

CHUẨN BỊ DỮ LIỆU TRONG VIỆC XÂY DỰNG CƠ SỞ DỮ LIỆU VIDEO SỐ

LƯƠNG XUÂN CƯƠNG¹, ĐỖ TRUNG TUẤN², ĐỖ XUÂN TIẾN¹

¹Học viện Kỹ thuật Quân sự, Email: lxcuong2002@yahoo.com; dxtienvxl@yahoo.com

²Đại học Quốc gia Hà Nội, Email tuandt@hep.edu.vn

Abstract. Preparing video data is an important step in the construction of digital video databases. This work demands segmenting a video sequence into shots, which are basic meaning units of video data. In this paper, a method to prepare video data is proposed consisting of a shot boundary detection technique on the basis of tracking camera motion through the background areas to the video frame (background tracking) and of some other steps to discard noise and to reduce false segmentation for increasing the accuracy of the video stream segmentation. Measures to be taken to overcome the problem of object's and camera's motion are proposed and an enhanced algorithm for segmenting video data is also reported.

Tóm tắt. Chuẩn bị dữ liệu video là một bước quan trọng trong quá trình xây dựng cơ sở dữ liệu (CSDL) video số. Nó luôn yêu cầu phân chia chuỗi video thành các lia (shot), là đơn vị cơ bản có nghĩa của dữ liệu video. Bài báo đề xuất một phương pháp chuẩn bị dữ liệu video, bao gồm: kỹ thuật phát hiện gianh giới giữa các lia video dựa trên đánh dấu vùng nền của các khung video, một số biện pháp giảm nhiễu và các phát hiện sai nhằm tăng khả năng phân chia một cách chính xác dòng video; một số biện pháp khắc phục vấn đề chuyển động của camera và đối tượng; thuật toán cho phép nâng cao khả năng phân đoạn dữ liệu video.

1. ĐẶT VẤN ĐỀ

Trong môi trường đa phương tiện, dữ liệu video là một dạng dữ liệu đặc biệt với những đặc tính nổi bật như kích thước lớn, nội dung ngữ nghĩa trừu tượng, có các quan hệ không gian - thời gian, tốc độ dòng dữ liệu cao... Hiện nay nhu cầu sử dụng video ngày càng tăng cao, đòi hỏi phải xây dựng các CSDL video. Quá trình thiết kế và xây dựng các CSDL video đòi hỏi rất nhiều công đoạn phức tạp. Một trong những bước rất quan trọng là chuẩn bị dữ liệu. Khác với loại dữ liệu truyền thống (dữ liệu số, văn bản), do đặc tính đặc biệt của mình, mô hình hóa dữ liệu video luôn được đặt ra như một yêu cầu tiên quyết. Mô hình hóa dữ liệu video liên quan đến vấn đề trình bày dữ liệu video, đó là việc thiết kế tóm tắt mức cao của video thô, tạo điều kiện cho các hoạt động khác như chèn, biên tập, chỉ số hóa, duyệt và truy vấn dữ liệu video. Vì vậy, mô hình hóa dữ liệu video thường là việc làm đầu tiên của quá trình chuẩn bị dữ liệu video. Hầu hết các quá trình chuẩn bị dữ liệu đòi hỏi dữ liệu video cần được phân chia theo thời gian thành các đoạn cơ sở có nghĩa. Sau đó là các bước tóm tắt mức cao, chỉ số hóa...

1.1. Mô hình hóa dữ liệu

Phân đoạn dữ liệu video theo thời gian nhằm mục đích phân chia dòng video liên tục thành các đơn vị cơ sở có ý nghĩa là các lia (shot) hoặc các cảnh (scene). Trong đó, lia được coi là một đơn vị vật lý của dòng video. Nó là một tập hợp các khung video liên tiếp thu được từ một hoạt động camera đơn. Như vậy giữa các khung có sự liên tục về mặt không gian-thời gian. Trong khi cảnh lại được hiểu là một đơn vị logic (về mặt ngữ nghĩa) của dòng video. Cảnh có thể gồm một hoặc một số lia có sự liên quan về mặt không gian và/hoặc thời gian cùng mang đến một ngữ nghĩa nào đó. Phân đoạn video thực chất là phát hiện trong một chuỗi video (sequence) ranh giới giữa các lia hoặc các cảnh. Các kỹ thuật phân đoạn dữ liệu video có thể chia ra thành nhiều lớp, tùy theo việc lựa chọn (i) các đặc trưng của video để so sánh, (ii) các phương pháp so sánh, (iii) đối tượng video thực hiện phân đoạn. Các kỹ thuật phân đoạn dữ liệu video chủ yếu dựa vào nội dung hình ảnh - âm thanh của video, thực hiện đo đạc sự khác nhau giữa các khung để phát hiện ranh giới giữa các lia (các chuyển lia).

1.2. Mục đích của bài báo

Bài báo trình bày một phương pháp chuẩn bị dữ liệu video đã được sử dụng cho xây dựng CSDL video trong các hệ thống camera giám sát an ninh. Các hệ thống này sử dụng nhiều loại camera: loại đứng yên quan sát trong một không gian cố định; loại chuyển động với nhiều kiểu quét, quay, zoom... khác nhau. Dòng dữ liệu video thu được từ hệ thống camera quan sát chịu ảnh hưởng của chuyển động của camera và/hoặc đối tượng. Dữ liệu video trong các hệ thống này có các đặc điểm riêng: số lượng dữ liệu lớn; cấu trúc dữ liệu theo chuẩn riêng; dòng video có phản ánh các chuyển động phức tạp; yêu cầu xây dựng mô hình dữ liệu riêng; các yêu cầu truy vấn, tìm kiếm đối tượng video là chủ yếu. Do đó cần chú ý nhiều đến chuẩn bị dữ liệu trước khi thu thập dữ liệu vào CSDL. Trong đó đặc biệt nhằm vào một số kỹ thuật mới và các biện pháp cải tiến nhằm nâng cao hiệu quả phân đoạn dữ liệu video. Các phần tiếp theo của bài báo trình bày một số giải pháp: (i) Phương pháp phát hiện cắt lia dựa trên đánh dấu vùng nền.

(ii) Khắc phục vấn đề chuyển động của camera và đối tượng.

(iii) Nâng cao hiệu quả phân đoạn video.

2. PHƯƠNG PHÁP PHÁT HIỆN CẮT LIA DỰA TRÊN ĐÁNH DẤU VÙNG NỀN

2.1. Cụ thể hóa yêu cầu của vấn đề

Một dòng video gồm n khung (f_1, f_2, \dots, f_n) có thể xem như một chuỗi các khung ảnh tĩnh (frame) nối tiếp nhau và cường độ sáng của một điểm ảnh cũng là một hàm của thời gian $f(x, y, t)$. Các chuyển lia thể hiện đã có một sự thay đổi nào đó của không gian và/hoặc thời gian giữa các lia ([4]). Có hai loại chuyển lia là các cắt lia (chuyển lia đột ngột) và các chuyển lia từ từ.

Các thuật toán phát hiện cắt lia đều có chung mục đích là phát hiện tính không liên tục thời gian trong nội dung hình ảnh của một chuỗi video. Trong đó vấn đề then chốt là chọn đặc tính (các đặc tính) để biểu diễn nội dung hình ảnh. Đặc tính được chọn cần biểu thị

các thuộc tính giống nhau nếu các ảnh nằm trong cùng một lia, nhưng sẽ thay đổi rõ rệt nếu chúng nằm trong các lia khác nhau.

Hiện nay đã có một số thuật toán phát hiện cắt lia ([2, 7, 11]), như các thuật toán dựa trên so sánh ở mức pixel; thuật toán dựa trên so sánh đặc tính màu, biểu đồ màu; các thuật toán dựa trên đặc tính cạnh; các thuật toán dựa trên chuyển động của đối tượng; các thuật toán dựa trên các tham số của video nén; ngoài ra, còn có một số phương pháp nhằm vào các đặc tính hình ảnh mức cao của video...

Các tác giả đã đề xuất một phương pháp phát hiện gianh giới lia (PRL) ([9]) dựa trên việc đánh dấu đường di chuyển camera qua vùng nền khung video, cùng với các kỹ thuật nhằm giảm nhiễu để tăng khả năng phát hiện chính xác gianh giới các lia. Kỹ thuật PRL này dựa trên hai cơ sở sau:

1. Lựa chọn các vùng đặc biệt trên các khung ảnh video để đánh dấu đường di chuyển của camera nhằm tăng hiệu quả cho thuật toán so sánh giữa các khung trong PRL;
2. Vận dụng kỹ thuật nén ảnh theo phương pháp Tháp Gauss (Gaussian Pyramid) [6, 8] để giảm nhỏ kích thước ảnh phục vụ cho PRL.

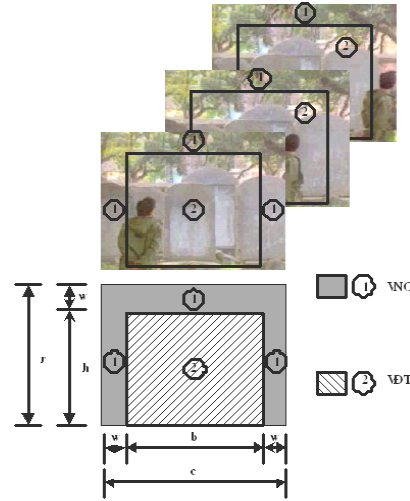
Khi một lia được thực hiện từ một hoạt động camera, theo vết di chuyển của camera là một cách để xác định gianh giới giữa các lia. Có thể làm

được việc này bằng cách đánh dấu vùng nền trong các khung video. Vùng nền (VNC) được định nghĩa như Hình 1. Lí do chọn VNC trong các khung video có dạng chữ π : (i) trước hết là do phần đáy của một khung thường là phần thể hiện các đối tượng hay còn gọi là vùng cận cảnh, hay vùng đối tượng (VĐT); (ii) vùng nền chữ nhật nhỏ trên đỉnh phản ánh chuyển động ngang của camera; (iii) hai cột trái/ phải phản ánh chuyển động dọc của camera. Kết hợp giữa vùng đỉnh và các cột có thể vạch ra chuyển động bất kỳ theo hướng đường chéo.

Dựa trên cách xác định vùng nền, ta có thể thực hiện kỹ thuật PRL thông qua việc xác định mức độ phân biệt VNC của hai khung liền kề. Mặc dù độ phức tạp của thuật toán có tăng, nhưng bù lại phương pháp này lại tỏ ra hiệu quả trên các phương diện sau:

1. Số lượng phép so sánh giảm do kích thước vùng cần so sánh không phải là toàn bộ khung.
2. Giảm được phát hiện sai đáng kể so với một số phương pháp khác do các biện pháp lọc nhiễu (Mục 3.1) và phân biệt khá tốt được tính liên tục thời gian của dòng video với các chuyển động của camera và/hoặc đối tượng.

Để tiện so sánh các vùng nền, nên sử dụng các hình chuyển dạng do kéo thẳng VNC. Vùng nền kéo thẳng của VNC được gọi là vùng nền chuyển đổi (kí hiệu là VNCĐ); và nó là một mảng các pixel hai chiều, như trong Hình 2.



Hình 1. Vùng nền cố định



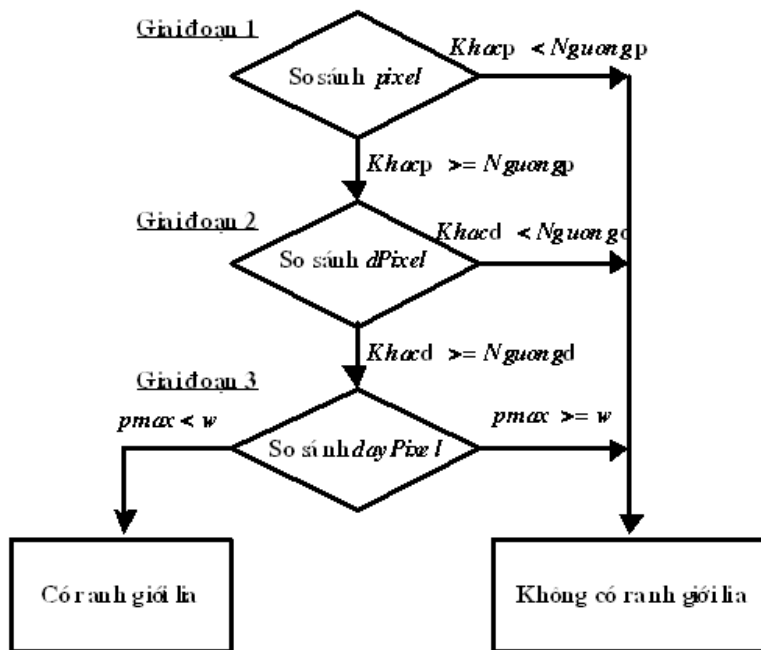
Hình 2. Vùng nền chuyển đổi

Để so sánh nội dung hai vùng nền, trong một số trường hợp cần dựa vào việc so sánh trung bình của chúng. Chúng tôi đã dùng phương pháp Tháp Gauss, quen thuộc trong kỹ thuật nén ảnh. Nó cho phép chuyển một VNCĐ hai chiều thành một dãy các pixel (kí hiệu là *dayPixel*) và cuối cùng sẽ chuyển thành một pixel đơn (kí hiệu là *dPixel*). Theo hình 2, vùng rộng $L \times w$ sẽ được chia thành các cửa sổ; qua một số lần áp dụng phương pháp này, tính được dãy các pixel đặc trưng cho trung bình màu, để tính *dPixel*.

2.2. Thuật toán

Thuật toán sử dụng *dayPixel* và *dPixel* để PRL thể hiện trên hình 3. Hai giai đoạn đầu thường được sử dụng để kiểm tra nhanh cho những trường hợp rõ ràng. Chỉ khi nào không phát hiện được mới sử dụng đến giai đoạn 3. Thuật toán được mô tả chi tiết như sau:

Giai đoạn 1. So sánh pixel: Một pixel trong mỗi VNC được so với pixel ở cùng vị trí và một số pixel lân cận (Vungp) trong VNC của khung liền kề nó. Nếu tổng số pixel trong hai VNC khác nhau *Khacp* nhỏ hơn một giá trị ngưỡng ($Khacp < Nguongp$), hai khung tương ứng với hai VNC coi như cùng một lia và xét tới khung tiếp theo. Ngược lại, sang giai đoạn 2.



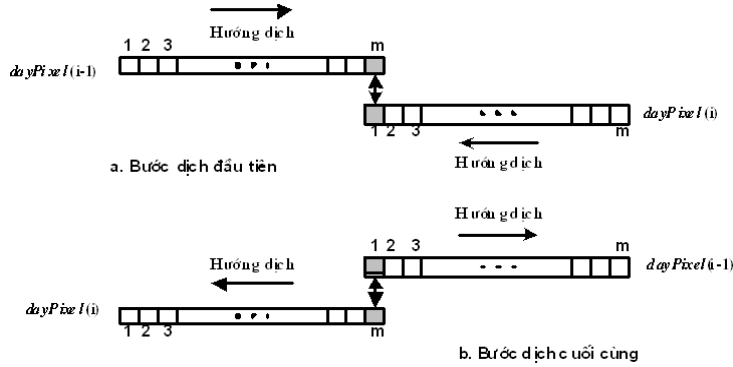
Hình 3. Mô tả thuật toán

2. *Giai đoạn 2.* So sánh các trung bình màu *dPixel*: Nếu sự khác nhau của *dPixel* của hai khung (gọi là *Khacd*) chưa vượt quá ngưỡng ($Khacd < Nguongd$), hai khung coi như cùng một lia và sẽ xét tới khung tiếp theo. Ngược lại, sang giai đoạn sau.

Các ngưỡng *Nguongp* và *Nguongd* được chọn theo kinh nghiệm; tùy theo nội dung video. Chẳng hạn khi xử lý dữ liệu về thông tin quảng cáo, $Nguongp = 0, 15$, $Nguongd = 0, 20$.

Giai đoạn 3. So sánh *dayPixel*: Tính toán xem hai khung chia sẻ bao nhiêu phần nền chung. Hai *dayPixel* được so sánh bằng cách dịch chúng đi theo chiều ngược lại như trên hình 4. Mỗi bước dịch là một pixel và sẽ tính được số điểm khớp nhau *p* tương ứng (số các giá trị khớp nhau liên tiếp *p* của *dayPixel* của khung *i* (*dayPixel_i*) và của khung *i* - 1 (*dayPixel_{i-1}*)).

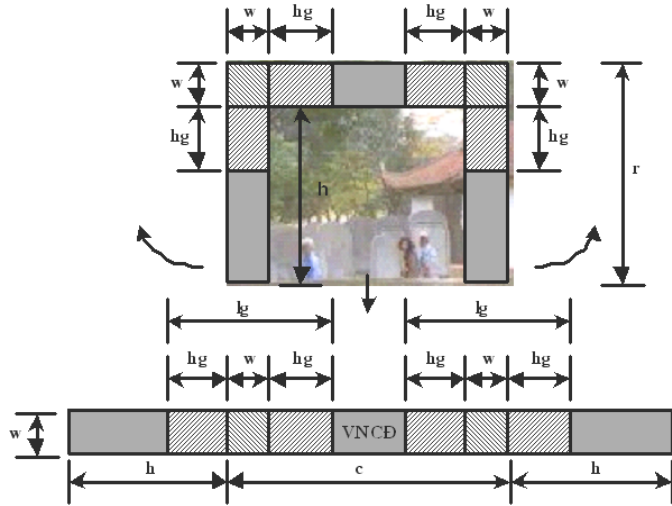
Ví dụ trong bước dịch đầu tiên, giá trị pixel đầu tiên của *dayPixeli* được so sánh với giá trị pixel cuối *m* của *dayPixeli-1*. Giai đoạn này kết thúc khi giá trị pixel cuối *m* của *dayPixeli* được so với giá trị pixel đầu của *dayPixeli-1*. Nếu số cực đại của tất cả các giá trị điểm khớp (*pmax*) lớn hơn chiều rộng của VNC (*w*), hai khung coi như cùng một lia và sẽ xét tới khung tiếp theo. Ngược lại, đã phát hiện một chuyển đổi lia giữa hai khung.



Hình 4. Bước dịch đầu tiên và cuối cùng của dayPixel

2.3. Biện pháp lọc nhiễu

Trong mỗi khung video có thể nhận thấy rằng các phần trung tâm của VNC là rất nhạy với nhiễu. Đặc biệt trong các trường hợp ngắt cảnh vì sự di chuyển của đối tượng vào VNC hoặc đột ngột xuất hiện và bành trướng về kích thước (cận cảnh). Bằng việc chọn tính toán thêm các vùng góc của VNC như trên Hình 5, sẽ có thêm các thông tin để lọc nhiễu, giảm được phân đoạn sai. Ngoài số đo giá trị màu trung bình xấp xỉ của toàn bộ VNCD là *dPixel*, còn tính thêm các giá trị tương ứng cho góc trái và góc phải (*dPixelTrai* và *dPixelPhai*).



Hình 5. Lựa chọn các vùng góc nền của khung video

Trong thuật toán đã trình bày (ii), bổ sung thêm Giai đoạn 4 để lọc nhiễu. Sau khi đã thực hiện xong Giai đoạn 3 và phát hiện ra hai khung *i* và *j* có khả năng là một gianh giới giữa hai lia. Đặt *KhacTrai* và *KhacPhai* là các sự khác nhau của hai giá trị *dPixelTrai* và của hai giá trị *dPixelPhai* giữa hai khung *i* và *j*. Chúng được tính như sau

$$KhacTrai = \frac{100|dPixelPhai_i - dPixelPhai_j|}{2^n} \% \tag{1}$$

trong đó, *n* là số bit màu, ở đây ta chọn *n* = 8.

Trong tự tính được *KhacPhai*. Từ đó xác định được giá trị khác nhau nhỏ nhất *KhacGoc*

$$KhacGoc = \min(KhacTrai, KhacPhai). \quad (2)$$

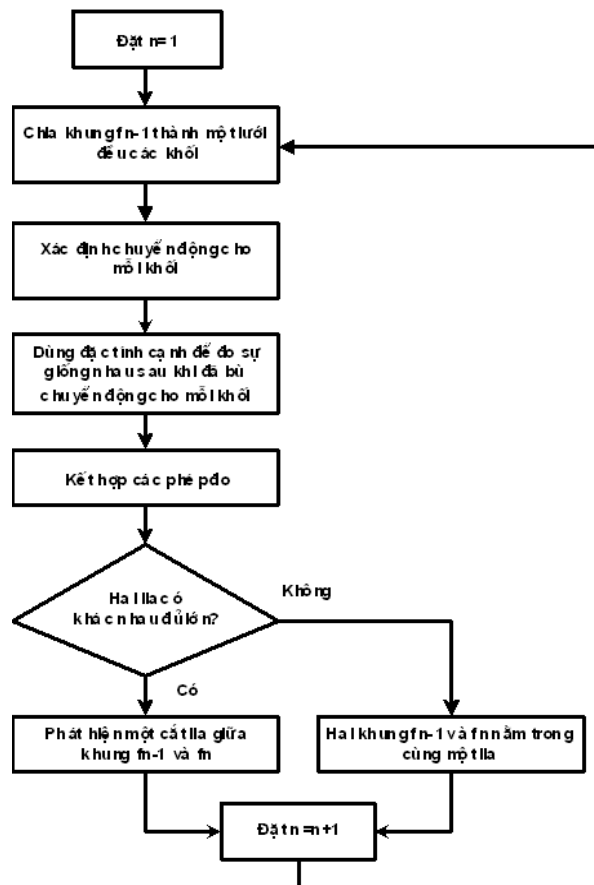
Nếu $KhacGoc$ lớn hơn 10% giá trị khác nhau của các $dPixel$ có thể có, lúc này thực sự xuất hiện một gianh giới lia giữa hai khung i và j . Ngược lại, hai khung coi như nằm cùng trong một lia.

3. TINH CHỈNH DỮ LIỆU, XỬ LÝ CHUYỂN ĐỘNG CỦA ĐỐI TƯỢNG VÀ CAMERA

3.1. Tóm tắt thuật toán

Hiện nay, kỹ thuật phát hiện cắt lia đã đạt được các kết quả khá tốt. Ngược lại, kỹ thuật phát hiện chuyển lia phức tạp vẫn gặp phải rất nhiều khó khăn. Nguyên nhân chính là do khó có thể phân biệt chính xác những thay đổi nội dung hình ảnh trong các khung liên kề do chuyển lia và do chuyển động của camera và các đối tượng gây ra. Chúng tôi cũng đã đề xuất một phương pháp khắc phục vấn đề chuyển động của camera và đối tượng trong phát hiện các chuyển lia video ([3, 10]).

Thuật toán dựa trên so sánh các đặc tính toàn cục có thể khắc phục vấn đề chuyển động của camera và đối tượng, do đối tượng toàn cục ít thay đổi. Đối với trường hợp hai lia có các đặc tính toàn cục thể hiện các thuộc tính giống nhau nhưng lại có bố trí không gian khác nhau, cần chia khối để so sánh các đặc tính của các vùng tương ứng trong các khung. Phương pháp này loại trừ sự khác nhau sinh ra bởi chuyển động trước khi tính toán sai lệch giữa hai khung. Như vậy sự khác nhau nào đó còn lại trong nội dung hình ảnh chỉ là do một cắt lia gây nên. Hai lia nối nhau bởi một chuyển lia có thể có phân bố màu và/hoặc bố trí không gian khác nhau. Do đó một thay đổi trong bố trí cấu trúc của một ảnh thường biểu thị sự không liên tục hình ảnh. Các đặc tính cạnh phản ánh nhiều về cấu trúc trong một ảnh, nghĩa là phân bố các cạnh trong hai khung nằm ở gianh giới lia sẽ rất khác nhau. Tư tưởng chính của phương pháp là sử dụng sự tương quan của các đặc tính cạnh giữa các vùng tương ứng để đo sự giống nhau nội dung hình ảnh giữa hai khung. Phương pháp được mô tả tóm tắt như Hình 4. Với mỗi cặp khung trong chuỗi video,



Hình 6. Các bước phát hiện cắt lia

khung thứ nhất f_{n-1} được chia thành các khối bằng một lưới đều. Sau đó đo sự giống nhau giữa mỗi cặp khung bằng cách so sánh các đặc tính cạnh trong mỗi khối. Bước tiếp theo là ước lượng chuyển động cho mỗi khối giữa hai khung để bù sự khác nhau sinh ra bởi chuyển động của camera và đối tượng.

Như vậy sự khác nhau còn lại trong nội dung hình ảnh của cặp khung sẽ chỉ là kết quả của một cắt lia. Mỗi khối trong khung thứ nhất sẽ làm mẫu để tìm các khối khớp với nó trong một vùng lân cận xung quanh khối tương ứng với nó trong khung thứ hai f_n . Vị trí của khối khớp nhất sẽ xác định độ dịch chuyển của mỗi khối giữa hai khung, từ đó tính một số đo sự giống nhau để bù chuyển động. Sử dụng một hàm độ khớp giữa hai khối xác định hệ số tương quan giữa chúng. Giá trị hệ số tương quan lớn nhất được sử dụng làm số đo khớp cố định với mỗi khối. Hai khung thuộc về cùng một lia phân bố không gian của các đặc tính cạnh sẽ giống nhau và khớp cố định với đa số các khối lớn sẽ biểu thị một khớp tốt. Nếu phần lớn các khớp không tốt sẽ cho thấy phân bố không gian của các đặc tính cạnh bị thay đổi, biểu hiện sẽ có một cắt lia. Kết hợp số đo khớp cố định của tất cả các khối sẽ cho một số đo sự giống nhau cho mỗi cặp khung. Bước cuối cùng là sử dụng số đo giống nhau này để quyết định sự xuất hiện của một cắt lia.

3.2. Chia khối cho các khung nhằm bù chuyển động

Bước đầu tiên trong phương pháp là ước lượng chuyển động cho mỗi khối giữa hai khung để loại trừ sự khác nhau sinh ra bởi các chuyển động. Giả thiết rằng khung ảnh gồm các khối dịch chuyển và chuyển động trong mỗi khối pixel là không thay đổi. Theo đó khối thứ i là $x_i(\xi)$ trong khung f_{n-1} được làm mẫu như là một phiên bản được dịch đi của một khối cùng kích thước $x_j(\xi)$ trong khung f_n như sau

$$x_i(\xi) = x_j(\xi + d) \quad (3)$$

trong đó ξ là vector tọa độ không gian và d là vector dịch chuyển xác định nhờ tìm vị trí của khối khớp nhất trong lân cận khối tương ứng trong khung f_n . Giả sử x_i được biểu diễn dưới dạng một ma trận các điểm ảnh trong không gian rời rạc 2-D, thì các vector tọa độ không gian ξ và các vector dịch chuyển d có thể được biểu diễn dưới dạng các ma trận cột trong mặt phẳng xy . Các thuật toán khớp khối biến đổi tùy theo tiêu chuẩn khớp và phương pháp tìm kiếm.

Đơn giản nhất là phương pháp đo sự giống nhau giữa các mẫu mức xám chứa trong $x_i(\xi)$ với một khối có khả năng khớp nhất trong f_n ([1]). Các phương pháp đo thường tính tổng sai lệch tuyệt đối hoặc tổng sai lệch bình phương. Dựa vào $x_i(\xi)$ và $y_j(\xi)$, ở đây $y_j(\xi)$ là một khối lân cận khối tương ứng trong khung f_n , tương quan của hai hàm 2-D được định nghĩa

$$\rho(\zeta) = (x_i \times y_j)(\zeta) = \sum_{\xi} x_i(\xi + \zeta) y_j(\xi) \quad (4)$$

trong đó ζ là các vector dịch chuyển thay đổi trong quá trình tính hàm tương quan cho hai khối điểm ảnh. Vì là ảnh số, cho nên ứng với mỗi độ dịch rời rạc có khả năng ζ sẽ có một giá trị giống nhau tỷ lệ với mức độ khớp, khi áp dụng sự dịch chuyển tương ứng ζ tới vùng x_i . Nói cách khác, bằng việc tính $\rho(\zeta)$ trên tất cả các dịch chuyển có khả năng (tất cả các ví dụ ở đây x_i và y_j che lấp nhau), vùng tương quan đạt giá trị đỉnh nguyên, tại dịch chuyển gần nhất với dịch chuyển đúng của x_i giữa f_{n-1} và f_n . Nói cách khác, nếu công thức (3) luôn thoả mãn cho vùng x_i thì $\rho(\zeta)$ có giá trị đỉnh của nó tại $\zeta = d$ và độ dịch đã được xác định.

Tuy nhiên, khi sử dụng công thức (4) để ước lượng chuyển động sẽ gặp một số bất tiện như dải của $\rho(\zeta)$ phụ thuộc vào kích thước x_i và nó không phải luôn là hằng số trong toàn bộ chuỗi ảnh ([5]).

Để khắc phục được nhược điểm trên sử dụng tương quan chéo tiêu chuẩn hóa, định nghĩa như sau

$$\gamma(\zeta) = \frac{\sum_{\xi}(x_i(\xi + \zeta) - \bar{x}_i(\xi))(y_j(\xi) - \bar{y}_j(\zeta))}{[\sum_{\xi}(x_i(\xi + \zeta) - \bar{x}_i(\xi))^2 \sum_{\xi}(y_j(\xi) - \bar{y}_j(\zeta))^2]^{1/2}} \quad (5)$$

trong đó $\bar{x}_i(\xi)$ là trung bình của khối x_i và $\bar{y}_j(\zeta)$ là trung bình của y_j trong vùng che lấp bên dưới x_i . Phép đo thống kê này có đặc điểm là nó đo tương quan trên tỷ lệ chia tuyệt đối và chỉ thị một cách tuyến tính sự giống nhau giữa các khối. Sau đó giá trị của hệ số tương quan lớn nhất được sử dụng làm số đo độ khớp cố định. Hơn nữa, nhờ tiêu chuẩn hóa hàm tương quan chéo, phép đo trở nên bất biến trước thay đổi của độ lớn kích thước ảnh ([4]).

Tuy nhiên, việc tính toán tương quan được tiêu chuẩn hóa trong miền không gian tương đối khó khăn, trừ khi các vùng là nhỏ. Giả sử rằng x_i và y_j là các vùng vuông có kích thước $N \times N$ và $M \times M$, thông thường $N \leq M$, tương quan được tiêu chuẩn hóa giữa hai vùng có thể được tính hiệu quả hơn trong miền tần số khi N tiến gần M và với N và M lớn hơn ([3, 4]). Tương quan được tiêu chuẩn hóa trong miền tần số được định nghĩa như sau

$$\rho(\xi) = \frac{\mathfrak{F}^{-1}\{x_i(\omega)\hat{y}_j^*(\omega)\}}{\sqrt{\int |x_i(\omega)|^2 d\omega \cdot \int |\hat{y}_j(\omega)|^2 d\omega}} \quad (6)$$

trong đó ξ và ω là các vector tọa độ không gian và tọa độ tần số không gian, $x_i(\omega)$ biểu thị biến đổi Fourier của khối $x_i(\xi)$, \mathfrak{F}^{-1} biểu thị hàm Fourier ngược và là liên hợp phức. Vì lý do để giảm sai số, với một khối kích thước M dải của ước lượng là $[-M/2 + 1, M/2]$. Thậm chí, với một khối có cạnh M , chỉ có ước lượng bên trong dải $[-M/4 + 1, M/4]$ được quan tâm đến. Phương pháp ước lượng chuyển động giữa vùng quan tâm x_i có cạnh $N \times N$ trong khung với một vùng tương ứng trong khung như sau

1. Tách các khối x'_i và x'_j cả hai có kích thước $M \times M$ trong đó $M = 2N$.
2. Thực hiện tương quan tiêu chuẩn hóa sử dụng công thức (6).
3. Chỉ tìm đỉnh tương quan bên trong ước lượng dịch chỗ $[-N/2 + 1, N/2]$.

3.3. Kết hợp sử dụng các đặc tính cạnh và biểu đồ màu nhằm tăng độ tin cậy

Quá trình khớp khối dựa trên giả thiết phân bố các đặc tính cạnh thay đổi trong một chuyển lia. Do đó cần phải tính đến tương quan của các đặc tính cạnh giữa hai ảnh (sử dụng tương quan với mẫu mức xám chứa trong mỗi khối là không đủ chính xác). Vì đặc tính cạnh miêu tả hiện tượng ảnh tần số cao, do đó sẽ sử dụng một bộ lọc thông cao cho mỗi ảnh trước khi thực hiện tương quan. Một vùng tương quan nhận được từ các vùng thông cao sẽ chứa nhiều đỉnh có thể tìm thấy hơn, ngược lại những vùng tương quan nhận được từ những vùng thông thấp sẽ cho một miền tương quan phẳng dễ gây nên sai số. Kết quả có được khi áp dụng bộ lọc thông cao là trung bình cục bộ của cường độ ảnh được loại bỏ, nghĩa là tương quan giữa các khối là bất biến trước biến đổi cường độ trung bình.

Nếu chỉ dùng phép đo giống nhau dựa trên tương quan các đặc tính cạnh, ngay cả với kích thước các khối biến đổi thích nghi, cũng không đủ phản ánh hết các thay đổi nội dung hình ảnh trong khung video. Ví dụ trong trường hợp các khung video có cấu trúc đơn giản,

trong khung có nhiều đối tượng di chuyển hoặc các chuyển động quá phức tạp đều dẫn đến tương quan kém. Do đó cần sử dụng kết hợp với so sánh phân bố màu để tăng độ chính xác trong phát hiện cắt lia.

Hai lia có thể có phân bố màu khác nhau hoặc giống nhau nhưng cấu trúc không gian khác nhau. Rõ ràng hai đặc trưng quan trọng cần so sánh để phát hiện cắt lia là (i) phân bố màu và (ii) cấu trúc bên trong. Do đó, thuật toán phát hiện cắt lia nêu trên được mở rộng kết hợp với so sánh các phân bố màu của các vùng tương ứng (chia khối). Trong đó, mỗi khối được bù chuyển động trước khi so sánh biểu đồ màu của các vùng tương ứng. Nói cách khác, biểu đồ màu của mỗi khối trong khung f_{n-1} được so sánh với biểu đồ của khối khớp nhất trong khung f_n . Kết hợp so sánh biểu đồ màu với so sánh các đặc trưng cạnh sẽ cho một vector giống nhau ứng với với mỗi cặp khung, trong đó có hai thành phần là:

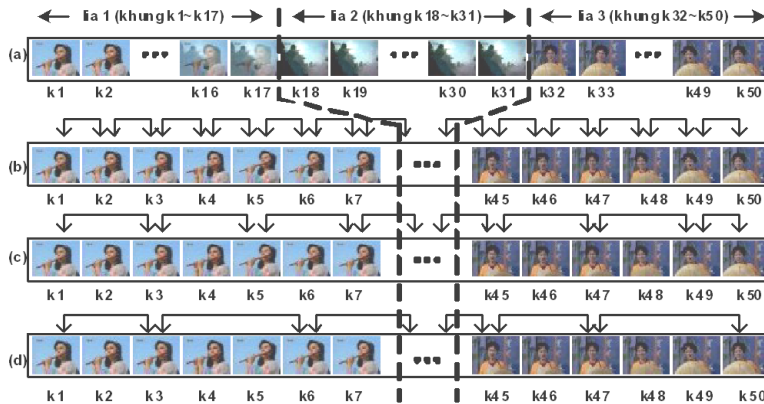
1. Số đo giống nhau đã bù chuyển động của phân bố các đặc trưng cạnh trong mỗi khung.
2. Số đo khác nhau đã bù chuyển động của phân bố màu trong mỗi khung.

Việc kết hợp này dựa trên giả thiết lý tưởng là các khung liền kề của cùng một lia sẽ có cấu trúc không gian và phân bố màu giống nhau, ngược lại, hai khung tách biệt bởi một cắt lia phải có cạnh và phân bố màu khác nhau. Sử dụng kết hợp hai phương pháp có ưu điểm là khi một phép đo chỉ thị khả năng có một cắt lia, thì phép đo khác sẽ được dùng để xác minh lại.

4. NÂNG CAO KHẢ NĂNG PHÂN ĐOẠN DỮ LIỆU VIDEO

4.1. Nhu cầu phân đoạn

Các thuật toán phân đoạn video hiện nay đều thực hiện so sánh ở mức khung giữa các khung liền kề và cho các kết quả chấp nhận được. Chúng quét toàn bộ dòng video và xử lý các khung theo trình tự tuyến tính. Do tính tương quan cao của các khung video trong cùng một lia, cho nên trong trường hợp này rất nhiều phép so sánh thực sự không cần thiết, làm cho các thuật toán phức tạp hơn và khả năng thực hiện gặp nhiều khó khăn, nhất là trong trường hợp kích thước video lớn ([7]). Xuất phát từ nhận xét đó, chúng tôi đề xuất phương pháp nhằm cải tiến các thuật toán phân đoạn dữ liệu video bằng cách áp dụng các kỹ thuật nhảy đều và nhảy thích ứng. Thực chất là nhờ đó, có thể bỏ qua việc so sánh một số khung khi chúng ở trong cùng một lia mà không sợ sai sót.



Hình 7. Minh họa phương pháp so sánh không tuyến tính

Chuỗi video mẫu dùng để phân tích như trên Hình 7a. Giả sử chuỗi video này gồm 50 khung và có 3 lia. Gianh giới của lia 1 và lia 2 nằm giữa các khung k_{17} và k_{18} và chuyển lia này là một chồng mờ; gianh giới của lia 2 và lia 3 nằm giữa các khung k_{31} và k_{32} và chuyển lia này là một cắt lia. Các kỹ thuật phân đoạn dữ liệu video trước đây thực hiện các phép so sánh sự khác nhau giữa các khung liên kề, nếu chúng sử dụng các thuật toán tuyến tính, chúng sẽ quét lần lượt toàn bộ các khung như mô tả trên Hình 7b.

4.2. Kỹ thuật nhảy đều

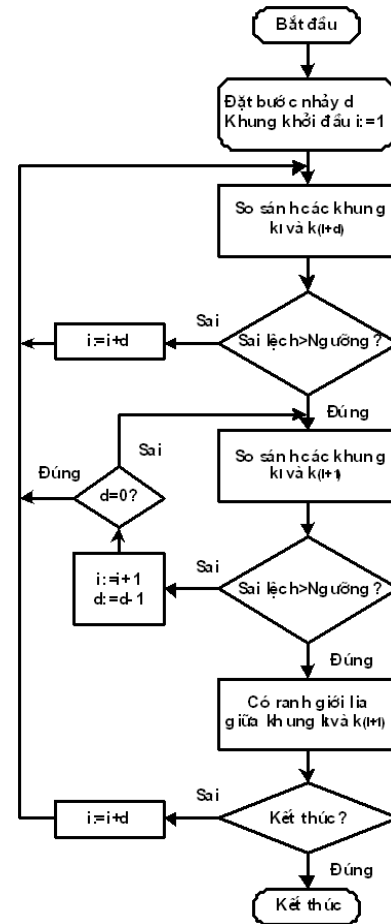
Không như các thuật toán tuyến tính, kỹ thuật nhảy đều không thực hiện các phép so sánh giữa các khung liên kề mà thực hiện so sánh cách đều một số khung. Mỗi phép so sánh sẽ được thực hiện cách đều d khung (bước nhảy d). Thí dụ trường hợp $d = 2$ như trên Hình 7c. Các phép so sánh thực hiện giữa các khung sau. k_1 với $k_{1+d}(k_3)$, k_3 với $k_{3+d}(k_5)$, k_5 với $k_{5+d}(k_7), \dots$

Quá trình so sánh cứ tiếp tục như vậy. Khi phát hiện ra hai khung k_i và k_{i+2} nằm trong hai lia khác nhau, sẽ tiến hành so sánh khung k_i với k_{i+1} để xác định gianh giới các lia. Nếu khung k_i và k_{i+1} nằm trong hai lia khác nhau, thì gianh giới lia nằm giữa hai khung k_i và k_{i+1} . Ngược lại, nếu khung k_i và k_{i+1} nằm trong cùng một lia, thì chắc chắn gianh giới lia nằm giữa hai khung k_{i+1} và k_{i+2} . Trường hợp mỗi lia có tối thiểu 2 khung sẽ không cần so sánh hai khung k_{i+1} với k_{i+2} . Khi đã xác định được gianh giới lia, trình tự được lặp lại cho lia tiếp theo. Lưu đồ thuật toán của kỹ thuật nhảy đều được thể hiện trên Hình 8. Chú ý trong phép kiểm tra để kết thúc vẫn đặt lại số thứ tự khung là $i := i + d$, vì sau vòng lặp so sánh giữa các khung k_i và k_{i+1} , số đếm i tăng lên 1 đơn vị thì ngược lại số đếm d lại giảm 1 đơn vị.

Kỹ thuật nhảy đều có thể cho phép giảm gần một nửa (với $d = 2$) số phép so sánh so với khi chỉ sử dụng các thuật toán tuyến tính. Có thể tổng quát hóa trong trường hợp bước nhảy d lớn hơn để cho kết quả tốt hơn. Số phép so sánh của quá trình phân đoạn luôn bị giới hạn bởi khả năng xử lý của máy tính. Với các máy tính có tốc độ không cao ý nghĩa của kỹ thuật này càng thể hiện rõ.

4.3. Kỹ thuật nhảy động

Giá trị bước nhảy d nói trên luôn phụ thuộc vào các dòng video. Trong kỹ thuật nhảy thích ứng (nhảy động), giá trị d được xác định động, biến đổi một cách thích ứng để đảm bảo là tối ưu. Mỗi lần lặp, thuật toán tính lại bước nhảy d cho thích hợp, bằng cách so sánh khung hiện tại với khung vừa được kiểm tra như trên Hình 7d.



Hình 8. Lưu đồ thuật toán nhảy đều

Nếu sự giống nhau của phép so sánh này nhiều hơn phép so sánh trước, ở phép so sánh kế tiếp, giá trị d sẽ được tăng lên và ngược lại. Mặt khác nếu phép so sánh hiện tại cho biết 2 khung nằm ở 2 lia khác nhau, sẽ quay lại thực hiện theo xu hướng lùi dần sử dụng các bước nhảy đều.

Khi phát hiện chính xác gianh giới lia, sẽ trở lại tiến trình nhảy động thuận bình thường. Kỹ thuật nhảy động sử dụng bước nhảy thích ứng cho các kết quả khá tốt.

5. KẾT LUẬN

Bài báo đã giới thiệu việc chuẩn bị dữ liệu trong xây dựng các hệ CSDL video số. Một số giải pháp đều nhằm mục đích phân đoạn dữ liệu video một cách chính xác và hiệu quả. Phương pháp này đã được ứng dụng trong chuẩn bị dữ liệu video khi tổ chức các CSDL video trong hệ thống camera quan sát an ninh.

Thuật toán phát hiện cắt lia cho thấy rõ hiệu quả, vì đặc điểm của các vùng nền và các dạng chuyển động trong video nhận được từ hệ thống này là dễ nhận diện. Đặc biệt thuật toán nhảy động cho phép cải thiện đáng kể khả năng phân đoạn dữ liệu video.

TÀI LIỆU THAM KHẢO

- [1] A. Giachetti, Matching techniques to compute image motion, *Image and Vision Computing* **18** (3) (2000) 247–260.
- [2] Đỗ Trung Tuấn, Lương Xuân Cường, Khun Piseth, Nguyễn Văn Tào, Về xử lý dữ liệu video, Toàn văn các báo cáo khoa học, *Hội thảo Quốc gia lần thứ VI: Một số vấn đề chọn lọc của công nghệ thông tin và truyền thông*, Thái Nguyên, NXB Khoa học Kỹ thuật, tháng 8, 2003 (454–459).
- [3] E. Saez et al, Combining luminescence and edge based metrics for robust temporal video segmentation, *IEEE Int'l. Conf. on Image Processing (ICIP2004)*, Singapore, Oct. 24 – 27, 2004 (126–132).
- [4] H. Yu, W. Wolf, A visual search system for video and image databases, *Proc. IEEE Int'l Conf. on Multimedia Computing and Systems*, Ottawa - Canada, Jun., 1997 (517–524).
- [5] J.P. Lewis, Fast normalized cross-correlation, *Vision Interface (1995)*, trang 120–123, <http://www.idiom.com/zilla/Work/nvisionInterface/nip.html>.
- [6] JungHwan Oh Kien et al, A content-based scene change detection and classification technique using background tracking, *SPIE Conf. on Multimedia Computing and Networking 2000*, San Jose, CA, Jan., 2000 (233–245).
- [7] JungHwan Oh Kien et al, “Efficient and cost-effective techniques for browsing and indexing large video databases, computer science program”, School of EECS University of Central Florida, Orlando, (2001) (2328–2362).
- [8] Lương Mạnh Bá, Nguyễn Thanh Thủy, *Nhập môn xử lý ảnh số*, NXB Khoa học Kỹ thuật, Hà Nội, 1998.
- [9] Lương Xuân Cường, Đỗ Trung Tuấn, Thuật toán cho phép nâng cao khả năng phân đoạn dữ liệu video, *Tạp chí khoa học - Chuyên san Khoa học Tự nhiên và Công nghệ*, Đại học Quốc gia Hà Nội, **XXII** (3) (2006) 13–24.

- [10] Lương Xuân Cường, Đỗ Xuân Tiến, Đỗ Trung Tuấn, Một kỹ thuật chỉ số hóa tự động dữ liệu video dựa trên đánh dấu vùng nền, *Tạp chí khoa học - Khoa học Tự nhiên và Công nghệ*, Đại học Quốc gia Hà Nội, **XXII** (2) (2006) 1–11.
- [11] Lương Xuân Cường, Đỗ Xuân Tiến, Đỗ Trung Tuấn, Phương pháp phát hiện các chuyển lia phức tạp dựa trên ước lượng chuyển động trong các khung video, *Tạp chí Chuyên san Các công trình nghiên cứu - triển khai viễn thông và công nghệ thông tin* (17) (12 - 2006) 15–25.
- [12] S. V. Porter et al, Temporal video segmentation and classification of edit effects, *Image and Vision Computing* **21** (2003) 1097–1106.

Nhận bài ngày 26 - 1 - 2007

Nhận lại sau sửa ngày 23 - 3 - 2007