Frame Fusion for Video Copy Detection

Yao Zhao
Institute of Information Science
Beijing Jiaotong University

Outline

一、 Introduction
二、 Video Copy Detection Based on Frame Fusion
三、 Experiments
四、 Conclusion
It’s easy to make and distribute digital production with the popularity of video editing tools and video sharing websites.

Copyright protection and media tracking become more and more difficult.

Huge amounts of video data in internet make the problem more difficult.

Content based Copy Detection (CBCD) plays an important role in copyright protection
Introduction—
What is Content based Copy Detection

- Content based Copy Detection (CBCD) refers to judging whether a query video contains any content originated from copyright protected video via some content analysis techniques.

Query video

Copyright protected video clip

- CBCD plays an important role in copyright protection, media tracking and law enforcement investigation, and so on.
- CBCD offers an alternative to the watermarking technique.
Introduction—
Challenges to CBCD

- Dealing with various video transformations
- Precise localization
- Real-time requirement
Introduction—Challenges to CBCD

- Video Transformations

Picture in Picture 1

Strong Re-encoding

Picture in Picture 2

Change of Gamma
Introduction—
Challenges to CBCCD

- Video Transformations
  - Blur
  - Letterbox
  - Noise
  - Crop
Outline

一、Introduction
二、Video Copy Detection Based on Frame Fusion
三、Experiments
四、Conclusion
Video Copy Detection Based on Frame Fusion—Objectives & Solution

- **Objectives**
  - Precisely localize the distorted and short video copies in long (or unbounded) video stream
  - Deal with various video transformations

- **Solution**
  - Frame fusion based copy detection
  - Temporal consistent constraint
  - Convert copy detection problem to HMM decoding problem
  - Use Viterbi-Like algorithm to solve HMM decoding problem specific to video copy detection
Video Copy Detection Based on Frame Fusion—System Framework

System Framework of proposed video copy detection system
Video Copy Detection Based on Frame Fusion—Keyframe Selection

- Keyframe selection is a necessary step for improving the efficiency of copy detection

- Uniformly sample each shot at a fixed sampling rate (3f/s in our experiment)
Video Copy Detection Based on Frame Fusion—Feature Extraction

- Extract affine-invariant key points for each frame
- Represent each key points with SIFT descriptor
- Use hierarchy-based Bag-of-Features to calculate a word histogram for each frame
Video Copy Detection Based on Frame Fusion—Similarity Measurement

- Okapi BM25 score ranking:

\[
s(q, d) = \sum_{i=1}^{m} RSJ(w_i) \cdot \frac{f(w_i, d) \cdot (k + 1)}{f(w_i, d) + k \cdot \left(1 - b + b \cdot \frac{|d|}{\text{avd}}\right)}
\]

- For each query keyframe, we got a list of similar frames in reference video.
The purpose of frame fusion is to determine whether the query video contains a copy derived from the reference video by fusing the returned reference frames.

A possible copy clip may be constructed by concatenating the similar frames one by one.
Video Copy Detection Based on Frame Fusion—Temporal Consistent

- Temporal Consistent: If two clips are similar, their frames should be successively similar.
Video Copy Detection Based on Frame Fusion—Viterbi-Like Decoding

- Convert the frame fusion problem into HMM (Hidden Markov Model) decoding problem
- In particular, the query subsequence can be directly treated as the emission sequence of a HMM, and the reference frame constitute the state set

![Diagram](image)

E = \{q_1, q_2, q_3, q_4\} = \{e_1, e_2, e_3, e_4\}
S = \text{Unique}\{L_1, L_2, L_3, L_4\}
S = \text{Unique}\{(F_1, F_2, F_3), (F_4, F_5, F_6), (F_4, F_2, F_5), (F_4, F_5, F_6)\}
S = \{F_1, F_2, F_3, F_4, F_5, F_6\}
S = \{s_1, s_2, s_3, s_4, s_5, s_6\}
Video Copy Detection Based on Frame Fusion—Viterbi-Like Decoding

- The Viterbi algorithm avoids tracking all possible paths by keeping only the most likely path (the partial best path) for each state
- Automatically localize the boundaries of possible copied clips in a long video stream
- Speedup the frame fusion process by a dynamic-programming procedure
Outline

一、 Introduction

二、 Video Copy Detection Based on Frame Fusion

三、 Experiments

四、 Conclusion
Experiments

- Datasets
  - Sound & Vision dataset used in TRECVID 2008 CBCD task
- Evaluation Criteria
  - Copy Overlap Degree: measures the overlap degree between the detected copy and its ground truth
  - Reference Overlap Degree: measure the overlap degree between the asserted reference clip of a copy and its ground truth
## Experiments

- **Transformations**

<table>
<thead>
<tr>
<th>Type</th>
<th>Decrease in Quality (Ta)</th>
<th>Post-Production (Tb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Transformations</td>
<td>Blur</td>
<td>Crop</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>Shift</td>
</tr>
<tr>
<td></td>
<td>Frame dropping</td>
<td>Contrast</td>
</tr>
<tr>
<td></td>
<td>Contrast</td>
<td>Caption (text insertion)</td>
</tr>
<tr>
<td></td>
<td>Compression</td>
<td>Flip (vertical mirroring)</td>
</tr>
<tr>
<td></td>
<td>Ratio</td>
<td>Insertion of pattern</td>
</tr>
<tr>
<td></td>
<td>Noise</td>
<td>Picture in picture</td>
</tr>
</tbody>
</table>
Experiments

- Transformations

Original

T1: Cam Cording

T2: Picture in Picture

T3: Insertions of Pattern

T4: Strong Reencoding

T5: Change of Gamma

T6: Combining 3 transformations of the type Ta

T7: Combining 5 transformations of the type Ta

T8: Combining 3 transformations of the type Tb

T9: Combining 5 transformations of the type Tb

T10: Combining all transformations
Experiments

Comparison with state-of-the-art copy detection systems on copy overlap degree with varied transformations
Experiments

Comparison with state-of-the-art copy detection systems on reference overlap degree with varied transformations
Outline

一、Introduction
二、Video Copy Detection Based on Frame Fusion
三、Experiments
四、Conclusion
Conclusion

- Frame fusion framework for handling copy detection problem in unbounded query stream
- Convert copy detection to HMM decoding problem
- Viterbi-like algorithm for high efficient frame fusion and boundary decision
Thank you !