Slide Image Retrieval: A Preliminary Study

Guo Min Liew and Min-Yen Kan

National University of Singapore
Web IR / NLP Group (WING)
Beyond Text in Digital Libraries

Papers don’t store all of the information about a discovery

- Dataset
- Tools
- Implementation details / conditions

They also don’t help a person learn the research

- Textbooks
- Slides

We’ll focus on this
What is Slide Image Retrieval?

- Hypothesis: image and presentation features can enhance text-based features
**Slidir Architecture**

**Training**
1. Pre-Processing
2. Feature Extraction
3. Machine Learning
4. Image Ranking

**Testing**
- Presentation Slides
- Relevance Judgments
- Queries
- Text + Image + Presentation Features
- Model

**Presentation Slides**
1. Preprocessing

Generally consistent background → Perform collective separation

Simple preprocessing, not meant to be state-of-the-art

Top down approach using projection profile cuts with thresholds from PPT template
Motivation

- Problem
  - ECC is very computation-intensive
  - 8-bit micro-controllers like 8051 are too slow for ECC

- Goal
  - Shorten time it takes to perform ECC
  - Lower cost hardware accelerator (minimal area)

- Hardware/software co-design
  - Field arithmetic in SW, rest in HW
  - 922-bit Arithmetic Unit (approx. 11k gates)
  - Addition and Multiplication in GF(2^m)
  - Scalar multiplication over GF(2^{193}) in 120 msec @ 12 MHz

Experimental Paradigm

Motivation

- Problem
  - ECC is very computation-intensive
  - 8-bit micro-controllers like 8051 are too slow for ECC

- Goal
  - Shorten time it takes to perform ECC
  - Lower cost hardware accelerator (minimal area)

- Hardware/software co-design
  - Field arithmetic in SW, rest in HW
  - 922-bit Arithmetic Unit (approx. 11k gates)
  - Addition and Multiplication in GF(2^m)
  - Scalar multiplication over GF(2^{193}) in 120 msec @ 12 MHz
2. Feature Extraction

Utilizes textual, image and presentation features

• Textual Features
  – Main component of feature set
  – Most images identified by surrounding and embedded text

• Image Features
  – Supplement to textual features
  – Useful in ranking images with similar textual feature scores

• Presentation Features
  – Features unique to presentation slides
Sample Features

Textual

1. All text
2. Slide text
3. Next slide’s text
4. Previous slide’s text
5. Slide Title
6. Presentation Title
7. Slide Image Text

Image

- Image Size
- Image Colors
- NPIC image type (CIVR, 06)

Presentation

1. Slide Order
2. Image position on Slide
3. Number of Images on Slide
Machine Learning

Regression Model

– Relevance judgments and feature vectors input into machine learner
– Linear regression used to produce continuous output value for image ranking

Human Relevance Judgments

– 9 participants
– Four presentation sets of 10 queries each
– Participants rank up to five most relevant images to the query
Regression Analysis

- 10-fold cross validation
- System with fielded text features able to replicate human upper bound
- Image and presentation features further improve correlation

<table>
<thead>
<tr>
<th></th>
<th>All text baseline</th>
<th>Fielded Text</th>
<th>With Presentation and Image Features</th>
<th>Inter-annotator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean absolute error</td>
<td>19.36</td>
<td>13.72</td>
<td>13.81</td>
<td>9.40</td>
</tr>
<tr>
<td>(lower is better)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>-0.015</td>
<td>0.565</td>
<td>0.581</td>
<td>0.531</td>
</tr>
<tr>
<td>(higher is better)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Still a gap to improve

Comparison of mean absolute error and correlation

Looks pretty good
Binary Classification

- Human participants judgments converted to binary
- Evaluated using SVM and J48 decision tree classifier
- General accuracy fairly good
- Decision tree algorithm slightly more precise → sequential testing adequate

<table>
<thead>
<tr>
<th></th>
<th>SVM (SMO)</th>
<th>J48 Decision Tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>90.9%</td>
<td>92.5%</td>
</tr>
<tr>
<td>Kappa</td>
<td>.54</td>
<td>.59</td>
</tr>
<tr>
<td>Relevant Class Precision</td>
<td>.59</td>
<td>.70</td>
</tr>
<tr>
<td>Recall</td>
<td>.58</td>
<td>.58</td>
</tr>
<tr>
<td>$F_1$</td>
<td>.59</td>
<td>.63</td>
</tr>
</tbody>
</table>

10-fold performance over 8062 instances
Feature Analysis

- Human subject study with 20 volunteers
- Subjects select most and least relevant images

Sample Human Subject Survey

<table>
<thead>
<tr>
<th>No.</th>
<th>Query</th>
<th>Most Relevant Image by Rank</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Perfect Graphs</td>
<td>A C E B na</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Odd Hole</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2-join</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>double star</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Xuming Liu</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### SLIDIR Features

<table>
<thead>
<tr>
<th>Rank</th>
<th>SLIDIR Features</th>
<th>Regression Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Slide Title</td>
<td>-234.39</td>
</tr>
<tr>
<td>2</td>
<td>Slide Image Type</td>
<td>-124.09</td>
</tr>
<tr>
<td>3</td>
<td>Slide Text</td>
<td>-83.86</td>
</tr>
<tr>
<td>4</td>
<td>Slide Image Text</td>
<td>-83.64</td>
</tr>
<tr>
<td>5</td>
<td>Next Slide’s Text</td>
<td>-29.42</td>
</tr>
<tr>
<td>6</td>
<td>Presentation Title</td>
<td>-9.85</td>
</tr>
<tr>
<td>7</td>
<td>Number of Images on Slide</td>
<td>-1.89</td>
</tr>
<tr>
<td>8</td>
<td>Slide Order</td>
<td>-0.674</td>
</tr>
<tr>
<td>9</td>
<td>Image Size</td>
<td>0.0</td>
</tr>
<tr>
<td>10</td>
<td>Slide Image Position</td>
<td>0.0</td>
</tr>
<tr>
<td>11</td>
<td>Number of Colors</td>
<td>0.0004</td>
</tr>
<tr>
<td>12</td>
<td>Previous Slide’s Text</td>
<td>31.70</td>
</tr>
</tbody>
</table>

**Textual features most important, image and presentation features still play significant role**
Image Ranking Evaluation

- Online survey with 15 volunteers
- 8 given queries and sampled dataset of 100 images
- Volunteers indicate relevant images to query
- 53% of top 10 images are relevant (max of 15 relevant)
- Average precision of 0.74

<table>
<thead>
<tr>
<th>n</th>
<th>Precision at n</th>
<th>Recall at n</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = top 10</td>
<td>.53</td>
<td>.73</td>
</tr>
<tr>
<td>n = top 20</td>
<td>.38</td>
<td>.95</td>
</tr>
<tr>
<td>n = top 30</td>
<td>.27</td>
<td>.98</td>
</tr>
<tr>
<td>n = top 40</td>
<td>.21</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Average Precision .74
Future Work

Limitations
• Very small data set – 100 images from an AI course

Preprocessing
• Improve segmentation and collective extraction

Field Implementation – Ongoing work
• Incorporate with previous work on SlideSeer (JCDL, 2007) that aligns presentation and document pairs

More information:
See you in Singapore!
Next month or next year!

Questions?