

CẢI TIẾN HÀM TÍNH ĐỘ MÉO TRONG QUÁ TRÌNH TỐI ƯU ĐỘ MÉO-TỐC ĐỘ BIT CHO BỘ MÃ HÓA VIDEO H.266/VCC

Vũ Hữu Tiến, Thippaphone Sisouvong

Học viện Công nghệ Bưu chính Viễn thông

Tóm tắt: Mã hóa video linh hoạt (VVC) hay còn được gọi là H.266 là chuẩn mã hóa video mới nhất và là phiên bản kế thừa của mã hóa video hiệu suất cao (HEVC). Cũng như tên gọi của mình, VVC là chuẩn mã hóa hiệu quả với nhiều nội dung video và ứng dụng. Trong VVC có nhiều công cụ mã hóa mới so với HEVC, tuy nhiên chưa thực sự có nhiều nghiên cứu về tối ưu tốc độ - độ méo (RDO) – một kỹ thuật quan trọng trong mã hóa video để xác định chế độ mã hóa cho một đơn vị mã hóa nào đó để đạt được hiệu năng tốc độ - độ méo tối ưu. Bài báo này đề xuất một thuật toán ước lượng độ méo của các khối hình trong quá trình tối ưu RDO dựa trên đại lượng đo khách quan kết hợp nhiều phương pháp đánh giá video (VMAF). So với thuật toán RDO sử dụng đại lượng đo khách quan phổ biến thường được dùng trong các chuẩn mã hóa video là tổng các sai số bình phương (SSE), thuật toán đề xuất đã đạt được mức tiết kiệm tốc độ bit lên tới 8.41% khi thực hiện trên bộ mã hóa H266/VVC.

Từ khóa: VMAF, RDO, H266/VVC.

I. GIỚI THIỆU

Các ứng dụng video độ trung thực cao ngày càng trở nên phổ biến dẫn đến nhu cầu về các kỹ thuật mã hóa video hiệu quả cao hơn nữa. Mã hóa video đa năng (VVC) hay còn được biết đến là H.266/VVC là chuẩn mã hóa video mới nhất được thiết kế với mục đích cung cấp khả năng mã hóa hiệu quả đáp ứng một dải rộng các ứng dụng và nội dung video. Các nghiên cứu [1] [2] cho thấy VVC có sự cải thiện chất lượng từ 32% đến 43% so với HEVC tùy thuộc vào độ phân giải của nội dung video. Để đạt được điều đó, chuẩn mã hóa video này gồm nhiều phương pháp mới bao gồm việc đưa ra nhiều chế độ mã hóa, tối ưu quá trình lựa chọn tham số lượng tử, lọc nhiễu.

Trong các hệ thống mã hóa video thực tế, kiểm soát tốc độ đóng một vai trò quan trọng vì có những hạn chế về băng thông hoặc lưu trữ trong các ứng dụng cụ thể. Mục đích của tối ưu hóa tốc độ bit – độ méo là tối ưu hóa chất lượng video, giảm thiểu sự biến dạng của video tùy thuộc vào các hạn chế về tốc độ bit. Các yêu cầu và phương pháp của RDO không mang tính chuẩn mực trong tất cả các tiêu chuẩn mã hóa video nhưng đây là yếu tố quan trọng ảnh hưởng đến hiệu suất của bộ mã hóa. Nhiều tính năng nâng cao (ví dụ lựa chọn kích thước khối) trong các chuẩn mã

hóa video sẽ không mang lại hiệu quả nếu không có quá trình RDO. RDO phải chọn các chế độ và bộ tham số tốt nhất cho mỗi khối hình video. Đối với chuỗi video được chia thành các đơn vị mã hóa, bài toán RDO có thể được biểu diễn như sau:

$$\min \sum_{i=1}^N D_i \text{ với điều kiện } \sum_{i=1}^N R_i \leq R_T \quad (1)$$

trong đó D_i và R_i là độ méo và tốc độ bit của đơn vị mã hóa thứ i . N là tổng số các đơn vị mã hóa và R_T là tổng tốc độ bit hiện có. Bài toán (1) được quy về bài toán tìm cách phân bổ tốc độ bit tối ưu cho mỗi đơn vị mã hóa với ràng buộc về tốc độ bit sao cho độ méo của chuỗi video nên đạt giá trị nhỏ nhất có thể.

Trong các bộ mã hóa của các chuẩn mã hóa video, bài toán tối ưu trong biểu thức (1) được chuyển thành bài toán tối ưu không ràng buộc bằng cách sử dụng nhân tử Lagrange λ và mỗi đơn vị mã hóa được xử lý riêng biệt và độc lập theo thứ tự mã hóa. Nhân tử Lagrange sẽ quyết định trọng số tương đối giữa tốc độ bit và độ méo trong việc lựa chọn chế độ mã hóa và điều này sẽ ảnh hưởng rất nhiều đến chất lượng hình ảnh và tốc độ bit. Rõ ràng λ cần được xác định trước khi tính toán độ méo D_i và tốc độ bit R_i thực tế. Do đó điều quan trọng là phải ước lượng được độ méo D_i sao cho phù hợp mới chức chất lượng hình ảnh. Trong các chuẩn mã hóa video hiện nay, độ méo thường được đo bằng các phương pháp đo khách quan. Để cải thiện hiệu năng mã hóa video, rất nhiều phương pháp đã được đề xuất thay thế đại lượng đánh giá chất lượng khách quan trong quá trình RDO bằng một đại lượng khác gần với chủ quan cảm nhận của mắt người hơn.

Mặc dù phương pháp đánh giá khách quan dễ triển khai nhưng cách tin cậy nhất để đánh giá chất lượng hình ảnh là đánh giá chủ quan bởi vì người xem là người cảm nhận cuối cùng trong hầu hết các ứng dụng. Phương pháp điểm số ý kiến trung bình (MOS) [3] là đại lượng đo chất lượng chủ quan được thực hiện bằng cách lấy ý kiến của một số người quan sát. Tuy nhiên, phương pháp MOS này cần nhiều thời gian đo và khó tích hợp vào trong các ứng dụng. Do vậy, mục tiêu của việc nghiên cứu các phương pháp đo chất lượng hình ảnh khách quan là thiết kế các phương pháp đo chất lượng có thể dự đoán được chất lượng hình ảnh nhận được một cách tự động. Phương pháp đo khách quan thường được sử dụng trong các chuẩn mã hóa video là tổng bình phương sai số (SSE) đo đơn giản trong tính toán, có ý nghĩa vật lý rõ ràng và dễ dàng về mặt toán học để giải quyết bài toán tối ưu. Tuy nhiên phương pháp này không phản ánh đúng chất lượng cảm nhận của mắt người.

Đã có rất nhiều nghiên cứu về các phương pháp đo chất lượng hình ảnh khách quan kết hợp với các đại lượng đo chất lượng cảm nhận bằng cách xem xét đặc tính của hệ

Tác giả liên hệ: Vũ Hữu Tiến

Email: tienvh@ptit.edu.vn

Đến tòa soạn: 9/2024, chỉnh sửa: 10/2024, chấp nhận đăng: 10/2024

thống thị giác người (HVS). Một số đại lượng đo đã được đề xuất gồm chỉ số tương đồng về mặt cấu trúc (SSIM) [4], đại lượng đo chất lượng video (VQM) [5]. SSIM ước tính sự suy giảm của tính tương đồng về cấu trúc dựa trên các thuộc tính thống kê của thông tin cục bộ giữa ảnh tham chiếu và ảnh bị méo. Đây là phiên bản cải tiến của phương pháp chỉ số chất lượng hình ảnh phổ quát trước đó và kết hợp ba đại lượng đo tính tương đồng cục bộ dựa trên độ sáng, độ tương phản và cấu trúc. Trong bài báo [9], đại lượng SSIM được sử dụng để đo độ méo trong quá trình RDO. Các phương pháp trong [10] [8] đề xuất sử dụng SSIM để tính các hệ số trong quá trình biến đổi DCT. VQM của Pinson và Wolf dự đoán một cách chặt chẽ các xếp hạng chất lượng chủ quan. Phương pháp này dựa trên các bộ lọc suy giảm, các bộ lọc này kết hợp các phép đo độ mờ, giật, nhiễu toàn cục, độ méo khối và độ méo màu. VQM tính toán bảy tham số dựa trên các đặc trưng chất lượng khác nhau. Những tham số này thu được bằng cách lọc các video bị suy giảm và video tham chiếu để trích xuất các thuộc tính quan tâm. Các tính năng không gian-thời gian sau đó được trích xuất và bằng cách so sánh các số liệu thống kê của vùng video gốc và video bị suy giảm đã được lọc, ta sẽ thu được tham số chất lượng cho tính năng này.

Hoàn toàn khác biệt với các phương pháp này là một đề xuất bởi các nhà nghiên cứu Netflix có tên gọi Kết hợp nhiều phương pháp đánh giá video (VMAF) [12]. Đây là phương pháp dự đoán chất lượng chủ quan dựa trên học máy, kết hợp một số đại lượng đo bao gồm: đại lượng đo mất chi tiết (DLM), VIF (ở bốn thang đo khác nhau) và sự sai khác khung hình trung bình theo thời gian bằng cách sử dụng phương pháp học máy vector hỗ trợ (SVM). Một số nghiên cứu [13] - [15] đã cho thấy VMAF đạt kết quả tốt hơn so với các phương pháp đo khách quan trước đó và nó có mối tương quan lớn với MOS. Trong một số nghiên cứu gần đây [16] - [18], các tác giả đề xuất các phương pháp sử dụng VMAF vào trong quá trình tối ưu RDO bằng phương pháp học máy. Tuy nhiên, quá trình sử dụng mô hình học máy để ước lượng giá trị lượng tử có thể làm tăng thời gian mã hóa của bộ mã hóa. Trong bài báo [19], mô hình mô tả mối quan hệ giữa VMAF và SSE được sử dụng để mã hóa các khối hình video. Tuy nhiên, phương pháp này đòi hỏi phải tính VMAF ngoại tuyến trước khi mã hóa. Vì vậy, phương pháp này đòi hỏi mất nhiều thời gian hơn so với phương pháp thông thường. Sở dĩ việc phải xây dựng mô hình mô tả mối quan hệ giữa VMAF với một đại lượng khách quan khác là do không thể đo chất lượng bằng VMAF trực tiếp trong quá trình RDO. Vì vậy, các phương pháp này đều phải đo giá trị VMAF thông qua một đại lượng khách quan khác để có thể áp dụng công thức tính toán trong quá trình RDO.

Trong bài báo này, mô hình mô tả mối tương quan giữa VMAF và PSNR (Peak Signal Noise Rate) được đề xuất nhằm đánh giá chất lượng video được chính xác hơn, từ đó làm tăng hiệu năng cho quá trình RDO. Cụ thể, phương pháp đề xuất sử dụng đại lượng VMAF thay cho PSNR là một đại lượng khách quan thường được dùng trong mã hóa video truyền thống để tìm mối tương quan giữa Rate-QP và Distortion-QP. Từ đó, hàm chi phí trong quá trình RDO được tính và lựa chọn giá trị QP tối ưu sao cho chi phí mã hóa là nhỏ nhất.

Phần tiếp theo của bài báo được cấu trúc như sau. Phần II giới thiệu về kỹ thuật RDO trong mã hóa video và tham số đánh giá chất lượng video VMAF. 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

A. RDO

Bắt đầu từ tiêu chuẩn mã hóa video H.264, thuật toán RDO được sử dụng đã đạt được những ưu điểm vượt trội so với các chuẩn mã hóa video trước đó về hiệu năng mã hóa [15]. Thuật toán RDO giúp bộ mã hóa lựa chọn chế độ mã hóa tối ưu nhất trong số rất nhiều các chế độ mã hóa. Cụ thể, quá trình RDO giúp tối thiểu hóa độ méo hình ảnh (D) ứng với một giá trị tốc độ bit cho trước R_c bằng cách lựa chọn các tham số mã hóa phù hợp. Bài toán trên được đưa về bài toán tìm cực trị của D với điều kiện ràng buộc là tốc độ bit $R \leq R_c$.

Để giải bài toán trên, phương pháp nhân tử Lagrange được sử dụng. Ý tưởng cơ bản của phương pháp này là tìm nhân tử Lagrange và các giá trị D, R sao hàm chi phí đạt giá trị nhỏ nhất:

$$J = D + \lambda \cdot R \quad (1)$$

Trong đó J là hàm chi phí Lagrange và λ là nhân tử Lagrange. Khi mối quan hệ giữa R-D là hàm lồi, R và D khả vi tại mọi điểm thì J đạt cực tiểu khi:

$$\frac{dJ}{dR} = \frac{dD}{dR} + \lambda = 0 \quad (2)$$

Trong các chuẩn mã hóa video, mối quan hệ giữa R và D được biểu diễn bằng biểu thức:

$$R(D) = a \log_2 \left(\frac{b}{D} \right) \quad (3)$$

Trong đó a và b là các hệ số không đổi. Mô hình của độ méo D được biểu diễn bằng biểu thức:

$$D = \frac{QP^2}{3} \quad (4)$$

Trong đó QP là tham số lượng tử. Thay (3), (4) vào (2) ta có:

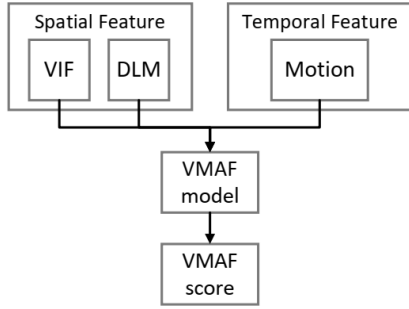
$$\lambda = -\frac{dD}{dR} = c \cdot QP^2 \quad (5)$$

Với c là hằng số và có giá trị bằng 0.57 trong tiêu chuẩn mã hóa video H.266/VVC. Để tìm giá trị nhỏ nhất của hàm chi phí, bộ mã hóa sẽ mã hóa nhiều chế độ khác nhau bao gồm chế độ liên ảnh, nội ảnh, các kích thước khối khác nhau và các giá trị QP khác nhau. Với mỗi chế độ mã hóa sẽ có một bộ ba giá trị R, D và λ được tìm và hàm chi phí J được tính theo công thức (2). Chế độ nào đạt được giá trị J nhỏ nhất sẽ được lựa chọn. Trong hầu hết các bộ mã hóa hiện nay, giá trị D được tìm bằng cách tính giá trị khác biệt giữa ảnh gốc và hình ảnh được giải nén. Mặc dù việc đánh giá độ méo dựa vào phương pháp khách quan có ưu điểm là hiệu quả trong tính toán nhưng có nhược điểm là độ chính xác không cao so với phương pháp chủ quan. Vì vậy, trong bài báo này, phương pháp đo độ méo sẽ sử dụng tham số VMAF là tham số kết hợp cả phương pháp khách quan và chủ quan.

B. VMAF

Đánh giá chất lượng là một nhu cầu thiết yếu trong nhiều dịch vụ video, cùng với sự phát triển của một số chỉ số đo chất lượng để tận dụng lợi thế của đánh giá tự động. Để có được một chỉ số chất lượng cảm nhận phù hợp, Netflix đã phát triển một chỉ số chất lượng cảm nhận có tên là Kết hợp nhiều phương pháp đánh giá video (VMAF) để cung cấp điểm số chính xác cho các nội dung khác nhau

như chương trình truyền hình từng đoạt giải thưởng, phim, phim hoạt hình, phim tài liệu.



Hình 1. Minh họa sự kết hợp các chỉ số đo trong VMAF

Ý tưởng về VMAF là sự kết hợp của một số các chỉ số để duy trì các điểm mạnh như được mô tả trong Hình 1. Có nhiều tính năng và chỉ số đã được sử dụng rộng rãi, nhưng ba chỉ số cơ bản là VIF, DLM và Motion, được áp dụng trong phiên bản VMAF hiện tại. Để giải thích rõ hơn, VIF định lượng sự mất mát của thông tin hình ảnh từ quan điểm của lý thuyết thông tin, trong khi DLM đo lường sự mất mát chi tiết ảnh hưởng đến hiển thị nội dung. Cả hai chỉ số đều nắm bắt được đặc điểm không gian của video, và chỉ số còn lại - Motion bao gồm đặc điểm thời gian. Sự sai khác về pixel được tính toán giữa các khung hình liên tiếp cho chỉ số Motion. Hơn nữa, VMAF sử dụng lợi thế của thuật toán học máy hiệu suất vượt trội thông qua học có giám sát (tức là hồi quy SVM) thay vì đánh trọng số truyền thống. Ban đầu, để tạo ra một bộ dữ liệu cho đánh giá chủ quan được tiêu chuẩn hóa, video với nhiều tính năng khác nhau được mã hóa ở các độ phân giải và tốc độ bit khác nhau, sau đó người quan sát đưa ra điểm số bằng cách so sánh những khiếm khuyết của video clip bị méo. Cuối cùng, mô hình VMAF sử dụng hồi quy SVM được đào tạo dựa trên điểm số thu thập được từ những người quan sát để gán trọng số phù hợp cho các chỉ số cơ bản. Dựa trên kết quả thực nghiệm cho thấy thang đo VMAF đạt độ chính xác cao trong các trường hợp video có độ phân giải cao.

III. PHƯƠNG PHÁP ĐỀ XUẤT

Trong một số phương pháp trước đây, các mô hình D-Q và R-Q được đề xuất sử dụng VMAF ở mức mã hóa khối hình (macroblock). Tuy nhiên, do kích thước khối hình nhỏ nên việc sử dụng VMAF để đo chất lượng sẽ không còn chính xác. Vì vậy, thông thường, giá trị VMAF sẽ được ước lượng thông qua đại lượng khách quan. Với cách tương tự, trong bài báo này, hàm số mô tả mối quan hệ giữa VMAF và SSE được xây dựng để tính độ méo của các khối hình. Từ đó, hàm R-Q và D-Q được thiết lập dựa trên VMAF.

Để thiết lập hàm R-D và D-Q, căn cứ trên mối quan hệ giữa độ méo và chất lượng tỷ lệ nghịch với nhau, độ méo của mỗi khung hình được tính như sau:

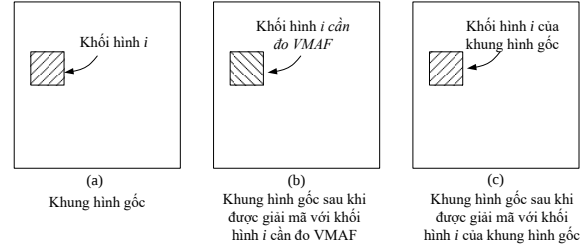
$$D = \frac{1}{VMAF} \quad (6)$$

Khi đó, công thức tính hàm chi phí (1) có dạng như sau:

$$J = \frac{1}{VMAF} + \lambda \cdot R \quad (7)$$

Để ước lượng giá trị VMAF thông qua giá trị SSE, một hàm gần đúng mô tả mối tương quan giữa VMAF và SSE được xây dựng dựa trên kết quả đo VMAF và SSE. Cụ thể, sau khi mã hóa 05 chuỗi video với độ phân giải 352x288, 1280x720 và 1920x1080 với các mức lượng tử QP khác

nhau, giá trị VMAF và SSE của các khối hình kích thước 64x64, 128x128 được ước lượng. Dựa trên tập dữ liệu này, đường cong mô tả mối quan hệ giữa SSE và VMAF được xây dựng.



Hình 2. Phương pháp ước lượng VMAF cho các khối hình.

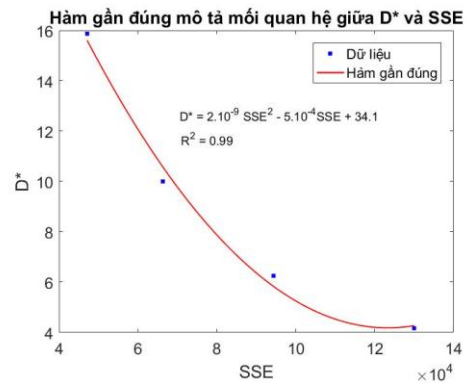
Hình 2 mô tả quá trình ước lượng giá trị VMAF cho một khối hình thứ i trong khung hình sau giải mã. Giả sử khối hình thứ i có kích thước bất kỳ trong khung hình sau giải mã cần đo VMAF. Khung hình sau giải mã có mức chất lượng là $VMAF_{rec}$. Sau khi thay khối hình i của khung hình gốc vào khung hình sau giải mã, mức chất lượng của khung hình sau khi được giải mã là $VMAF'_{rec}$. Khi đó, giá trị ước lượng VMAF của khung hình i sẽ là:

$$VMAF_i = |VMAF_{rec} - VMAF'_{rec}| \quad (8)$$

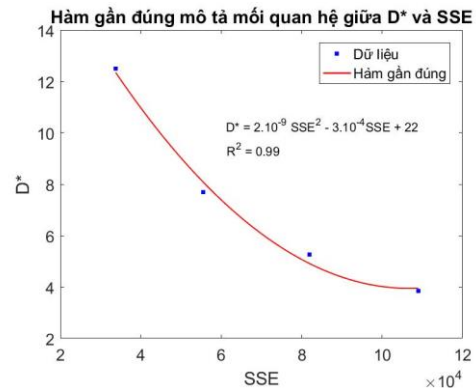
Để ước lượng độ méo dựa trên VMAF, từ (6) và (8) ta có độ méo của khối hình thứ i sẽ là:

$$D_i = \frac{1}{|VMAF_{rec} - VMAF'_{rec}|} \quad (9)$$

Hình 3 mô tả giá trị D_i của các khối hình và giá trị SSE_i tương ứng của khối hình đó tại 4 mức QP bằng 26, 30, 34 và 38.



Hình 3. Mối quan hệ giữa D^* và SSE của chuỗi BQMall



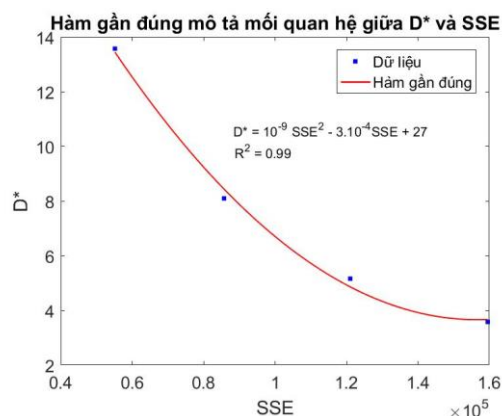
Hình 4. Mối quan hệ giữa D^* và SSE của chuỗi Basketball

Bảng 1. Độ chính xác của hàm gần đúng mô tả mối quan hệ giữa D^* và SSE

Chuỗi video	R square của hàm VMAF-PSNR
Basketball	0.99
Partyscene	0.99
BQMall	0.99
RacingHorse	0.97
Campus	0.98

Từ kết quả đo ta thấy các hàm gần đúng mô tả mối quan hệ giữa D^* và SSE có độ chính xác lên tới 99%. Vì vậy, hàm gần đúng tổng quát được xây dựng dựa trên giá trị trung bình của các giá trị D^* và SSE của 05 chuỗi video thử nghiệm. Sau khi lấy trung bình giá trị độ méo dựa trên VMAF của các khối hình (D^*) và độ méo dựa trên SSE của các khối hình tại mỗi mức QP, ta có hàm gần đúng mô tả mối quan hệ độ méo của khối hình dựa trên VMAF và giá trị SSE như sau:

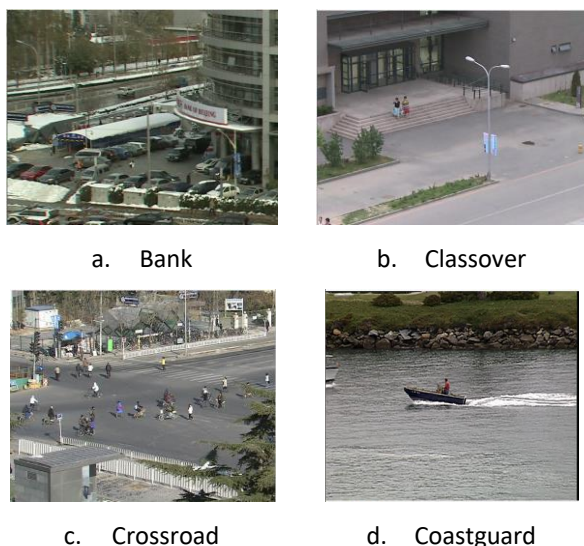
$$D^* = 10^{-9}SSE^2 - 3.10^{-4}SSE + 27 \quad (10)$$



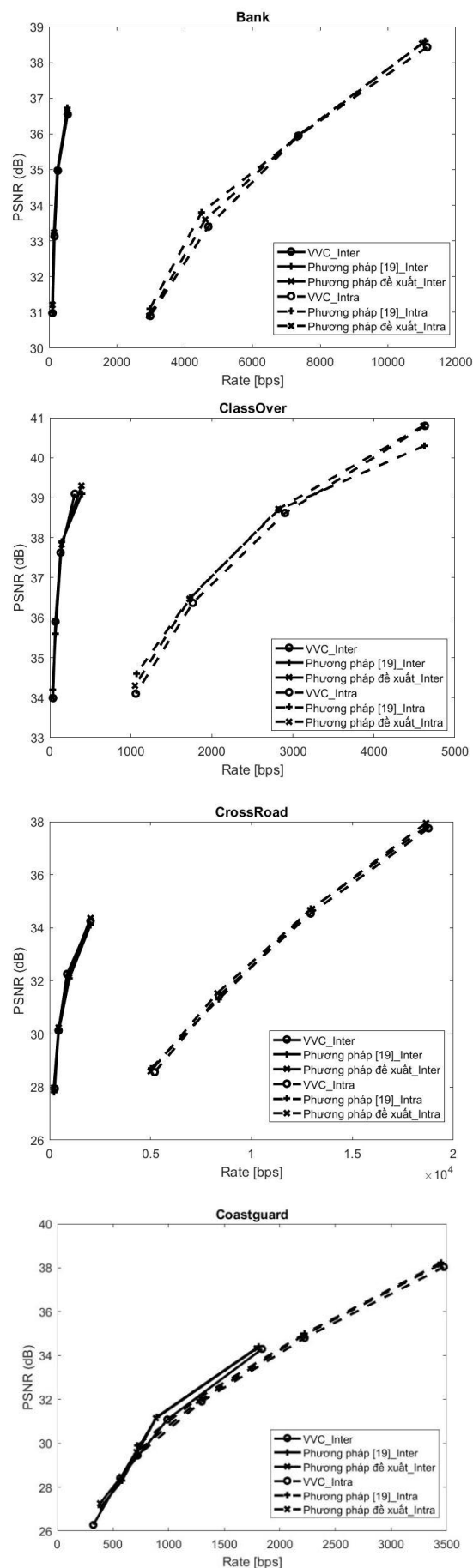
Hình 5. Hàm gần đúng mô tả mối quan hệ giữa D^* và SSE

Thay giá trị độ méo trong công thức (10) vào công thức (1) ta có hàm tối ưu mới như sau:

$$J = D^* + \lambda. R \quad (11)$$



Hình 6. Khung hình đầu tiên của 4 chuỗi video thử nghiệm



Hình 7. So sánh hiệu năng giữa phương pháp đề xuất với chuẩn H.266/VVC trong chế độ mã hóa nội khung (Inter coding mode) và liên khung (Intra coding mode)

Bảng 2. So sánh hiệu năng Độ méo – Tốc độ bit phương pháp đề xuất với H.266 và phương pháp [19]

Video	QP	H.266/VVC		Phương pháp [19]		Phương pháp đề xuất		Phương pháp đề xuất và H.266		Phương pháp đề xuất và [19]	
		Bitrate	PSNR	Bitrate	PSNR	Bitrate	PSNR	BD-Rate	BD-PSNR	BD-Rate	BD-PSNR
Bank_Inter coding mode	26	546.1	36.55	535.1	36.73	530.8	36.61	-8.41	0.24	-1.48	0.14
	30	255.3	34.97	250.9	35	240.1	34.98				
	34	156.7	33.13	140.7	33.2	143.4	33.24				
	38	97	30.98	105	31	95.1	31.21				
Bank_Intra coding mode	26	11159	38.43	11100	38.6	11009	38.53	-2.68	0.16	2.35	-0.13
	30	7359.5	35.95	7320	35.9	7339.5	35.95				
	34	4698	33.4	4500	33.8	4610	33.6				
	38	2983	30.9	2970	31.1	2953	30.9				
ClassOver_Inter	26	308.6	39.1	396.2	39.1	390.1	39.3	1.41	-0.01	-1.75	0.04
	30	132.5	37.63	148	37.9	145.5	37.83				
	34	69.7	35.9	70.1	35.6	73.7	35.9				
	38	39.1	33.99	35.2	34.2	34.1	34				
ClassOver_Intra	26	4630.1	40.8	4624	40.3	4610.1	40.8	-4.24	0.20	-0.71	0.04
	30	2902.4	38.62	2823	38.7	2820.4	38.72				
	34	1765	36.37	1730	36.5	1725	36.47				
	38	1061.4	34.1	1071	34.6	1051.4	34.3				
CrossRoad_Inter	26	2026.9	34.26	2020.7	34.1	2024.2	34.37	-0.34	-0.03	-7.12	0.21
	30	862.1	32.26	980.9	32.1	966.4	32.27				
	34	444.2	30.13	453.2	30.1	448.1	30.24				
	38	256.3	27.94	219.1	27.81	210.7	27.98				
CrossRoad_Intra	26	18765.3	37.75	18651.1	37.8	18645.3	37.95	-2.45	0.18	-1.43	0.13
	30	12929.2	34.54	12940.3	34.72	12900.1	34.67				
	34	8383	31.44	8380	31.31	8309	31.54				
	38	5201.3	28.56	5021.3	28.7	5001.3	28.59				
Coastguard_Inter	26	1840.6	34.3	1810.1	34.41	1791.8	34.32	-8.32	0.43	-1.50	0.04
	30	988.13	31.08	893	31.2	885.5	31.14				
	34	564.14	28.41	582	28.3	574.2	28.48				
	38	322.8	26.28	390.1	27.1	382.9	27.26				
Coastguard_Intra	26	3475.5	38.03	3450.3	38.23	3440.5	38.14	-2.75	0.15	2.22	-0.12
	30	2223.5	34.8	2223.1	35	2210	34.91				
	34	1298	31.9	1281.1	32.1	1270.6	31.92				
	38	720.13	29.44	717.1	29.9	710.13	29.6				

IV. ĐIỀU KIỆN THỬ NGHIỆM VÀ PHÂN TÍCH KẾT QUẢ

A. Điều kiện thử nghiệm

Để đánh giá hiệu năng của phương pháp đề xuất, 4 chuỗi video bao gồm các chuỗi *Bank 720x576*, *Classover 720x576*, *Crossroad 720x576* và *Coastguard 352x288* được sử dụng. Độ dài của mỗi chuỗi là 50 khung hình.

Hình 6 mô tả khung hình đầu tiên của các chuỗi video thử nghiệm. 4 chuỗi video dung thử nghiệm đại diện cho 4 loại video có nhiều chi tiết, ít chi tiết, chuyển động chậm và chuyển động nhanh [20]. Trong bài báo này, bộ mã hóa

video VVC được sử dụng để mã hóa các chuỗi video [21]. Kết quả của phương pháp đề xuất được so sánh với kết quả của bộ mã hóa VVC và phương pháp trong tài liệu [19] trong hai chế độ mã hóa là liên khung (Inter coding) và nội khung (Intra coding). Tham số được sử dụng để so sánh hiệu năng của các phương pháp bao gồm BD-Rate và BD-PSNR trong đó BD-Rate dùng để so sánh tốc độ bit và BD-PSNR dùng để so sánh chất lượng của video [22].

B. Kết quả mô phỏng

Hình 7 và Bảng 2 so sánh hiệu năng của phương pháp đề xuất với chuẩn VVC và phương pháp đề xuất trong tài liệu [19]. Kết quả cho thấy bitrate của phương pháp đề xuất

đạt được thấp hơn các phương pháp còn lại từ 0.71% đến 8.41% với các giá trị QP 26, 30, 34 và 38. Tuy nhiên, trong trường hợp chuỗi Classover chế độ mã hóa liên khung, giá trị BD-rate của phương pháp đề xuất thấp hơn của VVC và trong trường hợp chuỗi Bank, Coastguard chế độ mã hóa nội khung, giá trị BD-rate của phương pháp đề xuất thấp hơn của phương pháp [19]. Điều này là do trong trường hợp chuỗi video có chuyển động nhanh, việc đánh giá độ méo thông qua VMAF chưa đạt được hiệu quả cao như SSE. Tuy nhiên, trong hầu hết các trường hợp còn lại, phương pháp đề xuất đều thu được chất lượng cao hơn trong khi bitrate của phương pháp đề xuất đều thấp hơn của các phương pháp còn lại. Kết quả này thu được là do trong quá trình RDO, phương pháp đề xuất sử dụng VMAF để ước lượng độ méo cho các khối hình chính xác hơn trường hợp sử dụng SSE. Do đó, quá trình tối ưu độ méo-tốc độ bit của phương pháp đề xuất đạt hiệu quả cao hơn, dẫn đến bộ mã hóa có chất lượng hình ảnh tốt hơn cũng như tốc độ bit thấp hơn.

V. KẾT LUẬN

Bài báo đã đề xuất một phương pháp ước lượng giá trị VMAF thông qua giá trị SSE để áp dụng vào quá trình tối ưu hóa tốc độ và độ méo. Cụ thể, phương pháp xây dựng hàm mô tả mối quan hệ giữa VMAF và SSE, từ đó áp dụng vào quá trình RDO để dự đoán giá trị QP cho từng khối hình. Kết quả mô phỏng cho thấy phương pháp đề xuất đạt được hiệu năng nén cao hơn chuẩn mã hóa VVC và phương pháp được đề xuất trước đó xét theo giá trị BD-PSNR và BD-Rate. Cụ thể, giá trị BD-Rate của phương pháp đề xuất thấp hơn VVC và phương pháp trước đó lần lượt là 8.41% và 7,12%, BD-PSNR của phương pháp đề xuất cao hơn của VVC là 0.43% và của phương pháp trước đó là 0,21%.

TÀI LIỆU THAM KHẢO

- [1] S. Bouaafia, R. Khemiri and F. E. Sayadi, "Rate-Distortion Performance Comparison: VVC vs. HEVC," *2021 18th International Multi-Conference on Systems, Signals & Devices (SSD)*, Monastir, Tunisia, 2021, pp. 440-444, doi: 10.1109/SSD52085.2021.9429377.
- [2] B. Bross, J. Chen, J. -R. Ohm, G. J. Sullivan and Y. -K. Wang, "Developments in International Video Coding Standardization After AVC, With an Overview of Versatile Video Coding (VVC)," in *Proceedings of the IEEE*, vol. 109, no. 9, pp. 1463-1493, Sept. 2021, doi: 10.1109/JPROC.2020.3043399.
- [3] S. Bianco, L. Celona, P. Napoletano, and R. Schettini, "On the use of deep learning for blind image quality assessment," *Signal, Image Video Process.*, vol. 12, no. 2, pp. 355-362, 2018, doi: 10.1007/s11760-017-1166-8.
- [4] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, Apr. 2004, doi: 10.1109/TIP.2003.819861.
- [5] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312-322, 2004, doi: 10.1109/TBC.2004.834028.
- [9] Y. H. Huang, T. S. Ou, P. Y. Su, and H. H. Chen, "Perceptual rate-distortion optimization using structural similarity index as quality metric," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1614-1624, 2010, doi: 10.1109/TCSVT.2010.2087472.
- [10] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-motivated rate-distortion optimization for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 516-529, 2012, doi: 10.1109/TCSVT.2011.2168269.
- [11] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "Perceptual video coding based on SSIM-inspired divisive normalization," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1418-1429, 2013, doi: 10.1109/TIP.2012.2231090.
- [12] A. M. Z. Li, C. Bampis, J. Novak, A. Aaron, K. Swanson and J. Cock, "Vmaf: The journey continues," *Netflix Technol. Blog*, vol. 25, 2018.
- [13] R. Rassool, "VMAF reproducibility: Validating a perceptual practical video quality metric," *IEEE Int. Symp. Broadband Multimed. Syst. Broadcast. BMSB*, 2017, doi: 10.1109/BMSB.2017.7986143.
- [14] C. Lee, S. Woo, S. Baek, J. Han, J. Chae, and J. Rim, "Comparison of objective quality models for adaptive bit-streaming services," *2017 8th Int. Conf. Information, Intell. Syst. Appl. IISA 2017*, vol. 2018-Janua, pp. 1-4, 2018, doi: 10.1109/IISA.2017.8316385.
- [15] N. Barman, S. Schmidt, S. Zadtootaghaj, M. G. Martini, and S. Möller, "An evaluation of video quality assessment metrics for passive gaming video streaming," *Proc. 23th ACM Work. Pack. Video, PV 2018*, pp. 7-12, 2018, doi: 10.1145/3210424.3210434.
- [16] C. Zhu, Y. Huang, R. Xie and L. Song, "HEVC VMAF-oriented Perceptual Rate Distortion Optimization using CNN," *2021 Picture Coding Symposium (PCS)*, Bristol, United Kingdom, 2021, pp. 1-5, doi: 10.1109/PCS50896.2021.9477459.
- [17] T. H. Vu, H. P. Cong, T. Sisouvang, X. HoangVan, S. NguyenQuang and M. DoNgoc, "VMAF based quantization parameter prediction model for low resolution video coding," *2022 International Conference on Advanced Technologies for Communications (ATC)*, Ha Noi, Vietnam, 2022, pp. 364-368, doi: 10.1109/ATC55345.2022.9942982.
- [18] Vu, T.H.; Do, M.N.; Nguyen, S.Q.; PhiCong, H.; Sisouvang, T.; HoangVan, X. Learning Adaptive Quantization Parameter for Consistent Quality Oriented Video Coding. *Electronics* 2023, 12, 4905. <https://doi.org/10.3390/electronics12244905>
- [19] S. Deng, J. Han, and Y. Xu, "VMAF Based Rate-Distortion Optimization for Video Coding," *IEEE 22nd Int. Work. Multimed. Signal Process. MMSP 2020*, 2020, doi: 10.1109/MMSP48831.2020.9287114.
- [20] Xiph.org Video Test Media, [Online]. Available: <https://media.xiph.org/video/derf/>
- [21] VVCSoftware_VTM-12.1. [Online]. Available: https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tree/VTM-12.1
- [22] G. Bjontegaard, Calculation of average PSNR differences between RD-curves (VCEG-M33).

IMPROVING DISTORTION ESTIMATION IN RDO PROCESS FOR H.266/VVC VIDEO CODING

Abstract: Versatile Video Coding (VVC) also known as H.266 is the latest video coding standard and the successor of High Efficiency Video Coding (HEVC). As its name suggests, VVC is an efficient encoding standard for various video contents and applications. In VVC, many new techniques compared to HEVC are applied but there is few research on rate-distortion optimization (RDO) - an important technique in video encoding to determine the coding mode for coding units. In this article, to improve RDO for VVC, a new algorithm using objective quality assessment metric to estimate distortion for coding units is proposed. Then, the estimated distortion is applied in RDO to improve Rate-Distortion performance of encoder. The

proposed algorithm achieved bit rate savings up to 8.41% when implemented on H266/VVC encoder.

Keyword: VMAF, RDO, H.266/VVC.



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.



Thippaphone Sisouvong tốt nghiệp đại học và cao học ngành Công nghệ Thông tin vào các năm 2004, 2015 tại Đại học Quốc gia Lào và Học viện Công nghệ Bưu chính Viễn thông. Hiện đang là nghiên cứu sinh tại 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.