

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

NGUYỄN NGỌC SINH

**PHƯƠNG PHÁP XÁC ĐỊNH CHUYỂN ĐỘNG DỰA TRÊN
ĐẶC TÍNH CỦA MẮT NGƯỜI ĐỐI VỚI THÀNH PHẦN
XÁM TRONG VIỆC LOẠI BỎ DƯ THỪA VỀ MẶT THỜI
GIAN CỦA MÃ HÓA VIDEO MJPEG**

LUẬN VĂN THẠC SĨ Công nghệ Kỹ thuật Điện tử Truyền thông

Hà Nội - 2016

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

NGUYỄN NGỌC SINH

**PHƯƠNG PHÁP XÁC ĐỊNH CHUYỂN ĐỘNG DỰA TRÊN
ĐẶC TÍNH CỦA MẮT NGƯỜI ĐỐI VỚI THÀNH PHẦN
XÁM TRONG VIỆC LOẠI BỎ DƯ THỪA VỀ MẶT THỜI
GIAN CỦA MÃ HÓA VIDEO MJPEG**

Ngành: Công nghệ Kỹ thuật Điện tử Truyền thông

Chuyên ngành: Kỹ thuật Điện tử

Mã số: 60520203

LUẬN VĂN THẠC SĨ Công nghệ Kỹ thuật Điện tử Truyền thông

NGƯỜI HƯỚNG DẪN KHOA HỌC : PGS.TS Trần Xuân Tú

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn tốt nghiệp: **“Phương pháp xác định chuyển động dựa trên đặc tính của mắt người đối với thành phần xám trong việc loại bỏ dư thừa về mặt thời gian của mã hóa video MJPEG”** là công trình nghiên cứu của riêng tác giả. Các số liệu, kết quả trình bày trong luận văn là hoàn toàn trung thực, chưa từng được công bố trong các bất kỳ công trình nào khác.

Trong luận văn có dùng một số tài liệu tham khảo như đã nêu trong phần tài liệu tham khảo.

Tác giả luận văn

Nguyễn Ngọc Sinh

LỜI CẢM ƠN

Để hoàn thành luận văn tốt nghiệp này, trước tiên, tôi xin gửi lời cảm ơn chân thành và sâu sắc tới người Thầy của tôi PGS.TS Trần Xuân Tú. Thầy là người đã luôn theo sát tôi, tận tình chỉ bảo, góp ý và hướng dẫn, định hướng cho tôi trong suốt quá trình làm luận văn này tại PTN Trọng điểm Hệ thống Tích hợp Thông minh, Đại học Công nghệ. Tôi không chỉ được học ở Thầy phương pháp luận nghiên cứu khoa học, tôi còn tích lũy được rất nhiều bài học quý báu về cách làm việc chuyên nghiệp, lối tư duy đánh giá sự việc, những kinh nghiệm làm việc rất quan trọng cho tôi trong công việc sau này. Em cảm ơn Thầy rất nhiều!

Tiếp theo, tôi cũng xin gửi lời cảm ơn sâu sắc tới các Thầy, các Cô và các anh chị em trong khoa Điện tử - Viễn thông và PTN Trọng điểm Hệ thống Tích hợp Thông minh đã luôn sẵn sàng giúp đỡ tạo điều kiện tốt nhất cho tôi trong quá trình làm luận văn.

Cuối cùng, tôi xin gửi những lời cảm ơn chân thành nhất tới gia đình tôi, những người luôn luôn ủng hộ, động viên tôi cả về vật chất lẫn tinh thần để tôi có thể hoàn thành luận văn tốt nhất.

Mặc dù có nhiều cố gắng, song luận văn vẫn còn những hạn chế nhất định. Tôi rất mong nhận được các góp ý, chỉ bảo của các thầy, cô để hoàn thiện hơn luận văn của mình. Tôi xin chân thành cảm ơn.

Hà Nội, Ngày 28 tháng 12 năm 2016

Học viên

Nguyễn Ngọc Sinh

MỤC LỤC

LỜI CAM ĐOAN	I
LỜI CẢM ƠN	II
MỤC LỤC	1
DANH MỤC VIẾT TẮT	4
DANH SÁCH HÌNH VẼ	5
DANH SÁCH BẢNG	6
MỞ ĐẦU	8
TÓM TẮT	8
CHƯƠNG 1. MÃ HÓA VIDEO TRONG XU THẾ IoT VÀ CƠ HỘI CHO MÃ HÓA MJPEG	10
1.1. Tổng quan mã hóa video	10
1.2. Mã hóa video trong xu thế IoT	13
1.3. Mã hóa video MJPEG	14
1.3.1. Mã hóa ảnh JPEG	14
1.3.2. MJPEG trong mạng cảm biến không dây	17
1.4. Kết luận chương 1	18
CHƯƠNG 2. PHƯƠNG PHÁP XÁC ĐỊNH CHUYỂN ĐỘNG DỰA TRÊN ĐẶC TÍNH CỦA MẮT NGƯỜI ĐỐI VỚI THÀNH PHẦN XÁM TRONG VIỆC LOẠI BỎ DƯ THỪA VỀ MẶT THỜI GIAN CỦA MÃ HÓA MJPEG	19
2.1. Dư thừa mặt thời gian	19
2.2. Các kỹ thuật loại bỏ dư thừa về mặt thời gian	19
2.2.1. Dự đoán bằng khung hình liền kề trước	20
2.2.2. Dự đoán bằng mô hình ước lượng chuyển động của các khối	21
2.2.3. Đồng bộ khung tham chiếu giữa mã hóa và giải mã	21
2.3. Kỹ thuật loại bỏ dư thừa thời gian đề xuất dựa trên phân tách chuyển động	23
2.3.1. Thuật toán xác định chuyển động	24
2.3.2. Đặc tính của mắt người đối với thành phần xám và khả năng áp dụng cho phân tách chuyển động	24
2.4. Phương pháp xác định chuyển động dựa trên đặc tính của mắt người đối với thành phần xám trong việc loại bỏ dư thừa về mặt thời gian của mã hóa MJPEG ...	27
2.4.1. Nguyên lý hoạt động	28
2.4.2. Sơ đồ mã hóa chi tiết	29
2.5. Kết luận chương 2	32

CHƯƠNG 3. MÔ PHỎNG VÀ ĐÁNH GIÁ PHƯƠNG ÁN ĐỀ XUẤT	33
3.1. Thông số đánh giá	33
3.1.1. Tỷ lệ nén	33
3.1.2. Chất lượng hình ảnh giải mã PSNR	33
3.1.3. Độ phức tạp	34
3.2. Mô phỏng và đánh giá kết quả	34
3.2.1. Kịch bản mô phỏng	35
3.2.2. Kết quả mô phỏng	36
KẾT LUẬN	44
DANH MỤC CÁC CÔNG TRÌNH CỦA TÁC GIẢ LIÊN QUAN ĐẾN LUẬN VĂN	45
TÀI LIỆU THAM KHẢO	46

DANH MỤC VIẾT TẮT

4:2:0	Phương pháp lấy mẫu theo tỷ lệ 4:2:0; thành phần màu ở cả chiều ngang và chiều dọc được lấy mẫu bằng 1/2 của thành phần xám.
4:2:2	Phương pháp lấy mẫu mà thành phần màu ở chiều dọc được lấy mẫu bằng 1/2 của thành phần xám.
4:4:4	Phương pháp lấy mẫu mà thành phần màu được lấy mẫu bằng với số mẫu ở thành phần xám.
DCT	Discrete Cosine Transform.
GOP	Group of Picture, nhóm các khung hình đã được mã hóa.
H.261	Chuẩn mã hóa video H.261.
H.264/AVC	H.264 Advance Video Codec.
HEVC	High Efficient Video Codec.
IOT	Internet of Thing.
ISO	International Standards Organization. Một liên đoàn trên toàn thế giới của các cơ quan tiêu chuẩn từ khoảng 140 quốc gia. Homepage at: http://www.iso.ch .
ITU	International Telecommunications Union. Tổ chức quốc tế thuộc hệ thống Liên Hợp Quốc, nơi các chính phủ và khu vực tư nhân phối hợp các mạng và dịch vụ viễn thông toàn cầu. Homepage at: http://www.itu.int .
JPEG	Joint Photographic Experts Group. Cộng đồng phát triển chuẩn nén ảnh JPEG và JPEG2000. Homepage at: http://www.jpeg.org .
MJPEG	Motion Joint Photographic Experts Group.

MPEG Motion Picture Experts Group. Cộng đồng phát triển MPEG. Homepage at: <http://www.chiariglione.org/mpeg>.

PNG Portable Network Graphics.

PSNR Peak Signal to Noise Ratio.

UBICOMP Ubiquitous computing.

DANH SÁCH HÌNH VẼ

<i>Hình 1.1. Tỷ lệ nén của một số chuẩn mã hóa video thông dụng.</i>	11
<i>Hình 1.2. Sơ đồ lịch sử phát triển của nén video từ 1990 [1].</i>	12
<i>Hình 1.3. Giải pháp tiền xử lý cho camera trong mạng cảm biến không dây.</i>	13
<i>Hình 1.4. Mã hóa JPEG.</i>	14
<i>Hình 1.5. Phân bố năng lượng của ma trận 8x8 sau DCT [2]</i>	16
<i>Hình 2.1. Ba khung hình liên tiếp trong video.</i>	19
<i>Hình 2.2. Khung hình 1(trái), 2(phải).</i>	20
<i>Hình 2.3. Khung hình sai khác.</i>	20
<i>Hình 2.4. Sơ đồ đồng bộ khung tham chiếu giữa mã hóa và giải mã</i>	22
<i>Hình 2.5. Nguyên lý đề xuất</i>	24
<i>Hình 2.6. Cấu trúc mắt người (Encyclopedia Britannica, 1994)</i>	25
<i>Hình 2.7. Toàn bộ sắc thái xám của ảnh 5 bit và 8 bit</i>	26
<i>Hình 2.8. Kết quả sử dụng đặc tính của mắt người</i>	27
<i>Hình 2.9. Nguyên lý mã hóa</i>	28
<i>Hình 2.10. Nguyên lý giải mã</i>	28
<i>Hình 2.11. Sơ đồ mã hoá chi tiết</i>	29
<i>Hình 3.1. Sơ đồ cho kịch bản đánh giá</i>	35
<i>Hình 3.2. Khung hình mã hóa - giải mã video akiyo</i>	36
<i>Hình 3.3. Khung hình mã hóa - giải mã video container</i>	37
<i>Hình 3.4. Khung hình mã hóa - giải mã video foreman</i>	37
<i>Hình 3.5. Khung hình mã hóa -giải mã video hall</i>	38
<i>Hình 3.6. Khung hình mã hóa -giải mã video news</i>	38
<i>Hình 3.7. Khung hình mã hóa -giải mã video soccer</i>	39
<i>Hình 3.8. Kết quả nén video mẫu tại QP[5-95]</i>	40
<i>Hình 3.9. Tỷ lệ nén inter-intra của video mẫu</i>	41
<i>Hình 3.10. Kết quả PSNR và tỷ lệ nén của các video mẫu</i>	42

DANH SÁCH BẢNG

<i>Bảng 3.1. Môi trường mô phỏng</i>	34
<i>Bảng 3.2. Video tham gia mô phỏng</i>	34
<i>Bảng 3.3. Các ký hiệu sử dụng chung trong phần mô phỏng</i>	36
<i>Bảng 3.4. Số phép tính của các thuật toán Zipfian, biến đổi DCT thuận, Lượng tử và phương pháp đề xuất trên từng khối 8x8 [3]</i>	43

TÓM TẮT

Ngày nay chúng ta đang trong thời kỳ hậu PC, dần tiến đến thời kỳ mạng internet kết nối vạn vật (IoT - Internet Of Things) và tính toán phân tán rộng khắp (ubicom). Trong thời đại IoT, camera sẽ được lắp đặt với một số lượng lớn trong các mạng cảm biến không dây. Khi đó, các camera sẽ đóng vai trò như đôi mắt của IoT và thực hiện các hoạt động an ninh, giám sát. Tuy nhiên, do đặc tính giới hạn về năng lực tính toán, năng lượng và băng thông của các camera trong mạng cảm biến không dây mà không thể áp dụng các phương pháp nén video tiên tiến. Những mã hóa video đòi hỏi các hệ thống năng lực tính toán và nguồn năng lượng lớn. Trong nội dung luận văn, học viên đã nghiên cứu và phân tích các chuẩn mã hóa video cùng với đặc tính của hệ thống thị giác của con người. Qua đó, học viên đề xuất một phương pháp mã hóa video áp dụng cho các mạng cảm biến không dây: **“Phương pháp xác định chuyển động dựa trên đặc tính của mắt người đối với thành phần xám trong việc loại bỏ dư thừa về mặt thời gian của mã hóa video MJPEG”**.

Mô hình nén video được đề xuất trong luận văn bao gồm các giai đoạn như sau: phân tách chuyển động, mã hóa khối tĩnh - động, giải mã khung tham chiếu. Phân tách chuyển động được thực thi dựa trên đặc tính của mắt người trong việc nhận biết và phân tách các màu xám. Mã hóa khối tĩnh - động bao gồm hai phần mã hóa các khối tĩnh và các khối động. Mã hóa các khối động được thực hiện tương tự như mã hóa các khối bình thường trong mã hóa JPEG. Mã hóa các khối tĩnh dựa trên đặc tính sai khác bằng không mà thực hiện rút gọn các bước không cần thiết: DCT, lượng tử hóa. Phần giải mã khung tham chiếu nhằm đảm bảo dữ liệu tham chiếu của hai phần mã hóa và giải mã giống nhau. Thực tế, giải mã khung tham chiếu chỉ bao gồm giải mã các khung động.

Phương án đề xuất đảm bảo được tính đơn giản nhưng hiệu quả trong việc tăng tỷ lệ nén của mã hóa MJPEG. Với chất lượng video sau mã hóa tương đương, phương pháp đề xuất có thể đạt tỷ lệ nén gấp đôi so với MJPEG truyền thống và luôn cao hơn so với các phương pháp loại bỏ dư thừa thời gian đã được nêu trong luận văn: nén sai khác MJPEG, MJPEG với thuật toán xác định chuyển động Zipfian. Về mặt độ phức tạp, phương án đề xuất sử dụng một lượng phép tính ít gấp 3 lần so với phương án áp dụng MJPEG với thuật toán xác định chuyển động Zipfian. Điều này đảm bảo phương án đề xuất có thể đạt được tốc độ mã hóa cao hơn so với các phương án mã hóa xây dựng từ MJPEG được đề cập cho các video tồn tại ít chuyển động. Về mặt khả năng ứng dụng, mã hóa đề xuất hoàn toàn phù hợp cho các camera trong mạng cảm biến không dây bởi mã hóa đề xuất có độ phức tạp thấp và cho phép linh động GOP.

MỞ ĐẦU

Ngày này, chúng ta đang trong thời kỳ hậu PC, dần tiến đến thời kỳ IoT và tính toán phân tán rộng khắp (ubicom). Trong thời đại IoT, camera sẽ được lắp đặt với một số lượng lớn trong các mạng cảm biến không dây. Khi đó, các camera sẽ đóng vai trò như đôi mắt của IoT để thực hiện các hoạt động an ninh, giám sát. Tuy nhiên, do đặc tính giới hạn về năng lực tính toán, năng lượng và băng thông của các camera trong mạng cảm biến không dây mà không thể áp dụng các phương pháp nén video tiên tiến. Trong nội dung luận văn này, học viên đã nghiên cứu và phân tích các chuẩn mã hóa video cùng với đặc tính của hệ thống thị giác của con người. Qua đó, học viên đề xuất một phương mã hóa video áp dụng cho các mạng cảm biến không dây: **“Phương pháp xác định chuyển động dựa trên đặc tính của mắt người đối với thành phần xám trong việc loại bỏ dư thừa về mặt thời gian của mã hóa video MJPEG ”**

Mô hình nén video được đề xuất trong luận văn bao gồm 3 giai đoạn như sau: phân tách chuyển động, mã hóa khối tĩnh - động, giải mã khung tham chiếu. Phân tách chuyển động được thực thi dựa trên đặc tính của mắt người trong việc nhận biết và phân tách các màu xám. Mã hóa khối tĩnh - động bao gồm hai phần mã hóa các khối tĩnh và các khối động. Mã hóa các khối động được thực hiện tương tự như mã hóa các khối bình thường trong mã hóa JPEG. Mã hóa các khối tĩnh dựa trên đặc tính sai khác bằng không mà thực hiện rút gọn các bước không cần thiết: DCT, lượng tử hóa. Phần giải mã khung tham chiếu là phần đảm bảo dữ liệu tham chiếu của hai phần mã hóa và giải mã giống nhau. Thực tế, giải mã khung tham chiếu chỉ bao gồm giải mã các khung động.

Phương án đề xuất đảm bảo được tính đơn giản nhưng hiệu quả trong việc tăng tỷ lệ nén của mã hóa MJPEG. Với chất lượng video sau mã hóa tương đương, phương pháp đề xuất có thể đạt tỷ lệ nén gấp đôi so với MJPEG truyền thống và luôn cao hơn so với các phương pháp loại bỏ dư thừa thời gian đã được nêu trong luận văn: nén sai khác MJPEG, MJPEG với thuật toán xác định chuyển động Zipfian.

Luận văn được chia thành các phần như sau:

Chương 1: Mã hóa video trong xu thế IoT và cơ hội cho mã hóa MJPEG

Chương này trình bày hai nội dung kiến thức chính. Đầu tiên, học viên nhắc lại các nguyên tắc cơ bản trong mã hóa video, lịch sử của mã hóa video và những yêu cầu mới đối với mã hóa video trong thời kỳ IoT. Tiếp đó, học viên trình bày các hiểu biết của học viên đối với mã hóa video MJPEG. Đồng thời học viên cũng phân tích, làm rõ những lợi thế và hạn chế của mã hóa video MJPEG khi áp dụng cho các mạng cảm biến không dây.

Chương 2: Phương pháp xác định chuyển động dựa trên đặc tính của mắt người đối với thành phần xám trong việc loại bỏ dư thừa về mặt thời gian của mã hóa

mjpeg

Trong mã hóa video, loại bỏ dư thừa về mặt thời gian là chìa khóa của việc tăng tỷ lệ nén lên nhiều lần. Chương 2 trình bày những kiến thức cơ bản về dư thừa thời gian, các kỹ thuật loại bỏ dư thừa về mặt thời gian đang được sử dụng. Tiếp đó, học viên phân tích những điểm mạnh yếu của các kỹ thuật loại bỏ dư thừa về mặt thời gian. Qua đó, học viên đề xuất một phương án loại bỏ dữ liệu về mặt thời gian cho MJPEG và có thể áp dụng cho mạng cảm biến không dây. Cuối cùng, chương 2 trình bày chi tiết về sự kết hợp giữa MJPEG và kỹ thuật loại bỏ dư thừa đề xuất dựa trên đặc tính của mắt người đối với thành phần xám.

Chương 3: Mô phỏng và đánh giá phương án đề xuất

Trong chương này trình bày kết quả mô phỏng của phương án mã hóa video đề xuất và so sánh kết quả thu được của phương pháp đề xuất với các phương pháp mã hóa video khác dựa trên MJPEG: mã hóa MJPEG truyền thống, mã hóa sai khác MJPEG, mã hóa MJPEG kết hợp với thuật toán xác định chuyển động.

Chương 1

MÃ HÓA VIDEO TRONG XU THẾ IOT VÀ CƠ HỘI CHO MÃ HÓA MJPEG

1.1 Tổng quan mã hóa video

Nén hay mã hóa video là việc giảm thiểu và loại bỏ các thành phần dư thừa trong video qua đó giúp cho lưu trữ video trên ổ đĩa và truyền qua mạng một cách hiệu quả. Nén video có thể chia ra làm hai dạng chính: nén mất mát và nén không mất mát. Nén không mất mát là phương pháp nén mà video được giải mã giống với video trước khi mã hóa. Phương pháp này cho chất lượng hình ảnh cao nhất nhưng phải đánh đổi bằng việc tỷ lệ nén ở mức độ thấp. Nén mất mát có nghĩa là video sau khi được giải mã có sai khác với video gốc. Ưu điểm của phương pháp này là tỷ lệ nén cao hơn rất nhiều so với nén không mất mát. Trong nén mất mát, các dữ liệu mất mát là những dữ liệu dư thừa, dữ liệu này thường là mắt người không thể hoặc ít có khả năng cảm nhận. Dữ liệu dư thừa được chia làm 4 loại chính: dư thừa về khả năng nhận thức (perceptual redundancies), dư thừa về mặt không gian (spatial redundancies), dư thừa về mặt thời gian (temporal redundancies), dư thừa về mặt thống kê (statistical redundancies).

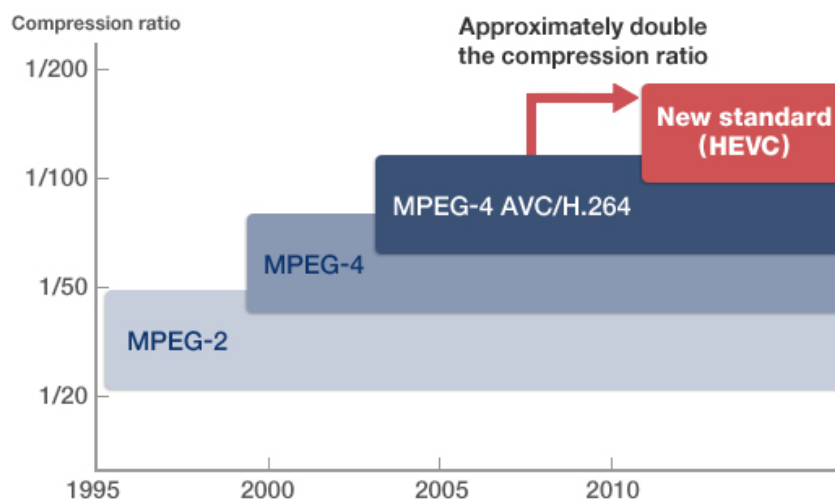
Dư thừa về mặt nhận thức là những chi tiết của bức ảnh mà mắt người không thể cảm nhận được. Bất cứ dữ liệu nào mà mắt người không thể cảm nhận được đều có thể loại bỏ được. Loại bỏ dư thừa về khả năng nhận thức được thực hiện bởi quá trình lấy mẫu các thành phần màu.

Dư thừa về mặt thời gian: Video là tập hợp chuỗi các khung hình liên tiếp, những khung hình liên tiếp có tỷ lệ lớn các thành phần giống nhau. Do đó, việc loại bỏ dữ liệu giống nhau (dư thừa về mặt thời gian) giữa các khung hình giúp tăng tỷ lệ giá trị không và lân cận không. Trong các mã hóa video, xác định và loại bỏ dư thừa về mặt thời gian chính là chìa khóa chính để tăng tỷ lệ nén lên nhiều lần.

Dư thừa về mặt không gian là sự thay đổi trong cấp độ của ảnh. Độ nhạy cảm của mắt người tỷ lệ nghịch với tần số của tín hiệu, điều này có nghĩa là khả năng phân biệt sự thay đổi của mắt người cũng giảm dần đối với các tần số cao. Tính chất này của mắt người được gọi là tích hợp không gian (spatial integration). Dựa vào đặc tính này các phương pháp mã hóa video có thể loại bỏ hoặc giảm các tần số cao mà không ảnh hưởng đến cảm nhận của mắt người. Việc này được thực hiện bởi hai quá trình: biến đổi sang miền tần số và lượng tử hóa.

Dư thừa về mặt thống kê: Các giá trị sau lượng tử, véc-tơ chuyển động và các dữ liệu khác được mã hóa bằng các chuỗi nhị phân tại bước cuối cùng của nén video. Phương pháp đơn giản nhất để mã hóa các giá trị này là sử dụng phương pháp mã hóa có độ dài cố định - fixed length codes. Tuy nhiên, do đặc tính của video là các giá trị không tuân theo quy luật phân bố đồng đều. Vậy nên, việc mã hóa có độ dài cố định sẽ cho kết quả chuỗi mã hóa có kích thước lớn hơn nhiều so với khả năng có thể. Sử dụng mã hóa có độ dài thay đổi (variable length codes) là giải pháp cho những trường hợp các ký hiệu (symbol) có phân bố không đồng đều. Những ký hiệu có tỷ lệ xuất hiện lớn hơn sẽ được biểu diễn bởi những đoạn bit có kích thước nhỏ hơn. Trong nén video, loại bỏ dư thừa này được thực hiện bởi mã hóa Entropy (Entropy Coding).

Sự hiệu quả của một phương pháp mã hóa video được đánh giá dựa trên hai phương diện chính: tỷ lệ nén trên cùng một chất lượng ảnh (khả năng loại bỏ các tín hiệu dư thừa) và tốc độ mã hóa. Phần lớn việc lựa chọn phương pháp mã hóa phụ thuộc vào yêu cầu về thời gian mã hóa của từng ứng dụng: mã hóa thời gian thực hay chấp nhận độ trễ của mã hóa để đạt tỷ lệ nén cao.

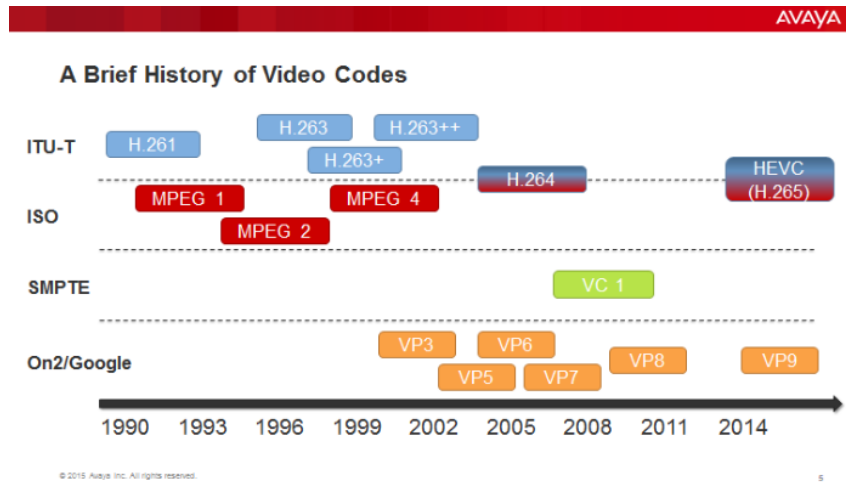


Hình 1.1: Tỷ lệ nén của một số chuẩn mã hóa video thông dụng.

Hình 1.1 so sánh tỷ lệ nén giữa một số chuẩn mã hóa video thông dụng hiện nay. Áp dụng các phương pháp nén video giúp tiết kiệm ít nhất 20 lần dung lượng lưu trữ cũng như băng thông khi truyền tải. Tỷ lệ nén này có thể tăng lên đến gần 200 lần đối với chuẩn mã hóa video tiên tiến nhất hiện nay là H.265/HEVC.

Hình 1.2 trình bày lịch sử phát triển của các chuẩn mã hóa video từ những năm 1990. Như Hình 1.2 việc phát triển các chuẩn mã hóa video cũng như phương pháp được xây dựng bởi một số tổ chức lớn như: ITU-T, ISO, SMPTE, On2/Google.

Theo thời gian, càng ngày càng nhiều chuẩn mã hóa mới được phát minh. Theo Hình 1.2 có 16 chuẩn mã hóa video phổ biến đã và đang được sử dụng cho tới nay. Những chuẩn mới hơn, tiên tiến hơn cho kết quả tốt hơn cả về chất lượng lẫn tỷ lệ nén. Để có



Hình 1.2: Sơ đồ lịch sử phát triển của nén video từ 1990 [1].

những ưu điểm đó, các chuẩn mã hóa mới cũng yêu cầu các hệ thống có năng lực tính toán cao hơn và tiêu thụ năng lượng nhiều hơn.

H.261 [4] là một chuẩn mã hóa video cho hội nghị truyền hình được công bố bởi ITU vào năm 1990. Chuẩn này được thiết kế cho dữ liệu ở các tốc độ bằng cấp số nhân của 64Kbit/s. Cùng với công bố chuẩn H.261, sơ đồ mã hóa Hybrid cũng được hoàn thiện và giới thiệu. Sau này, sơ đồ Hybrid trở thành sơ đồ mã hóa cho các chuẩn mã hóa tiên tiến khác như H.264/AVC, HEVC. Trong sơ đồ Hybrid, các khung hình mã hóa được chia ra làm hai loại khung hình intra và khung hình inter. Khung hình inter là các khung hình mà dữ liệu được mã hóa phụ thuộc vào các khung hình khác (reference frame - khung hình tham chiếu) trong video. Dữ liệu mã hóa của các khung hình inter được tính toán từ quá trình phỏng đoán liên ảnh (Inter Prediction). Các khung hình tham chiếu này có thể xuất hiện trước hoặc sau khung hình hiện tại trong thứ tự sắp xếp của video. Khung hình intra là những khung hình mà dữ liệu mã hóa được nội suy từ chính trong khung hình. Dữ liệu nội suy được thực hiện bởi quá trình phỏng đoán nội ảnh (Intra Prediction).

H.264/AVC [5] [6] là kết quả của sự phối hợp giữa hai tổ chức lớn là ITU và MPEG. H.264 được đặt tên bởi ITU-T, trong khi đó ISO/IEC gọi là MPEG-4 Part 10/AVC. Chuẩn mã hóa video H.264/AVC được thiết kế nhằm vượt qua những hạn chế của các chuẩn mã hóa video trước đó như đạt tỷ lệ nén gấp đôi với cùng chất lượng ảnh, giảm lỗi do quá trình truyền tải qua mạng...

H.265/HEVC [7] [8] chuẩn mã hóa tiên tiến nhất hiện nay được thiết kế cho video có độ phân giải cao như fullHD, 2K, 4K, và có thể đạt được tỷ lệ nén gấp đôi so với H.264/AVC.

Ngoài ra, một số chuẩn mã hóa video khác như MPEG1, MPEG2, H262, H263 đã được trình bày tại báo cáo [9]

1.2 Mã hóa video trong xu thế IoT

IoT và mô hình tính toán phân tán rộng khắp (ubicomp) là xu thế tất yếu khi mà các vấn đề tiên quyết như: hiệu năng và kích thước CPU, năng lượng, băng thông đã và đang được giải quyết. Trong thời đại IoT, camera sẽ được nhúng xung quanh môi trường sống của con người và đóng vai trò là đôi mắt cho mô hình IoT. Những camera này sẽ giúp thực hiện các ứng dụng khác nhau như an ninh, giám sát giao thông, nông nghiệp ...

Mạng cảm biến không dây là một phần cơ bản trong mô hình IoT. Nó được sử dụng trong những trường hợp cần sử dụng các cảm biến một cách linh động, ẩn giấu đối với người dùng cuối và đặc biệt sử dụng cho những khu vực xa và nguy hiểm, không thể cung cấp năng lượng một cách liên tục và thường xuyên.

Khi camera tham gia vào các mạng cảm biến không dây thì bắt buộc bị giới hạn về mặt năng lực tính toán, năng lượng, bộ nhớ và băng thông. Trong khi đó, thông tin truyền tải giữa camera và trung tâm gần nhất vẫn phải đảm bảo yêu cầu thời gian thực. Điều này đặt ra một thách thức lớn trong việc áp dụng các chuẩn nén video tiên tiến và hiệu quả cho các camera trong mạng cảm biến không dây.

Một giải pháp đã được đề xuất cho việc xử lý dữ liệu trong các mạng cảm biến không dây là chỉ thực hiện giai đoạn tiền xử lý hay xử lý thô tại các cảm biến. Dữ liệu sau tiền xử lý sẽ được truyền về và thực hiện công việc xử lý phức tạp tại các trung tâm có năng lực tính toán cao hơn và không bị giới hạn về năng lượng.

Đối với các camera trong mạng cảm biến không dây thì phương án tiền xử lý là phương án khả thi có thể đáp ứng được yêu cầu về thời gian thực mặc dù bị giới hạn về năng lực tính toán và năng lượng. Tiền xử lý ở đây có nghĩa là sử dụng các thuật toán nén video đơn giản. Hình 1.3 trình bày mô hình chi tiết. Các video từ camera sẽ được nén bằng các thuật toán nén video đơn giản và truyền về các trung tâm xử lý. Tại đây, các video đã được mã hóa tại các camera sẽ được giải mã và thực hiện mã hóa lại bằng các phương pháp/chuẩn mã hóa tiên tiến. Như vậy, vừa có thể đảm bảo được tính chất thời gian thực của hệ thống vừa đảm bảo yêu cầu hiệu quả khi lưu trữ và truyền tải.



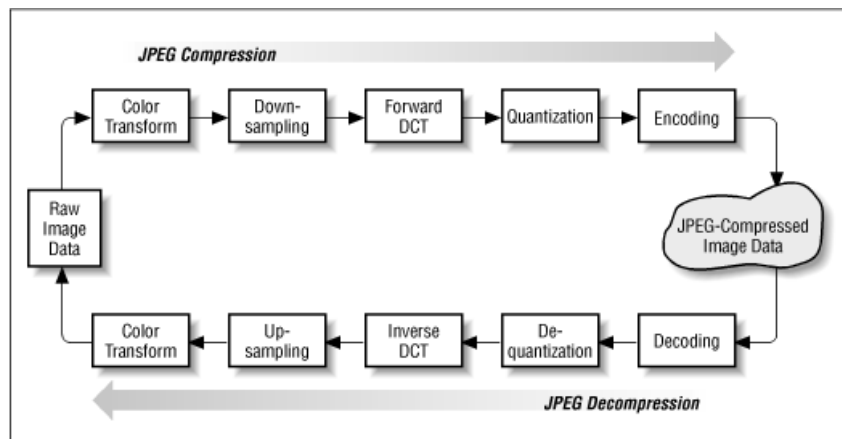
Hình 1.3: Giải pháp tiền xử lý cho camera trong mạng cảm biến không dây.

1.3 Mã hóa video MJPEG

Mã hóa MJPEG là phương pháp mã hóa các khung hình trong video một cách riêng lẻ thành các chuỗi bit JPEG. MJPEG được đánh giá là phương pháp mã hóa video đơn giản, yêu cầu ít năng lực tính toán nhất do không sử dụng bất kỳ kỹ thuật nào để loại bỏ dư thừa về mặt thời gian. Cũng chính vì vậy mà tỷ lệ nén của MJPEG đạt được là thấp nhất so với các phương pháp mã hóa video khác như H.263, H.264/AVC, HEVC.

1.3.1 Mã hóa ảnh JPEG

Mã hóa JPEG [10] được xây dựng dựa trên việc loại bỏ những dư thừa về mặt: khả năng nhận thức, không gian và thống kê thông qua các quá trình biến đổi cosin rời rạc (DCT - Discrete Cosine Transform), lượng tử hóa và mã hóa Entropy.



Hình 1.4: Mã hóa JPEG.

Hình 1.4 trình bày các bước để mã hóa và giải mã các khung hình trong MJPEG. Quá trình “biến đổi không gian màu (color transform)” sẽ chuyển đổi không gian màu của ảnh gốc sang không gian màu YCbCr. Sau đó thực hiện “lấy mẫu (downsampling)” cho các thành phần màu Cb và Cr. Biến đổi “biến đổi DCT thuận (forward DCT)” sẽ chuyển các khối dữ liệu ảnh sang miền tần số và thực hiện quá trình “lượng tử hóa (quantization)” để loại bỏ bớt thành phần có tần số cao. Cuối cùng, quá trình “mã hóa Entropy (Entropy coding)” sẽ thực hiện mã hóa dữ liệu sau lượng tử để được chuỗi mã hóa JPEG.

Biến đổi không gian màu - Color Transform

Mã hóa JPEG có khả năng mã hóa trong tất cả các không gian màu. Tuy nhiên, JPEG cũng như các thuật toán mã hóa ảnh hay video khác đều đạt được tỷ lệ nén cao nhất với không gian màu có thành phần chói như YCbCr. Nguyên nhân bắt nguồn từ đặc tính sinh

lý của mắt người. Hệ thống thị giác của mắt người nhạy cảm nhất đối với thành phần chói (Y trong YCbCr) và ít nhạy cảm hơn đối với các thành phần khác (Cb, Cr). Chính vì vậy cho phép ta loại bỏ nhiều dữ liệu hơn trong các thành phần Cb, Cr mà không ảnh hưởng đến khả năng cảm nhận của mắt người.

Lấy mẫu Down sampling - Up sampling

Như đã thảo luận trong phần biến đổi không gian màu, hệ thống thị giác của mắt người có những đặc tính nhất định đối với từng thành phần màu khác nhau. Điều này cho phép giảm thiểu dữ liệu của những thành phần màu ít tác động đến thị giác. Nhiệm vụ của quá trình “lấy mẫu” là loại bỏ bớt dữ liệu màu. Trong không gian màu YCbCr có 3 tỷ lệ lấy mẫu thông dụng YCbCr 4:4:4, YCbCr 4:2:2, YCbCr 4:2:0. Ngược lại với “lấy mẫu” là “upsampling”, quá trình này nhận dữ liệu đã được lấy mẫu và khôi phục thành dữ liệu YCbCr 4:4:4.

Biến đổi DCT

DCT là quá trình biến đổi dữ liệu ảnh giữa miền không gian và miền tần số. Trong mã hóa JPEG, biến đổi DCT sẽ được thực hiện trên các khối ảnh 8x8.

$$\mathbf{G}_{u,v} = \frac{C(u)}{2} \frac{C(v)}{2} \sum_{i=0}^7 \sum_{j=0}^7 (\mathbf{g})_{i,j} \cos \frac{(2i+1)u\pi}{16} \cos \frac{(2j+1)v\pi}{16}, \quad (1.1)$$

Trong đó;

$$C(u) = \begin{cases} \frac{1}{\sqrt{2}} & u = 0 \\ 1 & u > 0 \end{cases}. \quad (1.2)$$

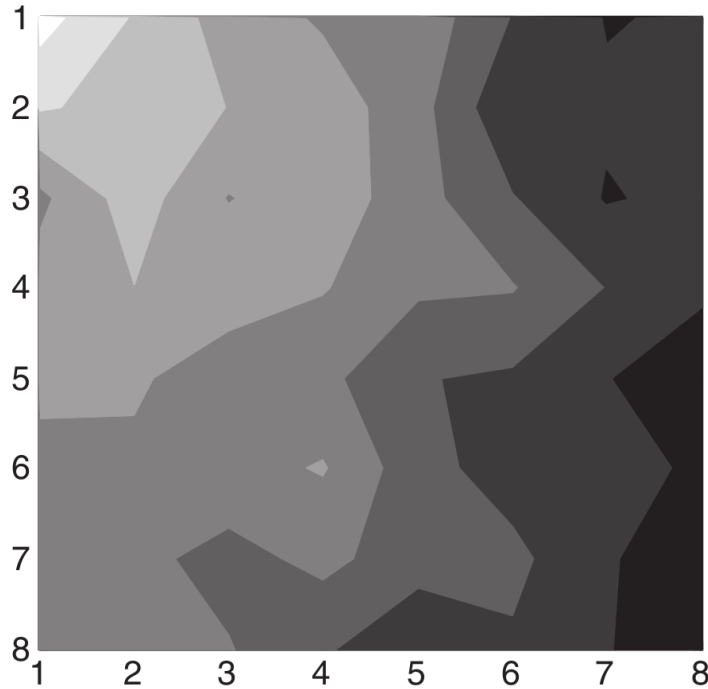
$$0 \leq u, v < 8.$$

$g_{i,j}$ là dữ liệu ảnh đầu vào tại cột i , hàng j .

$G_{u,v}$ là dữ liệu sau biến đổi DCT thuận tại cột u , hàng v .

Công thức 1.1 trình bày biến đổi DCT thuận. Dữ liệu đầu vào $g_{i,j}$ là các khối ảnh 8x8 đã được dịch sang vùng giá trị xoay quanh giá trị không. Giả dụ, khối ảnh 8-bit sẽ được dịch từ miền giá trị [0:255] sang miền giá trị [-127:128]. Dữ liệu đầu ra là khối ảnh ở miền tần số $G_{u,v}$. Trên miền tần số thì năng lượng của bức ảnh sẽ tập trung tại các tần số thấp. Điều này tạo tiền đề cho quá trình lượng tử hóa.

Hình 1.5 trình bày năng lượng của các điểm ảnh sau quá trình biến đổi DCT thuận. Những thành phần sáng màu hơn có năng lượng cao hơn và tập trung ở các điểm nằm trên góc trên bên trái. Năng lượng tập trung nhiều nhất ở điểm (1,1) hay DC và ít nhất ở



Hình 1.5: Phân bố năng lượng của ma trận 8x8 sau DCT [2]

điểm (8,8). Càng xa điểm DC năng lượng càng nhỏ và giá trị tiến dần tới 0. Dựa vào bản đồ năng lượng này mà xây dựng được thứ tự zigzag trong phần mã hóa Entropy.

$$\mathbf{g}_{i,j} = \frac{C(u)}{2} \frac{C(v)}{2} \sum_{i=0}^7 \sum_{j=0}^7 (\mathbf{G})_{u,v} \cos \frac{(2i+1)u\pi}{16} \cos \frac{(2j+1)v\pi}{16}, \quad (1.3)$$

Trong đó:

$$0 \leq u, v < 8,$$

$$C(u) = \begin{cases} \frac{1}{\sqrt{2}} & u = 0 \\ 1 & u > 0 \end{cases}.$$

$g_{i,j}$ là dữ liệu ảnh đầu vào tại cột i , hàng j .

$G_{u,v}$ là dữ liệu sau biến đổi DCT thuận tại cột u , hàng v .

Công thức 1.3 trình bày biến đổi DCT ngược, cho phép chuyển dữ liệu ảnh từ miền tần số sang miền không gian.

Lượng tử hóa - Giải lượng tử

Lượng tử hóa là quá trình giảm dữ liệu ảnh trong miền tần số cao. Điều này cho phép tăng tỷ lệ các thành phần có giá trị không và lân cận giá trị không. Công thức 1.4 trình bày cách tính dữ liệu ảnh sau lượng tử.

$$\mathbf{B}_{i,j} = \text{round}\left(\frac{G_{i,j}}{Q_{i,j}}\right) \quad (1.4)$$

Trong đó,

$B_{i,j}$ là dữ liệu sau lượng tử tại cột i , hàng j ,

$G_{i,j}$ là dữ liệu tại miền tần số tại cột i , hàng j ,

$Q_{i,j}$ là hệ số lượng tử tại cột i , hàng j .

Lấy mẫu và lượng tử hóa là hai quá trình làm mất mát dữ liệu trong ảnh dựa trên đặc tính của hệ thống thị giác của con người. Dữ liệu mất mát này không thể khôi phục trong quá trình giải mã. Nhưng nhờ hai quá trình này mà JPEG có thể đạt được tỷ lệ nén cao gấp nhiều lần so với các chuẩn nén ảnh không mất mát như PNG.

Mã hóa Entropy

Mã hóa Entropy là bước cuối cùng trong mã hóa JPEG dùng để loại bỏ dư thừa về mặt thông kê. Mã hóa Entropy trong JPEG là một phương pháp mã hóa dữ liệu không mất mát được xây dựng trên cơ sở các thuật toán: thứ tự zigzag, mã hóa runlength, mã hóa Huffman. Quá trình mã hóa Entropy cho phép những ký tự (symbol) có tỷ lệ xuất hiện lớn hơn được biểu diễn bằng chuỗi bit 0-1 có kích thước nhỏ hơn.

1. Thứ tự zigzag là việc sắp xếp lại thứ tự các giá trị ảnh sau lượng tử hóa sang chuỗi số 1 chiều. Kết quả là các giá trị có năng lượng thấp (giá trị lân cận 0 và 0) nằm ở cuối chuỗi số. Thứ tự zigzag được xuất phát từ bản đồ phân bố năng lượng sau biến đổi thuận và lượng tử.
2. Mã hóa runlength [11] là việc sử dụng một ký tự đặc biệt (EOB - End Of Block) cho tất cả các giá trị không cuối cùng của chuỗi số sau khi sắp xếp lại theo thứ tự zigzag.
3. Mã hóa Huffman [12] là phương pháp mã hóa dữ liệu dựa trên tỷ lệ xuất hiện của các ký tự. Ký tự có tỷ lệ xuất hiện nhiều hơn sẽ được mã hóa bằng chuỗi bit 0-1 ngắn hơn.

1.3.2 MJPEG trong mạng cảm biến không dây

MJPEG là một trong những phương pháp nén video đơn giản, yêu cầu năng lực tính toán ít nhất so với các phương pháp mã hóa video khác. Nó hoàn toàn có thể đáp ứng được yêu cầu tính toán thời gian thực trong các điều kiện bị hạn chế về năng lực tính toán. Tuy nhiên, vấn đề tỷ lệ nén thấp là rào cản lớn nhất trong việc áp dụng MJPEG. Khi tỷ lệ nén thấp có nghĩa là các camera sẽ chiếm nhiều băng thông. Điều này sinh ra tình trạng tắc nghẽn băng thông hoặc sẽ phải giảm thiểu số lượng camera trong mạng

hoặc giảm tốc độ khung hình của các camera. Bên cạnh đó, truyền tải lượng dữ liệu lớn cũng ảnh hưởng không nhỏ đến việc tiêu thụ năng lượng của các camera.

Tỷ lệ nén thấp của MJPEG có nguyên nhân trực tiếp từ việc MJPEG không loại bỏ bất kỳ một dữ liệu dư thừa nào về mặt thời gian. Do vậy, áp dụng các phương pháp loại bỏ dư thừa về mặt thời gian đơn giản là giải pháp khả thi cho vấn đề này. Một trong những phương pháp đơn giản nhất là sử dụng dữ liệu sai khác giữa hai khung hình liên tiếp để loại một phần dư thừa về mặt thời gian.

Một ưu điểm khi sử dụng nén sai khác trong MJPEG cho mã hóa video trong các mạng cảm biến không dây là các khung hình được mã hóa bằng JPEG không bị ràng buộc bởi thông số GOP như H.264/AVC, HEVC. Điều này cho phép MJPEG linh động trong việc lựa chọn kỹ thuật mã hóa inter hay intra cho từng khung hình. Bất kỳ khung hình nào cũng có thể trở thành khung hình intra hoặc khung hình inter. Linh động trong lựa chọn kỹ thuật nén có một ý nghĩa đặc biệt quan trọng trong việc áp dụng cho các camera cố định. Nó cho phép gửi đi liên tục những khung hình inter khi mà không có sự sai khác giữa các khung hình liên tiếp.

1.4 Kết luận chương 1

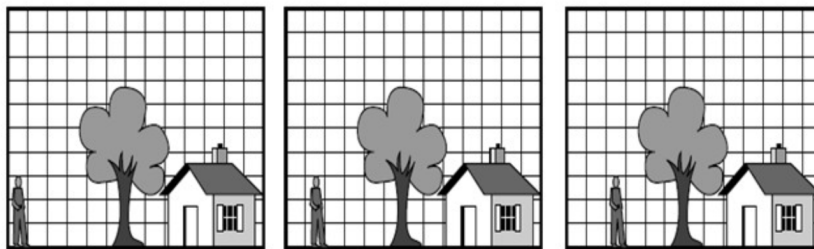
Trong chương, học viên đã trình bày những kiến thức cơ bản trong mã hóa video, sơ lược về lịch sử phát triển của mã hóa video và những yêu cầu mới cho mã hóa video trong xu thế IoT. Đồng thời, học viên đã trình bày nguyên lý cơ bản của phương pháp mã hóa video MJPEG. Qua đó, học viên đã đưa ra những phân tích, làm rõ những lợi thế (độ phức tạp, tính linh động và sự phổ biến) và hạn chế về mặt tỷ lệ nén của mã hóa video MJPEG khi áp dụng cho các mạng cảm biến không dây. Chương tiếp theo sẽ thảo luận về vấn đề dư thừa về mặt thời gian và việc áp dụng một kỹ thuật loại bỏ dư thừa thời gian được đề xuất cho mã hóa video MJPEG.

Chương 2

PHƯƠNG PHÁP XÁC ĐỊNH CHUYỂN ĐỘNG DỰA TRÊN ĐẶC TÍNH CỦA MẮT NGƯỜI ĐỐI VỚI THÀNH PHẦN XÁM TRONG VIỆC LOẠI BỎ DƯ THỪA VỀ MẶT THỜI GIAN CỦA MÃ HÓA MJPEG

2.1 Dư thừa mặt thời gian

Video số là tập hợp những bức ảnh/khung hình số liên tiếp. Những khung hình liên tiếp này có một tỷ lệ lớn các thành phần giống nhau. Việc truyền tải và lưu trữ các thành phần giống nhau giữa các khung hình làm lãng phí băng thông và bộ nhớ. Do vậy, những dữ liệu giống nhau này được định nghĩa là dư thừa về mặt thời gian trong video.



Hình 2.1: Ba khung hình liên tiếp trong video.

Để làm rõ hơn về dư thừa thời gian, Hình 2.1 trình bày một video mẫu bao gồm 3 khung hình liên tiếp. Các chi tiết về ngôi nhà, cái cây, bầu trời là tĩnh, và được đánh giá là những dư thừa về mặt thời gian. Vật thể duy nhất chuyển động trong 3 khung hình là con người. Tuy nhiên, việc truyền tải nguyên vẹn dữ liệu về con người trong ba khung hình vẫn tồn tại dư thừa về thời gian. Nguyên nhân xuất phát từ việc vật thể con người đã xuất hiện ngay từ khung hình đầu. Do đó, dữ liệu cần thiết để mã hóa chỉ bao gồm vị trí hay dữ liệu chuyển động của người trong các khung hình kế tiếp.

2.2 Các kỹ thuật loại bỏ dư thừa về mặt thời gian

Trong nén video việc loại bỏ dư thừa về mặt thời gian được thực hiện dựa trên mô hình dự đoán/tiên đoán về mặt thời gian (temporal prediction). Mục tiêu của mô hình

dự đoán này là loại bỏ dư thừa thời gian bằng việc xây dựng một khung hình tiên đoán cho khung hình hiện tại và thực hiện trừ hai khung hình. Khung hình tiên đoán được xây dựng từ những khung hình đã được mã hóa (khung hình tham chiếu) trước đó. Kết quả của quá trình là một khung hình sai khác có năng lượng ít hơn (nhiều giá trị trùng lặp và gần với giá trị 0) và hiệu quả hơn để loại bỏ dư thừa về mặt thống kê so với khung hình gốc.

2.2.1 Dự đoán bằng khung hình liền kề trước

Mô hình tiên đoán thời gian đơn giản nhất là sử dụng trực tiếp khung hình vừa được mã hóa trước đó làm khung hình tiên đoán. Ưu điểm của phương pháp này là sự đơn giản một cách tối đa trong quá trình tìm kiếm và xác định khung hình/khối tiên đoán. Tuy nhiên, dữ liệu sai khác cần mã hóa vẫn còn tồn tại nhiều thông tin dư thừa.



Hình 2.2: Khung hình 1(trái), 2(phải).



Hình 2.3: Khung hình sai khác.

Hình 2.3 trình bày kết quả của mô hình dự đoán dựa trên khung hình liền kề phía trước. Trong đó, khung hình 1 được sử dụng làm khung hình tiên đoán, khung hình 2 là khung hình cần được mã hóa. Hình 2.3 là khung hình sai khác, những màu xám biểu diễn cho giá trị của sự sai khác là 0, màu trắng biểu diễn sự sai khai lớn hơn 0, ngược lại màu đen thể hiện sai khác nhỏ hơn 0. Giá trị tuyệt đối của sự sai khác càng lớn thì màu sắc càng tiến gần tới trắng hoặc đen. Hạn chế lớn nhất của mô hình này là chỉ có khả năng loại bỏ dư thừa về mặt thời gian của những vật thể hay khối tĩnh. Khung hình sai khác vẫn còn tồn tại rất nhiều năng lượng, được tập trung ở màu đen và trắng. Năng lượng này xuất phát từ vật thể chuyển động giữa hai khung hình liên tiếp 1 và 2. Điều

này đồng nghĩa với việc vẫn còn một lượng lớn thông tin ở khung hình sai khác cần được mã hóa. Những thông tin này vẫn có thể lược giản để đạt được tỷ lệ nén cao hơn.

2.2.2 Dự đoán bằng mô hình ước lượng chuyển động của các khối

Trong video số, sự sai khác giữa các khung hình bao gồm các chuyển động, thay đổi ánh sáng và những vùng mới. Dư thừa về mặt thời gian của những sai khác này không thể loại bỏ bởi kỹ thuật dự đoán chỉ dựa trên khung hình liền kề phía trước 2.2.1. Kỹ thuật dự đoán dựa trên ước lượng chuyển động của các khối được phát triển để khắc phục nhược điểm phương pháp 2.2.1. Nguyên tắc hoạt động của kỹ thuật này được miêu tả vắn tắt như sau:

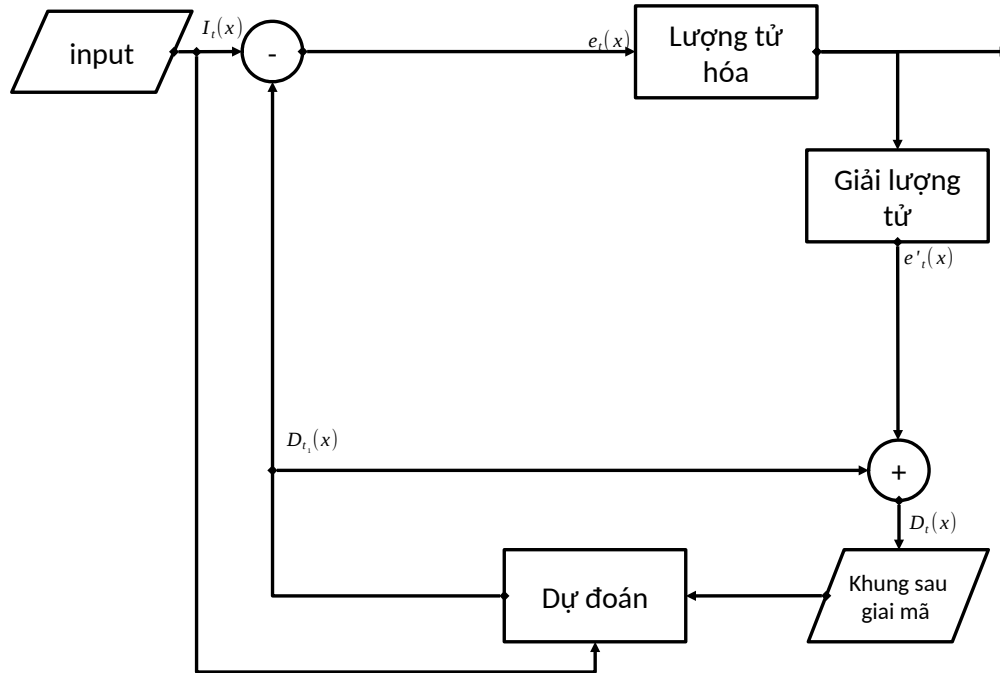
1. Chia khung hình mã hóa hiện tại thành các khối $M \times N$.
2. Tìm kiếm một khối $M \times N$ giống nhất của khối cần được mã hóa trong các khung hình tham chiếu D (tập hợp khung hình sau giải mã). Khung hình tham chiếu D có thể xuất hiện trước hoặc sau trong thứ tự sắp xếp của video. Phương pháp dễ hiểu nhất để tìm kiếm là tính sai khác giữa khối cần được mã hóa với các khối khác trong các khung hình tham chiếu. Sau đó chọn ra một khối có sự sai khác là nhỏ nhất và được gọi là khối giống nhất (best match). Quá trình tìm kiếm khối giống nhất được gọi là ước lượng chuyển động (motion estimation).
3. Khối được lựa chọn sẽ trở thành khối tiên đoán và sự sai khác giữa hai khối này sẽ được mã hóa. Quá trình này được gọi là bù đắp chuyển động (motion compensation)
4. Cùng với sự sai khác giữa hai khối thì dữ liệu về vị trí của khối tiên đoán sẽ được mã hóa cùng. Các thông số chỉ tới khối tiên đoán được gọi là véc-tơ chuyển động.

Ưu điểm của phương pháp này là khả năng loại bỏ dư thừa về mặt thời gian của các vật thể chuyển động. Lấy một ví dụ, vật thể A di chuyển từ vị trí có tọa độ (x_1, y_1) sang vị trí có tọa độ (x_2, y_2) . Phương pháp này truyền đi năng lượng dư thừa gần như bằng 0 cùng với một véc-tơ mô tả chuyển động từ vị trí (x_1, y_1) sang (x_2, y_2) . Với một video như Hình 2.1, dữ liệu sai khác cần mã hóa khi áp dụng kỹ thuật này chỉ bao gồm dữ liệu về vị trí và hướng chuyển động của con người. Hạn chế của phương pháp này là số lượng phép tính lớn để có thể tìm ra khối giống nhất và sự phụ thuộc lớn giữa các khung hình. Để tìm ra khối tiên đoán, kỹ thuật này phải thực hiện tính toán và so sánh sự sai khác giữa khối cần mã hóa với toàn bộ các khối trong vùng tìm kiếm.

2.2.3 Đồng bộ khung tham chiếu giữa mã hóa và giải mã

Trong mã hóa video, các khung hình tham chiếu không nằm trong tập hợp của khung hình gốc mà thuộc trong tập hợp các khung hình sau giải mã. Điều này nhằm đảm bảo

rằng khung hình tham chiếu được sử dụng trong bộ mã hóa và bộ giải mã là giống nhau. Qua đó, sự sai khác không mong muốn giữa khung hình sau mã hóa và khung hình gốc luôn nằm trong tầm kiểm soát.



Hình 2.4: Sơ đồ đồng bộ khung tham chiếu giữa mã hóa và giải mã

Hình 2.4 trình bày sơ đồ xây dựng dữ liệu khung tham chiếu đồng bộ giữa mã hóa và giải mã. Sơ đồ này là sơ đồ DPMC (Difference Pulse Coding Module) và được áp dụng trong sơ đồ mã hóa video Hybrid. Để đơn giản hóa thì quá trình biến đổi và biến đổi ngược được lược giản.

$$\text{Sai số lượng tử hóa: } Qe_t(x) = e'_t(x) - e_t(x) \quad (2.1)$$

$$\text{Khung hình giải mã: } D_t(x) = D_{t_1}(x) + e'_t(x) = D_{t_1}(x) + Qe_t(x) + e_t(x) \quad (2.2)$$

$$\text{Sai khác mong muốn: } EXP(x) = I_t(x) - I_{t_1}(x) \quad (2.3)$$

$$\text{Sai khác thực tế: } R(x) = I_t(x) - D_{t_1}(x) \quad (2.4)$$

Các công thức 2.1 - 2.4 trình bày hệ quả của mô hình khi đảm bảo sự đồng bộ khung tham chiếu giữa mã hóa và giải mã. Sai khác thực tế khi tham gia quá trình mã hóa Entropy bao gồm cả sai khác mong muốn và sai khác do sai số cặp quá trình lượng tử

hóa - giải lượng tử. Hệ quả này làm giảm tỷ lệ nén của ảnh sau mã hóa (tăng sự sai khác giữa khung hình sau mã hóa và khung hình gốc) và tăng kích thước của video sau mã hóa. Để sự sai khác thực tế không trở nên quá lớn và mất kiểm soát thì sau một lượng khung hình được áp dụng các kỹ thuật loại bỏ dư thừa về mặt thời gian, một khung hình intra sẽ được sử dụng. Khoảng cách giữa hai khung hình intra được gọi là GOP.

Giá trị GOP giúp kiểm soát sự sai khác do sai số lượng tử. Tuy nhiên, sử dụng GOP cũng có tác động rất lớn đối với tỷ lệ nén đặc biệt là trong các ứng dụng an ninh hay giám sát. Nếu có thể kéo dài tối đa chỉ số GOP (sử dụng linh động GOP) cho phép tỷ lệ nén video tăng lên một cách đáng kể. Điều này được làm rõ như sau:

1. Các khung hình inter thường có tỷ lệ nén gấp hàng chục thậm chí hàng trăm lần so với các khung hình intra. Tỷ lệ này càng tăng đối với các chuẩn mã hóa tiên tiến và đặc thù của các ứng dụng an ninh.
2. Camera an ninh thường xuyên ghi lại những khung hình liên tiếp giống nhau (khung hình không chuyển động) và thậm chí những khung hình chuyển động vẫn thường có tỷ lệ khối tĩnh lớn.
3. Giả sử một ví dụ lý tưởng về sự ảnh hưởng của GOP như sau: Video gồm 80 khung hình. Kích thước của khung hình intra là x . Kích thước của khung hình inter là y và $x = 10y$. Kích thước của video sau mã hóa là s_n với n khung hình intra:

$$s_{11} = 11x + 69y = 179y \text{ (GOP = 10)}$$

$$s_{10} = 10x + 70y = 170y = 95\%s_{11}$$

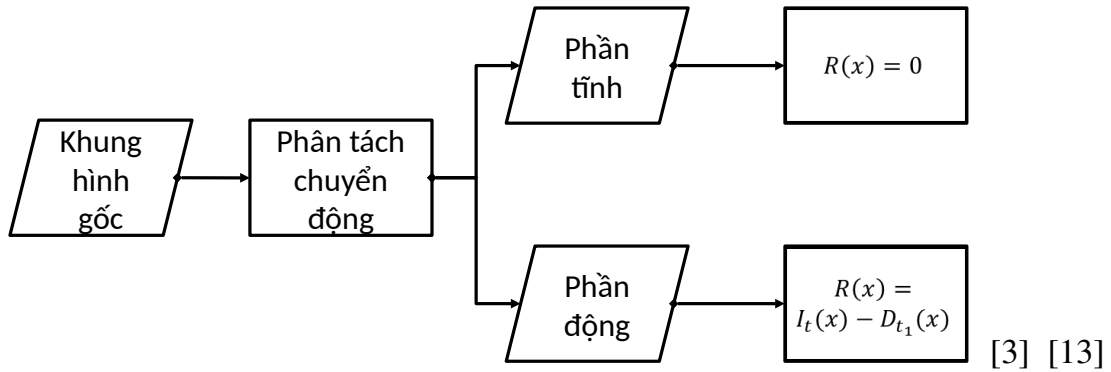
$$s_9 = 9x + 71y = 97y = 90\%s_{11}$$

$$s_1 = 1x + 79y = 179y = 50\%s_{11}$$

2.3 Kỹ thuật loại bỏ dư thừa thời gian đề xuất dựa trên phân tách chuyển động

Đồng bộ dữ liệu khung hình tham chiếu là điều cần thiết xong nó cũng mang lại những hạn chế nhất định 2.2.3. Trong phần này, học viên trình bày một phương án đề xuất để hạn chế hệ quả tiêu cực của việc đồng bộ dữ liệu khung tham chiếu trong mã hóa các phần tĩnh và cho phép không giới hạn chỉ số GOP.

Hình 2.5 trình bày sơ đồ nguyên lý phương án đề xuất. Các khung hình gốc trong video sẽ được đưa vào quá trình phân tách chuyển động. Quá trình này dựa trên dữ liệu của các khung hình gốc và có nhiệm vụ tách khung hình mã hóa hiện tại thành 2 phần riêng biệt: các khối chuyển động và các khối tĩnh. Sai khác của các khối tĩnh và khối



Hình 2.5: Nguyên lý đề xuất

động sẽ được tính như công thức 2.5 - 2.6

$$\text{Sai khác khối tĩnh: } I_t(x) = I_{t-1}(x) \Leftrightarrow R_t(x) = EXP(x) = 0 \Leftrightarrow D_t(x) = D_{t-1}(x) \quad (2.5)$$

$$\text{Sai khác khối động: } R(x) = I_t(x) - D_{t-1}(x) \quad (2.6)$$

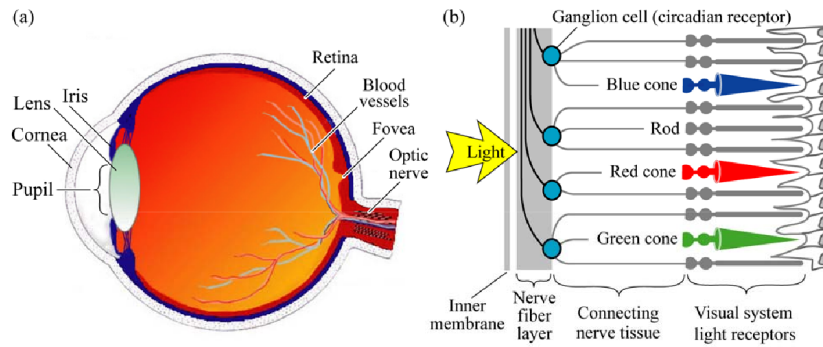
Với việc tác khung hình gốc thành hai cho phép thực hiện tính sai khác của các thành phần tĩnh chính xác như giá trị mong muốn, giá trị 0. Giá trị 0 này giúp tăng hiệu quả nén của mã hóa Entropy cũng như tăng chất lượng ảnh của các vật thể tĩnh. Hơn nữa, kết quả này cho phép không giới hạn số lượng khung hình inter khi không tồn tại các chuyển động trong các khung hình liên tiếp.

2.3.1 Thuật toán xác định chuyển động

Áp dụng các thuật toán xác định chuyển động là phương án cho kỹ thuật phân tách chuyển động và đã đạt được một số thành quả nhất định. Các báo cáo [13] [3] đã trình bày chi tiết việc áp dụng thuật toán xác định chuyển động Sigma-Delta và Zipfian. Tuy nhiên, xây dựng khung hình tham chiếu (chỉ sử dụng để tính trạng thái điểm ảnh) bằng các thuật toán trên đòi hỏi số lượng lớn phép tính trên một điểm ảnh. Trong báo cáo [13], với mỗi khối ảnh 8x8 tiêu tốn 200 phép tính cộng/trừ, con số này tương đương với mỗi điểm cần thêm khoảng 3 phép tính. Đồng thời, khả năng bám theo các vật thể chuyển động nhanh là rất hạn chế. Do vậy, yêu cầu đặt ra là phải phát triển một phương pháp phân tách chuyển động đơn giản hơn, hiệu quả hơn và có khả năng bám chuyển động tốt hơn.

2.3.2 Đặc tính của mắt người đối với thành phần xám và khả năng áp dụng cho phân tách chuyển động

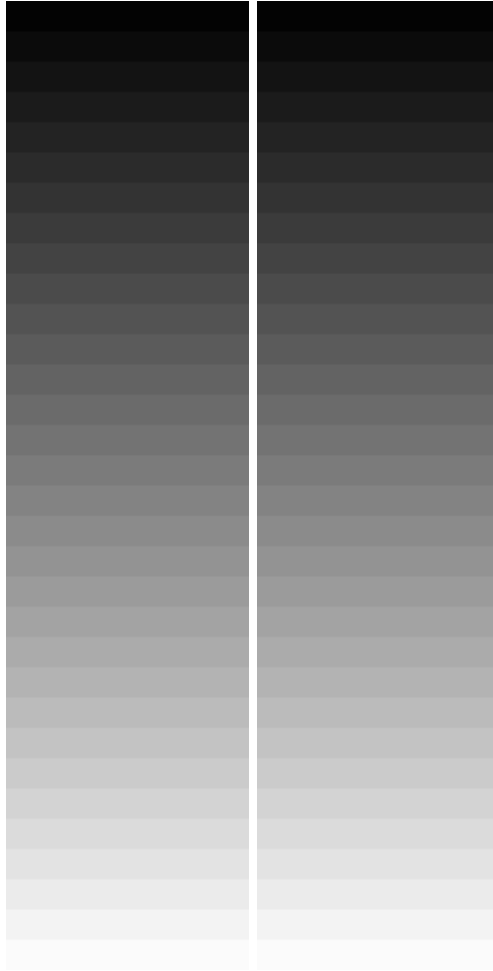
Hình 2.6 trình bày cấu trúc cơ bản của mắt người. Hình 2.6(b) trình bày cấu trúc của võng mạc, bao gồm: tế bào que (rod cell) và tế bào nón (cone cells). Tế bào que có khả



Hình 2.6: Cấu trúc mắt người (Encyclopedia Britannica, 1994)

năng phân biệt cấp độ sáng của ảnh thu nhận được, hay phân biệt cấp độ của ảnh xám. Tế bào nón hoạt động và nhận biết cấp độ sáng trong toàn bộ dải phổ nhìn thấy của mắt người. Bên cạnh tế bào que, có 3 loại tế bào nón riêng biệt, mỗi loại lại có khả năng nhạy cảm với dải phổ của các thành phần màu khác nhau: đỏ (red), xanh lá cây (green) và xanh dương (blue).

Tại một báo cáo trên trang "Journal of the Royal Society Interface", Eric Kreit và những đồng sự của mình tại đại học Cincinnati đã trình bày [14] rằng hệ thống thị giác của con người có khả năng phân biệt được khoảng 10 triệu màu sắc khác nhau. Tuy nhiên, đối với thành phần xám, mắt người chỉ có khả năng phân biệt khoảng 30 sắc thái khác nhau mặc dù nhạy cảm nhất đối với thành phần này. Một ví dụ phổ biến của việc áp dụng tính chất của hệ thống thị giác con người đối với ảnh xám là các sản phẩm máy đọc sách (ereader). Các thiết bị này thay vì sử dụng 256 sắc thái của màu xám thì chỉ sử dụng duy nhất 16 sắc thái khác nhau. Hình 2.7 trình bày tất cả các sắc thái của ảnh xám có thể được biểu diễn bằng 5-bit (trái) và 8-bit (phải) dữ liệu. Mỗi bức ảnh có 32 hàng. Một hàng trong ảnh 5-bit chỉ biểu diễn duy nhất một cấp độ của màu xám, trong khi đó ảnh 8-bit biểu diễn 8 cấp độ liên tiếp khác nhau. Kết quả của nghiên cứu [14] mở ra cơ



Hình 2.7: Toàn bộ sắc thái xám của ảnh 5 bit và 8 bit

hội áp dụng đặc tính độ nhạy cảm của mắt người đối với thành phần xám cho phân tách chuyển động. Nguyên tắc hoạt động được trình bày như sau: "Một điểm ảnh là chuyển động khi thành phần xám của điểm ảnh có sự thay đổi nằm trong khả năng nhận biết của hệ thống thị giác của con người. Một điểm ảnh là tĩnh khi thành phần xám của điểm ảnh không có sự thay đổi hoặc sự thay đổi nằm ngoài khả năng nhận biết của hệ thống thị giác của con người." Trong ảnh số, giới hạn này được tính như công thức sau 2.7:

$$pthresh \leq 2^{b-5} \quad (2.7)$$

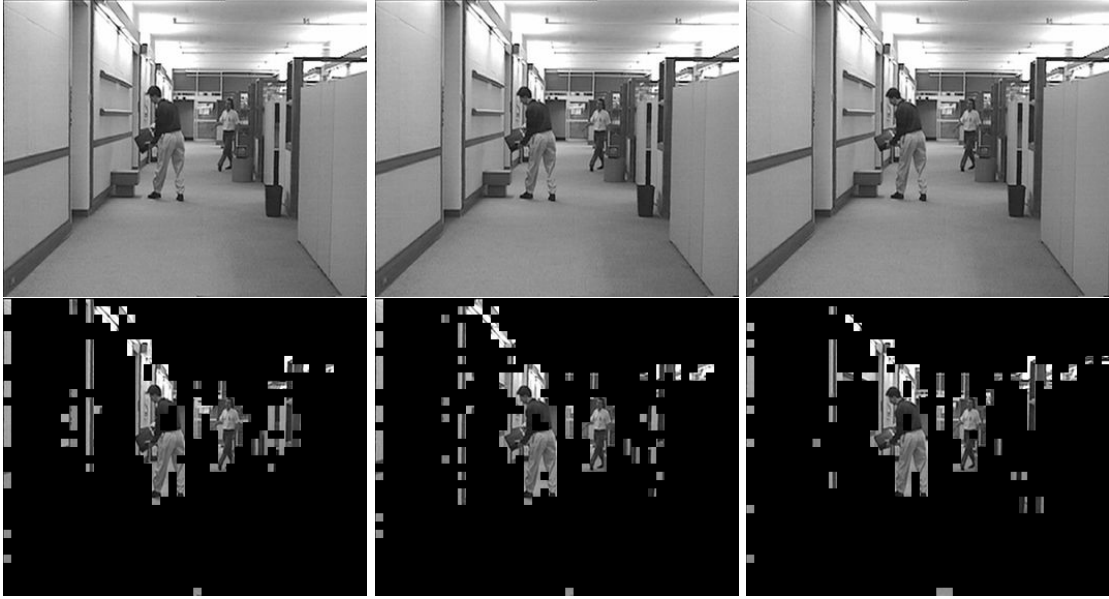
Trong đó:

$pthresh$ là ngưỡng sai khác của ảnh xám.

b là số bit biểu diễn một điểm ảnh xám trong ảnh gốc.

5 là số bit điều khiển 32 mức độ xám khác nhau phù hợp với giới hạn mắt của người.

Hình 2.8 trình bày các khung ảnh gốc và ảnh chuyển động trong các khung hình từ 93-95 trong video mẫu hall bằng việc áp dụng đặc tính của mắt người đối với thành phần xám. Hàng đầu tiên là các khung hình gốc, hàng thứ 2 là các khung hình với các khối chuyển động với màu sắc khác nhau, các khối tĩnh với màu đen. Hình ảnh từ hàng thứ 2



Hình 2.8: Kết quả sử dụng đặc tính của mắt người

xuất phát từ công thức 2.8. Kết quả trình bày các vật thể động một cách chính xác, đảm bảo bao trùm toàn bộ vật thể chuyển động. Điều này tạo tiền đề cho việc áp dụng đặc tính của mắt người đối với thành phần xám trong kỹ thuật phân tách chuyển động.

$$I_t(x) = \begin{cases} 0, & \text{nếu } I_t(x) - I_{t-1}(x) < pthresh \\ I_t(x) & \end{cases} \quad (2.8)$$

Trong đó:

I là khung hình đầu vào

t là thứ tự khung hình

x là vị trí điểm ảnh

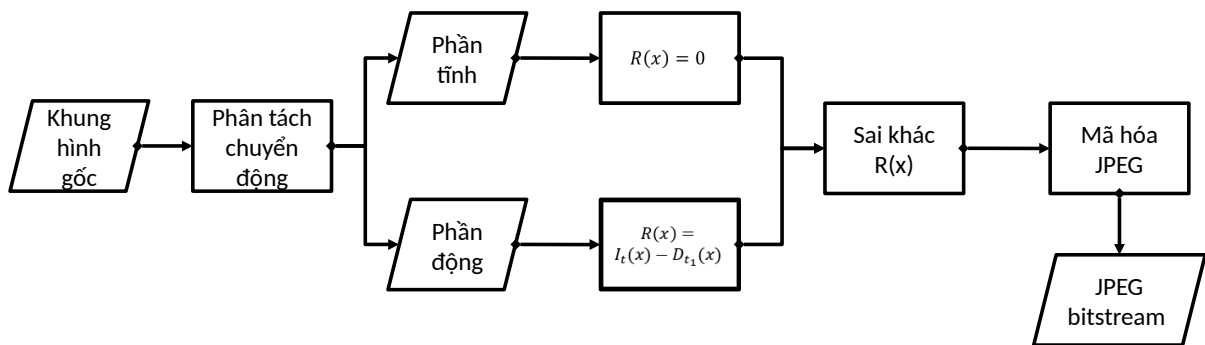
2.4 Phương pháp xác định chuyển động dựa trên đặc tính của mắt người đối với thành phần xám trong việc loại bỏ dư thừa về mặt thời gian của mã hóa MJPEG

Mục tiêu của nghiên cứu là xây dựng một phương pháp mã hóa video cho các camera trong các mạng cảm biến không dây. Phương pháp mã hóa video có khả năng hoạt động trong các điều kiện bị giới hạn về năng lực tính toán, năng lượng và băng thông. Đồng thời phương pháp mã hóa đảm bảo tính linh hoạt và có khả năng hoạt động trên nhiều nền tảng IoT khác nhau.

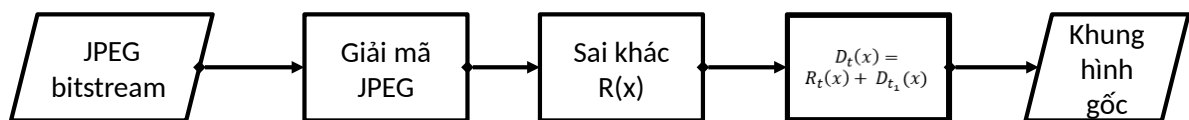
Ý tưởng về sự kết hợp giữa phương pháp mã hóa video MJPEG cùng với đặc tính của mắt người đối với thành phần xám là kết quả của quá trình nghiên cứu và phân tích các chuẩn, phương pháp mã hóa video, xu thế IoT và đặc tính của mắt người đối với ảnh số. Mã hóa MJPEG đảm bảo khả năng hoạt động trong các điều kiện bị giới hạn về năng lượng và năng lực tính toán. MJPEG được xây dựng dựa trên mã hóa ảnh JPEG, được sử dụng và hỗ trợ rộng rãi trên nhiều nền tảng khác nhau. Áp dụng tính chất của mắt người đối với thành phần xám trong kỹ thuật phân tách chuyển động là phương án hiệu quả, đơn giản giúp MJPEG có thể loại bỏ một phần dư thừa về thời gian và đặc biệt cho phép linh động GOP.

2.4.1 Nguyên lý hoạt động

Hình 2.9 trình bày sơ đồ nguyên lý hoạt động của việc áp dụng kỹ thuật loại bỏ dư thừa thời gian dựa trên phân tách chuyển động cho MJPEG. Các khung hình gốc sẽ được đưa vào quá trình phân tách chuyển động để chia từng khung hình ra làm hai phần riêng biệt: phần chuyển động và phần tĩnh. Đối với phần tĩnh, dữ liệu sai khác được gán bằng 0. Đối với thành phần động, dữ liệu sai khác được tính như trong kỹ thuật loại bỏ dư thừa bằng việc sử dụng khối tiên đoán là khối có cùng vị trí và nằm tại khung hình liền kề phía trước như trong mục 2.2.1. Tổng hợp sai khác của cả phần tĩnh và phần động chính là khung hình sai khác $R(x)$ và sẽ được thực hiện mã hóa bằng JPEG.



Hình 2.9: Nguyên lý mã hóa

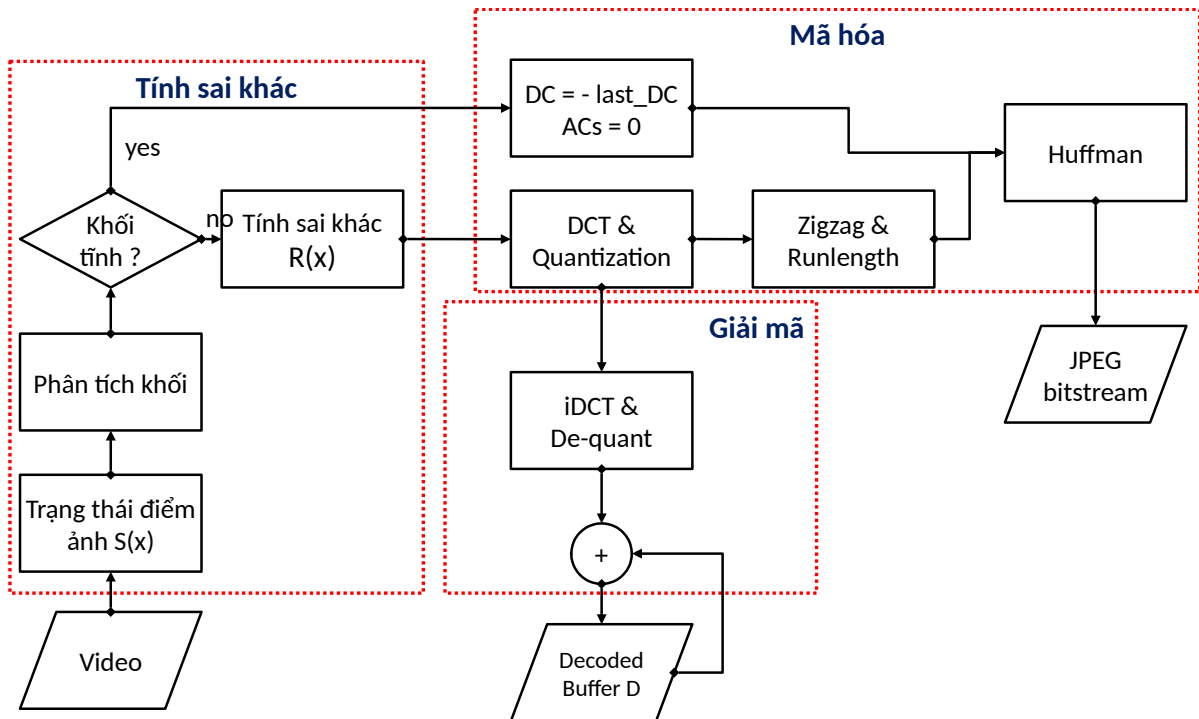


Hình 2.10: Nguyên lý giải mã

Hình 2.10 trình bày sơ đồ giải mã phương án đề xuất. Thành phần chính trong sơ đồ này là bộ giải mã JPEG. Bộ giải mã này sẽ giải mã dữ liệu nén sai khác từ sơ đồ Hình 2.9. Sau đó, dữ liệu sai khác $R(x)$ được cộng với khung tham chiếu tương ứng để nhận

được dữ liệu tương đương với khung hình gốc. Sơ đồ giải mã này cũng chính là sơ đồ giải mã cho phương pháp nén sai khác dựa vào khung phía trước trong nén MJPEG truyền thống.

2.4.2 Sơ đồ mã hóa chi tiết



Hình 2.11: Sơ đồ mã hoá chi tiết

Hình 2.11 trình bày sơ đồ mã hóa chi tiết đã được tối ưu hóa cho áp dụng loại bỏ dư thừa thời gian trong MJPEG. Sơ đồ đề xuất bao gồm các giai đoạn chính theo thứ tự như sau: tính sai khác, mã hóa khối động - tĩnh, giải mã khung tham chiếu. Giai đoạn tính sai khác thực hiện tính toán trạng thái của từng khối trong khung hình và tính toán khung hình sai khác. Giai đoạn này được thực hiện bởi 3 quá trình chính: tính trạng thái điểm ảnh, phân tích khối và tính sai khác. Sau khi có dữ liệu về trạng thái khối và sai khác, giai đoạn mã hóa thực thi việc nén dữ liệu dựa trên loại bỏ 3 loại dư thừa: dư thừa về nhận thức, dư thừa về mặt không gian và dư thừa về mặt thống kê bằng mã hóa JPEG. Giai đoạn cuối cùng là giải mã nhằm đảm bảo sự đồng bộ khung tham chiếu giữa hai bộ mã hóa và giải mã.

Trạng thái của điểm ảnh

Tính trạng thái của từng điểm ảnh $S(x)$ là bước đầu tiên của quá trình mã hóa. Từng điểm ảnh trong các khối sẽ được kiểm tra trạng thái: chuyển động hay tĩnh. Trạng thái của điểm ảnh phụ thuộc vào mức độ thay đổi của màu xám của điểm ảnh và đặc tính của

mất người đối sự thay đổi đó như trong 2.3.2. Công thức 2.9 trình bày công thức tổng quát của quá trình này.

$$S(x) = \begin{cases} 0, & \text{nếu } Y(I_t(x)) - Y(I_{t-1}(x)) < pthresh \Leftrightarrow \text{Điểm tĩnh} \\ 1 & \Leftrightarrow \text{Điểm động} \end{cases} \quad (2.9)$$

Trong đó,

$S(x)$ là trạng thái của điểm ảnh x .

$Y(I_t(x))$ là ảnh xám gốc tại điểm x .

Phân tích khối

Sau quá trình 2.4.2 trạng thái các điểm ảnh $S(x)$ đã được cập nhật. Phần này thực hiện phân tích dữ liệu trạng thái của tất cả các điểm ảnh $S(x)$ trong từng khối $B(i)$ và kết luận trạng thái của từng khối. Tương tự như trạng thái điểm ảnh, các khối cũng có duy nhất hai trạng thái là tĩnh và động. Tiếp đó, tùy thuộc vào trạng thái của khối mà việc tính sai khác cũng khác nhau.

Phương pháp đánh giá trạng thái của một khối được dựa trên xác suất và phân bố của các điểm tĩnh và động trong khối. Công thức 2.10 - 2.11 trình bày một ví dụ về cách tính trạng thái của khối ảnh. Đối với những khối tĩnh, tỷ lệ các điểm động thông thường nhỏ hơn 10%.

$$B(i) = \begin{cases} 0, & \text{nếu } \sum_{x \in B(i)} S(x) < bthresh \Leftrightarrow \text{Khối tĩnh} \\ 1 & \Leftrightarrow \text{Khối động} \end{cases} \quad (2.10)$$

$$bthresh = 64 \times p \quad (2.11)$$

Trong đó:

$B(i)$ là khối thứ i trong khung hình,

$S(x)$ là trạng thái điểm ảnh x ,

$bthresh$ là số điểm ảnh chuyển động tối đa trong khối,

p là tỷ lệ số điểm ảnh động tối đa trong khối, $p \leq 10\%$

Tính sai khác

Tính sai khác của từng loại khối khác nhau được trình bày như công thức 2.12

$$R(x) = \begin{cases} 0, & \text{nếu điểm ảnh thuộc khối tĩnh} \\ I_t(x) - D_{t-1}(x), & \text{nếu điểm ảnh thuộc khối động} \end{cases} \quad (2.12)$$

Trong đó,

I là khung ảnh gốc.

D là khung ảnh sau giải mã.

x là vị trí điểm ảnh.

t là thứ tự khung hình trong video.

Mã hóa khối tĩnh - động

Trạng thái của các khối ảnh đại diện một phần cho dữ liệu sai khác trong khối. Khối tĩnh đại diện cho toàn bộ dữ liệu sai khác bên trong bằng 0. Chính đặc điểm không thay đổi này cho phép thiết kế bộ mã hóa thu gọn cho các khối tĩnh.

Như Hình 2.11, các khối tĩnh được mã hóa trực tiếp bằng dữ liệu sai khác DC và dữ liệu 0 của toàn bộ thành phần AC mà không cần thực hiện qua bước biến đổi DCT và lượng tử hóa. Để giải thích cho sự lược giản này, phương trình 2.13 trình bày kết quả thực hiện DCT và lượng tử của ma trận không - khối tĩnh. Kết quả này hoàn toàn không phụ thuộc vào các giá trị của ma trận lượng tử.

$$\text{Sai khác: } \mathbf{R} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \quad (2.13)$$

$$\text{Biến đổi DCT: } \mathbf{G} = \text{DCT}(\mathbf{R}) = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \quad (2.14)$$

$$\text{Lượng tử: } \mathbf{B}_{i,j} = \text{round}\left(\frac{G_{i,j}}{Q_{i,j}}\right) = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \quad (2.15)$$

Đối với các khối động, giai đoạn mã hóa được diễn ra như bình thường của mã hóa JPEG: biến đổi DCT, lượng tử hóa và mã hóa Entropy.

Giải mã khung tham chiếu

Giải mã khung tham chiếu được xây dựng đảm bảo đồng bộ khung tham chiếu giữa bộ mã hóa và bộ giải mã. Hình 2.11 trình bày các bước xây dựng khung hình sau giải mã D_t .

Đối với các khối động, dữ liệu sau giải mã được xây dựng bằng việc thực hiện tuần tự các quá trình sau: giải lượng tử, biến đổi DCT ngược và xây dựng khung hình giải

mã. Quá trình giải lượng tử và biến đổi DCT ngược nhằm tính toán dữ liệu sai khác $R'_t(x)$. Tiếp đó, dữ liệu sai khác $R'_t(x)$ sẽ được cộng với dữ liệu của khung hình phía trước $D_{t-1}(x)$ để khôi phục dữ liệu khung ảnh $D_t(x)$

Trong bộ giải mã không bao gồm các quá trình để giải mã các khối tnh. Điều này xuất phát từ thực tế dữ liệu sai khác $R(x) \equiv R'(x) = 0$. Các công thức 2.16 - 2.19 sau trình bày chi tiết hơn về vấn đề này. Đặc biệt, công thức 2.19 là khởi nguồn trực tiếp cho việc áp dụng kỹ thuật linh động GOP trong mã hóa đề xuất và lược giản giải mã khối tnh.

$$\text{Sau lượng tử: } \mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \quad (2.16)$$

$$\text{Giải lượng tử: } (\mathbf{B} \times \mathbf{Q}) = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \quad (2.17)$$

$$\text{Sai khác: } \mathbf{R}' = \mathbf{iDCT}(\mathbf{b} \times \mathbf{Q}) = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \quad (2.18)$$

$$\text{Khung hình sau giải mã: } \mathbf{D}_t = \mathbf{R}' + \mathbf{D}_{t-1} = \mathbf{D}_{t-1} \quad (2.19)$$

2.5 Kết luận chương 2

Chương 2 đã trình bày những kiến thức cơ bản về dư thừa thời gian và các kỹ thuật loại bỏ thời gian đang được sử dụng từ đơn giản đến phức tạp: kỹ thuật dựa trên khung hình liền kề phía trước, kỹ thuật đánh giá chuyển động dựa trên khối, và kỹ thuật đề xuất dựa trên phân tách chuyển động. Trong đó, với kỹ thuật đề xuất dựa trên phân tách chuyển động xây dựng trên thuật toán Zipfian đã đạt được những thành quả nhất định. Tuy nhiên, việc áp dụng các thuật toán xác định chuyển động vẫn yêu cầu số lượng lớn các phép tính trên từng điểm ảnh. Do đó, phương pháp phân tách chuyển động dựa trên đặc tính của mặt người đối với thành phần xám đã được học viên đề xuất và giới thiệu.

Từ việc phân tích các kỹ thuật loại bỏ dư thừa về mặt thời gian, chương 2 đã trình bày rõ ràng cách áp dụng đặc tính của mặt người trong kỹ thuật phân tách chuyển động cho loại bỏ dư thừa về mặt thời gian của mã hóa MJPEG.

Chương 3

MÔ PHỎNG VÀ ĐÁNH GIÁ PHƯƠNG ÁN ĐỀ XUẤT

3.1 Thông số đánh giá

3.1.1 Tỷ lệ nén

Khả năng nén dữ liệu là một trong những thông số quan trọng nhất trong việc đánh giá hiệu quả của các thuật toán/ phương pháp mã hóa video. Phương pháp mã hóa video càng hiệu quả thì cho tỷ lệ nén càng cao. Đối với một số chuẩn mã hóa tiên tiến thì tỷ lệ nén so với file video gốc có thể lên đến 100 như H.264/AVC hay 200 đối với HEVC. Công thức tính tỷ lệ nén của các thuật toán được trình bày như sau:

$$\text{Tỷ lệ nén} = \frac{\text{Kích thước file gốc}}{\text{Kích thước file sau mã hóa}} \quad (3.1)$$

3.1.2 Chất lượng hình ảnh giải mã PSNR

Để có thể so sánh được hiệu quả của các phương pháp mã hóa video thì phải đánh giá được tỷ lệ nén của từng phương pháp trên cùng một chất lượng ảnh/video sau giải mã. Đánh giá chất lượng hình ảnh thì thông số thông dụng là PSNR (Peak Signal to Noise Ratio - Tỷ số tín hiệu cực đại trên nhiễu).

PSNR [15] hay còn gọi là tỷ số tín hiệu cực đại trên nhiễu. Thông số này đại diện cho tỷ lệ giữa giá trị năng lượng tối đa của một tín hiệu và năng lượng nhiễu ảnh hưởng đến độ chính xác của thông tin. Do đó, PSNR được dùng để đo chất lượng của các tín hiệu khôi phục sau một quá trình nén có mất mát. Tỷ số PSNR càng cao tương ứng với sự sai khác giữa ảnh gốc và ảnh khôi phục càng nhỏ. Công thức 3.2 trình bày cách tính thông số PSNR.

$$MSE = \frac{1}{m \times n} \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \left(I_{(i,j)} - D_{(i,j)} \right)^2 \quad (3.2)$$

$$PSNR = 10 \times \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \quad (3.3)$$

Trong đó,

$m \times n$ là kích thước ảnh

I và D lần lượt là khung ảnh gốc và khung ảnh được giải mã.

MSE là mean squared signal,
 MAX_I là giá trị tối đa của điểm ảnh, ảnh 8-bit là 255.

3.1.3 Độ phức tạp

Độ phức tạp của thuật toán mã hóa video là một thông số rất quan trọng khi sử dụng cho các hệ thống bị giới hạn về năng lực tính toán. Số lượng phép tính và tốc độ mã hóa của thuật toán là những thông số đại diện cho khả năng thực thi của cũng như độ phức tạp của mã hóa.

3.2 Mô phỏng và đánh giá kết quả

Bảng 3.1: Môi trường mô phỏng

Thông số	Ghi chú
Thư viện JPEG	libjpeg-6b [16]
Hệ điều hành	Centos 6.5 x32
GNU version	g++/gcc 4.4.7
Matlab	7.9.0.529
CPU	Intel(R) Core(TM)2 Duo CPU E4400 @ 2.00GHz

Bảng 3.2: Video tham gia mô phỏng

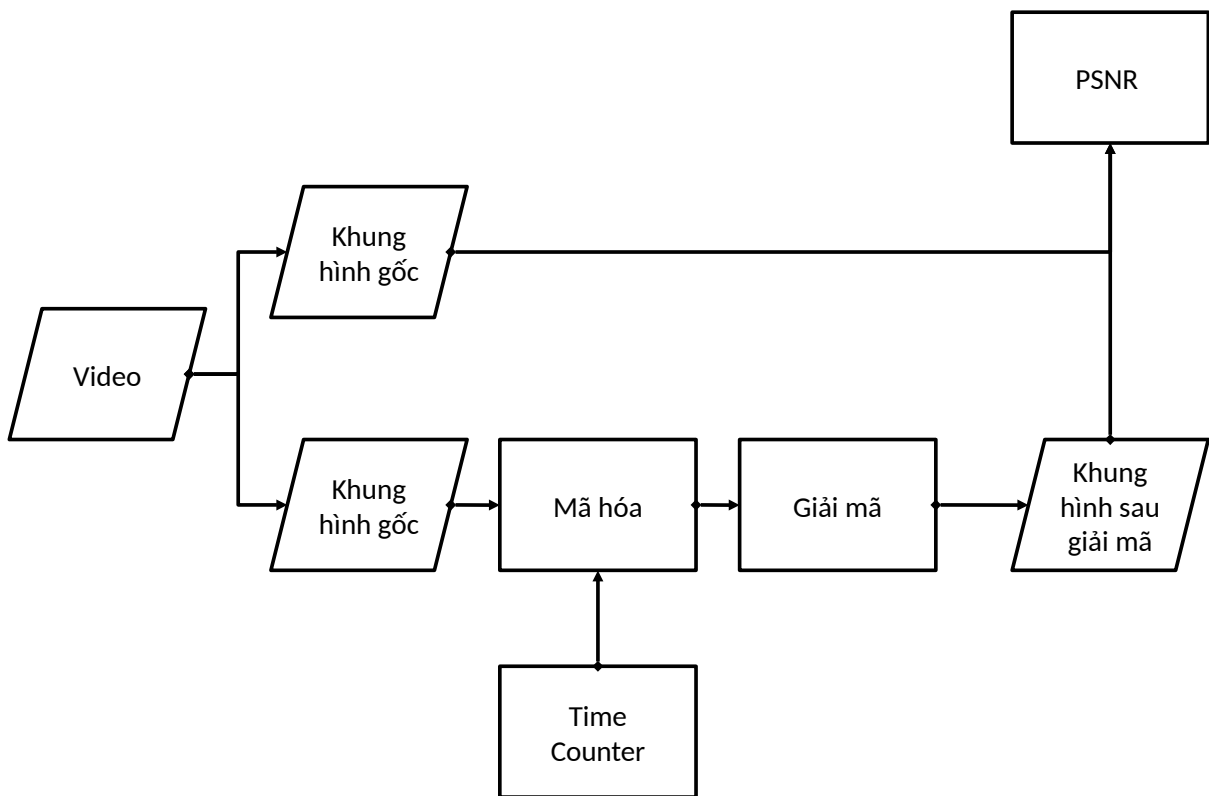
Video	Kích thước	FPS	Nhóm	Ghi chú
Akiyo	CIF	30	A	Chương trình thời sự
Container	CIF	30	A	Video ngoài bến cảng
Foreman	CIF	30	B	Người nói trước máy quay
Hall	CIF	30	A	Camera giám sát hành lang
News	CIF	30	B	Chương trình thời sự với nhiều chuyển động
Soccer	CIF	30	C	Trận đấu bóng

Nhóm A có độ chi tiết thấp và số lượng chuyển động nhỏ Nhóm B có số lượng chuyển động ở mức trung bình. Nhóm C có số lượng chuyển động ở mức cao và liên tục.

Bảng 3.1 - 3.2 trình bày các điều kiện của môi trường mô phỏng và đặc điểm của các video tham gia mô phỏng.

3.2.1 Kịch bản mô phỏng

Như đã trình bày trong phần 3.1 có ba thông số chính để đánh giá khả năng của các phương pháp mã hóa video: PSNR, tỷ lệ nén, tốc độ mã hóa. Mô hình đánh giá kết quả thuật toán được hình bày như Hình 3.1. Trong đó phần đánh giá PSNR sẽ được sử dụng bằng phần mềm thứ 3, được xây dựng trên các hàm có sẵn của Matlab. Phần đánh giá tốc độ mã hóa sẽ được thực hiện bằng các hàm đếm thời gian trong hệ điều hành. Các phương pháp tham gia mã hóa bao gồm: MJPEG, nén sai khác trong MJPEG, MJPEG kết hợp với thuật toán xác định chuyển động và phương pháp đề xuất.



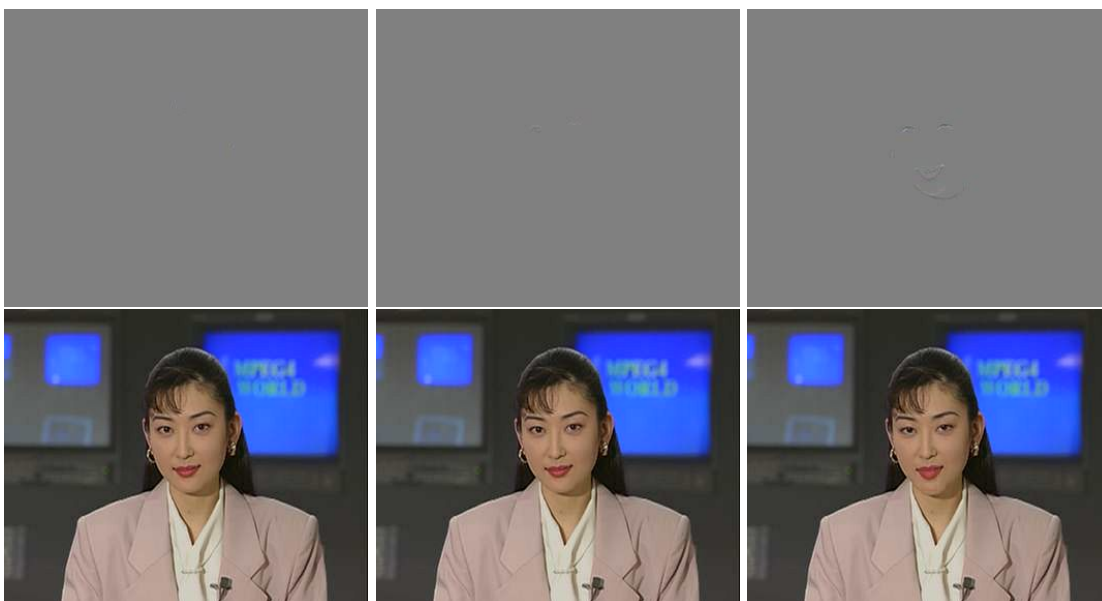
Hình 3.1: Sơ đồ cho kịch bản đánh giá

3.2.2 Kết quả mô phỏng

Ký hiệu	Ý nghĩa
conv	Phương pháp nén MJPEG truyền thống
diff	Phương pháp nén sai khác trong MJPEG
modi	Phương pháp nén sử dụng MJPEG kết hợp thuật toán xác định chuyển động zipfian
xbit	Phương pháp nén đề xuất sử dụng $const = 8$
nth-frame	Thứ tự khung hình trong video, bắt đầu từ 0
QP	Giá trị tính ma trận lượng tử trong mã hóa JPEG
Videos	Akiyo, container, foreman, hall, news, soccer [17]

Bảng 3.3: Các ký hiệu sử dụng chung trong phần mô phỏng

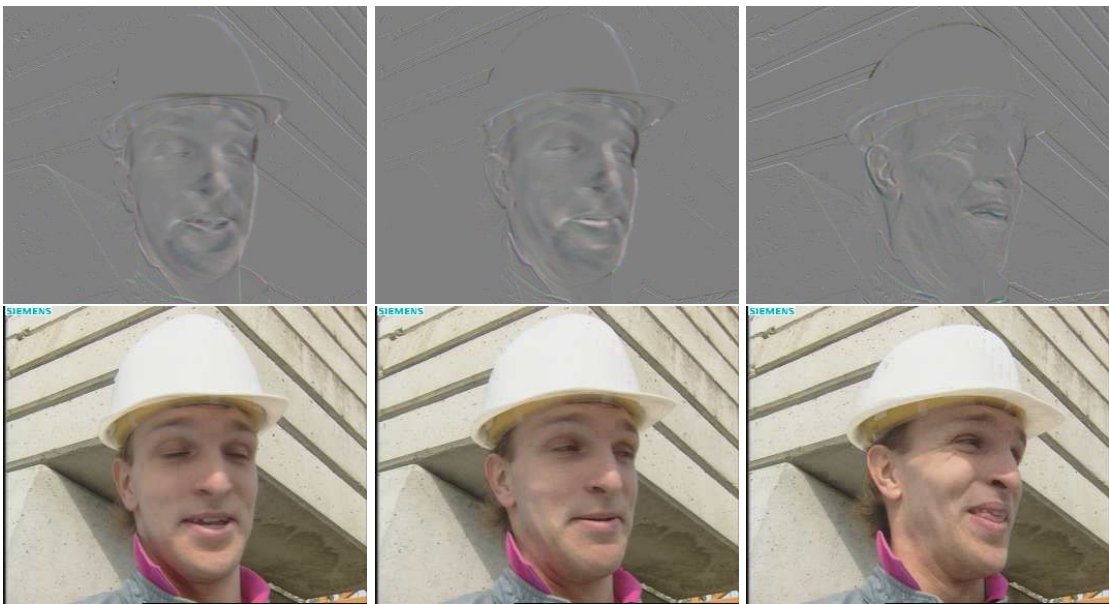
Khung hình mã hóa - giải mã



Hình 3.2: Khung hình mã hóa - giải mã video akiyo



Hình 3.3: Khung hình mã hóa - giải mã video container



Hình 3.4: Khung hình mã hóa - giải mã video foreman



Hình 3.5: Khung hình mã hóa -giải mã video hall



Hình 3.6: Khung hình mã hóa -giải mã video news

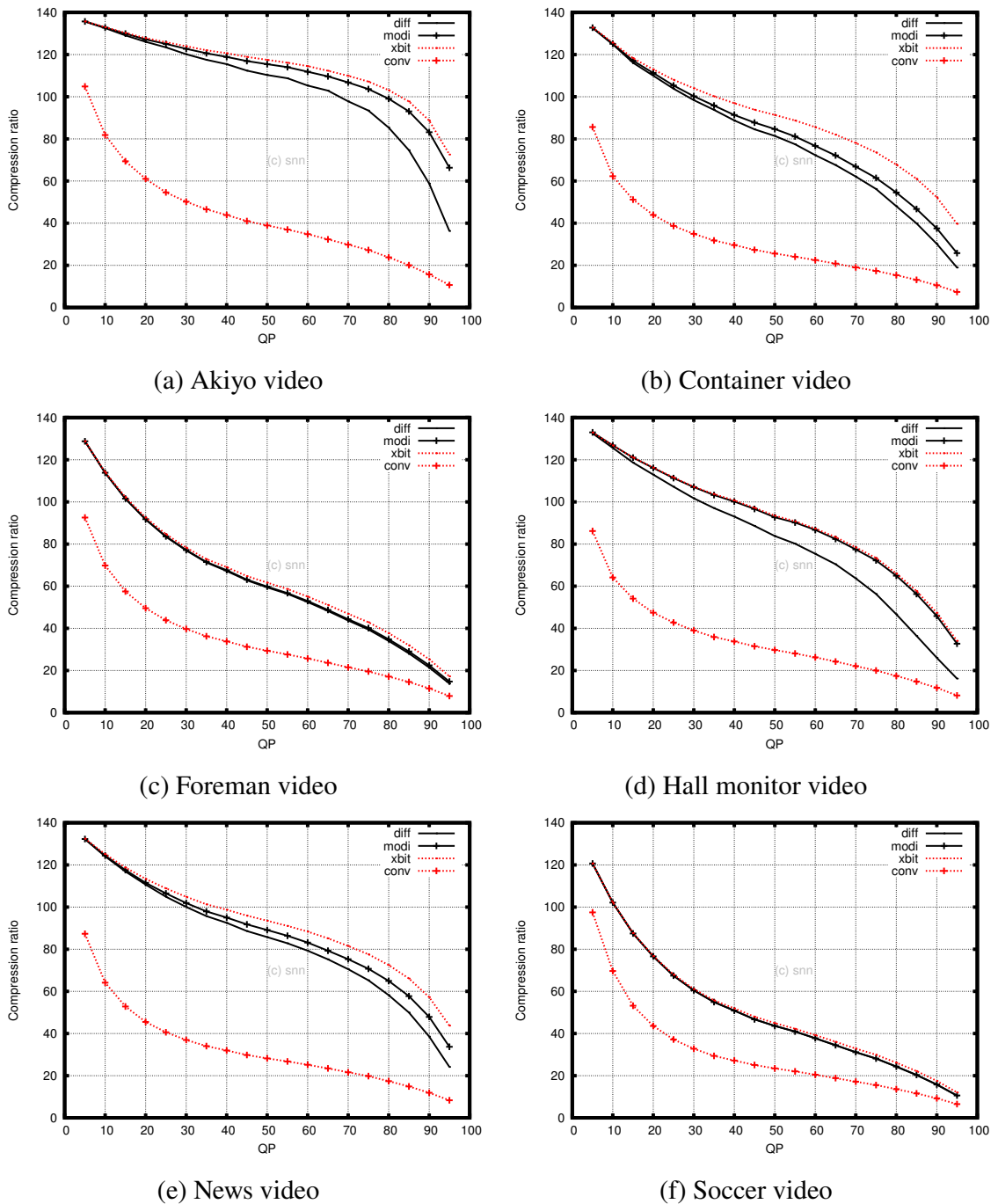


Hình 3.7: Khung hình mã hóa -giải mã video soccer

Hình 3.2 - Hình 3.7 trình bày các khung hình mã hóa bằng phương pháp đề xuất ở hàng đầu tiên. Hàng thứ 2 là các khung hình đã được giải mã, và sử dụng chung bộ giải mã với phương pháp nén sai khác MJPEG và MJPEG kết hợp với thuật toán xác định chuyển động. Các khung hình sử dụng nằm lần lượt ở vị trí thứ 2, 4 và 10 trong chuỗi mã hóa 10 khung hình đầu tiên. Tại các khu vực nền, mã hóa đề xuất cho một hình ảnh rất mịn bởi hoàn toàn không có sai số lượng tử. Tuy nhiên, giống như phương pháp dựa trên khung hình liền kề phía trước, phương pháp đề xuất không có khả năng loại bỏ dư thừa của các vật thể chuyển động. Năng lượng dư thừa này có thể thấy rõ trong các video loại B *Foreman* và loại C *Soccer*.

Tỷ lệ nén

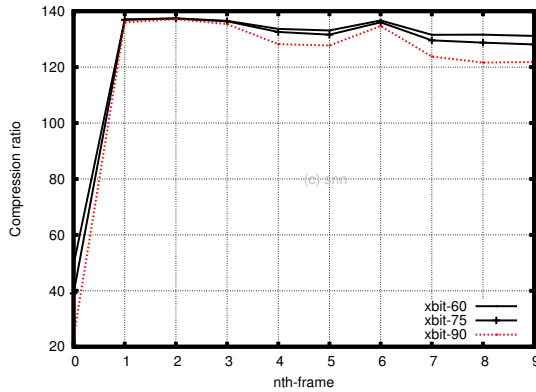
Hình 3.8 trình bày tỷ lệ nén trung bình của 10 khung hình đầu tiên trong các video mẫu với các phương pháp mã hóa khác nhau tại QP [5-90]. Trong 10 khung hình, khung hình đầu tiên được nén theo phương pháp intra, các khung hình còn lại được nén theo phương pháp inter sử dụng các kỹ thuật loại bỏ dư thừa thời gian khác nhau. Trong các phương án được đề cập, MJPEG cho kết quả tỷ lệ nén thấp hơn hẳn so với các phương pháp khác. Cùng với một hằng số QP thì các phương pháp sử dụng kỹ thuật loại bỏ dư thừa về mặt thời gian có thể cho hiệu quả nén tăng gấp đôi. Nguyên nhân xuất phát từ thực tế mã hóa MJPEG không sử dụng các kỹ thuật loại bỏ dư thừa về mặt thời gian. Trong các phương án có loại bỏ dư thừa về mặt thời gian, tỷ lệ nén tăng dần theo thứ tự: nén sai khác MJPEG, MJPEG và thuật toán Zipfian, MJPEG và đặc tính của mắt người. Sự chênh lệch về tỷ lệ nén này tăng dần theo chỉ số QP và phụ thuộc vào tỷ lệ các thành phần tĩnh trong video. Tại các chỉ số QP thấp, các hệ số lượng tử trong ma trận lượng



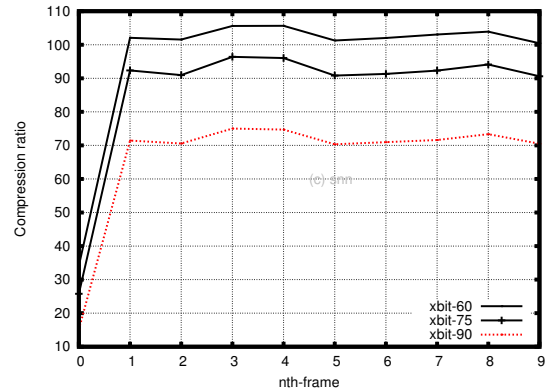
Hình 3.8: Kết quả nén video mẫu tại QP[5-95]

tử có giá trị cao và các thành phần AC sau lượng tử tiến gần giá trị 0 hơn. Trong ba mã hóa kể trên, nén sai khác trong MJPEG có tỷ lệ nén kém nhất xuất phát từ sai số lượng tử hóa. Mã hóa MJPEG với Zipfian có tỷ lệ nén kém hơn MJPEG với đặc tính của mắt người bắt nguồn từ việc các thuật toán xác định chuyển động luôn cố gắng bao trùm toàn bộ vật thể mặc dù nội tại không chuyển động. Tỷ lệ thành phần tĩnh cũng tác động không nhỏ đến sự chênh lệch, với những video có số lượng chuyển động lớn và nhanh thì sự sai khác giữa các mã hóa MJPEG có loại bỏ sai khác càng nhỏ hơn.

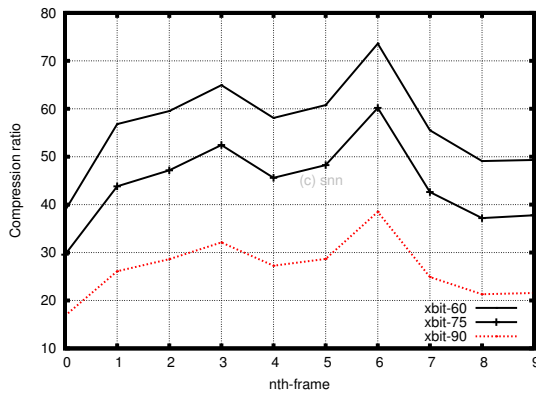
Tỷ lệ nén của khung hình intra, inter



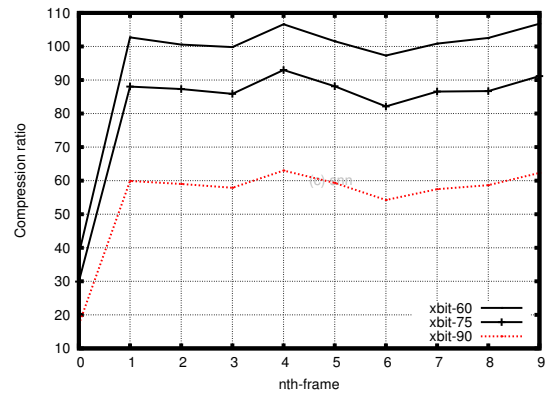
(a) Akiyo video



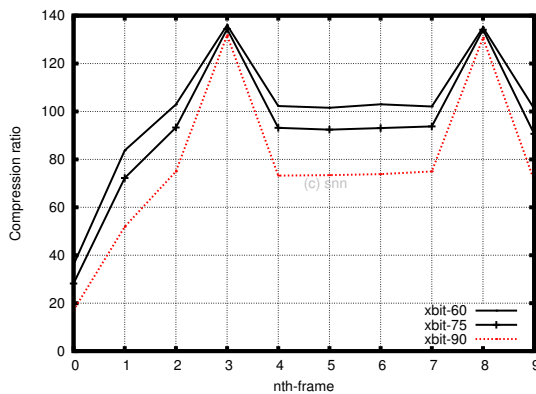
(b) Container video



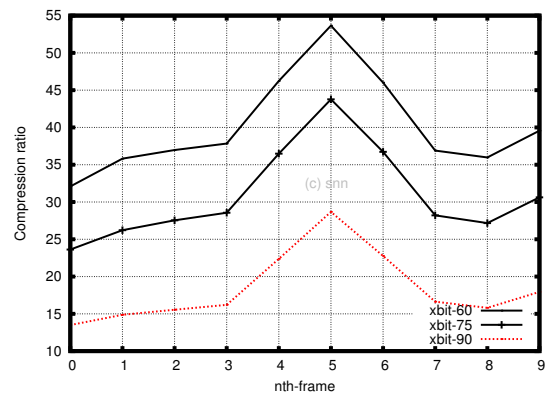
(c) Foreman video



(d) Hall monitor video



(e) News video



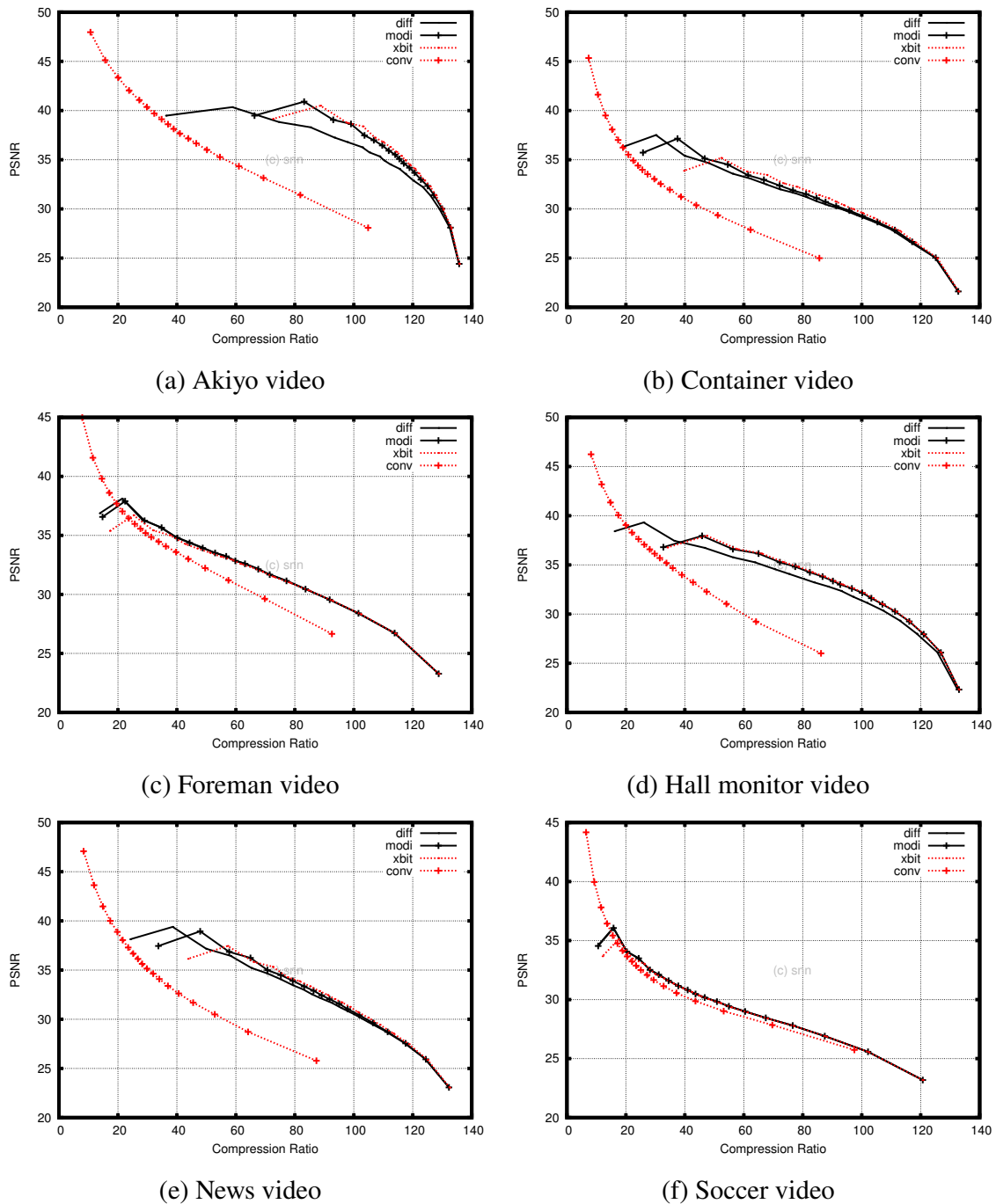
(f) Soccer video

Hình 3.9: Tỷ lệ nén inter-intra của video mẫu

Hình 3.9 trình bày tỷ lệ nén của từng khung hình trong 10 khung hình được sử dụng bằng phương pháp đề xuất tại các giá trị QP phổ biến: 60-75-90. Riêng khung hình đầu tiên là khung hình intra, 9 khung hình tiếp theo là những khung hình inter. Trong hình vẽ, các khung hình inter có tỷ lệ nén biến động rất lớn. Sự biến động này xuất phát từ việc tỷ lệ nén của các khung hình inter phụ thuộc rất nhiều vào tỷ lệ chuyển động và số lượng chuyển động trong từng khung hình là ngẫu nhiên. Tuy nhiên, tỷ lệ nén của khung

hình inter luôn luôn cao hơn so với khung hình intra. Đối với video tồn tại nhiều chuyển động như *Soccer*, khung hình inter chỉ đạt được tỷ lệ nén gấp 1.5 lần so với khung hình intra. Tuy nhiên, tỷ lệ này có thể đạt được vài chục lần đến một trăm lần trong có video tồn tại ít chuyển động như *Akiyo*.

Chỉ số PSNR và tỷ lệ nén tương ứng của các phương pháp mã hóa MJPEG



Hình 3.10: Kết quả PSNR và tỷ lệ nén của các video mẫu

Hình 3.10 trình bày PSNR và tỷ lệ nén tương ứng của 10 khung hình đầu với các

thuật toán nén video khác nhau. Bên cạnh việc so sánh tỷ lệ nén của các phương pháp mã hóa dựa trên MJPEG tại những điểm QP giống nhau thì việc so sánh tỷ lệ nén tại cùng chất lượng ảnh (ví dụ cùng PSNR) là rất quan trọng. Bởi vì các mã hóa này sử dụng những kỹ thuật khác nhau để đạt được tỷ lệ nén mong muốn nên dư thừa được loại bỏ cũng khác nhau. Hơn nữa, mục tiêu cuối cùng của mã hóa là đạt được tỷ lệ nén cao mà vẫn đảm bảo tính đúng đắn của video sau giải mã. Theo Hình 3.10, MJPEG có khả năng đạt được PSNR cao nhất. Tất cả điều này xuất phát từ thực tế MJPEG không loại bỏ dư thừa về mặt thời gian nên kết quả video sau giải mã chỉ chịu ảnh hưởng bởi lượng tử hóa. Tuy nhiên, khi so sánh cùng một chỉ số PSNR, MJPEG vẫn có được tỷ lệ nén thấp nhất. Các mã hóa có kỹ thuật loại bỏ dư thừa về thời gian có tỷ lệ nén cao hơn tại cùng một PSNR là bởi có nhiều dư thừa hơn đã được loại bỏ và có khả năng khôi phục tốt hơn (dư thừa về mặt thời gian). Những kỹ thuật loại bỏ được nhiều dư thừa hơn càng có tỷ lệ nén cao hơn tại cùng PSNR.

Độ phức tạp

Bảng 3.4: Số phép tính của các thuật toán Zipfian, biến đổi DCT thuận, Lượng tử và phương pháp đề xuất trên từng khối 8x8 [3]

Phép tính	Zipfian	FDCT	Lượng tử hóa	Đề xuất
Cộng/Trừ	200	928	64	≤ 64
Nhân	0	192	64	0
Dịch	20	128	0	0

Bảng 3.4 trình bày số lượng các phép tính sử dụng cho một khối 8x8. Trong đó, Zipfian và số lượng các phép tính của DCT, lượng tử được trích xuất từ bài báo [3]. Giá trị số lượng phép tính cộng/trừ của phương án đề xuất nhỏ hơn 64 phép tính xuất phát từ công thức 2.9 - 2.10 và số lượng điểm ảnh trong mỗi khối là 64. Theo như công thức 2.9, mỗi điểm ảnh chỉ tiêu tốn một phép trừ để tính sai khác. Công thức 2.10 chỉ ra rằng trạng thái của khối có thể được xác định mà không cần đến việc tính sai khác cho toàn bộ điểm ảnh. Như trình bày trong Bảng 3.4, số lượng phép tính cho từng khối của thuật toán Zipfian ít hơn hẳn so với biến đổi DCT thuận và lượng tử. Trong báo cáo [3], những video có số lượng khối tĩnh nhiều hơn 60%, MJPEG và Zipfian có số lượng phép tính toán nhỏ hơn so với MJPEG truyền thống. Mặc dù vậy, Zipfian vẫn yêu cầu một số lượng phép tính nhiều hơn hẳn phương án đề xuất, con số này là khoảng 3 lần.

KẾT LUẬN

Trong suốt thời gian nghiên cứu thực hiện luận văn, với sự hướng dẫn tận tình của PGS.TS Trần Xuân Tú, cùng với những cố gắng và nỗ lực của bản thân, toàn bộ nội dung của luận văn đã hoàn thiện và đáp ứng được các yêu cầu đặt ra.

Luận văn đã đề xuất, thiết kế "**Phương pháp xác định chuyển động dựa trên đặc tính của mắt người đối với thành phần xám trong việc loại bỏ dư thừa về mặt thời gian của mã hóa video MJPEG**" phù hợp với camera trong mạng cảm biến không dây.

Kiến trúc đề xuất xuất phát từ thực tế về các đặc tính và giới hạn của mắt người đối với thành phần xám. Trong kiến trúc mã hóa đề xuất, để mã hóa được một khung hình, mã hóa được chia làm ba giai đoạn: phân tách chuyển động dựa trên đặc tính của mắt người, mã hóa khối tĩnh-động và giải mã. Giai đoạn phân tách chuyển động có vai trò phân chia các thành phần tĩnh động trong khung hình. Phân tách hỗ trợ thực hiện giai đoạn mã hóa tĩnh - động. Trong đó, mã hóa các khối tĩnh được lược giản bớt các quá trình không cần thiết: biến đổi DCT và lượng tử hóa. Giai đoạn cuối cùng đảm bảo đồng bộ khung tham chiếu giữa bộ giải mã và mã hóa. Dựa vào đặc tính của các khối tĩnh mà toàn bộ phần giải mã khối tĩnh được lược giản.

Kết quả mô phỏng và đánh giá chỉ ra rằng mã hóa đề xuất đảm bảo được tính đơn giản nhưng hiệu quả trong việc tăng tỷ lệ nén của mã hóa MJPEG. Với chất lượng video sau mã hóa tương đương, phương pháp đề xuất có thể đạt tỷ lệ nén gấp đôi so với MJPEG truyền thống và luôn cao hơn so với các phương pháp loại bỏ dư thừa thời gian đã được nêu trong luận văn: nén sai khác MJPEG, MJPEG với thuật toán xác định chuyển động Zipfian. So sánh độ phức tạp, phương án đề xuất sử dụng một lượng phép tính ít gấp 3 lần so với phương án áp dụng MJPEG với thuật toán xác định chuyển động Zipfian. Điều này đảm bảo phương án đề xuất có thể đạt được tốc độ mã hóa cao hơn so với các phương án mã hóa xây dựng từ MJPEG được đề cập cho các video tồn tại ít chuyển động. Về mặt khả năng ứng dụng, mã hóa đề xuất hoàn toàn phù hợp cho các camera trong mạng cảm biến không dây bởi mã hóa đề xuất có độ phức tạp thấp và cho phép linh động GOP.

Trong tương lai học viên mong muốn được áp dụng các kỹ thuật loại bỏ dư thừa về mặt thời gian đề xuất cho các chuẩn mã hóa tiên tiến hơn như HEVC.

DANH MỤC CÁC CÔNG TRÌNH CỦA TÁC GIẢ LIÊN QUAN ĐẾN LUẬN VĂN

1. Ngoc-Sinh Nguyen, Duy-Hieu Bui, and Xuan-Tu Tran. Reducing temporal redundancy in mjpeg using zipfian estimation techniques. In *2014: the 12th of the biennial IEEE Asia Pacific Conference on Circuits and Systems (IEEE APCCAS 2014)*, pages 65–68, November 2014. Outstanding Student Paper Award & Travel Support Grant Award.
2. Ngoc-Sinh Nguyen, Duy-Hieu Bui, and Xuan-Tu Tran. An Efficient Video Coding Algorithm Targeting Low Bitrate Stationary IoT Cameras. In *2013 IEICE International Conference on Integrated Circuits, Devices, and Verification (ICDV 2013)*, pages 127–132, November 2014.

TÀI LIỆU THAM KHẢO

- [1] “Brief history of video codecs.” <https://www.linkedin.com/pulse/brief-history-video-codecs-yoav-nativ>.
- [2] I. E. Richardson, *The H.264 Advanced Video Coding Standard*. second ed.
- [3] N.-S. Nguyen, D.-H. Bui, and X.-T. Tran, “Reducing temporal redundancy in mjpeg using zipfian estimation techniques,” in *2014: the 12th of the biennial IEEE Asia Pacific Conference on Circuits and Systems (IEEE APCCAS 2014)*, pp. 65–68, November 2014. Outstanding Student Paper Award & Travel Support Grant Award.
- [4] ITU-T, “Video codec for audiovisual services at p x 64 kbit/s,” tech. rep., International Telecommunication Union, March 1993.
- [5] ITU-T, “H.264/AVC,” recommendation, International Telecommunication Union, 2003.
- [6] Thomas Wiegand, Gary J. Sullivan, Gisle Bjøntegaard and Ajay Luthra, “Overview of the H.264/AVC Video Coding Standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, July 2003.
- [7] ITU-T, “H.265/HEVC,” recommendation, International Telecommunication Union, April 2013.
- [8] TGary J. Sullivan, Jens-Rainer Ohm, Woo-Jin Han and Thomas Wiegand, “Overview of the High Efficiency Video Coding (HEVC) Standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, December 2012.
- [9] S. Ponlatha and R. S. Sabeenian, “Comparison of video compression standards,” in *December 2013 International Journal of Computer and Electrical Engineering*, December 2013.
- [10] ITU-T, “Recommendation T.81,” recommendation, The International Telegraph and Telephone Consultative Committee, September 1992.
- [11] “Runlength.” https://en.wikipedia.org/wiki/Run-length_encoding.
- [12] “Huffman coding.” https://en.wikipedia.org/wiki/Huffman_coding.

- [13] N.-S. Nguyen, D.-H. Bui, and X.-T. Tran, “An efficient video coding algorithm targeting low bitrate stationary cameras,” in *2013 IEICE International Conference on Integrated Circuits, Devices, and Verification (ICDV 2013)*, pp. 127–132, November 2014.
- [14] Eric Kreit, Lydia M. Mäthger, Roger T. Hanlon, Patrick B. Dennis, Rajesh R. Naik, Eric Forsythe, Jason Heikenfeld, “Biological versus electronic adaptive coloration: how can one inform the other?,” *Journal of The Royal Society Interface* 10(78), September 2012.
- [15] “Peak Signal to Noise Ratio.” https://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio.
- [16] “JPEG library libjpeg.” <http://www.ijg.org/>.
- [17] “Video samples.” <http://media.xiph.org/video/derf/>.