



(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)



1-0048258

(51)^{2020.01} H04N 19/00

(13) B

(21) 1-2021-06304

(22) 11/03/2020

(86) PCT/US2020/022136 11/03/2020

(87) WO2020/185922 17/09/2020

(30) 62/816,749 11/03/2019 US; 62/832,132 10/04/2019 US

(45) 25/07/2025 448

(43) 27/12/2021 405A

(73) HUAWEI TECHNOLOGIES CO., LTD. (CN)

Huawei Administration Building, Bantian, Longgang District, Shenzhen, Guangdong
518129, China

(72) WANG, Ye-Kui (US); HENDRY, Fnu (ID).

(74) Công ty Luật TNHH T&G (TGVN)

(54) PHƯƠNG PHÁP GIẢI MÃ, PHƯƠNG PHÁP MÃ HÓA, THIẾT BỊ TẠO MÃ
VIDEO, PHƯƠNG TIỆN ĐỌC ĐƯỢC BẰNG MÁY TÍNH, BỘ GIẢI MÃ VÀ BỘ
MÃ HÓA

(21) 1-2021-06304

(57) Sáng chế đề cập đến cơ chế tạo mã video. Cụ thể, sáng chế đề cập đến phương pháp giải mã, phương pháp mã hóa, thiết bị tạo mã video, phương tiện đọc được bằng máy tính, bộ giải mã và bộ mã hóa. Cơ chế này bao gồm việc nhận dòng bit bao gồm các hình ảnh con được liên kết với hình ảnh và cờ. Các hình ảnh con có trong các đơn vị lớp trừu tượng mạng (NAL) của lớp tạo mã video (VCL). Giá trị loại đơn vị NAL là giống nhau đối với tất cả các đơn vị NAL VCL được liên kết với hình ảnh khi cờ được thiết đặt ở giá trị thứ nhất. Khi cờ được thiết đặt ở giá trị thứ hai, giá trị loại đơn vị NAL thứ nhất đối với các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh khác với giá trị loại đơn vị NAL thứ hai đối với các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh. Các hình ảnh con được giải mã dựa trên các giá trị loại đơn vị NAL. Các hình ảnh con được chuyển tiếp để hiển thị dưới dạng một phần của chuỗi video được giải mã.

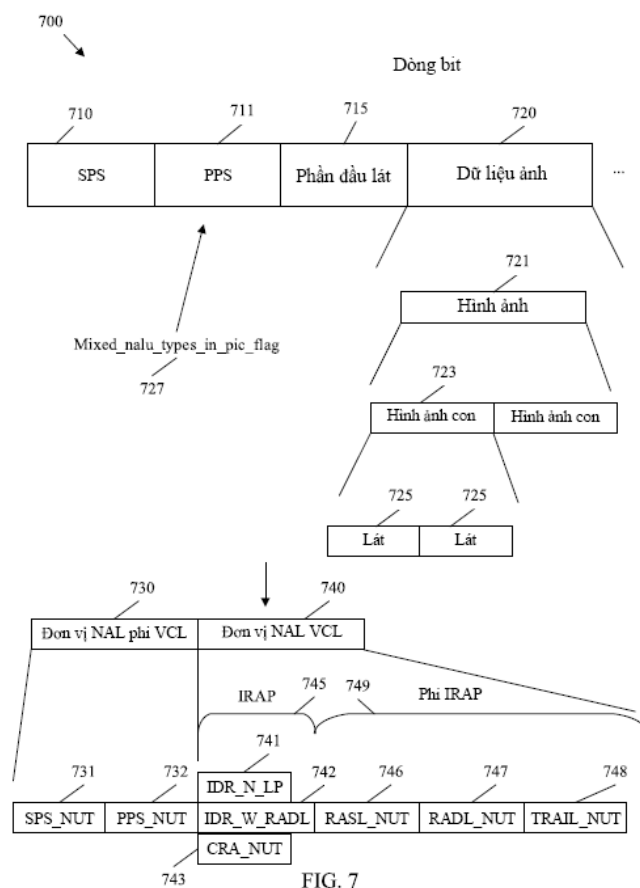


FIG. 7

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập chung đến việc tạo mã video, và cụ thể là đề cập đến việc tạo mã các hình ảnh con của các hình ảnh khi tạo mã video.

Tình trạng kỹ thuật của sáng chế

Lượng dữ liệu video cần để mô tả thậm chí video tương đối ngắn có thể là quan trọng, mà có thể dẫn đến những khó khăn khi dữ liệu cần được tạo luồng hoặc nếu không thì được truyền thông qua mạng truyền thông với dung lượng dải thông hạn chế. Vì vậy, dữ liệu video thường được nén trước khi được truyền thông qua các mạng viễn thông đương thời. Kích thước của video cũng có thể là vấn đề khi video được lưu trữ trên thiết bị lưu trữ do các tài nguyên bộ nhớ có thể bị giới hạn. Các thiết bị nén video thường sử dụng phần mềm và/hoặc phần cứng ở nguồn để tạo mã dữ liệu video trước khi truyền hoặc lưu trữ, nhờ đó làm giảm lượng dữ liệu cần để đại diện cho các ảnh video số. Dữ liệu đã nén sau đó được nhận tại điểm đến bởi thiết bị giải nén video để giải mã dữ liệu video. Với các tài nguyên mạng hạn chế và các nhu cầu luôn tăng lên về chất lượng video cao hơn, các kỹ thuật nén và giải nén cải tiến để nâng cao tỷ lệ nén với việc đánh đổi một ít cho đến không phải đánh đổi về chất lượng ảnh luôn được mong đợi.

Bản chất kỹ thuật của sáng chế

Theo một phương án, sáng chế bao gồm phương pháp được thực hiện trong bộ giải mã, phương pháp này bao gồm các bước: nhận, bởi bộ nhận của bộ giải mã, dòng bit bao gồm các hình ảnh con được liên kết với hình ảnh và cờ, trong đó các hình ảnh con có trong các đơn vị lớp trừu tượng mạng (network abstraction layer, NAL) của lớp tạo mã video (video coding layer, VCL) (NAL VCL); xác định, bởi bộ xử lý, giá trị loại đơn vị NAL thứ nhất là giống nhau đối với tất cả các đơn vị NAL VCL được liên kết với hình ảnh khi cờ được thiết đặt ở giá trị thứ nhất; xác định, bởi bộ xử lý, giá trị loại đơn vị NAL thứ nhất đối với các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh khác với giá trị loại đơn vị NAL thứ hai đối với các đơn vị NAL VCL chứa một hoặc nhiều hình

ảnh con của hình ảnh khi cờ được thiết đặt ở giá trị thứ hai; và giải mã, bởi bộ xử lý, một hoặc nhiều hình ảnh con dựa trên giá trị loại đơn vị NAL thứ nhất hoặc giá trị loại đơn vị NAL thứ hai.

Hình ảnh có thể được phân chia thành nhiều hình ảnh con. Các hình ảnh con này có thể được tạo mã thành các dòng bit con riêng rẽ, mà sau đó có thể được hợp nhất thành dòng bit để truyền đến bộ giải mã. Ví dụ, các hình ảnh con có thể được sử dụng cho các ứng dụng thực tế ảo (virtual reality, VR). Như một ví dụ cụ thể, người dùng có thể chỉ xem một phần của hình ảnh VR vào thời điểm bất kỳ. Theo đó, các hình ảnh con khác nhau có thể được truyền ở các độ phân giải khác nhau sao cho nhiều dải thông hơn có thể được cấp phát cho các hình ảnh con mà rất có khả năng được hiển thị và các hình ảnh con mà ít có khả năng được hiển thị có thể được nén để làm tăng hiệu suất tạo mã. Hơn nữa, các luồng video có thể được mã hóa bằng cách sử dụng các hình ảnh điểm truy cập ngẫu nhiên nội bộ (intra-random access point, IRAP). Hình ảnh IRAP được tạo mã theo dự đoán trong ảnh và có thể được giải mã mà không tham chiếu đến các hình ảnh khác. Các hình ảnh phi IRAP có thể được tạo mã theo dự đoán liên ảnh và có thể được giải mã bằng cách tham chiếu các hình ảnh khác. Các hình ảnh phi IRAP ít hơn đáng kể so với các hình ảnh IRAP. Tuy nhiên, chuỗi video phải bắt đầu giải mã với hình ảnh IRAP do hình ảnh IRAP chứa đủ dữ liệu cần được giải mã mà không tham chiếu các hình ảnh khác. Các hình ảnh IRAP có thể được sử dụng trong các hình ảnh con, và có thể cho phép các sự thay đổi độ phân giải động. Theo đó, hệ thống video có thể truyền nhiều hình ảnh IRAP hơn đối với các hình ảnh con mà có nhiều khả năng được xem hơn (ví dụ, dựa trên công quan sát hiện hành của các người dùng) và ít hình ảnh IRAP hơn đối với các hình ảnh con mà ít có khả năng được xem để tăng hiệu suất tạo mã thêm nữa. Tuy nhiên, các hình ảnh con là một phần của cùng một hình ảnh. Theo đó, sơ đồ này có thể dẫn đến hình ảnh mà chứa của hình ảnh con IRAP và hình ảnh con phi IRAP. Một số hệ thống video không được trang bị để xử lý hình ảnh trộn với cả các vùng IRAP lẫn phi IRAP. Sáng chế bao gồm cờ chỉ báo liệu hình ảnh có được trộn và vì vậy chứa cả các thành phần IRAP và phi IRAP hay không. Dựa trên cờ này, bộ giải mã có thể xử lý các hình ảnh con khác nhau một cách khác nhau khi giải mã để giải mã và hiển thị hình ảnh/các hình ảnh con một cách chính xác. Cờ này có thể được lưu trữ trong PPS và có thể được gọi là `mixed_nalu_types_in_pic_flag`. Như vậy, các cơ chế được bộc lộ cho phép thực hiện chức

năng bổ sung. Hơn nữa, các cơ chế được bộc lộ cho phép các sự thay đổi độ phân giải động khi sử dụng các dòng bit hình ảnh con. Vì vậy, các cơ chế được bộc lộ cho phép các dòng bit hình ảnh con độ phân giải thấp hơn được truyền khi tạo luồng video VR mà không làm giảm đáng kể trải nghiệm của người dùng. Như vậy, các cơ chế được bộc lộ làm tăng hiệu suất tạo mã, và vì vậy làm giảm việc sử dụng các tài nguyên mạng, các tài nguyên bộ nhớ, và/hoặc các tài nguyên xử lý tại bộ mã hóa và bộ giải mã.

Tùy chọn là, theo khía cạnh bất kỳ trong số các khía cạnh trên đây, một cách thực hiện khác của khía cạnh này đề xuất, trong đó dòng bit bao gồm tập hợp thông số hình ảnh (picture parameter set, PPS) bao gồm cờ.

Tùy chọn là, theo khía cạnh bất kỳ trong số các khía cạnh trên đây, một cách thực hiện khác của khía cạnh này đề xuất, trong đó giá trị loại đơn vị NAL thứ nhất chỉ báo hình ảnh chứa hình ảnh con điểm truy cập ngẫu nhiên nội bộ (IRAP), và trong đó giá trị loại đơn vị NAL thứ hai chỉ báo hình ảnh chứa hình ảnh con phi IRAP.

Tùy chọn là, theo khía cạnh bất kỳ trong số các khía cạnh trên đây, một cách thực hiện khác của khía cạnh này đề xuất, trong đó giá trị loại đơn vị NAL thứ nhất bằng làm mới quá trình giải mã tức thời (Instantaneous Decoding Refresh, IDR) với hình ảnh chủ đạo giải mã được truy cập ngẫu nhiên (IDR_W_RADL), IDR không có các hình ảnh chủ đạo (IDR_N_LP), hoặc loại đơn vị NAL truy cập ngẫu nhiên sạch (clean random access, CRA) (CRA_NUT).

Tùy chọn là, theo khía cạnh bất kỳ trong số các khía cạnh trên đây, một cách thực hiện khác của khía cạnh này đề xuất, trong đó giá trị loại đơn vị NAL thứ hai bằng loại đơn vị NAL hình ảnh kế tiếp (TRAIL_NUT), loại đơn vị NAL hình ảnh chủ đạo giải mã được truy cập ngẫu nhiên (RADL_NUT), hoặc loại đơn vị NAL hình ảnh chủ đạo bỏ qua truy cập ngẫu nhiên (random access skipped leading picture, RASL) (RASL_NUT).

Tùy chọn là, theo khía cạnh bất kỳ trong số các khía cạnh trên đây, một cách thực hiện khác của khía cạnh này đề xuất, trong đó cờ là `mixed_nalu_types_in_pic_flag`.

Tùy chọn là, theo khía cạnh bất kỳ trong số các khía cạnh trên đây, một cách thực hiện khác của khía cạnh này đề xuất, trong đó `mixed_nalu_types_in_pic_flag` bằng một khi quy định rằng hình ảnh tham chiếu đến PPS có nhiều hơn một trong số các đơn vị

NAL VCL và các đơn vị NAL VCL không có cùng giá trị của loại đơn vị NAL (`nal_unit_type`), và trong đó `mixed_nalu_types_in_pic_flag` bằng không (zero) khi quy định hình ảnh tham chiếu đến PPS có một hoặc nhiều đơn vị NAL VCL và các đơn vị NAL VCL có cùng giá trị của `nal_unit_type`.

Theo một phương án, sáng chế bao gồm phương pháp được thực hiện trong bộ mã hóa, phương pháp này bao gồm các bước: xác định, bởi bộ xử lý, liệu hình ảnh có chứa các hình ảnh con thuộc các loại khác nhau hay không; mã hóa, bởi bộ xử lý, các hình ảnh con của hình ảnh thành các đơn vị NAL VCL trong dòng bit; mã hóa thành dòng bit, bởi bộ xử lý, cờ được thiết đặt ở giá trị thứ nhất khi giá trị loại đơn vị NAL thứ nhất là giống nhau đối với tất cả các đơn vị NAL VCL được liên kết với hình ảnh và được thiết đặt ở giá trị thứ hai khi giá trị loại đơn vị NAL thứ nhất đối với các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh khác với giá trị loại đơn vị NAL thứ hai đối với các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh; và lưu trữ, bởi bộ nhớ được ghép nối với bộ xử lý, dòng bit để truyền thông đến bộ giải mã.

Hình ảnh có thể được phân chia thành nhiều hình ảnh con. Các hình ảnh con này có thể được tạo mã thành các dòng bit con riêng rẽ, mà sau đó có thể được hợp nhất thành dòng bit để truyền đến bộ giải mã. Ví dụ, các hình ảnh con có thể được sử dụng cho các ứng dụng thực tế ảo (virtual reality, VR). Như một ví dụ cụ thể, người dùng có thể chỉ xem một phần của hình ảnh VR vào thời điểm bất kỳ. Theo đó, các hình ảnh con khác nhau có thể được truyền ở các độ phân giải khác nhau sao cho nhiều dải thông hơn có thể được cấp phát cho các hình ảnh con mà rất có khả năng được hiển thị và các hình ảnh con mà ít có khả năng được hiển thị có thể được nén để làm tăng hiệu suất tạo mã. Hơn nữa, các luồng video có thể được mã hóa bằng cách sử dụng các hình ảnh điểm truy cập ngẫu nhiên nội bộ (intra-random access point, IRAP). Hình ảnh IRAP được tạo mã theo dự đoán trong ảnh và có thể được giải mã mà không tham chiếu đến các hình ảnh khác. Các hình ảnh phi IRAP có thể được tạo mã theo dự đoán liên ảnh và có thể được giải mã bằng cách tham chiếu các hình ảnh khác. Các hình ảnh phi IRAP sát hơn đáng kể so với các hình ảnh IRAP. Tuy nhiên, chuỗi video phải bắt đầu giải mã với hình ảnh IRAP do hình ảnh IRAP chứa đủ dữ liệu cần được giải mã mà không tham chiếu các hình ảnh khác. Các hình ảnh IRAP có thể được sử dụng trong các hình ảnh con, và có thể cho phép các sự thay đổi độ phân giải động. Theo đó, hệ thống video có thể truyền nhiều hình ảnh IRAP

hơn đối với các hình ảnh con mà có nhiều khả năng được xem hơn (ví dụ, dựa trên công quan sát hiện hành của các người dùng) và ít hình ảnh IRAP hơn đối với các hình ảnh con mà ít có khả năng được xem để tăng hiệu suất tạo mã thêm nữa. Tuy nhiên, các hình ảnh con là một phần của cùng một hình ảnh. Theo đó, sơ đồ này có thể dẫn đến hình ảnh mà chứa của hình ảnh con IRAP và hình ảnh con phi IRAP. Một số hệ thống video không được trang bị để xử lý hình ảnh trộn với cả các vùng IRAP lẫn phi IRAP. Sáng chế bao gồm cờ chỉ báo liệu hình ảnh có được trộn và vì vậy chứa cả các thành phần IRAP và phi IRAP hay không. Dựa trên cờ này, bộ giải mã có thể xử lý các hình ảnh con khác nhau một cách khác nhau khi giải mã để giải mã và hiển thị hình ảnh/các hình ảnh con một cách chính xác. Cờ này có thể được lưu trữ trong PPS và có thể được gọi là `mixed_nalu_types_in_pic_flag`. Như vậy, các cơ chế được bộc lộ cho phép thực hiện chức năng bổ sung. Hơn nữa, các cơ chế được bộc lộ cho phép các sự thay đổi độ phân giải động khi sử dụng các dòng bit hình ảnh con. Vì vậy, các cơ chế được bộc lộ cho phép các dòng bit hình ảnh con độ phân giải thấp hơn được truyền khi tạo luồng video VR mà không làm giảm đáng kể trải nghiệm của người dùng. Như vậy, các cơ chế được bộc lộ làm tăng hiệu suất tạo mã, và vì vậy làm giảm việc sử dụng các tài nguyên mạng, các tài nguyên bộ nhớ, và/hoặc các tài nguyên xử lý tại bộ mã hóa và bộ giải mã.

Tùy chọn là, theo khía cạnh bất kỳ trong số các khía cạnh trên đây, một cách thực hiện khác của khía cạnh này đề xuất, còn bao gồm mã hóa PPS thành dòng bit, trong đó cờ được mã hóa thành PPS.

Tùy chọn là, theo khía cạnh bất kỳ trong số các khía cạnh trên đây, một cách thực hiện khác của khía cạnh này đề xuất, trong đó giá trị loại đơn vị NAL thứ nhất chỉ báo hình ảnh chứa hình ảnh con IRAP, và trong đó giá trị loại đơn vị NAL thứ hai chỉ báo hình ảnh chứa hình ảnh con phi IRAP.

Tùy chọn là, theo khía cạnh bất kỳ trong số các khía cạnh trên đây, một cách thực hiện khác của khía cạnh này đề xuất, trong đó giá trị loại đơn vị NAL thứ nhất bằng `IDR_W_RADL`, `IDR_N_LP`, hoặc `CRA_NUT`.

Tùy chọn là, theo khía cạnh bất kỳ trong số các khía cạnh trên đây, một cách thực hiện khác của khía cạnh này đề xuất, trong đó giá trị loại đơn vị NAL thứ hai bằng `TRAIL_NUT`, `RADL_NUT`, hoặc `RASL_NUT`.

Tùy chọn là, theo khía cạnh bất kỳ trong số các khía cạnh trên đây, một cách thực hiện khác của khía cạnh này đề xuất, trong đó cờ là `mixed_nalu_types_in_pic_flag`.

Tùy chọn là, theo khía cạnh bất kỳ trong số các khía cạnh trên đây, một cách thực hiện khác của khía cạnh này đề xuất, trong đó `mixed_nalu_types_in_pic_flag` bằng một khi quy định rằng hình ảnh tham chiếu đến PPS có nhiều hơn một trong số các đơn vị NAL VCL và các đơn vị NAL VCL không có cùng giá trị của `nal_unit_type`, và trong đó `mixed_nalu_types_in_pic_flag` bằng không (zero) khi quy định hình ảnh tham chiếu đến PPS có một hoặc nhiều đơn vị NAL VCL và các đơn vị NAL VCL có cùng giá trị của `nal_unit_type`.

Theo một phương án, sáng chế bao gồm thiết bị tạo mã video bao gồm: bộ xử lý, bộ nhận được ghép nối với bộ xử lý, bộ nhớ được ghép nối với bộ xử lý, và bộ truyền được ghép nối với bộ xử lý, trong đó bộ xử lý, bộ nhận, bộ nhớ, và bộ truyền được tạo cấu hình để thực hiện phương pháp theo khía cạnh bất kỳ trong số các khía cạnh trên đây.

Theo một phương án, sáng chế bao gồm phương tiện đọc được bằng máy tính lâu dài bao gồm sản phẩm chương trình máy tính để sử dụng bởi thiết bị tạo mã video, sản phẩm chương trình máy tính bao gồm các lệnh thực hiện được bằng máy tính được lưu trữ trên phương tiện đọc được bằng máy tính lâu dài sao cho khi được thực hiện bởi bộ xử lý sẽ làm cho thiết bị tạo mã video thực hiện phương pháp theo khía cạnh bất kỳ trong số các khía cạnh trên đây.

Theo một phương án, sáng chế bao gồm bộ giải mã bao gồm: phương tiện nhận để nhận dòng bit bao gồm các hình ảnh con được liên kết với hình ảnh và cờ, trong đó các hình ảnh con có trong các đơn vị NAL VCL; phương tiện xác định để: xác định giá trị loại đơn vị NAL thứ nhất là giống nhau đối với tất cả các đơn vị NAL VCL được liên kết với hình ảnh khi cờ được thiết đặt ở giá trị thứ nhất; và xác định giá trị loại đơn vị NAL thứ nhất đối với các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh khác với giá trị loại đơn vị NAL thứ hai đối với các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh khi cờ được thiết đặt ở giá trị thứ hai; và phương tiện giải mã để giải mã một hoặc nhiều hình ảnh con dựa trên giá trị loại đơn vị NAL thứ nhất hoặc giá trị loại đơn vị NAL thứ hai.

Tùy chọn là, theo khía cạnh bất kỳ trong số các khía cạnh trên đây, một cách thực hiện khác của khía cạnh này đề xuất, trong đó bộ giải mã còn được tạo cấu hình để thực hiện phương pháp theo khía cạnh bất kỳ trong số các khía cạnh trên đây.

Theo một phương án, sáng chế bao gồm bộ mã hóa bao gồm: phương tiện xác định để xác định liệu hình ảnh có chứa các hình ảnh con thuộc các loại khác nhau hay không; phương tiện mã hóa để: mã hóa các hình ảnh con của hình ảnh thành các đơn vị NAL VCL trong dòng bit; và mã hóa thành dòng bit còn được thiết đặt ở giá trị thứ nhất khi giá trị loại đơn vị NAL thứ nhất là giống nhau đối với tất cả các đơn vị NAL VCL được liên kết với hình ảnh và được thiết đặt ở giá trị thứ hai khi giá trị loại đơn vị NAL thứ nhất đối với các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh khác với giá trị loại đơn vị NAL thứ hai đối với các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh; và phương tiện lưu trữ để lưu trữ dòng bit để truyền thông đến bộ giải mã.

Tùy chọn là, theo khía cạnh bất kỳ trong số các khía cạnh trên đây, một cách thực hiện khác của khía cạnh này đề xuất, trong đó bộ mã hóa còn được tạo cấu hình để thực hiện phương pháp theo khía cạnh bất kỳ trong số các khía cạnh trên đây.

Nhằm mục đích làm rõ, phương án bất kỳ trong số các phương án trên đây có thể được kết hợp với một hoặc nhiều phương án bất kỳ trong số các phương án khác trên đây để tạo phương án mới nằm trong phạm vi của sáng chế.

Các dấu hiệu này và các dấu hiệu khác sẽ được hiểu rõ ràng hơn từ phần mô tả chi tiết dưới đây kết hợp với các hình vẽ kèm theo và yêu cầu bảo hộ.

Mô tả vắn tắt các hình vẽ

Để hiểu đầy đủ hơn về sáng chế, có thể tham khảo phần mô tả vắn tắt sau đây, kết hợp với các hình vẽ kèm theo và phần mô tả chi tiết, trong đó các số chỉ dẫn giống nhau đại diện cho các phần giống nhau.

Fig.1 là lưu đồ của phương pháp tạo mã tín hiệu video làm ví dụ.

Fig.2 là hình vẽ giản lược của hệ thống tạo mã và giải mã (codec) làm ví dụ để tạo mã video.

Fig.3 là hình vẽ giản lược minh họa bộ mã hóa video làm ví dụ.

Fig.4 là hình vẽ giản lược minh họa bộ giải mã video làm ví dụ.

Fig.5 là hình vẽ giản lược minh họa chuỗi video được tạo mã làm ví dụ.

Fig.6 là hình vẽ giản lược minh họa các luồng video hình ảnh con được chia tách từ luồng video hình ảnh thực tế ảo (VR).

Fig.7 là hình vẽ giản lược minh họa dòng bit làm ví dụ chứa các hình ảnh với các loại đơn vị lớp trừu tượng (NAL) trộn.

Fig.8 là hình vẽ giản lược của thiết bị tạo mã video làm ví dụ.

Fig.9 là lưu đồ của phương pháp làm ví dụ để mã hóa chuỗi video chứa hình ảnh với các loại đơn vị NAL trộn thành dòng bit.

Fig.10 là lưu đồ của phương pháp làm ví dụ để giải mã chuỗi video chứa hình ảnh với các loại đơn vị NAL trộn từ dòng bit.

Fig.11 là hình vẽ giản lược của hệ thống làm ví dụ để tạo mã chuỗi video chứa hình ảnh với các loại đơn vị NAL trộn thành dòng bit.

Mô tả chi tiết sáng chế

Nên được hiểu từ đầu rằng mặc dù cách thực hiện minh họa của một hoặc nhiều phương án được đưa ra dưới đây, nhưng các hệ thống và/hoặc các phương pháp được bộc lộ có thể được thực hiện nhờ sử dụng số lượng kỹ thuật bất kỳ, liệu hiện đã được biết đến hay đang tồn tại. Sáng chế sẽ hoàn toàn không bị giới hạn ở các cách thực hiện minh họa, các hình vẽ, và các kỹ thuật được minh họa dưới đây, bao gồm các thiết kế và các cách thực hiện làm ví dụ được minh họa và được mô tả ở đây, mà có thể được cải biến nằm trong phạm vi của yêu cầu bảo hộ kèm theo cùng với phạm vi tương đương đầy đủ của chúng.

Các từ viết tắt sau đây được sử dụng ở đây, chuỗi video được tạo mã (Coded Video Sequence, CVS), bộ đệm hình ảnh được giải mã (Decoded Picture Buffer, DPB), làm mới quá trình giải mã tức thời (IDR), điểm truy cập ngẫu nhiên nội bộ (IRAP), bit có trọng số nhỏ nhất (LSB), bit có trọng số lớn nhất (MSB), lớp trừu tượng mạng (NAL), đếm thứ tự hình ảnh (Picture Order Count, POC), phân tải dữ liệu chuỗi bai thô (Raw Byte Sequence

Payload, RBSP), tập hợp thông số chuỗi (Sequence Parameter Set, SPS), và bản phác thảo công việc (Working Draft, WD).

Nhiều kỹ thuật nén video có thể được sử dụng để làm giảm kích thước của các tệp video với lượng mất dữ liệu tối thiểu. Ví dụ, các kỹ thuật nén video có thể bao gồm việc thực hiện dự đoán không gian (ví dụ, trong hình ảnh) và/hoặc dự đoán thời gian (ví dụ, liên hình ảnh) để làm giảm hoặc loại bỏ sự dư thừa dữ liệu trong các chuỗi video. Đối với việc tạo mã video dựa trên khối, lát video (ví dụ, hình ảnh video hoặc một phần của hình ảnh video) có thể được phân chia thành các khối video, mà cũng có thể được gọi là các khối cây (treeblocks), các khối cây tạo mã (coding tree block, CTB), các đơn vị cây tạo mã (coding tree unit, CTU), các đơn vị tạo mã (coding unit, CU), và/hoặc các nút tạo mã. Các khối video trong lát được tạo mã trong ảnh (I) của hình ảnh được tạo mã sử dụng dự đoán không gian đối với các mẫu tham chiếu trong các khối lân cận trong cùng một hình ảnh. Các khối video trong lát dự đoán đơn hướng (P) hoặc dự đoán hai hướng (B) được tạo mã liên ảnh của hình ảnh có thể được tạo mã bằng cách sử dụng dự đoán không gian đối với các mẫu tham chiếu trong các khối lân cận trong cùng một hình ảnh hoặc dự đoán thời gian đối với các mẫu tham chiếu trong các hình ảnh tham chiếu khác. Các hình ảnh có thể được gọi là các khung và/hoặc các ảnh, và các hình ảnh tham chiếu có thể được gọi là các khung tham chiếu và/hoặc các ảnh tham chiếu. Dự đoán không gian hoặc thời gian dẫn đến khối dự đoán đại diện cho khối ảnh. Dữ liệu dư thể hiện các sự chênh lệch điểm ảnh giữa khối ảnh gốc và khối dự đoán. Do đó, khối được tạo mã liên ảnh được mã hóa theo vector chuyển động mà trỏ đến khối của các mẫu tham chiếu tạo thành khối dự đoán, và dữ liệu dư chỉ báo sự chênh lệch giữa khối được tạo mã và khối dự đoán. Khối được tạo mã trong ảnh được mã hóa theo chế độ tạo mã trong ảnh và dữ liệu dư. Để nén thêm, dữ liệu dư có thể được biến đổi từ miền điểm ảnh sang miền biến đổi. Điều này tạo ra các hệ số biến đổi dư, mà có thể được lượng tử hóa. Ban đầu, các hệ số biến đổi được lượng tử hóa có thể được sắp xếp trong mảng hai chiều. Các hệ số biến đổi được lượng tử hóa có thể được quét để tạo ra vector một chiều của các hệ số biến đổi. Tạo mã entropi có thể được áp dụng để thu được kết quả nén thậm chí là tốt hơn. Các kỹ thuật nén video này được thảo luận chi tiết hơn dưới đây.

Để đảm bảo video đã mã hóa có thể được giải mã chính xác, video được mã hóa và được giải mã theo các tiêu chuẩn tạo mã video tương ứng. Các tiêu chuẩn tạo mã video

bao gồm lĩnh vực chuẩn hóa (Standardization Sector) thuộc Hiệp hội viễn thông quốc tế (International Telecommunication Union, ITU) (ITU-T) H.261, nhóm chuyên gia hình ảnh động (Motion Picture Experts Group, MPEG) của Tổ chức chuẩn hóa quốc tế (International Organization for Standardization)/Ủy ban kỹ thuật điện quốc tế (International Electrotechnical Commission) (ISO/IEC) MPEG-1 phần 2, ITU-T H.262 hoặc ISO/IEC MPEG-2 phần 2, ITU-T H.263, ISO/IEC MPEG-4 phần 2, Tạo mã video tiên tiến (Advanced Video Coding, AVC), cũng được biết đến với tên gọi là ITU-T H.264 hoặc ISO/IEC MPEG-4 phần 10, và Tạo mã video hiệu suất cao (High Efficiency Video Coding, HEVC), cũng được biết đến với tên gọi là ITU-T H.265 hoặc MPEG-H phần 2. AVC bao gồm các loại mở rộng như tạo mã video thay đổi được (Scalable Video Coding, SVC), tạo mã video đa cảnh nhìn (Multiview Video Coding, MVC) và tạo mã video đa cảnh nhìn cộng độ sâu (Multiview Video Coding plus Depth, MVC+D), và AVC ba chiều (3D) (3D-AVC). HEVC bao gồm các loại mở rộng như HEVC thay đổi được (SHVC), HEVC đa cảnh nhìn (MV-HEVC), và 3D HEVC (3D-HEVC). Nhóm chuyên gia video chung (joint video experts team, JVET) của ITU-T và ISO/IEC đã bắt đầu phát triển tiêu chuẩn tạo mã video được gọi là tạo mã video đa năng (VVC). VVC có trong bản phác thảo công việc (WD), mà bao gồm JVET-M1001-v6 mà cung cấp phần mô tả thuật toán, phần mô tả phía bộ mã hóa của VVC WD, và phần mềm tham chiếu.

Các hệ thống tạo mã video có thể mã hóa video bằng cách sử dụng các hình ảnh IRAP và các hình ảnh phi IRAP. Các hình ảnh IRAP là các hình ảnh được tạo mã theo dự đoán trong ảnh mà dùng làm các điểm truy cập ngẫu nhiên đối với chuỗi video. Các khối dự đoán trong ảnh của hình ảnh được tạo mã bằng cách tham chiếu đến các khối khác trong cùng một hình ảnh. Điều này ngược lại với các hình ảnh phi IRAP mà sử dụng dự đoán liên ảnh. Trong dự đoán liên ảnh, các khối của hình ảnh hiện hành được tạo mã bằng cách tham chiếu đến các khối khác trong hình ảnh tham chiếu mà khác với hình ảnh hiện hành. Do hình ảnh IRAP được tạo mã mà không tham chiếu đến các hình ảnh khác, nên hình ảnh IRAP có thể được giải mã mà không phải là trước tiên giải mã các hình ảnh khác bất kỳ. Theo đó, bộ giải mã có thể bắt đầu giải mã chuỗi video tại hình ảnh IRAP bất kỳ. Ngược lại, hình ảnh phi IRAP được tạo mã khi tham chiếu đến các hình ảnh khác, và vì vậy bộ giải mã thường không thể bắt đầu giải mã chuỗi video tại hình ảnh phi IRAP. Các hình ảnh IRAP cũng làm mới DPB. Điều này là do hình ảnh IRAP là điểm bắt đầu đối với

CVS, và các hình ảnh trong CVS không tham chiếu đến các hình ảnh trong CVS trước. Như vậy, các hình ảnh IRAP cũng có thể dùng các lỗi tạo mã liên quan đến dự đoán liên ảnh do các lỗi này không thể lan truyền qua hình ảnh IRAP. Tuy nhiên, các hình ảnh IRAP lớn hơn đáng kể so với các hình ảnh phi IRAP từ quan điểm kích thước dữ liệu. Như vậy, chuỗi video thường bao gồm nhiều hình ảnh phi IRAP với số lượng nhỏ hơn của các hình ảnh IRAP rải rác để cân bằng hiệu suất tạo mã với chức năng. Ví dụ, sáu mươi khung CVS có thể bao gồm một hình ảnh IRAP và năm mươi chín hình ảnh phi IRAP.

Trong một số trường hợp, các hệ thống tạo mã video có thể được sử dụng để tạo mã video thực tế ảo (VR), mà cũng có thể được gọi là video 360 độ. Video VR có thể bao gồm hình cầu của nội dung video được hiển thị như thể người dùng ở tâm của hình cầu. Chỉ một phần của hình cầu, được gọi là cổng quan sát, được hiển thị cho người dùng. Ví dụ, người dùng có thể sử dụng thiết bị hiển thị lắp trên đầu (HMD) mà chọn và hiển thị cổng quan sát của hình cầu dựa trên chuyển động đầu của người dùng. Điều này mang lại ấn tượng về sự có mặt về mặt vật lý trong không gian ảo như được thể hiện bởi video. Để đạt được kết quả này, mỗi hình ảnh của chuỗi video bao gồm toàn bộ hình cầu của dữ liệu video ở thời điểm tương ứng. Tuy nhiên, chỉ một phần nhỏ (ví dụ, một cổng quan sát đơn lẻ) của hình ảnh được hiển thị cho người dùng. Phần còn lại của hình ảnh được loại bỏ mà không được thực hiện. Toàn bộ hình ảnh thường được truyền sao cho cổng quan sát khác nhau có thể được chọn theo cách động và được hiển thị đáp lại chuyển động đầu của người dùng. Cách tiếp cận này có thể dẫn đến các kích thước tệp video rất lớn.

Để nâng cao hiệu suất tạo mã, một số hệ thống chia các hình ảnh thành các hình ảnh con. Hình ảnh con là vùng không gian xác định của hình ảnh. Mỗi hình ảnh con chứa cổng quan sát tương ứng của hình ảnh. Video có thể được mã hóa ở hai hoặc nhiều hơn hai độ phân giải. Mỗi độ phân giải được mã hóa thành một dòng bit con khác nhau. Khi người dùng tạo luồng video VR, hệ thống tạo mã có thể hợp nhất các dòng bit con thành dòng bit để truyền dựa trên cổng quan sát hiện hành khi sử dụng bởi người dùng. Cụ thể, cổng quan sát hiện hành được thu nhận từ dòng bit con độ phân giải cao và các cổng quan sát mà không được xem được thu nhận từ (các) dòng bit độ phân giải thấp. Theo cách này, video chất lượng cao nhất được hiển thị cho người dùng và video chất lượng thấp hơn được loại bỏ. Trong trường hợp người dùng chọn cổng quan sát mới, video độ phân giải thấp hơn được trình diễn cho người dùng. Bộ giải mã có thể yêu cầu rằng cổng quan sát

mới nhận video độ phân giải cao hơn. Sau đó bộ mã hóa có thể thay đổi quy trình hợp nhất theo đó. Một khi hình ảnh IRAP được đạt đến, bộ giải mã có thể bắt đầu giải mã chuỗi video độ phân giải cao hơn tại cổng quan sát mới. Cách tiếp cận này làm tăng đáng kể độ nén video mà không tác động tiêu cực đến trải nghiệm xem của người dùng.

Một vấn đề với cách tiếp cận trên đây là độ dài của việc thời gian cần để thay đổi các độ phân giải dựa trên độ dài của thời gian cho đến khi hình ảnh IRAP được đạt đến. Điều này là do bộ giải mã không thể bắt đầu giải mã chuỗi video khác nhau ở hình ảnh phi IRAP như được mô tả trên đây. Một cách tiếp cận để làm giảm độ trễ này là để bao gồm nhiều hình ảnh IRAP hơn. Tuy nhiên, điều này dẫn đến việc tăng kích thước tệp. Để cân bằng chức năng với hiệu suất tạo mã, các cổng quan sát/các hình ảnh con khác nhau có thể bao gồm các hình ảnh IRAP ở các tần số khác nhau. Ví dụ, các cổng quan sát mà có nhiều khả năng được xem hơn có thể có nhiều hình ảnh IRAP hơn so với các cổng quan sát khác. Ví dụ, trong ngữ cảnh bóng rổ, các cổng quan sát liên quan đến các rổ và/hoặc sân trung tâm có thể bao gồm các hình ảnh IRAP ở tần số lớn hơn so với các cổng quan sát mà xem các khán đài hoặc trần mà các cổng quan sát này ít có khả năng được xem bởi người dùng.

Cách tiếp cận này dẫn đến các vấn đề khác. Cụ thể, các hình ảnh con mà chứa các cổng quan sát là một phần của hình ảnh đơn lẻ. Khi các hình ảnh con khác nhau có các hình ảnh IRAP ở các tần số khác nhau, một số trong số các hình ảnh bao gồm cả các hình ảnh con IRAP lẫn các hình ảnh con phi IRAP. Đây là vấn đề do các hình ảnh được lưu trữ trong dòng bit bằng cách sử dụng các đơn vị NAL. Đơn vị NAL là bộ phận lưu trữ mà chứa tập hợp thông số hoặc lát của hình ảnh và phần đầu lát tương ứng. Đơn vị truy cập là đơn vị mà chứa toàn bộ hình ảnh. Như vậy, đơn vị truy cập chứa tất cả các đơn vị NAL liên quan đến hình ảnh. Các đơn vị NAL cũng chứa loại mà chỉ báo loại của hình ảnh mà bao gồm lát. Trong một số hệ thống video, tất cả các đơn vị NAL liên quan đến một hình ảnh đơn lẻ (ví dụ, có trong cùng đơn vị truy cập) được yêu cầu phải có cùng loại. Như vậy, cơ chế lưu trữ đơn vị NAL có thể dừng hoạt động một cách chính xác khi hình ảnh bao gồm cả các hình ảnh con IRAP và các hình ảnh con phi IRAP.

Được bộc lộ ở đây là các cơ chế để điều chỉnh sơ đồ lưu trữ NAL để hỗ trợ các hình ảnh mà bao gồm cả các hình ảnh con IRAP và các hình ảnh con phi IRAP. Từ đó, điều

này cho phép video VR mà bao gồm các tần số hình ảnh con IRAP khác nhau đối với các công quan sát khác nhau. Trong ví dụ thứ nhất, được bộc lộ ở đây là cờ mà chỉ báo liệu hình ảnh có được trộn hay không. Ví dụ, cờ này có thể chỉ báo rằng hình ảnh chứa cả các hình ảnh con IRAP và phi IRAP. Dựa trên cờ này, bộ giải mã có thể xử lý các loại hình ảnh con khác nhau một cách khác nhau khi giải mã để giải mã và hiển thị hình ảnh/các hình ảnh con một cách chính xác. Cờ này có thể được lưu trữ trong tập hợp thông số hình ảnh (PPS) và có thể được gọi là `mixed_nalu_types_in_pic_flag`.

Trong ví dụ thứ hai, được bộc lộ ở đây là cờ mà chỉ báo liệu hình ảnh có được trộn hay không. Ví dụ, cờ này có thể chỉ báo rằng hình ảnh chứa cả các hình ảnh con IRAP và phi IRAP. Hơn nữa, cờ này ràng buộc hình ảnh sao cho hình ảnh trộn chứa chính xác hai loại đơn vị NAL bao gồm một loại IRAP và một loại phi IRAP. Ví dụ, hình ảnh có thể chứa các đơn vị NAL IRAP bao gồm một và chỉ một loại trong số làm mới quá trình giải mã tức thời (IDR) với hình ảnh chủ đạo giải mã được truy cập ngẫu nhiên (IDR_W_RADL), IDR không có các hình ảnh chủ đạo (IDR_N_LP), hoặc loại đơn vị NAL truy cập ngẫu nhiên sạch (CRA) (CRA_NUT). Hơn nữa, hình ảnh có thể chứa các đơn vị NAL phi IRAP bao gồm một và chỉ một loại trong số loại đơn vị NAL hình ảnh kế tiếp (TRAIL_NUT), loại đơn vị NAL hình ảnh chủ đạo giải mã được truy cập ngẫu nhiên (RADL_NUT), hoặc loại đơn vị NAL hình ảnh chủ đạo bỏ qua truy cập ngẫu nhiên (RASL) (RASL_NUT). Dựa trên cờ này, bộ giải mã có thể xử lý các hình ảnh con khác nhau một cách khác nhau khi giải mã để giải mã và hiển thị hình ảnh/các hình ảnh con một cách chính xác. Cờ này có thể được lưu trữ trong PPS và có thể được gọi là `mixed_nalu_types_in_pic_flag`.

Fig.1 là lưu đồ của phương pháp hoạt động 100 làm ví dụ để tạo mã tín hiệu video. Cụ thể, tín hiệu video được mã hóa tại bộ mã hóa. Quy trình mã hóa nén tín hiệu video bằng cách sử dụng các cơ chế khác nhau để làm giảm kích thước tệp video. Kích thước tệp nhỏ hơn cho phép tệp video đã nén được truyền đến người dùng, trong khi làm giảm lượng thông tin chiếm dải thông thừa được liên kết. Sau đó, bộ giải mã giải mã tệp video đã nén để tái tạo tín hiệu video gốc để hiển thị cho người dùng cuối. Quy trình giải mã thường phản chiếu quy trình mã hóa để cho phép bộ giải mã tái tạo tín hiệu video một cách thích hợp.

Ở bước 101, tín hiệu video được đưa vào bộ mã hóa. Ví dụ, tín hiệu video có thể là tệp video không nén được lưu trữ trong bộ nhớ. Theo một ví dụ khác, tệp video có thể thu được bởi thiết bị thu video, như camera video, và được mã hóa để hỗ trợ phát sóng trực tiếp (live streaming) video. Tệp video có thể bao gồm cả thành phần audio lẫn thành phần video. Thành phần video chứa một loạt khung ảnh mà, khi được xem liên tục, mang lại cảm giác thị giác về chuyển động. Các khung chứa các điểm ảnh mà được biểu diễn trên phương diện ánh sáng, ở đây được gọi là các thành phần độ chói (hoặc các mẫu độ chói), và màu, mà được gọi là các thành phần sắc độ (hoặc các mẫu màu). Trong một số ví dụ, các khung cũng có thể chứa các giá trị độ sâu để hỗ trợ xem ba chiều.

Ở bước 103, video được phân chia thành các khối. Việc phân chia bao gồm chia nhỏ các điểm ảnh trong mỗi khung thành các khối hình vuông và/hoặc hình chữ nhật để nén. Ví dụ, trong tạo mã video hiệu suất cao (High Efficiency Video Coding, HEVC) (cũng được biết đến là H.265 và MPEG-H phần 2), trước tiên, khung có thể được chia thành các đơn vị cây tạo mã (CTU), mà là các khối có kích thước được xác định trước (ví dụ, sáu mươi tư điểm ảnh nhân sáu mươi tư điểm ảnh). Các CTU chứa cả các mẫu độ chói lẫn sắc độ. Các cây tạo mã có thể được sử dụng để chia các CTU thành các khối và sau đó chia nhỏ các khối theo cách đệ quy cho đến khi thu được các cấu hình mà hỗ trợ việc mã hóa thêm nữa. Ví dụ, các thành phần độ chói của khung có thể được chia nhỏ thêm cho đến khi các khối riêng chứa các giá trị chiếu sáng đồng nhất tương đối. Hơn nữa, các thành phần sắc độ của khung có thể được chia nhỏ thêm cho đến khi các khối riêng chứa các giá trị màu đồng nhất tương đối. Theo đó, các cơ chế phân chia thay đổi phụ thuộc vào nội dung của các khung video.

Ở bước 105, các cơ chế nén khác nhau được sử dụng để nén các khối ảnh được phân chia ở bước 103. Ví dụ, dự đoán liên ảnh và/hoặc dự đoán trong ảnh có thể được sử dụng. Dự đoán liên ảnh được thiết kế để tận dụng việc các đối tượng trong cảnh chung có xu hướng xuất hiện trong các khung liên tiếp. Theo đó, khối thể hiện đối tượng trong khung tham chiếu không cần được mô tả lặp lại trong các khung liên kế. Cụ thể, đối tượng, như bàn, có thể ở yên một vị trí không đổi trong nhiều khung. Vì vậy, bàn được mô tả một lần và các khung liên kế có thể tham chiếu lại khung tham chiếu. Các cơ chế so khớp mẫu hình có thể được sử dụng để so khớp các đối tượng trong nhiều khung. Hơn nữa, việc di chuyển các đối tượng có thể được biểu diễn qua nhiều khung, ví dụ do di chuyển đối tượng

hoặc di chuyển camera. Theo một ví dụ cụ thể, video có thể thể hiện ô tô mà di chuyển qua màn hình trên nhiều khung. Các vector chuyển động có thể được sử dụng để mô tả sự di chuyển đó. Vector chuyển động là vector hai chiều để tạo ra độ lệch từ các tọa độ của đối tượng trong khung đến các tọa độ của đối tượng trong khung tham chiếu. Như vậy, dự đoán liên ảnh có thể mã hóa khối ảnh trong khung hiện hành dưới dạng tập hợp các vector chuyển động chỉ báo độ lệch từ khối tương ứng trong khung tham chiếu.

Dự đoán trong ảnh mã hóa các khối trong khung chung. Dự đoán trong ảnh tận dụng việc các thành phần độ chói và sắc độ có xu hướng gộp lại trong khung. Ví dụ, một vết màu lục ở một phần của cây có xu hướng được định vị liên kề các vết màu lục tương tự. Dự đoán trong ảnh sử dụng nhiều chế độ dự đoán có hướng (ví dụ, ba mươi ba trong HEVC), chế độ phẳng, và chế độ dòng điện một chiều (DC). Các chế độ có hướng chỉ báo rằng khối hiện hành là tương tự/giống các mẫu của khối lân cận theo hướng tương ứng. Chế độ phẳng chỉ báo rằng một loạt khối dọc theo hàng/cột (ví dụ, mặt phẳng) có thể được nội suy dựa trên các khối lân cận tại các cạnh của hàng. Trên thực tế, chế độ phẳng chỉ báo sự chuyển tiếp trơn nhẵn của ánh sáng/màu qua hàng/cột bằng cách sử dụng độ nghiêng không đối xứng khi thay đổi các giá trị. Chế độ DC được sử dụng để để làm trơn nhẵn biên và chỉ báo rằng khối tương tự/giống với giá trị trung bình được liên kết với các mẫu của tất cả các khối lân cận được liên kết với các hướng góc của các chế độ dự đoán có hướng. Theo đó, các khối dự đoán trong ảnh có thể đại diện cho các khối ảnh như là các giá trị chế độ dự đoán liên quan khác nhau thay vì các giá trị thực tế. Hơn nữa, các khối dự đoán liên ảnh có thể đại diện cho các khối ảnh như là các giá trị vector chuyển động thay vì các giá trị thực tế. Trong trường hợp khác, các khối dự đoán có thể không đại diện chính xác các khối ảnh trong một số trường hợp. Các sự chênh lệch bất kỳ được lưu trữ trong các khối dư. Các sự biến đổi có thể được áp dụng vào các khối dư để nén tệp thêm nữa.

Ở bước 107, các kỹ thuật lọc khác nhau có thể được áp dụng. Trong HEVC, các bộ lọc được áp dụng theo sơ đồ lọc nội vòng. Việc dự đoán dựa trên khối được thảo luận trên đây có thể dẫn đến tạo các ảnh có khối ở bộ giải mã. Hơn nữa, sơ đồ dự đoán dựa trên khối có thể mã hóa khối và sau đó tái tạo khối được mã hóa để sử dụng sau đó dưới dạng khối tham chiếu. Sơ đồ lọc nội vòng áp dụng lặp lại các bộ lọc triệt nhiễu âm, các bộ lọc giải khối, các bộ lọc vòng thích ứng, và các bộ lọc độ lệch tương thích mẫu (sample

adaptive offset, SAO) vào các khối/các khung. Các bộ lọc này làm giảm bớt các vật tạo tác tạo khối như vậy sao cho tệp được mã hóa có thể được tái tạo chính xác. Hơn nữa, các bộ lọc này làm giảm bớt các vật tạo tác trong các khối tham chiếu được tái tạo sao cho các vật tạo tác ít có khả năng tạo các vật tạo tác bổ sung trong các khối tiếp theo mà được mã hóa dựa trên các khối tham chiếu được tái tạo.

Một khi tín hiệu video đã được phân chia, được nén, và được lọc, dữ liệu tạo thành được mã hóa trong dòng bit ở bước 109. Dòng bit bao gồm dữ liệu được thảo luận trên đây cũng như dữ liệu báo hiệu bất kỳ được mong muốn hỗ trợ việc tái tạo tín hiệu video chính xác tại bộ giải mã. Ví dụ, dữ liệu này có thể bao gồm dữ liệu phân chia, dữ liệu dự đoán, các khối dư, và các cờ khác nhau cung cấp các lệnh tạo mã cho bộ giải mã. Dòng bit có thể được lưu trữ trong bộ nhớ để truyền đến bộ giải mã khi có yêu cầu. Dòng bit cũng có thể là phát rộng và/hoặc phát đa phương đến các bộ giải mã. Việc tạo dòng bit là quy trình lặp lại. Theo đó, các bước 101, 103, 105, 107, và 109 có thể xảy ra liên tiếp và/hoặc đồng thời trên nhiều khung và khối. Thứ tự được thể hiện trên Fig.1 được trình bày cho rõ ràng và dễ thảo luận, và không được dự định để giới hạn quy trình tạo mã video ở thứ tự cụ thể.

Bộ giải mã nhận dòng bit và bắt đầu quy trình giải mã ở bước 111. Cụ thể, bộ giải mã sử dụng sơ đồ giải mã entropi để chuyển đổi dòng bit thành cú pháp và dữ liệu video tương ứng. Bộ giải mã sử dụng dữ liệu cú pháp từ dòng bit để xác định các phân chia đối với các khung ở bước 111. Việc phân chia nên khớp với các kết quả của việc phân chia khối ở bước 103. Lúc này, việc mã hóa/giải mã entropi như được sử dụng ở bước 111 được mô tả. Bộ mã hóa thực hiện nhiều sự lựa chọn trong suốt quá trình nén, như chọn các sơ đồ phân chia khối từ một số sự lựa chọn khả thi dựa trên việc định vị không gian các giá trị trong (các) ảnh được đưa vào. Việc báo hiệu các sự lựa chọn chính xác có thể sử dụng một số lượng học (bin) lớn. Như được sử dụng ở đây, học (bin) là giá trị nhị phân mà được xử lý như là biến số (ví dụ, giá trị bit mà có thể thay đổi phụ thuộc vào ngữ cảnh). Tạo mã entropi cho phép bộ mã hóa loại bỏ các tùy chọn bất kỳ mà rõ ràng là không khả thi đối với trường hợp cụ thể, để lại tập hợp các tùy chọn chấp nhận được. Sau đó, mỗi tùy chọn chấp nhận được được ấn định một từ mã. Độ dài của các từ mã dựa trên số lượng tùy chọn chấp nhận được (ví dụ, một học (bin) cho hai tùy chọn, hai học cho từ ba đến bốn tùy chọn, v.v.) Bộ mã hóa sau đó mã hóa từ mã cho tùy chọn đã chọn. Sơ đồ này làm

giảm kích thước của các từ mã do các từ mã lớn như mong muốn để chỉ báo duy nhất một sự lựa chọn từ tập con nhỏ của các tùy chọn chấp nhận được ngược với việc chỉ báo duy nhất sự lựa chọn từ tập hợp lớn tiềm tàng của tất cả các tùy chọn khả thi. Sau đó, bộ giải mã giải mã sự lựa chọn bằng cách xác định tập hợp các tùy chọn chấp nhận được theo cách tương tự với bộ mã hóa. Bằng cách xác định tập hợp các tùy chọn chấp nhận được, bộ giải mã có thể đọc từ mã và xác định sự lựa chọn được thực hiện bởi bộ mã hóa.

Ở bước 113, bộ giải mã thực hiện việc giải mã khối. Cụ thể, bộ giải mã sử dụng biến đổi ngược để tạo các khối dư. Sau đó, bộ giải mã sử dụng các khối dư và các khối dự đoán tương ứng để tái tạo các khối ảnh theo sự phân chia. Các khối dự đoán có thể bao gồm cả các khối dự đoán trong ảnh lẫn các khối dự đoán liên ảnh như được tạo ra tại bộ mã hóa ở bước 105. Sau đó, các khối ảnh đã tái tạo được định vị vào các khung của tín hiệu video đã tái tạo theo dữ liệu phân chia được xác định ở bước 111. Cú pháp đối với bước 113 cũng có thể được báo hiệu trong dòng bit thông qua tạo mã entropi như được thảo luận trên đây.

Ở bước 115, việc lọc được thực hiện trên các khung của tín hiệu video đã tái tạo theo cách tương tự như bước 107 ở bộ mã hóa. Ví dụ, các bộ lọc triệt nhiễu âm, các bộ lọc giải khối, các bộ lọc vòng thích ứng, và các bộ lọc SAO có thể được áp dụng vào các khung để loại bỏ các vật tạo tác tạo khối. Một khi các khung được lọc, tín hiệu video có thể được kết xuất đến bộ hiển thị ở bước 117 để người dùng cuối xem.

Fig.2 là hình vẽ giản lược của hệ thống tạo mã và giải mã (codec) 200 làm ví dụ để tạo mã video. Cụ thể, hệ thống codec 200 cung cấp chức năng để hỗ trợ việc thực hiện phương pháp hoạt động 100. Hệ thống codec 200 được tổng quát hóa để thể hiện các thành phần được sử dụng trong cả bộ mã hóa lẫn bộ giải mã. Hệ thống codec 200 nhận và phân chia tín hiệu video như được thảo luận liên quan đến các bước 101 và 103 trong phương pháp hoạt động 100, mà dẫn đến tín hiệu video đã phân chia 201. Sau đó, hệ thống codec 200 nén tín hiệu video đã phân chia 201 thành dòng bit được tạo mã khi đóng vai trò là bộ mã hóa như được thảo luận liên quan đến các bước 105, 107, và 109 trong phương pháp 100. Khi đóng vai trò là bộ giải mã, hệ thống codec 200 tạo ra tín hiệu video đầu ra từ dòng bit như được thảo luận liên quan đến các bước 111, 113, 115, và 117 trong phương pháp hoạt động 100. Hệ thống codec 200 bao gồm thành phần điều khiển bộ tạo mã chung

211, thành phần định tỷ lệ biến đổi và lượng tử hóa 213, thành phần ước lượng trong hình ảnh 215, thành phần dự đoán trong hình ảnh 217, thành phần bù chuyển động 219, thành phần ước lượng chuyển động 221, thành phần định tỷ lệ và biến đổi ngược 229, thành phần phân tích điều khiển bộ lọc 227, thành phần các bộ lọc nội vòng 225, thành phần bộ đệm hình ảnh được giải mã 223, và thành phần tạo mã số học nhị phân thích ứng ngữ cảnh (CABAC) và định dạng phần đầu 231. Các thành phần này được ghép nối như được thể hiện. Trên Fig.2, các đường liền nét chỉ báo sự di chuyển của dữ liệu cần được mã hóa/được giải mã trong khi các đường đứt nét chỉ báo sự di chuyển của dữ liệu điều khiển mà điều khiển hoạt động của các thành phần khác. Các thành phần của hệ thống codec 200 có thể đều có trong bộ mã hóa. Bộ giải mã có thể bao gồm tập con của các thành phần của hệ thống codec 200. Ví dụ, bộ giải mã có thể bao gồm thành phần dự đoán trong hình ảnh 217, thành phần bù chuyển động 219, thành phần định tỷ lệ và biến đổi ngược 229, thành phần các bộ lọc nội vòng 225, và thành phần bộ đệm hình ảnh được giải mã 223. Lúc này, các thành phần này được mô tả.

Tín hiệu video đã phân chia 201 là chuỗi video thu được mà đã được phân chia thành các khối của các điểm ảnh bởi cây tạo mã. Cây tạo mã sử dụng các chế độ chia tách khác nhau để chia nhỏ khối của các điểm ảnh thành các khối nhỏ hơn của các điểm ảnh. Sau đó, các khối này có thể còn được chia nhỏ thêm thành các khối nhỏ hơn. Các khối này có thể được gọi là các nút trên cây tạo mã. Các nút cấp trên (parent node) lớn hơn được chia tách thành các nút con nhỏ hơn. Số lần một nút được chia nhỏ thêm được gọi là độ sâu của nút/cây tạo mã. Các khối được chia có thể có trong các đơn vị tạo mã (CU) trong một số trường hợp. Ví dụ, CU có thể là một phần con của CTU mà chứa khối độ chói, (các) khối sắc độ chênh lệch đỏ (Cr), và (các) khối sắc độ chênh lệch lam (Cb) cùng với các chỉ dẫn cú pháp tương ứng đối với CU. Các chế độ chia tách có thể bao gồm cây nhị phân (BT), cây tam phân (TT), và cây tứ phân (QT) được sử dụng để phân chia nút lần lượt thành hai, ba, hoặc bốn nút con có các hình dạng thay đổi tùy thuộc vào các chế độ chia tách được sử dụng. Tín hiệu video đã phân chia 201 được chuyển tiếp đến thành phần điều khiển bộ tạo mã chung 211, thành phần định tỷ lệ biến đổi và lượng tử hóa 213, thành phần ước lượng trong hình ảnh 215, thành phần phân tích điều khiển bộ lọc 227, và thành phần ước lượng chuyển động 221 để nén.

Thành phần điều khiển bộ tạo mã chung 211 được tạo cấu hình để quyết định liên quan đến việc tạo mã các ảnh của chuỗi video thành dòng bit theo các ràng buộc ứng dụng. Ví dụ, thành phần điều khiển bộ tạo mã chung 211 quản lý việc tối ưu hóa tốc độ bit/kích thước dòng bit so với chất lượng tái tạo. Các quyết định này có thể được thực hiện dựa trên tính khả dụng của không gian lưu trữ/dải thông và các yêu cầu độ phân giải ảnh. Thành phần điều khiển bộ tạo mã chung 211 cũng quản lý việc sử dụng bộ đệm xét về tốc độ truyền để giảm bớt các vấn đề chạy ngầm và tràn của bộ đệm. Để quản lý các vấn đề này, thành phần điều khiển bộ tạo mã chung 211 quản lý việc phân chia, dự đoán, và lọc bởi các thành phần khác. Ví dụ, thành phần điều khiển bộ tạo mã chung 211 có thể làm tăng độ phức tạp nén theo cách động để làm tăng độ phân giải và làm tăng lượng sử dụng dải thông hoặc làm giảm độ phức tạp nén để làm giảm độ phân giải và lượng sử dụng dải thông. Vì vậy, thành phần điều khiển bộ tạo mã chung 211 điều khiển các thành phần khác của hệ thống codec 200 để cân bằng chất lượng tái tạo tín hiệu video với các vấn đề về tốc độ bit. Thành phần điều khiển bộ tạo mã chung 211 tạo dữ liệu điều khiển, mà điều khiển hoạt động của các thành phần khác. Dữ liệu điều khiển cũng được chuyển tiếp đến thành phần CABAC và định dạng phần đầu 231 để được mã hóa trong dòng bit để báo hiệu các thông số để giải mã tại bộ giải mã.

Tín hiệu video đã phân chia 201 cũng được gửi đến thành phần ước lượng chuyển động 221 và thành phần bù chuyển động 219 để dự đoán liên ảnh. Khung hoặc lát của tín hiệu video đã phân chia 201 có thể được chia thành nhiều khối video. Thành phần ước lượng chuyển động 221 và thành phần bù chuyển động 219 thực hiện việc tạo mã dự đoán liên ảnh đối với khối video nhận được so với một hoặc nhiều khối trong một hoặc nhiều khung tham chiếu để cung cấp dự đoán thời gian. Hệ thống codec 200 có thể thực hiện nhiều quá trình tạo mã, ví dụ, để chọn chế độ tạo mã thích hợp cho mỗi khối của dữ liệu video.

Thành phần ước lượng chuyển động 221 và thành phần bù chuyển động 219 có thể được tích hợp ở mức cao, nhưng được minh họa riêng rẽ nhằm các mục đích liên quan đến khái niệm. Ước lượng chuyển động, được thực hiện bởi thành phần ước lượng chuyển động 221, là quy trình tạo các vector chuyển động, mà ước lượng chuyển động đối với các khối video. Vector chuyển động, ví dụ, có thể chỉ báo sự dịch chuyển của đối tượng được tạo mã so với khối dự đoán. Khối dự đoán là khối mà được thấy là so khớp gần với khối

cần được tạo mã, trên phương diện chênh lệch điểm ảnh. Khối dự đoán cũng có thể được gọi là khối tham chiếu. Sự chênh lệch điểm ảnh như vậy có thể được xác định bởi tổng sai phân tuyệt đối (sum of absolute difference, SAD), tổng sai phân bình phương (sum of square difference, SSD), hoặc các chuẩn đo chênh lệch khác nhau. HEVC sử dụng một số đối tượng được tạo mã bao gồm CTU, các khối cây tạo mã (coding tree block, CTB), và các CU. Ví dụ, CTU có thể được chia thành các CTB, mà sau đó có thể được chia thành các CB để có trong các CU. CU có thể được mã hóa dưới dạng đơn vị dự đoán (PU) chứa dữ liệu dự đoán và/hoặc đơn vị biến đổi (TU) chứa dữ liệu dư đã biến đổi đối với CU. Thành phần ước lượng chuyển động 221 tạo ra các vector chuyển động, các PU, và các TU bằng cách sử dụng phân tích tỷ lệ méo (rate-distortion) như là một phần của quy trình tối ưu hóa tỷ lệ méo. Ví dụ, thành phần ước lượng chuyển động 221 có thể xác định nhiều khối tham chiếu, nhiều vector chuyển động, v.v. đối với khối/khung hiện hành, và có thể chọn các khối tham chiếu, các vector chuyển động, v.v. có các đặc tính tỷ lệ méo tốt nhất. Các đặc tính tỷ lệ méo tốt nhất làm cân bằng cả chất lượng của video tái tạo (ví dụ, lượng dữ liệu bị tổn thất do nén) với hiệu suất tạo mã (ví dụ, kích thước của việc mã hóa cuối cùng).

Trong một số ví dụ, hệ thống codec 200 có thể tính toán các giá trị đối với các vị trí điểm ảnh nguyên con của các hình ảnh tham chiếu được lưu trữ trong thành phần bộ đệm hình ảnh được giải mã 223. Ví dụ, codec video hệ thống 200 có thể nội suy các giá trị của các vị trí điểm ảnh một phần tư, các vị trí điểm ảnh một phần tám, hoặc các vị trí điểm ảnh phân số khác của hình ảnh tham chiếu. Do đó, thành phần ước lượng chuyển động 221 có thể thực hiện việc tìm kiếm chuyển động so với các vị trí điểm ảnh đầy đủ và các vị trí điểm ảnh phân số và kết xuất vector chuyển động với độ chính xác điểm ảnh thập phân. Thành phần ước lượng chuyển động 221 tính toán vector chuyển động đối với PU của khối video trong lát được tạo mã liên ảnh bằng cách so sánh vị trí của PU với vị trí của khối dự đoán của hình ảnh tham chiếu. Thành phần ước lượng chuyển động 221 kết xuất vector chuyển động tính toán được dưới dạng dữ liệu chuyển động đến thành phần CABAC và định dạng phân đầu 231 để mã hóa và chuyển động đến thành phần bù chuyển động 219.

Bù chuyển động, được thực hiện bởi thành phần bù chuyển động 219, có thể bao gồm việc tìm hoặc tạo khối dự đoán dựa trên vector chuyển động được xác định bởi thành

phần ước lượng chuyển động 221. Mặt khác, thành phần ước lượng chuyển động 221 và thành phần bù chuyển động 219 có thể được tích hợp về mặt chức năng trong một số ví dụ. Khi nhận vector chuyển động đối với PU của khối video hiện hành, thành phần bù chuyển động 219 có thể định vị khối dự đoán mà vector chuyển động trỏ đến. Sau đó, khối video dư được tạo ra bằng cách trừ các giá trị điểm ảnh của khối dự đoán từ các giá trị điểm ảnh của khối video hiện hành đang được tạo mã, tạo thành các giá trị chênh lệch điểm ảnh. Nói chung, thành phần ước lượng chuyển động 221 thực hiện ước lượng chuyển động so với các thành phần độ chói, và thành phần bù chuyển động 219 sử dụng các vector chuyển động được tính toán dựa trên các thành phần độ chói đối với cả các thành phần sắc độ lẫn các thành phần độ chói. Khối dự đoán và khối dư được chuyển tiếp đến thành phần định tỷ lệ biến đổi và lượng tử hóa 213.

Tín hiệu video đã phân chia 201 cũng được gửi đến thành phần ước lượng trong hình ảnh 215 và thành phần dự đoán trong hình ảnh 217. Như với thành phần ước lượng chuyển động 221 và thành phần bù chuyển động 219, thành phần ước lượng trong hình ảnh 215 và thành phần dự đoán trong hình ảnh 217 có thể được tích hợp ở mức cao, nhưng được minh họa riêng rẽ nhằm các mục đích liên quan đến khái niệm. Thành phần ước lượng trong hình ảnh 215 và thành phần dự đoán trong hình ảnh 217 dự đoán nội bộ khối hiện hành so với các khối trong khung hiện hành, như là một sự lựa chọn đối với việc dự đoán liên ảnh được thực hiện bởi thành phần ước lượng chuyển động 221 và thành phần bù chuyển động 219 giữa các khung, như được mô tả trên đây. Cụ thể, thành phần ước lượng trong hình ảnh 215 xác định chế độ dự đoán trong ảnh để sử dụng để mã hóa khối hiện hành. Trong một số ví dụ, thành phần ước lượng trong hình ảnh 215 chọn chế độ dự đoán trong ảnh thích hợp để mã hóa khối hiện hành từ nhiều chế độ dự đoán trong ảnh đã kiểm tra. Sau đó, các chế độ dự đoán trong ảnh đã chọn được chuyển tiếp đến thành phần CABAC và định dạng phần đầu 231 để mã hóa.

Ví dụ, thành phần ước lượng trong hình ảnh 215 tính toán các giá trị tỷ lệ méo nhờ sử dụng phân tích tỷ lệ méo đối với các chế độ dự đoán trong ảnh đã kiểm tra khác nhau, và chọn chế độ dự đoán trong ảnh có các đặc tính tỷ lệ méo tốt nhất trong số các chế độ đã kiểm tra. Việc phân tích tỷ lệ méo xác định chung lượng méo (hoặc sai số) giữa khối được mã hóa và khối chưa được mã hóa gốc mà đã được mã hóa để tạo ra khối được mã hóa, cũng như tốc độ bit (ví dụ, số lượng bit) được sử dụng để tạo khối được mã hóa.

Thành phần ước lượng trong hình ảnh 215 tính toán các tỷ lệ từ các phần méo và các tỷ lệ đối với các khối được mã hóa khác nhau để xác định chế độ dự đoán trong ảnh nào thể hiện giá trị tỷ lệ méo tốt nhất đối với khối. Ngoài ra, thành phần ước lượng trong hình ảnh 215 có thể được tạo cấu hình để tạo mã các khối độ sâu của bản đồ độ sâu map nhờ sử dụng chế độ tạo mô hình độ sâu (depth modeling mode, DMM) dựa trên tối ưu hóa tỷ lệ méo (RDO).

Thành phần dự đoán trong hình ảnh 217 có thể tạo khối dư từ khối dự đoán dựa trên các chế độ dự đoán trong ảnh đã chọn được xác định bởi thành phần ước lượng trong hình ảnh 215 khi được thực hiện trên bộ mã hóa hoặc đọc khối dư từ dòng bit khi được thực hiện trên bộ giải mã. Khối dư bao gồm sự chênh lệch về các giá trị giữa khối dự đoán và khối gốc, được thể hiện dưới dạng ma trận. Sau đó, khối dư được chuyển tiếp đến thành phần định tỷ lệ biến đổi và lượng tử hóa 213. Thành phần ước lượng trong hình ảnh 215 và thành phần dự đoán trong hình ảnh 217 có thể hoạt động trên cả các thành phần độ chói lẫn sắc độ.

Thành phần định tỷ lệ biến đổi và lượng tử hóa 213 được tạo cấu hình để nén khối dư thêm nữa. Thành phần định tỷ lệ biến đổi và lượng tử hóa 213 áp dụng biến đổi, như biến đổi cosin rời rạc (DCT), biến đổi sin rời rạc (DST), hoặc biến đổi tương tự về mặt khái niệm, vào khối dư, tạo ra khối video bao gồm các giá trị hệ số biến đổi dư. Biến đổi sóng nhỏ (Wavelet), biến đổi nguyên, biến đổi dải con hoặc các loại biến đổi khác cũng có thể được sử dụng. Việc biến đổi có thể chuyển đổi thông tin còn lại từ miền giá trị điểm ảnh sang miền biến đổi, như miền tần số. Thành phần định tỷ lệ biến đổi và lượng tử hóa 213 cũng được tạo cấu hình để định tỷ lệ thông tin còn lại được biến đổi, ví dụ dựa trên tần số. Việc định tỷ lệ này bao gồm áp dụng bội suất vào thông tin còn lại sao cho thông tin tần số khác nhau được lượng tử hóa ở các mức phân mảnh khác nhau, mà có thể tác động đến chất lượng trực quan cuối cùng của video được tái tạo. Thành phần định tỷ lệ biến đổi và lượng tử hóa 213 cũng được tạo cấu hình để lượng tử hóa các hệ số biến đổi để làm giảm tốc độ bit thêm nữa. Quy trình lượng tử hóa có thể làm giảm độ sâu bit được liên kết với một số hoặc tất cả hệ số. Mức độ lượng tử hóa có thể được thay đổi bằng cách điều chỉnh thông số lượng tử hóa. Trong một số ví dụ, thành phần định tỷ lệ biến đổi và lượng tử hóa 213 sau đó có thể thực hiện việc quét ma trận bao gồm các hệ số biến đổi

được lượng tử hóa. Các hệ số biến đổi được lượng tử hóa được chuyển tiếp đến thành phần CABAC và định dạng phần đầu 231 cần được mã hóa trong dòng bit.

Thành phần định tỷ lệ và biến đổi ngược 229 áp dụng công đoạn ngược của thành phần định tỷ lệ biến đổi và lượng tử hóa 213 để hỗ trợ ước lượng chuyển động. Thành phần định tỷ lệ và biến đổi ngược 229 áp dụng định tỷ lệ ngược, biến đổi ngược, và/hoặc lượng tử hóa ngược để tái tạo khối dư trong miền điểm ảnh, ví dụ, để sử dụng sau này làm khối tham chiếu mà có thể trở thành khối dự đoán cho một khối hiện hành khác. Thành phần ước lượng chuyển động 221 và/hoặc thành phần bù chuyển động 219 có thể tính toán khối tham chiếu bằng cách thêm khối dư trở lại khối dự đoán tương ứng để sử dụng khi ước lượng chuyển động khối/khung sau này. Các bộ lọc được áp dụng vào các khối tham chiếu được tái tạo để giảm bớt các vật tạo tác được tạo ra trong suốt quá trình định tỷ lệ, lượng tử hóa, và biến đổi. Nếu không, các vật tạo tác này có thể gây ra dự đoán (và tạo các vật tạo tác bổ sung) không chính xác khi các khối tiếp theo được dự đoán.

Thành phần phân tích điều khiển bộ lọc 227 và thành phần các bộ lọc nội vòng 225 áp dụng các bộ lọc vào các khối dư và/hoặc vào các khối ảnh đã tái tạo. Ví dụ, khối dư đã biến đổi từ thành phần định tỷ lệ và biến đổi ngược 229 có thể được kết hợp với khối dự đoán tương ứng từ thành phần dự đoán trong hình ảnh 217 và/hoặc thành phần bù chuyển động 219 để tái tạo khối ảnh gốc. Sau đó, các bộ lọc có thể được áp dụng vào khối ảnh đã tái tạo. Trong một số ví dụ, các bộ lọc có thể được áp dụng vào các khối dư để thay thế. Như với các thành phần khác trên Fig.2, thành phần phân tích điều khiển bộ lọc 227 và thành phần các bộ lọc nội vòng 225 được thích hợp ở mức cao và có thể được thực hiện cùng nhau, nhưng được thể hiện riêng rẽ nhằm các mục đích liên quan đến khái niệm. Các bộ lọc được áp dụng vào các khối tham chiếu được tái tạo được áp dụng vào các vùng không gian cụ thể và bao gồm nhiều thông số để điều chỉnh cách mà các bộ lọc này được áp dụng. Thành phần phân tích điều khiển bộ lọc 227 phân tích các khối tham chiếu được tái tạo để xác định các bộ lọc này nên được áp dụng ở đâu và thiết đặt các thông số tương ứng. Dữ liệu này được chuyển tiếp đến thành phần CABAC và định dạng phần đầu 231 như là dữ liệu điều khiển bộ lọc để mã hóa. Thành phần các bộ lọc nội vòng 225 áp dụng các bộ lọc này dựa trên dữ liệu điều khiển bộ lọc. Các bộ lọc có thể bao gồm bộ lọc giải khối, bộ lọc triệt nhiễu âm, bộ lọc SAO, và bộ lọc vòng thích ứng. Các bộ lọc này có thể

được áp dụng trong miền không gian/điểm ảnh (ví dụ, trên khối điểm ảnh đã tái tạo) hoặc trong miền tần số, tùy thuộc vào ví dụ.

Khi hoạt động như là bộ mã hóa, khối ảnh đã tái tạo, khối dư, và/hoặc khối dự đoán được lọc được lưu trữ trong thành phần bộ đệm hình ảnh được giải mã 223 để sử dụng về sau khi ước lượng chuyển động như được thảo luận trên đây. Khi hoạt động như là bộ giải mã, thành phần bộ đệm hình ảnh được giải mã 223 lưu trữ và chuyển tiếp các khối đã tái tạo và đã lọc đến bộ hiển thị như là một phần của tín hiệu video đầu ra. Thành phần bộ đệm hình ảnh được giải mã 223 có thể là thiết bị nhớ bất kỳ có khả năng lưu trữ các khối dự đoán, các khối dư, và/hoặc các khối ảnh đã tái tạo.

Thành phần CABAC và định dạng phần đầu 231 nhận dữ liệu từ các thành phần khác nhau của hệ thống codec 200 và mã hóa dữ liệu này thành dòng bit được tạo mã để truyền đến bộ giải mã. Cụ thể, thành phần CABAC và định dạng phần đầu 231 tạo ra các phần đầu khác nhau để mã hóa dữ liệu điều khiển, như dữ liệu điều khiển chung và dữ liệu điều khiển bộ lọc. Hơn nữa, dữ liệu dự đoán, bao gồm dữ liệu chuyển động và dự đoán trong ảnh, cũng như dữ liệu dư ở dạng dữ liệu hệ số biến đổi được lượng tử hóa đều được mã hóa trong dòng bit. Dòng bit cuối cùng bao gồm tất cả thông tin được mong muốn bởi bộ giải mã để tái tạo tín hiệu video đã phân chia gốc 201. Thông tin này cũng có thể bao gồm các bảng chỉ số chế độ dự đoán trong ảnh (cũng được gọi là các bảng ánh xạ từ mã), các định nghĩa về các ngữ cảnh mã hóa đối với các khối, các chỉ báo khác nhau của các chế độ dự đoán trong ảnh có khả năng nhất, chỉ báo về thông tin phân chia, v.v.. Dữ liệu này có thể được mã hóa bằng cách sử dụng việc tạo mã entropi. Ví dụ, thông tin có thể được mã hóa bằng cách sử dụng tạo mã độ dài biến số thích ứng ngữ cảnh (context adaptive variable length coding, CAVLC), CABAC, tạo mã số học nhị phân thích ứng ngữ cảnh dựa trên cú pháp (syntax-based context-adaptive binary arithmetic coding, SBAC), tạo mã entropi phân chia khoảng xác suất (probability interval partitioning entropy, PIPE), hoặc kỹ thuật tạo mã entropi khác. Tiếp theo tạo mã entropi, dòng bit đã tạo mã có thể được truyền đến thiết bị khác (ví dụ, bộ giải mã video) hoặc thu được để truyền hoặc truy hồi sau này.

Fig.3 là sơ đồ khối minh họa bộ mã hóa video 300 làm ví dụ. Bộ mã hóa video 300 có thể được sử dụng để thực hiện các chức năng mã hóa của hệ thống codec 200 và/hoặc

thực hiện các bước 101, 103, 105, 107, và/hoặc 109 của phương pháp hoạt động 100. Bộ mã hóa 300 phân chia tín hiệu video đầu vào, tạo thành tín hiệu video đã phân chia 301, mà gần như tương tự như tín hiệu video đã phân chia 201. Sau đó, tín hiệu video đã phân chia 301 được nén và được mã hóa thành dòng bit bởi các thành phần của bộ mã hóa 300.

Cụ thể, tín hiệu video đã phân chia 301 được chuyển tiếp đến thành phần dự đoán trong hình ảnh 317 để dự đoán trong ảnh. Thành phần dự đoán trong hình ảnh 317 có thể gần như tương tự như thành phần ước lượng trong hình ảnh 215 và thành phần dự đoán trong hình ảnh 217. Tín hiệu video đã phân chia 301 cũng được chuyển tiếp đến thành phần bù chuyển động 321 để dự đoán liên ảnh dựa trên các khối tham chiếu trong thành phần bộ đệm hình ảnh được giải mã 323. Thành phần bù chuyển động 321 có thể gần như tương tự như thành phần ước lượng chuyển động 221 và thành phần bù chuyển động 219. Các khối dự đoán và các khối dư từ thành phần dự đoán trong hình ảnh 317 và thành phần bù chuyển động 321 được chuyển tiếp đến thành phần biến đổi và lượng tử hóa 313 để biến đổi và lượng tử hóa các khối dư. Thành phần biến đổi và lượng tử hóa 313 có thể gần như tương tự như thành phần định tỷ lệ biến đổi và lượng tử hóa 213. Các khối dư đã biến đổi và đã lượng tử hóa và các khối dự đoán tương ứng (cùng với dữ liệu điều khiển được liên kết) được chuyển tiếp đến thành phần tạo mã entropi 331 để tạo mã thành dòng bit. Thành phần tạo mã entropi 331 có thể gần như tương tự như thành phần CABAC và định dạng phân đầu 231.

Các khối dư đã biến đổi và đã lượng tử hóa và/hoặc các khối dự đoán tương ứng cũng được chuyển tiếp từ thành phần biến đổi và lượng tử hóa 313 đến thành phần biến đổi ngược và lượng tử hóa ngược 329 để tái tạo thành các khối tham chiếu để sử dụng bởi thành phần bù chuyển động 321. Thành phần biến đổi ngược và lượng tử hóa ngược 329 có thể gần như tương tự như thành phần định tỷ lệ và biến đổi ngược 229. Các bộ lọc nội vòng trong thành phần các bộ lọc nội vòng 325 cũng được áp dụng vào các khối dư và/hoặc các khối tham chiếu được tái tạo, tùy thuộc vào ví dụ. Thành phần các bộ lọc nội vòng 325 có thể gần như tương tự như thành phần phân tích điều khiển bộ lọc 227 và thành phần các bộ lọc nội vòng 225. Thành phần các bộ lọc nội vòng 325 có thể bao gồm nhiều bộ lọc như được thảo luận liên quan đến thành phần các bộ lọc nội vòng 225. Các khối được lọc sau đó được lưu trữ trong thành phần bộ đệm hình ảnh được giải mã 323 để sử dụng làm các khối tham chiếu bởi thành phần bù chuyển động 321. Thành phần bộ

đệm hình ảnh được giải mã 323 có thể gần như tương tự như thành phần bộ đệm hình ảnh được giải mã 223.

Fig.4 là sơ đồ khối minh họa bộ giải mã video 400 làm ví dụ. Bộ giải mã video 400 có thể được sử dụng để thực hiện các chức năng giải mã của hệ thống codec 200 và/hoặc thực hiện các bước 111, 113, 115, và/hoặc 117 của phương pháp hoạt động 100. Bộ giải mã 400 nhận dòng bit, ví dụ từ bộ mã hóa 300, và tạo ra tín hiệu video đầu ra đã tái tạo dựa trên dòng bit để hiển thị cho người dùng cuối.

Dòng bit được nhận bởi thành phần giải mã entropi 433. Thành phần giải mã entropi 433 được tạo cấu hình để thực hiện sơ đồ giải mã entropi, như CAVLC, CABAC, SBAC, tạo mã PIPE, hoặc các kỹ thuật tạo mã entropi khác. Ví dụ, thành phần giải mã entropi 433 có thể sử dụng thông tin phần đầu để cung cấp ngữ cảnh để dịch dữ liệu bổ sung được mã hóa dưới dạng các từ mã trong dòng bit. Thông tin đã giải mã bao gồm thông tin mong muốn bất kỳ để giải mã tín hiệu video, như dữ liệu điều khiển chung, dữ liệu điều khiển bộ lọc, thông tin phân chia, dữ liệu chuyển động, dữ liệu dự đoán, và các hệ số biến đổi được lượng tử hóa từ các khối dư. Các hệ số biến đổi được lượng tử hóa được chuyển tiếp đến thành phần biến đổi ngược và lượng tử hóa ngược 429 để tái tạo thành các khối dư. Thành phần biến đổi ngược và lượng tử hóa ngược 429 có thể tương tự như thành phần biến đổi ngược và lượng tử hóa ngược 329.

Các khối dư được tái tạo và/hoặc các khối dự đoán được chuyển tiếp đến thành phần dự đoán trong hình ảnh 417 để tái tạo thành các khối ảnh dựa trên các hoạt động dự đoán trong ảnh. Thành phần dự đoán trong hình ảnh 417 có thể tương tự như thành phần ước lượng trong hình ảnh 215 và thành phần dự đoán trong hình ảnh 217. Cụ thể, thành phần dự đoán trong hình ảnh 417 sử dụng các chế độ dự đoán để định vị khối tham chiếu trong khung và áp dụng khối dư vào kết quả để tái tạo các khối ảnh được dự đoán trong ảnh. Các khối ảnh được dự đoán trong ảnh đã tái tạo và/hoặc các khối dư và dữ liệu dự đoán liên ảnh tương ứng được chuyển tiếp đến thành phần bộ đệm hình ảnh được giải mã 423 qua thành phần các bộ lọc nội vòng 425, mà lần lượt có thể gần như tương tự như thành phần bộ đệm hình ảnh được giải mã 223 và thành phần các bộ lọc nội vòng 225. Thành phần các bộ lọc nội vòng 425 bộ lọc các khối ảnh đã tái tạo, các khối dư và/hoặc các khối dự đoán, và thông tin này được lưu trữ trong thành phần bộ đệm hình ảnh được giải mã

423. Các khối ảnh đã tái tạo từ thành phần bộ đệm hình ảnh được giải mã 423 được chuyển tiếp đến thành phần bù chuyển động 421 để dự đoán liên ảnh. Thành phần bù chuyển động 421 có thể gần như tương tự như thành phần ước lượng chuyển động 221 và/hoặc thành phần bù chuyển động 219. Cụ thể, thành phần bù chuyển động 421 sử dụng các vector chuyển động từ khối tham chiếu để tạo khối dự đoán và áp dụng khối dư vào kết quả để tái tạo khối ảnh. Các khối được tái tạo thành cũng có thể được chuyển tiếp qua thành phần các bộ lọc nội vòng 425 đến thành phần bộ đệm hình ảnh được giải mã 423. Thành phần bộ đệm hình ảnh được giải mã 423 tiếp tục lưu trữ các khối ảnh đã tái tạo bổ sung, mà có thể được tái tạo thành các khung qua thông tin phân chia. Các khung này cũng có thể được đặt trong một chuỗi. Chuỗi này được kết xuất đến bộ hiển thị dưới dạng tín hiệu video đầu ra đã tái tạo.

Fig.5 là hình vẽ giản lược minh họa CVS 500 làm ví dụ. Ví dụ, CVS 500 có thể được mã hóa bởi bộ mã hóa, như hệ thống codec 200 và/hoặc bộ mã hóa 300, theo phương pháp 100. Hơn nữa, CVS 500 có thể được giải mã bởi bộ giải mã, như hệ thống codec 200 và/hoặc bộ giải mã 400. CVS 500 bao gồm các hình ảnh được tạo mã theo thứ tự giải mã 508. Thứ tự giải mã 508 là thứ tự trong đó các hình ảnh được định vị trong dòng bit. Các hình ảnh của CVS 500 sau đó được kết xuất theo thứ tự trình diễn 510. Thứ tự trình diễn 510 là thứ tự trong đó các hình ảnh sẽ được hiển thị bởi bộ giải mã để làm cho video tạo thành để hiển thị một cách chính xác. Ví dụ, các hình ảnh của CVS 500 có thể được định vị chung theo thứ tự trình diễn 510. Tuy nhiên, các hình ảnh nhất định có thể được di chuyển vào các vị trí khác nhau để làm tăng hiệu suất tạo mã, ví dụ bằng cách đặt các hình ảnh tương tự ở vị trí lân cận gần hơn để hỗ trợ dự đoán liên ảnh. Sự di chuyển các hình ảnh như vậy theo cách này dẫn đến thứ tự giải mã 508. Trong ví dụ được thể hiện, các hình ảnh được đánh chỉ số theo thứ tự giải mã 508 từ không (zero) đến bốn. Theo thứ tự trình diễn 510, các hình ảnh ở chỉ số hai và chỉ số ba đã được di chuyển trước hình ảnh ở chỉ số không (zero).

CVS 500 bao gồm hình ảnh IRAP 502. Hình ảnh IRAP 502 là hình ảnh được tạo mã theo dự đoán trong ảnh mà dùng làm điểm truy cập ngẫu nhiên đối với CVS 500. Cụ thể, các khối của hình ảnh IRAP 502 được tạo mã bằng cách tham chiếu đến các khối khác của hình ảnh IRAP 502. Do hình ảnh IRAP 502 được tạo mã mà không tham chiếu đến các hình ảnh khác, nên hình ảnh IRAP 502 có thể được giải mã mà không phải là

trước tiên giải mã các hình ảnh khác bất kỳ. Theo đó, bộ giải mã có thể bắt đầu giải mã CVS 500 ở hình ảnh IRAP 502. Hơn nữa, hình ảnh IRAP 502 có thể làm DPB được làm mới. Ví dụ, các hình ảnh được trình diễn sau khi hình ảnh IRAP 502 có thể không dựa trên các hình ảnh trước hình ảnh IRAP 502 (ví dụ, chỉ số hình ảnh không (zero)) để dự đoán liên ảnh. Như vậy, bộ đệm hình ảnh có thể được làm mới một khi hình ảnh IRAP 502 được giải mã. Điều này có hiệu quả dừng các lỗi tạo mã liên quan đến dự đoán liên ảnh bất kỳ do các lỗi này không thể lan truyền qua hình ảnh IRAP 502. Hình ảnh IRAP 502 có thể bao gồm các loại hình ảnh khác nhau. Ví dụ, hình ảnh IRAP có thể được tạo mã như là IDR hoặc CRA. IDR là hình ảnh được tạo mã trong ảnh mà bắt đầu CVS 500 mới và làm mới bộ đệm hình ảnh. CRA là hình ảnh được tạo mã trong ảnh mà đóng vai trò là điểm truy cập ngẫu nhiên mà không bắt đầu CVS 500 mới hoặc làm mới bộ đệm hình ảnh. Theo cách này, các hình ảnh chủ đạo 504 được liên kết với CRA có thể tham chiếu các hình ảnh trước CRA, trong khi các hình ảnh chủ đạo 504 được liên kết với IDR có thể không tham chiếu các hình ảnh trước IDR.

CVS 500 cũng bao gồm các hình ảnh phi IRAP khác nhau. Các hình ảnh này bao gồm các hình ảnh chủ đạo 504 và các hình ảnh kế tiếp 506. Hình ảnh chủ đạo 504 là hình ảnh được định vị sau hình ảnh IRAP 502 theo thứ tự giải mã 508, nhưng được định vị trước hình ảnh IRAP 502 theo thứ tự trình diễn 510. Các hình ảnh kế tiếp 506 được định vị sau hình ảnh IRAP 502 theo cả thứ tự giải mã 508 lẫn thứ tự trình diễn 510. Các hình ảnh chủ đạo 504 và các hình ảnh kế tiếp 506 đều được tạo mã theo dự đoán liên ảnh. Các hình ảnh kế tiếp 506 được tạo mã khi tham chiếu đến hình ảnh IRAP 502 hoặc các hình ảnh được định vị sau hình ảnh IRAP 502. Vì vậy, các hình ảnh kế tiếp 506 có thể luôn được giải mã một khi hình ảnh IRAP 502 được giải mã. Các hình ảnh chủ đạo 504 có thể bao gồm các hình ảnh chủ đạo bỏ qua truy cập ngẫu nhiên (random access skipped leading, RASL) và chủ đạo giải mã được truy cập ngẫu nhiên (random access decodable leading, RADL). Hình ảnh RASL được tạo mã bằng cách tham chiếu đến các hình ảnh trước hình ảnh IRAP 502, nhưng được tạo mã ở vị trí sau hình ảnh IRAP 502. Do các hình ảnh RASL dựa trên các hình ảnh trước đó, nên hình ảnh RASL không thể được giải mã khi bộ giải mã bắt đầu giải mã ở hình ảnh IRAP 502. Theo đó, các hình ảnh RASL được bỏ qua và không được giải mã khi hình ảnh IRAP 502 được sử dụng làm điểm truy cập ngẫu nhiên. Tuy nhiên, các hình ảnh RASL được giải mã và được hiển thị khi bộ giải mã sử dụng hình

ảnh IRAP trước đó (trước chỉ số không (zero) và không được thể hiện) làm điểm truy cập ngẫu nhiên. Các hình ảnh RADL được tạo mã khi tham chiếu đến hình ảnh IRAP 502 và/hoặc các hình ảnh theo sau hình ảnh IRAP 502, nhưng được định vị trước hình ảnh IRAP 502 theo thứ tự trình diễn 510. Do các hình ảnh RADL không dựa trên các hình ảnh trước hình ảnh IRAP 502, các hình ảnh RADL có thể được giải mã và được hiển thị khi hình ảnh IRAP 502 là điểm truy cập ngẫu nhiên.

Các hình ảnh từ CVS 500, từng hình ảnh có thể được lưu trữ trong đơn vị truy cập. Hơn nữa, các hình ảnh có thể được phân chia thành các lát, và các lát có thể có trong các đơn vị NAL. Đơn vị NAL là bộ phận lưu trữ mà chứa tập hợp thông số hoặc lát của hình ảnh và phần đầu lát tương ứng. Các đơn vị NAL là các loại được ấn định để chỉ báo cho bộ giải mã loại của dữ liệu có trong đơn vị NAL. Ví dụ, các lát từ hình ảnh IRAP 502 có thể có trong đơn vị NAL của IDR với RADL (IDR_W_RADL), đơn vị NAL của IDR không có các hình ảnh chủ đạo (IDR_N_LP), đơn vị NAL CRA, v.v. Đơn vị NAL IDR_W_RADL chỉ báo hình ảnh IRAP 502 là hình ảnh IDR mà được liên kết với hình ảnh chủ đạo RADL 504. Đơn vị NAL IDR_N_LP chỉ báo hình ảnh IRAP 502 là hình ảnh IDR mà không được liên kết với các hình ảnh chủ đạo 504 bất kỳ. Đơn vị NAL CRA chỉ báo hình ảnh IRAP 502 là hình ảnh CRA mà có thể được liên kết với các hình ảnh chủ đạo 504. Các lát của các hình ảnh phi IRAP cũng có thể được đặt vào các đơn vị NAL. Ví dụ, các lát của các hình ảnh kế tiếp 506 có thể được đặt trong loại đơn vị NAL hình ảnh kế tiếp (TRAIL_NUT), mà chỉ báo các hình ảnh kế tiếp 506 được tạo mã dự đoán liên ảnh các hình ảnh. Các lát của các hình ảnh chủ đạo 504 có thể có trong loại đơn vị NAL RASL (RASL_NUT) và/hoặc loại đơn vị NAL RADL (RADL_NUT), mà có thể chỉ báo các hình ảnh tương ứng được tạo mã dự đoán liên ảnh các hình ảnh chủ đạo 504 của các loại tương ứng. Bằng cách báo hiệu các lát của các hình ảnh trong các đơn vị NAL tương ứng, bộ giải mã có thể xác định một cách dễ dàng các cơ chế giải mã thích hợp để áp dụng vào mỗi hình ảnh/lát.

Fig.6 là hình vẽ giản lược minh họa các luồng video hình ảnh con 601, 602, và 603 được chia tách từ luồng video hình ảnh VR 600. Ví dụ, mỗi luồng trong số các luồng video hình ảnh con 601- 603 và/hoặc luồng video hình ảnh VR 600 có thể được tạo mã trong CVS 500. Theo đó, các luồng video hình ảnh con 601- 603 và/hoặc luồng video hình ảnh VR 600 có thể được mã hóa bởi bộ mã hóa, như hệ thống codec 200 và/hoặc bộ

mã hóa 300, theo phương pháp 100. Hơn nữa, các luồng video hình ảnh con 601-603 và/hoặc luồng video hình ảnh VR 600 có thể được giải mã bởi bộ giải mã, như hệ thống codec 200 và/hoặc bộ giải mã 400.

Luồng video hình ảnh VR 600 bao gồm các hình ảnh được trình diễn theo thời gian. Cụ thể, VR hoạt động bằng cách tạo mã hình cầu của nội dung video, mà có thể được hiển thị như thể người dùng ở trung tâm của hình cầu. Mỗi hình ảnh bao gồm toàn bộ hình cầu. Trong khi đó, chỉ một phần của hình ảnh, được biết đến là cổng quan sát, được hiển thị cho người dùng. Ví dụ, người dùng có thể sử dụng thiết bị hiển thị lắp trên đầu (HMD) mà chọn và hiển thị cổng quan sát của hình cầu dựa trên chuyển động đầu của người dùng. Điều này mang lại ấn tượng về sự có mặt về mặt vật lý trong không gian ảo như được thể hiện bởi video. Để đạt được kết quả này, mỗi hình ảnh của chuỗi video bao gồm toàn bộ hình cầu của dữ liệu video ở thời điểm tương ứng. Tuy nhiên, chỉ một phần nhỏ (ví dụ, một cổng quan sát đơn lẻ) của hình ảnh được hiển thị cho người dùng. Phần còn lại của hình ảnh được loại bỏ mà không được thực hiện. Toàn bộ hình ảnh thường được truyền sao cho cổng quan sát khác nhau có thể được chọn theo cách động và được hiển thị đáp lại chuyển động đầu của người dùng.

Trong ví dụ được thể hiện, các hình ảnh của luồng video hình ảnh VR 600, mỗi hình ảnh có thể được chia nhỏ thành các hình ảnh con dựa trên các cổng quan sát khả dụng. Theo đó, mỗi hình ảnh và hình ảnh con tương ứng bao gồm vị trí thời gian (ví dụ, thứ tự hình ảnh) như là một phần của việc trình diễn theo thời gian. Các luồng video hình ảnh con 601- 603 được tạo ra khi sự chia nhỏ được áp dụng một cách thích hợp theo thời gian. Sự chia nhỏ thích hợp này tạo các luồng video hình ảnh con 601-603 trong đó mỗi luồng chứa tập hợp các hình ảnh con có kích thước, hình dạng, và vị trí không gian định trước so với các hình ảnh tương ứng trong luồng video hình ảnh VR 600. Hơn nữa, tập hợp các hình ảnh con trong luồng video hình ảnh con 601-603 thay đổi theo vị trí thời gian trong thời gian trình diễn. Như vậy, các hình ảnh con của các luồng video hình ảnh con 601-603 có thể được căn thẳng trong miền thời gian dựa trên vị trí thời gian. Sau đó, các hình ảnh con từ các luồng video hình ảnh con 601-603 tại mỗi vị trí thời gian có thể được hợp nhất trong miền không gian dựa trên vị trí không gian định trước để tái tạo luồng video hình ảnh VR 600 để hiển thị. Cụ thể, các luồng video hình ảnh con 601-603, mỗi luồng có thể được mã hóa thành các dòng bit con riêng rẽ. Khi các dòng bit con này được

hợp nhất với nhau, chúng dẫn đến dòng bit mà bao gồm toàn bộ tập hợp các hình ảnh theo thời gian. Dòng bit tạo thành có thể được truyền đến bộ giải mã để giải mã và hiển thị dựa trên công quan sát được chọn hiện hành của người dùng.

Một trong số các vấn đề với video VR là việc tất cả các luồng video hình ảnh con 601-603 có thể được truyền đến người dùng ở chất lượng cao (ví dụ, độ phân giải cao). Điều này cho phép bộ giải mã chọn theo cách động công quan sát hiện hành của người dùng và hiển thị (các) hình ảnh con từ các luồng video hình ảnh con 601-603 tương ứng theo thời gian thực. Tuy nhiên, người dùng có thể chỉ xem một công quan sát đơn lẻ, ví dụ từ luồng video hình ảnh con 601, trong khi các luồng video hình ảnh con 602-603 được loại bỏ. Như vậy, việc truyền các luồng video hình ảnh con 602-603 ở chất lượng cao có thể làm lãng phí một lượng đáng kể. Để nâng cao hiệu suất tạo mã, video VR có thể được mã hóa thành các luồng video 600 trong đó mỗi luồng video 600 được mã hóa ở chất lượng/độ phân giải khác nhau. Theo cách này, bộ giải mã có thể truyền yêu cầu luồng video hình ảnh con 601 hiện hành. Đáp lại, bộ mã hóa (hoặc bộ tạo lát trung gian hoặc máy chủ nội dung khác) có thể chọn luồng video hình ảnh con 601 chất lượng cao hơn từ luồng video 600 chất lượng cao hơn và các luồng video hình ảnh con 602-603 chất lượng thấp hơn từ luồng video 600 chất lượng thấp hơn. Bộ mã hóa sau đó có thể hợp nhất các dòng bit con này với nhau thành dòng bit được mã hóa hoàn chỉnh để truyền đến bộ giải mã. Theo cách này, bộ giải mã nhận một loạt hình ảnh nơi mà công quan sát hiện hành là chất lượng cao hơn và các công quan sát khác là chất lượng thấp hơn. Hơn nữa, các hình ảnh con chất lượng cao nhất thường được hiển thị cho người dùng (chuyển động đầu lơ đãng) và các hình ảnh con chất lượng thấp hơn thường được loại bỏ, mà làm cân bằng chức năng với hiệu suất tạo mã.

Trong trường hợp người dùng chuyển từ xem luồng video hình ảnh con 601 sang luồng video hình ảnh con 602, bộ giải mã yêu cầu luồng video hình ảnh con 602 hiện hành mới được truyền ở chất lượng cao hơn. Sau đó bộ mã hóa có thể thay đổi cơ chế hợp nhất theo đó. Như được lưu ý trên đây, bộ giải mã có thể chỉ bắt đầu giải mã CVS 500 mới ở hình ảnh IRAP 502. Theo đó, luồng video hình ảnh con 602 được hiển thị ở chất lượng thấp hơn cho đến khi hình ảnh IRAP/hình ảnh con được đạt đến. Hình ảnh IRAP sau đó có thể được giải mã ở chất lượng cao hơn để bắt đầu giải mã phiên bản chất lượng cao

hơn của luồng video hình ảnh con 602. Cách tiếp cận này làm tăng đáng kể độ nén video mà không tác động tiêu cực đến trải nghiệm xem của người dùng.

Một vấn đề với cách tiếp cận trên đây là độ dài của việc thời gian cần để thay đổi các độ phân giải dựa trên độ dài của thời gian cho đến khi hình ảnh IRAP được đạt đến trong luồng video. Điều này là do bộ giải mã không thể bắt đầu giải mã phiên bản khác của luồng video hình ảnh con 602 ở hình ảnh phi IRAP. Một cách tiếp cận để làm giảm độ trễ này là để bao gồm nhiều hình ảnh IRAP hơn. Tuy nhiên, điều này dẫn đến việc tăng kích thước tệp. Để cân bằng chức năng với hiệu suất tạo mã, các công quan sát/các luồng video hình ảnh con 601-603 khác nhau có thể bao gồm các hình ảnh IRAP ở các tần số khác nhau. Ví dụ, các công quan sát/các luồng video hình ảnh con 601-603 mà có nhiều khả năng được xem hơn có thể có nhiều hình ảnh IRAP hơn so với các công quan sát/các luồng video hình ảnh con 601-603 khác. Ví dụ, trong ngữ cảnh bóng rổ, các công quan sát/các luồng video hình ảnh con 601-603 liên quan đến các rổ và/hoặc sân trung tâm có thể bao gồm các hình ảnh IRAP ở tần số lớn hơn so với các công quan sát/các luồng video hình ảnh con 601-603 mà xem các khán đài hoặc trần do các công quan sát/các luồng video hình ảnh con 601-603 này ít có khả năng được xem bởi người dùng.

Cách tiếp cận này dẫn đến các vấn đề thêm nữa. Cụ thể, các hình ảnh con từ các luồng video hình ảnh con 601-603 mà chia sẻ POC là một phần của hình ảnh đơn lẻ. Như được lưu ý trên đây, các lát từ một hình ảnh được bao gồm trong đơn vị NAL dựa trên loại hình ảnh. Trong một số hệ thống tạo mã video, tất cả các đơn vị NAL liên quan đến một hình ảnh đơn lẻ được ràng buộc để bao gồm cùng một loại đơn vị NAL. Khi các luồng video hình ảnh con 601-603 khác nhau có các hình ảnh IRAP ở các tần số khác nhau, một số trong số các hình ảnh này bao gồm cả các hình ảnh con IRAP lẫn các hình ảnh con phi IRAP. Điều này vi phạm ràng buộc mà mỗi hình ảnh đơn lẻ sẽ sử dụng chỉ các đơn vị NAL của cùng một loại.

Sáng chế khắc phục vấn đề này bằng cách loại bỏ ràng buộc là tất cả các đơn vị NAL đối với các lát trong hình ảnh sử dụng cùng một loại đơn vị NAL. Ví dụ, một hình ảnh có trong một đơn vị truy cập. Bằng cách loại bỏ ràng buộc này, đơn vị truy cập có thể bao gồm cả các loại đơn vị NAL IRAP và các loại đơn vị NAL phi IRAP. Hơn nữa, cờ có thể được mã hóa để chỉ báo khi hình ảnh/đơn vị truy cập bao gồm hỗn hợp các loại đơn

vị NAL IRAP và các loại đơn vị NAL phi IRAP. Trong một số ví dụ, cờ là các loại đơn vị NAL trộn trong cờ hình ảnh (`mixed_nalu_types_in_pic_flag`). Ngoài ra, ràng buộc có thể được áp dụng để đòi hỏi rằng một hình ảnh trộn/đơn vị truy cập đơn lẻ có thể chỉ chứa một loại của đơn vị NAL IRAP và một loại của đơn vị NAL phi IRAP. Điều này ngăn chặn các hỗn hợp loại đơn vị NAL không dự tính trước xảy ra. Nếu các hỗn hợp này được phép, thì bộ giải mã sẽ phải được thiết kế để quản lý các hỗn hợp này. Điều này sẽ làm tăng một cách không cần thiết độ phức tạp phần cứng được yêu cầu mà không mang lại lợi ích thêm nữa cho quy trình tạo mã. Ví dụ, hình ảnh trộn có thể bao gồm một loại đơn vị NAL IRAP được chọn từ `IDR_W_RADL`, `IDR_N_LP`, hoặc `CRA_NUT`. Hơn nữa, hình ảnh trộn có thể bao gồm một loại đơn vị NAL phi IRAP được chọn từ `TRAIL_NUT`, `RADL_NUT`, và `RASL_NUT`. Các cách thực hiện ví dụ của sơ đồ này được thảo luận chi tiết hơn dưới đây.

Fig.7 là hình vẽ giản lược minh họa dòng bit 700 làm ví dụ chứa các hình ảnh với các loại đơn vị NAL trộn. Ví dụ, dòng bit 700 có thể được tạo ra bởi hệ thống codec 200 và/hoặc bộ mã hóa 300 để giải mã bởi hệ thống codec 200 và/hoặc bộ giải mã 400 theo phương pháp 100. Hơn nữa, dòng bit 700 có thể bao gồm luồng video hình ảnh VR 600 được hợp nhất từ nhiều luồng video hình ảnh con 601-603 ở các độ phân giải video trong đó mỗi luồng video hình ảnh con chứa CVS 500 ở vị trí không gian khác nhau.

Dòng bit 700 bao gồm tập hợp thông số chuỗi (SPS) 710, các tập hợp thông số hình ảnh (PPS) 711, các phần đầu lát 715, và dữ liệu ảnh 720. SPS 710 chứa dữ liệu chuỗi chung cho tất cả các hình ảnh trong chuỗi video có trong dòng bit 700. Dữ liệu này có thể bao gồm sự định cỡ hình ảnh, độ sâu bit, các thông số công cụ tạo mã, các giới hạn tốc độ bit, v.v.. PPS 711 chứa các thông số mà áp dụng vào toàn bộ hình ảnh. Vì vậy, mỗi hình ảnh trong chuỗi video có thể tham chiếu đến PPS 711. Cần lưu ý rằng, mặc dù mỗi hình ảnh tham chiếu đến PPS 711, nhưng một PPS 711 có thể chứa dữ liệu đối với nhiều hình ảnh trong một số ví dụ. Ví dụ, nhiều hình ảnh tương tự có thể được tạo mã theo các thông số tương tự. Trong trường hợp như vậy, một PPS 711 có thể chứa dữ liệu đối với các hình ảnh tương tự này. PPS 711 có thể chỉ báo các công cụ tạo mã khả dụng đối với các lát trong các hình ảnh tương ứng, các thông số lượng tử hóa, các độ lệch, v.v.. Phần đầu lát 715 chứa các thông số mà là đặc trưng cho mỗi lát trong một hình ảnh. Vì vậy, có thể có một phần đầu lát 715 cho mỗi lát trong chuỗi video. Phần đầu lát 715 có thể chứa thông

tin loại lát, các số đếm thứ tự hình ảnh (picture order count, POC), các danh sách hình ảnh tham chiếu, các trọng số dự đoán, các điểm mục nhập ảnh xếp lớp (tile), các thông số giải khối, v.v.. Cần lưu ý rằng phần đầu lát 715 cũng có thể được gọi là phần đầu nhóm ảnh xếp lớp trong một số ngữ cảnh.

Dữ liệu ảnh 720 chứa dữ liệu video được mã hóa theo dự đoán liên ảnh và/hoặc dự đoán trong ảnh cũng dữ liệu dư như được biến đổi và được lượng tử hóa tương ứng. Ví dụ, chuỗi video bao gồm các hình ảnh 721 được tạo mã như là dữ liệu ảnh 720. Hình ảnh 721 là một khung đơn lẻ của chuỗi video và vì vậy, thường được hiển thị như là một bộ phận đơn lẻ khi hiển thị chuỗi video. Tuy nhiên, các hình ảnh con 723 có thể được hiển thị để thực hiện các kỹ thuật nhất định như thực tế ảo. Các hình ảnh 721, mỗi hình ảnh tham chiếu PPS 711. Các hình ảnh 721 có thể được chia thành các hình ảnh con 723, các ảnh xếp lớp, và/hoặc các lát. Hình ảnh con 723 là vùng không gian của hình ảnh 721 mà được áp dụng một cách thích hợp trong chuỗi video được tạo mã. Theo đó, hình ảnh con 723 có thể được hiển thị bởi HMD trong ngữ cảnh VR. Hơn nữa, hình ảnh con 723 với POC được quy định có thể được thu nhận từ luồng video hình ảnh con 601-603 ở độ phân giải tương ứng. Hình ảnh con 723 có thể tham chiếu SPS 710. Trong một số hệ thống, các lát 725 được gọi là các nhóm ảnh xếp lớp chứa các ảnh xếp lớp. Các lát 725 và/hoặc các nhóm ảnh xếp lớp của các ảnh xếp lớp tham chiếu phần đầu lát 715. Lát 725 có thể được xác định là số lượng nguyên của các ảnh xếp lớp hoàn chỉnh hoặc số lượng nguyên của các hàng CTU hoàn chỉnh liên tiếp trong ảnh xếp lớp của hình ảnh 721 mà được bao gồm một cách riêng biệt trong một đơn vị NAL riêng lẻ. Vì vậy, các lát 725 được chia thêm nữa thành các CTU và/hoặc các CTB. Các CTU/các CTB được chia thêm nữa thành các khối tạo mã dựa trên các cây tạo mã. Sau đó, các khối tạo mã có thể được mã hóa/được giải mã theo các cơ chế dự đoán.

Các tập hợp thông số và/hoặc các lát 725 được tạo mã trong các đơn vị NAL. Đơn vị NAL có thể được xác định là cấu trúc cú pháp chứa chỉ báo về loại dữ liệu để tuân theo và các bai (byte) chứa dữ liệu đó ở dạng RBSP được đặt rải rác khi cần với các bai ngăn chặn mô phỏng. Cụ thể hơn, đơn vị NAL là bộ phận lưu trữ mà chứa tập hợp thông số hoặc lát 725 của hình ảnh 721 và phần đầu lát 715 tương ứng. Cụ thể, các đơn vị NAL VCL 740 là các đơn vị NAL chứa lát 725 của hình ảnh 721 và phần đầu lát 715 tương ứng. Hơn nữa, các đơn vị NAL phi VCL 730 chứa các tập hợp thông số, như SPS 710 và

PPS 711. Một số loại đơn vị NAL có thể được sử dụng. Ví dụ, SPS 710 và PPS 711 có thể lần lượt có trong loại đơn vị NAL SPS (SPS_NUT) 731 và loại đơn vị NAL PPS (PPS_NUT) 732, mà đều là các đơn vị NAL phi VCL 730.

Như được lưu ý trên đây, các hình ảnh IRAP, như hình ảnh IRAP 502, có thể có trong các đơn vị NAL IRAP 745. Các hình ảnh phi IRAP, như các hình ảnh chủ đạo 504 và các hình ảnh kế tiếp 506, có thể có trong các đơn vị NAL phi IRAP 749. Cụ thể, đơn vị NAL IRAP 745 là đơn vị NAL bất kỳ chứa lát 725 được lấy từ hình ảnh IRAP hoặc hình ảnh con. Đơn vị NAL phi IRAP 749 là đơn vị NAL bất kỳ chứa lát 725 được lấy từ hình ảnh bất kỳ mà không phải là hình ảnh IRAP hoặc hình ảnh con (ví dụ, các hình ảnh chủ đạo hoặc các hình ảnh kế tiếp). Các đơn vị NAL IRAP 745 và các đơn vị NAL phi IRAP 749 đều là các đơn vị NAL VCL 740 do chúng đều chứa dữ liệu lát. Theo một phương án làm ví dụ, đơn vị NAL IRAP 745 có thể bao gồm các lát 725 từ hình ảnh IDR mà không có các hình ảnh chủ đạo hoặc IDR được liên kết với các hình ảnh RADL trong đơn vị NAL IDR_N_LP 741 hoặc đơn vị NAL IDR_w_RADL 742 một cách tương ứng. Hơn nữa, đơn vị NAL IRAP 745 có thể bao gồm các lát 725 từ hình ảnh CRA trong CRA_NUT 743. Theo một phương án làm ví dụ, đơn vị NAL phi IRAP 749 có thể bao gồm các lát 725 từ hình ảnh RASL, hình ảnh RADL, hoặc hình ảnh kế tiếp, trong RASL_NUT 746, RADL_NUT 747, hoặc TRAIL_NUT 748 một cách tương ứng. Theo một phương án làm ví dụ, danh sách hoàn chỉnh của các đơn vị NAL khả thi được thể hiện dưới đây như được phân loại bởi loại đơn vị NAL.

nal_unit_type	Tên của nal_unit_type	Nội dung của đơn vị NAL và cấu trúc cú pháp RBSP	Lớp loại đơn vị NAL
0	TRAIL_NUT	Lát được tạo mã của hình ảnh kế tiếp slice_layer_rbsp()	VCL
1	STSA_NUT	Lát được tạo mã của hình ảnh STSA slice_layer_rbsp()	VCL

2	RADL_NUT	Lát được tạo mã của hình ảnh RADL slice_layer_rbsp()	VCL
3	RASL_NUT	Lát được tạo mã của hình ảnh RASL slice_layer_rbsp()	VCL
4..6	RSV_VCL_4.. RSV_VCL_6	Các loại đơn vị NAL VCL phi IRAP được dành riêng	VCL
7 8	IDR_W_RADL IDR_N_LP	Lát được tạo mã của hình ảnh IDR slice_layer_rbsp()	VCL
9	CRA_NUT	Lát được tạo mã của hình ảnh CRA silce_layer_rbsp()	VCL
10	GDR_NUT	Lát được tạo mã của hình ảnh GDR slice_layer_rbsp()	VCL
11 12	RSV_IRAP_11 RSV_IRAP_12	Các loại đơn vị NAL VCL IRAP được dành riêng	VCL
13	DCI_NUT	Thông tin khả năng giải mã decoding_capability_information_rbsp()	phi VCL
14	VPS_NUT	Tập hợp thông số video video_parameter_set_rbsp()	phi VCL
15	SPS_NUT	Tập hợp thông số chuỗi seq_parameter_set_rbsp()	phi VCL
16	PPS_NUT	Tập hợp thông số hình ảnh pic_parameter_set_rbsp()	phi VCL
17 18	PREFIX_APS_NUT SUFFIX_APS_NUT	Tập hợp thông số thích ứng adaptation_parameter_set_rbsp()	phi VCL
19	PH_NUT	Phần đầu hình ảnh picture_header_rbsp()	phi VCL
20	AUD_NUT	Dấu tách AU access_unit_delimiter_rbsp()	phi VCL

21	EOS_NUT	Hết chuỗi end_of_seq_rbsp()	phi VCL
22	EOB_NUT	Hết dòng bit end_of_bitstream_rbsp()	phi VCL
23 24	PREFIX_SEI_ NUT SUFFIX_SEI_ NUT	Thông tin nâng cao phụ sei_rbsp()	phi VCL
25	FD_NUT	Dữ liệu điền đầy filler_data_rbsp()	phi VCL
26 27	RSV_NVCL_2 6 RSV_NVCL_2 7	Các loại đơn vị NAL phi VCL được dành riêng	phi VCL
28..31	UNSPEC_28.. UNSPEC_31	Các loại đơn vị NAL phi VCL không được quy định	phi VCL

Như được lưu ý trên đây, luồng video VR có thể bao gồm các hình ảnh con 723 với các hình ảnh IRAP ở các tần số khác nhau. Điều này cho phép ít hình ảnh IRAP hơn được sử dụng cho các vùng không gian mà người dùng ít có khả năng nhìn vào và nhiều hình ảnh IRAP hơn được sử dụng cho các vùng không gian mà người dùng rất có khả năng xem thường xuyên. Theo cách này, vùng không gian mà người dùng rất có khả năng chuyển về thường xuyên có thể được điều chỉnh nhanh chóng đến độ phân giải cao hơn. Khi cách tiếp cận này dẫn đến hình ảnh 721 mà bao gồm cả các đơn vị NAL IRAP 745 lẫn các đơn vị NAL phi IRAP 749, thì hình ảnh 721 được gọi là hình ảnh trộn. Điều kiện này có thể được báo hiệu bởi các loại đơn vị NAL trộn trong cờ hình ảnh (mixed_nalu_types_in_pic_flag) 727. mixed_nalu_types_in_pic_flag 727 có thể được thiết đặt trong PPS 711. Hơn nữa, mixed_nalu_types_in_pic_flag 727 có thể được thiết đặt bằng một khi quy định rằng mỗi hình ảnh 721 tham chiếu đến PPS 711 có nhiều hơn một đơn vị NAL VCL 740 và các đơn vị NAL VCL 740 không có cùng giá trị của loại đơn vị NAL (nal_unit_type). Hơn nữa, mixed_nalu_types_in_pic_flag 727 có thể được thiết đặt bằng không (zero) khi mỗi hình ảnh 721 tham chiếu đến PPS 711 có một hoặc

nhiều đơn vị NAL VCL 740 và các đơn vị NAL VCL 740 của mỗi hình ảnh 721 tham chiếu đến PPS 711 đều có cùng giá trị của `nal_unit_type`.

Hơn nữa, ràng buộc có thể được sử dụng sao cho các đơn vị NAL VCL 740 của một hoặc nhiều hình ảnh con 723 của hình ảnh 721 đều có giá trị cụ thể thứ nhất của loại đơn vị NAL và các đơn vị NAL VCL 740 khác trong hình ảnh 721 đều có giá trị cụ thể thứ hai khác của loại đơn vị NAL khi `mixed_nalu_types_in_pic_flag` 727 được thiết đặt. Ví dụ, ràng buộc này có thể đòi hỏi rằng hình ảnh trộn 721 chứa một loại đơn vị NAL IRAP 745 và một loại đơn vị NAL phi IRAP 749. Ví dụ, hình ảnh 721 có thể bao gồm một hoặc nhiều đơn vị NAL IDR_N_LP 741, một hoặc nhiều đơn vị NAL IDR_w_RADL 742, hoặc một hoặc nhiều CRA_NUT 743, nhưng không bao gồm sự kết hợp bất kỳ của các đơn vị NAL IRAP 745 đó. Hơn nữa, hình ảnh 721 có thể bao gồm một hoặc nhiều RASL_NUT 746, một hoặc nhiều RADL_NUT 747, hoặc một hoặc nhiều TRAIL_NUT 748, nhưng không bao gồm sự kết hợp bất kỳ của các đơn vị NAL IRAP 745 đó.

Theo một cách thực hiện ví dụ, các loại hình ảnh được sử dụng để xác định các quy trình giải mã. Các quy trình này bao gồm việc dẫn xuất nhận dạng hình ảnh, ví dụ nhờ đếm thứ tự hình ảnh (POC), đánh dấu trạng thái của các hình ảnh tham chiếu trong bộ đệm hình ảnh được giải mã (DPB), kết xuất các hình ảnh từ DPB, v.v.. Hình ảnh có thể được nhận dạng bởi loại dựa trên loại đơn vị NAL mà chứa tất cả hình ảnh được tạo mã hoặc phần con của nó. Trong một số hệ thống tạo mã video, các loại hình ảnh có thể bao gồm các hình ảnh làm mới quá trình giải mã tức thời (IDR) và các hình ảnh phi IDR. Trong các hệ thống tạo mã video khác, các loại hình ảnh có thể bao gồm các hình ảnh kế tiếp, các hình ảnh truy cập lớp con thời gian (temporal sub-layer access, TSA), các hình ảnh truy cập lớp con thời gian theo bước (step-wise temporal sub-layer access, STSA), các hình ảnh chủ đạo giải mã được truy cập ngẫu nhiên (RADL), các hình ảnh chủ đạo bỏ qua truy cập ngẫu nhiên (RASL), các hình ảnh truy cập liên kết gãy (broken-link access, BLA), các hình ảnh truy cập ngẫu nhiên tức thời, và các hình ảnh truy cập ngẫu nhiên sạch. Các loại hình ảnh này có thể còn được phân biệt dựa trên việc liệu hình ảnh này là hình ảnh được tham chiếu lớp con hoặc hình ảnh không được tham chiếu lớp con. Hình ảnh BLA có thể còn được phân biệt là BLA có hình ảnh chủ đạo, BLA có hình ảnh RADL,

và BLA không có hình ảnh chủ đạo. Hình ảnh IDR có thể còn được phân biệt dưới dạng IDR có hình ảnh RADL và IDR không có hình ảnh chủ đạo.

Các loại hình ảnh này có thể được sử dụng để thực thi các chức năng liên quan đến video khác nhau. Ví dụ, các hình ảnh IDR, BLA, và/hoặc CRA có thể được sử dụng để thực thi hình ảnh IRAP. Hình ảnh IRAP có thể cung cấp các chức năng / các lợi ích sau đây. Sự có mặt của hình ảnh IRAP có thể chỉ báo rằng quy trình giải mã có thể được khởi tạo từ hình ảnh đó. Chức năng này cho phép thực thi dấu hiệu truy cập ngẫu nhiên trong đó quy trình giải mã bắt đầu ở vị trí được chỉ định trong dòng bit miễn là hình ảnh IRAP có mặt ở vị trí đó. Vị trí này không nhất thiết phải ở đầu của dòng bit. Sự có mặt của hình ảnh IRAP cũng làm mới quy trình giải mã sao cho các hình ảnh được tạo mã bắt đầu tại hình ảnh IRAP, không bao gồm các hình ảnh RASL, được tạo mã mà không cần sự tham chiếu bất kỳ đến các hình ảnh được định vị trước hình ảnh IRAP. Theo đó, hình ảnh IRAP được định vị trong dòng bit dừng việc lan truyền các lỗi giải mã. Vì vậy, các lỗi giải mã của các hình ảnh được tạo mã được định vị trước hình ảnh IRAP không thể lan truyền qua hình ảnh IRAP và vào các hình ảnh mà theo sau hình ảnh IRAP theo thứ tự giải mã.

Các hình ảnh IRAP mang lại các chức năng khác nhau, nhưng tạo ra sự bất lợi cho hiệu suất nén. Theo đó, sự có mặt của hình ảnh IRAP có thể gây ra tăng đột biến về tốc độ bit. Bất lợi này đối với hiệu suất nén có nhiều nguyên nhân khác nhau. Ví dụ, hình ảnh IRAP là hình ảnh được dự đoán trong ảnh mà được biểu diễn bởi nhiều bit hơn đáng kể so với các hình ảnh được dự đoán liên ảnh được sử dụng làm các hình ảnh phi IRAP. Hơn nữa, sự có mặt của hình ảnh IRAP phá vỡ dự đoán thời gian được sử dụng trong dự đoán liên ảnh. Cụ thể, hình ảnh IRAP làm mới quy trình giải mã bằng cách loại bỏ các hình ảnh tham chiếu trước đó khỏi DPB. Việc loại bỏ các hình ảnh tham chiếu trước đó làm giảm tính khả dụng của các hình ảnh tham chiếu để sử dụng khi tạo mã các hình ảnh mà theo sau hình ảnh IRAP theo thứ tự giải mã, và vì vậy làm giảm hiệu suất của quy trình này.

Các hình ảnh IDR có thể sử dụng các quy trình báo hiệu và dẫn xuất khác với các loại hình ảnh IRAP khác. Ví dụ, các quy trình báo hiệu và dẫn xuất liên quan đến IDR có thể thiết đặt phần bit có trọng số lớn nhất (MSB) của POC ở không (zero) thay vì dẫn xuất MSB từ hình ảnh khóa trước đó. Hơn nữa, phần đầu lát của hình ảnh IDR có thể không chứa thông tin được sử dụng để hỗ trợ khi quản lý hình ảnh tham chiếu. Trong khi đó, các

loại hình ảnh khác, như CRA, kế tiếp, TSA, v.v., có thể chứa thông tin hình ảnh tham chiếu như tập hợp hình ảnh tham chiếu (reference picture set, RPS) hoặc danh sách hình ảnh tham chiếu, mà có thể được sử dụng để thực thi quy trình đánh dấu các hình ảnh tham chiếu. Quy trình đánh dấu các hình ảnh tham chiếu là quy trình xác định trạng thái của các hình ảnh tham chiếu trong DPB là được sử dụng để tham chiếu hoặc không được sử dụng để tham chiếu. Đối với các hình ảnh IDR, thông tin này có thể không được báo hiệu do sự có mặt của IDR chỉ báo rằng quy trình giải mã sẽ đánh dấu một cách đơn giản tất cả các hình ảnh tham chiếu trong DPB là không được sử dụng để tham chiếu.

Ngoài các loại hình ảnh, việc nhận dạng hình ảnh bởi POC cũng được sử dụng cho nhiều mục đích, như để sử dụng các hình ảnh tham chiếu quản lý trong việc dự đoán liên ảnh, để kết xuất các hình ảnh từ DPB, để định tỷ lệ các vectơ chuyển động, để dự đoán có trọng số, v.v.. Ví dụ, trong một số hệ thống tạo mã video, các hình ảnh trong DPB có thể được đánh dấu là được sử dụng để tham chiếu ngắn hạn, được sử dụng để tham chiếu dài hạn, hoặc không được sử dụng để tham chiếu. Một khi hình ảnh đã được đánh dấu là không được sử dụng để tham chiếu thì hình ảnh có thể không được sử dụng để dự đoán nữa. Khi hình ảnh này không còn cần để kết xuất thì hình ảnh này có thể được loại bỏ khỏi DPB. Trong các hệ thống tạo mã video khác, các hình ảnh tham chiếu có thể được đánh dấu là ngắn hạn và dài hạn. Hình ảnh tham chiếu có thể được đánh dấu là không được sử dụng để tham chiếu khi hình ảnh không còn cần để tham chiếu dự đoán. Sự chuyển đổi giữa các trạng thái này có thể được điều khiển bởi quy trình đánh dấu hình ảnh tham chiếu được giải mã. Quy trình của sổ trượt ngầm và/hoặc quy trình hoạt động điều khiển quản lý bộ nhớ (memory management control operation, MMCO) rõ ràng có thể được sử dụng làm các cơ chế đánh dấu hình ảnh tham chiếu được giải mã. Quy trình của sổ trượt đánh dấu hình ảnh tham chiếu ngắn hạn là không được sử dụng để tham chiếu khi số lượng khung tham chiếu bằng số lớn nhất được quy định được biểu thị là `max_num_ref_frames` trong SPS. Các hình ảnh tham chiếu ngắn hạn có thể được lưu trữ theo cách vào trước ra trước (first-in first-out) sao cho các hình ảnh ngắn hạn được giải mã gần nhất được giữ lại trong DPB. Quy trình MMCO rõ ràng có thể bao gồm nhiều lệnh MMCO. Lệnh MMCO có thể đánh dấu một hoặc nhiều hình ảnh tham chiếu ngắn hạn hoặc dài hạn là không được sử dụng để tham chiếu, đánh dấu tất cả các hình ảnh là không được sử dụng để tham chiếu, hoặc đánh dấu hình ảnh tham chiếu hiện hành hoặc

hình ảnh tham chiếu ngắn hạn sẵn có là dài hạn và ấn định chỉ số hình ảnh dài hạn cho hình ảnh tham chiếu dài hạn.

Trong một số hệ thống tạo mã video, các hoạt động đánh dấu hình ảnh tham chiếu cũng như các quy trình để kết xuất và loại bỏ các hình ảnh khỏi DPB được thực hiện sau khi hình ảnh đã được giải mã. Các hệ thống tạo mã video khác sử dụng RPS để quản lý hình ảnh tham chiếu. Sự khác biệt cơ bản nhất giữa cơ chế RPS và quy trình MMCO/cửa sổ trượt là đối với mỗi lát cụ thể, RPS cung cấp tập hợp hoàn chỉnh của các hình ảnh tham chiếu mà được sử dụng bởi hình ảnh hiện hành hoặc hình ảnh sau đó bất kỳ. Vì vậy, tập hợp hoàn chỉnh của tất cả các hình ảnh mà sẽ được giữ trong DPB để sử dụng bởi hình ảnh hiện hành hoặc tương lai được báo hiệu trong RPS. Điều này khác với sơ đồ MMCO/cửa sổ trượt trong đó chỉ các sự thay đổi tương đối với DPB được báo hiệu. Với cơ chế RPS, không có thông tin nào từ các hình ảnh trước đó theo thứ tự giải mã là cần để duy trì trạng thái đúng của các hình ảnh tham chiếu trong DPB. Thứ tự của các hoạt động giải mã hình ảnh và DPB được thay đổi trong một số hệ thống tạo mã video để lợi dụng các ưu điểm của RPS và nâng cao tính bền vững với lỗi. Trong một số hệ thống tạo mã video, các hoạt động đánh dấu hình ảnh và đệm bao gồm cả kết xuất lẫn loại bỏ các hình ảnh được giải mã khỏi DPB có thể được áp dụng sau khi hình ảnh hiện hành đã được giải mã. Trong các hệ thống tạo mã video khác, RPS trước tiên được giải mã từ phần đầu lát của hình ảnh hiện hành, và sau đó các hoạt động đánh dấu hình ảnh và đệm có thể được áp dụng trước khi giải mã hình ảnh hiện hành.

Trong VVC, các tiếp cận quản lý hình ảnh tham chiếu có thể được tổng kết như sau. Hai danh sách hình ảnh tham chiếu, danh sách (list) 0 và danh sách 1 được biểu thị, được báo hiệu và dẫn xuất trực tiếp. Chúng không dựa trên RPS hoặc cửa sổ trượt cộng quy trình MMCO như được thảo luận trên đây. Việc đánh dấu hình ảnh tham chiếu dựa trực tiếp trên các danh sách hình ảnh tham chiếu 0 và 1 sử dụng cả các mục nhập hoạt động và không hoạt động trong các danh sách hình ảnh tham chiếu, trong khi chỉ các mục nhập hoạt động có thể được sử dụng làm các chỉ số tham chiếu khi dự đoán liên ảnh các CTU. Thông tin để dẫn xuất hai danh sách hình ảnh tham chiếu được báo hiệu bởi các phần tử cú pháp và các cấu trúc cú pháp trong SPS, PPS, và phần đầu lát. Các cấu trúc RPL định trước được báo hiệu trong SPS để sử dụng bằng cách tham chiếu trong phần đầu lát. Hai danh sách hình ảnh tham chiếu được tạo ra cho tất cả các loại lát bao gồm các

lát dự đoán liên ảnh hai hướng (B), các lát dự đoán liên ảnh đơn hướng (P), và các lát dự đoán trong ảnh (I). Hai danh sách hình ảnh tham chiếu có thể được xây dựng mà không sử dụng quy trình khởi tạo danh sách hình ảnh tham chiếu hoặc quy trình thay đổi danh sách hình ảnh tham chiếu. Các hình ảnh tham chiếu dài hạn (LTRP) được nhận dạng bởi các POC LSB. Các chu kỳ POC MSB đenta có thể được báo hiệu đối với các LTRP như được xác định bởi cơ sở hình ảnh cạnh hình ảnh.

Để tạo mã ảnh video, trước tiên, ảnh được phân chia, và các phần chia được tạo mã thành dòng bit. Các sơ đồ phân chia hình ảnh khác nhau là khả dụng. Ví dụ, ảnh có thể được phân chia thành các lát đều, các lát phụ thuộc, các ảnh xếp lớp, và/hoặc theo việc xử lý song song mặt sóng (Wavefront Parallel Processing, WPP). Để đơn giản, HEVC hạn chế các bộ mã hóa sao cho chỉ các lát đều, các lát phụ thuộc, các ảnh xếp lớp, WPP, và các sự kết hợp của chúng có thể được sử dụng khi phân chia lát thành các nhóm của các CTB để tạo mã video. Việc phân chia này có thể được áp dụng để hỗ trợ việc so khớp kích thước đơn vị truyền tối đa (Maximum Transfer Unit, MTU), việc xử lý song song, và trễ đầu-cuối giảm bớt. MTU biểu thị lượng dữ liệu tối đa mà có thể được truyền trong một gói đơn lẻ. Nếu phần tải dữ liệu gói vượt quá MTU, thì phần tải dữ liệu đó được chia tách thành hai gói qua quy trình được gọi là phân mảnh.

Lát đều, cũng được gọi đơn giản là lát, là phần được phân chia của hình ảnh mà có thể được tái tạo một cách độc lập từ các lát đều khác trong cùng một hình ảnh, mặc dù có một số sự phụ thuộc lẫn nhau do các hoạt động lọc vòng. Mỗi lát đều được đóng gói trong chính đơn vị lớp trừu tượng mạng (NAL) của nó để truyền. Hơn nữa, dự đoán trong hình ảnh (dự đoán trong mẫu, dự đoán thông tin chuyển động, dự đoán chế độ tạo mã) và sự phụ thuộc vào mã hóa entropi qua các biên lát có thể không được phép hỗ trợ việc tái tạo độc lập. Việc tái tạo độc lập này hỗ trợ song song hóa. Ví dụ, song song hóa dựa trên lát đều sử dụng hoạt động truyền thông liên bộ xử lý hoặc liên lõi tối thiểu. Tuy nhiên, do mỗi lát đều là độc lập, nên mỗi lát được liên kết với phần đầu lát riêng rẽ. Việc sử dụng các lát đều có thể gặp phải một lượng thông tin thừa tạo mã đáng kể do chi phí bit của phần đầu lát đối với mỗi lát và do thiếu dự đoán qua các biên lát. Hơn nữa, các lát đều có thể được sử dụng để hỗ trợ việc so khớp đối với các yêu cầu kích thước MTU. Cụ thể, do lát đều được đóng gói trong một đơn vị NAL riêng lẻ và có thể được tạo mã một cách độc lập, nên mỗi lát đều sẽ nhỏ hơn MTU trong các sơ đồ MTU để tránh phá vỡ lát thành

nhiều gói. Như vậy, mục tiêu song song hóa và mục tiêu so khớp kích thước MTU có thể đặt ra các nhu cầu trái ngược cho bố cục lát trong hình ảnh.

Các lát phụ thuộc tương tự như các lát đều, nhưng có các phần đầu lát được rút ngắn và cho phép phân chia các biên khối cây ảnh mà không phá vỡ dự đoán trong hình ảnh. Theo đó, các lát phụ thuộc cho phép lát đều được phân đoạn thành nhiều đơn vị NAL, mà cung cấp trễ đầu-cuối giảm bớt bằng cách cho phép một phần của lát đều được gửi ra ngoài trước khi việc mã hóa toàn bộ lát đều hoàn thành.

Các hình ảnh có thể được chia thành các nhóm ảnh xếp lớp/các lát và các ảnh xếp lớp. Ảnh xếp lớp là một chuỗi của các CTU mà bao trùm vùng hình chữ nhật của hình ảnh. Nhóm ảnh xếp lớp/lát chứa số lượng ảnh xếp lớp của hình ảnh. Chế độ nhóm ảnh xếp lớp quét kiểu mảnh và chế độ nhóm ảnh xếp lớp hình chữ nhật có thể được sử dụng để tạo các ảnh xếp lớp. Trong chế độ nhóm ảnh xếp lớp quét kiểu mảnh, nhóm ảnh xếp lớp chứa chuỗi của các ảnh xếp lớp khi quét kiểu mảnh ảnh xếp lớp của hình ảnh. Trong chế độ nhóm ảnh xếp lớp hình chữ nhật, nhóm ảnh xếp lớp chứa một số ảnh xếp lớp của hình ảnh mà tạo chung thành vùng hình chữ nhật của hình ảnh. Các ảnh xếp lớp trong nhóm ảnh xếp lớp hình chữ nhật gần như giống quét kiểu mảnh ảnh xếp lớp của nhóm ảnh xếp lớp. Ví dụ, ảnh xếp lớp có thể là phần được phân chia của ảnh được tạo ra bởi các biên ngang và thẳng đứng mà tạo ra các cột và các hàng của các ảnh xếp lớp. Các ảnh xếp lớp có thể được tạo mã theo thứ tự quét kiểu mảnh (từ phải sang trái và từ trên xuống dưới). Thứ tự quét của các CTB là cục bộ trong ảnh xếp lớp. Theo đó, các CTB trong ảnh xếp lớp thứ nhất được tạo mã theo thứ tự quét kiểu mảnh, trước khi tiến hành các CTB trong ảnh xếp lớp tiếp theo. Tương tự như các lát đều, các ảnh xếp lớp phá vỡ các sự phụ thuộc dự đoán trong hình ảnh cũng như các sự phụ thuộc giải mã entropi. Tuy nhiên, các ảnh xếp lớp có thể không có trong các đơn vị NAL cá nhân, và vì vậy các ảnh xếp lớp không thể được sử dụng để so khớp kích thước MTU. Mỗi ảnh xếp lớp có thể được xử lý bởi một bộ xử lý/lỗi, và hoạt động truyền thông liên bộ xử lý/liên lỗi được sử dụng để dự đoán trong hình ảnh giữa các bộ phận xử lý giải mã các ảnh xếp lớp lân cận có thể được giới hạn để truyền phần đầu lát được chia sẻ (khi các ảnh xếp lớp liên kề ở trong cùng một lát), và thực hiện việc chia sẻ dựa trên lọc vòng các mẫu tái tạo và siêu dữ liệu. Khi nhiều hơn một ảnh xếp lớp có trong lát, độ lệch bai điểm mục nhập đối với mỗi ảnh xếp lớp không phải là độ lệch điểm mục nhập đầu tiên trong lát có thể được báo hiệu trong phần

đầu lát. Đối với mỗi lát và ảnh xếp lớp, ít nhất một trong các điều kiện sau đây nên được đáp ứng: 1) tất cả các khối cây được tạo mã trong lát thuộc về cùng ảnh xếp lớp; và 2) tất cả các khối cây được tạo mã trong ảnh xếp lớp thuộc về cùng lát.

Trong WPP, ảnh được phân chia thành các hàng đơn lẻ của các CTB. Các cơ chế giải mã entropi và dự đoán có thể sử dụng dữ liệu từ các CTB trong các hàng khác. Việc xử lý song song được làm cho khả thi qua giải mã song song các hàng CTB. Ví dụ, hàng hiện hành có thể được giải mã song song với hàng trước đó. Tuy nhiên, việc giải mã hàng hiện hành bị trễ từ quy trình giải mã các hàng trước đó bởi hai CTB. Trễ này đảm bảo rằng dữ liệu liên quan đến CTB bên trên và CTB bên trên và bên phải của CTB hiện hành trong hàng hiện hành là khả dụng trước khi CTB hiện hành được tạo mã. Cách tiếp cận này xuất hiện dưới dạng mặt sóng khi được biểu diễn bằng đồ họa. Sự bắt đầu so le này cho phép việc song song hóa với nhiều bộ xử lý/lỗi như ảnh chứa các hàng CTB. Do dự đoán trong hình ảnh giữa các hàng khối cây lân cận trong hình ảnh được phép, hoạt động truyền thông liên bộ xử lý/liên lỗi để cho phép dự đoán trong hình ảnh có thể là quan trọng. Việc phân chia WPP không xem xét các kích thước đơn vị NAL. Vì vậy, WPP không hỗ trợ so khớp kích thước MTU. Tuy nhiên, các lát đều có thể được sử dụng kết hợp với WPP, với lượng thông tin thừa tạo mã nhất định, để thực hiện so khớp kích thước MTU như mong muốn. Cuối cùng, phân đoạn mặt sóng có thể chứa chính xác một hàng CTB. Hơn nữa, khi sử dụng WPP và khi lát bắt đầu trong một hàng CTB, thì lát sẽ kết thúc trong cùng hàng CTB đó.

Các ảnh xếp lớp cũng có thể bao gồm các tập hợp ảnh xếp lớp ràng buộc chuyển động. Tập hợp ảnh xếp lớp ràng buộc chuyển động (motion constrained tile set, MCTS) là tập hợp ảnh xếp lớp được thiết kế sao cho các vector chuyển động liên kết bị hạn chế trở đến các vị trí mẫu đầy đủ bên trong MCTS và đến các vị trí mẫu thập phân mà đòi hỏi chỉ các vị trí mẫu đầy đủ bên trong MCTS để nội suy. Hơn nữa, việc sử dụng các ứng viên vector chuyển động để dự đoán vector chuyển động thời gian thu được từ các khối bên ngoài MCTS là không được phép. Theo cách này, mỗi MCTS có thể được giải mã một cách độc lập mà không cần có các ảnh xếp lớp trong MCTS. Các tin nhắn thông tin nâng cao phụ (supplemental enhancement information, SEI) của các MCTS thời gian có thể được sử dụng để chỉ báo sự có mặt của các MCTS trong dòng bit và báo hiệu các MCTS. Tin nhắn SEI của các MCTS cung cấp thông tin phụ mà có thể được sử dụng khi trích

xuất dòng bit con MCTS (được quy định là một phần của các ngữ nghĩa của tin nhắn SEI) để tạo dòng bit phù hợp cho tập hợp MCTS. Thông tin này bao gồm số lượng tập hợp thông tin trích xuất, mỗi tập hợp xác định số số lượng tập hợp MCTS và chứa các bai của phần tải dữ liệu chuỗi bai thô (RBSP) của tập hợp thông số video (video parameter set, VPS) thay thế, các tập hợp thông số chuỗi (SPS), và các tập hợp thông số hình ảnh (PPS) được sử dụng trong suốt quá trình trích xuất dòng bit con MCTS. Khi trích xuất dòng bit con theo quy trình trích xuất dòng bit con MCTS, các tập hợp thông số (VPS, SPS, và PPS) có thể được ghi lại hoặc thay thế, và các phần đầu lát có thể được cập nhật do một hoặc tất cả các phần tử cú pháp liên quan đến địa chỉ lát (bao gồm `first_slice_segment_in_pic_flag` và `slice_segment_address`) có thể sử dụng các giá trị khác nhau trong dòng bit con được trích xuất.

Các ứng dụng VR, cũng được gọi là các ứng dụng video 360 độ, có thể hiển thị chỉ phần của hình cầu hoàn chỉnh và do đó chỉ tập con của toàn bộ hình ảnh. Cơ chế truyền phụ thuộc công quan sát 360 nhờ tạo luồng thích ứng động qua giao thức truyền tải siêu văn bản (DASH) có thể được sử dụng để làm giảm tốc độ bit và hỗ trợ truyền video 360 độ qua các cơ chế tạo luồng. Cơ chế này chia hình cầu / hình ảnh được chiếu thành nhiều MCTS, ví dụ bằng cách sử dụng việc chiếu lập bản đồ khối lập phương (cubemap projection, CMP). Hai hoặc nhiều hơn hai dòng bit có thể được mã hóa với các độ phân giải hoặc các chất lượng không gian khác nhau. Khi truyền dữ liệu đến bộ giải mã, (các) MCTS từ dòng bit độ phân giải / chất lượng cao hơn được truyền đối với công quan sát cần được hiển thị (ví dụ, công quan sát trước). Các MCTS từ các dòng bit độ phân giải / chất lượng thấp hơn được truyền đối với các công quan sát khác. Các MCTS này được đóng gói theo cách nhất định và sau đó được gửi đến bộ nhận để được giải mã. Điều mong đợi là công quan sát được nhìn bởi người dùng được biểu diễn bởi MCTS độ phân giải / chất lượng cao để tạo trải nghiệm xem tích cực. Khi đầu của người dùng quay để xem công quan sát khác (ví dụ, công quan sát bên trái hoặc bên phải), nội dung được hiển thị đến từ công quan sát độ phân giải / chất lượng thấp hơn trong một khoảng thời gian ngắn trong khi hệ thống đang tìm các MCTS có độ phân giải cao / chất lượng cao cho công quan sát mới. Khi đầu của người dùng quay để xem công quan sát khác, thì có trễ giữa thời gian của đầu người dùng quay và thời gian khi sự biểu diễn độ phân giải / chất lượng cao hơn của công quan sát được nhìn thấy. Trễ này phụ thuộc vào việc hệ thống có thể

tìm thấy các MCTS có độ phân giải / chất lượng cao hơn cho công quan sát đó nhanh đến mức nào, mà theo đó, phụ thuộc vào chu kỳ IRAP. Chu kỳ IRAP là khoảng thời gian giữa các sự xuất hiện của hai IRAP. Trễ này liên quan đến chu kỳ IRAP do các MCTS của công quan sát mới có thể chỉ giải mã được bắt đầu từ hình ảnh IRAP.

Ví dụ, nếu chu kỳ IRAP được tạo mã từng một giây thì phần sau đây áp dụng. Tình huống tốt nhất đối với trễ giống như trễ khứ hồi của mạng nếu đầu của người dùng quay để xem công quan sát mới ngay trước khi hệ thống bắt đầu tìm phân đoạn / chu kỳ IRAP mới. Trong trường hợp này, hệ thống sẽ có thể yêu cầu các MCTS có độ phân giải/chất lượng cao hơn cho công quan sát mới ngay tức thì, và vì vậy chỉ trễ là trễ khứ hồi của mạng, mà là trễ của yêu cầu tìm cộng thời gian truyền của các MCTS được yêu cầu, giả sử rằng trễ đệm tối thiểu có thể được thiết đặt ở khoảng không (zero) và trễ cảm biến là nhỏ và không đáng kể. Trễ khứ hồi của mạng có thể vào khoảng hai trăm mili giây chẳng hạn. Tình huống xấu nhất đối với trễ là chu kỳ IRAP + trễ khứ hồi của mạng nếu đầu của người dùng quay để xem công quan sát mới ngay sau khi hệ thống đã yêu cầu phân đoạn tiếp theo. Các dòng bit có thể được mã hóa với các hình ảnh IRAP thường xuyên hơn sao cho chu kỳ IRAP là ngắn hơn để cải thiện tình huống xấu nhất trên đây do nó làm giảm toàn bộ trễ. Tuy nhiên, cách tiếp cận này làm tăng các yêu cầu dải thông do hiệu suất nén bị giảm bớt.

Theo một cách thực hiện ví dụ, các hình ảnh con của cùng một hình ảnh được tạo mã được phép chứa các giá trị `nal_unit_type` khác nhau. Cơ chế này được mô tả như sau. Các hình ảnh có thể được chia thành các hình ảnh con. Hình ảnh con là tập hợp hình chữ nhật của các nhóm ảnh xếp lớp/các lát mà bắt đầu với nhóm ảnh xếp lớp mà có `tile_group_address` bằng không (zero). Mỗi hình ảnh con có thể tham chiếu đến PPS tương ứng và vì vậy có thể có sự phân chia ảnh xếp lớp riêng rẽ. Sự có mặt của các hình ảnh con có thể được chỉ báo trong PPS. Mỗi hình ảnh con được xử lý như hình ảnh trong quy trình giải mã. Lọc nội vòng qua các biên hình ảnh con có thể luôn không được phép. Độ rộng và độ cao của hình ảnh con có thể được quy định theo các đơn vị của các kích thước CTU độ chói. Vị trí của hình ảnh con trong hình ảnh có thể không được báo hiệu, nhưng có thể được dẫn xuất nhờ sử dụng quy tắc sau đây. Hình ảnh con lấy vị trí chưa bị chiếm tiếp theo trong thứ tự quét kiểu mảnh CTU trong hình ảnh mà đủ lớn để chứa hình ảnh con trong các biên hình ảnh. Các hình ảnh tham chiếu để giải mã mỗi hình ảnh con được tạo

ra bằng cách trích xuất khu vực cùng chỗ với hình ảnh con hiện hành từ các hình ảnh tham chiếu trong bộ đệm hình ảnh được giải mã. Khu vực được trích xuất là hình ảnh con được giải mã, và vì vậy việc dự đoán liên ảnh xảy ra giữa các hình ảnh con có cùng kích thước và cùng vị trí trong hình ảnh. Trong trường hợp như vậy, việc cho phép các giá trị `nal_unit_type` khác nhau trong hình ảnh được tạo mã cho phép các hình ảnh con bắt nguồn từ hình ảnh truy cập ngẫu nhiên và các hình ảnh con bắt nguồn từ hình ảnh truy cập phi ngẫu nhiên được hợp nhất vào cùng một hình ảnh được tạo mã mà không gặp khó khăn đáng kể (ví dụ, không thay đổi mức VCL). Lợi ích như vậy cũng dành cho việc tạo mã dựa trên MCTS.

Việc cho phép các giá trị `different nal_unit_type` trong hình ảnh được tạo mã có thể có lợi ích trong các tình huống khác. Ví dụ, người dùng có thể xem một số khu vực của nội dung video 360 độ thường xuyên hơn các khu vực khác. Để tạo sự cân bằng tốt hơn giữa hiệu suất tạo mã và độ trễ chuyển mạch công quan sát chất lượng so sánh được trung bình trong video 360 độ phụ thuộc công quan sát dựa trên MCTS/hình ảnh con truyền các hình ảnh IRAP thường xuyên hơn có thể được tạo mã đối với các khu vực được xem chung nhiều hơn so với các khu vực khác. Độ trễ chuyển mạch công quan sát chất lượng so sánh được là độ trễ được trải nghiệm bởi người dùng khi chuyển từ công quan sát thứ nhất sang công quan sát thứ hai cho đến khi chất lượng trình diễn của công quan sát thứ hai đạt đến chất lượng trình diễn so sánh được với công quan sát thứ nhất.

Một cách thực hiện khác sử dụng các giải pháp sau đây để hỗ trợ các loại đơn vị NAL trộn trong hình ảnh bao gồm dẫn xuất POC và quản lý hình ảnh tham chiếu. Cờ (`sps_mixed_tile_groups_in_pic_flag`) có mặt trong tập hợp thông số mà được tham chiếu đến một cách trực tiếp hoặc gián tiếp bởi các nhóm ảnh xếp lớp để quy định liệu có thể có các hình ảnh với IRAP trộn và các hình ảnh con phi IRAP hay không. Đối với đơn vị NAL chứa nhóm ảnh xếp lớp IDR, cờ (`poc_msb_reset_flag`) có mặt trong phần đầu nhóm ảnh xếp lớp tương ứng để quy định liệu POC MSB có được thiết đặt lại trong việc dẫn xuất POC đối với hình ảnh hay không. Biến số được gọi là `PicRefreshFlag` được xác định và được liên kết với hình ảnh. Cờ này quy định liệu việc dẫn xuất POC và trạng thái DPB có nên được làm mới khi giải mã hình ảnh hay không. Giá trị của `PicRefreshFlag` được dẫn xuất như sau. Nếu nhóm ảnh xếp lớp hiện hành có trong đơn vị truy cập thứ nhất trong dòng bit, thì `PicRefreshFlag` được thiết đặt bằng một. Còn không nếu nhóm ảnh xếp lớp

hiện hành là nhóm ảnh xếp lớp IDR, thì PicRefreshFlag được thiết đặt bằng `sps_mixed_tile_groups_in_pic_flag ? poc_msb_reset_flag : 1`. Còn không nếu nhóm ảnh xếp lớp hiện hành là nhóm ảnh xếp lớp CRA, thì phần sau đây áp dụng. Nếu đơn vị truy cập hiện hành là đơn vị truy cập thứ nhất của chuỗi được tạo mã, thì PicRefreshFlag được thiết đặt bằng một. Đơn vị truy cập hiện hành là đơn vị truy cập thứ nhất của chuỗi được tạo mã khi đơn vị truy cập theo ngay sau phần cuối của đơn vị NAL chuỗi hoặc biến số liên kết `HandleCraAsFirstPicInCvsFlag` được thiết đặt bằng một. Nếu không, PicRefreshFlag được thiết đặt bằng không (zero) (ví dụ, nhóm ảnh xếp lớp hiện hành không thuộc về đơn vị truy cập thứ nhất trong dòng bit và không phải là nhóm ảnh xếp lớp IRAP).

Khi PicRefreshFlag bằng một, giá trị của POC MSB (`PicOrderCntMsb`) được thiết đặt lại về bằng không (zero) trong suốt quá trình dẫn xuất POC cho hình ảnh. Thông tin được sử dụng để quản lý hình ảnh tham chiếu như tập hợp hình ảnh tham chiếu (RPS) hoặc danh sách hình ảnh tham chiếu (RPL) được báo hiệu trong nhóm ảnh xếp lớp/phần đầu lát mà không quan tâm đến loại đơn vị NAL tương ứng. Các danh sách hình ảnh tham chiếu được xây dựng khi bắt đầu giải mã mỗi nhóm ảnh xếp lớp mà không quan tâm đến loại đơn vị NAL. Các danh sách hình ảnh tham chiếu có thể bao gồm `RefPicList[ 0 ]` và `RefPicList[ 1 ]` đối cách tiếp cận RPL, `RefPicList0[ ]` và `RefPicList1[ ]` đối với cách tiếp cận RPS, hoặc các danh sách tương tự chứa các hình ảnh tham chiếu dùng cho các hoạt động dự đoán liên ảnh đối với hình ảnh. Khi PicRefreshFlag bằng một, trong suốt quy trình đánh dấu hình ảnh tham chiếu, tất cả các hình ảnh tham chiếu trong DPB được đánh dấu là không được sử dụng để tham chiếu.

Các cách thực hiện này được liên kết với các vấn đề nhất định. Ví dụ, khi trộn các giá trị `nal_unit_type` trong hình ảnh không được phép và khi dẫn xuất việc liệu hình ảnh là hình ảnh IRAP và việc dẫn xuất biến số `NoRaslOutputFlag` được mô tả ở mức hình ảnh, bộ giải mã có thể thực hiện các sự dẫn xuất này sau khi nhận đơn vị NAL VCL thứ nhất của hình ảnh bất kỳ. Tuy nhiên, do hỗ trợ các loại đơn vị NAL trộn trong hình ảnh, bộ giải mã sẽ phải chờ các đơn vị NAL VCL khác của hình ảnh đến trước khi thực hiện các sự dẫn xuất trên đây. Trong trường hợp xấu nhất, bộ giải mã sẽ phải chờ đơn vị NAL VCL cuối cùng của hình ảnh đến. Hơn nữa, các hệ thống này có thể báo hiệu cờ trong các phần đầu nhóm ảnh xếp lớp của các đơn vị NAL IDR để quy định liệu POC MSB có được thiết

đặt lại khi dẫn xuất POC đối với hình ảnh hay không. Cơ chế này có các vấn đề sau đây. Trường hợp các loại đơn vị NAL CRA và các loại đơn vị NAL phi IRAP được trộn sẽ không được hỗ trợ bởi cơ chế này. Hơn nữa, việc báo hiệu thông tin này trong nhóm ảnh xếp lớp/phần đầu lát của đơn vị NAL VCL sẽ đòi hỏi giá trị được thay đổi trong suốt quá trình trích xuất hoặc hợp nhất dòng bit khi sự thay đổi đối với trạng thái của việc liệu các đơn vị NAL IRAP (IDR hoặc CRA) có được trộn với các đơn vị NAL phi IRAP trong hình ảnh hay không. Cách ghi lại như vậy của các phần đầu lát sẽ xảy ra bất cứ khi nào người dùng yêu cầu video, và vì vậy đòi hỏi các tài nguyên phần cứng đáng kể. Hơn nữa, một số cách trộn khác của các loại đơn vị NAL khác nhau trong hình ảnh không phải là việc trộn loại đơn vị NAL IRAP cụ thể và loại đơn vị NAL phi IRAP cụ thể là được phép. Sự linh hoạt như vậy không mang lại sự hỗ trợ cho các trường hợp sử dụng thực tế trong khi chúng làm phức tạp thiết kế của codec, mà làm tăng một cách không cần thiết độ phức tạp của bộ giải mã và vì vậy làm tăng các chi phí thực hiện được liên kết.

Nói chung, sáng chế mô tả các kỹ thuật để hỗ trợ truy cập ngẫu nhiên dựa trên MCTS hoặc hình ảnh con khi tạo mã video. Cụ thể hơn, sáng chế mô tả các thiết kế cải tiến để hỗ trợ các loại đơn vị NAL trộn trong hình ảnh, mà được sử dụng để hỗ trợ truy cập ngẫu nhiên dựa trên MCTS hoặc hình ảnh con. Phần mô tả của các kỹ thuật này dựa trên tiêu chuẩn VVC, mà còn áp dụng cho các đặc tả codec video/phương tiện khác.

Để giải quyết các vấn đề trên đây, các cách thực hiện làm ví dụ sau đây được bộc lộ. Các cách thực hiện này có thể được áp dụng một cách riêng lẻ hoặc kết hợp. Trong một ví dụ, mỗi hình ảnh được liên kết với chỉ báo về việc liệu hình ảnh có chứa các giá trị `nal_unit_type` trộn hay không. Chỉ báo này được báo hiệu trong PPS. Chỉ báo này hỗ trợ việc xác định liệu có thiết đặt lại POC MSB và/hoặc thiết đặt lại DPB hay không bằng cách đánh dấu tất cả các hình ảnh tham chiếu là không được sử dụng để tham chiếu. Khi chỉ báo này được báo hiệu trong PPS, thay đổi về giá trị trong PPS có thể được tiến hành trong suốt quá trình hợp nhất hoặc trích xuất riêng rẽ. Tuy nhiên, điều này có thể chấp nhận được do các PPS được ghi và được thay thế bởi các cơ chế khác trong suốt quá trình trích xuất hoặc hợp nhất dòng bit đó.

Theo cách khác, chỉ báo này có thể được báo hiệu trong phần đầu nhóm ảnh xếp lớp nhưng được đòi hỏi phải giống nhau đối với tất cả các nhóm ảnh xếp lớp của hình

ảnh. Tuy nhiên, trong trường hợp này, giá trị có thể cần được thay đổi trong suốt quá trình trích xuất dòng bit con của các MCTS/các chuỗi hình ảnh con. Theo cách khác, chỉ báo này có thể được báo hiệu trong phần đầu đơn vị NAL nhưng được đòi hỏi phải giống nhau đối với tất cả các nhóm ảnh xếp lớp của hình ảnh. Tuy nhiên, trong trường hợp này, giá trị có thể cần được thay đổi trong suốt quá trình trích xuất dòng bit con của các MCTS/các chuỗi hình ảnh con. Theo cách khác, chỉ báo này có thể được báo hiệu bằng cách xác định các loại đơn vị NAL VCL bổ sung này mà, khi được sử dụng cho hình ảnh, tất cả các đơn vị NAL VCL của hình ảnh sẽ có cùng giá trị loại đơn vị NAL. Tuy nhiên, trong trường hợp này, giá trị loại đơn vị NAL của các đơn vị NAL VCL có thể cần được thay đổi trong suốt quá trình trích xuất dòng bit con của các MCTS/các chuỗi hình ảnh con. Theo cách khác, chỉ báo này có thể được báo hiệu bằng cách xác định các loại đơn vị NAL VCL IRAP bổ sung này mà, khi được sử dụng cho hình ảnh, tất cả các đơn vị NAL VCL của hình ảnh sẽ có cùng giá trị loại đơn vị NAL. Tuy nhiên, trong trường hợp này, giá trị loại đơn vị NAL của các đơn vị NAL VCL có thể cần được thay đổi trong suốt quá trình trích xuất dòng bit con của các MCTS/các chuỗi hình ảnh con. Theo cách khác, mỗi hình ảnh mà có ít nhất một đơn vị NAL VCL với loại bất kỳ trong số các loại đơn vị NAL IRAP có thể được liên kết với chỉ báo về việc liệu hình ảnh có chứa các giá trị loại đơn vị NAL trộn hay không.

Hơn nữa, ràng buộc có thể được áp dụng sao cho việc trộn các giá trị `nal_unit_type` trong hình ảnh là được phép theo kiểu có giới hạn bằng cách chỉ cho phép các loại đơn vị NAL IRAP và phi IRAP trộn lẫn. Đối với hình ảnh cụ thể bất kỳ, hoặc tất cả các đơn vị NAL VCL có cùng loại đơn vị NAL hoặc một số đơn vị NAL VCL có loại đơn vị NAL IRAP cụ thể và số còn lại có loại đơn vị NAL VCL phi IRAP cụ thể. Nói cách khác, các đơn vị NAL VCL của hình ảnh cụ thể bất kỳ không thể có nhiều hơn một loại đơn vị NAL IRAP và không thể có nhiều hơn một loại đơn vị NAL phi IRAP. Hình ảnh có thể được coi là hình ảnh IRAP chỉ khi hình ảnh không chứa các giá trị `nal_unit_type` trộn và các đơn vị NAL VCL có loại đơn vị NAL IRAP. Đối với đơn vị NAL IRAP bất kỳ (bao gồm IDR) mà không thuộc về hình ảnh IRAP, POC MSB có thể không được thiết đặt lại. Đối với đơn vị NAL IRAP bất kỳ (bao gồm IDR) mà không thuộc về hình ảnh IRAP, DPB không được thiết đặt lại, và vì vậy việc đánh dấu tất cả các hình ảnh tham chiếu là không được sử dụng để tham chiếu không được thực hiện. `TemporalId` có thể được thiết đặt bằng

không (zero) đối với hình ảnh nếu ít nhất một đơn vị NAL VCL của hình ảnh là đơn vị NAL IRAP.

Phần sau đây là cách thực hiện cụ thể của một hoặc nhiều khía cạnh được mô tả trên đây. Hình ảnh IRAP có thể được xác định là hình ảnh được tạo mã mà giá trị của `mixed_nalu_types_in_pic_flag` bằng không (zero) và mỗi đơn vị NAL VCL có `nal_unit_type` nằm trong đoạn từ `IDR_W_RADL` đến `RSV_IRAP_VCL13`. Cú pháp PPS và các ngữ nghĩa làm ví dụ như sau.

<code>pic_parameter_set_rbsp() {</code>	Bộ mô tả
<code>pps_pic_parameter_set_id</code>	<code>ue(v)</code>
<code>pps_seq_parameter_set_id</code>	<code>ue(v)</code>
<code>mixed_nalu_types_in_pic_flag</code>	<code>ue(v)</code>
<code>single_tile_in_pic_flag</code>	<code>u(1)</code>
<code>...</code>	

`mixed_nalu_types_in_pic_flag` được thiết đặt bằng không (zero) để quy định rằng mỗi hình ảnh tham chiếu đến PPS có nhiều đơn vị NAL VCL và các đơn vị NAL này không có cùng giá trị của `nal_unit_type`. `mixed_nalu_types_in_pic_flag` được thiết đặt bằng không (zero) để quy định rằng các đơn vị NAL VCL của mỗi hình ảnh tham chiếu đến PPS có cùng giá trị của `nal_unit_type`.

Cú pháp nhóm ảnh xếp lớp/phần đầu lát làm ví dụ như sau.

<code>tile_group_header() {</code>	Bộ mô tả
<code>tile_group_pic_parameter_set_id</code>	<code>ue(v)</code>
<code>if(rect_tile_group_flag NumTilesInPic > 1)</code>	
<code>tile_group_address</code>	<code>u(v)</code>
<code>if(!rect_tile_group_flag && !single_tile_per_tile_group_flag)</code>	
<code>num_tiles_in_tile_group_minus1</code>	<code>ue(v)</code>

tile_group_type	ue(v)
tile_group_pic_order_cnt_lsb	u(v)
for(i = 0; i < 2; i++) {	
if(num_ref_pic_lists_in_sps[i] > 0 && (i == 0 (i == 1 && rpl1_idx_present_flag)))	
ref_pic_list_sps_flag[i]	u(1)
if(ref_pic_list_sps_flag[i]) {	
if(num_ref_pic_lists_in_sps[i] > 1 && (i == 0 (i == 1 && rpl1_idx_present_flag)))	
ref_pic_list_idx[i]	u(v)
} else	
ref_pic_list_struct(i, num_ref_pic_lists_in_sps[i])	
for(j = 0; j < NumLtrpEntries[i][RplsIdx[i]]; j++) {	
delta_poc_msb_present_flag[i][j]	u(1)
if(delta_poc_msb_present_flag[i][j])	
delta_poc_msb_cycle_lt[i][j]	ue(v)
}	
}	
if(tile_group_type == P tile_group_type == B) {	
num_ref_idx_active_override_flag	u(1)
if(num_ref_idx_active_override_flag)	
for(i = 0; i < (tile_group_type == B ? 2 : 1); i++)	
if(num_ref_entries[i][RplsIdx[i]] > 1)	
num_ref_idx_active_minus1[i]	ue(v)
}	

...	
-----	--

Các ngữ nghĩa phần đầu đơn vị NAL làm ví dụ như sau. Đối với các đơn vị NAL VCL của hình ảnh cụ thể bất kỳ, một trong số hai điều kiện sau đây sẽ được thỏa mãn. Tất cả các đơn vị NAL VCL có cùng giá trị của `nal_unit_type`. Một số đơn vị NAL VCL có giá trị loại đơn vị NAL IRAP cụ thể (tức là, giá trị của `nal_unit_type` nằm trong đoạn từ `IDR_W_RADL` đến `RSV_IRAP_VCL13`), trong khi tất cả các đơn vị NAL VCL khác có loại đơn vị NAL VCL phi IRAP cụ thể (tức là, giá trị của `nal_unit_type` nằm trong đoạn từ `TRAIL_NUT` đến `RSV_VCL_7`, hoặc nằm trong đoạn từ `RSV_VCL14` đến `RSV_VCL15`). `nuh_temporal_id_plus1` trừ 1 quy định bộ nhận dạng thời gian đối với đơn vị NAL. Giá trị của `nuh_temporal_id_plus1` sẽ không bằng không (zero).

Biến số `TemporalId` được dẫn xuất như sau:

$$\text{TemporalId} = \text{nuh_temporal_id_plus1} - 1 \quad (7-1)$$

Khi `nal_unit_type` nằm trong đoạn từ `IDR_W_RADL` đến `RSV_IRAP_VCL13`, đối với đơn vị NAL VCL của hình ảnh, mà không quan tâm đến giá trị `nal_unit_type` của các đơn vị NAL VCL khác của hình ảnh, `TemporalId` sẽ bằng không (zero) đối với tất cả các đơn vị NAL VCL của hình ảnh. Giá trị của `TemporalId` sẽ là giống nhau đối với tất cả các đơn vị NAL VCL của đơn vị truy cập. Giá trị của `TemporalId` của hình ảnh được tạo mã hoặc đơn vị truy cập là giá trị của `TemporalId` của các đơn vị NAL VCL của hình ảnh được tạo mã hoặc đơn vị truy cập.

Quy trình giải mã làm ví dụ đối với hình ảnh được tạo mã như sau. Quy trình giải mã hoạt động như sau đối với hình ảnh hiện hành `CurrPic`. Việc giải mã các đơn vị NAL được quy định ở đây. Các quy trình giải mã sau đây sử dụng các phần tử cú pháp trong lớp phần đầu nhóm ảnh xếp lớp và phía trên. Các biến số và các hàm số liên quan đến việc đếm thứ tự hình ảnh được dẫn xuất như được quy định ở đây. Điều này được viện dẫn chỉ đối với nhóm ảnh xếp lớp/lát thứ nhất của hình ảnh. Ở đầu quy trình giải mã đối với mỗi nhóm ảnh xếp lớp/lát, quy trình giải mã để xây dựng các danh sách hình ảnh tham chiếu được lấy để dẫn xuất danh sách hình ảnh tham chiếu 0 (`RefPicList[ 0 ]`) và danh sách hình ảnh tham chiếu 1 (`RefPicList[ 1 ]`). Nếu hình ảnh hiện hành là hình ảnh IDR,

thì quy trình giải mã để xây dựng các danh sách hình ảnh tham chiếu có thể được lấy nhằm mục đích kiểm tra sự phù hợp dòng bit, nhưng có thể không cần để giải mã hình ảnh hiện hành hoặc các hình ảnh theo sau hình ảnh hiện hành theo thứ tự giải mã.

Quy trình giải mã để xây dựng các danh sách hình ảnh tham chiếu như sau. Quy trình này được lấy ở đầu quy trình giải mã đối với mỗi nhóm ảnh xếp lớp. Các hình ảnh tham chiếu được đánh địa chỉ qua các chỉ số tham chiếu. Chỉ số tham chiếu là chỉ số vào danh sách hình ảnh tham chiếu. Khi giải mã nhóm ảnh xếp lớp I, không có danh sách hình ảnh tham chiếu nào được sử dụng khi giải mã dữ liệu nhóm ảnh xếp lớp. Khi giải mã nhóm ảnh xếp lớp P, chỉ danh sách hình ảnh tham chiếu không (zero) (`RefPicList[ 0 ]`) được sử dụng khi giải mã dữ liệu nhóm ảnh xếp lớp. Khi giải mã nhóm ảnh xếp lớp B, cả danh sách hình ảnh tham chiếu không (zero) và danh sách hình ảnh tham chiếu một (`RefPicList[ 1 ]`) được sử dụng khi giải mã dữ liệu nhóm ảnh xếp lớp. Ở đầu quy trình giải mã đối với mỗi nhóm ảnh xếp lớp, các danh sách hình ảnh tham chiếu `RefPicList[ 0 ]` và `RefPicList[ 1 ]` được dẫn xuất. Các danh sách hình ảnh tham chiếu được sử dụng khi đánh dấu các hình ảnh tham chiếu hoặc khi giải mã dữ liệu nhóm ảnh xếp lớp. Đối với nhóm ảnh xếp lớp bất kỳ của hình ảnh IDR hoặc nhóm ảnh xếp lớp I của hình ảnh phi IDR, `RefPicList[ 0 ]` và `RefPicList[ 1 ]` có thể được dẫn xuất nhằm mục đích kiểm tra sự phù hợp dòng bit, nhưng sự dẫn xuất của chúng không cần để giải mã hình ảnh hiện hành hoặc hình ảnh theo sau hình ảnh hiện hành theo thứ tự giải mã. Đối với nhóm ảnh xếp lớp P, `RefPicList[ 1 ]` có thể được dẫn xuất nhằm mục đích kiểm tra sự phù hợp dòng bit, nhưng sự dẫn xuất không cần để giải mã hình ảnh hiện hành hoặc các hình ảnh theo sau hình ảnh hiện hành theo thứ tự giải mã.

Fig.8 là hình vẽ giản lược của thiết bị tạo mã video 800 làm ví dụ. Thiết bị tạo mã video 800 thích hợp để thực hiện các ví dụ/các phương án được bộc lộ như được mô tả ở đây. Thiết bị tạo mã video 800 bao gồm các cổng xuôi dòng 820, các cổng ngược dòng 850, và/hoặc các bộ phận thu-phát (Tx/Rx) 810, bao gồm các bộ truyền và/hoặc các bộ nhận để truyền thông dữ liệu ngược dòng và/hoặc xuôi dòng qua mạng. Thiết bị tạo mã video 800 cũng bao gồm bộ xử lý 830 bao gồm bộ phận logic và/hoặc bộ xử lý trung tâm (CPU) để xử lý dữ liệu và bộ nhớ 832 để lưu trữ dữ liệu. Thiết bị tạo mã video 800 cũng có thể bao gồm các thành phần điện, quang sang điện (optical-to-electrical, OE), các thành phần điện sang quang (electrical-to-optical, EO), và/hoặc các thành phần truyền thông vô

tuyến được ghép nối với các cổng ngược dòng 850 và/hoặc các cổng xuôi dòng 820 để truyền thông dữ liệu qua các mạng truyền thông điện, quang, hoặc vô tuyến. Thiết bị tạo mã video 800 cũng có thể bao gồm các thiết bị đầu vào và/hoặc đầu ra (I/O) 860 để truyền thông dữ liệu đến và từ người dùng. Các thiết bị I/O 860 có thể bao gồm thiết bị đầu ra như bộ hiển thị để hiển thị dữ liệu video, loa để kết xuất dữ liệu audio, v.v.. Các thiết bị I/O 860 cũng có thể bao gồm các thiết bị đầu vào, như bàn phím, chuột, bi xoay, v.v., và/hoặc các giao diện tương ứng để tương tác với thiết bị đầu ra đó.

Bộ xử lý 830 được thực thi bởi phần cứng và phần mềm. Bộ xử lý 830 có thể được thực thi như là một hoặc nhiều chip CPU, nhân (ví dụ, bộ xử lý nhiều nhân), mảng cổng lập trình được dạng trường (field-programmable gate array, FPGA), mạch tích hợp chuyên dụng (application specific integrated circuit, ASIC), và bộ xử lý tín hiệu số (digital signal processor, DSP). Bộ xử lý 830 truyền thông với các cổng xuôi dòng 820, Tx/Rx 810, các cổng ngược dòng 850, và bộ nhớ 832. Bộ xử lý 830 bao gồm môđun tạo mã 814. Môđun tạo mã 814 thực thi các phương án được bộc lộ được mô tả ở đây, như các phương pháp 100, 900, và 1000, mà có thể sử dụng CVS 500, luồng video hình ảnh VR 600, và/hoặc dòng bit 700. Môđun tạo mã 814 cũng có thể thực thi phương pháp/cơ chế khác bất kỳ được mô tả ở đây. Hơn nữa, môđun tạo mã 814 có thể thực thi hệ thống codec 200, bộ mã hóa 300, và/hoặc bộ giải mã 400. Ví dụ, môđun tạo mã 814 có thể thiết đặt cờ trong PPS để chỉ báo khi hình ảnh chứa cả các đơn vị NAL IRAP lẫn phi IRAP và hạn chế các hình ảnh này để chứa chỉ một loại đơn vị NAL IRAP và một loại đơn vị NAL phi IRAP. Vì vậy, môđun tạo mã 814 làm cho thiết bị tạo mã video 800 cung cấp chức năng và/hoặc hiệu suất tạo mã bổ sung khi tạo mã dữ liệu video. Như vậy, môđun tạo mã 814 cải thiện chức năng của thiết bị tạo mã video 800 cũng như các địa chỉ các vấn đề mà là đặc trưng cho các kỹ thuật tạo mã video. Hơn nữa, môđun tạo mã 814 thực hiện sự biến đổi thiết bị tạo mã video 800 sang trạng thái khác. Theo cách khác, môđun tạo mã 814 có thể được thực thi dưới dạng các lệnh được lưu trữ trong bộ nhớ 832 và được thực hiện bởi bộ xử lý 830 (ví dụ, dưới dạng sản phẩm chương trình máy tính được lưu trữ trên phương tiện lưu trữ).

Bộ nhớ 832 bao gồm một hoặc nhiều loại bộ nhớ như các đĩa, các ổ băng, các ổ trạng thái rắn, bộ nhớ chỉ đọc (ROM), bộ nhớ truy cập ngẫu nhiên (RAM), bộ nhớ tốc độ nhanh (flash memory), bộ nhớ khả lập địa chỉ nội dung bậc ba (ternary content-

addressable memory, TCAM), bộ nhớ truy cập ngẫu nhiên tĩnh (SRAM), v.v.. Bộ nhớ 832 có thể được sử dụng làm thiết bị lưu trữ dữ liệu tràn, để lưu trữ các chương trình khi các chương trình này được chọn để thực hiện, và để lưu trữ các lệnh và dữ liệu được đọc trong suốt quá trình thực hiện chương trình.

Fig.9 là lưu đồ của phương pháp ví dụ 900 để mã hóa chuỗi video, như CVS 500, chứa hình ảnh với các loại đơn vị NAL trộn thành dòng bit, như dòng bit 700 bao gồm luồng video hình ảnh VR 600 được hợp nhất từ nhiều luồng video hình ảnh con 601-603 ở nhiều độ phân giải video. Phương pháp 900 có thể được sử dụng bởi bộ mã hóa, như hệ thống codec 200, bộ mã hóa 300, và/hoặc thiết bị tạo mã video 800 khi thực hiện phương pháp 100.

Phương pháp 900 có thể bắt đầu khi bộ mã hóa nhận chuỗi video bao gồm các hình ảnh, như các hình ảnh VR, và xác định để mã hóa chuỗi video đó thành dòng bit, ví dụ dựa trên đầu vào người dùng. Ở bước 901, bộ mã hóa xác định liệu hình ảnh hiện hành có chứa các hình ảnh con thuộc các loại khác nhau hay không. Các loại này có thể bao gồm ít nhất một lát của hình ảnh chứa một phần của hình ảnh con IRAP và ít nhất một lát của hình ảnh chứa một phần của hình ảnh con NAL phi IRAP. Ở bước 903, bộ mã hóa mã hóa các lát của các hình ảnh con của hình ảnh thành các đơn vị NAL VCL trong dòng bit. Các đơn vị NAL VCL này có thể bao gồm một hoặc nhiều đơn vị NAL IRAP và một hoặc nhiều đơn vị NAL phi IRAP. Ví dụ, bước mã hóa có thể bao gồm việc hợp nhất các dòng bit con có các độ phân giải khác nhau thành một dòng bit để truyền thông đến bộ giải mã.

Ở bước 905, bộ mã hóa mã hóa PPS thành dòng bit và mã hóa cờ thành PPS trong dòng bit. Như một ví dụ cụ thể, việc mã hóa PPS có thể bao gồm việc thay đổi PPS được mã hóa trước đó để bao gồm giá trị cờ, ví dụ đáp lại việc hợp nhất các dòng bit con. Cờ có thể được thiết đặt ở giá trị thứ nhất khi giá trị loại đơn vị NAL là giống nhau đối với tất cả các đơn vị NAL VCL được liên kết với hình ảnh. Cờ cũng có thể được thiết đặt ở giá trị thứ hai khi giá trị loại đơn vị NAL thứ nhất đối với các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh khác với giá trị loại đơn vị NAL thứ hai đối với các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh. Ví dụ, giá trị loại đơn vị NAL thứ nhất có thể chỉ báo hình ảnh chứa hình ảnh con IRAP, và giá trị loại đơn vị NAL thứ hai có thể chỉ báo rằng hình ảnh cũng chứa hình ảnh con phi IRAP. Hơn nữa,

giá trị loại đơn vị NAL thứ nhất có thể bằng một trong số IDR_W_RADL, IDR_N_LP, hoặc CRA_NUT. Ngoài ra, giá trị loại đơn vị NAL thứ hai có thể bằng một trong số TRAIL_NUT, RADL_NUT, hoặc RASL_NUT. Như một ví dụ cụ thể, cờ có thể là `mixed_nalu_types_in_pic_flag`. Theo một ví dụ cụ thể, `mixed_nalu_types_in_pic_flag` có thể được thiết đặt bằng một để quy định rằng mỗi hình ảnh mà liên quan đến PPS chứa cờ có nhiều hơn một đơn vị NAL VCL. Hơn nữa, cờ quy định rằng các đơn vị NAL VCL được liên kết với hình ảnh tương ứng không phải tất cả đều có cùng giá trị của loại đơn vị NAL (`nal_unit_type`). Theo một ví dụ cụ thể khác, `mixed_nalu_types_in_pic_flag` có thể được thiết đặt bằng không (zero) để quy định rằng mỗi hình ảnh mà liên quan đến PPS chứa cờ có một hoặc nhiều đơn vị NAL VCL. Hơn nữa, cờ quy định rằng tất cả các đơn vị NAL VCL của hình ảnh tương ứng có cùng giá trị của `nal_unit_type`.

Ở bước 907, bộ mã hóa có thể lưu trữ dòng bit để truyền thông đến bộ giải mã.

Fig.10 là lưu đồ của phương pháp ví dụ 1000 để giải mã chuỗi video, như CVS 500, chứa hình ảnh với các loại đơn vị NAL trộn từ dòng bit, như dòng bit 700 bao gồm luồng video hình ảnh VR 600 được hợp nhất từ nhiều luồng video hình ảnh con 601-603 ở nhiều độ phân giải video. Phương pháp 1000 có thể được sử dụng bởi bộ giải mã, như hệ thống codec 200, bộ giải mã 400, và/hoặc thiết bị tạo mã video 800 khi thực hiện phương pháp 100.

Phương pháp 1000 có thể bắt đầu khi bộ giải mã bắt đầu nhận dòng bit của dữ liệu được tạo mã đại diện cho chuỗi video, ví dụ là kết quả của phương pháp 900. Ở bước 1001, bộ giải mã nhận dòng bit. Dòng bit bao gồm các hình ảnh con được liên kết với hình ảnh và cờ. Như một ví dụ cụ thể, dòng bit có thể bao gồm PPS mà bao gồm cờ. Hơn nữa, các hình ảnh con có trong các đơn vị NAL VCL. Ví dụ, các lát được liên kết với các hình ảnh con có trong các đơn vị NAL VCL.

Ở bước 1003, bộ giải mã xác định giá trị loại đơn vị NAL là giống nhau đối với tất cả các đơn vị NAL VCL được liên kết với hình ảnh khi cờ được thiết đặt ở giá trị thứ nhất. Hơn nữa, bộ giải mã xác định giá trị loại đơn vị NAL thứ nhất đối với các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh khác với giá trị loại đơn vị NAL thứ hai đối với các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh khi cờ được thiết đặt ở giá trị thứ hai. Ví dụ, giá trị loại đơn vị NAL thứ nhất có thể

chỉ báo hình ảnh chứa hình ảnh con IRAP, và giá trị loại đơn vị NAL thứ hai có thể chỉ báo rằng hình ảnh cũng chứa hình ảnh con phi IRAP. Hơn nữa, giá trị loại đơn vị NAL thứ nhất có thể bằng một trong số IDR_W_RADL, IDR_N_LP, hoặc CRA_NUT. Ngoài ra, giá trị loại đơn vị NAL thứ hai có thể bằng một trong số TRAIL_NUT, RADL_NUT, hoặc RASL_NUT. Như một ví dụ cụ thể, cờ có thể là `mixed_nalu_types_in_pic_flag`. `mixed_nalu_types_in_pic_flag` có thể được thiết đặt bằng một khi quy định rằng mỗi hình ảnh tham chiếu đến PPS có nhiều hơn một đơn vị NAL VCL và các đơn vị NAL VCL không có cùng giá trị của loại đơn vị NAL (`nal_unit_type`). Ngoài ra, `mixed_nalu_types_in_pic_flag` có thể được thiết đặt bằng không (zero) khi mỗi hình ảnh tham chiếu đến PPS có một hoặc nhiều đơn vị NAL VCL và các đơn vị NAL VCL của mỗi hình ảnh tham chiếu đến PPS có cùng giá trị của `nal_unit_type`.

Ở bước 1005, bộ giải mã có thể giải mã một hoặc nhiều hình ảnh con dựa trên các giá trị loại đơn vị NAL. Bộ giải mã cũng có thể chuyển tiếp một hoặc nhiều hình ảnh con để hiển thị dưới dạng một phần của chuỗi video được giải mã ở bước 1007.

Fig.11 là hình vẽ giản lược của hệ thống ví dụ 1100 để tạo mã chuỗi video, như CVS 500, chứa hình ảnh với các loại đơn vị NAL trộn thành dòng bit, như dòng bit 700 bao gồm luồng video hình ảnh VR 600 được hợp nhất từ nhiều luồng video hình ảnh con 601-603 ở nhiều độ phân giải video. Hệ thống 1100 có thể được thực thi bởi bộ mã hóa và bộ giải mã như hệ thống codec 200, bộ mã hóa 300, bộ giải mã 400, và/hoặc thiết bị tạo mã video 800. Hơn nữa, hệ thống 1100 có thể được sử dụng khi thực thi phương pháp 100, 900, và/hoặc 1000.

Hệ thống 1100 bao gồm bộ mã hóa video 1102. Bộ mã hóa video 1102 bao gồm môđun xác định 1101 để xác định liệu hình ảnh có chứa các hình ảnh con thuộc các loại khác nhau hay không. Bộ mã hóa video 1102 còn bao gồm môđun mã hóa 1103 để mã hóa các hình ảnh con của hình ảnh thành các đơn vị NAL VCL trong dòng bit. Môđun mã hóa 1103 còn để mã hóa thành dòng bit cờ được thiết đặt ở giá trị thứ nhất khi giá trị loại đơn vị NAL là giống nhau đối với tất cả các đơn vị NAL VCL được liên kết với hình ảnh và được thiết đặt ở giá trị thứ hai khi giá trị loại đơn vị NAL thứ nhất đối với các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh khác với giá trị loại đơn vị NAL thứ hai đối với các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình

ảnh. Bộ mã hóa video 1102 còn bao gồm môđun lưu trữ 1105 để lưu trữ dòng bit để truyền thông đến bộ giải mã. Bộ mã hóa video 1102 còn bao gồm môđun truyền 1107 để truyền dòng bit đến bộ giải mã video 1110. Bộ mã hóa video 1102 có thể còn được tạo cấu hình để thực hiện bước bất kỳ trong số các bước của phương pháp 900.

Hệ thống 1100 cũng bao gồm bộ giải mã video 1110. Bộ giải mã video 1110 bao gồm môđun nhận 1111 để nhận dòng bit bao gồm các hình ảnh con được liên kết với hình ảnh và cờ, trong đó các hình ảnh con có trong các đơn vị NAL VCL. Bộ giải mã video 1110 còn bao gồm môđun xác định 1113 để xác định giá trị loại đơn vị NAL là giống nhau đối với tất cả các đơn vị NAL VCL được liên kết với hình ảnh khi cờ được thiết đặt ở giá trị thứ nhất. Môđun xác định 1113 còn để xác định giá trị loại đơn vị NAL thứ nhất đối với các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh khác với giá trị loại đơn vị NAL thứ hai đối với các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh khi cờ được thiết đặt ở giá trị thứ hai. Bộ giải mã video 1110 còn bao gồm môđun giải mã 1115 để giải mã một hoặc nhiều hình ảnh con dựa trên các giá trị loại đơn vị NAL. Bộ giải mã video 1110 còn bao gồm môđun chuyển tiếp 1117 để chuyển tiếp một hoặc nhiều hình ảnh con để hiển thị dưới dạng một phần của chuỗi video được giải mã. Bộ giải mã video 1110 còn có thể được tạo cấu hình để thực hiện bước bất kỳ trong số các bước của phương pháp 1000.

Thành phần thứ nhất được ghép nối trực tiếp với thành phần thứ hai khi không có các thành phần xen giữa, trừ đường dây, tuyến đi dây, hoặc môi trường khác giữa thành phần thứ nhất và thành phần thứ hai. Thành phần thứ nhất được ghép nối gián tiếp với thành phần thứ hai khi có các thành phần xen giữa, khác với đường dây, tuyến đi dây, hoặc môi trường khác giữa thành phần thứ nhất và thành phần thứ hai. Thuật ngữ “được ghép nối” và các biến thể của nó bao gồm cả được ghép nối trực tiếp và được ghép nối gián tiếp. Việc sử dụng thuật ngữ “khoảng” có nghĩa là phạm vi bao gồm $\pm 10\%$ của số tiếp theo đó trừ khi được nêu ngược lại.

Cũng cần hiểu rằng các bước của các phương pháp làm ví dụ được nêu ở đây không nhất thiết phải được thực hiện theo thứ tự được mô tả, và thứ tự của các bước của các phương pháp này nên được hiểu là chỉ làm ví dụ. Tương tự, các bước bổ sung có thể

có trong các phương pháp này, và các bước nhất định có thể được bỏ qua hoặc được kết hợp, trong các phương pháp phù hợp với các phương án khác nhau của sáng chế.

Mặc dù một số phương án đã được đề xuất trong sáng chế, nhưng có thể hiểu rằng các hệ thống và các phương pháp được bộc lộ có thể có trong nhiều dạng cụ thể khác mà không nằm ngoài mục đích hoặc phạm vi của sáng chế. Các ví dụ này cần được coi là minh họa và không phải giới hạn, và sáng chế không bị giới hạn ở các nội dung chi tiết được đưa ra ở đây. Ví dụ, các phần tử hoặc các thành phần khác nhau có thể được kết hợp hoặc được tích hợp trong hệ thống khác hoặc các dấu hiệu nhất định có thể được bỏ qua, hoặc không được thực hiện.

Ngoài ra, các kỹ thuật, các hệ thống, các hệ thống con, và các phương pháp được mô tả và được minh họa trong các phương án khác nhau rời rạc hoặc riêng rẽ có thể được kết hợp hoặc được tích hợp với các hệ thống, các thành phần, các kỹ thuật, hoặc các phương pháp khác mà không nằm ngoài phạm vi của sáng chế. Các ví dụ khác về các sự thay đổi, thay thế, và cải biến có thể biết được bởi người có hiểu biết trung bình trong lĩnh vực tương ứng và có thể được tạo ra mà không nằm ngoài mục đích và phạm vi được bộc lộ ở đây.

YÊU CẦU BẢO HỘ

1. Phương pháp giải mã được thực hiện trong bộ giải mã, phương pháp này bao gồm các bước:

nhận, bởi bộ nhận của bộ giải mã, dòng bit bao gồm các hình ảnh con được liên kết với hình ảnh và cờ, trong đó các hình ảnh con có trong các đơn vị lớp trừu tượng mạng (network abstraction layer, NAL) của lớp tạo mã video (video coding layer, VCL) (NAL VCL), trong đó cờ là `mixed_nalu_types_in_pic_flag`;

xác định, bởi bộ xử lý của bộ giải mã, rằng giá trị loại đơn vị NAL thứ nhất mà là giống nhau đối với tất cả các đơn vị NAL VCL được liên kết với hình ảnh khi cờ được thiết đặt ở giá trị thứ nhất;

xác định, bởi bộ xử lý, rằng giá trị loại đơn vị NAL thứ nhất đối với tập hợp thứ nhất của các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh khác với giá trị loại đơn vị NAL thứ hai đối với tập hợp thứ hai của các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh khi cờ được thiết đặt ở giá trị thứ hai; và

giải mã, bởi bộ xử lý, một hoặc nhiều hình ảnh con dựa trên giá trị loại đơn vị NAL thứ nhất hoặc giá trị loại đơn vị NAL thứ hai.

2. Phương pháp theo điểm 1, trong đó dòng bit bao gồm tập hợp thông số hình ảnh (PPS) bao gồm cờ.

3. Phương pháp theo điểm 1 hoặc 2, trong đó giá trị loại đơn vị NAL thứ nhất chỉ báo hình ảnh chứa hình ảnh con điểm truy cập ngẫu nhiên nội bộ (intra-random access point, IRAP), và trong đó giá trị loại đơn vị NAL thứ hai chỉ báo hình ảnh chứa hình ảnh con phi IRAP.

4. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 3, trong đó giá trị loại đơn vị NAL thứ nhất bằng làm mới quá trình giải mã tức thời (Instantaneous Decoding Refresh, IDR) với hình ảnh chủ đạo giải mã được truy cập ngẫu nhiên (IDR_W_RADL), IDR không có các hình ảnh chủ đạo (IDR_N_LP), hoặc loại đơn vị NAL truy cập ngẫu nhiên sạch (clean random access, CRA) (CRA_NUT).

5. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 4, trong đó giá trị loại đơn vị NAL thứ hai bằng

loại đơn vị NAL hình ảnh kế tiếp (TRAIL_NUT), loại đơn vị NAL hình ảnh chủ đạo giải mã được truy cập ngẫu nhiên (RADL_NUT), hoặc loại đơn vị NAL hình ảnh chủ đạo bỏ qua truy cập ngẫu nhiên (RASL) (RASL_NUT).

6. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 5, trong đó `mixed_nalu_types_in_pic_flag` bằng một khi quy định rằng hình ảnh tham chiếu đến PPS có nhiều hơn một trong số các đơn vị NAL VCL và các đơn vị NAL VCL không có cùng giá trị của loại đơn vị NAL (`nal_unit_type`), và trong đó `mixed_nalu_types_in_pic_flag` bằng không (zero) khi quy định hình ảnh tham chiếu đến PPS có một hoặc nhiều đơn vị NAL VCL và các đơn vị NAL VCL có cùng giá trị của `nal_unit_type`.

7. Phương pháp mã hóa được thực hiện trong bộ mã hóa, phương pháp này bao gồm các bước:

xác định, bởi bộ xử lý của bộ mã hóa, liệu hình ảnh có chứa các hình ảnh con thuộc các loại khác nhau hay không;

mã hóa, bởi bộ xử lý, các hình ảnh con của hình ảnh thành các đơn vị lớp trừu tượng mạng (NAL) của lớp tạo mã video (VCL) trong dòng bit;

mã hóa thành dòng bit, bởi bộ xử lý, cờ được thiết đặt ở giá trị thứ nhất khi giá trị loại đơn vị NAL thứ nhất là giống nhau đối với tất cả các đơn vị NAL VCL được liên kết với hình ảnh và được thiết đặt ở giá trị thứ hai khi giá trị loại đơn vị NAL thứ nhất đối với tập hợp thứ nhất của các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh khác với giá trị loại đơn vị NAL thứ hai đối với tập hợp thứ hai của các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh, trong đó cờ là `mixed_nalu_types_in_pic_flag`; và

lưu trữ, bởi bộ nhớ được ghép nối với bộ xử lý, dòng bit để truyền thông đến bộ giải mã.

8. Phương pháp theo điểm 7, phương pháp này còn bao gồm bước mã hóa tập hợp thông số hình ảnh (picture parameter set, PPS) thành dòng bit, trong đó cờ được mã hóa thành PPS.

9. Phương pháp theo điểm 7 hoặc 8, trong đó giá trị loại đơn vị NAL thứ nhất chỉ báo hình ảnh chứa hình ảnh con điểm truy cập ngẫu nhiên nội bộ (intra-random access point,

IRAP), và trong đó giá trị loại đơn vị NAL thứ hai chỉ báo hình ảnh chứa hình ảnh con phi IRAP.

10. Phương pháp theo điểm bất kỳ trong số các điểm từ 7 đến 9, trong đó giá trị loại đơn vị NAL thứ nhất bằng làm mới quá trình giải mã tức thời (Instantaneous Decoding Refresh, IDR) với hình ảnh chủ đạo giải mã được truy cập ngẫu nhiên (IDR_W_RADL), IDR không có các hình ảnh chủ đạo (IDR_N_LP), hoặc loại đơn vị NAL truy cập ngẫu nhiên sạch (clean random access, CRA) (CRA_NUT).

11. Phương pháp theo điểm bất kỳ trong số các điểm từ 7 đến 10, trong đó giá trị loại đơn vị NAL thứ hai bằng loại đơn vị NAL hình ảnh kế tiếp (TRAIL_NUT), loại đơn vị NAL hình ảnh chủ đạo giải mã được truy cập ngẫu nhiên (RADL_NUT), hoặc loại đơn vị NAL hình ảnh chủ đạo bỏ qua truy cập ngẫu nhiên (RASL) (RASL_NUT).

12. Phương pháp theo điểm bất kỳ trong số các điểm từ 7 đến 11, trong đó `mixed_nalu_types_in_pic_flag` bằng một khi quy định rằng hình ảnh tham chiếu đến PPS có nhiều hơn một trong số các đơn vị NAL VCL và các đơn vị NAL VCL không có cùng giá trị của loại đơn vị NAL (`nal_unit_type`), và trong đó `mixed_nalu_types_in_pic_flag` bằng không (zero) khi quy định hình ảnh tham chiếu đến PPS có một hoặc nhiều đơn vị NAL VCL và các đơn vị NAL VCL có cùng giá trị của `nal_unit_type`.

13. Thiết bị tạo mã video bao gồm:

bộ xử lý, bộ nhận được ghép nối với bộ xử lý, bộ nhớ được ghép nối với bộ xử lý, và bộ truyền được ghép nối với bộ xử lý, trong đó bộ xử lý, bộ nhận, bộ nhớ, và bộ truyền được tạo cấu hình để thực hiện phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 6.

14. Thiết bị tạo mã video bao gồm:

bộ xử lý, bộ nhận được ghép nối với bộ xử lý, bộ nhớ được ghép nối với bộ xử lý, và bộ truyền được ghép nối với bộ xử lý, trong đó bộ xử lý, bộ nhận, bộ nhớ, và bộ truyền được tạo cấu hình để thực hiện phương pháp theo điểm bất kỳ trong số các điểm từ 7 đến 12.

15. Phương tiện đọc được bằng máy tính lâu dài bao gồm sản phẩm chương trình máy tính để sử dụng bởi thiết bị tạo mã video, sản phẩm chương trình máy tính bao gồm các

lệnh thực hiện được bằng máy tính được lưu trữ trên phương tiện đọc được bằng máy tính lâu dài sao cho khi được thực hiện bởi bộ xử lý sẽ làm cho thiết bị tạo mã video thực hiện phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 6.

16. Phương tiện đọc được bằng máy tính lâu dài bao gồm sản phẩm chương trình máy tính để sử dụng bởi thiết bị tạo mã video, sản phẩm chương trình máy tính bao gồm các lệnh thực hiện được bằng máy tính được lưu trữ trên phương tiện đọc được bằng máy tính lâu dài sao cho khi được thực hiện bởi bộ xử lý sẽ làm cho thiết bị tạo mã video thực hiện phương pháp theo điểm bất kỳ trong số các điểm từ 7 đến 12.

17. Bộ giải mã bao gồm:

phương tiện nhận để nhận dòng bit bao gồm các hình ảnh con được liên kết với hình ảnh và cờ, trong đó các hình ảnh con có trong các đơn vị lớp trừu tượng mạng (NAL) của lớp tạo mã video (VCL), trong đó cờ là `mixed_nalu_types_in_pic_flag`;

phương tiện xác định để:

xác định giá trị loại đơn vị NAL thứ nhất là giống nhau đối với tất cả các đơn vị NAL VCL được liên kết với hình ảnh khi cờ được thiết đặt ở giá trị thứ nhất; và

xác định giá trị loại đơn vị NAL thứ nhất đối với tập hợp thứ nhất của các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh khác với giá trị loại đơn vị NAL thứ hai đối với tập hợp thứ hai của các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh khi cờ được thiết đặt ở giá trị thứ hai; và

phương tiện giải mã để giải mã một hoặc nhiều hình ảnh con dựa trên giá trị loại đơn vị NAL thứ nhất hoặc giá trị loại đơn vị NAL thứ hai.

18. Bộ giải mã theo điểm 17, trong đó bộ giải mã còn được tạo cấu hình để thực hiện phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 6.

19. Bộ mã hóa bao gồm:

phương tiện xác định để xác định liệu hình ảnh có chứa các hình ảnh con thuộc các loại khác nhau hay không;

phương tiện mã hóa để:

mã hóa các hình ảnh con của hình ảnh thành các đơn vị lớp trừu tượng mạng (NAL) của lớp tạo mã video (VCL) trong dòng bit; và

mã hóa, thành dòng bit, cờ được thiết đặt ở giá trị thứ nhất khi giá trị loại đơn vị NAL thứ nhất là giống nhau đối với tất cả các đơn vị NAL VCL được liên kết với hình ảnh và được thiết đặt ở giá trị thứ hai khi giá trị loại đơn vị NAL thứ nhất đối với tập hợp thứ nhất của các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh khác với giá trị loại đơn vị NAL thứ hai đối với tập hợp thứ hai của các đơn vị NAL VCL chứa một hoặc nhiều hình ảnh con của hình ảnh; và phương tiện lưu trữ để lưu trữ dòng bit để truyền thông đến bộ giải mã;

trong đó cờ là `mixed_nalu_types_in_pic_flag`.

20. Bộ mã hóa theo điểm 19, trong đó bộ mã hóa còn được tạo cấu hình để thực hiện phương pháp theo điểm bất kỳ trong số các điểm từ 7 đến 12.

1/11

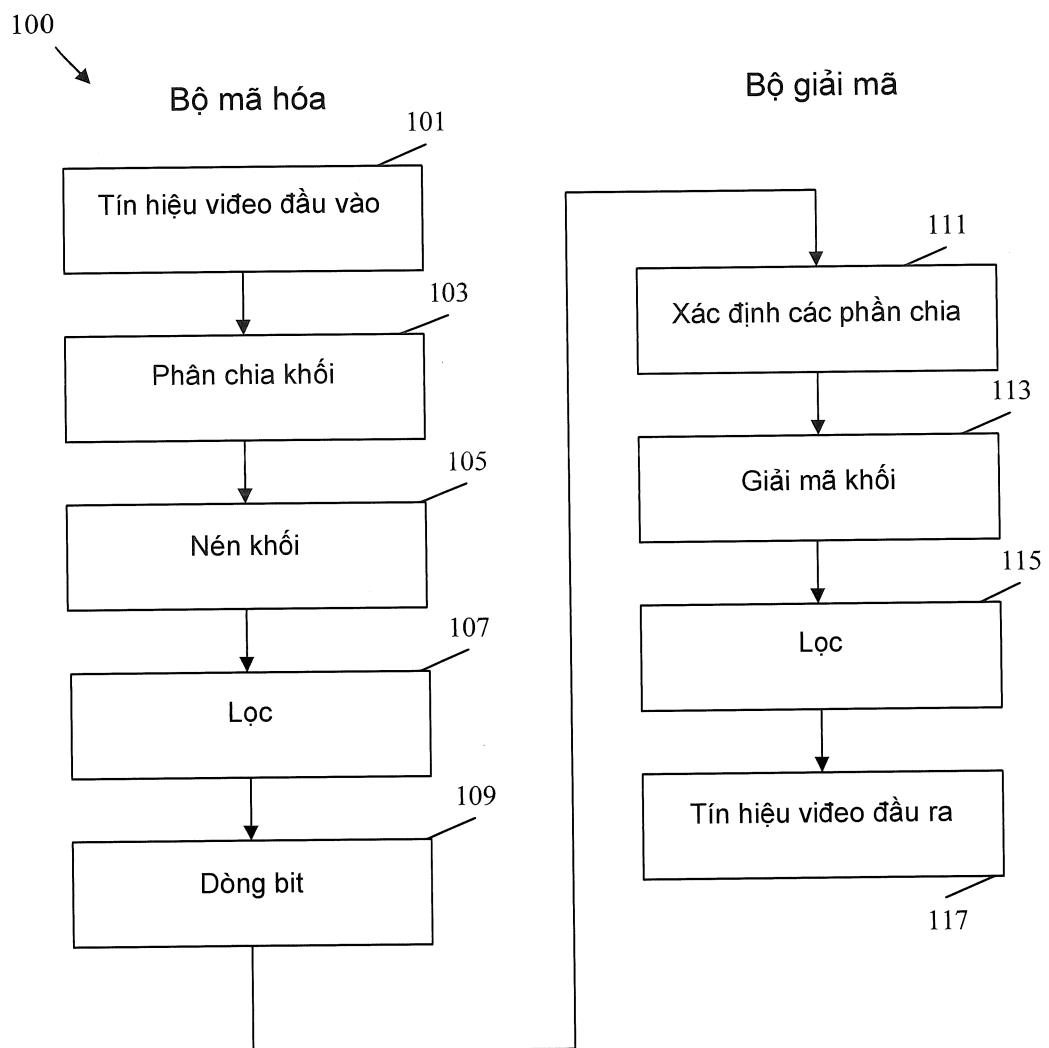


FIG. 1

2/11

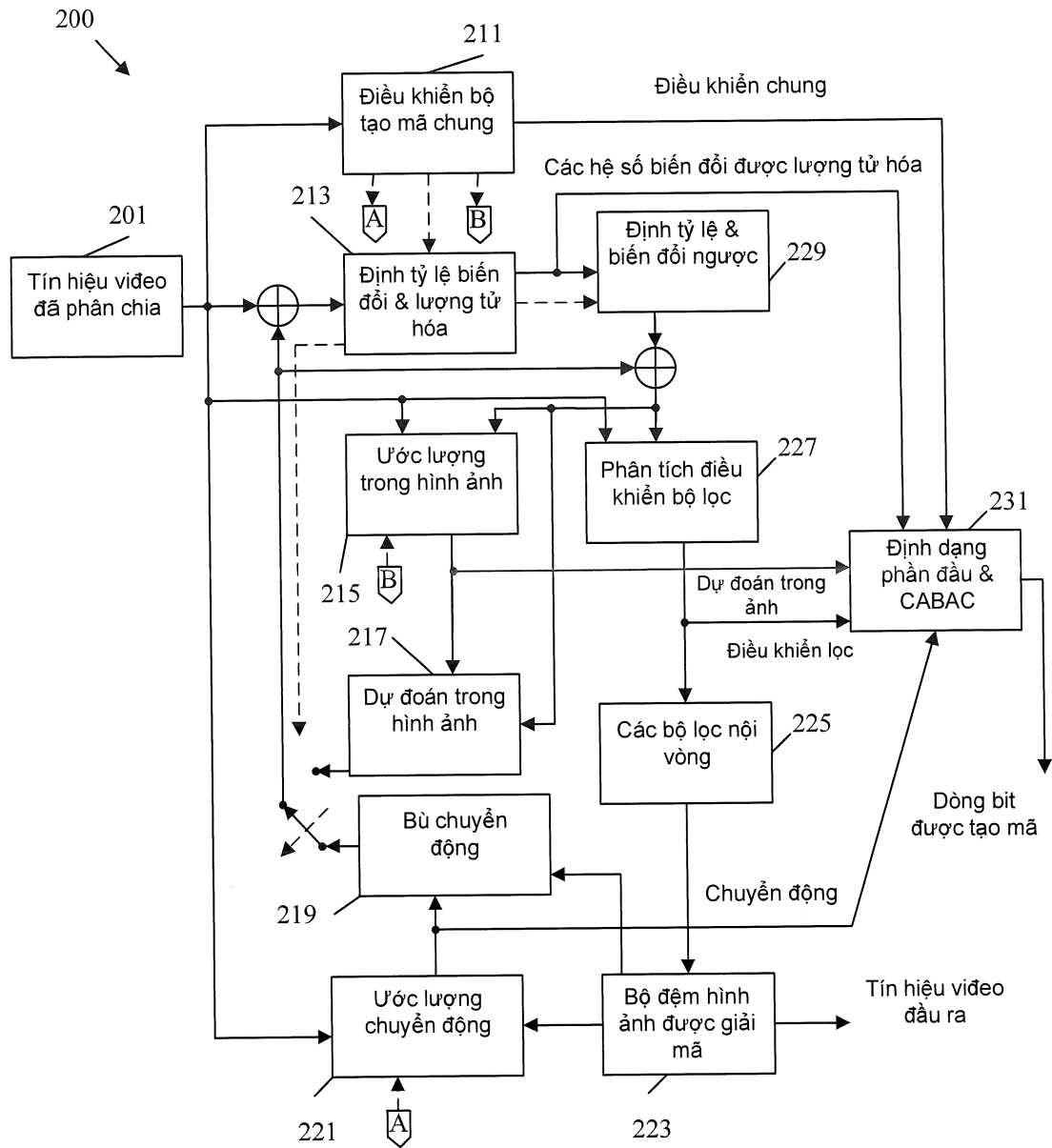


FIG. 2

3/11

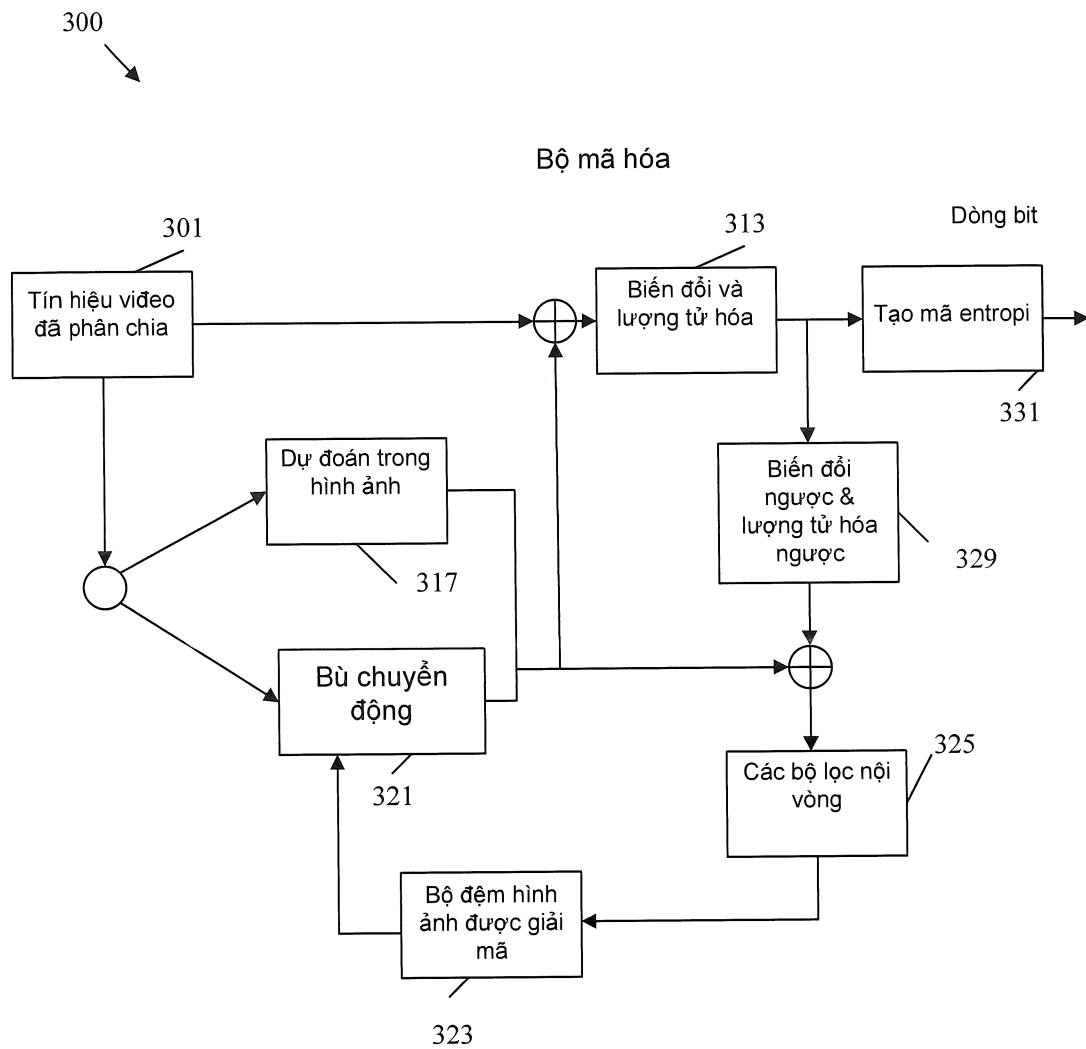


FIG. 3

4/11

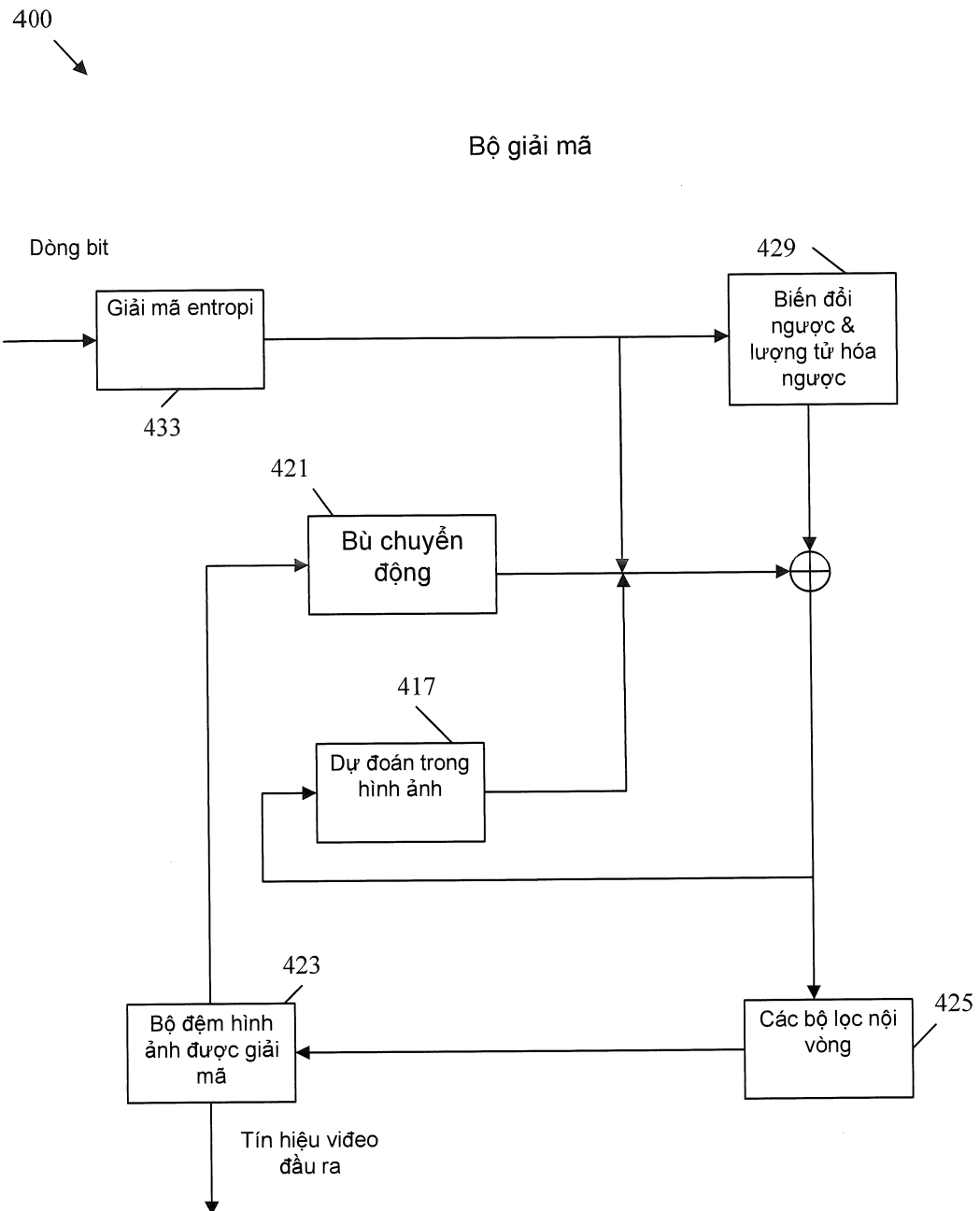


FIG. 4

5/11

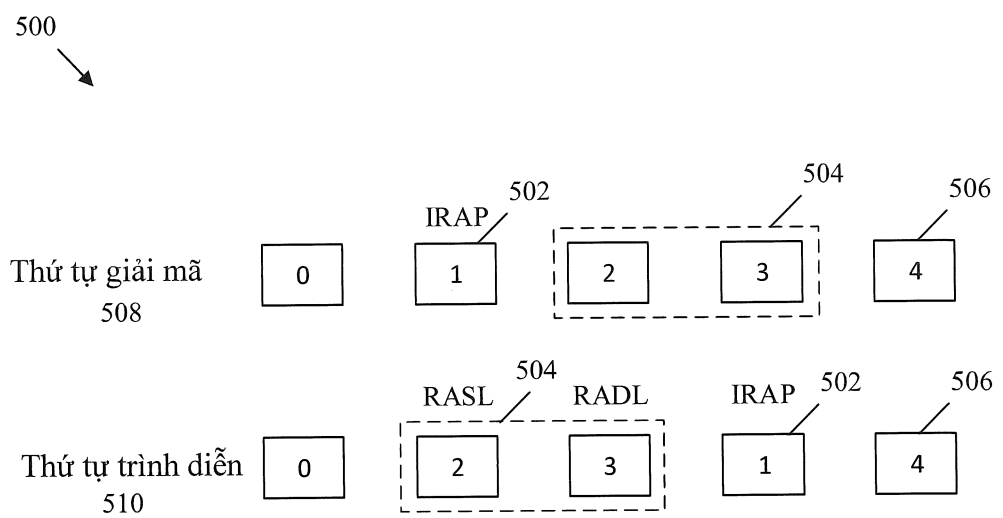


FIG. 5

6/11

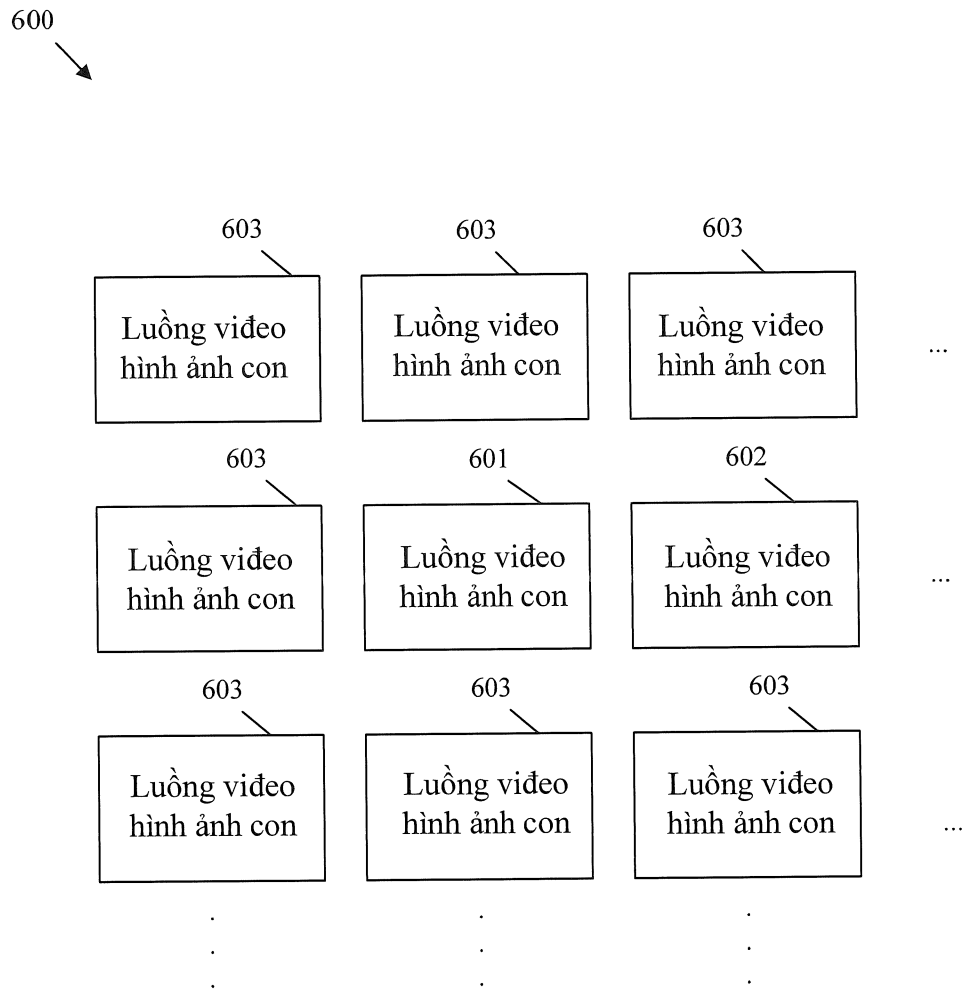
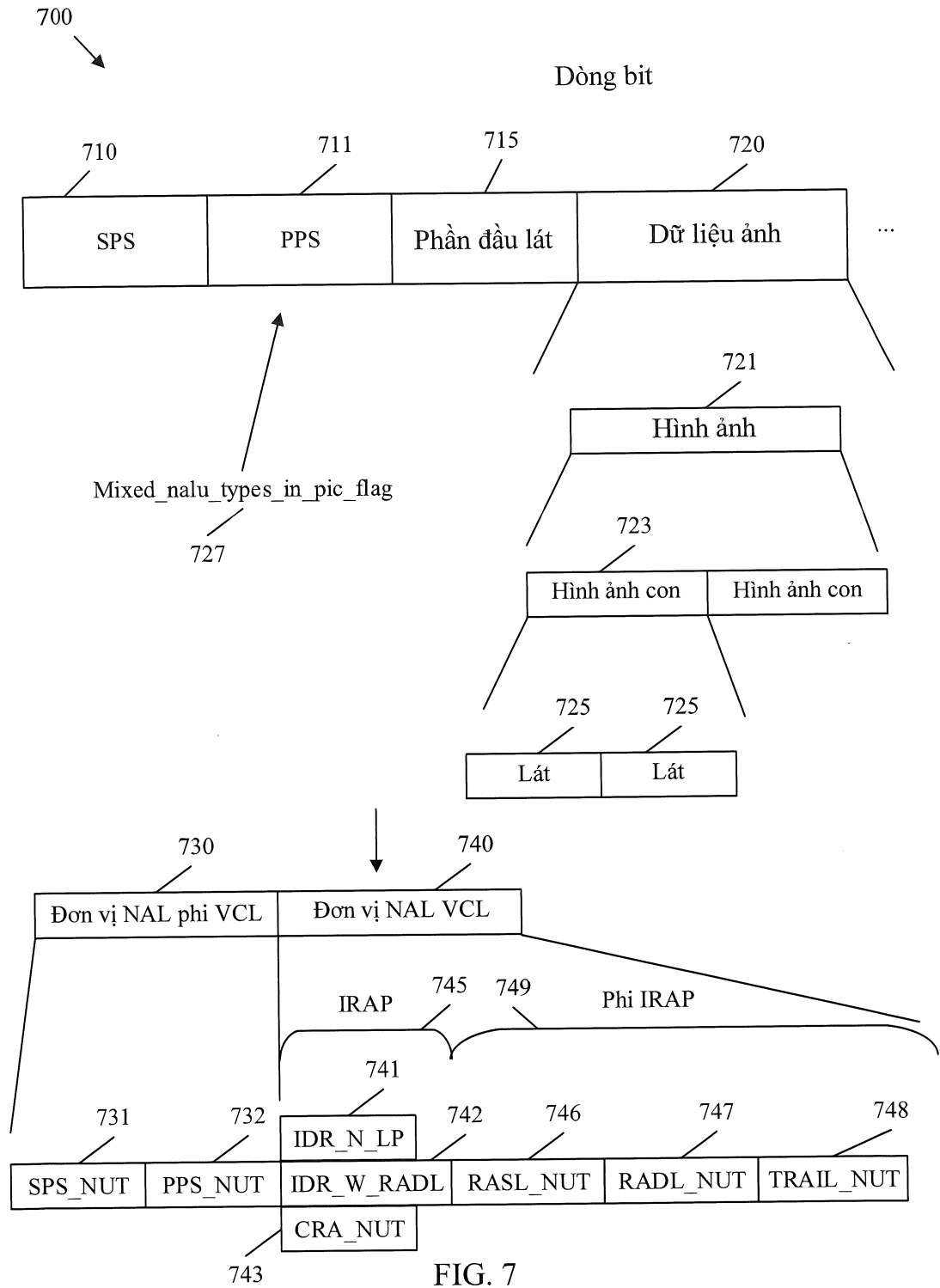


FIG. 6

7/11



8/11

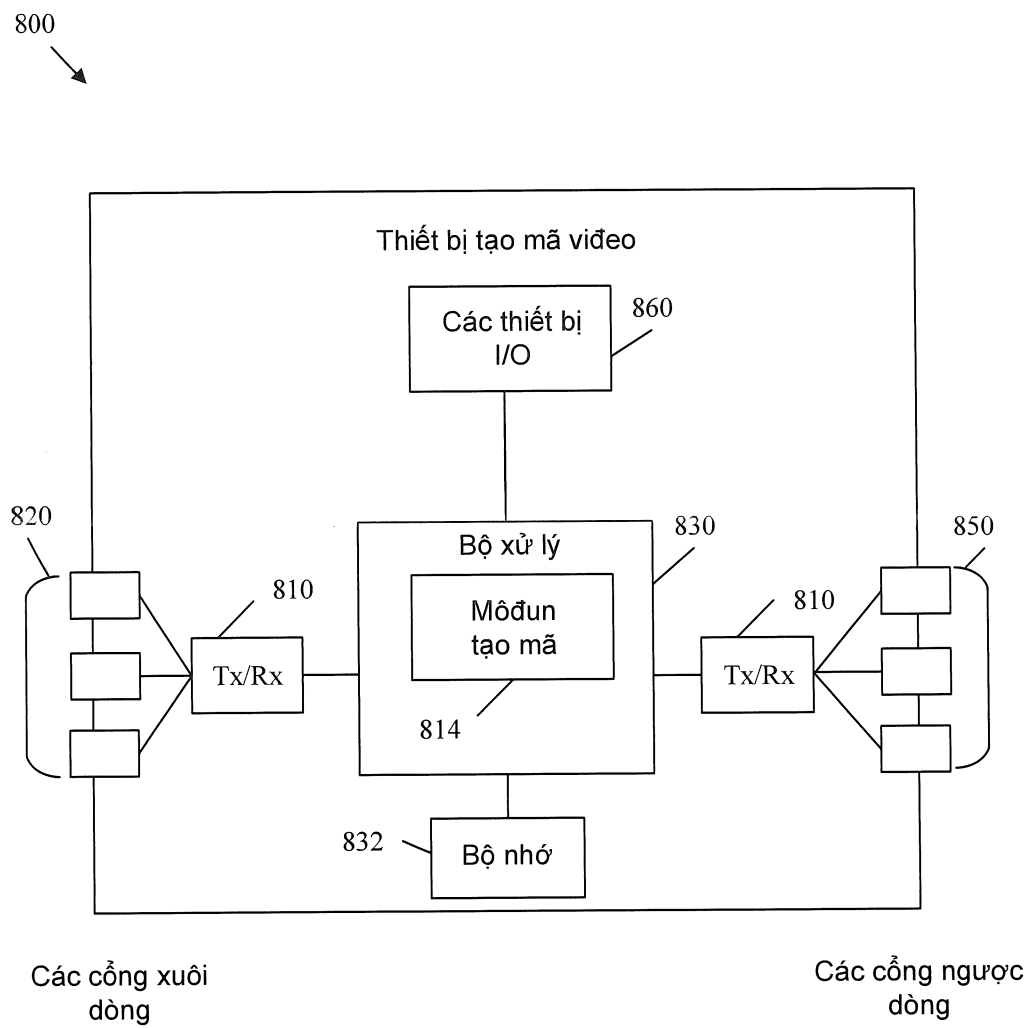


FIG. 8

9/11

900

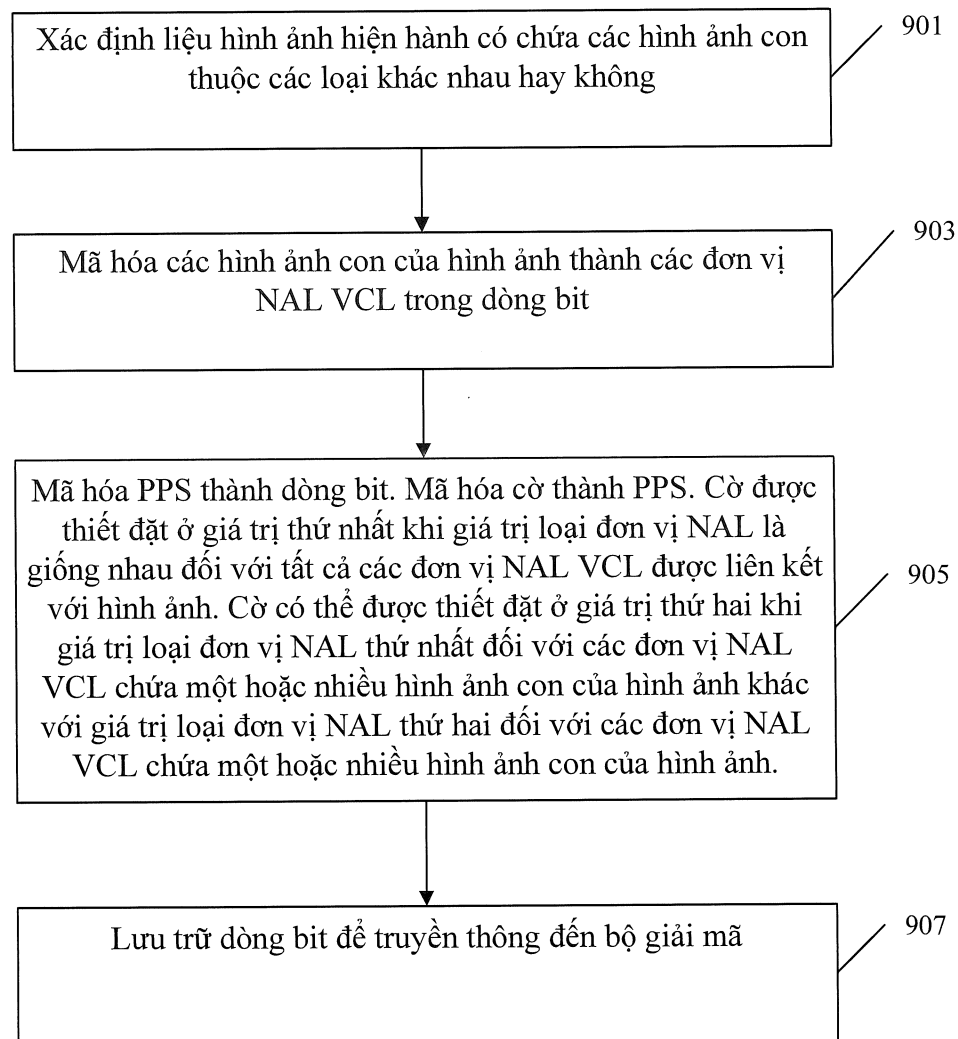


FIG. 9

10/11

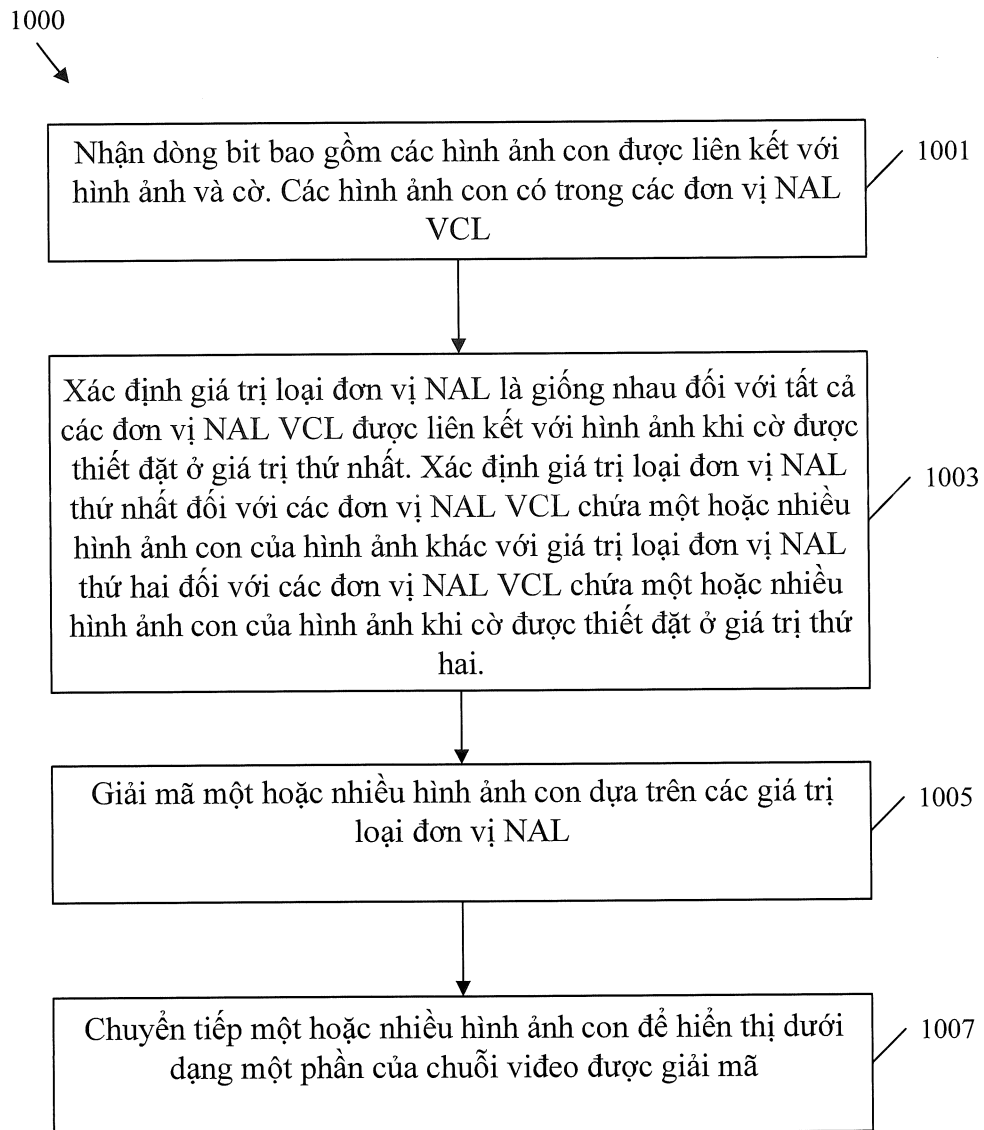


FIG. 10

11/11

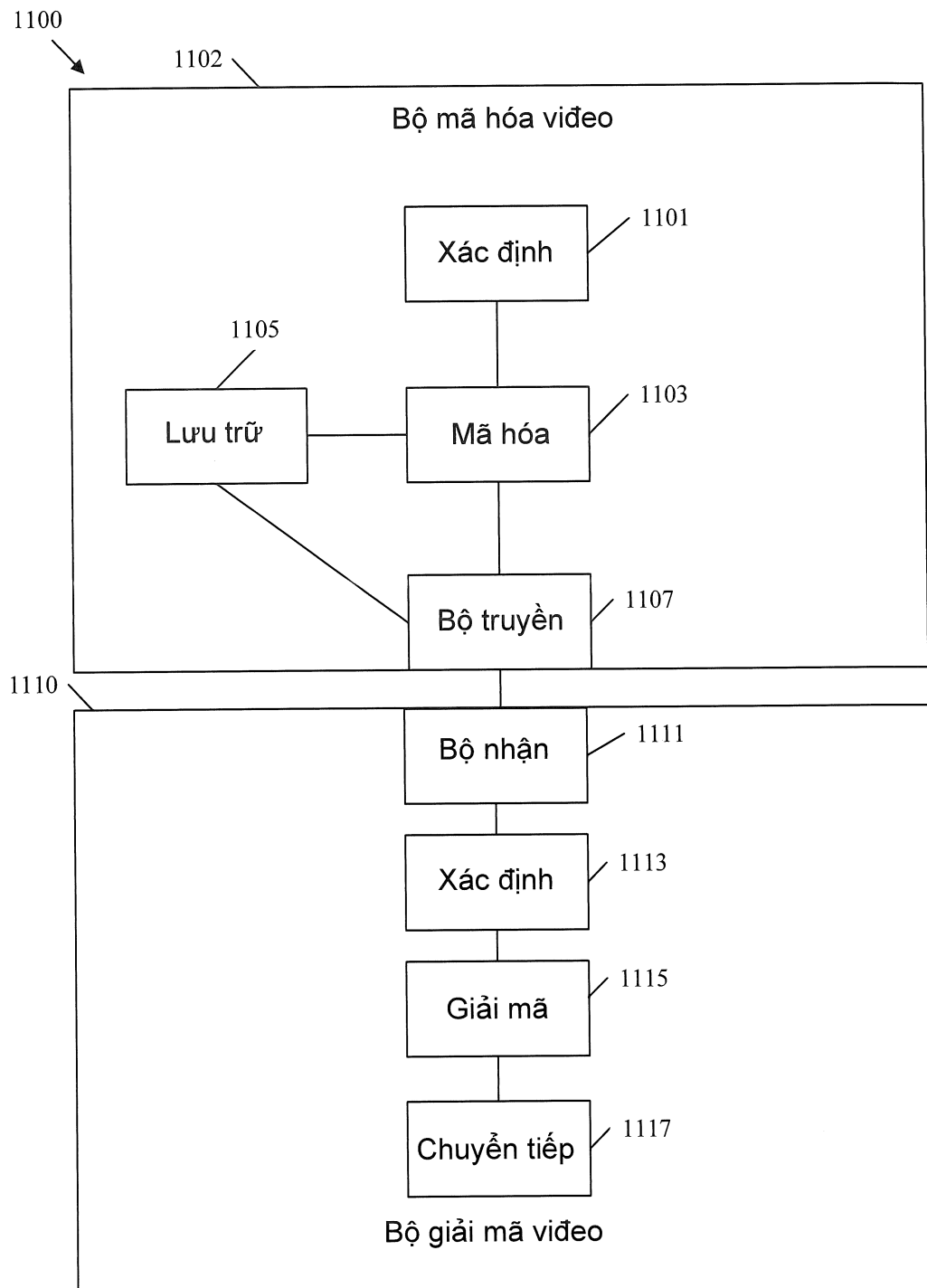


FIG. 11