

Video Summarization

HUYNH TIEN NAM, University of Science, VNU – HCMC

Video Summarization

Categories and Subject Descriptors: Computer Vision

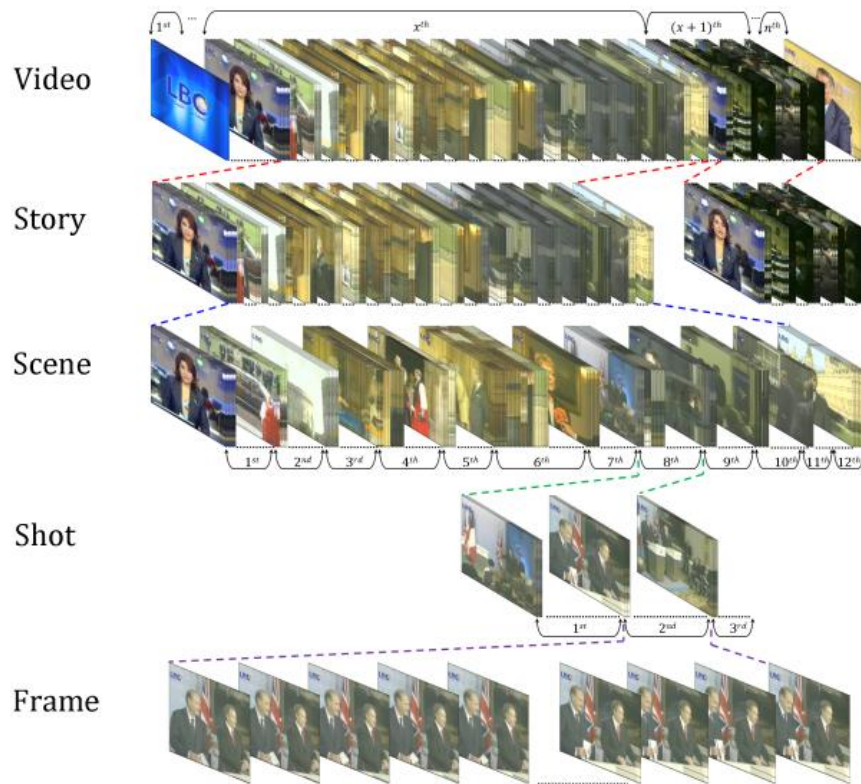
General Terms: Video Summarization

Additional Key Words and Phrases: shot boundary detection; temporal video segmentation; video content analysis; video browsing; video retrieval ...

1. Giới thiệu

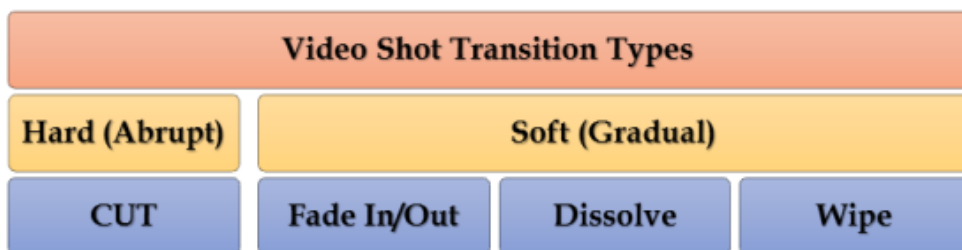
- Sự gia tăng gần đây về số lượng video trong mạng không gian mạng là do sự sẵn có của các thiết bị đa phương tiện, công nghệ truyền thông phát triển cao, và các thiết bị lưu trữ chi phí thấp.
- Những video này chỉ được lưu trữ trong cơ sở dữ liệu thông qua chú thích văn bản. Việc duyệt và truy xuất video dựa trên nội dung không hiệu quả do phương pháp dùng để lưu trữ các video trong cơ sở dữ liệu. Cơ sở dữ liệu video có kích thước lớn và chứa nhiều thông tin, và các đặc điểm này nhấn mạnh sự cần thiết của các phân tích cấu trúc video tự động.
- Shot boundary detection (SBD) được xem là một quá trình quan trọng của việc duyệt và truy xuất video. SBD nhằm mục tiêu phát hiện quá trình chuyển đổi và ranh giới giữa các shot liên tiếp, do đó, những shot giàu thông tin được sử dụng trong việc lập chỉ mục và truy xuất video dựa trên nội dung.
- Video bao gồm một câu chuyện duy nhất (như một trận bóng đá) hay nhiều câu chuyện khác nhau (như tin tức). Một câu chuyện được định nghĩa là một clip thu được một loạt các sự kiện (event) hoặc một hành động liên tục và nó có thể bao gồm một vài cảnh (scene). Một cảnh (scene) là một tập hợp các shot liên tiếp về mặt ngữ nghĩa và theo thời gian được chụp ở nhiều góc máy.
- Shot là phần xây dựng của video, nó là một tập hợp một hoặc nhiều khung hình được lấy liên tục (không gián đoạn) bởi một thiết bị ghi duy nhất và những khung hình này tượng trưng cho một hành động không ngừng trong thời gian và không gian thể hiện một hành động hoặc sự kiện nhất định.

Video Summarization



2. Các loại chuyển đổi trên video:

- Biên giới giữa hai shot được gọi là ranh giới (boundary) hoặc chuyển đổi (transition). Ghép nối giữa hai hoặc nhiều shot được triển khai trong quá trình chỉnh sửa video để tạo video trong quá trình sản xuất video.
- Gồm 2 loại chuyển đổi chính:
 - + Hard Transition.
 - + Soft Transition.



3. Các hướng tiếp cận việc phát hiện ranh giới các shot (Shot Boundary Detection)

A. Pixel-based approach:

- Hướng tiếp cận dựa trên pixel liên quan đến tính toán chênh lệch giữa hai điểm ảnh tương ứng (ở vị trí x và y) của hai frame video liên tiếp (f_n và f_{n+1}). Tiếp theo, tổng chênh lệch pixel được xác định và so sánh với ngưỡng. Một ranh giới được khai báo nếu tổng vượt quá ngưỡng đã chọn.

$$D(f_n, f_{n+1}) = \frac{1}{N_x N_y} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} |P(f_n, x, y) - P(f_{n+1}, x, y)|$$

$$DP(f_n, f_{n+1}, x, y) = \begin{cases} 1 & \text{if } |P(f_n, x, y) - P(f_{n+1}, x, y)| > T_1 \\ 0 & \text{otherwise} \end{cases}$$

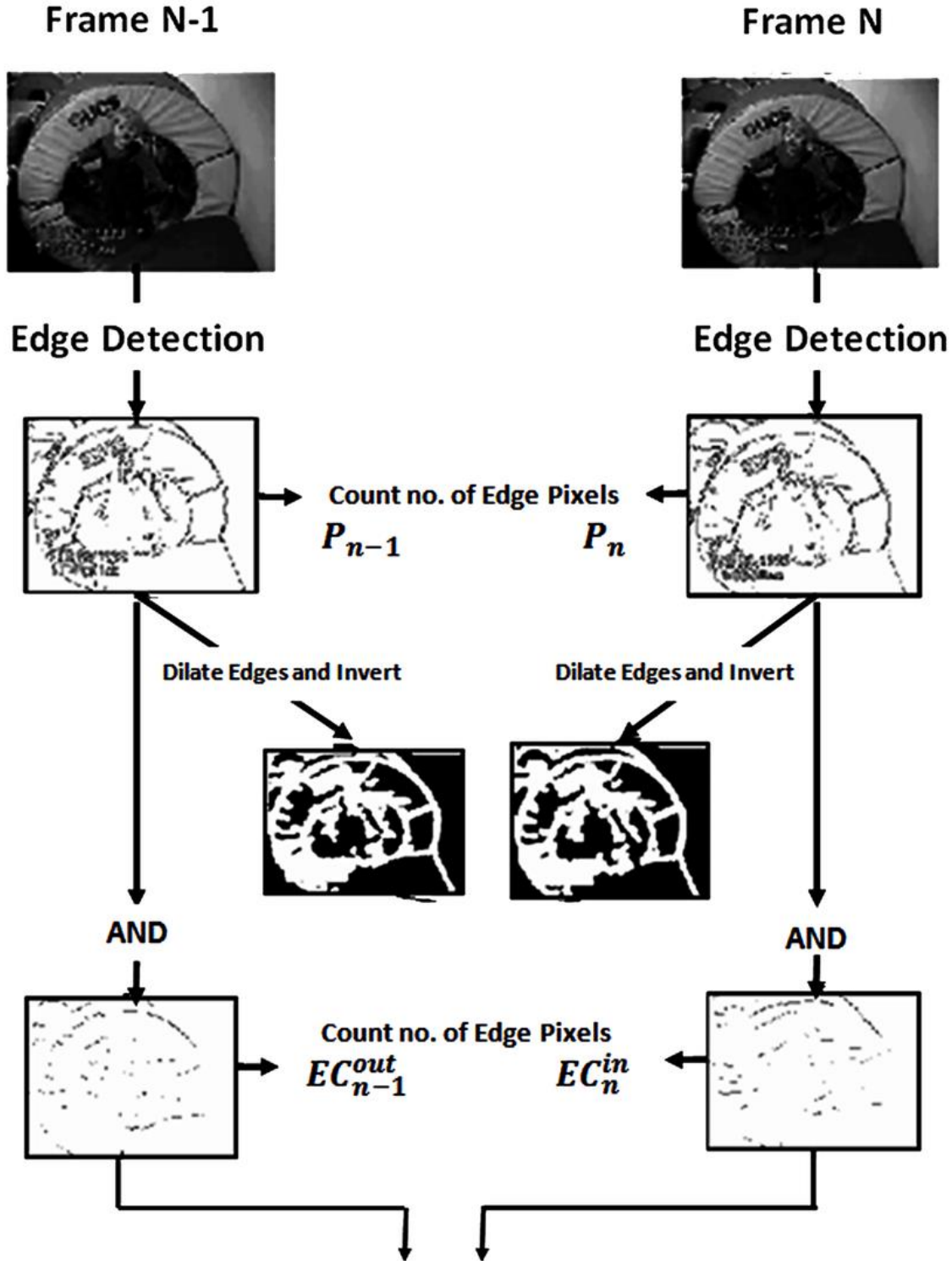
B. Histogram-Based Approach:

- Hướng tiếp cận dựa trên histogram liên quan đến tính toán chênh lệch giữa hai histogram ảnh mức xám của hai frame video liên tiếp (f_n và f_{n+1}). Tiếp theo, chênh lệch histogram được xác định và so sánh với ngưỡng. Một ranh giới được khai báo nếu tổng vượt quá ngưỡng đã chọn.

$$HDM(f_n, f_{n+1}) = \sum_{j=1}^{N_H} |H(f_n, j) - H(f_{n+1}, j)|$$

C. Edge Change Ratio (ECR):

- Một đặc trưng được chứng minh là hữu ích trong việc phát hiện shot boundary là cạnh. Zabih et al đã đề xuất cách tiếp cận ECR. Các cạnh của đối tượng trong frame cuối trước khi hard cut không thể tìm thấy trong frame đầu sau hard cut và ngược lại các cạnh của vật thể trong frame đầu sau hard cut cũng không thể tìm thấy trong frame cuối trước hard cut.
- Nếu ECR này lớn hơn ngưỡng được xác định trước thì nó được coi là cắt cảnh giữa các frame. Điều này được lặp lại cho tất cả các frame trong video và những hard cut có thể được phát hiện.



$$ECR_n = \max \left(ECR_n^{in}, ECR_{n-1}^{out} \right) = \max \left(\frac{\sum EC_n^{in}}{\sum p_n}, \frac{\sum EC_{n-1}^{out}}{\sum p_{n-1}} \right)$$

4. Thực nghiệm:

A. Pixel-based approach:

- i. Tính độ dị biệt giữa các frame:
 - + Dùng thư viện opencv để đọc video, chuyển hết các frame sang ảnh mức xám, sau đó dùng lệnh absdiff để trừ 2 frame liên tiếp nhau.
- ii. Phát hiện các frame có tiềm năng:
 - + Tạo cửa sổ bao gồm 30 frames tuần tự và chọn frame có sự khác biệt lớn nhất, được gọi là max_diff_frame.
 - + Kiểm tra xem sự khác biệt của frame đã chọn có chênh lệch trung bình gấp 3 lần so với các khung khác trong cửa sổ hay không. Nếu thoả mãn được điều kiện ở trên, hãy xem max_diff_frame như một trong những frames có tiềm năng.
 - + Chọn thêm 30 frames tuần tự sau max_diff_frame cộng 8 frame, vì ta xem xét một shot không nên ngắn hơn 8 khung hình.
 - + Lặp lại 2 bước trên cho đến khi tất cả các frame được đưa vào cửa sổ.
- iii. Tối ưu hoá các frame có tiềm năng:
 - + Kiểm tra xem sự khác biệt của khung tiềm năng có lớn hơn hoặc bằng 10 không.
 - + Đối với mỗi frame tiềm năng, hãy kiểm tra xem sự khác biệt có nhiều hơn hai lần so với mức trung bình của 10 khung hình trước nó và 10 khung tiếp theo không. Nếu yêu cầu nêu trên được đáp ứng, hãy xem frame tiềm năng là một trong các shot boundary frame.
- Link code colab trên tập RAI: https://colab.research.google.com/drive/1UsYH2WAPV_S5UalY89ln7kQncTqCU9mX?usp=sharing
- Link kết quả trên tập RAI: <https://drive.google.com/drive/folders/10IDbVxXZsrY4V-jL0CIPxkuMuj3ci8-W?usp=sharing>

Recall: 77,92 %

Precision: 81,03 %

Video Summarization

- Link code colab trên tập BBC:
<https://colab.research.google.com/drive/1Id3aOyzWXJDyJMuF7lwgxSlFepYIBtMK?usp=sharing>
- Link kết quả trên tập BBC:
https://drive.google.com/drive/folders/1nKFq_vXaQKY4kXUPkKDt8Nt6i6P-dE9K?usp=sharing

Recall: 84,5 %

Precision: 96,08 %

B. Histogram-based approach:

- Dùng thư viện opencv để đọc video, chuyển hết các frame sang ảnh mức xám, tiếp theo ta tính histogram cho từng frame. Sau đó ta tiến hành trừ 2 histogram của 2 frame liên tiếp.
- Ta so sánh kết quả thu được với threshold.
- Nếu vượt ngưỡng, ta lưu lại chỉ số frame đó làm frame kết thúc shot và lưu frame kế làm frame đầu cho shot tiếp theo.
- Link code colab trên tập RAI:
https://colab.research.google.com/drive/1ia5FuQ_dm6J8gOQ0iQkFqtT5LJ0JCogX?usp=sharing
- Link kết quả trên tập RAI:
<https://drive.google.com/drive/folders/1jylNtFyeT1826JwoQAhJHWiNKyT4eRY3?usp=sharing>

Recall: 82,78 %

Precision: 76,7 %

- Link code colab trên tập BBC: <https://colab.research.google.com/drive/1j3eaAQ-uqsZT7HyMuo2RTPzG4r-LfbZ?usp=sharing>
- Link kết quả trên tập BBC:
https://drive.google.com/drive/folders/17RnraovhxqtJmNnc3KSJYkdZMHWvwz_V?usp=sharing

Recall: 89,96 %

Video Summarization

Precision: 87,01 %

C. ECR approach:

- Dùng thư viện opencv để đọc video, chuyển hết các frame sang ảnh mức xám. Tiếp theo ta tính thông số ECR cho 2 frame liên tiếp.
- Ta so sánh ECR thu được với threshold và lưu kết quả.
- Nếu vượt ngưỡng, ta lưu lại chỉ số frame đó làm frame kết thúc shot và lưu frame kế làm frame đầu cho shot tiếp theo.
- Link code colab trên tập RAI: https://colab.research.google.com/drive/1g-6ZrDMZAb5LWeC_es1fWaBgRCnKfGiZ?usp=sharing
- Link kết quả trên tập RAI: <https://drive.google.com/drive/folders/1uwxSSiUw8y1F6K5EsKzSvRWUElHWxXaw?usp=sharing>

Recall: 71,41 %

Precision: 52,04 %

- Link code colab trên tập BBC: <https://colab.research.google.com/drive/1ahkfvL4gnZfrqxCreQ3iHguP9KNpaxZb?usp=sharing>
- Link kết quả trên tập BBC: https://drive.google.com/drive/folders/14weDTfqw_0pDrmRgImQtLCsHOVY_EdvC?usp=sharing

Recall: 49,88 %

Precision: 37,54 %

D. Tính độ chính xác:

Recall = Correct / (Correct + Missed)

Precision = Correct / (Correct + False)

Sai số cho phép: $\varepsilon = 5$.

Video Summarization

Link code colab:

https://colab.research.google.com/drive/1kLn6i3cbYJbdtP9aMk2VA8gYzhHzin_p?usp=sharing

Tài liệu tham khảo:

Mann, Jaspreet Kaur and Navjot Kaur. "Key Frame Extraction from a Video using Edge Change Ratio." (2015).

Abdulhussain SH, Ramli AR, Saripan MI, Mahmmod BM, Al-Haddad SAR, Jassim WA. Methods and Challenges in Shot Boundary Detection: A Review. Entropy (Basel). 2018 Mar 23;20(4):214. doi: 10.3390/e20040214. PMID: 33265305; PMCID: PMC7512729.

Jacobs, Arne, Andrea Miene, George T. Ioannidis and Otthein Herzog. "Automatic Shot Boundary Detection Combining Color, Edge, and Motion Features of Adjacent Frames." TRECVID (2004).

github.com/yonatankatz/edge-change-ratio-example