

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



NGUYỄN VIỆT QUÂN

**NGHIÊN CỨU VÀ ĐÁNH GIÁ HIỆU NĂNG CỦA
CHUẨN NÉN VIDEO H.265/HEVC VỚI CÁC CHUỖI
VIDEO ĐẦU VÀO CÓ ĐẶC TÍNH THAY ĐỔI**

TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

Hà Nội – 2020

Luận văn được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học:

TS. VŨ HỮU TIẾN

Phản biện 1:

Phản biện 2:

Luận văn sẽ được bảo vệ trước Hội đồng chấm luận văn thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông:

Vào lúc: giờ ngày tháng năm 2020

Có thể tìm hiểu luận văn tại:

1. Thư viện Học viện Công nghệ Bưu chính Viễn thông

MỞ ĐẦU

Trong những năm gần đây khi mà lĩnh vực công nghệ và truyền thông ngày càng phát triển cùng với sự phát triển mạnh của internet toàn cầu thì việc truyền tải và lưu trữ video cũng ngày càng trở nên phổ biến hơn. Cùng với đó, chất lượng của phần cứng ngày càng được nâng cao và giá thành lại ngày càng hợp lí giúp cho việc tải hay trải nghiệm một chương trình, bộ phim ngày càng trở nên dễ dàng nhanh chóng và mượt mà hơn.

Tuy nhiên chỉ xem thôi thì chưa đủ, khi chất lượng cuộc sống ngày càng nâng cao kéo theo nhu cầu nghe nhìn vì đó cũng ngày càng khắt khe hơn. Nếu như vài năm trước việc thưởng thức một bộ phim ở độ phân giải 480p hay HD 720p đã đủ sức làm thỏa mãn người xem, thì giờ đây phim với độ phân giải Full HD 1080p đã trở nên quá đỗi bình thường. Và khi những màn hình 4K được cho ra đời với độ phân giải siêu cao gấp 4 lần Full HD thì nhu cầu trải nghiệm của người dùng cũng ngày càng thể hiện mạnh mẽ hơn và cũng ngày càng khắt khe hơn. Cùng với màn hình 4K, phim 4K cũng được cho ra đời mang đến cho người dùng những trải nghiệm cực kì ấn tượng, cực kì kinh ngạc với chất lượng hình ảnh siêu thực làm cho người xem như lạc vào một thế giới khác, thế giới 4K.

Nhu cầu sử dụng nội dung video chất lượng của con người ngày càng cao dẫn đến sự ra đời của các chuẩn video mới có chất lượng cao như 2K, 4K, 8K. Bên cạnh đó, nhu cầu sử dụng video hiện nay còn phổ biến trong điều kiện truyền tải online, không chỉ qua các kênh truyền hữu tuyến chất lượng cao như cáp quang, mà cả kênh truyền vô tuyến WLAN, mạng di động 3G, 4G cũng được khai thác triệt để mọi lúc mọi nơi. Những yêu cầu này dẫn đến việc phải xử lý dữ liệu video trước khi truyền tải nhằm nâng cao chất lượng hình ảnh đồng thời giảm dung lượng truyền dẫn là một yêu cầu cấp thiết cần giải quyết.

Tuy nhiên, để có được những trải nghiệm tuyệt vời đó thì người dùng cũng sẽ phải đầu tư thêm chi phí cho việc nâng cấp các thiết bị lưu trữ cũng như đường truyền internet. Vì thế nhu cầu một chuẩn nén video mới hiệu quả hơn nhằm giảm tải áp lực cho hệ thống ngày càng bức thiết hơn. Và H.265/HEVC đã được cho ra đời nhằm đáp ứng những nhu cầu trên.

H.265/HEVC là một chuẩn nén video hoàn toàn mới với tên gọi đầy đủ là High Efficiency Video Coding (HEVC) được Ủy ban Viễn thông Quốc tế ITU-T thông qua và bắt đầu được các nhà phát triển đưa vào sản phẩm thương mại. H.265 hứa hẹn mang lại

khả năng nén cao gấp đôi so với người tiền nhiệm H.264/AVC (Advance Video Coding) hiện đang được dùng phổ biến khi chiếm tới 80% các nội dung video lưu trữ hiện nay.

HEVC có tỷ lệ nén dữ liệu gấp đôi so với H.264/MPEG-4 AVC ở cùng một mức độ chất lượng video. Nói cách khác nó có thể được sử dụng để cung cấp chất lượng hình ảnh được cải thiện đáng kể khi xét cùng một bitrate. H.265/HEVC cũng mang lại những cải tiến về âm thanh, không gian, màu sắc và quan trọng nhất nó có khả năng giám sát một phạm vi hoạt động nâng cao.

H.265/HEVC cũng mang lại những cải tiến về âm thanh, không gian, màu sắc và quan trọng nhất nó có khả năng giám sát một phạm vi hoạt động nâng cao. Với chuẩn mã hóa này trên các camera thì việc xuất hiện camera 4K-UltraHD và 8k-UltraHD là điều có thể xảy trong tương lai, khi đó video chất lượng cao không chỉ còn là một loại hình phương tiện giải trí, mà nó còn đóng góp giá trị khai thác thực tiễn rất lớn cho lĩnh vực giám sát, giao thông và quốc phòng an ninh.

Chính vì những lý do trên, học viên quyết định chọn đề tài: ***“Nghiên cứu và đánh giá hiệu năng của chuẩn nén video H.265/HEVC với các chuỗi video đầu vào có đặc tính thay đổi”*** làm luận văn thạc sỹ. Trong suốt quá trình nghiên cứu và triển khai đề tài, học viên nhận thấy hiệu quả nén của chuẩn mới tăng đáng kể so với những chuẩn cũ, trong nội dung chương 3 của luận văn học viên đã đi sâu và phân tích đánh giá hiệu năng của H.265/HEVC dựa trên phần mềm mã nguồn mở HEVC Test Model.

Nội dung nghiên cứu của luận văn bao gồm các phần:

CHƯƠNG I: Tổng quan về mã hóa Video.

CHƯƠNG II: Tổng quan về mã hóa, giải mã video độ phân giải cao H.265/HEVC.

CHƯƠNG III: Phân tích đánh giá hiệu năng của chuẩn nén Video HEVC

CHƯƠNG 1. TỔNG QUAN VỀ MÃ HÓA VIDEO

1.1 Giới thiệu chung

Nguyên tắc cơ bản để tạo ra video từ trước đến nay là phát liên tục nhiều hình ảnh trong một giây. Khi những hình ảnh chuyển động liên tục nhau (motion picture) được phát liên tục với tốc độ trình chiếu từ 18 hình/giây trở lên, khi đó mắt của chúng ta ghi nhận đó là một chuỗi hình ảnh liên tục, tạo thành một đoạn phim.

Người ta đã chuẩn hóa video phổ biến ở ba chuẩn khung hình gồm 24 fps, 25 fps và 30 fps (fps - frame per second - số khung hình/giây). Tại Việt Nam và châu Âu các nhà làm phim và truyền hình thường dùng chuẩn 25 hình/giây (gọi là hệ PAL). Tại Mỹ thường dùng chuẩn là 30 hình/giây (gọi là hệ NTSC).

Kỹ thuật nén ảnh số đang đóng một vai trò cực kỳ quan trọng trong các hệ thống viễn thông và đa phương tiện để giải quyết vấn đề băng thông của đường truyền. Các kỹ thuật mã hóa video đều cố gắng làm giảm lượng thông tin không cần thiết cho một chuỗi các bức ảnh mà không làm giảm chất lượng của nó đối với người xem. Nói chung, tín hiệu video thường chứa đựng một lượng lớn các thông tin dư thừa. Chúng thường được chia thành hai loại: dư thừa về mặt không gian (bên trong từng frame) và dư thừa về mặt thời gian (giữa các frame trong chuỗi video). Mục đích của mã hóa video là nhằm làm giảm số bit khi lưu trữ và khi truyền bằng cách phát hiện để loại bỏ các lượng thông tin dư thừa này và dùng các kỹ thuật mã hoá entropy để tối thiểu hoá lượng tin quan trọng cần giữ lại.

Mã hóa video cho phép nén video có dung lượng hàng chục GB giảm xuống chỉ còn vài trăm MB thuận tiện cho việc quản lý và lưu trữ trên các phương tiện giải trí đa dạng theo tỷ lệ mã hóa. Một vấn đề khác cần quan tâm đó là bản thân quá trình tính toán cũng cần tài nguyên nên điều này cần được xem xét trong các tình huống cụ thể, ví dụ trong trường hợp một ứng dụng thời gian thực với yêu cầu độ trễ thấp thì một thuật toán có tỷ số mã hóa rất cao nhưng phức tạp và tốn thời gian giải mã sẽ làm hỏng mục đích của việc mã hóa. Tỷ lệ mã hóa cao về kích thước dữ liệu sẽ đồng nghĩa với việc phải trả giá bằng chất lượng suy giảm và thời gian thực hiện mã hóa lâu hơn.

1.2 Nguyên tắc mã hóa video.

1.2.1. Nguyên tắc chính của mã hóa video.

Các nghiên cứu trước đây đã chỉ ra rằng giữa các frame trong chuỗi video và giữa các pixel trong cùng một frame có một mối tương quan nhất định. Dựa vào các mối tương quan này chúng ta có thể thực hiện việc mã hóa tín hiệu video mà không làm ảnh

hưởng tới độ phân giải của ảnh. Ngoài ra, khai thác đặc điểm của mắt người là kém nhạy cảm với một số thông tin hình ảnh theo không gian và thời gian nên có thể loại bỏ thông tin này trong quá trình mã hóa. Đây chính là kỹ thuật mã hóa tổn hao để tiết kiệm băng thông trong khi vẫn đảm bảo chất lượng video ở mức có thể chấp nhận được.

Nguyên lý của việc mã hóa video dựa trên các kỹ thuật giảm các dư thừa thông tin sau:

- Dư thừa thông tin trong miền không gian (Spatial redundancy): Dư thừa thông tin trong miền không gian xuất hiện giữa các pixel trong cùng một khung hình (ví dụ sự tương đồng giữa các pixel). Thông tin dư thừa được loại bỏ bằng kỹ thuật mã hóa biến đổi (transform coding).
- Dư thừa thông tin trong miền thời gian (Temporal redundancy): Loại thông tin dư thừa này xuất hiện khi giữa các khung ảnh liên tiếp có những thông tin tương đồng. Để giảm dư thừa này người ta dùng kỹ thuật mã hóa sự khác biệt giữa các frame.
- Dư thừa thông tin trong dữ liệu ảnh sau khi mã hóa: Để loại bỏ dư thừa này người ta dùng mã entropy, cụ thể là mã có độ dài thay đổi (Variable Length Coding).

1.2.2 Kỹ thuật giảm dư thừa thông tin trong miền không gian.

a) Mã hóa dự báo

Tại thời điểm ban đầu, phương pháp giảm dư thừa không gian được đưa ra dựa trên việc dự báo giá trị của các pixel hiện tại dựa vào giá trị của pixel đã được mã hóa trước đó. Phương pháp này được gọi là “Điều chế xung mã sai phân” (Differential Pulse Code Modulation – DPCM). hình 1.1 mô tả sơ đồ khối của bộ mã hóa này. Theo sơ đồ, sự sai khác giữa giá trị của pixel hiện tại và giá trị dự báo từ bộ dự báo được lượng tử và mã hóa trước khi truyền đi. Tại phía giải mã, sự sai khác này được cộng với giá trị dự báo từ bộ dự báo để khôi phục lại giá trị đúng của pixel hiện tại. Trong trường hợp bộ lượng tử không được sử dụng thì kỹ thuật này được gọi là mã hóa không tổn hao (loss-less coding).

Bộ dự báo cho kết quả tốt nhất nếu quá trình dự báo được dựa trên những giá trị của các pixel liên kế đã được mã hóa trước đó. Các pixel liên kế có thể nằm trong cùng frame (mã hóa nội ảnh) hoặc có thể nằm trong frame trước (mã hóa liên ảnh). Nếu bộ dự báo sử dụng cả hai kỹ thuật trên thì được gọi là “mã hóa dự báo lai” (Hybrid predictive coding).

b) Mã hóa biến đổi

Mục đích của việc sử dụng mã hóa biến đổi là nhằm loại bỏ dư thừa thông tin theo không gian. Quá trình mã hóa được thực hiện bằng cách biến đổi giá trị của các pixel sang miền không gian khác trước khi loại bỏ những dữ liệu không cần thiết. Ý tưởng chính trong việc áp dụng mã hóa biến đổi vào việc mã hóa ảnh là năng lượng của hầu hết các bức ảnh tập trung chủ yếu tại miền tần số thấp và được biểu diễn bởi các hệ số sau biến đổi. Quá trình lượng tử hóa được sử dụng với mục đích giữ lại các hệ số trong miền tần số thấp và loại bỏ các hệ số trong miền tần số cao. Quá trình này sẽ không làm ảnh hưởng nhiều tới chất lượng của ảnh vì các hệ số trong miền tần số cao lưu giữ ít thông tin của ảnh hơn. Tuy nhiên quá trình này ít nhiều vẫn làm mất thông tin của ảnh gốc.

Hiện nay có nhiều phép biến đổi được đưa ra và đã thể hiện được các ưu điểm của nó như phép biến đổi Wavelet, biến đổi Karhune-Loève, biến đổi Cosin rời rạc (Discrete Cosine Transform – DCT). Trong các chuỗi ảnh video, các pixel có mối tương quan theo không gian (các pixel trong một ảnh) và thời gian (các pixel trong các ảnh liên tục). Do vậy, việc lựa chọn biến đổi DCT 3 chiều là phù hợp nhất. Tuy nhiên, trong thực tế, do phải lưu trữ các frame ảnh trong bộ nhớ đệm khi thực hiện biến đổi DCT 3 chiều nên sẽ không phù hợp trong truyền thông thời gian thực. Vì vậy phép biến đổi DCT 2 chiều thường được sử dụng hơn.

c) Lượng tử hóa các hệ số DCT

Chúng ta lưu ý rằng phép biến đổi các giá trị từ miền không gian pixel sang miền không gian khác chưa phải là mã hóa tín hiệu. Một khối 64 pixel qua phép biến đổi DCT ta thu được 64 hệ số DCT. Do tính trực giao của phép biến đổi, năng lượng trên cả hai miền pixel và miền không gian sau biến đổi đều bằng nhau. Tuy nhiên, phép biến đổi làm cho năng lượng của ảnh được tập trung chủ yếu trong vùng của các hệ số DCT gần với hệ số 1 chiều (vùng tần số thấp) trong khi phần lớn các hệ số khác mang năng lượng ít hơn. Hệ số DC có giá trị cao nhất. Các hệ số AC càng xa hệ số DC thì giá trị càng bé. Dựa vào đặc điểm của mắt người là ít nhạy cảm với các hình ảnh bị méo ở tần số cao nên bộ lượng tử được sử dụng nhằm loại bỏ đi các hệ số AC tại vùng tần số này. Hình 1.2 biểu diễn các giá trị đầu vào và đầu ra của bộ lượng tử. Như ta thấy, với các hệ số có giá trị nhỏ sau khi qua bộ lượng tử sẽ bị mã hóa về mức 0. Đây chính là quá trình loại bỏ những hệ số DCT ở vùng tần số cao.

1.2.3 Kỹ thuật giảm dư thừa thông tin trong miền thời gian.

Kỹ thuật giảm dư thừa thông tin trong miền thời gian được thực hiện dựa trên việc tìm ra sự khác nhau giữa các khung hình liên tiếp. Đây chính là thuật toán mã hóa liên

ảnh. Đối với các đối tượng tĩnh trong ảnh, sự khác biệt gần như bằng 0. Do vậy những đối tượng này không cần nhiều thông tin để mã hóa. Ngược lại, đối với các đối tượng chuyển động nhiều, sự khác biệt giữa các khung hình là rất lớn. Điều này đồng nghĩa với việc chúng ta cần nhiều thông tin để mã hóa. Để làm giảm lượng thông tin này, người ta tiến hành thêm một bước trung gian gọi là *ước lượng chuyển động* (*motion estimation*) cho các đối tượng trong hình. Quá trình ước lượng chuyển động sẽ cho kết quả là các vector chuyển động. Dựa vào các vector này và khung hình trước đó, khung hình hiện tại sẽ được dự đoán. Quá trình này được gọi là “*bù chuyển động*” (*motion compensated*). Như vậy, sự khác biệt giữa khung hình hiện tại và khung hình dự đoán sẽ được giảm đi so với sự khác biệt giữa khung hình hiện tại và khung hình trước đó.

a) Ước lượng chuyển động

Trong các tiêu chuẩn mã hóa video, thuật toán *BMA* (*Block Matching Algorithm*) thường được sử dụng để ước lượng chuyển động. Trong thuật toán này, một khung hình được chia thành các khối có kích thước $N \times N$ pixel. Mỗi khối pixel này sẽ được di chuyển quanh vị trí ban đầu một khoảng w pixel để tìm ra vị trí của khối trong khoảng $(N+2w) \times (N+2w)$ có cùng tọa độ nhưng ở khung hình trước đó giống với nó nhất. Khoảng cách từ tâm hình vuông $(N+2w) \times (N+2w)$ tới tọa độ khối tìm được chính là khoảng chuyển động của khối pixel $N \times N$.

Để tìm và đánh giá mức độ giống nhau giữa hai khối pixel, chúng ta sử dụng một số phương pháp như sử dụng hàm tương quan chéo (*Crosscorrelation function - CCF*), hàm trung bình bình phương lỗi (*Mean Square Error – MSE*) và hàm trung bình tuyệt đối của lỗi (*Mean absolute error – MAE*). Khối pixel giống với khối ban đầu sẽ có CCF lớn nhất hoặc có MSE và MAE nhỏ nhất. Trong thực tế, các chuẩn mã hóa video thường sử dụng phương pháp MSE hoặc MAE:

b) Bù chuyển động

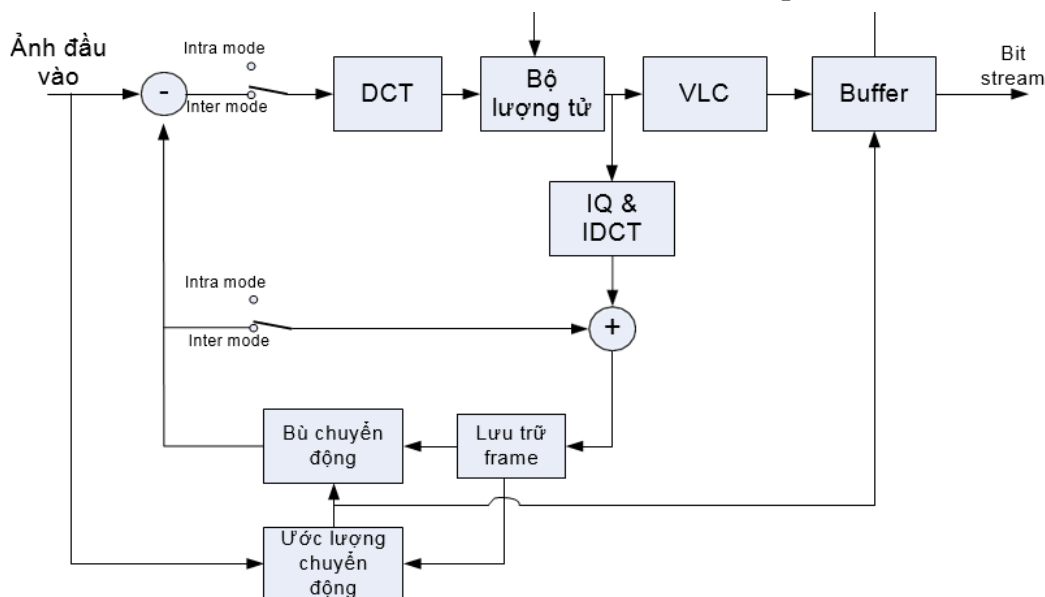
Sau quá trình ước lượng chuyển động ta có được vector chuyển động của các đối tượng trong khối pixel $N \times N$. Dựa vào các vector chuyển động, các khối pixel $N \times N$ trong khung hình trước đó được dịch chuyển theo hướng và độ lớn của vector chuyển động đã chỉ ra. Quá trình này được gọi là bù chuyển động. Kết quả của quá trình này là một khung hình mới được cho là giống với khung hình hiện tại. Tuy nhiên, đây vẫn chỉ là quá trình dự đoán nên khung hình dự đoán sẽ không thể giống hoàn toàn với khung hình hiện tại.

c) Mã có chiều dài thay đổi

Trong trường hợp chúng ta muốn giảm số bit mã hóa tại đầu ra của bộ mã hóa, các hệ số DCT và vector chuyển động sẽ được mã hóa bằng mã có chiều dài thay đổi (*Variable Length Coding – VLC*). Nguyên lý của việc mã hóa này là các từ mã ngắn sẽ được gán cho các thông tin có xác suất xuất hiện lớn, các từ mã dài được gán cho các thông tin có xác suất xuất hiện bé. Như vậy, độ dài của các từ mã sẽ biến đổi tỷ lệ nghịch với xác suất xuất hiện của ký hiệu cần được mã hóa. Theo định luật Shannon, số bit tối thiểu cần thiết để mã hóa một ký hiệu có xác suất xuất hiện p là $\log_2 p$.

1.2.4 Sơ đồ tổng quát của mã hóa video.

Hình 1.6 mô tả sơ đồ tổng quát của bộ mã hóa video được sử dụng trong các chuẩn mã hóa như *H.261*, *H.264*, *MPEG-1*, *MPEG-2* và *H.264/MPEG-4 part 10*.



Hình 1.6 Sơ đồ nguyên lý tổng quát của bộ mã hóa video

a) Intraframe/Interframe loop

Trong chế độ mã hóa liên ảnh, các giá trị khác biệt giữa các pixel của khung hình hiện tại và khung hình dự đoán dựa vào khung hình trước đó được mã hóa và truyền đi. Tại phía thu, sau khi giải mã, các giá trị khác biệt này sẽ được cộng với khung hình dự đoán do bên thu đưa ra để xây dựng lên khung hình hiện tại. Như vậy quá trình dự đoán đóng vai trò rất quan trọng vì nếu dự đoán càng chính xác thì giá trị khác biệt sẽ càng nhỏ và do đó cần ít thông tin để truyền đi. Đối với trường hợp chuỗi ảnh tĩnh, giá trị dự đoán của các pixel trong khung hình hiện tại sẽ là các pixel cùng tọa độ trong khung hình trước đó.

Trong chế độ mã hóa nội ảnh, giá trị dự đoán được dựa trên các khối hoặc các pixel liền kề trước đó trong cùng một khung hình. Chế độ này được áp dụng cho khung hình

đầu tiên của một nhóm các khung hình hoặc áp dụng trong việc mã hóa ảnh tĩnh JPEG sẽ được đề cập chi tiết trong phần sau.

b) Ước lượng chuyển động

Lượng dữ liệu truyền đi sẽ rất lớn nếu với mỗi pixel đều có một vector chuyển động. Thay vào đó, người ta sẽ gán cho mỗi nhóm pixel (block) một vector chuyển động. Trong các chuẩn mã hóa, mỗi nhóm pixel là một hình vuông có kích thước 16x16 (được gọi là một *Macroblock - MB*) được ước lượng chuyển động và được bù chuyển động. Thông thường, việc ước lượng chuyển động được thực hiện trên kênh chói (kênh Y) của các khung hình.

c) Inter/Intra switch

Inter/Intra switch có tác dụng chuyển đổi giữa hai chế độ liên ảnh và nội ảnh. Trong chuẩn mã hóa ảnh JPEG, tất cả các MB trong ảnh được mã hóa ở chế độ nội ảnh. Với các chuẩn mã hóa video như H.26x và MPEG, một số loại frame được mã hóa ở chế độ nội ảnh, một số loại frame được mã hóa ở chế độ liên ảnh. Ngoài ra, trong cùng một frame liên ảnh, một số MB được mã hóa ở chế độ nội ảnh để tối ưu hóa tỷ lệ mã hóa.

d) DCT

Trên kênh Y và hai kênh màu U, V, mỗi MB được chia thành các khối nhỏ hơn có kích thước 8x8. Như vậy, sẽ có 4 khối 8x8 thuộc kênh Y và một số khối thuộc kênh màu tùy theo độ phân giải của ảnh. Mỗi khối 8x8 này sẽ được biến đổi DCT để có được một ma trận 8x8 các hệ số của biến đổi DCT.

d) Lượng tử hóa

Có hai loại lượng tử hóa. Một loại có dead zone được sử dụng cho các hệ số AC và hệ số DC của các MB mã hóa liên ảnh. Loại còn lại không có dead zone được sử dụng cho các hệ số DC của MB nội ảnh. Khoảng giá trị của các hệ số được lượng tử có thể từ -2047 đến +2047. Với bộ lượng tử có dead zone, nếu trị tuyệt đối của các hệ số nhỏ hơn bước lượng tử q thì sẽ được gán bằng 0. Ngược lại, giá trị của các hệ số sau lượng tử sẽ trong khoảng từ 1 đến 31.

e) Variable Length Coding

Các hệ số sau khi được lượng tử sẽ được mã hóa bằng mã có độ dài thay đổi. Ngoài ra, giá trị của vector chuyển động cũng được mã hóa bằng mã này cùng với các hệ số lượng tử.

f) Giải lượng tử và biến đổi DCT ngược

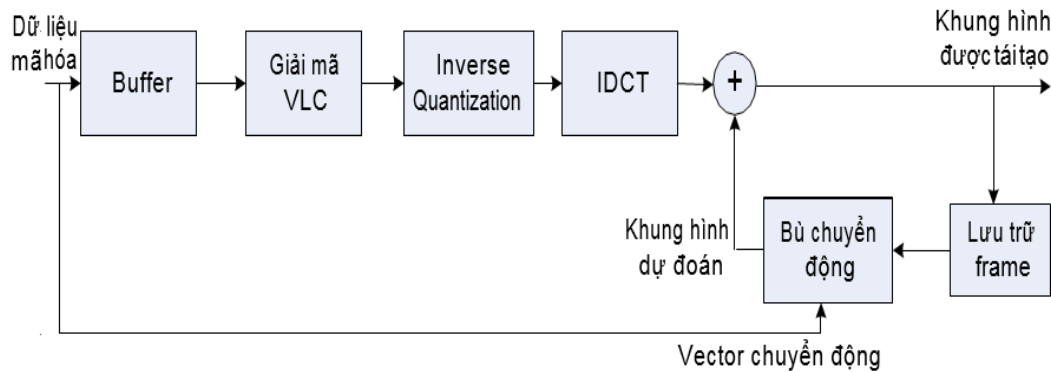
Để tái tạo khung hình hiện thời, các hệ số DCT sau lượng tử được giải lượng tử và biến đổi DCT ngược. Sau đó, các giá trị này được cộng với khung hình trước đó đang được bộ mã hóa lưu giữ để tái tạo lại khung hình hiện thời. Khung hình hiện thời này sẽ lại được lưu giữ để dùng cho quá trình dự đoán khung hình tiếp theo.

g) Bộ đệm

Tốc độ bit được tạo ra bởi bộ mã hóa sẽ liên tục thay đổi vì tốc độ bit phụ thuộc vào mức độ chuyển động của các đối tượng trong video. Vì vậy, trước khi truyền tín hiệu video trên các kênh truyền có băng thông cố định thì các bit được lưu trong bộ đệm để điều tiết việc truyền đi. Bộ đệm trong trường hợp này chính là bộ nhớ có hai cổng ghi và đọc. Cổng ghi có nhiệm vụ nhận dữ liệu là các bit sau khi được mã hóa. Cổng đọc có nhiệm vụ đọc ra dữ liệu với một tốc độ ổn định. Mỗi khi dữ liệu được đọc thì bộ đệm sẽ giải phóng dữ liệu đó trong bộ đệm. Trong trường hợp có dữ liệu ghi vào nhiều hơn dữ liệu đọc ra (với những video có nhiều chuyển động) thì bộ đệm gửi thông tin phản hồi tới bộ lượng tử để tăng mức lượng tử. Khi đó dữ liệu ra của bộ mã hóa sẽ ít đi vì các hệ số DCT bị giảm đi. Ngược lại, với những video có ít chuyển động thì bộ lượng tử sẽ giảm mức lượng tử để cải thiện chất lượng video. Quá trình này được gọi là ổn định tốc độ bit (*Constant Bit Rate*). Với những bộ mã hóa có tốc độ bit thay đổi (*Variable Bit Rate—VBR*) thì bộ đệm không được sử dụng và mức lượng tử được giữ nguyên trong suốt quá trình mã hóa.

1.3 Giải mã hóa video.

Về cơ bản, quá trình giải mã hóa bao gồm các bước giống như quá trình mã hóa nhưng thứ tự ngược lại. hình 1.7 mô tả quá trình giải mã hóa tín hiệu video. Ban đầu dữ liệu nhận được là các tín hiệu mã hóa được tách thành hai phần: dữ liệu mã hóa giá trị khác biệt và dữ liệu mã hóa giá trị vector chuyển động. Vector chuyển động sẽ được đưa vào bộ dự đoán để dự đoán khung hình hiện tại. Dữ liệu mã hóa giá trị khác biệt giữa khung hình hiện tại và khung hình trước đó lần lượt được giải mã entropy, giải lượng tử, biến đổi DCT ngược. Sau khi biến đổi DCT ngược ta sẽ được khung hình là hiệu số của khung hình hiện tại và khung hình dự đoán. Sau khi được cộng với khung hình dự đoán, ta sẽ thu được khung hình tái tạo của khung hình hiện tại.



Hình 1.7 Sơ đồ giải mã hóa tín hiệu video

1.4 Kết luận chương.

Chương đầu luận văn đã tìm hiểu nguyên tắc chung của các bộ mã hóa video và các nguyên tắc được sử dụng trong các chuẩn mã hóa video như H.261, H.263, MPEG -1, 2 và 4

Sau khi mã hóa video, các dòng bit hoặc sẽ được đóng gói vào các gói tin để truyền qua mạng hoặc được đóng gói vào các file để lưu trữ. Quá trình giải mã hóa bao gồm các bước giống như quá trình mã hóa nhưng thứ tự ngược lại

Công nghệ không ngừng phát triển. Điều này cũng không ngoại lệ khi nói đến quá trình mã hóa và phát video. Các kỹ sư công nghệ luôn cố gắng không ngừng nghỉ nhằm cải thiện cách thức thu, mã hóa và phát hình ảnh nhằm tối đa hiệu quả lưu trữ và thời gian truyền tải. Vào năm 2013, Liên hiệp truyền thông quốc tế ITU đã giới thiệu chuẩn mã hóa mang tên H.265 hay còn gọi High Efficiency Video Coding – HEVC. Cũng giống như trước đó, H.265 là thành quả của sự kết hợp giữa ITU-T VCEG và ISO/IEC MPEG.

Chuẩn H.265 hứa hẹn là chuẩn mã hóa thay thế xứng đáng cho chuẩn H.264 vốn đang rất phổ biến cho đến thời điểm hiện tại. Với chuẩn H.265 này khi xem video online sẽ giảm được $\frac{1}{2}$ băng thông tải và không phải chi trả quá nhiều tiền cho việc xem video với kết nối 3G/4G mà vẫn thưởng thức được video chất lượng cao, thời gian tải video giảm đáng kể cũng giúp chúng ta tiết kiệm được rất nhiều thứ. Nội dung tiếp theo của đề tài sẽ giới thiệu đầy đủ hơn về chuẩn mã hóa tiên tiến này, đồng thời trình bày thuật toán mã hóa video với cấu trúc nhóm ảnh tùy biến theo nội dung của video có thể ứng dụng và tối ưu hiệu quả của chuẩn mã hóa H.265.

CHƯƠNG 2. TỔNG QUAN VỀ MÃ HÓA VIDEO, GIẢI MÃ VIDEO ĐỘ PHÂN GIẢI CAO H.265/HEVC

2.1 Tổng quan về mã hóa video độ phân giải cao.

2.1.1 Giới thiệu chuẩn nén H.265.

Chuẩn nén H.265 là một chuẩn mã hóa/giải mã video và định dạng video tốt nhất hiện nay. Chuẩn nén H.265 có khả năng nén gấp đôi so với chuẩn trước (chuẩn H.264) đó mà chất lượng không đổi. Chuẩn này có thể hỗ trợ độ phân giải lên đến 8192 x 4320, bao gồm cả chất lượng 8k Ultra HD.

Các ưu điểm của chuẩn nén H.265:

- Được trang bị tính năng hỗ trợ xử lý song song hiệu quả.
- Phân vùng linh hoạt hơn, với kích cỡ các vùng từ lớn đến nhỏ tùy theo nội dung của khung hình.
- Chế độ dò tìm/dự báo và phát tín hiệu, chuyển động của các vector phức tạp hơn.
- Bộ lọc nội suy và bộ lọc khử nhiễu cho các khối ảnh tinh vi hơn.
- Linh hoạt hơn trong các chế độ dự báo và chuyển đổi kích thước khối.

2.1.2 Cấu trúc mã hoá.

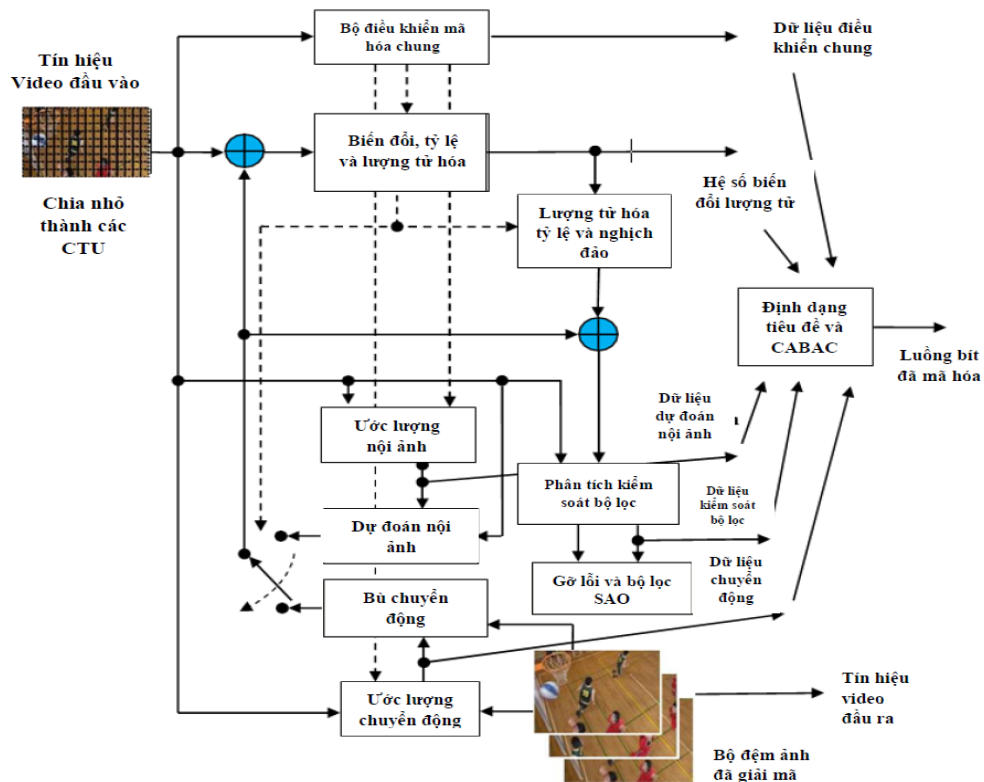
Lớp mã hóa video của HEVC thực hiện cách tiếp cận lai (dự đoán trong ảnh/liên ảnh và mã hóa biến đổi 2D) như được sử dụng trong tất cả các tiêu chuẩn nén video kể từ H.261. Sơ đồ khối của bộ mã hóa video HEVC được thể hiện như ở Hình 2.3.

Mỗi ảnh đầu vào được chia thành các khối ảnh, sau đó được mã hóa và được truyền tải đến các bộ giải mã. Ảnh đầu tiên của một chuỗi video được mã hoá chỉ sử dụng dự đoán trong ảnh. Đối với các ảnh còn lại của chuỗi sử dụng các chế độ mã hóa dự đoán liên ảnh theo thời gian. Quá trình mã hóa cho dự đoán liên ảnh sẽ lựa chọn dữ liệu chuyển động bao gồm các ảnh tham chiếu và vector chuyển động (MV) dùng để dự đoán các mẫu của mỗi khối ảnh. Các bộ mã hóa và giải mã tạo ra tín hiệu dự đoán liên ảnh giống nhau bằng cách sử dụng kỹ thuật bù chuyển động (MC) và thông tin phụ (side information) là dữ liệu quyết định chế độ.

Các tín hiệu dư thừa của dự đoán trong ảnh và liên ảnh sẽ được biến đổi bởi phép biến đổi không gian tuyến tính. Sau đó các hệ số biến đổi được định cỡ, lượng tử hóa, mã hóa Entropy, và được truyền cùng với các thông tin dự đoán.

Đồng thời, bộ mã hóa cũng sao chép lại mạch vòng xử lý giải mã sao cho cả hai phía mã hóa và giải mã cùng tạo ra các dự đoán giống nhau đối với ảnh kế tiếp. Do đó, các hệ

số biến đổi lượng tử hóa sẽ được tạo lại qua định cỡ ngược và sau đó là biến đổi ngược để sao lại gần đúng tín hiệu dư thừa. Sau đó các tín hiệu dư thừa này cộng với các tín hiệu dự đoán và được đưa vào một hoặc hai bộ lọc để làm trơn ảnh. Ảnh biểu diễn cuối cùng (một bản sao của đầu ra của bộ giải mã) được lưu trữ trong một bộ đệm ảnh giải mã và được sử dụng để dự đoán các ảnh tiếp theo. Nói chung, thứ tự của tiến trình mã hóa hoặc giải mã ảnh thường khác so với thứ tự ảnh đến từ nguồn, đòi hỏi phải có sự phân biệt giữa thứ tự giải mã (thứ tự dòng bit) và thứ tự đầu ra (thứ tự hiển thị) của một bộ giải mã. Sau đây là mô tả chi tiết các thành phần trong sơ đồ bộ mã hóa HEVC.



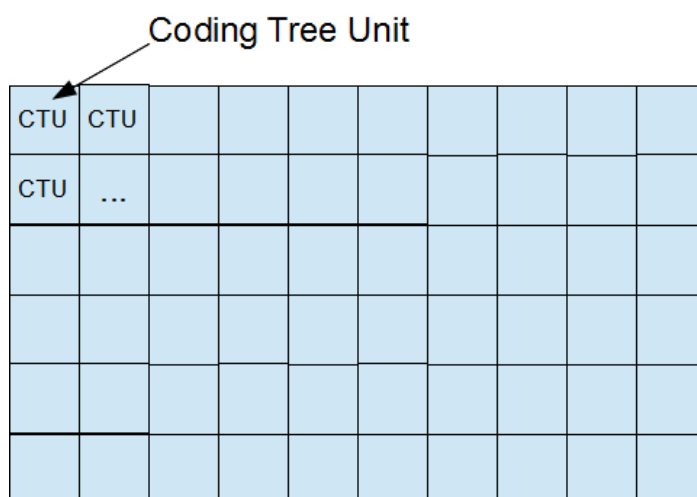
Hình 2.1 Sơ đồ khối bộ mã hóa HEVC

a) Cấu trúc đơn vị cây mã hóa (CTU) và khối cây mã hóa (CTB)

Cốt lõi của lớp mã hóa trong các tiêu chuẩn video trước đây là macroblock, bao gồm một khối 16x16 các mẫu thành phần chói (chói) và hai khối 8x8 các mẫu thành phần màu (trường hợp lấy mẫu 4:2:0).

Tuy nhiên, sự tăng trưởng nhanh chóng về nhu cầu trải nghiệm video chất lượng cao đã dẫn đến sự ra đời của các đơn vị hình ảnh với độ phân giải cực lớn. Các sản phẩm với độ phân giải màn hình 4K trở nên phổ thông và mọi người bắt đầu nói về 8K. Ngay cả thiết bị di động cũng có kích thước khung hình HD cao hơn 2048x1530. Do đó mà chúng

ta cần các macroblock lớn hơn để mã hóa hiệu quả vector chuyển động cho các kích thước khung hình này. Mặt khác, các chi tiết nhỏ cũng vẫn quan trọng và đôi khi chúng ta cần thực hiện dự đoán và chuyển đổi ở mức chi tiết đến 4x4.



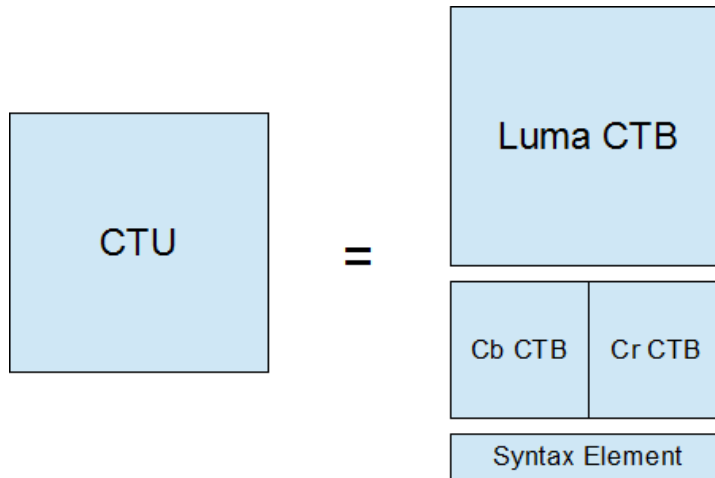
Hình 2.2 Cấu trúc CTU

Vậy làm cách nào chúng ta có thể hỗ trợ nhiều loại kích thước theo khối như vậy một cách hiệu quả. Đó là một thách thức mà HEVC đã cố gắng giải quyết bằng cách định nghĩa ra các đơn vị cấu hình mới, trong đó có CTU (*Coding Tree Unit – đơn vị cây mã hóa*) là đơn vị ở cấp cao nhất. Giả sử chúng ta có một hình ảnh để mã hóa, HEVC sẽ chia hình ảnh thành các CTU như Hình 2.4.

Chiều rộng và chiều cao của CTU được cấu hình trong bộ tham số, nghĩa là tất cả các CTU trong chuỗi video sẽ có cùng kích thước: 64x64, 32x32, 16x16 và các kích thước này được hiểu là kích thước logic, không chứa đơn vị đo lường cụ thể.

CTU là một đơn vị logic và nó thường bao gồm ba khối: một mẫu chói (Y), hai mẫu màu (Cb và Cr) và các phần tử cú pháp liên quan. Mỗi khối được gọi là CTB (*Coding Tree Block – Khối cây mã hóa*).

Mỗi CTB cũng có cùng kích thước với CTU là: 64x64, 32x32, 16x16, trong đó các kích thước lớn hơn sẽ cho phép nén tốt hơn.



Hình 2.3 Cấu trúc CTB

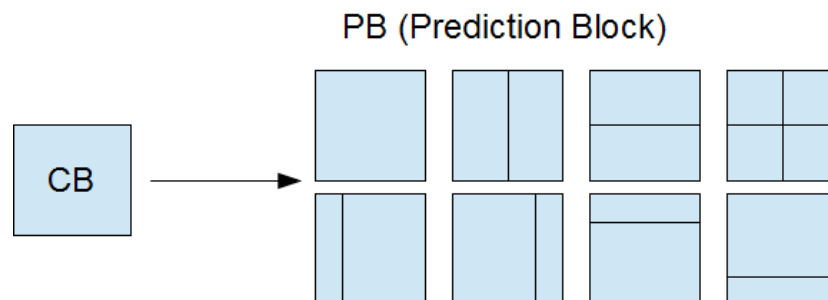
b) Đơn vị mã hóa (CU) và khối mã hóa (CB)

Tuy nhiên, tùy thuộc vào một phần của khung hình video, CTB có thể quá lớn để quyết định xem có nên thực hiện dự đoán liên ảnh hay dự đoán nội ảnh hay không. Do đó, mỗi CTB có thể được chia thành nhiều CB (*Coding Block – Khối mã hóa*) khác nhau. Ví dụ: một số CTB được chia thành 16x16 CB trong khi các CTB khác được chia thành 8x8 CB. Hình 2.6 minh họa cách chia nhỏ thành cấu trúc CB của các CTB 64x64.

Trong đó, CB là điểm quyết định xem có có thực hiện dự đoán liên ảnh hoặc nội ảnh hay không, và kiểu dự đoán sẽ được mã hóa trong CU (*Coding Unit – Đơn vị mã hóa*). Cú pháp cây tứ phân (*Quadtree Syntax*) chứa trong CTU cho phép chia các CB có kích cỡ và vị trí thích hợp dựa trên đặc tính tín hiệu của vùng tạo bởi CTB. Do đó gốc của cây tứ phân là CTU. Thông thường, kích cỡ của CB được xác định là: $8 \times 8 < \text{kích cỡ CB size} < \text{kích cỡ CTB}$. Khi đó, đơn vị mã hóa CU sẽ gồm 1 CB chói và 2 CB màu, cùng cú pháp đi kèm. Một CTB có thể chứa chỉ một CU hoặc được phân chia thành nhiều CU, và mỗi CU có một phân vùng liên quan đến các đơn vị dự báo (PUs) và một cây các đơn vị biến đổi (TUs).

c) Đơn vị dự đoán (PU) và khối dự đoán (PB)

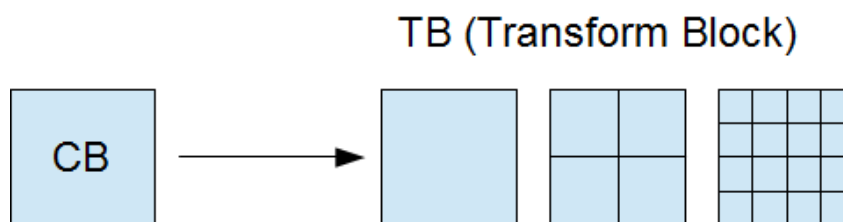
Đơn vị CB là đủ tốt cho việc đưa ra quyết định sử dụng phương pháp dự đoán liên ảnh hay nội ảnh, nhưng nó vẫn có thể quá lớn để lưu trữ các vector chuyển động. Ví dụ, một đối tượng rất nhỏ như tuyết rơi có thể di chuyển ở giữa CB kích thước 8x8 sẽ không thể được xử lý chính xác. Vì vậy, HEVC cần một đơn vị với cấu trúc nhỏ hơn gọi là PB như Hình 2.7. Mỗi CB có thể được chia thành các PB (*Prediction Block – Khối dự đoán*) khác nhau tùy thuộc vào phương pháp dự đoán theo thời gian hay không gian. HEVC hỗ trợ các kích cỡ PB khác nhau từ 4x4 cho đến 64x64.



Hình 2.5 Cấu trúc PB

d) Đơn vị biến đổi (TU) và khối biến đổi (TB)

Một khi dự đoán được thực hiện, chúng ta cần phải mã hóa phần dư (là sự khác biệt giữa hình ảnh dự đoán và hình ảnh thực tế) với phép biến đổi DCT. Trong trường hợp này, cấu trúc CB vẫn có thể là quá lớn cho việc thực hiện thủ tục này bởi một CB có thể chứa cả một phần ảnh với mức độ chi tiết cao (tín hiệu cao tần) và một phần ảnh trơn phẳng (tín hiệu thấp tần). Do đó HEVC chia mỗi CB thành các TB (*Transform Block – Khối biến đổi*) khác nhau và các TB không cần thiết phải tương ứng với mỗi PB trong cấu trúc ảnh. TB có thể thực hiện phép biến đổi đơn trên phần dư của nhiều PB khác nhau và ngược lại.



Hình 2.6 Cấu trúc TB

e) Tín hiệu hóa vector chuyển động

Dự đoán vector chuyển động tiên tiến (AMVP) được sử dụng, bao gồm cả dẫn xuất của nhiều ứng viên có khả năng nhất dựa trên dữ liệu từ các PB liền kề và hình ảnh tham chiếu. Một chế độ kết hợp cho mã hóa MV cũng có thể được sử dụng, cho phép thừa kế của các MV từ các PBs lân cận theo thời gian hoặc không gian. Hơn nữa, so với H.264 / MPEG-4 AVC, suy luận bỏ qua được cải thiện và suy luận chuyển động trực tiếp cũng được quy định.

f) Bù chuyển động

Độ chính xác $\frac{1}{4}$ -mẫu được sử dụng cho các MV, và các bộ lọc 7-tap hoặc 8-tap được sử dụng cho phép nội suy của các vị trí mẫu-phân đoạn (được so sánh với việc lọc 6-tap của vị trí $\frac{1}{2}$ -mẫu theo sau bởi phép nội suy tuyến tính cho vị trí $\frac{1}{4}$ -mẫu trong H.264 /

MPEG-4 AVC). Tương tự như H.264 / MPEG-4 AVC, nhiều ảnh tham chiếu được sử dụng. Đối với mỗi PB, một hoặc hai vectơ chuyển động có thể được truyền đi với mã hóa dự đoán đơn hoặc mã hóa dự đoán đối tượng. Như trong H.264 / MPEG-4 AVC, một hoạt động định cỡ (scale) và bù đắp được áp dụng cho các tín hiệu dự đoán gọi là dự đoán có trọng số.

g) Dự đoán trong ảnh

Các mẫu biên được giải mã của các khối liên kề là dữ liệu tham chiếu cho dự đoán theo không gian trong các vùng dự đoán liên ảnh không được thực hiện. Dự đoán trong ảnh hỗ trợ 33 chế độ có hướng (so với 8 chế độ trong H.264 / MPEG-4 AVC), cộng với chế độ dự đoán mặt phẳng và DC. Các chế độ dự đoán trong ảnh được chọn sẽ được mã hóa bằng dẫn xuất chế độ có thể xảy ra nhất (ví dụ, các hướng dự đoán) dựa trên những PB lân cận giải mã trước đây đó.

h) Điều khiển lượng tử hóa

Như trong H.264 / MPEG-4 AVC, lượng tử hóa tái thiết đồng nhất (URQ) được dùng trong HEVC, với các ma trận định cỡ (scale) lượng tử hóa được hỗ trợ cho các kích thước khối biến đổi khác nhau.

i) Mã hóa Entropy

Mã hóa số học nhị phân thích ứng ngữ cảnh (CABAC) được sử dụng để mã hóa Entropy. Nó tương tự như các sơ đồ CABAC trong H.264 / MPEG-4 AVC, nhưng có thêm một số cải tiến để cải thiện tốc độ lưu lượng (đặc biệt là cho các kiến trúc xử lý song song), hiệu suất nén, và để giảm bộ nhớ ngữ cảnh của nó.

j) Bộ lọc bỏ khối trong-vòng

Một bộ lọc bỏ khối giống như bộ lọc trong H.264 / MPEG-4 AVC được dùng trong vòng lặp dự đoán liên ảnh. Tuy nhiên, thiết kế được đơn giản hóa về các quá trình đưa ra quyết định và lọc, và nó được làm thuận lợi hơn để xử lý song song.

k) Bù thích ứng mẫu (SAO)

Một ánh xạ biên độ phi tuyến tính được đưa vào trong vòng lặp dự đoán liên ảnh đằng sau bộ lọc bỏ khối. Mục đích là để tái tạo tốt hơn các biên độ tín hiệu ban đầu bằng cách sử dụng một bảng look-up, bảng này được mô tả bằng một vài thông số bổ sung mà chúng có thể được xác định bằng cách phân tích biểu đồ ở phía bộ mã hóa.

2.1.2 Các kỹ thuật mã hóa video HEVC

a) Biểu diễn lấy mẫu của ảnh

Để biểu diễn các tín hiệu video màu, HEVC sử dụng một không gian 3 màu cơ bản YCbCr với lấy mẫu 4:2:0. Các ảnh video lấy mẫu tăng lên với kích thước ảnh hình chữ nhật $W \times H$, trong đó W là chiều rộng và H là chiều cao của ảnh theo quan điểm lấy mẫu chói, còn với lấy mẫu màu 4:2:0 sẽ là $W/2 \times H/2$.

b) Phân chia hình ảnh thành đơn vị cây mã hóa

Một ảnh được phân chia thành các đơn vị mã hóa cây (CTU), mỗi CTU chứa các CTB chói và CTB màu. Một CTB chói lấy mẫu $L \times L$ cho các thành phần chói và lấy mẫu $L/2 \times L/2$ cho mỗi một trong hai thành phần màu. $L=16,32,64$ được xác định bởi một yếu tố cú pháp mã hóa được quy định trong SPS.

c) Phân chia CTB thành CB

Các khối CTB chói và màu có thể được sử dụng trực tiếp như các CB hoặc có thể phân chia tiếp thành nhiều CB. Việc phân chia được thực hiện bằng cách sử dụng cấu trúc cây. Việc phân chia cấu trúc cây trong HEVC thường được áp dụng đồng thời cho cả khối chói và màu, mặc dù các ngoại lệ áp dụng khi tiến tới kích thước tối thiểu nhất định.

CTU chứa cú pháp cây tứ phân cho phép chia các CB đến một kích thước phù hợp dựa trên các đặc điểm tín hiệu của vùng mà nó được bao phủ bởi các CTB. Quá trình chia tách cây tứ phân có thể được lặp đến khi một CB chói đạt đến một kích thước tối thiểu cho phép. Kích thước này được chọn bởi bộ mã hóa sử dụng cú pháp trong SPS và luôn là 8×8 hoặc lớn hơn.

d) Khối dự đoán (PB) và đơn vị dự đoán (PU)

Một PB của phân chia bất đối xứng có chiều cao hoặc chiều rộng là $M/4$ và chiều rộng hoặc chiều cao là M và các PB khác lấp đầy phần còn lại của CB bởi 1 chiều cao hoặc chiều rộng của $3M/4$ và chiều rộng hoặc chiều cao M . Mỗi PB dự đoán liên ảnh được gán một hoặc hai vectơ chuyển động và các chỉ số ảnh tham chiếu. Để tối thiểu băng thông bộ nhớ trường hợp xấu nhất, các PB của thành phần chói 4×4 là không được phép cho dự đoán liên ảnh, và các PB của thành phần chói 4×8 và 8×4 được giới hạn cho mã hóa dự đoán đơn. Các PB thành phần chói và màu kết hợp với cú pháp dự đoán tạo thành PU.

e) Phân chia cây cấu trúc thành các đơn vị biến đổi và các khối biến đổi

Việc phân chia được báo hiệu bằng một cây tứ phân dư thừa.

Phân chia CB và TB, nơi mà một khối có thể được chia 1 cách đệ quy thành các góc phân tư, như minh họa trong Hình 2.10. Đối với một CB thành phần chói $M \times M$, một cờ báo hiệu có được chia thành bốn khối có kích thước $M/2 \times M/2$ không. Độ sâu tối đa của

cây tứ phân dư thừa trong SPS sẽ báo hiệu cho việc có chia tách thêm nữa không. Mỗi góc phần tư có một cờ chỉ định việc nó có được chia thành bốn góc phần tư không. Các khối nút lá từ cây tứ phân dư thừa là các khối biến đổi được xử lý tiếp bằng mã hóa biến đổi. Các bộ mã hóa chỉ ra kích thước TB thành phần chói tối thiểu và tối đa mà nó sẽ sử dụng. Việc chia tách là ngầm mặc định khi kích thước CB lớn hơn so với kích thước TB tối đa. Và việc không chia tách là ngầm mặc định khi chia tách sẽ cho kết quả là kích thước TB thành phần chói nhỏ hơn kích thước tối thiểu. Kích thước TB thành phần màu bằng một nửa kích thước TB thành phần chói ở mỗi chiều, ngoại trừ trường hợp kích thước TB chói là 4x4, trong đó trường hợp đó, một TB màu 4x4 duy nhất được bao phủ bởi bốn TB chói 4x4. Trong trường hợp các CU dự đoán trong ảnh, các mẫu được giải mã của các TB lân cận gần nhất (trong hoặc ngoài CB) được sử dụng như là dữ liệu tham chiếu cho dự đoán trong ảnh.

f) Mảng (slice) và Tile

Mảng là một chuỗi các CTU được xử lý theo thứ tự của quét mảnh. Một ảnh được chia thành một hoặc một vài mảnh như thể hiện trong Hình 2.11(a) vì thế một ảnh là một tập hợp của một hoặc nhiều mảnh. Mỗi mảnh có thể được mã hóa sử dụng các loại mã hóa khác nhau như sau.

- *Mảng I*: tất cả CU của mảng được mã hóa chỉ sử dụng dự đoán trong ảnh.
- *Mảng P*: Ngoài các loại mã hóa của mảng I, một số CU của mảng P cũng có thể được mã hóa bằng dự đoán liên ảnh với nhiều nhất một tín hiệu dự đoán bù chuyển động cho mỗi PB (dự đoán đơn). Mảng P chỉ sử dụng danh sách 0 ảnh tham chiếu.
- *Mảng B*: Bên cạnh các loại mã hóa có sẵn trong một mảng P, một số CU của mảng B cũng có thể được mã hóa bằng dự đoán liên ảnh với nhiều nhất hai tín hiệu dự đoán bù chuyển động cho mỗi PB (dự đoán đôi). Mảng B sử dụng cả hai danh sách 0 và 1 ảnh tham chiếu.

Mục đích chính của các mảng là đồng bộ lại sau khi dữ liệu mất. Hơn nữa, các mảng thường được hạn chế để sử dụng số bit tối đa. Vì vậy, các mảng thường có thể chứa số CTU thay đổi mức độ cao/mảnh.

Tile là các vùng hình chữ nhật của ảnh có khả năng giải mã độc lập. Mục đích chính của tile là cho phép sử dụng các kiến trúc xử lý song song đối với mã hóa và giải mã. Nhiều tile có thể chia sẻ thông tin tiêu đề được chứa trong cùng mảnh. Nói cách khác,

một tile đơn lẻ có thể chứa nhiều mảng. Tile gồm một nhóm các CTU xếp hình chữ nhật như Hình 2.11(b).

Để hỗ trợ việc đóng gói dữ liệu, các mảng phụ thuộc được bổ sung. Cuối cùng, với WPP, một mảng được chia thành các hàng CTU. Việc giải mã của mỗi hàng có thể được bắt đầu ngay sau một vài quyết định dự báo và thích ứng của các bộ mã hóa Entropy được tạo ra ở hàng trước. Điều này hỗ trợ xử lý song song các hàng của các CTU bằng cách sử dụng một số luồng xử lý trong bộ mã hóa hoặc giải mã (hoặc cả hai). Một ví dụ được trình bày trong Hình 2.11(c). Đối với thiết kế đơn giản, WPP không được sử dụng kết hợp với các Tile.

g) Dự đoán trong ảnh

Dự đoán trong ảnh hoạt động theo kích thước TB, và các mẫu biên giải mã trước đó từ những TB lân cận về mặt không gian được sử dụng để tạo ra tín hiệu dự đoán. Dự đoán trong ảnh hỗ trợ 33 hướng khác nhau với các kích thước TB từ 4x4 đến 32x32. Các hướng dự đoán được thể hiện trong Hình 2.12. Ngoài ra dự đoán mặt phẳng và dự đoán DC cũng có thể được sử dụng. Đối với thành phần màu, chế độ dự đoán ngang, dọc, mặt phẳng và DC có thể được báo hiệu một cách rõ ràng, hoặc chế độ dự đoán thành phần màu được chỉ định giống các chế độ dự đoán thành phần chói.

Mỗi CB được mã hóa bằng một trong những loại mã hóa, tùy thuộc vào loại mảng. Tương tự như H.264/MPEG-4/AVC, mã hóa dự đoán trong ảnh được hỗ trợ trong tất cả các loại mảng. HEVC hỗ trợ các phương pháp mã hóa dự đoán trong ảnh khác nhau gọi là *Intra_Angular*, *Intra_Planar*, và *Intra-DC*. Sau đây trình bày một lời giải thích ngắn gọn về những điều này và một số kỹ thuật được áp dụng chung.

Phân chia khối dự đoán (PB)

Một CB dự đoán trong ảnh có kích thước $M \times M$ có thể có một trong hai loại phân chia PB gọi là $PART_2N \times 2N$ và $PART_N \times N$, loại đầu tiên chỉ ra rằng các CB không phân chia và loại thứ hai chỉ ra rằng CB được chia thành bốn PB kích thước bằng nhau. ($N=M/2$). Tuy nhiên, nó có khả năng biểu diễn các vùng tương tự mà có thể được xác định bởi bốn PB bằng cách sử dụng bốn CB nhỏ hơn khi kích thước của các CB hiện tại lớn hơn so với kích thước CU tối thiểu. Vì vậy, HEVC chỉ cho phép kiểu phân chia $PART_N \times N$ được sử dụng khi kích thước CB hiện tại bằng với kích thước CU tối thiểu. Điều này có nghĩa rằng kích thước PB luôn bằng với kích thước CB khi CB được mã hóa bằng chế độ dự đoán trong ảnh và kích thước CB là không bằng với kích thước CU tối

thiếu. Dù chế độ dự đoán trong ảnh ở mức PB, nhưng các quá trình dự đoán thực tại lại hoạt động riêng biệt cho từng TB.

Dự đoán Intra_Angular

Khi sử dụng một chế độ Intra_Angular, mỗi TB được dự đoán là có hướng từ các mẫu lân cận theo không gian đã được tái tạo lại (nhưng chưa được lọc bởi bộ lọc trong vòng) trước khi được sử dụng cho dự đoán này. Đối với một TB có kích thước $N \times N$, tổng $4N+1$ mẫu lân cận theo không gian được sử dụng để dự đoán, như trong Hình 2.5. Khi hoạt động giải mã là có trước, các mẫu từ các TB phía dưới bên trái được sử dụng để dự đoán trong HEVC bên cạnh các mẫu từ các TB ở bên trái, phía trên, và phía trên bên phải của TB hiện tại.

Quá trình dự đoán của chế độ Intra_Angular có thể liên quan đến mẫu ngoại suy từ vị trí mẫu tham chiếu dự đoán theo một hướng đã cho. Để loại bỏ sự chuyển đổi mẫu-theo-mẫu giữa các bộ đệm hàng và cột tham chiếu, cho Intra_Angular [k] với $k=2...17$, các mẫu nằm ở hàng trên được dự tính như mẫu bổ sung nằm ở cột bên trái và với $k = 18...34$, các mẫu ở cột bên trái được dự tính nằm cùng với các mẫu hàng trên.

Để cải thiện tính chính xác dự đoán trong ảnh, vị trí mẫu tham chiếu dự tính được tính toán với độ chính xác $1/32$ mẫu. Phép nội suy song tuyến tính được sử dụng để có giá trị của mẫu tham chiếu dự tính sử dụng hai mẫu tham chiếu gần nhất nằm ở vị trí số nguyên.

Quá trình dự đoán của các chế độ Intra_Angular là nhất quán trên tất cả các kích thước khối và hướng dự đoán, trong khi H.264/MPEG-4/AVC sử dụng các phương pháp khác nhau cho các kích thước 4×4 , 8×8 và 16×16 . Thiết kế này là đặc biệt, vì HEVC hỗ trợ nhiều hơn về kích cỡ TB và tăng đáng kể các hướng dự đoán so với H.264/MPEG-4 AVC.

Dự đoán Intra-Planar và Intra-DC

Bên cạnh dự đoán Intra_Angular nhắm tới các vùng với các biên có hướng mạnh mẽ, HEVC hỗ trợ hai phương pháp dự đoán khác, Intra_Planar và Intra_DC, 2 chế độ tương tự được quy định trong H.264/MPEG-4/AVC. Trong khi dự đoán Intra_DC sử dụng giá trị trung bình của các mẫu tham chiếu cho các dự đoán, giá trị trung bình của hai dự đoán tuyến tính sử dụng bốn mẫu tham chiếu góc được sử dụng trong dự đoán Intra_Planar để ngăn chặn sự gián đoạn dọc theo biên khối. Các chế độ dự đoán Intra_Planar được hỗ trợ ở tất cả các kích thước khối trong HEVC, trong khi

H.264/MPEG-4 AVC hỗ trợ dự đoán mặt phẳng chỉ khi PB thành phần chói là 16x16, và dự đoán mặt phẳng hoạt động hơi khác so với dự đoán mặt phẳng trong HEVC.

h) Dự đoán liên ảnh

Phân chia PB

So với các CB dự đoán trong ảnh, HEVC hỗ trợ nhiều hình dạng phân chia PB cho các CB dự đoán liên ảnh. Các chế độ phân chia của PART_2Nx2N, PART_2NxN, và PART_Nx2N chỉ ra các trường hợp khi CB không được chia, chia tách thành hai PB kích thước bằng nhau theo chiều ngang, và chia thành hai PBs kích thước bằng nhau theo chiều dọc. PART_NxN xác định rằng các CB được chia thành bốn PB kích thước bằng nhau, nhưng chế độ này chỉ được hỗ trợ khi kích thước CB bằng với kích thước CB cho phép nhỏ nhất. Ngoài ra, có bốn loại phân chia có hỗ trợ chia tách CB thành hai PB có kích cỡ khác nhau: PART_2NxN_U, PART_2NxN_D, PART_nLx2N, và PART_nRx2N. Những loại này được gọi là phân chia chuyển động không đối xứng.

Phép nội suy mẫu phân đoạn

Các phép nội suy mẫu phân đoạn cho mẫu chói trong HEVC sử dụng ứng dụng riêng biệt: một bộ lọc tám-chạm cho các vị trí $\frac{1}{2}$ -mẫu và một bộ lọc bảy-chạm cho các vị trí $\frac{1}{4}$ -mẫu. Điều này trái ngược với quá trình được sử dụng trong H.264/MPEG-4 AVC, ở đó áp dụng một quá trình nội suy hai giai đoạn, đầu tiên tạo ra các giá trị của một hoặc hai mẫu lân cận ở vị trí $\frac{1}{2}$ -mẫu sử dụng bộ lọc sáu-chạm, làm tròn kết quả trung gian, và sau đó tính trung bình hai giá trị tại vị trí số nguyên hoặc $\frac{1}{2}$ -mẫu. HEVC sử dụng một phép nội suy phân tách phù hợp duy nhất tạo ra tất cả các vị trí phân đoạn mà không có hoạt động làm tròn trung gian, trong đó cải thiện độ chính xác và đơn giản hóa kiến trúc của phép nội suy mẫu phân đoạn. Độ chính xác nội suy cũng được cải thiện trong HEVC bằng cách sử dụng các bộ lọc dài hơn, tức là, bộ lọc bảy chạm hoặc tám chạm chứ không phải là bộ lọc sáu-chạm được sử dụng trong H.264/MPEG-4/AVC. Bằng cách chỉ sử dụng bảy-chạm chứ không phải tám-chạm cho các vị trí $\frac{1}{2}$ -mẫu là đủ cho các vị trí phép nội suy $\frac{1}{4}$ -mẫu từ các vị trí $\frac{1}{4}$ mẫu là tương đối gần vị trí số nguyên-mẫu, vì vậy các mẫu xa nhất trong một phép nội suy tám-chạm sẽ hiệu quả xa hơn trong trường hợp $\frac{1}{2}$ -mẫu. Các giá trị chạm bộ lọc của bộ lọc phép nội suy được bắt nguồn một phần từ phương trình hàm cơ sở DCT.

Trong H.264/MPEG-4/AVC, tăng đến ba giai đoạn của hoạt động làm tròn để đạt được mỗi mẫu dự đoán (cho mẫu nằm tại các vị trí $\frac{1}{4}$ -mẫu). Nếu dự đoán đôi được sử dụng, tổng các hoạt động làm tròn là bảy trong trường hợp xấu nhất. Trong HEVC, nhiều

nhất hai hoạt động làm tròn để đạt được mỗi mẫu nằm tại các vị trí $\frac{1}{4}$ mẫu, năm hoạt động làm tròn là đủ trong trường hợp xấu nhất khi dự đoán đôi được sử dụng. Hơn nữa, trong sử dụng phổ biến nhất, nơi mà độ sâu bit $B=8b$, tổng của các hoạt động làm tròn trong trường hợp xấu nhất là tiếp tục giảm xuống còn 3. Do ít các hoạt động làm tròn, lỗi làm tròn số bị tích lũy là giảm và có tính linh hoạt hơn đến cách thức thực hiện các hoạt động cần thiết trong các bộ giải mã.

g) Mã hóa Entropy

HEVC chỉ rõ chỉ có một phương thức mã hóa Entropy, CABAC [2] không phải là hai như trong H.264 / MPEG-4 AVC. Các thuật toán cốt lõi của CABAC là không thay đổi, và các phần dưới đây trình diễn một số khía cạnh của nó trong HEVC.

Mô hình hóa ngữ cảnh

Lựa chọn phù hợp của mô hình hóa ngữ cảnh là một yếu tố quan trọng để nâng cao hiệu quả của mã hóa CABAC. Trong HEVC, độ sâu chia tách của cây mã hóa hoặc cây biến đổi được khai thác để dẫn xuất các chỉ số mô hình ngữ cảnh của các yếu tố cú pháp khác nhau, thêm vào những lân cận không gian sử dụng trong H.264/AVC. Mặc dù số lượng các ngữ cảnh sử dụng trong HEVC là ít hơn so với H.264/MPEG-4/AVC, thiết kế mã hóa Entropy thực sự cung cấp độ nén tốt hơn sẽ là một phần mở rộng thẳng của sơ đồ H.264/MPEG-4/AVC. Hơn nữa, việc sử dụng rộng rãi hơn được tạo ra trong HEVC của chế độ đường vòng của các hoạt động CABAC để tăng thông lượng bằng cách giảm lượng dữ liệu được mã hoá sử dụng ngữ cảnh CABAC. Phụ thuộc giữa các dữ liệu được mã hóa cũng được xem xét để cho phép tối đa hóa thông lượng hơn nữa.

Quét hệ số thích ứng

Quét hệ số được thực hiện trong các khối con 4×4 cho tất cả các kích cỡ TB. Ba phương pháp quét hệ số, đường chéo lên-bên phải, theo chiều ngang, và theo chiều dọc như Hình 2.14, được lựa chọn mặc nhiên để mã hóa các biến đổi hệ số của TB 4×4 và 8×8 trong vùng dự đoán trong ảnh. Việc lựa chọn các trình tự quét hệ số phụ thuộc vào các hướng của dự đoán trong ảnh. Quét theo chiều dọc được sử dụng khi hướng dự đoán gần với chiều ngang và quét chiều ngang được sử dụng khi các hướng dự đoán gần thẳng đứng. Quét đường chéo lên-bên phải được sử dụng cho các hướng dự đoán khác.

Mã hóa hệ số

Tương tự như H.264/MPEG-4/AVC, HEVC truyền đi vị trí của hệ số biến đổi khác-0 cuối cùng, 1 ánh xạ quan trọng, các bit dấu và các mức của các hệ số biến đổi. Tuy nhiên,

những thay đổi khác nhau cho từng phần đã được tạo ra, đặc biệt là xử lý tốt hơn việc tăng kích thước đáng kể các TB.

h) Bộ lọc trong vòng

Trong HEVC, hai bước xử lý là bộ lọc bỏ khối (bung khối) (DBF) tiếp theo sau là một bộ lọc bù thích ứng mẫu (SAO), được áp dụng cho các mẫu được khôi phục trước khi ghi chúng vào bộ đệm ảnh giải mã trong vòng lặp bộ giải mã. DBF làm giảm những thành phần khối do mã hóa dạng khối tạo ra. Các DBF tương tự như DBF của chuẩn H.264/MPEG-4/AVC, trong khi SAO mới được giới thiệu trong HEVC. Trong khi DBF chỉ được áp dụng cho các mẫu nằm ở biên khối, bộ lọc SAO được áp dụng thích nghi với tất cả các mẫu đáp ứng các điều kiện nhất định, ví dụ, dựa trên gradient. Trong sự phát triển của HEVC, nó cũng đã được xem xét để vận hành một bước xử lý thứ ba gọi là bộ lọc vòng lặp thích ứng (ALF) sau khi lọc SAO; Tuy nhiên, tính năng ALF chưa có trong thiết kế cuối cùng của HEVC.

Bộ lọc bỏ khối

Bộ lọc bỏ khối được áp dụng cho tất cả các mẫu liền kề với biên PU hoặc biên TU trừ trường hợp khi các biên cũng là một biên ảnh, hoặc khi bỏ khối bị vô hiệu hóa ngang qua các biên mảng hoặc Tile. Cần lưu ý rằng cả hai biên PU và TU cần được xem xét vì các biên PU không phải luôn luôn bằng nhau với biên TU trong một số trường hợp của các CB dự đoán liên ảnh. Các yếu tố cú pháp trong SPS và các điều khiển phần đầu mảng có hay chẳng bộ lọc bỏ khối được áp dụng qua các biên mảng và Tile.

Sức mạnh của bộ lọc bỏ khối được điều khiển bởi các giá trị của một số yếu tố cú pháp tương tự như trong H.264/MPEG-4AVC, nhưng chỉ có ba cái mạnh được sử dụng hơn là năm. P và Q cho trước là hai khối liền kề với một biên lưới 8x8 chung, sức mạnh bộ lọc bằng 2 được gán vào khi một trong những khối là dự đoán trong ảnh. Mặt khác, sức mạnh bộ lọc bằng 1 được gán vào nếu bất kỳ điều kiện sau đây được thỏa mãn:

- P hay Q có ít nhất một hệ số biến đổi khác-0
- Các chỉ số tham chiếu của P và Q không bằng nhau.
- Các vector chuyển động của P và Q không bằng nhau.
- Sự khác biệt giữa một thành phần vector chuyển động của P và Q là lớn hơn hoặc bằng một mẫu-số nguyên.

SAO

SAO là một quá trình sửa đổi các mẫu được giải mã bằng cách thêm một giá trị bù đắp vào mỗi mẫu sau ứng dụng các bộ lọc bỏ khối, dựa trên các giá trị trong bảng look-

up truyền bởi bộ mã hóa. Lọc SAO được thực hiện trên vùng, dựa trên một loại lọc lựa chọn mỗi CTB bởi một yếu tố cú pháp sao_type_idx. Giá trị 0 cho sao_type_idx chỉ ra các bộ lọc SAO không áp dụng cho CTB, và các giá trị 1 và 2 báo hiệu sử dụng lọc bù băng tần và bù cạnh.

2.2 So sánh hiệu năng nén của h.265 với một số chuẩn nén trước.

Về mặt lý thuyết H.265/HEVC được đánh giá là hiệu quả hơn H.264/AVC khoảng 25-35% . Ở độ phân giải càng cao, thì hiệu suất nén của H.265/HEVC càng cao. H.265/HEVC cơ bản có cấu trúc tương tự như các tiêu chuẩn trước đó như MPEG-2 và H.264/AVC.

- Chất lượng H.265/HEVC so với H.264/AVC, hình bên dưới là một ví dụ so sánh H.265/HEVC và H.264/AVC đây là hai khung hình video ở độ phân giải HD 720p. Bên trái là sử dụng chuẩn nén H.264, bên phải sử dụng chuẩn nén H.265 với bitrate 420kbps. Có thể thấy chất lượng của video H.265 tốt hơn video H.264 khi các vùng màu thể hiện rõ ràng và sắc nét hơn, vùng tóc cũng thể hiện mượt mà hơn.

- Thời gian tải video H.265 so với video H.264 với cùng một tốc độ mạng. Chúng ta có thể thấy rõ được video chuẩn nén H.265 lợi như nào.

- Kết quả trên hình chỉ ra rằng so với các chuẩn trước đây, H.265 cần số lượng bit ít hơn để mã hóa nhưng đồng thời vẫn đảm bảo chất lượng hình ảnh tốt hơn. Ví dụ, với cùng tốc độ bit là 1000kb/s, chất lượng đạt được của H.265 là 37.5dB trong khi của H.264 chỉ là 35dB.

- H265 nén mất khá nhiều thời gian hơn, tính toán lâu hơn. Tuy nhiên, bù lại, H265 có dung lượng nén gấp ~ 2.3 lần của H264.

2.3 Kết luận chương.

Trong phần này, các kỹ thuật giải mã video sẽ được nghiên cứu nhằm tìm ra bộ giải mã phù hợp nhất trong hệ thống video streaming. Cụ thể, các kỹ thuật giải mã video nói chung và các kỹ thuật tiên tiến được sử dụng trong chuẩn nén H.265 được giới thiệu. Ngoài ra, một số kỹ thuật giải mã trong truyền dẫn video qua mạng quang vô tuyến cũng đã được giới thiệu trong phần này. Các kỹ thuật này được đánh giá và so sánh với kỹ thuật trước đây như H.264 để thấy rõ các cải tiến của H.265. Cụ thể, các kỹ thuật áp dụng trong mỗi giai đoạn giải mã sẽ được so sánh và đánh giá ưu nhược điểm so với các kỹ thuật trong các chuẩn nén H.264 trước đây.

Các nghiên cứu trong phần này đã chỉ ra rằng, chất lượng video tại phía thu trong chuẩn nén H.265 có thể được cải thiện so với các chuẩn nén trước đây. Ngoài ra hiệu quả

nén của H.265 cũng được tăng lên đáng kể. Tuy nhiên, cấu trúc của bộ giải mã H.265 phức tạp hơn so với H.264. Điều này cũng dẫn đến thời gian giải mã của H.265 cũng tăng lên so với H.264. Tuy nhiên, trong một số chế độ mã hóa của H.265, việc giải mã đạt được hiệu quả cao nhưng thời gian giải mã không tăng lên nhiều so với các chuẩn trước đây. Vì vậy, việc lựa chọn bộ mã hóa H.265 để triển khai trong hệ thống truyền dẫn video thời gian thực qua môi trường quang vô tuyến là phù hợp.

CHƯƠNG 3. PHÂN TÍCH ĐÁNH GIÁ HIỆU NĂNG CỦA CHUẨN NÉN VIDEO HEVC TRÊN PHẦN MỀM HEVC

3.1. Giới thiệu phần mềm mô phỏng HM và các tham số mô phỏng.

Để đánh giá hiệu năng nén của chuẩn HEVC, luận văn sử dụng phần mềm mã nguồn mở HM (HEVC Test Model). Đây là phần mềm tham chiếu được triển khai tham chiếu tiêu chuẩn HEVC.

3.2. Các chuỗi video dùng để đánh giá.

Trong luận văn này sử dụng 9 chuỗi video chuẩn QCIF (Quarter Common Intermediate File Format) với nội dung chuyển động khác nhau để so sánh hiệu năng của các thuật toán trong các chế độ mã hóa khác nhau. Các chuỗi video này được phân loại thành 3 loại: Loại A, loại B, loại C với độ phức tạp chuyển động tăng dần. Các chuỗi video loại A có chuyển động thấp hay chậm. Các chuỗi loại B có chuyển động trung bình và chuỗi loại C có chuyển động phức tạp hay chuyển động nhanh. Loại A bao gồm các chuỗi: Akiyo, Hall và Mother-daughter. Các frame đầu tiên của chuỗi này được mô tả trong hình 3.2. Các frame đầu tiên của chuỗi loại B là Mobile, News, và Coastguard với chuyển động trung bình trình bày trong hình 3.3. Tương tự, hình 3.4 mô tả các frame đầu tiên của chuỗi thuộc loại C gồm Foreman, Carphone và Soccer với sự chuyển đổi đối tượng nhanh và tốc độ chuyển động cao.



Silent



Hall



Mother-daughter

Hình 3.2: Các frame đầu tiên của chuỗi Akiyo, Claire, Mother-daughter



Mobile



News



Coastguard

Hình 3.3: Các frame đầu tiên của chuỗi News, Hall, Coastguard



Foreman



Carphone



Soccer

Hình 3.4: Các frame đầu tiên của chuỗi Foreman, Carphone, Soccer

3.3. Các tham số đánh giá hiệu năng nén.

Để đánh giá hiệu năng của chuẩn nén HEVC đối với các chuỗi video có đặc tính khác nhau, hai tham số PSNR (Peak Signal-to-Noise Ratio) và số lượng bit dùng để mã hóa trong một giây (Bitrate) được sử dụng để so sánh hiệu năng nén. Tham số PSNR được sử dụng để đo mức độ giống nhau giữa hai chuỗi video trước khi nén và sau khi giải nén. Nói cách khác, PSNR phản ánh chất lượng của chuỗi video sau khi giải mã. Tham số này càng lớn đồng nghĩa với chất lượng chuỗi video sau khi giải nén càng giống với chất lượng chuỗi video gốc. Thông thường, chuỗi video có PSNR từ 30 dB trở lên sẽ được coi là có chất lượng tốt. PSNR được tính dựa trên công thức sau:

$$PSNR = 20 \cdot \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right) \quad (3.1)$$

Trong đó MAX_I là giá trị tối đa của điểm ảnh trên ảnh. Khi một điểm ảnh được biểu diễn bởi 8 bit thì giá trị của nó sẽ là 255. Tổng quát khi tín hiệu được biểu diễn bởi b bit trên một đơn vị lấy mẫu thì MAX_I bằng $2^b - 1$

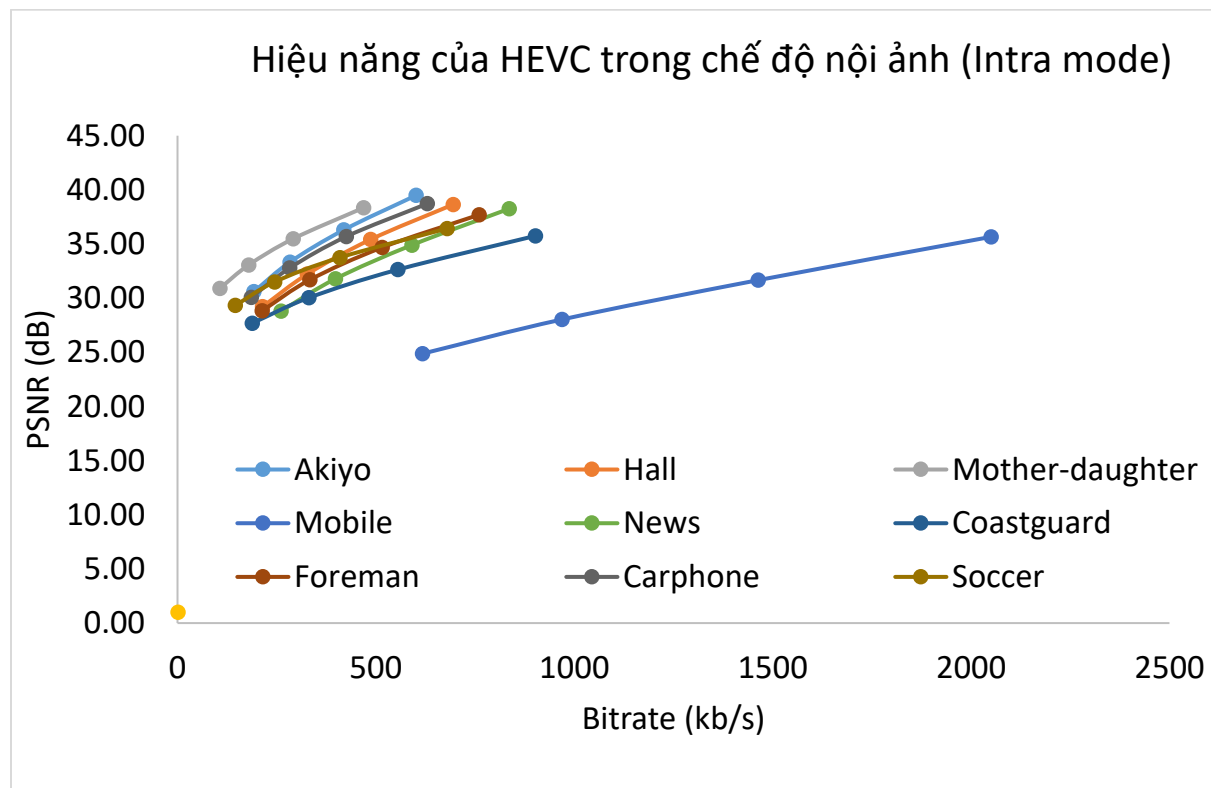
MSE (Mean Squared Error) là sai số toàn phương trung bình dùng cho ảnh hai chiều có kích thước là $M \times N$. Trong đó Org là ảnh gốc còn Rec là ảnh được khôi phục tương ứng. MSE được tính qua công thức:

$$MSE = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (Org(i,j) - Rec(i,j))^2 \quad (3.2)$$

Bitrate là tham số phản ánh số lượng bit trung bình dùng để mã hóa các khung hình video trong một giây. Bitrate càng lớn nghĩa là số bit dùng để mã hóa video lớn. Nếu cùng một giá trị PSNR, chuỗi video nào có bitrate càng nhỏ nghĩa là hiệu quả nén chuỗi video đó càng cao.

3.4. Phân tích kết quả mô phỏng.

3.4.1 Mã hóa video trong chế độ nội ảnh (Intra mode).



Hình 3.5. So sánh hiệu năng nén của HEVC cho các chuỗi khác nhau

Bảng 3.2. So sánh chất lượng ảnh (PSNR-dB) sau giải nén của các chuỗi với các tham số lượng tử (QP) khác nhau

	QP=28	QP=32	QP=36	QP=40
Akiyo	39,50	36,31	33,33	30,61
Hall	38,65	35,41	32,23	29,22
Mother-daughter	38,36	35,49	33,07	30,91
Mobile	35,66	31,66	28,03	24,88

News	38,26	34,90	31,79	28,82
Coastguard	35,75	32,63	30,04	27,69
Foreman	37,70	34,68	31,70	28,84
Carphone	38,72	35,70	32,79	30,07
Soccer	36,44	33,75	31,49	29,33

Bảng 3.3. So sánh Bitrate (Kbps) của các chuỗi với các tham số lượng tử (QP) khác nhau

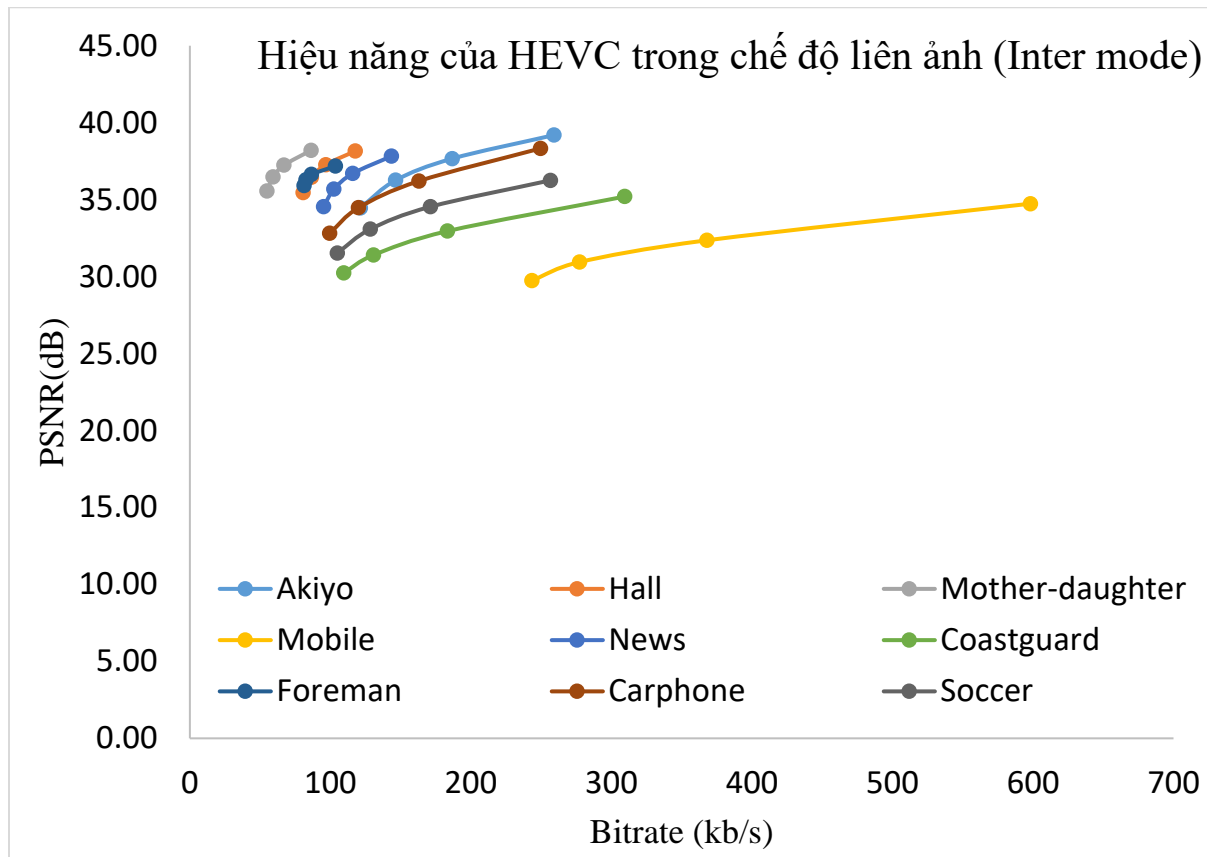
	QP=28	QP=32	QP=36	QP=40
Akiyo	601,77	419,36	283,3	192,72
Hall	695,09	486,5	327,44	214,55
Mother-daughter	469,45	291,81	179,67	106,68
Mobile	2050,54	1463,23	969,2	617,9
News	836,42	590,57	398,46	261,06
Coastguard	902,22	555,07	331,09	188,1
Foreman	760,5	515,8	333,8	213,2
Carphone	629,42	425,85	283,13	186,2
Soccer	679,98	409,08	244,58	145,88

Như kết quả đã chỉ ra trong hình 3.5 và các bảng 3.2, 3.3, ta thấy rằng trong chế độ nén nội ảnh các chuỗi video có chuyển động chậm và ít chuyển động (các chuỗi loại A) sẽ có chất lượng cao hơn trong khi cần ít bit hơn để mã hóa. Ngược lại, các chuỗi ở lớp B (có chuyển động phức tạp như Mobile) sẽ cần nhiều bit để mã hóa trong khi chất lượng hình ảnh đạt được không cao. Cụ thể, hình 3.2 cho thấy ứng với mỗi giá trị lượng tử QP, chuỗi Mobile có Bitrate lớn nhất trong khi PSNR đạt được là nhỏ nhất. Trong khi đó chuỗi Mother-daughter có Bitrate nhỏ nhất và PSNR đạt được chỉ sau chuỗi Akiyo. Như vậy, có thể thấy rằng, trong chế độ nội ảnh, các chuỗi có chuyển động chậm và ít chuyển động sẽ đạt được hiệu quả nén tốt nhất.

3.4.2 Mã hóa video trong chế độ liên ảnh (Inter mode).

Hình 3.6 và bảng 3.4, 3.5 chỉ ra hiệu năng nén của HEVC trong chế độ liên ảnh đối với các chuỗi video có đặc tính nén khác nhau. Bảng 3.5 cho thấy chuỗi Mobile vẫn là chuỗi cần nhiều bit để mã hóa nhất so với các chuỗi còn lại. Tuy nhiên, trong chế độ liên ảnh số lượng bit cần ít hơn nhiều so với chế độ nội ảnh. Bảng 3.4 cho thấy chất lượng trung bình của các video loại A vẫn cao hơn so với các video loại B & C. Với

những video chuyển động nhanh như Mobile và Coastguard thì chất lượng thu được sẽ kém nhất. Lý do là vì đối với loại video chuyển động nhanh, ước lượng chuyển động và bù chuyển động trong quá trình mã hóa sẽ kém chính xác so với video chuyển động chậm. Do đó, chất lượng ảnh tại phía giải mã sẽ thấp.



Hình 3.6. So sánh hiệu năng nén của HEVC cho các chuỗi khác nhau

Bảng 3.4. So sánh chất lượng ảnh (PSNR-dB) sau giải nén của các chuỗi với các tham số lượng tử (QP) khác nhau

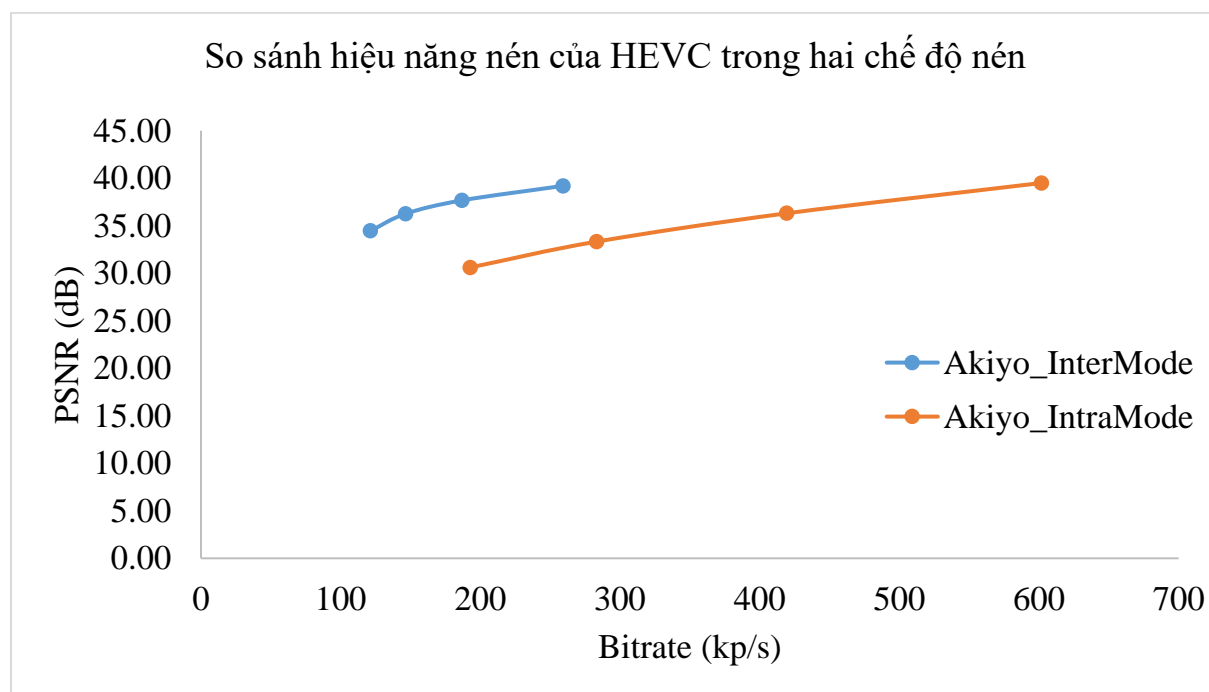
	QP=28	QP=32	QP=36	QP=40
Akiyo	39,20	37,67	36,26	34,46
Hall	38,15	37,27	36,45	35,45
Mother-daughter	38,21	37,24	36,47	35,57
Mobile	34,74	32,36	30,95	29,74
News	37,82	36,70	35,68	34,55
Coastguard	35,20	32,96	31,40	30,24
Foreman	37,18	36,64	36,29	35,92

Carphone	38,33	36,21	34,47	32,81
Soccer	36,25	34,54	33,08	31,54

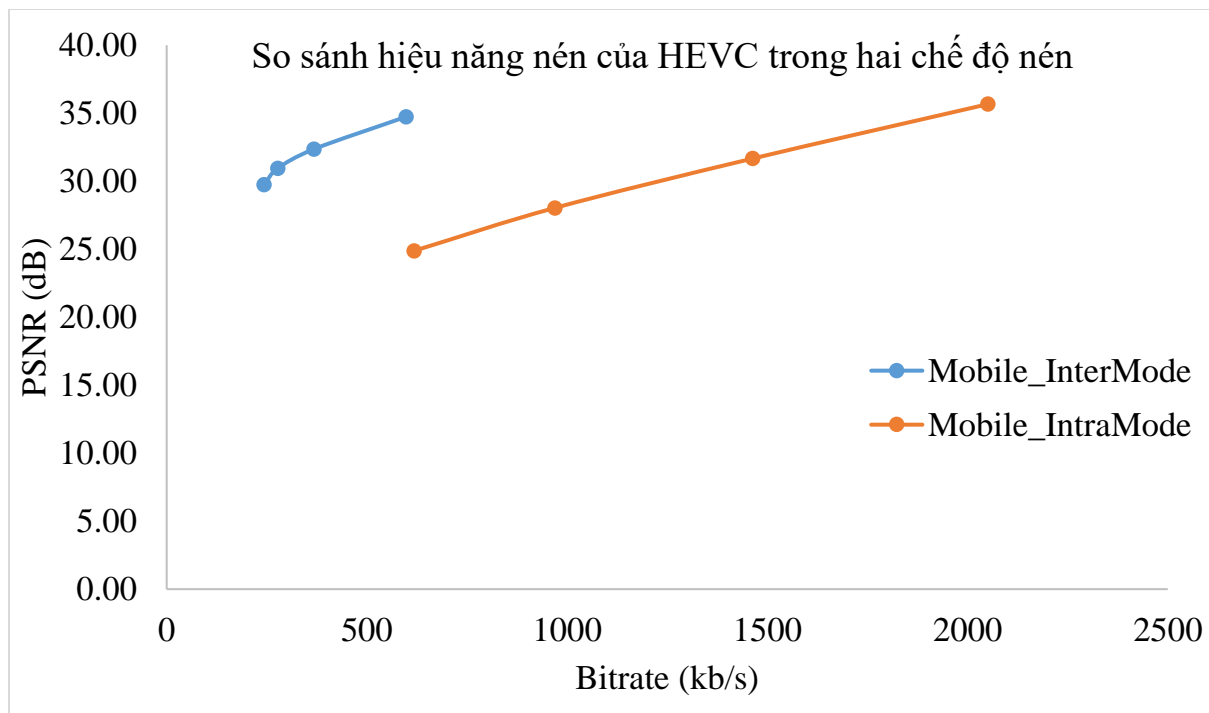
Bảng 3.5. So sánh Bitrate (Kbps) của các chuỗi với các tham số lượng tử (QP) khác nhau

	QP=28	QP=32	QP=36	QP=40
Akiyo	259,21	186,72	146,44	121,27
Hall	117,62	96,71	86,46	80,64
Mother-daughter	86,29	66,97	59,13	54,91
Mobile	598,23	367,88	277,28	243,45
News	143,45	115,9	102,53	95,15
Coastguard	309,35	183,32	130,7	109,55
Foreman	103,61	86,55	82,54	81,29
Carphone	249,49	163,1	120,04	99,39
Soccer	256,51	171,25	128,42	104,95

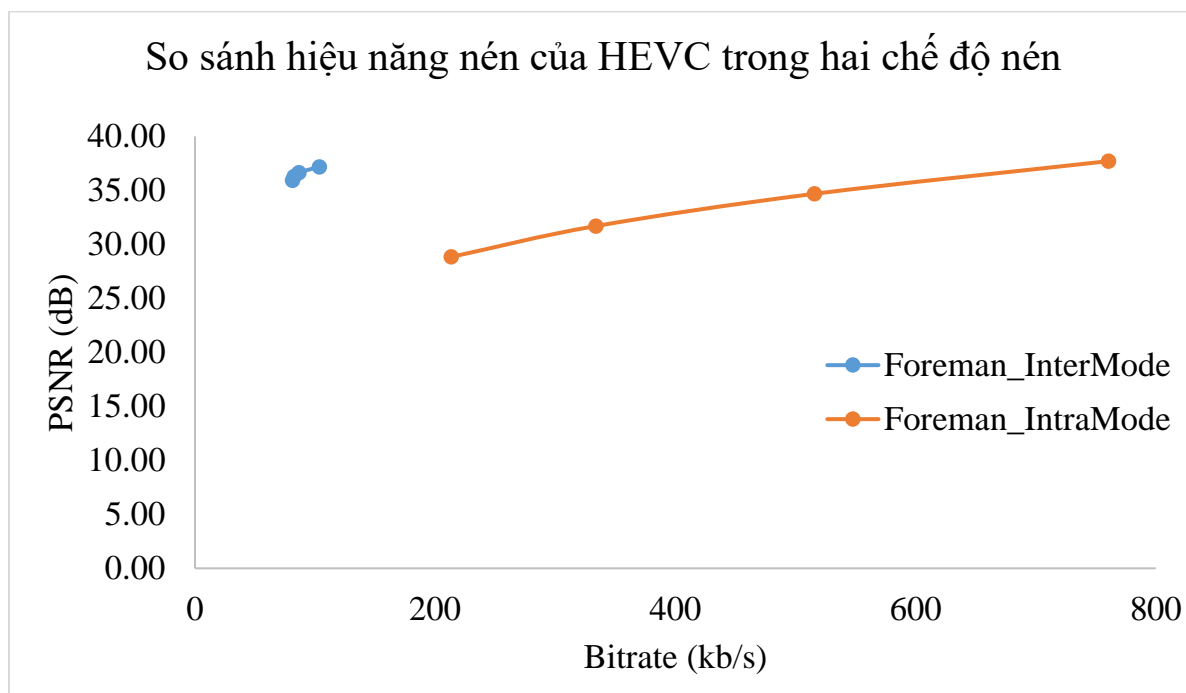
3.5. So sánh hai chế độ liên ảnh.



Hình 3.7. So sánh hiệu năng nén của HEVC trong hai chế độ nén của chuỗi Akiyo



Hình 3.8. So sánh hiệu năng nén của HEVC trong hai chế độ nén của chuỗi Mobile



Hình 3.9. So sánh hiệu năng nén của HEVC trong hai chế độ nén của chuỗi Foreman

Hình 3.7, 3.8 và 3.9 so sánh hiệu năng nén của HEVC trong hai chế độ liên ảnh và nội ảnh đối với 3 loại chuỗi: Ít chuyển động (Akiyo), chuyển động vừa (Mobile) và chuyển động nhanh (Foreman). Ta có thể thấy đối với chuỗi ít chuyển động, sự khác biệt giữa chế độ không lớn bằng hai loại chuỗi còn lại. Ngoài ra, trong chế độ nội ảnh (Intra mode), sự khác biệt về giá trị lượng tử sẽ dẫn đến sự thay đổi rõ rệt về lượng bit cần để mã hóa và chất lượng ảnh tại phía giải mã.

3.6. Chế độ mã hóa Inter với kích thước chuỗi khác nhau

Để kiểm tra chế độ mã hóa với các kích thước chuỗi khác nhau, bốn chuỗi video được sử dụng để đánh giá bao gồm Coastguard, Suzie, Pamphlet và Harbor với các tham số được tóm tắt trong bảng 3.5. Khung hình đầu tiên của bốn chuỗi video này được hiển thị trong hình 3.5.

Test sequences	Spatial resolution	Number of frames	Quantization parameters
Coastguard	176x144	300	{26,30,34,38}
Suzie		150	{25,29,34,40}
Pamphlet		150	{25,29,34,40}
Harbour		150	{25,29,34,40}

Bảng 3. 6. Các tham số của chuỗi video thử nghiệm



Hình 3.10 Frame đầu tiên của các chuỗi video thử nghiệm

Để đánh giá thuật toán được đề xuất, chỉ số BD-PSNR (Bjntegaard-Delta PSNR) được sử dụng để so sánh, với BD-PSNR để cung cấp độ lợi tương đối giữa hai phương pháp, bằng cách đo sự khác biệt trung bình giữa hai đường cong đánh giá tỷ lệ sai lệch RD (RD – Rate-Distortion) với đường cong RD được chọn làm đường cong cơ sở. Nếu BD-PSNR là dương, nó có nghĩa là đường cong thứ hai tốt hơn đường cong cơ sở. Trong đánh giá này, các đường cong RD của GOP4 và phương pháp được đề xuất có tên là Adaptive GOP được so sánh với đường cong cơ sở GOP2.

Kết quả đánh giá hiệu năng về tỷ lệ sai lệch RD cho bốn chuỗi video thử nghiệm được trình bày trong bảng 3.7.

Theo kết quả biểu diễn trong bảng 3.7, các giá trị PSNR của phương pháp đề xuất tốt hơn các giá trị của GOP 4 và tương đương với các giá trị của GOP2. Giá trị bitrate của phương pháp được đề xuất cao hơn giá trị của GOP4 và thấp hơn giá trị của GOP2. Do đó, việc lựa chọn giữa GOP2 và GOP4 phụ thuộc vào sự cân bằng giữa PSNR và Bitrate. Kết quả cho thấy mức độ suy giảm chất lượng của video (về giá trị PSNR) trong phương pháp đề xuất là không đáng kể trong khi tiết kiệm được khá nhiều bitrate. Bảng 3.2 cho thấy khả năng tối ưu bitrate của phương pháp đề xuất là 3,37% và 9,62% so với GOP2 và GOP4, tương ứng.

Sequence	QP	GOP2		GOP4		Adaptive GOP	
		Bitrate	PSNR	Bitrate	PSNR	Bitrate	PSNR
Coastguard	26	27760	38.18	28242	34.65	27735	38.14
	30	17131	34.87	16140	32.48	17058	34.84
	34	9838	31.88	8228	30.36	9760	31.85
	38	5256	29.14	3781	28.23	5199	29.12
Average		14996.25	33.52	14097.75	31.43	14938	33.49
Suzie	26	18424	41.58	19719	41.26	18565	41.34
	30	10869	38.56	11172	38.23	10530	38.26
	34	5725	35.41	5588	35.15	5283	35.29
	38	2667	32.24	2353	32.04	2270	32.19
Average		9421.25	36.95	9708	36.67	9162	36.77
Pamphlet	26	23893.93	41.15	23128.28	41.35	22453.65	41.37
	30	15669.9	37.42	14900.7	37.51	14504.5	37.56
	34	9013.55	33.18	8567.73	33.24	8349.78	33.29
	38	3897.73	28.86	3667.88	28.91	3587.02	28.95
Average		13118.78	35.15	12566.15	35.25	12223.74	35.29
Harbour	26	45656.58	38.04	45680.28	37.62	45337.92	37.81
	30	29713.93	34.18	28617.86	33.73	28830.11	33.96
	34	16805.14	30.36	15471.99	30.03	15889.86	30.23

	38	7646.22	26.24	6768.94	26.09	7082.92	26.22
Average		24955.47	32.2	24134.77	31.86	24285.2	32.06

Bảng 3.7 Hiệu năng của các chuỗi video thử nghiệm trong thuật toán đề xuất

3.7. Kết luận chương

Chương 3 đã trình bày quá trình mô phỏng đo và đánh giá hiệu năng nén của chuẩn nén video HEVC đối với các chuỗi video có đặc tính khác nhau. Phần mềm mô phỏng được sử dụng là phần mềm HM của Nhóm hợp tác chung về mã hóa video (JCT-VC) từ các chuyên gia ITU-T SG 16 và ISO / IEC SC29 WG11. Các chuỗi video được thử nghiệm là các chuỗi video có các đặc tính thay đổi từ chuyển động chậm đến chuyển động nhanh, từ ít chuyển động đến nhiều chuyển động. Kết quả mô phỏng cho thấy chuẩn nén HEVC đạt được hiệu quả cao nhất đối với các chuỗi chuyển động chậm và ít chuyển động cả về chất lượng và hiệu năng nén.

KẾT LUẬN

Luận văn đã tìm hiểu tổng quan về mã hóa và giải mã video các kỹ thuật giải mã video nói chung và các kỹ thuật tiên tiến được sử dụng trong chuẩn nén H.265. Các kỹ thuật này được đánh giá và so sánh với kỹ thuật trước đây như H.264 để thấy rõ các cải tiến của H.265. Cụ thể, các kỹ thuật áp dụng trong mỗi giai đoạn giải mã sẽ được so sánh và đánh giá ưu nhược điểm so với các kỹ thuật trong các chuẩn nén H.264 trước đây.

Để đánh giá hiệu năng nén của chuẩn HEVC, luận văn sử dụng phần mềm mã nguồn mở HM. Các nghiên cứu sử dụng phần mềm mô phỏng trong chương 3 đã chỉ ra rằng, chất lượng video tại phía thu trong chuẩn nén H.265 có thể được cải thiện so với các chuẩn nén trước đây. Ngoài ra hiệu quả nén của H.265 cũng được tăng lên đáng kể. Tuy nhiên, cấu trúc của bộ giải mã H.265 phức tạp hơn so với H.264. Điều này cũng dẫn đến thời gian giải mã của H.265 cũng tăng lên so với H.264. Tuy nhiên, trong một số chế độ mã hóa của H.265, việc giải mã đạt được hiệu quả cao nhưng thời gian giải mã không tăng lên nhiều so với các chuẩn trước đây.

TÀI LIỆU THAM KHẢO

Tiếng Việt

- [1] Vũ Hữu Tiến, Hà Đình Dũng, “Bài giảng: Xử lý và Truyền thông Đa phương tiện”, Học viện Công nghệ Bưu chính Viễn thông, 2016.
- [2] Đỗ Hoàng Tiến, Dương Thanh Phương, “Kỹ thuật truyền hình”, NXB Khoa học và Kỹ thuật 2004.
- [3] Luận văn Vũ Thị Quỳnh Anh, “Triển mô hình tập âm tương quan nhằm nâng cao hiệu năng mã hóa của kỹ thuật mã hóa Video phân tán DVC” Trang 13-23, 2019

Tiếng Anh

- [4] J. R. Jain, A. K. Jam, "Displacement measurement and its application in interframe image coding", IEEE Trans. Commun., vol. COM-29, pp. 1799-1808, Dec. 1981.
- [5] Mohammed Ghanbari, “ Video coding: an introduction to standard codecs”, The Institution of Electrical Engineers, 1999.
- [6] R. Srinivasan, K.R. Rao, “Predoctove coding based on efficient motion estimation”, IEEE Trans. Commun., pp. 888-896, 2012.
- [7] R. Li, B. Zeng, M. L. Liou, "A new three-step search algorithm for block motionestimation", IEEE Trans. Circuits Syst. Video Technol., vol. 4, no. 4, pp. 438-442, Aug. 1994.
- [8] JVT Editors (T. Wiegand, G. Sullivan, A. Luthra), Draft ITUT Recommendation and final draft international standard of joint video specification (ITU-T Rec.H.264 |ISO/IEC 14496- 10 AVC), JVT-G050r1, Geneva, May 2003.
- [9] L. Natrio, C. Brites, J. Ascenso, and F. Pereira, Side information extrapolation for low-delay pixel-domain distributed video coding, International Workshop on Very Low Bitrate Video, 2005.
- [10] A. Aaron and B. Girod, , Wyner-Ziv video coding with low-encoder complexity, Picture Coding Symposium, 2004.
- [11] [5] D. Slepian and J. Wolf, Noiseless Coding of Correlated Information Sources, IEEE Transactions on Information Theory, vol. 19, no. 4, pp.471-480, July 1973.
- [12] A. Wyner and J. Ziv, The Rate-Distortion Function for Source Coding with Side Information at the Decoder, IEEE Transactions on Information Theory, vol. 22, no. 1, pp.1-10, January 1976.

- [13] R. Puri and K. Ramchandran, PRISM: A new robust video coding architecture based on distributed compression principles, 40th Allerton Conf. Communication, Control and Computing,, Allerton, IL, USA, 2002.
- [14] A. Aaron, R. Thang, and B. Girod (2002) “Wyner-Ziv Coding of Motion Video”, in Proc. Asilomar Conference on Signals and Systems, Pacific Grove, CA, USA, November 2002.

Trang Web

- [15] <https://hevc.hhi.fraunhofer.de/>