

BIỂU DIỄN VIDEO RÚT GỌN BẰNG CÁC KHUNG ĐẠI DIỆN NHỜ VÀO KỸ THUẬT PHÂN ĐOẠN VIDEO VÀ RÚT TRÍCH KHUNG ĐẠI DIỆN

Nông Thị Hoa^{*}, Nguyễn Văn Tào

Trường Đại học Công nghệ thông tin & Truyền thông - ĐH Thái Nguyên

TÓM TẮT

Hiện nay, các video chất lượng cao xuất hiện ngày càng nhiều nên các công nghệ nén dữ liệu cũng được phát triển để giảm dung lượng lưu trữ. Tuy nhiên, các CSDL video lớn dùng cho mục đích phân tích nội dung video và rút trích thông tin có ích từ video vẫn có nhu cầu rút gọn nội dung và dung lượng video để vừa tăng tốc độ tìm kiếm vừa giảm dung lượng lưu trữ. Để giải quyết nhu cầu này, các nghiên cứu về phân đoạn video để rút gọn nội dung và dung lượng video bằng cách khung đại diện (key-frame) đã được phát triển. Trong bài báo này, chúng tôi trình bày nguyên tắc xây dựng ứng dụng biểu diễn video rút gọn bằng các khung đại diện. Dựa vào đặc trưng của các video phổ biến, một ứng dụng thực đã được xây dựng để thể hiện các video rút gọn với các lựa chọn thích hợp về kỹ thuật phân đoạn và kỹ thuật rút trích khung đại diện. Các thử nghiệm được làm trên tập videothu thập từ hai kênh tin tức nổi tiếng NHK và CNN có định dạng MP4 để đánh giá hiệu quả của việc dùng các khung đại diện thay thế cho video cả về dung lượng lưu trữ và về tốc độ duyệt nội dung video. Kết quả thực nghiệm cho thấy một lựa chọn tốt về kỹ thuật phân đoạn và kỹ thuật rút trích khung đại diện đã mang lại hiệu quả đáng kể trong việc giảm dung lượng lưu trữ và giảm thời gian duyệt video.

Từ khóa: phân đoạn video, rút trích khung đại diện, rút gọn video, phân tích video

GIỚI THIỆU CHUNG

Hiện nay, sự ra đời của nhiều thiết bị số hiện đại đã tạo ra một số lượng lớn các video chất lượng cao. Do các video này có thể cung cấp nhiều tri thức hữu ích nên việc phân tích nội dung video và rút trích thông tin có ích từ video ngày càng trở thành nhu cầu cấp thiết cho nhiều ứng dụng thực. Tuy nhiên, các video chất lượng cao thường đòi hỏi dung lượng lưu trữ lớn nên các công nghệ nén video đã ra đời.

Theo cấu tạo của video, mỗi video gồm nhiều đoạn (segment). Mỗi đoạn gồm nhiều khung (frame) tương tự nhau để thể hiện nội dung của đoạn. Do tính tương tự của các khung trong một đoạn, chúng ta có thể chọn một khung đại diện để thể hiện thay cho nội dung cả đoạn. Điều này nghĩa là có thể dùng khung đại diện của các đoạn để thể hiện nội dung cho toàn bộ video. Kết quả là dung lượng dùng để thể hiện nội dung của video giảm và thời gian để duyệt nội dung video trên các khung đại diện cũng giảm. Hai ưu điểm này là

yêu cầu cấp thiết của các CSDL video lớn dùng cho nhiệm vụ phân tích và rút trích thông tin có ích từ nội dung video.

Trong bài báo này, chúng tôi trình bày ngắn gọn nguyên tắc xây dựng một ứng dụng biểu diễn video rút gọn bằng các khung đại diện. Chúng tôi đưa ra một lựa chọn thích hợp về kỹ thuật phân đoạn video và kỹ thuật rút trích khung đại diện để xây dựng ứng dụng biểu diễn video rút gọn dựa vào các đặc trưng của các video phổ biến hiện nay. Các thử nghiệm được làm trên tập video thu thập trên Internet từ hai kênh tin tức nổi tiếng NHK và CNN để đánh giá tính hiệu quả của ứng dụng. Các kết quả thực nghiệm được trình bày và tổng hợp để đánh giá hiệu quả của việc dùng các khung đại diện thay thế cho video cả về dung lượng lưu trữ và về tốc độ duyệt nội dung video.

Cấu trúc của bài báo gồm 4 phần. Phần II tổng kết các nghiên cứu liên quan. Trong phần III, nguyên tắc phát triển một ứng dụng trình bày video rút gọn bằng các khung đại diện được trình bày. Tiếp theo, các bước chi tiết trong xây dựng ứng dụng được mô tả trong

^{*} Tel: 01238 492484, Email: nongthihoaa@gmail.com

Phần IV, Phần V trình bày về kết quả thực nghiệm. Cuối cùng, một vài kết luận và hướng phát triển được nêu ra.

CÁC NGHIÊN CỨU LIÊN QUAN

Phân đoạn video là phân chia video thành các đoạn liên tục dựa vào nội dung video. Nhiều kết quả nghiên cứu đã được công bố trong thời gian qua. Costas Cotsaces và đồng nghiệp [4] đã tổng hợp các thao tác khai thác các thông tin cơ bản từ video. Các tổng kết tập trung vào phát hiện ranh giới chuyển đoạn và biểu diễn video thu gọn. S.V. Porter và đồng nghiệp [10] đã giới thiệu một phương pháp mới để phát hiện sự chuyển đoạn trong video bằng việc dùng hệ số tương quan khung trung bình và ước lượng chuyển động dựa trên khối ảnh. O. Chum và đồng nghiệp [3] đã đề xuất hai phương thức mới cho hình ảnh từng lập gần và phát hiện đoạn mới của video. Cách đầu dựa trên biểu đồ màu và dùng bảng băm để phân lớp. Cách thứ hai sử dụng bộ mô tả đặc trưng SIFT và tính sự giao thoa giữa các khung bằng thuật toán min-Hash. Z. Cernekova và đồng nghiệp [2] đã trình bày một số kỹ thuật phát hiện biên đoạn dựa trên biểu đồ màu, biến đổi cosin rời rạc, vector chuyển động, và các phương pháp khối phù hợp. A. Hanjalic [6] trình bày một giải pháp để phát hiện biên đoạn dựa trên sự cực tiểu hóa xác suất phát hiện lỗi trung bình. T. Vasileios và đồng nghiệp [12] đề xuất phương thức phân cụm các khung dựa trên sự tương đồng của các khung. Nhóm ước tính số lượng cụm và áp dụng thuật toán k-means trong bước phân cụm dựa vào vector riêng của ma trận sự tương đồng. Y. Alper, A.S. Mubarak [1] trình bày một thuật toán dùng dạng biến đổi của không gian màu RGB để mô tả các khung trong một hệ tọa độ mô tả nhiều hơn.

Việc rút trích khung đại diện từ video là tìm các điểm nổi bật trong nội dung của video. Mục đích là tìm ra các khung chứa thông tin đại diện nhất cho nội dung video. C. Kim, J.N. Hwang đã dùng kỹ thuật dựa vào đối

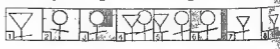
tượng [7],[8]. Ban đầu, nhóm lấy ra các đối tượng từ mỗi khung bằng phương pháp tách cạnh của Canny. Các đối tượng ở hai khung được gắn với nhau bởi khoảng cách tính từ tâm của khung và khoảng cách lớn nhất giữa các đối tượng trong hai khung đang xét. Một khung được gán nhãn là khung đại diện nếu khoảng cách của khung này so với khung trước vượt quá một ngưỡng đã chọn. Z. Li và đồng nghiệp [9] đã tìm điểm nổi bật trên các video nén. Việc tìm sự nổi bật được thực hiện theo hai cách gồm giảm thiểu sự sai khác với tỷ lệ tổng hợp, hoặc giảm thiểu tỷ lệ tổng hợp với mức sai khác định sẵn. X. Zhu và đồng nghiệp [11] chia video thành các đoạn dựa vào sự tương đồng về biểu đồ màu và nền giữa các khung. Các điểm nổi bật được xây dựng bằng chọn khung đại diện ứng với mỗi đoạn video. A. Ferman, A. Tekalp [5] thực hiện tìm điểm nổi bật qua hai bước. Đầu tiên, nhóm chia video thành các đoạn dựa vào biểu đồ màu trung bình. Thuật toán phân cụm mờ được thực hiện trên các khung. Với mỗi cụm, khung nằm ở trung tâm cụm sẽ được chọn làm khung đại diện.

NGUYÊN TẮC XÂY DỰNG ỨNG DỤNG BIỂU DIỄN VIDEO RÚT GỌN BẰNG KHUNG ĐẠI DIỆN

Ứng dụng biểu diễn video rút gọn bằng các khung đại diện cần trải qua các bước sau.

- Bước 1:** Phân chia video thành các đoạn con.
- Bước 2:** Rút trích khung đại diện của mỗi đoạn thu được ở Bước 1.
- Bước 3:** Biểu diễn các khung đại diện của toàn bộ video.

Hình 1 mô tả rõ hơn các thành phần của một video mẫu. Video trong Hình 1 có 5 đoạn. Chọn khung cuối của mỗi đoạn làm khung đại diện thì khung đại diện của các đoạn trong video mẫu là khung 1, 3, 5, 6, và 8. Khi đó, video rút gọn chỉ gồm có 5 khung đại diện này.



Hình 1. Cấu trúc của một video mẫu

Các kỹ thuật phân đoạn video

Phân đoạn video là chia video thành các đoạn có nội dung khác nhau. Nói cách khác, chúng ta cần tìm ra điểm khác biệt giữa các khung để biết các khung ở hai đoạn khác nhau. Vì vậy, phân đoạn video có hai bước gồm rút trích các đặc trưng của các khung và dùng các đặc trưng để phân đoạn video.

Việc rút trích các đặc trưng của khung thường chia làm hai nhóm. Nhóm đầu gồm các đặc trưng dựa vào màu sắc [4] như thành phần màu của các điểm ảnh theo một mô hình màu thích hợp, biểu đồ histogram của cả khung. Nhóm hai gồm các đặc trưng dựa vào các thông tin tổng hợp của khung [4] như hình ảnh các cạnh nổi bật, giá trị entropy, giá trị chuyển đổi cosin rời rạc, giá trị chuyển đổi Fourier....

Việc phân đoạn video có thể làm theo hai cách chính [4] gồm: 1) Tính sự tương đồng giữa các đặc trưng của các khung được chọn. Nếu độ tương đồng nhỏ hơn một ngưỡng thích hợp thì các khung ở hai đoạn khác nhau. Thức đo sự tương đồng thường dùng là các chuẩn L_n (L_n norm). 2) Dùng các đặc trưng của các khung làm dữ liệu vào cho một kỹ thuật phân lớp hay kỹ thuật phân cụm để thu được nhãn đoạn của các khung.

Các kỹ thuật rút trích khung đại diện

Sau khi chia video thành các đoạn gồm các khung có tính tương đồng cao, việc rút trích khung đại diện [4] được làm bằng một trong các cách sau:

- Lấy khung ở trung tâm của đoạn
- Lấy khung chứa nhiều thông tin nhất của đoạn thông qua thước đo entropy.
- Lấy khung đầu tiên của đoạn
- Lấy khung cuối cùng của đoạn

XÂY DỰNG ỨNG DỤNG BIỂU DIỄN VIDEO RÚT GỌN

Lựa chọn kỹ thuật phân đoạn video và kỹ thuật rút trích khung đại diện

Hiện nay, đa số các video được đăng tải trên Internet đều có chất lượng hình ảnh cao cả về

màu sắc và độ nét. Vì vậy, sự khác biệt các thành phần màu của các điểm ảnh có thể phân ánh tốt sự thay đổi về nội dung giữa các khung.

Thông số màu của từng điểm trên khung được thể hiện theo mô hình RGB. Vì vậy, có ba ma trận có cùng kích thước với khung thể hiện giá trị của các màu đỏ, xanh lá cây và xanh lục. Để đo sự tương đồng giữa 2 khung được chọn, thực hiện trừ ma trận màu thể hiện hai khung đó theo chuẩn L_1 . Giả sử, thành phần màu của điểm ảnh i trong khung a là (x, y, z) và trong khung b là (x', y', z') Công thức L_1 dùng cho một điểm ảnh được tính như sau:

$$L_{ab}(i) = |x - x'| + |y - y'| + |z - z'| \quad (1)$$

với $|x|$ là lấy giá trị tuyệt đối của x .

Đối với từng điểm ảnh, nếu tổng sự chênh lệch ba kênh màu của điểm lớn hơn 30 thì coi như là có sự khác biệt nội dung tại điểm đó. Lý do chọn 30 là để tránh việc thay đổi nhỏ của ánh sáng cũng tạo ra sự khác biệt màu và có thể được coi là thay đổi nội dung [4]. Đối với cả khung, nếu số điểm có sự khác biệt nội dung lớn hơn 1/5 số điểm trên toàn khung thì coi như hai khung ở hai phân đoạn khác nhau và rút trích khung đại diện.

Giảm thời gian tính toán

Để giảm thời gian tính toán, chúng tôi lấy các khung ở các vị trí 1, 40, 80, 120,... và khung cuối cùng để rút trích khung đại diện. Số hiệu khung cách nhau 40 giá trị có nghĩa là cứ khoảng hơn 1 giây là kiểm tra xem video có chuyển sang nội dung mới không vì một giây của video được biểu diễn bằng 24 khung. Cách làm cụ thể được mô tả như sau:

- Ban đầu, lựa chọn khung thứ 20 của video làm khung đại diện cho đoạn đầu tiên. Điều này nghĩa là lấy hình ảnh trong 1 giây đầu của video.

- Đầu tiên, xét khung số 20 và khung số 40, nếu hai khung có độ tương đồng cao hơn ngưỡng chọn trước thì kết luận hai khung này ở cùng một phân đoạn. Ngược lại, lấy khung số 40 làm khung đại diện cho phân đoạn 2. Tiếp tục, xét với cặp khung (40, 80), (80, 120),....

- Lặp lại quá trình trên cho đến khi xét đến khung cuối cùng của video.

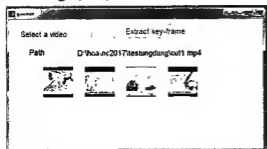
Xây dựng chương trình

Chương trình được viết bằng Matlab. Giao diện của chương trình gồm các thao tác sau:

- Chọn một file video từ thư mục trong máy tính bằng cách bấm vào nút lệnh có dấu ba chấm. Sau khi chọn, đường dẫn đến file sẽ hiện ở bên dưới.

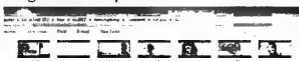
- Chọn nút lệnh **Extract key-frame** để rút trích khung đại diện và hiện kết quả ở bên dưới.

Hình 2 là một giao diện thể hiện kết quả rút trích khung đại diện của một video mẫu.



Hình 2. Kết quả rút trích khung đại diện của một video mẫu

Hình 3 dưới đây là kết quả rút trích khung đại diện của video có tên "Andrea aybar - shot clip" với 7 khung đại diện được hiển thị trong Windows Explorer.



Hình 3. Kết quả rút trích khung đại diện của video có tên "Andrea aybar - shot clip"

KẾT QUẢ THỬ NGHIỆM

Chúng tôi thu thập 50 video trên mạng Internet từ các kênh tin tức của NHK Nhật Bản và CNN Mỹ với nhiều nội dung khác nhau như tin tức, giải trí, thể thao... Các video này đều có 3 đặc trưng của video phổ biến hiện nay như đã trình bày trong phần A về lý do lựa chọn các kỹ thuật dùng trong ứng dụng. Thời lượng các video từ 1 đến 2 phút và dung lượng là từ 0.5Mb đến 16Mb. Để đánh giá hiệu quả việc giảm dung lượng và tốc độ, các thử nghiệm sẽ được làm với các

video ở định dạng MP4 và có dạng chuyển đoạn là trực tiếp.

Kết quả giảm dung lượng lưu trữ

Việc so sánh dung lượng giảm được tính thông qua phép đo dung lượng của Windows Explorer. Dung lượng đầy đủ của video là dung lượng video do Windows Explorer đo được. Dung lượng rút gọn là tổng dung lượng ảnh của các khung đại diện.

Bảng 1 thống kê việc giảm dung lượng lưu trữ tính bằng Mb của một số video dùng trong các thử nghiệm. Tỷ lệ % giảm dung lượng bình quân là từ 85% trở nên. Kết quả cho thấy dung lượng lưu trữ đã giảm rất nhiều. Điều này cho thấy việc lựa chọn kỹ thuật phân đoạn và rút trích khung đại diện cho ứng dụng là thích hợp.

Kết quả giảm thời gian duyệt video

Để việc so sánh tốc độ được chính xác, chúng tôi tính toán thời gian đọc các khung từ file ảnh vào biến. Vì các khung đại diện được lưu ra file ảnh dạng .JPG nên các khung của video được ghi ra file có đuôi .JPG. Dùng lệnh đọc ảnh và cặp lệnh **tic** và **toc** của Matlab để đo thời gian đọc tập ảnh của cả video và thời gian đọc tập ảnh của các khung đại diện.

Vì việc giảm dung lượng sẽ dẫn đến việc giảm tốc độ duyệt video nên chúng tôi chọn 5 video ngắn nhất của hai nhóm video để lấy số liệu về việc giảm thời gian duyệt video. Dữ liệu trong Bảng I cho thấy thời gian đọc các khung đại diện giảm từ 96% trở lên. Kết quả này cho thấy việc rút gọn video bằng các khung đại diện sẽ giảm mạnh thời gian duyệt nội dung video. Điều này cũng minh chứng cho việc lựa chọn kỹ thuật phân đoạn và rút trích khung đại diện cho tập video thử nghiệm là hợp lý và mang lại hiệu quả cao.

KẾT LUẬN

Trong bài báo này, một ứng dụng biểu diễn video rút gọn bằng các khung đại diện được trình bày để giảm dung lượng lưu trữ video trong CSDL dùng cho phân tích và rút trích thông tin có ích từ video. Các thực nghiệm

được làm trên các video phổ biến cho thấy việc lựa chọn kỹ thuật phân đoạn và kỹ thuật rút trích khung đại diện là thích hợp và mang lại hiệu quả tốt. Kết quả thực nghiệm cho thấy dung lượng video giảm nhiều (trên 85%) và giảm đáng kể (trên 96%) thời gian duyệt nội dung video.

Trong thời gian tới, chúng tôi sẽ tiến hành thử nghiệm với CSDL lớn hơn và so sánh hiệu quả với nhiều kỹ thuật rút trích khung đại diện mới.

LỜI CẢM ƠN

Bài báo này là sản phẩm của đề tài có mã số T2017-07-02, được tài trợ bởi kinh phí của trường Đại học Công nghệ Thông tin và Truyền thông.

TÀI LIỆU THAM KHẢO

1. Y. Alper, A. Mubarak, "Shot detection using principal coordinate system", IASTED International Conference on Internet and Multimedia Systems and Applications, 2000.
2. Z. Cernekova, I. Pitas, C. Nikou, "Information Theory-Based Shot Cut/Fade Detection and Video Summarization", IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, Issue 1, pp. 82 – 91, 2005.
3. O. Chum, J. Philbin, M. Isard, A. Zisserman, "Scalable Near Identical Image and Shot Detection", International Conference on Image and Video Retrieval, pp. 549-556, 2007.
4. C. Cotsaces, N. Nikolaidis, I. Pitas, "Video Shot Boundary Detection and Condensed Representation: A Review", IEEE Signal

Processing Magazine, vol. 23, Issue. 2, pp. 28-37, 2006

5. A. Ferman, A. Tekalp, "Two-stage hierarchical video summary extraction to match low-level user browsing preferences," IEEE Trans. Multimedia, vol. 5, no. 3, pp. 244 – 256, June 2003.
6. A. Hanjalic, "Shot-Boundary Detection: Unraveled and Resolved?," IEEE Transtraction on Circuits and Systems for video technology, vol. 12, no. 2, pp. 90-104, 2002
7. C. Kim, J.N. Hwang, "Object-based video abstraction for video surveillance systems," IEEE Trans. Circuits Syst. Video Technol., vol. 12, no. 12, pp. 1128 – 1138, Dec. 2002.
8. C. Kim, J.N. Hwang, "Fast and automatic video object segmentation and tracking for content-based applications," IEEE Trans. Circuits Syst. Video Technol., vol. 12, no. 2, pp. 122 – 129, Feb. 2002.
9. Z. Li, G. M. Schuster, A. K. Katsaggelos, "Minmax optimal video summarization," IEEE Trans. Circuits Syst. Video Technol., vol. 15, no. 10, pp. 1245–1256, Oct. 2005.
10. SV. Porter, M. Mirmehdi, BT. Thomas, "Detection and Classification of Shot Transitions", British Machine Vision Conference, pp. 73-82, 2001.
11. X. Zhu, J. Fan, A. K. Elmagarmid, and X. Wu, "Hierarchical video summarization and content description joint semantic and visual similarity." ACM Multimedia Systems, vol. 9, no. 1, July 2003.
12. T. Vasileios, C. Aristidis, P. Nikolaos, "Scene Detection in Videos Using Shot Clustering and Sequence Alignment", IEEE transtraction on Multimedia, vol. 11, no. 1, 2009, pp. 89-100, 2009.

Bảng 1. Thống kê việc giảm dung lượng và giảm thời gian duyệt của một số video dùng trong thử nghiệm

TÊN VIDEO	Dung lượng đầy đủ (MB)	Dung lượng rút gọn (MB)	Số khung đại diện	Tỷ lệ % giảm dung lượng	Thời gian đọc video đầy đủ (s)	Thời gian đọc video rút gọn (s)	Tỷ lệ % giảm thời gian (s)
Andrea aybar - shot clip	8.27	0.70	7	91.592	87.16	0.80	99.07
Caught on camera Teenager protects brothers during home invasion	8.77	1.28	35	85.405	78.62	0.91	98.83
Distraction- Magician sneezes his head off	7.51	1.13	35	84.953	38.07	0.54	98.58
Hotel worker gives account of Vegas shooting	8.14	0.59	6	92.790	178.55	0.35	99.80
How to make - Corkscrew spin serve + Sidespin serve	7.41	0.47	18	93.595	57.46	0.44	99.22
Abe Planning To Visit US in April	2.02	0.17	9	91.588	13.41	0.18	98.64
Crazy Ping Pong Shot	0.53	0.01	1	97.677	1.50	0.09	93.66
Little Kid Dancing on Marlins JumboTron Fan Cam	1.10	0.25	8	77.006	5.09	0.17	96.62
Mike Tyson pranks Dana White	1.17	0.17	10	85.644	5.52	0.17	96.80
This Lion Cub Trying to Roar is the Cutest Thing You Will See This Week!	1.02	0.09	4	91.278	5.74	0.12	97.84

ABSTRACT

APPLYING KEY-FRAME EXTRACTION FOR STORING AND BROWSING VIDEOS

Nong Thi Hoa*, Nguyen Van Tao

University of Information and Communication Technology - TNU

Nowaday, the number of videos increase sharply based on widely developing of digital devices. Therefore, the need of extracting useful information and analysing content of videos is very essential. To solve these problems, methods of presenting condensed videos by key-frames have been developed to decrease both both the capacity of videos and the time for browsing content of videos. In this paper, an applications of presenting condensed videos by key-frames is presented. Experiments have conducted on a video dataset collecting from NHK and CNN channel to prove the effectiveness of this application. Results show that choosing methods of segmenting videos and extracting key-frames is suitable and drops significantly both the capacity of videos and the time for browsing content of videos.

Keywords: video segmentation, extract key-frame, video analysis, video storage, browse video

Ngày nhận bài: 23/10/2017; Ngày phân biệt: 06/11/2017; Ngày duyệt đăng: 30/11/2017

* Tel: 01238 492484, Email. nongthihoah@gmail.com