A Multimodal and Multilevel Ranking Scheme for Large-Scale Video Retrieval

Steven C. H. Hoi, Member, IEEE

Michael R. Lyu, Fellow, IEEE

IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 10, NO. 4, JUNE 2008

Speaker: Ming-Hsun Ko
Date: 2009/05/07
Outline

• Introduction
• Overview of Framework
• Ranking methods
• Visual Feature
• Experimental results
• Conclusions and future works
Introduction

- Multimodal and multilevel (MMML) ranking framework
- Effectiveness of retrieval performance
- Reduce the computational cost
- Large-scale video retrieval tasks
Overview of approach

- User's Query
  - return top K shots
  - Semi-Supervised Ranking ↓ SSR
    - Top N4 ranked shots
  - Supervised Ranking ↑ SVM
    - Top N3 ranked shots
  - NN Ranking of Text & Visual
    - Top M ranked stories
      - Text
    - Top N1 ranked shots
  - Video Processing
  - Video Stories
  - Image Processing
  - Video Shots
  - Multimodal Fusion
Ranking methods

- Text-based
- Nearest Neighbour
- Support Vector Machine (SVM)
- Semi-Supervised Ranking (SSR)
Story 1: Local time 15th morning while Brazil visit of Chinese state president Hu Jintao arrived at the located Sao Paulo suburbs of Brazil countries attack research institute....

Story 2: Economic development of the accomplishment a number of new rich people between different parts by young people in majority they have antiseptic car majority of...

Story 3: Hu Jintao during his speech said in Argentina are Latin America in the region central Asian countries in Latin America between the 2 although remote heavy foreign...
Nearest Neighbour

- Combining visual and textual information

- Normalized retrieval data 0-1

\[ f_{u}^{NN} = \rho f_{u}^{txt} + (1 - \rho) f_{u}^{EU} \]

- Fixed to 0.5 in the experiments
Support Vector Machine (SVM)

- Supervised reranking method

- Users set positive and negative query example

\[ \tilde{G} = (\tilde{V}, \tilde{E}) \quad \text{vertex set} \quad V = L \cup U \]
Multimodal Semi-supervised

• Unsupervised method

\[ G = (V, E) \quad \text{vertex set} \quad V = L \cup U \]

• Semi-supervised method

\[ G' = (V', E'), \text{ where } V' = V \cup L_T \cup U_T \]

• \( L_T \) textual nodes of the query
Visual Feature

- Color:
  - 3 *3 equal grids
  - color space (HSV)
  - 81-dimensional
Visual Feature

• Edge direction histogram
  – histogram >>36 bins of 10 degrees

• Gabor representation

• Five scale levels and eight orientations
Experimental results

- TRECVID 2005
- About 50 hours of rushes on vacation spots provided by the BBC Archive
## Experimental results

<table>
<thead>
<tr>
<th>Methods</th>
<th>MAP TOP1000</th>
<th>Prec. TOP10</th>
<th>Prec. TOP15</th>
<th>Prec. TOP20</th>
<th>Prec. TOP30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>0.0902 (+0.00%)</td>
<td>0.1333</td>
<td>0.2000</td>
<td>0.1917</td>
<td>0.1847</td>
</tr>
<tr>
<td>Text+NN</td>
<td>0.1046 (+15.96%)</td>
<td>0.2583</td>
<td>0.2444</td>
<td>0.2375</td>
<td>0.2097</td>
</tr>
<tr>
<td>Text+SVM</td>
<td>0.1171 (+29.82%)</td>
<td>0.3250</td>
<td>0.3000</td>
<td>0.2806</td>
<td>0.2562</td>
</tr>
<tr>
<td>MMML</td>
<td>0.1267 (+40.47%)</td>
<td>0.3708</td>
<td>0.3333</td>
<td>0.3042</td>
<td>0.2681</td>
</tr>
</tbody>
</table>

**Mean average precision (MAP)**
Experimental results

Fig. 7. Comparison of our solutions to other results in TRECVID 2005.
Conclusions and future works

- Retrieval performance better than TRECVID 2005
- Good balance
  - retrieval performance
  - computational efficiency
- Query-class dependent weighting methods
~~The End~~
Thank you