



- (12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ
(19) Cộng hòa xã hội chủ nghĩa Việt Nam (VN) (11) 
CỤC SỞ HỮU TRÍ TUỆ
1-0045140
(51)^{2020.01} **H04N 19/573**; H04N 19/132; H04N 19/513; H04N 19/105; H04N 19/176 (13) **B**

-
- (21) 1-2021-06729 (22) 27/04/2020
(86) PCT/US2020/030155 27/04/2020 (87) WO2020/220048 29/10/2020
(30) 62/838,939 25/04/2019 US
(45) 25/04/2025 445 (43) 25/01/2022 406A
(71) BEIJING DAJIA INTERNET INFORMATION TECHNOLOGY CO., LTD. (CN)
Room 101D1-7, 1st Floor, Building 1, No. 6, Shangdi West Road, Haidian District,
Beijing, Beijing 100085, China
(72) XIU, Xiaoyu (CN); CHEN, Yi-Wen (TW); WANG, Xianglin (US); YE, Shuiming
(CN); MA, Tsung-Chuan (CN); JHU, Hong-Jheng (CN).
(74) Công ty Luật TNHH T&G (TGVN)
-

- (54) PHƯƠNG PHÁP TÍNH CHỈNH DỰ ĐOÁN BẰNG LUỒNG QUANG HỌC VÀ
THIẾT BỊ TÍNH TOÁN

(21) 1-2021-06729

(57) Sáng chế đề xuất phương pháp tinh chỉnh dự đoán bằng luồng quang học (prediction refinement with optical flow - PROF) và thiết bị tính toán. Phương pháp này bao gồm bước thu hình ảnh tham chiếu thứ nhất được liên kết với khối video trong tín hiệu video và vectơ chuyển động (motion vector - MV) thứ nhất, từ khối video trong hình ảnh hiện tại đến khối tham chiếu trong hình ảnh tham chiếu thứ nhất, thu các mẫu dự đoán thứ nhất $I(i,j)$ của khối video, điều khiển các độ sâu bit bên trong của các thông số PROF bên trong, thu các giá trị tinh chỉnh dự đoán cho các mẫu dự đoán thứ nhất $I(i,j)$ dựa trên các giá trị gradien theo phương ngang và theo phương dọc và các chênh lệch chuyển động theo phương ngang và theo phương dọc, thu các mẫu dự đoán thứ hai $I'(i,j)$ được liên kết với MV thứ hai và các giá trị tinh chỉnh dự đoán tương ứng cho các mẫu dự đoán thứ hai $I'(i,j)$, và thu các mẫu dự đoán cuối cùng dựa trên sự kết hợp của các mẫu dự đoán thứ nhất $I(i,j)$, các mẫu dự đoán thứ hai $I'(i,j)$, và các giá trị tinh chỉnh dự đoán.

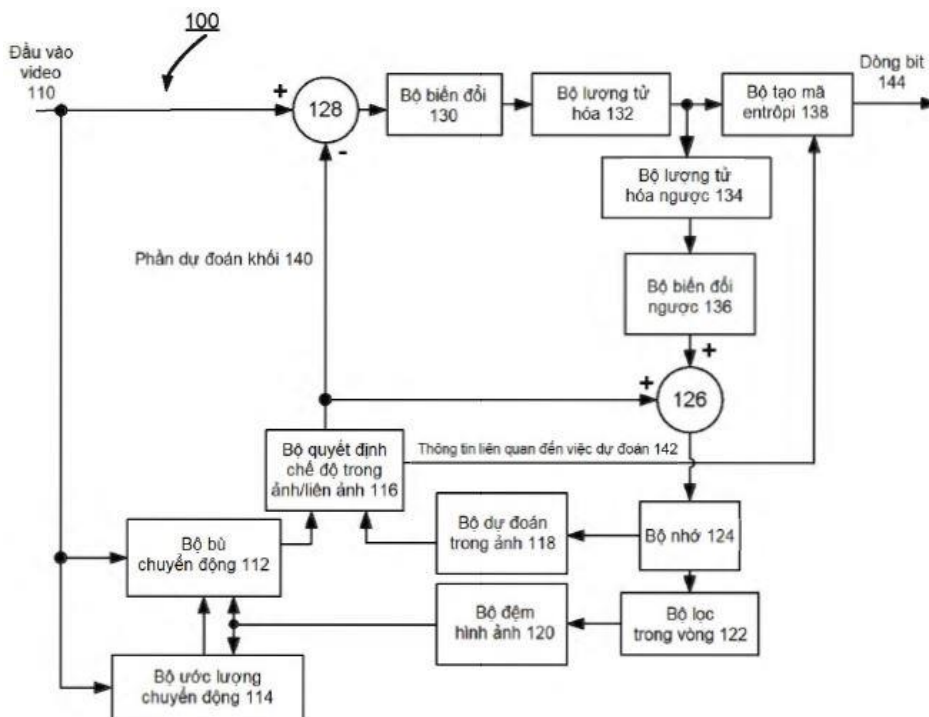


FIG. 1

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến việc tạo mã và nén video. Cụ thể hơn, sáng chế đề cập đến các phương pháp và thiết bị về hai công cụ dự đoán liên ảnh được nghiên cứu trong tiêu chuẩn tạo mã video vạn năng (versatile video coding - VVC), cụ thể là, việc tinh chỉnh dự đoán bằng luồng quang học (prediction refinement with optical flow - PROF) và luồng quang học hai chiều (bi-directional optical flow - BDOF).

Tình trạng kỹ thuật của sáng chế

Các kỹ thuật tạo mã video khác nhau có thể được sử dụng để nén dữ liệu video. Việc tạo mã video được thực hiện theo một hoặc nhiều tiêu chuẩn tạo mã video. Ví dụ, các tiêu chuẩn tạo mã video bao gồm tạo mã video vạn năng (VVC), mô hình thử nghiệm thăm dò chung (Joint Exploration Test Model - JEM), tạo mã video hiệu suất cao (High-Efficiency Video Coding - H.265/HEVC), tạo mã video cải tiến (Advanced Video Coding - H.264/AVC), tạo mã nhóm chuyên gia hình ảnh chuyển động (Moving Picture Experts Group - MPEG), hoặc loại tương tự. Việc tạo mã video thường tận dụng các phương pháp dự đoán (ví dụ, dự đoán liên ảnh, dự đoán trong ảnh, hoặc dạng tương tự) mà chúng tận dụng ưu điểm của sự dư thừa có mặt trong các hình ảnh hoặc các chuỗi video. Mục tiêu quan trọng của các kỹ thuật tạo mã video là nhằm nén dữ liệu video thành dạng sử dụng tốc độ bit thấp hơn, trong khi loại bỏ hoặc tối thiểu hóa các suy giảm đối với chất lượng video.

Bản chất kỹ thuật của sáng chế

Các ví dụ của sáng chế cung cấp các phương pháp và thiết bị biểu diễn độ sâu bit của việc tinh chỉnh dự đoán bằng luồng quang học.

Theo khía cạnh thứ nhất của sáng chế, sáng chế đề xuất phương pháp biểu diễn độ sâu bit của việc tinh chỉnh dự đoán bằng luồng quang học (PROF) cho việc giải mã tín hiệu video. Phương pháp này có thể bao gồm bước thu hình ảnh tham chiếu thứ nhất được liên kết với khối video trong tín hiệu video và vectơ chuyển động (motion vector

- MV) thứ nhất, từ khối video trong hình ảnh hiện tại đến khối tham chiếu trong hình ảnh tham chiếu thứ nhất. Hình ảnh tham chiếu thứ nhất có thể bao gồm nhiều khối video không chồng lấp nhau và ít nhất một khối video có thể được liên kết với ít nhất một MV. Phương pháp này có thể còn bao gồm bước thu các mẫu dự đoán thứ nhất $I(i, j)$ của khối video được tạo ra từ khối tham chiếu trong hình ảnh tham chiếu thứ nhất. i và j có thể thể hiện tọa độ của một mẫu với khối video. Phương pháp này có thể bao gồm bước điều khiển các độ sâu bit bên trong của các thông số PROF bên trong. Các thông số PROF bên trong có thể bao gồm các giá trị gradien theo phương ngang, các giá trị gradien theo phương dọc, các chênh lệch chuyển động theo phương ngang, và các chênh lệch chuyển động theo phương dọc được suy ra cho các mẫu dự đoán $I(i, j)$. Phương pháp này thêm vào đó có thể bao gồm bước thu các giá trị tinh chỉnh dự đoán cho các mẫu dự đoán thứ nhất $I(i, j)$ dựa trên các giá trị gradien theo phương ngang và theo phương dọc và các chênh lệch chuyển động theo phương ngang và theo phương dọc. Phương pháp này có thể bao gồm bước thu, khi khối video có thể chứa MV thứ hai, các mẫu dự đoán thứ hai $I'(i, j)$ được liên kết với MV thứ hai và các giá trị tinh chỉnh dự đoán tương ứng cho các mẫu dự đoán thứ hai $I'(i, j)$ này. Phương pháp này có thể bao gồm bước thu các mẫu dự đoán cuối cùng của khối video dựa trên sự kết hợp của các mẫu dự đoán thứ nhất $I(i, j)$, các mẫu dự đoán thứ hai $I'(i, j)$, và các giá trị tinh chỉnh dự đoán.

Theo khía cạnh thứ hai của sáng chế, sáng chế đề xuất phương pháp biểu diễn độ sâu bit của luồng quang học hai chiều (BDOF) cho việc giải mã tín hiệu video. Phương pháp này có thể bao gồm bước thu hình ảnh tham chiếu thứ nhất $I^{(0)}$ và hình ảnh tham chiếu thứ hai $I^{(1)}$ được liên kết với khối video. Hình ảnh tham chiếu thứ nhất $I^{(0)}$ có thể là ở trước hình ảnh hiện tại và hình ảnh tham chiếu thứ hai $I^{(1)}$ có thể là ở sau hình ảnh hiện tại trong thứ tự hiển thị. Phương pháp này cũng có thể bao gồm bước thu các mẫu dự đoán thứ nhất $I^{(0)}(i, j)$ của khối video từ khối tham chiếu trong hình ảnh tham chiếu thứ nhất $I^{(0)}$. i và j có thể thể hiện tọa độ của một mẫu với hình ảnh hiện tại. Phương pháp này có thể bao gồm bước thu các mẫu dự đoán thứ hai $I^{(1)}(i, j)$ của khối video từ khối tham chiếu trong hình ảnh tham chiếu thứ hai $I^{(1)}$. Phương pháp này có thể bao gồm bước áp dụng BDOF đối với khối video dựa trên các mẫu dự đoán thứ nhất $I^{(0)}(i, j)$

và các mẫu dự đoán thứ hai $I^{(1)}(i, j)$. Phương pháp này có thể bao gồm bước thu các giá trị gradien theo phương ngang và theo phương dọc của mẫu dự đoán thứ nhất $I^{(0)}(i, j)$ và các mẫu dự đoán thứ hai $I^{(1)}(i, j)$ dựa trên các mẫu dự đoán đã được đệm thêm. Phương pháp này thêm vào đó có thể bao gồm bước thu các tinh chỉnh về chuyển động cho các mẫu trong khối video dựa trên BDOF đang được áp dụng đối với khối video và các giá trị gradien theo phương ngang và theo phương dọc. Phương pháp này có thể bao gồm bước thu các mẫu dự đoán đôi của khối video dựa trên các tinh chỉnh về chuyển động này.

Theo khía cạnh thứ ba của sáng chế, sáng chế đề xuất thiết bị tính toán. Thiết bị tính toán này có thể bao gồm một hoặc nhiều bộ xử lý, bộ nhớ đọc được bằng máy tính phi chuyển tiếp lưu trữ các lệnh có thể thực thi được bởi một hoặc nhiều bộ xử lý này. Một hoặc nhiều bộ xử lý này có thể được cấu hình để thu hình ảnh tham chiếu thứ nhất được liên kết với khối video trong tín hiệu video và MV thứ nhất, từ khối video trong hình ảnh hiện tại đến khối tham chiếu trong hình ảnh tham chiếu thứ nhất. Hình ảnh tham chiếu thứ nhất có thể bao gồm nhiều khối video không chồng lấp nhau và ít nhất một khối video có thể được liên kết với ít nhất một MV. Một hoặc nhiều bộ xử lý này có thể còn được cấu hình để thu các mẫu dự đoán thứ nhất $I(i, j)$ của khối video được tạo ra từ khối tham chiếu trong hình ảnh tham chiếu thứ nhất. i và j thể hiện tọa độ của một mẫu với khối video. Một hoặc nhiều bộ xử lý này có thể được cấu hình để điều khiển các độ sâu bit bên trong của các thông số PROF bên trong. Các thông số PROF bên trong có thể bao gồm các giá trị gradien theo phương ngang, các giá trị gradien theo phương dọc, các chênh lệch chuyển động theo phương ngang, và các chênh lệch chuyển động theo phương dọc được suy ra cho các mẫu dự đoán $I(i, j)$. Một hoặc nhiều bộ xử lý này thêm vào đó có thể được cấu hình để thu các giá trị tinh chỉnh dự đoán cho các mẫu dự đoán thứ nhất $I(i, j)$ dựa trên các giá trị gradien theo phương ngang và theo phương dọc và các chênh lệch chuyển động theo phương ngang và theo phương dọc. Một hoặc nhiều bộ xử lý này thêm vào đó có thể được cấu hình để thu, khi khối video có thể chứa MV thứ hai, các mẫu dự đoán thứ hai $I'(i, j)$ được liên kết với MV thứ hai và các giá trị tinh chỉnh dự đoán tương ứng cho các mẫu dự đoán thứ hai $I'(i, j)$ này. Một hoặc nhiều bộ xử lý này có thể được cấu hình để thu các mẫu dự đoán cuối cùng của khối video dựa trên sự kết hợp của các mẫu dự đoán thứ nhất $I(i, j)$, các mẫu dự

đoán thứ hai $I'(i, j)$, và các giá trị tinh chỉnh dự đoán.

Theo khía cạnh thứ tư của sáng chế, sáng chế đề xuất thiết bị tính toán. Thiết bị tính toán này có thể bao gồm một hoặc nhiều bộ xử lý, bộ nhớ đọc được bằng máy tính phi chuyên tiếp lưu trữ các lệnh có thể thực thi được bởi một hoặc nhiều bộ xử lý này. Một hoặc nhiều bộ xử lý này có thể được cấu hình để thu hình ảnh tham chiếu thứ nhất $I^{(0)}$ và hình ảnh tham chiếu thứ hai $I^{(1)}$ được liên kết với khối video. Hình ảnh tham chiếu thứ nhất $I^{(0)}$ có thể là ở trước hình ảnh hiện tại và hình ảnh tham chiếu thứ hai $I^{(1)}$ có thể là ở sau hình ảnh hiện tại trong thứ tự hiển thị. Một hoặc nhiều bộ xử lý này có thể còn được cấu hình để thu các mẫu dự đoán thứ nhất $I^{(0)}(i, j)$ của khối video từ khối tham chiếu trong hình ảnh tham chiếu thứ nhất $I^{(0)}$. i và j có thể thể hiện tọa độ của một mẫu với hình ảnh hiện tại. Một hoặc nhiều bộ xử lý này có thể được cấu hình để thu các mẫu dự đoán thứ hai $I^{(1)}(i, j)$ của khối video từ khối tham chiếu trong hình ảnh tham chiếu thứ hai $I^{(1)}$. Một hoặc nhiều bộ xử lý này có thể được cấu hình để áp dụng BDOF đối với khối video dựa trên các mẫu dự đoán thứ nhất $I^{(0)}(i, j)$ và các mẫu dự đoán thứ hai $I^{(1)}(i, j)$. Một hoặc nhiều bộ xử lý này có thể được cấu hình để thu các giá trị gradien theo phương ngang và theo phương dọc của mẫu dự đoán thứ nhất $I^{(0)}(i, j)$ và các mẫu dự đoán thứ hai $I^{(1)}(i, j)$ dựa trên các mẫu dự đoán đã được đệm thêm. Một hoặc nhiều bộ xử lý này thêm vào đó có thể được cấu hình để thu các tinh chỉnh về chuyển động cho các mẫu trong khối video dựa trên BDOF đang được áp dụng đối với khối video và các giá trị gradien theo phương ngang và theo phương dọc. Một hoặc nhiều bộ xử lý này có thể được cấu hình để thu các mẫu dự đoán đôi của khối video dựa trên các tinh chỉnh về chuyển động này

Cần hiểu rằng cả phần mô tả tổng quát nêu trên lẫn phần mô tả chi tiết sau đây đều chỉ là các ví dụ và không mang tính giới hạn đối với sáng chế.

Mô tả vắn tắt các hình vẽ

Các hình vẽ kèm theo, được tích hợp trong, và tạo thành một phần, của bản mô tả này, minh họa các ví dụ nhất quán với nội dung của sáng chế, và cùng với phần mô tả, có chức năng giải thích các nguyên lý của sáng chế.

Fig.1 là sơ đồ khối của bộ mã hóa, theo ví dụ của sáng chế.

Fig.2 là sơ đồ khối của bộ giải mã, theo ví dụ của sáng chế.

Fig.3A là sơ đồ minh họa các phần chia khối trong cấu trúc cây nhiều loại, theo ví dụ của sáng chế.

Fig.3B là sơ đồ minh họa các phần chia khối trong cấu trúc cây nhiều loại, theo ví dụ của sáng chế.

Fig.3C là sơ đồ minh họa các phần chia khối trong cấu trúc cây nhiều loại, theo ví dụ của sáng chế.

Fig.3D là sơ đồ minh họa các phần chia khối trong cấu trúc cây nhiều loại, theo ví dụ của sáng chế.

Fig.3E là sơ đồ minh họa các phần chia khối trong cấu trúc cây nhiều loại, theo ví dụ của sáng chế.

Fig.4 là sơ đồ minh họa mô hình luồng quang học hai chiều (BDOF), theo ví dụ của sáng chế.

Fig.5A là hình vẽ minh họa mô hình afin, theo ví dụ của sáng chế.

Fig.5B là hình vẽ minh họa mô hình afin, theo ví dụ của sáng chế.

Fig.6 là hình vẽ minh họa mô hình afin, theo ví dụ của sáng chế.

Fig.7 là hình vẽ minh họa việc tinh chỉnh dự đoán bằng luồng quang học (PROF), theo ví dụ của sáng chế.

Fig.8 là tiến trình công việc của BDOF, theo ví dụ của sáng chế.

Fig.9 là tiến trình công việc của PROF, theo ví dụ của sáng chế.

Fig.10 là phương pháp biểu diễn độ sâu bit của PROF, theo nội dung của sáng chế.

Fig.11 là phương pháp biểu diễn độ sâu bit của BDOF, theo nội dung của sáng chế.

Fig.12 là hình vẽ minh họa tiến trình công việc của PROF cho dự đoán đôi, theo ví dụ của sáng chế.

Fig.13 là hình vẽ minh họa các giai đoạn trong đường ống của quá trình BDOF và PROF, theo nội dung của sáng chế.

Fig.14 là hình vẽ minh họa phương pháp suy diễn gradient của BDOF, theo nội dung của sáng chế.

Fig.15 là hình vẽ minh họa phương pháp suy diễn gradient của PROF, theo nội dung của sáng chế.

Fig.16 là sơ đồ minh họa môi trường điện toán được ghép nối với giao diện người dùng, theo ví dụ của sáng chế.

Mô tả chi tiết sáng chế

Sau đây, sáng chế sẽ được mô tả chi tiết đến các phương án thực hiện mang tính ví dụ, các ví dụ của chúng này được minh họa trong các hình vẽ kèm theo. Phần mô tả sau đây liên quan đến các hình vẽ kèm theo trong đó các số chỉ dẫn giống nhau trong các hình vẽ khác nhau thể hiện các thành phần giống hoặc tương tự nhau trừ khi được biểu thị khác đi. Các cách triển khai được nêu trong phần mô tả sau đây của các phương án không thể hiện tất cả các cách triển khai phù hợp với nội dung của sáng chế. Thay vào đó, chúng chỉ là các ví dụ về các thiết bị và các phương pháp phù hợp với các khía cạnh liên quan đến sáng chế như được nêu trong các yêu cầu bảo hộ kèm theo.

Thuật ngữ được sử dụng trong sáng chế này chỉ nhằm mục đích mô tả các phương án cụ thể và không được dự tính nhằm giới hạn nội dung bộc lộ của sáng chế. Như được sử dụng trong phần mô tả sáng chế và các yêu cầu bảo hộ kèm theo, các dạng số ít cũng được dự tính để bao gồm các dạng số nhiều, trừ khi ngữ cảnh nêu rõ khác đi. Cũng cần hiểu rằng, thuật ngữ “và/hoặc” được sử dụng ở đây là nhằm biểu thị và bao gồm bất kỳ hoặc tất cả các tổ hợp khả thi của một hoặc nhiều mục được liệt kê liên quan.

Cần hiểu rằng, mặc dù các thuật ngữ “thứ nhất”, “thứ hai”, “thứ ba”, v.v. có thể được dùng ở đây để mô tả các thông tin khác nhau, thông tin đó không bị hạn chế bởi các thuật ngữ này. Các thuật ngữ này chỉ được sử dụng để phân biệt một phân loại của thông tin so với phân loại của thông tin khác. Ví dụ, không tách rời khỏi phạm vi của sáng chế này, thông tin thứ nhất có thể được đặt là thông tin thứ hai; và tương tự, thông tin thứ hai cũng có thể được đặt là thông tin thứ nhất. Như được sử dụng ở đây, thuật ngữ “nếu” có thể được hiểu nghĩa là “khi” hoặc “lúc” hoặc “để đáp ứng lại một phán quyết”, tùy thuộc vào ngữ cảnh.

Phiên bản thứ nhất của tiêu chuẩn HEVC được hoàn thành vào tháng 10 năm 2013, phiên bản này đề xuất việc tiết kiệm xấp xỉ 50% tốc độ bit hoặc chất lượng cảm nhận tương đương khi so với tiêu chuẩn tạo mã video thế hệ trước H.264/MPEG AVC. Mặc dù tiêu chuẩn HEVC cung cấp các cải tiến đáng kể đối với việc tạo mã so với các tiêu chuẩn tiền nhiệm của nó nhưng có bằng chứng là có thể đạt được hiệu quả tạo mã vượt trội bằng các công cụ tạo mã bổ sung trên HEVC. Dựa trên điều đó, cả VCEG và MPEG đều bắt đầu công việc khám phá các kỹ thuật tạo mã mới cho việc chuẩn hóa tạo mã video trong tương lai. Nhóm khám phá video hợp tác (Joint Video Exploration Team -

JVET) được hình thành vào tháng 10 năm 2015 bởi ITU-T VCEG và ISO/IEC MPEG để bắt đầu nghiên cứu quy mô về các công nghệ cải tiến có khả năng cho phép tăng cường hiệu quả tạo mã một cách thực chất. Một phần mềm tham chiếu được gọi là mô hình khám phá hợp tác (Joint Exploration Model - JEM) đã được duy trì bởi JVET bằng cách tích hợp nhiều công cụ tạo mã bổ sung trên đỉnh của mô hình thử nghiệm HEVC (HEVC test Model - HM).

Vào tháng 10 năm 2017, lời kêu gọi chung yêu cầu các đề xuất (Call for Proposal - CfP) về việc nén video với khả năng vượt qua HEVC đã được phát hành bởi ITU-T và ISO/IEC. Vào tháng 4 năm 2018, 23 phản hồi đối với CfP được nhận và được đánh giá tại phiên họp JVET lần thứ 10, các phản hồi này đã cho thấy hiệu quả nén thu được cao hơn HEVC khoảng quanh mức 40%. Dựa trên các kết quả đánh giá này, JVET đã triển khai dự án mới để phát triển tiêu chuẩn tạo mã video thế hệ mới có tên là tạo mã video vận năng (VVC). Cũng trong tháng đó, một cơ sở mã phần mềm tham chiếu, được gọi là mô hình thử nghiệm VVC (VVC test model - VTM), được thiết lập để thể hiện cách triển khai tham chiếu của tiêu chuẩn VVC.

Giống như HEVC, VVC được xây dựng trên khung tạo mã video lai trên cơ sở khối. Fig.1 thể hiện sơ đồ chung của bộ mã hóa video trên cơ sở khối cho VVC. Cụ thể, Fig.1 thể hiện bộ mã hóa 100 điển hình. Bộ mã hóa 100 có đầu vào video 110, bộ bù chuyển động 112, bộ ước lượng chuyển động 114, bộ quyết định chế độ trong ảnh/liên ảnh 116, phần dự đoán khối 140, bộ cộng 128, bộ biến đổi 130, bộ lượng tử hóa 132, thông tin liên quan đến việc dự đoán 142, bộ dự đoán trong ảnh 118, bộ đệm hình ảnh 120, bộ lượng tử hóa ngược 134, bộ biến đổi ngược 136, bộ cộng 126, bộ nhớ 124, bộ lọc trong vòng 122, bộ tạo mã entropi 138, và dòng bit 144.

Trong bộ mã hóa 100, khung video được phân chia thành nhiều khối video để xử lý. Với mỗi khối video đã cho, dự đoán được tạo ra dựa trên hoặc là cách tiếp cận dự đoán liên ảnh hoặc là cách tiếp cận dự đoán trong ảnh.

Phần dư của việc dự đoán, thể hiện chênh lệch giữa khối video hiện tại, là một phần của đầu vào video 110, và phần dự đoán của nó, là một phần của phần dự đoán khối 140, được gửi đến bộ biến đổi 130 từ bộ cộng 128. Các hệ số biến đổi sau đó được gửi từ bộ biến đổi 130 tới bộ lượng tử hóa 132 để làm giảm entropi. Các hệ số được lượng tử hóa sau đó được nạp vào bộ mã hóa entropy 138 để tạo ra dòng bit video được

nén. Như được thể hiện trên Fig.1, thông tin liên quan đến việc dự đoán 142 từ bộ dự đoán trong ảnh/liên ảnh 116, chẳng hạn như thông tin phân chia khối video, các véctor chuyển động (các MV), chỉ số hình ảnh tham chiếu, và chế độ dự đoán trong ảnh, cũng được nạp qua bộ mã hóa entropy 138 và được lưu vào trong dòng bit được nén 144. Dòng bit được nén 144 bao gồm dòng bit video.

Trong bộ mã hóa 100, các mạch liên quan đến bộ giải mã cũng là cần thiết để tái tạo các điểm ảnh cho mục đích dự đoán. Đầu tiên, phần dư của việc dự đoán được tái tạo qua bộ lượng tử hóa ngược 134 và bộ biến đổi ngược 136. Phần dư của việc dự đoán được tái tạo này được kết hợp với phần dự đoán khối 140 để tạo ra các điểm ảnh được tái tạo chưa được lọc cho khối video hiện tại.

Việc dự đoán theo không gian (hoặc “việc dự đoán trong ảnh”) sử dụng các điểm ảnh từ các mẫu của các khối lân cận đã được tạo mã sẵn (được gọi là các mẫu tham chiếu) trong cùng một khung video với khối video hiện tại để dự đoán khối video hiện tại.

Việc dự đoán theo thời gian (cũng được gọi là “dự đoán liên ảnh”) sử dụng các điểm ảnh được tái tạo từ các hình ảnh video đã được tạo mã sẵn để dự đoán khối video hiện tại. Việc dự đoán theo thời gian làm giảm dư thừa về mặt thời gian cố hữu trong tín hiệu video. Tín hiệu dự đoán theo thời gian cho đơn vị tạo mã (coding unit - CU) hoặc khối tạo mã đã cho thường được phát tín hiệu bởi một hoặc nhiều MV, các MV này chỉ ra lượng và hướng chuyển động giữa CU hiện tại và tham chiếu theo thời gian của nó. Hơn nữa, nếu có hỗ trợ nhiều hình ảnh tham chiếu, một chỉ số hình ảnh tham chiếu sẽ được gửi bổ sung, chỉ số này được dùng để nhận dạng xem tín hiệu dự đoán theo thời gian đến từ hình ảnh tham chiếu nào trong bộ lưu trữ hình ảnh tham chiếu.

Bộ ước lượng chuyển động 114 lấy đầu vào video 110 và tín hiệu từ bộ đệm hình ảnh 120 và đưa ra, đến bộ bù chuyển động 112, tín hiệu ước lượng chuyển động. Bộ bù chuyển động 112 lấy đầu vào video 110, tín hiệu từ bộ đệm hình ảnh 120, và tín hiệu ước lượng chuyển động từ bộ ước lượng chuyển động 114 và đưa ra, đến bộ quyết định chế độ trong ảnh/liên ảnh 116, tín hiệu bù chuyển động.

Sau khi việc dự đoán theo không gian và/hoặc theo thời gian được thực hiện, bộ quyết định chế độ trong ảnh/liên ảnh 116 trong bộ mã hóa 100 chọn chế độ dự đoán tốt nhất, ví dụ, dựa trên phương pháp tối ưu hóa đối với độ méo theo tốc độ. Phần dự đoán

khối 140 sau đó được trừ từ khối video hiện tại và phần dư của việc dự đoán được tạo thành được giải tương liên nhờ sử dụng bộ biến đổi 130 và bộ lượng tử hóa 132. Các hệ số phần dư được lượng tử hóa được tạo thành được lượng tử hóa ngược bởi bộ lượng tử hóa ngược 134 và được biến đổi ngược bởi bộ biến đổi ngược 136 để tạo thành phần dư được tái tạo, phần dư này sau đó được cộng lại vào khối dự đoán để tạo thành tín hiệu được tái tạo của CU. Việc lọc trong vòng 122 tiếp theo, chẳng hạn như bộ lọc khử khối, độ lệch thích ứng mẫu (sample adaptive offset - SAO), và/hoặc bộ lọc trong vòng thích ứng (adaptive in-loop filter - ALF), có thể được áp dụng trên CU được tái tạo trước khi nó được nhập trong bộ lưu trữ hình ảnh tham chiếu của bộ đệm hình ảnh 120 và được sử dụng để tạo mã cho các khối video tương lai. Để tạo thành dòng bit video đầu ra 144, chế độ tạo mã (liên ảnh hoặc trong ảnh), thông tin chế độ dự đoán, thông tin chuyển động, và các hệ số phần dư được lượng tử hóa tất cả đều được gửi tới đơn vị tạo mã entropi 138 để còn được nén và đóng gói để tạo thành dòng bit.

Ví dụ, bộ lọc khử khối là có sẵn trong AVC, HEVC, cũng như phiên bản hiện tại của VVC. Trong HEVC, bộ lọc trong vòng bổ sung được gọi là SAO (sample adaptive offset - độ lệch thích ứng mẫu) được định nghĩa để tiếp tục cải thiện hơn nữa hiệu suất tạo mã. Trong phiên bản hiện tại của tiêu chuẩn VVC, thêm một bộ lọc trong vòng khác được gọi là ALF (adaptive loop filter - bộ lọc vòng thích ứng) đang được chủ động nghiên cứu và có thể được chứa trong sản phẩm cuối cùng.

Các thao tác của bộ lọc trong vòng này là tùy chọn. Việc thực hiện các thao tác này giúp cải thiện hiệu suất mã hóa và chất lượng trực quan. Chúng cũng có thể được tắt do quyết định được đưa ra bởi bộ mã hóa 100 để tiết kiệm độ phức tạp tính toán.

Cần lưu ý rằng việc dự đoán trong ảnh thường được dựa trên các điểm ảnh được tái tạo không được lọc, trong khi việc dự đoán liên ảnh được dựa trên các điểm ảnh được tái tạo được lọc nếu các tùy chọn bộ lọc này được bật lên bởi bộ mã hóa 100.

Tín hiệu video đầu vào được xử lý theo từng khối (được gọi là các đơn vị tạo mã (CU)). Trong VTM-1.0, CU có thể lên tới 128x128 điểm ảnh. Tuy nhiên, khác với HEVC, vốn phân chia các khối chỉ dựa trên các cây tứ phân, trong VVC, một đơn vị cây tạo mã (Coding Tree Unit - CTU) được chia tách thành các CU, để thích ứng với các đặc tính cục bộ luôn biến đổi, dựa trên cây tứ phân/nhị phân/tam phân. Thêm vào đó, khái niệm loại đơn vị nhiều phần chia trong HEVC được loại bỏ, cụ thể là, việc tách rời

CU, đơn vị dự đoán (Prediction Unit - PU) và đơn vị biến đổi (Transform Unit - TU) không còn tồn tại trong VVC nữa; thay vào đó, mỗi CU luôn được sử dụng như là đơn vị cơ bản cho cả việc dự đoán lẫn việc biến đổi mà không có thêm sự phân chia nào nữa. Trong cấu trúc cây nhiều loại này, một CTU đầu tiên được phân chia bởi cấu trúc cây tứ phân. Sau đó, mỗi nút lá cây tứ phân có thể còn được phân chia bởi cấu trúc cây nhị phân hoặc tam phân.

Như được thể hiện trên Fig.3A, Fig.3B, Fig.3C, Fig.3D, và Fig.3E (được mô tả dưới đây), có năm loại chia tách, phân chia tứ phân, phân chia nhị phân theo phương ngang, phân chia nhị phân theo phương dọc, phân chia tam phân theo phương ngang, và phân chia tam phân theo phương dọc.

Fig.3A thể hiện sơ đồ minh họa sự phân chia tứ phân khối trong cấu trúc cây nhiều loại, theo nội dung của sáng chế.

Fig.3B thể hiện sơ đồ minh họa sự phân chia nhị phân theo phương dọc khối trong cấu trúc cây nhiều loại, theo nội dung của sáng chế.

Fig.3C thể hiện sơ đồ minh họa sự phân chia nhị phân theo phương ngang khối trong cấu trúc cây nhiều loại, theo nội dung của sáng chế.

Fig.3D thể hiện sơ đồ minh họa sự phân chia tam phân theo phương dọc khối trong cấu trúc cây nhiều loại, theo nội dung của sáng chế.

Fig.3E thể hiện sơ đồ minh họa sự phân chia tam phân theo phương ngang khối trong cấu trúc cây nhiều loại, theo nội dung của sáng chế.

Trên Fig.1, việc dự đoán theo không gian và/hoặc việc dự đoán theo thời gian có thể được thực hiện. Việc dự đoán theo không gian (hoặc “việc dự đoán trong ảnh”) sử dụng các điểm ảnh từ các mẫu của các khối lân cận đã được tạo mã sẵn (được gọi là các mẫu tham chiếu) trong cùng một hình ảnh/phiên video để dự đoán khối video hiện tại. Việc dự đoán theo không gian làm giảm dư thừa về mặt không gian cố hữu trong tín hiệu video. Việc dự đoán theo thời gian (cũng được gọi là “dự đoán liên ảnh” hoặc “dự đoán được bù chuyển động”) sử dụng các điểm ảnh được tái tạo từ các hình ảnh video đã được tạo hóa sẵn để dự đoán khối video hiện tại. Việc dự đoán theo thời gian làm giảm dư thừa về mặt thời gian cố hữu trong tín hiệu video. Tín hiệu dự đoán theo thời gian cho CU đã cho thường được phát tín hiệu bởi một hoặc nhiều MV, các MV này chỉ ra lượng và hướng chuyển động giữa CU hiện tại và tham chiếu theo thời gian của nó.

Cũng như vậy, nếu có hỗ trợ nhiều hình ảnh tham chiếu, một chỉ số hình ảnh tham chiếu sẽ được gửi bổ sung, chỉ số này được dùng để nhận dạng xem tín hiệu dự đoán theo thời gian đến từ hình ảnh tham chiếu nào trong bộ lưu trữ hình ảnh tham chiếu. Sau khi dự đoán theo không gian và/hoặc theo thời gian, khối quyết định chế độ trong bộ mã hóa chọn chế độ dự đoán tốt nhất, ví dụ, dựa trên phương pháp tối ưu hóa độ méo theo tốc độ. Sau đó, khối dự đoán được trừ từ khối video hiện tại, và phần dư của việc dự đoán được giải tương liên nhờ sử dụng bộ biến đổi và được lượng tử hóa. Các hệ số phần dư được lượng tử hóa sẽ được lượng tử hóa ngược và biến đổi ngược để tạo thành phần dư được tái tạo, phần dư này sau đó được cộng lại vào khối dự đoán để tạo thành tín hiệu được tái tạo của CU. Hơn nữa, việc lọc trong vòng, chẳng hạn như bộ lọc khử khối, độ lệch thích ứng mẫu (SAO), và bộ lọc trong vòng thích ứng (ALF), có thể được áp dụng trên CU được tái tạo trước khi nó được nhập trong bộ lưu trữ hình ảnh tham chiếu và được sử dụng để tạo mã cho các khối video tương lai. Để tạo thành dòng bit video đầu ra, chế độ tạo mã (liên ảnh hoặc trong ảnh), thông tin chế độ dự đoán, thông tin chuyển động, và các hệ số phần dư được lượng tử hóa tất đều được gửi tới đơn vị tạo mã entropi để còn được nén và đóng gói để tạo thành dòng bit.

Fig.2 thể hiện sơ đồ khối chung của bộ giải mã video cho VVC. Cụ thể, Fig.2 thể hiện sơ đồ khối bộ giải mã 200 điển hình. Bộ giải mã 200 có dòng bit 210, bộ giải mã entropi 212, bộ lượng tử hóa ngược 214, bộ biến đổi ngược 216, bộ cộng 218, bộ lựa chọn chế độ trong ảnh/liên ảnh 220, bộ dự đoán trong ảnh 222, bộ nhớ 230, bộ lọc trong vòng 228, bộ bù chuyển động 224, bộ đệm hình ảnh 226, thông tin liên quan đến việc dự đoán 234, và đầu ra video 232.

Bộ giải mã 200 là tương tự như phần liên quan tới việc tái tạo nằm trong bộ mã hóa 100 của Fig.1. Trong bộ giải mã 200, luồng bit video đi vào 210 đầu tiên được giải mã qua bộ giải mã entropi 212 để suy ra các mức hệ số được lượng tử hóa và thông tin liên quan tới việc dự đoán. Các mức hệ số được lượng tử hóa sau đó được xử lý qua bộ lượng tử hóa ngược 214 và bộ biến đổi ngược 216 để thu được phần dư của việc dự đoán được tái tạo. Cơ chế bộ dự đoán khối, được triển khai trong bộ phận lựa chọn chế độ trong ảnh/liên ảnh 220, được cấu hình để thực hiện hoặc là việc dự đoán trong ảnh 222, hoặc là việc bù chuyển động 224, dựa trên thông tin dự đoán đã được giải mã. Tập hợp của các điểm ảnh được tái tạo chưa được lọc được thu bằng cách lấy tổng phần dư dự

đoán được tái tạo từ bộ biến đổi ngược 216 và đầu ra dự đoán được tạo ra bởi cơ chế bộ dự đoán khối, sử dụng bộ lấy tổng 218.

Khối được tái tạo có thể còn đi qua bộ lọc trong vòng 228 trước khi nó được lưu trữ trong bộ đệm hình ảnh 226 vốn hoạt động như là bộ lưu trữ hình ảnh tham chiếu. Video được tái tạo trong bộ đệm hình ảnh 226 có thể được gửi ra để điều vận thiết bị hiển thị, cũng như được sử dụng để dự đoán các khối video tương lai. Trong các tình huống mà bộ lọc trong vòng 228 được bật lên, thao tác lọc được thực hiện trên các điểm ảnh được tái tạo này để suy ra đầu ra video được tái tạo cuối cùng 232.

Trên Fig.2, dòng bit video đầu tiên được giải mã entropy tại đơn vị giải mã entropy. Chế độ tạo mã và thông tin dự đoán được gửi đến hoặc là đơn vị dự đoán theo không gian (nếu được tạo mã trong ảnh) hoặc là đơn vị dự đoán theo thời gian (nếu được tạo mã liên ảnh) để tạo thành khối dự đoán. Các hệ số biến đổi phần dư được gửi tới đơn vị lượng tử hóa ngược và đơn vị biến đổi ngược để tái tạo khối phần dư. Sau đó, khối dự đoán và khối phần dư được cộng lại với nhau. Khối được tái tạo có thể còn trải qua việc lọc trong vòng trước khi nó được lưu trữ trong bộ lưu trữ hình ảnh tham chiếu. Sau đó, video được tái tạo trong bộ lưu trữ ảnh tham chiếu được gửi ra để điều vận thiết bị hiển thị, cũng như được sử dụng để dự đoán các khối video tương lai.

Nói chung, các kỹ thuật dự đoán liên ảnh cơ bản được áp dụng trong VCC đều được giữ cho giống như HEVC ngoại trừ việc một số môđun còn được mở rộng và/hoặc tăng cường. Cụ thể, đối với tất cả các tiêu chuẩn video có trước, một khối tạo mã chỉ có thể được liên kết với một MV duy nhất khi khối tạo mã được dự đoán đơn hoặc hai MV khi khối tạo mã được dự đoán đôi. Do giới hạn như vậy đối với việc bù chuyển động trên cơ sở khối thông thường, chuyển động nhỏ có thể vẫn còn lại bên trong các mẫu dự đoán sau khi bù chuyển động, do đó ảnh hưởng tiêu cực đến hiệu suất tổng thể của việc bù chuyển động. Để cải thiện cả độ chi tiết và độ chính xác của các MV, hai phương pháp tinh chỉnh trên phạm vi mẫu dựa trên luồng quang học, cụ thể là luồng quang học hai chiều (BDOF) và tinh chỉnh dự đoán bằng luồng quang học (PROF) cho chế độ afin, hiện đang được nghiên cứu cho tiêu chuẩn VVC. Các khía cạnh kỹ thuật chính của hai công cụ tạo mã liên ảnh này sẽ được xem xét ngắn gọn trong phần sau đây.

Luồng quang học hai chiều

Trong VVC, BDOF được áp dụng để tinh chỉnh các mẫu dự đoán của các khối tạo

mã được dự đoán đôi. Cụ thể là, như được thể hiện trên Fig.4, hình vẽ này thể hiện minh họa về BDOF, BDOF là việc tinh chỉnh về chuyển động trên phạm vi mẫu được thực hiện trên đỉnh của các dự đoán được bù chuyển động trên cơ sở khối khi dự đoán đôi được sử dụng. Việc tinh chỉnh về chuyển động (v_x, v_y) của mỗi khối con 4×4 được tính toán bằng cách tối thiểu hóa chênh lệch giữa các mẫu dự đoán L0 và L1 sau khi BDOF được áp dụng bên trong một cửa sổ 6×6 xung quanh khối con đó. Cụ thể, giá trị của (v_x, v_y) được suy ra dưới dạng

$$\begin{aligned} v_x &= S_1 > 0 ? \text{clip3} \left(-th_{BDOF}, th_{BDOF}, -((S_3 \cdot 2^3) \gg \lfloor \log_2 S_1 \rfloor) \right) : 0 \\ v_y &= S_5 > 0 ? \text{clip3} \left(-th_{BDOF}, th_{BDOF}, - \left((S_6 \cdot 2^3 \right. \right. \\ &\quad \left. \left. - ((v_x S_{2,m}) \ll n_{S_2} + v_x S_{2,s}) / 2 \right) \gg \lfloor \log_2 S_5 \rfloor \right) : 0 \end{aligned} \quad (1)$$

trong đó $\lfloor \cdot \rfloor$ là hàm làm tròn xuống (floor); $\text{clip3}(\min, \max, x)$ là hàm xén giá trị cho trước x bên trong khoảng giới hạn $[\min, \max]$; ký hiệu $\gg$ thể hiện thao tác dịch phải bit; ký hiệu $\ll$ thể hiện thao tác dịch trái bit; th_{BDOF} là ngưỡng tinh chỉnh chuyển động để ngăn chặn các lỗi lan truyền do chuyển động cục bộ bất thường, được tính bằng $1 \ll \max(5, \text{bit-depth} - 7)$, trong đó bit-depth là độ sâu bit bên trong. Trong (1), $S_{2,m} = S_2 \gg n_{S_2}$, $S_{2,s} = S_2 \& (2^{n_{S_2}} - 1)$.

Các giá trị của S_1, S_2, S_3, S_5 và S_6 được tính bằng

$$\begin{aligned} S_1 &= \sum_{(i,j) \in \Omega} \psi_x(i,j) \cdot \psi_x(i,j) \\ S_2 &= \sum_{(i,j) \in \Omega} \psi_x(i,j) \cdot \psi_y(i,j) \\ S_3 &= \sum_{(i,j) \in \Omega} \theta(i,j) \cdot \psi_x(i,j) \\ S_5 &= \sum_{(i,j) \in \Omega} \psi_y(i,j) \cdot \psi_y(i,j) \end{aligned} \quad (2)$$

$$S_6 = \sum_{(i,j) \in \Omega} \theta(i,j) \cdot \psi_y(i,j)$$

trong đó

$$\begin{aligned} \psi_x(i,j) &= \left(\frac{\partial I^{(1)}}{\partial x}(i,j) + \frac{\partial I^{(0)}}{\partial x}(i,j) \right) \gg \max(1, bit-depth - 11) \\ \psi_y(i,j) &= \left(\frac{\partial I^{(1)}}{\partial y}(i,j) + \frac{\partial I^{(0)}}{\partial y}(i,j) \right) \gg \max(1, bit-depth - 11) \\ \theta(i,j) &= \left(I^{(1)}(i,j) \gg \max(4, bit-depth - 8) \right) \\ &\quad - \left(I^{(0)}(i,j) \gg \max(4, bit-depth - 8) \right) \end{aligned} \quad (3)$$

trong đó $I^{(k)}(i,j)$ là giá trị mẫu tại tọa độ (i,j) của tín hiệu dự đoán trong danh sách k , $k = 0,1$, được tạo ra với độ chính xác cao trung gian (cụ thể là, 16 bit); $\frac{\partial I^{(k)}}{\partial x}(i,j)$ và $\frac{\partial I^{(k)}}{\partial y}(i,j)$ là các gradient theo phương ngang và theo phương dọc của mẫu thu được bằng cách tính toán trực tiếp sự chênh lệch giữa hai mẫu lân cận của nó, cụ thể là

$$\begin{aligned} \frac{\partial I^{(k)}}{\partial x}(i,j) &= \left(I^{(k)}(i+1,j) - I^{(k)}(i-1,j) \right) \gg \max(6, bit-depth - 6) \\ \frac{\partial I^{(k)}}{\partial y}(i,j) &= \left(I^{(k)}(i,j+1) - I^{(k)}(i,j-1) \right) \gg \max(6, bit-depth - 6) \end{aligned} \quad (4)$$

Dựa trên sự tinh chỉnh chuyển động được suy ra trong (1), các mẫu dự đoán đôi cuối cùng của CU được tính toán bằng cách nội suy các mẫu dự đoán L0/L1 dọc theo quỹ đạo chuyển động dựa trên mô hình luồng quang học, như được chỉ ra bởi

$$pred_{BDOF}(x,y) = \left(I^{(0)}(x,y) + I^{(1)}(x,y) + b + o_{offset} \right) \gg shift \quad (5)$$

$$b = rnd \left(\left(v_x \left(\frac{\partial I^{(1)}(x, y)}{\partial x} - \frac{\partial I^{(0)}(x, y)}{\partial x} \right) \right) / 2 \right) \\ + rnd \left(\left(v_y \left(\frac{\partial I^{(1)}(x, y)}{\partial y} - \frac{\partial I^{(0)}(x, y)}{\partial y} \right) \right) / 2 \right)$$

trong đó *shift* và *offset* là giá trị dịch phải và giá trị độ lệch được áp dụng để kết hợp các tín hiệu dự đoán L0 và L1 cho việc dự đoán đôi, mà chúng tương ứng bằng $15 - BD$ và $1 \ll (14 - BD) + 2 \cdot (1 \ll 13)$. Dựa trên phương pháp điều khiển độ sâu bit nêu trên, độ sâu bit tối đa của các thông số trung gian của toàn bộ quá trình BDOF sẽ không vượt quá 32-bit, và đầu vào lớn nhất đối với phép nhân này là trong vòng 15-bit, cụ thể là số nhân 15-bit là đủ cho các cách triển khai BDOF.

Chế độ affin

Trong HEVC, chỉ mô hình chuyển động tịnh tiến là được áp dụng cho việc dự đoán được bù chuyển động. Trong khi đó trong thực tế, có rất nhiều loại chuyển động, ví dụ như phóng to/thu nhỏ, quay, các chuyển động phối cảnh, và các chuyển động bất thường khác. Trong VVC, việc dự đoán được bù chuyển động affin được áp dụng bằng cách phát tín hiệu của một cờ cho mỗi khối tạo mã liên ảnh để chỉ ra rằng liệu mô hình chuyển động tịnh tiến hay chuyển động affin được áp dụng cho việc dự đoán liên ảnh. Trong thiết kế VVC hiện tại, hai chế độ affin, bao gồm chế độ affin 4 thông số và chế độ affin 6 thông số, được hỗ trợ đối với một khối tạo mã affin.

Mô hình affin 4 thông số có các thông số sau đây: hai thông số cho chuyển động tịnh tiến tương ứng theo các phương ngang và dọc, một thông số cho chuyển động thu-phóng và một thông số cho chuyển động quay theo cả hai phương. Thông số thu-phóng theo phương ngang là bằng với thông số thu-phóng theo phương dọc. Thông số quay theo phương ngang là bằng với thông số quay theo phương dọc. Để đạt được hiệu quả thích ứng tốt hơn đối với các MV và thông số affin, trong VVC, các thông số affin đó được diễn giải thành hai MV (các MV này cũng được gọi là các vectơ chuyển động của điểm điều khiển (control point motion vector - CPMV)) được đặt ở góc trên bên trái và góc trên bên phải khối hiện tại. Như được thể hiện trên các hình vẽ Fig.5A và Fig.5B, trường chuyển động affin của khối được mô tả bởi hai MV điểm điều khiển (V_0, V_1).

Fig.5A thể hiện minh họa về mô hình afin 4 thông số. Fig.5B thể hiện minh họa về mô hình afin 4 thông số. Dựa trên chuyển động của điểm điều khiển, trường chuyển động (v_x, v_y) của một khối được tạo mã afin được mô tả dưới dạng

$$\begin{aligned} v_x &= \frac{(v_{1x} - v_{0x})}{w} x - \frac{(v_{1y} - v_{0y})}{w} y + v_{0x} \\ v_y &= \frac{(v_{1y} - v_{0y})}{w} x + \frac{(v_{1x} - v_{0x})}{w} y + v_{0y} \end{aligned} \quad (6)$$

Chế độ afin 6 thông số có các thông số sau đây: hai thông số cho chuyển động tịnh tiến tương ứng theo các phương ngang và dọc, một thông số cho chuyển động thu-phóng và một thông số cho chuyển động quay theo phương ngang, một thông số cho chuyển động thu-phóng và một thông số cho chuyển động quay theo phương dọc. Mô hình chuyển động afin 6 thông số được tạo mã bằng ba MV tại ba CPMV.

Fig.6 thể hiện minh họa về mô hình afin 6 thông số. Như được thể hiện trên Fig.6, ba điểm điều khiển của một khối afin 6 thông số được đặt ở góc trên bên trái, góc trên bên phải, và góc dưới bên trái của khối đó. Chuyển động tại điểm điều khiển góc trên bên trái được liên hệ với chuyển động tịnh tiến, và chuyển động tại điểm điều khiển góc trên bên phải được liên hệ với chuyển động quay và thu-phóng theo phương ngang, và chuyển động tại điểm điều khiển góc dưới bên trái được liên hệ với chuyển động quay và thu-phóng theo phương dọc. So với mô hình chuyển động afin 4 thông số, chuyển động quay và thu-phóng theo phương ngang của mô hình 6 thông số có thể không giống như các chuyển động đó theo phương dọc. Giả sử rằng (V_0, V_1, V_2) là các MV của các góc trên bên trái, trên bên phải, và dưới bên trái của khối hiện tại trên Fig.6, MV của mỗi khối con (v_x, v_y) được suy ra nhờ sử dụng ba MV tại các điểm điều khiển như sau:

$$\begin{aligned} v_x &= v_{0x} + (v_{1x} - v_{0x}) * \frac{x}{w} + (v_{2x} - v_{0x}) * \frac{y}{h} \\ v_y &= v_{0y} + (v_{1y} - v_{0y}) * \frac{x}{w} + (v_{2y} - v_{0y}) * \frac{y}{h} \end{aligned} \quad (7)$$

PROF cho chế độ afin

Để cải thiện độ chính xác của việc bù chuyển động afin, PROF hiện đang được nghiên cứu trong VVC hiện tại, mà nó tinh chỉnh việc bù chuyển động afin dựa trên khối con, dựa trên mô hình luồng quang học. Cụ thể, sau khi thực hiện việc bù chuyển động

afin dựa trên khối con, mẫu dự đoán độ chói của một khối afin được biến đổi bởi một giá trị tinh chỉnh mẫu được suy ra dựa trên phương trình luồng quang học. Cụ thể, các thao tác của PROF có thể được tổng kết thành bốn bước như sau:

Bước một: Việc bù chuyển động afin trên cơ sở khối con được thực hiện để tạo ra dự đoán khối con $I(i, j)$ nhờ sử dụng các MV khối con như được suy ra trong (6) cho mô hình afin 4 thông số và (7) cho mô hình afin 6 thông số.

Bước hai: Các gradien theo không gian $g_x(i, j)$ và $g_y(i, j)$ của mỗi mẫu trong các mẫu dự đoán được tính toán dưới dạng

$$\begin{aligned} g_x(i, j) &= (I(i+1, j) - I(i-1, j)) \gg (\max(2, 14 - \text{bit-depth}) - 4) \\ g_y(i, j) &= (I(i, j+1) - I(i, j-1)) \gg (\max(2, 14 - \text{bit-depth}) - 4) \end{aligned} \quad (8)$$

Để tính các gradien này, một cột/hàng các mẫu dự đoán bổ sung cần được tạo ra trên mỗi phía của một khối con. Để giảm băng thông bộ nhớ và độ phức tạp, các mẫu trên các biên mở rộng này được sao chép từ vị trí điểm ảnh gần nhất trong hình ảnh tham chiếu để tránh các quá trình nội suy bổ sung.

Bước ba: Giá trị tinh chỉnh dự đoán độ chói được tính bởi

$$\Delta I(i, j) = g_x(i, j) * \Delta v_x(i, j) + g_y(i, j) * \Delta v_y(i, j) \quad (9)$$

trong đó $\Delta v(i, j)$ là chênh lệch giữa MV điểm ảnh được tính cho vị trí mẫu (i, j) , được biểu thị bởi $v(i, j)$, và MV khối con của khối con nơi đặt điểm ảnh (i, j) .

Bước bốn: Trong thiết kế PROF hiện tại, sau khi cộng tinh chỉnh dự đoán vào mẫu dự đoán ban đầu, một thao tác xén được thực hiện để xén giá trị của mẫu dự đoán đã tinh chỉnh để nó nằm trong 15-bit, cụ thể là

$$\begin{aligned} I^r(i, j) &= I(i, j) + \Delta I(i, j) \\ I^r(i, j) &= \text{clip3}(-2^{14}, 2^{14} - 1, I^r(i, j)) \end{aligned}$$

trong đó $I(i, j)$ và $I^r(i, j)$ tương ứng là các mẫu dự đoán ban đầu và đã tinh chỉnh tại vị trí (i, j) .

Fig.7 thể hiện minh họa về quá trình PROF cho chế độ afin.

Vì các thông số mô hình afin và vị trí điểm ảnh so với tâm của khối con là không

đổi từ khối con này qua khối con khác, $\Delta v(i, j)$ có thể được tính toán cho khối con thứ nhất, và được tái sử dụng cho các khối con khác trong cùng CU. Đặt Δx và Δy là độ lệch theo phương ngang và theo phương dọc từ vị trí mẫu (i, j) đến tâm của khối con mà mẫu thuộc về khối con đó, $\Delta v(i, j)$ có thể được suy ra thành

$$\begin{aligned}\Delta v_x(i, j) &= c * \Delta x + d * \Delta y \\ \Delta v_y(i, j) &= e * \Delta x + f * \Delta y\end{aligned}\tag{10}$$

Dựa trên các phương trình suy diễn MV của khối con afin (6) và (7), có thể suy diễn ra được chênh lệch MV $\Delta v(i, j)$. Cụ thể, đối với mô hình afin 4 thông số,

$$\begin{cases} c = f = \frac{v_{1x} - v_{0x}}{w} \\ e = -d = \frac{v_{1y} - v_{0y}}{w} \end{cases}$$

Đối với mô hình afin 6 thông số,

$$\begin{cases} c = \frac{v_{1x} - v_{0x}}{w} \\ d = \frac{v_{2x} - v_{0x}}{h} \\ e = \frac{v_{1y} - v_{0y}}{w} \\ f = \frac{v_{2y} - v_{0y}}{h} \end{cases}$$

trong đó (v_{0x}, v_{0y}) , (v_{1x}, v_{1y}) , (v_{2x}, v_{2y}) là MV điểm điều khiển góc trên bên trái, góc trên bên phải và góc dưới bên trái của khối tạo mã hiện tại, w và h là chiều rộng và chiều cao của khối đó. Trong thiết kế PROF đã có, chênh lệch MV Δv_x và Δv_y luôn luôn được suy ra với độ chính xác 1/32-pel.

Hiệu suất tạo mã của chế độ afin

Mặc dù PROF có thể tăng cường hiệu suất tạo mã của chế độ afin, thiết kế của nó vẫn còn có thể được cải tiến. Đặc biệt, sự thật là cả PROF và BDOF đều được xây dựng dựa trên khái niệm luồng quang học nên việc cân đối các thiết kế của PROF và BDOF càng nhiều càng tốt rất được mong muốn để PROF có thể tận dụng tối đa các logic hiện có của BDOF để tạo thuận lợi cho các cách triển khai phần cứng. Dựa trên những cân nhắc như thế, các vấn đề dưới đây về tương tác giữa các thiết kế PROF và BDOF hiện

tại được xác định trong sáng chế này.

Như được mô tả trong phần “PROF cho chế độ afin,” trong phương trình (8), độ chính xác của các gradien được xác định dựa trên độ sâu bit bên trong. Mặt khác, chênh lệch MV, cụ thể là, Δv_x và Δv_y , luôn luôn được suy ra với độ chính xác 1/32-pel. Tương ứng, dựa trên phương trình (9), độ chính xác của việc tinh chỉnh PROF được suy diễn ra là phụ thuộc vào độ sâu bit bên trong. Tuy nhiên, tương tự như BDOF, PROF được áp dụng trên đỉnh của các giá trị mẫu dự đoán ở độ sâu bit cao trung gian (cụ thể là, 16-bit) để giữ được độ chính xác suy diễn PROF cao hơn. Do đó, bất kể độ sâu bit tạo mã bên trong, độ chính xác của các tinh chỉnh dự đoán suy ra bởi PROF sẽ khớp với độ chính xác của các mẫu dự đoán trung gian, cụ thể là 16-bit. Nói cách khác, độ sâu bit biểu diễn của chênh lệch MV và các gradien trong thiết kế PROF hiện có không khớp nhau hoàn toàn để suy ra được các tinh chỉnh dự đoán chính xác so với độ chính xác mẫu dự đoán (cụ thể là 16-bit). Trong khi đó, dựa trên việc so sánh các phương trình (1), (4) và (8), PROF và BDOF hiện có sử dụng các độ chính xác khác nhau để biểu diễn các gradien mẫu và chênh lệch MV. Như đã được chỉ ra trên đây, thiết kế không thống nhất như vậy là không mong muốn đối với phần cứng bởi vì logic BDOF hiện có không thể được tái sử dụng.

Như được mô tả trong phần “PROF cho chế độ afin”, khi một khối afin hiện tại được dự đoán đôi, PROF được áp dụng cho các mẫu dự đoán trong danh sách L0 và L1 tách biệt nhau; sau đó, các tín hiệu dự đoán L0 và L1 đã được tăng cường được lấy trung bình để tạo ra tín hiệu dự đoán đôi cuối cùng. Trái lại, thay vì suy diễn tinh chỉnh PROF tách biệt cho mỗi hướng dự đoán, BDOF suy diễn tinh chỉnh dự đoán một lần, tinh chỉnh này sau đó được áp dụng để tăng cường cho tín hiệu dự đoán L0 và L1 đã được kết hợp với nhau.

Các hình vẽ Fig.8 và Fig.9 (được mô tả dưới đây) so sánh tiến trình công việc của BDOF và PROF hiện tại đối với dự đoán đôi. Trong thiết kế đường ống phần cứng codec thực tế, nó thường phân công các môđun mã hóa/giải mã lớn khác nhau cho mỗi giai đoạn trong đường ống sao cho có thể xử lý song song nhiều khối tạo mã hơn. Tuy nhiên, do sự khác biệt giữa các tiến trình công việc của BDOF và PROF, điều này có thể dẫn tới việc khó có một thiết kế đường ống chung có thể chia sẻ bởi BDOF và PROF, điều này là không thân thiện đối với việc triển khai codec thực tế.

Fig.8 thể hiện tiến trình công việc của BDOF. Tiến trình công việc 800 bao gồm phần bù chuyển động L0 810, phần bù chuyển động L1 820, và BDOF 830. Ví dụ, phần bù chuyển động L0 810 có thể là danh sách các mẫu bù chuyển động lấy từ hình ảnh tham chiếu phía trước. Hình ảnh tham chiếu phía trước là hình ảnh tham chiếu đứng trước hình ảnh hiện tại trong khối video. Ví dụ, phần bù chuyển động L1 820 có thể là danh sách các mẫu bù chuyển động lấy từ hình ảnh tham chiếu tiếp theo. Hình ảnh tham chiếu tiếp theo là hình ảnh tham chiếu đứng sau hình ảnh hiện tại trong khối video. BDOF 830 lấy các mẫu bù chuyển động từ phần bù chuyển động L1 810 và phần bù chuyển động L1 820 và đưa ra các mẫu dự đoán, như được mô tả đối với Fig.4 trên đây.

Fig.9 thể hiện tiến trình công việc của PROF hiện có. Tiến trình công việc 900 bao gồm phần bù chuyển động L0 910, phần bù chuyển động L1 920, PROF L0 930, PROF L1 940, và phần lấy trung bình 960. Ví dụ, phần bù chuyển động L0 910 có thể là danh sách các mẫu bù chuyển động lấy từ hình ảnh tham chiếu phía trước. Hình ảnh tham chiếu phía trước là hình ảnh tham chiếu đứng trước hình ảnh hiện tại trong khối video. Ví dụ, phần bù chuyển động L1 920 có thể là danh sách các mẫu bù chuyển động lấy từ hình ảnh tham chiếu tiếp theo. Hình ảnh tham chiếu tiếp theo là hình ảnh tham chiếu đứng sau hình ảnh hiện tại trong khối video. PROF L0 930 lấy các mẫu bù chuyển động L0 từ phần bù chuyển động L0 910 và đưa ra các giá trị tinh chỉnh về chuyển động, như được mô tả đối với Fig.7 trên đây. PROF L1 940 lấy các mẫu bù chuyển động L1 từ phần bù chuyển động L1 920 và đưa ra các giá trị tinh chỉnh về chuyển động, như được mô tả đối với Fig.7 trên đây. Phần lấy trung bình 960 lấy trung bình các đầu ra giá trị tinh chỉnh về chuyển động của PROF L0 930 và PROF L1 940.

Với cả BDOF và PROF, các gradien đều cần được tính toán cho mỗi mẫu bên trong khối tạo mã hiện tại, điều này yêu cầu việc tạo ra một hàng/cột các mẫu dự đoán bổ sung trên mỗi phía của khối đó. Để tránh độ phức tạp tính toán bổ sung của phép nội suy mẫu, các mẫu dự đoán trong vùng đã mở rộng xung quanh khối được sao chép trực tiếp từ các mẫu tham chiếu ở các vị trí nguyên (cụ thể là không có phép nội suy). Tuy nhiên, theo thiết kế hiện có, các mẫu nguyên tại các vị trí khác nhau được lựa chọn để tạo ra các giá trị gradien của BDOF và PROF. Cụ thể, đối với BDOF, mẫu tham chiếu nguyên được đặt ở bên trái của mẫu dự đoán (đối với các gradien theo phương ngang) và bên trên mẫu dự đoán (đối với các gradien theo phương dọc) được sử dụng; đối với PROF,

mẫu tham chiếu nguyên gần với mẫu dự đoán nhất được sử dụng cho việc tính toán các gradien. Tương tự như vấn đề về việc biểu diễn độ sâu bit, phương pháp tính toán gradien không thống nhất như vậy cũng là không mong muốn đối với các cách triển khai codec phần cứng.

Như đã được chỉ ra trên đây, động lực của PROF đó là bù lại chênh lệch MV nhỏ giữa MV của mỗi mẫu và MV khối con được suy ra tại tâm của khối con mà mẫu thuộc về khối con đó. Theo thiết kế PROF hiện tại, PROF luôn luôn được gọi ra khi một khối tạo mã được dự đoán bởi chế độ affin. Tuy nhiên, như được xác định trong các phương trình (6) và (7), các MV khối con của một khối affin được suy ra từ các MV điểm điều khiển. Do đó, khi chênh lệch giữa các MV điểm điều khiển là tương đối nhỏ, các MV tại mỗi vị trí mẫu sẽ là nhất quán. Trong trường hợp như vậy, vì lợi ích của việc áp dụng PROF có thể rất giới hạn, có thể không đáng để thực hiện PROF khi cân nhắc việc đánh đổi về hiệu năng/độ phức tạp.

Cải thiện hiệu suất của chế độ affin sử dụng PROF

Trong sáng chế này, các phương pháp được cung cấp để cải thiện và đơn giản hóa thiết kế PROF hiện có để tạo thuận lợi cho các cách triển khai codec phần cứng. Cụ thể, sáng chế đặc biệt chú ý đến việc cân đối các thiết kế của BDOF và PROF để chia sẻ tối đa các logic BDOF hiện có với PROF. Nói chung, khía cạnh chính của các công nghệ được đề xuất trong sáng chế này được tổng kết như sau.

Fig.10 thể hiện phương pháp biểu diễn độ sâu bit của PROF để giải mã tín hiệu video theo nội dung của sáng chế.

Trong bước 1010, thu hình ảnh tham chiếu thứ nhất được liên kết với khối video trong tín hiệu video và MV thứ nhất, từ khối video trong hình ảnh hiện tại đến khối tham chiếu trong hình ảnh tham chiếu thứ nhất. Hình ảnh tham chiếu thứ nhất bao gồm nhiều khối video không chồng lấp nhau và ít nhất một khối video được liên kết với ít nhất một MV. Ví dụ, các hình ảnh tham chiếu có thể là các hình ảnh video lân cận với hình ảnh hiện tại đang được mã hóa.

Trong bước 1012, thu các mẫu dự đoán thứ nhất $I(i, j)$ của khối video được tạo ra từ khối tham chiếu trong ảnh tham chiếu thứ nhất. i và j có thể thể hiện tọa độ của một mẫu với khối video. Ví dụ, các mẫu dự đoán $I(i, j)$ có thể là mẫu dự đoán sử dụng MV trong danh sách L0 của hình ảnh tham chiếu phía trước trong thứ tự hiển thị.

Trong bước 1014, điều khiển các độ sâu bit bên trong của các thông số PROF bên trong. Các thông số PROF bên trong bao gồm các giá trị gradien theo phương ngang, các giá trị gradien theo phương dọc, các chênh lệch chuyển động theo phương ngang, và các chênh lệch chuyển động theo phương dọc được suy ra cho các mẫu dự đoán $I(i, j)$.

Trong bước 1016, thu các giá trị tinh chỉnh dự đoán cho các mẫu dự đoán thứ nhất $I(i, j)$ dựa trên các giá trị gradien theo phương ngang và theo phương dọc và các chênh lệch chuyển động theo phương ngang và theo phương dọc.

Trong bước 1018, thu, khi khôi video chứa MV thứ hai, các mẫu dự đoán thứ hai $I'(i, j)$ được liên kết với MV thứ hai và các giá trị tinh chỉnh dự đoán tương ứng cho các mẫu dự đoán thứ hai $I'(i, j)$ này.

Trong bước 1020, thu các mẫu dự đoán cuối cùng của khối video dựa trên sự kết hợp của các mẫu dự đoán thứ nhất $I(i, j)$, các mẫu dự đoán thứ hai $I'(i, j)$, và các giá trị tinh chỉnh dự đoán.

Thứ nhất, để cải thiện hiệu suất tạo mã của PROF trong khi vẫn đạt được một thiết kế thống nhất hơn, một phương pháp được đề xuất để thống nhất độ sâu bit biểu diễn của các gradien mẫu và chênh lệch MV được dùng bởi BDOF và PROF.

Thứ hai, để tạo thuận lợi cho thiết kế đường ống phần cứng, việc cân đối tiến trình công việc của PROF với tiến trình công việc của BDOF cho dự đoán đôi được đề xuất. Cụ thể, không giống như PROF hiện có mà nó suy diễn các tinh chỉnh dự đoán tách biệt cho L0 và L1, phương pháp được đề xuất suy diễn tinh chỉnh dự đoán một lần, suy diễn này được áp dụng cho tín hiệu dự đoán L0 và L1 đã được kết hợp với nhau.

Thứ ba, hai phương pháp được đề xuất để cân đối việc suy diễn các mẫu tham chiếu nguyên để tính các giá trị gradien được dùng bởi BDOF và PROF.

Thứ tư, để làm giảm độ phức tạp tính toán, các phương pháp kết thúc sớm được đề xuất để tùy nghi mà vô hiệu hóa quá trình PROF cho các khối tạo mã afin khi một số điều kiện được thỏa mãn.

Thiết kế biểu diễn độ sâu bit được cải thiện của các gradien PROF và chênh lệch MV

Như được phân tích trong phần “Cải thiện hiệu suất của chế độ afin sử dụng PROF”, các độ sâu bit biểu diễn của chênh lệch MV và các gradien mẫu trong PROF hiện tại không thống nhất để suy diễn được các tinh chỉnh dự đoán chính xác. Hơn nữa, độ sâu

bit biểu diễn của các gradien mẫu và chênh lệch MV là không thống nhất giữa BDOF và PROF, điều này là không thân thiện đối với phần cứng. Trong phần này, một phương pháp biểu diễn độ sâu bit cải tiến được đề xuất bằng cách mở rộng phương pháp biểu diễn độ sâu bit của BDOF thành PROF. Cụ thể là, trong phương pháp được đề xuất này, các gradient theo phương ngang và theo phương dọc tại mỗi vị trí mẫu được tính toán dưới dạng

$$\begin{aligned} g_x(i, j) &= (I(i+1, j) - I(i-1, j)) \gg \max(6, \text{bit-depth} - 6) \\ g_y(i, j) &= (I(i, j+1) - I(i, j-1)) \gg \max(6, \text{bit-depth} - 6) \end{aligned} \quad (11)$$

Thêm vào đó, giả sử rằng Δx và Δy là độ lệch theo phương ngang và theo phương dọc được biểu diễn ở độ chính xác $\frac{1}{4}$ -pel, từ một vị trí mẫu đến tâm của khối con mà mẫu này thuộc về khối con đó, chênh lệch MV của PROF $\Delta v(x, y)$ tương ứng tại vị trí mẫu này được suy ra là

$$\begin{aligned} \Delta v_x(i, j) &= (c * \Delta x + d * \Delta y) \gg (13 - dMvBits) \\ \Delta v_y(i, j) &= (e * \Delta x + f * \Delta y) \gg (13 - dMvBits) \end{aligned} \quad (12)$$

trong đó $dMvBits$ là độ sâu bit của các giá trị gradien được dùng bởi quá trình BDOF, cụ thể là $dMvBits = \max(5, (\text{bit-depth} - 7)) + 1$. Trong các phương trình (11) và (12), c, d, e và f là các thông số afin được suy ra dựa trên các MV điểm điều khiển afin. Cụ thể, đối với mô hình afin 4 thông số,

$$\begin{cases} c = f = \frac{v_{1x} - v_{0x}}{w} \\ e = -d = \frac{v_{1y} - v_{0y}}{w} \end{cases}$$

Đối với mô hình afin 6 thông số,

$$\begin{cases} c = \frac{v_{1x} - v_{0x}}{w} \\ d = \frac{v_{2x} - v_{0x}}{h} \\ e = \frac{v_{1y} - v_{0y}}{w} \\ f = \frac{v_{2y} - v_{0y}}{h} \end{cases}$$

trong đó (v_{0x}, v_{0y}) , (v_{1x}, v_{1y}) , (v_{2x}, v_{2y}) là các MV điểm điều khiển góc trên bên trái, góc trên bên phải, và góc dưới bên trái của khối tạo mã hiện tại, được biểu diễn với độ chính xác 1/16-pel, và w và h là chiều rộng và chiều cao của khối đó.

Các tiến trình công việc được cân đối của BDOF và PROF cho dự đoán đôi

Như được mô tả trên đây, khi một khối tạo mã afin được dự đoán đôi, PROF hiện tại được áp dụng một chiều. Cụ thể hơn, các tinh chỉnh mẫu PROF được suy diễn và áp dụng tách biệt cho các mẫu dự đoán trong danh sách L0 và L1. Sau đó, các tín hiệu dự đoán đã tinh chỉnh, tương ứng từ danh sách L0 và L1, được lấy trung bình để tạo ra tín hiệu dự đoán đôi cuối cùng của khối. Điều này là ngược lại với thiết kế của BDOF, trong đó các tinh chỉnh mẫu được suy ra và áp dụng cho tín hiệu dự đoán đôi. Sự khác biệt như vậy giữa các tiến trình công việc dự đoán đôi của BDOF và PROF có thể là không thân thiện đối với thiết kế đường ống codec thực tế.

Fig.11 thể hiện phương pháp biểu diễn độ sâu bit của BDOF để giải mã tín hiệu video theo nội dung của sáng chế.

Trong bước 1110, thu hình ảnh tham chiếu thứ nhất $I^{(0)}$ và hình ảnh tham chiếu thứ hai $I^{(1)}$ được liên kết với khối video. Hình ảnh tham chiếu thứ nhất $I^{(0)}$ là ở trước hình ảnh hiện tại và hình ảnh tham chiếu thứ hai $I^{(1)}$ là ở sau hình ảnh hiện tại trong thứ tự hiển thị. Ví dụ, các hình ảnh tham chiếu có thể là các hình ảnh video lân cận với hình ảnh hiện tại đang được mã hóa.

Trong bước 1112, thu các mẫu dự đoán thứ nhất $I^{(0)}(i, j)$ của khối video từ khối tham chiếu trong hình ảnh tham chiếu thứ nhất $I^{(0)}$. i và j có thể thể hiện tọa độ của một mẫu với hình ảnh hiện tại.

Trong bước 1114, thu các mẫu dự đoán thứ hai $I^{(1)}(i, j)$ của khối video từ khối tham chiếu trong hình ảnh tham chiếu thứ hai $I^{(1)}$.

Trong bước 1116, áp dụng BDOF đối với khối video dựa trên các mẫu dự đoán thứ nhất $I^{(0)}(i, j)$ và các mẫu dự đoán thứ hai $I^{(1)}(i, j)$.

Trong bước 1118, thu các giá trị gradien theo phương ngang và theo phương dọc của mẫu dự đoán thứ nhất $I^{(0)}(i, j)$ và các mẫu dự đoán thứ hai $I^{(1)}(i, j)$ dựa trên các mẫu dự đoán đã được đệm thêm.

Trong bước 1120, thu các tinh chỉnh về chuyển động cho các mẫu trong khối video

dựa trên BDOF đang được áp dụng đối với khối video và các giá trị gradien theo phương ngang và theo phương dọc.

Trong bước 1122, thu các mẫu dự đoán đôi của khối video dựa trên các tinh chỉnh chuyển động.

Để tạo thuận lợi cho thiết kế đường ống phần cứng, một phương pháp đơn giản hóa, theo nội dung của sáng chế, đó là biến đổi quá trình dự đoán đôi của PROF sao cho các tiến trình công việc của hai phương pháp tinh chỉnh dự đoán được cân đối. Cụ thể, thay vì áp dụng tách biệt việc tinh chỉnh cho mỗi hướng dự đoán, phương pháp PROF được đề xuất suy diễn các tinh chỉnh dự đoán một lần dựa trên các MV điểm điều khiển của danh sách L0 và L1; các tinh chỉnh dự đoán được suy ra này sau đó được áp dụng cho tín hiệu dự đoán L0 và L1 đã được kết hợp với nhau để tăng cường chất lượng. Cụ thể, dựa trên chênh lệch MV như được suy ra trong phương trình (12), các mẫu dự đoán đôi cuối cùng của một khối tạo mã afin được tính toán bởi phương pháp được đề xuất dưới dạng

$$\begin{aligned} pred_{PROF}(i, j) &= (I^{(0)}(i, j) + I^{(1)}(i, j) + \Delta I(i, j) + o_{offset}) \gg shift \\ \Delta I(i, j) &= (g_x(i, j) * \Delta v_x(i, j) + g_y(i, j) * \Delta v_y(i, j) + 1) \gg 1 \\ I^r(i, j) &= I(i, j) + \Delta I(i, j) \end{aligned} \quad (13)$$

trong đó $shift$ và o_{offset} là giá trị dịch phải và giá trị độ lệch được áp dụng cho các tín hiệu dự đoán L0 và L1 được kết hợp với nhau để dự đoán đôi, tương ứng bằng $(15 - \text{bit-depth})$ và $1 \ll (14 - \text{bit-depth}) + (2 \ll 13)$. Hơn nữa, như được thể hiện trên phương trình (13), thao tác xén trong thiết kế PROF hiện có (như được thể hiện trên phương trình (9)) được loại bỏ trong phương pháp được đề xuất này.

Fig.12 thể hiện minh họa về quá trình PROF khi phương pháp PROF dự đoán đôi được đề xuất này được áp dụng. Quá trình PROF 1200 bao gồm phần bù chuyển động L0 1210, phần bù chuyển động L1 1220, và PROF dự đoán đôi 1230. Ví dụ, phần bù chuyển động L0 1210 có thể là danh sách các mẫu bù chuyển động lấy từ hình ảnh tham chiếu phía trước. Hình ảnh tham chiếu phía trước là hình ảnh tham chiếu đứng trước hình ảnh hiện tại trong khối video. Ví dụ, phần bù chuyển động L1 1220 có thể là danh sách các mẫu bù chuyển động lấy từ hình ảnh tham chiếu tiếp theo. Hình ảnh tham chiếu

tiếp theo là hình ảnh tham chiếu đứng sau hình ảnh hiện tại trong khối video. PROF dự đoán đôi 1230 lấy các mẫu bù chuyển động từ phần bù chuyển động L1 1210 và phần bù chuyển động L1 1220 và đưa ra các mẫu dự đoán đôi, như được mô tả trên đây.

Fig.13 thể hiện minh họa về giai đoạn trong đường ống ví dụ khi cả BDOF và PROF được đề xuất được áp dụng. Fig.13 cho thấy lợi ích tiềm năng của phương pháp được đề xuất đối với thiết kế đường ống phần cứng. Giai đoạn trong đường ống 1300 bao gồm MV phân tích cú pháp/giải mã và các mẫu tham chiếu tìm nạp 1310, phần bù chuyển động 1320, BDOF/PROF 1330. Giai đoạn trong đường ống 1300 sẽ mã hóa các khối video BLK0, BKL1, BKL2, BKL3, và BLK4. Mỗi khối video sẽ bắt đầu trong MV phân tích cú pháp/giải mã và các mẫu tham chiếu tìm nạp 1310 và chuyển đến phần bù chuyển động 1320 và sau đó phần bù chuyển động 1320, BDOF/PROF 1330, lần lượt như vậy. Điều này có nghĩa là BLK0 sẽ không bắt đầu trong quá trình giai đoạn trong đường ống 1300 cho đến khi BLK0 di chuyển lên trên phần bù chuyển động 1320. Điều này cũng áp dụng cho tất cả các giai đoạn và các khối video khi thời gian đi từ T0 đến T1, T2, T3, và T4.

Trên Fig.13, quá trình giải mã của một khối liên ảnh chủ yếu bao gồm ba bước:

Thứ nhất, phân tích cú pháp/giải mã các MV của khối tạo mã và tìm nạp các mẫu tham chiếu.

Thứ hai, tạo ra các tín hiệu dự đoán L0 và/hoặc L1 của khối tạo mã.

Thứ ba, thực hiện tinh chỉnh trên phạm vi mẫu cho các mẫu dự đoán đôi được tạo ra dựa trên BDOF khi khối tạo mã được dự đoán bởi một chế độ phi afin hoặc PROF khi khối tạo mã được dự đoán bởi chế độ afin.

Như được thể hiện trên Fig.13, sau khi phương pháp cân đối được đề xuất được áp dụng, cả BDOF và PROF đều được áp dụng trực tiếp cho các mẫu dự đoán đôi. Đã cho trước rằng BDOF và PROF được áp dụng cho các loại khối tạo mã khác nhau (cụ thể là BDOF được áp dụng cho các khối phi afin, và PROF được áp dụng cho các khối afin), hai công cụ tạo mã này không thể bị gọi ra cùng lúc. Do đó, các quá trình giải mã tương ứng của chúng có thể được thực hiện bằng cách chia sẻ cùng giai đoạn trong đường ống. Điều này là hiệu quả hơn so với thiết kế PROF hiện có, trong đó rất khó phân công cùng giai đoạn trong đường ống cho cả BDOF và PROF do sự khác nhau về tiến trình công việc dự đoán đôi của chúng.

Trong phần mô tả trên đây, phương pháp được đề xuất chỉ xem xét việc cân đối các tiến trình công việc của BDOF và PROF. Tuy nhiên, theo các thiết kế hiện có, đơn vị thao tác cơ bản cho hai công cụ tạo mã này cũng được thực hiện ở những kích thước khác nhau. Cụ thể, đối với BDOF, một khối tạo mã được chia tách thành nhiều khối con với kích thước $W_s \times H_s$, trong đó $W_s = \min(W, 16)$ và $H_s = \min(H, 16)$, trong đó W và H là chiều rộng và chiều cao của khối tạo mã. Các thao tác của BDOF, chẳng hạn như tính toán gradien và suy diễn tinh chỉnh mẫu, được thực hiện độc lập cho mỗi khối con. Mặt khác, như được mô tả trên đây, khối tạo mã afin được chia thành các khối con 4×4 , với mỗi khối con được phân công một MV riêng rẽ được suy ra dựa trên các mô hình afin hoặc là 4 thông số hoặc là 6 thông số. Vì PROF chỉ được áp dụng cho khối afin, đơn vị thao tác cơ bản của nó là khối con 4×4 . Tương tự như vấn đề về tiến trình công việc dự đoán đôi, việc sử dụng kích thước đơn vị thao tác cơ bản khác nhau cho PROF so với BDOF cũng không thân thiện với các cách triển khai phần cứng và khiến cho BDOF và PROF khó chia sẻ chung giai đoạn trong đường ống của toàn bộ quá trình giải mã. Để giải quyết vấn đề như vậy, trong một phương án, việc thống nhất kích thước khối con của chế độ afin cho giống với kích thước khối con của BDOF được đề xuất. Cụ thể, theo phương pháp được đề xuất, nếu một khối tạo mã được tạo mã bằng chế độ afin, nó sẽ được chia tách thành các khối con với kích thước $W_s \times H_s$, trong đó $W_s = \min(W, 16)$ và $H_s = \min(H, 16)$, trong đó W và H là chiều rộng và chiều cao của khối tạo mã. Mỗi khối con được phân công một MV riêng rẽ và được coi là một đơn vị thao tác PROF độc lập. Điều đáng nói là đơn vị thao tác PROF độc lập đảm bảo rằng thao tác PROF trên đỉnh của nó được thực hiện mà không tham chiếu thông tin từ các đơn vị thao tác PROF lân cận. Cụ thể, chênh lệch MV của PROF tại một vị trí mẫu được tính là chênh lệch giữa MV tại vị trí mẫu đó và MV tại tâm của đơn vị thao tác PROF mà mẫu được đặt trong đó; các gradien được dùng bởi việc suy diễn PROF được tính bằng cách đệm thêm các mẫu dọc theo mỗi đơn vị thao tác PROF. Những lợi ích được khẳng định của phương pháp được đề xuất chủ yếu bao gồm các khía cạnh sau: 1) kiến trúc đường ống được đơn giản hóa với kích thước đơn vị thao tác cơ bản thống nhất cho cả việc bù chuyển động lẫn việc tinh chỉnh BDOF/PROF; 2) việc sử dụng băng thông bộ nhớ giảm nhờ kích thước khối con được mở rộng cho việc bù chuyển động afin; 3) độ phức tạp tính toán trên mỗi mẫu của phép nội suy mẫu phân số giảm.

Vì độ phức tạp tính toán giảm (cụ thể là mục 3), với phương pháp được đề xuất, điều kiện ràng buộc bộ lọc nội suy 6 điểm nổi hiện có cho các khối tạo mã afin có thể được loại bỏ. Thay vào đó, phép nội suy 8 điểm nổi mặc định cho các khối tạo mã phi afin cũng được sử dụng cho các khối tạo mã afin. Trong trường hợp này, độ phức tạp tính toán tổng thể vẫn có thể có lợi thế so sánh so với thiết kế PROF hiện có được dựa trên khối con 4x4 với bộ lọc nội suy 6 điểm nổi.

Việc cân đối suy diễn gradien cho BDOF và PROF

Như được mô tả trên đây, cả BDOF và PROF đều tính gradien của mỗi mẫu trong khối tạo mã hiện tại, việc này truy cập một hàng/cột các mẫu dự đoán bổ sung trên mỗi phía của khối đó. Để tránh độ phức tạp nội suy bổ sung, các mẫu dự đoán cần thiết trong vùng mở rộng xung quanh biên của khối được sao chép trực tiếp từ các mẫu tham chiếu nguyên. Tuy nhiên, như được chỉ ra trong phần “xác định vấn đề,” các mẫu nguyên tại các vị trí khác nhau được dùng để tính các giá trị gradien của BDOF và PROF.

Để đạt được một thiết kế thống nhất hơn, hai phương pháp được bọc lộ sau đây nhằm thống nhất các phương pháp suy diễn gradient được dùng bởi BDOF và PROF. Trong phương pháp thứ nhất, việc thống nhất phương pháp suy diễn gradient của PROF cho giống với phương pháp của BDOF được đề xuất. Cụ thể, với phương pháp thứ nhất, vị trí nguyên được dùng để tạo ra các mẫu dự đoán trong vùng mở rộng được xác định bằng cách làm tròn xuống vị trí mẫu phân số, cụ thể là vị trí mẫu nguyên được lựa chọn được đặt ở bên trái của vị trí mẫu phân số (đối với các gradien theo phương ngang) và ở trên vị trí mẫu phân số (đối với các gradien theo phương dọc).

Trong phương pháp thứ hai, việc thống nhất phương pháp suy diễn gradient của BDOF cho giống với phương pháp của PROF được đề xuất. Cụ thể hơn, khi áp dụng phương pháp thứ hai, mẫu tham chiếu nguyên gần với mẫu dự đoán nhất được dùng cho việc tính toán các gradien.

Fig.14 thể hiện ví dụ về việc sử dụng phương pháp suy diễn gradient của BDOF, trong đó các hình tròn rỗng biểu diễn các mẫu tham chiếu 1410 tại các vị trí nguyên, các hình tam giác biểu diễn các mẫu dự đoán phân số 1430 của khối hiện tại, và các hình tròn xám biểu diễn các mẫu tham chiếu nguyên 1420 được dùng để điền vào vùng mở rộng của khối hiện tại.

Fig.15 thể hiện ví dụ về việc sử dụng phương pháp suy diễn gradient của PROF,

trong đó các hình tròn rỗng biểu diễn các mẫu tham chiếu 1510 tại các vị trí nguyên, các hình tam giác biểu diễn các mẫu dự đoán phân số 1530 của khối hiện tại, và các hình tròn xám biểu diễn các mẫu tham chiếu nguyên 1520 được dùng để điền vào vùng mở rộng của khối hiện tại.

Fig.14 và Fig.15 minh họa các vị trí mẫu nguyên tương ứng được sử dụng cho việc suy diễn các gradien cho BDOF và PROF tương ứng khi phương pháp thứ nhất (Fig.14) và phương pháp thứ hai (Fig.15) được áp dụng. Trên các hình vẽ Fig.14 và Fig.15, các hình tròn rỗng biểu diễn các mẫu tham chiếu tại các vị trí nguyên, các hình tam giác biểu diễn các mẫu dự đoán phân số của khối hiện tại, và các hình tròn xám biểu diễn các mẫu tham chiếu nguyên được dùng để điền vào vùng mở rộng của khối hiện tại cho việc suy diễn gradient.

Việc kết thúc sớm PROF dựa trên chênh lệch MV điểm điều khiển

Theo thiết kế PROF hiện tại, PROF luôn luôn được gọi ra khi một khối tạo mã được dự đoán bởi chế độ afin. Tuy nhiên, như được xác định trong các phương trình (6) và (7), các MV khối con của một khối afin được suy ra từ các MV điểm điều khiển. Do đó, khi các chênh lệch giữa các MV điểm điều khiển là tương đối nhỏ, các MV tại mỗi vị trí mẫu sẽ là nhất quán. Trong trường hợp như vậy, lợi ích của việc áp dụng PROF có thể rất hạn chế. Do đó, để làm giảm hơn nữa độ phức tạp tính toán trung bình của PROF, việc tùy nghi mà bỏ qua việc tinh chỉnh mẫu dựa trên PROF, dựa trên chênh lệch MV tối đa giữa MV trên phạm vi mẫu và MV trên phạm vi khối con, trong một khối con 4x4, được đề xuất. Vì các giá trị của chênh lệch MV của PROF của các mẫu bên trong một khối con 4x4 là đối xứng quanh tâm khối con, chênh lệch MV tối đa của PROF theo phương ngang và theo phương dọc có thể được tính dựa trên phương trình (10) dưới dạng

$$\begin{aligned}\Delta v_x^{max} &= 6 * (c + d) \\ \Delta v_y^{max} &= 6 * (e + f)\end{aligned}\tag{14}$$

Theo nội dung của sáng chế, các phép đo khác nhau có thể được sử dụng để xác định liệu chênh lệch MV đã đủ nhỏ để bỏ qua quá trình PROF hay chưa.

Trong một ví dụ, dựa trên phương trình (14), quá trình PROF có thể được bỏ qua khi tổng của chênh lệch MV theo phương ngang tối đa tuyệt đối và chênh lệch MV theo

phương dọc tối đa tuyệt đối là nhỏ hơn một ngưỡng định trước, cụ thể là

$$|\Delta v_x^{max}| + |\Delta v_y^{max}| \leq thres \quad (15)$$

Trong một ví dụ khác, nếu giá trị tối đa của $|\Delta v_x^{max}|$ và $|\Delta v_y^{max}|$ không lớn hơn một ngưỡng, quá trình PROF có thể được bỏ qua.

$$\text{MAX}(|\Delta v_x^{max}|, |\Delta v_y^{max}|) \leq thres \quad (16)$$

MAX(a, b) là hàm trả về giá trị lớn hơn giữa các giá trị đầu vào a và b.

Ngoài hai ví dụ trên, tinh thần của sáng chế cũng có thể được áp dụng cho các trường hợp khi các phép đo khác được dùng để xác định liệu chênh lệch MV đã đủ nhỏ để bỏ qua quá trình PROF hay chưa.

Trong phương pháp trên, PROF được bỏ qua dựa trên độ lớn của chênh lệch MV. Mặt khác, ngoài chênh lệch MV, việc tinh chỉnh mẫu PROF cũng được tính toán dựa trên thông tin gradien cục bộ tại mỗi vị trí mẫu trong một khối được bù chuyển động. Đối với các khối dự đoán chứa ít các chi tiết tần số cao hơn (ví dụ như vùng phẳng), các giá trị gradien có xu hướng là nhỏ sao cho các giá trị của các tinh chỉnh mẫu được suy ra sẽ là nhỏ. Cần nhắc đến điểm này, theo một khía cạnh khác của sáng chế, việc chỉ áp dụng PROF cho các mẫu dự đoán của các khối, mà các khối này chứa đủ thông tin tần số cao, được đề xuất.

Các phép đo khác nhau có thể được sử dụng để xác định liệu khối có chứa đủ thông tin tần số cao để quá trình PROF là đáng được gọi ra cho khối đó hay không. Trong một ví dụ, quyết định được đưa ra dựa trên độ lớn trung bình (cụ thể là giá trị tuyệt đối) của các gradien của các mẫu bên trong khối dự đoán đó. Nếu độ lớn trung bình này nhỏ hơn một ngưỡng, khối dự đoán sẽ được phân loại là vùng phẳng, và không nên áp dụng PROF; nếu không, khối dự đoán sẽ được coi là chứa đủ các chi tiết tần số cao nơi mà PROF vẫn có thể áp dụng được. Trong một ví dụ khác, độ lớn tối đa của các gradien của các mẫu bên trong khối dự đoán có thể được sử dụng. Nếu độ lớn tối đa này là nhỏ hơn một ngưỡng, PROF cần được bỏ qua cho khối đó. Trong thêm một ví dụ khác, chênh lệch giữa giá trị mẫu tối đa và giá trị mẫu tối thiểu, $I_{max} - I_{min}$, của khối dự đoán có thể được dùng để xác định liệu PROF có được áp dụng cho khối đó hay không. Nếu giá trị

chênh lệch này là nhỏ hơn một ngưỡng, PROF cần được bỏ qua cho khối đó. Điều đáng lưu ý là tinh thần của sáng chế cũng áp dụng được cho các trường hợp trong đó một số phép đo khác được sử dụng để xác định liệu một khối cho trước có chứa đủ thông tin tần số cao hay không.

Fig.16 thể hiện môi trường điện toán 1610 được ghép nối với giao diện người dùng 1660. Môi trường điện toán 1610 có thể là một phần của máy chủ xử lý dữ liệu. Môi trường điện toán 1610 gồm bộ xử lý 1620, bộ nhớ 1640, và giao diện I/O 1650.

Bộ xử lý 1620 thường điều khiển các hoạt động tổng thể của môi trường điện toán 1610, chẳng hạn như các hoạt động được liên kết với bộ hiển thị, việc thu thập dữ liệu, truyền thông dữ liệu, và xử lý hình ảnh. Bộ xử lý 1620 có thể bao gồm một hoặc nhiều bộ xử lý để thực thi các lệnh để thực hiện tất cả hoặc một số bước trong số các bước trong các phương pháp đã được mô tả trên đây. Hơn nữa, bộ xử lý 1620 có thể bao gồm một hoặc nhiều môđun mà chúng tạo thuận lợi cho sự tương tác giữa bộ xử lý 1620 và các thành phần khác. Bộ xử lý có thể là đơn vị xử lý trung tâm (Central Processing Unit - CPU), bộ vi xử lý, máy chip đơn lẻ, GPU, hoặc loại tương tự.

Bộ nhớ 1640 được cấu hình để lưu trữ các loại dữ liệu khác nhau để hỗ trợ các hoạt động của môi trường điện toán 1610. Bộ nhớ 1640 có thể bao gồm phần mềm định trước 1642. Các ví dụ về dữ liệu như vậy bao gồm các lệnh cho các ứng dụng hoặc các phương pháp bất kỳ mà được vận hành trên môi trường điện toán 1610, các bộ dữ liệu video, dữ liệu hình ảnh, v.v. Bộ nhớ 1640 có thể được triển khai bằng cách sử dụng loại bất kỳ trong số các thiết bị bộ nhớ khả biến hoặc bất biến, hoặc tổ hợp của chúng, chẳng hạn như bộ nhớ truy cập ngẫu nhiên tĩnh (Static Random Access Memory - SRAM), bộ nhớ chỉ đọc lập trình được xóa được bằng điện (Electrically Erasable Programmable Read-Only Memory - EEPROM), bộ nhớ chỉ đọc lập trình được xóa được (Erasable Programmable Read-Only Memory - EPROM), bộ nhớ chỉ đọc lập trình được (Programmable Read-Only Memory - PROM), bộ nhớ chỉ đọc (Read-Only Memory - ROM), bộ nhớ từ, bộ nhớ tốc độ nhanh, đĩa từ hoặc quang.

Giao diện I/O 1650 cung cấp giao diện giữa bộ xử lý 1620 và các môđun giao diện ngoại vi, chẳng hạn như bàn phím, bánh xe bấm được (click wheel), các nút, và dạng tương tự. Các nút có thể bao gồm nhưng không bị giới hạn ở, nút home (về giao diện chính), nút khởi động quét, và nút dừng quét. Giao diện I/O 1650 có thể được ghép nối

với bộ mã hóa và bộ giải mã.

Trong một phương án, sáng chế cũng đề xuất phương tiện lưu trữ đọc được bằng máy tính phi chuyển tiếp bao gồm nhiều chương trình, chẳng hạn như được chứa trong bộ nhớ 1640, có thể thực thi được bởi bộ xử lý 1620 trong môi trường điện toán 1610, để thực hiện các phương pháp đã được mô tả trên đây. Ví dụ, phương tiện lưu trữ đọc được bằng máy tính phi chuyển tiếp có thể là ROM, RAM, CD-ROM, băng từ, đĩa mềm, thiết bị lưu trữ dữ liệu quang hoặc dạng tương tự.

Phương tiện lưu trữ đọc được bằng máy tính phi chuyển tiếp có nhiều chương trình được lưu trữ trong đó để thực thi bởi thiết bị tính toán, có một hoặc nhiều bộ xử lý, trong đó nhiều chương trình này khi được thực thi bởi một hoặc nhiều bộ xử lý này sẽ khiến cho thiết bị tính toán thực hiện phương pháp đã được mô tả trên đây cho việc dự đoán chuyển động.

Trong một phương án, môi trường điện toán 1610 có thể được triển khai với một hoặc nhiều mạch tích hợp chuyên dụng (Application Specific Integrated Circuit - ASIC), bộ xử lý tín hiệu số (Digital Signal Processor - DSP), thiết bị xử lý tín hiệu số (Digital Signal Processing Device - DSPD), thiết bị logic lập trình được (Programmable Logic Device - PLD), mảng cổng lập trình được dạng trường (Field Programmable Gate Array - FPGA), đơn vị xử lý đồ họa (Graphical Processing Unit - GPU), bộ điều khiển, bộ vi điều khiển, bộ vi xử lý, hoặc các thành phần điện tử khác, để thực hiện các phương pháp được mô tả trên đây.

Phần mô tả của sáng chế đã được đưa ra nhằm mục đích minh họa và không được dự tính để trình bày cho đến cạn kiệt hoặc giới hạn sáng chế trong dạng đã bộc lộ. Nhiều biến thể, thay đổi, và cách thức triển khai thay thế sẽ trở nên rõ ràng với những người có trình độ trung bình trong lĩnh vực kỹ thuật này đã nhận được lợi ích từ những kiến thức được thể hiện trong các phần mô tả trên đây và các hình vẽ kèm theo.

Các ví dụ được chọn và được mô tả để giải thích các nguyên tắc của sáng chế và để cho phép người có hiểu biết trung bình trong lĩnh vực kỹ thuật này hiểu sáng chế đối với các cách triển khai khác nhau và để sử dụng tốt nhất các nguyên tắc cơ bản và các cách thức triển khai khác nhau với các biến thể khác nhau cho phù hợp với mục đích sử dụng cụ thể đã dự tính. Do đó, cần hiểu rằng phạm vi của sáng chế không bị giới hạn ở các ví dụ cụ thể của các cách thức triển khai đã được bộc lộ và rằng các biến thể và các

cách thức triển khai khác được dự tính để được chứa trong phạm vi của sáng chế.

YÊU CẦU BẢO HỘ

1. Phương pháp tinh chỉnh dự đoán bằng luồng quang học (prediction refinement with optical flow - PROF) để giải mã tín hiệu video, phương pháp này bao gồm các bước:

thu hình ảnh tham chiếu thứ nhất được liên kết với khối video trong hình ảnh hiện tại trong tín hiệu video và vectơ chuyển động (motion vector - MV) thứ nhất, từ khối video đó trong hình ảnh hiện tại đến khối tham chiếu trong hình ảnh tham chiếu thứ nhất, trong đó hình ảnh tham chiếu thứ nhất bao gồm nhiều khối video không chồng lấp nhau và ít nhất một khối video được liên kết với ít nhất một MV;

thu các mẫu dự đoán thứ nhất $I(i, j)$ của khối video được tạo ra từ khối tham chiếu trong hình ảnh tham chiếu thứ nhất, trong đó i và j thể hiện tọa độ của một mẫu trong khối video;

điều khiển các độ sâu bit bên trong của các thông số PROF bên trong, trong đó các thông số PROF bên trong bao gồm các giá trị gradien theo phương ngang, các giá trị gradien theo phương dọc, các chênh lệch chuyển động theo phương ngang, và các chênh lệch chuyển động theo phương dọc được suy ra cho các mẫu dự đoán thứ nhất $I(i, j)$;

thu các giá trị tinh chỉnh dự đoán cho các mẫu dự đoán thứ nhất $I(i, j)$ dựa trên các giá trị gradien theo phương ngang và theo phương dọc và các chênh lệch chuyển động theo phương ngang và theo phương dọc;

khi khối video chứa MV thứ hai, thu các mẫu dự đoán thứ hai $I'(i, j)$ được liên kết với MV thứ hai và các giá trị tinh chỉnh dự đoán tương ứng cho các mẫu dự đoán thứ hai $I'(i, j)$ này; và

thu các mẫu dự đoán cuối cùng của khối video dựa trên các mẫu dự đoán thứ nhất $I(i, j)$, các mẫu dự đoán thứ hai $I'(i, j)$, và các giá trị tinh chỉnh dự đoán.

2. Phương pháp theo điểm 1, trong đó bước điều khiển các độ sâu bit bên trong của các thông số PROF bên trong bao gồm việc, đối với một mẫu trong khối video:

thu giá trị gradien theo phương ngang dựa trên các mẫu dự đoán của hai mẫu liền kề với một mẫu đã nêu đó theo chiều ngang và giá trị dịch thứ nhất; và

thu giá trị gradien theo phương dọc dựa trên các mẫu dự đoán của hai mẫu liền kề với một mẫu đã nêu đó theo chiều dọc và giá trị dịch thứ nhất này.

3. Phương pháp theo điểm 2, trong đó giá trị dịch thứ nhất là bằng với giá trị tối đa giữa 6 và giá trị của độ sâu bit tạo mã trừ 6.

4. Phương pháp theo điểm 2, còn bao gồm các bước, đối với một mẫu đã nêu đó trong khối video:

thu các MV điểm điều khiển của khối tạo mã chứa khối video, trong đó các MV điểm điều khiển này bao gồm các MV của các khối góc trên bên trái, góc trên bên phải, và góc dưới bên trái của khối tạo mã;

thu các thông số mô hình afin được suy ra dựa trên các MV điểm điều khiển;

thu chênh lệch MV theo phương ngang $\Delta v_x(i, j)$ và chênh lệch MV theo phương dọc $\Delta v_y(i, j)$ dựa trên các thông số afin, độ lệch theo phương ngang, và độ lệch theo phương dọc;

dịch phải chênh lệch MV theo phương ngang $\Delta v_x(i, j)$ một lượng bằng giá trị dịch thứ hai; và

dịch phải chênh lệch MV theo phương dọc $\Delta v_y(i, j)$ một lượng bằng giá trị dịch thứ hai.

5. Phương pháp theo điểm 4, trong đó giá trị dịch thứ hai là bằng 13 trừ độ sâu bit độ chính xác của các giá trị gradien.

6. Phương pháp theo điểm 5, trong đó độ sâu bit độ chính xác của các giá trị gradien là bằng với giá trị tối đa giữa 6 và giá trị của độ sâu bit tạo mã trừ 6.

7. Phương pháp theo điểm 4, trong đó khi khối video bao gồm MV thứ hai, bước thu các mẫu dự đoán cuối cùng của khối video bao gồm việc:

thu các giá trị tinh chỉnh dự đoán thứ nhất $\Delta I(i, j)$ dựa trên các giá trị gradien theo phương ngang, các chênh lệch MV theo phương ngang $\Delta v_x(i, j)$, các giá trị gradien theo phương dọc, và các chênh lệch MV theo phương dọc $\Delta v_y(i, j)$ được tạo ra cho các mẫu dự đoán thứ nhất $I(i, j)$;

thu các giá trị tinh chỉnh dự đoán thứ hai $\Delta I'(i, j)$ dựa trên các giá trị gradien theo phương ngang, các chênh lệch chuyển động theo phương ngang $\Delta v_x(i, j)$, các giá trị gradien theo phương dọc, và các chênh lệch chuyển động theo phương dọc $\Delta v_y(i, j)$ được tạo ra cho các mẫu dự đoán thứ hai $I'(i, j)$;

thu các giá trị tinh chỉnh dự đoán bằng cách lấy trung bình của các giá trị tinh chỉnh dự đoán thứ nhất $\Delta I(i, j)$ và các giá trị tinh chỉnh dự đoán thứ hai $\Delta I'(i, j)$;

thu các mẫu dự đoán đôi dựa trên tổng của các mẫu dự đoán thứ nhất $I(i, j)$, các mẫu dự đoán thứ hai $I'(i, j)$, các giá trị tinh chỉnh dự đoán; và

dịch phải tổng này một lượng bằng giá trị dịch thứ ba.

8. Phương pháp theo điểm 6, trong đó giá trị dịch thứ ba là bằng 15 trừ độ sâu bit tạo mã.

9. Phương pháp theo điểm 2, trong đó việc thu các giá trị gradien theo phương ngang và theo phương dọc của mẫu dự đoán thứ nhất $I(i, j)$ còn bao gồm bước:

suy ra một hàng hoặc cột bổ sung của các mẫu dự đoán bên ngoài khối video mà nó liền kề với mỗi biên đỉnh, trái, đáy và phải của khối video cho các mẫu dự đoán thứ nhất $I(i, j)$.

10. Phương pháp theo điểm 9, trong đó bước suy ra cột hoặc hàng bổ sung của các mẫu dự đoán còn bao gồm việc:

suy ra các mẫu dự đoán bên ngoài khối video mà chúng liền kề với các biên trái và phải từ các mẫu tham chiếu nguyên trong hình ảnh tham chiếu thứ nhất ở bên trái của vị trí mẫu phân số; và

suy ra các mẫu dự đoán bên ngoài khối video mà chúng liền kề với các biên đỉnh và đáy từ các mẫu tham chiếu nguyên trong hình ảnh tham chiếu thứ nhất ở bên trên của vị trí mẫu phân số.

11. Phương pháp theo điểm 9, trong đó bước suy ra cột hoặc hàng bổ sung của các mẫu dự đoán còn bao gồm việc:

suy ra các mẫu dự đoán bên ngoài khối video mà chúng liền kề với các biên trái và phải từ các mẫu tham chiếu nguyên trong hình ảnh tham chiếu thứ nhất ở gần các vị trí mẫu phân số tương ứng nhất theo chiều ngang; và

suy ra các mẫu dự đoán bên ngoài khối video mà chúng liền kề với các biên đỉnh và đáy từ các mẫu tham chiếu nguyên trong hình ảnh tham chiếu thứ nhất ở gần các vị trí mẫu phân số tương ứng nhất theo chiều dọc.

12. Phương pháp theo điểm 1, trong đó bước thu các giá trị tinh chỉnh dự đoán cho các mẫu dự đoán thứ nhất $I(i, j)$ bao gồm việc:

thu giá trị tối đa thứ nhất của các giá trị tuyệt đối của các giá trị chênh lệch MV theo phương ngang;

thu giá trị tối đa thứ hai của các giá trị tuyệt đối của các giá trị chênh lệch MV theo

phương dọc; và

bỏ qua việc tính toán các giá trị tinh chỉnh dự đoán khi tổng của giá trị tối đa thứ nhất và giá trị tối đa thứ hai là nhỏ hơn ngưỡng định trước.

13. Phương pháp theo điểm 1, trong đó bước thu các giá trị tinh chỉnh dự đoán cho các mẫu dự đoán thứ nhất $I(i, j)$ bao gồm việc:

thu giá trị tối đa thứ nhất của các giá trị tuyệt đối của các giá trị chênh lệch MV theo phương ngang của khối video;

thu giá trị tối đa thứ hai của các giá trị tuyệt đối của các giá trị chênh lệch MV theo phương dọc của khối video; và

bỏ qua việc tính toán các giá trị tinh chỉnh dự đoán khi giá trị tối đa của giá trị tối đa thứ nhất và giá trị tối đa thứ hai là nhỏ hơn ngưỡng định trước.

14. Phương pháp theo điểm 1, trong đó bước thu các giá trị tinh chỉnh dự đoán cho các mẫu dự đoán thứ nhất $I(i, j)$ bao gồm việc:

thu giá trị tối đa thứ nhất của các giá trị tuyệt đối của các giá trị gradien theo phương ngang của khối video;

thu giá trị tối đa thứ hai của các giá trị tuyệt đối của các giá trị gradien theo phương dọc của khối video; và

bỏ qua việc tính toán các giá trị tinh chỉnh dự đoán khi giá trị tối đa của giá trị tối đa thứ nhất và giá trị tối đa thứ hai là nhỏ hơn ngưỡng định trước.

15. Thiết bị tính toán bao gồm:

một hoặc nhiều bộ xử lý;

phương tiện lưu trữ đọc được bằng máy tính phi chuyên tiếp lưu trữ các lệnh có thể thực thi được bởi một hoặc nhiều bộ xử lý, trong đó một hoặc nhiều bộ xử lý được cấu hình để:

thu hình ảnh tham chiếu thứ nhất được liên kết với khối video trong hình ảnh hiện tại trong tín hiệu video và vectơ chuyển động (MV) thứ nhất, từ khối video trong hình ảnh hiện tại đến khối tham chiếu trong hình ảnh tham chiếu thứ nhất, trong đó hình ảnh tham chiếu thứ nhất bao gồm nhiều khối video không chồng lấp nhau và ít nhất một khối video được liên kết với ít nhất một MV;

thu các mẫu dự đoán thứ nhất $I(i, j)$ của khối video được tạo ra từ khối tham chiếu trong hình ảnh tham chiếu thứ nhất, trong đó i và j thể hiện tọa độ của một

mẫu trong khối video;

điều khiển các độ sâu bit bên trong của các thông số PROF bên trong, trong đó các thông số PROF bên trong bao gồm các giá trị gradien theo phương ngang, các giá trị gradien theo phương dọc, các chênh lệch chuyển động theo phương ngang, và các chênh lệch chuyển động theo phương dọc được suy ra cho các mẫu dự đoán thứ nhất $I(i, j)$;

thu các giá trị tinh chỉnh dự đoán cho các mẫu dự đoán thứ nhất $I(i, j)$ dựa trên các giá trị gradien theo phương ngang và theo phương dọc và các chênh lệch chuyển động theo phương ngang và theo phương dọc;

khi khối video chứa MV thứ hai, thu các mẫu dự đoán thứ hai $I'(i, j)$ được liên kết với MV thứ hai và các giá trị tinh chỉnh dự đoán tương ứng cho các mẫu dự đoán thứ hai $I'(i, j)$ này; và

thu các mẫu dự đoán cuối cùng của khối video dựa trên các mẫu dự đoán thứ nhất $I(i, j)$, các mẫu dự đoán thứ hai $I'(i, j)$, và các giá trị tinh chỉnh dự đoán.

16. Thiết bị tính toán theo điểm 15, trong đó một hoặc nhiều bộ xử lý, được cấu hình để điều khiển các độ sâu bit bên trong của các thông số PROF bên trong, còn được cấu hình để, đối với một mẫu trong khối video:

thu giá trị gradien theo phương ngang dựa trên các mẫu dự đoán của hai mẫu liên kế với một mẫu đã cho đó theo chiều ngang và giá trị dịch thứ nhất; và

thu giá trị gradien theo phương dọc dựa trên các mẫu dự đoán của hai mẫu liên kế với một mẫu đã cho đó theo chiều dọc và giá trị dịch thứ nhất này.

17. Thiết bị tính toán theo điểm 16, trong đó giá trị dịch thứ nhất là bằng với giá trị tối đa giữa 6 và giá trị của độ sâu bit tạo mã trừ 6.

18. Thiết bị tính toán theo điểm 16, trong đó một hoặc nhiều bộ xử lý còn được cấu hình để, đối với một mẫu đã cho đó trong khối video:

thu các MV điểm điều khiển của khối tạo mã chứa khối video, trong đó các MV điểm điều khiển này bao gồm các MV của các khối góc trên bên trái, góc trên bên phải, và góc dưới bên trái của khối tạo mã;

thu các thông số mô hình afin được suy ra dựa trên các MV điểm điều khiển

thu chênh lệch MV theo phương ngang $\Delta v_x(i, j)$ và chênh lệch MV theo phương dọc $\Delta v_y(i, j)$ dựa trên các thông số afin, độ lệch theo phương ngang, và độ lệch theo

phương dọc;

dịch phải chênh lệch MV theo phương ngang $\Delta v_x(i, j)$ một lượng bằng giá trị dịch thứ hai; và

dịch phải chênh lệch MV theo phương dọc $\Delta v_y(i, j)$ một lượng bằng giá trị dịch thứ hai.

19. Thiết bị tính toán theo điểm 18, trong đó giá trị dịch thứ hai là bằng 13 trừ độ sâu bit độ chính xác của các giá trị gradien.

20. Thiết bị tính toán theo điểm 19, trong đó độ sâu bit độ chính xác của các giá trị gradien là bằng với giá trị tối đa giữa 6 và giá trị của độ sâu bit tạo mã trừ 6.

21. Thiết bị tính toán theo điểm 18, trong đó một hoặc nhiều bộ xử lý, được cấu hình để thu, khi khối video bao gồm MV thứ hai, các mẫu dự đoán cuối cùng của khối video, còn được cấu hình để:

thu các giá trị tinh chỉnh dự đoán thứ nhất $\Delta I(i, j)$ dựa trên các giá trị gradien theo phương ngang, các chênh lệch MV theo phương ngang $\Delta v_x(i, j)$, các giá trị gradien theo phương dọc, và các chênh lệch MV theo phương dọc $\Delta v_y(i, j)$ được tạo ra cho các mẫu dự đoán thứ nhất $I(i, j)$;

thu các giá trị tinh chỉnh dự đoán thứ hai $\Delta I'(i, j)$ dựa trên các giá trị gradien theo phương ngang, các chênh lệch chuyển động theo phương ngang $\Delta v_x(i, j)$, các giá trị gradien theo phương dọc, và các chênh lệch chuyển động theo phương dọc $\Delta v_y(i, j)$ được tạo ra cho các mẫu dự đoán thứ hai $I'(i, j)$

thu các giá trị tinh chỉnh dự đoán bằng cách lấy trung bình của các giá trị tinh chỉnh dự đoán thứ nhất $\Delta I(i, j)$ và các giá trị tinh chỉnh dự đoán thứ hai $\Delta I'(i, j)$;

thu các mẫu dự đoán đôi dựa trên tổng của các mẫu dự đoán thứ nhất $I(i, j)$, các mẫu dự đoán thứ hai $I'(i, j)$, các giá trị tinh chỉnh dự đoán; và

dịch phải tổng này một lượng bằng giá trị dịch thứ ba.

22. Thiết bị tính toán theo điểm 20, trong đó giá trị dịch thứ ba là bằng 15 trừ độ sâu bit tạo mã.

23. Thiết bị tính toán theo điểm 16, trong đó một hoặc nhiều bộ xử lý, được cấu hình để thu các giá trị gradien theo phương ngang và theo phương dọc của mẫu dự đoán thứ nhất $I(i, j)$, còn được cấu hình để:

suy ra một hàng hoặc cột bổ sung của các mẫu dự đoán bên ngoài khối video mà

nó liền kề với mỗi biên đỉnh, trái, đáy và phải của khối video cho các mẫu dự đoán thứ nhất $I(i, j)$.

24. Thiết bị tính toán theo điểm 23, trong đó một hoặc nhiều bộ xử lý, được cấu hình để suy ra một hàng hoặc cột bổ sung của các mẫu dự đoán, còn được cấu hình để:

suy ra các mẫu dự đoán bên ngoài khối video mà chúng liền kề với các biên trái và phải từ các mẫu tham chiếu nguyên trong hình ảnh tham chiếu thứ nhất ở bên trái của vị trí mẫu phân số; và

suy ra các mẫu dự đoán bên ngoài khối video mà chúng liền kề với các biên đỉnh và đáy từ các mẫu tham chiếu nguyên trong hình ảnh tham chiếu thứ nhất ở bên trên của vị trí mẫu phân số.

25. Thiết bị tính toán theo điểm 23, trong đó một hoặc nhiều bộ xử lý, được cấu hình để suy ra một hàng hoặc cột bổ sung của các mẫu dự đoán, còn được cấu hình để:

suy ra các mẫu dự đoán bên ngoài khối video mà chúng liền kề với các biên trái và phải từ các mẫu tham chiếu nguyên trong hình ảnh tham chiếu thứ nhất ở gần các vị trí mẫu phân số tương ứng nhất theo chiều ngang; và

suy ra các mẫu dự đoán bên ngoài khối video mà chúng liền kề với các biên đỉnh và đáy từ các mẫu tham chiếu nguyên trong hình ảnh tham chiếu thứ nhất ở gần các vị trí mẫu phân số tương ứng nhất theo chiều dọc.

26. Thiết bị tính toán theo điểm 15, trong đó một hoặc nhiều bộ xử lý, được cấu hình để thu các giá trị tinh chỉnh dự đoán cho các mẫu dự đoán thứ nhất $I(i, j)$, còn được cấu hình để:

thu giá trị tối đa thứ nhất của các giá trị tuyệt đối của các giá trị chênh lệch MV theo phương ngang;

thu giá trị tối đa thứ hai của các giá trị tuyệt đối của các giá trị chênh lệch MV theo phương dọc; và

bỏ qua việc tính toán các giá trị tinh chỉnh dự đoán khi tổng của giá trị tối đa thứ nhất và giá trị tối đa thứ hai là nhỏ hơn ngưỡng định trước.

27. Thiết bị tính toán theo điểm 15, trong đó một hoặc nhiều bộ xử lý, được cấu hình để thu các giá trị tinh chỉnh dự đoán cho các mẫu dự đoán thứ nhất $I(i, j)$, còn được cấu hình để:

thu giá trị tối đa thứ nhất của các giá trị tuyệt đối của các giá trị chênh lệch MV

theo phương ngang của khối video;

thu giá trị tối đa thứ hai của các giá trị tuyệt đối của các giá trị chênh lệch MV theo phương dọc của khối video; và

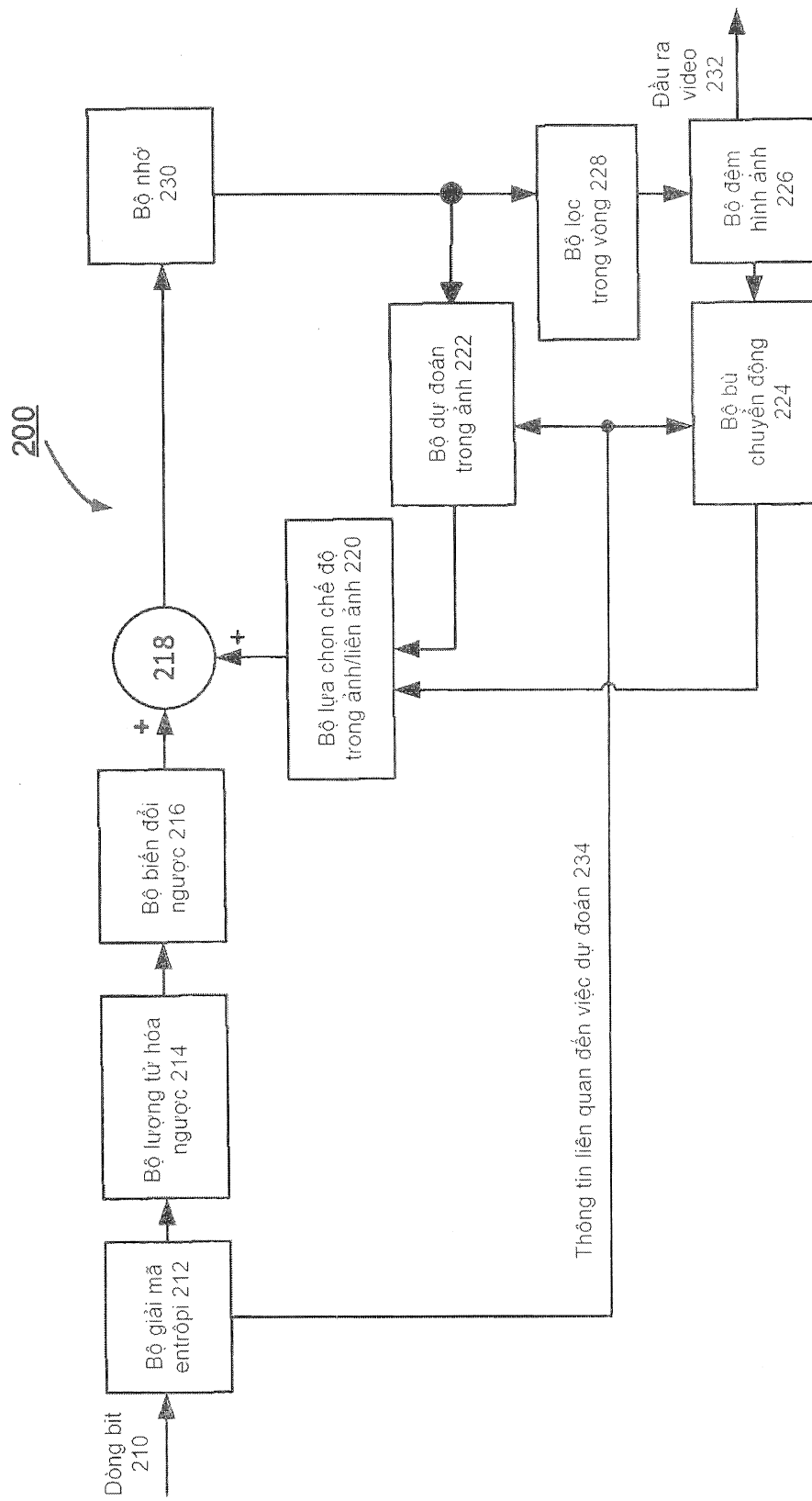
bỏ qua việc tính toán các giá trị tinh chỉnh dự đoán khi giá trị tối đa của giá trị tối đa thứ nhất và giá trị tối đa thứ hai là nhỏ hơn ngưỡng định trước.

28. Thiết bị tính toán theo điểm 15, trong đó một hoặc nhiều bộ xử lý, được cấu hình để thu các giá trị tinh chỉnh dự đoán cho các mẫu dự đoán thứ nhất $I(i, j)$, còn được cấu hình để:

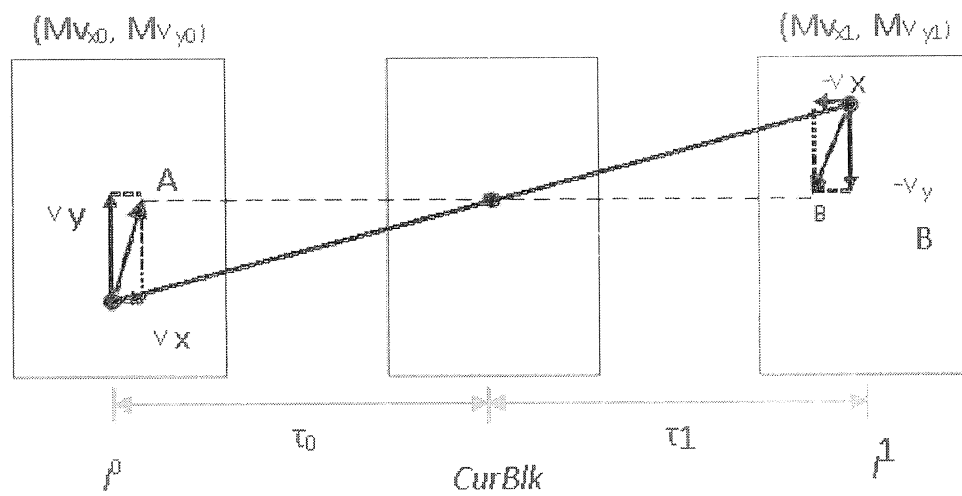
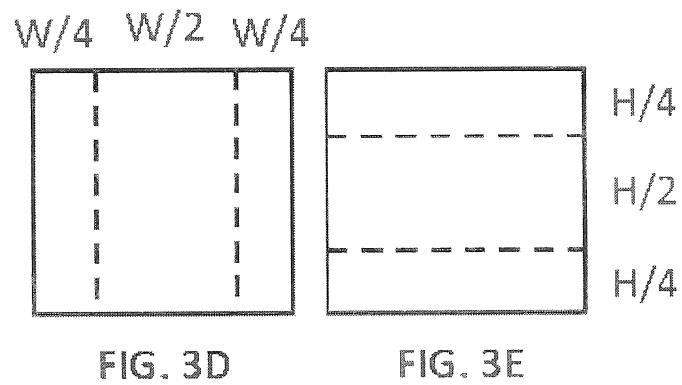
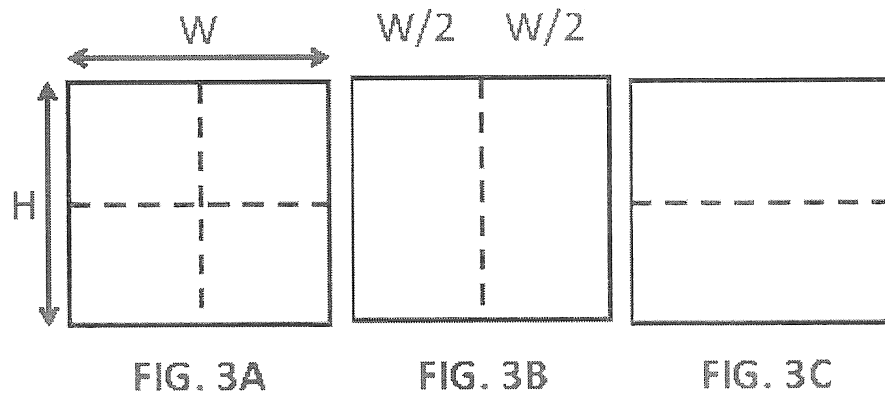
thu giá trị tối đa thứ nhất của các giá trị tuyệt đối của các giá trị gradien theo phương ngang của khối video;

thu giá trị tối đa thứ hai của các giá trị tuyệt đối của các giá trị gradien theo phương dọc của khối video; và

bỏ qua việc tính toán các giá trị tinh chỉnh dự đoán khi giá trị tối đa của giá trị tối đa thứ nhất và giá trị tối đa thứ hai là nhỏ hơn ngưỡng định trước.



254



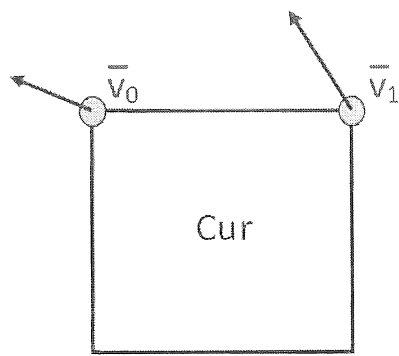


FIG. 5A

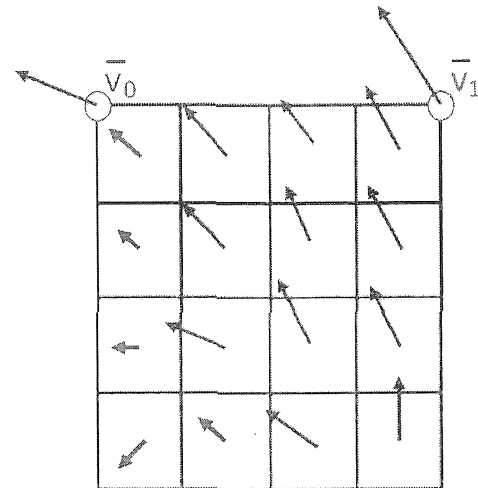


FIG. 5B

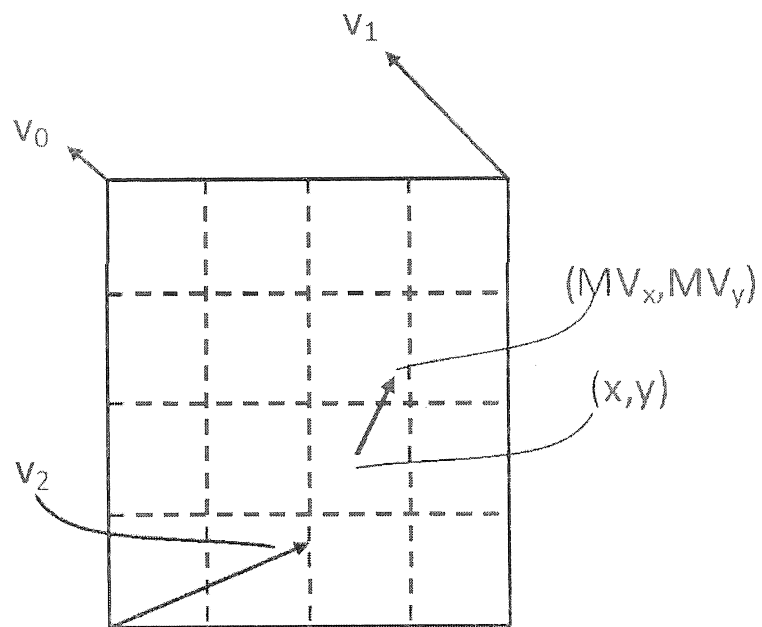


FIG. 6

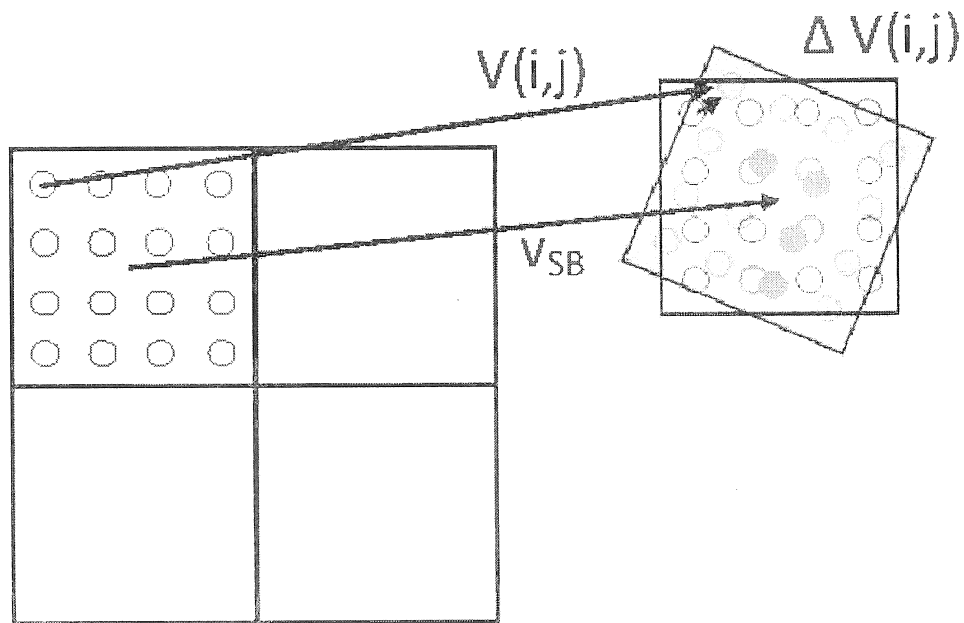


FIG. 7

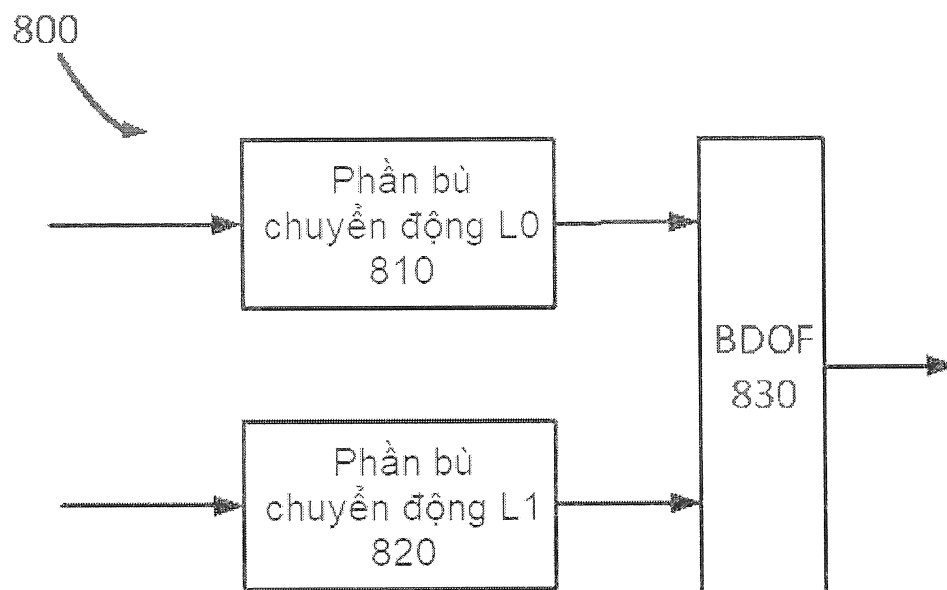


FIG. 8

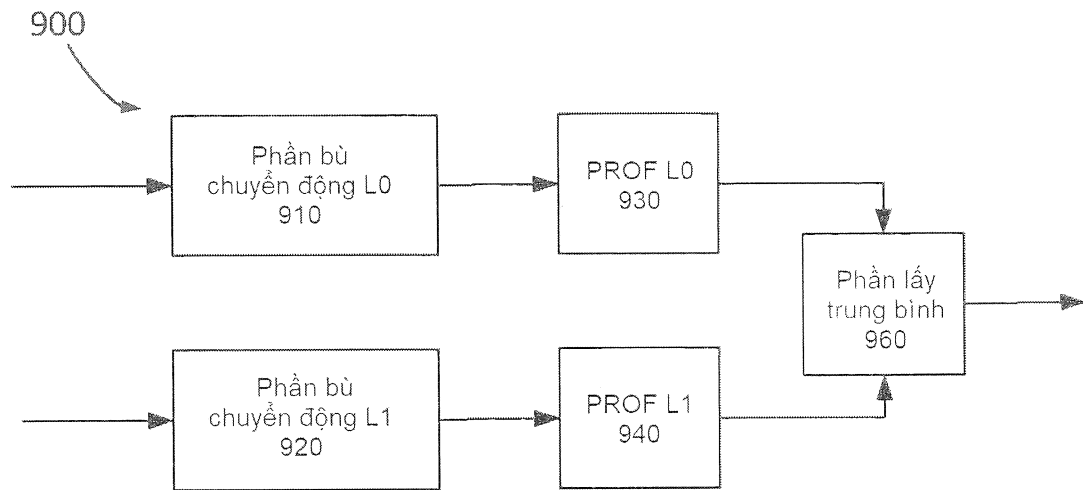


FIG. 9

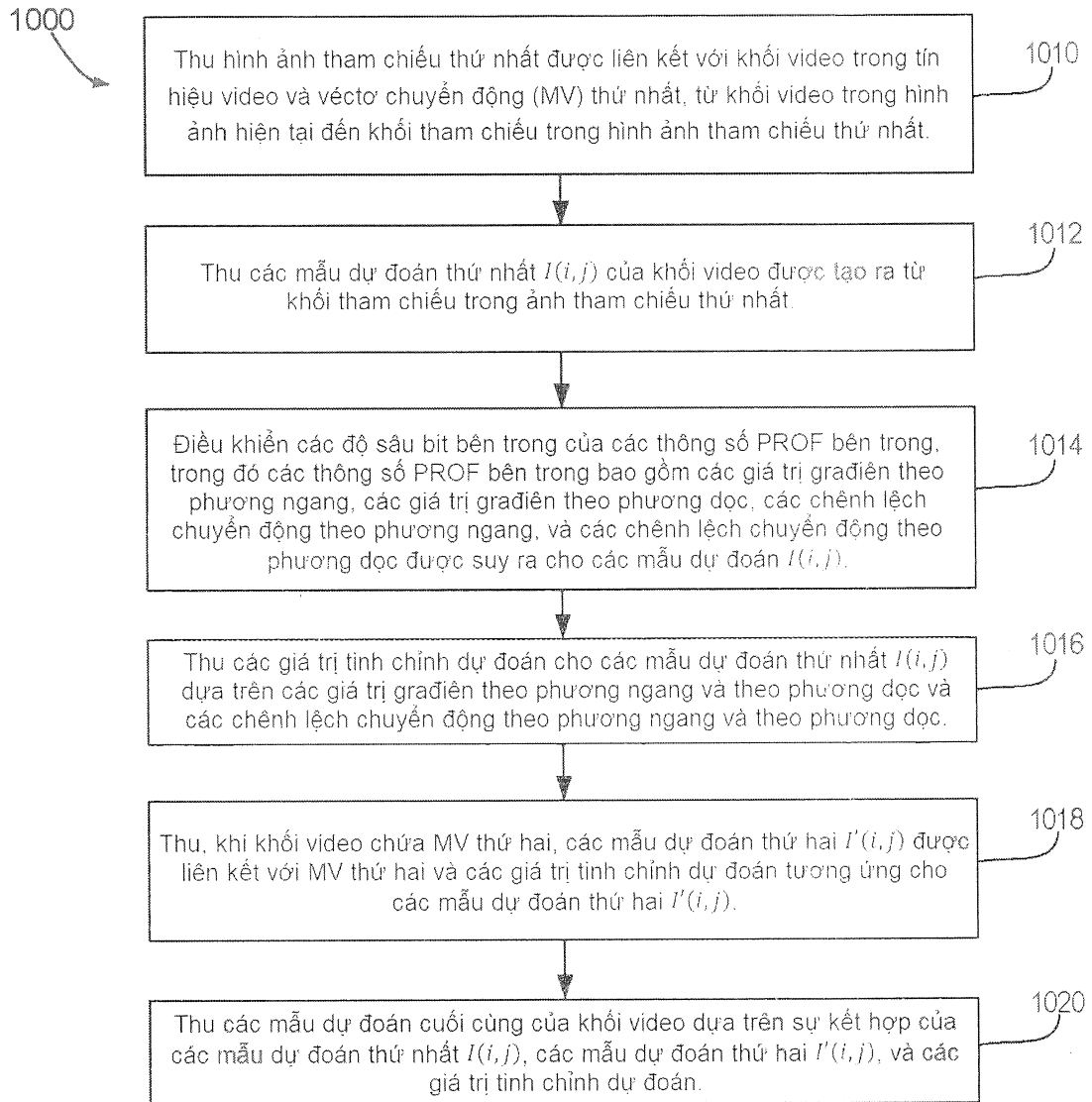


FIG. 10

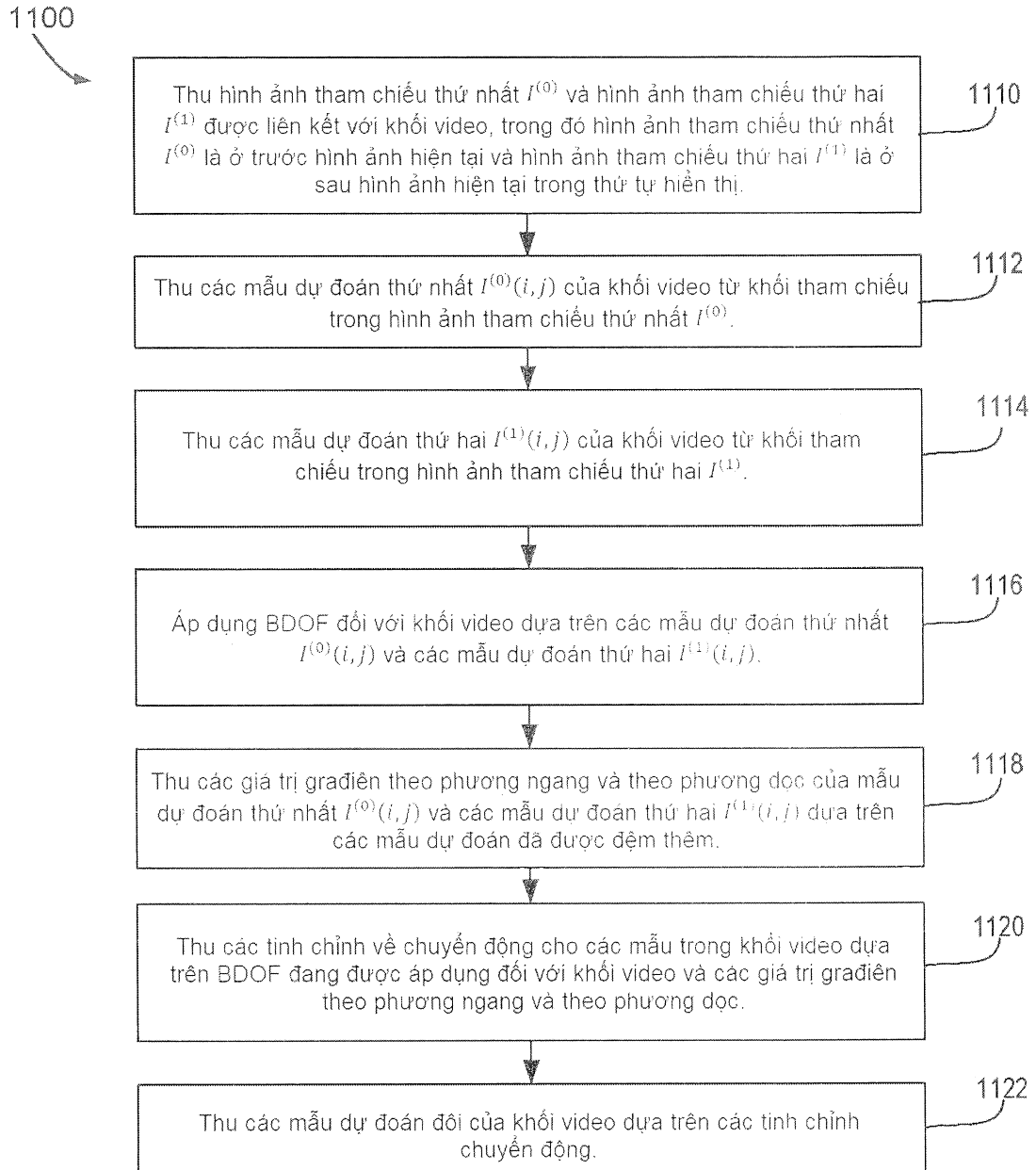


FIG. 11

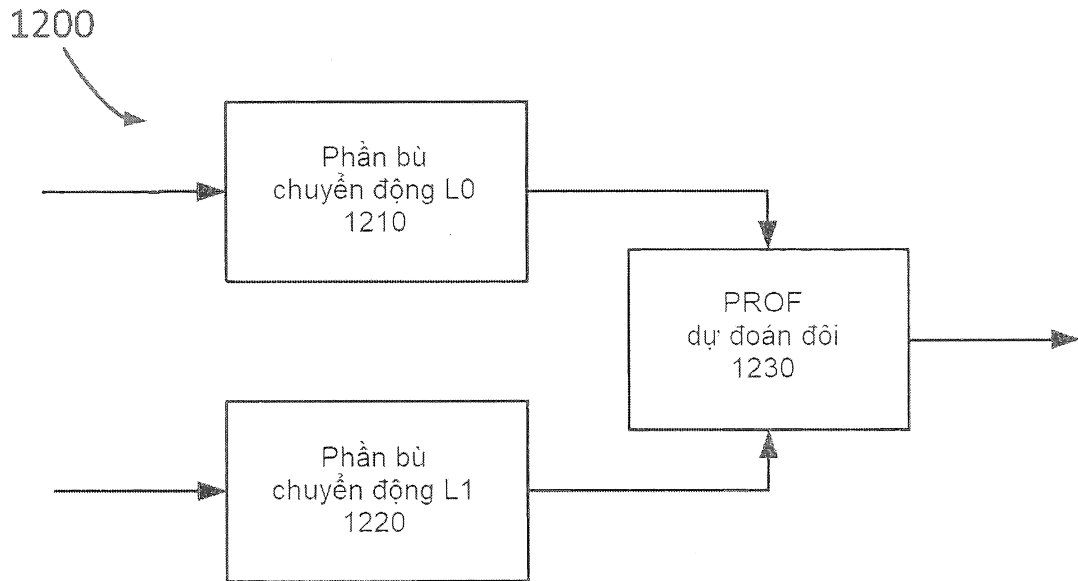


FIG. 12

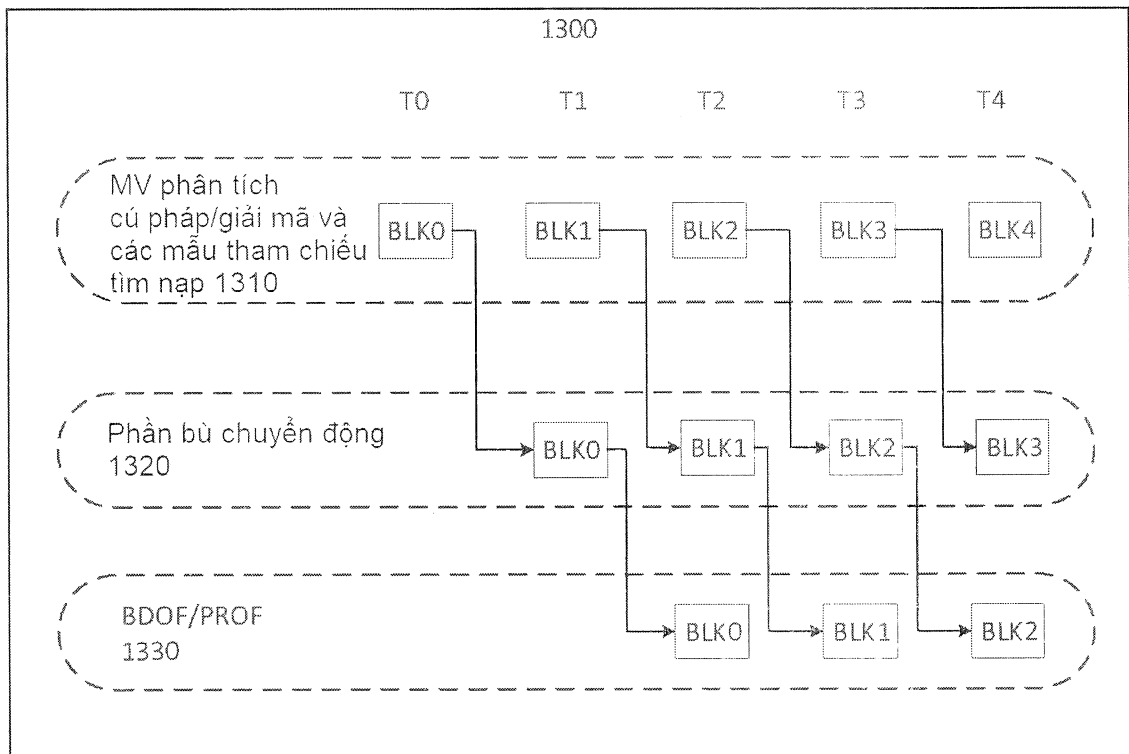
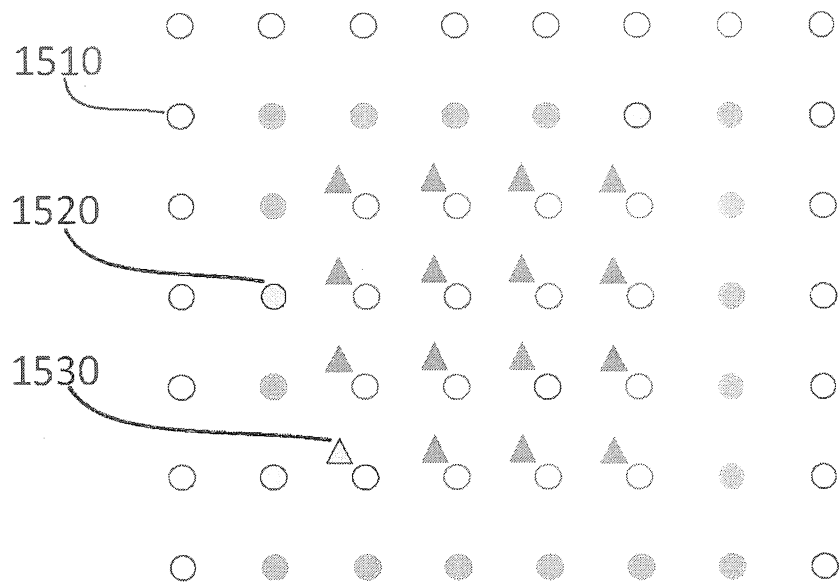
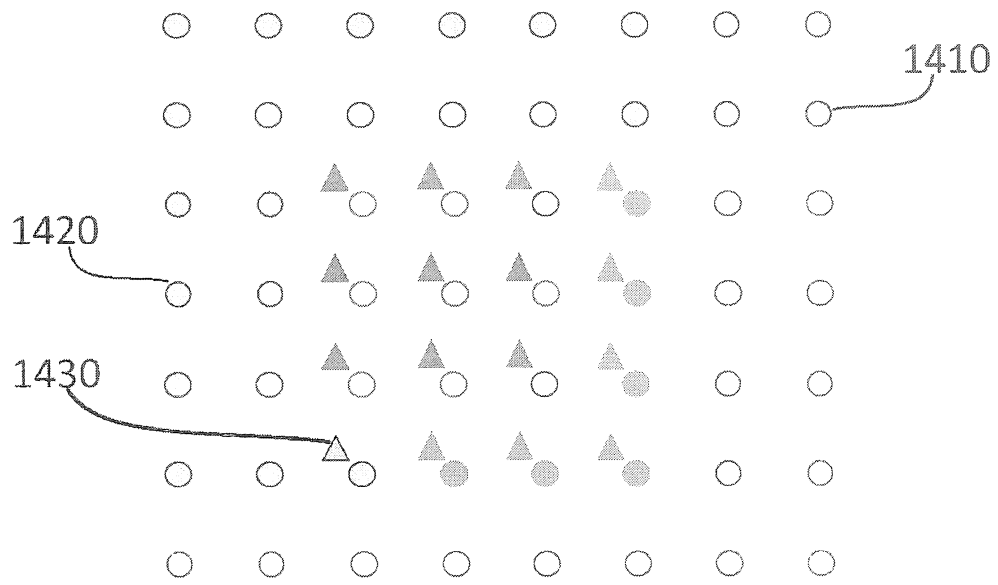
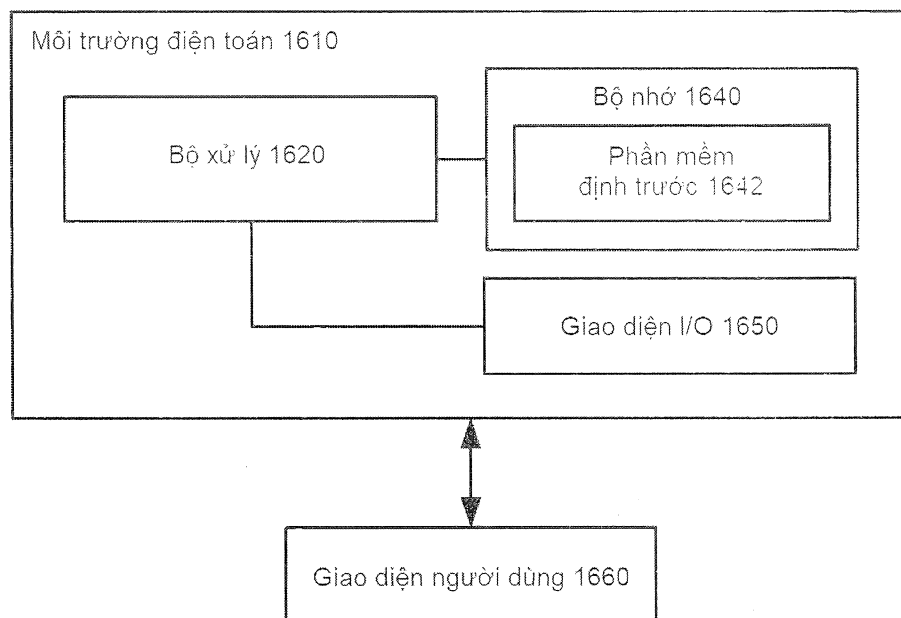


FIG. 13



**FIG. 16**