dsjose@cubs.buffalo.edu

# Agenda

- Importance of improving handwritten word recognition accuracy
- Phrase based direct model approach to improve accuracy
- Experiments
- Results

# Typical Documentary Analysis and Recognition System




The Handwriting Recognizer (HR) is a crucial component of any document analysis & retrieval system.

# Motivation

- Component systems often developed independently by different groups.
- Internals of one component not accessible to the developers of the next component in the pipeline.
- These components (e.g. HR) are treated as black boxes where only their output is observed.
- Output of these systems is error-prone.
- Word recognition is definitely a bottleneck.

| | writing style<br># of images | discrete<br>6 | cursive<br>5 | mixed<br>9 | total<br>20 |
|---|---|---|---|---|---|
| Line Separation | # of lines | 118 | 101 | 210 | 429 |
| | # of lines separated | 114<br>96.6% | 97<br>96.0% | 198<br>94.3% | 409<br>95.3% |
| Word Segmentation | # of words | 641 | 692 | 1,427 | 2,760 |
| | # of words segmented | 597<br>93.1% | 631<br>91.2% | 1,379<br>96.6% | 2,607<br>94.5% |
| Word Recognition | top 1 | 432<br>72.4% | 268<br>42.5% | 750<br>54.4% | 1,450<br>55.6% |
| | top 10 | 541<br>90.6% | 459<br>72.7% | 1.081<br>78.4% | 2,081<br>79.8% |

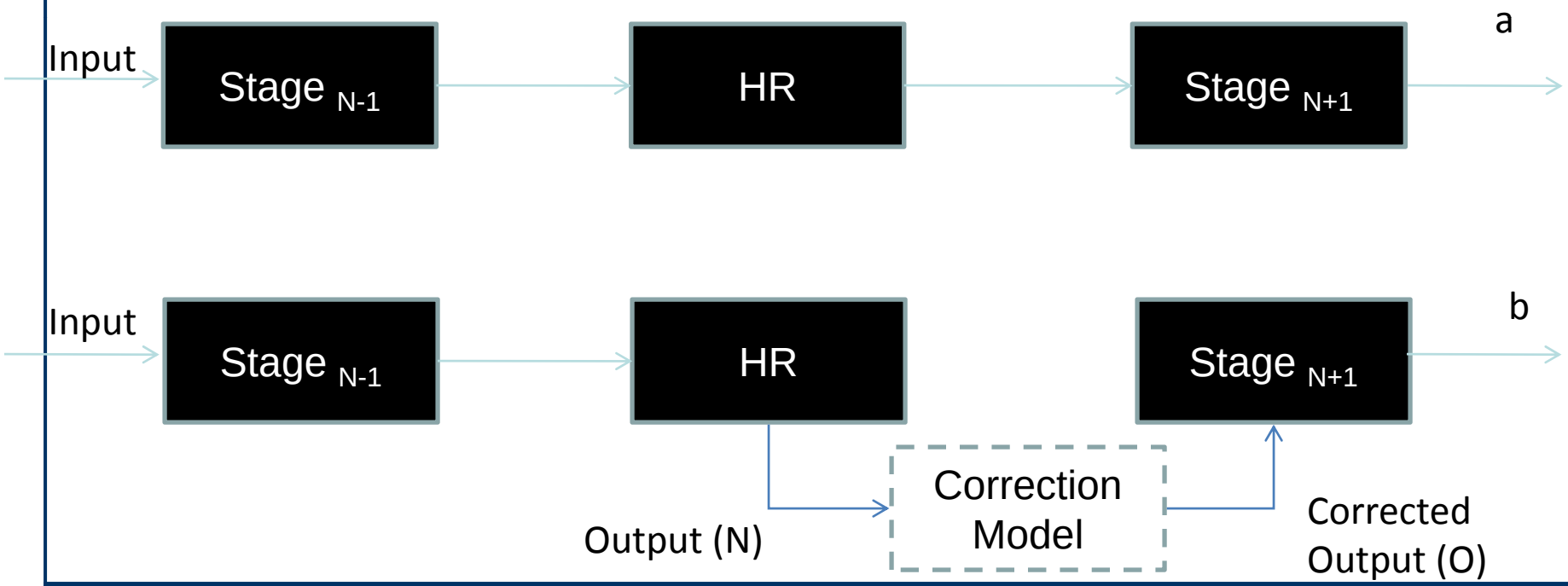
**Performance of line separation, word segmentation and word recognition on 20 document images of different writing styles.**

# Drawbacks of existing approaches in OCR post-processing

- Thus improving the performance of the recognizer will enrich the overall user experience.
  - Jones et al. [1] describe a multi-pass OCR post-processing system which carries out individual word corrections, combined edit distance corrections and bigram probability based correction in different passes.
  - Perez-Cortes et al. [2] use a stochastic finite state machine to test hypothesis of words. If the machine accepts the word, then no correction is made, otherwise smallest set of transitions that could not be traversed show the most similar string in the model.
  - Pal et al. [3] describe a method for OCR error correction of Devanagiri script using morphological parsing.
- Problems with these approaches include
  - Using features that are language dependent.
  - Application on machine print OCR that are conventionally “character-models” as opposed to HR systems that follow a word-based “multiple choice” paradigm.
  - Training the character confusion matrices is not straight forward.

# Proposed Approach

- Analogous to SMT the problem is viewed as a “direct phrase-based translation” task.
- HR output can be visualized as a noisy black-box through which the signal (truth) when passed gets corrupted and emerges out as the degraded output.
- We hope to model the inherent noise of the OCR and try to create an invertible transform to regenerate the truth from corrupt output



# Correction Model

Given sentence pairs in the source (Foreign/Corrupt) and the target (English/Truth) languages

- Align words in the source and target sentences (for e.g. using Levenshtein distance)
- Extract phrase pairs.
- Combine noise model with a n-gram language model to translate the source language into target language.

Given:

Target – Truth, Source - OCR output

P(*tgt,src*)

$$\hat{e} = \arg\max_{e} \left[ -\omega_{ph} \times \log_{10} P(src \mid tgt) - \omega_{lm} \times \log_{10} P(tgt) \right]$$

where:

e - current hypotheses,

$\hat{e}$ - extended hypotheses,

$w_{ph}$ - Phrase model weight,

$w_{lm}$ - Language model weight,

P(tgt) – Tri-gram Language model trained on Reuters data

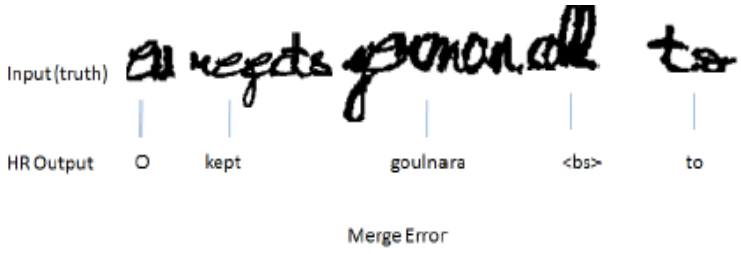
P(src|tgt) – Phrase Model trained on Conference on Computational Natural Language Learning 2003 data

# Steps Involved

- Handwritten words are generated from CONNL English text by concatenating character templates generated by the Blums MAT, followed by character auto-scaling, automatic baseline determination, ligature modeling, ligature joining, skeleton thickening and smoothing [4].

- In-house HR used for recognition is a lexicon driven HMM based word model recognizer. Alignments between input and output done using Levenshtein edit distance.

- Data is split into a training (75%) and test set (25%). Training and testing was done with a closed lexicon. 5% OOV's were present in the test set.





| | Documents | Lines | Words |
|---|---|---|---|
| Training | 1055 | 7878 | 127690 |
| Test | 340 | 2540 | 40789 |
| Total | 1395 | 10418 | 168479 |

# Phrase Model

Probability

| Recognized words | Hidden words | | | |
|---|---|---|---|---|
| pitcher | pitcher 1.00 | | | |
| pascolo | financially 0.40 | pascolo 0.20 | speculates 0.20 | poisonous 0.20 |
| notation | protection 0.50 | invitation 0.40 | motivation 0.05 | notation 0.05 |
| experts | experts 0.88 | expired 0.13 | | |
| updated | injunction 0.40 | uprooted 0.20 | infrastructural 0.40 | |

# Viterbi Decoding – Combining the Phrase and Language models

- To correct the OCR output for a given test sentence, we "translate" the sentence by decoding using two weighted components - the phrases obtained above and the language model.
- Formally, the final decoding e for the source f is the one that satisfies the following equation:

$$\hat{e} = \underset{e}{argmax} \left[ -w_{ph} \times log_{10} P(e|f) \right. \\ \left. -w_{lm} \times log_{10} P(e) \right]$$

  - Where P(e) is the trigram character language model probability and P(e|f) is the phrase-based direct model.
  - Weights $w_{ph}$ and $w_{lm}$ were chosen as ($w_{ph} + w_{lm} = 1$) for both mixture components.

- Whenever a test word is not found in the training model we utilize the top-10 unigram outputs from the word recognizer for that word image.

# Result of Viterbi decoding using the Phrase Model and Language Model:

Log probability

| Recognized words | Hidden words | | | | |
|---|---|---|---|---|---|
| om | the 0 | | | | |
| official | officials 7.89912 | official 6.79544 | | | |
| sort | said 15.2251 | | | | |
| om | the 24.5661 | | | | |
| accord | amanda 54.0033 | attackers 50.6336 | antara 57.4298 | | |
| had | had 86.9505 | batter 99.1431 | stated 97.6456 | | |
| Gerg | seized 143.344 | leaving 152.173 | tsang 152.516 | freeze 156.673 | |
| tra | two 239.679 | | | | |
| Roberta | kalashnikov 389.403 | nagatsuka 401.44 | | | |
| dealt | assault 639.717 | decades 649.419 | consistent 650.351 | nationwide 650.707 | |
| aples | wales 1043.42 | maybe 1043.54 | rifles 1036.11 | engineers 1046.72 | |
| ad | and 1679.26 | cash 1691.12 | seed 1691.66 | | |
| disappears | disappointed 2728.21 | disappeared 2721.85 | disappears 2731.83 | | |

**Decoded Sentence :**

**the official said the attackers had seized two kalashnikov assault rifles and disappeared**

# Results

- Raw, with LM and Noise corrected accuracies of the recognizer on the test set before and after the correction.

| | Raw | Top-10+3-gram LM | Corrected |
|---|---|---|---|
| Accuracy | 8.1% | 13.5% | 71.3% |

- We observe that there is a considerable increase in the accuracy after the “Noise” correction.

Advantages

- This technique is adaptable to other recognizers and even other scripts where training data is available.
- Fast Decoding - The Viterbi sentence decoding matrix shows the correction model options for the “observed” output from the recognizer with the corresponding probabilities.
- Models errors in the phrase context

Disadvantage

- Possibly over fitting  on synthetic data

# References


1. L. Bahl, F. Jelinek and R. Mercer, ”A Maximum Likelihood Approach to Continuous Speech Recognition”, IEEE Transactions on PAMI, 5(2):179–190, 1983.

2. H. Blum, ”A Transformation for Extracting New Descriptors of Shape”, Models for the perception of Speech and Visual Form, MIT Press, 1967, pp 362–380, Cambridge, MA.

3. A. Ittycheriah and S. Roukos, ”A Maximum Entropy Word Aligner for Arabic-English Machine Translation”, Proceedings of the Human Language Technology Conference (HLT-NAACL), 2005, Vancouver, Canada.

4. M. Jones, G. Story and B. Ballard, ”Integrating multiple knowledge sources in a a Bayesian OCR postprocessor”, International Conference on Document Analysis and Recognition, 1991, pp 925–933, St. Malo, France.

5. G. Kim, V. Govindaraju and S. Srihari, ”Architecture for handwriting recognition systems”, International Journal of Document Analysis and Recognition, 2(1):37–44, 1999.

# Thank You