

# **ĐẠI HỌC QUỐC GIA TP.HCM**

## **TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

Họ tên học viên cao học: Nguyễn Thành Công

Tên đề tài: Truy tìm frame ảnh trong cơ sở dữ liệu video

Chuyên ngành: Khoa học máy tính

Cán bộ hướng dẫn: TS. Trần Thái Sơn

**ĐẠI HỌC QUỐC GIA TP.HCM**  
**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

Họ tên HVCH: Nguyễn Thành Công

**ĐỀ CƯƠNG NGHIÊN CỨU ĐỀ TÀI**  
**ĐỒ ÁN THẠC SĨ**

Tên đề tài: Truy tìm frame ảnh trong cơ sở dữ liệu video

Chuyên ngành: Khoa học máy tính

Mã số chuyên ngành: 60 48 01 01

Xác nhận của cán bộ hướng dẫn  
(Ký tên và ghi rõ họ tên)

Họ tên:.....

Tp. HCM, tháng 05 năm 2015

## **Tổng quan**

Ngày nay, nhu cầu xử lý thông tin ảnh số và video càng ngày càng gia tăng do lượng thông tin lưu trữ trong đó có thể được khai thác và sử dụng cho nhiều mục đích khác nhau. Một trong số các ứng dụng đó là việc truy vấn ảnh và video trong cơ sở dữ liệu. Truy vấn ảnh và video hiện nay vẫn còn đang là một mảng mới đang được nghiên cứu, đòi hỏi có sự kết hợp giữa các mảng tri thức khác nhau như thống kê, thị giác máy tính, xử lý tín hiệu số, v.v... để giải quyết. Bài toán truy vấn có thể được chia thành nhiều bài toán con. Một trong số các bài toán con đó là bài toán truy tìm ảnh trong cơ sở dữ liệu video. Bằng việc áp dụng các thuật toán phân đoạn, gom nhóm và máy học trên dữ liệu video, ta có thể truy tìm ra các video có chứa frame ảnh gần giống với ảnh mà người dùng đưa vào. Giải quyết được bài toán này sẽ có lợi ích rất lớn khi áp dụng vào thực tế như tìm kiếm các phân cảnh quan trọng trong các bộ phim, tìm ra video quay được từ các camera quan sát cảnh vi phạm giao thông hoặc hành động phạm tội, v.v...

## **Mục đích nghiên cứu**

Đồ án này tập trung vào giải quyết bài toán truy tìm frame ảnh trong cơ sở dữ liệu video. Người dùng đưa vào một ảnh và hệ thống sẽ tìm kiếm trong cơ sở dữ liệu và trả ra kết quả là các video được cho là có chứa các frame ảnh gần giống nhất với ảnh được đưa vào. Ngoài ra, hệ thống còn cho biết phần trăm khả năng video có chứa ảnh đầu vào và vị trí của frame ảnh gần giống với ảnh đầu vào. Dữ liệu video được sử dụng trong đồ án là các video phim được thu thập từ Youtube.

Bài toán có thể được phân rã thành các bài toán con: phân đoạn video thành các shot, tìm kiếm các key frame đại diện cho các shot, rút trích đặc trưng và so khớp ảnh đầu vào với các key frame sử dụng một độ đo hay metric nào đó và xuất ra kết quả là các video đã được xếp hạng theo độ tương đồng hoặc độ dị biệt.

Giải quyết bài toán phân đoạn video thành các shot là bước đầu cơ bản để thực hiện các thao tác khác như đánh chỉ mục video hoặc truy vấn video dựa vào nội dung, v.v... Tác giả Nagasaka [1] đã mô tả phương pháp sử dụng độ chênh lệch màu sắc giữa các pixel tương ứng trong 2 khung ảnh liên kế nhau và phương pháp đặt ngưỡng để xác định thời điểm mà tại đó có sự thay đổi đột ngột trong nội dung của video. Trái ngược với việc sử dụng đặc trưng cục bộ trên

pixel, tác giả Tonomura [2] đã sử dụng lược đồ độ xám để tính độ dị biệt giữa các frame do tính chất toàn cục và các phương pháp dựa trên nền tảng lược đồ độ xám đều rất bền vững với các đối tượng chuyển động trong video quay từ camera. Tuy nhiên, phương pháp loại này đều gặp khó khăn khi có hai khung ảnh có cùng lược đồ độ xám và dẫn đến việc tách thiếu shot. Ngoài ra còn có các phương pháp tách shot video khác như tính độ tương đồng giữa 2 block con có cùng vị trí trong không gian ảnh trong 2 frame liên tiếp nhau (Kasturi [3]) hoặc sử dụng các đặc trưng cấp cao (moment [4], contour [5], v.v...) và các mô hình thống kê (Bayesian [6], Hidden Markov Model [7], v.v...) để nâng cao khả năng tách shot chính xác. Các đặc trưng chuyển động (motion-based) được rút trích trong miền không gian hoặc miền tần số cũng được áp dụng để phát hiện sự thay đổi đột ngột trong video như đặc trưng vector chuyển động [8], Optical Flow [9], v.v...

Tuy video đã được tách thành các shot và thu gọn lại nhưng số lượng frame ảnh vẫn còn rất lớn để có thể thực hiện truy vấn. Bài toán tiếp theo cần được giải quyết là bài toán rút trích key frame từ các shot để đại diện cho một video, giảm chi phí lưu trữ mà vẫn giữ được những thông tin quan trọng trong video. Do đó việc giải quyết bài toán này đóng vai trò chủ chốt trong việc xây dựng hầu hết các hệ thống xử lý video số. Đã có một số phương pháp được đề xuất để giải quyết bài toán này: Ajay [10] cho rằng càng nhiều thông tin chuyển động thì càng cần nhiều key frame để tổng quát hóa. Do đó, tác giả xác định key frame bằng cách chia nhỏ các shot theo thông tin chuyển động và chọn các frame nằm giữa các phân đoạn con đó. Nhằm bắt được nhu cầu sử dụng ảnh hoặc mẫu video nhỏ để truy vấn video trong cơ sở dữ liệu, tác giả Li Zhao, Wei Qi [11] đã tìm ra phương pháp coi các frame trong một shot video như các điểm trong không gian đặc trưng mà trong đó các điểm được nối với nhau bằng các đường cong. Các điểm đứt gãy (gọi là breakpoint) trên quỹ đạo đặc trưng là các key frame và đường thẳng đi qua các điểm breakpoint là biểu diễn của shot video đó. Tác giả trong Xianglin [12] lại đề xuất phương pháp xác định key frame dựa vào nhóm chiếm ưu thế và có thể xác định được số lượng key frames dựa vào độ phức tạp của video với chi phí tính toán thấp. Ngoài ra vẫn có một số phương pháp khác như sử dụng độ đo Hausdorff cải tiến để tính độ phân tán trực tiếp giữa 2 lược đồ của 2 frame cạnh nhau [13] hoặc các phương pháp khác dựa vào nội dung của video và đặc trưng chuyển động, v.v... Như vậy, ta xây dựng được một cơ sở dữ liệu các key frame cho mỗi video để lưu trữ và thực hiện các thao tác rút trích đặc trưng và đối sánh như đối với ảnh.

Ta có thể sử dụng các loại đặc trưng cấp thấp (màu, vân, dáng) hoặc các đặc trưng cấp cao (SIFT, SURF, v.v...) cùng mô hình giỏ đặc trưng và sử dụng các độ đo đối sánh thông dụng như Minkowski, Euclidean, v.v... để đối sánh ảnh đầu vào với các key frame để xác định được video có chứa thông tin tương đồng với ảnh đầu vào.

## **Thời gian thực hiện**

Thời gian thực hiện từ tháng 05/2015 đến tháng 10/2015.

## **Tài liệu tham khảo**

- [1] A. Nagasaka and Y. Tanaka. **Automatic video indexing and full-video search for object appearances**. In IFIP Working Conference on Visual Database Systems, pages 113–127, Budapest, Hungary, October 1991.
- [2] Y. Tonomura and S. Abe. **Content oriented visual interface using video icons for visual database systems**. Journal of Visual Languages and Computing, 1(2):183–198, June 1990.
- [3] R. Kasturi and R.C. Jain. **Dynamic vision**. In R. Kasturi and R.C. Jain, editors, Computer Vision: Principles, pages 469–480. IEEE Computer Society Press, Washington, 1991.
- [4] A. Dailianas, R.B. Allen, and P. England. **Comparison of automatic video segmentation algorithms**. In SPIE Conference on Integration Issues in Large Commercial Media Delivery Systems, volume 2615, pages 2–16, Philadelphia, PA, October 1995.
- [5] R. Zabih, J. Miller, and K. Mai. **A feature-based algorithm for detecting and classifying production effects**. Multimedia Systems, 7(2):119–128, March 1999.
- [6] N. Vasconcelos and A. Lippman. **Statistical models of video structure for content analysis and characterization**. IEEE Transactions on Image Processing, 9(1):3–19, January 2001.

- [7] S. Eickeler and S. Müller. **Content-based video indexing of tv broadcast news using hidden markov models**, International Conference on Acoustics, Speech, and Signal Processing, pages 2997–3000, Phoenix, AZ, March 1999.
- [8] A. Akutsu, Y. Tonomura, H. Hashimoto, and Y. Ohba. **Video indexing using motion vectors**, SPIE Conference on Visual Communications and Image Processing, volume 1818, pages 1522–1530, Boston, MA, November 1992.
- [9] O. Fatemi, S. Zhang, and S. Panchanathan. **Optical flow based model for scene cut detection**, Canadian Conference on Electrical and Computer Engineering, volume 1, pages 470–473, Calgary, Canada, May 1996.
- [10] Ajay divakaran, Regunathan Radhakrishnan, Kadir A Perker, **Motion activity based extraction of key frames from video shots**, IEEE 2002.
- [11] Li Zhao†\_, Wei Qi\*, Stan Z. Li\*, Shi-Qiang Yang†\_, H. J. Zhang\*, **Key frame extraction and shot retrieval using nearest feature line**, Microsoft Research China.
- [12] Xianglin Zeng, WeimingHu, Wanqing Liy, Xiaoqin Zhang, Bo Xu, **Key-frame extraction using dominant set clustering**, SCSSE, University of Wollongong, Australia.
- [13] Sang Hyun Kim, Rae-Hong Park, **An effective algorithm for video sequence matching using the modified Hausdorff distance and the directed divergence**, IEEE transactions on circuits and systems for video technology, vol. 1.