Typographical Features for Scene Text Recognition

Jerod Weinman

Department of Computer Science
Grinnell College
Scene Text: What Challenges?

- Small sample (< 50 chars)
- Out-of-plane distortion
- Uneven lighting
- Unusual fonts
- Word segmentation
- Character segmentation
What Assumptions?

- **Text line detection**
  
  - Jacobs et al. 05, Thillou et al. 05, Weinman et al. 06

- **Word segmentation**
  
  - Ohya et al. 94, McQueen & Mann 00, Thillou et al. 05

- **Binarization**
  
  - McQueen & Mann 00, Chen et al. 02, Thillou et al. 05, Weinman 06

- **Character segmentation**
  
  - Shi & Pavlidis 97, Edwards & Forsyth 05, Jacobs et al. 05, Thillou et al. 05
Overview

- Weinman et al. in ICPR 2008
  - Joint character segmentation and recognition
  - Open vocabulary; lexical bias
- Semi-Markov model
- Typographical regularities
- Experimental Results
- Conclusions
Semi-Markov Model

\[ p(y|x; \bar{\theta}) = \frac{1}{Z(x)} \exp\{U(y, x; \bar{\theta})\} \]

- **Appearance**
  \[ U^A(y_i, x) \]

- **Bigrams**
  \[ U^B(y_i, y_{i+1}) \]

- **Lexicon**
  \[ U^L \]

- **Parse**
  \[ U^P(y_i, y_{i+1}, x) \]

[Lafferty et al. 01]
[Sarawagi & Cohen 05]
Typography Observations

Highest Correlation
Width "g" vs. Width "6"
\[ \rho = 0.98 \]

Median Correlation
Width "m" vs. Width "A"
\[ \rho = 0.62 \]

Lowest Correlation
Width "e" vs. Width "l"

Histogram of Character Pairs

Correlation Coefficient \( \rho \)

Character Pairs

0  50  100  150  200

0  0.5  1
Character-Conditional Width Bigram

- Appearance
  \[ U^A(y_i, \mathbf{x}) \]
- Bigrams
  \[ U^B(y_i, y_{i+1}) \]
- Lexicon
  \[ U^L \]
- Parse
  \[ U^P(y_i, y_{i+1}, \mathbf{x}) \]
- Width Bigram
  \[ U^W(y_i, y_{i+1}, w_i, w_{i+1}) \]
Experimental Data

- Synthetic characters (1600 x 3 fonts)
- English text (82 books, Project Gutenberg)
- Lexicon (50th freq. percentile, SCOWL)
- 1140 char., 12 pixel x-height from 85 signs
Results

Character Error Rate (%)

- No Widths
- Widths

- Appearance: 22.2
- Bigrams (Open Vocab.): 17.4 (No Widths), 16.87 (Widths)
- Closed Vocab.: 19.41 (No Widths), 19.49 (Widths)
- Mixed Vocab.: 15.12 (No Widths), 13.99 (Widths)
Recognition Comparison v. OmniPage

DOUGLASS
Free checking

LIBRARY
AMHERST
TAVERN

Tradiliorpal Asi’?

uucL,ass
Free ehe in
LIBRFIRY
A),lHERbT
TAVF

Traditional Asian Healing Arts

Tradiliorpal Asi’?
HealiIN Arts
Comparison

<table>
<thead>
<tr>
<th>System</th>
<th>Character Error Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OmniPage</td>
<td>23.51</td>
</tr>
<tr>
<td>OmniPage + Binarized</td>
<td>16.61</td>
</tr>
<tr>
<td>Our System</td>
<td>13.99</td>
</tr>
</tbody>
</table>
Low Resolution

Resumes
Resumes
WM/ I WS
krauts
11.-4her
† tttA wms
Summary

- No binarization
- No prior word segmentation
- Joint character/word segmentation and recognition
- Open vocabulary with a lexical bias
- Learned typographical regularities
Thank you!
<table>
<thead>
<tr>
<th>System</th>
<th>Char Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>OmniPage</td>
<td>23.51%</td>
</tr>
<tr>
<td>OmniPage + Binarized</td>
<td>16.61%</td>
</tr>
<tr>
<td>Our System</td>
<td>13.99%</td>
</tr>
</tbody>
</table>
Interpretation Scoring
Recognition Information

- **Appearance**
  - $P(TH \mid \text{English}) = \frac{39}{1000}$
  - $P(QU \mid \text{English}) = \frac{1.4}{1000}$
  - $P(IN \mid \text{English}) = \frac{21}{1000}$
  - $P(QA \mid \text{English}) = \frac{0.001}{1000}$

- **Bigrams**

- **Lexicon**
  - $a$
  - Aaberg
  - Aachen
  - Zymurgy
  - Zyuganov
  - Zzz
Interpretation Graph

[Weinman et al. 2008]
[Jacobs et al. 2005]
[Lucas et al. 2003]