

# Các kỹ thuật chỉ số hóa và truy vấn dữ liệu Video

Nguyễn Đình Hóa  
[hoand@ptit.edu.vn](mailto:hoand@ptit.edu.vn) 0942807711

# Tóm tắt nội dung bài 12

---

- ▶ Chỉ số hóa và truy vấn dữ liệu âm thanh
  - ▶ Các thuộc tính miền thời gian
    - ▶ Năng lượng trung bình
    - ▶ Tốc độ đổi dấu của tín hiệu
    - ▶ Phần trăm của khoảng lặng
  - ▶ Các thuộc tính miền tần số
    - ▶ Phổ âm thanh
    - ▶ Băng thông
    - ▶ Phân bố năng lượng âm thanh
    - ▶ Độ điều hòa âm
    - ▶ Độ cao thấp của âm thanh
  - ▶ Âm phổ
  - ▶ Phân loại âm thanh
    - ▶ Phân loại theo từng bước
    - ▶ Phân loại theo vector thuộc tính
  - ▶ Nhận dạng và truy vấn âm thoại
  - ▶ Lưu trữ và truy vấn âm nhạc

# Đánh chỉ số và truy vấn video

---

- ▶ Dựa trên các thuộc tính chung (tiêu đề, tên tác giả, ngày sản xuất, thể loại,...): giống như các hệ CSDL thông thường (mô hình dữ liệu quan hệ).
- ▶ Dựa trên văn bản (phụ đề, thông tin miêu tả,...) kèm theo các nhãn thời gian. Truy vấn bằng hệ thống IR
- ▶ Dựa trên âm thanh đi kèm với video, có thêm thông tin về nhãn thời gian. Truy vấn giống với tín hiệu âm thanh.
- ▶ Dựa trên nội dung:
  - ▶ Cui video là một chuỗi các ảnh tĩnh. Quá trình truy vấn dựa trên truy vấn ảnh. Nhược điểm: mất liên kết thời gian giữa các khung hình, có thể phải xử lý quá nhiều ảnh.
  - ▶ Chia tệp video thành các đoạn nhỏ chứa các khung hình giống nhau. Truy vấn dựa trên các ảnh đại diện cho mỗi đoạn video.
- ▶ Phương pháp kết hợp.

# Đánh chỉ số và truy vấn dựa trên phân đoạn

---

- ▶ Video là chuỗi liên tục của các ảnh tĩnh (khung hình – frame)
- ▶ Video có thể chia nhỏ thành các mảnh (segments, shots) bao gồm chuỗi nhiều ảnh tĩnh liên tiếp có liên quan với nhau:
  - ▶ Mô tả về cùng một cảnh (scene)
  - ▶ Chứa thông tin về một thao tác của máy quay
  - ▶ Chứa thông tin về một hành động, sự kiện cụ thể của một thực thể nào đó
  - ▶ Các khung hình được lựa chọn như một công cụ duy nhất để đánh chỉ số và tra cứu.

# Đánh chỉ số và truy vấn dựa trên phân đoạn

---

## ► Các bước chính:

- Phân đoạn video (segmentation, shot detection, partition)
- Đánh chỉ số từng đoạn video: xác định khung hình đại diện của từng đoạn, áp dụng các phương pháp đánh chỉ số và truy vấn ảnh cho các khung hình này.
- Truy vấn dựa trên so sánh tương tự giữa câu truy vấn (hình ảnh hoặc thông tin về khung hình nào đó) với các khung hình đại diện của các đoạn video.

# Phân đoạn video (shot detection/segmentation)

---

- ▶ Các khung hình (frame) trong cùng một đoạn (lần bấm máy) thường giống nhau về một mặt nào đó
- ▶ Nếu có một sự khác nhau giữa hai khung hình liên tiếp (quá một ngưỡng cho trước) thì hai khung hình đó thuộc hai đoạn (lần bấm máy) khác nhau
- ▶ VD:
  - ▶ Chuyển cảnh giữa các máy quay (chuyển bất ngờ)
  - ▶ Các thao tác trên máy quay: vào hình, ra hình... (chuyển cảnh từ từ)

# Phân đoạn video

---

## ► Các kỹ thuật cơ bản:

- Tìm tổng tất cả các điểm ảnh khác nhau (tương ứng vị trí) giữa hai khung hình liên tiếp. So sánh với ngưỡng.
- Tìm sự khác nhau giữa các biểu đồ tần suất màu của hai khung hình. So sánh với ngưỡng.

$$SD_i = \sum_j |H_i(j) - H_{i+1}(j)|$$

- Dựa trên kiểm tra  $\chi^2$  ( $\chi^2$  test)

$$SD_i = \sum_j \frac{(H_i(j) - H_{i+1}(j))^2}{H_{i+1}(j)}$$

Vấn đề đặt ra là chọn các giá trị ngưỡng phù hợp.

# Phân đoạn video

---

- ▶ Phân đoạn với khung hình biến đổi từ từ
  - ▶ Sử dụng hai ngưỡng, một ngưỡng để bắt cảnh thay đổi đột ngột, một ngưỡng (thấp hơn) để bắt cảnh thay đổi từ từ.
    - ▶ Ngưỡng cao  $T_1$  được dùng để xác định cảnh thay đổi đột ngột
    - ▶ Ngưỡng thấp  $T_2$  dùng để đánh dấu khả năng cảnh thay đổi từ từ. Nếu tổng sự khác nhau giữa  $X$  khung hình liên tiếp vượt ngưỡng  $T_1$  thì xác định sự đổi cảnh.



# Phân đoạn video

- ▶ Giảm thiểu sai sót khi phân đoạn:
  - ▶ Phân biệt sự chuyển cảnh với các thao tác phóng to, thu nhỏ trên máy quay, hoặc thao tác lia hình bằng máy quay: sử dụng thuộc tính “dòng quang” (optical flow)
  - ▶ Phân biệt chuyển cảnh với sự thay đổi cường độ sáng của phim (do các hiệu ứng ánh sáng, hoặc do thay đổi nguồn sáng trong phim): chuẩn hóa các giá trị màu
    - ▶ Chuẩn hóa các giá trị màu thành phần:

$$R'_i = \frac{R_i}{\sqrt{\sum_{k=1}^N R_k^2}}, \quad G'_i = \frac{G_i}{\sqrt{\sum_{k=1}^N G_k^2}}, \quad B'_i = \frac{B_i}{\sqrt{\sum_{k=1}^N B_k^2}}$$

- ▶ Chuyển đổi sang độ màu:

$$r_i = \frac{R'_i}{R'_i + G'_i + B'_i}, \quad g_i = \frac{G'_i}{R'_i + G'_i + B'_i}$$

- ▶ Biểu đồ tần suất màu tổng hợp của r và g được xây dựng.

# Đánh chỉ số và truy vấn video dựa trên phân đoạn

---

- ▶ Mỗi đoạn video được đại diện bởi một hoặc một vài khung hình
- ▶ Các khung hình được coi là các ảnh tĩnh, việc đánh chỉ số và truy vấn giống như đối với ảnh
- ▶ Trong quá trình truy vấn, so sánh tương tự được thực hiện giữa nội dung truy vấn và các hình đại diện.
- ▶ Có nhiều cách để chọn khung hình đại diện cho từng đoạn video

# Đánh chỉ số và truy vấn dựa trên khung hình đại diện đoạn video

---

- ▶ Các cách chọn số khung hình đại diện cho từng phân đoạn video
  - ▶ Chọn một khung hình duy nhất đại diện cho từng phân đoạn video. Nhược điểm: mất thông tin về độ dài thời gian cũng như nội dung thay đổi trong đoạn video đó.
  - ▶ Chọn nhiều khung hình đại diện cho đoạn video, số khung hình dựa vào độ dài đoạn video đó. Thông thường cứ mỗi giây video chọn một khung hình.
  - ▶ Mỗi phân đoạn video lại được chia nhỏ thành các mảnh chứa các nội dung/cảnh khác nhau dựa trên vector chuyển động, dòng quang, hoặc sự khác biệt giữa các khung hình liên tiếp nhau. Một khung hình đại diện được chọn cho mỗi mảnh video.

# Đánh chỉ số và truy vấn dựa trên khung hình đại diện đoạn video

---

- ▶ Các cách chọn khung hình đại diện cho từng đoạn/mảnh video
  - ▶ Chọn khung hình đầu tiên của mỗi đoạn/mảnh video
  - ▶ Dựa trên khung hình trung bình của mỗi đoạn/mảnh video, trong đó mỗi điểm ảnh chứa giá trị trung bình giá trị màu của tất cả các điểm cùng vị trí trong các khung hình thuộc đoạn đó. Khung hình nào giống nhất với khung hình trung bình sẽ được chọn làm đại diện cho đoạn/mảnh video đó
  - ▶ Dựa trên biểu đồ tần suất màu trung bình của tất cả các khung hình trong đoạn đó. Khung hình nào có biểu đồ tần suất màu gần với biểu đồ trung bình thì sẽ được chọn.
  - ▶ ...

# Đánh chỉ số và truy vấn video

---

- ▶ Đánh chỉ số và truy vấn dựa trên thông tin về chuyển động: thông tin chuyển động được lấy từ dòng quang, hoặc các vector chuyển động.
  - ▶ Nội dung của chuyển động. VD: đoạn video hội thoại có ít nội dung chuyển động hơn đoạn phim hành động.
  - ▶ Tính đồng nhất của chuyển động: nói lên tính liên tục của các chuyển động
  - ▶ Chuyển động lia hình (theo chiều ngang)
  - ▶ Chuyển động nghiêng (theo chiều dọc)
- ✓ Các khung hình được đánh chỉ số và truy vấn dựa trên cả thông tin về nội dung lẫn thông tin về chuyển động

# Đánh chỉ số và truy vấn video

---

- ▶ Đánh chỉ số và truy vấn dựa trên các thực thể
  - ▶ Các thực thể trong video có thể được xác định dựa trên các điểm ảnh di chuyển giữa các khung hình liên tiếp
  - ▶ Các thực thể có thể được dùng để phân đoạn video và truy vấn video
- ▶ Đánh chỉ số và truy vấn video dựa trên siêu dữ liệu sử dụng hệ CSDL thông thường
- ▶ Đánh chỉ số và truy vấn video dựa trên nội dung văn bản sử dụng hệ thống IR
- ▶ Các phương pháp kết hợp.