

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

TRẦN DUY XUYÊN

**ĐÁNH GIÁ, PHÂN TÍCH VÀ SO SÁNH
HIỆU SUẤT CỦA HAI BỘ MÃ HOÁ
VIDEO H.265 VÀ H.264.**

**LUẬN VĂN THẠC SĨ
NGÀNH CÔNG NGHỆ KỸ THUẬT ĐIỆN TỬ, TRUYỀN THÔNG**

HÀ NỘI - 2016

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

TRẦN DUY XUYÊN

**ĐÁNH GIÁ, PHÂN TÍCH VÀ SO SÁNH
HIỆU SUẤT CỦA HAI BỘ MÃ HOÁ
VIDEO H.265 VÀ H.264.**

Ngành: Công nghệ Kỹ thuật Điện tử, Truyền thông
Chuyên ngành: Kỹ thuật Điện tử
Mã số: 60520203

**LUẬN VĂN THẠC SĨ
NGÀNH CÔNG NGHỆ KỸ THUẬT ĐIỆN TỬ, TRUYỀN THÔNG**

NGƯỜI HƯỚNG DẪN KHOA HỌC: TS. LÊ VŨ HÀ

HÀ NỘI - 2016

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn tốt nghiệp “*Đánh giá, phân tích và so sánh hiệu suất của hai bộ mã hoá video H.265 và H.264.*” là công trình nghiên cứu khoa học của riêng tôi được thực hiện dưới sự hướng dẫn của TS Lê Vũ Hà. Luận văn tốt nghiệp là kết quả của quá trình nghiên cứu độc lập, không sao chép công trình nghiên cứu của bất kỳ ai khác. Các số liệu trong luận văn được sử dụng trung thực, trích dẫn từ những nguồn hợp pháp và đáng tin cậy.

Hà Nội, ngày 19 tháng 07 năm 2016

Người thực hiện

Trần Duy Xuyên

MỤC LỤC

MỤC LỤC	1
DANH MỤC BẢNG BIỂU	3
DANH MỤC HÌNH VẼ	4
DANH MỤC THUẬT NGỮ VIẾT TẮT	5
TÓM TẮT.....	7
Chương 1. Giới thiệu	8
Chương 2. Tổng quan về H.265/HEVC	11
2.1. Cấu trúc bậc cao của H.265/HEVC	11
2.1.1. NAL và các loại hình ảnh	11
2.1.2. Các tập tham số.....	14
2.2. Cấu trúc mã hoá	15
2.3. Cấu trúc khối và kỹ thuật song song	16
2.3.1. Phân vùng khối	16
2.3.2. Phân vùng hình ảnh.....	21
2.3.3. Xử lý song song trong H.265/HEVC.....	22
2.4. Kết luận	27
Chương 3. Dự đoán nội ảnh và dự đoán liên ảnh.....	28
3.1. Dự đoán nội ảnh	28
3.1.1. Thiết lập mẫu tham chiếu	29
3.1.2. Dự đoán mẫu nội ảnh.....	33
3.1.3. Chế độ mã hóa nội ảnh	36
3.2. Dự đoán liên ảnh	38
3.2.1. Dự đoán vector chuyển động tiên tiến (AMVP).....	39
3.2.2. Nhập khối dự đoán liên ảnh.....	40
3.2.3. Nội suy mẫu phân số.....	41
3.3. Bộ lọc vòng trong.....	43
3.4. Kết luận	45
Chương 4. Những sửa đổi đề xuất về dự đoán nội ảnh	46
4.1. Chế độ dự đoán nội ảnh cơ bản trong H.265/HEVC	46
4.2. Thuật toán chọn chế độ nội ảnh nhanh	48

4.3. Quá trình dự đoán nội ảnh nhanh bằng kỹ thuật song song	49
Chương 5. Tổng kết.....	52
5.1. Kết luận	52
5.3. Hướng phát triển tiếp theo	53
DANH MỤC TÀI LIỆU THAM KHẢO	54

DANH MỤC BẢNG BIỂU

Bảng 2-1: Các loại gói Non-VCL NAL[1]	13
Bảng 2-2: Các loại gói VCL NAL[1].....	13
Bảng 2-3: Kích thước khối bù chuyển động đã được hỗ trợ trong H.265/HEVC và trong kỹ thuật trước đây[1].....	20
Bảng 3-1: Sự khác nhau cơ bản của dự đoán nội ảnh giữa H.265/HEVC và H.264/AVC[1]	29
Bảng 3-3: Giá trị của tham số B[1]	34
Bảng 3-2: Giá trị của tham số A[1].....	34
Bảng 3-4: Chế độ dự đoán nội ảnh màu dựa theo chế độ dự đoán nội ảnh[1]	38

DANH MỤC HÌNH VẼ

Hình 2-1: Cấu trúc tiêu đề gói NAL [1].....	12
Hình 2-2: Các tập tham số trong H.265/HEVC[1]	14
Hình 2-3: Cấu trúc mã hóa video của H.264/AVC.....	15
Hình 2-4: Cấu trúc mã hóa video Hybrid của phiên bản H.265/HEVC đầu tiên	15
Hình 2-6: Ví dụ về phân vùng hình ảnh thành nhiều CTU 64x64[1]	17
Hình 2-5: Ví dụ phân vùng hình ảnh thành nhiều khối macro 16x16[1].....	17
Hình 2-7: Ví dụ về phân vùng CTU theo thứ tự chiều sâu	18
Hình 2-8: Tất cả các kích thước của PU trong H.265/HEVC[1]	19
Hình 2-9: Ví dụ về chia một CTB thành nhiều TBs.....	20
Hình 2-10: Mối quan hệ giữa CU, PU và TU trong H.265/HEVC	21
Hình 2-11: Cấu trúc các mảnh độc lập trong H.265/HEVC[1]	21
Hình 2-12: Ví dụ về sử dụng tile chia hình ảnh thành 9 vùng, đường gạch thể hiện ranh giới các vùng. [1].....	25
Hình 2-13: Quy trình của tiles	25
Hình 2-14: Xử lý song song sóng trước	26
Hình 3-1: Ví dụ về các chế độ dự đoán nội ảnh của H.265/HEVC[1]	29
Hình 3-2: Quá trình thay thế mẫu (a) Mẫu tham chiếu trước khi thực hiện quá trình, các mẫu không có sẵn ký hiệu bằng màu xám	30
(b) Mẫu tham chiếu sau khi thực hiện quá trình[1]	30
Hình 3-3: Quá trình lọc mịn (a) Bước thứ nhất (b) bước thứ hai[1].....	32
Hình 3-4: Các chế độ góc trong dự đoán nội ảnh [1]	33
Hình 3-5: Ví dụ về việc đổi chỗ mẫu tham chiếu bên trái để mở rộng hàng tham chiếu dọc ở chế độ nội ảnh 23[1]	35
Hình 3-6: Ví dụ về việc sử dụng chế độ dự đoán planar nội ảnh	36
(a) Tính thành phần ngang (b) Tính thành phần dọc	36
(c) giá trị trung bình của (a) và (b) [1].....	36
Hình 3-7: Khái niệm cơ bản về dự đoán liên ảnh[1]	38
Hình 3-8: Quá trình dự đoán liên ảnh trong H.265/HEVC[1].....	39
Hình 3-9: Các khối có các ứng viên vector chuyển động[1].	40
Hình 3-10: (a) Các ứng viên có thể chọn trong danh sách sát nhập của khối X, chúng sáp nhập thành cùng dòng đậm, (b) Các bước kiểm tra dư để thêm ứng viên vào danh sách. [1]	41
Hình 3-11: Vị trí mẫu và phân số để nội suy [1]	42
Hình 3-12: De-blocking filter và SAO trong H.265/HEVC[1]	43
Hình 3-13: 4 mẫu và vị trí của chúng giữa vùng ranh giới của 2 khối P và Q[1]	44
Hình 4-1: Quá trình quyết định chế độ dự đoán nội ảnh trong H.265/HEVC[5]	46
Hình 4-2: Quá trình lựa chọn chế độ nội ảnh đã được sửa đổi.....	49
Hình 4-3: WPP bước thứ nhất	50
Hình 4-4: Luồng 4 bắt đầu quá trình đồng thời với bước đầu tiên.....	50
Hình 4-5: Luồng 4, mã hóa entropy đang chạy theo thứ tự quét.....	51

DANH MỤC THUẬT NGỮ VIẾT TẮT

A	
AMVP: Advance Motion Vector Prediction	Dự đoán vector chuyển động tiên tiến
AVC: Advance Video Coding (H.264/AVC)	Nén video tiên tiến
C	
CTU: Coding Tree Unit	Đơn vị mã hóa cây
CTB: Coding Tree Block	Khối mã hóa cây
CU: Coding Unit	Đơn vị mã hóa
CB: Coding Block	Khối mã hoá
CRA: Clean Random Access	Truy cập ngẫu nhiên
H	
HEVC: High Efficiency Video Coding	Mã hóa video hiệu suất cao
IDR: Instantaneous Decoding Refresh	Giải mã tức thời
IRAP: Intra Random Access Point	Điểm truy cập ngẫu nhiên nội ảnh
N	
NAL: Network Abstract Layer	Lớp trừu tượng mã hóa mạng
M	
MPM: Most Probable Mode	Chế độ dự đoán đúng nhất

P	
PB: Prediction Block	Khối dự đoán
PU: Prediction Unit	Đơn vị dự đoán
S	
STSA: Step-wise Temporal Sublayer Access	Truy cập tầng phụ tạm thời
T	
TB: Transform Block	Khối biến đổi
TU: Transform Unit	Đơn vị biến đổi
V	
VCL: Video Coding Layer	Lớp mã hóa video
W	
WPP: Wave-front Parallel Processing	Kỹ thuật xử lý song song sóng trước

TÓM TẮT

Video nếu không được nén sẽ có độ phân giải và chất lượng cao nhưng không thể chuyển tin hiệu qua mạng truyền thông bởi vì dung lượng quá lớn, trong khi nhu cầu xem, sử dụng video ngày càng cao. Do đó, các kỹ thuật mã hóa nén video đã được phát triển và cải thiện trong nhiều thập kỷ để khắc phục vấn đề này, giảm thiểu dung lượng video, giúp cho việc truyền dẫn video qua mạng đơn giản hơn. Hiện nay, H.264/AVC (Advance Video Coding - Nén video tiên tiến) là tiêu chuẩn nén mã hóa video được sử dụng phổ biến trên toàn thế giới. Tuy nhiên, nhu cầu về video chất lượng cao hơn, tốc độ khung hình cao hơn, chẳng hạn như video HD, UltraH, 4K và 8K, thì H.264/AVC vẫn chưa có khả năng nén tốt nhất. Vì vậy, một chuẩn nén video mới đã được phát triển, phiên bản đầu tiên của H.265/HEVC đã được công bố vào tháng 1 năm 2013, không chỉ có khả năng mã hoá tốt những video có độ phân giải cao mà còn giảm được dung lượng video mã hoá một nửa so với chuẩn nén H.264/AVC. Mục tiêu của luận án là tìm hiểu kỹ thuật nén video mới nhất hiện nay, đó là H.265/HEVC (High Efficiency Video Coding - Mã hóa video hiệu suất cao). Luận án sẽ đưa ra những so sánh về kỹ thuật nén H.265/HEVC và H.264/AVC, để biết lý do tại sao hiệu quả nén của H.265/HEVC tốt hơn 50% so với chuẩn nén trước đó (H.264/AVC). Ngoài ra, H.265/HEVC là một chuẩn nén mới và còn rất nhiều kỹ thuật có thể cải tiến tốt hơn, nên thông qua quá trình tìm hiểu so sánh, tôi sẽ tìm một số phương pháp để cải thiện hiệu suất hoặc giảm độ phức tạp, đặc biệt là trong kỹ thuật dự đoán nội ảnh của H.265/HEVC.

Chương 1

Giới thiệu

Ngày nay, nhu cầu sử dụng những dịch vụ truyền thông đa phương tiện ngày càng cao, đặc biệt là video. Cùng với đó, những yêu cầu về chất lượng, tốc độ truyền tải của video cũng ngày một khắt khe hơn. Từ đó, một vấn đề phát sinh, đó là vấn đề về dung lượng của một video. Video càng có độ phân giải cao, chất lượng cao thì đi kèm với nó là dung lượng cao. Một video chưa qua xử lý nén có dung lượng rất lớn. Ví dụ chỉ với video có độ phân giải là 720x576, 25 fps, độ sâu màu 8 bit cần tốc độ bit:

$$\begin{aligned} &720 \times 576 \times 25 \times 8 + 2 \times (360 \times 576 \times 25 \times 8) \\ &= 1,66 \text{ Mb/s (độ sáng + sắc độ)} \end{aligned}$$

Hay, với HDTV, một video có độ phân giải cao là 1920x1080, độ sâu ảnh 8 vắn 25fps, thì tốc độ bit là:

$$1920 \times 1080 \times 60 \times 8 + 2 \times (960 \times 1080 \times 60 \times 8) = 1,99 \text{ Gb/s}$$

Vậy nên cho dù là những hệ thống máy tính mạnh nhất, nếu không phải vấn đề băng thông thì quản lý lượng dữ liệu video lớn cũng là một vấn đề khó khăn. Tuy nhiên, video là một chuỗi ảnh mà sự khác nhau giữa các ảnh là rất nhỏ, do đó tồn tại một lượng thông tin thừa rất lớn nếu video được truyền tải toàn bộ. Thay vì truyền hai ảnh độc lập, ta có thể truyền chỉ một ảnh và sự khác nhau của nó với ảnh còn lại, qua đó giảm lượng dữ liệu cần truyền, giảm tốc độ bit. Nén video ra đời nhằm giải quyết vấn đề dung lượng và truyền tải video trên. Phát triển qua nhiều thập kỉ, nhiều chuẩn

nén video đã được công bố và sử dụng. Một số chuẩn nén video là H.261, H.262/MPEG-1, H.263, MPEG-4, và H.264/AVC.

Nén video đã phát triển được trong một thời gian dài, giúp giảm đáng kể dung lượng của một video, tăng khả năng truyền video qua các kênh truyền có băng thông khác nhau. Đến nay, H.264/AVC (Advance Video Coding), có tên đầy đủ là MPEG-4 Part 10, Advanced Video Coding (MPEG-4 AVC) và là kỹ thuật nén phổ biến nhất.

Mục tiêu của H.264/AVC là tập trung vào sự cải tiến trong mã hóa, kiến trúc dữ liệu và sự linh hoạt. H.264/AVC đã đạt được khả năng mã hóa tốt hơn so với chuẩn nén trước đó là H.263 và MPEG-2, giảm một nửa lượng bit-rate cần thiết để hiển thị một video trong khi người dùng cảm nhận về chất lượng video vẫn được đảm bảo. Hiểu theo một cách khác, đó là khả năng tăng gấp đôi số lượng kênh truyền video trong một mạng truyền video quảng bá. Video được nén qua H.264/AVC có thể truyền trong các kênh truyền khác nhau về các độ trễ cũng như khả năng mất dữ liệu. Ngoài ra, những video đó có khả năng truyền real-time (thời gian thực) qua mạng và người xem có thể truy cập ngẫu nhiên mà không cần phải truy cập từ đầu video.

Nhu cầu về xem, thưởng thức các video có chất lượng cao ngày càng lớn, chẳng hạn như độ phân giải cao (HD, ULTRA HD, 4K, 8K), tốc độ khung hình cao hơn và độ trung thực cao hơn, trong khi không có gì thay đổi về khả năng truyền tín hiệu qua mạng. Bên cạnh đó, video hiện nay ngày càng được sử dụng trong nhiều ứng dụng khác nhau, như ứng dụng di động, chat video theo thời gian thực, camera di động. Do đó, video có lượng truyền tải lớn nhất trên các mạng truyền thông, vì vậy chúng ta cần một chuẩn nén mới tốt hơn.

Mã hóa video hiệu suất cao (H.265/HEVC) là tiêu chuẩn nén mới nhất, nhằm mục đích cải thiện đáng kể hiệu suất nén, độ phân giải của tất cả các loại video, và các ứng dụng đa phương tiện. Chính thức ra phát triển từ tháng 1 năm 2010 với tên gọi là HEVC - High Efficiency Video Coding (Mã hóa video hiệu suất cao), phiên bản đầu tiên đã được hoàn thành, quy chuẩn H.265 và công bố vào tháng 1 năm 2013, bởi Nhóm các chuyên gia hình ảnh động ISO/IEC (MPEG) và Nhóm các chuyên gia mã

hóa video ITU-T (VCEG), những người đã phát triển Mã hóa video tiên tiến (H.264/AVC). H.264/AVC hiện tại đang thống trị các kỹ thuật mã hóa video trên toàn thế giới. Hiệu suất nén video tốt hơn khoảng 50% so với tiêu chuẩn trước (H.264/AVC), có nghĩa là tăng gấp đôi khả năng nén, hay giảm một nửa tốc độ bit trong khi dung lượng của video vẫn không đổi, qua đó H.265/HEVC sẽ sớm được sử dụng phổ biến. Hiệu suất nén video của H.265/HEVC cao, cho phép truyền các video chất lượng cao hơn, độ phân giải, tỷ lệ khung hình với dải động màu cao hơn. Do chỉ cung cấp hiệu suất mã hóa cho các bộ mã hóa, nên nó không đảm bảo video có chất lượng cụ thể, vì vậy H.265/HEVC hiện giờ vẫn còn trong quá trình phát triển và tối ưu hóa.

Mục đích của luận án này là thể hiện những khía cạnh quan trọng nhất và các thông tin, kỹ thuật quan trọng về H.265/HEVC, sau đó phân tích và so sánh với các tiêu chuẩn mã hóa trước đó, cụ thể là H.264/AVC để biết lý do tại sao H.265/HEVC có thể nén với hiệu suất cao như vậy. Nội dung luận án được bố trí như sau:

- Chương 2 mô tả nền tảng của H.265/HEVC và so sánh các tính năng chung về cấu trúc block với H.264/AVC và kỹ thuật song song mới.
- Chương 3 giải thích và so sánh kỹ thuật dự đoán nội ảnh và dự đoán liên ảnh trong H.265/HEVC, hai phần quan trọng nhất trong mã hóa nén video.
- Chương 4 giải thích một số kỹ thuật được đề xuất nhằm cải thiện phương pháp dự đoán intra prediction của HEVC.
- Chương 5 kết luận những mục tiêu đã đạt được sau khi hoàn thành khoá luận và những hướng phát triển sắp tới, có thể mở rộng.

Tổng quan về H.265/HEVC

2.1. Cấu trúc bậc cao của H.265/HEVC

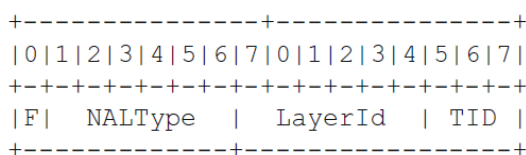
Cấu trúc, các thuật toán, kỹ thuật của H.265/HEVC được mở rộng, phát triển dựa trên chuẩn mã hóa H.264/AVC (Mã hóa video tiên tiến) trước đây. Nó hỗ trợ giao diện mạng, hệ thống, và tầng mã hóa video để truyền tín hiệu video. Các tập tham số chứa dữ liệu bậc cao, có thể truy cập dễ dàng khi mã hoá toàn bộ video, nhóm các hình ảnh liên tiếp nhau hoặc từng slice bên trong một bức ảnh. Tất cả các dữ liệu video đã mã hóa và giá trị của các tham số sử dụng trong quá trình mã hóa được đóng gói và lưu trong các gói tin của lớp trừu tượng hóa mạng (gói NAL – Network Abstraction Layer) trước khi truyền đến bộ giải mã. Từ H.264/AVC tới nay là H.265/HEVC, NAL được thiết kế “thân thiện với đường truyền”, nó có chức năng truyền tải dữ liệu dạng video trên các kênh truyền khác nhau, cho các ứng dụng khác nhau. Mỗi gói NAL chứa một số lượng byte hạn chế. Dựa trên 2 byte đầu tiên ở tiêu đề gói NAL, người nhận có thể xác định loại dữ liệu có trong gói NAL.

2.1.1. NAL và các loại hình ảnh

Trong H.265/HEVC, có tổng cộng 64 loại gói NAL, được chia thành hai loại: mã hóa video (VCL NAL) và mã hóa phi video (non-VCL NAL). Các gói VCL NAL mang một phần của một dữ liệu hình ảnh của video đã được mã hóa, trong khi các gói

non-VCL NAL chứa các dữ liệu kiểm soát, thông số sử dụng trong quá trình giải mã. Dữ liệu của một hình ảnh đã được mã hóa với non-VCL NAL (những thông số kiểm soát quá trình mã hoá hình ảnh đó) tạo ra một gói truy cập của H.265/HEVC. Do đó, gói truy cập của H.265/HEVC bao gồm nhiều hơn một gói VCL NAL, do đa số các trường hợp một gói NAL không thể chứa được dữ liệu mã hoá của một hình ảnh, trường hợp đặc biệt là một gói trong trường hợp hình ảnh được mã hóa không được phân vùng (nhiều slice).

Tất cả các gói NAL đều có một tiêu đề (header) và phần dữ liệu (payload). Tiêu đề của gói NAL có độ dài là 2 byte, được thiết kế để dễ dàng phân tích và xác định mục đích chính của gói NAL, hoặc tầng mã hoá video (layer) mà phần dữ liệu của nó mang theo. Tuy nhiên, bit đầu tiên là luôn được để "0" để tránh giả lập (emulation prevention bytes) hoặc ngăn ngừa việc một mẫu dữ liệu bị lặp lại một cách ngẫu nhiên bên trong phần dữ liệu tải. Sáu bit tiếp theo được sử dụng để xác định loại dữ liệu trong phần tải của gói NAL, sau đó là ID của tầng video đã được đề cập ở trên. Ba bit cuối cùng hiển thị một trong bảy giá trị nhận dạng tạm thời.



Hình 2-1: Cấu trúc tiêu đề gói NAL [1]

Loại gói VCL NAL được phân loại dựa trên các gói truy cập và loại hình ảnh. Có ba loại hình ảnh cơ bản trong H.265/HEVC: IRAP (Intra Random Access Point – tạm dịch: Hình ảnh điểm truy cập ngẫu nhiên nội ảnh), leading pictures and trailing pictures (tạm dịch: hình ảnh đầu tiên và hình ảnh cuối cùng của một nhóm ảnh trong một đoạn video); và một vài loại ảnh khác, STSA (Step-wise Temporal Sublayer Access – tạm dịch: hình ảnh truy cập tầng phụ tạm thời), TRAIL (Ordinary Trailing – tạm dịch: ảnh cuối gốc), IDR (Instantaneous Decoding Refresh – tạm dịch: ảnh giải mã tức thời), CRA (Clean Random Access – tạm dịch: hình ảnh truy cập ngẫu nhiên), vv. Tất cả các loại hình ảnh đều có vai trò riêng trong việc mã hóa video. Bên cạnh đó, các loại gói phi VCL NAL được phân loại dựa vào tập tham số, dấu phân tách, bộ lọc

dữ liệu, SEI (Supplemental Enhancement Information – tạm dịch: thông tin tăng cường bổ sung).

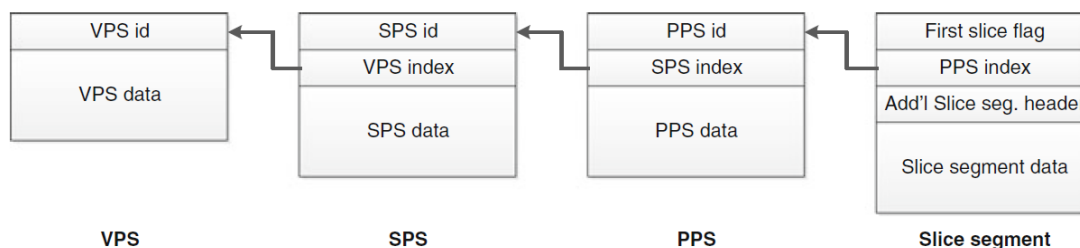
Non-VCL NAL unit types			
Parameter sets	32	VPS_NUT	Video parameter set
	33	SPS_NUT	Sequence parameter set
	34	PPS_NUT	Picture parameter set
Delimiters	35	AUD_NUT	Access unit delimiter
	36	EOS_NUT	End of sequence
	37	EOB_NUT	End of bitstream
Filler data	38	FD_NUT	Filler data
Supplemental enhancement information (SEI)	39	PREFIX_SEI_NUT	
	40	SUFFIX_SEI_NUT	
Reserved	41–47	RSV	
Unspecified	48–63	UNSPEC	

Bảng 2-1: Các loại gói Non-VCL NAL[1]

Trailing non-IRAP pictures			
Non-TSA, non-STSA trailing	0	TRAIL_N	Sub-layer non-reference
	1	TRAIL_R	Sub-layer reference
Temporal sub-layer access	2	TSA_N	Sub-layer non-reference
	3	TSA_R	Sub-layer reference
Step-wise temporal sub-layer	4	STSA_N	Sub-layer non-reference
	5	STSA_R	Sub-layer reference
Leading pictures			
Random access decodable	6	RADL_N	Sub-layer non-reference
	7	RADL_R	Sub-layer reference
Random access skipped leading	8	RASL_N	Sub-layer non-reference
	9	RASL_R	Sub-layer reference
Intra random access point (IRAP) pictures			
Broken link access	16	BLA_W_LP	May have leading pictures
	17	BLA_W_RADL	May have RADL leading
	18	BLA_N_LP	Without leading pictures
Instantaneous decoding refresh	19	IDR_W_RADL	May have leading pictures
	20	IDR_N_LP	Without leading pictures
Clean random access	21	CRA	May have leading pictures
Reserved			
Reserved non-IRAP	10–15	RSV	
Reserved IRAP	22–23	RSV	
Reserved non-IRAP	24–31	RSV	

Bảng 2-2: Các loại gói VCL NAL[1]

2.1.2. Các tập tham số



Hình 2-2: Các tập tham số trong H.265/HEVC[1]

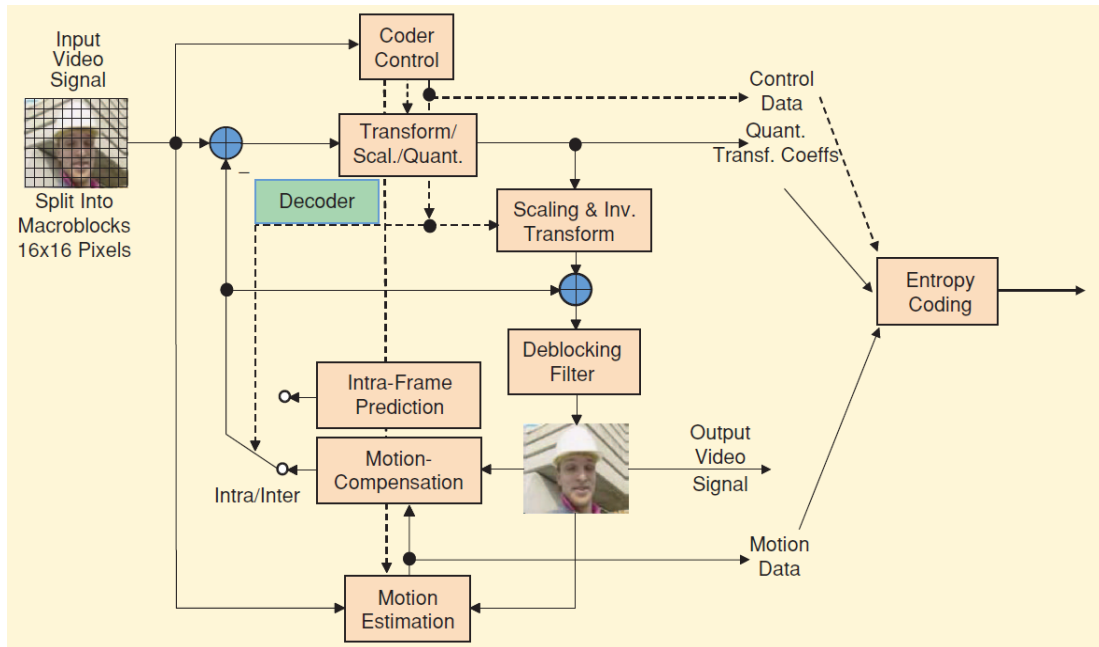
Các tập tham số trong H.265/HEVC về cơ bản là giống với các tập tham số trong H.264/AVC, ngoại trừ một tập mới được gọi là tập tham số video (VPS – Video Parameter Set), các tập khác giữ nguyên: tập tham số chuỗi (SPS – Sequence Parameter Set), tập tham số hình ảnh (PPS – Picture Parameter Set). Mục tiêu của việc sử dụng các tập tham số là làm tăng hiệu quả tỷ lệ bit, khả năng phục hồi lỗi, và cung cấp các giao diện lớp cho hệ thống. Cụ thể là những tham số có thể được sử dụng nhiều lần ở trong quá trình mã hoá những ảnh khác nhau, mảnh (slice) khác nhau... nên tùy vào phạm vi sử dụng của các tham số mà phân chia vào các tập khác nhau, tránh truyền lặp đi lặp lại nhiều lần – bit-rate tăng. Một mảnh hoặc mẫu (sample) có thể dễ dàng lấy được giá trị của tham số cần dùng bằng cách truy cập vào các tập tham số thông qua ID, như thể hiện trong hình trên (hình 2-2).

- Tập tham số video (VPS) là một tập mới được xác định trong H.265/HEVC. Các thông số của nó được áp dụng trong quá trình mã hoá toàn bộ video.

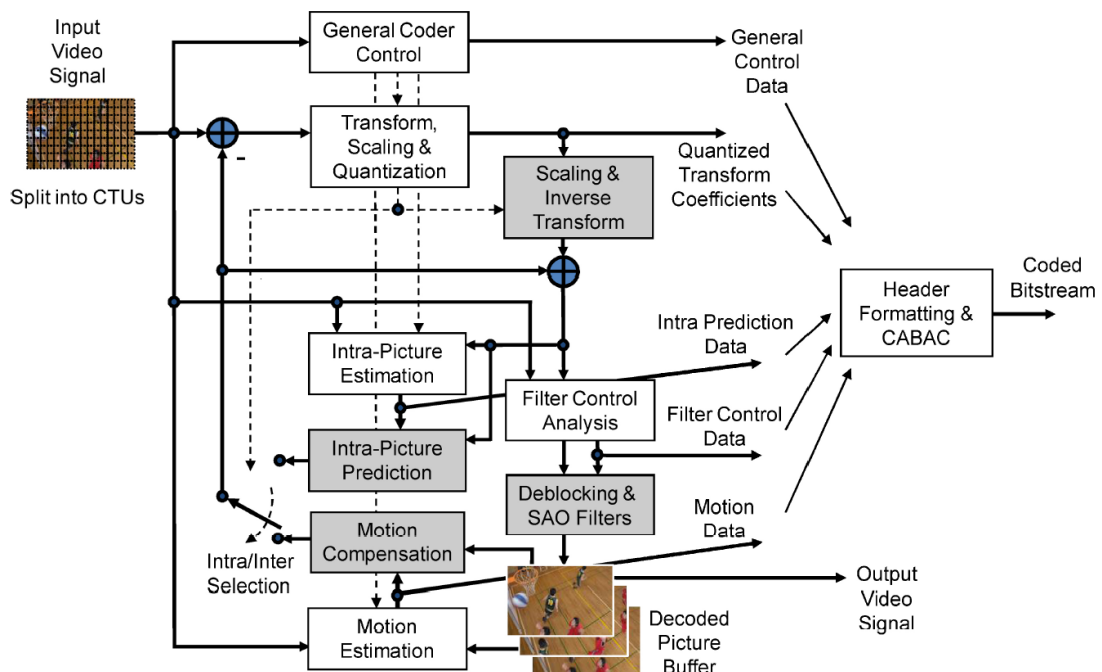
- Tập tham số chuỗi (SPS) chứa thông tin áp dụng cho quá trình mã hoá một nhóm các hình ảnh.

- Tập tham số hình ảnh (PPS) chứa các thông số áp dụng cho quá trình mã hoá một hình ảnh cụ thể. PPS thay đổi theo hình ảnh, nhưng những hình ảnh khác vẫn có thể tham chiếu tới PPS đó (thông qua ID).

2.2. Cấu trúc mã hoá



Hình 2-3: Cấu trúc mã hóa video của H.264/AVC



Hình 2-4: Cấu trúc mã hóa video Hybrid của phiên bản H.265/HEVC đầu tiên

Theo hình 2-4, ta có thể thấy rằng mã hoá H.265/HEVC được xây dựng dựa trên nền tảng cơ bản của H.264/AVC. Tất cả những bước xử lý chính đều được giữ nguyên, như mã hoá nội ảnh, mã hoá liên ảnh, các bộ lọc, lượng tử hoá.... Tuy nhiên bên cạnh đó, H.265/HEVC đã có những thay đổi. Dễ nhận biết nhất là sự tách biệt của mã hoá nội ảnh thành 2 khối là dự đoán nội ảnh (intra-picture prediction) và ước tính nội ảnh

(intra-picture estimation). Phần sơ đồ của H.265/HEVC chỉ có 1 xử lý mới duy nhất là: phân tích điều kiện bộ lọc (filter control analysis). Điều quan trọng ta rút ra được là có rất nhiều sự cải tiến trong H.265/HEVC từ H.264/AVC, nhưng nó chỉ được cải tiến về kỹ thuật trong từng bước xử lý của bộ mã hoá.

Cụ thể những sự thay đổi sẽ được giải thích chi tiết trong các phần sau. Sự tăng kích cỡ của khối và cấu trúc của nó sẽ được mô tả ngay phần tiếp theo. Sau đó sẽ là một trong những cải tiến quan trọng nhất của H.265/HEVC, đó là xử lý song song. Chương 3 sẽ mô tả về sự khác nhau của dự đoán nội ảnh và liên ảnh của hai chuẩn nén. Phần cuối cùng là những so sánh về bộ lọc trong, và chức năng khử rung, khử blocked của nó.

2.3. Cấu trúc khối và kỹ thuật song song

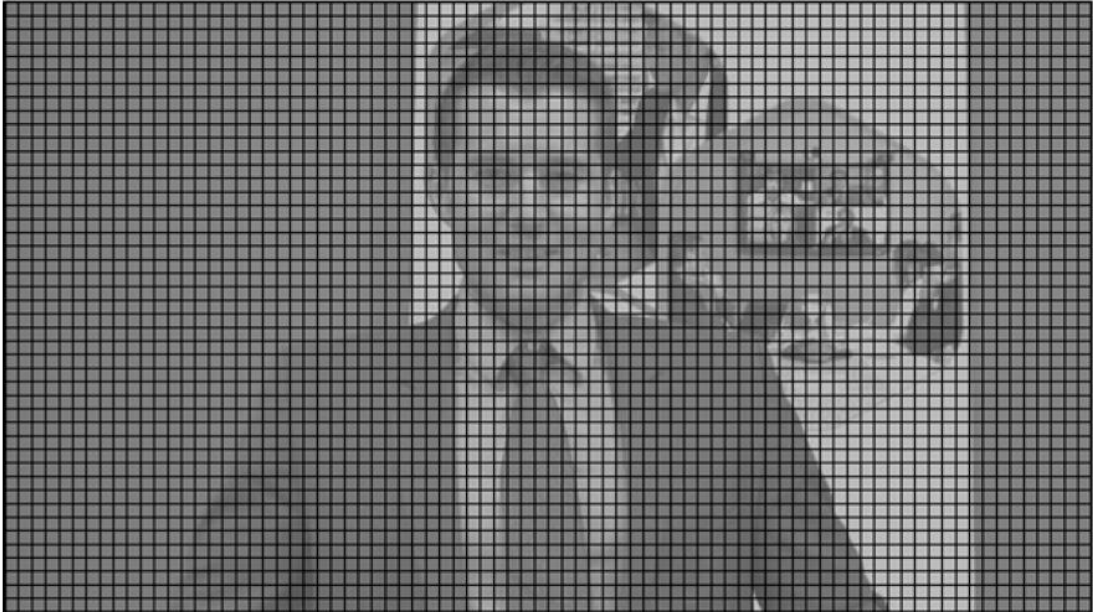
H.265/HEVC là một phương pháp mã hóa video hỗn hợp theo khối, nền tảng của mã hoá video, giống như những chuẩn mã hoá trước. Cũng như H.264/AVC, hình ảnh được chia thành nhiều khối. Tuy nhiên, H.265/HEVC đề xuất một cấu trúc dữ liệu, có thể cải thiện đáng kể khả năng dự đoán và chuyển đổi của tiêu chuẩn nén H.265/HEVC này.

2.3.1. Phân vùng khối

2.3.1.1. Khối mã hóa cây và đơn vị mã hóa cây

Thay vì sử dụng khối macroblock như H.264/AVC và tất cả các tiêu chuẩn mã hóa video trước, trong H.265/HEVC, một hình ảnh được phân chia thành nhiều khối vuông, gọi là khối mã hóa cây (CTB – Coding Tree Blocks), như thể hiện trong hình 2-5 (macroblock) và 2-6 (CTB). Những khối vuông CTB trong H.265/HEVC có kích thước từ 4x4 đến 64x64, lớn hơn so với kích thước lớn nhất của một macroblock (16x16) được sử dụng trong H.264/AVC. Một CTB, chính nó có thể chia ra các CTB khác dựa trên độ phức tạp của ảnh theo cấu trúc dữ liệu dạng cây (quad-tree structure). Những CTB nhỏ hơn này giống nhau về cả thành phần độ sáng (luma) và màu sắc (chroma). Do đó, một thành phần sáng CTB và 2 thành phần màu CTB tương ứng sẽ tạo ra một đơn vị mã hóa cây (CTU – Coding Tree Units) nếu định dạng nén video là

4:2:0, giống như macroblock, là đơn vị xử lý trong H.265/HEVC. Một nhóm các CTU liền kề cùng nhau cấu thành nên một mảnh, tương tự nhiều macroblock tạo nên mảnh trong H.264/AVC.



Hình 2-5: Ví dụ phân vùng hình ảnh thành nhiều khối macro 16x16[1]

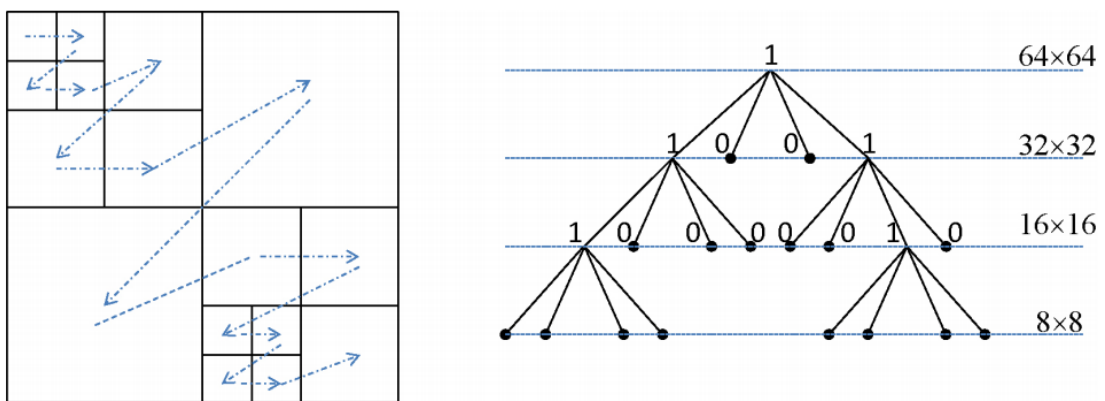


Hình 2-6: Ví dụ về phân vùng hình ảnh thành nhiều CTU 64x64[1]

Do kích thước khối lớn, việc mã hóa của H.265/HEVC trở lên hiệu quả hơn, nhưng đòi hỏi bộ nhớ tốt hơn, làm tăng độ trễ và sự phức tạp trong tính toán ở cả hai bộ mã hóa và giải mã. Tuy nhiên, kích thước khối lớn hơn cho phép cấu trúc mã hóa của H.265/HEVC phù hợp với đặc điểm nội dung video có độ phân giải cao, so với tất cả

các tiêu chuẩn mã hóa trước đó. Cụ thể như trong hình 2-6, hình ảnh có nhiều vùng giống nhau nên việc sử dụng khối kích thước lớn sẽ hiệu quả hơn, bao phủ được rộng hơn so với khối kích thước nhỏ, từ đó giảm số khối, giảm số xử lý mã hóa. Bộ mã hoá sẽ cân bằng giữa điều kiện phần cứng và yêu cầu của định dạng video mã hoá mà chia ảnh ra thành các CTB một cách hợp lý. Kích thước CTU được xác định bởi bộ mã hoá, sau đó truyền đến bộ giải mã thông qua các tập tham số chuỗi (SPS).

Cũng như CTB, CTU thể được phân chia thêm thành các đơn vị hình vuông nhỏ hơn, được gọi là đơn vị mã hóa (CU – Coding Unit) dựa theo cấu trúc dữ liệu dạng cây để giải quyết một số vấn đề: như việc lựa chọn chế độ dự đoán cho khối là nội ảnh (intra) hay liên ảnh (inter) là không thích hợp do gặp vấn đề bất lợi trong quá trình dự đoán tỷ lệ biến dạng (rate-distortion). Phân vùng đệ quy chia CTU thành nhiều CU có kích thước khối khác nhau, nhỏ hơn, làm cho H.265/HEVC có thể quyết định một cách linh hoạt và hiệu quả về dự đoán nội ảnh hay liên ảnh, đặc biệt là làm giảm mối tương quan giữa các khối trong cùng một ảnh.

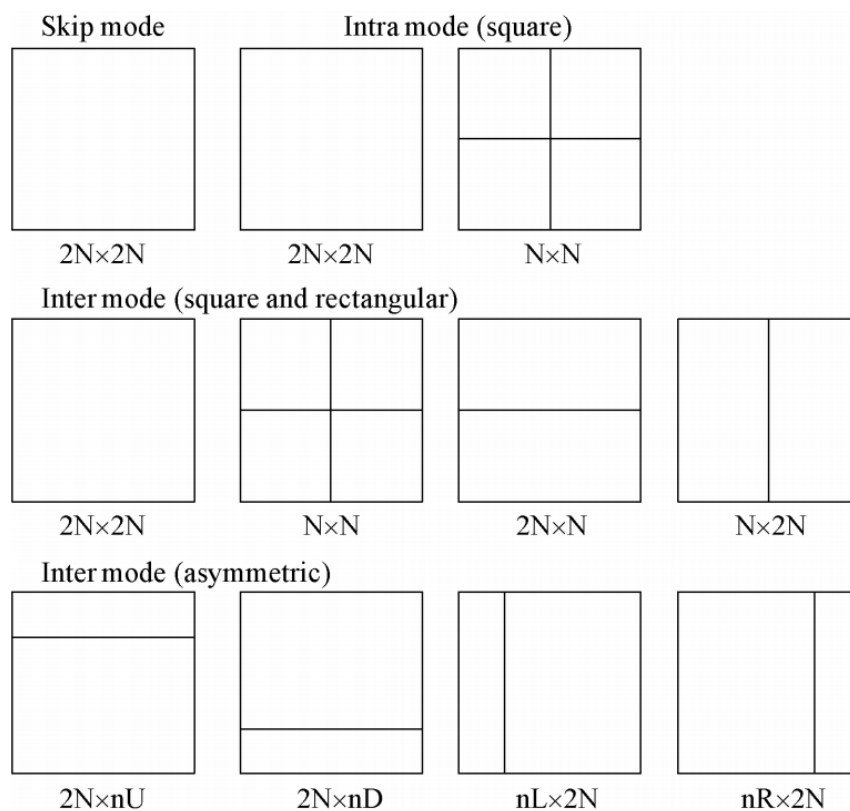


Hình 2-7: Ví dụ về phân vùng CTU theo thứ tự chiều sâu

2.3.1.2. Khối dự đoán và đơn vị dự đoán

Trường hợp CU được mã hóa theo dự đoán liên ảnh (inter-picture prediction), thành phần sáng (luma) và màu (chroma) của CU đó có thể được chia thành các đơn vị nhỏ hơn làm cơ sở để dự đoán, gọi là các khối dự đoán (PB – Prediction Blocks). Do đó, PB được định nghĩa là các mẫu chứa thành phần sáng hoặc màu mà sử dụng thông số chuyển động (motion parameters) giống nhau. Trong đó, các thông số chuyển động

bao gồm các vector chuyển động được dự đoán và những ảnh tham chiếu của các vector đó. Tương tự với cú pháp của CU, một đơn vị dự đoán (PU – Prediction Units) được tạo ra bởi một PB độ sáng và PB màu sắc tương ứng của nó. Một CU có thể chứa nhiều đơn vị dự đoán, trường hợp đặc biệt là khi kích thước của PU bằng với gốc CU nên CU chỉ có một PU. Bên trong một PU, tất cả quá trình dự đoán được thực hiện và thông tin được truyền đến bộ giải mã. Theo chế độ dự đoán liên ảnh, H.265/HEVC có tám hình dạng chia tách từ CU thành PU, xem hình 2-8. Với số lượng kích thước PU đa dạng hơn thì hiệu quả mã của của H.265/HEVC cũng cao hơn. Hiệu quả dự đoán bù chuyển động trong H.265/HEVC cao hơn. Tuy nhiên, có sự cân nhắc giữa số lượng nhỏ hơn các phương thức dự đoán và hiệu quả mã hóa. Kích thước khối tối thiểu của PU trong H.265/HEVC là 4x4.



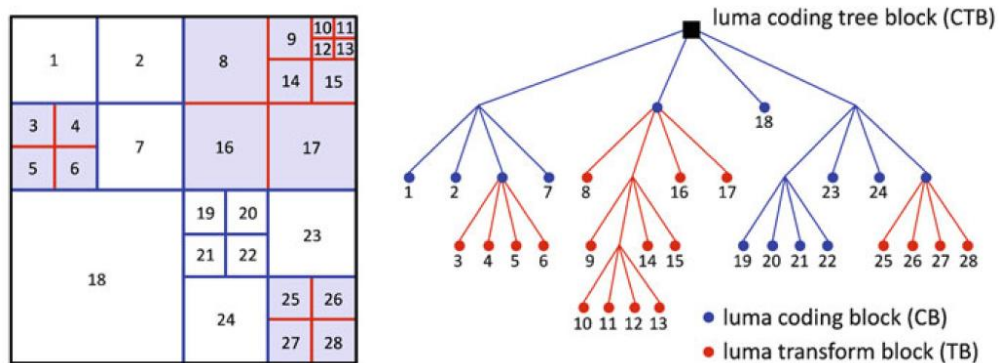
Hình 2-8: Tất cả các kích thước của PU trong H.265/HEVC[1]

Video coding standard	Supported block sizes for motion-compensated prediction
H.262 MPEG-2 Video	16×16
H.263	$16 \times 16, 8 \times 8$
MPEG-4 Visual	$16 \times 16, 8 \times 8$
H.264 MPEG-4 AVC	$16 \times 16, 16 \times 8, 8 \times 16, 8 \times 8, 8 \times 4, 4 \times 8, 4 \times 4$
HEVC	$64 \times 64, 64 \times 48, 64 \times 32, 64 \times 16, 48 \times 64, 32 \times 64, 16 \times 64, 32 \times 32, 32 \times 24, 32 \times 16, 32 \times 8, 24 \times 32, 16 \times 32, 8 \times 32, 16 \times 16, 16 \times 12, 16 \times 8, 16 \times 4, 12 \times 16, 8 \times 16, 4 \times 16, 8 \times 8, 8 \times 4, 4 \times 8$

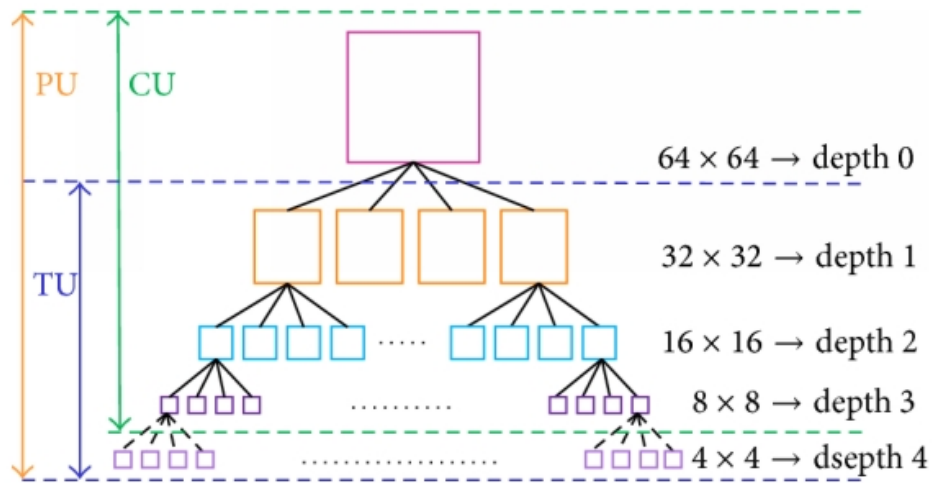
Bảng 2-3: Kích thước khối bù chuyển động đã được hỗ trợ trong H.265/HEVC và trong kỹ thuật trước đây[1]

2.3.1.3. Khối biến đổi và đơn vị biến đổi

Một block mã hóa (CB) có thể được chia thành nhiều khối biến đổi (TB – Transform Blocks). Một TBs được mô tả là một khối hình vuông chứa thành phần sáng hoặc màu, mà tại đó ánh xạ hai chiều được sử dụng để mã hóa. Việc chia CB thành nhiều TB được dựa trên cấu trúc dạng cây, như trên hình 2-9. Trong đó, vị trí gốc là CB và các lá là các TB. Một nút lá được xác định trong RQT (Residual Quad-tree Structure) khi kích thước khối biến đổi nhỏ nhất, tuy nhiên sự phân chia RQT bị hạn chế bởi độ sâu tối đa. Ví dụ, nếu độ sâu tối đa của RQT là 1, thì CB $2N \times 2N$ chỉ có thể được phân chia một lần thành 1 TB có cùng kích thước hoặc $4N \times N$. Trường hợp đặc biệt nếu độ sâu tối đa là 0, kích thước của CB hiện tại là 64×64 , trong khi kích thước biến đổi tối đa là 32. Tại thời điểm này, CB bị ép chia thành 4 TB 32×32 để đáp ứng giới hạn về kích thước biến đổi tối đa. Cũng như CB, TB sáng và các TB màu tương ứng cấu thành nên đơn vị chuyển đổi (TU – Transform Unit).



Hình 2-9: Ví dụ về chia một CTB thành nhiều TBs

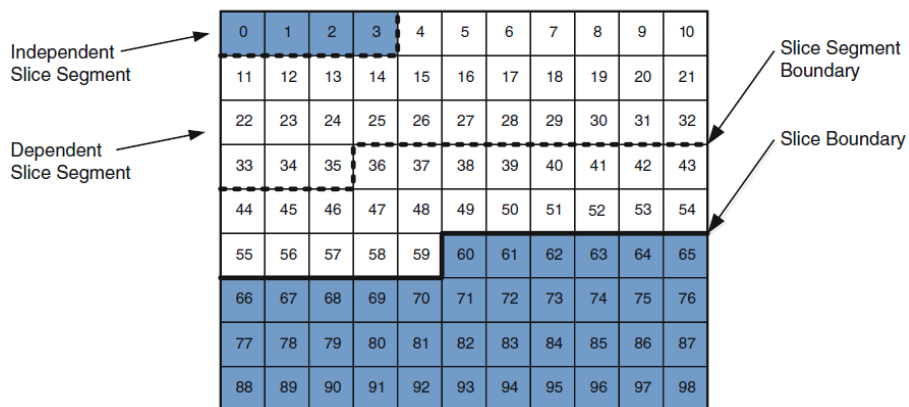


Hình 2-10: Mối quan hệ giữa CU, PU và TU trong H.265/HEVC

2.3.2. Phân vùng hình ảnh

2.3.2.1. Mảnh (slice), phân mảnh (slice fragmentation), phân đoạn mảnh (slice segments) và tập hợp phân đoạn mảnh(slice segment subsets)

Giống như tiêu chuẩn mã hóa H.264/AVC trước đây, trong H.265/HEVC, một hình ảnh được phân chia thành một hoặc nhiều mảnh (slice). Nó chứa một hoặc nhiều CTU, như macroblock trong H.264/AVC. Sự khác biệt trong H.265/HEVC là một mảnh có thể giải mã độc lập mà không phụ thuộc vào những mảnh khác trong cùng một ảnh. Các mảnh được chia sao cho thỏa mãn được ba mục đích: có khả năng khắc phục lỗi, khớp với kích thước đơn vị truyền tải tối đa (MTU – Maximum Transmission Units), và xử lý song song.



Hình 2-11: Cấu trúc các mảnh độc lập trong H.265/HEVC[1]

Một mảnh có hai phần: tiêu đề và dữ liệu. Tất cả các thông tin, thông số, hệ số liên quan tới quá trình giải mã của mảnh và CTU của nó chứa trong tiêu đề mảnh. Nhiều mảnh trong một hình ảnh có thể có cùng một header, giúp cho việc mã hóa hiệu quả hơn và giảm tốc độ bit. Như tên gọi của nó, dữ liệu mảnh chứa dữ liệu mã hóa của mảnh. Trong một mảnh, tất cả các CTU trực thuộc được mã hóa theo thứ tự quét mảnh, từ trái sang phải, từ trên xuống dưới. Cũng giống như H.264/AVC, có ba loại mảnh ứng với ảnh chứa nó: mảnh I (tất cả các CTU bên trong những mảnh này được mã hoá bằng cách chỉ sử dụng dự đoán nội ảnh), mảnh P (tất cả các CTU bên trong những mảnh này được mã hoá bằng cách sử dụng dự đoán liên ảnh (điều kiện: tất cả các hình ảnh tham chiếu của nó đều phải được mã hoá trước hình ảnh hiện tại, là hình ảnh trong danh mục danh sách hình ảnh tham chiếu 0 – list 0) và mảnh B (tương tự như mảnh P nhưng mảnh B có thể sử dụng hình ảnh trong danh sách 1 làm hình ảnh tham chiếu – list 1)).

Như đề cập ở trên, một mảnh có thể được chia thành nhiều mảnh nhỏ, gọi là các phân đoạn mảnh phụ thuộc. Tiêu đề của những mảnh này sử dụng chung với mảnh đầu tiên, qua đó mảnh đầu tiên bao giờ cũng là mảnh độc lập. Các đoạn mảnh khác là phụ thuộc, do đó không có phần tiêu đề. Trong cùng một mảnh, không có hạn chế về dự đoán và mã hóa entropy của phân phụ thuộc CTU qua các ranh giới phân khúc mảnh. Ngoài ra, một phần dữ liệu mảnh có thể được chia thành các phần nhỏ hơn, gọi là các nhóm phân đoạn mảnh (slice segments). Những nhóm này được chia sao cho thích hợp nhất đối với việc sử dụng các công cụ song song trong H.265/HEVC.

2.3.3. Xử lý song song trong H.265/HEVC

2.3.3.1. Mức độ song song

Song song là một trong những kỹ thuật quan trọng tạo nên sự khác biệt về hiệu quả mã hóa trong H.265/HEVC, so với những tiêu chuẩn mã hóa khác trước đó. Đa luồng không phải là khái niệm mới trong lập trình, nhưng đến giờ nó mới được áp dụng vào tiêu chuẩn nén video. Nhờ các bộ xử lý nhiều nhân trong một CPU, một hình ảnh, một mảnh hay một khối có thể được mã hóa một cách độc lập trong một lõi. Điều này làm cho quá trình mã hóa có thể thực hiện không đồng bộ, do đó làm giảm thời gian mã

hóa, giảm độ trễ mã hóa trong H.265/HEVC, so với quá trình mã hóa đồng bộ của các tiêu chuẩn cũ.

Trong H.265/HEVC, có ba cấp độ cấu trúc có thể thực hiện theo xử lý song song:

- Song song cấp hình ảnh: nhiều hình ảnh có thể được mã hóa cùng một lúc. Do đó, các thành phần phụ thuộc thời gian cho dự đoán bù chuyển động được đáp ứng. Mỗi lõi trong một bộ xử lý có trách nhiệm mã hóa một hình ảnh.
- Song song cấp mảnh: Như đã thảo luận, trong H.265/HEVC, một hình ảnh được phân chia thành nhiều mảnh. Những mảnh này có thể được mã hóa một cách độc lập với những mảnh khác trong cùng một hình ảnh, do đó các mảnh có thể được sử dụng để song song hóa. Tất cả các dữ liệu cần thiết để giải mã chứa trong tiêu đề mảnh, dữ liệu mảnh và các loại tập tham số.
- Song song cấp khối: dựa vào các quá trình độc lập của một khối mã hóa, song song cấp khối có nghĩa là một khối có thể được dự đoán trong một lõi, trong vòng lọc trong một lõi, và entropy được mã hóa trong một lõi khác. Mỗi bước mã hóa một khối có thể được xử lý đồng thời trên các lõi khác nhau.

Có ba mức mã hóa song song trong H.265/HEVC. Không phải tất cả đều hiệu quả mà vẫn có những hạn chế. Song song cấp hình ảnh chỉ có thể áp dụng tốt trong dự đoán nội ảnh, mà không hiệu quả trong dự đoán liên ảnh, vì nếu một hình ảnh tham chiếu của hình ảnh hiện tại không được mã hóa đúng thời gian, quá trình này phải dừng lại cho đến khi tất cả các hình ảnh tham chiếu đã sẵn sàng. Song song cấp hình ảnh làm tăng tốc độ xử lý nhưng không làm giảm thời gian chờ vì thời gian mã hóa / giải mã giữa các ảnh khác nhau. Song song cấp mảnh có bất lợi trong việc dự đoán, mã hóa entropy, xung đột ở danh giới các mảnh. Song song cấp khối làm tăng bộ nhớ, cũng như làm giảm thời gian giải mã đoạn mã hóa entrop. Phiên bản H.265/HEVC đầu tiên đã cung cấp hai công cụ song song để giải quyết các vấn đề trên: tile và song song sóng trước (WPP - Wave-front parallel processing)

2.3.3.2. Các công cụ song song cấp cao sử dụng trong H.265/HEVC

1. Tiles

Tile là một trong những tính năng mới được sử dụng trong H.265/HEVC. Nó có một số điểm tương đồng với nhóm mảnh, được sử dụng trong H.264/AVC. Tile là nhằm mục đích đạt được hiệu quả xử lý song song mà không ảnh hưởng nhiều đến chất lượng video, và kiểm soát cân bằng giữa các lỗi trong bộ mã hóa/giải mã.

Nếu tile được kích hoạt, hình ảnh sẽ được chia thành nhiều khu vực hình chữ nhật (phân nhóm hình ảnh). Dựa trên cơ chế phân vùng theo độ phức tạp của ảnh một cách linh hoạt, việc bố trí các khu vực hình chữ nhật (tile), bao gồm ranh giới của chúng, khác nhau giữa mỗi bức ảnh. Độ phức tạp của ảnh càng cao, càng cần nhiều xử lý, càng nhiều vùng hình chữ nhật được chia. Tuy nhiên, số tile là giới hạn, được xác định bởi một số biến được truyền trong PPS. Điểm đầu của mỗi tile cũng được truyền trong tiêu đề mảnh. Tile phải tuân theo ít nhất một trong hai điều kiện. Đầu tiên là tất cả CTU trong một phân đoạn mảnh phải thuộc về một tile, và thứ hai là tất cả CTU của một tile thuộc cùng một phân khúc mảnh.

Khi sử dụng tile, thứ tự mã hóa các CTU trong một bức ảnh thay đổi theo thứ tự quét mảnh dựa vào tile. Điều đó có nghĩa là CTU được xử lý theo hàng từ trái sang phải trong phạm vi tile. Do hai điều kiện trên, nếu điểm khởi đầu của một mảnh không giống với một tile, mảnh có thể không có nhiều tile bên trong.

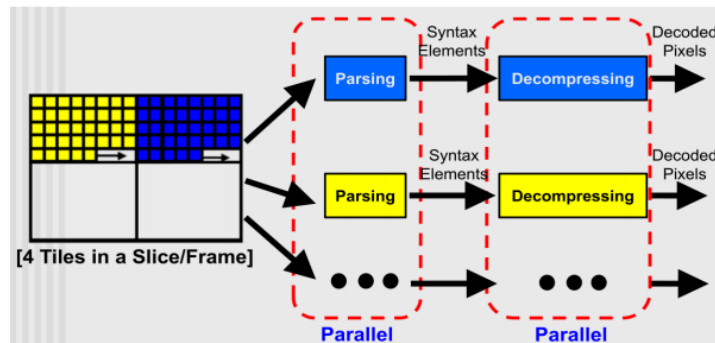
Mặc dù những loại tile này có thể được xử lý độc lập trong cả hai bộ mã hóa và giải mã, việc cài đặt lại mã hóa entropy tại các điểm đầu của tile, phá vỡ các thành phần phụ thuộc như trong dự đoán nội ảnh, có thể tạo ra các hiện vật trực quan tại danh giới Tile, và do đó làm giảm chất lượng mã hóa. Bộ lọc vòng (Bỏ chặn và SAO) có thể chặn các hiện vật tại danh giới Tile, các bộ lọc này sẽ được giải thích trong phần sau.

Sử dụng Tile, dự báo nội ảnh có cả lợi thế và bất lợi. Ưu điểm là giảm khoảng cách không gian, tăng hiệu suất khai thác các mối tương quan không gian giữa các mẫu,

CTU bên trong một gói. Hơn nữa, có thể giúp giảm các thông tin tiêu đề mảnh đã báo hiệu nếu không phải là phải sử dụng cơ chế mỗi mảnh mỗi Tile. Ngược lại, những bất lợi của việc sử dụng Tile là tạo ra hiện vật trực quan, phá vỡ sự phụ thuộc như đã đề cập ở trên, đặc biệt là nếu có nhiều Tile trong một hình ảnh, sẽ khởi tạo với nhiều tile.

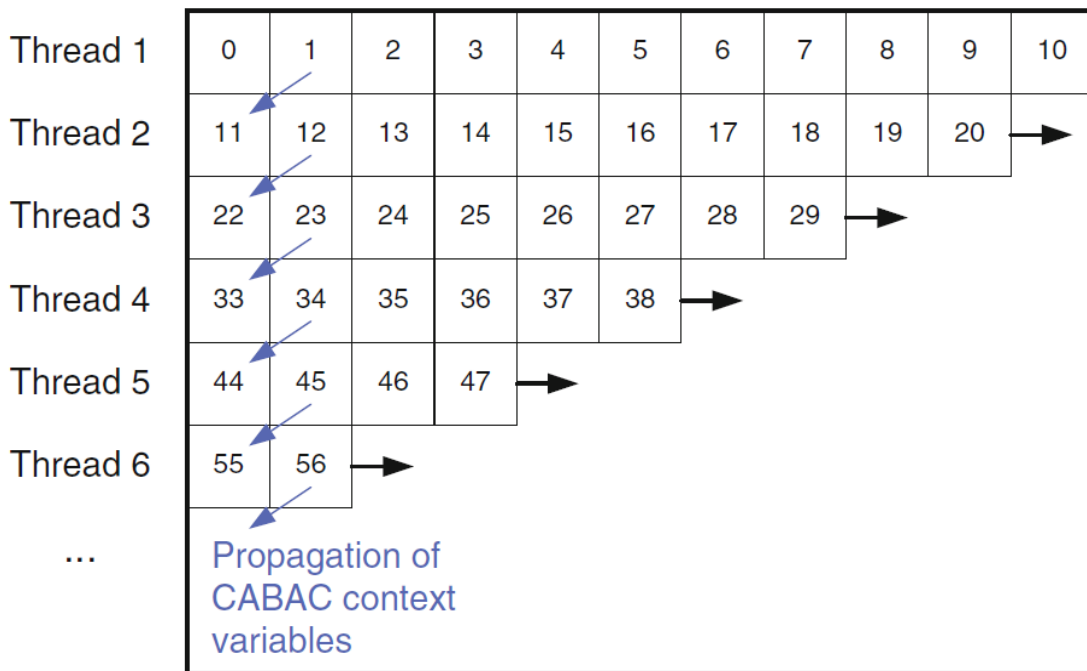
0	1	2	3	16	17	24	25	26	27	28
4	5	6	7	18	19	29	30	31	32	33
8	9	10	11	20	21	34	35	36	37	38
12	13	14	15	22	23	39	40	41	42	43
44	45	46	47	52	53	56	57	58	59	60
48	49	50	51	54	55	61	62	63	64	65
66	67	68	69	78	79	84	85	86	87	88
70	71	72	73	80	81	89	90	91	92	93
74	75	76	77	82	83	94	95	96	97	98

Hình 2-12: Ví dụ về sử dụng tile chia hình ảnh thành 9 vùng, đường gạch thể hiện ranh giới các vùng. [1]



Hình 2-13: Quy trình của tiles

2. Xử lý song song sóng trước (WPP - Wave-front parallel processing)



Hình 2-14: Xử lý song song sóng trước

Xử lý song song sóng trước (WPP – Wave-front parallel processing) là một công cụ song song chủ yếu, thực hiện trong H.265/HEVC. Khác với tile, WPP chia một hình ảnh thành các hàng riêng biệt. Những hàng này cũng có thể được mã hóa hay giải mã một cách độc lập, như tile. Nếu WPP được kích hoạt, mỗi hàng được xử lý trong theo thứ tự quét mảnh, và bắt đầu ngay sau khi hai CTU ở các dòng trước đó được mã hoá. WPP lưu trữ tốt hơn so với Tile vì nó không phá vỡ sự phụ thuộc như Tile. CABAC (Context-Adaptive Binary Arithmetic Coding) bị chấm dứt ở mỗi hàng để cho phép xử lý song song. Để giảm bớt những thiệt hại do quá trình khởi tạo CABAC truyền thống tại điểm bắt đầu một hàng, nội dung của CABAC đã được thích ứng được truyền từ bộ mã hóa hoặc giải mã tới cái khác. Do đó, tổn thất là nhỏ dựa trên WPP. So với tile, WPP đạt hiệu quả mã hóa cao hơn. Lý do là nó không phá vỡ sự phụ thuộc như trên, không thay đổi thứ tự mã hóa, và quá trình tái thiết. Số luồng bằng với số hàng trong một hình ảnh. Do đó, nếu có điều xấu xảy ra ở dòng trước, tất cả các hàng dưới dòng có thể không được mã. Đó là nhược điểm của việc sử dụng WPP.

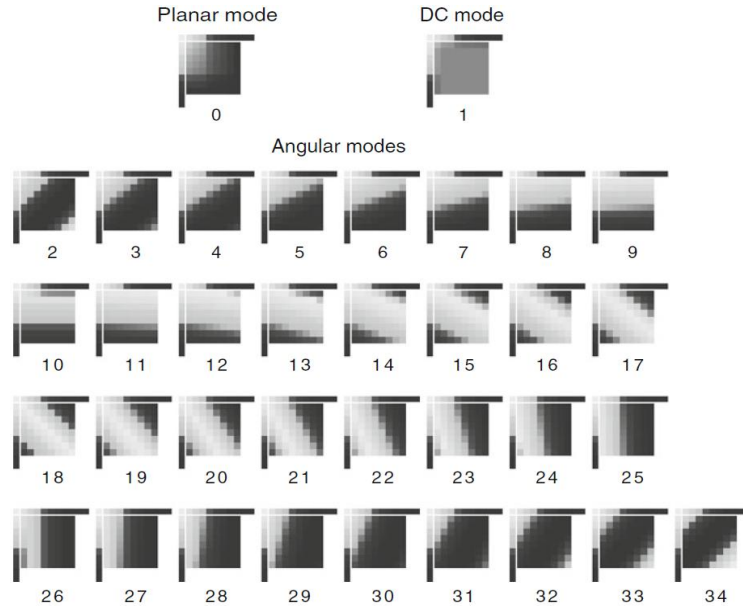
2.4. Kết luận

Về cơ bản, H.265/HEVC có phương pháp mã hóa giống với chuẩn H.264/AVC. Sự khác nhau nằm ở cấu trúc khối của H.265/HEVC đã có sự thay đổi hoàn toàn, từ macroblock chuyển sang CTU, PU, TU.... Thay đổi này nhắm tới những xử lý song song của H.265/HEVC mà nó sẽ được đề cập tới trong phần sau. Đặc biệt, kích thước khối đã được tăng lên từ 16x16 thành 64x64 là điểm nhấn trong chuẩn mã hóa mới này. Kích thước lớn giúp cho H.265/HEVC có khả năng mã hóa, nén những video có độ phân giải cao trong khi độ nén những video này với H.264/AVC chưa tốt. H.265/HEVC sẽ mã hóa nén video có độ phân giải cao tốt hơn H.264/AVC do kích thước khối cũng như kỹ thuật nén của H.265/HEVC phức tạp hơn. Tuy nhiên, với những video có độ phân giải dưới HD, H.264/AVC sẽ tốt hơn vì kích thước khối nhỏ hơn và xử lý ít phức tạp hơn H.265/HEVC. Do xử lý trong H.265/HEVC phức tạp, nên quá trình mã hóa nén sẽ kéo dài hơn H.264/AVC. Vì vậy, áp dụng xử lý song song sẽ giảm bớt đáng kể thời gian xử lý, nhưng yêu cầu phải có một hệ thống xử lý mạnh hơn.

Dự đoán nội ảnh và dự đoán liên ảnh

3.1. Dự đoán nội ảnh

Dự đoán nội ảnh là một trong những tính năng quan trọng nhất trong mã hóa video. Phương pháp chính của dự đoán nội ảnh là sử dụng các mẫu liền kề được mã hóa trong nội bộ ảnh để dự đoán mẫu hiện tại, dựa trên khái niệm về các mẫu lân cận có cấu trúc tương tự với nhau. Trong H.265/HEVC, với việc thực hiện kỹ thuật song song, dự đoán nội ảnh có 3 bước: xây dựng mảng mẫu tham chiếu, dự đoán mẫu, và xử lý sau dự đoán. Sau đó, dự đoán nội ảnh được chia thành 2 loại: phương pháp dự đoán góc (angular prediction), cung cấp khả năng dự đoán các mẫu chính xác bằng cách dựa trên các cạnh hướng; dự đoán phẳng (DC prediction) và dự đoán planar cung cấp các khả năng ước lượng, làm mịn ảnh. Số hướng góc trong H.265/HEVC là 34, nhiều hơn rất nhiều hướng so với 8 hướng trong H.264/AVC. Với kích thước CTU lớn hơn, điểm cốt lõi của thiết kế H.265/HEVC là nhằm đạt hiệu quả mã hóa cao hơn so với tiêu chuẩn nén trước đó thông qua giảm tương quan không gian. Tuy nhiên, những tính toán trong dự đoán nội ảnh trong H.265/HEVC phức tạp hơn nhiều.



Hình 3-1: Ví dụ về các chế độ dự đoán nội ảnh của H.265/HEVC[1]

Functionality	H.264/AVC	HEVC
Prediction block sizes	4×4 , 8×8 and 16×16	4×4 , 8×8 , 16×16 and 32×32
Luma intra prediction modes	9 (4×4 and 8×8), 4 (16×16)	35
Chroma intra prediction modes	4	4 + the luma mode
Reference sample smoothing	8×8	8×8 and above
Boundary smoothing	N/A	Used for directly horizontal, vertical and DC modes
Operation when reference samples missing	Use DC mode	Reference sample substitution
Number of most probable modes in mode coding	1	3

Bảng 3-1: Sự khác nhau cơ bản của dự đoán nội ảnh giữa H.265/HEVC và H.264/AVC[1]

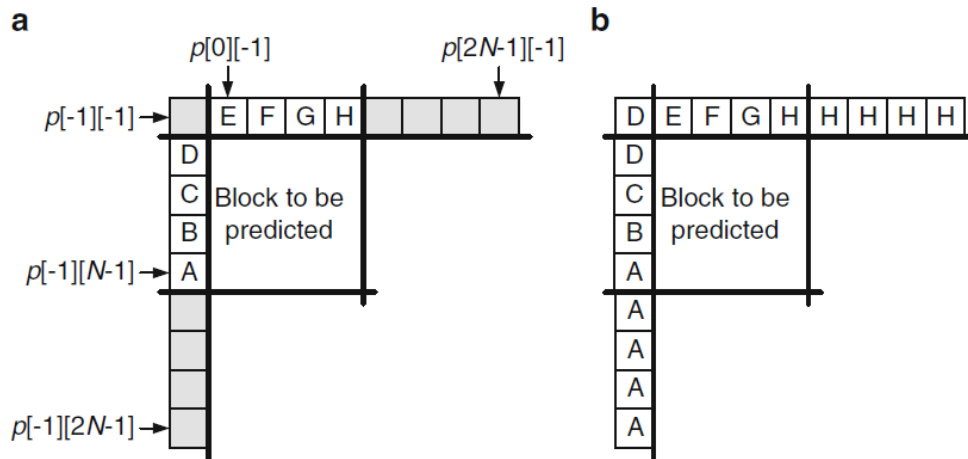
3.1.1. Thiết lập mẫu tham chiếu

Khác với H.264/AVC, H.265/HEVC giới thiệu một kỹ thuật mới được gọi là mẫu tham chiếu thay thế, tạo ra một bộ các chế độ dự đoán nội ảnh hoàn chỉnh sử dụng các mẫu tham chiếu lân cận. Do đó, để tăng số lượng các mẫu dự đoán, một quá trình lọc thích ứng được áp dụng, giúp lọc các mẫu tham chiếu theo phương thức dự đoán nội ảnh.

3.1.1.1. Thay thế mẫu tham chiếu

Dự đoán nội ảnh sử dụng nhiều mẫu liên kế để mẫu dự đoán hiện tại. Tuy nhiên, trong một số trường hợp, có các mẫu không có sẵn, chẳng hạn như mẫu bên ngoài hình ảnh, mảnh hoặc tile; hoặc mẫu thuộc về một PU với chế độ dự đoán liên ảnh. Tất cả các tham chiếu không có sẵn được thay thế bằng cách quét theo hướng kim đồng hồ và sử dụng các mẫu có sẵn mới nhất. Quá trình thay thế là:

- Nếu $p[-1][2N-1]$ không có, nó sẽ được thay thế bằng mẫu tham chiếu có sẵn đầu tiên khi quét các mẫu theo thứ tự chiều dọc $p[-1][2N-2]$ tới $p[-1][-1]$, sau đó theo hướng ngang từ $p[0][-1]$ tới $p[0][2N-1]$.
- Tất cả các mẫu không có sẵn theo chiều dọc của $p[-1][y]$ ($y = \{2N-2, \dots, -1\}$) được thay thế bằng các mẫu tham chiếu dưới đây $p[-1][y+1]$
- Tất cả các mẫu không có sẵn theo chiều ngang của $p[x][-1]$ ($x = \{0, \dots, 2N-1\}$) được thay thế bằng các mẫu tham chiếu dưới đây $p[x-1][-1]$



Hình 3-2: Quá trình thay thế mẫu (a) Mẫu tham chiếu trước khi thực hiện quá trình, các mẫu không có sẵn ký hiệu bằng màu xám (b) Mẫu tham chiếu sau khi thực hiện quá trình[1]

3.1.1.2. Quá trình lọc các mẫu tham chiếu

Để tránh các hướng không mong muốn giữa các mẫu được mã hóa nội ảnh, dựa trên phương thức dự đoán nội ảnh và kích thước của mẫu hiện tại, bộ lọc mịn được áp dụng trong H.265/HEVC. Nó cũng giống như khối dự đoán nội ảnh 8x8 của H.264.

Tuy nhiên, quá trình lọc được thực hiện với một số khối cụ thể và chế độ dự đoán nội ảnh. Trong trường hợp chế độ dự đoán nội ảnh là DC hoặc kích thước của khối là 4x4, quá trình này sẽ không xảy ra. Đối với các mẫu dự đoán 8x8, quá trình này chỉ áp dụng theo ba hướng: chế độ góc 2, 18 và 34. Đối với các khối dự đoán 16x16, quá trình này được áp dụng cho hầu hết các phương thức dự đoán nội ảnh, ngoại trừ: 9, 10, 11, 25, 26 và 27. Đối với các khối dự đoán 32x32, quá trình này không áp dụng cho hai chế độ dự đoán nội ảnh: hướng ngang (chế độ 10), và hướng dọc (chế độ 26).

Quá trình lọc có hai bước tùy thuộc kích thước khối và tính liên tục của các mẫu tham chiếu. Các $p[-1][2N-1]$ và $p[2N-1][1]$ không được thay đổi trong quá trình này. Tất cả các khối khác là các bộ lọc:

$$p[-1][-1] = (p[-1][0] + 2p[-1][-1] + p[0][-1] + 2) \gg 2$$

$$p[-1][y] = (p[-1][y+1] + 2p[-1][y] + p[-1][y-1] + 2) \gg 2$$

$$p[-1][x] = (p[x+1][-1] + 2p[x][-1] + p[x-1][-1] + 2) \gg 2$$

Trong đó $x = y = \{0, \dots, 2N-2\}$.

Đối với khối dự đoán 32x32, sau khi áp dụng quy trình đầu tiên, nếu các mẫu tham chiếu là phẳng, quá trình thứ hai được áp dụng. Độ phẳng được phát hiện bằng cách sử dụng phương trình này:

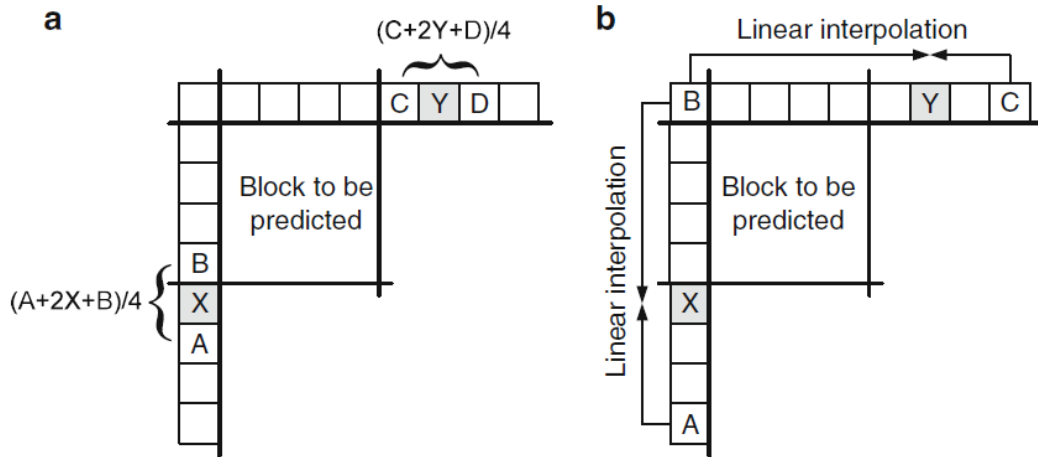
$$|p[-1][-1] + p[2N-1][-1] - 2p[N-1][-1]| < (1 \ll (b-5))$$

$$|p[-1][-1] + p[-1][2N-1] - 2p[1][N-1]| < (1 \ll (b-5))$$

Trong đó b là mẫu độ sâu bit. Nếu 2 phương trình trên là đúng. Các mẫu tham chiếu được tiếp tục sửa đổi như sau:

$$p[-1][y] = ((63 - y) * p[-1][-1] + (y + 1) * p[-1][63] + 32) \gg 6$$

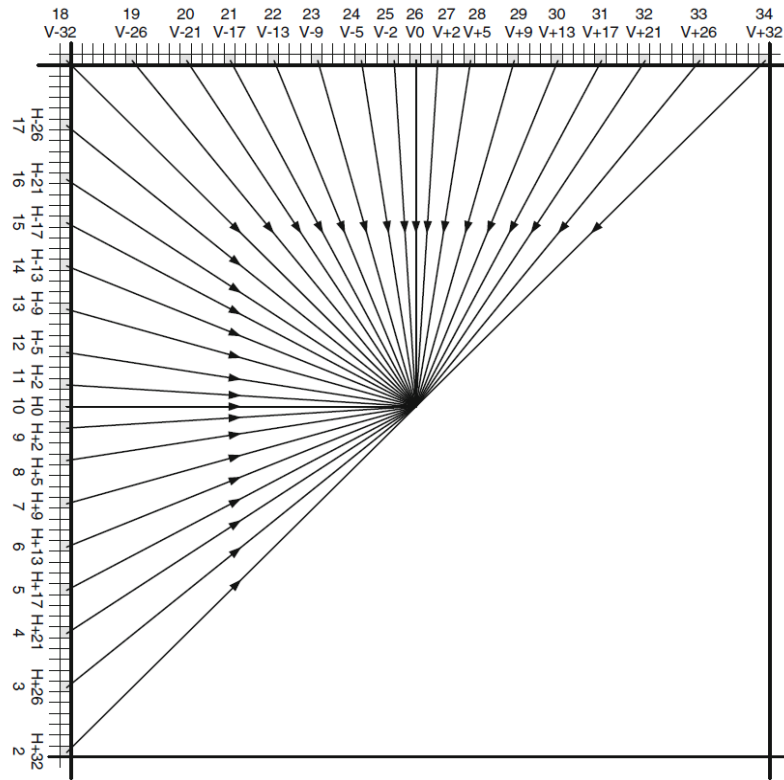
$$p[x][-1] = ((63 - x) * p[-1][-1] + (x + 1) * p[63][-1] + 32) \gg 6$$



Hình 3-3: Quá trình lọc mịn (a) Bước thứ nhất (b) bước thứ hai[1]

3.1.2. Dự đoán mẫu nội ảnh

3.1.2.1. Dự đoán góc



Hình 3-4: Các chế độ góc trong dự đoán nội ảnh [1]

H.265/HEVC cung cấp 33 chế độ dự đoán góc, với độ chính xác là 1/32 mẫu. Việc tăng số lượng các chế độ giúp tăng hiệu quả của H.265/HEVC trong việc dự đoán hướng phù hợp nhất cho các mẫu dự đoán, giảm dư thừa mẫu dự đoán góc.

Để giảm bớt sự phức tạp của dự đoán nội ảnh trong kỹ thuật H.265/HEVC, các mẫu tham chiếu trên $p[x][-1]$ và khối $p[-1][y]$ được sửa đổi, sau đó được sắp xếp thành mảng một chiều, bằng các phương trình thay thế:

- Chế độ dọc:

$$ref[x] = \begin{cases} p[-1+x][-1] & , \quad x \geq 0 \\ p[-1][-1 + ((x * B + 128))8] & , \quad x < 0 \end{cases}$$

- Chế độ ngang:

$$ref[y] = \begin{cases} p[-1][-1 + y] & , \quad y \geq 0 \\ p[-1 + ((y * B + 128))8][-1] & , \quad y < 0 \end{cases}$$

Trong đó B là góc nghịch đảo của tham số góc A. Cụ thể, tham số A cho biết số 1/32 đơn vị lưới mẫu mỗi hàng của mẫu phải được đổi chỗ với hàng trước đó.

Horizontal modes																
Mode	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
A	32	26	21	17	13	9	5	2	0	-2	-5	-9	-13	-17	-21	-26

Vertical modes																
Mode	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
A	-32	-26	-21	-17	-13	-9	-5	-2	0	2	5	9	13	17	21	26

Bảng 3-2: Giá trị của tham số A[1]

A	-32	-26	-21	-17	-13	-9	-5	-2
B	-256	-315	-390	-482	-630	-910	-1638	-4096

Bảng 3-3: Giá trị của tham số B[1]

Sau khi tạo ra mảng mẫu tham chiếu, quá trình tiếp theo là xây dựng các mẫu dự đoán. Sử dụng phương pháp nội suy, các mẫu dự đoán được tạo ra như sau:

- Chế độ ngang:

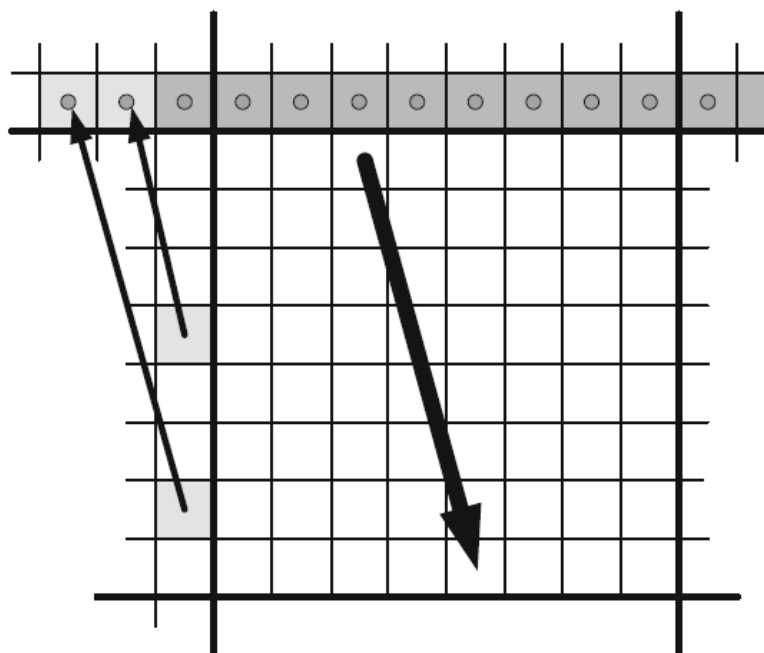
$$p[x][y] = ((32 - f) * ref[y + i + 1] + f * ref[y + i + 2] + 16) \gg 5$$

Trong đó $i = ((x + 1) * A) \gg 5$ và $f = ((x + 1) * A) \& 31$

- Chế độ dọc:

$$p[x][y] = ((32 - f) * ref[x + i + 1] + f * ref[x + i + 2] + 16) \gg 5$$

Trong đó $i = ((y + 1) * A) \gg 5$ và $f = ((y + 1) * A) \& 31$



Hình 3-5: Ví dụ về việc đổi chỗ mẫu tham chiếu bên trái đếm ở rộng hàng tham chiếu dọc ở chế độ nội ảnh 23[1]

3.1.2.2. Dự đoán DC và Dự đoán Planar

Nếu chế độ dự đoán DC được chọn, mẫu được dự đoán là giá trị trung bình của hai mẫu tham chiếu: ngay bên trái và phía trên của khối đang được dự đoán hiện tại. Mục đích của dự đoán DC là làm mềm các cạnh trái và đỉnh của khối đang được dự đoán.

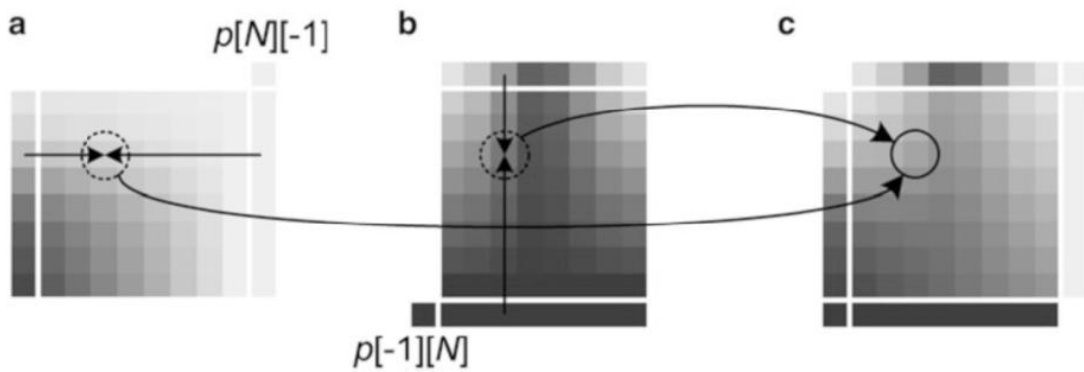
Dự đoán Planar của H.265/HEVC được thiết kế để làm mịn các khối cứng nhắc quan sát được khi chế độ dự đoán DC được áp dụng hoặc video có tốc độ bit thấp hơn. Phương pháp dự đoán planar là để dự đoán ra một bề mặt mà không làm ngắt quãng các ranh giới. Giá trị của các mẫu được tạo ra là giá trị trung bình, như sau:

$$p[x][y] = (p_h[x][y] + p_v[x][y] + N) \gg (\log_2(N) + 1)$$

Trong đó:

$$p_h[x][y] = (N - 1 - x) * p[-1][y] + (x + 1) * p[N][-1]$$

$$p_v[x][y] = (N - 1 - y) * p[x][-1] + (y + 1) * p[-1][N]$$



Hình 3-6: Ví dụ về việc sử dụng chế độ dự đoán planar nội ảnh

(a) Tính thành phần ngang (b) Tính thành phần dọc

(c) giá trị trung bình của (a) và (b) [1]

3.1.2.3. Hậu xử lý các mẫu dự đoán

Còn tồn tại một số ranh giới bị gián đoạn của một số mẫu dự đoán nội ảnh, đặc biệt là với chế độ dự đoán DC, dự đoán dọc trực tiếp (chế độ góc 26) hoặc dự đoán ngang (chế độ góc 10). Sự gián đoạn xảy ra dọc ranh giới khối. Hậu xử lý là quá trình này được thực hiện trong H.265/HEVC sau khi thực hiện dự đoán nội ảnh để giải quyết những vấn đề này.

3.1.3. Chế độ mã hóa nội ảnh

Dự đoán nội ảnh trong H.265/HEVC có không chỉ giúp dự đoán tốt hơn vì chế độ số hướng đã tăng lên 35, mà còn đảm bảo chế độ đã chọn được truyền một cách chính xác chỉ với lượng dữ liệu tối thiểu.

3.1.3.1. Chế độ dự đoán nội ảnh sáng

Đối với các thành phần sáng, so với H.264, H.265/HEVC có ba chế độ thay vì một. Ba chế độ được dựa trên các chế độ bên trái và PU lân cận. Nếu một trong các khối được mã hóa là chế độ tín hiệu, hoặc được mã hóa là chế độ điều chế xung mã (PCM – Pulse Code Modulation), khối hiện thời sẽ được thiết lập áp dụng chế độ DC. Giả sử rằng các chế độ dự đoán nội ảnh của các khối bên trái và trên là A và B, tương ứng. Trong trường hợp đó A hoặc B nằm ngoài CTU, khối được dự đoán với chế độ DC.

Nếu A khác với B, chế độ đầu tiên và thứ hai dễ xảy ra nhất được thiết lập bằng với A và B. Chế độ cuối cùng MPM [2] được xác định như sau:

- Chế độ Planar, nếu A hoặc B không ở chế độ Planar
- Chế độ DC, nếu A hoặc B không ở chế độ DC mode.
- Chế độ dọc (chế độ góc 26), nếu A hay B được dự đoán với chế độ DC và khối còn lại được dự đoán với chế độ planar.

Nếu A có chế độ dự đoán nội ảnh giống B thì 3 chế độ dự đoán có khả năng cao nhất được quy định như sau:

- Nếu A và B cả hai ở chế độ không góc cạnh. Thì:
 - MPM[0] là chế độ planar;
 - MPM[1] là chế độ DC;
 - MPM[2] là chế độ góc 26
- Nếu A và B cả hai ở chế độ góc. Thì:
 - $MPM[0] = A$
 - $MPM[1] = 2 + ((A - 2 - 1 + 32) \% 32)$
 - $MPM[2] = 2 + ((A - 2 - 1) \% 32)$

Sau khi tạo ra MPM, các chế độ này được sắp xếp theo thứ tự tăng dần của số chế độ. Nếu chế độ dự đoán nội ảnh của khối hiện thời bằng chính xác với một phần tử của MPM, chỉ có chỉ số của phần tử đó được truyền đến bộ giải mã. Nếu không, chế độ dự đoán được biến đổi thành một mã từ CABAC (Context-Adaptive Binary Arithmetic Coding) 5-bit, sau đó truyền đến bộ giải mã.

3.1.3.2. Chế độ dự đoán nội ảnh màu

Trong mã hóa video, gần như tất cả các trường hợp, chế độ dự đoán nội ảnh của mẫu màu diễn ra sau mẫu sáng. Theo kinh nghiệm, H.265/HEVC áp dụng một phương pháp mới để xác định chế độ mẫu màu. Cụ thể, chế độ dự đoán nội ảnh của mẫu màu là một trong năm chế độ: phẳng, góc 26, góc 10, DC hoặc chế độ dẫn xuất. Chế độ dẫn xuất là chế độ dự đoán nội ảnh sáng. Phương pháp này được sử dụng để tăng khả năng

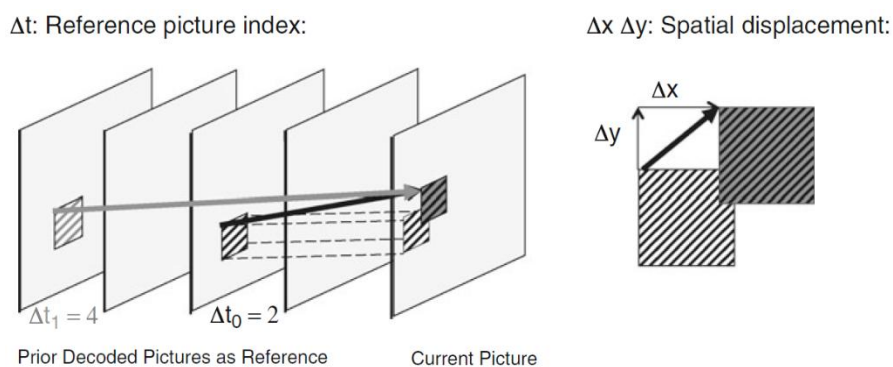
định hướng trong dự đoán màu nội ảnh và hạn chế tối đa các dữ liệu báo hiệu. Cụ thể hơn, nó thể hiện trong hình dưới đây:

Initial chroma intra mode	Final chroma intra mode when derived mode \neq initial chroma intra mode	Final chroma intra mode when derived mode = initial chroma intra mode
INTRA_PLANAR	INTRA_PLANAR	INTRA_ANGULAR[34]
INTRA_ANGULAR[26]	INTRA_ANGULAR[26]	INTRA_ANGULAR[34]
INTRA_ANGULAR[10]	INTRA_ANGULAR[10]	INTRA_ANGULAR[34]
INTRA_DC	INTRA_DC	INTRA_ANGULAR[34]
INTRA_DERIVED	Luma intra mode	N/A

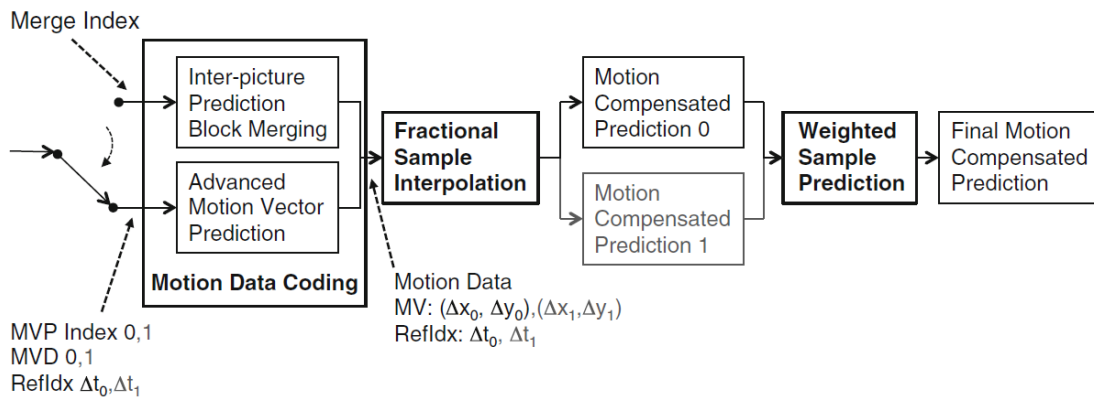
Bảng 3-4: Chế độ dự đoán nội ảnh màu dựa theo chế độ dự đoán nội ảnh[1]

3.2. Dự đoán liên ảnh

Trong khi dự đoán nội ảnh sử dụng các mối tương quan không gian của các khối liền kề để dự đoán khối hiện tại, dự đoán liên ảnh sử dụng các khối được mã hóa của các hình ảnh trước. Ví dụ một vật chuyển động trong một video, hình ảnh của nó là như nhau, nhưng nó lại ở những vị trí mới khi di chuyển từ hình ảnh này đến hình ảnh khác, vì vậy bộ giải mã có thể dự đoán chính xác vật chuyển động nếu nó có hình ảnh trước đó và một vector chuyển động. Do đó, dự đoán liên ảnh một khối là một quá trình lựa chọn các khối ứng viên phù hợp nhất và vector chuyển động liên quan của nó. Khái niệm cơ bản của dự đoán liên ảnh được giải thích trong hình bên dưới, trong đó Δt là chỉ số hình ảnh tham chiếu trước, Δx và Δy là các hệ số ngang và dọc của vector bù chuyển động, tương ứng.



Hình 3-7: Khái niệm cơ bản về dự đoán liên ảnh[1]



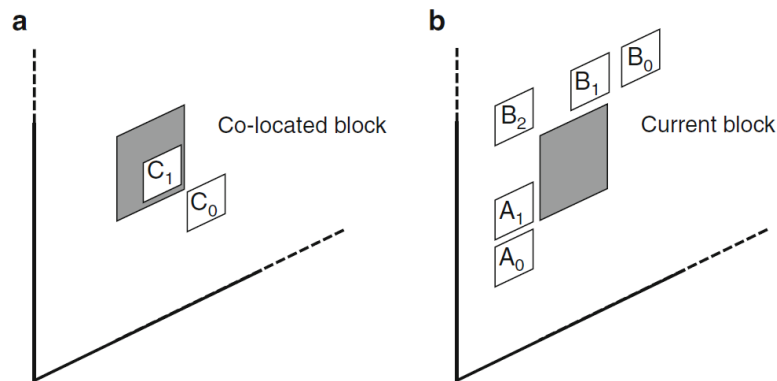
Hình 3-8: Quá trình dự đoán liên ảnh trong H.265/HEVC[1]

Mặc dù dự đoán liên ảnh trong H.265/HEVC về cơ bản giống như người tiền nhiệm H.264, nó vẫn có hai kỹ thuật cải tiến mới. Đầu tiên là cách tiếp cận mới, được gọi là dự đoán vector chuyển động tiên tiến (AMVP – Advance Motion Vector Prediction), là quá trình cải tiến về việc tạo ra vector chuyển động. Thứ hai là kỹ thuật nhập khối dự đoán liên ảnh, trong đó khối cử viên thay thế hoàn toàn khối hiện thời, nếu chế độ dự đoán của nó được quyết định là chế độ trực tiếp hoặc bỏ qua như H.264. Hình 24 thể hiện quá trình dự đoán liên ảnh trong H.265/HEVC.

3.2.1. Dự đoán vector chuyển động tiên tiến (AMVP)

Dự đoán vector chuyển động được sử dụng để có được dự đoán vector bù chuyển động (MCP – Motion Compensated Prediction). Như ví dụ trên về một đối tượng chuyển động, vì đối tượng này thường lớn hơn một khối, do đó đối tượng đang chuyển động có nghĩa là một nhóm các khối lân cận cũng đang chuyển động. Vì vậy, vector chuyển động của khối hiện tại là gần giống với khối liền kề của nó. Đó là lý do tại sao vector dự đoán chuyển động (MVP – Motion Vector Predictor) thường được tạo ra từ các khối liền kề về mặt không gian. Kỹ thuật dự đoán vector chuyển động đã được áp dụng trong hầu hết tất cả các tiêu chuẩn mã hóa trước. Trong H.265/HEVC, vector được dẫn xuất bằng cách thực hiện một kỹ thuật mới, gọi là cạnh tranh vector chuyển động. Bộ mã hóa tạo ra một danh sách các MVP, sau đó chọn một trong số chúng để truyền đến bộ giải mã. Dự đoán vector chuyển động tiên tiến (AMVP) tăng khả năng linh hoạt của kỹ thuật cạnh tranh vector chuyển động.

Bước đầu tiên trong AMVP là tạo ra một danh sách vector gồm 5 ứng viên. Những ứng viên được lựa chọn là: ba trong năm vector chuyển động của khối không gian lân cận ($A_0, A_1,$ and B_0, B_1, B_2), thứ 4 là vector trung bình của ba vector trên, và vector chuyển động của khối cùng vị trí với khối hiện tại nhưng của ảnh trước (C_0 or C_1). Những khối được đề cập ở trên nếu được mã hóa là dự đoán nội ảnh thì sẽ bị loại bỏ.

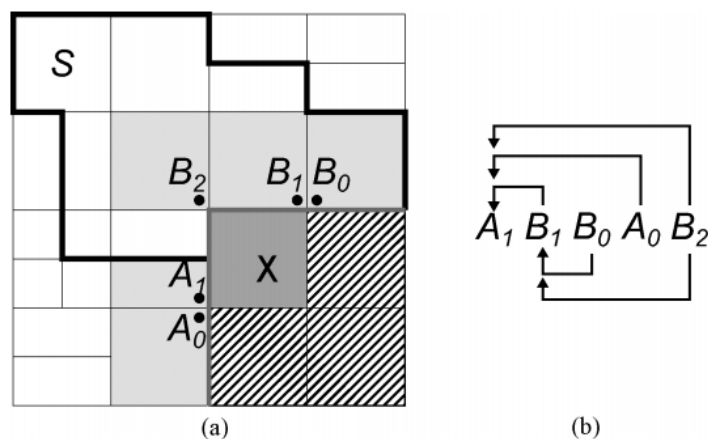


Hình 3-9: Các khối có các ứng viên vector chuyển động[1].

Sau đó, sau khi phân loại danh sách theo thứ tự khả năng chính xác, loại bỏ một số yếu tố ở vị trí sau trong danh sách, danh sách ứng viên chính thức có hai MVP, là một hoặc hai ứng viên không gian (vector chuyển động liên quan tới các khối lân cận), một trong những ứng viên thời gian (vector chuyển động liên quan tới khối cùng vị trí ở ảnh trước) nếu hai ứng viên không gian giống nhau hoặc không có sẵn, hoặc vector không nếu tất cả các vector ứng viên trước đều bị loại bỏ.

3.2.2. Nhập khối dự đoán liên ảnh

Do cấu trúc phân vùng dạng cây của H.265/HEVC, mỗi khối mã hóa có thông tin dự đoán của riêng mình, được truyền độc lập với bộ mã hóa đến bộ giải mã. Đây là nhược điểm của cấu trúc này. Vẫn tồn tại một số khối có chế độ dự đoán giống nhau, do đó sẽ không hiệu quả nếu bộ mã hóa ra hiệu lệnh cho tất cả thông tin. Để giảm bớt các tập tham số mã hóa dư thừa, kỹ thuật nhập khối dự đoán liên ảnh được áp dụng. Kỹ thuật này tập hợp tất cả các khối lân cận có cùng thông tin chuyển động. Quá trình sáp nhập ở chừng mực nào đó giống như AMVP. Bước đầu tiên là lựa chọn danh sách ứng viên hợp nhất, sau đó truyền những chỉ số xác định ứng viên đến bộ giải mã.



Hình 3-10: (a) Các ứng viên có thể chọn trong danh sách sát nhập của khối X, chúng sáp nhập thành cùng dòng đậm, (b) Các bước kiểm tra dư để thêm ứng viên vào danh sách. [1]

Danh sách ứng viên hợp nhất chỉ có một tính năng khác với các danh sách vector ứng viên chuyển động. Đó là thành phần ứng viên hợp nhất phải chứa tất cả các dữ liệu chuyển động có tất cả các thông tin dự đoán, chẳng hạn như loại dự đoán liên ảnh (P hoặc B), trong khi danh sách vector ứng viên chuyển động chỉ chứa các vector chuyển động. Các ứng viên hợp nhất được lựa chọn bởi:

- Tối đa bốn ứng viên được chọn từ năm khối lân cận không gian (A₁, B₁, B₀, A₀, B₂)
- Một ứng viên thời gian, giống như ứng viên thời gian trong danh sách ứng viên vector chuyển động.
- Các ứng viên bổ sung được tạo ra bằng cách kết hợp thông tin chuyển động đã tồn tại trong danh sách, hoặc không có ứng cử viên nào. Ví dụ về các ứng viên bổ sung được minh họa trong hình 2-20.
- Tất cả các ứng viên được mã hóa sử dụng dự đoán liên ảnh.

3.2.3. Nội suy mẫu phân số

Tương tự với H.264, H.265/HEVC hỗ trợ dự đoán với độ chính xác một phần tám pixel với thành phần màu, một phần tư pixel với thành phần sáng. Cụ thể, phép nội suy mẫu phân số được sử dụng để tạo ra các mẫu tham chiếu nếu các hệ số vectơ dự đoán chuyển động không phải là số nguyên.

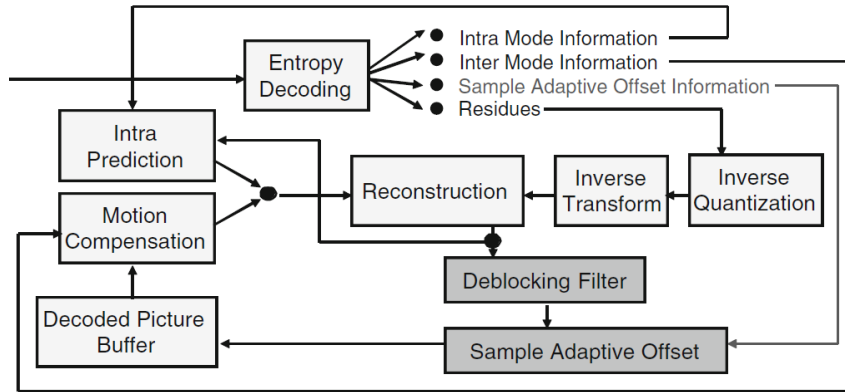
Quá trình nội suy tạo mẫu tứ điểm ảnh của H.264 có 3 giai đoạn. Giai đoạn đầu tiên là tạo ra nửa mẫu, sử dụng bộ lọc 6-tap, có độ trễ pha không đổi là 0,5, sau đó làm tròn kết quả trước khi trung bình nó với mẫu số nguyên gần nhất để tạo tứ điểm ảnh giữa chúng. Quá trình này được thay đổi trong H.265/HEVC. Các mẫu tứ điểm ảnh được tạo ra trực tiếp bằng cách sử dụng bộ lọc 7-tap hoặc 8-tap trên các điểm ảnh nguyên. Điều này giúp tăng hiệu quả mã hóa trong dự đoán, độ chính xác cao vì loại bỏ quá trình làm tròn, lỗi làm tròn.

$A_{-1,-1}$				$A_{0,-1}$	$a_{0,-1}$	$b_{0,-1}$	$c_{0,-1}$	$A_{1,-1}$				$A_{2,-1}$
$A_{-1,0}$				$A_{0,0}$	$a_{0,0}$	$b_{0,0}$	$c_{0,0}$	$A_{1,0}$				$A_{2,0}$
$d_{-1,0}$				$d_{0,0}$	$e_{0,0}$	$f_{0,0}$	$g_{0,0}$	$d_{1,0}$				$d_{2,0}$
$h_{-1,0}$				$h_{0,0}$	$i_{0,0}$	$j_{0,0}$	$k_{0,0}$	$h_{1,0}$				$h_{2,0}$
$n_{-1,0}$				$n_{0,0}$	$p_{0,0}$	$q_{0,0}$	$r_{0,0}$	$n_{1,0}$				$n_{2,0}$
$A_{-1,1}$				$A_{0,1}$	$a_{0,1}$	$b_{0,1}$	$c_{0,1}$	$A_{1,1}$				$A_{2,1}$
$A_{-1,2}$				$A_{0,2}$	$a_{0,2}$	$b_{0,2}$	$c_{0,2}$	$A_{1,2}$				$A_{2,2}$

Hình 3-11: Vị trí mẫu và phân số để nội suy [1]

Ví dụ, $a_{0,0}$ được tạo ra bằng cách sử dụng bộ lọc 7-tap, như $c_{0,0}$, $e_{0,0}$, $g_{0,0}$... trong khi $b_{0,0}$, $f_{0,0}$.. được tạo ra bằng cách sử dụng bộ lọc 8-tap, như thể hiện trong hình 2-21. Do sử dụng bộ lọc có số tap nhiều hơn, H.265/HEVC có thể dự đoán chính xác hơn so với bộ lọc 6-tap sử dụng trong H.264, nhưng nó làm tăng số lượng các dữ liệu cần phải được lưu cũng như sự phức tạp của bộ mã hóa. Vì thế. H.265/HEVC có một hạn chế để giảm băng thông bộ nhớ. Thay vì mẫu 4x4, kích thước khối dự đoán nhỏ nhất là 8x4 hoặc 4x8, để có thể được dự đoán đơn lẻ.

3.3. Bộ lọc vòng trong



Hình 3-12: De-blocking filter và SAO trong H.265/HEVC[1]

Bộ lọc vòng trong là một trong những phần quan trọng nhất của H.265/HEVC. Hai bộ lọc vòng trong là deblocking filter và độ lệch tương thích mẫu (SAO – Sample Adaptive Offset). Do H.265/HEVC là kỹ thuật mã hóa video hybrid dựa theo khối, nó tạo ra các khối và gián đoạn ở ranh giới dự đoán và khối trong suốt toàn bộ quá trình mã hóa. Bộ lọc deblocking làm giảm tất cả các nhiễu khối vuông, trong khi SAO làm giảm các nhiễu rung và thay đổi cường độ của một số mẫu. Cả hai đều được áp dụng như là bước cuối cùng sau quá trình tái xây dựng hình ảnh trước khi lưu vào bộ đệm hình ảnh được giải mã, theo thứ tự SAO được thực hiện sau bộ lọc deblocking filter. Ngoài ra, bộ lọc deblocking filter và SAO không phụ thuộc vào cấu hình của bộ mã hóa. Mục đích của bộ lọc vòng trong là nâng cao chất lượng của hình ảnh được giải mã, do đó nâng cao hiệu quả nén.

Trong H.265/HEVC, bộ lọc deblocking filter được sử dụng trên ranh giới giữa các đơn vị mã hóa (CU), các đơn vị dự đoán (PU) và các đơn vị biến đổi (TU). Việc quyết định liệu khối có được lọc hay không dựa vào chế độ dự đoán, các khối liền kề, các vector chuyển động, cụ thể hơn một biến gọi là sức mạnh ranh giới (Bs – Boundary strength).

- Bs = 2 nếu tồn tại ít nhất một trong các khối liền kề trong ảnh.
- Bs = 1 nếu:
 - o Một trong các khối liền kề có hệ số biến đổi khác không

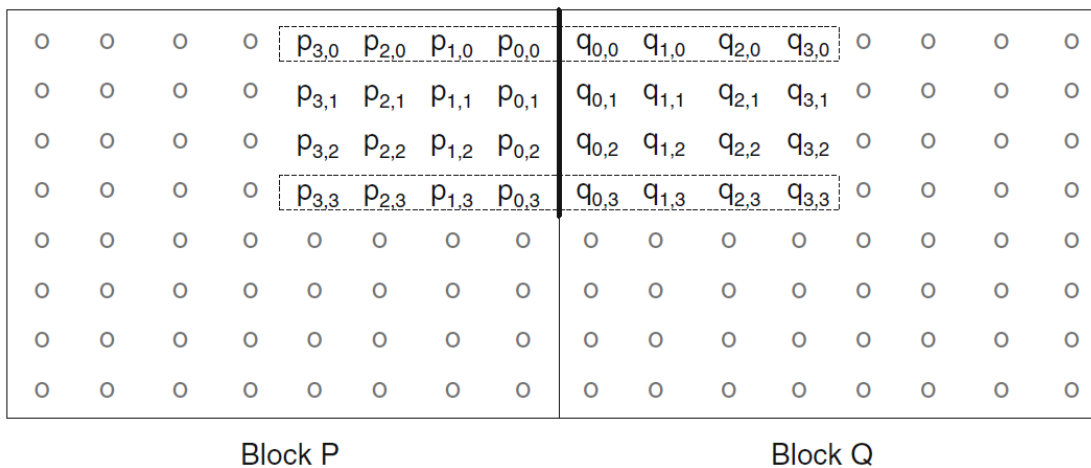
- Các khối liền kề có khác biệt tuyệt đối về vectơ chuyển động đó là lớn hơn một mẫu sáng số nguyên
 - Các khối liền kề có các vectơ chuyển động liên quan đến các hình ảnh khác nhau, hoặc số lượng vector giữa các khối không giống nhau.
- Nếu không, $B_s = 0$.

Trong H.264, B_s có thể có 5 mức $\{0, 1, 2, 3, 4\}$ trong nhiều hoàn cảnh, trong khi H.265/HEVC chỉ hỗ trợ 3 mức B_s . Ngoài ra, vì kích thước khối lớn hơn, có nghĩa là số khối nhỏ hơn, sự phức tạp của bộ lọc deblocking filter trong H.265/HEVC cũng giảm đi so với H.264.

Bộ lọc deblocking filter được sử dụng trên ranh giới của khối nếu B_s của các khối lớn hơn 0. Tuy nhiên, cần thực hiện một bước thẩm định trước khi sử dụng bộ lọc. Đối với các ranh giới khối chẵn, các căn cứ quyết định bộ lọc dựa trên mỗi phân đoạn bốn mẫu. Theo biểu thức sau:

$$|p_{2,0} - 2p_{1,0} + p_{0,0}| + |p_{2,3} - 2p_{1,3} + p_{0,3}| + |q_{2,0} - 2q_{1,0} + q_{0,0}| + |q_{2,3} - 2q_{1,3} + q_{0,3}| < \beta$$

Trong đó β là các căn cứ ngưỡng về tham số số lượng tử hóa.



Hình 3-13: 4 mẫu và vị trí của chúng giữa vùng ranh giới của 2 khối P và Q[1]

Có 2 chế độ lọc deblocking filter: chế độ bình thường và chế độ mạnh. Chế độ mạnh được áp dụng nếu cả ba biểu thức dưới đây là đúng:

$$|p_{2,i} - 2p_{1,i} + p_{0,i}| + |q_{2,i} - 2q_{1,i} + q_{0,i}| < \frac{\beta}{8}$$

$$|p_{3,i} - p_{0,i}| + |q_{0,i} - q_{3,i}| < \frac{\beta}{8}$$

$$|p_{0,i} - q_{0,i}| < 2.5t_c$$

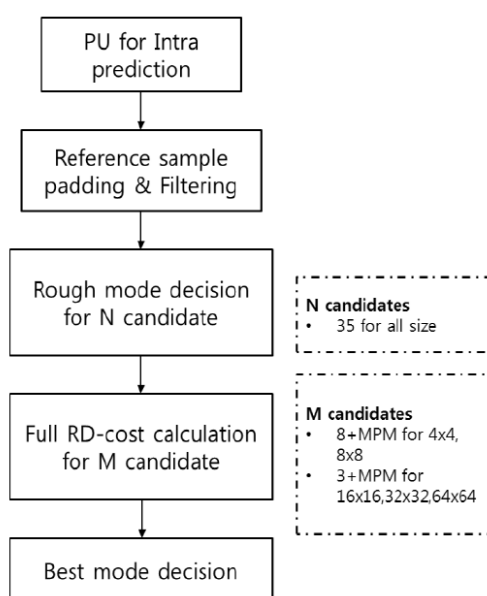
Chức năng chính của SAO là giảm nhiễu rung giữa các ảnh trước khi ảnh được lưu vào bộ đệm (Decoded Picture Buffer). Nếu bộ lọc SAO được kích hoạt, quá trình của nó sẽ được thực hiện sau bộ lọc de-blocking. Trong H.265/HEVC, bộ lọc SAO cũng có 2 chế độ là EO (Edge Offset) và BO (Band Offset).

3.4. Kết luận

Dự đoán nội ảnh và dự đoán trong ảnh là hai kỹ thuật xử lý quan trọng nhất trong nén video của tất cả các chuẩn từ trước tới nay. Trong H.265/HEVC, dự đoán nội ảnh đã được cải tiến hơn nhiều so với H.264/AVC để thích hợp với mã hóa nén video có độ phân giải cao. Đầu tiên, tập hợp kích thước của block dự đoán đã được mở rộng tới 32x32, trong khi H.264/AVC chỉ là 16x16. Cùng với cấu trúc khối thay đổi, khối dự đoán sẽ được tái cấu trúc giúp hình ảnh được trơn tru hơn. Thứ hai là số hướng trong dự đoán góc tăng từ 8 tới 33 hướng. Độ phức tạp tính toán cao hơn nhưng bù lại dự đoán nội ảnh trong H.265/HEVC sẽ dự đoán chuẩn xác hơn H.264/AVC. Nhiều block ở ranh giới khối cũng được xử lý tốt hơn với bộ lọc nhiều block, bộ lọc vòng trong. Xử lý thiết lập mẫu tham chiếu giúp cho dự đoán nội ảnh có thể được xử lý song song, đặc biệt là những mẫu ở ranh giới các tile khi sử dụng kỹ thuật song song tile. Dự đoán liên ảnh trong H.265/HEVC chỉ phát triển hơn một chút so với H.264/AVC. Vector chuyển động được dự đoán nhanh và chính xác hơn với kỹ thuật AMVP. Kỹ thuật nhập khối dự đoán liên ảnh giúp cho xử lý giải mã nhanh hơn và hình ảnh sẽ mượt mà hơn ở ranh giới khối khi đã được nhập. Cuối cùng là xử lý dự đoán trực tiếp mẫu phân số của H.265/HEVC giảm đáng kể sai số mẫu so với chuẩn mã hóa trước, H.264/AVC.

Những sửa đổi đề xuất về dự đoán nội ảnh

4.1. Chế độ dự đoán nội ảnh cơ bản trong H.265/HEVC



Hình 4-1: Quá trình quyết định chế độ dự đoán nội ảnh trong H.265/HEVC[5]

Phần này giải thích chi tiết quá trình lựa chọn chế độ dự đoán nội ảnh của một PU. Con số trên cho thấy toàn bộ bước quyết định chế độ dự đoán nội ảnh trong H.265/HEVC. Sau khi đã tạo ra các khối tham chiếu, tất cả 33 chế độ góc cạnh được móc nối lại để tính toán quyết định chế độ thô (RMD - Rough Mode Decision) bằng cách biến đổi Hadamard, trong khi H.264/AVC sử dụng cách biến đổi cosin rời rạc (DCT - Discrete Cosine Transform). Nó làm giảm sự phức tạp của thuật toán.

Trong quá trình RMD, biểu thức là:

$$C_{RMD} = HSAD + \lambda \cdot R_{mode}$$

Trong đó,

$$HSAD = \sum_i^W \sum_j^H |H(c_{ij} - p)H^T|$$

HSAD là tổng tuyệt đối của số dư biến đổi Hadamard, R_{mode} là các bit chế độ dự đoán, λ là Lagrange multiplier, c_{ij} là khối hiện tại và p là yếu tố dự đoán là điểm ảnh lân cận tương ứng với chế độ dự đoán. W và H là chiều rộng và chiều cao của khối hiện tại c_{ij} tương ứng. Ma trận H , là hạt nhân biến Hadamard, được xác định như sau:

$$H_{4 \times 4} = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{vmatrix}$$

$$H_{8 \times 8} = \begin{vmatrix} H_{4 \times 4} & H_{4 \times 4} \\ H_{4 \times 4} & -H_{4 \times 4} \end{vmatrix}$$

Bước cuối cùng là tính chi phí biến dạng theo tỷ lệ đủ (C_{FRD}):

$$C_{FRD} = SSD + \lambda \cdot R_{bits}$$

Trong đó SSD (Sum of Square Difference) là tổng chênh lệch vuông giữa khối đã được tái xây dựng và khối ban đầu, và R_{bits} là số bit của khối hiện tại. Sau khi thực hiện tất cả các bước, tất cả các giá trị CFRD được sắp xếp trong thứ tự tăng dần. Các chế độ tốt nhất được sắp xếp ở đầu danh sách. Có 8 chế độ ứng viên tốt nhất cho khối 8x8 và 4x4, các khối khác có 3 chế độ tốt nhất.

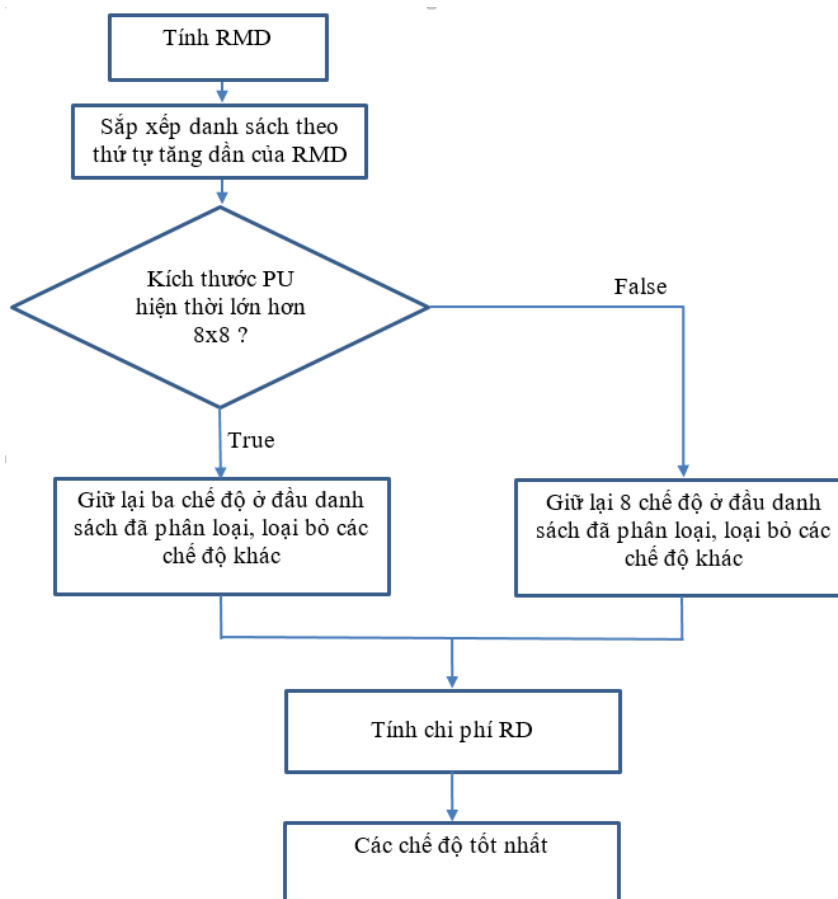
4.2. Thuật toán chọn chế độ nội ảnh nhanh

Phần 4.1 thể hiện tính phức tạp và mức tiêu thụ thời gian của chế độ dự đoán nội ảnh cơ bản của H.265/HEVC. Cụ thể, có hai lần 35 hướng. Vòng một là tính RMD, sau đó vòng hai là để đánh giá chi phí RD, và giữa mỗi vòng lặp là hai quá trình phân loại. Ngoài ra, sự phức tạp trong tính toán chi phí RD rất cao.

Do đó, để cải thiện tốc độ lựa chọn chế độ nội ảnh trong H.265/HEVC. Một lý tưởng đã được tìm ra. Trong [5] và [6] đều đồng ý rằng chế độ nội ảnh tốt nhất của một khối có mối quan hệ với danh sách đã được sắp xếp sau khi tính RMD. Tỷ lệ lựa chọn các chế độ tốt nhất trong danh sách là 90%.

Như đề cập ở trên, có 3 chế độ nội ảnh tốt nhất cho các khối, có kích thước là 8x8 và 4x4, và 8 chế độ cho các kích thước khác. Vì vậy, thuật toán cơ bản được thay đổi như sau: sau khi tính toán và phân loại tất cả 35 kết quả RMD, một số chế độ ở trên cùng của danh sách đã sắp xếp được giữ để đánh giá chi phí RD phụ thuộc vào kích thước của khối hiện tại (PU). Các chế độ khác đều bị loại.

Thuật toán đề xuất là:



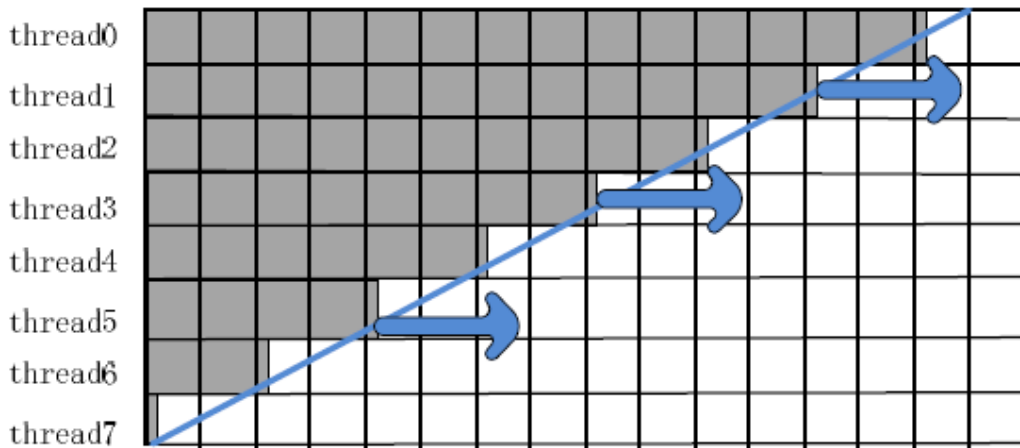
Hình 4-2: Quá trình lựa chọn chế độ nội ảnh đã được sửa đổi

Áp dụng thuật toán mới này, số lần lặp được giảm nhiều nhất có thể, giống như tính phức tạp của quá trình tính chi phí RD. Thuật toán này có thể đơn giản hơn nếu nó giữ 8 chế độ ứng viên mà không tính đến kích thước của PU, hoặc thay đổi số chế độ để tăng cơ hội tìm được chế độ tốt nhất.

4.3. Quá trình dự đoán nội ảnh nhanh bằng kỹ thuật song song

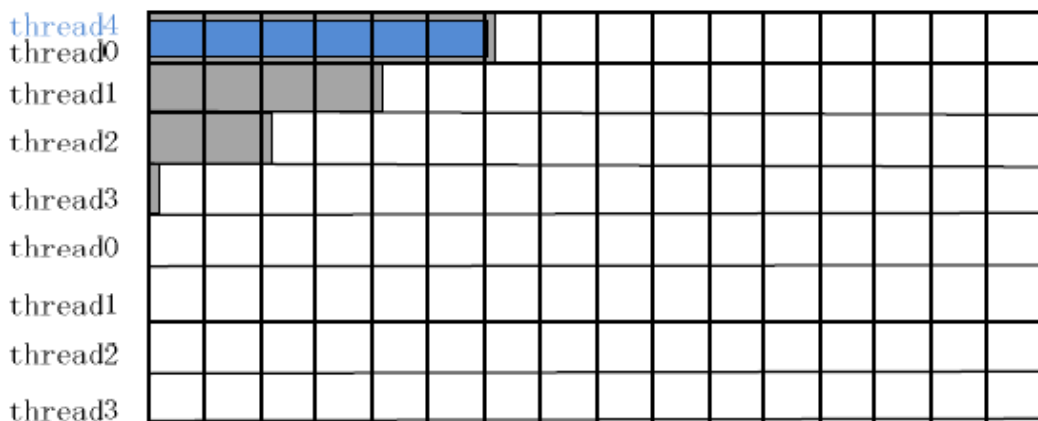
Trong phần 2.3.3, WPP được giải thích chi tiết. Nó là một trong những công cụ song song được sử dụng trong H.265/HEVC, và phù hợp với dự đoán nội ảnh. Một hình ảnh được chia thành nhiều hàng. Mỗi luồng nội ảnh được xử lý tất cả các CTU trong hàng đó, bắt đầu từ CTU ở bên trái sang cuối cùng bên phải. Do ràng buộc với khối tham chiếu trong ảnh, hàng được bắt đầu dự đoán ngay khi tất cả các khối trên của khối hiện thời đầu tiên kết thúc quá trình dự đoán. Ví dụ, tất cả các khối được chia

cùng kích thước, quá trình dự đoán hàng hiện tại có thể bắt đầu ở cùng thời điểm với khối thứ ba của hàng trên, xem hình 4-3.



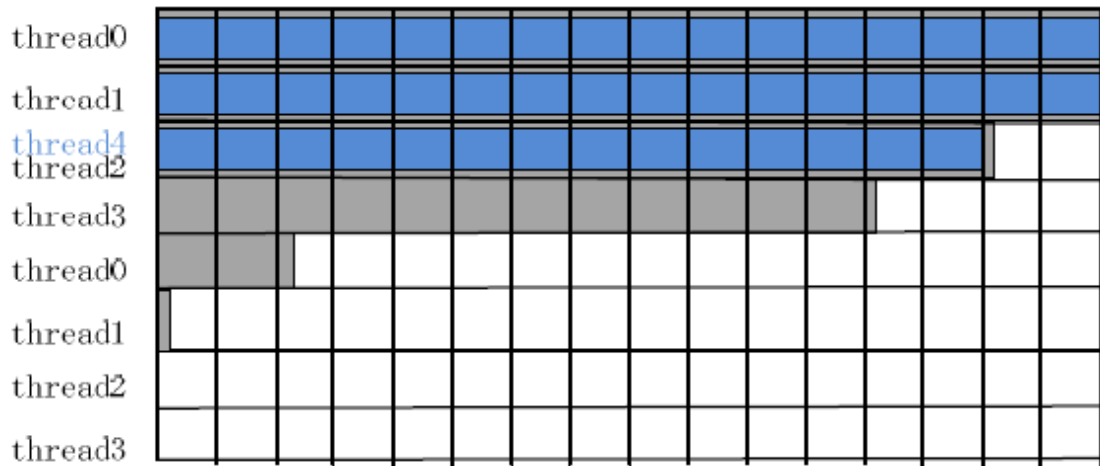
Hình 4-3: WPP bước thứ nhất

H.265/HEVC được thiết kế để thực hiện song song. Hầu hết tất cả các quy trình, đơn vị thuộc H.265/HEVC có thể thực hiện song song. Phần này giải thích quá trình song song. Quá trình này bao gồm 2 bước: xử lý song song và mã hóa entropy. Bước đầu tiên là WPP, là quá trình dự đoán nội ảnh của một hàng. Ngay sau khi hoàn thành bước đầu tiên, thông tin nội ảnh được lưu lại, sau đó bước thứ hai được bắt đầu bằng việc tạo ra một luồng mới để mã hóa entropy. Theo cách đơn giản, phương pháp kết hợp quá trình dự đoán nội ảnh với quá trình mã hóa entropy.



Hình 4-4: Luồng 4 bắt đầu quá trình đồng thời với bước đầu tiên

Thuật toán này có một số thuận lợi. Thuận lợi đầu tiên là tối đa hóa tỷ lệ tăng tốc ở bước WPP. Công cụ song song đẩy nhanh quá trình dự đoán nội ảnh. Thuận lợi thứ hai là giảm thiểu hao hụt hiệu suất. Quá trình mã hóa Entropy được xử lý theo thứ tự quét hàng tự để nó có thể được khởi động và chạy liên tục với các luồng WPP, do đó thuận lợi thứ ba là: tốc độ nhanh.



Hình 4-5: Luồng 4, mã hóa entropy đang chạy theo thứ tự quét

Chương 5

Tổng kết

5.1. Kết luận

Nói chung, tất cả các kỹ thuật thực hiện trong H.265/HEVC được cải tiến dựa trên H.264 AVC. Một số kỹ thuật phức tạp hơn so với kỹ thuật trước, ví dụ 64x64 CTU, dự đoán nội ảnh với 35 chế độ có sẵn, bộ lọc vòng trong có bộ lọc 7 hoặc 8 van, vv. Kích thước khối tăng lên đến tối đa 64x64, gấp 16 so với khối macro trong H264, là một trong những cải tiến quan trọng. Ngoài ra, H.265/HEVC được thiết kế để ứng dụng kỹ thuật song song. Tính năng mới nhất tuyệt vời nhất trong tiêu chuẩn nén video giúp đẩy nhanh thời gian mã hóa và giải mã, dựa trên công nghệ phần cứng đã được cải tiến. Hầu như tất cả các cú pháp và cấu trúc của H.265/HEVC đều hỗ trợ bộ mã hóa và giải mã một cách độc lập. Dù tính phức tạp trong thuật toán và tính toán, thời gian mã hóa và giải mã giảm nhờ kỹ thuật song song.

5.3. Hướng phát triển tiếp theo

Mặc dù H.265/HEVC được công bố vào đầu năm 2013, tiêu chuẩn mã hóa video phổ biến nhất vẫn là H.264. Gần đây, có rất nhiều nghiên cứu để nâng cấp H.265/HEVC, cả thuật toán cho phần mềm và phần cứng. Do đó, bài toán đặt ra là cần phải tìm ra một ý tưởng để phát triển H.265/HEVC trở lên tốt hơn, đặc biệt trong dự đoán nội ảnh là phần mà tôi tập trung phân tích.

Trong luận án, có những phân tích chỉ mang tính khái niệm mà không có mô hình và mô phỏng. Do đó, trong tương lai gần, điều phải làm là thực hiện những kỹ thuật đề xuất đã được đề cập trong chương trước. Đặc biệt, những bước phát triển sau, phần mềm mã nguồn mở HM (<https://hevc.hhi.fraunhofer.de/>) sẽ được sử dụng để thực hiện mô phỏng.

DANH MỤC TÀI LIỆU THAM KHẢO

- [1] . Vivienne Sze, Madhukar Budagavi, Gary J.Sullivan, “*High Efficiency Video Coding (HEVC) Algorithms and Architectures*”, USA.
- [2] . Gary J.Sullivan, Jens-Rainer Ohm, Woo-Jin Han, Thomas Weigand, “*Overview of the High Efficiency Video Coding (HEVC) standard*”, IEEE Transactions of recruiting.
- [3] . Marta Merlo Serrano, “*Improved Intra-Prediction for Video Coding*”, Master Thesis, Queen Mary University of London, July 2014.
- [4] . Frank Bossen, Benjamin Bross, Karsten Suhring, “*HEVC Complexity and Implementation Analysis*”, Circuits and Systems For Video Technology, Dec. 2012.
- [5] . Liang Zhao, Li Zang, Siwei Ma, Debin Zhao, “*Fast Mode Decision Algorithm for Intra Prediction in HEVC*”.
- [6] . Do Kyung Lee, Je-Chang Jeong, “*Fast Intra Coding by using RD Cost Candidate Elimination for High Efficiency Video Coding*”, the World Congress on Engineering and Computer science 2014, Vol. I, San Francisco, USA.
- [7] . Yanan Zhao, Li Song, Xiangwen Wang, Min Chen, Jia Wang, “*Efficient realization of parallel HEVC intra coding*”.
- [8] . Qin Yu, Liang Zhao, Siwei Ma, “*Parallel AMVP Candidate list construction for HEVC*”.
- [9] . Mauricio Alvarez Mesa, Chi Ching Chi, Thomas Schierl and Ben Juurlink, “*Evaluation of parallelization strategies for the emerging HEVC standard*”, Heinrich Hertz Institute, Berlin, Germany.
- [10] . H. Brahmasury Jain and K.R. Rao, “*Fast intra mode decision in High Efficiency Video Coding*”
- [11] . Heming Sun, Dajiang Zhou, Satoshi Goto, “*A low-complexity HEVC intra prediction algorithm based on level and mode filtering*”, IEEE International Conference on Multimedia and Expo, 2012.
- [12] . Roman I. Chernyak, “*Analysis of the intra prediction in H.265/HEVC*”, Mathematical Sciences, Vol. 8, 2014.
- [13] . Xu X., Cohen R., Vetro, Sun H., “*Predictive coding of intra prediction modes for High Efficiency Video Coding*”, Mistubishi electric research laboratories, May 2012.
- [14] . Haijun Lei, Zhongwang Yang, “*Fast intra prediction mode decision for High Efficiency Video Coding*”, 2nd International Symposium on Computer, Communication, Control and Automation, 2013.

- [15] . Shohei Matsuo, Seishi Takamura and Atsushi Shimizu, “*Modification of intra angular prediction in HEVC*”, NTT Media Intelligence Laboratories, NTT Corporation, Yokosuka, Japan.
- [16] . Younhee Kim, DongSan Jun, Soon-heung Jung, Jin Soo Choi, and Jinwoong Kim, “*A fast intra-prediction method in HEVC using rate-distortion estimation based on hadamard transform*”, ETRI Journal, Vol. 35, Num. 2, April 2013.
- [17] Thomas Wiegand, Gary J. Sullivan, “*The H.264/AVC Video Coding Standard*”, IEEE Signal processing magazine, 03/2007.