

PHƯƠNG PHÁP ĐẢM BẢO CHẤT LƯỢNG VIDEO DỰA TRÊN KỸ THUẬT DỰ ĐOÁN THAM SỐ LƯỢNG TỬ

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

Tóm tắt: Với sự bùng nổ của các dịch vụ xem video trực tuyến trên mạng viễn thông như Youtube, Netflix... trong khi cơ sở hạ tầng mạng là hữu hạn, việc đảm bảo chất lượng hình ảnh cho các dịch vụ này ngày càng cấp thiết. Hiện nay, có rất nhiều tham số đánh giá chất lượng hình ảnh video trong đó tham số đo tần suất thay đổi mức chất lượng hình ảnh là một trong những tham số quan trọng. Đối với người xem video, sự thay đổi chất lượng hình ảnh với tần suất cao sẽ gây ra sự khó chịu khi xem. Việc này có thể sẽ xảy ra thường xuyên khi chất lượng mạng không ổn định. Vì vậy, trong bài báo này, một phương pháp đảm bảo chất lượng video ổn định dựa trên kỹ thuật dự đoán tham số lượng tử được đề xuất. Cụ thể, dựa trên nội dung của từng khung hình video, mô hình học máy được đưa ra để dự đoán một mức lượng tử tương ứng cho từng khối hình trong khung hình nhằm đạt được mức chất lượng video mong muốn. Kết quả mô phỏng cho thấy phương pháp đề xuất đã đạt được hai tiêu chí đồng thời. Thứ nhất, phương pháp đề xuất đã đạt được mức chất lượng mong muốn cho video tái tạo. Thứ hai, hiệu năng nén của video trong phương pháp đề xuất tăng lên so với bộ mã hóa video tiêu chuẩn x.264.

Từ khóa: VMAF, Video coding, x.264.

I. GIỚI THIỆU

Với sự phát triển bùng nổ của dữ liệu trong các dịch vụ video trực tuyến trên mạng viễn thông, việc đảm bảo chất lượng trải nghiệm cho người xem là một trong những yêu cầu cấp thiết hiện nay. Chất lượng trải nghiệm được định nghĩa là mức độ hài lòng hoặc không hài lòng của người dùng khi sử dụng một dịch vụ hay ứng dụng nào đó [1]. Để đo mức độ hài lòng của người dùng có thể có rất nhiều tiêu chí như thời gian trễ khởi động, thời gian dừng hình, tần suất thay đổi mức chất lượng... Trong số các tiêu chí này, tần suất thay đổi chất lượng hình ảnh là một trong những tiêu chí quan trọng. Sự thay đổi chất lượng hình ảnh có thể xảy ra khi bộ mã hóa video điều chỉnh tham số lượng tử để đảm bảo tối ưu về tốc độ bit và độ méo (RDO). Tức là bộ mã hóa sẽ chọn ra giá trị tham số lượng tử tốt nhất sao cho tốc độ bit là nhỏ nhất và chất lượng hình ảnh là tốt nhất. Tuy nhiên, quá trình thực hiện RDO chỉ chú trọng tới mục tiêu để tối ưu giữa tốc độ bit và chất lượng nên sẽ dẫn đến tình trạng chất lượng ảnh đầu ra giữa các khung hình trong chuỗi video sẽ thay đổi

liên tục theo nội dung. Ngoài ra, trong quá trình RDO, chất lượng khung hình đều được đo bằng phương pháp khách quan, trong khi chất lượng hình ảnh video được người dùng cảm nhận là chủ quan. Vì vậy, việc đảm bảo tối ưu tốc độ bit và chất lượng, đồng thời đảm bảo chất lượng video tái tạo đạt mức chất lượng mong muốn là rất cần thiết.

Trong bài báo [2], phương pháp điều chỉnh hệ số Lagrangian λ theo nội dung video trong quá trình tối ưu tốc độ bit – độ méo (Rate-Distortion Optimization) để đảm bảo video sau khi khởi tạo luôn đạt được mức chất lượng nhất định. Cụ thể, hệ số Lagrangian và tham số lượng tử Q cho mỗi khung hình được tính sao cho giá trị khác biệt giữa chất lượng được dự đoán và chất lượng của khung hình sau giải mã là nhỏ nhất. Cùng với mục tiêu đạt được mức chất lượng không đổi giữa các khung hình, phương pháp [3] sử dụng hàm mật độ xác suất của các hệ số biến đổi để ước lượng độ sâu của cây mã hóa. Từ đó, chất lượng của các khối mã hóa được điều chỉnh để đảm bảo mức chất lượng ổn định. Trong bài báo [4] mô hình RDO thích ứng theo nội dung cho chuẩn mã hóa H.264 được sử dụng để ước lượng độ méo giữa khung hình gốc và khung hình sau giải mã. Dựa trên giá trị ước lượng, tham số QP cho mỗi khung hình được tìm để đạt được mức chất lượng mong muốn.

Trong các phương pháp trên, chất lượng video được đánh giá bằng tham số khách quan PSNR. Tuy nhiên, chất lượng tại phía người dùng được cảm nhận thông qua chủ quan của người dùng. Vì vậy, trong phương pháp đề xuất này, tham số VMAF kết hợp giữa hai phương pháp đánh giá khách quan và chủ quan được sử dụng để đánh giá chất lượng video.

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 [5], [6]. 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

Tác giả liên hệ: Vũ Hữu Tiến

Email: tienvh@ptit.edu.vn

Đến tòa soạn: 9/2022, chỉnh sửa: 10/2022, chấp nhận đăng: 10/2022

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:

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

Trong đó J là hàm chi phí Lagrangian và λ 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 [5], mối quan 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:

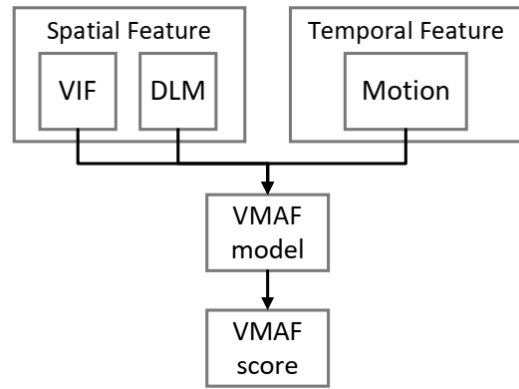
$$\lambda = -\frac{dD}{dR} = c.QP^2 \tag{5}$$

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 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 [7]. 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 đoạn giải thưởng, phim, phim hoạt hình, phim tài liệu.[8]

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 mát chi tiết ảnh hưởng đến hiển thị nội dung.



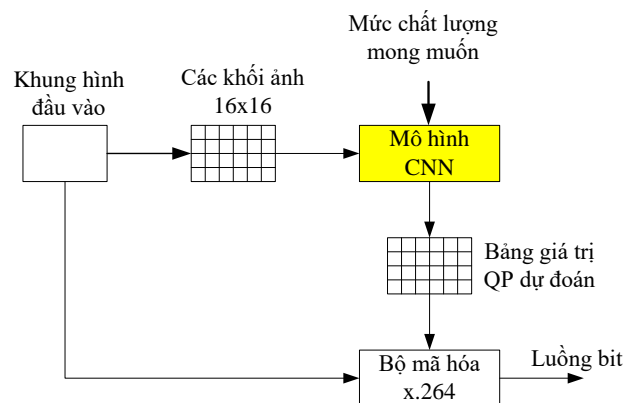
Hình 1. Minh họa sự kết hợp các chỉ số đo trong VMAF

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

A. Phương pháp đề xuất

Hình 2 cho thấy các bước tổng thể của thuật toán ước lượng bản đồ QP đề xuất. Ban đầu, khung hình hiện tại của chuỗi video được chia thành các mảng hình ảnh có kích thước 16 x 16. Mỗi mảng ghép hình ảnh được đưa vào một mô hình mạng nơ-ron tích chập (CNN) đã được huấn luyện cùng với điểm số VMAF dự kiến để ước tính giá trị QP. Sau khi ước tính giá trị QP cho tất cả các mảng ghép hình ảnh, bản đồ QP bao gồm tất cả các giá trị QP này được tạo và được đưa vào bộ mã hóa x.264 để mã hóa khung hình hiện tại. Đầu ra của bộ mã hóa là một khung hình được mã hóa có mức chất lượng tương ứng với điểm VMAF dự kiến. Trong đề xuất này, dải VMAF từ 40 đến 100 được chia thành 12 dải con (40 - 44, 45 - 49, ..., 90 - 94, 95 - 100). Mỗi dải con tương ứng với một mức chất lượng.



Hình 2. Các bước trong thuật toán đề xuất

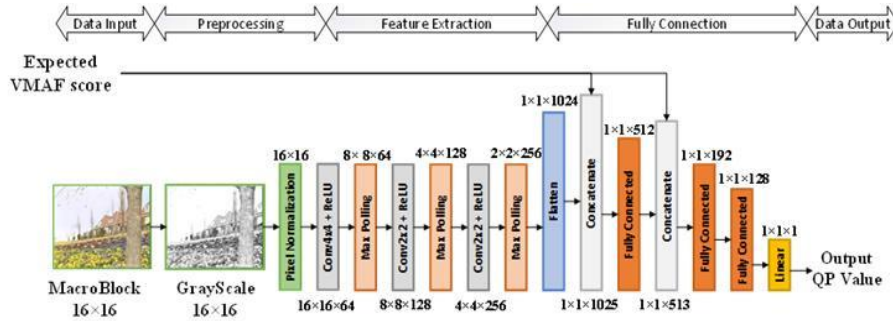
B. Mô Hình Dự Đoán QP

Trong phần này, chúng tôi giới thiệu mô hình học sâu để dự đoán QP của chế độ nội ảnh. Mô hình sử dụng các khối ảnh 16×16 (Macroblock – MB) làm đầu vào. Lý do sử dụng khối ảnh kích thước như vậy là do đây là kích thước lớn nhất bộ mã hóa x.264 có thể hỗ trợ. Cấu trúc mô hình (như được mô tả trong Hình 3) bao gồm:

Đề thu thập tập dữ liệu và bản đồ QP gốc phục vụ cho quá trình huấn luyện mô hình CNN, 15 video với độ phân giải 352×288 được nén ở 24 giá trị CRF (từ 22 đến 45).

Sau khi mã hóa, bản đồ QP chân lý cơ bản của từng mức chất lượng được trích xuất bằng quy trình sau:

- Đo điểm VMAF cho mỗi khung hình của video được



Hình 3. Kiến trúc mạng CNN đề xuất

- Các lớp tiền xử lý: MB được tiền xử lý trước bằng cách chuyển đổi thành hình ảnh xám và được chuẩn hóa thành giá trị trong dải 0-1.

- Các lớp nhân chập: Dữ liệu thông qua các lớp tiền xử lý sẽ được nhân chập với các nhân 4×4 ở lớp tích chập đầu tiên để trích xuất các đặc trưng mức thấp và nhân 2×2 cho các đặc trưng mức cao hơn.

- Các lớp kết nối đầy đủ: Các bản đồ đặc trưng từ các lớp tích chập được ghép nối với nhau và sau đó được làm phẳng thành một vector một cột. Và sau đó, tất cả các phân tử của vector được đưa qua ba lớp kết nối đầy đủ, lớp này dựa vào các đặc trưng được trích xuất bởi các lớp trước đó để dự đoán các giá trị QP. Hơn nữa, VMAF cũng có ảnh hưởng đáng kể đến tham số lượng tử hóa thích ứng. Khi VMAF giảm, QP có xu hướng tăng và ngược lại. Do đó, VMAF được bổ sung như một đặc trưng bên ngoài trong vector đặc trưng cho các lớp kết nối đầy đủ, cho phép mô hình tạo ra một tham số lượng tử hóa thích ứng.

- Các lớp khác: Trong giai đoạn huấn luyện, sau lớp tích chập, lớp pooling (gộp) đã được thêm vào để giảm kích thước của từng ma trận đặc trưng. Và sau đó, lớp chuẩn hóa được sử dụng để chuẩn hóa ma trận đặc trưng nhằm ổn định quá trình học và giảm số lượng các epochs huấn luyện. Bên cạnh đó, lớp dropout sẽ loại bỏ ngẫu nhiên các đặc trưng với xác suất 20%. Xác suất này được lựa chọn sau khi mô hình được thử nghiệm với nhiều giá trị dropout khác nhau và giá trị 20% có độ chính xác là cao nhất. Ngoài ra, mô hình cũng được thử nghiệm với một số kiến trúc mạng bao gồm 2, 3, 4, 5 và 6 lớp nhân chập. Trong số các kiến trúc trên, kiến trúc với 3 lớp nhân chập cho độ chính xác cao nhất. Vì vậy, kiến trúc này được sử dụng cho việc ước lượng giá trị lượng tử.

Trong mô hình đề xuất, sai số tuyệt đối trung bình (MAE) được sử dụng để đo độ chính xác. MAE được tính sử dụng công thức sau:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

trong đó \hat{y} là giá trị QP dự đoán, y là giá trị QP tham chiếu của mỗi mảnh hình ảnh, n là số lượng các khối ảnh dùng để huấn luyện.

tái tạo lại

- Nhóm các khung hình theo điểm số VMAF (mỗi nhóm tương ứng với một dải con của điểm số VMAF, như nêu trong Mục III.A).
- Trong mỗi nhóm khung hình, bản đồ QP của khung hình nào có giá trị J_{cost} nhỏ nhất được coi là bản đồ QP chân lý nền tảng cho mức chất lượng đó. Giá trị J_{cost} được tính dưới dạng hàm sau:

$$J_{cost} = \frac{100}{VMAF} + \lambda \cdot R \quad (7)$$

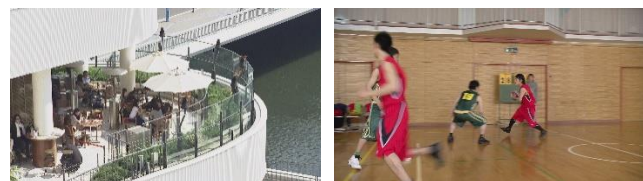
trong đó R là tốc độ bit của khung hình và λ là hệ số nhân Lagrange.

Sau khi có được bản đồ QP cơ bản cho từng nhóm, các khối hình của khung hình trong nhóm được sử dụng làm đầu vào cho mô hình CNN đề xuất. Nhãn của mỗi khối hình ảnh là giá trị QP có cùng vị trí với khối hình ảnh trong bản đồ QP. Để có được một mô hình CNN, 15 chuỗi video được sử dụng. Độ phân giải của mỗi khung hình là 355×288 và chiều dài là 50 khung hình. Sau khi huấn luyện với 100 epochs, MAE của mô hình là 2.11.

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 *BasketballDrive 1920x1080*, *BQTerrace 1920x1080*, *Kimono 1280x720* và *Cactus 1280x720* được sử dụng. Độ dài của mỗi chuỗi là 50 khung hình.



a. BQTerrace

b. BasketballDrive



c. Cactus

d. Kimono

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

V. KẾT LUẬN

Bài báo đã đề xuất một phương pháp dự đoán giá trị QP cho các khối ảnh trong khung hình để đảm bảo video có mức chất lượng như mong muốn tại đầu ra phía giải mã. Cụ thể, phương pháp đã cải tiến phương pháp RDO của H.264 và sử dụng mô hình mạng nơ-ron tích chập 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 bằng với giá trị mong muốn. Ngoài ra, hiệu năng của phương pháp đề xuất cũng cao hơn chuẩn mã hóa H.264 xét theo giá trị BD-VMAF và BD-Rate. Cụ thể, giá trị BD-VMAF của phương pháp đề xuất

Bảng 1. So sánh BDRate và BD-VMAF giữa H.264 và phương pháp đề xuất

Chuỗi Video		H.264		Phương pháp đề xuất		BDBR	BD-VMAF
		Bitrate	VMAF	Bitrate	VMAF		
BasketballDrive	crf 27	3510.08	99.00	2363.26	92.49	-25.98	16.69
	crf 29	2775.24	92.00	2107.84	88.25		
	crf 32	2001.74	71.00	1461.19	72.07		
	crf 37	1221.73	49.00	968.44	56.35		
BQTerrace	crf 32	1425.42	100.00	1008.06	100.00	-17.67	7.11
	crf 37	734.69	89.00	692.22	89.49		
	crf 42	422.18	68.00	320.24	73.94		
	crf 45	309.33	55.00	228.55	67.05		
Cactus	crf 29	2774.12	100.00	2558.08	100.00	-16.62	5.51
	crf 32	1909.05	93.00	1458.45	93.20		
	crf 37	1076.84	74.00	944.79	73.00		
	crf 42	625.63	52.00	432.76	50.79		
Kimono	crf 27	4573.45	96.00	2940.56	97.00	-23.25	7.41
	crf 29	3562.52	90.00	2634.68	92.00		
	crf 32	2466.17	79.00	1922.71	77.00		
	crf 37	1357.00	57.00	1195.70	55.00		
Trung bình		1921.57	79.00	1452.35	79.85	-20.88	9.18

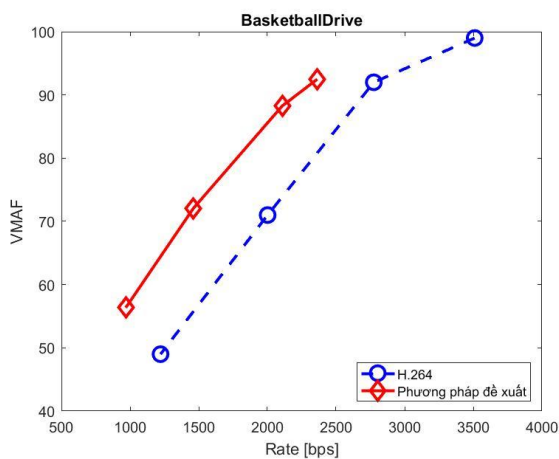
Hình 4 mô tả khung hình đầu tiên của các chuỗi video thử nghiệm. Trong bài báo này, bộ mã hóa video x.264 được sử dụng để mã hóa các chuỗi video. Kết quả của phương pháp đề xuất được so sánh với kết quả của bộ mã hóa x.264 trong trường hợp tốc độ bit không đổi.

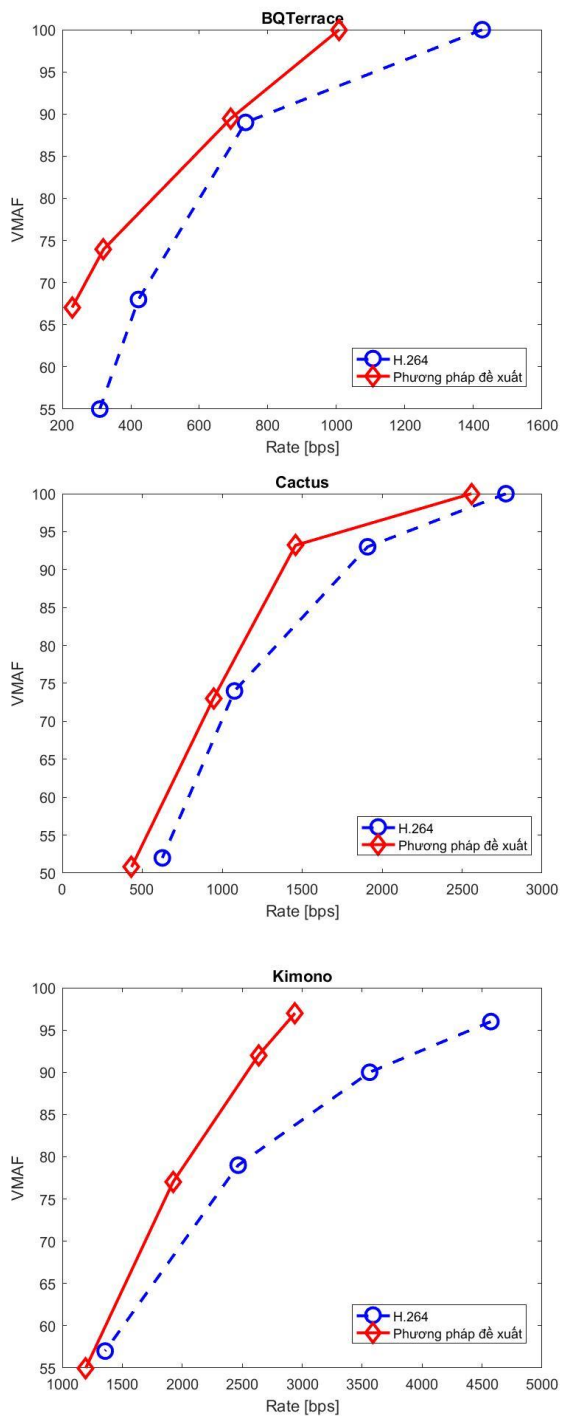
lớn hơn H.264 là 9.18 và BD-Rate của phương pháp đề xuất thấp hơn của H.264 là 20.88%.

B. Kết quả mô phỏng

Hình 5 và Bảng 1 mô tả kết quả so sánh giữa phương pháp đề xuất và chuẩn mã hóa video H.264. Kết quả cho thấy giá trị VMAF của phương pháp đề xuất đạt được xấp xỉ giá trị VMAF mong muốn. Cụ thể, Bảng 1 cho thấy các chuỗi hầu hết có giá trị VMAF bằng hoặc lớn hơn giá trị mong muốn trong trường hợp crf 28, 29. Trong khi đó, tại các giá trị crf 37, 42, giá trị VMAF nhỏ hơn giá trị mong muốn nhưng vẫn cùng mức chất lượng. Giá trị trung bình của VMAF của 4 chuỗi cho thấy phương pháp đề xuất đạt được mức chất lượng tương đương với mức chất lượng mong muốn trong khi tốc độ bit trung bình thấp hơn của H.264 là 4.6%.

Hình 5 cho thấy phương pháp đề xuất có hiệu năng tốt hơn chuẩn H.264 trên khía cạnh BD-Rate và BD-VMAF. Cụ thể, giá trị BD-Rate của phương pháp đề xuất giảm 20.88% so với BD-Rate của H.264. Ngoài ra, giá trị BD-VMAF của phương pháp đề xuất lớn hơn 9.18 so với BD-VMAF của H.264.





Hình 5. So sánh hiệu năng giữa phương pháp đề xuất và chuẩn H.264

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Học Viện Công nghệ Bưu chính Viễn thông trong đề tài mã số 01-HV-KTĐT1.

TÀI LIỆU THAM KHẢO

[1] K. Brunström *et al.*, “Qualinet White Paper on Definitions of Quality of Experience Qualinet White Paper on Definitions of Quality of Experience Output from the fifth Qualinet meeting, Novi Sad,” *Eur. Netw. Qual. Exp. Multimed. Syst. Serv. (COST Action IC 1003)*, no. March, p. 26, 2013.

[2] Q. Cai, Z. Chen, D. O. Wu, and B. Huang, “Real-Time Constant Objective Quality Video Coding Strategy in High Efficiency Video Coding,” *IEEE Trans. Circuits Syst.*

Video Technol., vol. 30, no. 7, pp. 2215–2228, 2020, doi: 10.1109/TCSVT.2019.2914100.

[3] C. W. Seo, J. H. Moon, and J. K. Han, “Rate control for consistent objective quality in high efficiency video coding,” *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2442–2454, 2013, doi: 10.1109/TIP.2013.2251647.

[4] H. Avc, I. Applications, C.-Y. Wu, and P. Su, “A Content-Adaptive Distortion – Quantization Model,” *IEEE Trans. Circuits Syst.*, vol. 24, no. 1, pp. 113–126, 2014, [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6562767>

[5] 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.

[6] D. K. Kwon, M. Y. Shen, and C. C. Jay Kuo, “Rate control for H.264 video with enhanced rate and distortion models,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 5, pp. 517–528, 2007, doi: 10.1109/TCSVT.2007.894053.

[7] T. T. Pham, X. Van Hoang, N. T. Nguyen, D. T. Dinh, and L. T. Ha, “End-to-End Image Patch Quality Assessment for Image/Video with Compression Artifacts,” *IEEE Access*, vol. 8, pp. 215157–215172, 2020, doi: 10.1109/ACCESS.2020.3040416.

[8] Z. Li, A. Aaron, A. Katsavounidis, Ioannis Moorthy, and M. Manohara, “Toward A Practical Perceptual Video Quality Metric,” *Netflix Blog*, 2016. <http://techblog.netflix.com/2016/06/toward-practical-perceptual-video.html>

CONTENT-BASED QUANTIZATION PREDICTION FOR CONSISTENT VIDEO QUALITY

Abstract: With the explosion of online video viewing services on telecommunications networks such as Youtube, Netflix, etc., while the network infrastructure is limited, the image quality for these services is an urgent requirement. Currently, there are many parameters to evaluate video image quality. The parameter measuring the frequency of change in image quality level is one of the important parameters. For video viewers, the high frequency change in image quality will cause annoyance. This can happen often when the network quality is unstable. Therefore, in this paper, a method to ensure consistent video quality based on quantization parameter prediction technique is proposed. Specifically, based on the content of each video frame, a machine learning model is proposed to predict a corresponding quantization level for each frame in the frame to achieve the expected video quality level. The simulation results show that the proposed method has achieved two criteria simultaneously. Firstly, the proposed method has achieved the expected quality level for the reconstructed video. Secondly, the compression performance of the video in the proposed method is increased compared to the standard x.264 video encoder.

Keyword: VMAF, Video coding, x.264.



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.



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.