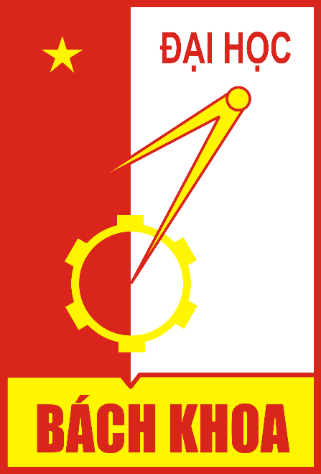
TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

─────── \* ───────



**BÀI TẬP LỚN**

MÔN: Hệ Cơ Sở Tri Thức – IT4361

**Đề Tài: Phân đoạn video đơn giản sử dụng đặc trưng toàn cục**

|  |  |  |
| --- | --- | --- |
| **Nhóm** | | 7 |
| **Thành viên** | Nguyễn Huy Phát  Nguyễn Hữu Phú  Trần Gia Nghĩa | |
| **Mã lớp:** | | 97544 – K59 |
| **Giảng viên hướng dẫn:** | | **TS.Nguyễn Thị Oanh** |

***Hà Nội, tháng 12 năm 2017***

**Lời nói đầu**

Hiện nay, phần lớn thông tin mà con người thu nhận được từ thế giới bên ngoài là thông qua thị giác, vì vậy những nghiên cứu về xử lý dạng thông tin này đều ít nhiều mang lại những lợi ích thiết thực cho cuộc sống. Với trình độ khoa học hiện nay, thông tin dạng này chủ yếu được số hóa dưới dạng video.

Trong quá trình tìm kiếm lời giải cho các bài toán xử lý video (ví dụ như truy vấn video, …) thì bài toán phân đoạn video xuất hiện một cách tự nhiên. Để có thể truy vấn video 1 cách hiệu quả, thì video phải được phân đoạn một cách hợp lý và thống nhất về mặt ngữ nghĩa. Sự phân đoạn bằng tay là cách thủ công và tốn nhiều thời gian, phụ thuộc vào kĩ năng và sự hiểu biết của người thực hiện.

Chính vì vậy, việc nghiên cứu về phân đoạn video một cách tự động là rất cần thiết và ngày càng trở nên cấp thiết hơn khi dữ liệu video ngày càng gia tăng. Và đó cũng là đề tài mà nhóm tôi nghiên cứu trong môn học này “phân đoạn video đơn giản bằng sử dụng đặc trưng toàn cục”.

**Phần I: Khái quát về video và bài toán phân loại video**

1. **Khái quát về video**

Video là sự tái tảo ảnh tự nhiên theo không gian hoặc thời gian hoặc là cả 2, thực chất là một dãy ảnh liên tục theo thời gian nhằm mô tả sự chuyển động.

Video là một tập hợp các khung hình (frame) mỗi khung hình là 1 ảnh. Shot là đơn vị cơ sở của 1 video. Một shot là 1 đơn vị vật lý của 1 video, gồm chuỗi các khung hình liên tiếp, ứng với 1 thao tác camera đơn.

Khi video được chiếu, các khung hình được hiển thị lần lượt với 1 tốc độ nhất định. Tốc độ thường thấy ở các video là 25 – 30 hình / giây.

Chuyển cảnh là một chuyển đổi đối với người xem từ shot này sang shot khác. Có 2 loại chuyển cảnh : chuyển cảnh đột ngột (không liên tục) và chuyển cảnh dần dần (liên tục)

1. **Bài toán phân loại video**

Các kĩ thuật phân đoạn chủ yếu dựa vào các khung hình của video. Vì vậy việc phân đoạn video cũng phụ thuộc nhiều vào các kĩ thuật xử lý ảnh.

Chúng đo sự khác nhau giữa các khung hình của video để phát hiện các biên của các shot.

Nhìn chung, hiệu quả của việc phân đoạn dựa trên các yếu tố sau :

* Các đặc tính để so sánh
* Các phương pháp so sánh
* Đối tượng video thực hiện phân đoạn

Và nhìn chung thì việc phát hiện chuyển cảnh từ từ khó hơn là phát hiện chuyển cảnh đột ngột

**Phần II: Phương pháp thực hiện phân đoạn video**

Có nhiều phương pháp phân đoạn video tương ứng với nhiều kí thuật xử lý ảnh khác nhau:

* Phân đoạn video dựa trên đặc trưng cục bộ
* Phân đoạn video dựa trên đặc trưng toàn cục

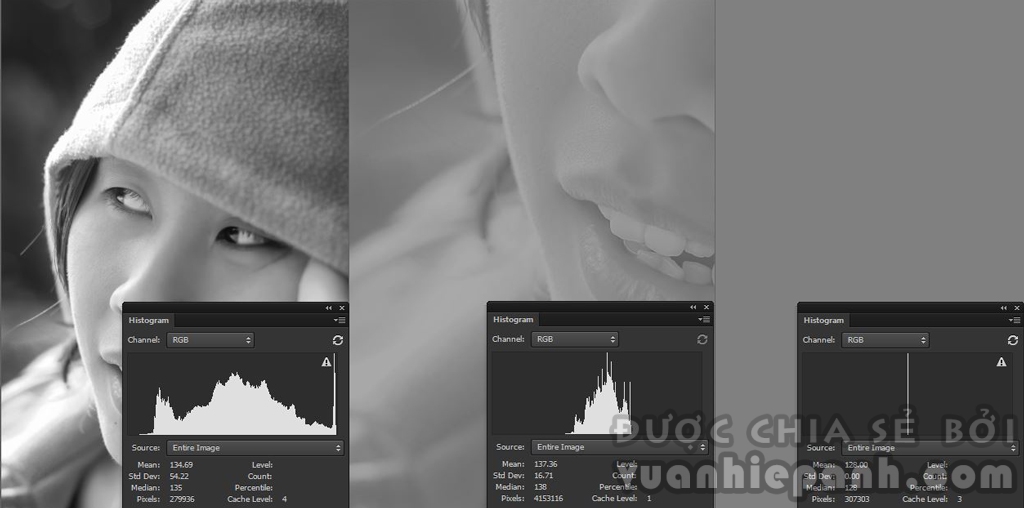
Bài báo cáo sẽ tập trung vào việc phân đoạn video bằng đặc trưng toàn cục, cụ thể ở đây là phân đoạn bằng histogram màu.

1. **Histogram là gì ?**

Histogram không hề xa lạ với những người chụp ảnh và xử lý ảnh số, ta có thể thấy Histogram trên máy ảnh, trong Photoshop (Ps), Camera RAW (ACR), Lightroom (Lr) hay bất kỳ phần mềm xư lý ảnh số nào.

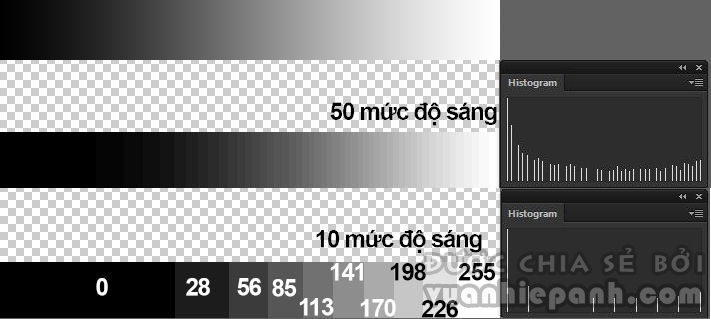
Một bức ảnh được tạo thành nhờ những điểm ảnh có mức độ sáng tối khác nhau, và càng có nhiều mức độ sáng tối càng có nhiều chi tiết trong ảnh, nếu tất cả các điểm ảnh đều có cùng một độ sáng, ảnh sẽ không có một chi tiết nào hết.





Có thể dễ dàng nhận thấy sự khác nhau về độ rộng của ba Histogram này. Histogram càng rộng (trải dài) ảnh càng chi tiết và ngược lại, càng hẹp thì càng ít chi tiết.

**Đọc hiểu Histogram**  
  
Nhìn vào một Histogram, sẽ thấy một phần giống như đồi núi nhấp nhô nằm trong một ô hình chữ nhật. Ta tạm gọi hình chữ nhật có chiều thẳng đứng là trục Y, chiều ngang là trục X. Chiều cao của phần đồi núi tương ứng với trục Y biểu thị số lượng điểm ảnh (Pixel), chiều dài của phần đồi núi tương ứng với trục X biểu thị dải sáng tối của bức ảnh (Tonal Range). Phần trục X biểu thị độ rộng của Histogram được chia thành 256 mức và được đánh số từ 0 đến 255 (0, 1, 2 …, 254, 255), điểm đầu 0 tương ứng với màu đen hoàn toàn, điểm cuối 255 tương ứng với trắng hoàn toàn, ở 2 đầu ảnh ảnh đều không có chi tiết mà chỉ là đen hoặc trắng.



Có thể thấy với hình càng ít mức độ sáng thì Histogram càng bị khuyết nhiều, ở hình dưới cùng mình đánh số của độ sáng tương ứng. Điểm 0 đen hoàn toàn nằm ở gốc bên trái Histogram, và có nhiều điểm ảnh có giá trị 0 (vùng đen rộng) nên điểm này cao nhất, các điểm còn lại tương tụ có chiều cao tương ứng với phần diện tích mà nó chiếm. Nhìn vào hình ở giữa với 50 mức độ sáng các bạn đã thấy có thể phân biệt được các vùng sáng khác nhau, hình dưới cùng có 10 mức thì phân biệt rõ ràng, hình trên cùng đủ dải từ 0 đến 255 nhìn mịn và mượt nhất. Như vậy nếu Histogram càng đều, liên tục thì độ chuyển sáng tối trong hình càng mượt đẹp.

1. **Phương pháp phân đoạn video bằng cách sử dụng histogram :**

* Tìm khoảng cách của 2 frame kế tiếp sử dụng histogram màu. Khoảng cách ở đây có thể là khoảng cách Euclid, khoảng cách cosine,… Trong project thì nhóm tôi sử dụng khoảng cách euclide.
* Sử dụng ngưỡng để xác định vị trí có sự khác biệt lớn để xác định biên của các shot. Cách xác định ngưỡng thì có thể là cài đặt cứng nhưng nhìn chung là không hiệu quả vì các video khác nhau, các cảnh khác nhau thì có các ngưỡng khác nhau. Vì vậy nhóm đề xuất các xác định ngưỡng như sau:
  + Tìm tất cả các khoảng cách giữa histogram của 2 frame liền nhau
  + Chọn ra số lượng các shot mà người dùng mong muốn chia video ra (giả sử là n)
  + Lấy n khoảng cách lớn nhất trong các khoảng cách đã tính ở trên. Nơi có n khoảng cách này là biên của các shot.
* Cách lấy keyframe của mỗi shot :
  + Tính histogram của mỗi khung hình trong shot đang xét
  + Lấy trung bình của các histogram này
  + Tìm khung hình có histogram gần với giá trị histogram trung bình nhất