Script Independent Keyword Spotting Using Moment Features

Venu Govindaraju

Venu@cubs.buffalo.edu
Keyword Spotting for Multi-script Documents

OCR - Based

Query

Query Subsystem

Transliteration

Class Labels

Document Image Corpus

OCR Subsystem

Preprocessing

Recognition Hypotheses

Matching Subsystem

Match in Feature Space

Image – Based

Query Image Provided OR Rendered

Center for Unified Biometrics and Sensors
University at Buffalo - The State University of New York
Multi-script Documents

Challenge: Script Invariant Word Image Representation
Devanagari OCR
(Block Adjacency Graph)

Branching
Merging

Sample stroke
Graph representation

Segmentation Hypothesis

BAG of alphabets
BAG of conjunct character
Recognition Methodology

• Language model to choose path - results:
  – Script writing grammar rules eliminate two choices:
  – Phonetic n-gram constraints remove another choice:
Word Spotting (GSC)

Previous Work

  - Matching GSC features of two word images.

- Corpus:
  9312 word images (3104 for queries and 6208 for tests) from 776 individuals

- Performance
  Report GSC outperforms DTW

| No. of Top Matches | Precision | | Recall |
|--------------------|-----------|------------------|
|                    | GSC       | DTW              | GSC    | DTW    |
| 1                  | 0.9919    | 0.9156           | 0.00064| 0.00059|
| 5                  | 0.9809    | 0.8845           | 0.00316| 0.00285|
| 10                 | 0.9742    | 0.8658           | 0.00628| 0.00558|
| 20                 | 0.9647    | 0.8419           | 0.01243| 0.01085|
| 50                 | 0.9455    | 0.7997           | 0.03046| 0.02576|
| 100                | 0.9218    | 0.7556           | 0.05939| 0.04869|
| 500                | 0.7858    | 0.5906           | 0.25316| 0.19025|
| 1000               | 0.6553    | 0.4967           | 0.42221| 0.32002|
| 1552               | 0.54549   | 0.43450          | 0.54549| 0.43450|
| 2000               | 0.4817    | 0.3988           | 0.62075| 0.51394|
| 3000               | 0.3892    | 0.3423           | 0.75225| 0.66165|
| 4000               | 0.3308    | 0.3024           | 0.85246| 0.77924|
Word Spotting (DTW)

Previous Work

- **Sequential Profile and DTW**
  [Rath et al, CVPR 2003]
  - **Corpus**
    Washington’s manuscripts
  - **Performance**
    Average precision: 67.92%
  - **Query**
    Image and Text

Word recognition probability

\[
\text{Pr}(\text{wrd} \mid \text{fv}) = \frac{\sum \text{Pr}(\text{wrd}, \text{fv})}{\text{Pr}(\text{wrd}, \text{fv})}
\]
Word Spotting (Gabor)

Previous Work

- Template Free Word Spotting in low-quality manuscripts [Huaigu, et al, ICAPR 2007]
  - Matching Gabor features of two word images.
  - Corpus: 12 medical forms containing 5295 character images.

Performance

Report probabilistic similarity performs better than Euclidean and WMR.

Feature Extraction

\[ V_w = [V_1^T \ V_2^T \ V_3^T \ V_4^T]^T \]

Probabilistic Similarity

\[ C_P(w, V_w) = -\frac{1}{n} \sum_{i=1}^{n} \ln(P(c_i | v_i)) \]
Issues

• Deal with
  – Complex characters (Devanagari)
  – Scale and translation
  – Multiple scripts

• Structural Features (GSC)
  – Script specific therefore ineffective (Srihari et al)
  – Profile features applicable only on long components (Manmatha et al)
Moment Features

Geometric Moments

\[ M_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) \, dx \, dy \]  \hspace{1cm} (1)

\[ M_{pq} = \sum_{X} \sum_{Y} x^p y^q f(x, y) \]  \hspace{1cm} (2)

Center of Gravity

\[ \bar{x} = \frac{M_{10}}{M_{00}}, \quad \bar{y} = \frac{M_{01}}{M_{00}}, \]  \hspace{1cm} (3)

Central Moments

\[ \bar{M}_{pq} = \sum_{X} \sum_{Y} (x - \bar{x})^p (y - \bar{y})^q f(x, y) \]  \hspace{1cm} (4)

Variance

\[ \sigma_x = \sqrt{\frac{\bar{M}_{20}}{M_{00}}}, \quad \sigma_y = \sqrt{\frac{\bar{M}_{02}}{M_{00}}}, \]  \hspace{1cm} (5)

\[ x^* = \frac{(x - \bar{x})}{\sigma_x}, \quad y^* = \frac{(y - \bar{y})}{\sigma_y}, \]  \hspace{1cm} (6)
Moment Features

• Pre-processing of document images

• Moments (up to 7th order) extracted from normalized word images

• Invariant to scale and translations

\[ \mathbf{m}_{pq} = \frac{\sum_X \sum_Y (x^*)^p (y^*)^q f(x, y)}{M_{00}} \]

(7)

Feature vector consists of 30 moment values
Construct for each word image and store in the index
Noise Sensitivity-High Order Moment Moments

Hindi Dataset

Average Precision vs higher order moments,
Apply relevance feedback to re-rank word images
Corpus

**English**: 707 handwritten word images extracted from IAM database and George Washington's historical manuscripts.

**Hindi**: 763 machine print word images extracted from Million Book Project documents.

**Sanskrit**: 693 machine print word images extracted from 5 documents downloaded from the URL: http://sanskrit.gde.to/
Corpus

• The test corpus consists of 5780 word images extracted from Million Book Project Documents.
Keyword Spotting

• ja.ngal → [ -3060.48 , 710.86 , 480388.32 , -43156.29 , ]

• kabUtar → [ 31000.55 , 2774.74 , 496660.19 , 7229.75 , -]

• Vachan → [ 8208.35 , -2379.97 , 146283.25 , 4141.59 , ]

\[
SIM(q, w) = \frac{\vec{q} \cdot \vec{w}}{|\vec{q}| \ast |\vec{w}|}
\]  

(8)
## Cosine Similarity

<table>
<thead>
<tr>
<th>Pair</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>0.9867</td>
<td>0.9932</td>
</tr>
<tr>
<td>B</td>
<td>0.9867</td>
<td>1</td>
<td>0.9467</td>
</tr>
<tr>
<td>C</td>
<td>0.9932</td>
<td>0.9467</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pair</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>1</td>
<td>0.9662</td>
<td>0.9312</td>
</tr>
<tr>
<td>E</td>
<td>0.9662</td>
<td>1</td>
<td>0.9187</td>
</tr>
<tr>
<td>F</td>
<td>0.9312</td>
<td>0.9984</td>
<td>1</td>
</tr>
</tbody>
</table>

A

**Commonwealth**

B Scale

**Commonwealth**

C Linear

D

**लक्ष्मण**

E scale

**लक्ष्मण**

F Linear

**लक्ष्मण**
## Mean Average Precision

<table>
<thead>
<tr>
<th>Query</th>
<th>OCR</th>
<th>GSC</th>
<th>Gradient</th>
<th>Gabor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baba</td>
<td>0.62</td>
<td>0.8</td>
<td>0.41</td>
<td>0.41</td>
</tr>
<tr>
<td>Bandar</td>
<td>1.0</td>
<td>0.092</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Bhakt</td>
<td>0.75</td>
<td>0.52</td>
<td>0.41</td>
<td>0.41</td>
</tr>
<tr>
<td>Jungal</td>
<td>0.65</td>
<td>0.74</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Machali</td>
<td>0.79</td>
<td>0.93</td>
<td>0.89</td>
<td>0.89</td>
</tr>
<tr>
<td>Mata</td>
<td>0.63</td>
<td>0.35</td>
<td>0.38</td>
<td>1.0</td>
</tr>
<tr>
<td>Naksha</td>
<td>1.0</td>
<td>0.77</td>
<td>1.0</td>
<td>0.41</td>
</tr>
<tr>
<td>Raat</td>
<td>0.70</td>
<td>0.72</td>
<td>0.41</td>
<td>0.38</td>
</tr>
<tr>
<td>Ramkumar</td>
<td>1.0</td>
<td>0.89</td>
<td>0.22</td>
<td>0.24</td>
</tr>
<tr>
<td>Vachan</td>
<td>0.86</td>
<td>0.67</td>
<td>0.24</td>
<td>0.22</td>
</tr>
<tr>
<td>Mean for 10</td>
<td>0.80</td>
<td>0.65</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>Mean for 20</td>
<td>0.67</td>
<td>0.60</td>
<td>0.29</td>
<td>0.29</td>
</tr>
</tbody>
</table>
Relevance Feedback

INDEXING

Word Features

Query Feature

Ranking

Feedback

query

results

Relevance

\[ SIM(q, w) = \frac{\vec{q} \cdot \vec{w}}{|\vec{q}| \cdot |\vec{w}|} \] (8)

\[ q_{new} = \gamma \cdot q_{old} + \frac{\alpha}{|R|} \cdot \sum_{i=1}^{\infty} d_i - \frac{\beta}{|NR|} \cdot \sum_{j=1}^{\infty} d_j \] (9)
Relevance Feedback on Vector Space

\[ R_1 = [12333.37, -12148.82] \]

\[ Q_{\text{new}} = [1288.43, -8450.10] \]

\[ Q_1 = [-3060.48, 710.86] \]

\[ NR_1 = [31000.55, 2774.75] \]

\[ NR_2 = [8208.35, -2379.98] \]
Relevance Feedback

\[ q_{new} = \gamma \cdot q_{old} + \frac{\alpha}{|R|} \cdot \sum_{i=1}^{i=R} d_i - \frac{\beta}{|NR|} \cdot \sum_{j=1}^{j=NR} d_j \]  

(9)

\[
\begin{align*}
Q1 &= [-3060.48, 710.86] \\
R1 &= [12333.37, -12148.82] \\
NR1 &= [31000.55, 2774.75] \\
NR2 &= [8208.35, -2379.98]
\end{align*}
\]

Qnew = \( Q1 + 0.75 \times (R1) - (0.25/2) \times (NR1 + NR2) \)

Qnew = [1288.43, -8450.10]

Cosine Similarity (Q1, R1) = -0.8527

Cosine Similarity (Qnew, R1) = 0.8011
## Average Precision

<table>
<thead>
<tr>
<th>Script</th>
<th>W/O Relevance Feedback %</th>
<th>Relevance Feedback %</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>66.30</td>
<td>69.20</td>
</tr>
<tr>
<td>Hindi</td>
<td>71.18</td>
<td>74.34</td>
</tr>
<tr>
<td>Sanskrit</td>
<td>87.88</td>
<td>92.33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Script</th>
<th>GSC 50% Recall</th>
<th>Moments 50% Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>60.0</td>
<td>71.6</td>
</tr>
<tr>
<td>Sanskrit</td>
<td>90.0</td>
<td>94.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Script</th>
<th>Gabor</th>
<th>Moments</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>56.15</td>
<td>66.30</td>
</tr>
<tr>
<td>Hindi</td>
<td>67.25</td>
<td>71.18</td>
</tr>
<tr>
<td>Sanskrit</td>
<td>79.10</td>
<td>87.88</td>
</tr>
</tbody>
</table>
Average Precision Curves for Few Queries

- **English**: 8 queries
- **Sanskrit**: 100 queries
- **Hindi**: 75 queries

**Average Precision**

*Graphs showing the average precision for English, Sanskrit, and Hindi queries.*
Summary

• Keyword Spotting Methods
  – OCR driven
  – Image based: Moments, GSC, Gabor

• Multi-script Documents
  – Moments
  – Relevance Feedback

• Future Work