**Báo cáo tiến độ công việc**

Sinh viên: ***Nguyễn Văn Toàn.***

1. **Các vấn đề đã đọc.**

* Các khái niệm, đặc trưng trong xử lý ảnh số và video và cấu trúc file video
  + ***Frame:*** các video được tạo nên từ nhiều ảnh tĩnh đặt liên tiếp nhau tạo nên chuyển động. Các ảnh tĩnh này gọi là frame.
  + Để tạo cảm giác chuyển động thì các video phải được quay với tốc độ 24frame/giây.
  + ***Đoạn cơ sở(shot):*** là một chuỗi các frame đặt liên tiếp nhau để mô tả hành động liên tục, được giới hạn bởi hai chuyển cảnh.
  + ***Chuyển cảnh*** là đường biên phân chia các đoạn cơ sở. có chuyển cảnh tức thì và chuyển cảnh tiệm tiến.
  + ***Cảnh video:***  là một tập các shot kề nhau (video scene). Có hai loại là physical boundaries và sematic boundaries.
  + ***Video group:*** nằm giữa physical boundaries và sematic boundaries.
* Biên và một số phương pháp phát hiện biên.(phương pháp gradient)
  + Biên là nơi có độ biến thiên về cường độ sang.
  + Có nhiều phương pháp phát hiện biên: gradient, laplace…
  + *Em đọc qua các công thức gradient nhưng chưa hiểu là sẽ làm như thế nào trong thực tế.*
* Đặc trưng chuyển động.
  + Là đặc trưng của video mà ảnh tĩnh không có.
  + Lược đồ chuyển động:
    - Phân phối chuyển động của các frame liên tiếp nhau.
    - Nó được thể hiện dưới các góc chuyển động.
    - Thuật toán tính lược đồ chuyển động:
      * Với mỗi khung hình định lượng các góc từ 0 – 360 thành 8 phần bằng nhau.
      * Bước 1: khởi tạo mảng các góc đã định lượng H[i] = 0 với *i = 0-7.*
      * Bước 2: xét một khối điển ảnh của frame hiện tại và tính độ dịch chuyển của nó bằng cách: trong frame tiếp theo, tìm khối có sự khác biệt đặc trưng nhỏ nhất so với khối đang xét, và sự khác biệt này cũng nhỏ hơn một ngưỡng định trước. Mục đích là để xem khối này dịch chuyển đến vị trí nào. Nếu không tìm thấy thì coi như không có sự di chuyển.
      * Bước 3: sau khi tính độ di chuyển, ta tính góc dịch chuyển và convert về giá trị a, với a nằm trong khoảng 0-7.
      * Bước 4: tang giá trị của H[a] = H[a]+1 cho đến khi tính được hết các khối điểm ảnh của khung hình.
* Cho ta biết hướng chuyển động của camera.
* Biên đoạn cơ sở.
  + Phát hiện những chuyển cảnh giữa các đoạn trong video. Xác định frame bắt đầu và frame kết thúc của đoạn cơ sở.
  + Ý tưởng để phân đoạn cơ sở là:
    - Rút trích đặc trưng của các frame trong video.
    - Tính toán sự khác biệt giữa các đặc trưng của các frame liên tiếp.
    - So sánh sự khác biệt này với ngưỡng cho trước. Nếu vượt ngưỡng thì phát hiện được đoạn cơ sở.

1. **Các vấn đề định hướng.**

* Trích chọn key frame: các key frame được chọn yêu cầu cần phải có khả năng biểu diễn nội dung của toàn bộ video. Một cách tiếp cận ban đầu là mỗi frame ban đầu của mỗi shot được chọn làm key frame. Tuy nhiên, có những thay đổi trong shot do chuyển động của máy quay hoặc đối tượng => cần tang số lượng các key frame trong một shot. Nếu độ dài nhỏ hơn 1 giây thì frame ở giữa sẽ được chọn. Nếu độ dài của shot đó dài hơn thì mỗi giây sẽ chọn ra key frame.
* Trích chọn đặc trưng của frame.
  + Đặc trưng màu sắc: mỗi điểm ảnh được biểu diễn như một điểm trong không gian màu sắc 3 chiều. Các không gian màu sắc thường dùng là: RBG, HSV, YUV, CMYK…
  + Đặc trưng kết cấu: đưa ra mô hình trực quan của ảnh và cách thức chúng được xác định trong không gian.
  + Đặc trưng hình dạng:
* phương pháp phát hiện biên đoạn cơ sở.
* Các phương pháp phát hiện biên chuyển cảnh giữa các shot.
* Phương pháp biểu diễn video.
  + Biểu diễn các key frame liên tiếp: điều kiện là đã có các shot và key frame. Các key frame được biểu diễn tuần tự từ trên xuống, từ trái qua phải. Kĩ thuật này chỉ hiệu quả với đoạn video ngắn.
  + Biểu diễn dựa vào nhóm(group based representation): ý tưởng ở đây là các shot có liên quan được gộp lại thành một group. Ban đầu chia dòng video(video stream) thành nhiều đoạn, mỗi đoạn chứa các shot liên tiếp bằng nhau. Mỗi đoạn lại được chia thành nhiều đoạn con (sub-segments) => xây dựng cây cấu trúc biểu diễn video.
  + Biểu diễn video dựa vào cảnh(scene):
  + Video Mosaic representation: ý tưởng là các thông tin hỗn tạp trong 1 shot được chia làm 3 thành phần:
    - Extended spatial information: ghi lại hiện trạng của toàn bộ nền trong shot và đc biểu diễn dưới dạng form mẫu.
    - Extended temporal information: ghi lại sự chuyển động của các vật thể độc lập(dynamic mosaic)
    - Geometric information: ghi lại sự biến đổi hình học qua sự chuyển động của camera.