PHƯƠNG PHÁP ĐIỀU CHỈNH HỆ SỐ LƯỢNG TỬ THEO NỘI DUNG CỤC BỘ KHUNG HÌNH VIDEO

Phạm Thanh Tùng[*], Vũ Hữu Tiến[**], Đinh Triều Dương[**], Lê Thanh Hà[***]
[*] Trường Đại học Phòng cháy chữa cháy
[**]Học viện Công nghệ Bưu chính Viễn thông
[***]Trường Đại học Công nghệ, VNU

Tóm tắt: Nội dung bài báo trình bày phương pháp điều chỉnh hệ số lượng tử nén video theo nội dung cục bộ trên từng khung hình. Phương pháp được phát triển dựa trên mô hình dự đoán chất lượng hình ảnh khi thay đổi hệ số lượng tử trích chọn đặc trưng nội dung bằng mạng Neural tích chập. Dữ liệu dùng để xây dựng mô hình dự đoán gồm hơn 40,000 điểm đánh giá chất lượng chủ quan trung bình (MOS) cục bộ theo khối ảnh. Phương pháp đề xuất áp dụng lựa chọn hệ số lượng tử (QP) trên bộ nén H.265/HEVC nhằm giảm tốc độ bit của video nén nhưng vẫn đảm bảo chất lượng chất lượng thị giác. Kết quả thực nghiệm cho thấy bộ nén cải tiến trên phiên bản HM 16.0 có hiệu năng cao hơn theo thang đo SSIM và thang đo chủ quan MOS so với bộ nén gốc. So sánh với các nghiên cứu có liên quan, bộ nén áp dụng phương pháp đề xuất cũng cho kết quả khả quan.

Từ khóa: Điều chỉnh hệ số lượng tử, bộ nén H.265/HEV, tăng hiệu năng nén.

I. GIỚI THIỆU

Hiện nay, bộ nén video hiệu suất cao (High Efficiency Video Coding - HEVC) đã được chấp nhận rộng rãi để đạt được hiệu năng nén tốt hơn so với H.264/AVC với cùng mức chất lượng hình ảnh [1]. Bộ nén có nhiều cấu hình nén khác nhau không chỉ áp dụng cho độ phân giải cao (FHD) mà còn cho 4K/8K ultra-HD (UHD) [2,3]. Hiệu năng nén của một bộ nén video là khả năng cân bằng giữa chất lượng video sau khi nén, tốc độ bit (bitrate) và chi phí tính toán. Các nghiên cứu nâng cao hiệu năng nén trên HEVC hiện nay tập trung vào bốn hướng chính gồm: (i) cải thiện chất lượng hình ảnh; (ii) giảm độ phức tạp tính toán; (iii) giảm tốc độ bit và (iv) phát triển các chuẩn nén thế hệ tiếp theo.

HEVC sử dụng cấu trúc phân vùng đơn vị mã hóa quadtree (CU) linh hoạt trong đó kích thước của mỗi CU có thể thay đổi từ $8x8$ đến $64x64$ pixel. Mỗi đơn vị cây mã hóa cấu trúc (CTU) sẽ được chia đệ quy thành CU theo kiểu quadtree trong quá trình tối ưu vòng lặp (RDO) để quyết định chế độ và cách phân vùng nén cuối cùng. Quá trình lượng tử được coi như là việc chia các hệ số DCT của khung hình cho hệ số lượng tử tương ứng, kết quả này sau đó sẽ được làm tròn xuống số nguyên gần nhất. Lựa chọn hệ số lượng tử (QP) là một giải pháp quan trọng để nâng cao hiệu năng nén video của từng đơn vị mã hóa cơ bản bằng cách chọn chế độ mã hóa tốt nhất dưới một giá trị QP nhất định, trong đó đơn vị cơ bản có thể là khung, lát (slice) hoặc đơn vị mã hóa (CU). Quá trình lượng tử hóa lượng tử hóa làm biến dạng biến dạng video ban đầu và chất lượng video tái tạo bị suy giảm. Do đó, việc nghiên cứu các thuật toán lượng tử hóa hiệu quả rất quan trọng để cải thiện hiệu năng mã hóa video.

Lựa chọn QP trong nén video có thể được mô tả như một bài toán RDO [4, 5] để tối thiểu hóa biến dạng D ở một tốc độ bit nhất định R_T như sau:

$$QP^* = (QP_i^*, ...QP_n^*) = \arg\min(QP)\sum_{i=1}^{N} D_i \,, \qquad (1)$$

$$s.t. \sum_{i=1}^{N} R_i \le R_T$$

trong đó N là số lượng CU, D_i và R_i tương ứng là biến dạng nén và tốc độ bit của CU thứ i. QP_i là trị số QP cho CU thứ i và $QP* = (QPi*, \cdots, QPN*)$ biểu diễn tập hợp các giá trị QP tùy chọn của tất cả các CU. Áp dụng phương pháp λ [6] đưa bài toán thành dạng tối ưu không điều kiện sau đây, phương trình (1) viết lại thành:

$$QP^* = \arg\min(QP)\{J\}\,, \qquad (2)$$

$$J = \sum_{i=1}^{N} (D_i + \lambda R_i)$$

với J là tổng hàm giá giữa biến dạng – tốc độ, và λ là tham số hiệu chỉnh giữa D_i và R_i. Trong quá trình tối ưu RDO, hệ số λ trong HEVC được chọn là:

$$\lambda = QP_{factor} \cdot 2^{QP/3}\,, \qquad (3)$$

trong đó QP là hệ số lượng tử, QP_{factor} là hằng số tùy theo cấu hình nén. Giá trị QP trong công thức (3) là một số nguyên thể hiện bước lượng tử hoá. Lựa chọn QP hiện tại trong HEVC như trên không phải là tối ưu do bước lượng tử hóa được giữ cố định để giảm độ phức tạp trong quy trình RDO. Việc áp dụng lược đồ QP cố định hoặc xác định trước có thể khiến tốc độ nén giảm đáng kể, trong khi chất

Tác giả liên hệ: Phạm Thanh Tùng,
Email: tungpt@daihocpccc.edu.vn
Đến tòa soạn: 9/2022, chỉnh sửa: 10/2022, chấp nhận đăng: 10/2022.

lượng thị giác do biến dạng nén phụ thuộc nội dung của các CU. Điều này trở thành một vấn đề cần giải quyết để phát triển phương pháp QP trong HEVC. Nhiều phương pháp điều chỉnh QP đã được nghiên cứu để đạt được hiệu năng mã hóa tốt như sử dụng mối quan hệ QP – λ để xác định giá trị λ theo QP ban đầu và hiệu chỉnh lại [7, 8]. Wang và cộng sự. [9] công bố phương pháp điều chỉnh QP có tính đến thông tin khối đã được mã hóa trước đó. Zhao và cộng sự. [10] đã đề xuất một sơ đồ cấu trúc QP thiết lập giá trị cho các lớp khung hình kế tiếp có thứ bậc khác nhau. Yeo và cộng sự. [11] cũng giới thiệu một thuật toán lựa chọn QP thích ứng theo khối. Thuật toán này dựa trên thuộc tính điểm ảnh theo không gian và thời gian của các khối CU nên cần thời gian mã hóa lâu hơn. Xiang và cộng sự. [12] đề xuất một phương pháp ước tính cảm nhận dịch chuyển thị giác bằng cách sử dụng mô hình không thời gian của biến dạng có thể nhận biết (JND) cho thiết kế bù QP.

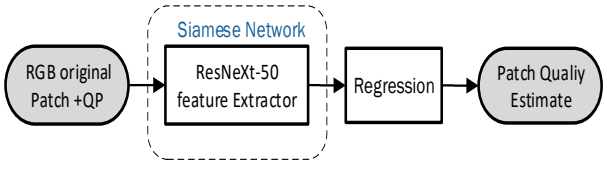
Một số nghiên cứu gần đây ứng dụng học sâu (DNN) cho hiệu quả khả quan đối với nén video [13-17]. Nổi bật trong đó là nghiên cứu sử dụng mô hình DNN trên HEVC điều chỉnh QP để giảm tốc độ bit trong khi vẫn duy trì chất lượng hình ảnh chủ quan [17]. Trong nghiên cứu trước đây [18] chúng tôi đã đề xuất mô hình học sâu sử dụng mạng Neural tích chập (CNN) dự đoán chất lượng hình ảnh được phát triển từ bộ dữ liệu thực nghiệm công bố trong [19]. Trên cơ sở mô hình này, chúng tôi tiếp tục ứng dụng nhằm xây dựng phương pháp điều chỉnh QP cục bộ nhằm giảm tốc độ bit nhưng vẫn đảm bảo chất lượng thị giác. Đồng thời, việc lựa chọn QP bằng thuật toán đề xuất không qua RDO giúp giảm thời gian tính toán.

Phần còn lại của bài báo được tổ chức như sau: phần II đề xuất phương pháp điều chỉnh hệ số lượng theo nội dung cục bộ khung hình video. Phần III cài đặt thử nghiệm và đánh giá hiệu năng phương pháp đề xuất cài đặt trên bộ nén H.265/HEVC khi nén một số mẫu video so với bộ nén chuẩn cũng như các nghiên cứu liên quan. Cuối cùng, chúng tôi kết luận bài báo trong phần IV.

II. ĐỀ XUẤT PHƯƠNG PHÁP ĐIỀU CHỈNH HỆ SỐ LƯỢNG THEO NỘI DUNG CỤC BỘ KHUNG HÌNH VIDEO

A. Mô hình dự đoán chất lượng hình ảnh khi nén

Trong nghiên cứu trước đây [18], chúng tôi đã xây dựng mô hình CNN dự đoán chất lượng ảnh chủ quan khi thay đổi QP theo nội dung cục bộ. Kiến trúc mô hình dự đoán biểu diễn như Hình 1. Dữ liệu đầu vào gồm 4 lớp có kích thước $128x128$ trong đó gồm ba lớp thành phần màu khối ảnh cần dự đoán chất lượng và lớp còn lại là lớp hệ lượng tử dùng để nén khối. Mạng ResNeXt-50 dùng để trích chọn đặc trưng thị giác theo nội dung khối ảnh. Vector đặc trưng kết quả được dùng làm đầu vào cho lớp hồi quy dự đoán chất lượng.



Hình 1. Kiến trúc mô hình dự đoán

Bộ dữ liệu [19] dùng để huấn luyện mô hình CNN được chúng tôi đề xuất trong [18] gồm 40286 mẫu. Mỗi mẫu trong bộ dữ liệu gồm 1 khối ảnh gốc kích thước $128x\ 128$, hệ số lượng tử QP dùng để nén khối gốc và kết quả chất lượng chủ quan DMOS. Bộ dữ liệu được chia thành 10 phần theo tỷ lệ 8:1:1 (tương ứng 32000: 4000: 4286 mẫu) dành lần lượt cho việc huấn luyện, hiệu chỉnh và đánh giá mô hình dự đoán chất lượng. Kết quả kiểm tra trên tập dữ liệu đánh giá cho kết quả tương quan tuyến tính Pearson (PLCC) giữa kết quả dữ đoán với kết quả thực nghiệm bằng 0.9505 và hệ số tương quan xếp hạng Spearman (SRCC) bằng 0.9088. So sánh với kết quả tương quan giữa hệ số lượng tử và chất lượng lần lượt $PLCC = -0{,}807$ và $SRCC = -0{,}8438$ cho thấy ảnh hưởng quan trọng của nội dung khối đến chất lượng hình ảnh biến dạng của nó khi nén.

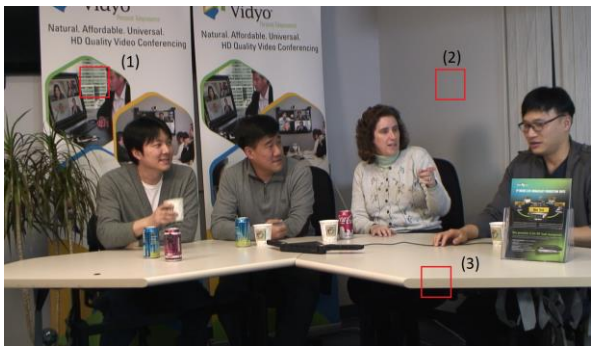
B. Đề xuất phương pháp điều chỉnh hệ số lượng tử

Trong nén video, có hai cách thiết lập QP chính: đồng nhất trên toàn bộ các khung hình video và tùy biến theo các đơn vị mã hóa cơ bản. Ở thiết lập thứ nhất, tất cả các khối trên tất cả các khung hình đều nén với cùng một hệ số lượng tử nên tốc độ nén nhanh hơn, nhưng có nhược điểm là chất lượng sau khi nén không đồng đều, dễ bị phát hiện biến dạng. Ở thiết lập thứ hai, bộ nén đặt ra giá trị QP ban đầu qp và ngưỡng điều chỉnh qp_{factor}. Khi đó, hệ số QP cho các đơn vị mã hóa được lựa chọn tối ưu theo RDO trong khoảng $[qp - qp_{factor}, qp + qp_{factor}]$. Do đó, trong nghiên cứu này chúng tôi đề xuất phương pháp điều chỉnh tăng QP trong khoảng $[qp, qp + qp_{factor}]$. Mục tiêu chính của phương pháp nhằm giảm tốc độ bit nhưng vẫn đảm bảo chất lượng theo cảm nhận thị giác. Một khối ảnh tham chiếu CU_{org} trong khung hình được nén bởi hệ số lượng tử qp có chất lượng dự đoán theo nhận thức thị giác là:

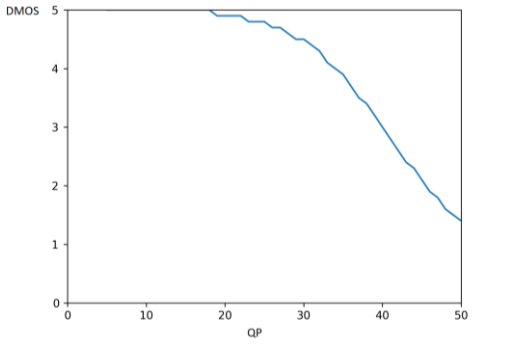
$$y = QIQA\left(CU_{org}, qp\right), \quad (4)$$

trong đó $QIQA$ là mô hình CNN dự đoán phương pháp đánh giá chất lượng hình ảnh trình bày bên trên.

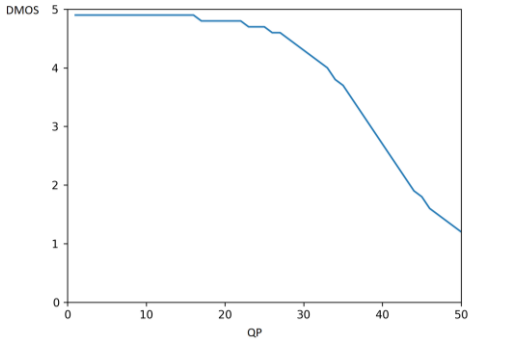
Các nghiên cứu theo [20] đã chỉ ra rằng với một ngưỡng thay đổi chất lượng nhất định, mắt người không nhận biết được sự thay đổi này. Đồng thời, tùy theo nội dung của khối ảnh, chất lượng hình ảnh suy giảm không giống nhau khi tăng hệ số lượng tử như kết quả nghiên cứu trong [17]. Áp dụng mô hình QIQA dự đoán chất lượng khung hình thứ nhất của video "FourPeople" đối với ba khối có nội dung khác nhau gồm: khối nhiều chi tiết (1), khối đồng nhất (2) và khối có cạnh tương phản cao (3). Kết quả cho thấy chất lượng thị giác khối (1) giảm chậm nhất trong khoảng lượng tử [25,40] tiếp theo là khối (2) trong khi chất lượng khối (3) suy giảm rất nhanh. Tốc độ suy giảm chất lượng thị giác trong khoảng [1,19] các khối có nhiều chi tiết gần như không đáng kể.



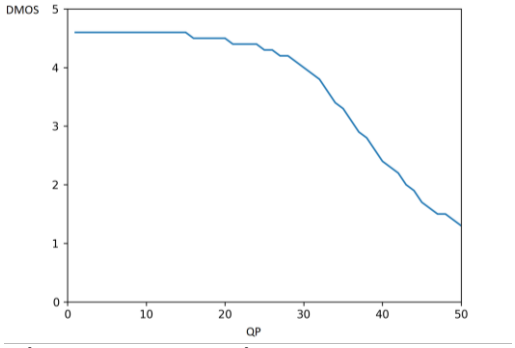
a. Khung hình đầu tiên video "FourPeople"



b. Chất lượng dự đoán khối nhiều chi tiết (1)



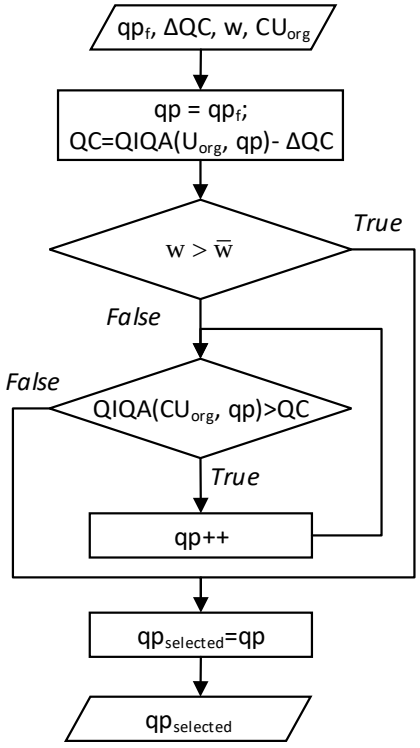
c. Chất lượng dự đoán khối đồng nhất (2)



d. Chất lượng dự đoán khối có cạnh tương phản cao (3)

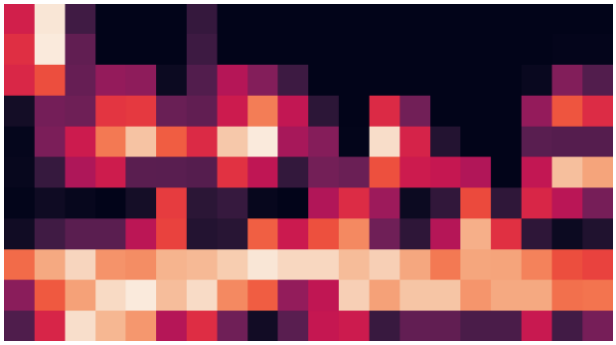
Hình 2. Chất lượng hình ảnh dự đoán một số khối khung hình đầu tiên video "FourPeople"

Dựa trên nhận định trên, chúng tôi đề xuất thuật toán nhằm lựa chọn hệ số lượng tử nhằm giảm tốc độ bit nhưng vẫn đảm bảo chất lượng thị giác như Hình 3. Thuật toán xác định hệ số lượng tử cho các khối của một khung hình video (thuật toán APQP) biểu diễn như Hình 3. Dữ liệu đầu vào của thuật toán bao gồm khối ảnh gốc CU_{org}, hệ số lượng tử qp và chất lượng điều chỉnh chấp nhận được ΔQC và trọng số chú ý w. Thuật toán đề xuất trong nghiên cứu tăng hệ số lượng tử cho đến khi chất lượng dự đoán còn nhỏ hơn được bằng chất lượng ban đầu cộng thêm ΔQC. Việc thay đổi hệ số lượng tử chỉ áp dụng đối với các khối có trọng số chú ý w xác định theo [15] thấp hơn trọng số trung bình của toàn khung hình. Kết quả của thuật toán là hệ số lượng tử ước lượng $qp_{selected}$ để chất lượng thị giác thay đổi trong giới hạn. Hình 4 cho thấy kết quả áp dụng thuật toán điều chỉnh hệ số lượng tử đối với khung hình đầu tiên video "FourPeople". Với hệ số lượng tử thiết lập ban đầu $qp = 27$, trọng số chú ý theo khối như Hình 4.b, các khối quan trọng thu hút chú ý giữ nguyên $qp = 27$, trong khi các khối còn lại qp điều chỉnh tăng nhằm giảm tốc độ bit cần thiết để nén khung hình. Kết quả dự đoán cho các khối kích thước 64×64 biểu diễn như Hình 4.c. Để sử dụng hệ số lượng tử cho các khối hình có kích thước nhỏ hơn, thuật toán sử dụng phương pháp nội suy. Ma trận lượng tử kết quả thu được biểu diễn trong Hình 4.d đã được nội suy từ kích thước khối 64×64 Hình 4.c thành 32×32.

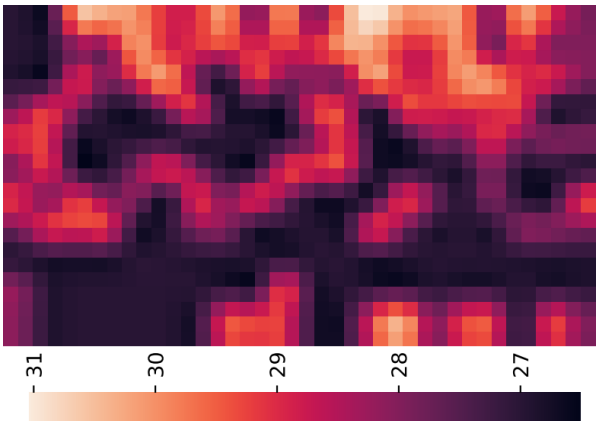


Hình 3. Sơ đồ thuật toán APQP

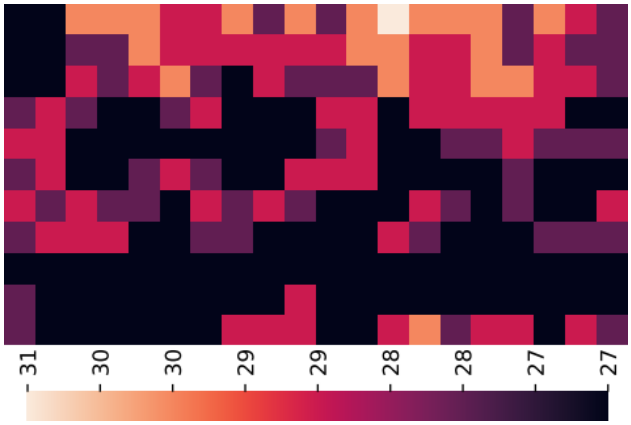
a. Phân vùng sự chú ý



b. Trọng số chú ý theo khối



c. Hệ số lượng tử điều chỉnh khối 64×64

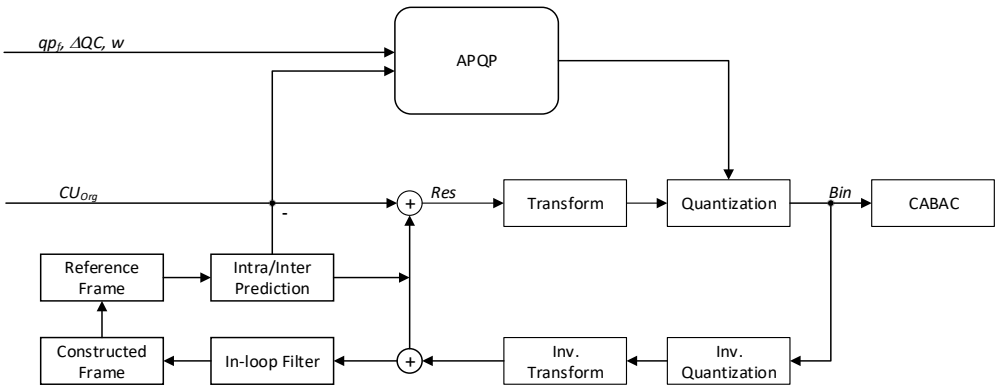


d. Hệ số lượng tử điều chỉnh khối 32×32

Hình 4. Hệ số lượng tử điều chỉnh trên khung hình đầu tiên video "FourPeople" với qp_f*=27*

III. ĐÁNH GIÁ HIỆU NĂNG PHƯƠNG PHÁP ĐỀ XUẤT

Để đánh giá hiệu quả của phương pháp lựa chọn hệ số lượng tử nhằm giảm tốc độ bit, chúng tôi cài đặt thuật toán của phương pháp vào bộ nén H.265/HEVC phiên bản HM 16.0 theo sơ đồ như Hình 5. Trong đó, mỗi khung hình gốc phân chia thành các khối kích thước 64×64 để lựa chọn hệ số lượng tử theo APQP. Đồng thời, khung hình gốc cũng được sử dụng để xác định trọng số thu hút sự chú ý w. Kết quả thu được là một ma trận hệ số lượng tử được xác định tương ứng với các khối của khung hình làm cơ sở cho quá trình lượng tử hoá.



Hình 5. Cài đặt trên bộ nén H.265/HEVC

Bộ nén có điều chỉnh trong nghiên cứu sử dụng thiết lập nén liên khung RA và LDP. Các mẫu video thử nghiệm để đánh giá hiệu năng gồm 2 video độ phân giải 2K (thuộc nhóm A), 5 video độ phân giải FHD (nhóm B), 4 video độ phân giải 832×480 (nhóm C), 4 video độ phân giải 416×240 (nhóm D) và 3 video độ phân giải HD (nhóm E). Mỗi video nén dưới 4 mức QP thông thường lần lượt là: $22, 27, 32$ và 37; giá trị ngưỡng chất lượng điều chỉnh $\Delta\, QC = 0.5$. Tương tự, bộ nén chuẩn HM 16.0 thực hiện nén liên khung các mẫu video thử nghiệm với cùng hệ số lượng tử như bộ nén đề xuất của nghiên cứu. Ngoài ra, để làm rõ hiệu năng nén theo nhận thức tri giác là mục tiêu chính của nghiên cứu, chúng tôi đã tiến hành thực nghiệm đánh giá chất lượng hình ảnh chủ quan các video tái tạo sau khi nén. Trong thực nghiệm, 20 người quan sát được yêu cầu đánh giá chất lượng video tái tạo theo hai bộ nén bằng thang đo MOS từ 1 đến 5 tương ứng với chất lượng tăng dần. Chất lượng chủ quan được tính bằng điểm trung bình của tất cả các người đánh giá.

Kết quả so sánh hiệu năng theo thang đo SSIM và đánh giá chủ quan giữa bộ nén đề xuất so với bộ nén gốc HM 16.0 trình bày trong Bảng 1. Trong đó, tốc độ bit giảm (Δ Birate) trung bình trên tất cả các mức QP đã nén là 18.83% tuy nhiên chất lượng cũng bị giảm theo. Để làm rõ hiệu năng của bộ nén đề xuất, chúng tôi sử dụng phương pháp tính BD-BR SSIM và BD-BR MOS theo Bjøntegaard [21]. Bảng kết quả cho thấy bộ nén đề xuất cho hiệu năng theo thang đo SSIM cơ bản tốt hơn so với bộ nén gốc trung bình

5.59% đối với cấu hình RA và 7.17% đối với cấu hình LDP. Theo thực nghiệm đo chất lượng chủ quan, bộ nén cải tiến cũng có hiệu suất tốt hơn cho các mẫu video với độ phân giải cao hơn (nhóm A và nhóm B). Đối với các video thử nghiệm có nhiều vùng phức tạp, nhiều chi tiết, thuật toán được đề xuất cho hiệu quả nén tốt hơn. Các đặc điểm hình ảnh của video thử nghiệm như vậy có thể thấy trong video 'ParkScene, 'Traffic', 'BQTerrace', 'PeopleOnStreet', 'BQSquare' trong đó tăng chất lượng đáng kể nhất về mặt cảm nhận. Mặt khác, phương pháp đề xuất hiệu quả kém trên 'KristenAndSara' và 'BasketballDrive' có nhiều đồng nhất và độ tương phản cao. Ví dụ trong Hình 6 cho thấy chất lượng hình ảnh khung hình đầu tiên video 'FourPeople' theo bộ nén đề xuất không có sự khác biệt về mặt thị giác với tốc độ bit thấp hơn trong khi hiệu năng nén đề xuất tốt hơn (Hình 7). Hiệu năng theo thang đo chất lượng chủ quan MOS cũng thể hiện rõ nét hơn so với thang đo SSIM.

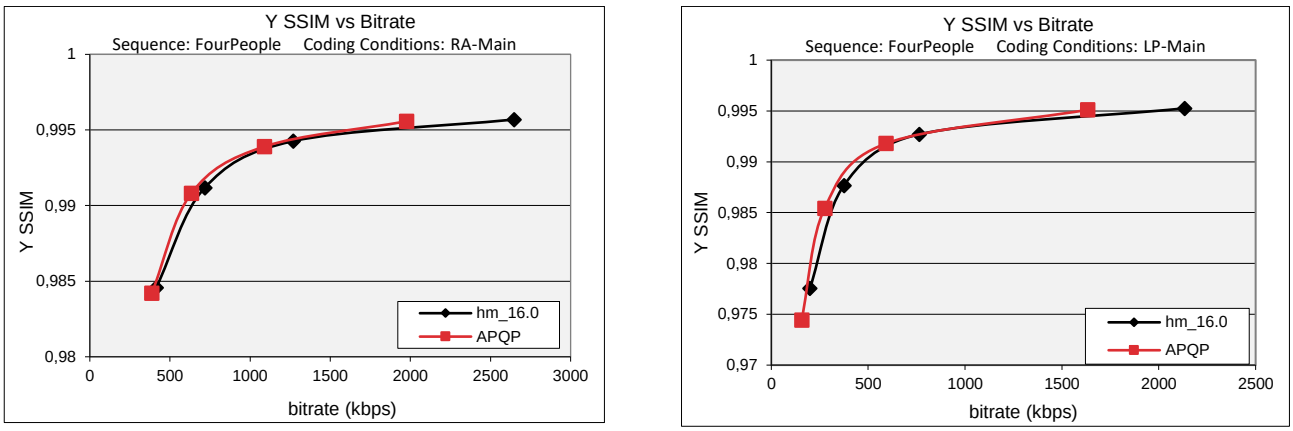
Để làm rõ thêm hiệu quả nén của phương pháp đề xuất, chúng tôi tiến hành so sánh hiệu năng với 3 nghiên cứu liên quan [11, 12, 17]. Kết quả theo Bảng 2 cho thấy hiệu năng nén trung bình với phương pháp đề xuất tốt hơn so với [11, 12] nhưng kém hơn so với [17]. Tuy nhiên, phương pháp đề xuất cho hiệu quả hơn đối với các video có độ phân giải cao (nhóm A) và các video có tỷ lệ khối nhiều chi tiết chiếm đa số.

Bảng 1. So sánh hiệu năng của bộ nén đề xuất so với bộ nén gốc

Nhóm	Video mẫu	Thiết lập nén RA			Thiết lập nén LDP		
		Δ Birate	BD-BR SSIM	BD-BR MOS	Δ Birate	BD-BR SSIM	BD-BR MOS
Class A	Traffic	-21.73%	-12.05%	-13.56%	-25.17%	-14.11%	-15.12%
	PeopleOnStreet	-26.34%	-14.35%	-14.62%	-30.29%	-12.93%	-16.12%
Class B	Kimono	-22.63%	-13.46%	-10.03%	-25.87%	-13.80%	-16.00%
	ParkScene	-24.32%	-17.75%	-15.62%	-25.38%	-17.77%	-17.22%
	Cactus	-15.04%	-4.31%	-2.56%	-17.05%	-3.64%	-3.43%
	BasketballDrive	-8.19%	3.90%	-2.40%	-10.33%	1.90%	-3.99%
	BQTerrace	-18.22%	-13.06%	-7.16%	-21.42%	-13.90%	-8.21%
Class C	BasketballDrill	-11.97%	-1.60%	-3.29%	-10.98%	-1.82%	-4.82%
	BQMall	-13.55%	-3.09%	-4.18%	-16.55%	-4.23%	-5.21%
	PartyScene	-24.73%	-6.63%	-6.71%	-24.17%	-5.89%	-5.81%
	RaceHorses	-21.43%	-8.34%	-10.02%	-25.17%	-11.03%	-10.21%
Class D	BasketballPass	-9.17%	-0.46%	-2.33%	-28.72%	0.90%	-4.12%
	BQSquare	-20.38%	-7.93%	-9.67%	-22.63%	-9.49%	-10.45%
	BlowingBubbles	-12.49%	-2.28%	-4.52%	-15.08%	-4.33%	-6.03%
	RaceHorses	-21.37%	-10.77%	-8.11%	-23.15%	-14.01%	-9.24%
Class E	FourPeople	-19.08%	-6.45%	-4.78%	-23.50%	-8.05%	-5.37%
	Johnny	-10.52%	0.25%	-0.90%	-12.88%	-0.04%	-1.16%
	KristenAndSara	-8.25%	1.68%	-2.52%	-10.17%	0.97%	-3.44%
	Class A	-24.04%	-13.20%	-14.09%	-27.73%	-13.52%	-15.62%
	Class B	-17.68%	-8.94%	-7.55%	-20.01%	-9.44%	-9.77%
	Class C	-17.92%	-4.92%	-6.05%	-19.22%	-5.74%	-6.51%
	Class D	-15.85%	-5.36%	-6.16%	-22.40%	-6.73%	-7.46%
	Class E	-12.62%	-1.51%	-2.73%	-15.52%	-2.37%	-3.32%
	All	**-17.19%**	**-6.48%**	**-6.83%**	**-20.47%**	**-7.29%**	**-8.11%**

a. Khung hình nén với qp_f=27 *b. Khung hình sau khi nén với bảng lượng tử điều chỉnh*

Hình 6. Khung hình đầu tiên video "FourPeople" nén với chất lượng thị giác không đổi



a. Thiết lập nén RA *b. Thiết lập nén LDP*

Hình 7. Đánh giá hiệu năng bộ nén đề xuất so với bộ nén HEVC

Bảng 2. So sánh hiệu năng của bộ nén đề xuất với các nghiên cứu liên quan

Nhóm	Video mẫu	Xiang [12]	Yeo [11]	Marzuki [17]	Đề xuất
Class A	Traffic	-4.32%	4.10%	-10.30%	**-12.05%**
	PeopleOnStreet	-6.51%	-7.46%	-13.12%	**-14.35%**
Class B	Kimono	-0.56%	-0.46%	-12.91%	**-13.46%**
	ParkScene	-5.38%	-5.51%	-17.00%	**-17.75%**
	Cactus	-1.41%	-2.77%	**-17.98%**	-4.31%
	BasketballDrive	-5.65%	-5.51%	**-15.13%**	3.90%
	BQTerrace	-2.07%	-6.64%	**-21.64%**	-13.06%
Class C	BasketballDrill	-11.78%	-8.95%	**-16.09%**	-1.60%
	BQMall	-2.37%	-3.46%	**-18.54%**	-3.09%
	PartyScene	-4.33%	-2.82%	**-19.26%**	-6.63%
	RaceHorses	-1.68%	-0.35%	**-16.58%**	-8.34%
Class D	BasketballPass	**-6.29%**	-5.63%	-4.74%	-0.46%
	BQSquare	-2.50%	-0.68%	**-16.24%**	-7.93%
	BlowingBubbles	-7.05%	-8.32%	**-17.15%**	-2.28%
	RaceHorses	-2.52%	-3.26%	-7.82%	**-10.77%**
	Class A	-5.42%	-1.68%	-11.71%	**-13.20%**
	Class B	-3.01%	-4.18%	**-16.93%**	-8.94%
	Class C	-5.04%	-3.90%	**-17.62%**	-4.92%
	Class D	-4.59%	-4.47%	**-11.49%**	-5.36%
	All	-4.51%	-3.56%	**-14.44%**	**-8.10%**

IV. KẾT LUẬN

Trong bài báo này, chúng tôi đề xuất phương pháp giảm tốc độ bit bằng cách điều chỉnh QP theo đặc trưng thị giác. Phương pháp áp dụng cho đơn vị mã hóa khối của khung hình video. Phương pháp đề xuất sử dụng một mô hình trên nền tảng mạng CNN Resnext-50 để trích xuất các đặc trưng dự đoán chất lượng hình ảnh sau nén của khối dưới một mức lượng tử xác định. Kết quả áp dụng với thiết lập nén liên khung RA và LDP trên H.265/HEVC cho thấy hiệu năng tăng 6.89% theo thang đo SSIM và 7.47 % theo thang đo chủ quan MOS. Đánh giá chất lượng chủ quan cho thấy rằng phương pháp được đề xuất có thể tạo ra chất lượng hình ảnh gần tương đương về nhận thức thị giác so với bộ nén gốc với cùng tốc độ bit giảm trung bình 18.83%. Trong hướng nghiên cứu tiếp theo, các tác giả sẽ tiếp tục đánh giá thêm về tốc độ thuật toán khi sử dụng phương pháp đề xuất trên bộ nén video.

REFERENCES

[1] G.-J. Sullivan, J.-R. Ohm, W.-Jin. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[2] I. Marzuki, Y. Ahn, and D. Sim, "Tile-level rate control for tileparallelization HEVC encoders," J. Real-Time Image Proc., vol. 16, no. 6, pp. 2107-2125, Dec. 2019, DOI: 10.1007/s11554-017-0720-5.

[3] JC. C. Chi, M. Alvarez-Mesa, B. Juurlink, G. Clare, F. Henry, S. Pateux, and T. Schierl, "Parallel scalability and efficiency of HEVC parallelization approaches," IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1827-1838, Dec. 2012.

[4] S. Li. C. Zhu, Y. Gao, Y. Zhou, F. Dufaux, and M.-T. Sun, "Lagrangian multiplier adaptation for rate–distortion optimization with inter-frame dependency," IEEE Trans.

Circuits Syst. Video Technol., vol. 26, no. 1, pp. 117–129, Jan. 2016.
[5] J. He, E.-H. Yang, F. Yang, and K. Yang, "Adaptive quantization parameter selection for H.265/HEVC by employing inter-frame dependency," IEEE Trans. Circuits Syst. Video Technol., vol. 28, no. 12, pp. 3424–3436, Dec. 2018.
[6] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," IEEE Signal Process. Mag., vol. 15, no. 6, pp. 23–50, Nov. 1998.
[7] B. Li, D. Zhang, H. Li, and J. Xu, QP determination by lambda value, document JCTVC-I0426, Joint Collaborative Team on Video Coding (JCTVC), 2012.
[8] B. Li, J. Xu, D. Zhang, and H. Li, "QP refinement according to Lagrange multiplier for High Efficiency Video Coding," in Proc. ISCAS, Beijing, China, 2013, pp. 447–480.
[9] M. Wang, K N. Ngan, H. Li, and H. Zeng, "Improved block level adaptive quantization for High Efficiency Video Coding," in Proc. ISCAS, Lisbon, Portugal, 2015, pp. 509-512.
[10] T. Zhao, Z. Wang, and C. W. Chen, "Adaptive quantization parameter cascading in HEVC hierarchical coding," IEEE Trans. Image Process., vol. 25, no. 7, pp. 2997–3009, Jul. 2016.
[11] C. Yeo, H. L. Tan, and Y. H. Tan, "SSIM-based adaptive quantization in HEVC," in Proc. ICASSP, Vancouver, BC, Canada, 2013, pp. 1690–1694.
[12] G. Xiang, H. Ja, M. Yang, X. Zhang, X. Huang, J. Liu, and X. Xie, "A perceptually temporal adaptive quantization algorithm for HEVC," J. Visual Communication and Image Representation, vol. 50, pp. 280-289, Jan. 2018.
[13] S. Ma, X. Zhang, C. Jia, Z. Zhao, S. Wang, and S. Wanga, "Image and video compression with neural networks: A review," IEEE Transactions on Circuits and Systems for Video Technology, vol. PP, pp. 1–1, 04 2019.
[14] Y. Li, B. Li, D. Liu, and Z. Chen, "A convolutional neural network-based approach to rate control in hevc intra coding," in 2017 IEEE Visual Communications and Image Processing (VCIP), 2017, pp. 1–4.
[15] S. Ki, S.-H. Bae, M. Kim, and H. Ko, "Learning-based just noticeable quantization distortion modeling for perceptual video coding," IEEE Transactions on Image Processing, vol. 27, no. 7, pp. 3178–3193, 2018.
[16] H. Choi and I. V. Baji´c, "Deep frame prediction for video coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 7, pp. 1843–1855, 2020.
[17] I. Marzuki and D. Sim, "Perceptual Adaptive Quantization Parameter Selection Using Deep Convolutional Features for HEVC Encoder," in IEEE Access, vol. 8, pp. 37052-37065, 2020, doi: 10.1109/ACCESS.2020.2976142.
[18] Thanh Tùng Phạm, Triều Dương Đinh, Văn Trọng Đặng and Thanh Hà Lê, Phương pháp lựa chọn hệ số lượng tử trên bộ nén H265/HEVC theo đặc trưng thị giác bằng mạng Neural tích chập, Hội nghị Quốc gia lần thứ 24 về Điện tử, Truyền thông và Công nghệ Thông tin (REV-ECIT 2021).
[19] T. P. Thanh, C. Ma Thi, T. N. Manh, L. Le Dinh, and H. Le Thanh, "Compression artifacts image patch database for perceptual quality assessment," in 2020 12th International Conference on Knowledge and Systems Engineering (KSE), 2020, pp. 55–60.
[20] N. Jayant, J. Johnston, R. Safranek, Signal compression based on models of human perception, Proc. IEEE 81 (10) (1993) 1385–1422.
[21] Bjøntegaard, G. (2001). Calculation of Average PSNR Differences between RD-curves.

ADAPTIVE QUANTIZATION PARAMETER SELECTION BY THE LOCAL CONTENT OF VIDEO FRAME

Abstract: This paper presents a method to adaptive quantization parameter (QP) in video compression by the local content on each frame. The proposed method is based on the image quality prediction model by the convolutional neural network when changing the quantization parameter. The dataset used to build the predictive model consists of more than 40,000 mean opinion scores (MOS) of image patches. The proposed method applies quantization parameter selecttion on the H.265/HEVC encoder to reduce the bitrate of compressed video but still ensure the perceptual quality. Experimental results show that the improved encoder based on version HM 16.0 has higher performance according to the SSIM and the MOS metric compared to the original encoder. Compared with related studies, the endcoder applying the proposed method also gives positive results.

Keywords: adaptive quantization parameter, H.265/HEVC encoder, improve compression efficiency.

Phạm Thanh Tùng tốt nghiệp đại học Bách Khoa Hà Nội năm 2007 và hiện là Nghiên cứu sinh chuyên ngành Khoa học máy tính tại trường Đại học Công nghệ, Đại học Quốc gia Hà Nội.
Đơn vị công tác: trường Đại học Phòng cháy chữa cháy.
Lĩnh vực nghiên cứu chính: nhận dạng đám cháy, nén video, thực tế ảo và thực tế tăng cường trong đào tạo, huấn luyện.

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.

Đinh Triều Dương tốt nghiệp đại học và cao học tại trường Đại học Công nghệ, tốt nghiệp tiến sĩ tại đại học Seul. Đơn vị công tác: Giảng viên khoa Điện tử viễn thông, đại học Công nghệ.
Lĩnh vực nghiên cứu: truyền thông, nén vieo.

Lê Thanh Hà tốt nghiệp chuyên ngành Công nghệ thông tin tại trường đại học Công nghệ, Đại học Quốc gia Hà Nội và nhận bằng tiến sĩ tại Khoa Kỹ thuật Điện tử, Đại học Hàn Quốc năm 2010. Đơn vị công tác: Khoa Công nghệ Thông tin, trường đại học Công nghệ, Đại học Quốc gia Hà Nội.
Lĩnh vực nghiên cứu: xử lý hình ảnh/ video, xử lý ảnh vệ tinh và thị giác máy tính.