



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

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

(11)



1-0049753

(51)^{2020.01} H04N 21/2343; H04N 21/81; H04N 13/161; H04N 13/178; H04N 13/243; H04N 19/154; H04N 19/59; H04N 19/597; H04N 21/218; H04N 21/2365; H04N 21/2662; H04N 21/4402; H04N 21/4728; G06N 3/08; H04N 13/139

(13) B

(21) 1-2022-02523

(22) 18/09/2020

(86) PCT/EP2020/076197 18/09/2020

(87) WO 2021/058402 01/04/2021

(30) 19199240.3 24/09/2019 EP

(45) 25/08/2025 449

(43) 25/07/2022 412A

(73) KONINKLIJKE PHILIPS N.V. (NL)

High Tech Campus 52, 5656 AG Eindhoven, Netherlands

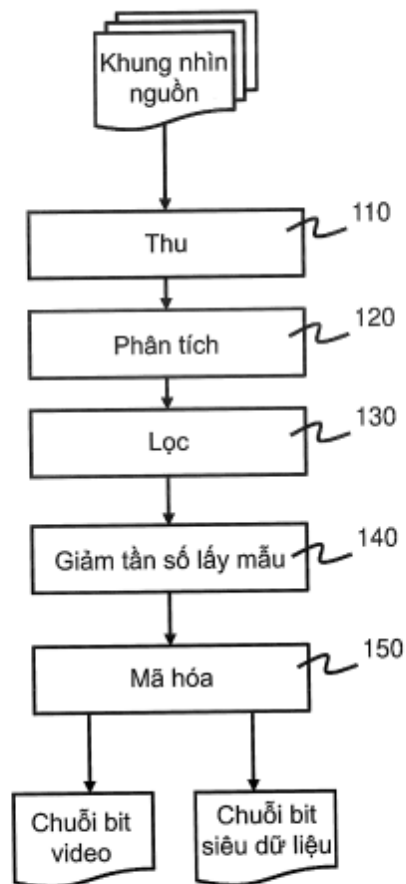
(72) KROON, Bart (NL); VAREKAMP, Christiaan (NL).

(74) Công ty Luật TNHH ROUSE Việt Nam (ROUSE LEGAL VIETNAM LTD.)

(54) PHƯƠNG PHÁP MÃ HÓA VÀ GIẢI MÃ VIDEO SỐNG ĐỘNG, BỘ MÃ HÓA VÀ BỘ GIẢI MÃ DÀNH CHO VIDEO SỐNG ĐỘNG, VÀ PHƯƠNG TIỆN ĐỌC ĐƯỢC BẰNG MÁY TÍNH

(21) 1-2022-02523

(57) Sáng chế đề xuất các phương pháp mã hóa và giải mã video sống động. Theo một phương pháp mã hóa, dữ liệu video nguồn bao gồm nhiều khung nhìn nguồn được mã hóa thành chuỗi bit video. Ít nhất một trong số các khung nhìn nguồn này được giảm tần số lấy mẫu trước khi mã hóa. Chuỗi bit siêu dữ liệu liên quan đến luồng video bao gồm siêu dữ liệu mô tả cấu hình giảm tần số lấy mẫu, để hỗ trợ bộ giải mã giải mã chuỗi bit video. Có thể tin rằng việc sử dụng khung nhìn được giảm tần số lấy mẫu có thể giúp giảm thành phần lậ mã hóa, so với phương pháp mã hóa dựa trên đoạn nối tạm. Sáng chế cũng đề xuất bộ mã hóa và bộ giải mã dành cho video sống động.



HÌNH 1

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến mã hóa video. Cụ thể, sáng chế đề cập đến phương pháp và thiết bị mã hóa và giải mã video sống động(immersive video).

Tình trạng kỹ thuật của sáng chế

Video sống động, còn được gọi là video sáu bậc tự do (6DoF), là video về cảnh ba chiều (3D) cho phép các khung nhìn của cảnh được dựng lại cho các điểm nhìn khác nhau về vị trí và hướng. Nó đại diện cho sự phát triển của video ba bậc tự do (3DoF), cho phép các khung nhìn được tái tạo cho các điểm nhìn với hướng tùy ý, nhưng chỉ tại một điểm cố định trong không gian. Trong 3DoF, bậc tự do là góc – cụ thể là dọc xuống (pitch), nghiêng (roll) và đảo lại (yaw). Video 3DoF hỗ trợ xoay đầu – nói cách khác, người dùng xem nội dung video có thể nhìn theo bất kỳ hướng nào trong cảnh, nhưng không thể di chuyển đến một nơi khác trong cảnh. Video 6DoF hỗ trợ xoay đầu và bổ sung hỗ trợ lựa chọn vị trí trong cảnh mà cảnh đó được xem.

Để tạo video 6DoF, cần có nhiều camera để ghi lại cảnh. Mỗi camera tạo ra dữ liệu hình ảnh (thường được gọi là dữ liệu kết cấu – texture data, trong ngữ cảnh này) và dữ liệu độ sâu tương ứng. Đối với mỗi pixel, dữ liệu độ sâu biểu thị độ sâu mà tại đó dữ liệu pixel hình ảnh tương ứng được quan sát, bằng camera cấp sẵn. Mỗi camera trong số nhiều camera cung cấp khung nhìn cảnh tương ứng. Phát tất cả dữ liệu kết cấu và dữ liệu độ sâu cho tất cả các khung nhìn có thể không thực tế hoặc không hiệu quả, trong nhiều ứng dụng.

Để giảm bớt sự dư thừa giữa các khung nhìn, người ta đề xuất cắt bớt các khung nhìn và đóng gói các khung nhìn này thành “tập bản ảnh kết cấu”, cho mỗi khung hình của luồng video. Phương pháp này cố gắng giảm hoặc loại bỏ các phần chồng lấn giữa nhiều khung nhìn và do đó cải thiện hiệu quả. Các phần không chồng lấn của các khung nhìn khác nhau, vẫn còn lại sau khi cắt, có thể được gọi là "mảnh". Ví dụ về phương pháp này được mô tả trong Alvaro Collet và cộng sự, “Video điểm nhìn tự do có thể phát chất lượng cao”, ACM Trans. Graphics (SIGGRAPH), 34(4), 2015.

Bản chất kỹ thuật của sáng chế

Điều mong muốn là cải thiện chất lượng và hiệu quả mã hóa video sống động. Phương pháp sử dụng cắt tia để tạo ra tập bản đồ kết cấu, như được mô tả ở trên, gặp vấn đề là các cạnh đoạn nối tạm có thể làm phát sinh các thành phần lạ có thể nhìn thấy trong các khung nhìn tái cấu trúc được kết xuất sau khi giải mã xong. Điều này đặc biệt xảy ra khi các cạnh đoạn nối tạm không thẳng hàng với các cạnh tự nhiên trong nội dung – ví dụ, nếu các đoạn nối tạm được chọn là hình chữ nhật. Các đoạn nối tạm hình chữ nhật có thể làm tăng các ranh giới hình zig-zag như một thành phần lạ có thể nhìn thấy trong đầu ra đã giải mã. Các điểm nổi bật cụ thể có thể gây ra cho các phương pháp mã hóa này khó khăn nhất định, vì vị trí của điểm nổi bật thay đổi theo từng khung nhìn, mặc dù vị trí của vật thể không thay đổi. Nếu một khung nhìn của vật thể được tái cấu trúc từ nhiều đoạn nối tạm được lấy từ nhiều khung nhìn, thì sự thay đổi giữa các khung nhìn có thể đặc biệt dễ nhìn thấy trong vùng lân cận của điểm nổi bật. Có thể nhìn thấy được điều này trong đầu ra đã giải mã.

Sáng chế được định nghĩa bởi các điểm yêu cầu bảo hộ.

Theo các ví dụ phù hợp với một phương án của sáng chế, phương pháp mã hóa video sống động được đề xuất bao gồm các bước:

bước thu dữ liệu video nguồn, bao gồm nhiều khung nhìn nguồn, mỗi khung nhìn nguồn bao gồm dữ liệu kết cấu và dữ liệu độ sâu;

bước xử lý dữ liệu video nguồn để tạo ra dữ liệu video nguồn đã xử lý, bước xử lý này bao gồm bước giảm tần số lấy mẫu một hoặc nhiều khung nhìn nguồn; và

bước mã hóa dữ liệu video nguồn đã xử lý để tạo ra một chuỗi bit video và chuỗi bit siêu dữ liệu,

trong đó chuỗi bit video bao gồm dữ liệu kết cấu được mã hóa và dữ liệu độ sâu được mã hóa, và chuỗi bit siêu dữ liệu bao gồm siêu dữ liệu để hỗ trợ bộ giải mã giải mã chuỗi bit video, trong đó siêu dữ liệu này mô tả cấu hình của bước xử lý được thực hiện trên một hoặc nhiều khung nhìn nguồn.

Theo các phương án của phương pháp này, có thể tránh việc cần cắt tia khung nhìn và tạo các đoạn nối tạm. Thay vì cắt tia và chuyển đổi khung nhìn nguồn thành đoạn nối tạm, mỗi khung nhìn nguồn có thể được mã hóa toàn bộ – nhưng với một hoặc nhiều

khung nhìn nguồn được lấy mẫu con (subsample) để giảm dư thừa và từ đó giảm tốc độ dữ liệu.

Dữ liệu kết cấu có thể bao gồm ít nhất thông tin về luma (độ sáng hoặc độ mạnh). Dữ liệu này cũng có thể bao gồm thông tin về chroma (màu sắc). Dữ liệu độ sâu có thể bao gồm dữ liệu độ sâu được chuẩn hóa, ví dụ bao gồm các giá trị độ sâu được chuẩn hóa được tính là $1/Z$, trong đó Z là độ sâu từ điểm tham chiếu hoặc mặt phẳng tham chiếu được xác định trước. Trong mỗi khung nhìn, dữ liệu kết cấu và/hoặc dữ liệu độ sâu có thể bao gồm mảng 2D gồm các pixel.

Quá trình xử lý này bao gồm bước giảm tần số lấy mẫu một hoặc nhiều khung nhìn nguồn. Mỗi khung nhìn nguồn thường bao gồm một mảng pixel hai chiều. Giảm tần số lấy mẫu nghĩa là kích thước của mảng được giảm xuống (trong một hoặc cả hai chiều) bằng cách giảm độ phân giải của khung nhìn (theo chiều tương ứng).

Ngoài bước giảm tần số lấy mẫu, quá trình xử lý này còn có thể bao gồm bước lọc. Ví dụ, quá trình xử lý có thể bao gồm bước lọc chống răng cưa, sau đó là giảm tần số lấy mẫu. Siêu dữ liệu này có thể mô tả cấu hình giảm tần số lấy mẫu, cấu hình lọc, hoặc cấu hình của cả lọc và giảm tần số lấy mẫu.

Bước lọc và giảm tần số lấy mẫu một hoặc nhiều khung nhìn nguồn có thể bao gồm bước lọc và giảm tần số lấy mẫu dữ liệu kết cấu hoặc dữ liệu độ sâu, hoặc cả hai dữ liệu. Bước giảm tần số lấy mẫu có thể được mô tả theo hệ số chia tỷ lệ – ví dụ, giảm tần số lấy mẫu theo hệ số 2, 4, hoặc 8 theo mỗi chiều. Lưu ý rằng dữ liệu kết cấu có thể được giảm tần số lấy mẫu theo cùng hệ số như dữ liệu độ sâu hoặc theo hệ số khác.

Bước mã hóa dữ liệu video nguồn đã xử lý có thể bao gồm phần nén bị mất của dữ liệu kết cấu và/hoặc dữ liệu độ sâu.

Siêu dữ liệu này có thể còn bao gồm các thông số của camera liên quan đến nhiều khung nhìn nguồn. Nói chung, thông số camera có thể được chỉ định theo cách làm cho chúng không phụ thuộc vào độ phân giải hình ảnh. Ví dụ, các thông số camera này có thể chỉ định tiêu cự liên quan đến chiều rộng cảm biến, hoặc có thể sử dụng tọa độ hình ảnh chuẩn hóa. Hoặc (ví dụ, khi tiêu cự được chỉ định liên quan đến cao độ pixel – còn được gọi là “đơn vị pixel”), các thông số camera này có thể liên quan đến hình ảnh nguồn đầy đủ độ phân giải hoặc hình ảnh được giảm tần số lấy mẫu. Thông thường, kích thước ảnh được xác định từ chuỗi bit video, nhưng cũng có thể là một phần của siêu dữ liệu. Kích thước ảnh cũng có thể là một phần của tập hợp thông số của camera.

Trong siêu dữ liệu, hệ số giảm tần số lấy mẫu có thể được chỉ định ngầm bằng cách xác định hai độ phân giải hình ảnh khác nhau (một trong chuỗi bit video và một trong các thông số camera). Bộ giải mã có thể hiểu theo thiết kế/đặc điểm kỹ thuật rằng cần tính toán hệ số tỷ lệ bằng cách so sánh các độ phân giải này.

Nói chung, bộ mã hóa xuất ra siêu dữ liệu cho phép bộ giải mã xác định lượng giảm tần số lấy mẫu đã được thực hiện.

Phương pháp này có thể bao gồm thêm bước phân tích dữ liệu video nguồn để chọn cấu hình cho quá trình xử lý, quá trình xử lý này bao gồm bước giảm tần số lấy mẫu (và có thể cả bước lọc) một hoặc nhiều khung nhìn nguồn theo cấu hình đã chọn.

Điều này có thể cho phép giảm tần số lấy mẫu (và có thể cả bước lọc) được điều chỉnh cho phù hợp với nội dung của dữ liệu video nguồn. Bước phân tích này có thể dựa trên các thông số của camera liên quan đến nhiều khung nhìn nguồn chẳng hạn. Hoặc hay ngoài ra, bước phân tích có thể dựa trên sự tương đồng (tức là dư thừa) nội dung trực quan giữa các khung nhìn.

Theo một số phương án, bước phân tích này được thực hiện trước khi xử lý dữ liệu video nguồn. Theo một số phương án khác, bước phân tích này liên quan đến hoặc phụ thuộc vào bước xử lý (lọc và giảm tần số lấy mẫu).

Bước phân tích này còn có thể bao gồm: bước phân tích sự tương đồng của nội dung trực quan giữa các khung nhìn; bước gán hệ số giảm tần số lấy mẫu thứ nhất cho khung nhìn thứ nhất; bước gán hệ số giảm tần số lấy mẫu thứ hai cho khung nhìn thứ hai; và bước gán hệ số giảm tần số lấy mẫu thứ ba cho khung nhìn thứ ba, trong đó hệ số giảm tần số lấy mẫu thứ hai cao hơn hệ số giảm tần số lấy mẫu thứ ba nếu sự tương đồng giữa khung nhìn thứ nhất và khung nhìn thứ hai lớn hơn sự tương đồng giữa khung nhìn thứ nhất và khung nhìn thứ ba.

Bước phân tích này có thể bao gồm: bước thu thập, dựa trên dữ liệu video nguồn, ít nhất một khung nhìn tham chiếu, bước lọc và bước giảm tần số lấy mẫu một hoặc nhiều khung nhìn nguồn theo nhiều cấu hình khác nhau, đối với mỗi cấu hình khác nhau: bước mã hóa dữ liệu video nguồn đã xử lý tương ứng để tạo chuỗi bit video ứng viên; bước giải mã từ chuỗi bit video ứng viên một hoặc nhiều khung nhìn nguồn; bước cấu trúc lại ít nhất một khung nhìn thử nghiệm từ một hoặc nhiều khung nhìn nguồn đã giải mã, ít nhất một khung nhìn thử nghiệm này tương ứng với ít nhất một khung nhìn tham chiếu đó; và bước so sánh khung nhìn thử nghiệm được tái cấu trúc với khung nhìn tham chiếu, bước phân

tích còn bao gồm thêm bước chọn cấu hình dựa trên kết quả so sánh, phương pháp này còn bao gồm thêm bước chọn, như chuỗi bit video, chuỗi bit video ứng viên liên quan đến cấu hình đã chọn.

Ít nhất một khung nhìn tham chiếu này có thể bao gồm một trong số các khung nhìn nguồn. Theo một số phương án, bước thu thập ít nhất một khung nhìn tham chiếu có thể bao gồm bước tổng hợp khung nhìn tham chiếu từ dữ liệu video nguồn. Khung nhìn tham chiếu được tổng hợp có thể tương ứng với cổng nhìn khác với khung nhìn nguồn.

Theo một số phương án, bước phân tích bao gồm bước chọn cấu hình có khung nhìn thử nghiệm được tái cấu trúc phù hợp nhất với khung nhìn tham chiếu. Nói cách khác, cấu hình này có thể được chọn dựa trên việc giảm hoặc giảm tối thiểu lỗi giữa khung nhìn tham chiếu và khung nhìn thử nghiệm được tái cấu trúc.

Phương pháp này còn có thể bao gồm bước thiết kế bộ phận lọc tái cấu trúc sử dụng thuật toán học máy, trong đó bộ phận lọc tái cấu trúc này có thể được bộ giải mã sử dụng khi tăng tần số lấy mẫu và lọc một hoặc nhiều khung nhìn đã giải mã, và siêu dữ liệu còn bao gồm mô tả bộ phận lọc tái cấu trúc này.

Theo một số phương án, thuật toán học máy này là mạng nơron tích chập (CNN), có thể tùy ý sử dụng học chuyển tiếp (transfer learning).

Mô tả bộ phận lọc tái cấu trúc có thể bao gồm tập hợp các hệ số (trọng số) cho một CNN. Tập hợp các trọng số này có thể là một phần (ví dụ, chỉ mô tả một số lớp của CNN) hoặc hoàn chỉnh.

Bước thiết kế bộ phận lọc tái cấu trúc có thể bao gồm bước đào tạo (training), bằng thuật toán học máy, hệ số lọc của bộ phận lọc tái cấu trúc để giảm lỗi tái cấu trúc. Lỗi tái cấu trúc có thể bao gồm khác biệt giữa khung nhìn đã tái cấu trúc và khung nhìn tham chiếu.

Phương pháp này có thể bao gồm: bước thu thập, dựa trên dữ liệu video nguồn, ít nhất một khung nhìn tham chiếu, bước giải mã từ chuỗi bit video một hoặc nhiều trong số nhiều khung nhìn nguồn, và bước xác định khung nhìn thử nghiệm tương ứng với khung nhìn tham chiếu, khung nhìn thử nghiệm đang được tạo lại từ một hoặc nhiều khung nhìn nguồn đã giải mã, trong đó việc thiết kế bộ lọc tái cấu trúc bao gồm các hệ số bộ lọc đào tạo của bộ lọc tái cấu trúc, sử dụng thuật toán học máy, để giảm lỗi giữa khung nhìn thử nghiệm và khung nhìn tham chiếu.

Thuật toán học máy này có thể được đào tạo để giảm thiểu lỗi giữa khung nhìn thử nghiệm và khung nhìn tham chiếu.

Khung nhìn thử nghiệm và khung nhìn tham chiếu có thể tương ứng với nhau ít nhất vì chúng có cùng các thông số công nhìn.

Trong một số trường hợp khung nhìn tham chiếu có thể chỉ là một trong các khung nhìn nguồn, và khung nhìn thử nghiệm là tái cấu trúc của khung nhìn nguồn. (Trong trường hợp này, bước tổng hợp khung nhìn tham chiếu có thể bao gồm chỉ bao gồm bước sao chép khung nhìn nguồn). Trong các trường hợp khác, khung nhìn tham chiếu có thể là khung nhìn tổng hợp khác với bất kỳ khung nhìn nguồn nào.

Sáng chế cũng đề xuất phương pháp giải mã video sống động, phương pháp này bao gồm các bước:

bước thu một chuỗi bit video và một chuỗi bit siêu dữ liệu, trong đó chuỗi bit video bao gồm dữ liệu kết cấu đã mã hóa và dữ liệu độ sâu đã mã hóa cho mỗi một trong số nhiều khung nhìn nguồn, và chuỗi bit siêu dữ liệu bao gồm siêu dữ liệu để hỗ trợ bước giải mã chuỗi bit video;

bước thu thập thông số công nhìn của ít nhất một khung nhìn đích được sẽ được kết xuất;

bước giải mã từ chuỗi bit video một hoặc nhiều khung nhìn nguồn; và

bước tái cấu trúc ít nhất một khung nhìn đích từ một hoặc nhiều khung nhìn nguồn đã giải mã, dựa trên các thông số công nhìn thu được,

trong đó siêu dữ liệu mô tả cấu hình của bước giảm tần số lấy mẫu được thực hiện khi mã hóa một hoặc nhiều khung nhìn nguồn, và

bước tái cấu trúc bao gồm bước lấy mẫu và bước lọc, dựa trên cấu hình được mô tả, ít nhất một trong số một hoặc nhiều khung nhìn nguồn đã giải mã hoặc nguồn gốc của chúng.

Các thông số công nhìn của một khung nhìn có thể bao gồm một vị trí và hướng xác định khung nhìn. Các thông số này có thể bao gồm thêm một trường nhìn.

Nguồn gốc của khung nhìn nguồn đã giải mã ít nhất này có thể là khung nhìn đích tổng hợp, có nguồn gốc từ khung nhìn nguồn (ví dụ, bằng cách chiếu/làm cong khung nhìn nguồn theo các thông số của công nhìn đích).

Phương pháp giải mã có thể bao gồm các bước: bước giải mã từ chuỗi bit video ít nhất hai trong số nhiều khung nhìn nguồn; và đối với mỗi khung nhìn trong số hai

khung nhìn nguồn đã giải mã, bước tổng hợp khung nhìn đích tổng hợp, dựa trên các thông số cổng nhìn, trong đó bước tái cấu trúc ít nhất một khung nhìn đích bao gồm bước tăng tần số lấy mẫu, bước lọc và bước kết hợp các khung nhìn đích tổng hợp, theo cấu hình đã mô tả trong siêu dữ liệu.

Bước tổng hợp từng khung nhìn đích tổng hợp có thể bao gồm bước chiếu hoặc làm cong khung nhìn nguồn đã giải mã tương ứng, theo các thông số cổng nhìn của khung nhìn đích.

Ít nhất hai khung nhìn nguồn đã giải mã có thể đã được giảm tần số lấy mẫu theo các hệ số khác nhau (1x, 2x, 4x, v.v.) tại bộ mã hóa. Chúng có thể có các kích thước/độ phân giải khác nhau.

Các bước tăng tần số lấy mẫu, bước lọc, và bước kết hợp có thể được thực hiện riêng biệt hoặc đồng thời. Ví dụ, các bước này có thể được triển khai tất cả cùng nhau bằng một mạng nơron như mạng CNN.

Các khung nhìn đích tổng hợp có thể có các tỷ lệ/độ phân giải khác nhau. Nghĩa là, chúng có thể yêu cầu tăng tần số lấy mẫu theo các hệ số khác nhau.

Bước tăng tần số lấy mẫu và bước lọc có thể lấy nhiều khung nhìn đích tổng hợp khác nhau ở các tỷ lệ/độ phân giải khác nhau, và tăng tần số lấy mẫu và lọc chúng để tái tạo lại ít nhất một khung nhìn đích. Bước tăng tần số lấy mẫu, bước lọc, và bước kết hợp có thể bao gồm việc sử dụng mạng nơron, hoạt động trên các khung nhìn đích tổng hợp làm đầu vào để tạo ra ít nhất một khung nhìn đích đầu ra.

Siêu dữ liệu có thể bao gồm mô tả về bộ lọc tái cấu trúc được sử dụng trong bước tăng tần số lấy mẫu và bước lọc, và bước tăng tần số lấy mẫu và bước lọc này có thể bao gồm việc sử dụng bộ lọc tái cấu trúc đã mô tả.

Mô tả này có thể xác định một trong số tập hợp các bộ lọc được xác định trước. Hoặc hay ngoài ra, mô tả này có thể bao gồm các hệ số bộ lọc.

Mô tả bộ lọc có thể bao gồm một tập hợp hoặc một phần tập hợp các trọng số cho mạng nơron và bước tăng tần số lấy mẫu và bước lọc lên có thể bao gồm bước xử lý một hoặc nhiều khung nhìn nguồn đã giải mã bằng cách sử dụng mạng nơron áp dụng các trọng số.

Chuỗi bit video có thể được thu từ bộ mã hóa từ xa, phương pháp này còn bao gồm bước phát các thông số cổng nhìn tới bộ mã hóa.

Các thông số công nhìn được phát có thể là thông số công nhìn của khung nhìn đích ít nhất, hoặc dự đoán về khung nhìn đích sẽ được kết xuất trong tương lai. Việc phát ngược tuyến một chỉ báo hoặc dự đoán về khung nhìn đích có thể tạo điều kiện cho việc mã hóa hiệu quả hơn tại bộ mã hóa, vì bộ mã hóa có thể điều chỉnh việc mã hóa cho khung nhìn đích mong muốn.

Sáng chế cũng đề xuất chương trình máy tính bao gồm mã máy tính làm cho hệ thống xử lý thực hiện phương pháp như đã tóm tắt ở trên khi chương trình này được chạy trên hệ thống xử lý này.

Chương trình máy tính có thể được lưu trữ trên phương tiện lưu trữ mà máy tính có thể đọc được. Đây có thể là phương tiện lưu trữ dài hạn.

Ngoài ra, sáng chế còn cung cấp thêm một bộ mã hóa video được tạo cấu hình để mã hóa video sống động, bộ mã hóa video này bao gồm:

đầu vào, được tạo cấu hình để thu dữ liệu video nguồn, bao gồm nhiều khung nhìn nguồn, mỗi khung nhìn nguồn bao gồm dữ liệu kết cấu và dữ liệu độ sâu;

bộ xử lý video, được tạo cấu hình để xử lý dữ liệu video nguồn để tạo ra dữ liệu video nguồn đã xử lý, bước xử lý này bao gồm bước giảm tần số lấy mẫu một hoặc nhiều khung nhìn nguồn;

bộ mã hóa, được tạo cấu hình để mã hóa dữ liệu video nguồn đã xử lý để tạo chuỗi bit video và chuỗi bit siêu dữ liệu; và

đầu ra, được tạo cấu hình để xuất chuỗi bit video và chuỗi bit siêu dữ liệu,

trong đó chuỗi bit video bao gồm dữ liệu kết cấu được mã hóa và dữ liệu độ sâu được mã hóa, và chuỗi bit siêu dữ liệu bao gồm siêu dữ liệu để hỗ trợ bộ giải mã giải mã chuỗi bit video, trong đó siêu dữ liệu này mô tả cấu hình của bước xử lý được thực hiện trên một hoặc nhiều khung nhìn nguồn.

Bộ mã hóa video có thể bao gồm thêm bộ phân tích video, được tạo cấu hình để phân tích dữ liệu video nguồn để chọn cấu hình cho bộ xử lý video, trong đó bộ xử lý video được tạo cấu hình để lọc và giảm tần số lấy mẫu một hoặc nhiều khung nhìn nguồn theo cấu hình đã chọn.

Bộ mã hóa video có thể bao gồm thêm bộ tạo khung nhìn tham chiếu và bộ giải mã tham chiếu, trong đó: bộ tạo khung nhìn tham chiếu được tạo cấu hình để tạo, dựa trên dữ liệu video nguồn, ít nhất một khung nhìn tham chiếu, bộ xử lý video được tạo cấu hình để lọc và giảm tần số lấy mẫu một hoặc nhiều khung nhìn nguồn theo nhiều

cấu hình khác nhau, đối với mỗi cấu hình trong số các cấu hình khác nhau: bộ mã hóa được tạo cấu hình để mã hóa dữ liệu video nguồn đã xử lý tương ứng để tạo chuỗi bit video ứng viên; và bộ giải mã tham chiếu được tạo cấu hình để: giải mã, từ chuỗi bit video ứng viên, một hoặc nhiều khung nhìn nguồn, tái cấu trúc ít nhất một khung nhìn thử nghiệm từ một hoặc nhiều khung nhìn nguồn đã giải mã, ít nhất một khung nhìn thử nghiệm này tương ứng với ít nhất một khung nhìn tham chiếu, và so sánh khung nhìn thử nghiệm được tái cấu trúc với khung nhìn tham chiếu, trong đó bộ phân tích được tạo cấu hình để chọn cấu hình dựa trên kết quả so sánh, và đầu ra được tạo cấu hình để xuất ra, như chuỗi bit video, video ứng viên chuỗi bit liên quan đến cấu hình đã chọn.

Bộ tạo khung nhìn tham chiếu có thể được tạo cấu hình để chọn một trong các khung nhìn nguồn làm khung nhìn tham chiếu ít nhất và/hoặc có thể được tạo cấu hình để tổng hợp khung nhìn tham chiếu từ dữ liệu video nguồn.

Bộ mã hóa video có thể bao gồm thêm hoặc thay thế với bộ tạo khung nhìn tham chiếu, bộ giải mã tham chiếu, và bộ đào tạo, trong đó:

bộ tạo khung nhìn tham chiếu được tạo cấu hình để tạo, dựa trên dữ liệu video nguồn, ít nhất một khung nhìn tham chiếu;

bộ giải mã tham chiếu được tạo cấu hình để giải mã, từ chuỗi bit video ứng viên, một hoặc nhiều trong số nhiều khung nhìn nguồn; và

bộ đào tạo được tạo cấu hình để:

xác định khung nhìn thử nghiệm tương ứng với khung nhìn tham chiếu, khung nhìn thử nghiệm này được tái cấu trúc từ một hoặc nhiều khung nhìn nguồn đã giải mã;

thiết kế, bằng cách sử dụng thuật toán học máy, bộ lọc tái cấu trúc được bộ giải mã sử dụng khi giảm tần số lấy mẫu và lọc một hoặc nhiều khung nhìn đã giải mã,

trong đó bộ lọc tái cấu trúc được thiết kế để giảm lỗi giữa khung nhìn thử nghiệm và khung nhìn tham chiếu.

Ngoài ra còn cung cấp thêm bộ giải mã video được tạo cấu hình để giải mã video sống động, bộ giải mã video bao gồm:

đầu vào chuỗi bit, được tạo cấu hình để thu chuỗi bit video và chuỗi bit siêu dữ liệu, trong đó chuỗi bit video bao gồm dữ liệu kết cấu đã mã hóa và dữ liệu độ sâu đã mã hóa cho mỗi trong số nhiều khung nhìn nguồn, và chuỗi bit siêu dữ liệu bao gồm siêu dữ liệu để hỗ trợ giải mã chuỗi bit video;

đầu vào điều khiển, được tạo cấu hình để thu thông số cổng nhìn của ít nhất một khung nhìn đích sẽ được kết xuất;

bộ giải mã, được tạo cấu hình để giải mã từ chuỗi bit video một hoặc nhiều khung nhìn nguồn; và

bộ xử lý tái cấu trúc, được tạo cấu hình để tái cấu trúc ít nhất một khung nhìn đích từ một hoặc nhiều khung nhìn nguồn đã giải mã, dựa trên các thông số cổng nhìn thu được, trong đó bước tái cấu trúc bao gồm bước tăng tần số lấy mẫu và bước lọc, dựa trên siêu dữ liệu, ít nhất một trong số một hoặc nhiều khung nhìn nguồn đã giải mã hoặc nguồn gốc của chúng.

Siêu dữ liệu có thể bao gồm mô tả của bộ lọc tái cấu trúc sẽ được sử dụng trong bước tăng tần số lấy mẫu và bước lọc, và bộ xử lý tái cấu trúc có thể được tạo cấu hình để tăng tần số lấy mẫu và lọc một trong một hoặc nhiều khung nhìn nguồn đã giải mã ít nhất, sử dụng bộ lọc tái cấu trúc đã mô tả.

Bộ xử lý tái cấu trúc có thể bao gồm bộ tổng hợp được tạo cấu hình để tổng hợp một khung nhìn đích tổng hợp từ mỗi khung nhìn trong số một hoặc nhiều khung nhìn nguồn; và bộ nâng cấp được tạo cấu hình để tăng lấy mẫu, lọc, và kết hợp các khung nhìn đích tổng hợp.

Ngoài ra, sáng chế còn đề xuất một chuỗi bit mã hóa video sống động, chuỗi bit này bao gồm:

chuỗi bit video, bao gồm dữ liệu kết cấu đã mã hóa và dữ liệu độ sâu đã mã hóa cho mỗi trong số nhiều khung nhìn nguồn; và

chuỗi bit siêu dữ liệu, bao gồm siêu dữ liệu để hỗ trợ bộ giải mã giải mã chuỗi bit video,

trong đó siêu dữ liệu này mô tả cấu hình bước giảm tần số lấy mẫu được thực hiện trên một hoặc nhiều khung nhìn nguồn trong quy trình mã hóa chuỗi bit video.

Cụ thể, siêu dữ liệu có thể bao gồm mô tả về bộ lọc tái cấu trúc sẽ được bộ giải mã sử dụng khi tăng tần số lấy mẫu và lọc một hoặc nhiều khung nhìn đã giải mã.

Chuỗi bit video sống động có thể được mã hóa và giải mã bằng các phương pháp như được tóm tắt bên trên. Chuỗi bit có thể được thể hiện trên một phương tiện máy tính có thể đọc được hoặc dưới dạng tín hiệu được điều chế thành sóng mang điện từ.

Các phương án thực hiện này và các phương án khác của sáng chế sẽ được thể hiện rõ ràng và được làm sáng tỏ theo các phương án được mô tả sau đây.

Mô tả vắn tắt các hình vẽ

Để hiểu rõ hơn về sáng chế cũng như để thể hiện rõ hơn cách sáng chế có thể được thực hiện, giờ đây, chỉ thông qua ví dụ, sẽ thực hiện tham chiếu đến các hình vẽ kèm theo, trong đó:

Hình 1 là lưu đồ về phương pháp mã hóa video 6DoF theo phương án thứ nhất;

Hình 2 là sơ đồ khối của bộ mã hóa 6DoF theo một phương án;

Hình 3 là lưu đồ về phương pháp mã hóa video 6DoF theo phương án thứ hai;

Hình 4 là lưu đồ về phương pháp mã hóa video 6DoF theo phương án thứ ba;

Hình 5 là lưu đồ minh họa phương pháp giải mã video 6DoF theo phương án thứ tư;

Hình 6 là sơ đồ khối về bộ giải mã 6DoF theo một phương án;

Hình 7 là lưu đồ minh họa phương pháp giải mã video 6DoF theo phương án thứ năm; và

Hình 8 là lưu đồ minh họa phương pháp giải mã video 6DoF theo phương án thứ sáu.

Mô tả chi tiết các phương án thực hiện sáng chế

Sáng chế sẽ được mô tả theo các hình vẽ.

Cần hiểu rằng mô tả chi tiết và các ví dụ cụ thể, trong khi biểu thị các phương án ví dụ về thiết bị, hệ thống và phương pháp, chỉ nhằm mục đích minh họa và không nhằm mục đích giới hạn phạm vi của sáng chế. Những điều này và các đặc điểm, phương án và ưu điểm khác của thiết bị, hệ thống và phương pháp của sáng chế sẽ được hiểu rõ hơn từ phần mô tả, điểm yêu cầu bảo hộ đi kèm cũng như hình vẽ đi kèm sau đây. Cần phải hiểu rằng các Hình chỉ là vẽ theo sơ đồ và không được vẽ theo tỷ lệ. Cũng nên hiểu rằng, các ký hiệu chỉ dẫn giống nhau sẽ được sử dụng trong các Hình để biểu thị các phần giống nhau hoặc tương tự nhau.

Các phương pháp mã hóa và giải mã video sống động được đề xuất. Theo một phương pháp mã hóa, dữ liệu video nguồn bao gồm nhiều khung nhìn nguồn được mã hóa thành chuỗi bit video. Ít nhất một trong các khung nhìn nguồn được lọc và giảm tần số lấy mẫu trước khi mã hóa. Chuỗi bit siêu dữ liệu liên quan đến luồng video bao gồm siêu dữ liệu mô tả cấu hình của bước lọc và bước giảm tần số lấy mẫu, để hỗ trợ bộ giải mã giải mã chuỗi bit video. Có thể tin rằng việc sử dụng khung nhìn được giảm

tần số lấy mẫu có thể giúp giảm thành phần lã mã hóa, so với phương pháp mã hóa dựa trên đoạn nối tạm. Sáng chế cũng đề xuất bộ mã hóa và bộ giải mã dành cho video sống động, và chuỗi bit video sống động.

Các phương án của sáng chế có thể thích hợp để triển khai một phần của tiêu chuẩn kỹ thuật, như MPEG-I Phần 12 Video sống động. Khi có thể, thuật ngữ được sử dụng trong tài liệu này được chọn để nhất quán với các thuật ngữ được sử dụng trong MPEG-I Phần 12. Tuy nhiên, cần hiểu rằng phạm vi của sáng chế không giới hạn ở MPEG-I Phần 12, cũng như bất kỳ tiêu chuẩn kỹ thuật nào khác.

Có thể hữu ích nếu đưa ra các định nghĩa/giải thích sau:

“Cảnh 3D” đề cập đến nội dung trực quan trong hệ tọa độ tham chiếu toàn cầu.

“Tập bản đồ” là tập hợp các đoạn nối tạm từ một hoặc nhiều biểu diễn cách trình bày sau quá trình hợp lại, thành một cặp hình ảnh chứa hình ảnh thành phần kết cấu và hình ảnh thành phần chiều sâu tương ứng.

“Thành phần tập bản đồ” là kết cấu hoặc thành phần chiều sâu của tập bản đồ.

“Thông số camera” xác định phép chiếu được sử dụng để tạo ra hình chiếu từ một cảnh 3D.

“Cắt tỉa” là một quá trình xác định và trích xuất các vùng bị che khuất trên các khung nhìn, dẫn đến các đoạn nối tạm.

“Trình kết xuất” là một phương án của quy trình tạo cổng nhìn hoặc khung nhìn đa hướng từ biểu diễn cảnh 3D, tương ứng với vị trí và hướng xem.

“Khung nhìn nguồn” là nguyên liệu video nguồn trước khi mã hóa tương ứng với định dạng của biểu diễn khung nhìn, có thể có được bằng cách chụp cảnh 3D bằng camera thực hoặc bằng cách chiếu bằng camera ảo lên bề mặt sử dụng các thông số camera nguồn.

“Khung nhìn đích” được định nghĩa là cổng nhìn phối cảnh hoặc khung nhìn đa hướng tại vị trí và hướng xem mong muốn.

“Biểu diễn khung nhìn” bao gồm các mảng mẫu 2D của thành phần kết cấu và thành phần chiều sâu tương ứng, thể hiện hình chiếu của cảnh 3D lên một bề mặt bằng cách sử dụng các thông số camera.

“Cổng nhìn” có nghĩa là hình chiếu của kết cấu lên bề mặt phẳng của trường nhìn của hình ảnh hoặc video đa chiều hoặc 3D phù hợp để hiển thị và để người dùng nhìn với một vị trí và hướng xem cụ thể.

Thuật toán học máy là bất kỳ thuật toán tự đào tạo nào xử lý dữ liệu đầu vào để tạo ra hoặc dự đoán dữ liệu đầu ra. Theo một số phương án của sáng chế, dữ liệu đầu vào này bao gồm một hoặc nhiều khung nhìn đã giải mã từ một chuỗi bit và dữ liệu đầu ra bao gồm dự đoán/tái cấu trúc khung nhìn đích.

Người có chuyên môn trong ngành sẽ biết rõ các thuật toán học máy phù hợp để được sử dụng trong sáng chế này. Các ví dụ về các thuật toán học máy phù hợp bao gồm thuật toán cây quyết định và mạng nơron nhân tạo. Các thuật toán học máy khác như hồi quy logistic, máy vectơ hỗ trợ hoặc mô hình Naïve Bayesian là những lựa chọn thay thế phù hợp.

Cấu trúc của mạng nơron nhân tạo (hay đơn giản là mạng nơron) được lấy cảm hứng từ bộ não con người. Mạng nơron bao gồm các lớp, mỗi lớp bao gồm nhiều nơron. Mỗi nơron bao gồm một phép toán. Cụ thể, mỗi nơron có thể bao gồm một tổ hợp có trọng số khác nhau của một kiểu biến đổi (ví dụ: cùng một kiểu biến đổi, sigmoid, v.v. nhưng có trọng số khác nhau). Trong quá trình xử lý dữ liệu đầu vào, phép toán của mỗi nơron được thực hiện trên dữ liệu đầu vào để tạo ra đầu ra dạng số và đầu ra của mỗi lớp trong mạng nơron được đưa vào một hoặc nhiều lớp khác (ví dụ: theo tuần tự). Lớp cuối cùng cung cấp đầu ra.

Phương pháp đào tạo một thuật toán học máy đã được biết đến phổ biến. Thông thường, các phương pháp này bao gồm bước thu thập bộ dữ liệu đào tạo, bao gồm mục nhập dữ liệu đầu vào đào tạo và mục nhập dữ liệu đầu ra đào tạo tương ứng. Một thuật toán học máy đã khởi tạo được áp dụng cho mỗi mục nhập dữ liệu đầu vào để tạo ra các mục nhập dữ liệu đầu ra dự đoán. Lỗi giữa các mục nhập dữ liệu đầu ra dự đoán và các mục nhập dữ liệu đầu ra đào tạo tương ứng được sử dụng để sửa đổi thuật toán học máy này. Quá trình này có thể được lặp lại cho đến khi lỗi hội tụ và các mục nhập dữ liệu đầu ra dự đoán đủ giống (ví dụ: chênh lệch $\pm 1\%$) với các mục nhập dữ liệu đầu ra đào tạo. Đây thường được biết đến như một kỹ thuật học có giám sát.

Ví dụ, khi thuật toán học máy được hình thành từ mạng nơron, (trọng số của) phép toán của từng nơron có thể được sửa đổi cho đến khi hội tụ sai số. Các phương pháp sửa đổi mạng nơron đã biết bao gồm giảm gradient, thuật toán lan truyền ngược, v.v..

Theo các phương án, các mục nhập dữ liệu đầu vào đào tạo có thể tương ứng với, ví dụ, các khung nhìn đã giải mã từ chuỗi video thử nghiệm, hoặc các khung nhìn

tổng hợp được tái cấu trúc từ các khung nhìn đã giải mã đó. Các mục nhập dữ liệu đầu ra đào tạo có thể tương ứng với các khung nhìn nguồn ban đầu (trước khi mã hóa) hoặc một khung nhìn tham chiếu được tổng hợp trực tiếp từ các khung nhìn nguồn ban đầu.

Mạng nơron tích chập (CNN, hoặc ConvNet) là một loại mạng nơron sâu, được áp dụng phổ biến nhất để phân tích hình ảnh trực quan. CNN là phiên bản chính quy hóa của nhận thức nhiều lớp.

Hình 1 là lưu đồ về phương pháp mã hóa video sống động theo phương án thứ nhất. Hình 2 minh họa sơ đồ khối về bộ mã hóa video 300 có thể thực hiện phương pháp của Hình 1.

Trong bước 110, đầu vào (bộ thu 310) của bộ mã hóa 300, thu dữ liệu video nguồn. Dữ liệu video nguồn này bao gồm nhiều khung nhìn nguồn, mỗi khung nhìn này bao gồm dữ liệu kết cấu và dữ liệu độ sâu. Bộ thu 310 này được ghép nối với bộ phân tích video 320. Trong bước 120, bộ phân tích video phân tích dữ liệu video nguồn để xác tạo cấu hình tốt nhất cho bước lọc và giảm tần số lấy mẫu tiếp theo. Phân tích này sẽ được mô tả chi tiết hơn bên dưới.

Bộ phân tích video 320 được ghép nối với bộ xử lý video 330. Bộ xử lý video cũng được ghép nối với bộ thu 310. Bộ xử lý video thu dữ liệu video nguồn từ bộ thu 310 và thu thông tin từ bộ phân tích video về cấu hình do nó xác định. Trong bước 130, bộ xử lý video lọc ít nhất một trong các khung nhìn nguồn và trong bước 140, bộ xử lý video giảm tần số lấy mẫu khung nhìn nguồn đã lọc. Bước lọc và giảm tần số lấy mẫu được thực hiện theo cấu hình được xác định bởi bộ phân tích video.

Đầu ra của bộ xử lý video được chuyển tới bộ mã hóa 340. Ở bước 150, bộ mã hóa mã hóa dữ liệu video nguồn đã xử lý để tạo một chuỗi bit video. Chuỗi bit video này bao gồm dữ liệu kết cấu được mã hóa và dữ liệu độ sâu được mã hóa. Bộ mã hóa 340 cũng tạo ra một chuỗi bit siêu dữ liệu, bao gồm siêu dữ liệu để hỗ trợ bộ giải mã giải mã chuỗi bit video. Cụ thể, chuỗi bit siêu dữ liệu bao gồm mô tả về bước lọc và giảm tần số lấy mẫu được thực hiện trên một hoặc nhiều khung nhìn nguồn.

Cả chuỗi bit video và chuỗi bit siêu dữ liệu đều được xuất qua đầu ra của bộ mã hóa video 300. Chúng có thể được phát tới bộ giải mã hoặc có thể được lưu trữ trên phương tiện lưu trữ mà máy có thể đọc được, để đọc và giải mã sau này. Bước giảm tần số lấy mẫu của ít nhất một trong các khung nhìn nguồn đạt được sự nén (giảm tốc độ bit). Đồng thời, vì phương pháp của phương án thứ nhất không dựa trên việc tập

hợp các khung nhìn nguồn vào tập bản đồ kết cấu dựa trên đoạn nổi tạm, nên tránh các thành phần lạ có thể được liên kết với các cạnh đoạn nổi tạm.

Hình 3 minh họa lưu đồ của phương pháp mã hóa video sống động theo phương án thứ hai. Các bước tương tự như phương án của Hình 1 được đưa ra giống như các số tham chiếu. Phương pháp của Hình 3 khác với phương pháp của Hình 1 về cơ bản ở chỗ nó sử dụng bộ giải mã tham chiếu trong bộ mã hóa video.

Như minh họa trong Hình 2, bộ mã hóa video 300 bao gồm bộ tạo khung nhìn tham chiếu 375. Bộ mã hóa video này cũng bao gồm bộ giải mã tham chiếu, bao gồm bộ giải mã 350, bộ chọn 360 và bộ tổng hợp 370. Bộ mã hóa video 300 này còn bao gồm bộ đào tạo 380, được ghép nối với bộ tạo khung nhìn tham chiếu 375 và đầu ra của bộ giải mã tham chiếu - cụ thể, là đầu ra của bộ tổng hợp 370. Trong bước 122, bộ tạo khung nhìn tham chiếu 375 tạo ra ít nhất một khung nhìn tham chiếu, từ dữ liệu video nguồn. Khung nhìn tham chiếu có thể là một trong các khung nhìn nguồn hoặc có thể là khung nhìn tổng hợp. Khung nhìn tham chiếu đã chọn được xác định bởi các thông số cổng nhìn thử nghiệm, nhập vào bộ mã hóa video 300.

Các bước phân tích 120a, bước lọc 130a và bước giảm tần số lấy mẫu 140a tương tự như các bước tương ứng 120, 130 và 140 của Hình 1, ngoại trừ, trong phương án này, bước lọc và bước giảm tần số lấy mẫu được thực hiện theo nhiều cấu hình khác nhau. Mục đích là để xác định cấu hình tốt nhất để lọc và giảm tần số lấy mẫu nội dung cụ thể của dữ liệu video nguồn. Khung nhìn tham chiếu sẽ được sử dụng làm tham chiếu hoặc "tiêu chuẩn vàng", để xác định cấu hình tốt nhất.

Đối với mỗi cấu hình lọc và giảm tần số lấy mẫu, bộ mã hóa 340 mã hóa dữ liệu video nguồn đã xử lý tương ứng, để tạo một chuỗi bit ứng viên (xem bước 150a). Các chuỗi bit ứng viên được chuyển từ bộ mã hóa 340 đến bộ giải mã tham chiếu, thông qua một cổng tắc 390. Trong bước 124, bộ giải mã tham chiếu giải mã từ mỗi chuỗi bit video ứng viên một hoặc nhiều trong số nhiều khung nhìn nguồn. Bước giải mã (bước 124) được bộ giải mã 350 thực hiện. Một hoặc nhiều khung nhìn nguồn có liên quan được bộ chọn 360 chọn theo cổng nhìn thử nghiệm. Từ các khung nhìn nguồn đã giải mã đã chọn, bộ tổng hợp 370 sẽ tái cấu trúc khung nhìn thử nghiệm theo cổng nhìn thử nghiệm (bước 126). Lưu ý rằng nếu cổng nhìn thử nghiệm tương ứng trực tiếp với một trong các khung nhìn nguồn, thì bộ tổng hợp chỉ chuyển khung nhìn nguồn (đã giải mã) đó đến bộ đào tạo 380. Tương tự như vậy, khung nhìn tham chiếu trong trường hợp này sẽ là khung

nhìn nguồn gốc. Tuy nhiên, nói chung, công nhìn thử nghiệm khác với bất kỳ khung nhìn nguồn nào. Trong trường hợp này, bộ tổng hợp 370 tái cấu trúc (bước 126) khung nhìn thử nghiệm tương ứng với công nhìn thử nghiệm từ một hoặc nhiều khung nhìn nguồn đã giải mã. Tương tự như vậy, bộ tạo khung nhìn tham chiếu 375 sẽ tổng hợp một khung nhìn tham chiếu từ các khung nhìn nguồn gốc.

Trên thực tế, bộ giải mã tham chiếu tái tạo các phép toán giải mã sẽ được thực hiện bởi bộ giải mã thực của chuỗi bit video. Do đó, đầu ra của bộ tổng hợp 370 là một luồng video được tái cấu trúc giống hệt với một luồng có thể được kết xuất ở bộ giải mã đầu xa. Trong bước 128, bộ đào tạo 380 so sánh khung nhìn thử nghiệm được tái tạo với khung nhìn tham chiếu. Kết quả của phép so sánh được chuyển đến bộ phân tích 320 và, bộ phân tích này chọn cấu hình lọc và giảm tần số lấy mẫu dẫn đến sai số thấp nhất giữa khung nhìn thử nghiệm tái cấu trúc và khung nhìn tham chiếu. Sau đó, chuỗi bit video ứng viên liên quan đến cấu hình này được chọn làm chuỗi bit video để xuất, qua bộ chuyển đổi 390.

Bằng cách này, phân tích tại bộ mã hóa có thể thử nghiệm với các cấu hình khác nhau - ví dụ: các nhân bộ lọc khác nhau và/hoặc tốc độ lấy mẫu giảm khác nhau, và có thể chọn cấu hình cung cấp hiệu suất tốt nhất (về lỗi mã hóa tối thiểu đối với một tốc độ bit cho trước) tại bộ giải mã.

Hình 4 là một lưu đồ khác minh họa phương pháp mã hóa theo phương án thứ ba. Phương án này giống với phương án thứ hai ở chỗ nó sử dụng bộ giải mã tham chiếu trong vòng lặp mã hóa. Tuy nhiên, nó khác với phương án thứ hai ở chỗ bộ đào tạo 380 thiết kế, ở bước 129, một bộ lọc tái cấu trúc sẽ được bộ giải mã sử dụng khi tăng tần số lấy mẫu và lọc các khung nhìn đã giải mã. Điều này có thể được thực hiện bằng cách sử dụng các kỹ thuật học máy. Tương tự như phương án thứ hai, kết quả là thu được khung nhìn tham chiếu và khung nhìn thử nghiệm tương ứng. Trong bước 129, bộ đào tạo đào tạo các hệ số lọc của bộ lọc tái cấu trúc, sử dụng thuật toán học máy, sao cho bộ lọc tái cấu trúc kết quả giảm thiểu lỗi giữa khung nhìn thử nghiệm và khung nhìn tham chiếu. Nói cách khác, trên các khung của dữ liệu video và/hoặc trên các pixel của mỗi khung, thuật toán học máy tập hợp hệ số mang đến hiệu suất tái cấu trúc tốt nhất. Trong mỗi lần lặp lại của thuật toán học máy (ví dụ: đối với mỗi khung của video), bộ tổng hợp 370 sẽ áp dụng tập hợp hệ số bộ lọc hiện tại, được suy ra từ thuật toán học máy cho đến thời điểm đó. Khi thuật toán học máy tối ưu hóa các hệ số (ví dụ: trên các khung hình liên tiếp), khung nhìn thử nghiệm tái cấu trúc sẽ trở thành gần như là khung nhìn tham chiếu.

Lưu ý rằng, không giống như phương án thứ hai, phương pháp của phương án thứ ba không yêu cầu phải mã hóa nhiều chuỗi bit video ứng viên. Tối ưu hóa giải mã, chứ không (cần) tối ưu hóa mã hóa. Khi dữ liệu video nguồn đã được mã hóa thành chuỗi bit video, và bước đào tạo bộ lọc tái cấu trúc hoàn tất, bộ mã hóa 340 sẽ nhúng mô tả của bộ lọc tái cấu trúc vào chuỗi bit siêu dữ liệu. Bộ giải mã sẽ sử dụng chuỗi này để tạo bộ lọc tái cấu trúc mà nó sẽ sử dụng để tạo cấu hình tăng tần số lấy mẫu và lọc trong quá trình giải mã.

Hình 5 là lưu đồ về phương pháp giải mã video sống động theo phương án thứ tư. Hình 6 minh họa sơ đồ khối của bộ giải mã video sống động tương ứng, có thể triển khai phương pháp giải mã, theo một phương án. Bộ giải mã video sống động 400 bao gồm đầu vào chuỗi bit (bộ thu 410); đầu vào điều khiển (bộ chọn 450); bộ giải mã 420; và bộ xử lý tái cấu trúc, bao gồm bộ tổng hợp 430 và bộ nâng cấp 440. Bộ giải mã video còn bao gồm bộ chọn 450.

Trong bước 210, đầu vào chuỗi bit thu một chuỗi bit video và chuỗi bit siêu dữ liệu do bộ mã hóa video 300 tạo ra như được mô tả ở trên. Bộ chọn 450 thu các thông số cổng nhìn của ít nhất một khung nhìn đích sẽ được kết xuất. Trong bước 220, bộ giải mã 420 giải mã từ chuỗi bit video một hoặc nhiều khung nhìn nguồn (có thể được bộ chọn 450 lựa chọn dựa trên các thông số cổng nhìn đích). Nếu bộ giải mã 420 có kiến thức hoặc dự đoán về cổng nhìn đích, nó có thể giải mã một cách có chọn lọc chuỗi bit video phù hợp - trước tiên là giải mã các khung nhìn nguồn liên quan nhất đến khung nhìn đích đó và tùy chọn bỏ qua bước giải mã các khung nhìn nguồn không liên quan đến khung nhìn đích.

Bộ tổng hợp 430 và bộ nâng cấp 440 tái cấu trúc ít nhất một khung nhìn đích từ một hoặc nhiều khung nhìn nguồn đã giải mã, dựa trên các thông số của cổng nhìn đích. Việc tái cấu trúc dựa trên chuỗi bit siêu dữ liệu, mô tả cấu hình lọc và giảm tần số lấy mẫu được thực hiện ở bộ mã hóa video 300. Việc tái cấu trúc bao gồm bước tăng tần số lấy mẫu 230 và lọc 240 một hoặc nhiều khung nhìn nguồn đã giải mã, tùy thuộc vào cấu hình được mô tả trong chuỗi bit siêu dữ liệu. Ví dụ, nếu một khung nhìn nguồn nhất định được giảm tần số lấy mẫu theo hệ số hai ở bộ mã hóa video 300, thì bước tăng tần số lấy mẫu 230 có thể tăng tần số lấy mẫu khung nhìn này theo hệ số hai sau khi giải mã. Ví dụ, khung nhìn đích tái cấu trúc có thể được xuất, ví dụ, ra bộ đệm khung sẵn sàng để hiển thị.

Bằng cách lọc và giảm tần số lấy mẫu một hoặc nhiều khung nhìn nguồn tại bộ mã hóa, và bước báo hiệu cho bộ giải mã - ví dụ - khung nhìn nào đã được giảm tần số lấy mẫu và bằng hệ số giảm tần số lấy mẫu nào, tốc độ bit của luồng video có thể được giảm xuống, trong khi cho phép bộ giải mã đạt được kết quả tái cấu trúc chất lượng cao. Theo một số phương án, chuỗi bit siêu dữ liệu bao gồm mô tả về bộ lọc tái tạo được thiết kế bởi bộ mã hóa để sử dụng tại bộ giải mã. Trong trường hợp này, bước 240 về lọc các khung nhìn nguồn đã giải mã có thể bao gồm bước lọc bằng cách sử dụng bộ lọc tái cấu trúc được xác định trong chuỗi bit siêu dữ liệu.

Hình 7 minh họa một phương án khác của phương pháp trong Hình 5, theo phương án thứ năm. Trong phương án này, các thông số cổng nhìn đích được phát ngược tuyến tới bộ mã hóa ở bước 250. Tại bộ mã hóa, kiến thức về cổng nhìn đích có thể được sử dụng để tối ưu hóa bước mã hóa - ví dụ: bằng cách phân bổ tốc độ bit cao hơn cho các khung nhìn phù hợp hơn để tái cấu trúc khung nhìn đích. Theo một ví dụ khác, bộ mã hóa có thể mã hóa từng khung nhìn ở nhiều hệ số tỷ lệ và dựa trên cổng nhìn, bộ phân tích này có thể quyết định chuỗi bit nào sẽ phát tới máy khách. (Quyết định này chỉ có thể thay đổi một lần trong một khoảng thời gian.) Điều thuận lợi là một khung nhìn tương tự như cổng nhìn được phát ở độ phân giải cao.

Hình 8 là lưu đồ của phương pháp giải mã theo phương án thứ sáu. Trong bước 220a, bộ giải mã 420 giải mã ít nhất hai khung nhìn nguồn từ chuỗi bit video. Trong bước 225, bộ tổng hợp 430 tổng hợp khung nhìn đích tổng hợp từ mỗi trong số các khung nhìn nguồn đã giải mã, dựa trên các thông số cổng nhìn của cổng nhìn đích. Bước này có thể bao gồm bước làm cong từng khung nhìn nguồn đã giải mã thành cổng nhìn đích, để tạo ra khung nhìn đích tổng hợp tương ứng. Trong ví dụ minh họa trong Hình 8, hai khung nhìn đích tổng hợp được tăng tần số lấy mẫu và lọc riêng biệt - một khung nhìn được tăng tần số lấy mẫu ở bước 232 và được lọc ở bước 242; và khung nhìn còn lại được tăng tần số lấy mẫu ở bước 234 và được lọc ở bước 244. Bước tăng tần số lấy mẫu và bước lọc có thể do bộ nâng cấp 440 thực hiện. Tuy nhiên, điều này không cần thiết, và thực sự các chức năng của bộ nâng cấp 440 và bộ tổng hợp 430 có thể được chia sẻ hoặc kết hợp với nhau. Trong bước 260, hai khung nhìn đích tổng hợp được kết hợp. Bước này có thể bao gồm tính trung bình hoặc nội suy giữa các khung nhìn đích tổng hợp khác nhau. Một lần nữa, theo phương án hiện tại, giả định rằng bước kết hợp (bước 260) được thực hiện theo bộ nâng cấp hoặc 440, nhưng điều này không cần thiết.

Theo một số phương án, bước tăng tần số lấy mẫu 230, 232, 234; bước lọc 240, 242, 244; và bước kết hợp 260 có thể được hợp nhất vào một quy trình duy nhất. Đặc biệt, đây có thể là trường hợp nếu một bộ lọc tái cấu trúc được thiết kế bằng cách sử dụng thuật toán học máy. Ví dụ, bộ lọc tái cấu trúc có thể bao gồm một tập hợp các trọng số/hệ số cho một mạng nơron. Mạng nơron này có thể được tạo cấu hình để tái cấu trúc khung nhìn đích từ nhiều khung nhìn đích tổng hợp, theo đó các bước tăng tần số lấy mẫu, bước lọc, và bước kết hợp đều được thực hiện cùng nhau bởi mạng nơron, lấy các khung nhìn đích tổng hợp (ở các quy mô/độ phân giải khác nhau) làm đầu vào và đầu ra của khung nhìn đích được tái cấu trúc.

Một lần nữa, theo Hình 2, các thành phần của bộ mã hóa video 300, và phép toán của chúng theo các phương án khác nhau, sẽ được mô tả chi tiết hơn như sau.

Dữ liệu video nguồn (được gắn nhãn “nguyên liệu nguồn” trong Hình 2) có thể là nguyên liệu nguồn 6DoF bao gồm các khung nhìn với các thành phần độ sáng, màu sắc và độ sâu. Các thông số camera có thể được cung cấp cho phép chiếu lại các điểm từ khung nhìn này sang khung nhìn khác. Ví dụ, có thể tổng hợp một khung nhìn từ khung nhìn khác và so sánh những điểm giống và khác nhau do phản xạ và sự bít kín.

Độ sâu có thể được biểu diễn dưới dạng bản đồ độ sâu dày đặc (tùy ý căn chỉnh với dữ liệu kết cấu) hoặc có thể là lưới 3D được liên kết với tọa độ 2D vào khung kết cấu. Dữ liệu kết cấu có thể bao gồm độ sáng và màu sắc, ví dụ: YUV hoặc RGB.

Nguyên liệu nguồn 6DoF có thể là kết quả của quá trình thu nhận và ước tính hoặc là kết quả của quá trình hậu sản xuất. Dữ liệu video nguồn có thể bao gồm một chuỗi các khung (video 6DoF). Nguồn này có thể chứa thông tin khác như dữ liệu âm thanh.

Bộ thu 310 có thể thu tất cả nguyên liệu cùng một lúc (xử lý ngoại tuyến), một đoạn các khung tại một thời điểm hoặc một khung tại một thời điểm (trực tuyến). Bộ thu có thể đệm thông tin và cung cấp các khung trong các phân đoạn cho bộ phân tích.

Bộ phân tích 320 có thể thực hiện tính toán để xác định các khung nhìn phù hợp để giảm quy mô và bất kỳ bước lọc nào khác. Bộ phân tích 320 có thể được cung cấp với ràng buộc, bao gồm nhưng không giới hạn ở: các ràng buộc về tốc độ bit hoặc tốc độ pixel; hoặc danh sách các phép lọc có thể có. Một đầu vào khác cho bộ phân tích có thể là cấu hình bộ giải mã bao gồm số lượt khung nhìn tối đa và/hoặc tổng số pixel mà bộ giải mã có khả năng giải mã. Đầu ra của bộ phân tích là 1) cấu hình cho các phép toán tiếp theo và 2) siêu dữ liệu sẽ được phát tới bộ giải mã.

Ví dụ, một danh sách các phép lọc có thể là {không, giảm quy mô 2x, giảm quy mô 4x, bỏ qua, v.v.} nhưng cũng có thể bao gồm các phép lọc khác như sẽ được mô tả thêm bên dưới. Một số khung nhìn có thể được gửi ở độ phân giải tạm thời thấp hơn - ví dụ: 15 fps/giây thay vì 120 fps/giây.

Bộ phân tích chỉ có thể nghiên cứu cấu hình camera (đặc biệt là vị trí tương đối của chúng trong không gian cảnh) để xác định các phép lọc phù hợp. Ví dụ, camera gần trung tâm nhất, hướng về phía trước có thể được giữ đầy đủ độ phân giải (không có bộ lọc), và sau đó các camera xung quanh được phân cụm lại. Từ mỗi cụm, camera trung tâm được giảm quy mô 2x. Sau đó, các camera còn lại trong mỗi cụm có thể được giảm quy mô 4x. Mô hình này có thể được tiếp tục, để có giảm tần số lấy mẫu 8x, 16x, (v.v.) cho một số camera. Ngoài ra, các khung nhìn ở cấp thấp nhất được bỏ qua. Các khung nhìn dạng thu nhỏ đại diện (thumbnail) được giảm tần số lấy mẫu 8x hoặc 16x này không mang nhiều thông tin cạnh, nhưng cung cấp thông tin về sự thay đổi màu sắc phụ thuộc vào góc nhìn và các điểm nổi bật (bán) di chuyển. Theo một số phương án của sáng chế, các bộ lọc được đào tạo có thể hợp nhất các cạnh chính xác của khung nhìn có độ phân giải cao với khung nhìn gần có độ phân giải thấp.

Khung nhìn nguồn có thể có trường nhìn, vị trí và hình chiếu khác nhau (ví dụ: mắt cá, phối cảnh, v.v.) và việc so sánh các máy ảnh là không quan trọng. Ở đây, khái niệm về sự giống nhau của cổng nhìn có thể mang lại lợi ích. Một tập hợp con các điểm (ở các giá trị độ sâu điển hình) từ cổng nhìn thứ nhất được chiếu lại đến cổng nhìn thứ hai và số điểm khớp với các điểm trong cổng nhìn thứ hai được đếm. Các tia có thể được làm tăng trọng số bằng góc hoặc độ phân giải. Nhiều phương pháp tính toán sự tương đồng cổng nhìn đã được biết đến trong ngành.

Theo một số phương án, bộ phân tích có thể thực hiện phép chiếu lại điểm hoặc các phép tính tổng hợp khung nhìn từ khung nhìn nguồn này sang khung nhìn nguồn khác, để tạo thành một bản đồ các điểm dư thừa giữa các khung nhìn. Dựa trên bản đồ này, các khung nhìn hữu ích nhất được chọn để có chất lượng tốt nhất. Từ các khung nhìn còn lại, những khung nhìn có nhiều thông tin nhất không có trong khung nhìn chất lượng cao có thể được giảm quy mô 2x và sau đó tất cả các khung nhìn khác có thể là 4x. Có thể một số khung nhìn được xác định là (gần như) hoàn toàn dư thừa và bộ phân tích có thể quyết định bỏ qua các khung nhìn này trong quá trình mã hóa tiếp theo.

Theo một số phương án, bộ phân tích thực hiện phân tích đa điểm, trong đó bộ giải mã tham chiếu 6DoF nằm trong vòng lặp với bộ mã hóa 6DoF và bộ phân tích thu được thông tin về chất lượng thực tế của các cổng nhìn thử nghiệm (có thể trùng với các khung nhìn nguồn). Dựa trên điều này, bộ phân tích có thể sửa cấu hình cho lần đi qua tiếp theo. Số lần đi qua có thể được xác định trước hoặc dựa trên mối tương quan giữa chất lượng mong đợi và thực tế như được xác định bằng cách chạy bộ giải mã 6DoF như một phần của bộ mã hóa 6DoF. Ví dụ, phương pháp này có thể được áp dụng theo phương án của Hình 4.

Bộ xử lý video 330 thu nguyên liệu nguồn và được bộ phân tích 320 tạo cấu hình. Tất cả các phép toán có thể được thực hiện trên mỗi khung nhìn hoặc thành phần khung nhìn. Ví dụ, kết cấu của một khung nhìn nhất định có thể không được lọc, nhưng độ sâu có thể được giảm tần số lấy mẫu theo hệ số 4x. Ít nhất một khung nhìn được chia tỷ lệ theo một số lượng nào đó.

Khi mục đích là có số lượng pixel tương ứng với 2-4 khung nhìn nguồn, nhưng số lượng khung nhìn đầu vào có thể là 10-100, thì giảm tần số lấy mẫu 2x là không đủ. Tuy nhiên, nếu một khung nhìn được giữ đầy đủ độ phân giải, và tất cả các khung nhìn khác được giảm tần số lấy mẫu 8x, thì có khả năng sẽ không có đủ thông tin độ phân giải cao để kết xuất các khung nhìn gần ngoại vi của không gian nhìn. Thay vào đó, dự kiến có thể có nhiều tùy chọn giảm tần số lấy mẫu, ví dụ: 2x, 4x, 8x, 16x.

Bộ xử lý video có thể được tạo cấu hình để làm biến dạng khung nhìn sao cho khung nhìn đã xử lý có độ phân giải tương đối cao hơn ở một số vùng (thường là những vùng được xác định là quan trọng hơn - ví dụ: ít dư thừa hơn) và độ phân giải tương đối thấp hơn ở các vùng khác (thông thường những vùng được xác định là ít quan trọng hơn - ví dụ như vì sự giống nhau của chúng với các khung nhìn khác). Tổng khung nhìn vẫn sẽ vừa với một hình chữ nhật nhất định và hình chữ nhật đó thường sẽ nhỏ hơn so với khung nhìn nguồn gốc (độ phân giải đầy đủ). Lưu ý rằng phép tính phân giải có thể thay đổi này không tạo ra các thành phần lạ có cạnh, không giống như cắt tỉa/hợp lại dựa trên đoạn nối tạm.

Bộ xử lý video có thể được tạo cấu hình để tổng hợp một khung nhìn ảo, như một khung nhìn trung tâm đa hướng. Nhìn chung, bộ giải mã có thể không biết khung nhìn đã mã hóa là thực hay ảo. Có thể có ích khi gắn cờ chỉ báo điều này trong siêu dữ liệu. Sau đó, cờ chỉ báo này có thể được sử dụng, ví dụ, bởi bộ lọc tái cấu trúc ở bộ giải mã, mà có thể được tạo cấu hình để ưu tiên các khung nhìn vật lý hơn đối với các cạnh vật thể.

Bộ xử lý video có thể được tạo cấu hình để loại bỏ các độ phân giải không gian nhất định. Ví dụ, thông tin độ phân giải thấp của khung nhìn độ phân giải cao có thể được loại bỏ vì có các khung nhìn độ phân giải thấp gần đó. Một ví dụ khác là sử dụng giảm tần số lấy mẫu để duy trì nhiều thông tin tần số cao hơn, đánh đổi lại là chấp nhận có hiện tượng răng cưa. Ví dụ, một khung nhìn có thể được lọc bằng thông hoặc lọc thông cao, để loại bỏ các tần số thấp và sau đó được giảm tần số lấy mẫu. Thông tin tần số cao là biệt hiệu, nhưng không bị mất vì nó không bị trộn lẫn với thông tin tần số thấp.

Một khung nhìn có thể được lọc theo nhiều cách, dẫn đến nhiều khung nhìn được lọc cho một khung nhìn nguồn duy nhất. Ví dụ, khung nhìn có thể được lọc bằng thông theo nhiều cách (tạo thành một bản lọc).

Thông thường, bộ mã hóa 340 sử dụng công nghệ nén video 2D ở lõi của nó, để nén nhiều khung nhìn ở các độ phân giải khác nhau. Một bộ mã hóa có thể mã hóa các khung nhìn dưới dạng các chuỗi bit riêng lẻ, được phát dưới dạng một chuỗi đa hướng – ví dụ, HEVC (H.265) hoặc HEVC nhiều lớp (H.265 Phụ lục F). Hoặc hay ngoài ra, bộ mã hóa có thể mã hóa các khung nhìn của cùng một góc nhìn nhưng với bước lọc khác nhau, sử dụng codec video có thể mở rộng, như HEVC có thể mở rộng (H.265 Phụ lục H). Một tùy chọn khác là một bộ mã hóa có thể áp dụng codec nhiều khung nhìn. Một bộ mã hóa như vậy có thể sử dụng các dự đoán giữa các khung nhìn. Một ví dụ về codec này là MV-HEVC (H.265 Phụ lục G). Một tùy chọn khác nữa là bản thân bộ mã hóa này có thể sử dụng bản đồ độ sâu để đưa ra các dự đoán giữa các khung nhìn. Một ví dụ về codec như vậy là 3D-HEVC (H.265 Phụ lục I). Bộ giải mã này có thể được cải thiện bằng cách sử dụng Bộ giải mã tham chiếu 6DoF theo một phương án của sáng chế để dự đoán khung nhìn.

Bộ mã hóa 340 có thể tập hợp một số khung nhìn lại với nhau thành một khung duy nhất – một tập bản đồ – để sử dụng tốt hơn các khả năng giải mã. Ví dụ, có quá nhiều bộ giải mã có thể dẫn đến sự cố đồng bộ hóa. Thông thường các tập bản đồ này phân tách các khung nhìn theo một dải tần an toàn và sắp xếp các khung nhìn trên toàn bộ các bộ phận mã hóa.

Bộ mã hóa có thể sử dụng các thông số lượng tử hóa (QP) khác nhau cho mỗi khung nhìn, và/hoặc có thể thực hiện phân bổ tốc độ bit dựa trên tốc độ bit đích cho mỗi vùng trong mỗi khung nhìn. Ví dụ, nếu một vùng trong khung nhìn được xác định là có ít tác động hơn đến chất lượng kết xuất cuối cùng, thì vùng này có thể được mã hóa với ít bit/pixel hơn mức trung bình.

Bộ mã hóa video 300 có thể bao gồm các thành phần của bộ giải mã tham chiếu (bộ giải mã 350, bộ chọn 360 và bộ tổng hợp 370). Bộ giải mã tham chiếu hoặc bộ lọc vòng lặp này có thể kết xuất các cổng nhìn thử nghiệm. Các cổng nhìn thử nghiệm này được so sánh với các khung nhìn tham chiếu. Một khung nhìn tham chiếu có thể là khung nhìn nguồn, hoặc có thể là khung nhìn được tổng hợp trực tiếp từ nguồn.

Lý do sử dụng các khung nhìn đã giải mã là bộ lọc tái cấu trúc có thể được đào tạo để bù đắp cho việc giảm chất lượng do nén các khung nhìn nguồn bị mất. Ngoài ra, có thể xảy ra trường hợp với các khung nhìn không nén, bộ phân tích 320 có thể đánh giá một khung nhìn cụ thể là dư thừa, vì nó có thể được dự đoán chính xác; tuy nhiên, khi sử dụng phiên bản đã giải mã của khung nhìn, do mất độ chính xác của độ sâu/kết cấu, có thể xác định rằng thực sự cần mã hóa khung nhìn (ngay cả khi không độ phân giải không đầy đủ).

Khung nhìn thử nghiệm được tái cấu trúc và khung nhìn tham chiếu được sử dụng làm đầu vào đào tạo để đào tạo hoặc cập nhật các hệ số lọc của bộ lọc tái cấu trúc. Theo một số phương án, các hệ số này được gửi trong chuỗi bit siêu dữ liệu tới bộ giải mã 6DoF 400.

Tuy nhiên, việc đào tạo/cập nhật trực tuyến các hệ số bộ lọc là không cần thiết. Theo một số phương án, tất cả các bộ lọc đều được đào tạo ngoại tuyến (và/hoặc thiết kế/tinh chỉnh bởi các chuyên gia là con người). Trong trường hợp này, có thể không có bộ đào tạo 380 trong bộ mã hóa video 300. Thay vào đó, lỗi giữa khung nhìn thử nghiệm và khung nhìn tham chiếu (trên mỗi khung nhìn hoặc theo không gian) được chuyển tiếp đến bộ phân tích 320, và bộ phân tích có thể sử dụng các bản đồ lỗi này để cải thiện cấu hình, bao gồm cả phân bổ tốc độ bit. Các bản đồ lỗi này có thể dựa trên các số liệu như PSNR, VIF, VMAF, v.v.

Một ví dụ của phương án thứ ba (xem Hình 4) sử dụng mạng nơ ron tích chập với học chuyển tiếp. Có thể đạt được những lợi thế đáng kể bằng cách sử dụng phương pháp học chuyển tiếp. Phương pháp học chuyển tiếp đào tạo một phần của mạng nơ ron ngoại tuyến, nhưng cập nhật việc đào tạo trực tuyến cho một phần khác – ví dụ: đào tạo trực tuyến một vài lớp cuối cùng của mạng nơ ron. Theo cách này, cần ít dữ liệu đào tạo hơn để điều chỉnh mạng nơ ron phù hợp với nội dung cụ thể và chỉ một phần nhỏ hệ số mạng nơ ron cần được phát tới bộ giải mã video 400 – do đó giảm tốc độ bit siêu dữ liệu.

Ngoại tuyến, bên ngoài vòng lặp mã hóa, nhiều mạng nơ-ron tích chập (CNN) có thể được đào tạo trước cho một tập hợp các tình huống dự đoán nhiều khung nhìn được cung cấp. Ví dụ, một CNN được đào tạo sử dụng hai khung nhìn tham chiếu để dự đoán, trong đó một tham chiếu có độ phân giải gốc và tham chiếu còn lại có độ phân giải bằng nửa độ phân giải gốc. CNN thứ hai được đào tạo để sử dụng ba khung nhìn tham chiếu để dự đoán với tham chiếu thứ ba có độ phân giải bằng một phần tư độ phân giải gốc. Đối với mỗi tình huống dự đoán nhiều khung nhìn, có một CNN được đào tạo. Bộ mã hóa có thể so sánh các tình huống và chọn tình huống tốt nhất. Lựa chọn cấu hình này sau đó được thêm vào dưới dạng siêu dữ liệu. Lưu ý rằng bản đồ kết cấu và độ sâu không cần phải có cùng độ phân giải – chúng có thể được lọc và giảm tần số lấy mẫu khác nhau.

Các CNN được đào tạo trước có thể được tinh chỉnh trong quá trình mã hóa, để có được mô hình tốt hơn cho một video, tập hợp các khung hình, hoặc một khung hình duy nhất được cung cấp sẵn, sử dụng phương pháp học chuyển tiếp. Để làm điều này, trọng số của phần lớn các bộ lọc (thường là các lớp đầu tiên trong mạng) vẫn cố định, nhưng trọng số của các bộ lọc sau được điều chỉnh. Các trọng số này đặc biệt phù hợp để mô hình hóa các đặc trưng của video đang được mã hóa. Lúc này, siêu dữ liệu chứa tham chiếu đến mạng đã được sử dụng cộng với trọng số video cụ thể cho các lớp đã sửa đổi. Điều này có thể cho phép tốc độ bit siêu dữ liệu thấp hơn nhiều mà không làm giảm chất lượng.

Một lần nữa, theo Hình 6, các thành phần của bộ mã hóa video 400, và phép toán của chúng theo các phương án khác nhau, sẽ được mô tả trong thông tin chi tiết bổ sung như sau.

Bộ giải mã 6DoF 400 thu chuỗi bit 6DoF (hoặc một phần của chuỗi bit này). Theo một số phương án, bộ thu 410 thu tất cả thông tin đã mã hóa, nhưng theo các phương án khác, dự đoán về các cổng nhìn đích được sử dụng để xác định tập con nào của chuỗi bit cơ bản là cần thiết, và thông qua giao tiếp máy khách-máy chủ, máy chủ sẽ chỉ truyền một tập con thông tin cho khách hàng (bộ giải mã video 400). Điều này thể hiện sự cải tiến của tính năng phát trực tuyến thích ứng như được sử dụng trước đây cho video 2D.

Theo một số phương án, bộ giải mã 420 chỉ đơn giản giải mã tất cả các chuỗi bit (sơ cấp) đã được thu. Tuy nhiên, có thể mong muốn áp dụng giải mã thích ứng. Vì độ trễ giữa bộ giải mã 420 và màn hình hiển thị nhỏ hơn nhiều so với giữa bộ mã hóa 6DoF 300 và bộ giải mã 6DoF 400, bộ giải mã có quyền truy cập vào các dự đoán cổng nhìn đích chính xác hơn. Do đó, bộ giải mã 420 có thể xác định rằng chỉ một

phần dữ liệu thu được phải được giải mã. Điều này có thể giúp giảm mức độ phức tạp và/hoặc mức sử dụng năng lượng của thiết bị khách (bộ giải mã 6DoF 400).

Bộ tổng hợp 430 có khung nhìn đã giải mã và cổng nhìn đích, và sử dụng tổng hợp khung nhìn để dự đoán khung nhìn đích. Theo một phương án, khung nhìn đích có thể được xác định theo hướng của tai nghe thực tế ảo (VR). Đối với tai nghe VR âm thanh nổi, có thể có một cổng nhìn đích cho mỗi mắt. Theo một phương án khác, cổng nhìn đích có thể được tính toán bằng cách sử dụng bộ phận đo lường quán tính (IMU) của thiết bị cầm tay như điện thoại thông minh hoặc máy tính bảng. Trong trường hợp này, cổng nhìn sẽ phụ thuộc vào tư thế của thiết bị. Ví dụ, phương pháp này có thể được sử dụng cho “thực tế tăng cường (AR) dựa trên máy tính bảng”.

Như đã đề cập ở trên, có thể thông tin độ sâu và thông tin kết cấu là từ các khung nhìn khác nhau, trong trường hợp đó độ sâu có thể được chiếu trước, và kết cấu được tìm nạp tiếp theo.

Trong quá trình tổng hợp khung nhìn, có thể xảy ra trường hợp một số thông tin bị thiếu trong khung nhìn đích (hiển thị), hoặc một số thông tin sẽ ở độ phân giải thấp hơn (kéo giãn). Bộ tổng hợp 430 có thể xuất ra các số liệu chất lượng như bản đồ độ tin cậy, hoặc các bản đồ cụ thể hơn như bản đồ độ phân giải, bản đồ góc tia (=bản đồ các góc giữa điểm, cổng nhìn nguồn và đích), bản đồ kéo dài, v.v. Sau đó, thông tin bổ sung này là đầu vào bộ nâng cấp 440, và tùy chọn bộ đào tạo 380 ở bộ mã hóa 6DoF 300.

Quá trình tổng hợp khung nhìn có thể thực hiện nâng cấp cùng với quá trình tổng hợp khung nhìn, hoặc có thể tạo các khung nhìn tổng hợp ở độ phân giải tương tự như khung nhìn đầu vào. Trong trường hợp thứ hai, việc nâng cấp được giao cho bộ nâng cấp 440.

Thông thường, bộ nâng cấp 440 bao gồm một bộ lọc hình ảnh. Bộ nâng cấp này thực hiện nâng cấp (nếu bộ tổng hợp 430 chưa thực hiện nâng cấp). Nó có thể bù đắp cho các tần số bị thiếu. Nó thường sẽ thực hiện một số pha trộn giữa các khung nhìn.

Bộ nâng cấp thường sẽ là một bộ lọc được đào tạo sử dụng các hệ số, được xác định trước và/hoặc được thu trong chuỗi bit siêu dữ liệu, để chuyển các khung nhìn nguồn tổng hợp thành một hình ảnh duy nhất trên mỗi cổng nhìn.

Bộ phân tích 320 tại bộ mã hóa 6DoF 300 có thể xác định rằng kết cấu của khung nhìn là cần thiết, nhưng thông tin về độ sâu từ khung nhìn gần đó là đủ. Trong trường hợp đó, thành phần độ sâu của khung nhìn có thể bị giảm. Trong một tình huống

khác, bộ phân tích 320 có thể xác định rằng thành phần độ sâu của một khung nhìn là cần thiết để lập mô hình chính xác các chuyển đổi tiền cảnh-hậu cảnh, nhưng kết cấu của khung nhìn gần đó đủ để tái cấu trúc bản đồ độ sâu tại bộ giải mã 6DoF 400. Trong cả hai trường hợp, bộ nâng cấp 440 được thiết kế sao cho có thể chấp nhận khung nhìn có các thành phần bị thiếu. Trong trường hợp xấu, chỉ có một bản đồ độ sâu hoặc lưới (đôi khi được gọi là proxy hình học) và nhiều kết cấu. (Ít nhất một trong các kết cấu sẽ có độ phân giải thấp hơn ít nhất một kết cấu khác).

Như đã đề cập trong mô tả của bộ thu 410 và bộ giải mã 420, chúng tôi hy vọng rằng chỉ một phần dữ liệu có sẵn sẽ được sử dụng trong bộ giải mã 6DoF 400. Các khung nhìn gần cổng nhìn đích mang nhiều thông tin hơn về cổng nhìn đích so với các khung nhìn ở xa hơn.

Độ phức tạp của bộ lọc đào tạo có thể được giảm bớt bằng cách chỉ cho phép số lượng khung nhìn cụ thể làm đầu vào. Ví dụ, một CNN có thể được đào tạo để có một khung nhìn đầy đủ, một được giảm quy mô 2x và một được giảm quy mô 4x. Có thể có nhiều mạng phụ thuộc vào việc điều chỉnh quy mô nào các trong số đó là gần nhất, gần thứ 2, v.v. với cổng nhìn đích.

Trong trường hợp này, ngay cả khi tạm thời có nhiều khung nhìn hơn, bộ chọn 450 vẫn sẽ chọn số lượng khung nhìn cố định. Bộ mã hóa 6DoF 300 có thể sử dụng thuộc tính này để đào tạo các bộ lọc cụ thể hơn.

Phương pháp codec có thể xác định tập hợp con các khung nhìn nào đang hoạt động trong một tình huống nhất định. Đây có thể là một thuật toán được triển khai trong bộ chọn 450 (ví dụ: quy tắc như "luôn lấy ba khung nhìn gần nhất"), hoặc nó có thể là thứ mà bộ phân tích 320 xác định và phát dưới dạng siêu dữ liệu. Ví dụ, siêu dữ liệu có thể có bất kỳ dạng nào sau đây:

- Chỉ định tập hợp con cho mỗi điểm lưới trong không gian 3D,
- Chia không gian 3D thành các ô trong đó mỗi ô có tập con là thuộc tính,
- Thuật toán (quy tắc lựa chọn) với các thông số được phát dưới dạng siêu dữ liệu.

Lưu ý rằng, mặc dù hầu hết các phương án được mô tả ở trên đã sử dụng tính năng lọc và giảm tần số lấy mẫu toàn bộ các khung nhìn như một phương án thay thế cho việc cắt tia dựa trên đoạn nối tạm, nhưng có thể các kỹ thuật này có thể được kết hợp, theo một số phương án. Ví dụ, bộ mã hóa 6DoF có thể triển khai cả hai kỹ thuật

và có thể chọn kỹ thuật nào trong số chúng mang lại hiệu quả mã hóa tốt nhất khi áp dụng cho một chuỗi video nhất định.

Các phương pháp mã hóa và giải mã theo Hình 1, 4, 5, 7 và 8, và bộ mã hóa và giải mã của theo Hình 2 và 6, có thể được triển khai trong phần cứng hoặc phần mềm, hoặc tổ hợp cả hai (ví dụ: như phần mềm điều khiển nhúng - firmware chạy trên thiết bị phần cứng). Trong phạm vi mà phương án được triển khai một phần hoặc toàn bộ trong phần mềm, các bước chức năng được minh họa trong lưu đồ quy trình có thể được thực hiện bởi các thiết bị tính toán vật lý được lập trình phù hợp, như một hoặc nhiều đơn vị xử lý trung tâm (CPU) hoặc đơn vị xử lý đồ họa (GPU), hoặc bộ gia tốc mạng nơron (NNA). Mỗi quy trình – và các bước thành phần riêng lẻ của nó như được minh họa trong các lưu đồ – có thể được thực hiện bởi các thiết bị tính toán giống nhau hoặc khác nhau. Theo các phương án, phương tiện lưu trữ có thể đọc được trên máy tính lưu trữ chương trình máy tính bao gồm mã chương trình máy tính được tạo cấu hình để làm cho một hoặc nhiều thiết bị tính toán vật lý thực hiện phương pháp mã hóa hoặc giải mã như được mô tả ở trên khi chương trình được chạy trên một hoặc nhiều thiết bị tính toán vật lý.

Phương tiện lưu trữ có thể bao gồm bộ nhớ máy tính cố định và không cố định như RAM, PROM, EPROM và EEPROM. Các phương tiện lưu trữ khác nhau có thể được cố định trong thiết bị tính toán hoặc có thể được vận chuyển, sao cho một hoặc nhiều chương trình được lưu trữ trên đó có thể được tải vào bộ xử lý.

Siêu dữ liệu theo một phương án có thể được lưu trữ trên phương tiện lưu trữ. Chuỗi bit theo một phương án có thể được lưu trữ trên cùng một phương tiện lưu trữ hoặc phương tiện lưu trữ khác. Siêu dữ liệu có thể được nhúng vào chuỗi bit nhưng điều này không cần thiết. Tương tự như vậy, siêu dữ liệu và/hoặc chuỗi bit (với siêu dữ liệu trong chuỗi bit hoặc tách biệt với nó) có thể được phát dưới dạng tín hiệu được điều chế thành sóng mang điện từ. Tín hiệu có thể được xác định theo tiêu chuẩn cho truyền thông kỹ thuật số. Sóng mang có thể là sóng mang quang, sóng tần số vô tuyến, sóng milimét hoặc sóng thông tin trường gần. Nó có thể có dây hoặc không dây.

Trong phạm vi mà phương án được thực hiện một phần hoặc toàn bộ trong phần cứng, các khối được minh họa trong sơ đồ khối trong Hình 2 và 6 có thể là các thành phần vật lý riêng biệt, hoặc phân chia logic của các thành phần vật lý đơn lẻ, hoặc tất cả có thể được thực hiện theo phương thức được tích hợp trong một thành phần vật lý. Các hàm của

một khối hiển thị trong hình vẽ có thể được phân chia giữa nhiều thành phần trong phương thức triển khai hoặc các hàm của nhiều khối minh họa trong hình vẽ có thể được kết hợp thành các thành phần đơn lẻ trong phương thức triển khai. Các thành phần phần cứng phù hợp để sử dụng theo các phương án của sáng chế bao gồm, nhưng không giới hạn ở, bộ vi xử lý thông thường, mạch tích hợp chuyên dụng (ASIC) và mảng cổng có thể lập trình trường (FPGA). Một hoặc nhiều khối có thể được thực hiện như một sự kết hợp của phần cứng chuyên dụng để thực hiện một số chức năng và một hoặc nhiều bộ vi xử lý được lập trình và hệ mạch liên kết để thực hiện các chức năng khác.

Nói chung, các ví dụ về các phương pháp mã hóa và giải mã dữ liệu, chương trình máy tính triển khai các phương pháp này, bộ mã hóa và giải mã video, và chuỗi bit được chỉ ra bởi các phương án bên dưới.

Phương án:

1. Phương pháp mã hóa video sống động, phương pháp này bao gồm các bước:
 - bước thu dữ liệu video nguồn, bao gồm nhiều khung nhìn nguồn, mỗi khung nhìn nguồn bao gồm dữ liệu kết cấu và dữ liệu độ sâu;
 - bước xử lý dữ liệu video nguồn để tạo ra dữ liệu video nguồn đã xử lý, bước xử lý này bao gồm bước giảm tần số lấy mẫu (140) một hoặc nhiều khung nhìn nguồn; và
 - bước mã hóa (150) dữ liệu video nguồn đã xử lý để tạo ra một chuỗi bit video và chuỗi bit siêu dữ liệu,
 - trong đó chuỗi bit video bao gồm dữ liệu kết cấu được mã hóa và dữ liệu độ sâu được mã hóa, và chuỗi bit siêu dữ liệu bao gồm siêu dữ liệu để hỗ trợ bộ giải mã giải mã chuỗi bit video, trong đó siêu dữ liệu này mô tả cấu hình của bước xử lý được thực hiện trên một hoặc nhiều khung nhìn nguồn.
2. Phương pháp theo phương án 1, còn bao gồm thêm bước phân tích (120) dữ liệu video nguồn để chọn cấu hình cho bước xử lý, bước xử lý này bao gồm bước lọc và giảm tần số lấy mẫu một hoặc nhiều khung nhìn nguồn theo cấu hình đã chọn.
3. Phương pháp theo phương án 2, trong đó bước phân tích (120) bao gồm:
 - bước phân tích sự tương đồng của nội dung trực quan giữa các khung nhìn;
 - bước gán hệ số giảm tần số lấy mẫu thứ nhất cho khung nhìn thứ nhất;
 - bước gán hệ số giảm tần số lấy mẫu thứ hai cho khung nhìn thứ hai; và
 - bước gán hệ số giảm tần số lấy mẫu cho khung nhìn thứ ba,

trong đó hệ số giảm tần số lấy mẫu thứ hai cao hơn hệ số giảm tần số lấy mẫu thứ ba nếu sự tương đồng giữa khung nhìn thứ nhất và khung nhìn thứ hai lớn hơn sự tương đồng giữa khung nhìn thứ nhất và khung nhìn thứ ba.

4. Phương pháp theo phương án 2 hoặc phương án 3, trong đó bước phân tích bao gồm:

bước thu thập (122), dựa trên dữ liệu video nguồn, ít nhất một khung nhìn tham chiếu,

bước lọc (130) và bước giảm tần số lấy mẫu (140) một hoặc nhiều khung nhìn nguồn theo nhiều cấu hình khác nhau,

đối với từng cấu hình trong nhiều cấu hình khác nhau:

bước mã hóa (150) dữ liệu video nguồn đã xử lý tương ứng để tạo ra chuỗi video ứng viên;

bước giải mã (124) từ chuỗi bit video ứng viên một hoặc nhiều khung nhìn nguồn;

bước tái cấu trúc (126) ít nhất một khung nhìn thử nghiệm từ một hoặc nhiều khung nhìn nguồn đã giải mã, ít nhất một khung nhìn thử nghiệm này tương ứng với ít nhất một khung nhìn tham chiếu; và

bước so sánh (128) khung nhìn thử nghiệm đã tái cấu trúc với khung nhìn tham chiếu,

bước phân tích (120a) còn bao gồm thêm bước chọn cấu hình dựa trên kết quả của so sánh,

phương pháp còn bao gồm bước chọn, như chuỗi bit video, chuỗi bit video ứng viên liên quan đến cấu hình đã chọn.

5. Phương pháp theo phương án bất kỳ trong số các phương án trên, còn bao gồm thêm bước thiết kế (129) bộ lọc tái cấu trúc sử dụng thuật toán học máy,

trong đó bộ lọc tái cấu trúc này có thể được bộ giải mã sử dụng khi giảm tần số lấy mẫu và lọc một hoặc nhiều khung nhìn đã giải mã, và

siêu dữ liệu còn bao gồm phần mô tả của bộ lọc tái cấu trúc.

6. Phương pháp theo điểm 5, bao gồm các bước:

bước thu thập (122), dựa trên dữ liệu video nguồn, ít nhất một khung nhìn tham chiếu,

bước giải mã (124) từ chuỗi bit video một hoặc nhiều khung nhìn nguồn, và

bước xác định khung nhìn thử nghiệm tương ứng với khung nhìn tham chiếu, khung nhìn thử nghiệm này được tái cấu trúc (126) từ một hoặc nhiều khung nhìn nguồn đã giải mã,

trong đó bước thiết kế bộ lọc tái cấu trúc bao gồm bước đào tạo hệ số bộ lọc của bộ lọc tái cấu trúc, sử dụng thuật toán học máy, để giảm lỗi giữa khung nhìn thử nghiệm và khung nhìn tham chiếu.

7. Phương pháp giải mã video sống động, phương pháp này bao gồm các bước:

bước thu (210) chuỗi bit video và chuỗi bit siêu dữ liệu, trong đó chuỗi bit video bao gồm dữ liệu kết cấu đã mã hóa và dữ liệu độ sâu đã mã hóa cho mỗi trong số nhiều khung nhìn nguồn, và chuỗi bit siêu dữ liệu này bao gồm siêu dữ liệu để hỗ trợ giải mã chuỗi bit video;

bước thu thập thông số công nhìn của ít nhất một khung nhìn đích được sẽ được kết xuất;

bước giải mã (220) từ chuỗi bit video một hoặc nhiều khung nhìn nguồn; và

bước tái cấu trúc ít nhất một khung nhìn đích từ một hoặc nhiều khung nhìn nguồn đã giải mã, dựa trên các thông số công nhìn thu được,

trong đó siêu dữ liệu mô tả cấu hình của bước giảm tần số lấy mẫu được thực hiện khi mã hóa một hoặc nhiều khung nhìn nguồn, và

bước tái cấu trúc bao gồm bước tăng tần số lấy mẫu (230) và bước lọc (240), dựa trên cấu hình được mô tả, ít nhất một trong số một hoặc nhiều khung nhìn nguồn đã giải mã hoặc nguồn gốc của chúng.

8. Phương pháp theo phương án 7, bao gồm:

bước giải mã (220a) từ chuỗi bit video ít nhất hai trong số nhiều khung nhìn nguồn; và

đối với mỗi khung nhìn trong số ít nhất hai khung nhìn nguồn đã giải mã, bước tổng hợp (225) khung nhìn đích tổng hợp, dựa trên các thông số công nhìn,

trong đó bước tái cấu trúc ít nhất một khung nhìn đích bao gồm bước tăng tần số lấy mẫu (232, 234), bước lọc (242, 244), và bước kết hợp (260) các khung nhìn đích tổng hợp, theo cấu hình đã mô tả trong siêu dữ liệu.

9. Phương pháp theo phương án 7 hoặc phương án 8, trong đó siêu dữ liệu bao gồm mô tả bộ lọc tái cấu trúc sẽ được sử dụng trong bước tăng tần số lấy mẫu và bước lọc,

và bước tăng tần số lấy mẫu (230, 232, 234) và bước lọc (240, 242, 244) bao gồm sử dụng bộ lọc tái cấu trúc đã mô tả.

10. Phương pháp theo phương án bất kỳ trong số các phương án từ 7 đến 9, trong đó chuỗi bit video thu được từ bộ mã hóa từ xa, phương pháp này còn bao gồm thêm bước phát (250) các thông số cổng nhìn cho bộ mã hóa này.

11. Chương trình máy tính bao gồm mã máy tính làm cho hệ thống xử lý thực hiện phương pháp theo phương án bất kỳ trong số các phương án từ 1 đến 10 khi chương trình này được chạy trên hệ thống xử lý này.

12. Bộ mã hóa video (300) được tạo cấu hình để mã hóa video sống động, bộ mã hóa này bao gồm:

đầu vào (310), được tạo cấu hình để thu dữ liệu video nguồn, bao gồm nhiều khung nhìn nguồn, mỗi khung nhìn nguồn bao gồm dữ liệu kết cấu và dữ liệu độ sâu;

bộ xử lý (300), được tạo cấu hình để xử lý dữ liệu video nguồn để tạo dữ liệu video nguồn đã xử lý, bước bao gồm bước giảm tần số lấy mẫu một hoặc nhiều khung nhìn nguồn;

bộ mã hóa (340), được tạo cấu hình để mã hóa dữ liệu video nguồn đã xử lý để tạo chuỗi bit video và chuỗi bit siêu dữ liệu; và

đầu ra, được tạo cấu hình để xuất chuỗi bit video và chuỗi bit siêu dữ liệu,

trong đó chuỗi bit video bao gồm dữ liệu kết cấu được mã hóa và dữ liệu độ sâu được mã hóa, và chuỗi bit siêu dữ liệu bao gồm siêu dữ liệu để hỗ trợ bộ giải mã giải mã chuỗi bit video, trong đó siêu dữ liệu này mô tả cấu hình của bước xử lý được thực hiện trên một hoặc nhiều khung nhìn nguồn.

13. Bộ mã hóa video của phương án 12, còn bao gồm thêm bộ phân tích (320), được tạo cấu hình để phân tích dữ liệu video nguồn để chọn cấu hình cho bộ xử lý video (330), trong đó bộ xử lý video được tạo cấu hình để lọc và giảm tần số lấy mẫu một hoặc nhiều khung nhìn nguồn theo cấu hình đã chọn.

14. Bộ giải mã video (400) được tạo cấu hình để giải mã video sống động, bộ giải mã video này bao gồm:

đầu vào chuỗi bit (410), được tạo cấu hình để thu chuỗi bit video và chuỗi bit siêu dữ liệu, trong đó chuỗi bit video bao gồm dữ liệu kết cấu đã mã hóa và dữ liệu độ sâu đã mã hóa cho mỗi trong số nhiều khung nhìn nguồn, và chuỗi bit siêu dữ liệu bao gồm siêu dữ liệu để hỗ trợ giải mã chuỗi bit video;

đầu vào điều khiển, được tạo cấu hình để thu thông số công nhìn của ít nhất một khung nhìn đích sẽ được kết xuất;

bộ giải mã (420), được tạo cấu hình để giải mã từ chuỗi bit video một hoặc nhiều khung nhìn nguồn; và

bộ xử lý tái cấu trúc (430, 440) được tạo cấu hình để tái cấu trúc ít nhất một khung nhìn đích từ một hoặc nhiều khung nhìn nguồn đã giải mã, dựa trên thông số công nhìn thu được, trong đó bước tái cấu trúc bao gồm bước tăng tần số lấy mẫu và bước lọc, dựa trên siêu dữ liệu, ít nhất là một trong số một hoặc nhiều khung nhìn nguồn đã giải mã hoặc nguồn gốc của nó.

15. Chuỗi bit mã hóa video sống động, chuỗi bit này bao gồm:

chuỗi bit video, bao gồm dữ liệu kết cấu đã mã hóa và dữ liệu độ sâu đã mã hóa cho mỗi trong số nhiều khung nhìn nguồn; và

chuỗi bit siêu dữ liệu, bao gồm siêu dữ liệu để hỗ trợ bộ giải mã giải mã chuỗi bit video,

trong đó siêu dữ liệu này mô tả cấu hình bước giảm tần số lấy mẫu được thực hiện trên một hoặc nhiều khung nhìn nguồn trong quy trình mã hóa chuỗi bit video.

Cụ thể hơn, sáng chế được xác định bởi các điểm yêu cầu bảo hộ kèm theo.

Các biến thể của các phương án được đề xuất có thể được những người có trình độ trong ngành hiểu và thực hành khi thực hiện sáng chế, từ việc nghiên cứu các hình vẽ, bản công bố và các điểm yêu cầu bảo hộ đi kèm. Trong các điểm yêu cầu bảo hộ, từ “bao gồm” không loại trừ các phần tử hoặc bước khác, và mào từ bất định “một” không loại trừ số nhiều. Bộ xử lý đơn hoặc bộ phận khác có thể hoàn thành được các chức năng của một số hạng mục được nêu trong các điểm yêu cầu bảo hộ. Việc một số biện pháp nhất định được nêu trong các điểm yêu cầu bảo hộ phụ thuộc khác nhau không có nghĩa là không được sử dụng kết hợp các biện pháp này nhằm mang lại lợi ích. Nếu chương trình máy tính được trình bày ở trên, nó có thể được lưu trữ/phân phối trên phương tiện phù hợp, chẳng hạn như phương tiện lưu trữ quang hoặc phương tiện thể rắn được cung cấp cùng hoặc như là một phần của phần cứng khác, nhưng cũng có thể được phân phối dưới các hình thức khác, chẳng hạn như thông qua Internet hoặc các hệ thống viễn thông có dây hoặc không dây khác. Nếu thuật ngữ “được điều chỉnh theo” được sử dụng trong các điểm này hoặc trong mô tả, lưu ý rằng thuật ngữ “được điều chỉnh theo” được nhằm

mục đích tương đương với thuật ngữ “được tạo cấu hình theo”. Không nên hiểu các ký hiệu chỉ dẫn bất kỳ trong các điểm yêu cầu bảo hộ là giới hạn phạm vi của sáng chế.

YÊU CẦU BẢO HỘ

1. Phương pháp mã hóa video sống động, phương pháp này bao gồm các bước:
 - bước thu (110) dữ liệu video nguồn, bao gồm nhiều khung nhìn nguồn bao gồm ít nhất ba khung nhìn nguồn, mỗi khung nhìn nguồn bao gồm dữ liệu kết cấu và dữ liệu độ sâu;
 - bước phân tích (120) dữ liệu video nguồn để chọn cấu hình cho bước xử lý dữ liệu video nguồn;
 - bước xử lý dữ liệu video nguồn để tạo dữ liệu video nguồn đã xử lý, trong đó bước xử lý này bao gồm bước giảm tần số lấy mẫu (140) một hoặc nhiều khung nhìn nguồn trong nhiều khung nhìn nguồn theo cấu hình đã chọn; và
 - bước mã hóa (150) dữ liệu video nguồn đã xử lý cho nhiều khung nhìn nguồn để tạo ra một chuỗi bit video và chuỗi bit siêu dữ liệu,
 - trong đó chuỗi bit video này bao gồm dữ liệu kết cấu đã mã hóa và dữ liệu độ sâu đã mã hóa, và chuỗi bit siêu dữ liệu bao gồm siêu dữ liệu để hỗ trợ bộ giải mã giải mã chuỗi bit video, trong đó siêu dữ liệu mô tả cấu hình của bước giảm tần số lấy mẫu được thực hiện trên một hoặc nhiều khung nhìn nguồn, và
 - trong đó bước phân tích (120) bao gồm:
 - bước phân tích sự tương đồng của nội dung trực quan giữa các khung nhìn nguồn;
 - bước gán hệ số giảm tần số lấy mẫu thứ nhất cho khung nhìn thứ nhất;
 - bước gán hệ số giảm tần số lấy mẫu thứ hai cho khung nhìn thứ hai; và
 - bước gán hệ số giảm tần số lấy mẫu thứ ba cho khung nhìn thứ ba,
 trong đó hệ số giảm tần số lấy mẫu thứ hai cao hơn hệ số giảm tần số lấy mẫu thứ ba nếu sự tương đồng của nội dung trực quan giữa khung nhìn thứ nhất và khung nhìn thứ hai lớn hơn sự tương đồng của nội dung trực quan giữa khung nhìn thứ nhất và khung nhìn thứ ba.

2. Phương pháp theo điểm 1, trong đó bước phân tích bao gồm các bước:
 - bước thu thập (122), dựa trên dữ liệu video nguồn, ít nhất một khung nhìn tham chiếu,
 - bước lọc (130) và bước giảm tần số lấy mẫu (140) một hoặc nhiều khung nhìn nguồn theo nhiều cấu hình khác nhau,

đối với từng cấu hình trong nhiều cấu hình khác nhau:

bước mã hóa (150) dữ liệu video nguồn đã xử lý tương ứng để tạo ra chuỗi bit video ứng viên;

bước giải mã (124) từ chuỗi bit video ứng viên một hoặc nhiều khung nhìn nguồn;

bước tái cấu trúc (126) ít nhất một khung nhìn thử nghiệm từ một hoặc nhiều khung nhìn nguồn đã giải mã, ít nhất một khung nhìn thử nghiệm này tương ứng với ít nhất một khung nhìn tham chiếu; và

bước so sánh (128) khung nhìn thử nghiệm đã tái cấu trúc với khung nhìn tham chiếu,

bước phân tích (120a) còn bao gồm thêm bước chọn cấu hình dựa trên các kết quả của bước so sánh,

phương pháp này còn bao gồm bước chọn, như chuỗi bit video, chuỗi bit video ứng viên liên quan đến cấu hình đã chọn.

3. Phương pháp theo điểm 1 hoặc 2, còn bao gồm thêm bước thiết kế (129) bộ lọc tái cấu trúc sử dụng thuật toán học máy,

trong đó bộ lọc tái cấu trúc này có thể được bộ giải mã sử dụng khi giảm tần số lấy mẫu và lọc một hoặc nhiều khung nhìn đã giải mã, và

siêu dữ liệu còn bao gồm phần mô tả của bộ lọc tái cấu trúc.

4. Phương pháp theo điểm 3, bao gồm các bước:

bước thu thập (122), dựa trên dữ liệu video nguồn, ít nhất một khung nhìn tham chiếu,

bước giải mã (124) từ chuỗi bit video một hoặc nhiều khung nhìn nguồn, và

bước xác định khung nhìn thử nghiệm tương ứng với khung nhìn tham chiếu, khung nhìn thử nghiệm này được tái cấu trúc (126) từ một hoặc nhiều khung nhìn nguồn đã giải mã,

trong đó bước thiết kế bộ lọc tái cấu trúc bao gồm bước đào tạo hệ số bộ lọc của bộ lọc tái cấu trúc, sử dụng thuật toán học máy, để giảm lỗi giữa khung nhìn thử nghiệm và khung nhìn tham chiếu.

5. Phương pháp giải mã video sống động, phương pháp này bao gồm các bước:

bước thu (210) chuỗi bit video và chuỗi bit siêu dữ liệu, trong đó chuỗi bit video bao gồm dữ liệu kết cấu đã mã hóa và dữ liệu độ sâu đã mã hóa cho mỗi trong số nhiều khung nhìn nguồn bao gồm ít nhất ba khung nhìn nguồn, và chuỗi bit siêu dữ liệu này bao gồm siêu dữ liệu để hỗ trợ giải mã chuỗi bit video;

bước thu thập thông số công nhìn của ít nhất một khung nhìn đích được sẽ được kết xuất;

bước giải mã (220a) từ chuỗi bit video ít nhất hai trong số nhiều khung nhìn nguồn; và

bước tái cấu trúc ít nhất một khung nhìn đích từ ít nhất hai khung nhìn nguồn đã giải mã trên, dựa trên các thông số công nhìn thu được,

trong đó siêu dữ liệu mô tả cấu hình bước giảm tần số lấy mẫu được thực hiện khi mã hóa ít nhất hai khung nhìn nguồn,

trong đó phương pháp này bao gồm, đối với mỗi khung nhìn trong số ít nhất hai khung nhìn đã giải mã, bước tổng hợp (225) khung nhìn đích tổng hợp, dựa trên các thông số công nhìn, và

trong đó bước tái cấu trúc ít nhất một khung nhìn đích bao gồm bước tăng tần số lấy mẫu (232, 234) và bước lọc (242, 244) ít nhất một trong số các khung nhìn đích tổng hợp, và bước kết hợp (260) các khung nhìn đích tổng hợp này, theo cấu hình đã mô tả trong siêu dữ liệu.

6. Phương pháp theo điểm 5, trong đó siêu dữ liệu bao gồm mô tả bộ lọc tái cấu trúc sẽ được sử dụng trong bước tăng tần số lấy mẫu và bước lọc, và bước tăng tần số lấy mẫu (230, 232, 234) và bước lọc (240, 242, 244) bao gồm bước sử dụng bộ lọc tái cấu trúc đã mô tả.

7. Phương pháp theo điểm bất kỳ trong số các điểm 5 hoặc 6, trong đó chuỗi bit video thu được từ bộ mã hóa từ xa, phương pháp này còn bao gồm thêm bước phát (250) các thông số công nhìn cho bộ mã hóa.

8. Phương tiện có thể đọc được bằng máy tính chứa chương trình máy tính bao gồm mã máy tính làm cho hệ thống xử lý thực hiện phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 7 khi chương trình này được chạy trên hệ thống xử lý này.

9. Bộ mã hóa video (300) được tạo cấu hình để mã hóa video sống động, bộ mã hóa này bao gồm:

đầu vào (310), được tạo cấu hình để thu dữ liệu video nguồn, bao gồm nhiều khung nhìn nguồn bao gồm ít nhất ba khung nhìn nguồn, mỗi khung nhìn nguồn bao gồm dữ liệu kết cấu và dữ liệu độ sâu;

bộ phân tích video (320);

bộ xử lý video (330);

bộ mã hóa (340); và

đầu ra,

bộ phân tích video (320) được tạo cấu hình để phân tích dữ liệu video nguồn để chọn cấu hình cho bộ xử lý video (330);

bộ xử lý video (330) được tạo cấu hình để xử lý dữ liệu video nguồn để tạo dữ liệu video nguồn đã xử lý, bước xử lý này bao gồm bước giảm tần số lấy mẫu một hoặc nhiều khung nhìn nguồn trong nhiều khung nhìn nguồn, trong đó bộ xử lý video được tạo cấu hình để giảm tần số lấy mẫu một hoặc nhiều khung nhìn nguồn theo cấu hình đã chọn;

bộ mã hóa (340) được tạo cấu hình để mã hóa dữ liệu video nguồn đã xử lý cho nhiều khung nhìn nguồn để tạo ra chuỗi bit video và chuỗi bit siêu dữ liệu; và

đầu ra được tạo cấu hình để xuất chuỗi bit video và chuỗi bit siêu dữ liệu,

trong đó chuỗi bit video này bao gồm dữ liệu kết cấu đã mã hóa và dữ liệu độ sâu đã mã hóa, và chuỗi bit siêu dữ liệu bao gồm siêu dữ liệu để hỗ trợ bộ giải mã giải mã chuỗi bit video, trong đó siêu dữ liệu mô tả cấu hình của bước giảm tần số lấy mẫu được thực hiện trên một hoặc nhiều khung nhìn nguồn,

trong đó bộ phân tích video (320), để phân tích dữ liệu video nguồn, được tạo cấu hình để:

phân tích sự tương đồng của nội dung trực quan giữa các khung nhìn nguồn;

gán hệ số giảm tần số lấy mẫu thứ nhất cho khung nhìn thứ nhất;

gán hệ số giảm tần số lấy mẫu thứ hai cho khung nhìn thứ hai; và

gán hệ số giảm tần số lấy mẫu thứ ba cho khung nhìn thứ ba,
 trong đó hệ số giảm tần số lấy mẫu thứ hai cao hơn hệ số giảm tần số lấy mẫu thứ ba nếu sự tương đồng của nội dung trực quan giữa khung nhìn thứ nhất và khung nhìn thứ hai lớn hơn sự tương đồng của nội dung trực quan giữa khung nhìn thứ nhất và khung nhìn thứ ba.

10. Bộ giải mã video (400) được tạo cấu hình để giải mã video sống động, bộ giải mã video này bao gồm:

đầu vào chuỗi bit (410), được tạo cấu hình để thu chuỗi bit video và chuỗi bit siêu dữ liệu, trong đó chuỗi bit video bao gồm dữ liệu kết cấu đã mã hóa và dữ liệu độ sâu đã mã hóa cho mỗi trong số nhiều khung nhìn nguồn bao gồm ít nhất ba khung nhìn nguồn, và chuỗi bit siêu dữ liệu bao gồm siêu dữ liệu để hỗ trợ giải mã chuỗi bit video;

đầu vào điều khiển, được tạo cấu hình để thu thông số cổng nhìn của ít nhất một khung nhìn đích sẽ được kết xuất;

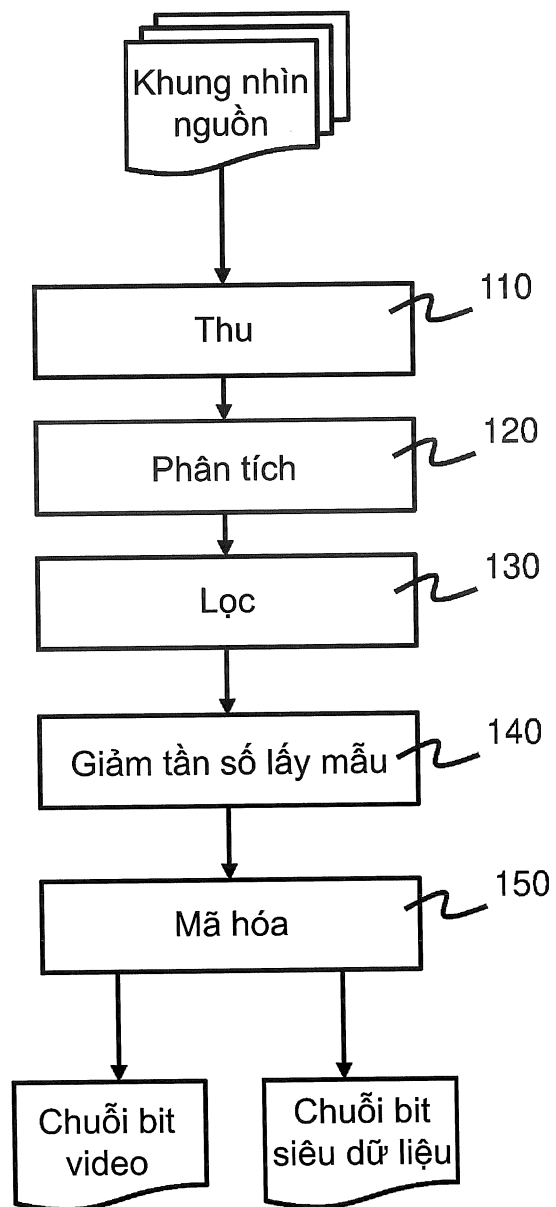
bộ giải mã (420), được tạo cấu hình để giải mã từ chuỗi bit video ít nhất hai trong số nhiều khung nhìn nguồn; và

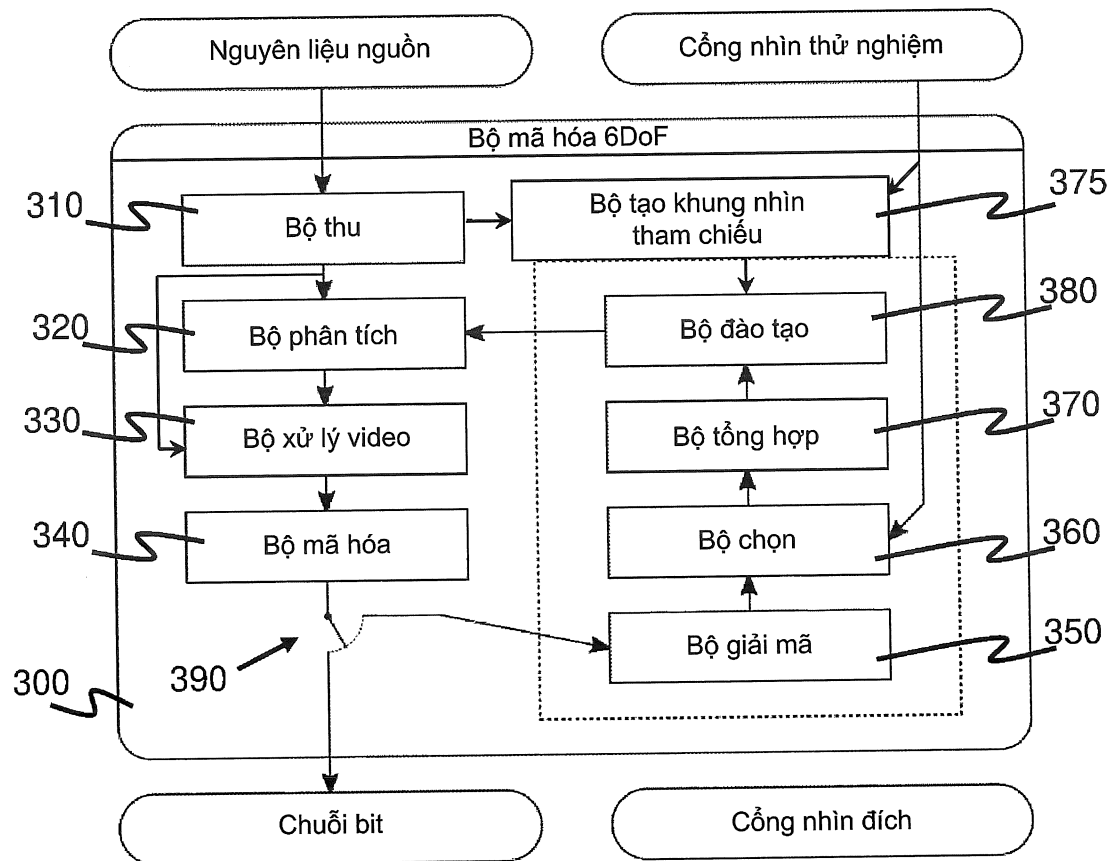
bộ xử lý tái cấu trúc (430, 440), được tạo cấu hình để tái cấu trúc ít nhất một khung nhìn đích từ ít nhất hai khung nhìn nguồn đã giải mã, dựa trên các thông số cổng nhìn thu được,

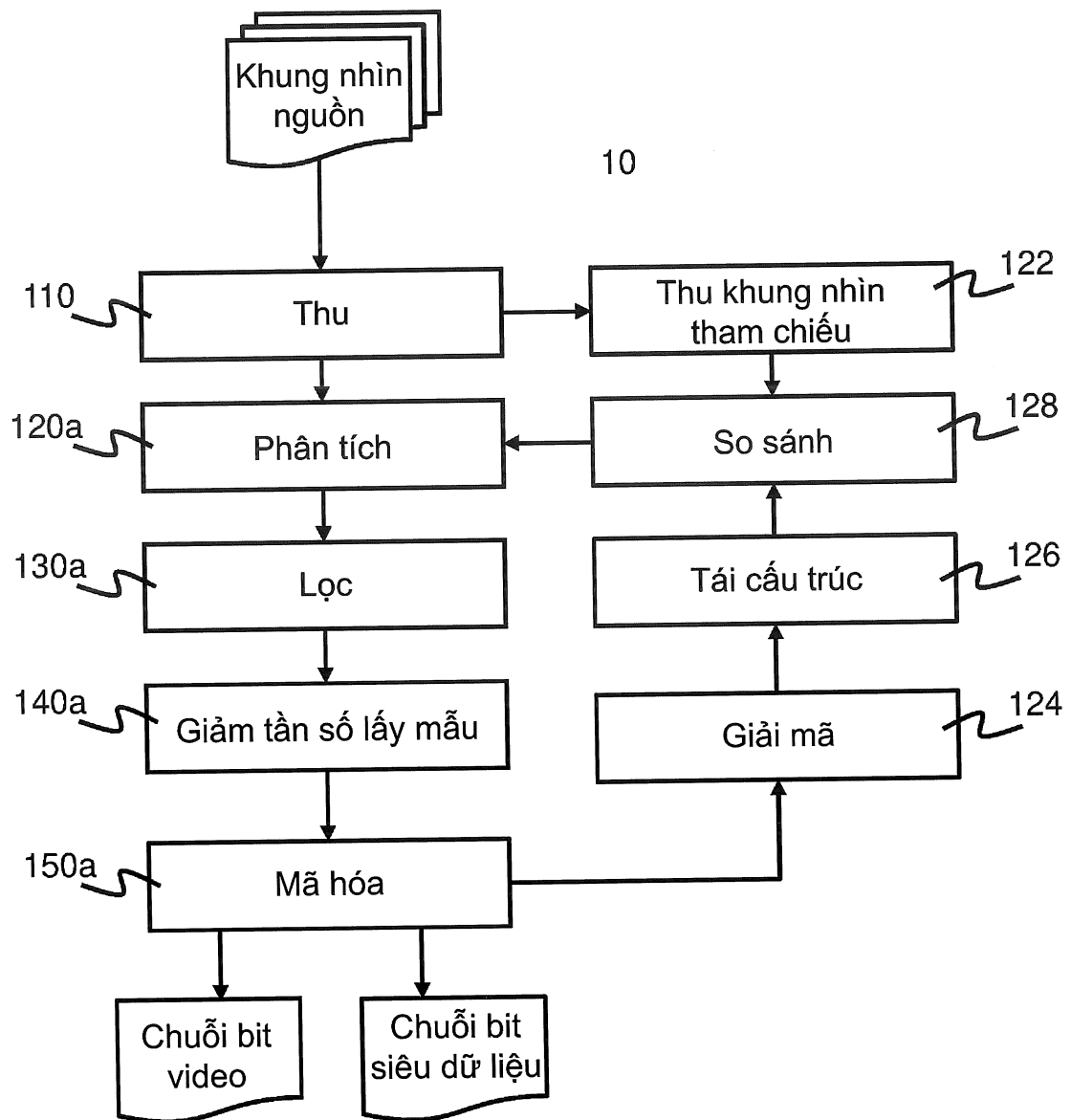
trong đó bộ xử lý tái cấu trúc bao gồm bộ tổng hợp được tạo cấu hình để tổng hợp khung nhìn đích tổng hợp từ mỗi khung nhìn trong số ít nhất hai khung nhìn nguồn đã giải mã; và

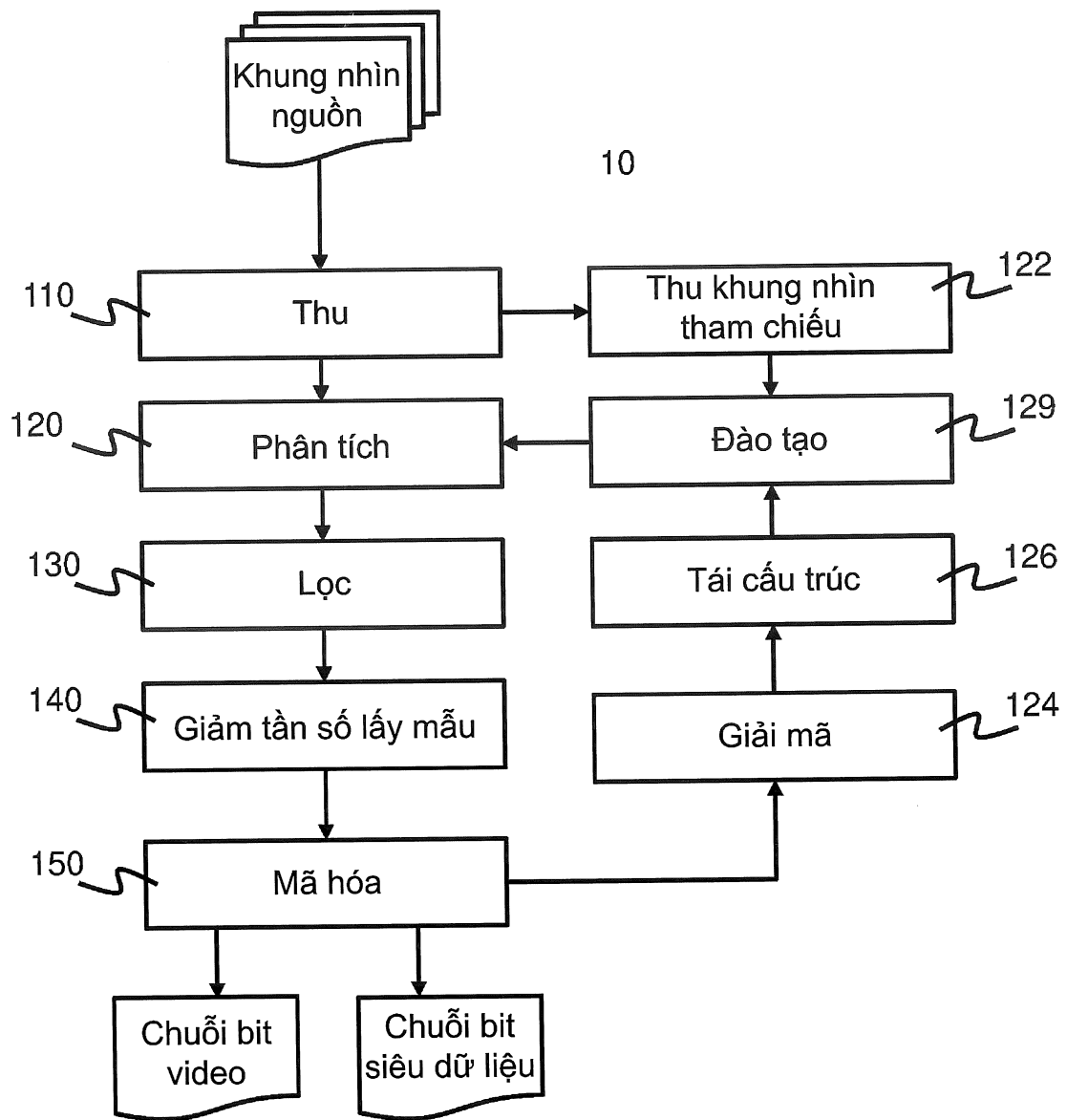
bộ nâng cấp được tạo cấu hình để tăng tần số lấy mẫu và lọc ít nhất một trong số các khung nhìn đích tổng hợp, và kết hợp các khung nhìn đích tổng hợp lại với nhau,

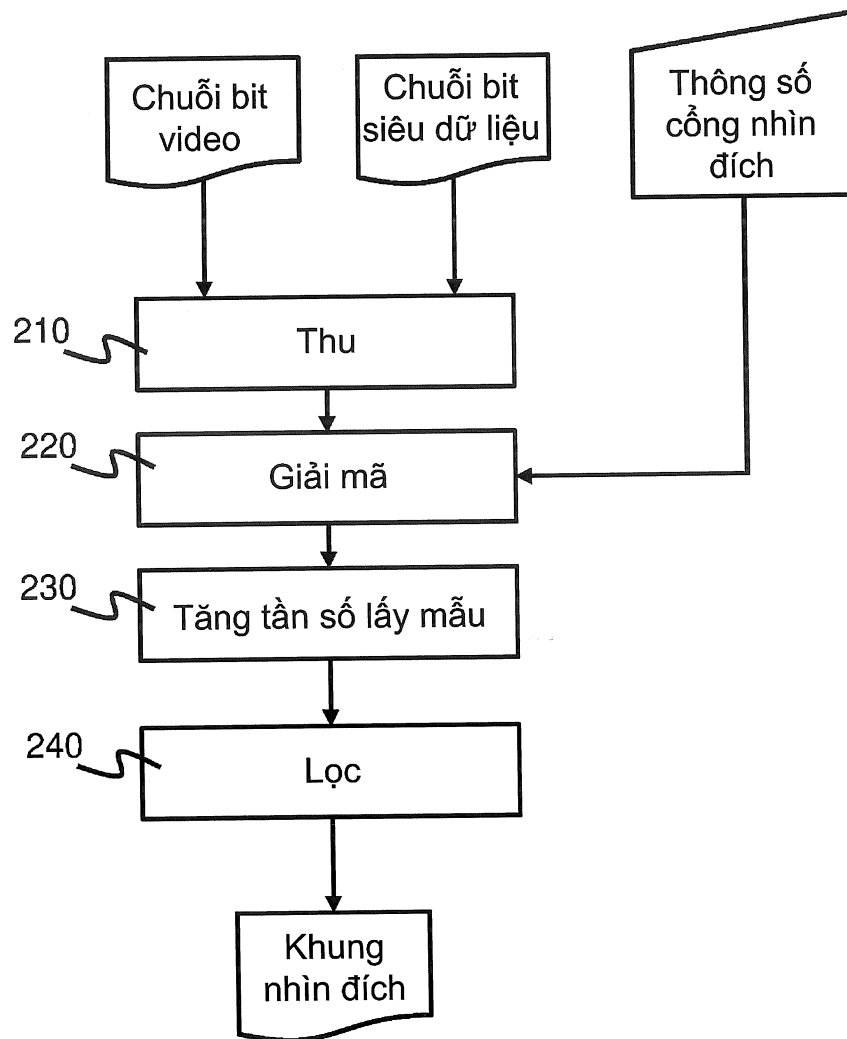
trong đó siêu dữ liệu mô tả cấu hình của việc giảm tần số lấy mẫu được thực hiện khi mã hóa ít nhất hai khung nhìn nguồn.

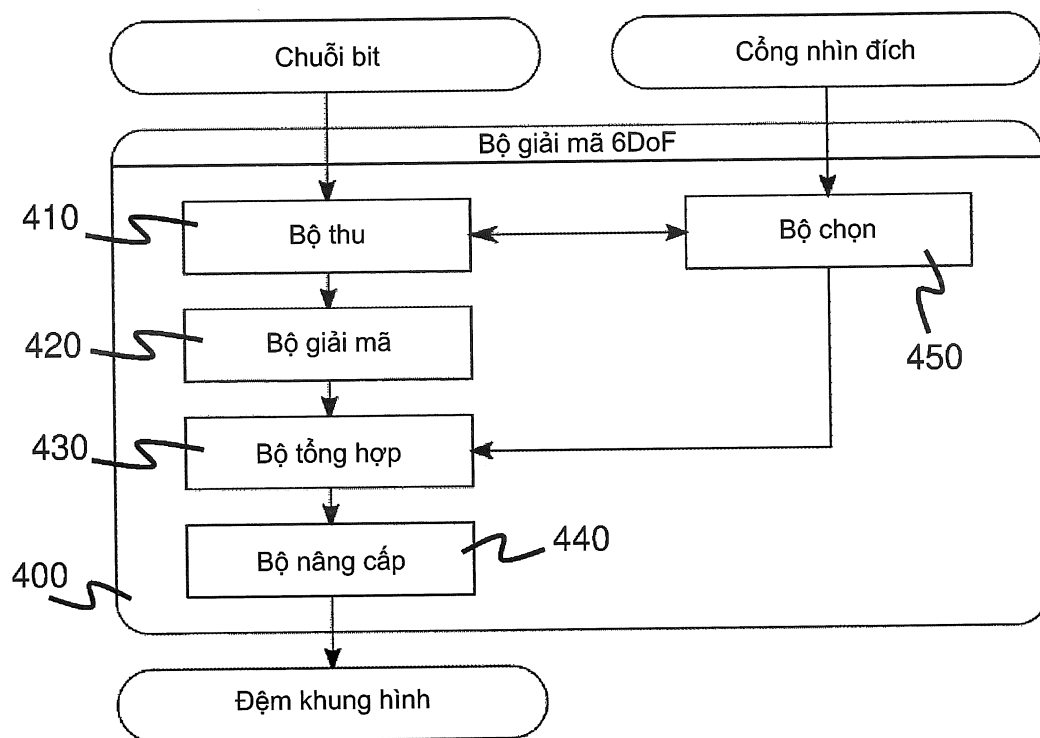
**HÌNH 1**

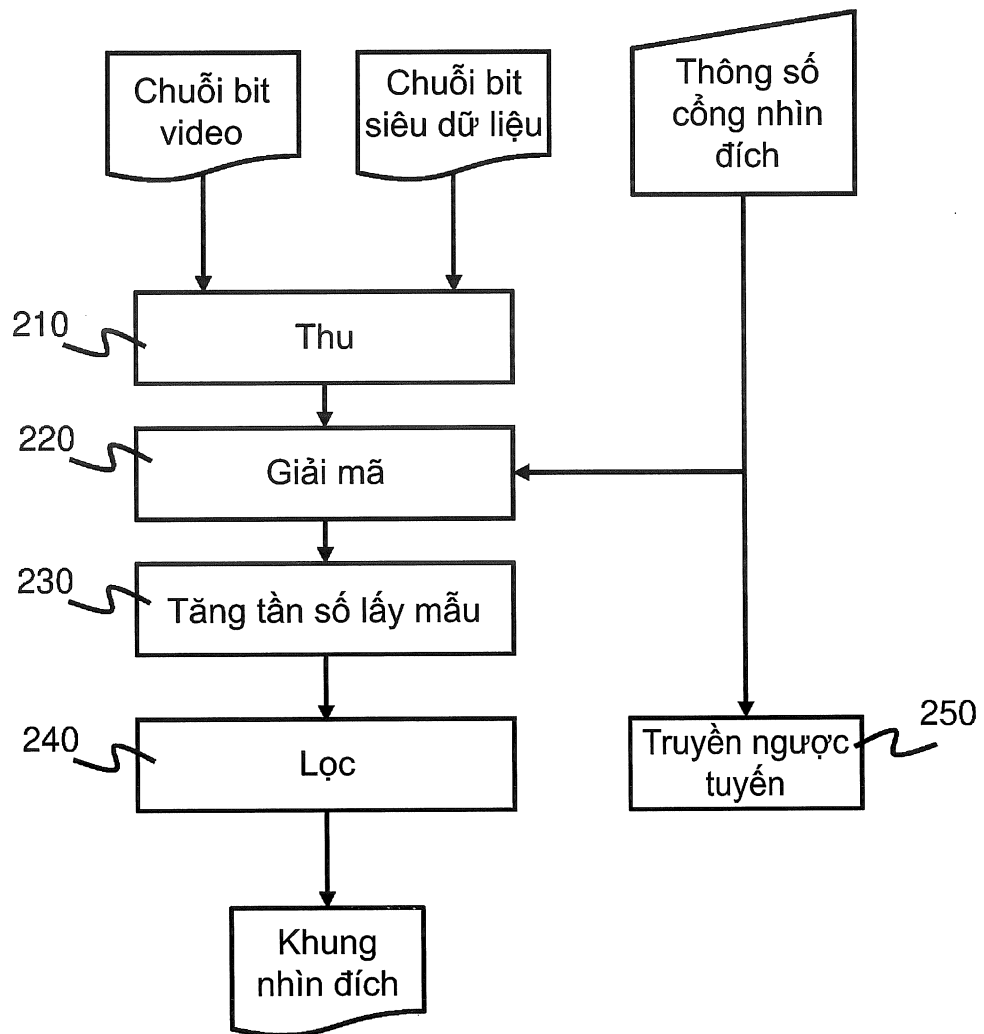
**HÌNH 2**

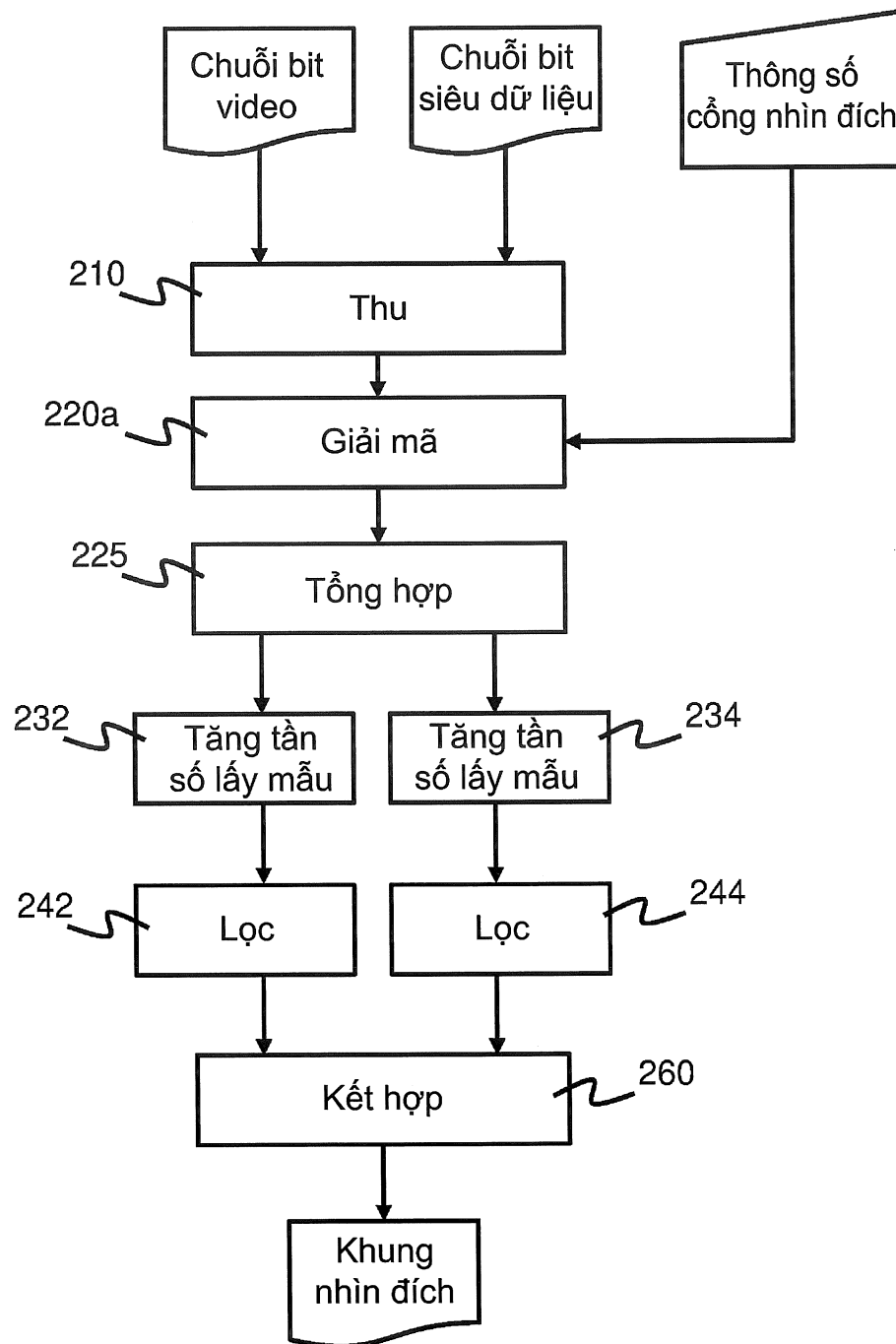
**HÌNH 3**

**HÌNH 4**

**HÌNH 5**

**HÌNH 6**

**HÌNH 7**

**HÌNH 8**