



(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ

(19) Cộng hòa xã hội chủ nghĩa Việt Nam (VN)
CỤC SỞ HỮU TRÍ TUỆ

(11)



1-0029717

(51)⁷ H04N 7/00; H04N 5/14; G06F 17/30; (13) B
G06T 7/20

(21) 1-2017-01657

(22) 03/05/2017

(45) 25/10/2021 403

(43) 26/11/2018 368A

(73) Công ty Cổ phần Nghiên cứu và Sản xuất Vinsmart (VN)

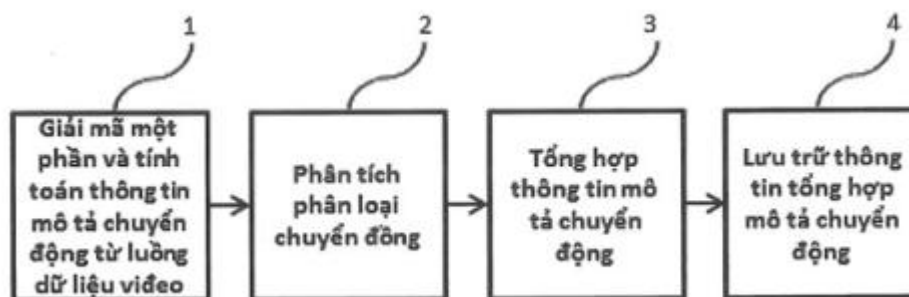
Lô CN1-06B-1&2 khu Công nghiệp Công nghệ cao 1, khu Công nghệ cao Hòa Lạc,
xã Hạ Bằng, huyện Thạch Thất, thành phố Hà Nội

(72) Nguyễn Đình Nam (VN); Nguyễn Thị Thủy (VN); Nguyễn Đỗ Văn (VN); Nguyễn
Minh Hòa (VN); Vương Tùng Long (VN); Lê Thanh Hà (VN).

(74) Công ty TNHH Tầm nhìn và Liên danh (VISION & ASSOCIATES CO.LTD.)

(54) PHƯƠNG PHÁP PHÂN TÍCH VÀ LƯU TRỮ THÔNG TIN MÔ TẢ CHUYỂN
ĐỘNG TRONG NỘI DUNG VIDEO

(57) Sáng chế liên quan đến các kỹ thuật xử lý để trích xuất và tổng hợp dữ liệu video, cụ thể là, sáng chế đề cập đến phương pháp phân tích và lưu trữ thông tin mô tả chuyển động trong nội dung video trong miền tín hiệu đã được nén, và hệ thống tổ chức lưu trữ thông tin tổng hợp mô tả chuyển động của video trên thiết bị lưu trữ vật lý nhằm hỗ trợ việc tìm kiếm video theo nội dung. Đầu vào của hệ thống là video đã được nén (bởi các chuẩn nén video công nghiệp như H.264/AVC, H.265/HEVC), trong đó chứa các thông tin mô tả chuyển động là thông tin quan trọng mô tả nội dung video. Luồng dữ liệu nhị phân video được giải nén một phần để trích rút và tính toán các thông tin mô tả chuyển động. Các thông tin mô tả chuyển động sau đó được phân tích và tổng hợp theo không gian (khung hình) và thời gian (đoạn video). Cuối cùng, thông tin tổng hợp mô tả chuyển động được lưu trữ để phục vụ việc tìm kiếm video.



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến các kỹ thuật xử lý để trích xuất và tổng hợp dữ liệu video trong miền tín hiệu đã được nén, và hệ thống tổ chức lưu trữ thông tin tổng hợp mô tả chuyển động của video trên thiết bị lưu trữ vật lý. Phương pháp và thiết bị được đề cập trong sáng chế nhằm hỗ trợ các hệ thống tìm kiếm nội dung video và truy xuất nhanh dữ liệu video đã được lưu trữ.

Tình trạng kỹ thuật của sáng chế

Sự phát triển của các thiết bị ghi hình hiện nay, đặc biệt là trong các ứng dụng dùng camera quan sát, tạo ra một lượng khổng lồ dữ liệu dưới dạng chuỗi ảnh số (video). Dữ liệu video thường được nén lại trước khi lưu trữ hay truyền gửi và sẽ được giải nén khi sử dụng.

Dữ liệu video từ camera quan sát trong một khoảng thời gian dài sẽ có dung lượng lớn, gây khó khăn trong khi sử dụng, đặc biệt là việc tìm kiếm các sự kiện quan trọng trong các tệp video lớn. Do đó, việc tóm lược nội dung của video và tổ chức lưu trữ hiệu quả các nội dung này là một bài toán quan trọng và có nhiều thách thức. Sáng chế này liên quan đến phương pháp tóm lược video dựa trên kỹ thuật trích xuất, tổng hợp thông tin chuyển động từ video và phương thức lưu trữ thông tin chuyển động nhằm hỗ trợ cho các hệ thống tìm kiếm sự kiện trong video theo nội dung video. Việc tóm lược nội dung video có thể được thực hiện thông qua giải nén một phần (một số công đoạn cuối cùng của quá trình nén) video, điều này giúp làm giảm một lượng lớn chi phí cho tính toán.

Thông thường, video sau khi được thu nhận từ cảm biến sẽ được nén theo một chuẩn nén video phổ dụng như H.261, H.263, MP4, H.264/AVC, H.265/HEVC, ... Nguyên lý căn bản của các hệ nén này là chia video thành các nhóm khung hình. Mỗi khung hình lại được chia thành các khối ảnh đơn vị (Ví dụ khối Macroblock 16x16 trong H.264/AVC). Quá trình nén từng khối ảnh được thực hiện thông qua việc dự đoán, lượng tử hóa phần dư, và nén entropy.

Khởi ảnh đơn vị Macroblock được dự đoán dựa trên thông tin đã được truyền từ bộ nén đến bộ giải nén. Thông thường có hai cách dự đoán là dự đoán nội khung hình và dự đoán liên khung hình. Dự đoán nội khung hình sử dụng dữ liệu hình ảnh đã được nén trong cùng khung hình với Macroblock đang được nén và dự đoán liên khung hình sử dụng dữ liệu hình ảnh ở những khung hình đã được nén trước đó. Dự đoán liên khung hình được thực hiện thông qua quá trình dự đoán và bù chuyển động trong đó bộ dự đoán chuyển động sẽ tìm ra các Macroblock trong khung hình tham chiếu gần giống nhất với Macroblock mới này và tính toán ra vectơ chuyển động (Motion Vector), vectơ này đặc trưng cho sự dịch chuyển của Macroblock mới cần mã hóa so với khung hình tham khảo.

Các Macroblock tham chiếu được gửi tới bộ trừ để trừ với Macroblock mới cần mã hóa nhằm tìm ra các sai số dự đoán (Error Prediction) hoặc tín hiệu dư, chúng sẽ đặc trưng cho sự sai khác giữa Macroblock cần dự đoán và Macroblock thực tế cần mã hóa. Tín hiệu dư hay sai số dự đoán này sẽ được biến đổi DCT và lượng tử hoá để làm giảm số lượng các bits cần lưu trữ hay truyền gửi. Các hệ số này cùng với các vectơ chuyển động sẽ được đưa tới bộ nén entropy và đưa ra dòng dữ liệu nhị phân (bitstream). Dòng dữ liệu nhị phân của video bao gồm: các hệ số biến đổi, thông tin dự đoán chuyển động, thông tin về cấu trúc dữ liệu nén và một số thông tin khác.

Trong quá trình giải nén, bộ giải nén video nhận dữ liệu video đã được nén là dòng chuỗi dữ liệu nhị phân, giải mã các phần tử và trích xuất ra các thông tin đã được mã hóa (như mô tả ở trên), gồm: các hệ số biến đổi, dung lượng bit, thông tin dự đoán chuyển động, v.v. và thực hiện các phép biến đổi ngược để khôi phục lại dữ liệu hình ảnh ban đầu.

Trong bài toán tóm tắt video phục vụ cho việc lưu trữ và tìm kiếm nội dung video, thông tin cần tìm kiếm trong video có thể là một hay nhiều đối tượng hay sự kiện đã xảy ra trong quá trình camera quan sát video, ví dụ một sự đột nhập trái phép. Trong thực tế với camera giám sát, các hiện tượng gây ra chuyển động đều đáng chú ý và có thể cung cấp thông tin quan trọng cho việc phát hiện và tìm kiếm sự kiện. Khi có sự chuyển động, vùng ảnh có chuyển động là vùng được quan tâm, các vùng tĩnh khác được gọi là nền. Khi nền là tĩnh (hầu như không có sự thay đổi về giá trị của các điểm ảnh) thì việc phát hiện chuyển động có thể được thực hiện đơn giản bằng cách trừ khung hình hiện tại với khung hình tham chiếu (khung hình nền đã được xác định trước). Tuy nhiên, trong thực tế nền ảnh thường có sự thay đổi do nhiễu hoặc những chuyển động không mong muốn

(như nhiễu do camera, lá cây rung hoặc ánh sáng ngoại lai). Do đó, việc phát hiện chuyển động thực có trong khung hình video để từ đó phát hiện và định vị sự kiện trong một đoạn xác định của tệp video dài là một bài toán có nhiều thách thức.

Bài toán tìm kiếm sự kiện trong video có dung lượng lớn, nhất là các video giám sát trong một thời gian dài, là một bài toán tiêu tốn thời gian và công sức của người dùng và thiết bị xử lý. Trong các nghiên cứu liên quan đã công bố đã có một số giải pháp cho việc tìm kiếm tự động để phát hiện và định vị thời điểm trong chuỗi video nơi xuất hiện sự kiện. Tuy nhiên, việc tìm kiếm nhanh và hiệu quả đoạn video có chứa sự kiện vẫn chưa được giải quyết thỏa đáng; quá trình xử lý dữ liệu video để tìm đến vị trí xảy ra sự kiện trong video vẫn còn nhiều hạn chế. Do đó, vấn đề xử lý phân tích, tổng hợp để tóm lược dữ liệu video sao cho việc tìm kiếm được tiện lợi và hiệu quả vẫn đòi hỏi những giải pháp tốt hơn.

Sáng chế US6697523, có tên là “Method for summarizing a video using motion and color descriptors”, liên quan đến phương pháp trích xuất thông tin chuyển động của video cho mục đích tóm lược tự động video. Phương pháp trong sáng chế này sử dụng dòng dữ liệu video được giải nén một phần và cả thông tin ảnh (giải nén toàn phần). Việc này làm tiêu tốn tài nguyên tính toán của thiết bị. Việc tóm lược video có thể được thực hiện đơn giản bằng cách lấy ra một khung hình đại diện cho một đoạn video để lưu lại, hoặc phân tích màu. Điều này làm hoặc mất đi thông tin, giảm độ chính xác của kết quả khi tìm kiếm, hoặc phải tính toán phức tạp trên miền đặc trưng của ảnh. Hơn nữa, sáng chế không đề xuất giải pháp lưu trữ hiệu quả thông tin đã được tổng hợp.

Sáng chế US5956026, có tên là “Method for hierarchical summarization and browsing of digital video”, liên quan đến phương pháp tóm tắt (summarization) và duyệt lướt (browsing) video bằng cách sinh ra một cấu trúc biểu diễn phân cấp video gián lược sử dụng một số khung hình đại diện. Trong đó mỗi ảnh đại diện cho một phân cảnh (shot) video, và hệ thống phải xác định được biên chuyển cảnh và số khung hình đại diện. Việc duyệt được thực hiện thông qua các khung hình đại diện. Sáng chế sử dụng cả dữ liệu video đã được giải nén để tính toán cho việc tóm lược video. Sáng chế không đề xuất giải pháp lưu trữ thông tin đã được gián lược của video.

Sáng chế số US7751632, có tên là “Intelligent, dynamic, long-term digital surveillance media storage system”, đề xuất phương pháp phân tích dòng dữ liệu đa phương tiện phục

vụ cho việc mã hóa và đánh chỉ mục lưu trữ dữ liệu theo yêu cầu thực của hệ thống giám sát. Trong đó, việc phân tích nội dung video được thực hiện dựa trên việc đánh giá phân loại dữ liệu chuyển động của mỗi khung hình. Từ đó hệ thống lựa chọn kỹ thuật mã hóa tối ưu cho từng khung hình để tạo ra các bộ mô tả riêng, nhằm phục vụ cho việc lưu trữ hiệu quả. Các đoạn video sau khi được phân tích để chọn ra cách mã hóa tối ưu thì sẽ bị xóa khỏi file video gốc, chỉ có bộ mô tả của nó được lưu lại. Sáng chế ^{này} không đề xuất việc phân tích tổng hợp thông tin chuyển động liên khung hình, không hỗ trợ tìm kiếm theo vùng chuyển động trong khung hình và không có hệ thống phân cấp lưu trữ thông tin tổng hợp chuyển động của video.

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế là đề xuất phương pháp trích xuất, tổng hợp, và lưu trữ thông tin chuyển động từ video và phương thức lưu trữ thông tin chuyển động nhằm hỗ trợ cho các hệ thống tìm kiếm sự kiện trong video theo nội dung video.

Theo một phương án, sáng chế đề xuất phương pháp phân tích và lưu trữ thông tin mô tả chuyển động trong nội dung video, phương pháp này bao gồm các bước: phân tích, phân loại chuyển động và tổng hợp thông tin mô tả chuyển động.

Theo một phương án khác, sáng chế đề xuất phương tiện lưu trữ dữ liệu tổng hợp mô tả chuyển động trong nội dung video theo cấu trúc dữ liệu phân cấp thành các mức.

Vì việc giải nén toàn bộ video để thu được hình ảnh có thể hiển thị được đòi hỏi thiết bị giải nén phải tiêu tốn một lượng lớn tài nguyên tính toán và bộ nhớ nên theo một khía cạnh của sáng chế, dữ liệu sử dụng để phân tích và tổng hợp được trích xuất từ khối dữ liệu thu được bằng cách giải nén một phần chuỗi nhị phân dữ liệu video đã nén (entropy). Dữ liệu này là các vector chuyển động của các khối ảnh và số lượng bit cần thiết để nén mỗi khối ảnh. Việc này cho phép giảm chi phí tính toán của thiết bị phần cứng trong việc tổng hợp và tìm kiếm sự kiện trong video.

Theo một khía cạnh khác của sáng chế, dữ liệu vector chuyển động được sử dụng để phân loại chuyển động thực trong video. Việc này cho phép tận dụng các vector chuyển động được nén trong luồng nhị phân video và tận dụng được các thông tin về khả năng xảy ra các sự kiện quan trọng thể hiện qua các vector chuyển động.

Dữ liệu video từ các nguồn camera quan sát thường rất dài, lưu trữ một video trong một khoảng thời gian dài trong một tệp (file) dữ liệu gây khó khăn trong việc thao tác và xử lý. Do đó, theo một khía cạnh khác của sáng chế, video được chia thành các phân đoạn theo các khoảng thời gian phù hợp để dễ dàng tổng hợp, lưu trữ và được sử dụng bởi các hệ thống hỗ trợ tìm kiếm video theo nội dung khác.

Theo một khía cạnh khác của sáng chế, thông tin chuyển động của video được phân thành bốn loại: không có chuyển động, ít chuyển động hoặc nhiều, có chuyển động, và có nhiều chuyển động.

Theo một khía cạnh khác của sáng chế, phân đoạn video được chia thành các lát cắt video theo không gian và thời gian. Trong đó, lát cắt theo thời gian của một phân đoạn video cho biết độ dài đơn vị của video được sử dụng cho việc phân tích và tổng hợp thông tin chuyển động. Lát cắt theo không gian của một phân đoạn video chia khung hình của video thành các khối ảnh đơn vị để phân tích và tổng hợp thông tin chuyển động. Việc này hỗ trợ các hệ thống tìm kiếm video theo nội dung nhằm cung cấp cho người dùng các lựa chọn để tìm kiếm video theo thời gian và không gian.

Theo một khía cạnh khác của sáng chế, dữ liệu chuyển động trong video đã được phân chia theo các lát cắt không gian và thời gian được lưu trữ theo một tổ chức phân cấp trong một thiết bị lưu trữ dữ liệu tương ứng với các mức phân chia video: phân đoạn video, lát cắt theo thời gian, lát cắt theo không gian. Việc này hỗ trợ các hệ thống tìm kiếm video truy xuất nhanh dữ liệu đã được lưu trữ.

Mô tả vắn tắt các hình vẽ

Các nội dung và các ưu điểm của sáng chế sẽ được mô tả và tham chiếu tới các hình vẽ sau:

Hình 1 thể hiện sơ đồ khối tổng quát của phương pháp phân tích và lưu trữ thông tin mô tả chuyển động trong nội dung video trên miền nén theo sáng chế;

Hình 2 thể hiện bước giải mã một phần và tính toán thông tin mô tả chuyển động từ luồng dữ liệu video tương ứng với bước 1 trên hình 1.

Hình 3 thể hiện bước phân tích, phân loại chuyển động tương ứng với bước 2 trên hình 1.

Hình 4 thể hiện bước tổng hợp thông tin mô tả chuyển động tương ứng với bước 3 trên hình 1.

Hình 5 thể hiện cấu trúc lưu trữ dữ liệu mô tả chuyển động trong nội dung video trên miền nén trong thiết bị lưu trữ là kết quả tương ứng với bước 4 trên hình 1.

Hình 6 thể hiện phương pháp phân chia dữ liệu mô tả chuyển động để lưu vào thiết bị lưu trữ với cấu trúc dữ liệu trên hình 5.

Mô tả chi tiết sáng chế

Hình 1 thể hiện sơ đồ khối tổng quát của phương pháp phân tích và lưu trữ thông tin mô tả chuyển động trong nội dung video trên miền nén theo sáng chế.

Như được thể hiện trên hình 1, phương pháp phân tích và lưu trữ thông tin mô tả chuyển động trong nội dung video trên miền nén theo sáng chế bao gồm các bước: bước 1 tính toán thông tin mô tả chuyển động từ luồng dữ liệu video; bước 2 phân tích, phân loại chuyển động; bước 3 tổng hợp thông tin mô tả chuyển động; bước 4 lưu trữ thông tin tổng hợp mô tả chuyển động.

Bước 1 tính toán thông tin mô tả chuyển động từ luồng video là việc trích xuất và tính toán thông tin cần thiết cho quá trình phân tích, phân loại chuyển động, bước 1 gồm các bước nhỏ: giải nén entropy luồng dữ liệu video đầu vào; tính ra các thông tin cơ bản của từng khối ảnh được giải nén; tính toán thông tin mô tả chuyển động; đưa ra dung lượng dùng để nén khối ảnh (S), tính ra độ dài (L) và hướng cơ bản (D) của các vectơ (mô tả) chuyển động cho bước tiếp theo. Sơ đồ chi tiết của bước này sẽ được thể hiện trên hình 2.

Bước 2 phân tích, phân loại chuyển động dựa vào các thông tin thu được ở bước 1 cho từng khung hình để xác định các vùng chuyển động trong khung hình đó. Sơ đồ chi tiết cho bước này sẽ được thể hiện trên hình 3.

Bước 3 tổng hợp thông tin mô tả chuyển động dựa trên việc phân tích phân loại chuyển động cho các khung hình. Bước tổng hợp thông tin mô tả chuyển động của các khung hình được thực hiện trong từng khoảng thời gian xác định T. Bước này sẽ phân loại mức độ chuyển động của từng vị trí trong khoảng thời gian T đó. Sơ đồ chi tiết cho bước 3 sẽ được thể hiện trên hình 4. Ở đây, tốt hơn là, khoảng thời gian T của video được xác định là một khoảng thời gian để video có ít nhất 2 khung hình và nhiều nhất là 300 khung hình.

Bước 4 lưu trữ thông tin tổng hợp mô tả chuyển động thu được, cụ thể là, lưu trữ dữ liệu thông tin thu được sau bước 3 theo cấu trúc phân cấp theo không gian và thời gian. Phương pháp phân chia dữ liệu phân cấp và cấu trúc dữ liệu trên thiết bị sẽ được thể hiện trên hình 5 và hình 6.

Hình 2 thể hiện bước trích giải mã một phần và tính toán thông tin mô tả chuyển động trong bước 1 ở hình 1. Bước này nhằm tính toán thông tin mô tả chuyển động từ luồng dữ liệu video, bắt đầu bằng việc giải nén entropy (bước 101) luồng dữ liệu video đã nén để thu được các thông tin cơ bản của một khối ảnh (bước 102) bao gồm: dung lượng dùng để nén của từng khối ảnh có kích thước, ví dụ, 16x16, các vector chuyển động của khối ảnh. Sau đó sẽ tính toán các thông tin mô tả chuyển động (bước 103) để đưa ra (bước 104) dung lượng (S) dùng để nén của từng khối ảnh, độ dài vector chuyển động (L), hướng của vector chuyển động (D) đã được chuẩn hóa về 8 hướng: {[hướng 1: $(0 - 45^\circ)$, hướng 2: $(45 - 90^\circ)$, hướng 3: $(90 - 135^\circ)$, hướng 4: $(135 - 180^\circ)$, hướng 5: $(180 - 225^\circ)$, hướng 6: $(225 - 270^\circ)$, hướng 7: $(270 - 315^\circ)$, hướng 8: $(315 - 360^\circ)$]} nhằm phục vụ cho bước phân tích, phân loại chuyển động (bước 2) trên hình 1. Ở đây, dung lượng dùng để nén một khối ảnh được tính là hiệu giữa vị trí bit khi bắt đầu và khi kết thúc giải nén một khối ảnh từ luồng dữ liệu nhị phân của video.

Hình 3 thể hiện bước phân tích, phân loại chuyển động (bước 2) trên hình 1.

Bước này bắt đầu từ bước 201 đánh giá dung lượng (S) dùng để nén từng khối ảnh nhằm xác định thông tin chuyển động có đủ lớn để coi là chuyển động thực. Việc đánh giá được thực hiện bằng cách so sánh dung lượng (S) dùng để nén khối ảnh với một ngưỡng Ts. Nếu dung lượng dùng để nén khối ảnh lớn hơn ngưỡng này thì khối được đánh dấu là có khả năng có chuyển động; ngược lại, đánh dấu là khối nền. Thông tin có hay không có chuyển động trong từng khối ảnh được tiếp tục sử dụng trong các bước tiếp theo. Ở đây, tốt hơn là, ngưỡng Ts được xác định là một giá trị trong khoảng từ 100 đến 200 hoặc người sử dụng có thể tự thiết lập tùy thuộc vào loại video nguồn.

Bước 202 gom các khối ảnh được đánh dấu là có khả năng chuyển động nằm liền kề (theo 8 hướng) thành các vùng và coi đó là những vùng có khả năng có chuyển động thực. Các vùng này sẽ được xem xét trong quá trình tiếp theo.

Bước 203 kiểm tra tính đồng hướng của các vector chuyển động của các khối ảnh trong từng vùng có khả năng chuyển động bằng cách tính tỉ lệ giữa tổng số các khối ảnh có

vector chuyển động có cùng hướng và tổng số các khối ảnh trong vùng này. Nếu tỉ lệ này là cao ($T_d = 60\% - 70\%$) thì đối tượng đó được coi là đối tượng chuyển động thật, và chuyển sang bước 204, ngược lại thì chuyển sang bước 205 để xác định có phải là chuyển động giả hay không.

Bước 204 mở rộng vùng khối ảnh có chuyển động bằng việc bổ sung các khối ảnh lân cận thuộc nền có các vector chuyển động cùng hướng và xấp xỉ về độ dài với các vector chuyển động của các khối ảnh trong vùng. Trong bước này, tốt hơn là, hai vector chuyển động được gọi là xấp xỉ về độ dài nếu độ dài của một vector lớn hơn 70% và nhỏ hơn 130% độ dài của vector còn lại. Tiếp theo, chuyển thông tin của vùng sang bước 206.

Bước 205 kiểm tra độ tập trung của các khối ảnh chuyển động trong vùng chuyển động bằng cách tính tỉ lệ giữa số khối ảnh trong vùng với số khối ảnh ở biên thuộc vùng. Theo một phương án, độ tập trung được tính bằng 2 thông số chu vi và diện tích. Tính chu vi của vùng chuyển động bằng cách xét các khối ảnh thuộc vùng, nếu khối ảnh nào liền kề với khối ảnh được đánh dấu là không có chuyển động thì chu vi sẽ tăng lên 1 còn diện tích là số khối ảnh trong vùng. Một vùng chuyển động có tỉ lệ diện tích trên chu vi lớn hơn ngưỡng T_c thì vùng đó được coi là có độ tập trung cao và được xác định là có chuyển động thật. Ngược lại thì đánh dấu vùng này là vùng nền.

Ở đây, tốt hơn là, T_c được xác định là một giá trị trong khoảng từ 1 đến 1,2.

Bước 206 đánh dấu vùng chuyển động từ thông tin của các bước 204 và 205. Thông tin vùng chuyển động được chuyển sang bước 208.

Bước 207 đánh dấu nền từ các thông tin của bước 205. Thông tin vùng nền được chuyển sang bước 208.

Bước 208 tổng hợp vùng chuyển động và không có chuyển động từ bước 206 và 207. Thể hiện thông tin kết quả của tiến trình phân tích, phân loại chuyển động. Kết quả phân tích của mỗi khung hình được thể hiện bởi một mảng hai chiều, mỗi vị trí trên mảng có giá trị 1 hoặc 0 thể hiện khối ảnh tại vị trí tương ứng có chuyển động (1) hay không có chuyển động (0). Thông tin này được sử dụng cho phương pháp tổng hợp thông tin mô tả chuyển động ở bước 3 trên hình 1.

Hình 4 thể hiện sơ đồ của bước 3 tổng hợp thông tin mô tả chuyển động trên hình 1.

Phương pháp này bắt đầu bằng bước tính trọng số chuyển động của các khối ảnh trong khoảng thời gian T (bước 301), đối với mỗi vị trí, trọng số chuyển động của các khối ảnh tại các thời điểm (tính theo khung hình) trong khoảng thời gian T được tính như sau:

nếu khối ảnh có chuyển động tại thời điểm đang xét thì trọng số chuyển động tại thời điểm đó bằng số đếm thời điểm có chuyển động liên tiếp trước đó,

ngược lại nếu khối ảnh tại thời điểm đang xét không có chuyển động thì trọng số bằng 0,

trọng số chuyển động của mỗi vị trí trên khung hình tổng hợp sau khoảng thời gian T bằng tổng cộng dồn theo thời gian các trọng số tại tất cả các thời điểm trong khoảng thời gian T đó.

Qua khảo sát trên các video, một vật thể chuyển động (người đi bộ, xe ô tô, xe máy..) đi qua 1 khối ảnh cần ít nhất từ 3 đến 5 khung hình. Như vậy nếu có một vật thể chuyển động đi qua một khối ảnh, tổng giá trị trọng số của khối ảnh đó trong khoảng thời gian T sẽ lớn. Trong khi đó, chuyển động giả gây ra bởi nhiễu làm cho tổng giá trị trọng số này nhỏ và có cách biệt lớn với trường hợp có chuyển động thật đi qua. Từ đó, có thể lựa chọn giá trị hợp lý để phân biệt được chuyển động thực và chuyển động gây ra bởi nhiễu.

Bước 302 đánh giá mức độ chuyển động để thực hiện phân lớp chuyển động cho từng vị trí trong khung hình tổng hợp sau khoảng thời gian T dựa trên trọng số tính được ở bước trước, mức độ chuyển động được chia thành 4 mức: không có chuyển động (00), ít chuyển động hoặc nhiễu (01), có chuyển động (10), và có nhiều chuyển động (11).

Bước 303 lưu các giá trị mức độ chuyển động từ bước 302 vào mảng hai chiều phục vụ cho các bước tiếp theo.

Hình 5 và Hình 6 thể hiện cấu trúc dữ liệu phân cấp và phương pháp phân chia dữ liệu lưu trữ trên phương tiện lưu trữ, hoặc thiết bị có chứa phương tiện lưu trữ, theo không gian và thời gian để chứa thông tin tổng hợp mô tả chuyển động. Cấu trúc dữ liệu này bao gồm: mức 1 là định danh nguồn video (401, 401a); mức 2 là định danh đoạn video (402, 402a); mức 3 là theo một trục không gian của khung hình (403, 403a); mức 4 là theo một trục không gian khác của khung hình (404, 404a), và mức 5 là theo thời gian của video (405, 405a).

Các khối 401, 401a là mức 1 định danh nguồn video được xác định là phương tiện lưu trữ hoặc luồng video trực tuyến. Ở đây, tốt hơn là, mức 1 là một thư mục trong hệ thống

tệp tin (file) của phương tiện lưu trữ có tên là tên của file video nguồn hoặc địa chỉ của nguồn video trực tuyến.

Các khối 402, 402a là mức 2 định danh đoạn video được xác định là đoạn video từ nguồn định danh ở mức gốc trong một khoảng thời gian nào đó. Ở đây, tốt hơn là, mức 2 là một thư mục trong hệ thống file của phương tiện lưu trữ có tên là chuỗi ký tự gồm tên của tệp tin (file) video nguồn hoặc địa chỉ của nguồn video trực tuyến cùng nối với chuỗi ký tự mô tả thời gian bắt đầu và thời gian thời gian kết thúc của đoạn video.

Các khối 403, 403a là mức 3 theo một trục không gian của khung hình trong đoạn video bao gồm các khối ảnh cùng chiều ngang. Ở đây, tốt hơn là, mức 3 là một thư mục trong hệ thống tệp tin của phương tiện lưu trữ có tên là chuỗi ký tự mô tả số thứ tự của lát cắt ngang.

Các khối 404, 404a là mức 4 theo một trục không gian khác của khung hình trong đoạn video bao gồm các khối theo chiều dọc của mức 2. Ở đây, tốt hơn là, mức 4 là một thư mục trong hệ thống file của phương tiện lưu trữ có tên là chuỗi ký tự mô tả số thứ tự của khối theo chiều thời gian của lát cắt dọc.

Các khối 405, 405a là mức 5 theo thời gian của video là một chuỗi giá trị biểu diễn mức độ chuyển động của khối ảnh: 00 (không có chuyển động) hoặc 01 (ít chuyển động hoặc nhiều) hoặc 10 (có chuyển động) hoặc 11 (có nhiều chuyển động). Mỗi giá trị biểu diễn mức độ chuyển động của khối ảnh trong 1 khung hình. Mỗi giá trị này được lưu trữ dưới dạng 02 bit nhị phân. Ở đây, tốt hơn là, mức 5 là một tệp trong hệ thống file của phương tiện lưu trữ chứa dữ liệu tổng hợp chuyển động của khối ảnh trong khoảng thời gian T, trong đó dữ liệu tổng hợp chuyển động là 4 mức độ chuyển động được biểu diễn dưới dạng nhị phân 2 bit tương ứng là 00, 01, 10, 11.

Yêu cầu bảo hộ

1. Phương pháp phân tích và lưu trữ thông tin mô tả chuyển động trong nội dung video, phương pháp này có thể thực hiện được trên các thiết bị tính toán, phương pháp này bao gồm các bước:

tính toán thông tin mô tả chuyển động từ luồng dữ liệu video (1), trong đó việc tính toán này bao gồm các bước:

giải nén entropy luồng dữ liệu dưới dạng chuỗi nhị phân của video (101), để

tính ra các thông tin cơ bản của từng khối ảnh (102) gồm dung lượng dùng để nén và các vectơ chuyển động, và

tính toán thông tin mô tả chuyển động (103), để

đưa ra dung lượng (S) dùng để nén của từng khối ảnh, độ dài vectơ chuyển động (L), hướng của vectơ chuyển động (D) đã được chuẩn hóa;

phân tích, phân loại chuyển động (2) bao gồm các bước:

đánh dấu khối ảnh là chuyển động nếu dung lượng dùng để nén khối ảnh này (S) lớn hơn một ngưỡng T_s , ngược lại thì đánh dấu khối ảnh thuộc nền,

gom các khối ảnh được đánh dấu là có khả năng chuyển động nằm liền kề (theo các hướng chuẩn) thành các vùng,

kiểm tra tính đồng hướng của các vectơ chuyển động của các khối ảnh trong từng vùng bằng cách tính tỉ lệ giữa tổng số các khối ảnh có vectơ chuyển động có cùng một hướng cơ bản và tổng số các khối ảnh trong vùng này, trong đó:

nếu tỷ lệ vectơ chuyển động trong một vùng có cùng một hướng cơ bản lớn hơn một ngưỡng T_d thì,

mở rộng vùng khối ảnh này bằng việc bổ sung các khối ảnh lân cận thuộc nền có các vectơ chuyển động cùng hướng và *xấp xỉ về độ dài* với các vectơ chuyển động của các khối ảnh trong vùng,

và đánh dấu vùng này là vùng chuyển động,

ngược lại kiểm tra độ tập trung của các khối ảnh chuyển động bằng cách tính tỉ lệ giữa số khối ảnh trong vùng với số khối ảnh ở biên thuộc vùng, nếu tỉ lệ này lớn hơn ngưỡng T_c thì đánh dấu vùng này là vùng chuyển động, ngược lại thì đánh dấu vùng này là vùng nền;

tổng hợp thông tin mô tả chuyển động (3) bao gồm các bước:

tính trọng số chuyển động của các khối ảnh trong khoảng thời gian T (301), đối với một vị trí, trọng số chuyển động của các khối ảnh tại các thời điểm (tính theo khung hình) trong khoảng thời gian T được tính như sau:

nếu khối ảnh có chuyển động tại thời điểm đang xét thì trọng số chuyển động tại thời điểm đó bằng số thời điểm có chuyển động liên tiếp trước đó,

ngược lại nếu khối ảnh tại thời điểm đang xét không có chuyển động thì trọng số bằng 0,

trọng số chuyển động của mỗi vị trí trên khung hình tổng hợp sau khoảng thời gian T bằng tổng cộng dồn theo thời gian các trọng số tại tất cả các thời điểm trong khoảng thời gian T đó,

đánh giá mức độ chuyển động (302) để thực hiện phân lớp chuyển động cho từng vị trí trong khung hình tổng hợp sau khoảng thời gian T dựa trên trọng số tính được ở bước trước, mức độ chuyển động được chia thành các mức: không có chuyển động (00), ít chuyển động hoặc nhiều (01), có chuyển động (10), và có nhiều chuyển động (11);

lưu các giá trị mức độ chuyển động từ bước đánh giá mức độ chuyển động (302) vào mảng hai chiều.

2. Phương pháp theo điểm 1, trong đó khối ảnh được xác định là vùng hình vuông trong khung hình của video có kích thước 16×16 điểm ảnh.

3. Phương pháp theo điểm 1, trong đó dung lượng dùng để nén một khối ảnh được tính là hiệu giữa vị trí bit khi bắt đầu và khi kết thúc giải nén một khối ảnh từ luồng dữ liệu nhị phân của video.

4. Phương pháp theo điểm 1, trong đó hướng cơ bản của vectơ chuyển động được xác định gồm 8 hướng theo hệ trục tọa độ Đề Các là: hướng 1: $(0 - 45^\circ]$, hướng 2: $(45 -$

$90^\circ]$, hướng 3: $(90 - 135^\circ]$, hướng 4: $(135 - 180^\circ]$, hướng 5: $(180 - 225^\circ]$, hướng 6: $(225 - 270^\circ]$, hướng 7: $(270 - 315^\circ]$, và hướng 8: $(315 - 360^\circ]$.

5. Phương pháp theo điểm 1, trong đó ngưỡng T_s được xác định là một giá trị trong khoảng từ 100 đến 200 hoặc người sử dụng có thể tự thiết lập tùy thuộc vào loại video nguồn.

6. Phương pháp theo điểm 1, trong đó ngưỡng T_d được xác định là một giá trị trong khoảng từ 60% đến 70%.

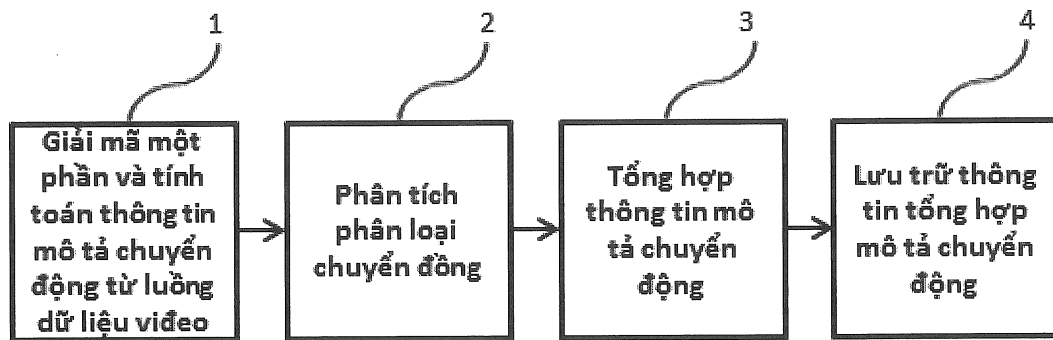
7. Phương pháp theo điểm 1, trong đó hai vectơ chuyển động được gọi là xấp xỉ về độ dài nếu độ dài của một vectơ lớn hơn 70% và nhỏ hơn 130% độ dài của vectơ còn lại.

8. Phương pháp theo điểm 1, trong đó ngưỡng T_c được xác định là một giá trị trong khoảng từ 1 đến 1,2.

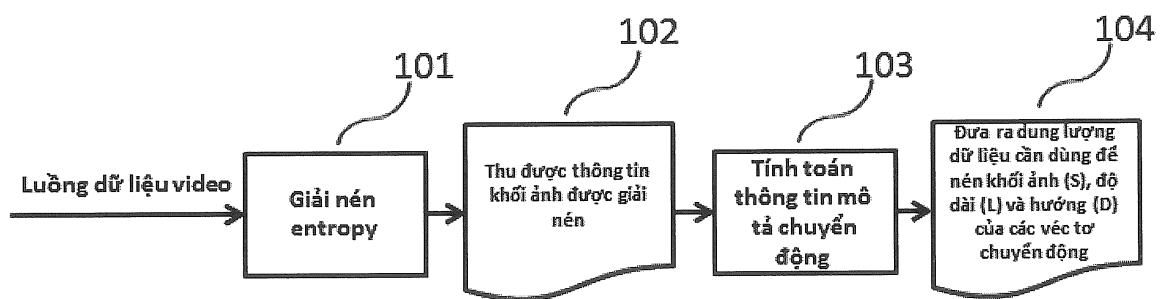
9. Phương pháp theo điểm 1, trong đó khoảng thời gian T của video được xác định là một khoảng thời gian để video có ít nhất 2 khung hình và nhiều nhất là 300 khung hình.

10. Phương pháp theo điểm 1, trong đó khung hình tổng hợp là một ma trận số hai chiều có kích thước bằng độ phân giải của một khung hình trong video chia cho kích thước của khối ảnh, trong đó mỗi phần tử của ma trận chứa dữ liệu tổng hợp chuyển động của khối ảnh trong khoảng thời gian T .

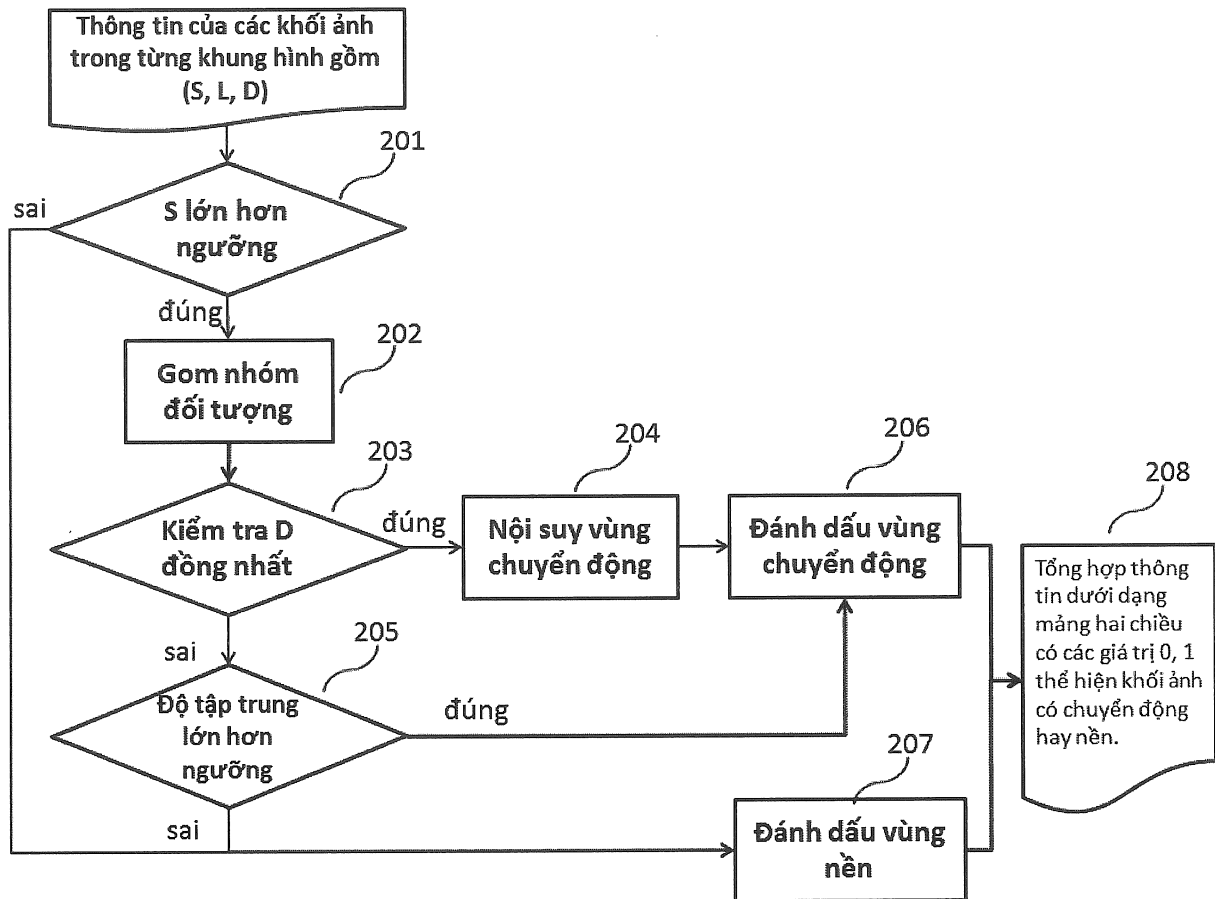
11. Phương pháp theo điểm 1, trong đó 4 mức độ chuyển động được biểu diễn dưới dạng nhị phân 2 bit tương ứng là 00, 01, 10, 11.



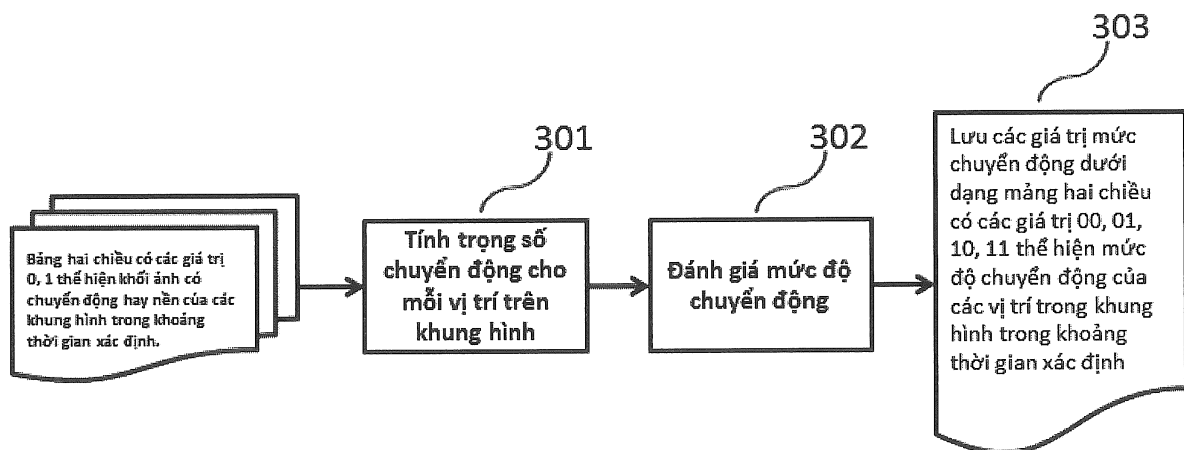
Hình 1



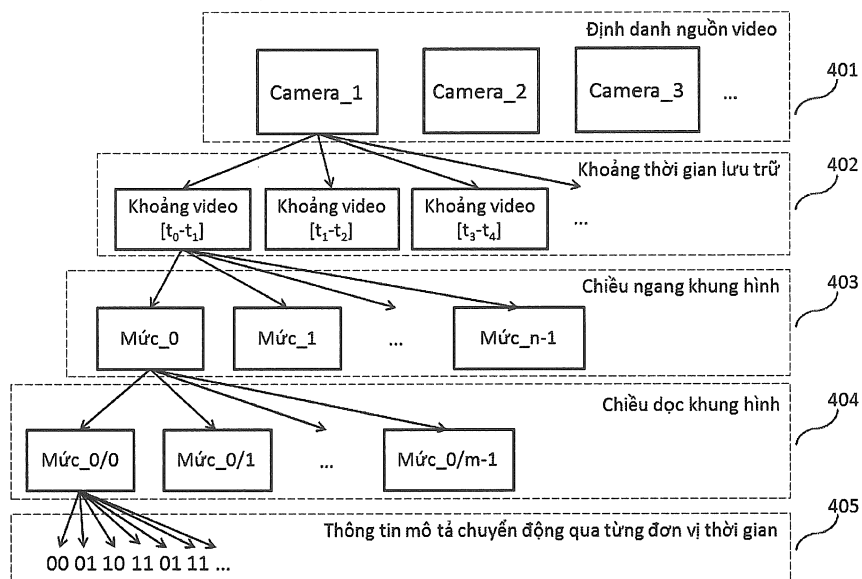
Hình 2



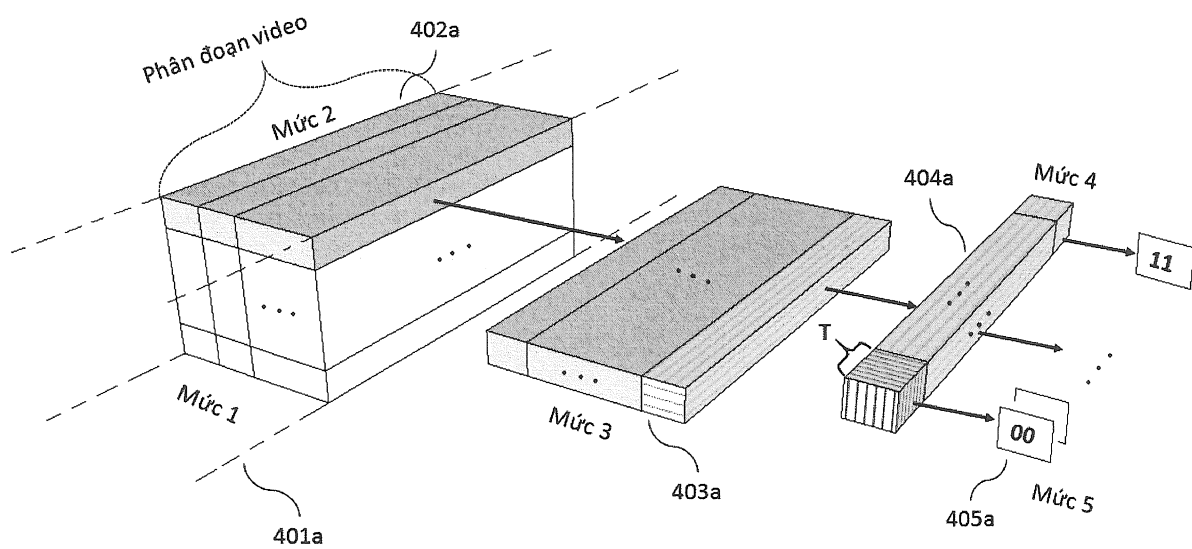
Hình 3



Hình 4



Hình 5



Hình 6