

# CẢI THIỆN HIỆU NĂNG RDO DỰA TRÊN THANG ĐO CHẤT LƯỢNG VMAF TRONG MÃ HÓA VIDEO HEVC

Nguyễn Thị Hương Thảo  
Học viện Công nghệ Bưu chính Viễn thông

**Tóm tắt:** Trong mã hóa video, tối ưu hóa tốc độ mã hóa - độ méo (Rate Distortion Optimization – RDO) là một trong những kỹ thuật chính giúp lựa chọn phương pháp nén tối ưu cho video. Cụ thể, dựa trên quá trình RDO, bộ mã hóa lựa chọn các tham số để mã hóa video sao cho tỷ lệ nén là cao nhất trong khi chất lượng video đạt được sau giải mã là tốt nhất có thể. Trong quá trình RDO, bộ mã hóa sử dụng các hàm mô tả mối quan hệ giữa tốc độ mã hóa (R) - tham số lượng tử (QP) và độ méo (D) – tham số lượng tử để tìm ra giá trị QP tối ưu sao cho tốc độ mã hóa là nhỏ nhất trong khi độ méo là nhỏ nhất. Các nghiên cứu trước đây thường sử dụng phương pháp đo khách quan để ước lượng độ méo của video trong quá trình này. Tuy nhiên, thang đo chất lượng video bằng phương pháp khách quan không phản ánh chính xác chất lượng của video so với phương pháp chủ quan. Vì vậy, trong bài báo này, quá trình RDO được cải tiến bằng cách dùng thang đo chủ quan có tên là Video Multiview Assessment Fusion (VMAF) để đánh giá độ méo. Kết quả cho thấy phương pháp đề xuất có thể tiết kiệm 4.89% bitrate trong khi chất lượng được tăng lên 5.98% so với một số phương pháp tham chiếu.

**Từ khóa:** VMAF, RDO, HEVC.

## I. GIỚI THIỆU

Mã hóa video hiệu năng cao (High Efficiency Video Coding – HEVC) được coi là một trong những tiêu chuẩn mã hóa video mới hiện nay cho video độ phân giải cao. Bộ mã hóa video HEVC có hiệu năng cao bởi sự mở rộng các tham số mã hóa trong quá trình RDO so với các chuẩn mã hóa trước đây. Quá trình này giúp HEVC lựa chọn được các tham số mã hóa tối ưu nhất trong quá trình mã hóa. Tuy nhiên kỹ thuật RDO truyền thống thường có chi phí tính toán cao bởi quá trình này cần phải khôi phục lại video để tính chi phí tốc độ bit – độ méo (RD). Chính vì vậy, độ phức tạp cao của quá trình RDO là một trong những hạn chế của bộ mã hóa khi triển khai trong các ứng dụng thời gian thực.

Trong mã hóa video, chất lượng (tỷ lệ nghịch với độ méo) của hình ảnh có thể được đo bằng phương pháp chủ quan hoặc khách quan. Phương pháp đánh giá chủ quan được thực hiện bằng cách tính trung bình các kết quả đánh giá của người xem [1]. Ưu điểm của phương pháp này là kết quả đo phản ánh chất lượng tương đối gần với chất

lượng do mắt người cảm nhận. Tuy nhiên, nhược điểm của nó là sự phức tạp trong triển khai đánh giá vì cần nhiều người và nhiều thời gian. Vì vậy, phương pháp đánh giá khách quan thường được sử dụng trong các ứng dụng thực tế. Đại lượng khách quan phổ biến được dùng trong mã hóa video là đại lượng tổng bình phương của lỗi (Sum of Squared Error – SSE). Tuy nhiên, giá trị SSE không phản ánh chính xác chất lượng do cảm nhận mắt người cảm nhận. Vì vậy, một số đại lượng được đề xuất bao gồm SSIM [2], VQM [3]. Trong thời gian gần đây, công ty cung cấp nội dung video Netflix đã phát triển một đại lượng lai ghép giữa phương pháp chủ quan và khách quan nhằm đánh giá chất lượng được gọi là VMAF [4]. Đại lượng VMAF ước lượng giá trị chất lượng do mắt người cảm nhận bằng mô hình học máy SVM kết hợp với một số phương pháp đo khách quan. Nhiều nghiên cứu [5][6][7] đã chỉ ra rằng VMAF đạt kết quả tốt hơn các phương pháp đo khách quan trước đây và nó có mối tương quan lớn so với thang đo đánh giá chủ quan của người xem (Mean Opinion Score – MOS).

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. Trong bài báo [8], đại lượng SSIM được sử dụng để đo độ méo trong quá trình RDO. Các phương pháp trong [9] [10] đề xuất sử dụng SSIM để tính các hệ số trong quá trình biến đổi DCT. Trong bài báo [11], 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 (offline) 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

Tác giả liên hệ: Nguyễn Thị Hương Thảo  
Email: thaonth@ptit.edu.vn

Đến tòa soạn: 10/2023, chỉnh sửa: 11/2023, chấp nhận đăng: 12/2023

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 [12][13]. 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 Lagrangian được sử dụng. Ý tưởng cơ bản của phương pháp này là tìm nhân tử Lagrangian và các giá trị D, R sao hàm chi phí đạt giá trị nhỏ nhất:

$$J = D + \lambda \cdot R \tag{1}$$

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

$$\frac{dJ}{dR} = \frac{dD}{dR} + \lambda = 0 \tag{2}$$

Trong tài liệu [13], mối quan hệ giữa R và D được biểu diễn bằng biểu thức:

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

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

$$D = \frac{QP^2}{3} \tag{4}$$

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

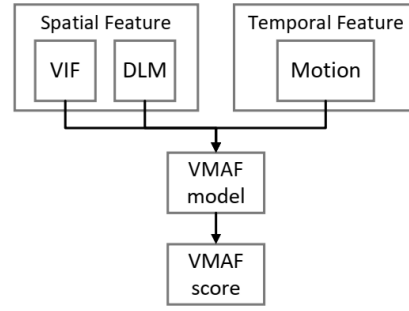
$$\lambda = -\frac{dD}{dR} = c \cdot QP^2 \tag{5}$$

Với c là hằng số và có giá trị bằng 0,136 trong tiêu chuẩn H.264. 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  $\lambda$  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 đánh giá rộng rãi trong năm qua, nhưng ba chỉ số cơ bản, tức 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 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 kế cho chỉ số Motion. Hơn nữa, VMAF sử dụng lợi thế của máy học 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, việc này gặp khó khăn trong việc tích hợp đại lượng chủ quan vào quá trình RDO vì không tồn tại công thức để đo chất lượng dựa trên chủ quan của mắt người. Vì vậy, để xây dựng hàm RDO, các phương pháp trước đây xây dựng công thức tính xấp xỉ VMAF dựa trên các đại lượng khách quan. Với cách tương tự, trong bài báo này, hàm số mô tả mối quan hệ giữa VMAF và PSNR được xây dựng để tính VMAF dựa trên PSNR. Từ đó, hàm R-Q và D-Q được thiết lập dựa trên VMAF.

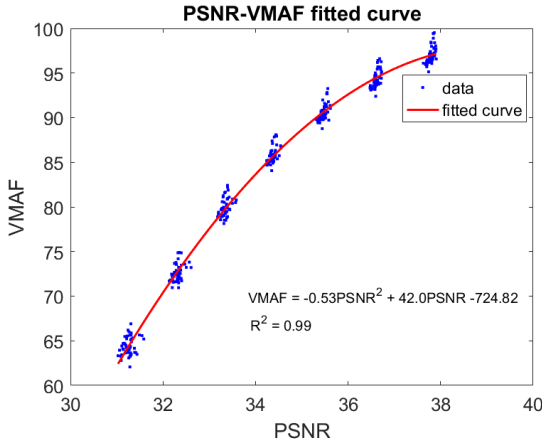
Để thiết lập hàm R-D và D-Q, căn cứ trên mối quan hệ giữa độ méo và chất lượng tỷ lệ nghịch với nhau, độ méo của mỗi khung hình được tính như sau:

$$D = \frac{1}{VMAF} \tag{6}$$

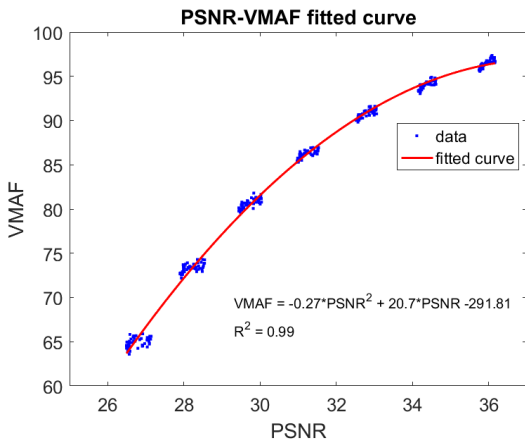
Khi đó, công thức tính hàm chi phí (1) có dạng như sau:

$$J = \frac{1}{VMAF} + \lambda \cdot R \tag{7}$$

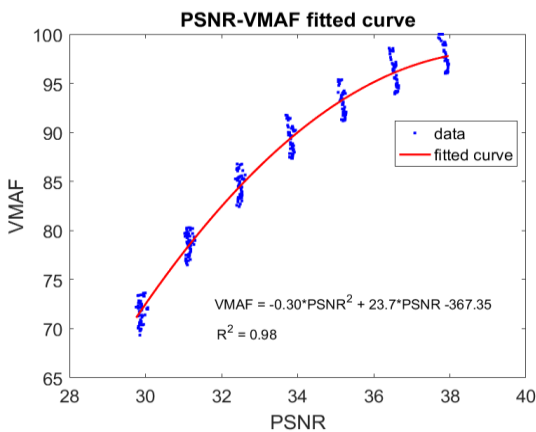
Để ước lượng giá trị VMAF thông qua giá trị PSNR, một hàm gần đúng mô tả mối tương quan giữa VMAF và PSNR được xây dựng dựa trên kết quả đo VMAF và PSNR. Cụ thể, sau khi mã hóa 10 chuỗi video với độ phân giải 352x288, 1280x720 và 1920x1080, chất lượng của các chuỗi video sau giải nén được đo ở thang đo PSNR và VMAF. Dựa trên tập dữ liệu này, đường cong mô tả mối quan hệ giữa PSNR và VMAF được xây dựng. Hình 2 mô tả mối quan hệ giữa PSNR và VMAF của 4 chuỗi video Basketball, PartyScene, BQMall, và Racing Horse và Bảng 1 mô tả độ chính xác của các hàm gần đúng.



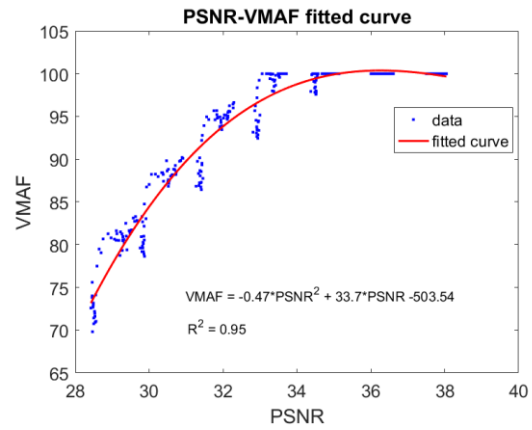
Hình 2. Hàm số mô tả mối quan hệ giữa VMAF và PSNR của chuỗi Basketball



Hình 3. Hàm số mô tả mối quan hệ giữa VMAF và PSNR của chuỗi PartyScene



Hình 4. Hàm số mô tả mối quan hệ giữa VMAF và PSNR của chuỗi BQMall



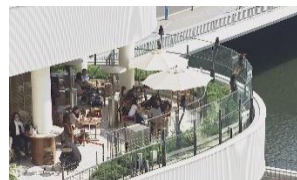
Hình 5. Hàm số mô tả mối quan hệ giữa VMAF và PSNR của chuỗi Racinghorse

Bảng 1. Độ chính xác của hàm gần đúng mô tả mối quan hệ giữa VMAF và PSNR

Chuỗi video	R square của hàm VMAF-PSNR
Hall	0.97
City	0.96
Foreman	0.93
Crew	0.87
Four-people	0.95
Ice	0.97
Basketball	0.99
Partyscene	0.99
BQMall	0.98
Racinghorse	0.95
<b>Average</b>	<b>0.96</b>

Từ kết quả đo ta thấy các hàm gần đúng mô tả mối quan hệ giữa VMAF và PSNR có độ chính xác tương đối cao (trung bình 96%). Vì vậy, hàm gần đúng tổng quát được xây dựng dựa trên giá trị trung bình của các hệ số của các hàm cho 10 chuỗi video thử nghiệm. Cụ thể, hàm gần đúng tổng quát có dạng như sau:

$$VMAF = -0.3 * PSNR + 25.3 * PSNR - 290.5 \quad (8)$$



a. BQTerrace



b. BasketballDrive

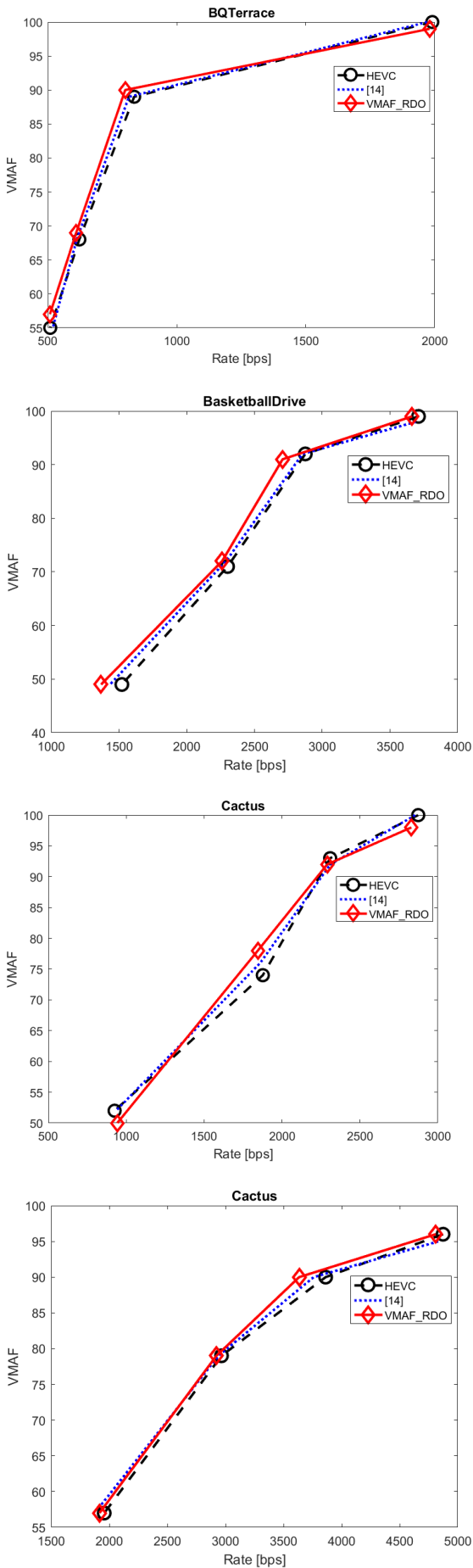


c. Cactus



d. Kimono

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 HEVC và phương pháp trong tài liệu [14]

#### 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 BQTerrace 1920x1080, BasketballDrive 1920x1080, Kimono 1280x720 và Cactus 1280x720 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. Trong bài báo này, bộ mã hóa video HEVC được sử dụng để mã hóa các chuỗi video. Kết quả của phương pháp đề xuất được so sánh với kết quả của bộ mã hóa HEVC trong một số chế độ mã hóa.

Trong bài báo này, để đánh giá hiệu năng của phương pháp đề xuất, kết quả BD-Rate và BD-VMAF của phương pháp đề xuất sẽ được so sánh với phương pháp đo VMAF trong tài liệu [14] và chuẩn mã hóa video HEVC.

Theo tài liệu [14], giá trị VMAF và PSNR tương ứng với 5 cấp độ đo của MOS như trong Bảng 2 sau:

Bảng 2. Mối tương quan giữa VMAF và PSNR với mức chất lượng theo thang đo MOS

Chất lượng video	PSNR	VMAF
Excellent	> 38	> 90
Good	35-38	74-90
Fair	33-35	58-74
Poor	30-33	38-58
Bad	< 30	< 38

Để chuyển đổi từ thang đo PSNR sang thang đo VMAF, các hàm tuyến tính mô tả mối quan hệ giữa PSNR và VMAF dựa trên các khoảng giá trị tương đồng trên mỗi mức chất lượng được tính như sau:

$$VMAF = \begin{cases} 1.3xPSNR & PSNR \leq 30 \\ 6.67xPSNR - 162 & 30 < PSNR \leq 33 \\ 8xPSNR - 206 & 33 < PSNR \leq 35 \\ 5.33xPSNR - 113 & 35 < PSNR \leq 38 \\ 5xPSNR - 100 & PSNR > 38 \end{cases} \quad (9)$$

Thay công thức (9) vào công thức (7) ta có công thức tính chi phí mã hóa dựa trên VMAF.

##### B. Kết quả mô phỏng

Hình 7 và Bảng 3 so sánh hiệu năng của phương pháp đề xuất với chuẩn HEVC và phương pháp đo VMAF trong tài liệu [14].

Kết quả cho thấy bitrate của phương pháp đề xuất đạt được thấp hơn các phương pháp còn lại từ 0.21 đến 4.89% với các giá trị QP 27, 30, 32 và 37. Trong khi đó chất lượng của phương pháp đề xuất cải thiện hơn so với HEVC và phương pháp trong tài liệu [14]. Tuy nhiên, trong trường hợp chuỗi Cactus, giá trị BD-rate của phương pháp đề xuất thấp hơn của phương pháp [14]. Điều này là do trong trường hợp QP=32, phương pháp đề xuất có bitrate cao hơn trong khi chất lượng kém hơn. Tuy nhiên, trong hầu hết các trường hợp còn lại, phương pháp đề xuất đều thu được chất lượng cao hơn hoặc bằng, trong khi bitrate của phương pháp đề xuất đều thấp hơn các phương pháp còn lại. Kết quả này thu được là do trong quá trình RDO, phương pháp đề xuất sử dụng thang đo VMAF thay vì

Bảng 3. So sánh hiệu năng giữa phương pháp được đề xuất với chuẩn HEVC và phương pháp trong tài liệu [14]

Video	QP	HEVC		[14]		VMAF_RDO		BD-Rate giữa VMAF_RDO và HEVC	BD-VMAF giữa VMAF_RDO và HEVC	BD-Rate giữa VMAF_RDO và [14]	BD-VMAF giữa VMAF_RDO và [14]
		Bitrate	VMAF	Bitrate	VMAF	Bitrate	VMAF				
BasketballDrive	27	3710.08	99.00	3690.01	98.00	3663.26	99.00	<b>-3.50</b>	<b>1.29</b>	<b>-2.41</b>	<b>0.55</b>
	30	2875.24	92.00	2855.30	92.00	2707.84	91.00				
	32	2301.74	71.00	2290.04	72.00	2261.19	72.00				
	37	1521.73	49.00	1434.02	49.00	1368.44	49.00				
BQTerrace	27	1990.42	100.00	1970.21	100.00	1980.00	99.00	<b>-4.89</b>	<b>3.63</b>	<b>-3.18</b>	<b>3.69</b>
	30	834.69	89.00	814.59	89.00	800.00	90.00				
	32	622.18	68.00	612.09	68.00	610.00	69.00				
	37	509.33	55.00	519.20	55.00	508.00	57.00				
Cactus	27	2874.12	100.00	2854.89	100.00	2830.00	98.00	<b>-0.24</b>	<b>5.98</b>	1.52	<b>1.74</b>
	30	2309.05	93.00	2310.19	92.00	2290.10	92.00				
	32	1876.84	74.00	1860.45	76.00	1844.79	78.00				
	37	925.63	52.00	934.90	52.00	942.76	50.00				
Kimono1	27	4873.45	96.00	4833.24	95.00	4810.56	96.00	<b>-0.21</b>	<b>0.91</b>	<b>-0.56</b>	<b>0.37</b>
	30	3862.52	90.00	3759.67	90.00	3634.68	90.00				
	32	2966.17	79.00	2940.04	79.00	2922.71	79.00				
	37	1957.00	57.00	1923.89	58.00	1915.70	57.00				

PSNR. Do đó, kết quả đầu ra bộ giải mã khi đo bằng VMAF cũng sẽ tốt hơn so với khi đo bằng PSNR.

### V. KẾT LUẬN

Bài báo đã đề xuất một phương pháp ước lượng giá trị VMAF thông qua giá trị PSNR 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 PSNR, 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 giá trị VMAF tại đầu ra của bộ giải mã lớn hơn hoặc bằng với các phương pháp tham chiếu. Ngoài ra, hiệu năng của phương pháp đề xuất cũng cao hơn chuẩn mã hóa HEVC và phương pháp tham chiếu xét theo giá trị BD-VMAF và BD-Rate. Cụ thể, giá trị BD-VMAF của phương pháp đề xuất lớn hơn các phương pháp còn lại là 5.98% và BD-Rate của phương pháp đề xuất thấp hơn của các phương pháp còn lại là 4.89%.

### TÀI LIỆU THAM KHẢO

[1] 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.

[2] 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.

[3] 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.

[4] A. M. Z. Li, C. Bampis, J. Novak, A. Aaron, K. Swanson and and J. Cock, "Vmaf: The journey continues," *Netflix Technol. Blog*, vol. 25, 2018.

[5] 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.

[6] 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.

[7] N. Barman, S. Schmidt, S. Zadtootaghaj, M. G. Martini, and S. Möller, "An evaluation of video ality assessment metrics for passive gaming video streaming," *Proc. 23th ACM Work. Pack. Video, PV 2018*, pp. 7–12, 2018, doi: 10.1145/3210424.3210434.

[8] 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.

[9] 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.

[10] 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.

- [11] 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.
- [12] S. Ma, W. Gao, and Y. Lu, "Rate-Distortion Analysis for H.264/AVC Video Coding and its Application to Rate Control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 12, pp. 1533–1543, 2005, doi: 10.1109/TCSVT.2005.857300.
- [13] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for: Video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, 1998, doi: 10.1109/79.733497.
- [14] A. Kruglov, "Interpretation of objective video quality metrics," [https://www.elecard.com/page/article\\_interpretation\\_of\\_metrics](https://www.elecard.com/page/article_interpretation_of_metrics), 2022.  
[https://www.elecard.com/page/article\\_interpretation\\_of\\_metrics](https://www.elecard.com/page/article_interpretation_of_metrics)

### IMPROVING VMAF-BASED RDO PROCESS IN VIDEO CODING

**Abstract:** In video encoding, Rate-Distortion optimization (RDO) is one of the main techniques to select an optimal coding mode for video. Specifically, based on the RDO process, the encoder selects the parameters to encode the coding units so that the bitrate is the lowest while quality of reconstructed video is highest. During the RDO process, the encoder uses models that describe the relationship between bitrate (R) - quantization parameter (QP) and distortion (D) - QP to find the optimal QP value. In previous studies, RDO often uses objective quality assessment methods to estimate video distortion during this process. However, the video quality assessment using the objective method does not accurately reflect the quality of the video compared to the subjective method. Therefore, in this paper, the RDO process is improved by using a subjective metric called Video Multiview Assessment Fusion (VMAF) to evaluate distortion. The results show that the proposed method can save 4.89% bitrate while the quality is increased by 5.98% compared to the benchmark methods.

**Keyword:** VMAF, RDO, HEVC.



**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.