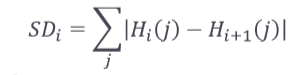
1. **Kỹ thuật xử lý video**
2. **Đánh chỉ số và truy vấn video**

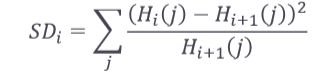
* Dựa trên các thuộc tính chung (tiêu đề, tên tác giả, ngày sản xuất, thể loại,…): giống như các hệ CSDL thông thường (mô hình dữ liệu quan hệ).
* Dựa trên văn bản (phụ đề, thông tin miêu tả,…) kèm theo các nhãn thời gian. Truy vấn bằng hệ thống IR.
* Dựa trên âm thanh đi kèm với video, có thêm thông tin về nhãn thời gian. Truy vấn giống với tín hiệu âm thanh.
* Dựa trên nội dung:
  + Coi 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.
* 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.
* 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.

1. **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).
* 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 mầu của hai khung hình. So sánh với ngưỡng.



* + Dựa trên kiểm tra 𝜒2 (𝜒2 test)



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

* 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 T1 được dùng để xác định cảnh thay đổi đột ngột.
    - Ngưỡng thấp T2 dùng để đánh dấu khả năng cảnh thay đổi từ từ. Nếu tổng sự khác nhau giữa các khung hình liên tiếp vượt ngưỡng T1 thì xác định sự đổi cảnh.
* 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 thác 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 mầu thành phần:



* + - Chuyển đổi sang độ mầu:



* + - Biểu đồ tần suất mầu tổng hợp của r và g được xây dựng.

1. **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.

1. 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 đoại 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.

1. 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 mầu 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 mầu trung bình của tất cả các khung hình trong đoạn đó. Khung hình nào có biểu đồ tần suất mầu gần với biểu đồ trung bình thì sẽ được chọn.

1. 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.

1. 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.

1. **Cây phân đoạn khung hình (frame segment tree)**
2. **Mô tả đầy đủ về cơ sở dữ liệu video**

Đầu tiên, chúng tôi giả định rằng OBJ là một tập hợp các đối tượng video và EVT là một tập hợp các sự kiện được tìm thấy trong cơ sở dữ liệu video. Do đó, ENT = OBJ∪EVT (ENT là tập hợp của tất cả các thực thể). Bản đồ liên kết λ ánh xạ các phần tử của ENT vào các tập hợp các chuỗi khung. Đối với mỗi sự kiện, ∈ EVT, ánh xạ nhiều-một ℵ ánh xạ mỗi sự kiện thành một loại hoạt động. Ví dụ: trong ví dụ phim của chúng tôi, (e1) = giết người. Tiếp theo, chúng tôi giả định rằng, với một loại hoạt động A, một ROLE (A) được thiết lập được liên kết với nó. Các thành viên của bộ ROLE (A) là các chuỗi biểu thị tên của các vai trò liên quan đến hoạt động A. Ví dụ: ROLE (giết người) = {nạn nhân, kẻ giết người, kẻ giết người} và ROLE ({đưa ra một bữa tiệc} = { máy chủ, khách}. Với bất kỳ sự kiện E nào thuộc loại hoạt động A, chúng tôi giả sử rằng có một PLAYERS ánh xạ ánh xạ ROLE (A) tới tập OBJ ∪ EVT ∪ STR; STR là tập hợp tất cả các chuỗi có thể. bản đồ PLAYERS chỉ định nhóm tham gia vào sự kiện này.

Định nghĩa 1. Cơ sở dữ liệu video là 9-tuple  
(RVD, OBJ, EVT,λ, ACT, R,,℘,ROLE,PLAYERS)  
trong đó - RVD là một bộ số nguyên {1, ..., n} cho sone n.  
- OBJ là một tập hợp các đối tượng. - EVT là một tập hợp các sự kiện.  
- là bản đồ liên kết gán một tập hợp các chuỗi khung vững chắc cho mỗi thực thể ent (OBJ∪EVT).  
- ACT là một tập hợp các loại hoạt động.  
- R là tập hợp các vai trò.  
- ℘ là bản đồ chỉ định loại hoạt động cho từng sự kiện.  
- ROLE là bản đồ lấy chuỗi khung [i, j) ∈ EVT∪So∈OBJ λ (o), một hoạt động A trong ℘([i, j)) và vai trò trong R (A) làm đầu vào và chỉ định một thành viên của OBJ làm đầu ra.  
- PLAYERS là một bản đồ lấy một sự kiện và loại hoạt động của nó làm đầu vào và trả về ánh xạ từ các vai trò của hoạt động đến các thực thể trong cơ sở dữ liệu và thành chuỗi làm đầu ra.

1. **Cấu trúc dữ liệu**

Chúng ta định nghĩa các cấu trúc dữ liệu cần thiết để lưu trữ dữ liệu video mà chúng ta quan tâm để xử lý. Có bốn loại cấu trúc cần được định nghĩa: cây phân đoạn khung cho phép chúng tôi chỉ định các sự kiện nào (và các trình phát, đội, v.v. có liên quan của chúng) và các đối tượng video xảy ra trong các phân đoạn của video; OBJECTARRAY cho phép truy cập các phân đoạn video bằng cách sử dụng các đối tượng làm khóa, ACTIVITYARRAY cho phép truy cập các phân đoạn video bằng cách sử dụng các loại hoạt động làm khóa và EVENTARRAY tạo điều kiện truy cập bằng các sự kiện làm khóa. Trong số này, thành phần chính là một loại cây phân đoạn đặc biệt.

* 1. **Cây phân đoạn khung hình**

Chúng ta định nghĩa một loại cây phân đoạn đặc biệt gọi là cây phân đoạn khung hinh được sử dụng cho dữ liệu đa phương tiện. Cấu trúc của các nút trong cây phân đoạn khung hình có thể được định nghĩa như sau. Lưu ý rằng trong cấu trúc dữ liệu này, trình phát bị giới hạn ở các đối tượng video. Nó là đơn giản để tăng cường các cấu trúc dữ liệu để cho phép người chơi thuộc bất kỳ loại nào.

**type treenode = record of**

**start : integer; /\* starting frame number \*/**

**finish : integer; /\* ending frame number \*/**

**objlist : ^objnode ; /\* list of objects associated with the node \*/ evtlist : ^evtnode; /\* list of activities associated with the node \*/**

**lchild : ^treenode ; /\* left child \*/ rchild : ^treenode; /\* right child \*/**

**end;**

**type objnode = record of**

**objid : integer ; /\* object number \*/**

**next : ^objnode; /\* next node in list \*/**

**end;**

**type evtnode = record of**

**evtid : integer; /\* activity number \*/**

**next : ^evtnode; /\* next node in activity list \*/**

**end;**

* 1. **Các mảng khác**

OBARARRAY, ACTIVITYARRAY và EVENTARRAY có thể được định nghĩa là các mảng với các phần tử có cấu trúc bản ghi tương ứng là objectrec, activityrec và eventrec. Chúng tôi giả sử rằng Val1, Val2, Val3, Val4 là các hằng số.

**type objectrec = record name : string; /\* name of object\*/ frames : ^frameseqlist;**

**/\* pointer to list of tree nodes \*/**

**end;**

**type activityrec = record**

**name : string;**

**events : ^evtnode;**

**end;**

**type eventrec = record**

**acttype: integer;**

**teamlist: ^playernode;**

**frames : ^frameseqlist;**

**end;**

**type playernode = record of**

**player: integer; /\* object number \*/**

**role : integer; /\* role number \*/**

**next : ^playernode; /\* next player in the team \*/**

**end;**

**type frameseqlist = record**

**segments : ^treenode; /\* points to a tree node \*/**

**next : ^frameseqlist; /\* next node in list \*/**

**end;**

**eventarray : array[1..Val1] of eventrec;**

**activityarray : array[1..Val2] of activityrec;**

**objectarray : array[1..Val3] of objectrec;**

**rolearray: array[1..Val4] of string;**

Với cấu trúc dữ liệu cây phân đoạn khung hình này, chúng có thể tiếp tục với:

1. Danh sách các truy vấn cơ bản có thể được đặt ra cho hệ thống
2. Phát triển các thuật toán để thực hiện các truy vấn này.
3. Phân tích độ phức tạp của các thuật toán cho mỗi truy vấn.
4. Kết xuất chính thức các thuật toán để cập nhật cấu trúc dữ liệu một cách hiệu quả khi các thực thể mới trong các video hiện có trở nên quan trọng.