AN EFFECTIVE VIDEO RETRIEVAL SYSTEM BY COMBINING VISUAL AND TEXTUAL MINING TECHNIQUES

Ja-Hwung Su, Hsin-Ho Yeh, Vincent S. Tseng

Department of Computer Science and Information Engineering
National Cheng Kung University, Tainan, Taiwan, R.O.C.

tseengsm@mail.ncku.edu.tw

2009/05/02
Outline

- Introduction
- System Architecture
- Experimental Evaluations
- Conclusions
- Future Work
Previous work on video retrieval

Textual search
- Rely on the video metadata heavily by exactly matching
- High cost by manual annotation

Visual search
- The traditional content-based video search is limited in the compound and complex visual contents in terms of effectiveness and efficiency.
Introduction (2)

- To retrieve users’ desired videos by combining textual- and visual-based mining

- Our advantages on solving previous problems
  1. For textual-based search, without annotating videos, the videos can be retrieved by the automated metadata we propose.
  2. Reduce semantic gap between video concepts and query terms
  3. The proposed approach can achieve high performance of visual-based search.
System Architecture

Target Videos

Target Video Categories

Visual Feature Extraction

Visual Analysis

Model_{Visual}

Query Refinement

Model_{Textual}

Query Phase

Preprocessing Phase

Textual Feature Collection

Textual Analysis

Feature Analysis

Query Video

Query Term

Relevant Videos
Visual Processing

Target Videos →

- Step 1. Shot Detection
- Step 2. Feature Extraction
- Step 3. Shot Clustering and Encoding
- Step 4. Temporal Pattern Generation
- Step 5. FPI Tree Construction

Model_{Visual} (FPI tree)
Visual Processing
—Shot Clustering and Encoding
Visual Processing
— Temporal-Pattern Generation

<table>
<thead>
<tr>
<th>Clip-id</th>
<th>Shot/Key-Frame Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clip 1</td>
<td>A, B, C, A</td>
</tr>
<tr>
<td>Clip 2</td>
<td>C, B, B, A, E, F</td>
</tr>
<tr>
<td>Clip 3</td>
<td>F, F, E, E, A, B, D, B, C, A, B</td>
</tr>
<tr>
<td>Clip 4</td>
<td>B, C, G, C, A, D, B</td>
</tr>
</tbody>
</table>

For Clip 1

<table>
<thead>
<tr>
<th></th>
<th>Two shot-patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A→B, A→C, A→A</td>
</tr>
<tr>
<td>B</td>
<td>B→C, B→A</td>
</tr>
<tr>
<td>C</td>
<td>C→A</td>
</tr>
</tbody>
</table>

Intelligent Database Laboratory, CSIE, NCKU

Start Point
winsize=3
### Visual Processing

#### Fast-Pattern-Index Tree Construction

<table>
<thead>
<tr>
<th>Clip1</th>
<th>Two shot-patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A→B, A→C, A→A, B→C, B→A, C→A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fast Pattern Index Tree**
### Visual Processing

**– Fast-Pattern-Index Tree Construction**

<table>
<thead>
<tr>
<th>Two shot-patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clip 1</strong></td>
</tr>
<tr>
<td><strong>Clip 2</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>1,2,3</td>
<td>1,2,3,4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>B</td>
<td>1,3,4</td>
<td>2,3</td>
<td>2,3,4</td>
<td>3,4</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>1,3,4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>3,4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fast Pattern Index Tree**
Textual Processing

Target Video Categories

- Step 1. Collect the relevant web URLs
- Step 2. Crawl the relevant web pages
- Step 3. Extract the keyword features
- Step 4. Calculate $tf$ and $idf$

Model$_{Texture}$
Textual Processing

Collect and Crawl The Relevant Web Pages

1. **Query Term**

2. **Google**

3. **Top10 URLs**

---

**Hayley Westenra** - Wikipedia, the free encyclopedia

Hayley Dee Westenra (born 10 April 1987 in Christchurch, New Zealand) [1] [2] is a New Zealand soprano. Her first internationally released album, Pure, ... en.wikipedia.org/wiki/Hayley_Westenra - 148k - Cached - Similar pages - 📤

**Hayley Westenra** - Official


www.hayleywestenra.com - 12k - Cached

**Hayley Westenra** - Wikipedia, the free encyclopedia

Hayley Dee Westenra (born 10 April 1987 in Christchurch, New Zealand) is a New Zealand soprano. Her first internationally released album, Pure, ... en.wikipedia.org/wiki/Hayley_Westenra

**Hayley Westenra** - Listen free at

Listen free to Hayley Westenra (Dark) on Last.fm. Hayley Westenra - 148k - Cached - Similar pages - 📤

---

Intelligent Database Laboratory, CSIE, NCKU
Hayley Dee Westenra is a New Zealand soprano. Her first internationally released album, Pure, reached No 1 on the UK classical charts in 2003 and has sold more than two million copies worldwide.
Textual Processing
– Match by Feature-Keywords

Baseball concept
player
score
rule
batter

Team
field
hit
pitch
run

query
War
policy
force
tactic
weapon
navy

TFIDF
Query Refinement

- Model_{Visual} → Refine the query
- Model_{Textual} → Relevant Videos
- Satisfied? → Yes
- Satisfied? → No
- No: Pick up the interested video
- Yes: Click the interested category

Intelligent Database Laboratory, CSIE, NCKU
Experiments

Dataset

Visual dataset
- 13 video concepts
- 258 video clips with 10464 shots
- Totally, the duration of video data is about 20 hours
- From each concept, 33% of visual videos are randomly selected as the testing data

Textual dataset
- For each concept, top 2000 keywords that came from top 10 search results by Google are selected as feature keywords.
- For each concept, we collect top 10 keywords from wikipedia as testing query.
Experiments – Measurement

- Visual search evaluation

<table>
<thead>
<tr>
<th></th>
<th>Returned</th>
<th>Non-returned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevant</td>
<td>Correct</td>
<td>Incorrect</td>
</tr>
<tr>
<td>Non-relevant</td>
<td>Incorrect</td>
<td>Incorrect</td>
</tr>
</tbody>
</table>

- Example
  - For 10 returned videos, the concepts of five clips are the same as that of query video. Also, there are 20 ground-truth videos.
    - precision = \( \frac{5}{10} \times 100\% = 50\% \)
    - Recall = \( \frac{5}{20} \times 100\% = 25\% \)

\[
\text{precision} = \frac{|\text{Correct}|}{|\text{Returned}|} \times 100\% \quad \text{recall} = \frac{|\text{Correct}|}{|\text{Relevant}|} \times 100\%
\]
Experiments
– Measurement

- Textual search evaluation
  - “Hit” represents the coverage for the correctly returned categories over the resulting ones

\[
Hit = \begin{cases} 
100\%, & \text{if the returned } k \text{ results contain the query term} \\
0, & \text{otherwise}
\end{cases}
\]

- For example
  - If now we have a query term, “homerun,” and 3 concepts (baseball, basketball, racing-car) are returned from our system. Then “homerun” hits the “baseball” concept.
Experiment (1) – Visual Search

- Precision

- Recall
Experiment (2)
- Textual Search

- Ratio of hit

![Bar Chart]

- Hit

- Top 1, Top 3, Top 5

Legend:
- baseball
- basketball
- billiardsball
- cartoon
- F1-racing-car
- racing-car
- military
- plane
- sea-world
- soccer
- surfing
- tennis
- volleyball
System Prototype
– Example

Concept Oriented Content-based Video Retrieval System

<table>
<thead>
<tr>
<th>Concept</th>
<th>Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>racing-car</td>
<td>1.15</td>
</tr>
<tr>
<td>F1-racing-car</td>
<td>0.84</td>
</tr>
<tr>
<td>volleyball</td>
<td>0.50</td>
</tr>
<tr>
<td>sea-world</td>
<td>0.66</td>
</tr>
<tr>
<td>soccer</td>
<td>0.62</td>
</tr>
<tr>
<td>High Level Concept</td>
<td></td>
</tr>
<tr>
<td>sport</td>
<td>3.30</td>
</tr>
<tr>
<td>seaworld</td>
<td>0.66</td>
</tr>
</tbody>
</table>
Conclusions

- We propose a hybrid approach with textual- and visual-based mining strategies

  - **Textual-based**
    - Using Google as our backend mediator to find the web pages that are most relevant to the target query term
    - Associating the users’ interests with the target concepts by feature-keyword matching

  - **Visual-based**
    - With the temporal properties, the proposed pattern-based index can accelerate the search.
    - By pattern-based matching, the user’s desired videos can be found effectively.
Future Work

- Pattern-based index can substantially reduce high dimensional complexity
  - Apply this index structure to different types of multimedia applications
  - Example
    - Music Retrieval, Multimedia Recommendation
Thank you for your attention