Fast and Robust Video Deduplication

[1]Chris Henry, [2]Rijun Liao, [3]Ruiyuan Lin, [4]Zhebin Zhang, [5]Hongyu Sun, & [6]Zhu Li

[1,2,6]University of Missouri-Kansas City

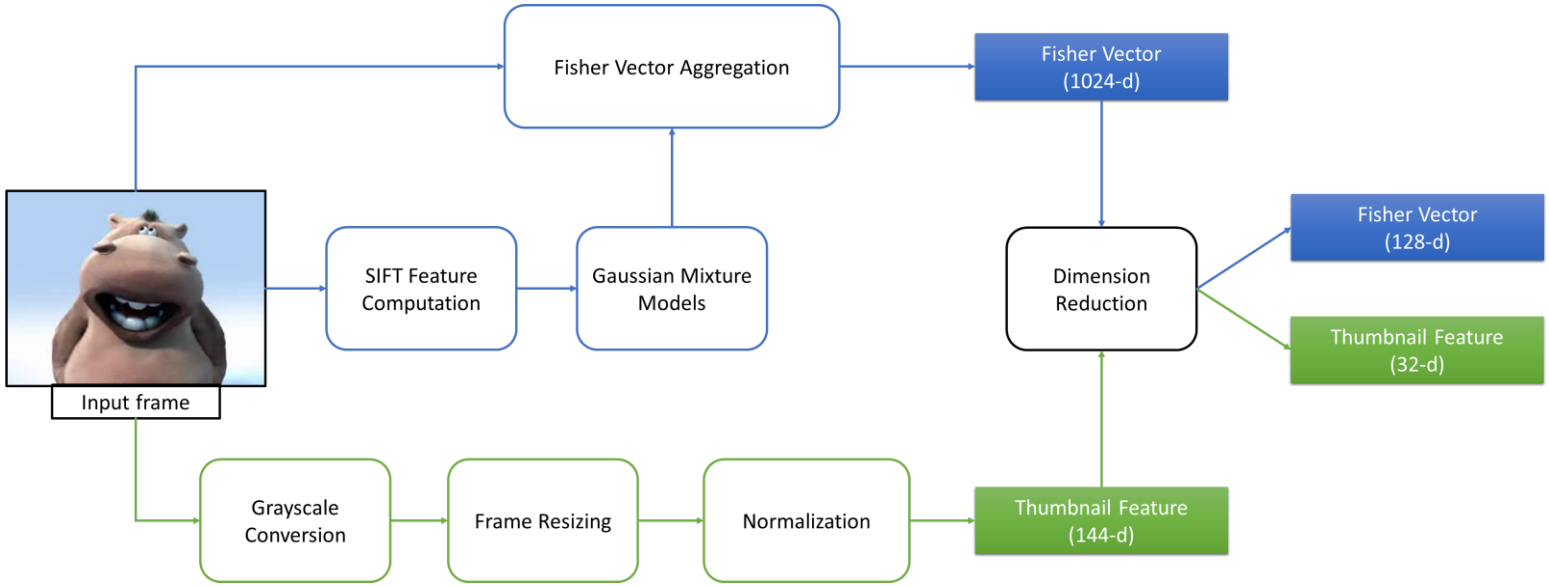
[3,4,5]InnoPeak Technology (Oppo US Research Center)

Introduction

- Social networks and mobile devices have caused a surge in the video content being shared online. Consequently, this has resulted in a rise in illegal pirate videos.
- These illicit copies have the same content as the original videos with slight modifications to dodge copy detection systems. Identifying these duplicates is known as video deduplication.
- Managing this vast amount of video data poses a significant challenge, and having knowledge of duplicate videos is crucial for applications like video content management, copyright protection, video surveillance, etc.
- In this work, we propose fast and robust video deduplication system. Given a query video, the system can retrieve duplicate video from a large 1-million frame (equivalent to 145 hours video) repository in around 460ms with a recall of 90.87%.

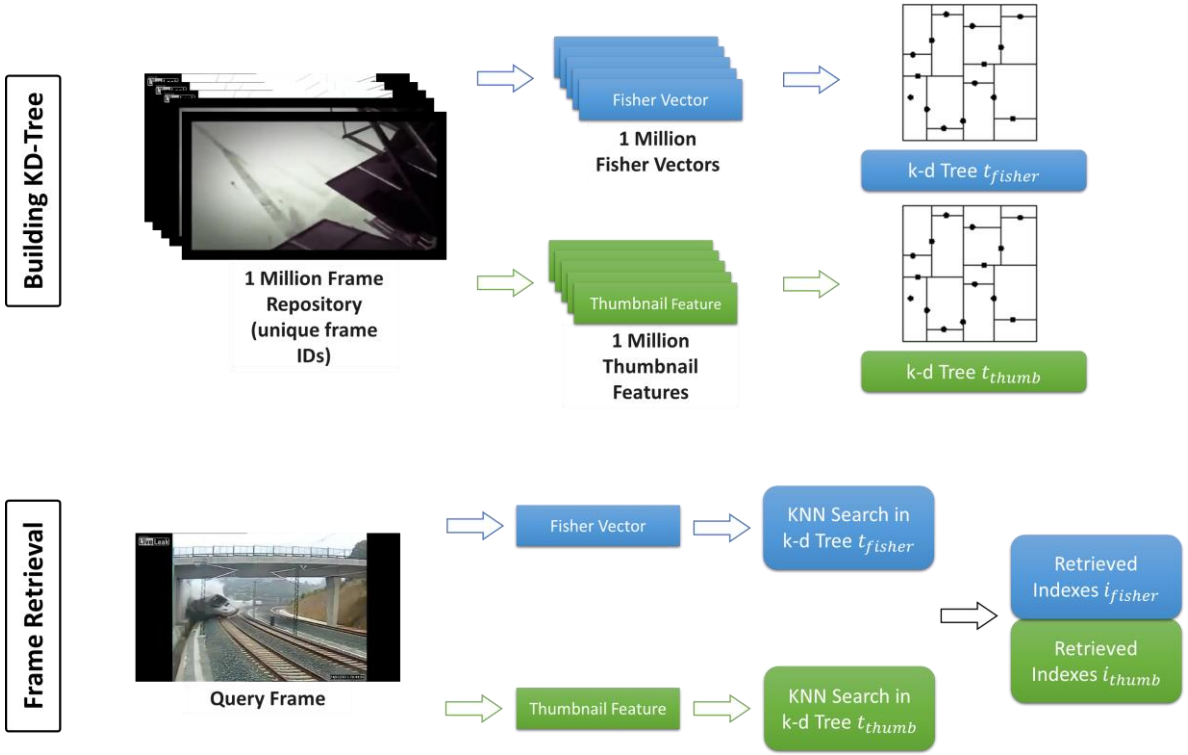
Proposed Method

1. Feature Generation



Proposed Method

2. Frame Retrieval



Proposed Method

3. Temporal Data Pruning: Each index is associated to a video ID and unique global timestamp.
   - Video ID Pruning
     - Compare the videos IDs of the retrieved frames in each set in $S_{id}^{q} = \left[S_{id}^{q(\tau_1)}, S_{id}^{q(\tau_2)}, \ldots, S_{id}^{q(\tau_m)}\right]$ where $S_{id}^{q(\tau_m)}$ contains indexes of the retrieved frames for query frame $q(\tau_m)$. Only video IDs which exists in all sets are kept while the remaining are pruned.
   - Timestamp Pruning
     - Compare the timestamp difference between the retrieved frames in each set in S_{id}^{q}. The timestamp difference between the frames in each set must be equal to τ which the sampling time while generating the query video.
   - The videos that satisfy the above criteria are retrieved.

Dataset Details

- FIVR-200K dataset (2019)
  - Contains data for Duplicate Scene Videos (DSV), Complementary Scene Videos (CSV), and Incident Scene Videos (ISV).
    - DSV: Videos that share at least one scene regardless of any applied transformation.
    - CSV: Videos that contain part of the same spatio-temporal segment but captured from different viewpoints.
    - ISV: Videos that capture the same incident, i.e., they are spatially and temporally close but have no overlap.
  - 225,960 video clips. Duplicate Scene Videos are 7,558 videos (only 5,108 were available for download).
  - 100 queries associated with 4,687 Wikipedia events
  - Extracted 2 frames per second from each DSV labelled video making 1-million frame repository (1, 044, 976-frames exactly) which is about 145 hours of videos.
  - Test data: Randomly selected about 493 videos from the repository and created augmented copies of these videos. Query video of duration 12 seconds was extracted from each of the 493 selected videos to be used as query video (represented as 4 frames).

Implementation Details

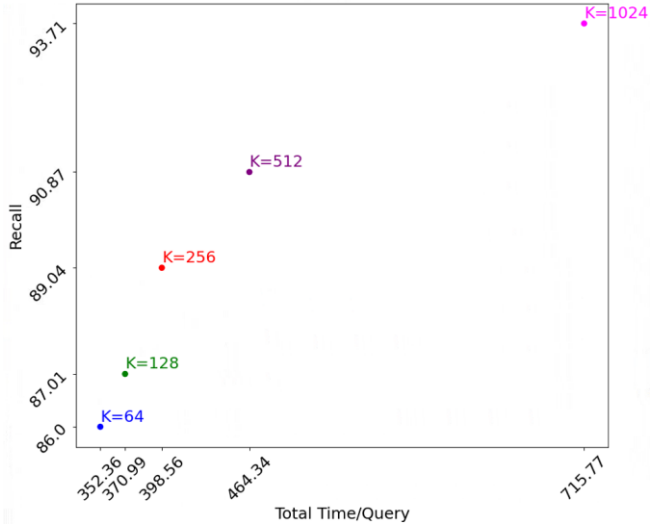
- Machine: Computer with Intel Core i7-12700 CPU with 64 gigabytes of RAM.
- Programming languages:
  - Python: Entire video deduplication system.
  - MATLAB: Extraction of SIFT keypoints features, GMM training, and computation of fisher vector.

Experimental Results

1. Recall
2. Time Consumption

Recall (K=64)	Recall (K=64128	Recall (K=256)	Recall (K=512)	Recall (K=1024)
86.00	87.01	89.04	90.87	93.71

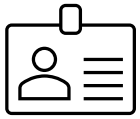
Time Consumption (ms)	K=64	K=128	K=256	K=512	K=1024
Frame Retrieval Time/Frame	87.51	91.20	94.87	98.38	101.83
Frame Retrieval Time/Query	350.07	364.83	379.50	393.53	407.34
Temporal Data Pruning/Query	2.21	6.06	18.94	70.63	308.13
Total Time/Frame	88.09	92.74	99.64	116.08	178.94
Total Time/Query	352.36	370.99	398.56	464.34	715.77



Conclusion

- A fast and robust video deduplication system is proposed.
- The system is able to retrieve a query video from a large 1-million frame repository around 460ms with a recall of 90.87%.
- The robustness of the system was tested on the FIVR-200K dataset. The results confirm that the proposed system is invariant to typical editing effects.

Thank You!



Chris Henry

chffn@umkc.edu

https://sites.google.com/view/chrishenry

Any Questions?