Stanford I2V: A News Video Dataset for Query-by-Image Experiments

André Araujo, J. Chaves, D. Chen, R. Angst, B. Girod

Stanford University
Motivation

Example: *Brand Monitoring*

Logo or product

Retrieval

* NBC, 11/18/2014, 7:35:33 PM
Motivation

Example: **Content Linking**

Armstrong’s Critics Largely Unmoved by Interview

By IAN AUSTEN
Published: January 18, 2013

Whatever its impact on the broader public, the first of Lance Armstrong’s two nights of televised confession appeared to have little positive effect on the cycling and antidoping communities.

Several members of both groups faulted Armstrong for the vagueness of his confession, particularly around sensitive matters, and its lack of apology, particularly toward people he had attacked for telling the truth in the past. Many characterized Armstrong’s interview with Oprah Winfrey as being more self-serving than revelatory.
Motivation

Example: Lecture search

Presentation slide

CS246, lecture 12
December 2, 2013
Online demo

http://videosearch.stanford.edu
Outline

- Related Work
- Stanford I2V Dataset
- Dataset Construction
- Baseline Experiments
Outline

- Related Work
  - Stanford I2V Dataset
  - Dataset Construction
  - Baseline Experiments
Related Work: Visual Search

**Query**

- **V2I: Augmented Reality**
  - TCD, Makar et al., 2012
  - Location Rec., Takacs et al., 2010

- **V2V: Content Tracking**
  - Frame Mat. + ST, Douze et al., 2010
  - TRECVID-CCD, Over et al., 2012

- **I2I: Traditional Visual Search**
  - FV, Jégou et al., 2012
  - SVT, Nistér et al., 2006
  - SIFT, Lowe, 2004

- **I2V: Video Search by Image**
  - BoW, Sivic et al., 2006
  - TRECVID-INS, Over et al., 2014
  - TAPS, Araujo et al., 2014
### Related Work: Existing I2V Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Size</th>
<th># Queries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sivic et al., Video-Google, 2006</td>
<td>2h</td>
<td>164</td>
</tr>
</tbody>
</table>
## Related Work: Existing I2V Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Size</th>
<th># Queries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sivic et al., Video-Google, 2006</td>
<td>2h</td>
<td>164</td>
</tr>
<tr>
<td>Over et al., TRECVID-INS, 2014</td>
<td>464h</td>
<td>30</td>
</tr>
</tbody>
</table>
## Related Work: Existing I2V Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Size</th>
<th># Queries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sivic et al., Video-Google, 2006</td>
<td>2h</td>
<td>164</td>
</tr>
<tr>
<td>Over et al., TRECVID-INS, 2014</td>
<td>464h</td>
<td>30</td>
</tr>
<tr>
<td>Araujo et al., CNN2h, 2014</td>
<td>2h</td>
<td>139</td>
</tr>
</tbody>
</table>
## Related Work: Existing I2V Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Size</th>
<th># Queries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sivic et al., Video-Google, 2006</td>
<td>2h</td>
<td>164</td>
</tr>
<tr>
<td>Over et al., TRECVID-INS, 2014</td>
<td>464h</td>
<td>30</td>
</tr>
<tr>
<td>Araujo et al., CNN2h, 2014</td>
<td>2h</td>
<td>139</td>
</tr>
<tr>
<td>Araujo et al., Stanford I2V, 2015</td>
<td>3,801h</td>
<td>229</td>
</tr>
</tbody>
</table>
Outline

- Related Work

- Stanford I2V Dataset

- Dataset Construction

- Baseline Experiments
Stanford I2V Dataset

Query images

Database videos (selected frames)
## Stanford I2V Dataset

<table>
<thead>
<tr>
<th></th>
<th>Full version</th>
<th>Light version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video hours</td>
<td>3.8k hours</td>
<td>1k hours</td>
</tr>
<tr>
<td>Video clips</td>
<td>84k</td>
<td>23k</td>
</tr>
<tr>
<td>Query images</td>
<td>229</td>
<td>78</td>
</tr>
<tr>
<td>Keyframes @1fps</td>
<td>14M</td>
<td>3.8M</td>
</tr>
<tr>
<td>Clip time</td>
<td>2.7 minutes</td>
<td>2.65 minutes</td>
</tr>
</tbody>
</table>
Evaluation Procedure

Query

1st stage: Retrieval of Clips

1

2

3

... 

2nd stage: Temporal Refinement

... 

Ranked retrieval measures:
- Average Precision (AP)
- Precision at 1 (p@1)

Unranked retrieval measure:
- Temporal Jaccard Index
Query/Annotation Viewer

Query image

Clip 1

Clip 2
Outline

- Related Work
- Stanford I2V Dataset
- Dataset Construction
- Baseline Experiments
Dataset Construction: Video Collection

News Videos Recording

Story Segmentation

Daneshi et al., 2013

Video clips
Dataset Construction: Query Set Collection

- Collected images from news websites
  - Used the Internet Archive Wayback Machine
  - Collected 805 candidate images from dates between October 1\textsuperscript{st} 2012 and September 30\textsuperscript{th} 2013

- Types of images:
  - Iconic images (events in the news)
  - Magazine covers (Time, Economist)
Outline

- Related Work
- Stanford I2V Dataset
- Dataset Construction
- Baseline Experiments
Example: Evaluation of Standard Technique

- SIFT descriptors + SCFV global signatures
  [Lowe, 2004] [Duan et al., 2014]

- Retrieval of Clips evaluation:
  - Compare query signature to video frames’ signatures (@1fps) from entire database
  - Evaluate performance over top 100 ranked clips

- Temporal Refinement evaluation:
  - Compare query signature to video frames’ signatures (@1fps) from each correct matching video
  - Feature matching + RANSAC between query and top 50 frames (consider a match if at least 8 inliers are found)
  - Evaluate Jaccard index between matches and ground-truth segments
Example: Evaluation of Standard Technique

Retrieval of Clips: results

Temporal Refinement results

- Light version
- Full version

Latency (secs)

mAP (%)

mJac (%)

Number of Gaussians

128 192 256 512
Summary

- Dataset for video retrieval using query images

- 3.8k hours of video and 229 queries – largest dataset yet

- First dataset to allow true large-scale experiments in this area

- Experiments using standard image retrieval technique were presented, serving as a baseline for future evaluations
Thank you! Questions?

Dataset webpage:

Online demo:
http://videosearch.stanford.edu

André Araujo
http://stanford.edu/~afaraujo
afaraujo@stanford.edu