

GIẢM THỜI GIAN MÃ HÓA CHO CHUẨN NÉN VIDEO HEVC DỰA TRÊN VIỆC CẢI TIẾN THUẬT TOÁN TÌM KIẾM TZ

Vũ Hữu Tiến, Nguyễn Thị Hương Thảo
Học viện Công nghệ Bưu chính Viễn thông

Tóm tắt: Nhu cầu sử dụng video độ phân giải cao và đặc biệt là sử dụng video chất lượng cao trên các thiết bị di động đã dẫn đến yêu cầu ngày càng lớn đối với các bộ mã hóa video hiệu suất cao. Đứng trước yêu cầu đó, các chuẩn mã hóa video mới được phát triển không ngừng. Chuẩn mã hóa video hiệu suất cao HEVC đạt được mức nén tăng 50% so với chuẩn H.264 trước đó với chất lượng video tương đương. Trong các chuẩn mã hóa video dự đoán này, ước lượng chuyển động (Motion Estimation - ME) là một trong các mô-đun quan trọng nhất quyết định đến hiệu suất nén bộ mã hóa video. Tuy nhiên nó cũng chiếm 50%-70% độ phức tạp mã hóa của toàn bộ bộ mã hóa. Sự tiêu tốn tài nguyên tính toán khá lớn làm giảm hiệu năng của bộ mã hóa, đặc biệt đối với video ultra HD và full HD ở các khía cạnh tốc độ mã hóa, tốc độ bit và chất lượng video. Do đó, việc giảm độ phức tạp tính toán và cải thiện thời gian tính toán cho thuật toán ước lượng chuyển động cho chuẩn HEVC là một nhiệm vụ rất được quan tâm. Trong bài báo này, chúng tôi đề xuất cải tiến kỹ thuật tìm kiếm TZ (Test Zone Search Algorithm) bằng cách sử dụng kỹ thuật học máy nhằm làm giảm thời gian tìm kiếm vector chuyển động. Các kết quả mô phỏng cho thấy phương pháp đề xuất giảm 7% thời gian tìm kiếm so với thuật toán tìm kiếm TZ chuẩn sử dụng trong chuẩn HEVC.

Từ khóa: HEVC, ước lượng chuyển động.

I. GIỚI THIỆU

Thế giới đã được chứng kiến sự phát triển vượt bậc của mã hóa video kể từ cuối những năm 1980. Các ứng dụng video như các cuộc gọi video, video hội nghị qua các mạng không dây và có dây, video phát trực tuyến, truyền hình video số độ nét cao... đã trở thành một phần không thể thiếu trong cuộc sống thường ngày. Tuy nhiên, việc truyền dẫn video không nén yêu cầu tốc độ dữ liệu khổng lồ. Ví dụ, một file video độ nét cao (1920 x 1080) 30 khung hình/s không nén sẽ yêu cầu tốc độ dữ liệu 1,5 Gb/s và bộ nhớ 1TB để lưu trữ đoạn video dài 90 phút. Điều này là phi thực tế khi sử dụng cơ sở hạ tầng đường truyền và bộ lưu trữ dữ liệu hiện thời. Vì vậy, việc nén video là rất quan trọng để truyền và lưu trữ lượng dữ liệu lớn như thế này.

Các kỹ thuật nén video khai thác các loại dư thừa thống kê có trong dữ liệu video. Cụ thể, hệ thống thị giác người không nhạy với độ méo nhưng nhạy với độ sáng của hình ảnh video. Dựa trên những đặc điểm này, các chuẩn mã hóa video thực hiện nén dữ liệu video bằng cách loại bỏ các dư thừa thời gian và dư thừa không gian có trong tín hiệu video. Rất nhiều chuẩn mã hóa video đã được đề xuất trong những năm qua như MPEG-1, H.262/MPEG-2, H.263,

H.264/AVC [1] và HEVC [2] trong đó chuẩn HEVC ra đời năm 2013 có thể tiết kiệm được 50% tốc độ bit so với H.264/AVC. Chuẩn HEVC hỗ trợ các độ phân giải lên tới 8192 x 4320 bao gồm cả 8K UHD.

Trong các chuẩn mã hóa video, dự đoán liên khung được coi là bước quan trọng của quá trình mã hóa và quyết định phần lớn đến hiệu suất nén của các chuẩn mã hóa video. Nó được sử dụng để loại bỏ dư thừa thời gian giữa các khung hình lân cận của một chuỗi video để cho phép tỷ lệ nén cao hơn. Trong dự đoán liên khung, mỗi khung hình được chia thành các khối. Sau đó, thay vì mã hóa trực tiếp mỗi khối, bộ mã hóa sẽ cố gắng tìm trong khung hình đã mã hóa trước đó một khối tương tự với khối đang cần mã hóa hiện thời. Quá trình này được thực hiện thông qua giải thuật đối sánh khối. Nếu giải thuật thành công, khối có thể được mã hóa bằng một vec-tơ gọi là vec-tơ chuyển động (Motion Vector - MV) với một sai số dự đoán nào đó. MV cho biết vị trí của khối phù hợp trong khung hình tham chiếu. Quá trình quyết định MV được gọi là quá trình ước lượng chuyển động.

Rất nhiều giải thuật ME đã được đề xuất từ những năm 1980 [3]. Giải thuật ME tốt nhất là giải thuật tìm kiếm đầy đủ (Full Search Algorithm - FSA) vì tất cả các vị trí trong cửa sổ tìm kiếm đều được kiểm tra và tính toán hàm chi phí tương ứng. Do đó giải thuật này luôn tìm ra khối phù hợp nhất trong khung hình tham chiếu. Tuy nhiên, giải thuật này cần nhiều tính toán nhất trong số các giải thuật ME. Để tăng tốc độ tìm kiếm, rất nhiều giải thuật ME nhanh đã được phát triển nhằm giảm thời gian mã hóa. Có thể kể đến các giải thuật tìm kiếm ba bước (Three Step Search - TSS) [4], giải thuật tìm kiếm ba bước mới (New TSS) [5,6], giải thuật tìm kiếm đơn giản và hiệu quả (Simple and Efficient TSS - SES) [7], giải thuật tìm kiếm bốn bước (4SS) [8], giải thuật tìm kiếm kim cương (Diamond Search - DS) [9]. Với mong muốn cải tiến hơn nữa tốc độ tìm kiếm để giảm thời gian mã hóa trong HEVC, bài báo này đề xuất một phương pháp cải tiến giải thuật tìm kiếm TZ bằng cách sử dụng kỹ thuật học máy.

Phần tiếp theo của bài báo được cấu trúc như sau. Phần II giới thiệu tóm tắt nguyên lý cơ bản của ước lượng chuyển động và giải thuật ước lượng chuyển động nhanh TZ. Phần III mô tả phương pháp đề xuất. Các tham số mô phỏng và kết quả mô phỏng được trình bày trong Phần IV và kết luận được đưa ra trong Phần V.

II. CÁC NGHIÊN CỨU LIÊN QUAN

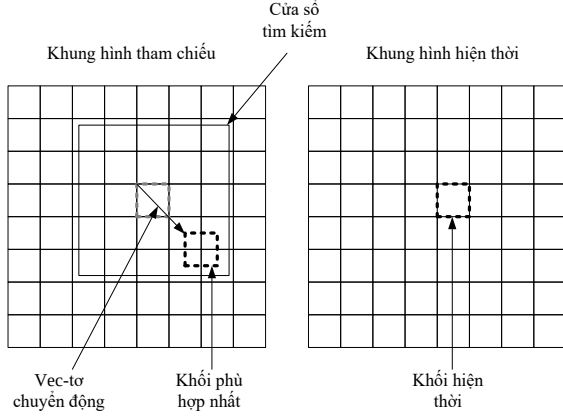
Nguyên lý cơ bản của ME được mô tả trong Hình 1. ME gồm hai bước là ước lượng chuyển động điểm ảnh nguyên (Integer Motion Estimation - IME) và ước lượng chuyển

Tác giả liên hệ: Vũ Hữu Tiến

Email: tienvh@ptit.edu.vn

Đến tòa soạn: 9/2021, chỉnh sửa: 10/2021, chấp nhận đăng: 10/2021

động điểm ảnh nội suy (Fractional Motion Estimation - FME). Trong IME, việc tìm kiếm được thực hiện tại các vị trí điểm ảnh nguyên và tạo ra MV nguyên. Trong FME, việc tìm kiếm được thực hiện tại các vị trí điểm ảnh nội suy sử dụng MV nguyên như là điểm bắt đầu.



Hình 1. Nguyên lý cơ bản của ước lượng chuyển động

Mục đích chính của ME là tìm trong cửa sổ tìm kiếm của khung hình tham chiếu khối phù hợp nhất với khối hiện thời và tạo ra vec-tơ chuyển động tốt nhất. Do đó, cần một tiêu chí đối sánh hay phương pháp tính toán độ méo để đo sự giống nhau giữa khối hiện thời và khối ứng viên. Tuy nhiên, rất khó để định lượng sự tương đồng hay độ méo trong nội dung hình ảnh bởi vì hệ thống thị giác người rất phức tạp và khó hiểu. Trong thực tế, đối với tín hiệu video số, có rất nhiều mô hình đo độ méo [3]. Trong số đó, sai số bình phương trung bình (Mean Square Error - MSE) và tổng các sai số tuyệt đối (Sum of Absolute Differences - SAD) là các mô hình phổ biến nhất. Giả sử $f(i, j, t)$ là cường độ sáng của điểm ảnh tại tọa độ (i, j) trong khung hình t và theo đó $f(i + MV_x, j + MV_y, t - 1)$ là cường độ sáng của điểm ảnh của khối ứng viên trong khung hình tham chiếu $t - 1$ với vec-tơ chuyển động (MV_x, MV_y) . Kích thước của khối cần mã hóa là $N \times N$ điểm ảnh và dải tìm kiếm là $\pm W$ theo mỗi hướng, vì vậy cửa sổ tìm kiếm có kích thước $(2W + N) \times (2W + N)$. Giả sử khối hiện thời tại vị trí (u, v) , thì mô hình đo độ méo MSE được định nghĩa như sau:

$$MSE(MV_x, MV_y) = \frac{1}{N \times N} \sum_{i=u}^{u+N-1} \sum_{j=v}^{v+N-1} [f(i, j, t) - f(i + MV_x, j + MV_y, t - 1)]^2 \quad (1)$$

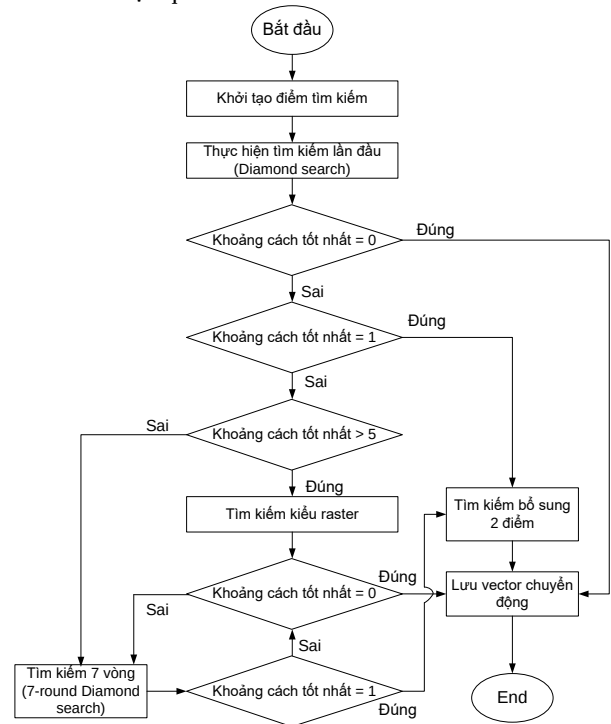
Giá trị MSE này được sử dụng để đánh giá sự giống nhau giữa hai khối. Để tìm ra khối phù hợp nhất, mô hình MSE được áp dụng cho từng khối ứng viên với các vec-tơ chuyển động khác nhau (MV_x, MV_y) bên trong cửa sổ tìm kiếm. Khối ứng viên nào có MSE nhỏ nhất sẽ được coi là khối phù hợp nhất.

Ngoài MSE còn có một mô hình đo độ méo khối tương tự là SAD, tham số này tính toán sự sai khác tuyệt đối giữa hai điểm ảnh của khối hiện thời và khối ứng viên thay vì bình phương sự sai khác trong MSE. SAD tại vị trí (u, v) với vec-tơ chuyển động (MV_x, MV_y) được định nghĩa như sau:

$$SAD(MV_x, MV_y) = \sum_{i=u}^{u+N-1} \sum_{j=v}^{v+N-1} |f(i, j, t) - f(i + MV_x, j + MV_y, t - 1)| \quad (2)$$

Tương tự MSE, khối nào có SAD nhỏ nhất được coi là khối phù hợp nhất.

Giải thuật FSA thực hiện tìm kiếm trên toàn bộ các vị trí trong dải tìm kiếm. Giải thuật này tối ưu theo khía cạnh nếu dải tìm kiếm được xác định đúng, nó sẽ đảm bảo tìm ra vị trí phù hợp nhất. Tuy nhiên, nếu dải tìm kiếm theo mỗi hướng là W với kích thước bước là 1 điểm ảnh thì có tổng cộng $(2W + 1)^2$ lần dịch chuyển để tìm ra vec-tơ chuyển động cho mỗi khối và yêu cầu một số lượng lớn các phép tính. Các yêu cầu tính toán lớn của FSA làm cho nó không khả thi cho các ứng dụng video thời gian thực. Các ứng dụng này thường sử dụng các chiến lược tìm kiếm nhanh và hiệu quả.



Hình 2. Sơ đồ tìm kiếm TZ

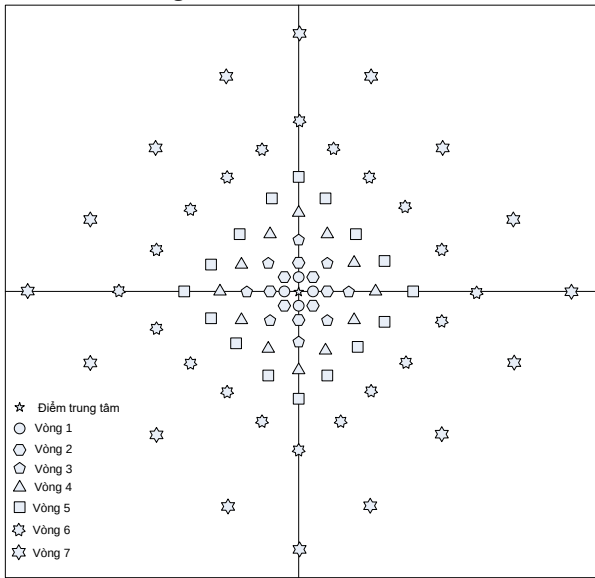
Nhiều giải thuật tìm kiếm nhanh đã được đề xuất để giảm độ phức tạp tính toán của FSA trong khi vẫn duy trì chất lượng dự đoán tương đương. Bằng cách chỉ tìm kiếm một số các ứng viên trong cửa sổ tìm kiếm nên các giải thuật nhanh sẽ giảm độ phức tạp và số lượng các phép toán. Tuy nhiên nó thường làm giảm chất lượng video và làm tăng tốc độ bit. Cùng với sự phát triển của các video độ nét cao, các giải thuật tìm kiếm đầy đủ rất ít khi được sử dụng trong các ứng dụng mã hóa vì độ phức tạp cực kỳ cao. Do đó các giải thuật tìm kiếm nhanh ngày càng được sử dụng rộng rãi. Một số giải thuật tìm kiếm nhanh hiện nay có thể tiệm cận chất lượng video như giải thuật tìm kiếm đầy đủ với tốc độ bit cao hơn một chút trong khi vẫn duy trì được độ phức tạp thấp. Và giải thuật tìm kiếm vùng kiểm tra (TZ) [10] là một trong số các giải pháp như thế.

Giải thuật tìm kiếm TZ là một giải thuật tìm kiếm nhanh mới được sử dụng rộng rãi và được chấp nhận bởi phần

mềm tham chiếu của chuẩn mã hóa video HEVC vì hiệu suất tốt [3]. Đây là thuật toán tìm kiếm lai ghép với sự kết hợp của tìm kiếm kim cương/hình vuông với tìm kiếm raster. Quá trình tìm kiếm TZ được mô tả trong Hình 2.

Tìm kiếm lần đầu

Trong phần mềm tham chiếu HEVC, đối với tìm kiếm lần đầu, có hai tùy chọn trong thuật toán tìm kiếm là tìm kiếm kim cương và tìm kiếm hình vuông, tuy nhiên tìm kiếm kim cương được đặt làm mặc định. Hình 3 minh họa mẫu tìm kiếm và chiến lược tìm kiếm kim cương 7 vòng, ở đó, ngôi sao ở giữa tượng trưng cho điểm bắt đầu và gọi là tâm tìm kiếm và các điểm khác đánh dấu các vị trí ứng viên. Một đặc điểm của thuật toán tìm kiếm kim cương là số vòng bị giới hạn bởi phạm vi tìm kiếm. Với dải tìm kiếm mặc định được sử dụng trong mô hình tham chiếu HEVC là ± 64 thì số vòng là 7.



Hình 3. Tìm kiếm kim cương 7 vòng trong tìm kiếm TZ

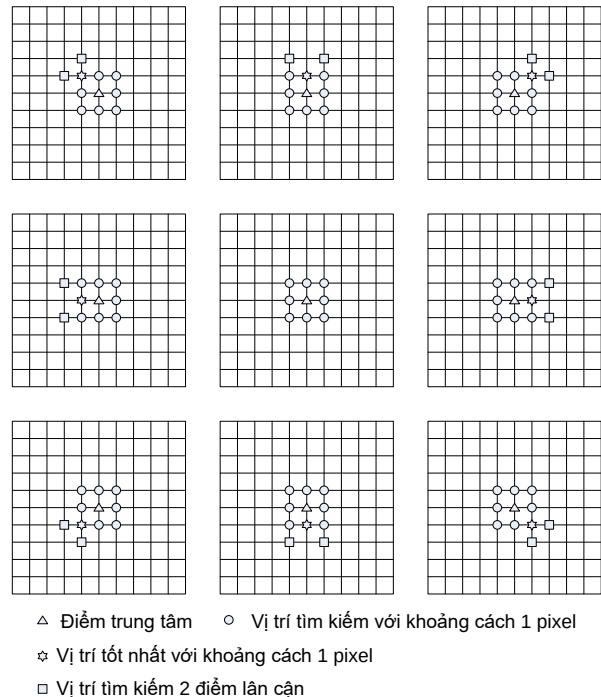
Như trong Hình 3, vòng đầu tiên của tìm kiếm kim cương có 4 mẫu xung quanh điểm bắt đầu với khoảng cách bằng 1, các vị trí này là trên, dưới, trái, phải. Từ vòng 2 đến vòng 4, số lượng các mẫu tìm kiếm cho mỗi vòng tăng lên 8 với 4 vị trí ở giữa và khoảng cách tương ứng của mỗi vòng là 2, 4 và 8. Tương tự từ vòng thứ 5 đến vòng thứ 7, có 8 vị trí một phần tư được thêm vào cho mỗi vòng và khoảng cách của mỗi vòng lần lượt là 16, 32, và 64. Do đó, số lượng mẫu tìm kiếm trong lần tìm kiếm đầu tiên này là 76 cho toàn bộ vùng tìm kiếm. Khi việc tìm kiếm cho toàn bộ các mẫu này kết thúc, có ba khả năng xảy ra dựa trên kết quả tìm kiếm của tìm kiếm lần đầu:

- 1) **Kết thúc:** Nếu khối phù hợp nhất có khoảng cách bằng 0, điều đó có nghĩa là dự đoán MV là đủ chính xác và không có khối ứng viên nào phù hợp hơn điểm bắt đầu, toàn bộ quá trình tìm kiếm TZ kết thúc ngay với điểm bắt đầu chính là MV cuối cùng tốt nhất.
- 2) **Tìm kiếm 2 điểm lân cận:** Nếu khoảng cách của khối phù hợp nhất bằng 1, bước tiếp theo của Tìm kiếm lần đầu là thực hiện tìm kiếm 2 điểm lân cận, bước này đóng vai trò là bước lọc tinh chỉnh cuối cùng trên các vị trí lân cận trước khi hoàn tất tìm kiếm TZ.

- 3) **Tìm kiếm raster:** Trong trường hợp khối phù hợp nhất cách điểm bắt đầu nhiều hơn 5 điểm ảnh, việc tìm kiếm raster trong toàn bộ cửa sổ tìm kiếm sẽ được thực hiện. Bởi vì khoảng cách tốt nhất lớn nhất 5 điểm ảnh cho thấy độ chính xác của dự đoán MV chưa tốt, việc tìm kiếm diện rộng như một biện pháp cần thiết để tìm ra khối phù hợp nhất.
- 4) **Tìm kiếm tinh chỉnh:** Khác với tất cả các trường hợp được đề cập trước đó, nếu khoảng cách nằm trong khoảng từ 2 đến 5 điểm ảnh, thay vì tìm kiếm tinh chỉnh trên các lân cận của nó, việc tìm kiếm tinh chỉnh trên toàn bộ dải tìm kiếm với tâm là vị trí của khối phù hợp nhất từ bước Tìm kiếm lần đầu sẽ được thực hiện. Đặc biệt, có một tiêu chí dừng sớm đối với tìm kiếm lần đầu. Bởi vì thứ tự tìm kiếm là từ vòng 1 đến vòng 7 với một khoảng cách tương ứng, nếu trong vòng thứ $(r \in [1,3])$, khối phù hợp tốt nhất tạm thời tìm được là MB_x và trong 3 vòng tiếp theo, từ vòng thứ $(r + 1)$ đến $(r + 3)$, không tồn tại ứng viên nào tốt hơn thì Tìm kiếm tinh chỉnh có thể kết thúc với MB_x là khối phù hợp nhất cho bước tiếp theo.

Tìm kiếm 2 điểm lân cận

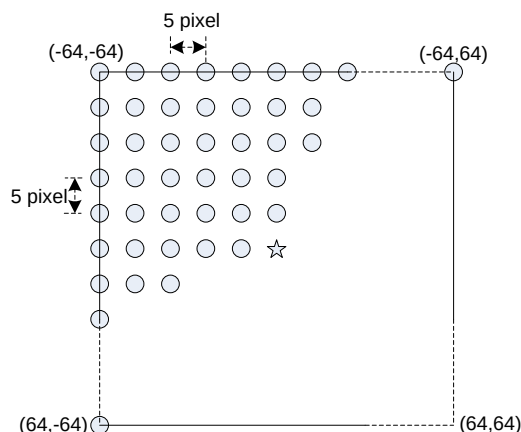
Bước tìm kiếm 2 điểm lân cận chỉ được yêu cầu khi khoảng cách của khối phù hợp nhất bằng 1. Khoảng cách bằng 1 cho thấy có 8 vị trí có thể đủ điều kiện như được mô tả trong cửa sổ trung tâm của Hình 4 trong đó 4 vị trí của vòng đầu tiên được đánh dấu bằng màu tím và 4 vị trí còn lại từ vòng thứ 2 (được đánh dấu bởi chấm màu xanh lam). Các mẫu tìm kiếm và chiến lược tìm kiếm 2 điểm lân cận được mô tả cho 8 trường hợp riêng biệt trong đó chấm đỏ cho biết nó là giá trị phù hợp nhất trong trường hợp đó và hai ngôi sao màu đen chỉ ra hai vị trí còn thiếu. Sau khi hoàn thành tìm kiếm 2 điểm còn thiếu kết thúc, vị trí dịch chuyển của khối phù hợp nhất cuối cùng sẽ được lưu lại và khi đó tìm kiếm TZ kết thúc.



Hình 4. Tìm kiếm 2 điểm lân cận

Tìm kiếm raster

Khi khoảng cách giữa điểm bắt đầu và khối phù hợp nhất trong bước Tìm kiếm lần đầu lớn hơn 5 điểm ảnh, tìm kiếm TZ sẽ bước vào giai đoạn tìm kiếm raster. Chiến lược của tìm kiếm raster rất đơn giản ở chỗ nó thực hiện tìm kiếm đầy đủ trong vùng cửa sổ tìm kiếm với khoảng cách tìm kiếm σ điểm ảnh theo cả hai hướng ngang và dọc, ở đó $\sigma = 5$ đối với phần mềm tham chiếu HEVC. Các mẫu tìm kiếm của tìm kiếm raster được mô tả trong Hình 5.



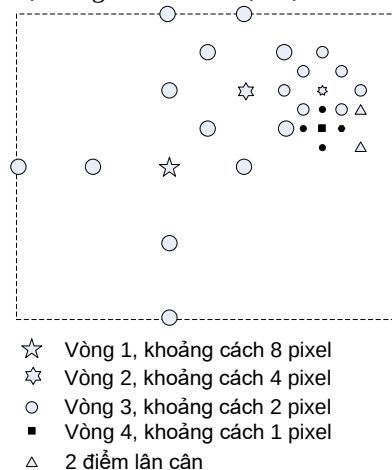
Hình 5. Các mẫu tìm kiếm của tìm kiếm raster với khoảng cách 5 điểm ảnh

Sau khi hoàn thành việc quét qua tất cả các mẫu tìm kiếm, nếu khối tối ưu được tạo ra bởi tìm kiếm raster trùng hợp với khối phù hợp nhất từ bước tìm kiếm lần đầu, vị trí dịch chuyển của nó được coi là MV tốt nhất và tìm kiếm TZ sẽ dừng ngay lập tức; nếu không thì tìm kiếm TZ sẽ được tiếp tục với việc tìm kiếm tinh chỉnh toàn bộ dải tìm kiếm.

Tìm kiếm tinh chỉnh

Tìm kiếm tinh chỉnh được thực hiện với hai trường hợp: (1) khoảng cách tốt nhất nằm giữa 2 và 5 điểm ảnh sau bước Tìm kiếm lần đầu; (2) Tìm kiếm raster tìm ra khối phù hợp tối ưu khác với kết quả của Tìm kiếm lần đầu. Trong tìm kiếm TZ thực hiện trong phần mềm tham chiếu HEVC, có hai tùy chọn đối với tìm kiếm tinh chỉnh: Tinh chỉnh hình sao và tinh chỉnh raster. Quá trình tinh chỉnh hình ngôi sao thực hiện tìm kiếm hình kim cương 7 vòng (được mô tả trong Hình 3) với dải tìm kiếm đầy đủ lấy khối phù hợp nhất từ giai đoạn trước làm tâm. Bởi vì Tìm kiếm hình sao chỉ dừng khi khoảng cách tốt nhất cuối cùng bằng 0 hoặc 1, nếu không nó sẽ tiếp tục thực hiện Tìm kiếm kim cương 7 vòng với tâm là vị trí tối ưu mới được tìm thấy ở tìm kiếm tinh chỉnh trước đó. Khi khoảng cách tốt nhất bằng 1, tìm kiếm hai điểm lân cận được thực hiện bổ sung trước khi dừng tìm kiếm TZ. Đối với Tinh chỉnh raster, khác với Tinh chỉnh hình sao, thay vì áp dụng Tìm kiếm kim cương 7 vòng, nó sử dụng Tìm kiếm kim cương 1 vòng với khoảng cách bằng một nửa khoảng cách tốt nhất của giai đoạn trước đó. Nó sẽ lặp lại quá trình này và sẽ không dừng cho đến khi hoặc là khoảng cách tốt nhất của vòng tìm kiếm nào đó bằng 0 hoặc khoảng cách tốt nhất giảm từ 1 hoặc 0 sau một vài vòng tìm kiếm. Một ví dụ về Tinh chỉnh raster được mô tả trong Hình 6. So với tinh chỉnh raster, tinh chỉnh hình sao sẽ tìm kiếm nhiều vị trí hơn bởi vì hệ thống tìm kiếm kim cương 7 vòng chắc chắn sẽ tạo ra MV chính xác hơn và vì vậy đạt được chất lượng video cao hơn.

Trong mô hình tham chiếu HM, tinh chỉnh hình sao được thiết lập là hệ thống tinh chỉnh mặc định.



Hình 6. Ví dụ về tìm kiếm tinh chỉnh raster

III. PHƯƠNG PHÁP ĐỀ XUẤT

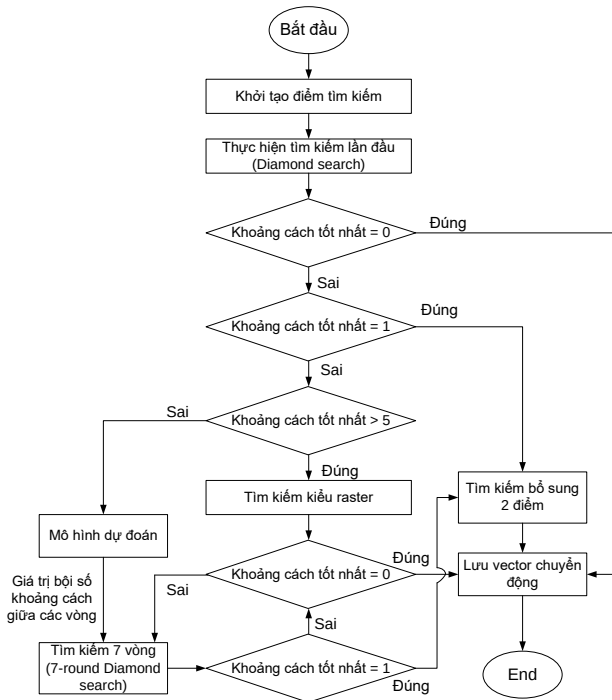
Như đã đề cập ở phần trên, trong trường hợp khoảng cách lớn hơn 5, bộ mã hóa sẽ thực hiện quá trình Tinh chỉnh hình sao. Quá trình này thực chất là thực hiện lại một lần nữa 7 vòng tính toán như hình trên nhưng điểm bắt đầu được thay bằng điểm vừa tìm được trong lần trước. Quá trình này chỉ dừng lại khi bộ mã hóa tìm được khối có khoảng cách bằng 0 hoặc 1. Khi khoảng cách bằng 0 thì dừng lại. Khi khoảng cách bằng 1 thì lại thực hiện tìm ở 2 vị trí điểm còn thiếu và dừng lại. Nếu khoảng cách lớn hơn 1 thì thực hiện tìm kiếm tại các điểm của vòng tiếp theo. Khoảng cách giữa các vòng sẽ tăng dần gấp đôi so với vòng trước. Tuy nhiên, theo quan sát qua quá trình thực nghiệm, lũy thừa của 2 có thể chỉ phù hợp với các video có chuyển động chậm. Trong trường hợp video chuyển động nhanh, bộ mã hóa cần nhiều vòng để tìm kiếm. Vì vậy, để tăng tốc độ tìm kiếm, bội số có thể tăng lên để phù hợp với nội dung video. Để thích ứng với nội dung video, mô hình mạng nơron được sử dụng để dự đoán bội số của khoảng cách giữa các vòng tìm kiếm khi khoảng cách tốt nhất khác 0 và 1 nhưng nhỏ hơn 5. Quá trình dự đoán khoảng cách bao gồm:

Bước 1: Xây dựng mô hình dự đoán khoảng cách

Trong phương pháp đề xuất, một mô hình mạng nơron được sử dụng để dự đoán bội số của khoảng cách giữa các vòng tìm kiếm thích ứng với nội dung của video. Cụ thể, kiến trúc mạng nơron được đề xuất bao gồm 3 lớp trong đó lớp đầu vào có 3 nơron nhận các giá trị đặc trưng của khung hình bao gồm giá trị SAD (giá trị khác biệt giữa khung hình hiện tại và khung hình trước đó), giá trị trung bình của hệ số DC và giá trị phương sai của hệ số DC của các khối trong một khung hình, lớp ẩn có 2 nơron và lớp đầu ra có 1 nơron chứa giá trị dự đoán của bội số. Hàm tích cực được sử dụng trong mạng là hàm sigmoid.

Dữ liệu được sử dụng để huấn luyện là tham số của các khung hình trong các chuỗi video "Harbor", "Husky", "Mobile", "News" và "Pamphlet" với độ phân giải không gian 176×144, số lượng các khung hình của mỗi chuỗi là 300. Các chuỗi được chọn có sự đa dạng về kết cấu ảnh và đặc tính chuyển động của chuỗi. Các chuỗi này được mã hóa với bốn tham số lượng tử từ lần lượt là 28, 32, 36 và 40. Kết quả của quá trình này là mô hình dự đoán khoảng cách giữa các vòng tìm kiếm.

Bước 2: Áp dụng mô hình dự đoán khoảng cách giữa các vòng tìm kiếm ở bước 1 vào cho các dữ liệu video mới.



Hình 7. Thuật toán tìm kiếm TZ cải tiến

Hình 7 mô tả các bước của thuật toán tìm kiếm TZ cải tiến. Sau khi thực hiện tìm kiếm theo lần đầu, nếu khoảng cách của vector chuyển động khác 0 và 1 nhưng nhỏ hơn 5, bộ mã hóa sẽ thực hiện tính giá trị SAD, giá trị trung bình và giá trị phương sai của các hệ số DC của các khối 8×8 của khung hình hiện thời. Các giá trị này sẽ là các giá trị đầu vào của mô hình dự đoán. Bộ dự đoán sẽ dự đoán giá trị bội số của khoảng cách giữa các vòng của quá trình tìm kiếm 7 vòng.

IV. ĐIỀU KIỆN THỬ NGHIỆM VÀ PHÂN TÍCH KẾT QUẢ

A. Điều kiện thử nghiệm

Để đánh giá hiệu năng của phương pháp tìm kiếm TZ cải tiến được đề xuất so với phương pháp tìm kiếm TZ của HEVC, bốn chuỗi video sau được sử dụng là “Akiyo”, “Carphone”, “Foreman” và “Coastguard” với độ phân giải 176x144 và các giá trị của tham số lượng tử lần lượt là 28, 32, 36 và 40.

B. Phân tích kết quả

Hiệu năng giữa phương pháp đề xuất với phương pháp TZ chuẩn được so sánh thông qua tham số tốc độ bit – độ biến dạng tín hiệu (RD) và giá trị Bjontegaard Rate [11]. Bảng I mô tả giá trị BD-Rate và BD-PSNR và Bảng II mô tả thời gian mã hóa của hai phương pháp. Hình 8 so sánh đường cong RD của hai phương pháp này.

Các kết quả trong Bảng I và Hình 8 cho thấy mức độ tiết kiệm tốc độ bit và mức chất lượng tín hiệu của hai phương pháp TZ chuẩn và TZ cải tiến là tương đương nhau. Phương pháp TZ cải tiến có lợi thế với hai chuỗi Foreman và Coastguard là những chuỗi có chuyển động nhanh và phức tạp trong khi phương pháp TZ chuẩn đạt kết quả tốt hơn với các chuỗi chuyển động chậm và ít phức tạp. Tuy nhiên, thời gian mã hóa của phương pháp TZ cải tiến giảm

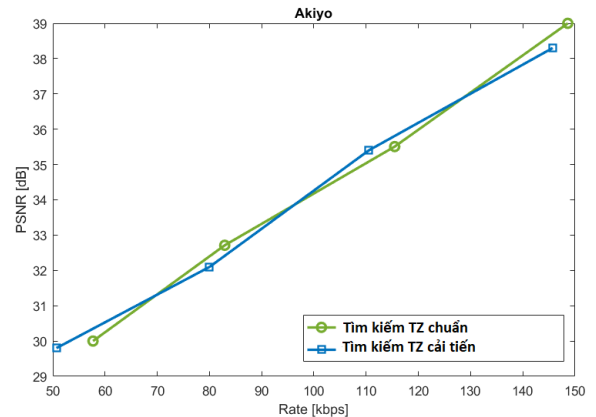
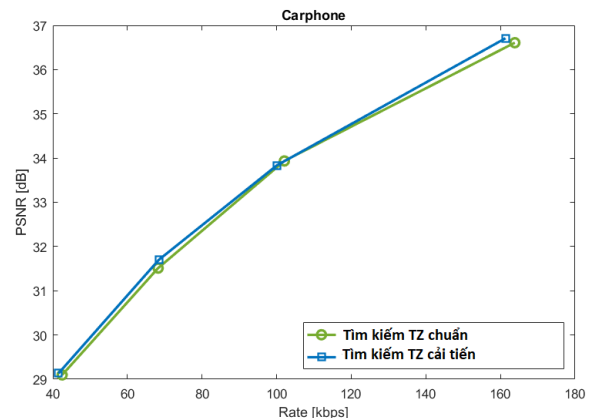
đi so với phương pháp TZ chuẩn đối với cả 4 chuỗi video thử nghiệm (Bảng II). Phương pháp TZ cải tiến trung bình giảm được 7% thời gian mã hóa so với phương pháp TZ chuẩn. Đó là nhờ việc lựa chọn khoảng cách trong tìm kiếm hình sao tùy thuộc vào nội dung chuỗi video để tối ưu thời gian mã hóa.

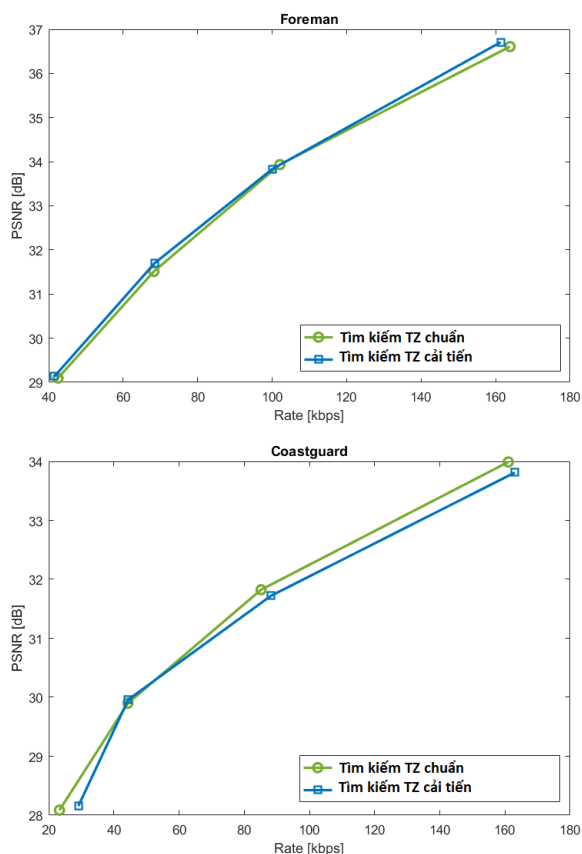
Bảng I. So sánh BD-Rate [%] và BD-PSNR [dB] của phương pháp tìm kiếm TZ cải tiến và phương pháp tìm kiếm TZ chuẩn

Chuỗi Video	BD-Rate	BD-PSNR
Carphone	2,00	- 0,11
Foreman	-1,58	0,06
Akiyo	0,57	0,08
Coastguard	-5,40	0,16
Trung bình	-1,10	0,05

Bảng II. So sánh thời gian mã hóa trung bình (s) cho 1 khung hình của bộ mã HEVC khi sử dụng phương pháp tìm kiếm TZ cải tiến và phương pháp tìm kiếm TZ chuẩn

Chuỗi Video	Tìm kiếm TZ chuẩn	Tìm kiếm TZ cải tiến
Carphone	9,08	8,55
Foreman	7,34	6,97
Akiyo	5,10	4,90
Coastguard	3,79	3,12
Trung bình	6,33	5,88





Hình 8. Đường cong RD của 4 chuỗi video Carphone, Akiyo, Foreman và Coastguard.

V. KẾT LUẬN

Bài báo đã đề xuất một phương pháp giảm thời gian mã hóa cho chuẩn HEVC bằng cách cải tiến thuật toán tìm kiếm TZ sử dụng kỹ thuật học máy. Khi khoảng cách tốt nhất khác 0 và 1 nhưng nhỏ hơn 5, việc dự đoán khoảng cách giữa các vòng tìm kiếm trong giai đoạn tinh chỉnh hình sao được thực hiện dựa trên mô hình mạng nơ-ron. Việc sử dụng mạng nơ-ron giúp tìm ra khoảng cách tốt nhất trong khoảng thời gian ngắn hơn để giúp giảm thời gian mã hóa. Các kết quả mô phỏng đã cho thấy phương pháp TZ cải tiến đã giảm thời gian mã hóa so với phương pháp TZ chuẩn là 7% trong khi vẫn giữ được mức chất lượng tương đương so với phương pháp TZ chuẩn.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Học Viện Công nghệ Bưu chính Viễn thông trong đề tài mã số 05-2021-HV-KTĐT1.

TÀI LIỆU THAM KHẢO

- [1] T. Wiegand, G. J. Sullivan, S. Member, G. Bjøntegaard, A. Luthra, and S. Member, "Overview of the H.264/AVC Video Coding Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, 2003.
- [2] H. Standard, G. J. Sullivan, J. Ohm, W. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [3] S. T. S. Metkar, *Motion Estimation Techniques for Digital Video Coding*. Springer briefs in applied

sciences and technology, 2013.

- [4] T. Koga, T. Ishiguro, Motion compensated inter-frame coding for video conferencing, *Proceedings of National Telecommunication Conference*, New Orleans, pp. G5.3.1–G5.3.5, Dec 1981.
- [5] Renxiang Li, Bing Zeng, and Ming L. Liou, A New 3-Step Search Algorithm for Block Motion Estimation, *IEEE Trans. Circuits And Systems For Video Technology*, vol 4., no. 4, pp. 438-442, August 1994.
- [6] SUN Ning-ning, FAN Chao, XIA Xu, "An Effective Three-step Search Algorithm for Motion Estimation", 2009 [7] A. Aaron, S. D. Rane, E. Setton, and B. Girod, "Transform-domain Wyner-Ziv codec for video," *Vis. Commun. Image Process.* 2004, vol. 5308, p. 520, 2004, doi: 10.1117/12.527204.
- [7] Jianhua Lu, and Ming L. Liou, "A Simple and Efficient Search Algorithm for Block-Matching Motion Estimation", *IEEE Trans. Circuits And Systems For Video Technology*, vol 7, no. 2, pp.429-433, April 1997
- [8] Lai-Man Po, and Wing-Chung Ma, "A Novel Four-Step Search Algorithm for Fast Block Motion Estimation", *IEEE Trans. Circuits And Systems For Video Technology*, vol 6, no. 3, pp. 313-317, June 1996
- [9] J.Y. Tham, S. Ranganath, M. Ranganath, A.A. Kassim, A novel unrestricted center-biased diamond search algorithm for block motion estimation. *IEEE Trans. Circuits Syst. Video Technol.* 8(4), 369–377 (1998)
- [10] X. Li, R. Wang, W. Wang, Z. Wang, S. Dong, "Fast Motion Estimation Methods for HEVC", *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, 2014.
- [11] G. Bjøntegaard, "Calculation of Average PSNR Differences between RD Curves", Document VCEG-M33, 13th ITU-T VCEG Meeting, Austin, TX, USA, 2001.

REDUCE ENCODING TIME FOR HEVC VIDEO CODING BY IMPROVING THE TZ SEARCH ALGORITHM

Abstract: The demand for high-resolution video, and especially the use of high-quality video on mobile devices, has led to an increasing demand for high-performance video encoders. To meet that requirement, new video coding standards are constantly being developed. The HEVC high-performance video coding standard achieves a 50% increase in compression compared to its predecessor H.264 with comparable video quality. In these predictive video coding standards, motion estimation (ME) is one of the most important modules that determines video encoder compression efficiency. However it also accounts for 50%-70% of the coding complexity of the entire encoder. The large computational resource consumption reduces the performance of the encoder, especially for ultra HD and full HD video in terms of encoder rate, bit rate and video

quality. Therefore, reducing the computational complexity and improving the computation time for the motion estimation algorithm for the HEVC standard is a task of great interest. In this paper, we improved TZ search technique by using machine learning techniques to reduce motion vector search time. The simulation results show that the proposed method reduces the encoding time by 7% compared to the HEVC standard TZ search algorithm.

Keyword: HEVC, Motion Estimation.



Nguyễn Thị Hương Thảo, Nhận bằng tốt nghiệp đại học và thạc sỹ Học viện Công nghệ Bưu chính Viễn thông vào các năm 2003 và 2010. Năm 2021 nhận bằng Tiến sĩ tại Học viện Công nghệ Bưu chính Viễn thông ngành Kỹ thuật điện tử. Hiện đang giảng dạy tại Khoa Kỹ thuật Điện tử 1 - Học viện Công nghệ Bưu chính Viễn thông. Lĩnh vực nghiên cứu: Xử lý tín hiệu Video, Xử lý Ảnh, Lý thuyết thông tin.



Vũ Hữu Tiến, Tốt nghiệp đại học và cao học ngành Điện tử Viễn thông vào các năm 2002, 2004 tại Đại học Bách khoa Hà nội. Năm 2010 nhận bằng Tiến sĩ tại Đại học Chulalongkorn (Thái Lan) ngành Điện tử Viễn thông. Hiện đang công tác tại Khoa Đa phương tiện, Học viện Công nghệ Bưu chính Viễn thông.

Lĩnh vực nghiên cứu hiện nay: Xử lý tín hiệu và truyền thông đa phương tiện, Phát triển ứng dụng đa phương tiện.