

**BỘ THÔNG TIN VÀ TRUYỀN THÔNG  
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

**NGUYỄN THỊ HƯƠNG THẢO**

**NGHIÊN CỨU CẢI TIẾN HIỆU NĂNG  
HỆ THỐNG MÃ HÓA/GIẢI MÃ VIDEO PHÂN TÁN**

**LUẬN ÁN TIẾN SĨ KỸ THUẬT**

**CHUYÊN NGÀNH: Kỹ thuật điện tử**

**MÃ SỐ: 9.52.02.03**

**NGƯỜI HƯỚNG DẪN KHOA HỌC:**

- 1. PGS.TS. Vũ Văn San**
- 2. TS. Nguyễn Ngọc Minh**

**Hà Nội - 2020**



## LỜI CAM ĐOAN

Tôi xin cam đoan dưới đây là luận án tốt nghiệp của riêng tôi dưới sự hướng dẫn của các giáo viên hướng dẫn. Tất cả các số liệu, các kết quả trình bày trong luận án hoàn toàn trung thực và có được từ những nghiên cứu mà tôi và nhóm nghiên cứu của tôi đã thực hiện trong quá trình làm luận án.

*Hà Nội, ngày      tháng      năm 2020*

Nghiên cứu sinh

## LỜI CẢM ƠN

Trong quá trình nghiên cứu, tác giả đã nhận được nhiều sự giúp đỡ quý giá. Tác giả xin gửi lời cảm ơn sâu sắc tới những người thầy, người đồng nghiệp PGS.TS Vũ Văn San và TS. Nguyễn Ngọc Minh đã tận tình hướng dẫn, định hướng nghiên cứu khoa học, giúp đỡ tác giả trong quá trình nghiên cứu và hoàn thành Luận án. Tác giả cũng xin gửi lời cảm ơn sâu sắc tới các thành viên trong Lab Công nghệ truyền thông đa phương tiện của Đại học Công nghệ - Đại học Quốc gia Hà Nội, đặc biệt TS. Hoàng Văn Xiêm đã hỗ trợ và có những góp ý khoa học quý báu cho nội dung luận án.

Tác giả xin gửi lời cảm ơn tới Lãnh đạo học viện Công nghệ bưu chính viễn thông, Hội đồng khoa học, Hội đồng Tiến sĩ, Khoa đào tạo sau đại học - Học viện Công nghệ bưu chính viễn thông đã tạo điều kiện thuận lợi cho tác giả hoàn thành luận án này.

Tác giả xin gửi lời cảm ơn tới các đồng nghiệp tại Khoa Kỹ thuật Điện tử 1 – Học viện Công nghệ Bưu chính Viễn thông đã tạo điều kiện về thời gian cũng như có các góp ý cho tác giả về nội dung luận án trong quá trình nghiên cứu và hoàn thiện luận án.

Tác giả xin gửi lời cảm ơn tới gia đình, bạn bè, đồng nghiệp đã luôn ủng hộ, động viên, chia sẻ, giúp đỡ tác giả hoàn thành luận án như ngày hôm nay.

*Hà Nội, ngày      tháng      năm 2020*  
Nghiên cứu sinh

## LỜI MỞ ĐẦU

### Lý do nghiên cứu

Video là một dạng dữ liệu đặc biệt, đầy tính hấp dẫn, dễ nhớ và phổ biến hơn bất cứ một loại dữ liệu nào khác. Các số liệu thống kê cho thấy nội dung video đang tăng mạnh hàng năm theo cấp số nhân. Điều đó cho thấy vai trò quan trọng của video trong tương lai. Cùng với đó là sự bùng nổ của các thiết bị điện tử mới có khả năng bắt giữ, chỉnh sửa, lưu trữ và chia sẻ nội dung video trên toàn thế giới đã dẫn đến một xu hướng mới mà ở đó việc xử lý thông tin được tích hợp triệt để vào các thiết bị và hoạt động hàng ngày. Để đạt được điều này, các thuật toán nén trở nên vô cùng quan trọng, đặc biệt là nén video. Mục đích của nén video là tạo ra một cách biểu diễn dữ liệu video sao cho chiếm ít dung lượng nhất. Vì dữ liệu video nén sẽ đòi hỏi ít không gian lưu trữ hơn và băng thông truyền tải nhỏ hơn, do đó nén video là một phần không thể thiếu của hầu hết các hệ thống lưu trữ, xử lý, truyền thông và hiển thị video. Đặc biệt là băng thông luôn là một yếu tố hạn chế trong nhiều ứng dụng.

Tuy nhiên, nén cũng đi đôi với những trả giá. Thứ nhất, nén càng nhiều sẽ càng làm giảm chất lượng hình ảnh. Thứ hai, độ phức tạp của các thuật toán nén sẽ tăng lên khi muốn tỷ lệ nén tốt hơn. Hơn nữa, tỷ lệ nén tốt nhất cũng phụ thuộc vào thiết bị, vị trí và ứng dụng cụ thể. Tuy nhiên, người dùng không cần phải biết đến các cấu hình phức tạp và các lựa chọn liên quan đến video và định dạng nén. Đó là lý do tại sao có nhiều chuẩn nén video, có thể phù hợp với vô số các kịch bản ứng dụng khác nhau. Các tiêu chuẩn này, còn được gọi là chuẩn mã hóa video, được sử dụng rộng rãi và phát triển không ngừng.

Trong suốt quá trình phát triển của mình, các tiêu chuẩn mã hóa video luôn tuân theo một quy tắc thống nhất. Đó là tiêu chuẩn mã hóa video mới luôn luôn phải làm giảm đáng kể tốc độ bit. Chẳng hạn, chuẩn mã hóa video hiệu suất cao HEVC [54] tiết kiệm tốc độ bit khoảng 50 % so với chuẩn H.264/AVC [65] trước đó với cùng một chất lượng hình ảnh. Tuy nhiên, để đạt được tỷ lệ nén tốt hơn với cùng một chất lượng tương đương sẽ phải trả giá bằng sự phức tạp của cả bộ mã hóa và bộ giải mã. Mặt khác, độ phức tạp bộ mã hóa và độ phức tạp của bộ giải mã cũng không giống nhau. Trong các chuẩn mã hóa video truyền thống, bộ mã hóa thường phức tạp hơn rất nhiều so với bộ giải mã. Cách thiết kế này thường phù hợp với các ứng dụng video truyền thống, ví dụ truyền hình quảng bá, nơi mà chương trình truyền hình được mã hóa một lần và sau đó được giải mã để xem bởi hàng triệu người dùng.

Với sự phát triển mạnh mẽ của mạng không dây, xuất hiện hàng loạt các ứng dụng mới mà ở đó không tuân thủ theo sơ đồ một - nhiều như các ứng dụng trước đó. Có thể kể đến mạng cảm biến video không dây (WVSN). Đối với các ứng dụng mới này, thường sẽ có nhiều bộ mã hóa gửi dữ liệu tới một vài bộ giải

mã trung tâm. Điều này cũng đi ngược lại với mục tiêu thiết kế ban đầu của các chuẩn nén video truyền thống. Trong các mạng này, có rất nhiều yêu cầu truyền dẫn đường lên. Do đó, các cảm biến camera cần có bộ mã hóa đơn giản trong khi các trạm gốc có thể giải mã với nhiều tài nguyên hơn hay nói cách khác bộ giải mã có thể có độ phức tạp cao. Vì vậy thách thức trong các hệ thống này là làm sao có thể thực hiện nén video trên các thiết bị có tài nguyên hạn chế hay nói cách khác các ứng dụng mới này yêu cầu bộ mã hóa nhẹ, hoặc ít nhất phải có sự phân chia độ phức tạp hệ thống một cách linh hoạt, có khả năng chống lại hiện tượng mất gói, hiệu suất nén cao và độ trễ thấp. Nói một cách khác, các ứng dụng mới này sẽ cần một sơ đồ mã hóa video mới có khả năng đáp ứng được các yêu cầu ở trên với hiệu suất mã hóa có thể tiệm cận với các hệ thống mã hóa video dự đoán hiện nay trong khi độ phức tạp bộ mã hóa thấp.

Một giải pháp thay thế có thể đáp ứng các yêu cầu này là mã hóa video phân tán (DVC). Trong mã hóa video phân tán, ước lượng chuyển động và do đó là độ phức tạp của nó được chuyển từ bộ mã hóa sang bộ giải mã. Mục tiêu trọng tâm giờ đây là một bộ mã hóa nhẹ, phù hợp với các thiết bị bị hạn chế về tài nguyên xử lý và cả thời gian xử lý. Ngược lại, bộ giải mã được coi là có tài nguyên phong phú. Một trường hợp ví dụ là quay video trên thiết bị bị hạn chế ví dụ như điện thoại cầm tay và sau đó giải mã trên máy tính ở nhà mà không có bất kỳ hạn chế thời gian nào. Phải nhấn mạnh rằng mục đích của DVC không phải là thay thế cho H.264/AVC hay HEVC bằng cách cung cấp khả năng nén tốt hơn. Trên thực tế, DVC tốt nhất được dự kiến là có thể thực hiện với hiệu năng nén tương đương.

Tuy nhiên bất kể các ưu điểm đầy tiềm năng của DVC như hiệu suất nén cao, công suất thấp, khả năng chống nhiễu nội tại như được chứng minh về mặt lý thuyết trong các nghiên cứu trước đây, cho đến hiện nay vẫn chưa có nền tảng nào thực hiện hoặc kiểm nghiệm DVC. Về mặt lý thuyết, DVC đã được chứng minh có thể đạt được hiệu năng tương đương như các chuẩn mã hóa video truyền thống. Tuy nhiên, các kết quả thực tế cho thấy DVC vẫn chưa đạt được hiệu suất mong muốn như các chuẩn nén video dự đoán hiện nay, đặc biệt trong bối cảnh giữ cho bộ mã hóa có độ phức tạp thấp. Vì vậy cho đến hiện nay, cộng đồng nghiên cứu vẫn đang nỗ lực để cải thiện hiệu năng méo tín hiệu - tốc độ bit (RD) và giải quyết các trở ngại chính của DVC trong các ứng dụng thực tế với các điều kiện bị ràng buộc. Tuy nhiên, trong thực tế vẫn còn một khoảng cách giữa hiệu suất mã hóa của DVC và các chuẩn mã hóa video truyền thống.

### **Mục tiêu nghiên cứu**

Các phân tích ở trên cho thấy rằng DVC đóng một vai trò quan trọng đối với các ứng dụng mới như mạng cảm biến hình ảnh hay mạng giám sát không dây. Tuy nhiên thách thức chính đối với DVC là phải đạt được hiệu năng nén ở mức tiệm cận với các hệ thống mã hóa video dự đoán hiện có. Vì vậy mục tiêu nghiên cứu của luận án là nghiên cứu đề xuất kỹ thuật cải tiến hiệu năng nén

của mã hóa video phân tán DVC và mã hóa video phân tán liên lớp DSVC.

### **Nội dung nghiên cứu**

Với mục tiêu nghiên cứu ở trên, phần nội dung nghiên cứu của Luận án tập trung vào các nội dung sau:

- Đề xuất các phương pháp mới nhằm cải tiến hiệu năng nén cho mã hóa video phân tán bao gồm phương pháp được thực hiện tại phía mã hóa và nhóm các phương pháp thực hiện tại phía giải mã.
- Đề xuất các kỹ thuật tạo thông tin phụ trợ mới để cải tiến hiệu năng nén cho mã hóa video phân tán liên lớp.

### **Đối tượng và phạm vi nghiên cứu**

Đối tượng nghiên cứu là mã hóa video phân tán và mở rộng của mã hóa video phân tán là mã hóa video phân tán liên lớp. Phạm vi nghiên cứu của luận án là các kỹ thuật nhằm nâng cao hiệu năng nén của mã hóa video phân tán và mã hóa video phân tán liên lớp.

### **Ý nghĩa khoa học và thực tiễn của đề tài nghiên cứu**

Ý nghĩa khoa học của luận án mà nghiên cứu sinh hướng đến là đề xuất các kỹ thuật mới nhằm nâng cao hiệu năng nén của mã hóa video phân tán và mã hóa video phân tán liên lớp.

Ý nghĩa thực tiễn của luận án mà nghiên cứu sinh hy vọng đạt được là những kết quả nghiên cứu của luận án góp phần vào việc rút ngắn khoảng cách về hiệu suất mã hóa của mã hóa video phân tán với hệ thống mã hóa video dự đoán hiện nay, sớm đưa mã hóa video phân tán vào các ứng dụng thực tế.

### **Phương pháp nghiên cứu**

Để thực hiện được các nội dung nghiên cứu đề ra, phương pháp nghiên cứu được sử dụng trong Luận án này bao gồm:

- Phương pháp nghiên cứu lý luận: Phân tích, tổng hợp và đánh giá các công trình nghiên cứu, sách chuyên khảo và các nguồn tài liệu khoa học có liên quan đến mã hóa video phân tán và mã hóa video phân tán liên lớp. Từ đó đề xuất các kỹ thuật mới nhằm cải tiến hiệu năng nén cho mã hóa video phân tán và mã hóa video phân tán liên lớp.
- Phương pháp mô phỏng: Các đề xuất cải tiến hiệu năng nén cho mã hóa video phân tán và mã hóa video phân tán liên lớp được thử nghiệm và đánh giá so sánh với các giải pháp trước đó trên các phần mềm tham chiếu.

### **Cấu trúc của luận án**

---

Ngoài phần mở đầu và phần kết luận, kiến nghị, Luận án được chia thành 3 chương với bố cục như sau:

**Chương 1: TỔNG QUAN VỀ MÃ HÓA VIDEO.**

Chương 1 giới thiệu tổng quan về mã hóa video nói chung và tập trung vào mã hóa video phân tán. Trong chương này giới thiệu một số các chuẩn nén video dự đoán, lý thuyết nền tảng của DVC và các bộ mã hóa video phân tán thực tế được nghiên cứu trong cộng đồng nghiên cứu DVC. Phần cuối của chương 1 đề cập đến các vấn đề đang được nghiên cứu trong DVC. Đây chính là các gợi mở cho các nghiên cứu được thực hiện ở các chương sau.

**Chương 2: CẢI TIẾN HIỆU NĂNG NÉN CHO MÃ HÓA VIDEO PHÂN TÁN.**

Chương 2 trình bày các kết quả nghiên cứu mới của Luận án cho mã hóa video phân tán. Phần đầu của chương này khảo sát một số các nghiên cứu có liên quan. Từ đó đề xuất một số phương pháp nhằm nâng cao hiệu năng nén cho mã hóa video phân tán bao gồm phương pháp thay đổi kích thước nhóm ảnh được thực hiện tại phía mã hóa, nhóm các phương pháp được thực hiện tại phía giải mã gồm phương pháp cải tiến chất lượng thông tin phụ trợ và phương pháp cải tiến độ chính xác của mô hình nhiễu tương quan.

Các kết quả nghiên cứu của chương 2 được công bố trong các bài báo số [1], [2], [3] trong danh mục các công trình công bố của tác giả.

**Chương 3: CẢI TIẾN HIỆU NĂNG NÉN CHO MÃ HÓA VIDEO PHÂN TÁN LIÊN LỚP.**

Chương 3 trình bày các kết quả nghiên cứu mới về cải tiến chất lượng thông tin phụ trợ cho mã hóa video phân tán liên lớp bao gồm kỹ thuật tạo thông tin phụ trợ dựa trên kết hợp ảnh và tạo thông tin phụ trợ dựa trên máy vecto hỗ trợ (SVM).

Các kết quả nghiên cứu của chương 3 được công bố trong các bài báo số [4],[5] trong danh mục các công trình công bố của tác giả.



# Mục lục

Mục lục . . . . .	viii
<b>Chương 1. Tổng quan về mã hóa video . . . . .</b>	<b>4</b>
1.1 Mã hóa video dự đoán . . . . .	7
1.1.1 Các tiêu chuẩn mã hóa video trước đây . . . . .	7
1.1.2 Chuẩn mã hóa video H.264/MPEG-4 AVC . . . . .	9
1.1.3 Chuẩn mã hóa video HEVC . . . . .	10
1.2 Mã hóa video phân tán . . . . .	11
1.3 Các kiến trúc mã hóa video phân tán . . . . .	15
1.3.1 Kiến trúc cơ bản của mã hóa video phân tán . . . . .	15
1.3.2 Kiến trúc mã hóa video Stanford . . . . .	16
1.3.3 Kiến trúc mã hóa video phân tán PRISM . . . . .	19
1.3.4 So sánh kiến trúc Stanford và kiến trúc PRISM . . . . .	21
1.3.5 Kiến trúc mã hóa video phân tán DISCOVER . . . . .	23
1.4 Các vấn đề đang được nghiên cứu trong mã hóa video phân tán . . . . .	28
1.4.1 Cải tiến hiệu năng nén . . . . .	28
1.4.2 Cung cấp khả năng chống lỗi . . . . .	32
1.4.3 Cung cấp khả năng liên lớp . . . . .	33
1.4.4 Mã hóa video phân tán đa góc nhìn . . . . .	33
1.5 Các hướng nghiên cứu trong luận án . . . . .	34
1.6 Kết luận chương . . . . .	35
<b>Chương 2. Cải tiến hiệu năng nén cho mã hóa video</b>	
<b>phân tán . . . . .</b>	<b>36</b>
2.1 Giới thiệu chung . . . . .	36
2.1.1 Các nghiên cứu về thông tin phụ trợ . . . . .	38

2.1.2	Các nghiên cứu về mô hình nhiễu tương quan . . . . .	40
2.1.3	Các nghiên cứu về thay đổi kích thước nhóm ảnh . . . . .	43
2.2	Đề xuất 1: Cải tiến hiệu năng nén cho DVC tại phía mã hóa . . . . .	45
2.2.1	Đặt vấn đề . . . . .	45
2.2.2	Phương pháp thay đổi kích thước GOP đề xuất . . . . .	47
2.2.3	Đánh giá hiệu năng của phương pháp đề xuất AGOP-DVC . . . . .	53
2.3	Đề xuất 2: Cải tiến hiệu năng nén cho DVC tại phía giải mã . . . . .	57
2.3.1	Đặt vấn đề . . . . .	57
2.3.2	Kiến trúc bộ mã hóa đề xuất Adv-DVC . . . . .	58
2.3.3	Các mô-đun mã hóa đề xuất mới . . . . .	59
2.3.4	Đánh giá hiệu năng . . . . .	66
2.4	Kết luận chương . . . . .	77

### **Chương 3. Cải tiến hiệu năng nén cho mã hóa video**

	<b>phân tán liên lớp . . . . .</b>	<b>79</b>
3.1	Giới thiệu chung . . . . .	79
3.1.1	Mã hóa video liên lớp . . . . .	81
3.1.2	Mã hóa video phân tán liên lớp . . . . .	85
3.2	Đề xuất 3: Tạo thông tin phụ trợ dựa trên kỹ thuật kết hợp ảnh cho DSVC88 . . . . .	88
3.2.1	Kiến trúc bộ mã hóa DSVC-Fusion . . . . .	88
3.2.2	Phương pháp tạo thông tin phụ trợ dựa trên kỹ thuật kết hợp ảnh . . . . .	92
3.2.3	Đánh giá hiệu năng . . . . .	94
3.3	Đề xuất 4: Tạo thông tin phụ trợ dựa trên kỹ thuật SVM . . . . .	97
3.3.1	Đặt vấn đề . . . . .	97
3.3.2	Kiến trúc bộ mã hóa DSVC-SVM . . . . .	97
3.3.3	Phương pháp tạo thông tin phụ trợ dựa trên kỹ thuật SVM . . . . .	100
3.3.4	Đánh giá hiệu năng . . . . .	109
3.4	Kết luận chương . . . . .	112
	<b>Kết luận . . . . .</b>	<b>116</b>
	<b>Các công trình khoa học . . . . .</b>	<b>118</b>
	<b>Tài liệu tham khảo . . . . .</b>	<b>118</b>

# Danh sách hình vẽ

1.1	Kiến trúc mã hóa video: a. Phương pháp mã hóa video dự đoán; b. Phương pháp mã hóa video phân tán . . . . .	6
1.2	Các chuẩn mã hóa video dự đoán . . . . .	7
1.3	Ví dụ quá trình mã hóa với thông tin phụ trợ . . . . .	12
1.4	Sơ đồ khối cơ bản của mã hóa video phân tán . . . . .	15
1.5	Kiến trúc mã hóa video phân tán Stanford . . . . .	17
1.6	Kiến trúc mã hóa video phân tán PRISM . . . . .	19
1.7	Kiến trúc mã hóa video phân tán DISCOVER . . . . .	23
1.8	Các dải hệ số DCT . . . . .	24
1.9	Tám ma trận lượng tử tương ứng với các điểm RD khác nhau . . . . .	25
1.10	Kiến trúc của kỹ thuật tạo thông tin phụ trợ bằng nội suy khung hình	26
2.1	Sáu khung hình đầu tiên của chuỗi video News . . . . .	46
2.2	(a) Đoạn chuyển động nhanh và (b) Đoạn chuyển động chậm . . . . .	46
2.3	Kiến trúc mã hóa video phân tán AGOP-DVC . . . . .	48
2.4	Một số đặc trưng của chuỗi Suzie . . . . .	50
2.5	Một số đặc trưng của chuỗi Coastguard . . . . .	50
2.6	Các khung hình đầu tiên của các chuỗi video thử nghiệm . . . . .	55
2.7	Kiến trúc mã hóa DVC đề xuất Adv-DVC . . . . .	59
2.8	Quá trình tạo SI . . . . .	60
2.9	Kiến trúc NN-CNM . . . . .	65
2.10	Các khung hình đầu tiên của các chuỗi video thử nghiệm . . . . .	66
2.11	So sánh tham số $\alpha$ của các phương pháp khác nhau . . . . .	70
2.12	So sánh PSNR [dB] của bộ mã hóa Adv-DVC với các bộ mã hóa khác	73
2.13	Hiệu năng RD đối với các chuỗi video . . . . .	75

---

3.1	Khả năng thích ứng của mã hóa video liên lớp . . . . .	81
3.2	Sơ đồ khối đơn giản của bộ mã hóa liên lớp gồm hai lớp . . . . .	83
3.3	Các ví dụ minh họa khả năng liên lớp: liên lớp thời gian, liên lớp không gian và liên lớp chất lượng . . . . .	84
3.4	Mô hình bộ mã hóa video phân tán liên lớp . . . . .	86
3.5	Sơ đồ kiến trúc bộ mã hóa và giải mã video phân tán liên lớp DSVC-Fusion . . . . .	89
3.6	PSNR của các khung hình tái tạo cho chuỗi Basketball Drill và BQMall . . . . .	96
3.7	Sơ đồ kiến trúc bộ mã hóa và bộ giải mã video phân tán liên lớp DSVC-SVM . . . . .	98
3.8	Kiến trúc tạo SI . . . . .	101
3.9	Quá trình huấn luyện trực tuyến SVM đề xuất cho mỗi chuỗi video .	107
3.10	Các khung hình đầu tiên của các chuỗi video thử nghiệm . . . . .	110

# Danh sách bảng

2.1	Điều kiện thử nghiệm . . . . .	54
2.2	Hiệu năng RD của các chuỗi video thử nghiệm . . . . .	56
2.3	Mức tiết kiệm BD-rate [%] . . . . .	57
2.4	So sánh các cấu trúc mạng nơ-ron . . . . .	64
2.5	Điều kiện thử nghiệm . . . . .	67
2.6	PSNR [dB] của phương pháp ReSI so với phương pháp MCTI . . . . .	68
2.7	PSNR [dB] của khung hình giải mã của bộ mã hóa Adv-DVC so với các bộ mã hóa khác . . . . .	72
2.8	So sánh BD-Rate [%] và BD-PSNR [dB] giữa Adv-DVC với HEVC Intra . . . . .	74
2.9	So sánh BD-Rate [%] và BD-PSNR [dB] giữa Adv-DVC với các bộ mã hóa DVC khác . . . . .	76
3.1	Đặc điểm của các chuỗi video thử nghiệm . . . . .	95
3.2	PSNR [dB] của các phương pháp tạo SI . . . . .	95
3.3	Chất lượng của các khung hình giải mã [dB] . . . . .	96
3.4	Độ chính xác của các mô hình huấn luyện . . . . .	108
3.5	Đặc điểm của các chuỗi video thử nghiệm . . . . .	110
3.6	Chất lượng SI tạo bằng kỹ thuật SVM so với các phương pháp khác [dB] . . . . .	111

# Danh mục viết tắt

Từ viết tắt	Nghĩa tiếng Anh	Nghĩa tiếng Việt
AC	Alternating Current	Hệ số xoay chiều
AMV	Average of Motion Vector	Trung bình của các vector chuyển động
AVC	Advanced Video Coding	Mã hóa video tiên tiến
BL	Base Layer	Lớp cơ sở
CBP	Constrained Base Profile	Chế độ mã hóa ràng buộc
CNM	Correlation Noise Model	Mô hình nhiễu tương quan
CRC	Cyclic Redundancy Code	Mã dư thừa vòng
DC	Direct Current	Hệ số một chiều
DCT	Discrete Cosine Transform	Biến đổi Cosin rời rạc
DoH	Difference of Histogram	Sai khác của lược đồ xám
DSC	Distributed Source Coding	Mã hóa nguồn phân tán
DSVC	Distributed Scalable Video Coding	Mã hóa video liên lớp phân tán
DVC	Distributed Video Coding	Mã hóa video phân tán
EL	Enhancement Layer	Lớp tăng cường
FVC	Future Video Coding	Chuẩn mã hóa video tương lai
GOP	Group of Picture	Nhóm ảnh
HDTV	High Definition Television	Truyền hình độ nét cao
HEVC	High Efficiency Video Coding	Chuẩn mã hóa video hiệu năng cao
HVS	Human Vision System	Hệ thống thị giác người
IDCT	Inverse Discrete Cosine Transform	Biến đổi DCT ngược
KF	Key Frame	Khung hình chính

LDPC	Low-Density Parity-Check	Mã kiểm tra chẵn lẻ mật độ thấp
LDPCA	Low-Density Parity-Check Accumulate	Mã kiểm tra chẵn lẻ mật độ thấp tích lũy
LSB	Least Significant Bit	Bit có trọng số nhỏ nhất
MAD	Mean of Absolutely Difference	Trung bình sai khác tuyệt đối
MAE	Mean Absolute Error	Sai số tuyệt đối trung bình
MCTI	Motion Compensated Temporal Interpolation	Nội suy thời gian bù chuyển động
MSB	Maximum Significant Bit	Bit có trọng số lớn nhất
MSE	Mean Square Error	Trung bình bình phương sai số
MV	Motion Vector	Vector chuyển động
NN	Neural Network	Mạng nơ-ron
NMV	Number of Motion Vector	Số lượng các vector chuyển động
PDWZ	Partially Decoded Wyner-Ziv	Khung hình WZ được giải mã một phần
PSNR	Peak Signal-to-Noise Rate	Tỷ lệ tín hiệu trên nhiễu
RD	Rate Distortion	Tốc độ - độ méo tín hiệu
SAD	Sum of Absolutely Difference	Tổng các sai số tuyệt đối
SHVC	Scalable High Efficiency Video Coding	Chuẩn mã hóa HEVC liên lớp
SI	Side Information	Thông tin phụ trợ
SIB	Side Information Block	Khối thông tin phụ trợ
SIF	Side Information Frame	Khung hình thông tin phụ trợ
SR	Search Range	Dải tìm kiếm
SVM	Support Vector Machine	Máy vector hỗ trợ
TDWZ	Transform Domain Wyner-Ziv	Mã hóa Wyner-Ziv miền biến đổi
WVSN	Wireless Video Sensor Network	Mạng cảm biến video không dây
WZF	Wyner-Ziv Frame	Khung hình Wyner-Ziv

Chương 1

# Tổng quan về mã hóa video

Lĩnh vực nén video đã có sự phát triển nhanh chóng trong suốt ba mươi năm qua với nhiều chuẩn mã hóa video khác nhau. Mục tiêu chính của việc chuẩn hóa video là đạt được tốc độ bit thấp cho việc lưu trữ và truyền dữ liệu trong khi vẫn duy trì được độ méo tín hiệu ở mức có thể chấp nhận được.

Sự ra đời của máy tính bán dẫn vào năm 1960 đã đánh dấu một kỷ nguyên mới trong xử lý và truyền thông kỹ thuật số. Vào đầu những năm 1980, nhờ sự phát triển của phần cứng, đặc biệt là sự tiến độ đáng kể của vi mạch với mật độ tích hợp rất cao (VLSI) đã cho ra đời các thiết bị video số. Việc số hóa video giúp cho việc xử lý dễ dàng hơn và thúc đẩy sự phát triển nhanh chóng của truyền hình quảng bá video.

Ngày nay, hình ảnh và video có thể được tìm thấy ở mọi nơi trong cuộc sống hàng ngày của con người. Trung bình có khoảng 30.000 giờ video được tải lên Youtube mỗi giờ trên toàn thế giới và một người phải mất 82 năm để xem toàn bộ lượng video được tải lên Youtube trong một giờ. Ngoài ra, độ phân giải của video cũng tăng lên đáng kể từ  $100 \times 100$  trong những năm 1960 cho đến nay có thể lên đến  $7680 \times 4320$  hay còn gọi là độ phân giải 8K. Tuy nhiên, kích thước của dữ liệu thô có thể rất lớn đến mức cần phải có những tài nguyên khổng lồ để lưu trữ và truyền tải. Ví dụ, dung lượng của một bộ phim màu dài 150 phút

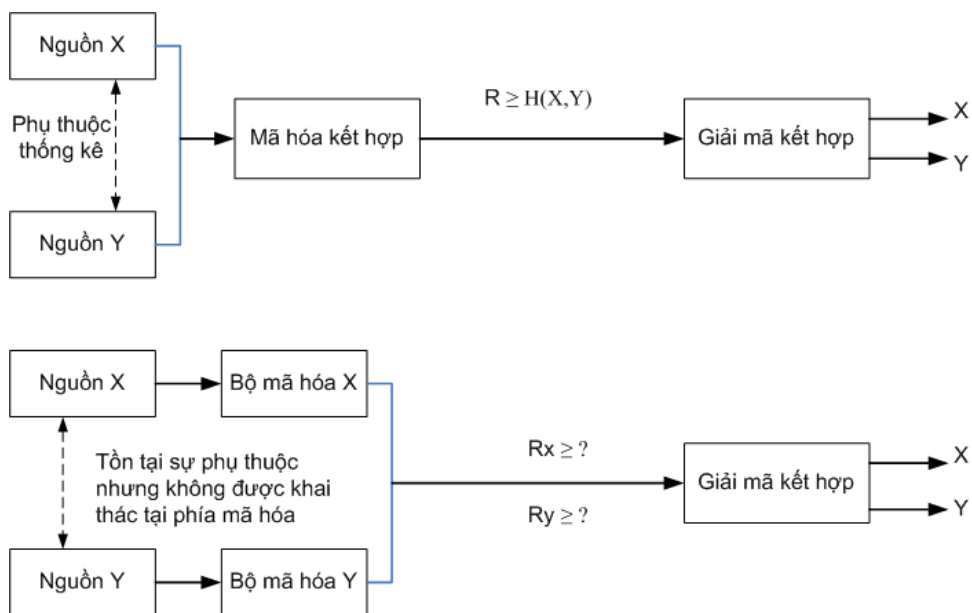


với 30 khung hình/giây và độ phân giải  $720 \times 480$  có thể tới 280 GB khi chưa nén, chưa kể đến trường hợp phim cần truyền qua mạng Internet nơi mà băng thông có thể chưa đến 100 Mbit/s. Do đó, công nghệ nén video là hết sức cần thiết mặc dù sức mạnh của máy tính, bộ lưu trữ và băng thông mạng đã tăng lên đáng kể.

Nén video đề cập đến quá trình trong đó lượng dữ liệu để biểu diễn tín hiệu video đầu vào được giảm xuống một mức độ nhất định để đạt được hiệu quả cao hơn trong việc lưu trữ và truyền tải. Nén có thể đạt được bằng cách giảm các loại dư thừa vốn có trong video bao gồm dư thừa về không gian, thời gian, thống kê và dư thừa do thị giác người.

Phương pháp nén được phân thành hai loại, nén không tổn thất và nén có tổn thất. Phương pháp nén không tổn thất sẽ nén dữ liệu theo cách mà tín hiệu được tái tạo giống hệt như tín hiệu ban đầu. Theo định lý mã hóa nguồn của Shannon, khi tốc độ mã hóa (số bit trung bình trên mỗi ký hiệu) lớn hơn entropy của tín hiệu nguồn, tín hiệu được tái tạo sẽ không bị biến dạng. Tuy nhiên, trong các trường hợp thực tế, biến dạng ở mức độ nào đó lại có thể được chấp nhận để có thể giảm tốc độ bit hơn nữa, do đó nén có tổn hao sẽ hấp dẫn hơn. Ví dụ, đối với giám sát video, nén có tổn thất được ưu tiên vì có thể loại bỏ các chi tiết không cần thiết để tiết kiệm tài nguyên lưu trữ và truyền tải. Vấn đề quan trọng trong việc nén tổn hao là làm cho tốc độ mã hóa càng nhỏ càng tốt trong khi vẫn duy trì cùng chất lượng hình ảnh.

Cụ thể, chúng ta xác định bài toán như sau: Giả sử  $X$  và  $Y$  là hai chuỗi video có phân bố độc lập, giống nhau, tương quan thống kê từ hai bộ mã hóa riêng biệt nhưng đều biết về sự tồn tại của nhau [29]. Bộ giải mã cũng có đầy đủ thông tin về các bộ mã hóa. Bài toán là xác định tốc độ bit mã hóa tối thiểu cho mỗi nguồn video sao cho giải mã kết hợp tại bộ giải mã có thể tái tạo lại mỗi nguồn với đủ độ chính xác. Bài toán này có thể được giải quyết bằng cách sử dụng entropy kết hợp vì các chuỗi video  $X$  và  $Y$  là tương quan thống kê. Có hai phương pháp để tái tạo chúng như trong Hình 1.1:



Hình 1.1: Kiến trúc mã hóa video: a. Phương pháp mã hóa video dự đoán; b. Phương pháp mã hóa video phân tán

*Phương pháp mã hóa video dự đoán (PVC):* mã hóa kết hợp, giải mã kết hợp

Nếu hai chuỗi video phụ thuộc thống kê  $X$  và  $Y$  được mã hóa cùng nhau để khai thác sự phụ thuộc thống kê của chúng, tốc độ mã hóa không tổn thất tối thiểu là entropy kết hợp của chúng  $H(X, Y)$ :

$$R_{(X,Y)} = H(X, Y) \quad (1.1)$$

*Phương pháp mã hóa video phân tán (DVC):* mã hóa độc lập, giải mã kết hợp.

Nếu các chuỗi video  $X$  và  $Y$  được mã hóa độc lập, tốc độ mã hóa tương ứng của chúng là:

$$R_X \geq H(X); R_Y \geq H(Y) \quad (1.2)$$

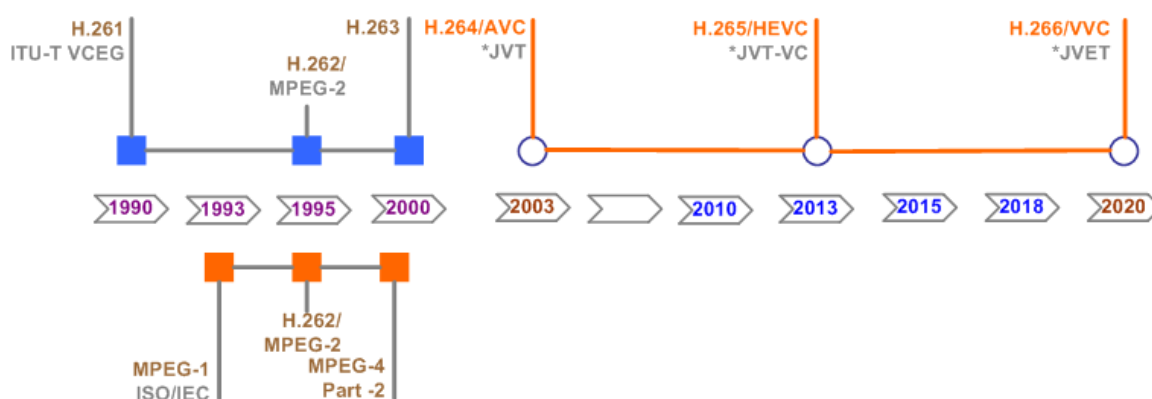
ở đó  $H(X)$  và  $H(Y)$  là các entropy của  $X$  và  $Y$  tương ứng. Do đó tốc độ mã hóa yêu cầu là:

$$R_X + R_Y \geq R_{(X,Y)} \quad (1.3)$$

Nguyên lý cơ bản và lịch sử phát triển của các phương pháp mã hóa này sẽ được trình bày chi tiết ở các phần tiếp theo.

## 1.1 Mã hóa video dự đoán

Mã hóa video dự đoán đã có bước phát triển mạnh mẽ trong vài năm trở lại đây, nhiều thuật toán mã hóa khác nhau được đề xuất và cải tiến. Sự phát triển của các chuẩn nén quốc tế đã thúc đẩy sự cải tiến của các ứng dụng mã hóa video. Các chuẩn nén video dự đoán bao gồm MPEG-1 [63], MPEG-2 [12], MPEG-4 [47], H.261 [33], H.263 [19], H.264 / MPEG-4 AVC [65] và tiêu chuẩn mã hóa video hiệu suất cao HEVC [54]. Hình 1.2 tóm tắt sự phát triển của các tiêu chuẩn mã hóa video của Liên minh Viễn thông Quốc tế (ITU-T) và Tổ chức Tiêu chuẩn quốc tế (ISO) và sự hợp tác của hai tổ chức này từ năm 1984.



Hình 1.2: Các chuẩn mã hóa video dự đoán

### 1.1.1 Các tiêu chuẩn mã hóa video trước đây

Trước khi H.264 / MPEG-4 AVC và HEVC được phát triển, các tiêu chuẩn nén video do nhóm các chuyên gia về mã hóa video (VCEG) của ITU-T và nhóm các chuyên gia về hình ảnh động (MPEG) của ISO/IEC sản xuất bao gồm H.261, MPEG-1, H.262 / MPEG-2, H.263 và MPEG-4. H.261, một trong những tiêu chuẩn H.26x đầu tiên trong lĩnh vực của VCEG, khởi đầu quá trình

tiêu chuẩn hóa cho mã hóa video kỹ thuật số thực tế. Sơ đồ mã hóa tuân theo cách tiếp cận mã hóa video kết hợp dựa trên khối. Khái niệm macroblock (MB) đóng vai trò là đơn vị xử lý cơ bản lần đầu tiên xuất hiện trong H.261. Dự đoán liên ảnh làm giảm dư thừa theo thời gian với các vec-tơ chuyển động có giá trị nguyên. Mặc dù hiện nay đã có những cải tiến đáng kể về khả năng nén, H.261 vẫn đóng vai trò là một cột mốc quan trọng trong lĩnh vực mã hóa video.

Năm 1992, MPEG-1 được MPEG thiết kế cho ứng dụng CD-ROM. MPEG-1 dựa trên mã hóa biến đổi bù chuyển động, trong đó các công cụ mã hóa chính bao gồm chuyển đổi màu sang YUV và lấy mẫu xuống trong miền UV, khử tương quan không gian với biến đổi cosine rời rạc (DCT)  $8 \times 8$  và giảm dư thừa không gian với dự đoán liên ảnh song hướng. MPEG-1 từng được sử dụng phổ biến bởi nhiều dịch vụ truyền hình cáp / vệ tinh kỹ thuật số trước khi bị thay thế bởi H.262 / MPEG-2, được thiết kế bởi MPEG và VCEG để đáp ứng nhu cầu truyền hình giải trí cho các phương tiện truyền dẫn như vệ tinh và CATV, và phương tiện lưu trữ kỹ thuật số như DVD. Để nén hiệu quả video kỹ thuật số xen kẽ ở chất lượng phát sóng, nhiều tính năng mới đã được phát triển. Các công cụ liên lớp được coi là các công cụ chức năng lần đầu tiên được định nghĩa. MPEG-2 cung cấp sự hỗ trợ đối với phát video kỹ thuật số chất lượng truyền hình và cũng hỗ trợ định dạng phim và video khác trên DVD và các đĩa tương tự.

Năm 1995, VCEG đã phát triển H.263 cho hội nghị truyền hình ở tốc độ bit thấp đối với kịch bản truyền thông không dây di động. Cho đến nay, H.263 vẫn đóng vai trò chủ đạo trong bộ mã hóa hội nghị truyền hình và điện thoại di động. Cấu hình cơ bản dựa trên H.261 áp dụng kết hợp dự đoán liên ảnh sử dụng bù chuyển động  $16 \times 16 / 8 \times 8$  với độ phân giải nửa pixel và mã hóa biến đổi với biến đổi cosine rời rạc (DCT)  $8 \times 8$ . H.263 vượt trội so với các tiêu chuẩn mã hóa video trước đây tại các tốc độ bit khác nhau.

Được bắt đầu vào năm 1998, MPEG-4 được tạo ra để xử lý nội dung đa phương tiện theo hướng dựa trên đối tượng, tương tác và phi tuyến tính. MPEG-

4 bao gồm các cấu hình và cấp độ mới với nhiều công cụ mã hóa mới lạ như đồ họa tương tác, mã hóa đối tượng và hình dạng, mã hóa hình ảnh tĩnh dựa trên biến đổi sóng con, mô hình hóa khuôn mặt, mã hóa video liên lớp và đồ họa 3D. MPEG-4 cung cấp các chức năng như: cải thiện hiệu quả mã hóa so với MPEG-2, khả năng mã hóa dữ liệu đa phương tiện hỗn hợp (video, âm thanh, thoại), khả năng phục hồi lỗi để cho phép truyền mạnh mẽ và khả năng tương tác với cảnh nghe nhìn được tạo ra tại bộ thu.

### 1.1.2 Chuẩn mã hóa video H.264/MPEG-4 AVC

Năm 2003, phiên bản đầu tiên của H.264 / MPEG-4 Part 10 hay mã hóa video cải tiến (AVC) đã được hoàn thiện và hiện là một trong những tiêu chuẩn mã hóa phổ biến nhất để ghi, nén và phân phối video độ nét cao. Chuẩn này được phát triển bởi ITU-T VCEG cùng với ISO / IEC MPEG hay còn được biết đến với tên nhóm chung JVT. Tiêu chuẩn này nhằm mục đích nén video với một nửa tốc độ bit so với của các tiêu chuẩn video trước đó (MPEG-2, H.263 và MPEG-4) trong khi vẫn giữ được chất lượng tốt. H.264 MPEG-4 AVC bao gồm một số tính năng mới để nén video hiệu quả hơn nhiều và cung cấp tính linh hoạt hơn cho các ứng dụng trong các môi trường mạng khác nhau. Cụ thể, các tính năng bao gồm: bù chuyển động kích thước khối có thể thay đổi với kích thước khối lớn như  $16 \times 16$  và nhỏ đến  $4 \times 4$  cho phép phân đoạn chính xác các vùng chuyển động, độ chính xác  $1/4$  pixel để bù chuyển động và điều này cho phép mô tả chính xác sự dịch chuyển của các vùng chuyển động, một bộ lọc giảm nhiễu khối trong vòng lặp giúp loại bỏ các nhiễu khối, mã hóa số học nhị phân thích ứng theo ngữ cảnh (CABAC), mã hóa độ dài thay đổi thích ứng theo ngữ cảnh (CAVLC).

Một trong những tính năng chính được thêm vào tiêu chuẩn là mã hóa video liên lớp (SVC). Nó cho phép mã hóa các dòng bit chất lượng cao có chứa các dòng bit con. Bằng cách loại bỏ bớt các gói từ video có độ phân giải cao, một luồng bit video con được tạo ra và làm giảm băng thông cần thiết, dẫn đến giảm

độ phân giải theo thời gian và không gian. Một tính năng chính khác là mã hóa video đa góc nhìn (MVC). Nó cho phép mã hóa dòng bit thể hiện nhiều hơn một góc nhìn của một cảnh video, ví dụ: mã hóa video 3D lập thể.

H.264 hoạt động tốt hơn hẳn so với video MPEG-2 với cùng chất lượng tại một nửa tốc độ bit hoặc thấp hơn, đặc biệt trong các tình huống tốc độ bit cao và độ phân giải cao. Vì vậy, nó trở thành tiêu chuẩn hàng đầu và đã được áp dụng trong nhiều ứng dụng mã hóa video như iPod cũng như trong các tiêu chuẩn phát sóng truyền hình như DVB-H và DMB.

### 1.1.3 Chuẩn mã hóa video HEVC

Vào đầu năm 2013, phiên bản đầu tiên của tiêu chuẩn mã hóa video hiệu suất cao HEVC được phê chuẩn là sự kế thừa của chuẩn H.264 / MPEG-4 AVC. HEVC được tạo ra để cung cấp khả năng nén cao hơn để đáp ứng nhu cầu về truyền liên tục video chất lượng cao và cũng tập trung vào nhu cầu ngày càng tăng của xử lý song song. Bộ mã hóa HEVC vẫn dựa trên phương pháp mã hóa biến đổi và dự đoán kết hợp, với dự đoán bù chuyển động, dự đoán nội ảnh, lọc trong vòng, biến đổi 2D và mã hóa entropy thích ứng.

So với các tiêu chuẩn nén video trước đây, kích thước cấu trúc đơn vị mã hóa lớn hơn với phân vùng con phức tạp hơn được giới thiệu trong HEVC. Cấu trúc đơn vị mã hóa (CU) thay thế cấu trúc macroblock và mỗi CU chứa một hoặc một vài đơn vị dự đoán có kích thước khối thay đổi (PU) và (các) đơn vị chuyển đổi (TU). Dự đoán vectơ chuyển động cải tiến (AMVP) được sử dụng để phát hiện hầu hết các ứng cử viên có khả năng nhất từ các PU lân cận và hình ảnh tham khảo. Các véc-tơ chuyển động có độ chính xác 1/4 pixel được sử dụng để nội suy các vị trí phân số. Một bộ lọc giảm nhiễu khối trong vòng lặp với xử lý song song được sử dụng. Độ lệch thích ứng mẫu (SAO) cũng được sử dụng trong vòng lặp dự đoán liên ảnh để tái tạo biên độ tín hiệu ban đầu tốt hơn. HEVC có tỷ lệ nén dữ liệu gấp 2 lần so với H.264 / MPEG-4 AVC ở cùng mức chất lượng video, trong khi đó, HEVC còn hỗ trợ độ nét cực cao (UHD 8K) và

độ phân giải lên đến  $8192 \times 4320$ .

Các chuẩn mã hóa video dự đoán trên đây có hiệu năng nén cao, phù hợp với rất nhiều ứng dụng. Tuy nhiên cũng có thể thấy để đạt được hiệu năng nén như vậy, bộ mã hóa sẽ phải có độ phức tạp rất lớn vì đây là nơi phải thực hiện nhiệm vụ quan trọng nhất là ước lượng và bù chuyển động. Cách thiết kế như vậy sẽ trở nên không phù hợp cho các ứng dụng mới đường lên có đặc điểm là nhiều bộ mã hóa tương ứng với một vài bộ giải mã. Trong các ứng dụng mới này, bộ mã hóa nhẹ và đơn giản là một yêu cầu quan trọng. Vì vậy, giải pháp tiềm năng cho các ứng dụng này là cách tiếp cận mã hóa video thứ hai: mã hóa video phân tán. Phần tiếp theo sẽ trình bày chi tiết về nguyên lý cũng như các vấn đề đang được nghiên cứu đối với mã hóa video phân tán.

## 1.2 Mã hóa video phân tán

Mã hóa nguồn phân tán (DSC) là cách xử lý nhiều nguồn thông tin có tương quan với nhau trong bối cảnh phân tán. Mã hóa video phân tán là một trường hợp đặc biệt của mã hóa nguồn phân tán đã thu hút sự quan tâm của cộng đồng nghiên cứu trong nhiều năm qua. Trong phần này giới thiệu hai định lý trong lý thuyết thông tin là cơ sở của mã hóa nguồn phân tán và cả mã hóa video phân tán là định lý Slepian-Wolf [53] và định lý Wyner-Ziv [66]. Định lý Slepian-Wolf phát biểu rằng, khi mã hóa không tổn thất cho các nguồn có tương quan với nhau thì tốc độ tối ưu đạt được bằng mã hóa kết hợp và giải mã kết hợp cũng có thể đạt được bằng cách mã hóa phân tán và giải mã kết hợp. Định lý Wyner-Ziv mở rộng kết quả này cho việc mã hóa có tổn thất đối với các tập dữ liệu có tương quan khi mã hóa phân tán và giải mã kết hợp được thực hiện mà chỉ cần sử dụng mối tương quan giữa các nguồn ở phía bộ giải mã.

Như đã đề cập trong Hình 1.1, đối với phương pháp mã hóa phân tán, một câu hỏi đặt ra là liệu có thể tái tạo lại chuỗi video với xác suất lỗi nhỏ tùy ý với tốc độ mã hóa nhỏ hơn các đại lượng entropy  $H(X)$  và  $H(Y)$  hay không. Mã

hóa nguồn phân tán chính là câu trả lời cho bài toán này.

### Định lý Slepian-Wolf đối với nén không tổn thất:

Với hai chuỗi ngẫu nhiên  $X$  và  $Y$  phụ thuộc thống kê, độc lập và có phân bố giống nhau, tốc độ tối thiểu để mã hóa độc lập hai nguồn tương quan bằng tốc độ tối thiểu khi mã hóa kết hợp với xác suất lỗi nhỏ tùy ý.

$$R_X + R_Y \geq H(X, Y) \quad (1.4)$$

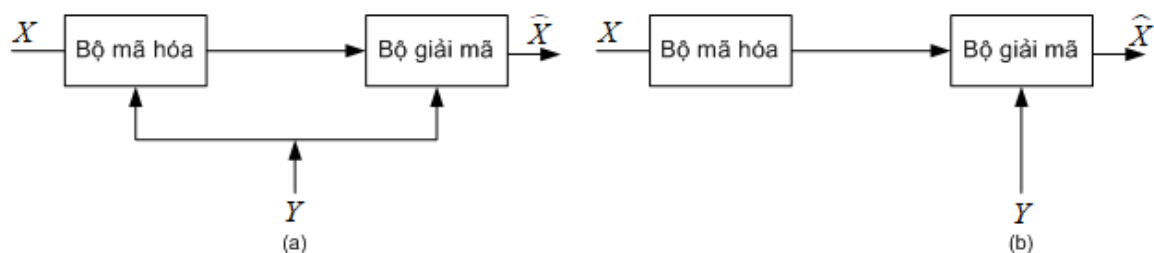
### Định lý Wyner-Ziv đối với nén có tổn thất:

Wyner-Ziv đề xuất mở rộng định lý Slepian-Wolf bằng cách định nghĩa một kịch bản tương tự về việc mã hóa độc lập nhưng trong ngữ cảnh nén có tổn thất và nguồn  $Y$  (thông tin phụ trợ) chỉ sẵn có ở bộ giải mã. Định lý phát biểu rằng khi thực hiện mã hóa độc lập  $X$  với thông tin phụ trợ  $Y$  với một số điều kiện:  $X$  và  $Y$  là các nguồn Gauss kết hợp, không nhớ và có xét đến giá trị sai khác bình phương trung bình (MSE) thì sẽ không có tổn thất về hiệu suất mã hóa so với trường hợp mã hóa kết hợp.

Định lý Wyner-Ziv còn được gọi là định lý mã hóa video phân tán Wyner-Ziv.

### Ví dụ minh họa quá trình mã hóa với thông tin phụ trợ:

Để hiểu hơn về quá trình mã hóa với thông tin phụ trợ có thể xem xét ví dụ dưới đây [22]. Cho  $X$  và  $Y$  là dữ liệu nhị phân chiều dài 3 bit nhận một trong



Hình 1.3: Ví dụ quá trình mã hóa với thông tin phụ trợ

8 trường hợp đồng xác suất, riêng biệt nhau và có tương quan với nhau. Tương quan giữa chúng phải thỏa mãn khoảng cách Hamming giữa  $X$  và  $Y$  lớn nhất bằng 1.  $Y$  sẵn có ở bộ giải mã sử dụng 3 bit. Trường hợp (a) cả bộ mã hóa và



bộ giải mã sử dụng thông tin phụ trợ  $Y$  có tương quan với  $X$ . Khi đó có thể mã hóa  $X$  bằng 2 bit. Trường hợp (b), chỉ bộ giải mã truy nhập được  $Y$ . Tuy nhiên trong trường hợp này,  $X$  cũng có thể được mã hóa bằng 2 bit. Cụ thể, với  $X$  cho trước (ví dụ,  $[010]$ ,  $X$  hoặc là giống  $Y$  ( $[010]$  hoặc ( $[110]$ ), ( $[000]$ ), ( $[011]$ ). Mục tiêu là mã hóa  $X$  một cách hiệu quả trong hai kịch bản được mô tả trong Hình 1.3 sao cho nó có thể tái tạo hoàn hảo ở bộ giải mã.

**Kịch bản 1:** Trong kịch bản đầu tiên (xem Hình 1.3 (a)),  $Y$  xuất hiện ở cả hai phía mã hóa và giải mã ( $Y$  có thể sẵn có ở phía giải mã sử dụng 3 bit). Ở đây  $X$  có thể được dự đoán từ  $Y$ . Dư thừa  $X \oplus Y$  hoặc mẫu sai khác của  $X$  tương ứng với  $Y$  nhận 4 giá trị riêng biệt và do đó có thể được mã hóa bằng 2 bit. Đây là tỉ lệ ít nhất có thể (tốt nhất) để mã hóa  $X$ . Bộ giải mã có thể kết hợp mẫu sai khác với  $Y$  để giải mã ra  $X$ .  $X$  tương tự như khối video hiện thời cần mã hóa,  $Y$  tương tự như dự đoán từ bộ nhớ các khung hình, tương quan giữa  $X$  và  $Y$  tương tự như tương quan thời gian giữa các khung hình video liên tiếp, và vì vậy chế độ mã hóa này tương tự như mã hóa dự đoán.

**Kịch bản 2:** Trong kịch bản thứ 2 (xem Hình 1.3 (b)),  $Y$  sẵn có ở phía giải mã sử dụng 3 bit nhưng để mã hóa cho  $X$  không có sự xuất hiện của  $Y$ . Tuy nhiên, cấu trúc tương quan giữa  $X$  và  $Y$  đã biết và bộ giải mã có khả năng truy nhập đến  $Y$ . Điều tốt nhất có thể được làm trong trường hợp này là gì? Vì kịch bản này tồi hơn so với kịch bản đầu tiên, hiệu năng của nó bị giới hạn so với kịch bản thứ nhất. Tuy nhiên, câu trả lời đáng ngạc nhiên là thậm chí trong kịch bản dường như tồi hơn này, chúng ta có thể đạt được cùng hiệu năng như trong kịch bản đầu tiên. Đó là,  $X$  cũng có thể được mã hóa bởi 2 bit.

Điều này có thể làm được sử dụng cách tiếp cận sau. Không gian các từ mã của  $X$  được chia thành 4 tập con, mỗi tập chứa 2 từ mã có khoảng cách Hamming bằng 3, và gọi là Coset1 ( $[0\ 0\ 0]$  và  $[1\ 1\ 1]$ ), Coset 2 ( $[0\ 0\ 1]$  và  $[1\ 1\ 0]$ ), Coset 3 ( $[0\ 1\ 0]$  và  $[1\ 0\ 1]$ ) và Coset 4 ( $[1\ 0\ 0]$  và  $[0\ 1\ 1]$ ). Bộ mã hóa cho  $X$  xác định tập chứa từ mã đối với  $X$  và gửi chỉ số của tập đó thay vì từ mã riêng biệt. Bởi vì có 4 tập nên chúng được đánh số bằng 2 bit. Bộ giải mã, khi nhận

được chỉ số của coset, sử dụng  $Y$  để xác định  $X$  đúng từ tập đó bằng cách chọn từ mã gần với  $Y$  nhất là đáp án cuối cùng.

Chú ý rằng khoảng cách giữa  $X$  và  $Y$  lớn nhất bằng 1, và khoảng cách giữa 2 từ mã trong mỗi tập đều bằng 3. Vì vậy, giải mã có thể được thực hiện một cách hoàn hảo (ví dụ, nếu  $Y$  là [001] và  $X$  là [011], thì bộ mã hóa gửi chỉ số của Coset 4. Bộ giải mã nhận được chỉ số này, tính khoảng cách giữa [001] và [100] bằng 2, giữa [001] và [011] bằng 1. Bởi vì nó biết rằng khoảng cách giữa  $X$  và  $Y$  lớn nhất bằng 1, nên [011] được giải mã là từ mã phát đi. Chế độ mã hóa mà bộ giải mã có khả năng truy nhập tới thông tin phụ trợ tương quan được gọi là chế độ mã hóa thông tin phụ trợ (SI). Các định lý Slepian-Wolf và Wyner-Ziv đã chứng minh rằng về mặt lý thuyết, hiệu năng của hệ thống mã hóa phân tán có thể đạt được giống như hiệu năng của hệ thống dựa trên mã hóa dự đoán. Nói ngắn gọn, tương quan giữa  $X$  và  $Y$  có thể giúp giảm tốc độ truyền dẫn.

Trong thực tế, việc phân chia không gian từ mã nguồn và gán nhãn cho các coset kết quả (chỉ số được gán cho các cosets còn được gọi là syndromes) có thể được làm theo một cách rất hiệu quả về mặt tính toán thông qua khung làm việc của các mã coset. Vì vậy, bộ mã hóa trong hệ thống mã hóa thông tin phụ trợ có độ phức tạp mã hóa thấp.

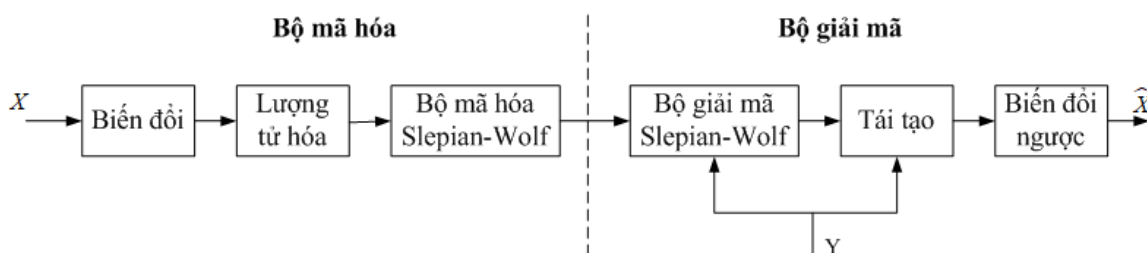
Chú ý rằng việc phân chia  $X$  này cũng mang tính phổ quát. Nghĩa là, cùng một cách phân chia  $X$  cho tất cả các  $Y$  không quan tâm đến giá trị của  $Y$ , chỉ cần cả  $X$  và  $Y$  thỏa mãn cấu trúc tương quan. Ví dụ, nếu  $X$  là [010] thì cùng một cách mã hóa  $X$  (chỉ số của coset 3) sẽ được áp dụng cho tất cả các trường hợp của  $Y$  nghĩa là, [010], [110], [000], [011]. Vì vậy nếu  $Y$  nhận giá trị [010] và trong quá trình truyền dẫn tới bộ giải mã, nó trở thành [110] để bộ giải mã có một phiên bản lỗi của  $Y$ ,  $X$  vẫn có thể được khôi phục lại một cách chính xác. Đó là bởi vì phiên bản lỗi của  $Y$  cũng thỏa mãn cấu trúc tương quan. Vì vậy, không giống với mã hóa dự đoán, sẽ không có sự phụ thuộc giữa việc mã hóa  $X$  và giá trị của thông tin tương quan  $Y$ , do đó cung cấp khả năng chống lỗi cao.

## 1.3 Các kiến trúc mã hóa video phân tán

Dựa trên các định lý Slepian-Wolf và định lý Wyner-Ziv, một số các kiến trúc mã hóa video phân tán được đề xuất. Phần tiếp theo dưới đây giới thiệu kiến trúc cơ bản của mã hóa video phân tán và một số kiến trúc mã hóa video thực tế.

### 1.3.1 Kiến trúc cơ bản của mã hóa video phân tán

Hình 1.4 mô tả kiến trúc cơ bản của mã hóa video phân tán trong miền biến đổi. Có thể tóm tắt quá trình mã hóa và giải mã video phân tán như sau:



Hình 1.4: Sơ đồ khối cơ bản của mã hóa video phân tán

- Các khung hình của chuỗi video  $X$  đầu tiên được biến đổi từ miền không gian sang miền khác, ví dụ miền tần số, cho phép biểu diễn nguồn  $X$  nhỏ gọn hơn và do đó sẽ làm giảm tốc độ bit.
- Các hệ số biến đổi tương ứng sau đó được lượng tử hóa để tạo ra luồng ký tự lượng tử hóa nhằm khai thác các hạn chế của hệ thống thị giác người.
- Quá trình mã hóa video phân tán cơ bản kết thúc bằng mô-đun mã hóa Slepian-Wolf. Như đã được đề cập ở phần trước, bộ giải mã video phân tán có sẵn một ước lượng hay còn gọi là phiên bản “lỗi” của chuỗi  $X$ , được mô tả trong Hình 1.4 là  $Y$ ; nếu các thông tin dư thừa về chuỗi  $X$  được truyền đến bộ giải mã, các “lỗi” trong chuỗi  $Y$  có thể được sửa. Do đó, mô-đun

bộ mã hóa Slepian-Wolf trong Hình 1.4 đóng vai trò là bộ tạo ra thông tin dư thừa; nói cách khác, bằng cách sử dụng các loại mã kênh, bộ mã hóa Slepian-Wolf tạo ra thông tin dư thừa (có nghĩa là các bit chẵn lẻ) từ luồng ký tự đã lượng tử hóa.

- Tại bộ giải mã, luồng ký hiệu đã lượng tử hóa được giải mã thông qua việc giải mã kênh - nguồn với sự trợ giúp của chuỗi thông tin phụ trợ  $Y$ . Chuỗi  $Y$  (thông thường là ước lượng về chuỗi  $X$  tại bộ giải mã) được coi là sẵn có tại bộ giải mã.
- Luồng ký tự lượng tử đã được giải mã được sử dụng cùng với thông tin phụ trợ  $Y$  trong mô đun tái tạo để ước lượng các hệ số biến đổi. Bước cuối cùng để tái tạo lại chuỗi  $X$  là chuỗi  $\hat{X}$  bằng cách thực hiện biến đổi ngược.

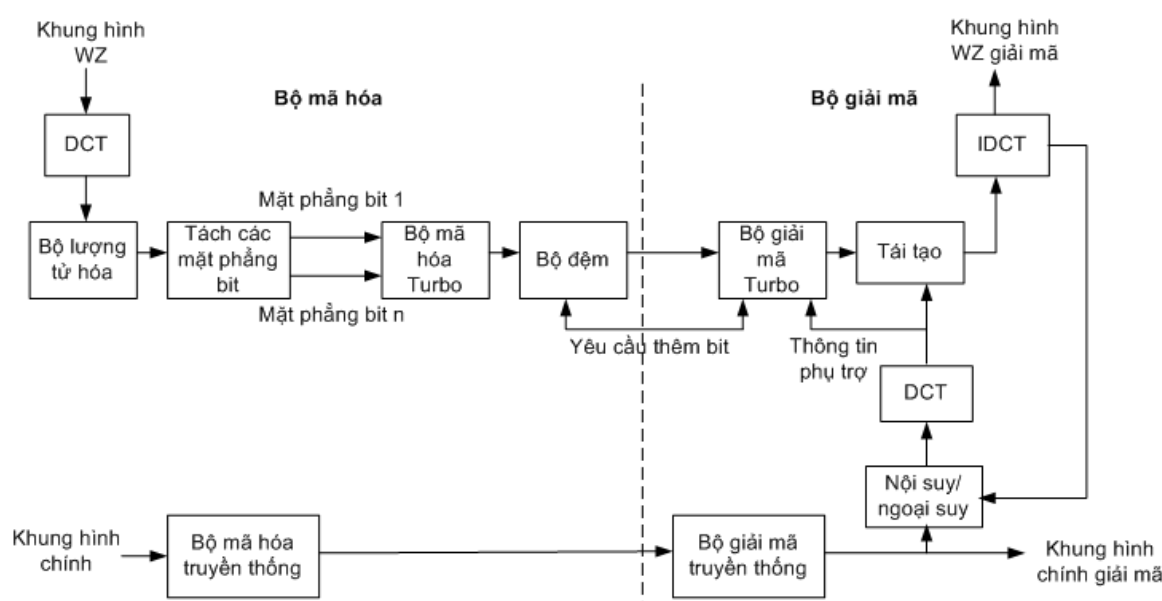
Dựa trên kiến trúc mã hóa video cơ bản, việc thiết kế các hệ thống DVC thực tế đã được đề xuất trong những năm gần đây [44][22][9]. Đóng góp quan trọng nhất là của hai nhóm nghiên cứu chính: nhóm của giáo sư Bernd Girod tại Đại học Stanford hay còn gọi là kiến trúc Stanford [2][22] mà sau này được cải tiến thành kiến trúc DISCOVER [9] và nhóm của giáo sư Kannan Ramchandran tại Berkeley (Đại học California) còn được biết đến là kiến trúc PRISM [44]. Phần tiếp theo sẽ giới thiệu các kiến trúc này.

### 1.3.2 Kiến trúc mã hóa video Stanford

Kiến trúc mã hóa video phân tán Stanford [2] lần đầu được đề xuất vào năm 2002 cho miền pixel, và sau đó được mở rộng sang miền biến đổi [22]. Hình 1.5 mô tả kiến trúc video WZ Stanford miền biến đổi.

#### Quá trình mã hóa

- *Chia tách khung hình:* Chuỗi video được chia thành các khung hình Wyner-Ziv (WZF) và khung hình chính (KF). Các KF được mã hóa intra bởi các



Hình 1.5: Kiến trúc mã hóa video phân tán Stanford

chuẩn mã hóa video truyền thống còn các WZF được mã hóa theo nguyên tắc mã hóa video phân tán.

- *Mã hóa biến đổi:* Các WZF được mã hóa biến đổi, thông thường là biến đổi cosine rời rạc (DCT) dựa trên khối. Các hệ số DCT của toàn bộ WZF được nhóm lại với nhau, theo vị trí của mỗi hệ số DCT trong một khối, tạo thành các dải hệ số DCT.
- *Lượng tử hóa:* Mỗi dải DCT được lượng tử hóa đồng nhất với số lượng các mức tùy thuộc vào chất lượng yêu cầu. Đối với một dải DCT cho trước, các bit của các ký tự đã lượng tử hóa được nhóm với nhau tạo thành các mặt phẳng bit. Các mặt phẳng bit này sau đó được mã hóa turbo độc lập.
- *Mã hóa turbo:* Mã hóa turbo cho mỗi dải DCT bắt đầu với mặt phẳng bit quan trọng nhất (MSB). Các thông tin chẵn lẻ được tạo ra cho mỗi mặt phẳng bit sau đó được lưu trong bộ đệm và được gửi theo từng cụm tùy theo yêu cầu từ phía bộ giải mã thông qua kênh phản hồi.

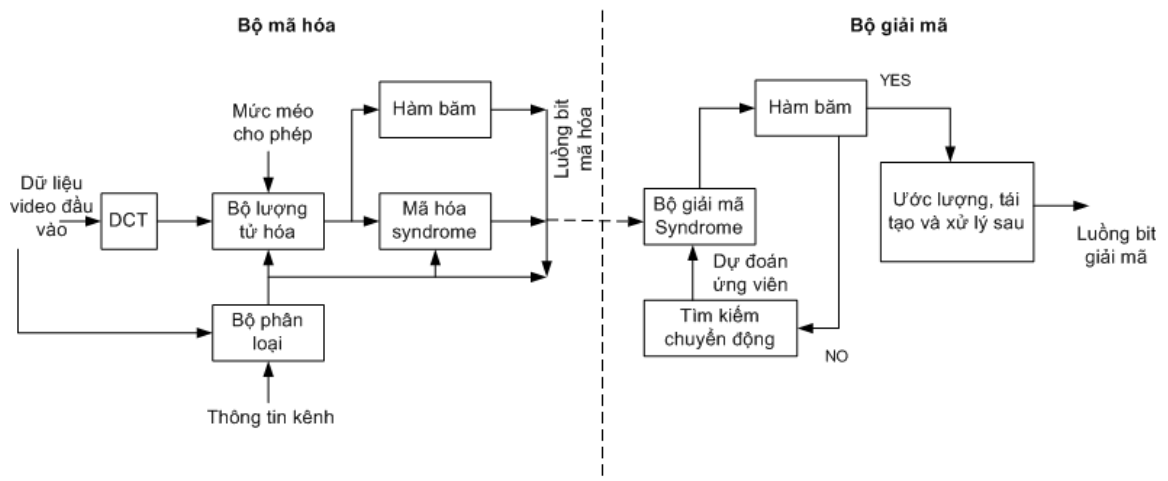
## Quá trình giải mã

- *Tạo thông tin phụ trợ:* Bộ giải mã tạo ra thông tin phụ trợ cho mỗi WZF bằng cách nội suy khung hình bù chuyển động (hoặc ngoại suy) sử dụng các khung hình đã giải mã gần nhất về mặt thời gian. Khung hình thông tin phụ trợ (SIF) tương ứng với mỗi WZF được coi như một ước lượng (hay phiên bản nhiễu) của khung hình WZ gốc. Chất lượng ước lượng SIF càng tốt thì số lượng các “lỗi” mà bộ giải mã turbo phải sửa càng ít và vì vậy, số lượng các bit chẵn lẻ (hay tốc độ bit) càng nhỏ.
- *Mô hình hóa nhiễu tương quan:* Nhiễu tương quan (CNM) là sự sai khác giữa khung hình WZ gốc tại phía mã hóa và khung hình thông tin phụ trợ tương ứng tại phía giải mã. Thống kê dư thừa giữa các hệ số DCT tương ứng của WZF và SIF được giả định mô hình hóa bởi phân bố Laplace mà ở đó các tham số của phân bố này được ước lượng ban đầu bằng cách huấn luyện ngoại tuyến. Đây là cách tiếp cận không thực tế vì giả định sẵn có hoặc là dữ liệu gốc ở phía giải mã hoặc thông tin phụ trợ ở phía mã hóa. Tuy nhiên gần đây đã có nhiều giải pháp được phát triển để khắc phục giả định không thực tế này.
- *Giải mã Turbo:* Khi biết các hệ số DCT của SIF và thống kê dư thừa đối với một dải hệ số DCT nào đó thì mỗi mặt phẳng bit của dải DCT sẽ được giải mã turbo (bắt đầu từ vị trí bit MSB). Bộ giải mã turbo nhận từ bộ mã hóa các cụm liên tiếp các bit chẵn lẻ, theo yêu cầu được tạo ra qua kênh phản hồi. Để quyết định có cần gửi thêm các bit chẵn lẻ để giải mã thành công cho một mặt phẳng bit nào đó, bộ giải mã sử dụng tiêu chuẩn dừng yêu cầu. Sau khi giải mã turbo thành công cho mặt phẳng bit MSB của một dải DCT, bộ giải mã turbo tiếp tục theo cách tương tự với các mặt phẳng bit còn lại của cùng dải DCT đó. Khi tất cả các mặt phẳng bit của một dải DCT được giải mã turbo thành công, bộ giải mã turbo bắt đầu giải mã cho dải DCT tiếp theo.

- *Tái tạo*: Sau khi giải mã turbo, tất cả các mặt phẳng bit của mỗi dải DCT được nhóm lại với nhau để tạo thành một luồng ký tự lượng tử hóa và được giải mã cho mỗi dải đó. Khi có được tất cả các ký tự đã lượng tử hóa, có thể tái tạo lại tất cả các hệ số DCT đã truyền với sự giúp đỡ của SIF tương ứng. Các dải hệ số DCT mà không có bit WZ (bit chẵn lẻ) nào được truyền sẽ được thay thế bằng dải DCT tương ứng của SIF.
- *Biến đổi ngược*: Sau khi tái tạo lại tất cả các dải DCT, thực hiện phép biến đổi cosine rời rạc ngược (IDCT) để thu được khung hình WZ giải mã.
- *Sắp xếp lại khung hình*: Cuối cùng, để thu được chuỗi video giải mã, các KF và WZF được kết hợp với nhau để tạo ra chuỗi video ban đầu.

### 1.3.3 Kiến trúc mã hóa video phân tán PRISM

Kiến trúc bộ mã hóa video phân tán PRISM [44] được mô tả trong Hình 1.6, nhấn mạnh đến các khối mã hóa chính. Trước tiên, các khung hình được chia thành các khối  $8 \times 8$  không chồng lấn nhau. Mỗi khối  $8 \times 8$  sau đó được mã hóa độc lập theo nguyên tắc phân tán.



Hình 1.6: Kiến trúc mã hóa video phân tán PRISM

## Quá trình mã hóa

- *Biến đổi DCT*: Quá trình này tạo ra ma trận các hệ số biến đổi đối với mỗi khối (được quét zig-zag) để tạo ra luồng mã hóa.
- *Lượng tử hóa*: Lượng tử hóa vô hướng được áp dụng lên các hệ số DCT thu được từ khối mã hóa trước. Bước lượng tử cho mỗi hệ số được lựa chọn dựa trên chất lượng tái tạo mong muốn trong khi vẫn xem xét đến độ nhạy khác nhau của hệ thống thị giác người (HVS) đối với các tần số không gian khác nhau.
- *Bộ phân loại*: Trước khi mã hóa, mỗi khối  $8 \times 8$  được phân loại thành một trong số các lớp xác định trước tùy thuộc vào mức tương quan giữa khối hiện thời và khối dự đoán trong khung hình tham chiếu. Quá trình này sẽ xác định lớp mã hóa nào sẽ được sử dụng cho mỗi khối: Intra, SKIP hay Syndrome. Chế độ Intra có nghĩa là khối đó sẽ được mã hóa nội ảnh, chế độ SKIP có nghĩa là khối đó sẽ không được mã hóa mà sẽ sử dụng nguyên vẹn khối thông tin phụ trợ (SIB) tương ứng ở bộ giải mã. Nếu chế độ mã hóa Syndrome được chọn, các bit của mỗi hệ số sẽ được gán vào một số loại, ví dụ bit nào sẽ được rút ra tại bộ giải mã và bit nào sẽ được gửi từ bộ mã hóa.
- *Mã hóa syndrome*: Sau khi phân loại khối, mỗi khối được mã hóa theo lớp mã hóa của nó. Nếu một khối rơi vào lớp mã hóa Syndrome thì chỉ có các bit ít quan trọng nhất (LSB) của hệ số DCT lượng tử được mã hóa Syndrome bởi vì nó được giả định rằng các MSB có thể rút ra từ các SIB. Một phần của các bit LSB được mã hóa coset và phần còn lại được mã hóa entropy.
- *Bộ tạo băm*: Đối với mỗi khối, bộ mã hóa gửi tổng kiểm tra dư thừa vòng (CRC) 16 bit như chữ ký tương ứng của khối. Điều này là cần thiết để lựa chọn ra SIB ứng viên tốt nhất tại bộ giải mã.



## Quá trình giải mã

- *Tìm kiếm chuyển động*: Bộ giải mã tạo ra các SIB ứng viên là tất cả các khối dịch chuyển nửa pixel trong khung hình tham chiếu trong một vùng cửa sổ quanh tâm của khối cần giải mã.
- *Giải mã syndrome*: Mỗi SIB ứng viên sẽ đóng vai trò là SI cho giải mã syndrome. Quá trình này gồm hai bước: 1) thực hiện giải mã coset cho mỗi SIB và 2) thực hiện giải mã entropy cho các LSB đối với các khối đã thỏa mãn đối sánh hàm băm.
- *Kiểm tra băm*: Mỗi SIB ứng viên sẽ dẫn đến một khối giải mã. Để lựa chọn ra một trong số các ứng viên và phát hiện quá trình giải mã thành công (nghĩa là đạt được xác suất sai lỗi nhỏ), các khối giải mã sẽ được kiểm tra hàm băm với giá trị băm từ bộ mã hóa gửi tới. Nếu không thỏa mãn thì một SIB ứng viên mới lại tiếp tục được giải mã. Ngược lại, nếu thỏa mãn thì quá trình giải mã syndrome đi đến kết thúc.
- *Tái tạo và IDCT*: Chuỗi lượng tử sau khi khôi phục được sử dụng cùng với SIB tương ứng để tạo ra khối tái tạo tốt nhất.

### 1.3.4 So sánh kiến trúc Stanford và kiến trúc PRISM

Từ quan điểm kỹ thuật có thể thấy sự khác nhau cơ bản về chức năng giữa hai kiến trúc này như sau:

- Stanford là kiến trúc mã hóa dựa trên khung hình còn PRISM là kiến trúc mã hóa dựa trên khối. Có thể thấy rằng, kiến trúc PRISM có khả năng thích ứng mã hóa với các thay đổi của chuỗi video hơn kiến trúc Stanford.
- Stanford là kiến trúc điều khiển tốc độ tại bộ giải mã trong khi PRISM thực hiện điều khiển tốc độ tại bộ mã hóa. Kiến trúc Stanford yêu cầu

kênh phản hồi và do đó có hạn chế trong các ứng dụng thời gian thực, tuy nhiên lại làm cho bài toán điều khiển tốc độ dễ dàng hơn.

- Kiến trúc Stanford có bộ mã hóa cực đơn giản trong khi kiến trúc PRISM có bộ mã hóa phức tạp hơn và thông minh hơn. Đó là bởi vì bộ mã hóa của kiến trúc PRISM thực hiện một số các tác vụ liên ảnh và điều này cho phép kết hợp một số chế độ mã hóa thích ứng với sự thay đổi của chuỗi video.
- Kiến trúc Stanford sử dụng các mã kênh phức tạp, có thể kể đến mã turbo hay LDPC. Ngược lại, kiến trúc PRISM sử dụng các loại mã kênh đơn giản (ví dụ mã BCH).
- Kiến trúc Stanford không sử dụng dữ liệu bổ sung còn bộ mã hóa của kiến trúc PRISM tạo mã băm và gửi sang phía giải mã trong quá trình ước lượng chuyển động.
- Kiến trúc PRISM có khả năng chống nhiễu tốt hơn bởi vì kiến trúc này thực hiện tìm kiếm chuyển động tại bộ giải mã, điều này cho phép tìm ra SIB ít sai lỗi hơn so với kiến trúc Stanford.

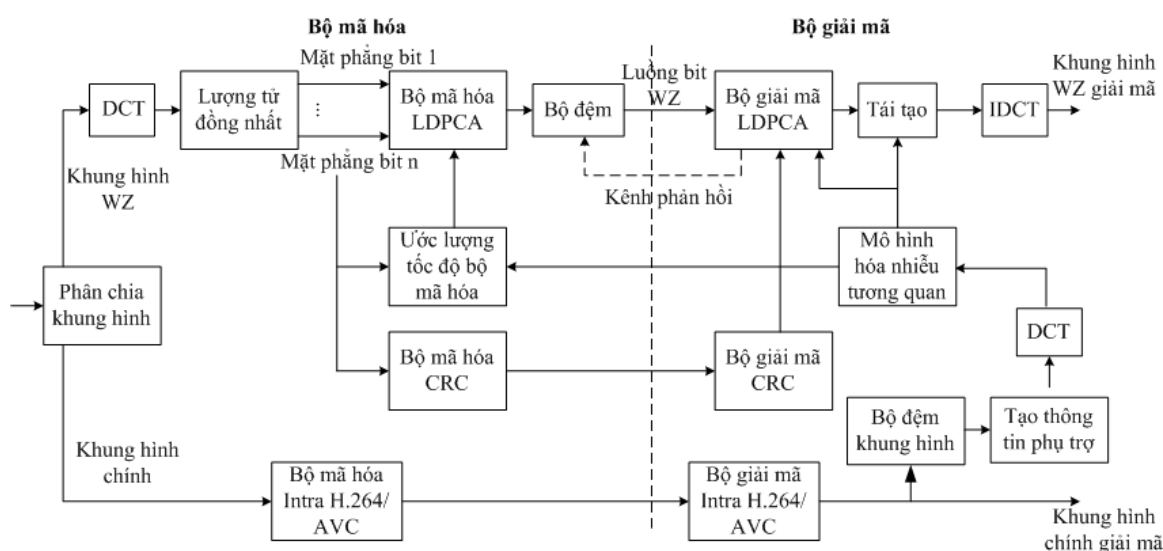
Theo thời gian, sự khác nhau giữa hai kiến trúc ban đầu này cũng có sự giảm bớt. Ví dụ, giờ đây có nhiều kiến trúc mã hóa video dựa trên Stanford cũng sử dụng mã hóa Intra dựa trên khối có lựa chọn, bộ mã hóa cũng sử dụng hàm băm và không cần kênh phản hồi. Tuy nhiên, sau một vài năm, khoảng cách hiệu năng giữa hai giải pháp mã hóa video phân tán này dường như càng tăng lên, ít nhất là ở khía cạnh hiệu năng nén trong trường hợp kênh không nhiễu.

Năm 2007, dự án DISCOVER [9] được xây dựng dựa trên kiến trúc Stanford cho thấy hiệu năng nén trong trường hợp kênh không nhiễu vượt trội so với H.264/AVC Intra với độ phức tạp mã hóa thấp hơn. Cũng trong năm 2007, nhóm của đại học Berkeley cũng giới thiệu một kiến trúc cải tiến có hiệu năng

tốt hơn H.263+ và thậm chí cả H.263+ kết hợp với mã sửa lỗi trước trong điều kiện mất gói.

### 1.3.5 Kiến trúc mã hóa video phân tán DISCOVER

Trong cộng đồng nghiên cứu DVC, kiến trúc DISCOVER thường được sử dụng như một tham chiếu trong các nghiên cứu. Trong luận án này cũng sử dụng kiến trúc DISCOVER làm cơ sở để kiểm chứng các đề xuất nên phần này xin được trình bày chi tiết về hoạt động của kiến trúc này. Kiến trúc DISCOVER được mô tả chi tiết như trong Hình 1.7.

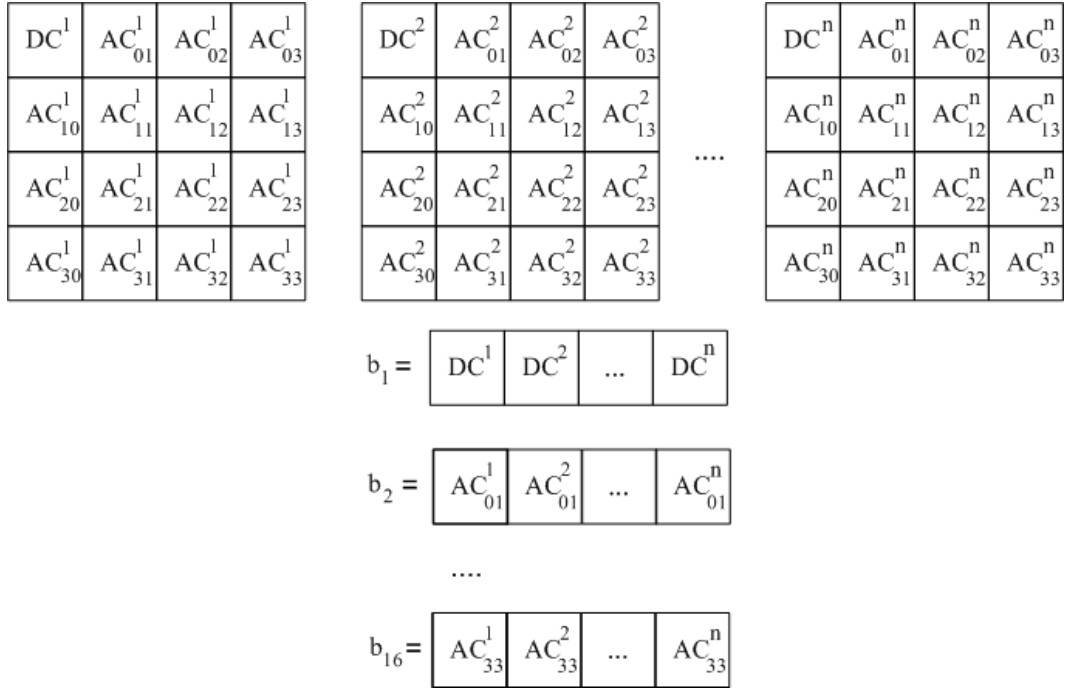


Hình 1.7: Kiến trúc mã hóa video phân tán DISCOVER

#### Quá trình mã hóa

- *Phân chia khung hình:* Chuỗi video được chia thành các WZF và KF. Khi sử dụng kích thước nhóm ảnh (GOP) bằng 2 nghĩa là các khung hình lẻ và chẵn là các KF và WZF tương ứng.
- *Biến đổi DCT:* Biến đổi DCT dựa trên khối  $4 \times 4$  được áp dụng cho các WZF. Các hệ số DCT của toàn bộ WZF được nhóm với nhau theo cùng

vị trí trong mỗi khối để tạo thành 16 dải hệ số DCT ký hiệu là  $b_k$  ( $k = 1, 2, \dots, 16$ ). Hình 1.8 mô tả cách tạo các dải hệ số DCT đối với WZF gồm  $n$  khối  $4 \times 4$ .



Hình 1.8: Các dải hệ số DCT

- *Lượng tử hóa:* Sau mã hóa biến đổi, mỗi dải hệ số DCT  $b_k$  được lượng tử hóa đồng nhất với  $2^{M_k}$  mức, ở đó số các mức phụ thuộc vào dải hệ số DCT  $b_k$ . Việc thay đổi số lượng các mức lượng tử khác nhau nhằm đạt được các chất lượng giải mã khác nhau. Thông thường có 8 ma trận lượng tử  $4 \times 4$  khác nhau tương ứng với 8 điểm tốc độ-méo tín hiệu (RD). Hình 1.9 dưới đây mô tả các ma trận lượng tử khác nhau tương ứng với tốc độ bit và chất lượng khác nhau. Trong ma trận lượng tử  $4 \times 4$ , giá trị tại mỗi vị trí để chỉ số các mức lượng tử liên quan đến dải hệ số DCT  $b_k$ . Ví dụ, giá trị 16 tại vị trí (0,0) trong ma trận (a) nghĩa là có  $2^{M_k} = 16$  mức lượng tử hay có 4 bit được gửi đi cho mỗi hệ số trong dải  $b_1$ . Giá trị  $2^{M_k} = 0$  nghĩa là không có bit WZ nào được truyền cho dải tương ứng.

Các chỉ số lượng tử hóa của mỗi dải DCT  $b_k$  sau đó được tổ chức lại thành

16	8	0	0
8	0	0	0
0	0	0	0
0	0	0	0

(a)

32	8	0	0
8	0	0	0
0	0	0	0
0	0	0	0

(b)

32	8	4	0
8	4	0	0
4	0	0	0
0	0	0	0

(c)

32	16	8	4
16	8	4	0
8	4	0	0
4	0	0	0

(d)

32	16	8	4
16	8	4	4
8	4	4	0
4	4	0	0

(e)

64	16	8	8
16	8	8	4
8	8	4	4
8	4	4	0

(f)

64	32	16	8
32	16	8	4
16	8	4	4
8	4	4	0

(g)

128	64	32	16
64	32	16	8
32	16	8	4
16	8	4	0

(h)

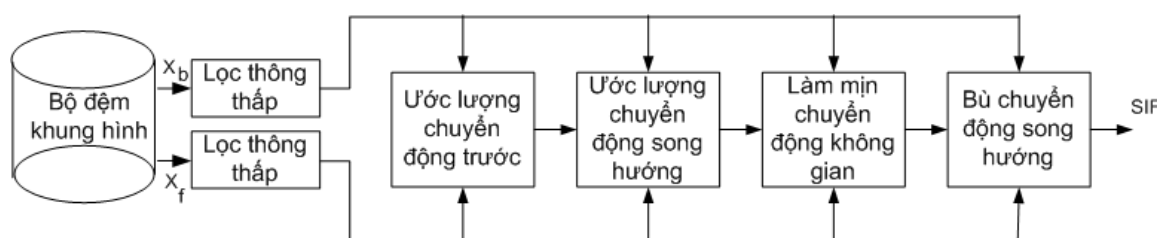
Hình 1.9: Tám ma trận lượng tử tương ứng với các điểm RD khác nhau

các mặt phẳng bit. Đối với một dải cho trước, các bit có mức độ quan trọng giống nhau sẽ được nhóm với nhau tạo thành mặt phẳng bit tương ứng và sau đó chúng sẽ được mã hóa LDPCA độc lập.

- *Mã hóa LDPCA*: Mỗi mặt phẳng bit trong dải hệ số DCT  $b_k$  sẽ được mã hóa bởi bộ mã hóa LDPCA để tạo ra thông tin chẵn lẻ. Các bit chẵn lẻ này được lưu giữ trong bộ đệm và được gửi theo từng cụm tới bộ giải mã tùy theo yêu cầu được gửi tới thông qua kênh phản hồi. Để hỗ trợ cho bộ giải mã phát hiện các lỗi còn sót lại, tổng kiểm tra CRC-8 bit của mặt phẳng bit tương ứng cũng được gửi tới bộ giải mã.
- *Ước lượng tốc độ bộ mã hóa*: Để hạn chế các yêu cầu từ bộ giải mã và vì vậy giảm độ phức tạp giải mã và độ trễ đường truyền, bộ mã hóa ước lượng một số bit ban đầu cần gửi cho mỗi mặt phẳng bit trước khi nhận được yêu cầu. Số bit này phải nhỏ hơn số lượng bit cuối cùng cần gửi để đảm bảo không có tổn thất về hiệu năng RD.

## Quá trình giải mã

- *Tạo thông tin phụ trợ*: SIF tương ứng với WZF được tạo ra bằng kỹ thuật nội suy khung hình bù chuyển động, sử dụng các khung hình tham chiếu trước và sau. Nếu ước lượng SIF càng chính xác, số lượng các “sai lỗi” mà bộ giải mã LDPCA cần phải sửa càng ít và tốc độ bit cần thiết cho quá trình giải mã thành công càng nhỏ (nghĩa là với xác suất lỗi nhỏ). Sơ đồ nội suy bù chuyển động (MCI) [17] được sử dụng trong kiến trúc DISCOVER được mô tả trong Hình 1.10 dưới đây:



Hình 1.10: Kiến trúc của kỹ thuật tạo thông tin phụ trợ bằng nội suy khung hình

Trước tiên, cả hai KF tham chiếu được lọc thông thấp và được sử dụng như các tham chiếu trong thuật toán ước lượng chuyển động tìm kiếm đầy đủ sử dụng tiêu chí đối sánh cải tiến. Tiêu chí đối sánh này ưu tiên các vector chuyển động ở gần tâm. Sau đó thực hiện ước lượng chuyển động song hướng. Trong bước này, các vector chuyển động tương ứng với mỗi SIB được lựa chọn và lọc. Cách tiếp cận phân cấp được sử dụng với các kích thước khối  $16 \times 16$  hoặc  $8 \times 8$ . Tiếp theo, bộ lọc vector trung vị có trọng số được sử dụng để cải thiện độ mịn không gian của trường chuyển động. Đối với mỗi SIB, các vector chuyển động ứng viên ở các khối lân cận thậm chí có thể hiển thị phép chiếu chuyển động tốt hơn. Bộ lọc này cũng có thể được điều chỉnh bởi một tập các trọng số (ví dụ độ mịn của trường chuyển động) tùy thuộc vào mức méo khối đối với mỗi vector chuyển động ứng viên.

Cuối cùng, sử dụng vector chuyển động có được, SIF được tạo ra bằng cách bù chuyển động trên các KF trước và sau.

- *Ước lượng DCT*: Thực hiện biến đổi DCT  $4 \times 4$  dựa trên khối cho SIF để tạo ra các hệ số DCT là ước lượng của hệ số DCT của WZF.
- *Mô hình hóa nhiễu tương quan*: Thống kê dư thừa giữa các hệ số DCT của WZF và SIF được giả định tuân theo phân bố Laplace. Tham số phân bố Laplace được ước lượng trực tuyến tại bộ giải mã ở các mức khác nhau, thường là mức dải DCT và mức hệ số.
- *Giải mã LDPCA*: Bộ giải mã nhận các dải DCT của WZF và thống kê dư thừa tương ứng với một dải hệ số DCT  $b_k$  và các cụm bit chẵn lẻ do bộ mã hóa gửi tới để giải mã ra khoảng lượng tử hóa của dải DCT đó. Sau khi giải mã LDPCA thành công cho mặt phẳng bit quan trọng nhất của dải  $b_k$ , bộ giải mã LDPCA tiếp tục theo cách tương tự cho các mặt phẳng bit còn lại của dải đó. Khi tất cả các mặt phẳng bit của dải hệ số DCT  $b_k$  được giải mã LDPCA thành công, bộ giải mã LDPCA bắt đầu giải mã cho dải  $b_{k+1}$ . Quá trình này được lặp lại cho đến khi tất cả các dải hệ số DCT được giải mã thành công.
- *Tiêu chí dừng gửi yêu cầu*: Để quyết định có cần gửi thêm bit cho quá trình giải mã một mặt phẳng bit nào đó thành công, bộ giải mã sử dụng tiêu chuẩn dừng gửi yêu cầu đơn giản. Đó là kiểm tra tất cả các biểu thức kiểm tra chẵn lẻ cho mã LDPC có được thỏa mãn đối với từ mã đã giải mã (quyết định cứng). Nếu không cần thêm bit để giải mã cho mặt phẳng bit, quá trình giải mã mặt phẳng bit hoặc dải tiếp theo có thể bắt đầu; ngược lại, giải mã LDPC cho mặt phẳng bit phải tiếp tục với một yêu cầu mới và nhận một cụm bit chẵn lẻ khác.
- *Kiểm tra CRC*: Bởi vì vẫn có một số lỗi dư thừa ngay cả khi tất cả các biểu thức kiểm tra chẵn lẻ được thỏa mãn và có thể gây ảnh hưởng tiêu

cực lên chất lượng khung hình giải mã, tổng kiểm tra CRC được bộ mã hóa gửi tới để giúp bộ giải mã phát hiện và sửa các lỗi còn lại trong mỗi mặt phẳng bit. Bởi vì CRC được kết hợp với tiêu chí dừng yêu cầu nên nó không cần phải quá mạnh để đảm bảo xác suất sai xấp xỉ bằng 0 cho mỗi mặt phẳng bit được giải mã. Do đó, tổng kiểm tra CRC-8 cho mỗi mặt phẳng bit được coi là đủ mạnh cho trường hợp này và nó chỉ cần thêm một lượng bit tối thiểu (8 bit).

- *Tái tạo*: Khi tất cả các mặt phẳng bit liên quan đến mỗi dải hệ số DCT được giải mã thành công, các mặt phẳng bit được nhóm lại với nhau và sẽ cho biết khoảng lượng tử cho mỗi hệ số DCT. Đại lượng này cùng với mô hình nhiễu tương quan sẽ hướng dẫn cho bộ giải mã biết giá trị hệ số DCT gốc tương ứng.
- *IDCT*: Thực hiện biến đổi IDCT  $4 \times 4$  để tái tạo WZF miền pixel.

## 1.4 Các vấn đề đang được nghiên cứu trong mã hóa video phân tán

Trong những năm qua đã có rất nhiều các công trình khoa học được thực hiện nhằm cải thiện hiệu năng cho mã hóa video phân tán. Về cơ bản, có thể tóm tắt các nội dung chính được tập trung nghiên cứu trong DVC bao gồm: cải thiện hiệu năng nén, cải thiện độ phức tạp mã hóa, cung cấp khả năng chống lỗi, cung cấp khả năng liên lớp và ứng dụng DVC trong các hệ thống đa góc nhìn.

### 1.4.1 Cải tiến hiệu năng nén

Mặc dù về mặt lý thuyết, hiệu năng nén của DVC có thể đạt mức tương đương với hệ thống mã hóa video truyền thống như H.264/AVC, HEVC nhưng các kết quả thực nghiệm cho thấy vẫn còn một khoảng cách khá lớn giữa hai sơ



đồ mã hóa này. Do đó hầu hết các nghiên cứu về DVC tập trung vào việc cải tiến hiệu năng nén, đặc biệt trong ngữ cảnh giữ cho bộ mã hóa độ có độ phức tạp thấp. Cải tiến hiệu năng nén của DVC thường được thực hiện trên các khối riêng lẻ: cải tiến bộ mã hóa Slepian-Wolf, cải tiến chất lượng thông tin phụ trợ, cải tiến tính chính xác của mô hình hóa nhiễu tương quan, cải thiện khối tái tạo và lựa chọn chế độ mã hóa.

### **Cải tiến bộ mã hóa Slepian-Wolf**

Bởi vì mã hóa Slepian-Wolf là cốt lõi trong mã hóa video phân tán, và mã hóa kênh đóng vai trò chính trong mã hóa Slepian-Wolf nên sự phát triển của mã hóa kênh sẽ góp phần quan trọng không chỉ về mặt hiệu năng nén mà còn cả khía cạnh độ phức tạp của bộ mã hóa. Có rất nhiều loại mã hóa kênh đã được sử dụng trong mã hóa video phân tán, ví dụ các mã khối được sử dụng trong PRISM [44], mã turbo sử dụng trong kiến trúc Stanford [22] hay mã LDPC [9].

Bên cạnh mã turbo, mã được sử dụng nhiều nhất trong mã hóa video phân tán là mã LDPC do hiệu năng của chúng. Các nghiên cứu cho thấy mã LDPC có thể tiệm cận dung lượng của nhiều loại kênh truyền thông khác nhau so với mã turbo. Do đó, mã LDPC được sử dụng nhiều hơn trong mã hóa video phân tán. Mã LDPC có sự cân bằng tốt giữa độ trễ, độ phức tạp và hiệu năng hệ thống. Trong thực tế, độ phức tạp giải mã LDPC dựa trên thuật toán tổng-tích thấp hơn so với mã turbo. Hơn nữa, độ phức tạp mã hóa cũng có thể đạt được bằng cách thiết kế mã LDPC phù hợp. Thông thường, hiệu năng của các mã này phụ thuộc vào chiều dài khối. Lớp mã LDPC thích ứng tốc độ cho mã hóa nguồn phân tán LDPCA có hiệu năng tốt hơn mã turbo đối với tốc độ vừa và tốc độ cao.

### **Tạo thông tin phụ trợ**

SIF có thể được xem là phiên bản nhiễu của WZF gốc, vì vậy mục đích của việc tạo SIF là tạo ra một ước lượng của WZF sao cho càng giống càng tốt. Chất

lượng của SIF có ảnh hưởng trực tiếp đến hiệu năng nén cuối cùng cũng như độ phức tạp giải mã bởi vì nếu SIF có chất lượng tốt thì sẽ chứa ít lỗi hơn và vì vậy số bit chẵn lẻ để giải mã sẽ ít hơn. SIF thường được tạo ra bằng cách nội suy hoặc ngoại suy các khung hình tham chiếu. Tuy nhiên, cách ước lượng này cũng gặp nhiều khó khăn bởi vì thông tin chuyển động thường không đồng nhất theo thời gian và sự chuyển cảnh hoặc chuyển động nhanh có thể ảnh hưởng nghiêm trọng đến độ chính xác của các phương pháp dựa trên nội suy hoặc ngoại suy. Ngoài ra, việc mã hóa với các kích thước GOP dài cũng có thể dẫn đến việc suy giảm chất lượng SIF.

Các bộ mã hóa DVC miền biến đổi thường biến đổi các khung hình thành các dải DCT  $b_k$  và việc giải mã được thực hiện theo từng dải. Khi một dải được giải mã thành công, nó sẽ cung cấp thêm thông tin cho bộ giải mã. Với sự trợ giúp của thông tin mới này, SIF có thể được lọc dần dần và vì vậy cải thiện hiệu suất mã hóa để giải mã phần còn lại của mỗi khung hình. Quá trình lọc rất quan trọng, đặc biệt khi chuyển động nhanh hoặc khi xảy ra chuyển cảnh bởi vì kỹ thuật nội suy/ngoại suy để tạo SIF ban đầu thường có kết quả kém trong các tình huống này. Tuy nhiên, việc lựa chọn các vùng trong SIF để lọc là rất cần thiết vì không cần phải cập nhật toàn bộ SIF, tức là một số vùng sẽ không thay đổi theo thời gian và mọi cập nhật trên vùng này có thể làm cho SIF thậm chí còn kém hơn và cùng lúc đó làm tăng độ phức tạp giải mã.

Tuy nhiên, những cải tiến về chất lượng SI trong quá trình giải mã không phải lúc nào cũng làm tăng hiệu suất nén cuối cùng nếu các mô-đun liên quan khác không được phối hợp tốt.

### **Mô hình nhiễu tương quan**

Bởi vì mục tiêu của mã hóa video phân tán là mã hóa có tổn thất về sự sai khác giữa dữ liệu gốc WZF và thông tin phụ trợ tương ứng của nó SIF, nên điều cần thiết để đạt được hiệu năng nén hiệu quả là bộ mã hóa và bộ giải mã phải có thông tin về mối tương quan thống kê giữa dữ liệu thông tin ban đầu

và thông tin phụ trợ.

Đối với kiến trúc Stanford với điều khiển tốc độ tại bộ giải mã thuần túy, chỉ bộ giải mã cần biết về CNM giữa WZF và SIF của nó. CNM này phải được ước lượng một cách thực tế, có nghĩa là tại bộ giải mã và không có quyền truy cập đến dữ liệu gốc. Mô hình hóa nhiều tương quan có thể được thực hiện ở nhiều mức độ khác nhau (ví dụ: mức dải hoặc mức hệ số), cho phép mô hình có sự thích ứng linh hoạt với sự tương quan thời gian và không gian khác nhau. Trong [16] [14] [18], Brites và các cộng sự đã đề xuất các giải pháp CNM cho kiến trúc Stanford miền pixel và miền biến đổi. Các kết quả cho thấy có mô hình hóa nhiều ở mức hệ số (miền biến đổi) và mức pixel (miền pixel) cho kết quả tốt nhất.

### **Lựa chọn chế độ mã hóa**

Mặc dù kiến trúc PRISM hoạt động dựa trên mức khối và kiến trúc Stanford hoạt động dựa trên mức khung hình, trong nhiều đề xuất cải thiện hiệu năng nén cho kiến trúc Stanford thực hiện thêm khối phân loại chế độ mã hóa dựa trên đặc tính của chuỗi video đầu vào. Các đặc tính chuyển động được xem xét dựa trên sự tương quan thời gian giữa các khung hình lân cận và sự tương quan không gian trong khung hình hiện thời.

Các nghiên cứu trong [57] [59] đề xuất thay đổi chế độ mã hóa cho mỗi khối hiện thời trong WZF để cải thiện chất lượng khung hình giải mã ở phía thu. Một hướng nghiên cứu khác tập trung vào việc thay đổi kích thước GOP [5] [6] [21] [62] tùy thuộc vào đặc tính của chuỗi video hiện thời. Các giải pháp này thường tiết kiệm được tốc độ bit và do đó cải thiện được hiệu năng nén của hệ thống.

### **Tái tạo**

Mô-đun cuối cùng trong kiến trúc mã hóa video phân tán thường là mô-đun tái tạo với mục tiêu là chuyển đổi các ký hiệu hoặc khoảng lượng tử đã giải

mã thành giá trị thực, hoặc giá trị độ chói cho WZF trong miền pixel hoặc hệ số biến đổi đối với kiến trúc miền biến đổi. Giải pháp ban đầu được giới thiệu trong [2], trong đó giá trị giải mã là thông tin phụ trợ nếu nó nằm trong khoảng đã giải mã hoặc các giới hạn biên gần với thông tin phụ trợ nếu nó rơi ra bên ngoài khoảng. Phương pháp này đã được sử dụng trong rất nhiều nghiên cứu về DVC. Gần đây, một giải pháp tái tạo mới đã được phát triển bởi Kubasov và các cộng sự [31] trong đó các giá trị giải mã được xây dựng lại bằng cách sử dụng phương pháp dựa trên MSE tối ưu sử dụng các biểu thức dạng đóng xuất phát từ mô hình tương quan Laplace.

### 1.4.2 Cung cấp khả năng chống lỗi

Các bộ mã hóa video truyền thống rất nhạy với lỗi đường truyền, vì vậy để truyền dữ liệu video qua kênh có tổn thất thường sử dụng hai cách: sửa lỗi trước (FEC) và tự động yêu cầu lặp lại (ARQ) hoặc kết hợp cả hai. ARQ có ưu điểm là chỉ truyền lại các gói bị mất nên không lãng phí băng thông. Tuy nhiên, ARQ lại không phù hợp cho kịch bản thời gian thực bởi vì nó yêu cầu độ trễ phù hợp giữa quá trình mã hóa và giải mã, khi truyền lại các gói bị mất (có thể nhiều lần). Ngược lại FEC lại dễ dàng sửa lỗi bằng cách truyền thêm các thông tin chẵn lẻ.

Không giống các kiến trúc video truyền thống, DVC lại có khả năng chống nhiễu nội tại do không có vòng lặp dự đoán như trong các bộ mã hóa của bộ mã hóa video dự đoán. Khi xảy ra lỗi, SI có thể không chính xác, và vì vậy, DVC sẽ hoạt động giống như một mã nguồn-kênh kết hợp chứ không chỉ đơn giản là một mã nguồn. Hầu hết các nghiên cứu về DVC khi tập trung vào khả năng chống lỗi đều nỗ lực tăng khả năng chống lỗi của mã hóa video dự đoán bằng cách thêm kênh phụ trợ với thông tin dư thừa được mã hóa theo nguyên tắc phân tán.

Trong [52], các tác giả sử dụng mã hóa phân tán để mã hóa cho dữ liệu dư thừa của một số khung hình để ngăn chặn sự lan truyền lỗi tại bộ giải mã. Thay

vì gửi định kỳ các khung hình được mã hóa Intra, bộ mã hóa sẽ gửi một lượng thông tin nhỏ gọi là thông tin coset tới bộ giải mã. Thông tin coset này được sử dụng để sửa các sai lỗi mà bộ mã hóa không cần phải có hiểu biết chính xác về gói hay thông tin đã bị mất.

Nghiên cứu về hiệu năng chống lỗi của kiến trúc DISCOVER cũng được thực hiện trong [8]. Các kết quả thực nghiệm cho thấy DISCOVER thực hiện tốt hơn H.264/AVC trong trường hợp kênh có lỗi và kích thước GOP nhỏ. Điều này càng xác nhận thêm khả năng chống lỗi nội tại của mã hóa video phân tán với sự có mặt của các loại mã hóa kênh, ví dụ mã turbo.

### 1.4.3 Cung cấp khả năng liên lớp

Khả năng liên lớp là một tính năng rất quan trọng khi truyền dẫn video trong môi trường mạng hỗn tạp – nơi có sự đa dạng về chất lượng đường truyền, khả năng xử lý và hiển thị của các thiết bị đầu cuối. Đối với các kiến trúc mã hóa video liên lớp hiện nay, các lớp cơ sở và tăng cường dựa trên cách tiếp cận dự đoán mà ở đó, bộ mã hóa sử dụng các khung hình đã giải mã ở lớp trước làm tham chiếu để tạo ra các lớp trên.

Với cách tiếp cận phân tán, chỉ cần biết tương quan giữa một lớp và SI được tạo ra từ lớp trước là đủ. Trong một số đề xuất các kiến trúc video liên lớp có sử dụng nguyên tắc mã hóa phân tán [51] [56] [64] [68], lớp cơ sở được mã hóa bằng các chuẩn mã hóa video truyền thống còn lớp tăng cường được mã hóa theo nguyên tắc phân tán. Các kết quả thực nghiệm cho thấy trong điều kiện đường truyền mất gói, hiệu năng nén các mã hóa video phân tán liên lớp vượt trội so với các mã hóa video dự đoán truyền thống.

### 1.4.4 Mã hóa video phân tán đa góc nhìn

Khi có nhiều góc nhìn khác nhau về cùng một hình ảnh được mã hóa bởi các nút camera khác nhau, ví dụ trong mạng cảm biến hình ảnh, thì việc mã

hóa giữa các cách nhìn có thể cải thiện hiệu năng mã hóa bằng cách khai thác dư thừa giữa các camera. Nếu sử dụng mã hóa video đa góc nhìn (MDVC) kiểu dự đoán thì cần phải có truyền tin giữa các camera. Các khung hình được giải mã thuộc về các góc nhìn khác sẽ được chèn vào danh sách khung hình tham chiếu và được sử dụng cho việc ước lượng/bù chuyển động sự chênh lệch giữa các góc nhìn khác nhau. Cách tiếp cận này yêu cầu việc truyền tin giữa các camera để cho phép một camera sử dụng các khung hình của camera khác cho việc bù chênh lệch.

Với kịch bản đa góc nhìn sử dụng DVC [23], mỗi camera có thể mã hóa độc lập các khung hình và bộ giải mã sẽ khai thác tương quan giữa các góc nhìn. Các bộ mã hóa MDVC sẽ khai thác đồng thời tương quan thời gian của cùng một góc nhìn và tương quan giữa các góc nhìn khác nhau. Trong [39][49], các tác giả đề xuất các cách tạo thông tin phụ trợ khác nhau cho hệ thống MDVC và các kết quả thực nghiệm cho thấy bằng việc kết hợp hai loại thông tin phụ trợ khác nhau trong MDVC có thể cải thiện đáng kể hiệu năng nén của hệ thống.

## 1.5 Các hướng nghiên cứu trong luận án

Với các hướng nghiên cứu DVC được trình bày ở mục 1.4, nghiên cứu sinh nhận thấy rằng việc cải tiến hiệu năng nén cho DVC được coi là quan trọng nhất vì các bộ mã hóa DVC thực tế vẫn chưa đạt được hiệu năng nén như các chuẩn video truyền thống mặc dù về mặt lý thuyết nó có hiệu năng tương đương.

Về mặt số lượng các nghiên cứu về DVC, hầu hết các nghiên cứu đều tập trung vào vấn đề cải tiến hiệu năng nén của hệ thống DVC, trong số đó nhiều nhất là các nghiên cứu về cải thiện thông tin phụ trợ và mô hình hóa nhiễu tương quan vì đây là các mô-đun quan trọng nhất quyết định hiệu năng nén của hệ thống.

Với mục tiêu cải tiến hiệu năng nén cho mã hóa video phân tán trong khi vẫn duy trì độ phức tạp thấp cho bộ mã hóa, nghiên cứu sinh tập trung vào các

vấn đề sau:

- Nghiên cứu đề xuất phương pháp thực hiện tại phía mã hóa để cải tiến hiệu năng nén của mã hóa video phân tán nhưng không làm tăng đáng kể độ phức tạp mã hóa.
- Nghiên cứu đề xuất các phương pháp có thể phức tạp hơn thực hiện tại phía giải mã để cải tiến hiệu năng nén cho mã hóa video phân tán.
- Nghiên cứu đề xuất các phương pháp cải tiến hiệu năng nén cho mở rộng của mã hóa video phân tán là mã hóa video phân tán liên lớp.

## 1.6 Kết luận chương

Chương này giới thiệu về mã hóa video nói chung, và đặc biệt là cung cấp nền tảng lý thuyết và các kết quả nghiên cứu về DVC. Nền tảng lý thuyết của DVC được xây dựng dựa trên các kết quả của mã hóa nguồn phân tán, trong đó các chuỗi tương quan được mã hóa bằng cách mã hóa độc lập và giải mã kết hợp. Hiệu suất nén của các bộ mã hóa nguồn như vậy được bắt nguồn từ định lý Slepian-Wolf cho mã hóa nguồn không tổn thất và định lý Wyner-Ziv mở rộng cho mã hóa nguồn có tổn thất. Dựa trên các kết quả lý thuyết này, một số các kiến trúc DVC thực tế được giới thiệu bao gồm: kiến trúc PRISM, kiến trúc Stanford và kiến trúc DISCOVER. Phần cuối là một số các nội dung nghiên cứu được thực hiện trong cộng đồng nghiên cứu DVC bao gồm: cải thiện hiệu năng mã hóa, lựa chọn chế độ mã hóa, cung cấp khả năng chống lỗi, cung cấp khả năng liên lớp và video phân tán đa góc nhìn. Dựa trên những khảo sát cũng như đánh giá sơ bộ về tình hình nghiên cứu DVC nói chung hiện nay, nghiên cứu sinh đã đưa ra hướng nghiên cứu cụ thể nhằm cải tiến hiệu năng nén cho mã hóa video phân tán và mã hóa video phân tán liên lớp.

## Chương 2

# Cải tiến hiệu năng nén cho mã hóa video phân tán

Như đã trình bày ở chương 1, cải tiến hiệu năng nén là một mục tiêu quan trọng đối với mã hóa video phân tán. Vì vậy, chương 2 giới thiệu các đề xuất liên quan đến vấn đề này bao gồm: (1) cải tiến hiệu năng nén cho DVC tại phía mã hóa sử dụng phương pháp thay đổi kích thước nhóm ảnh; (2) cải tiến hiệu năng nén cho DVC tại phía giải mã với mô-đun cải tiến phương pháp tạo thông tin phụ trợ và mô-đun ước lượng tham số  $\alpha$  cho mô hình nhiễu tương quan. Trước khi đi vào chi tiết các đề xuất này, phần Giới thiệu chung sẽ khảo sát một số nghiên cứu có liên quan.

## 2.1 Giới thiệu chung

Trong các chuẩn mã hóa video dự đoán hiện nay, hiệu năng nén đạt được bằng cách tăng độ phức tạp bộ mã hóa trong khi duy trì độ phức tạp bộ giải mã ở mức nhỏ nhất có thể. Độ phức tạp bộ mã hóa chủ yếu nằm ở quá trình ước lượng chuyển động phức tạp để tạo ra các khung hình dự đoán chính xác hay nói cách khác để làm giảm dư thừa giữa khung hình gốc và khung hình dự



đoán. Điều này giúp tạo ra tốc độ bit thấp hơn với cùng chất lượng video đầu ra.

Ngược lại, DVC chủ yếu thực hiện ước lượng chuyển động ở phía giải mã để tạo ra SIF, do đó bộ mã hóa được giữ đơn giản và vì vậy có độ phức tạp thấp. Đây cũng là phần quyết định nhiều nhất đến hiệu năng nén của DVC. Có thể phân loại các nhóm phương pháp cải tiến hiệu năng nén cho DVC thành nhóm các phương pháp được thực hiện tại phía mã hóa và nhóm các phương pháp thực hiện tại phía giải mã. Nhóm các phương pháp thực hiện tại phía mã hóa thường ít được quan tâm hơn vì DVC yêu cầu bộ mã hóa có độ phức tạp thấp. Tuy nhiên, nếu khai thác tốt các loại tương quan trong chuỗi video tại phía mã hóa bằng các công cụ phù hợp có thể tăng hiệu năng nén của bộ mã hóa DVC.

Nhóm các phương pháp thực hiện tại phía giải mã là nhóm mang lại mức cải thiện hiệu năng nén nhiều nhất cho DVC. Đó là bởi vì các công việc chính của DVC nằm ở phía giải mã và bộ giải mã DVC cho phép độ phức tạp cao hơn so với bộ mã hóa rất nhiều.

Cho đến hiện nay, có rất nhiều nghiên cứu đề xuất cải tiến hiệu năng nén cho DVC với các giải pháp đa dạng tập trung vào các phương pháp ở cả hai phía mã hóa và giải mã. Với mục tiêu cải tiến hơn nữa hiệu năng nén của DVC, chương này nghiên cứu và đề xuất một phương pháp thực hiện tại phía mã hóa và hai phương pháp được thực hiện tại phía giải mã nhằm cải tiến hiệu năng nén của DVC. Đó là phương pháp thay đổi kích thước nhóm ảnh tại phía mã hóa và nhóm phương pháp thực hiện tại phía giải mã bao gồm tạo thông tin phụ trợ và ước lượng tham số cho mô hình hóa nhiễu tương quan.

Trước khi đi vào chi tiết của các đề xuất, phần tiếp theo giới thiệu một số nghiên cứu trước đây có liên quan đến các nội dung được đề xuất trong chương này. Đó là các nghiên cứu về vấn đề cải tiến chất lượng thông tin phụ trợ, cải tiến tính chính xác của mô hình nhiễu tương quan và một số giải pháp thay đổi kích thước nhóm ảnh.

### 2.1.1 Các nghiên cứu về thông tin phụ trợ

Nếu như hiệu năng nén của các chuẩn mã hóa video dự đoán chủ yếu nằm ở các kỹ thuật ước lượng và bù chuyển động hiệu quả thì đối với DVC, chìa khóa là các kỹ thuật tạo SIF ở phía bộ giải mã. Bài toán ở đây là phải xây dựng SIF ở phía giải mã bằng cách sử dụng các khung hình đã giải mã trước đó sao cho càng giống WZF gốc ở phía mã hóa càng tốt. Nói cách khác, SIF ở phía giải mã được coi là một phiên bản “nhiều” của WZF gốc ở phía bộ mã hóa. Để hỗ trợ sửa các “lỗi” trong SIF, phía mã hóa sẽ gửi các thông tin chẩn lẻ sang phía giải mã. Cách tiếp cận này rất giống với mã hóa kênh, ở đó thông tin sai cần phải sửa nằm ở phía bộ giải mã. Nếu SIF được ước lượng tốt hay nói cách khác nếu nó càng giống WZF gốc thì số “sai lỗi” cần sửa càng ít và số bit kiểm tra bộ mã hóa cần gửi sang phía bộ giải mã càng ít.

Trong mã hóa video dự đoán, các kỹ thuật ước lượng và bù chuyển động được thực hiện ở phía bộ mã hóa với điều kiện các khung hình gốc sẵn có ở phía bộ mã hóa. Ngược lại, đối với DVC, một loạt các kỹ thuật ước lượng chuyển động được thực hiện ở phía giải mã nhằm tạo ra SIF mà không có sự xuất hiện của WZF gốc ở phía giải mã. Tuy nhiên, ước lượng gặp rất nhiều khó khăn vì thông tin chuyển động luôn thay đổi và không đồng nhất dọc theo chuỗi, và sự thay đổi cảnh hay sự bất thường về độ chói có thể làm giảm độ chính xác của các kỹ thuật tạo SIF.

Có rất nhiều đề xuất tạo SI, tuy nhiên có thể phân loại các kỹ thuật tạo SI thành hai loại chính là kỹ thuật tạo SI cơ bản và kỹ thuật tạo SI nâng cao. Các kỹ thuật tạo SI cơ bản thường tạo ra SIF có thể sử dụng trực tiếp trong quá trình giải mã. Ngược lại, các kỹ thuật tạo SI nâng cao thường thực hiện cải tiến chất lượng của SIF ban đầu được tạo ra bằng các kỹ thuật tạo SI cơ bản.

Với các kỹ thuật tạo SI cơ bản thường phát triển theo hai hướng chính. Cách tiếp cận thứ nhất tạo ra SIF cho mỗi WZF tại bộ giải mã trước khi quá trình giải mã Slepian-Wolf bắt đầu. SIF được tạo ra dựa trên các khung hình đã giải mã

trước đó mà không cần bất cứ dữ liệu bổ sung nào từ phía bộ mã hóa [3][10][32]. Các giải pháp này có ưu điểm là bộ mã hóa không cần tạo và gửi các thông tin bổ sung về khung hình hiện thời nhưng có nhược điểm là khó ước lượng chuyển động một cách chuẩn xác khi kích thước GOP lớn cũng như tại các vùng có chuyển động bất thường hay chuyển động nhanh. Cách tiếp cận này thường sử dụng trong các kiến trúc tuân theo sơ đồ Stanford.

Ngược lại, cách tiếp cận thứ hai trong nhóm các kỹ thuật tạo SI cơ bản thường sử dụng trong các kiến trúc tuân theo sơ đồ PRISM. Trong cách tiếp cận này, SI được tạo ra bằng cách thử, trên cơ sở từng khối dựa vào các khung hình đã giải mã trước đó và một số dữ liệu điều khiển từ bộ mã hóa gửi tới. Theo cách tiếp cận này, bộ giải mã Slepian-Wolf sẽ chạy một số lần cho mỗi ứng viên SI được tạo ra tại bộ mã hóa và ứng viên tốt nhất sẽ được chọn dựa theo tiêu chí đã định trước [44][43]. Có thể thấy cách tiếp cận này có ưu điểm là có khả năng thích ứng tốt với nội dung video do có thực hiện một số kỹ thuật ước lượng đơn giản. Tuy nhiên nó lại làm tăng thêm độ phức tạp cho bộ mã hóa cũng như cần phải truyền thêm dữ liệu điều khiển sang phía giải mã.

Để cải thiện hơn nữa chất lượng của SI, các giải pháp tạo SI nâng cao được đề xuất. Các SI vẫn có thể được tạo ra trước khi quá trình giải mã Slepian-Wolf bắt đầu nhưng có sử dụng thêm một số thông tin bổ sung hay còn gọi là các gợi ý thêm từ bộ mã hóa gửi tới [1][42]. Trong các giải pháp này, bộ mã hóa thường tạo mã băm cho mỗi khối và gửi tới bộ giải mã để hỗ trợ quá trình giải mã nhằm cải thiện chất lượng SI. Tuy nhiên, nhược điểm của các giải pháp này là độ phức tạp của bộ mã hóa sẽ tăng lên.

Một hướng tiếp cận khác là SI ban đầu được lọc liên tục qua nhiều vòng lặp trong quá trình giải mã Slepian-Wolf bằng cách sử dụng các dữ liệu đã giải mã trước đó [35] [4] [34]. Với cách tiếp cận này, sự tương đồng giữa SIF và WZF sẽ tăng liên tục trong quá trình giải mã và vì vậy sẽ cải thiện chất lượng cuối cùng và giảm tốc độ bit yêu cầu. Hơn nữa, do không cần phải tạo và gửi thông tin thêm về khung hình hiện thời nên bộ mã hóa vẫn giữ được độ phức tạp thấp.

Nhược điểm của các cách tiếp cận này là độ phức tạp bộ giải mã sẽ tăng lên so với các giải pháp trước đó.

Các khảo sát về SI ở trên cho thấy rằng nếu SI được lọc liên tục trong quá trình giải mã có thể cải thiện đáng kể chất lượng của nó mà không ảnh hưởng đến độ phức tạp mã hóa. Vì vậy, trong phạm vi luận án này, một phương pháp tạo SI bằng cách lọc liên tục trong suốt quá trình giải mã được đề xuất. So với các cách tạo SI cơ bản, cách tạo SI dựa trên việc lọc cho thấy các cải tiến đáng kể. Đó là do việc tận dụng các thông tin đã giải mã trước đó để cải thiện liên tục chất lượng của SI trong quá trình giải mã. SI được xây dựng càng chính xác sẽ giúp làm giảm lượng bit cần truyền cũng như nâng cao chất lượng của khung hình được giải mã.

Trong đề xuất này, SIF ban đầu ( $SI_{MCTI}$ ) được tạo ra bằng kỹ thuật MCTI. Sau mỗi mặt phẳng bit hoặc dải được giải mã, các hệ số tái tạo được thay thế vào  $SI_{MCTI}$  ban đầu để tạo ra SI mới được gọi là SI được giải mã một phần (PDWZ). Sau đó thực hiện ước lượng chuyển động trên ba khung hình tham chiếu là  $SI_{MCTI}$ , khung hình chính đã giải mã trước và sau để tìm ra ba khối ứng viên trên ba khung hình tham chiếu tương ứng với ba vector chuyển động. Sau đó thực hiện một quá trình kết hợp ba khối ứng viên này để tạo ra SI cuối cùng có chất lượng tốt nhất. Đóng góp này đã được trình bày tại Hội thảo Quốc tế về Truyền thông và Công nghệ thông tin năm 2019 (ISCIT 2019) [Bài báo 2].

### 2.1.2 Các nghiên cứu về mô hình nhiễu tương quan

Như đã trình bày ở phần trên, SI đóng vai trò quan trọng đối với hiệu năng của hệ thống. Vì vậy, nếu SIF được ước lượng càng giống WZF, số lượng các sai lỗi cần phải sửa bởi bộ giải mã càng ít. Vì vậy, việc ước lượng nhiễu tương quan giữa SIF và WZF gốc rất quan trọng đối với hiệu năng nén của mã hóa video phân tán.

Trong các nghiên cứu về CNM trong DVC, nhiễu này thường được mô hình hóa bằng phân bố Laplace [18][28]. Để khai thác nhiều hơn thông tin tương

quan, một số các phân bố khác đã được kiểm chứng trong một số nghiên cứu như mô hình hàm mũ trong [37] hay mô hình nhiễu tương quan thích ứng kết hợp cả phân bố Laplace và phân bố Gauss trong [45]. Tuy nhiên, phân bố Laplace thường được sử dụng vì nó cân bằng giữa tính chính xác mô hình và độ phức tạp tính toán. Khung hình dư thừa  $R = WZF(x, y) - SIF(x, y)$  được mô hình hóa bởi phân bố Laplace như trong biểu thức 2.1 dưới đây:

$$f_R(r) = \frac{\alpha}{2} e^{-\alpha|r|} \quad (2.1)$$

ở đó  $f_R(r)$  là hàm phân bố xác suất và tham số phân bố Laplace,  $\alpha$  được tính bởi:

$$\alpha = \sqrt{\frac{2}{\sigma^2}} \quad (2.2)$$

ở đó  $\sigma^2$  là phương sai của khung hình dư thừa  $R$ .

Tham số phân bố Laplace  $\alpha$  có thể được ước lượng ở các mức khác nhau: mức khung hình, mức dải DCT hay mức hệ số. Trong các nghiên cứu về CNM, các tham số của CNM có thể được ước lượng dựa trên quá trình xử lý trực tuyến hoặc ngoại tuyến. Ước lượng tham số CNM ngoại tuyến [9][3][38] là các giải pháp mà ở đó tham số CNM được ước lượng tại bộ mã hóa sử dụng WZF gốc và ước lượng tham số CNM trực tuyến [14][15][36] nghĩa là các tham số CNM được ước lượng tại phía bộ giải mã mà không cần sử dụng WZF gốc. Mặc dù các cách tiếp cận ngoại tuyến cho hiệu năng nén tốt hơn so với cách tiếp cận trực tuyến nhưng nó không nhận được nhiều sự quan tâm bởi vì đây là kịch bản không mong muốn. Bộ mã hóa phải thực hiện các ước lượng chuyển động phức tạp để tạo ra SIF như tại bộ giải mã nên bộ mã hóa không giữ được độ phức tạp thấp như mục tiêu xây dựng ban đầu.

Một hướng tiếp cận khác đề xuất ước lượng tham số mô hình nhiễu tương quan tại cả phía bộ mã hóa và giải mã [60][61][25]. Trong các nghiên cứu này, mô hình nhiễu tương quan thực hiện xác định số lượng các mặt phẳng bit ít quan trọng nhất ( $n_{LSB}$ ) cần được mã hóa và truyền tới bộ giải mã và giá trị này được tính toán tại cả hai phía. Trong khi [60] đề xuất giải pháp CNM bất đối

xúng ở đó  $n_{LSB}$  được tính riêng biệt tại cả bộ mã hóa và giải mã với các phương pháp tạo SI khác nhau thì đề xuất trong [61] lại sử dụng cùng một phương pháp tạo thông tin tương quan tại bộ mã hóa và giải mã. Có thể thấy rằng [60] có thể tạo ra các thông tin tương quan khác nhau tại hai phía thì đề xuất trong [61] lại làm tăng độ phức tạp bộ mã hóa. Để khắc phục các nhược điểm này, giải pháp CNM thích ứng được đề xuất trong [25] sử dụng cách tiếp cận tối ưu RD. Các kết quả cho thấy giải pháp này cho phép duy trì độ phức tạp thấp cho bộ mã hóa trong khi vẫn cải thiện được hiệu năng RD của hệ thống.

Để ước lượng tham số CNM một cách chính xác hơn, các tham số CNM có thể được cập nhật liên tục sau mỗi lần giải mã một mặt phẳng bit hay một dải [34][41]. Đó là bởi vì sau khi giải mã các mặt phẳng bit có thể nhận được nhiều thông tin hơn và thông tin này được khai thác để giải mã cho các mặt phẳng bit hay dải tiếp theo.

Gần đây, mạng nơ ron đã được áp dụng và đạt được nhiều thành công đáng kể trong nhiều lĩnh vực bao gồm cả nén video. Đối với các nén video truyền thống, đã có một số các phương pháp dựa trên mạng nơ ron được đề xuất cho các mô-đun cụ thể như dự đoán nội ảnh và mã hóa dư thừa, mã hóa entropy để cải thiện hiệu năng của hệ thống. Đối với DVC, đã có một số giải pháp tạo SI dựa trên học sâu được đề xuất [20][58].

Nhận thấy các kết quả đáng chú ý của việc sử dụng mạng nơ ron trong nén video, và để khai thác thế mạnh của mạng nơ ron để cải tiến hơn nữa hiệu năng nén của DVC, một kỹ thuật ước lượng tham số CNM dựa trên mạng nơ-ron được đề xuất trong luận án này. Trong đề xuất này, quá trình học được thực hiện trên khung hình dư thừa  $R$  – khung hình được tạo ra từ các khung hình chính đã giải mã trước đó tại bộ giải mã. Các kết quả thực nghiệm cho thấy giải pháp này đạt được hiệu năng nén tốt hơn so với các giải pháp có liên quan, có thể tiết kiệm đến 35% tốc độ bit khi so với kiến trúc DISCOVER và khoảng 22% tốc độ bit khi so với HEVC Intra trong khi vẫn cung cấp chất lượng khung hình tái tạo tương đương. Đóng góp này đã được chấp nhận đăng trên Tạp chí

Điện tử và Truyền thông (REV) năm 2020 (Bài báo 3).

### 2.1.3 Các nghiên cứu về thay đổi kích thước nhóm ảnh

Không phải là các mô-đun chính như tạo SI hay CNM nhưng các nghiên cứu về lựa chọn chế độ mã hóa trong kiến trúc DISCOVER cũng nhận được nhiều quan tâm trong cộng đồng nghiên cứu DVC.

Như đã trình bày ở trên, SI và CNM là hai mô-đun quan trọng nhất ảnh hưởng đến hiệu năng nén của hệ thống. Cả hai thành phần này đều phụ thuộc vào khoảng cách giữa hai khung hình chính hay kích thước nhóm ảnh (GOP). Tuy nhiên, sử dụng kích thước GOP cố định cho toàn bộ chuỗi video có thể không hiệu quả bởi vì tương quan thời gian không được khai thác đầy đủ khi nội dung chuỗi video thay đổi. Đối với khung hình chuyển động nhanh, tương quan thời gian nhỏ nên cần phải chọn kích thước GOP nhỏ. Ngược lại, đối với các khung hình có chuyển động chậm hoặc trung bình, tương quan thời gian lớn và trong trường hợp này nên chọn kích thước GOP lớn hơn.

Các đề xuất trong [6][21][62][11] thực hiện điều khiển kích thước GOP theo sự thay đổi của chuyển động trong chuỗi video. Trong [11], các tác giả sử dụng các đặc trưng liên quan đến lược đồ ảnh và phương sai khối để phân loại mức độ hoạt động của chuỗi video. Một ý tưởng khác từ [6] sử dụng đáp ứng của hệ thống trong quá khứ để lựa chọn kích thước GOP. Ban đầu, một tập nhỏ bao gồm  $N$  kích thước GOP khác nhau được tạo ra. Hiệu năng mã hóa của mỗi kích thước GOP được tính toán dựa trên tỷ số của PSNR ước lượng trung bình và tốc độ mã hóa trung bình. Kích thước GOP nào cho tỷ lệ cao nhất sẽ được chọn làm kích thước GOP tiếp theo. Krishna R.V và cộng sự trong [62] đề xuất một thuật toán điều khiển kích thước GOP đơn giản trong đó các khối trong một khung hình được phân loại thành các khối Intra, khối SKIP và khối WZ. Khung hình hiện thời được phân loại thành WZF hay KF tùy thuộc vào số lượng các khối SKIP. Các kết quả cho thấy giải thuật này cho kết quả khá tốt với độ phức tạp bộ mã hóa tăng không đáng kể.

Sau khi khảo sát về tình hình nghiên cứu của việc lựa chọn chế độ mã hóa trong DVC, nghiên cứu sinh nhận thấy rằng có thể cải thiện hiệu năng nén của hệ thống bằng cách thay đổi kích thước GOP. Với lựa chọn thay đổi kích thước GOP, có thể đạt được hiệu năng nén tốt hơn nếu có thể phân loại chính xác kích thước GOP theo nội dung video. Vì vậy, nghiên cứu sinh đề xuất sử dụng một thuật toán học máy để lựa chọn kích thước GOP một cách hiệu quả cho mỗi đoạn video.

Trong đề xuất này, các khung hình của chuỗi video đầu vào trước khi mã hóa được đưa vào khối trích chọn đặc trưng để trích xuất đặc tính chuyển động của chuỗi. Các đặc trưng này được đưa vào huấn luyện và phân loại bằng kỹ thuật học máy để quyết định khung hình hiện tại là chuyển động nhanh hay chuyển động chậm. Nếu là chuyển động chậm thì sẽ lựa chọn  $GOP = 4$ . Ngược lại thì sẽ chọn  $GOP = 2$ . Đóng góp này đã được báo cáo tại Hội nghị Quốc tế về các Công nghệ tiên tiến trong Truyền thông năm 2018 (ATC 2018) [Bài báo 1].

Phần tiếp theo của chương này sẽ trình bày chi tiết các đóng góp được giới thiệu ở trên và được cấu trúc như sau. Phần 2.2 trình bày các kết quả đạt được khi thực hiện thay đổi kích thước GOP tại phía mã hóa. Phần 2.3 đề xuất các phương pháp thực hiện tại phía giải mã để cải thiện hiệu năng nén cho DVC với phương pháp tạo thông tin phụ trợ bằng cách lọc liên tục trong quá trình giải mã và cải tiến tính chính xác của tham số  $\alpha$  của mô hình hóa nhiễu tương quan dựa trên mạng nơ-ron. Cuối cùng là kết luận chương được tóm tắt trong phần 2.4.

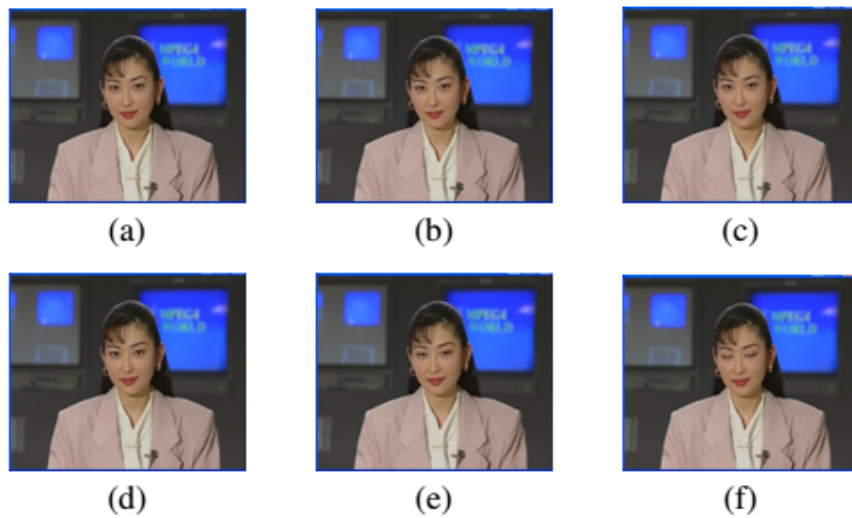


## 2.2 Đề xuất 1: Cải tiến hiệu năng nén cho DVC tại phía mã hóa

### 2.2.1 Đặt vấn đề

Trong DVC, SI và CNM là hai thành phần chức năng chính quyết định chủ yếu đến hiệu năng mã hóa. Mặt khác, kết quả của cả hai quá trình này lại phụ thuộc vào GOP - số lượng khung hình giữa hai KF liên tiếp nhau. Khi  $GOP = 2$  nghĩa là giữa hai KF có một WZF và khi  $GOP = 4$  nghĩa là giữa hai KF có ba WZF. Khi kích thước GOP càng lớn, việc dự đoán SIF từ các KF đã giải mã càng khó khăn vì khoảng cách giữa chúng càng lớn. Cùng với đó, việc ước lượng CNM cũng trở nên khó chính xác hơn. Vì vậy, trong DVC, GOP thông thường được chọn bằng 2.

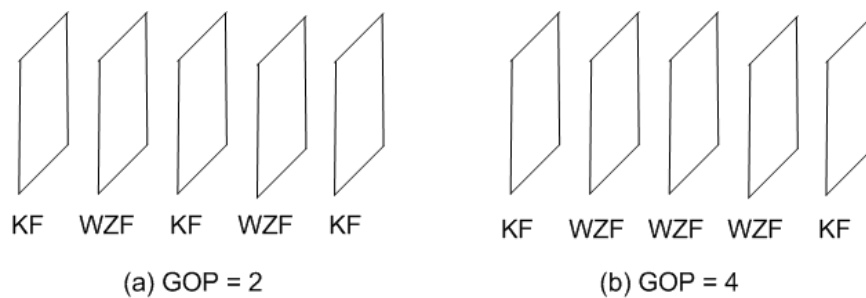
Tuy nhiên, đối với các đoạn video có chuyển động chậm, tương quan giữa các khung hình cao thì việc ước lượng SI cũng như CNM dễ đạt được kết quả chính xác. Để trực quan có thể xem xét Hình 2.1 dưới đây. Hình này gồm sáu khung hình đầu tiên của chuỗi *News*. Đây là chuỗi được coi là chuyển động chậm với vùng chuyển động chủ yếu ở phần mặt và với vùng nền tương đối tĩnh. Có thể thấy rằng các khung hình này gần như giống nhau, ngoại trừ vùng mắt mở trong khung hình (a) và mắt nhắm trong khung hình (f). Nếu sử dụng kích thước  $GOP = 2$  nghĩa là các khung hình (a), (c) và (e) sẽ được phân loại là các KF. Tuy nhiên, khi quan sát bằng mắt có thể nhận thấy các khung hình này khá giống nhau và có thể chỉ thiết lập các khung hình (a) và (f) là KF là đủ hay nói cách khác kích thước GOP có thể tăng lên bằng 5. Giải pháp này có thể tiết kiệm một số lượng lớn các bit và cải thiện hiệu năng nén đáng kể. Một số nghiên cứu thay đổi kích thước GOP tùy thuộc vào nội dung chuỗi video cho thấy có thể đạt được kết quả khá tốt với độ phức tạp mã hóa tăng không đáng kể. Tuy nhiên, các giải thuật điều chỉnh kích thước GOP này phụ thuộc chủ yếu vào một số giả định có tính xác định. Do đó mức cải thiện hiệu năng nén chưa



Hình 2.1: Sáu khung hình đầu tiên của chuỗi video News

cao.

Mục đích của đề xuất mới là phân loại kích thước GOP một cách chính xác hơn dựa trên nội dung video. Do nội dung của dữ liệu video rất đa dạng nên đề xuất này sử dụng một thuật toán học máy để lựa chọn kích thước GOP cho mỗi phân đoạn video gồm năm khung hình một cách hiệu quả. Như mô tả trong hình 2.2, khi đoạn video là chuyển động nhanh thì phân đoạn đó có hai WZF và khi chuyển động chậm thì có ba WZF.



Hình 2.2: (a) Đoạn chuyển động nhanh và (b) Đoạn chuyển động chậm

Để quyết định chính xác loại của một phân đoạn video bằng phương pháp học máy, cần phải lựa chọn các đặc trưng mô tả chính xác về nội dung chuyển động nhanh hay chậm, phức tạp hay đơn giản của phân đoạn đó và sử dụng một

thuật toán học máy phù hợp để phân loại chuyển động.

Tuy nhiên, việc quyết định kích thước GOP cần phải được thực hiện ở bước đầu tiên của quá trình mã hóa WZF tại phía mã hóa nên cần lựa chọn một thuật toán học máy hiệu quả nhưng không quá phức tạp. Điều này là cần thiết nhằm đảm bảo bộ mã hóa giữ được độ phức tạp thấp như ban đầu. Các đặc trưng được chọn ngoài việc phản ánh được nội dung chuyển động cũng như cấu trúc bề mặt của đoạn video cũng cần đảm bảo tính toán nhanh và đơn giản. Quá trình huấn luyện được thực hiện ngoại tuyến để đảm bảo bộ mã hóa đơn giản.

Tập dữ liệu huấn luyện đã được thử nghiệm với một số thuật toán học máy để phân loại bao gồm cây quyết định, Gaussian Naive Bayes và hồi quy logistic để so sánh độ chính xác. Các kết quả cho thấy cây quyết định cho độ chính xác cao nhất. Mặc dù thuật toán này đơn giản nhưng đây là một thuật toán phù hợp với nhiều loại dữ liệu và không cần chuẩn hóa dữ liệu đầu vào. Do đó thuật toán cây quyết định mà cụ thể là C4.5 [46] được lựa chọn để phân loại kích thước GOP.

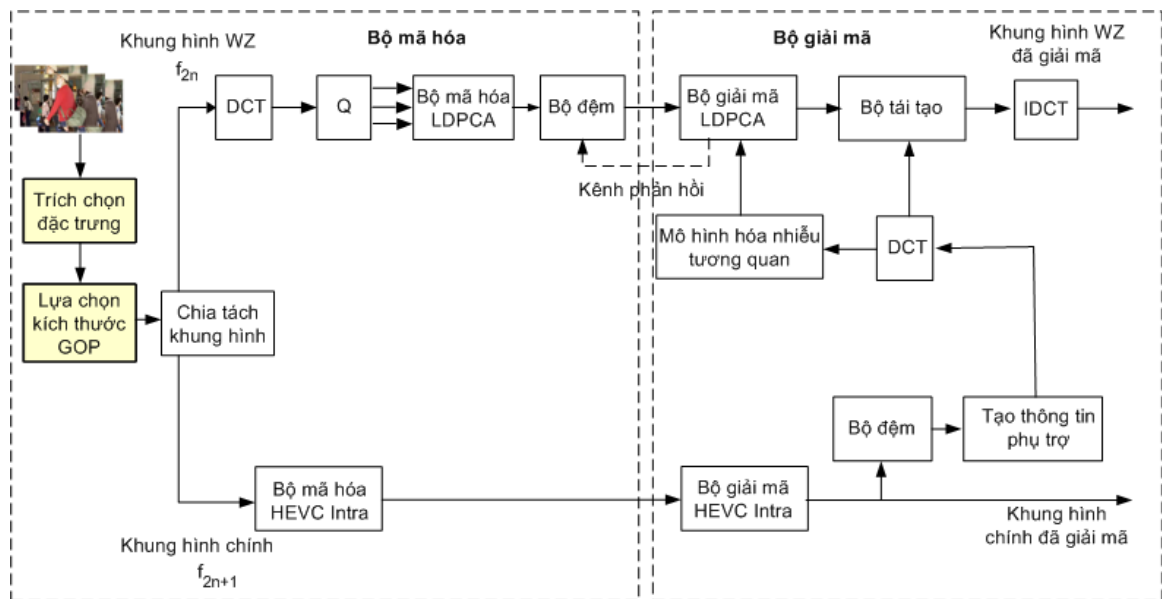
Các đặc trưng sử dụng và quá trình huấn luyện, phân loại được mô tả chi tiết ở mục 2.2.2.

### 2.2.2 Phương pháp thay đổi kích thước GOP đề xuất

Phương pháp thay đổi kích thước GOP đề xuất được thực hiện tại bộ mã hóa DVC. Vì vậy trước khi giới thiệu chi tiết về đề xuất này, phần tiếp theo sẽ giới thiệu tóm tắt kiến trúc bộ mã hóa đề xuất với mô-đun mới thay đổi kích thước GOP.

#### A. Kiến trúc bộ mã hóa AGOP-DVC đề xuất

Kiến trúc mã hóa video phân tán AGOP-DVC với khối đề xuất mới được tô màu đậm được mô tả trong hình 2.3 dưới đây. Quá trình mã hóa và giải mã có thể tóm tắt như sau:



Hình 2.3: Kiến trúc mã hóa video phân tán AGOP-DVC

### Quá trình mã hóa

Chuỗi video được chia thành hai loại khung hình là KF và WZF. Trong đề xuất này, kích thước GOP bằng 2 hoặc 4 tùy thuộc vào nội dung của chuỗi video. Chi tiết về quá trình lựa chọn kích thước GOP sẽ được trình bày ở phần sau. Các KF được mã hóa Intra bằng chuẩn mã hóa video truyền thống HEVC Intra. Các WZF được mã hóa theo nguyên tắc phân tán. Mỗi WZF được chia thành các khối có kích thước  $4 \times 4$  và mỗi khối được biến đổi DCT  $4 \times 4$ . Các hệ số DCT của mỗi WZF được sắp xếp thành 16 dải trong đó các hệ số có cùng vị trí từ các khối khác nhau sẽ nằm trong cùng một dải. Các dải DCT này sau đó được lượng tử vô hướng đồng nhất. Các ma trận lượng tử tương ứng với các tốc độ khác nhau như trong [56]. Các dải DCT sau lượng tử được nhị phân hóa và các bit với cùng độ quan trọng được nhóm lại thành các mặt phẳng bit. Các mặt phẳng bit này được đưa vào bộ mã hóa LDPC để tạo ra các bit chẵn lẻ. Các bit chẵn lẻ được lưu giữ trong bộ đệm và sẽ gửi tới bộ giải mã tùy theo yêu cầu của bộ giải mã.

### Quá trình giải mã

Các KF được giải mã được sử dụng để tạo nên SI. SI này được biến đổi DCT  $4 \times 4$  để tạo thành một ước lượng của các hệ số DCT của WZF. Các hệ số DCT của SI này cùng với nhiễu tương quan được đưa vào bộ giải mã LDPCA. Bộ giải mã này sẽ sửa các "sai lỗi" trong SI bằng cách sử dụng các bit chẵn lẻ bên mã hóa gửi sang. Sau khi đã giải mã LDPCA, bộ tái tạo sẽ cùng với các hệ số DCT của SI để khôi phục lại hệ số DCT của WZF ban đầu. Cuối cùng, biến đổi DCT ngược được thực hiện để khôi phục lại WZF miền pixel.

## B. Phương pháp thay đổi kích thước GOP đề xuất

Trong đề xuất này, mỗi phân đoạn gồm 5 khung hình video đầu vào được đánh giá để lựa chọn kích thước GOP phù hợp bằng cách trích chọn các đặc trưng. Các đặc trưng này được đưa vào cây quyết định để phân loại mỗi phân đoạn này là chuyển động nhanh hay chuyển động chậm để lựa chọn kích thước GOP phù hợp. Để tiết kiệm thời gian xử lý, quá trình huấn luyện được thực hiện ngoại tuyến trước khi sử dụng cho quá trình mã hóa.

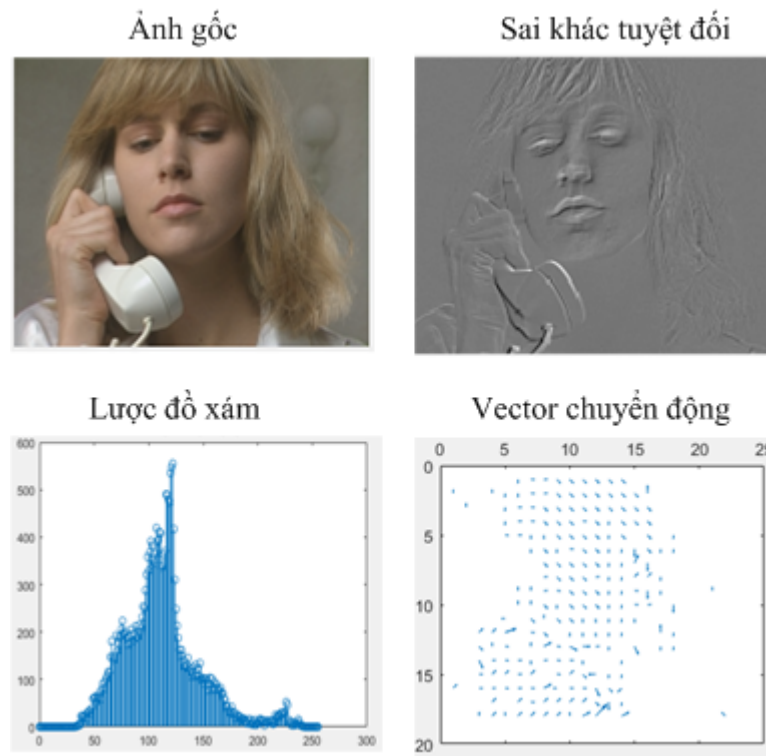
### 1) Xác định đặc trưng

Như đề cập ở trên, các đặc trưng được lựa chọn phải phản ánh đầy đủ bản chất của nội dung video. Vì vậy, một số đặc trưng được lựa chọn liên quan đến chuyển động trong khi một số khác lại liên quan đến kết cấu.

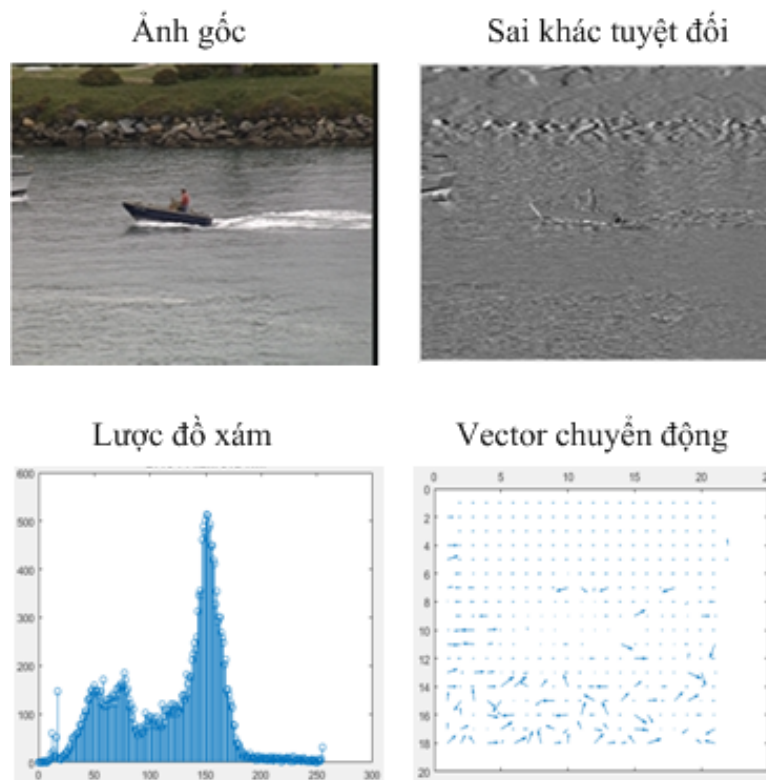
Hình 2.4 và hình 2.5 mô tả một số đặc trưng của hai chuỗi video *Suzie* và *Coastguard*. Các hình vẽ cho thấy với chuỗi chuyển động chậm *Suzie*, đặc trưng SAD ít có sự thay đổi, lược đồ xám đơn giản và trường véc-tơ chuyển động ít. Ngược lại, với chuỗi chuyển động nhanh *Coastguard*, đặc trưng SAD có nhiều sự thay đổi, lược đồ xám có hình dạng phức tạp hơn và trường véc-tơ chuyển động nhiều gần như trên toàn bộ khung hình. Điều đó cho thấy mỗi đặc trưng đều đóng vai trò mô tả khá chính xác nội dung chuyển động và kết cấu của mỗi chuỗi video.

Các đặc trưng nhằm phát hiện các thay đổi về chuyển động bao gồm:

- Sự sai khác của lược đồ xám (DoH).



Hình 2.4: Một số đặc trưng của chuỗi Suzie



Hình 2.5: Một số đặc trưng của chuỗi Coastguard

- Trung bình của các vector chuyển động (AMV)
- Số lượng các vector chuyển động (NMV)
- Phương sai trung bình của phân đoạn (ASV)
- Giá trị trung bình của phân đoạn (ASM)

Các đặc trưng phản ánh cấu trúc bề mặt của chuỗi video bao gồm:

- Tổng các sai số tuyệt đối (SAD)
- Phương sai giá trị DC (DCV)
- Trung bình giá trị DC (DCM)
- Phương sai giá trị AC (ACV)
- Trung bình giá trị AC (ACM)

Công thức xác định các đặc trưng như sau:

$$DoH = \frac{1}{N} \sum_{k=1}^{N-1} \left( \frac{1}{H.W} \sum_{i=0}^{L-1} |h(KF_{k+1}(i)) - h(KF_{k-1}(i))| \right) \quad (2.3)$$

$$AMV = \frac{1}{N-1} \sum_{k=1}^{N-1} MV(KF_{k+1}, KF_{k-1}) \quad (2.4)$$

$$NMV = \frac{1}{N-1} \sum_{k=1}^{N-1} NMV(KF_{k+1}, KF_{k-1}) \quad (2.5)$$

$$ASV = \frac{1}{N} \sum_{k=1}^N \sigma^2(KF_k) \quad (2.6)$$

$$ASM = \frac{1}{N} \sum_{k=1}^N \left( \frac{1}{H.W} \sum_{x=1}^H \sum_{y=1}^W KF_k(x, y) \right) \quad (2.7)$$

$$SAD = \frac{1}{N-1} \sum_{k=1}^{N-1} \left( \sum_{x=1}^H \sum_{y=1}^W |KF_{k+1}(x, y) - KF_{k-1}(x, y)| \right) \quad (2.8)$$

$$DCV = \sigma_{DC}^2 \quad (2.9)$$

$$DCM = \frac{1}{N} \sum_{k=1}^N DC(KF_k) \quad (2.10)$$

$$ACV = \frac{1}{N} \sum_{k=1}^N \sigma_{AC}^2(KF_k) \quad (2.11)$$

$$ACM = \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^{H.W-1} AC_i(KF_k) \quad (2.12)$$

ở đó:

- $KF_k$ : KF thứ  $k$ .
- $N$ : số lượng KF trong phân đoạn 5 khung hình ( $N = 3$ ).
- $H, W$  là kích thước của khung hình.
- $h(KF_k(i))$ : lược đồ xám của KF thứ  $k$  tại mức xám  $i$ .
- $MV(KF_{k+1}, KF_{k-1})$ : chiều dài tổng của vector chuyển động giữa hai KF.
- $NMV(KF_{k+1}, KF_{k-1})$ : số lượng vector chuyển động giữa hai KF.
- $\sigma^2$ : giá trị phương sai.
- $DC(KF_k)$ : thành phần DC của KF thứ  $k$ .
- $AC_i(KF_k)$ : thành phần AC thứ  $i$  của KF thứ  $k$ .

## 2) Quá trình huấn luyện

Tập dữ liệu sử dụng cho quá trình huấn luyện của đề xuất này được trích xuất từ năm chuỗi video *Foreman*, *Soccer*, *Hall Monitor*, *Akiyo*, *Carphone* và *News* với các đặc tính chuyển động đa dạng. Tập dữ liệu gồm 352 phân đoạn video, mỗi phân đoạn gồm 5 khung hình.

Nhãn GOP2 và GOP4 tương ứng cho mỗi phân đoạn được xác định như sau:



- Với mỗi phân đoạn 5 khung hình thực hiện mã hóa DVC với hai kích thước GOP=2 và GOP=4.
- Tính BD-PSNR tương ứng với hai cách mã hóa.
- Kích thước GOP nào tương ứng BD-PSNR lớn hơn sẽ được chọn là nhãn tương ứng cho phân đoạn đó.

Các đặc trưng cùng với nhãn tương ứng của 352 phân đoạn được đưa vào để huấn luyện bằng thuật toán tiêu biểu của cây quyết định là C4.5 [46].

### 3) Áp dụng vào quá trình lựa chọn kích thước GOP

Sau quá trình huấn luyện, mô hình cây quyết định xây dựng ở trên sẽ được sử dụng để lựa chọn kích thước GOP. Chuỗi video đầu vào sẽ được chia thành các phân đoạn gồm 5 khung hình và các đặc trưng tương ứng với mỗi phân đoạn được trích xuất. Dựa trên các đặc trưng này, sử dụng mô hình cây quyết định đã được huấn luyện để chọn ra nhãn tương ứng (GOP2 hay GOP4) cho mỗi phân đoạn. Sau khi lựa chọn được kích thước GOP phù hợp sẽ thực hiện chia tách các khung hình thành WZF và KF tùy theo kích thước GOP.

## 2.2.3 Đánh giá hiệu năng của phương pháp đề xuất AGOP-DVC

Phương pháp thay đổi kích thước GOP được đánh giá trong ngữ cảnh áp dụng vào kiến trúc mã hóa video phân tán AGOP-DVC. Hiệu năng được đánh giá ở đây là hiệu năng RD của AGOP-DVC với hai tham số PSNR (dB) và Bitrate (bps). Hai đại lượng BD-PSNR và BD-Rate [13] cũng được sử dụng trong các đánh giá.

Đại lượng BD-PSNR cho biết mức cải thiện tương đối giữa hai phương pháp bằng cách đo lường sự sai khác trung bình giữa hai đường cong RD trong đó một đường cong RD được chọn làm đường cong cơ sở. Nếu BD-PSNR dương, điều đó có nghĩa là đường cong thứ hai tốt hơn đường cong cơ sở và ngược lại.

Bảng 2.1: Điều kiện thử nghiệm

Chuỗi video	Độ phân giải không gian	Số lượng khung hình	Tham số lượng tử
Coastguard	176x144	300	{25,29,34,40}
Suzie		150	{25,29,34,40}
Pamphlet		150	{25,29,34,40}
Harbour		150	{25,31,36,44}

Tương tự như vậy, đại lượng BD-Rate cho biết mức độ tiết kiệm tốc độ bit giữa hai phương pháp. Nếu BD-Rate âm, điều đó có nghĩa là đường cong thứ hai tốt hơn đường cong cơ sở và ngược lại.

Bộ mã hóa tham chiếu được sử dụng để so sánh ở đây dựa trên kiến trúc DISCOVER. Đây là bộ mã hóa video phân tán DISCOVER [9] với KF được mã hóa bằng chuẩn mã hóa HEVC. Bộ mã hóa này khi sử dụng kích thước cố định  $GOP=2$  và  $GOP=4$  có tên gọi lần lượt là GOP2-DVC và GOP4-DVC. Hai bộ mã hóa này sẽ được so sánh với bộ mã hóa đề xuất AGOP-DVC trên khía cạnh hiệu năng RD.

### A. Điều kiện thử nghiệm

Việc đánh giá hiệu năng nén được thực hiện trên bốn chuỗi video thử nghiệm *Coastguard*, *Suzie*, *Pamphlet* và *Harbour*. Các chuỗi này được lựa chọn vì tính đa dạng về chuyển động và kết cấu trong mỗi chuỗi. Bảng 2.1 mô tả các đặc điểm chính của các chuỗi video này. Hình 2.6 mô tả các khung hình đầu tiên của các chuỗi video này.

### B. Đánh giá hiệu năng RD

Hiệu năng RD được đo bằng tốc độ bit (bps) và PSNR (dB) đối với bốn chuỗi video được trình bày trong Bảng 2.2. Như đã dự đoán trước, quan sát cho thấy các giá trị PSNR của phương pháp đề xuất AGOP-DVC tốt hơn các giá trị PSNR của GOP4-DVC và xấp xỉ các giá trị PSNR của GOP2-DVC. Các giá trị bitrate của phương pháp đề xuất cao hơn các giá trị của GOP4-DVC và thấp



Hình 2.6: Các khung hình đầu tiên của các chuỗi video thử nghiệm

hơn các giá trị của GOP2-DVC. Nói cách khác, các kết quả cho thấy chất lượng của các khung hình video của phương pháp đề xuất giảm không đáng kể (theo khía cạnh PSNR) trong khi mức tiết kiệm bitrate lại khá cao. Bảng 2.3 cho thấy mức tiết kiệm tốc độ bit của phương pháp đề xuất trung bình là 3,37 % và 9,62 % khi so với GOP2-DVC và GOP4-DVC.

Bảng 2.2: Hiệu năng RD của các chuỗi video thử nghiệm

Chuỗi video	QP	GOP2-DVC		GOP4-DVC		AGOP-DVC	
		Bitrate	PSNR	Bitrate	PSNR	Bitrate	PSNR
Coastguard	26	27760	38,18	28242	34,65	27735	38,14
	30	17131	34,87	16140	32,48	17058	34,84
	34	9838	31,88	8228	30,36	9760	31,85
	38	5256	29,14	3781	28,23	5199	29,12
<b>Trung bình</b>		<b>14996,25</b>	<b>33,52</b>	<b>14097,75</b>	<b>31,43</b>	<b>14938</b>	<b>33,49</b>
Suzie	26	18424	41,58	19719	41,26	18565	41,34
	30	10869	38,56	11172	38,23	10530	38,26
	34	5725	35,41	5588	35,15	5283	35,29
	38	2667	32,24	2353	32,04	2270	32,19
<b>Trung bình</b>		<b>9421,25</b>	<b>36,95</b>	<b>9708,00</b>	<b>36,67</b>	<b>9162,00</b>	<b>36,77</b>
Pamphlet	26	23893,93	41,15	23128,28	41,35	22453,65	41,37
	30	15669,90	37,42	14900,70	37,51	14504,50	37,56
	34	9013,55	33,18	8567,73	33,24	8349,78	33,29
	38	3897,73	28,86	3667,88	28,91	3587,02	28,95
<b>Trung bình</b>		<b>13118,78</b>	<b>35,15</b>	<b>12566,15</b>	<b>35,25</b>	<b>12223,74</b>	<b>35,29</b>
Harbour	26	45656,58	38,04	45680,28	37,62	45337,92	37,81
	30	29713,93	34,18	28617,86	33,73	28830,11	33,96
	34	16805,14	30,36	15471,99	30,03	15889,86	30,23
	38	7646,22	26,24	6768,94	26,09	7082,92	26,22
<b>Trung bình</b>		<b>24955,47</b>	<b>32,20</b>	<b>24134,77</b>	<b>31,86</b>	<b>24285,20</b>	<b>32,06</b>

Bảng 2.3: Mức tiết kiệm BD-rate [%]

Chuỗi video	AGOP-DVC vs. GOP2-DVC	AGOP-DVC vs. GOP4-DVC
Coastguard	-0,04	-26,24
Suzie	-2,28	-7,52
Pamphlet	-9,04	-3,26
Harbour	-2,12	-1,48
<b>Trung bình</b>	<b>-3,37</b>	<b>-9,62</b>

## 2.3 Đề xuất 2: Cải tiến hiệu năng nén cho DVC tại phía giải mã

### 2.3.1 Đặt vấn đề

Trong DVC, SIF tương ứng với WZF gốc được tạo ra tại phía giải mã bằng cách sử dụng các KF đã giải mã. Vì vậy, hiệu năng nén của DVC phụ thuộc rất nhiều các mô-đun tại bộ giải mã như mô-đun tạo SI và mô hình nhiễu tương quan.

Tạo SI là phần trung tâm của DVC nên có rất nhiều các đề xuất liên quan đến việc cải thiện SI. DISCOVER sử dụng kỹ thuật ước lượng thông tin phụ trợ cải tiến. Kỹ thuật này bao gồm ước lượng chuyển động, nội suy chuyển động, làm mịn chuyển động không gian và bù chuyển động. Trong giải pháp tạo SI này cũng như rất nhiều các giải pháp tạo SI về sau, ví dụ [32] thường cố gắng để tìm được ước lượng tốt nhất về WZF gốc trước khi bắt đầu quá trình giải mã. Và trong quá trình giải mã tất cả các mặt phẳng bit của tất cả các hệ số DCT, SI ban đầu này được giữ nguyên không thay đổi.

Tuy nhiên, trong codec DISCOVER cũng như các codec theo hướng này, lần lượt các mặt phẳng bit rồi lần lượt các hệ số được giải mã hay nói cách khác, SI được “sửa sai” dần dần. Trong trường hợp đó, ta sẽ có thêm nhiều thông tin

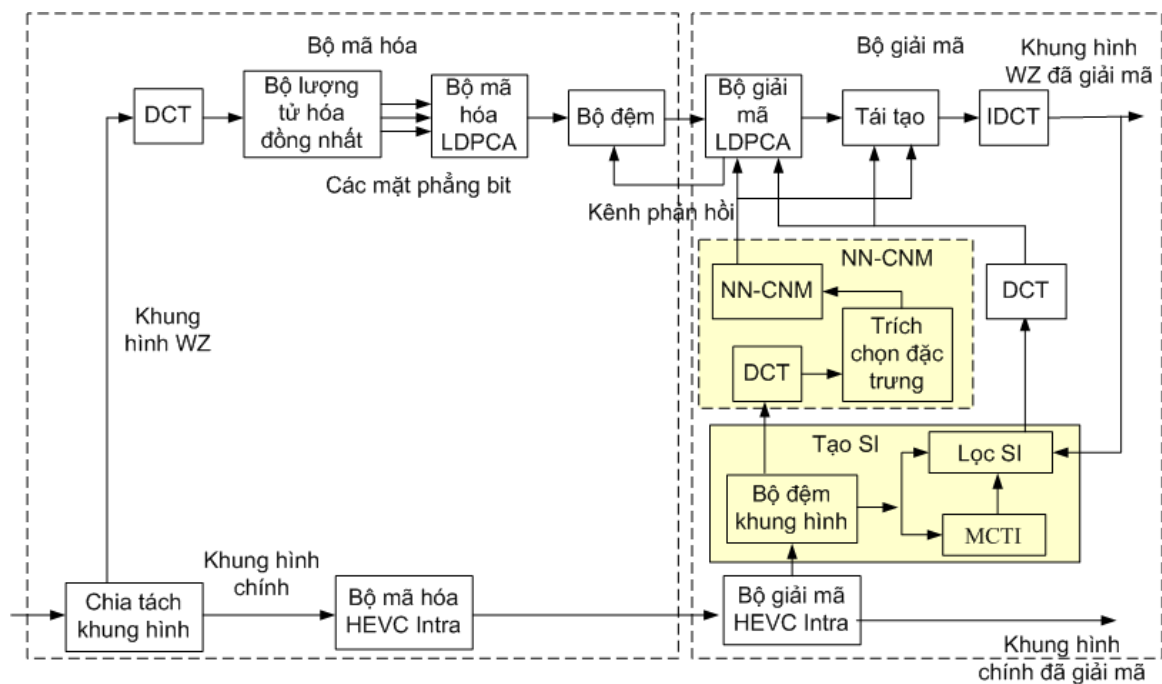
mới tại bộ giải mã trong suốt quá trình giải mã. Các thông tin này có thể được khai thác để cải thiện thông tin phụ trợ để giải mã cho các hệ số còn lại. Liên quan đến hướng nghiên cứu này, có một số các đề xuất để tiếp tục cải thiện chất lượng của SI khi nhận thêm nhiều thông tin mới trong quá trình giải mã [35][4]. Trong [35], để việc tìm kiếm nhanh, khung hình được tái tạo mới được lọc trên SI ban đầu. Còn trong [4], việc lọc được thực hiện trên các KF trước và sau. Để cải thiện hơn nữa chất lượng của SI, trong đề xuất này các khung hình mới tái tạo một phần được lọc trên cả ba khung hình: SIF ban đầu và các KF trước và sau. Sau quá trình lọc, sử dụng cơ chế kết hợp để xây dựng SI mới có chất lượng tốt hơn và sử dụng SI mới này để giải mã cho các mặt phẳng bit kế tiếp

Cùng với SI, CNM cũng đóng vai trò quan trọng đối với hiệu năng nén của DVC. Trong các nghiên cứu trước đây, CNM thường được mô hình hóa với phân bố Laplace vì phân bố này cân bằng giữa tính chính xác mô hình và độ phức tạp tính toán. Tuy nhiên, tham số phân bố  $\alpha$  của phân bố Laplace như trong biểu thức 2.1 thường được ước lượng từ các KF đã giải mã trước đó. Các tham số  $\alpha$  này có thể rất khác so với giá trị  $\alpha$  thực được tính từ WZF và SIF. Với mục tiêu ước lượng được tham số  $\alpha$  gần với giá trị thực hơn nữa, luận án này sử dụng mạng nơ-ron để ước lượng giá trị  $\alpha$  của mô hình hóa nhiều tương quan.

Chi tiết của các cải tiến mới về SI và CNM tại bộ giải mã được trình bày chi tiết trong mục 2.3.2 và các đánh giá được giới thiệu trong mục 2.3.3.

### 2.3.2 Kiến trúc bộ mã hóa đề xuất Adv-DVC

Sơ đồ kiến trúc mã hóa DVC đề xuất (Adv-DVC) với hai công cụ mã hóa mới được tô đậm là tạo thông tin phụ trợ và mô hình hóa nhiều tương quan được mô tả trong Hình 2.7. Nguyên lý hoạt động của kiến trúc này cũng tương tự như kiến trúc DVC được giới thiệu trong Hình 2.3 với một số thay đổi như sau: 1) kích thước GOP=2; 2) Mô-đun Tạo SI tạo thông tin phụ trợ bằng phương pháp lọc liên tục trong quá trình giải mã; và 3) Mô-đun NN-CNM ước lượng tham số của CNM dựa trên mạng nơ-ron.



Hình 2.7: Kiến trúc mã hóa DVC đề xuất Adv-DVC

### 2.3.3 Các mô-đun mã hóa đề xuất mới

Trong phần này giới thiệu hai mô-đun mã hóa mới thực hiện tại bộ giải mã và được sử dụng trong kiến trúc DVC đề xuất bao gồm:

- Tạo thông tin phụ trợ bằng cách lọc liên tục trong suốt quá trình giải mã.
- Ước lượng tham số  $\alpha$  cho mô hình nhiễu tương quan dựa trên mạng nơ-ron.

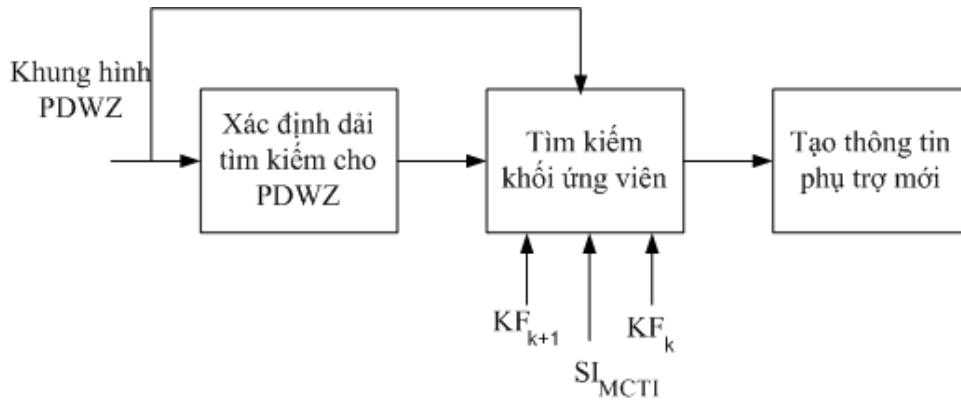
#### A. Tạo thông tin phụ trợ bằng cách lọc liên tục

Trong mô-đun tạo SI này, SIF ban đầu được tạo ra sử dụng kỹ thuật MCTI sử dụng các KF đã giải mã trước đó gồm  $KF_{k+1}$ ,  $KF_{k-1}$ . SIF ban đầu này được gọi là  $SI_{MCTI}$ .

Mỗi khi một mặt phẳng bit được giải mã LDPCA thành công, chúng được đưa tới mô-đun tái tạo để khôi phục lại hệ số DCT gốc với sự trợ giúp của SIF tương ứng. Các hệ số DCT đã tái tạo sẽ thay thế cho các hệ số cùng vị trí trong SIF trước đó để tạo ra SIF mới được gọi là WZF được giải mã một phần

(PDWZ).

Quá trình tạo SI bằng cách lọc liên tục được mô tả trong Hình 2.8 dưới đây.



Hình 2.8: Quá trình tạo SI

### 1) Xác định dải tìm kiếm cho PDWZ

Sau khi giải mã xong một mặt phẳng bit, khung hình PDWZ mới được tạo ra. Khung hình PDWZ này được ước lượng chuyển động trong các vùng tìm kiếm khác nhau tùy thuộc vào nội dung chuyển động của khung hình đó. Trong đề xuất này, mức độ chuyển động của khung hình được đánh giá dựa trên tham số biên độ vector chuyển động trung bình.

$$MV_{avg} = \frac{1}{n} \sum_{i=1}^n \|\overrightarrow{MV}_i\| \quad (2.13)$$

với  $\overrightarrow{MV}_i$  là vector chuyển động trong PDWZ và  $n$  là số lượng các vector chuyển động.

Dải tìm kiếm ( $SR$ ) được xác định cho mỗi PDWZ phụ thuộc vào điều kiện mà  $\overrightarrow{MV}_i$  thỏa mãn trong công thức 2.14.

$$SR = \begin{cases} 14 \times 14 & MV_{avg} < T_1 \\ 16 \times 16 & T_1 \leq MV_{avg} < T_2 \\ 18 \times 18 & MV_{avg} \geq T_2 \end{cases} \quad (2.14)$$

ở đó  $T_1, T_2$  là các ngưỡng. Trong trường hợp này,  $T_1, T_2$  được chọn thực nghiệm bằng 5 và 10 tương ứng.



### 2) Tìm kiếm khối ứng viên

Khung hình PDWZ sẽ được chia thành các khối kích thước  $4 \times 4$  và việc tìm kiếm ứng viên sẽ được thực hiện cho từng khối trên ba khung hình tham chiếu (RF) gồm  $KF_{k+1}$ ,  $KF_{k-1}$  và  $SI_{MCTI}$ .

Gọi  $bPDWZ_n$  là khối  $n$  trong PDWZ và  $bRF_k^{d(k)}$  là khối  $k$  trong RF với khoảng dịch chuyển  $d(k)$  so với khối  $bPDWZ_n$ .

Với mỗi khối ứng viên  $bRF_k^{d(k)}$  sẽ thực hiện tính tổng các trị tuyệt đối sai số (MAD) so với  $bPDWZ_n$  như trong biểu thức 2.15. Khối nào cho MAD nhỏ nhất sẽ được coi là giống với khối  $bPDWZ_n$  nhất.

$$MAD = \alpha_n(RF) = \sum_{x=0}^3 \sum_{y=0}^3 |bRF_k^{d(k)}(x, y) - bPDWZ_n(x, y)| \quad (2.15)$$

với  $(x, y)$  là tọa độ pixel trong mỗi khối.

Giả sử sau bước tìm kiếm tìm ra ba khối ứng viên tương ứng với  $bPDWZ_n$  trong ba khung hình tham chiếu là  $bKF_{k+1}$ ,  $bKF_{k-1}$ ,  $bSI_{MCTI}$  với giá trị MAD tương ứng là  $\alpha_n(KF_{k+1})$ ,  $\alpha_n(KF_{k-1})$ ,  $\alpha_n(SI_{MCTI})$ . Các ứng viên này được đưa vào bước kết hợp để tạo ra khối SI mới.

### 3) Kết hợp các khối ứng viên

Sau bước 2) ở trên nhận được ba khối ứng viên trong ba khung hình tham chiếu tương ứng với ba giá trị MAD. Theo lẽ tự nhiên, sai số đối sánh càng nhỏ chứng tỏ khối đó có khả năng là một ứng viên tốt và có thể đóng vai trò lớn hơn trong việc tạo SI cuối cùng và ngược lại. Vì vậy, trọng số  $\beta_n(RF) = 1/\alpha_n(RF)$  được sử dụng để để xác định đóng góp của mỗi khối ứng viên trong cơ chế kết hợp. Khối SI mới được tính theo biểu thức 2.16 dưới đây:

$$bSI_n = \frac{bKF_{k+1} \cdot \beta_n(KF_{k+1}) + bKF_{k-1} \cdot \beta_n(KF_{k-1}) + bSI_{MCTI} \cdot \beta_n(SI_{MCTI})}{\beta_n(KF_{k+1}) + \beta_n(KF_{k-1}) + \beta_n(SI_{MCTI})} \quad (2.16)$$

SIF mới được tạo ra lại được sử dụng để giải mã cho mặt phẳng bit kế tiếp. Khi tất cả các dải hệ số DCT được giải mã, SI cuối cùng sẽ được tạo ra để thực hiện tái tạo lần cuối cùng để nhận được WZF gốc.

## B. Ước lượng tham số $\alpha$ cho mô hình nhiễu tương quan

Tham số  $\alpha$  của mô hình nhiều tương quan nếu được ước lượng đúng phải sử dụng WZF gốc tại phía mã hóa và SIF tại phía giải mã. Tuy nhiên, tại phía giải mã không sẵn có WZF nên nhiều tương quan được xấp xỉ bằng cách sử dụng các KF đã giải mã và được bù chuyển động:

$$R_k(x, y) = KF'_{k+1}(x, y) - KF'_{k-1}(x, y) \quad (2.17)$$

Tuy nhiên không như các nghiên cứu khác, tham số  $\alpha$  được ước lượng trực tiếp từ  $R_k(x, y)$ , trong luận án này sử dụng mạng nơ-ron để ước lượng tham số  $\alpha$  dựa trên các đặc trưng của  $R_k(x, y)$ .

Giá trị  $\alpha$  được ước lượng ở mức dải DCT, do đó các đặc trưng được trích xuất trong miền DCT.

### 1) Xác định đặc trưng

Các đặc trưng của mỗi dải DCT được tính như sau:

- **Bước 1:** Biến đổi khung hình dư thừa  $R_k(x, y)$  sang miền DCT:

$$T_k(u, v) = DCT [R_k(x, y)] \quad (2.18)$$

Các hệ số DCT của khung hình  $T_k(u, v)$  được nhóm thành 16 dải trong đó  $T_{k,0}$  gồm các hệ số DC và  $T_{k,b}$  ( $b = \overline{1, 15}$ ) là các dải chứa các hệ số AC.

- **Bước 2:** Trong mỗi dải, bốn đặc trưng *Min*, *Max*, *Trung bình*, *Phương sai* được tính như sau:

$$\begin{aligned} X_1 &= \text{Min} \{T_{k,b}(i)\} \\ X_2 &= \text{Max} \{T_{k,b}(i)\} \\ X_3 &= \frac{1}{N} \sum_{i=1}^N T_{k,b}(i) \\ X_4 &= \frac{1}{N} \sum_{i=1}^N T_{k,b}^2(i) - \left( \frac{1}{N} \sum_{i=1}^N T_{k,b}(i) \right)^2 \end{aligned} \quad (2.19)$$

với  $N$  là số lượng các hệ số DCT trong mỗi dải DCT.  $T_{k,b}(i)$  là hệ số thứ  $i$  của dải DCT  $b$  trong khung hình  $k$ .

### 2) Xây dựng tập dữ liệu huấn luyện

Để xây dựng tập dữ liệu huấn luyện cho mạng nơ-ron, luận án sử dụng 10 chuỗi video *Coastguard, Hall-Monitor, News, Container, Flower Garden, Mobile, Mother, Claire, Grandma, Harbour* với độ phân giải không gian  $176 \times 144$  (QCIF) và số lượng các khung hình cho mỗi chuỗi là 300. Lý do để lựa chọn các chuỗi video này để huấn luyện là sự đa dạng của các đặc tính chuyển động của chuỗi. Các chuỗi này được mã hóa và giải mã HEVC Intra với bốn tham số lượng tử và xây dựng được 1490 khung hình dư thừa.

Đối với mỗi dải DCT sẽ có 1490 mẫu tương ứng, mỗi mẫu có 4 đặc trưng. Đầu ra tương ứng, hay giá trị  $\alpha$  tham chiếu, được gán cho mỗi mẫu được tính dựa trên khung hình dư thừa đúng theo các bước sau:

- **Bước 1:** Tính khung hình dư thừa tham chiếu:

$$\bar{R}_k(x, y) = WZF(x, y) - SIF(x, y) \quad (2.20)$$

- **Bước 2:** Biến đổi khung hình  $\bar{R}_k(x, y)$  sang miền DCT như biểu thức 2.18 với  $R_k(x, y)$  được thay bằng  $\bar{R}_k(x, y)$ .
- **Bước 3:** Tính phương sai trung bình  $\bar{\sigma}_{k,b}$ . Phương sai trung bình  $\bar{\sigma}_{k,b}$  cho dải  $b$  của khung hình thứ  $k$  được tính như trong công thức 2.21:

$$\bar{\sigma}_{k,b}^2 = \frac{1}{N} \sum_{i=1}^N \bar{T}_{k,b}^2(i) - \left( \frac{1}{N} \sum_{i=1}^N \bar{T}_{k,b}(i) \right)^2 \quad (2.21)$$

với chiều dài của dải DCT  $N = 1584$ .

- **Bước 4:** Tính giá trị  $\alpha$  hay đầu ra tham chiếu.

$$\bar{\alpha}_{k,b} = \sqrt{\frac{2}{\bar{\sigma}_{k,b}^2}} \quad (2.22)$$

### 3) Kiến trúc mạng nơ-ron

Bảng 2.4: So sánh các cấu trúc mạng nơ-ron

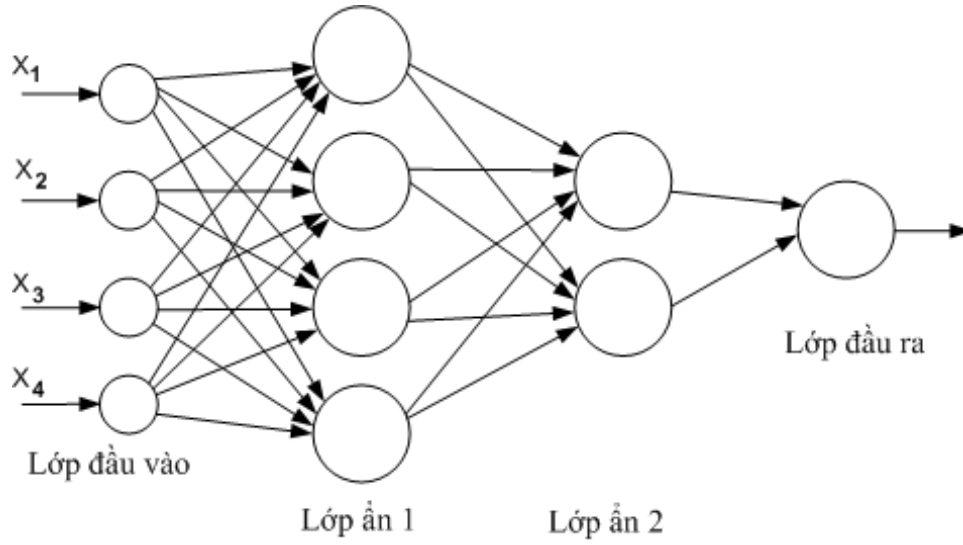
Số lớp	Cấu trúc mạng nơ-ron	Sai số tuyệt đối trung bình (MAE)
3	{4,4,1}	0,20
4	{4,4,4,1}	0,18
5	{4,4,4,4,1}	0,17
6	{4,4,4,4,4,1}	0,08
6	{4,4,4,3,2,1}	0,08
5	{4,4,3,2,1}	0,12
4	{4,4,3,1}	0,10
<b>4</b>	<b>{4,4,2,1}</b>	<b>0,04</b>
3	{4,3,1}	0,24
3	{4,2,1}	0,35

Để huấn luyện cho tập dữ liệu xây dựng ở trên, trong nghiên cứu này sử dụng một mạng nơ-ron nhiều lớp. Để quyết định cấu trúc mạng nơ-ron với số lớp ẩn và số nơ-ron trong mỗi lớp, nhiều cấu trúc mạng nơ-ron khác nhau đã được thử nghiệm. Cấu trúc nào có độ chính xác cao nhất sẽ được sử dụng cho thuật toán. Bảng 2.4 chỉ ra kết quả thử nghiệm của một số cấu trúc mạng. Trong bảng này, cột cấu trúc mạng được ký hiệu  $\{x_1, x_2, \dots, x_n\}$  tương ứng với cấu trúc mạng nơ-ron gồm  $n$  lớp trong đó lớp 1 là lớp đầu vào gồm  $x_1$  nơ-ron, lớp  $n$  là lớp đầu ra gồm  $x_n$  nơ-ron, các lớp ẩn có số nơ-ron tương ứng là  $x_2, \dots, x_{n-1}$ .

Sai số tuyệt đối trung bình (MAE) là sự sai khác giữa giá trị  $\alpha$  tham chiếu trong công thức 2.22 và giá trị  $\alpha$  dự đoán ở đầu ra mạng nơ-ron. Nếu MAE càng nhỏ nghĩa là dự đoán càng chính xác.

Dựa trên độ chính xác của các mô hình, cấu trúc mạng nơ-ron 4,4,2,1 được chọn. Mạng nơ-ron này gồm có một lớp đầu vào, hai lớp ẩn và một lớp đầu ra với tên gọi NN-CNM như mô tả trong Hình 2.9.

Lớp đầu vào với bốn giá trị  $X_1, X_2, X_3, X_4$  tương ứng là bốn đặc trưng Min, Max, Trung bình và Phương sai của các dải DCT của khung hình dư thừa. Tất



Hình 2.9: Kiến trúc NN-CNM

cả các lớp trong mạng được kết nối đầy đủ. Trong các lớp ẩn 1 và 2, hàm kích hoạt ReLU được sử dụng. Giả sử rằng  $X$  là tập gồm bốn đầu vào,  $\widehat{Y}_{i,k}$  là đầu ra tại nơ-ron thứ  $k$  của lớp thứ  $i$ . Đầu ra tại một nơ-ron được tính như sau:

$$\widehat{Y}_{1,k} = g(W_{1,k} * X + B_1), \quad k = \overline{1,4} \quad (2.23)$$

$$\widehat{Y}_{2,k} = g(W_{2,k} * \widehat{Y}_1 + B_2), \quad k = \overline{1,2} \quad (2.24)$$

ở đó  $W_{i,k}$  và  $B_i$  là các ma trận trọng số và các hệ số bias của lớp  $i$ ,  $\widehat{Y}_1$  là ma trận của đầu ra tại lớp ẩn 1.  $g(\cdot)$  là hàm kích hoạt ReLU và được xác định như sau:

$$g(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases} \quad (2.25)$$

Tại lớp đầu ra, hàm kích hoạt tuyến tính được sử dụng để dự đoán ra giá trị  $\widehat{\alpha}$  cho dải hệ số DCT như sau:

$$\widehat{\alpha} = W_3 * \widehat{Y}_2 + B_3 \quad (2.26)$$

ở đó  $W_3$  là ma trận trọng số và  $\widehat{Y}_2$  là ma trận đầu ra tại lớp ẩn 2.

#### 4) Huấn luyện mô hình

Tập dữ liệu chuẩn bị ở trên được chia thành tập huấn luyện và tập kiểm tra theo tỷ lệ 70/30 để huấn luyện cho mạng nơ-ron 2 lớp ẩn. Kết quả của quá trình huấn luyện là tập các trọng số tương ứng.

#### 5) Áp dụng mô hình đã huấn luyện cho bộ mã hóa DVC

Tại phía bộ giải mã DVC, các KF đã giải mã được sử dụng để tạo ra khung hình nhiều tương quan như biểu thức 2.17. Các đặc trưng tương ứng cho mỗi dải DCT của khung hình này được trích xuất như các công thức 2.18 và 2.19. Sử dụng tập các trọng số của mô hình huấn luyện để áp dụng cho tập đặc trưng này để dự đoán giá trị đầu ra  $\hat{\alpha}$ .

### 2.3.4 Đánh giá hiệu năng

Phần này giới thiệu và phân tích các kết quả về hiệu năng nén của các công cụ mã hóa đề xuất so với các kiến trúc mã hóa video có liên quan.

#### A. Điều kiện thử nghiệm

Đánh giá hiệu năng được thực hiện trên bốn chuỗi video *Akiyo*, *Carphone*, *Foreman* và *Soccer*. Các chuỗi này được lựa chọn vì tính đa dạng về nội dung chuyển động và các cấu trúc bề mặt.



Hình 2.10: Các khung hình đầu tiên của các chuỗi video thử nghiệm

Trong phần đánh giá này sử dụng các bộ mã hóa sau:

- **HEVC Intra:** Bộ mã hóa này sử dụng phần mềm tham chiếu HM với chế độ mã hóa Intra. Đây là bộ mã hóa đại diện cho mã hóa video dự đoán được lựa chọn để so sánh vì có độ phức tạp bộ mã hóa thấp.

Bảng 2.5: Điều kiện thử nghiệm

Chuỗi video	Độ phân giải không gian	Số lượng khung hình	Tham số lượng tử
Akiyo	176x144	300	{25,29,34,40}
Carphone		300	{25,29,34,40}
Foreman		300	{25,29,34,40}
Soccer		300	{25,31,36,44}

- **DISCOVER-HEVC**: Đây là kiến trúc DISCOVER với khung hình chính được mã hóa bởi HEVC Intra thay vì H.264/AVC Intra.
- **ReSI-DVC**: Bộ mã hóa này là bộ mã hóa DISCOVER-HEVC với mô-đun tạo SI cải tiến được đề xuất trong phần A của mục 2.3.3.
- **Adv-DVC**: Đây là bộ mã hóa được xây dựng trên DISCOVER-HEVC với mô-đun tạo SI cải tiến và mô-đun NN-CNM mới như mô tả trong hình 2.7.

Phần tiếp theo sẽ đánh giá chất lượng SI được tạo ra theo phương pháp lọc liên tục, tiếp đó là độ chính xác của phương pháp ước lượng tham số  $\alpha$  đề xuất. Cuối cùng là đánh giá về hiệu năng nén tổng thể của kiến trúc DVC đề xuất khi thực hiện đồng thời cả hai cải tiến về tạo SI và CNM so với các bộ mã hóa khác có liên quan.

## B. Chất lượng của SI theo phương pháp đề xuất

Phần này đánh giá chất lượng của SI theo phương pháp lọc đề xuất bằng cách so sánh với giải pháp tạo SI được sử dụng nhiều trong kiến trúc DVC là MCTI.

Bảng 2.6 mô tả kết quả về chất lượng SI đạt được bằng phương pháp đề xuất ReSI và mức cải tiến so với phương pháp MCTI trước đó. Cột ReSI mô tả giá trị PSNR [dB] của SI được tạo ra bởi phương pháp lọc liên tục và cột ReSI so với MCTI mô tả mức tăng PSNR [dB] của phương pháp tạo SI bằng cách lọc liên tục so với phương pháp MCTI. Có thể rút ra một số nhận xét như sau:

Bảng 2.6: PSNR [dB] của phương pháp ReSI so với phương pháp MCTI

Chuỗi Video	QP	ReSI	ReSI so với MCTI
Akiyo	25	43,68	4,05
	29	39,82	4,18
	34	35,25	3,31
	40	30,66	2,54
Mức cải thiện SI trung bình			<b>3,52</b>
Foreman	25	40,97	0,28
	29	37,47	0,26
	34	33,75	0,19
	40	29,72	0,11
Mức cải thiện SI trung bình			<b>0,21</b>
Carphone	25	39,62	2,16
	29	36,22	2,44
	34	32,77	2,43
	40	28,98	2,43
Mức cải thiện SI trung bình			<b>2,37</b>
Soccer	25	39,93	0,25
	31	35,94	0,23
	36	32,68	0,21
	44	28,85	0,12
Mức cải thiện SI trung bình			<b>0,20</b>
Mức cải thiện SI trung bình tổng cộng			<b>1,51</b>



- Phương pháp tạo SI đề xuất ReSI đạt được chất lượng SI tốt hơn so với MCTI cho tất cả các tham số lượng tử của các chuỗi video thử nghiệm.
- Mức cải tiến không đồng đều giữa các chuỗi có nội dung chuyển động khác nhau. Mức cải tiến lớn nhất đạt được đối với chuỗi chuyển động chậm *Akiyo* và mức cải thiện ít hơn đối với các chuỗi chuyển động nhanh như *Soccer*. Đó là bởi vì với các chuỗi chuyển động nhanh, khó tạo ra SI chất lượng tốt vì sự sai khác giữa các KF là rất lớn so với trường hợp chuỗi chuyển động chậm.
- Mặc dù phương pháp đề xuất cho chất lượng SI tốt hơn so với MCTI cho cả bốn tham số lượng tử nhưng kết quả thường tốt hơn đối với các QP thấp. Đó là bởi vì với QP thấp, các KF thường có chất lượng tốt hơn và điều đó giúp tạo ra các SI có chất lượng tốt hơn.

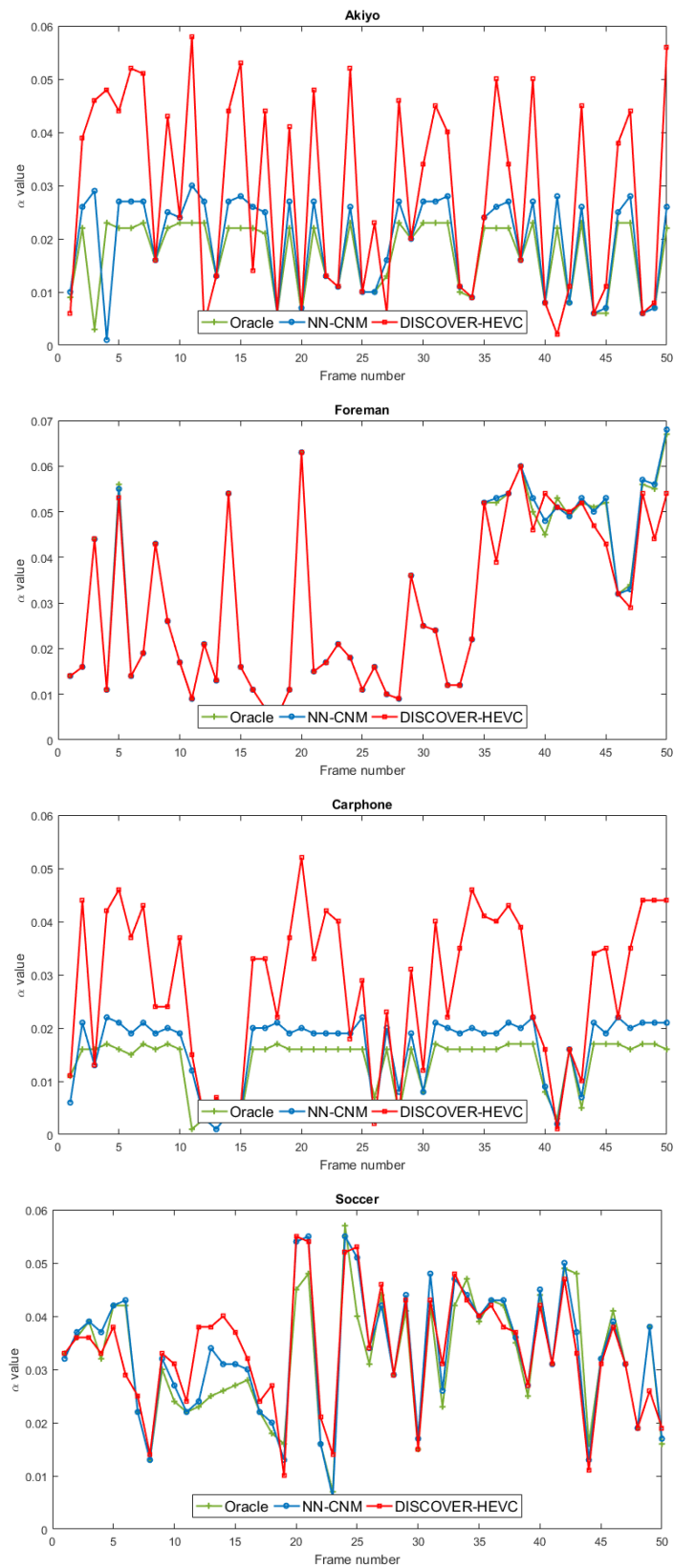
### C. Độ chính xác của phương pháp ước lượng tham số $\alpha$ đề xuất

Trong tiểu mục này, tham số  $\alpha$  được ước lượng bởi phương pháp NN-CNM được so sánh với tham số  $\alpha$  được tính như trong DISCOVER-HEVC. Quy ước tên gọi các phương pháp như sau:

- **NN-CNM**: giá trị  $\alpha$  được ước lượng bằng mạng nơ-ron.
- **DISCOVER-HEVC**: giá trị  $\alpha$  được ước lượng như trong kiến trúc DISCOVER-HEVC.
- **Oracle**: giá trị  $\alpha$  tham chiếu hay giá trị  $\alpha$  được tính dựa trên nhiều tương quan thực như trong biểu thức 2.25.

Nếu tham số ước lượng nào (NN-CNM hoặc DISCOVER-HEVC) càng gần giá trị tham chiếu (Oracle), việc ước lượng được coi là càng chính xác.

Như được chỉ ra trong các hình 2.11, NN-CNM gần với Oracle hơn so với DISCOVER-HEVC, đặc biệt với các chuỗi video chuyển động chậm như *Akiyo*

Hình 2.11: So sánh tham số  $\alpha$  của các phương pháp khác nhau

và *Carphone*. Điều này cho thấy phương pháp ước lượng tham số  $\alpha$  đề xuất dựa trên mạng nơ-ron đã cải thiện tính chính xác cho CNM so với phương pháp trước đó được sử dụng trong DISCOVER.

#### D. Đánh giá chất lượng của khung hình giải mã

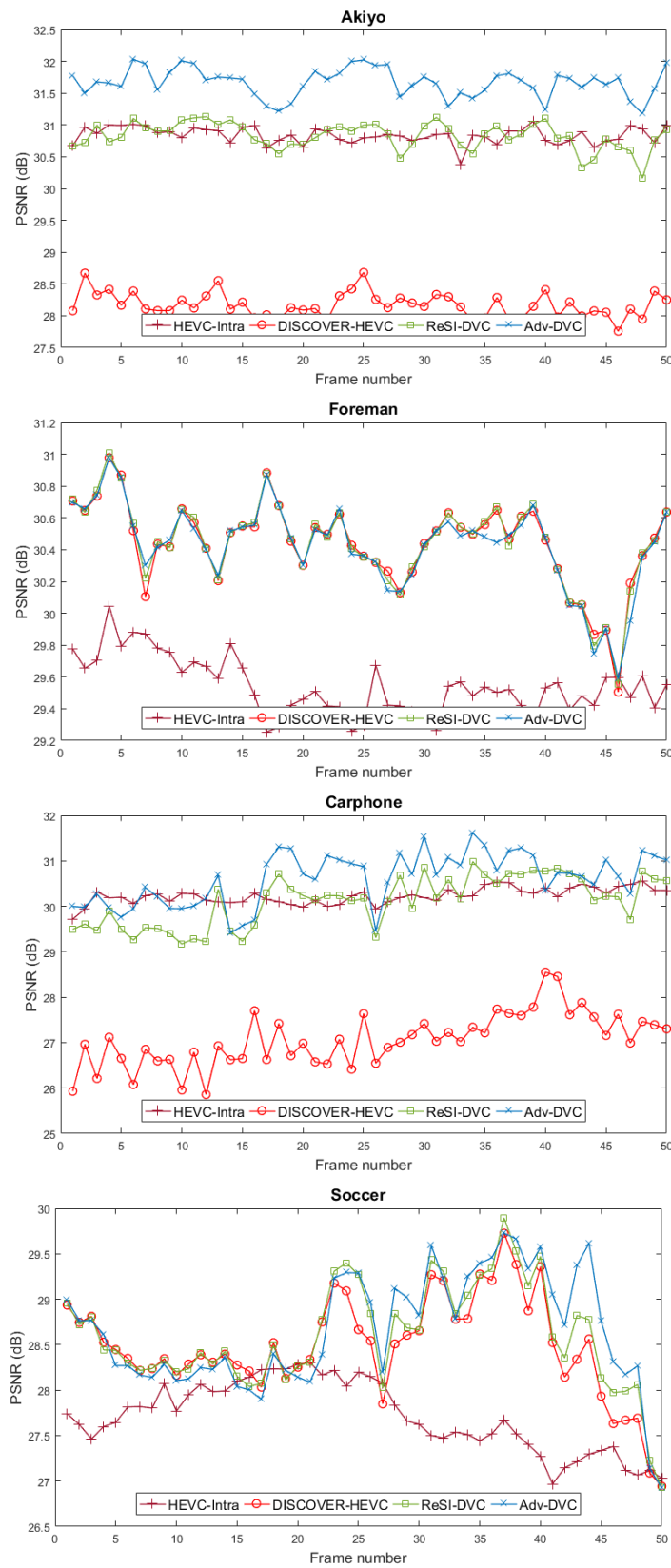
Trong phần này, chất lượng của khung hình giải mã được tạo ra bởi bộ mã hóa đề xuất Adv-DVC được so sánh với khung hình giải mã của các bộ mã hóa ReSI-DVC, DISCOVER-HEVC và HEVC-Intra. So sánh được thực hiện với cả bốn tham số lượng tử cho mỗi chuỗi video. Các kết quả được giới thiệu trong Bảng 2.7 và minh họa trong Hình 2.12. Trong Bảng 2.7, cột Adv-DVC là giá trị PSNR [dB] của khung hình giải mã được tạo ra bởi bộ mã hóa Adv-DVC. Các cột tiếp theo lần lượt là mức tăng PSNR[dB] của cột Adv-DVC khi so với giá trị PSNR[dB] của khung hình giải mã được tạo ra bởi các bộ mã hóa ReSI-DVC, DISCOVER-HEVC và HEVC-Intra.

Từ các kết quả trong Bảng 2.7 có thể rút ra một số nhận xét như sau:

- **Adv-DVC với các bộ mã hóa DVC khác:** Các bộ mã hóa DVC khác ở đây bao gồm DISCOVER-HEVC và ReSI-DVC. Kiến trúc đề xuất Adv-DVC đạt được kết quả tốt hơn so với tất cả các bộ mã hóa DVC khác cho tất cả các chuỗi video thử nghiệm. So với DISCOVER-HEVC, PSNR của Adv-DVC có mức cải tiến lên đến 3,55 dB, ví dụ chuỗi *Akiyo* và mức cải tiến ít hơn khi gặp chuỗi chuyển động nhanh và phức tạp như *Soccer*. So với ReSI-DVC, Adv-DVC cũng đạt được những mức cải tiến đáng kể.
- **Adv-DVC với HEVC Intra:** HEVC Intra được sử dụng để so sánh vì đây là bộ mã hóa video dự đoán có độ phức tạp thấp. Như được mô tả trong Bảng 2.7, Adv-DVC đạt được giá trị PSNR cao hơn so với HEVC Intra cho hầu hết các chuỗi ngoại trừ chuỗi *Carphone*. Các mức cải tiến đối với các chuỗi chuyển động chậm và nhanh cũng khác nhau. Đối với các chuỗi chuyển động chậm, ví dụ *Akiyo*, mức cải thiện PSNR có thể lên đến 1,37 dB nhưng kết quả không được tốt đối với các chuỗi chuyển động nhanh

Bảng 2.7: PSNR [dB] của khung hình giải mã của bộ mã hóa Adv-DVC so với các bộ mã hóa khác

Chuỗi video	QP	Adv-DVC	với ReSI-DVC	với DISCOER-HEVC	với HEVC-Intra
Akiyo	25	43,91	0,17	3,36	1,94
	29	40,46	0,48	3,78	1,48
	34	36,39	0,86	3,60	1,18
	40	31,80	0,83	3,46	0,88
<b>Mức cải thiện trung bình</b>			<b>0,59</b>	<b>3,55</b>	<b>1,37</b>
Foreman	25	40,92	-0,06	0,00	1,21
	29	37,74	0,25	0,32	1,09
	34	33,97	0,18	0,26	0,89
	40	29,97	0,20	0,28	0,79
<b>Mức cải thiện trung bình</b>			<b>0,14</b>	<b>0,21</b>	<b>1,00</b>
Carphone	25	39,64	-0,04	1,25	-1,16
	29	36,39	0,05	1,41	-1,34
	34	33,22	0,21	1,68	-0,82
	40	29,79	0,48	3,10	-0,15
<b>Mức cải thiện trung bình</b>			<b>0,18</b>	<b>1,86</b>	<b>-0,87</b>
Soccer	25	39,91	0,03	0,10	0,44
	31	35,93	0,03	0,10	0,61
	36	32,67	0,01	0,07	0,22
	44	28,87	0,00	0,04	0,65
<b>Mức cải thiện trung bình</b>			<b>0,02</b>	<b>0,08</b>	<b>0,48</b>
<b>Mức cải thiện trung bình tổng cộng</b>			<b>0,23</b>	<b>1,43</b>	<b>0,49</b>



Hình 2.12: So sánh PSNR [dB] của bộ mã hóa Adv-DVC với các bộ mã hóa khác

Bảng 2.8: So sánh BD-Rate [%] và BD-PSNR [dB] giữa Adv-DVC với HEVC Intra

Chuỗi video	Adv-DVC và HEVC Intra	
	BD-Rate	BD-PSNR
Akiyo	-57,34	6,58
Foreman	-50,59	4,00
Carphone	-17,99	0,94
Soccer	37,88	-1,62
<b>Trung bình</b>	<b>-22,01</b>	<b>2,47</b>

*Carphone*. Lý do là vì chuỗi *Carphone* được coi là chuỗi chuyển động phức tạp với sự thay đổi bất thường về nội dung. Cụ thể, trong chuỗi này, sự thay đổi cảnh xảy ra tại các WZF thứ 89 và 115. Điều này dẫn tới sự sụt giảm chất lượng của SI và tính chính xác của CNM. Do đó, giá trị PSNR giảm đột ngột tại các khung hình này. Với chế độ mã hóa Intra của HEVC, do không khai thác tương quan thời gian nên chất lượng của khung hình giải mã có chất lượng tốt hơn.

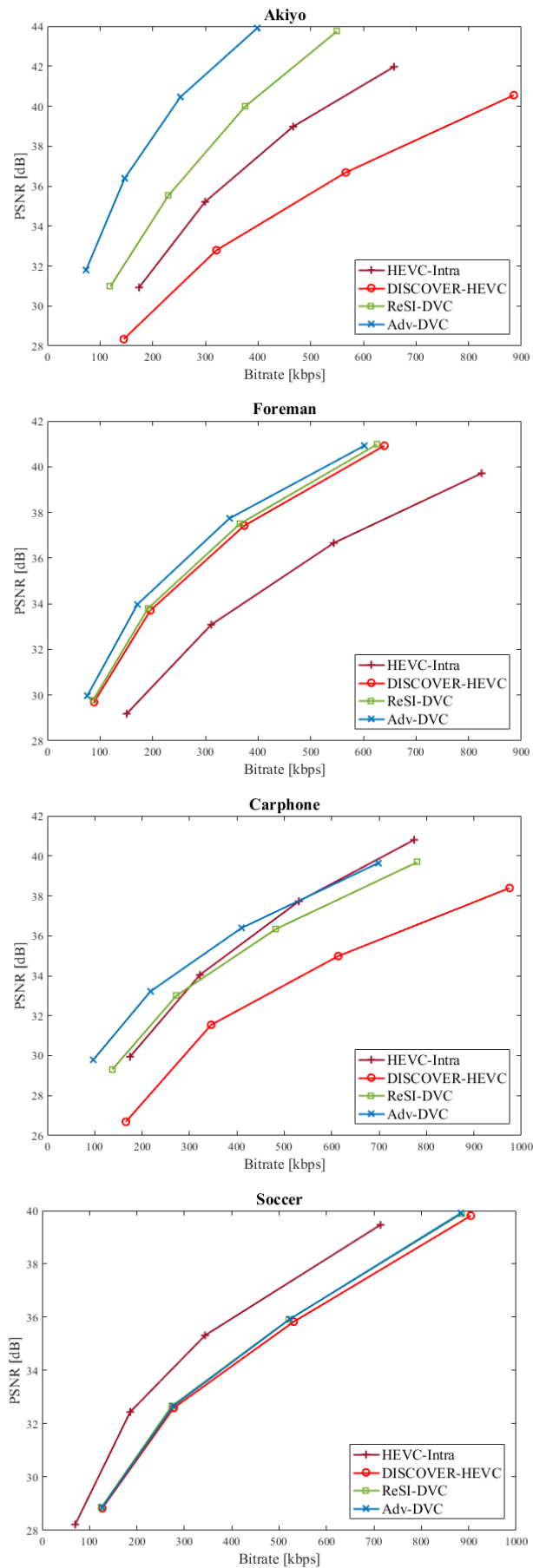
- Tổng thể, kiến trúc mã hóa đề xuất Adv-DVC cho chất lượng khung hình giải mã tốt hơn so với cả ba bộ mã hóa còn lại ReSI-DVC, DISCOVER-HEVC và HEVC-Intra.

### E. Hiệu năng nén của bộ mã hóa DVC đề xuất Adv-DVC

Trong phần này, hiệu năng nén của bộ mã hóa Adv-DVC được so sánh với các bộ mã hóa có liên quan. Các kết quả được trình bày trong Bảng 2.8 và Bảng 2.9 và minh họa trong Hình 2.13.

Từ Bảng 2.8 và Bảng 2.9 và Hình 2.13 có thể rút ra một số nhận xét sau:

- **Adv-DVC so với HEVC Intra:** Hiệu năng RD của Adv-DVC tốt hơn so với HEVC Intra cho hầu hết các chuỗi ngoại trừ chuỗi chuyển động phức tạp *Soccer*. Đối với chuỗi chuyển động chậm, Adv-DVC cho kết quả tốt hơn



Hình 2.13: Hiệu năng RD đối với các chuỗi video

Bảng 2.9: So sánh BD-Rate [%] và BD-PSNR [dB] giữa Adv-DVC với các bộ mã hóa DVC khác

Chuỗi video	với DISCOVER- HEVC		với DVC-ReSI	
	BD-Rate	BD-PSNR	BD-Rate	BD-PSNR
Akiyo	-72,76	8,94	-52,62	5,37
Foreman	-14,46	0,86	-11,24	0,65
Carphone	-51,46	4,15	-20,79	1,25
Soccer	-2,43	0,14	0,52	-0,03
<b>Trung bình</b>	<b>-35,27</b>	<b>3,52</b>	<b>-21,03</b>	<b>1,81</b>

bởi vì SI có chất lượng tốt và CNM chính xác. Khi đo bằng đại lượng BD Rate, Adv-DVC có thể tiết kiệm 57,3 % đối với chuỗi chuyển động chậm, ví dụ chuỗi *Akiyo*. Tổng thể, Adv-DVC đạt được mức tiết kiệm tốc độ bit 22,01 % và mức cải thiện BD-PSNR 2,47 dB.

- **Adv-DVC với các bộ mã hóa DVC khác:** Hiệu năng RD của Adv-DVC được cải tiến đáng kể so với các bộ mã hóa DVC khác đối với tất cả các chuỗi video thử nghiệm. Mức cải tiến đối với các chuỗi chuyển động chậm cao hơn so với các chuỗi chuyển động phức tạp. Khi so với codec DISCOVER-HEVC, mức cải tiến BD-PSNR có thể lên tới 8,94 dB và mức tiết kiệm tốc độ bit đạt tới 72,76 % đối với chuỗi *Akiyo*. Đối với các chuỗi chuyển động phức tạp và nhanh thì gặp khó khăn trong việc tạo ra SI chất lượng tốt và CNM chính xác. Vì vậy khó có thể đạt được mức cải tiến lớn như đối với chuỗi chuyển động chậm. Tuy nhiên, về tổng thể, Adv-DVC đạt được mức giảm tốc độ bit trung bình là 35,27 % khi so với DISCOVER-HEVC và 21,03 % khi so sánh với ReSI-DVC.



## 2.4 Kết luận chương

Chương này đã thực hiện khảo sát một số các nghiên cứu có liên quan đến cải tiến hiệu năng nén của DVC bao gồm giải pháp lựa chọn chế độ mã hóa được thực hiện tại phía mã hóa và nhóm giải pháp được thực hiện tại phía giải mã bao gồm cải tiến chất lượng của thông tin phụ trợ và cải tiến tính chính xác của mô hình nhiễu tương quan. Trên cơ sở đó đã đề xuất và phân tích đánh giá các giải pháp đề xuất. Các đóng góp được trình bày trong chương này gồm có: phương pháp thay đổi kích thước nhóm ảnh được thực hiện tại phía mã hóa; và nhóm giải pháp được thực hiện tại phía giải mã bao gồm tạo SI bằng cách lọc liên tục và ước lượng tham số  $\alpha$  cho mô hình hóa nhiễu tương quan dựa trên mạng nơ-ron.

Các kết quả thu được cho thấy các đề xuất này đã cải tiến đáng kể hiệu năng nén so với các bộ mã hóa video có liên quan khác đặc biệt là đối với các chuỗi video chuyển động chậm. Cụ thể như sau: (1) Đề xuất 1: Cải tiến hiệu năng nén cho DVC tại phía mã hóa hay phương pháp thay đổi kích thước GOP đề xuất có mức tiết kiệm bit so với khi sử dụng kích thước cố định  $GOP = 2$  là 3.37% và  $GOP = 4$  là 6,2%. (2) Đề xuất 2: Cải tiến hiệu năng nén cho DVC tại phía giải mã với hai mô-đun mới gồm tạo thông tin phụ trợ bằng cách lọc liên tục và ước lượng tham số mô hình nhiễu tương quan bằng mạng nơ-ron có thể đạt được mức tiết kiệm bit trung bình 22,1% so với các bộ mã hóa video dự đoán và 35,27 % so với bộ mã hóa DISCOVER ban đầu.

Điều này một lần nữa khẳng định tiềm năng ứng dụng DVC trong các ứng dụng có sử dụng các chuỗi video ít chuyển động ví dụ như mạng video giám sát.

### Các bài báo đã công bố:

[1] (2018) T. N. Thi Huong, H. Phi Cong, T. V. Huu, and X. Hoang Van, *Artificial Intelligence Based Adaptive GOP Size Selection for Effective Wyner-Ziv Video Coding*, International Conference on Advanced Technologies for Communications, vol. 2018-October, pp. 120–124, 2018, doi: 10.1109/ATC.2018.8587573.

[2] (2019) T. V. Huu, T. Nguyen Thi Huong, M. N. Ngoc, and X. Hoang-van, *Improving performance of distributed video coding by consecutively refining of side information and correlation noise model*, Proceedings - 2019 19th International Symp Symposium on Communications and Information Technologies (ISCIT 2019), pp. 502–506, 2019, doi: 10.1109/ISCIT.2019.8905187.

[3] (2020) Tien Vu Huu, Thao Nguyen Thi Huong, Xiem Hoang Van, San Vu Van, *Improving TDWZ Correlation Noise Estimation: A Deep Learning based Approach*, REV Journal on Electronics and Communications: Article scheduled for publication in Vol. 10, No. 1–2, January–June, pp. 11-20, 2020, ISSN 1859 – 378X

# Cải tiến hiệu năng nén cho mã hóa video phân tán liên lớp

Trong mã hóa video, khả năng liên lớp là một giải pháp hữu ích giúp cho việc truyền các luồng video qua mạng hỗn tạp. Nguyên lý mã hóa video phân tán của DVC sẽ có nhiều lợi thế khi áp dụng vào kiến trúc mã hóa video liên lớp. Vì vậy trong chương 3 sẽ giới thiệu sơ lược về mã hóa video phân tán liên lớp và các đề xuất để tăng cường hiệu năng nén của kiến trúc này bao gồm: (1) tạo thông tin phụ trợ dựa trên kỹ thuật kết hợp ảnh; và (2) tạo thông tin phụ trợ dựa trên SVM. Trước khi đi vào chi tiết của các đề xuất, phần Giới thiệu chung sẽ trình bày một số nghiên cứu liên quan đến mã hóa video liên lớp.

## 3.1 Giới thiệu chung

Sự phát triển của kỹ thuật video số và các cải tiến không ngừng trong cơ sở hạ tầng truyền thông đã thúc đẩy hàng loạt các ứng dụng đa phương tiện có khả năng tương tác như hội nghị video thời gian thực, streaming video và truyền hình di động. Thị trường sử dụng video tương tác thu hút nhiều người sử dụng hơn và chính điều này đặt ra yêu cầu chất lượng video luôn phải tốt

nhất mọi lúc và mọi nơi với mạng của người sử dụng.

Nói cách khác, video truyền đi phải tương thích được với các đặc tính của máy thu như tốc độ bit yêu cầu, độ phân giải và tốc độ khung hình. Chất lượng video cung cấp luôn phải có chất lượng tốt nhất trong điều kiện mạng và máy thu có những hạn chế nhất định. Bên cạnh đó, thông thường cùng một đường truyền được sử dụng để truyền tới các thiết bị có tính năng hạn chế khác nhau như điện thoại di động cũng như các thiết bị có hiệu năng cao, ví dụ như máy thu truyền hình độ nét cao HDTV. Cuối cùng, luồng video phải thích ứng với các mạng không dây có tổn thất.

Với các lý do này, có thể thấy các mạng có tính chất hỗn tạp và không xác định đặt ra một bài toán lớn đối với các bộ mã hóa video truyền thống khi không thể cung cấp khả năng thích ứng đối với tình hình mạng và thiết bị thu đa dạng.

Trong tình hình đó, mã hóa video liên lớp là một giải pháp hấp dẫn cho những thách thức đặt ra bởi các đặc điểm của các ứng dụng video hiện đại. Trong ngữ cảnh này, khả năng liên lớp nói đến đặc tính của luồng bit video cho phép loại bỏ một phần của luồng bit để điều chỉnh nó theo nhu cầu của người dùng đầu cuối cũng như khả năng của thiết bị nhận hoặc điều kiện mạng, sao cho luồng bit kết quả vẫn tuân thủ các tiêu chuẩn mã hóa video được sử dụng.

Ngoài mã hóa video liên lớp, có một số lựa chọn thay thế là chuyển mã và simulcast. Chuyển mã liên quan đến việc giải mã một luồng bit và mã hóa lại nó với các tham số mã hóa khác, và đôi khi bao gồm cả giảm độ phân giải video sau khi giải mã và trước khi mã hóa lại để cung cấp luồng bit video với độ phân giải thấp hơn. Chuyển mã video thường có độ trễ và phức tạp về mặt tính toán.

Simulcast thực hiện mã hóa tách biệt video ở các tốc độ bit khác nhau và có thể tại độ phân giải, tốc độ khung hình khác nhau và được truyền đồng thời. Bộ giải mã có thể chọn một lớp nào đó của luồng bit simulcast để yêu cầu hoặc giải mã.

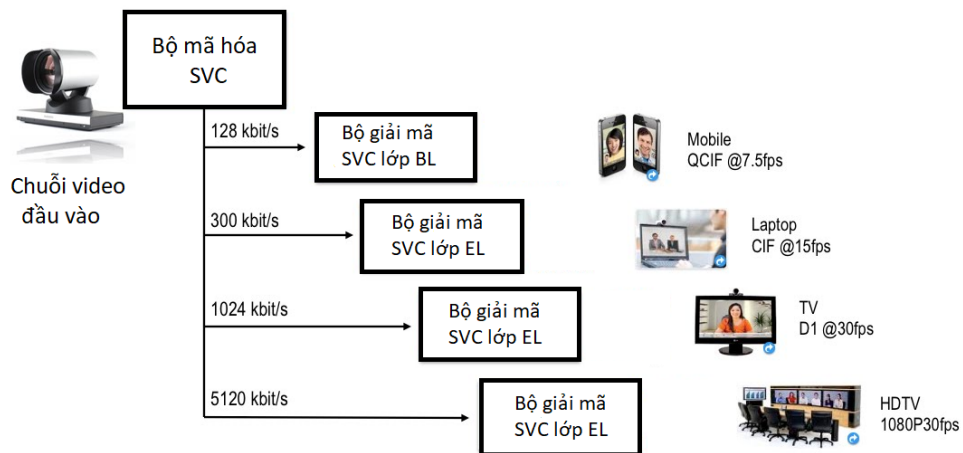
Một lựa chọn khác là truyền tốc độ bit thích ứng, trong đó bộ mã hóa tạo

và lưu trữ nhiều bản sao của cùng một nội dung video (với các tham số mã hóa khác nhau), nhưng chỉ có một cách mã hóa được truyền một cách hiệu quả đến phía thu.

Rõ ràng, mã hóa video liên lớp phải cạnh tranh với các lựa chọn thay thế này. Cụ thể, mã hóa video liên lớp chỉ hữu ích nếu nó cung cấp hiệu quả mã hóa cao hơn so với các giải pháp khác. Do đó, mục tiêu của các chuẩn mở rộng khả năng liên lớp cho một tiêu chuẩn mã hóa video là cho phép tạo ra một luồng video có chứa một hoặc nhiều luồng bit con có thể được giải mã với độ phức tạp và chất lượng tái tạo tương đương với trường hợp sử dụng mã hóa một lớp với cùng một lượng dữ liệu như trong luồng bit con.

### 3.1.1 Mã hóa video liên lớp

Nguyên tắc của bộ mã hóa video liên lớp là chia video luồng đơn truyền thống thành nhiều luồng thường gọi là các lớp. Hình 3.1 mô tả khái niệm này với một bộ mã hóa chuỗi video đầu vào thành bốn lớp. Vì vậy, các máy thu có thể lựa chọn và giải mã số lớp khác nhau – tương ứng với chất lượng video khác nhau – tùy theo khả năng xử lý của mạng và bản thân thiết bị. Cấu trúc lớp

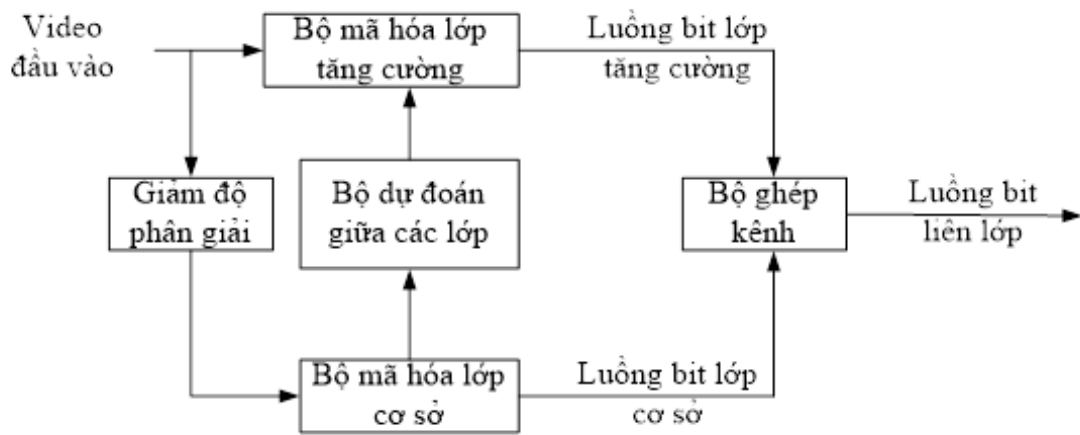


Hình 3.1: Khả năng thích ứng của mã hóa video liên lớp

của nội dung video liên lớp có thể được định nghĩa là sự kết hợp của lớp cơ sở

(BL) và một vài lớp tăng cường phụ thêm (EL). Lớp BL tương ứng với hiệu năng video thấp nhất, các lớp EL cho phép cải thiện chất lượng BL cho trước. Có 3 loại liên lớp chính: không gian, thời gian và chất lượng. Khả năng liên lớp không gian và thời gian mô tả các trường hợp trong đó một luồng bit con thể hiện nội dung nguồn với kích thước hình ảnh (hay độ phân giải không gian) và tốc độ khung hình (hoặc độ phân giải thời gian) giảm, tương ứng. Với khả năng mở rộng chất lượng, còn được gọi là khả năng mở rộng tỷ lệ tín hiệu / nhiễu (SNR) hoặc khả năng mở rộng độ trung thực, dòng bit phụ cung cấp cùng một độ phân giải không gian và thời gian như luồng bit hoàn chỉnh, nhưng với chất lượng tái tạo thấp hơn và do đó, tốc độ bit thấp hơn.

Các BL đại diện cho chất lượng thấp nhất. Một hoặc nhiều lớp EL có thể được mã hóa bằng cách tham chiếu đến các lớp thấp hơn và cung cấp chất lượng video tốt hơn. Nếu so với việc giải mã đầy đủ luồng bit, giải mã một số các lớp sẽ cho video với chất lượng thấp hơn nhưng vẫn chấp nhận được. Mặc dù chất lượng bị giảm nhưng khi đặt trong tình huống tốc độ bit giảm, nó vẫn được coi là tốt hơn khi so với trường hợp sử dụng luồng bit video không liên lớp. Thông thường khi đó chất lượng video không liên lớp sẽ bị giảm nghiêm trọng và đôi khi nhanh chóng xuống mức chất lượng không thể xem được. Tuy nhiên, mã hóa video liên lớp thường tốn nhiều bit hơn để đạt được chất lượng video tương đương khi so với mã hóa video không liên lớp. Hình 3.2 mô tả sơ đồ khối đơn giản hóa cho mã hóa liên lớp không gian và chất lượng với hai lớp. Có thể thấy bộ mã hóa liên lớp bao gồm hai bộ mã hóa, mỗi bộ cho một lớp. Trong mã hóa liên lớp không gian, video đầu vào được giảm kích cỡ và đưa vào bộ mã hóa BL, trong khi video đầu vào có kích thước ban đầu là đầu vào của bộ mã hóa lớp EL. Trong mã hóa liên lớp chất lượng, cả hai bộ mã hóa đều sử dụng cùng một tín hiệu đầu vào. Bộ mã hóa BL tuân thủ tiêu chuẩn mã hóa video một lớp, để đạt được khả năng tương thích ngược với mã hóa một lớp; bộ mã hóa lớp EL thường bao gồm các tính năng mã hóa bổ sung. Đầu ra của cả hai bộ mã hóa được ghép kênh để tạo thành luồng bit liên lớp.

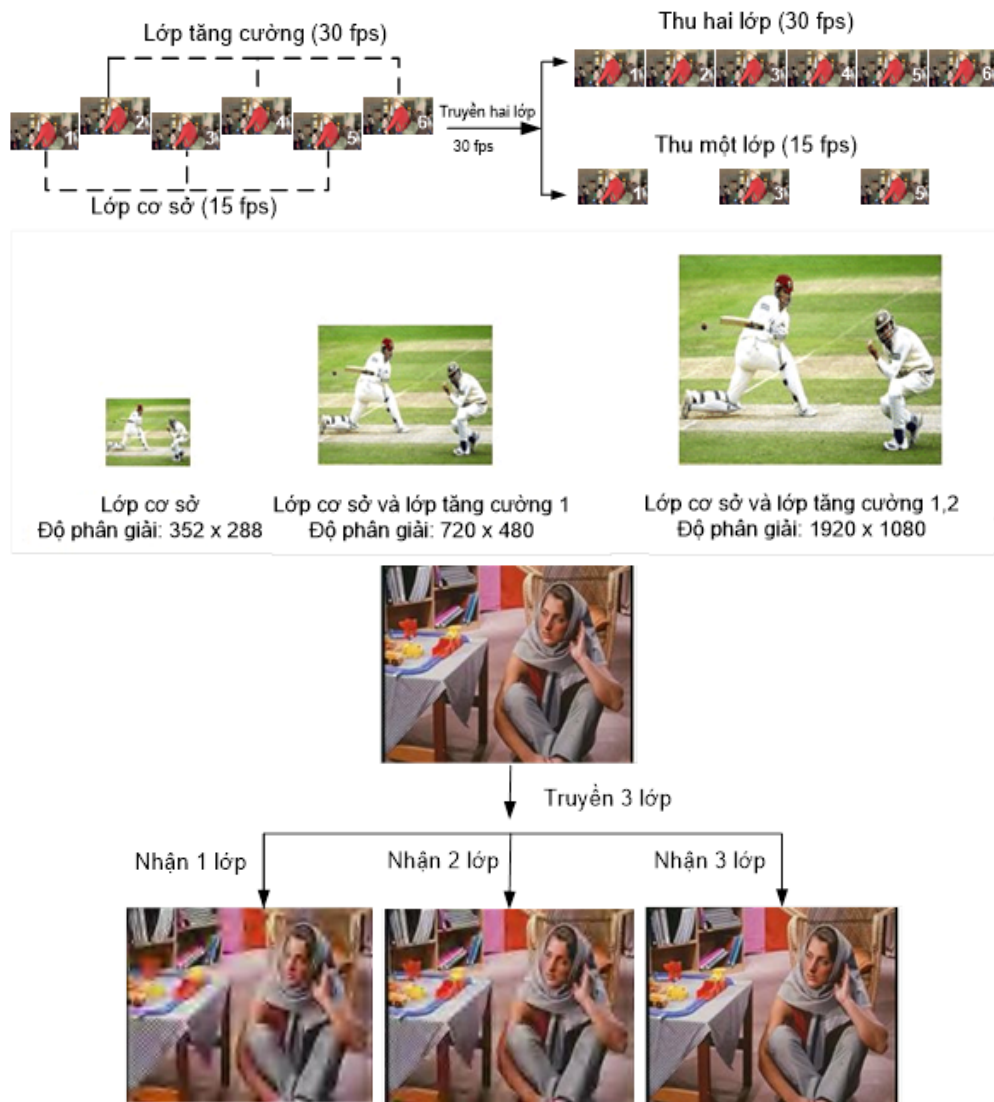


Hình 3.2: Sơ đồ khối đơn giản của bộ mã hóa liên lớp gồm hai lớp

Để minh họa cho khả năng liên lớp thời gian, Hình 3.3(a) giới thiệu một kịch bản đơn giản ở đó lớp BL gồm một nhóm nhỏ các khung hình và lớp EL gồm một nhóm khác. Một máy thu giả định trong mạng độ rộng băng thông thấp sẽ chỉ nhận BL, vì vậy tạo ra video bị giật (15 khung hình trong một giây) so với bình thường. Ngược lại, máy thu thứ hai (có độ rộng băng thông cao hơn) có thể xử lý và kết hợp hai lớp, vì vậy tạo ra video tốc độ khung hình đầy đủ (30 khung hình trong một giây) và video được tái tạo sẽ mịn hơn.

Tiếp đó, Hình 3.3(b) minh họa một ví dụ về khả năng liên lớp không gian mà ở đó kết quả của các lớp EL cải thiện độ phân giải của chuỗi video được giải mã. Như đã thấy trong hình, nếu phía thu nhận càng nhiều lớp, độ phân giải của chuỗi video giải mã càng cao. Cuối cùng, Hình 3.3(c) mô tả khái niệm liên lớp chất lượng, ở đó các lớp EL cải thiện chất lượng SNR của luồng video nhận được. Hiển nhiên, nếu phía thu có càng nhiều lớp, chất lượng trải nghiệm của người dùng càng tốt.

Với vai trò quan trọng của khả năng liên lớp, hiện nay, các chuẩn mã hóa video đều có thêm mở rộng liên lớp. Các tiêu chuẩn mã hóa video quốc tế H.262, H.263, và H.264, HEVC bao gồm một số công cụ có thể hỗ trợ mã hóa video liên lớp. Trong khi các cấu hình liên lớp của H.262, H.263 hiếm khi được sử dụng



Hình 3.3: Các ví dụ minh họa khả năng liên lớp: liên lớp thời gian, liên lớp không gian và liên lớp chất lượng



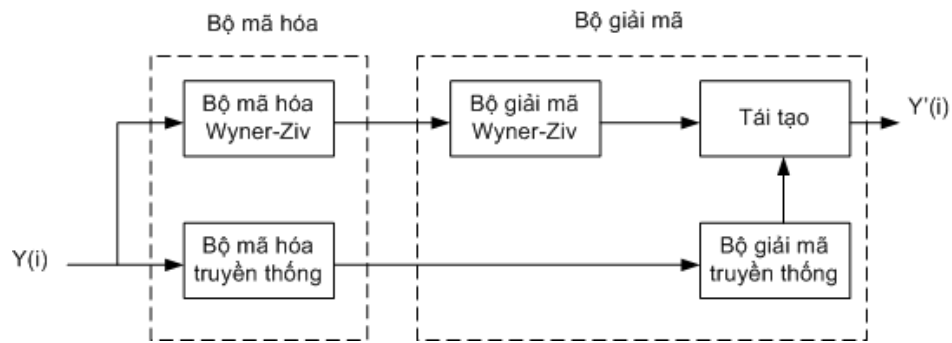
trong thực tế, phần mở rộng liên lớp SVC của H.264 | MPEG-4 AVC [50], hay SHVC của HEVC [7] được sử dụng thành công trong một số ứng dụng hội nghị truyền hình. Các chuẩn mở rộng sau này đều được thiết kế như một bộ mã hóa video liên lớp đầy linh hoạt và mạnh mẽ, có khả năng cung cấp một mức chất lượng cho trước với tỉ lệ nén tương đương với độ phức tạp giải mã thấp hơn khi so với các chuẩn không liên lớp tương ứng. Tất cả ba loại khả năng mở rộng này đều được hỗ trợ trong cả SVC và SHVC. Các kết quả trong [7] cho thấy đối với các ứng dụng video giải trí (cấu hình truy nhập ngẫu nhiên), khi so với HEVC Simulcast, SHVC có mức giảm BD-rate là 16,5%, 27% và 21% tương ứng với các khả năng liên lớp không gian tăng 2 lần, 1,5 lần và liên lớp SNR trong trường hợp tính tổng số bit EL + BL. Khi so với SVC, SHVC có mức giảm BD-rate là từ 50% đến 63% tùy thuộc vào loại liên lớp và trường hợp sử dụng.

### 3.1.2 Mã hóa video phân tán liên lớp

Đối với hệ thống mã hóa video phân tán liên lớp, thực tế cho đến nay chưa có một định nghĩa chính xác cho hệ thống này, tuy nhiên có thể coi hệ thống mã hóa video phân tán liên lớp (DSVC) là hệ thống mã hóa video mà ở đó có sử dụng nguyên tắc mã hóa video phân tán trong hệ thống đồng thời cung cấp khả năng liên lớp. Để đáp ứng khả năng tương thích với các chuẩn mã hóa video hiện thời, các hệ thống DSVC hầu hết đều mã hóa lớp BL bằng các chuẩn mã hóa video truyền thống như H264/AVC hoặc H265/HEVC còn các lớp EL được mã hóa bằng phương pháp mã hóa video phân tán. Sự kết hợp giữa hai phương pháp mã hóa video này giúp tận dụng ưu điểm hiệu suất nén cao của mã hóa video dự đoán và ưu điểm độ phức tạp bộ mã hóa thấp và khả năng chống lỗi tốt của mã hóa video phân tán.

Hình 3.4 mô tả chung về hệ thống DSVC với một lớp BL và một lớp EL. Lớp BL được mã hóa bằng bộ mã hóa truyền thống và lớp EL được mã hóa theo nguyên tắc mã hóa phân tán.

Dựa trên mô hình mã hóa này, nhiều đề xuất bộ mã hóa phân tán liên lớp



Hình 3.4: Mô hình bộ mã hóa video phân tán liên lớp

được giới thiệu. Tagliasacchi và các cộng sự trong [55] đã thực hiện một phiên bản PRISM liên lớp. Kiến trúc này đề xuất khả năng liên lớp thời gian – không gian với lớp BL được mã hóa bằng AVC/H.264 và lớp EL được mã hóa theo nguyên tắc PRISM. Kết quả cho thấy bộ mã hóa PRISM liên lớp này có hiệu năng vượt trội so với bộ mã hóa H.264/AVC. Ngay cả trong trường hợp kênh truyền bị mất gói hay lớp cơ sở bị mất, PRISM liên lớp vẫn cho chất lượng tái tạo tốt. Đây là một kết quả rất đáng khích lệ nếu so sánh với các bộ mã hóa video liên lớp tiêu chuẩn dựa trên dự đoán. Khi lớp BL bị mất, chất lượng video thường bị ảnh hưởng nghiêm trọng.

Một giải pháp cho bài toán mã hóa video liên lớp khác được giới thiệu trong [51]. Bộ mã hóa này tập trung vào liên lớp chất lượng. Lớp BL vẫn được mã hóa bằng chuẩn H.264 và các khung hình của lớp EL được chia nhỏ và mã hóa bằng nhiều bộ mã hóa LDPC. Kết quả cho thấy bộ mã hóa đề xuất có thể cải thiện đến 4 dB so với bộ mã hóa liên lớp dựa trên nguyên tắc mã hóa dự đoán.

Trong [40], các tác giả đề xuất hệ thống mã hóa video dựa trên nguyên tắc DVC với các khả năng mở rộng thời gian, không gian và chất lượng. Trong đó, mã hóa video truyền thống vẫn được sử dụng để mã hóa cho lớp BL và mã hóa WZ tạo ra các lớp EL. Tại bộ giải mã, lớp BL được sử dụng để tạo ra SI cho quá trình giải mã DVC. Các kết quả cho thấy bộ mã hóa DVC liên lớp này có hiệu năng nén thấp hơn mã hóa video liên lớp truyền thống trong trường hợp không có nhiễu. Tuy nhiên, trong điều kiện có nhiễu, hệ thống dựa trên DVC cho thấy

khả năng chống nhiễu tốt hơn so với mã hóa video liên lớp truyền thống.

Xiem H.V và các cộng sự trong [61][25] đã đề xuất giải pháp mã hóa video liên lớp có độ phức tạp thấp và có khả năng tương thích ngược với lớp BL được mã hóa bằng chuẩn HEVC và sử dụng cách tiếp cận DVC cho các lớp EL. Bộ mã hóa này cung cấp khả năng liên lớp không gian và thời gian. Trong [61] các tác giả đề xuất một giải pháp tạo SI mới trong đó tận dụng cả thông tin lớp BL được giải mã và thông tin về khung hình chính tại lớp EL. Các kết quả cho thấy giải pháp có thể đạt được hiệu năng RD tốt hơn so với cấu hình SHVC tương đương cũng như HEVC simulcasting khi sử dụng cấu hình mã hóa nội ảnh. Ngoài ra có thể kể đến một ưu điểm khác của hệ thống là vẫn giữ được độ phức tạp mã hóa thấp khi so với các bộ mã hóa video dự đoán.

Sakomizu và cộng sự trong [48] đề xuất một giải pháp ở đó hai phương pháp tạo SI được chấp nhận bao gồm một phương pháp tạo SI đơn giản ở phía mã hóa để ước lượng tốc độ mã hóa và một phương pháp tạo SI hai bước ở phía giải mã để tái tạo thông tin lớp EL.

Các nghiên cứu ở trên cho thấy mã hóa video liên lớp dựa trên nguyên lý phân tán hứa hẹn trở thành một giải pháp mã hóa video liên lớp tiềm năng với độ phức tạp mã hóa thấp và khả năng chống nhiễu tăng cường trong khi vẫn cung cấp các khả năng thích ứng theo nhu cầu cho các mạng hỗn tạp và các loại thiết bị đầu cuối với các cách sử dụng khác nhau. Tuy nhiên, có thể thấy rằng, hiệu suất nén của các giải pháp DSVC vẫn còn hạn chế, đặc biệt khi so với tiêu chuẩn SHVC cho mã hóa video dự đoán. Vì vậy, mục tiêu của chương này là đề xuất các cải tiến đối với quá trình tạo SI nhằm cải thiện hiệu năng nén cho DSVC với lớp BL được mã hóa bằng chuẩn HEVC-Intra.

Phần tiếp theo của chương này sẽ được cấu trúc như sau. Phần 3.2 giới thiệu về phương pháp tạo SI dựa trên kỹ thuật kết hợp ảnh cho DSVC. Phương pháp tạo SI dựa trên kỹ thuật SVM cho hệ thống DSVC được trình bày trong phần 3.3. Cuối cùng là phần kết luận chương được giới thiệu trong phần 3.4.

## 3.2 Đề xuất 3: Tạo thông tin phụ trợ dựa trên kỹ thuật kết hợp ảnh cho DSVC

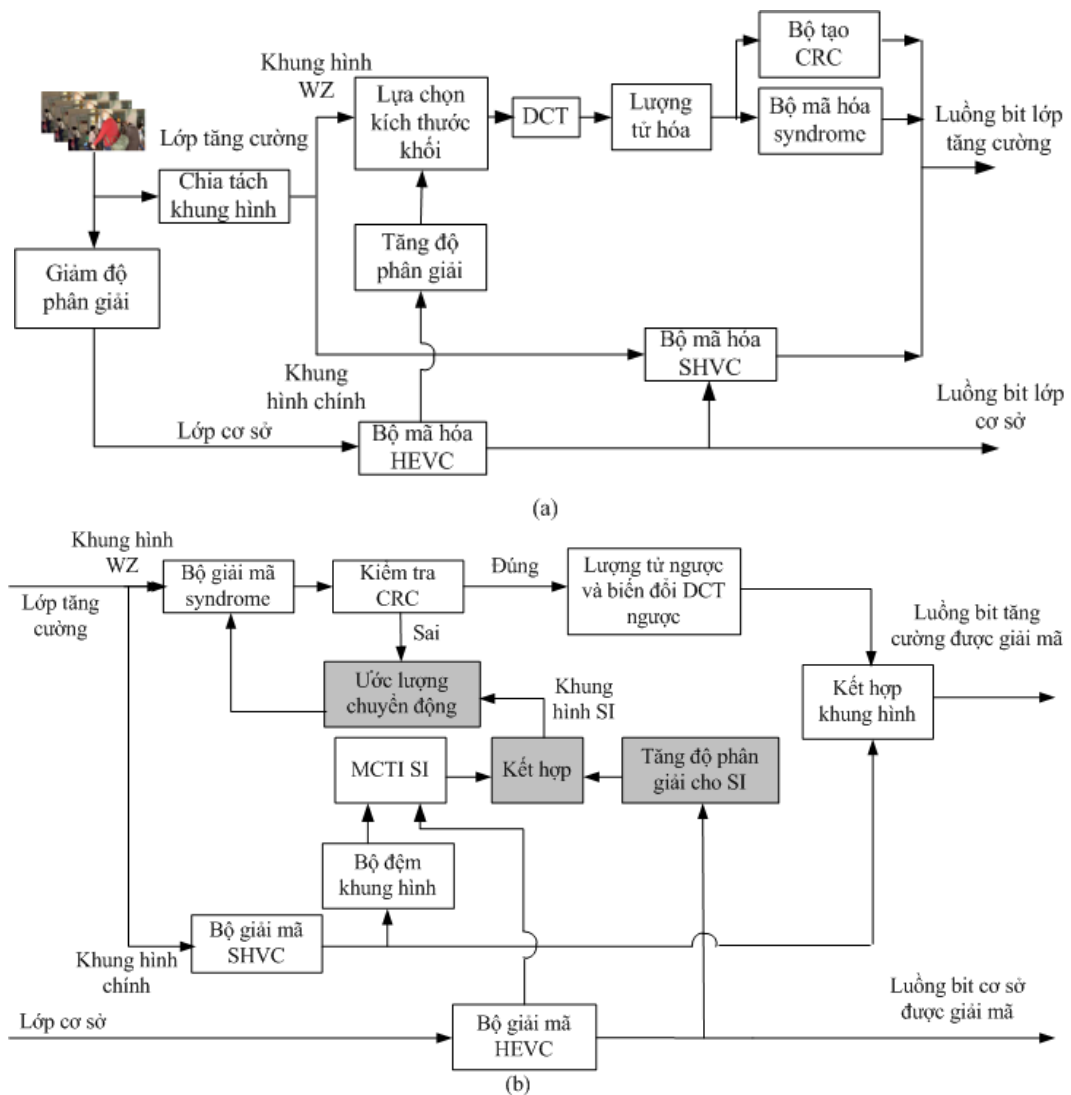
Để đáp ứng nhu cầu của các ứng dụng mới như đã trình bày ở trên, việc xây dựng bộ mã hóa DSVC phải đảm bảo một số yêu cầu sau:

- Khả năng liên lớp: Khả năng thích ứng với các mạng hỗn tạp và các thay đổi động của mạng và thiết bị người dùng là các yêu cầu tối quan trọng. Trong phần này tập trung vào khả năng liên lớp thời gian và không gian.
- Khả năng tương thích ngược: Đây là một yêu cầu cơ bản đối với mã hóa video liên lớp bởi vì nó cho phép đảm bảo sự chuyển tiếp suôn sẻ và sự chấp nhận của thị trường đối với các giải pháp không liên lớp. Vì vậy, DSVC cần phải có khả năng tương thích ngược với HEVC.
- Hiệu năng nén: Hiệu năng nén luôn luôn cần phải tối đa nhất có thể trong khi vẫn đảm bảo được các tiêu chí ban đầu như độ phức tạp mã hóa thấp.
- Độ phức tạp mã hóa: Các bộ mã hóa DVC thường hướng đến các ứng dụng mới như mạng video giám sát, mạng cảm biến, do đó ngoài việc đảm bảo khả năng tương thích ngược, các bộ mã hóa DSVC cũng phải đảm bảo độ phức tạp mã hóa thấp.

Để đáp ứng các yêu cầu này, bộ mã hóa DSVC đề xuất chấp nhận mã hóa HEVC-Intra để mã hóa cho lớp BL và mã hóa phân tán cho lớp EL. Điều này để đảm bảo độ phức tạp mã hóa vẫn được giữ ở mức thấp trong khi cung cấp hiệu năng nén cao.

### 3.2.1 Kiến trúc bộ mã hóa DSVC-Fusion

Kiến trúc bộ mã hóa video phân tán liên lớp đề xuất với tên gọi DSVC-Fusion được mô tả trong Hình 3.5. Kiến trúc này có hai lớp gồm một lớp BL và một



Hình 3.5: Sơ đồ kiến trúc bộ mã hóa và giải mã video phân tán liên lớp DSVC-Fusion

lớp EL. Bộ mã hóa DSVC ở đây cung cấp khả năng mở rộng không gian. Vì vậy mỗi khung hình trong chuỗi video gốc được mã hóa hai lần ở lớp BL và lớp EL. Đối với lớp BL, trước khi đưa vào bộ mã hóa HEVC Intra, khung hình gốc sẽ được giảm độ phân giải không gian.

Nguyên tắc mã hóa video phân tán trong bộ mã hóa này dựa trên kiến trúc PRISM đã được giới thiệu trong chương 1. Quá trình mã hóa và giải mã cho lớp EL có thể được tóm tắt ở dưới đây.

### A. Quá trình mã hóa

1) *Chia tách khung hình*: Các khung hình video EL được chia thành các KF và WZF luân phiên nhau. Trong trường hợp này, kích thước GOP=2. Các KF được mã hóa bằng bộ mã hóa SHVC Intra và các WZF được mã hóa với công cụ mã hóa phân tán và có sử dụng thông tin của khung hình lớp BL tương ứng.

2) *Lựa chọn kích thước khối*: Mỗi khung hình WZ được chia thành các khối kích thước  $4 \times 4$  pixel hoặc  $8 \times 8$  tùy thuộc vào mức tương quan giữa WZF đang mã hóa với KF ngay trước đó. Khi mức tương quan giữa hai khung hình là lớn thì WZF sẽ được chia thành các khối kích thước  $8 \times 8$  và ngược lại, khi mức tương quan giữa hai khung hình nhỏ thì WZF sẽ được chia thành các khối kích thước  $4 \times 4$ . Đại lượng được chọn để đánh giá mức độ tương quan giữa hai khung hình được dựa trên đại lượng MAD như trong công thức 3.1:

$$MAD = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N |WZF(x, y) - KF(x, y)| \quad (3.1)$$

ở đó  $M, N$  là kích thước của khung hình.  $(x, y)$  là tọa độ các pixel. Giá trị MAD càng nhỏ nghĩa là sự tương quan giữa hai khung hình này càng lớn và ngược lại.

Để phân loại khung hình thành loại chuyển động nhanh hay chậm, ta cần phải xác định được ngưỡng (Thr) phù hợp. Trong phương pháp này, ngưỡng được tính là trung bình của MAD của các khung hình trước đó. Dựa trên ngưỡng đã xác định, kích thước khối mã hóa cho mỗi WZF được xác định trong công thức 3.2:

$$S = \begin{cases} 4 \times 4 & MAD > Thr \\ 8 \times 8 & MAD \leq Thr \end{cases} \quad (3.2)$$

ở đó  $S$  là kích thước khối. Nếu  $MAD > Thr$ , tương quan được coi là thấp và vì vậy để có thể khôi phục lại chính xác WZF tại bộ giải mã, ta sẽ sử dụng kích thước khối  $4 \times 4$ . Nếu  $MAD \leq Thr$ , điều đó có nghĩa là tương quan cao và vì vậy sử dụng kích thước khối  $8 \times 8$ .

3) *Biến đổi DCT*: Sau khi lựa chọn kích thước khối của mỗi WZF, khung hình đó được chia thành các khối kích thước  $4 \times 4$  hoặc  $8 \times 8$  tùy thuộc vào kết quả của bước trước và áp dụng biến đổi DCT vào mỗi khối.

4) *Bộ lượng tử hóa*: Tiếp theo áp dụng lượng tử hóa vô hướng lên các hệ số DCT thu được từ module mã hóa trước. Bước lượng tử hóa được lựa chọn cho mỗi hệ số được xác định bởi chất lượng tái tạo mong muốn trong khi vẫn xem xét đến độ nhạy khác nhau của hệ thống thị giác người (HVS) đối với các tần số không gian khác nhau. Trong bộ mã hóa này sử dụng bộ lượng tử hóa như trong chuẩn JPEG.

5) *Bộ mã hóa syndrome*: Sau lượng tử, giá trị syndrome được tính cho mỗi hệ số DCT đã lượng tử dựa trên thống kê tương quan của mô-đun ước lượng chuyển động. Với giả định các bit MSB có thể rút ra từ SI tốt nhất, phía mã hóa chỉ gửi số lượng bit LSB ( $n_{LSB}$ ) của hệ số DCT đã lượng tử tới bộ giải mã.  $n_{LSB}$  được tính như sau:

$$n_{LSB} = \begin{cases} 2 + \left\lfloor \log_2 \left( \frac{|(X_q \cdot \Delta) - X_p|}{\Delta} \right) \right\rfloor, & d > \Delta \\ 0, & d \leq \Delta \end{cases} \quad (3.3)$$

ở đó  $d = \min \{|(X_q \cdot \Delta) - X_p|, |X - X_p|\}$ ,  $\Delta$  là kích thước bước lượng tử đối với hệ số DCT đó.  $X_p$  là hệ số DCT chưa lượng tử của khối dự đoán.

Số lượng bit  $n_{LSB}$  và giá trị của nó được gửi tới bộ giải mã như giá trị syndrome của hệ số DCT đó.

6) *Bộ tạo CRC*: Bộ mã hóa này sử dụng mã dư thừa vòng có mục đích tạo ra một chữ ký nhị phân để kiểm tra khối được giải mã, từ đó lựa chọn ra được khối giải mã đúng.

## B. Quá trình giải mã

1) *Tạo thông tin phụ trợ*: Mô-đun này tạo ra một ước lượng cho WZF gốc bằng cách sử dụng các KF đã giải mã và khung hình lớp BL đã giải mã. Chi tiết của quá trình tạo thông tin phụ trợ được trình bày trong mục 3.2.2 dưới đây.

2) *Bộ giải mã syndrome*: Mô-đun này có trách nhiệm lựa chọn các từ mã lượng tử hóa từ các coset trong khi thực hiện khai thác thông tin phụ trợ được gửi tới từ mô-đun tìm kiếm chuyển động. Dựa trên chỉ số coset, bộ giải mã syndrome tìm trong coset từ mã nào gần với SIB nhất. Khối được giải mã này được gửi tới mô-đun tính toán CRC để kiểm tra tiếp.

3) *Tìm kiếm chuyển động*: Mô-đun tìm kiếm chuyển động có mục tiêu cung cấp một phiên bản bù chuyển động của khối hiện thời cho bộ giải mã syndrome. Trong thực tế, mô-đun này phải tạo ra các ứng viên SIB. Bộ giải mã tìm kiếm SIB trong một cửa sổ  $16 \times 16$  quanh khối hiện thời và gửi SIB này tới bộ giải mã syndrome. Khi quá trình giải mã thành công thì quá trình tìm kiếm chuyển động sẽ dừng lại.

4) *Tính CRC*: Vì với mỗi khối ứng viên SIB kích thước  $4 \times 4$  hoặc  $8 \times 8$ , sau khi được giải mã syndrome thì giá trị CRC sẽ được tính dựa trên các bit MSB của các hệ số DCT lượng tử. Trong trường hợp giá trị CRC tính được phù hợp với giá trị CRC từ bên mã hóa gửi tới thì khối đó được coi là giải mã thành công. Ngược lại thì khối tìm kiếm chuyển động lại tìm kiếm một ứng viên mới và lặp lại toàn bộ quá trình kiểm tra ở trên.

5) *Tái tạo*: Mô-đun này có mục đích lấy lại được giá trị DCT tương ứng với mỗi hệ số DCT lượng tử hóa, vì vậy tái tạo lại nguồn bằng một phiên bản xấp xỉ của khối các hệ số DCT đã mã hóa.

6) *Biến đổi DCT ngược (IDCT)*: Khi tất cả các hệ số đã được giải lượng tử, thực hiện quét zig-zag đã được thực hiện tại bộ mã hóa sẽ được làm ngược lại để lấy được khối hai chiều các hệ số được tái tạo. Các hệ số biến đổi sau đó được biến đổi DCT ngược để thu được các pixel ban đầu.

### 3.2.2 Phương pháp tạo thông tin phụ trợ dựa trên kỹ thuật kết hợp ảnh

Đối với bộ mã hóa DSVC-Fusion, SIB có thể được tạo ra tại bộ giải mã bằng cách sử dụng thêm thông tin từ khung hình độ phân giải thấp ở lớp BL thay vì chỉ sử dụng KF giải mã. Nói cách khác, trong trường hợp này cả tương quan thời gian và tương quan không gian được sử dụng để ước lượng SIB. Điều này có thể giúp cải thiện chất lượng của SIB trong bộ mã hóa DSVC.

#### A. Tạo các ứng viên thông tin phụ trợ

Trong bộ mã hóa DSVC-Fusion này, có hai ứng viên SIB được tạo ra qua



khai thác thông tin tương quan thời gian và tương quan không gian. Bằng cách khai thác tương quan thời gian, SI thứ nhất được tạo ra sử dụng kỹ thuật nội suy thời gian bù chuyển động (MCTI) được gọi là  $SIB_{MCTI}$ . Khai thác tương quan không gian, SI thứ hai được tạo ra bằng cách tăng độ phân giải của khung hình lớp BL đã giải mã với tên gọi là  $SIB_{UpBL}$ .

Quá trình tạo ra  $SIB_{MCTI}$  bao gồm các bước sau:

- 1) *Lọc thông thấp*: Các KF trước và sau được lọc thông thấp để cải thiện độ tin cậy của vector chuyển động.
- 2) *Ước lượng chuyển động trước*: Một vector chuyển động được ước lượng cho mỗi khối trong KF sau với tham chiếu đến KF trước.
- 3) *Ước lượng chuyển động song hướng*: Sử dụng các hình chiếu của vector chuyển động, một vector chuyển động đối xứng được lựa chọn cho mỗi khối trong khung hình SI bằng cách lựa chọn vector nào đi qua gần với tâm khối nhất. Sau đó, vector chuyển động đối xứng được chia tách thành vector chuyển động trước và vector chuyển động sau với giả định chuyển động không đổi. Hai vector chuyển động được lọc tiếp trong một vùng nhỏ để giữ tính đối xứng trong quá trình lọc này.
- 4) *Lọc không gian*: Để loại bỏ các vector chuyển động ngoại lai, bộ lọc trung vị có trọng số được áp dụng cho hai trường vector chuyển động.
- 5) *Bù chuyển động song hướng*: Hai khối bù chuyển động trong các KF trước và sau được lấy trung bình để tạo ra SIB.

$SIB_{UpBL}$  được tạo ra bằng cách sử dụng cùng một bộ lọc tăng độ phân giải trong SHVC [7].

### **B. Cải thiện chất lượng cho $SIB_{MCTI}$**

Ứng viên  $SIB_{MCTI}$  được cải thiện chất lượng bằng cách sử dụng  $SIB_{UpBL}$  như một khung hình tham chiếu. Trong phương pháp này,  $SIB_{UpBL}$  được chọn làm khung hình tham chiếu vì hai lý do sau: (1) chất lượng của  $SIB_{UpBL}$  cao hơn so với khung hình MCTI và (2) khung hình  $SIB_{UpBL}$  sẵn có tại bộ giải mã.

Sau khi khung hình MCTI được tạo ra từ các KF trước và sau đã bù chuyển

động song hướng, các pixel của ba khung hình được so sánh với các pixel tương ứng trong khung hình đã tăng độ phân giải. Pixel ở khối nào có sai khác nhỏ nhất sẽ được coi là giá trị pixel mới của khung hình MCTI.

Giả định rằng  $P_{x,y}^B$ ,  $P_{x,y}^F$ ,  $P_{x,y}^M$ ,  $P_{x,y}^U$  là các giá trị pixel ở vị trí  $(x, y)$  tương ứng của các khối trong KF trước và sau,  $SIB_{MCTI}$  và  $SIB_{UpBL}$ . Giá trị mới của pixel trong khung hình MCTI,  $P_{x,y}^{New-M}$ , được tính như sau:

$$M = \min \{ |P_{x,y}^B - P_{x,y}^U|, |P_{x,y}^F - P_{x,y}^U|, |P_{x,y}^M - P_{x,y}^U| \} \quad (3.4)$$

$$P_{x,y}^{New-M} = \begin{cases} P_{x,y}^B & \text{ khi } M = P_{x,y}^B - P_{x,y}^U \\ P_{x,y}^F & \text{ khi } M = P_{x,y}^F - P_{x,y}^U \end{cases} \quad (3.5)$$

### C. Kết hợp các SIB ứng viên

Sau khi cải thiện chất lượng của  $SIB_{MCTI}$ , khối mới  $SIB_{New\_MCTI}$  này sẽ được kết hợp với  $SIB_{UpBL}$  để tạo ra khối mới  $SIB_{Fusion}$ . Cụ thể, các pixel của khối này được tính như sau:

$$P_{x,y}^{Fusion} = \alpha P_{x,y}^{New-M} + (1 - \alpha) P_{x,y}^U \quad (3.6)$$

Trong công thức 3.6,  $\alpha$  là trọng số để đo mức đóng góp của  $SIB_{New\_MCTI}$  và  $SIB_{UpBL}$  vào kết quả cuối cùng  $SIB_{Fusion}$ . Trong đề xuất này, giá trị  $\alpha$  được chọn qua thực nghiệm bằng 0,1.

### 3.2.3 Đánh giá hiệu năng

Phần này giới thiệu kết quả mô phỏng của phương pháp tạo SI đề xuất được thực hiện trên bốn chuỗi video thử nghiệm.

#### A. Điều kiện thử nghiệm

Việc đánh giá hiệu năng được thực hiện trên bốn chuỗi video với sự đa dạng về đặc tính chuyển động và cấu trúc bề mặt. Bảng 3.1 tổng kết một số đặc điểm chính của các chuỗi video này và các tham số lượng tử sử dụng cho BL và EL.

Bảng 3.1: Đặc điểm của các chuỗi video thử nghiệm

Chuỗi video	Độ phân giải không gian	Độ phân giải thời gian	Số lượng khung hình
Basketball Drill	832 × 480	50	50
BQMall		60	50
PartyScene		50	50
RaceHorses		30	50
Tham số lượng tử	Lớp BL: 32; Lớp EL: 30		

Bảng 3.2: PSNR [dB] của các phương pháp tạo SI

Chuỗi video	$SI_{MCTI}$	$SI_{UpBL}$	$SI_{Fusion}$
BasketballDrill	26,33	31,67	31,95
BQMall	27,70	28,12	28,36
PartyScene	25,27	24,99	25,31
RaceHorses	21,84	29,94	29,97

### B. Đánh giá chất lượng SI

Để đánh giá hiệu năng của phương pháp đề xuất, PSNR của khung hình SI được tạo ra bằng phương pháp đề xuất  $SI_{Fusion}$  được so sánh với khung hình SI được tạo ra bằng kỹ thuật MCTI,  $SI_{MCTI}$  và khung hình SI được tạo ra bởi khung hình lớp BL sau khi tăng độ phân giải  $SI_{UpBL}$ .

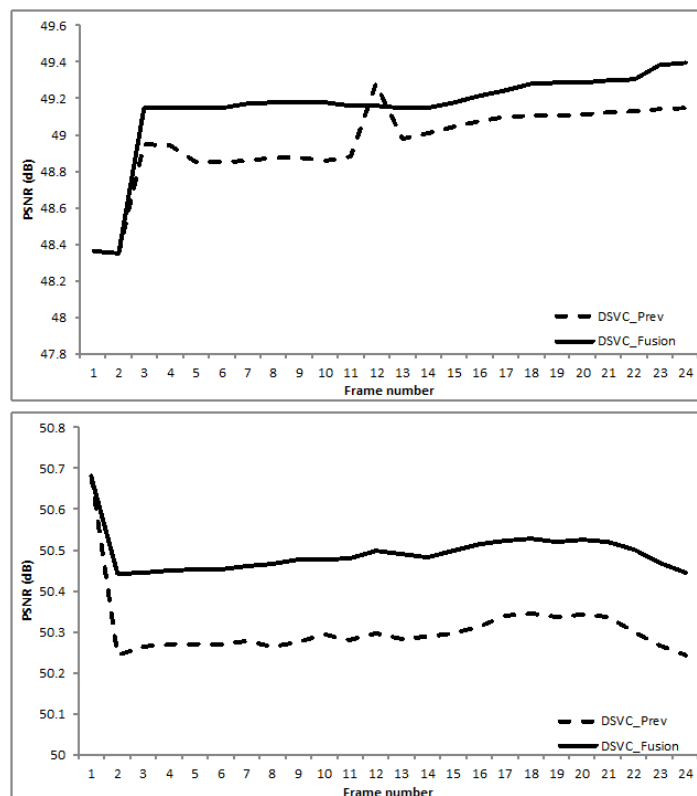
Các kết quả được trình bày trong Bảng 3.2. Kết quả cho thấy chất lượng của  $SI_{UpBL}$  thường tốt hơn chất lượng của  $SI_{MCTI}$  trừ chuỗi *PartyScene*. Sau quá trình kết hợp, chất lượng của  $SI_{Fusion}$  đạt được tốt hơn so với các SI ứng viên khác cho tất cả các chuỗi video thử nghiệm.

### C. Đánh giá chất lượng khung hình giải mã

PSNR của WZF tái tạo khi sử dụng các phương pháp tạo SI khác nhau được trình bày trong Bảng 3.3 và mô tả trong Hình 3.6. Các kết quả chỉ ra rằng  $DSVC_{Fusion}$  tốt hơn  $DSVC_{Previous}$  cho hầu hết các chuỗi ngoại trừ chuỗi

Bảng 3.3: Chất lượng của các khung hình giải mã [dB]

Chuỗi video	$DSVC_{Previous}$	$DSVC_{Fusion}$
BasketballDrill	48,95	49,14
BQMall	50,30	50,49
PartyScene	49,91	50,04
RaceHorses	51,92	51,79



Hình 3.6: PSNR của các khung hình tái tạo cho chuỗi BasketBall Drill và BQMall

*RaceHorse*. Tổng thể, phương pháp đề xuất đạt được mức cải thiện trung bình tốt hơn. Vì thế, chất lượng của khung hình tái tạo được cải thiện.

### 3.3 Đề xuất 4: Tạo thông tin phụ trợ dựa trên kỹ thuật SVM

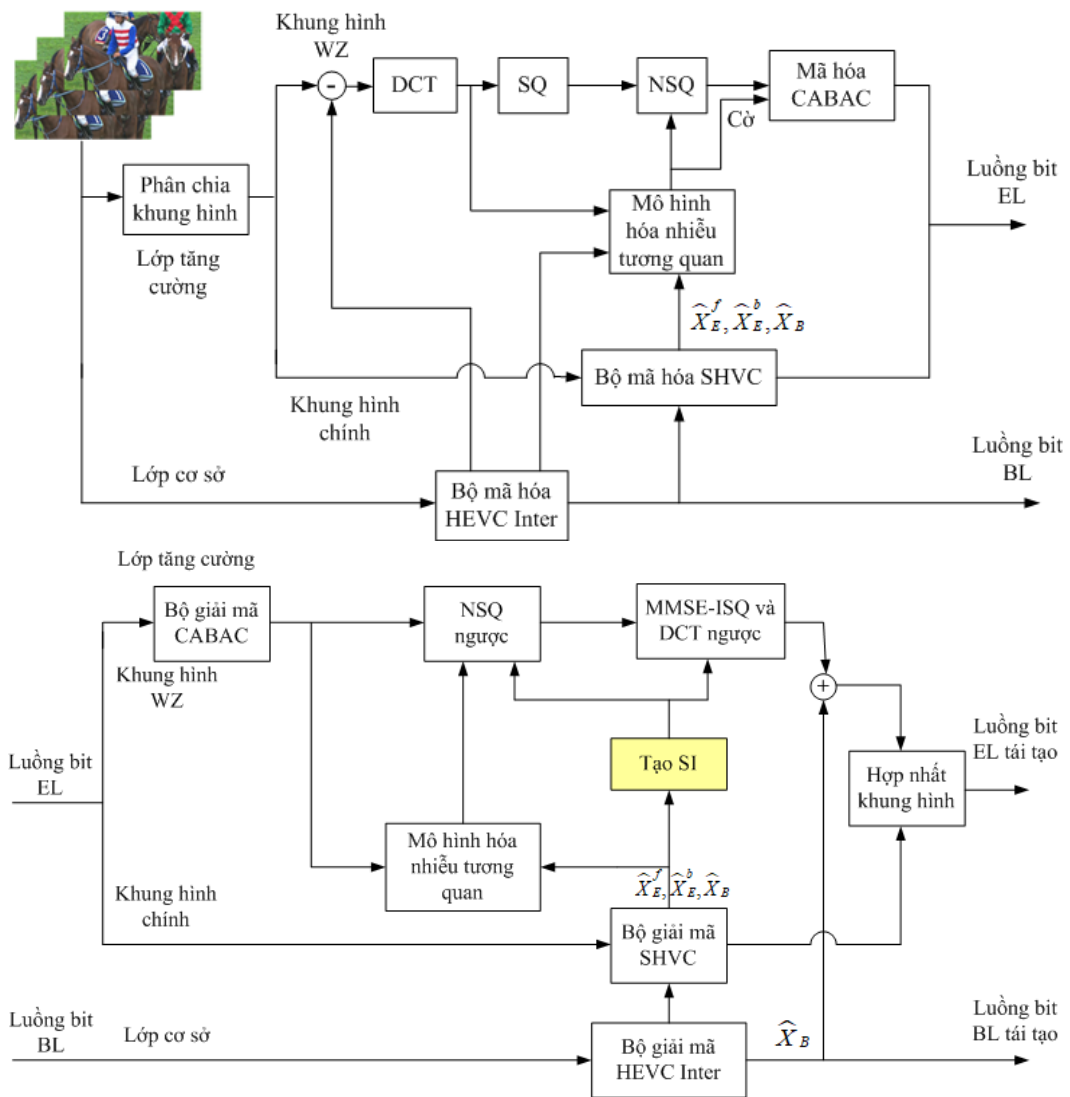
#### 3.3.1 Đặt vấn đề

Tạo thông tin phụ trợ vẫn là một trong những mô-đun quan trọng nhất của mã hóa video phân tán liên lớp vì nó quyết định đến chất lượng tái tạo lớp EL. Chất lượng SI càng cao thì càng cải thiện được hiệu năng nén của DSVC. Trong ngữ cảnh của mã hóa video liên lớp, có thể tạo ra SI với chất lượng tốt hơn nếu khai thác được đầy đủ các thông tin sẵn có như trường chuyển động BL, và các khung hình BL và EL đã giải mã.

Vì vậy trong phần này đề xuất một phương pháp tạo SI mới bằng cách khai thác một cách hiệu quả thông tin chuyển động BL và khung hình EL giải mã. Các thông tin này được sử dụng linh hoạt để tạo ra ba SI ứng viên. Các SI ứng viên này thể hiện chất lượng vượt trội khác nhau trong các ngữ cảnh khác nhau, do đó cần một cơ chế phân loại thích hợp để chọn ra một SI ứng viên phù hợp trong hoàn cảnh cụ thể. Kỹ thuật phân loại được lựa chọn ở đây là kỹ thuật máy vec-tơ hỗ trợ (SVM) với quá trình huấn luyện trực tuyến. Điều này giúp cho việc phân loại thích ứng được với nội dung chuỗi video và đạt được kết quả chính xác hơn.

#### 3.3.2 Kiến trúc bộ mã hóa DSVC-SVM

Trước khi đi vào chi tiết của phương pháp tạo SI đề xuất cho DSVC, phần này sẽ giới thiệu kiến trúc tổng thể của bộ mã hóa đề xuất DSVC-SVM. Trong kiến trúc này có hai lớp, một lớp BL và một lớp EL. Các khung hình lớp BL được nén bằng bộ mã hóa HEVC Inter. Các khung hình BL giải mã cùng với dư



Hình 3.7: Sơ đồ kiến trúc bộ mã hóa và bộ giải mã video phân tán liên lớp DSVC-SVM

thừa lớp BL và thông tin chuyển động lớp BL được sử dụng để nén các khung hình EL. Trong kiến trúc này, dư thừa EL tương ứng với sự sai khác giữa WZF gốc và khung hình BL giải mã. Chỉ có phần dư thừa này được mã hóa và truyền tới bộ giải mã. Lượng dư thừa EL được mã hóa là số lượng các bit LSB,  $n_{LSB}$ , đây là phần không thể rút ra từ SI tại bộ giải mã. Quá trình mã hóa và giải mã có thể tóm tắt như dưới đây.

**A. Quá trình mã hóa**

- *Chia tách khung hình*: Các khung hình EL được chia thành WZF và KF. Các KF được nén bằng bộ mã hóa SHVC và các WZF được nén bằng nguyên lý phân tán.
- *Biến đổi DCT và lượng tử vô hướng (SQ)*: Đối với các WZF, dư thừa EL được tạo ra bằng cách trừ khung hình gốc cho khung hình BL giải mã. Dư thừa này được biến đổi DCT và lượng tử hóa với kích thước bước lượng tử EL tương ứng.
- *Mô hình hóa nhiễu tương quan tại bộ mã hóa*: Thông tin tương quan giữa dư thừa EL tại bộ mã hóa và dư thừa SI tại bộ giải mã được mô tả bằng số lượng bit LSB,  $n_{LSB}$  cho mỗi hệ số. Chi tiết về cách tính  $n_{LSB}$  được mô tả trong [25].
- *Lượng tử vô hướng lồng nhau (NSQ)*: Dựa trên  $n_{LSB}$  và giá trị dư thừa EL của mỗi hệ số DCT, kỹ thuật NSQ [38] được sử dụng để tạo ra syndrome (S) tương ứng.
- *Mã hóa CABAC*: Syndrome được tạo ra ở bước trước cùng với thông tin về CNM được mã hóa bằng mã entropy CABAC của HEVC.
- Cuối cùng, hai luồng bit WZF và KF được đóng gói độc lập để cung cấp khả năng liên lớp thời gian. Chúng kết hợp lại với nhau thành luồng bit EL.

## B. Quá trình giải mã

Các khung hình BL và EL với mức chất lượng khác nhau được giải mã như sau: trước tiên, khung hình BL được giải mã với bộ giải mã HEVC Inter. Sau đó, KF của lớp EL được giải mã bằng bộ giải mã SHVC. Cuối cùng, các WZF của lớp EL được giải mã bằng bộ giải mã theo các bước sau:

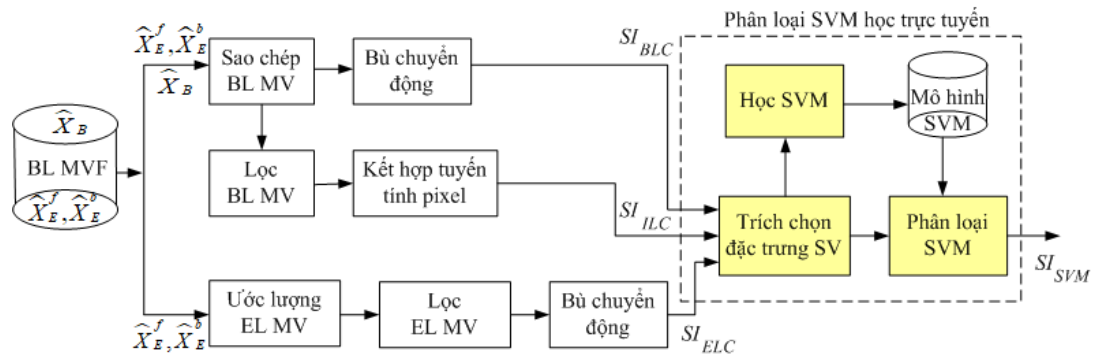
- *Giải mã CABAC*: Luồng bit WZ của EL được giải mã entropy bằng bộ giải mã CABAC. Thông tin giải mã bao gồm giá trị syndrome và một số thông tin bổ sung về mô hình hóa nhiễu tương quan.

- *Tạo dư thừa SI (SIR)*: Mô-đun tạo dư thừa SI đóng vai trò quan trọng trong DSVC bởi vì dư thừa SI được sử dụng để tái tạo lại dư thừa EL và dư thừa EL được cộng với khung hình BL để tạo ra khung hình WZ của EL. Khi chất lượng của dư thừa SI cao thì số lượng bit LSB cần mã hóa và gửi tới bộ giải mã càng ít và vì vậy làm tăng hiệu năng nén của DSVC. Trong kiến trúc DSVC-SVM này, một phương pháp tạo SI mới dựa trên SVM [24] được đề xuất. Chi tiết về phương pháp này sẽ được giới thiệu ở phần 3.3.3.
- *Mô hình hóa nhiều tương quan tại bộ giải mã*: Mô-đun này thực hiện tính giá trị  $n_{LSB}$  mà bên mã hóa đã gửi tới và về mặt lý tưởng, phải không có sự sai khác nào giữa hai mô-đun CNM tại bộ mã hóa và bộ giải mã. Trong kiến trúc này sử dụng cờ về chế độ CNM [25] mà bộ mã hóa gửi tới để tăng độ chính xác của quá trình dự đoán  $n_{LSB}$  tại bộ giải mã.
- *NSQ ngược*: NSQ ngược được sử dụng để tái tạo lại các hệ số đã lượng tử vô hướng (đã loại bỏ các MSB). Kiến trúc này sử dụng mô-đun NSQ ngược như trong [38].
- *Lượng tử hóa ngược và DCT ngược*: Trong bước này, các hệ số DCT được tái tạo sử dụng phương pháp sai số bình phương trung bình tối thiểu (MMSE) [67]. Sau đó, các hệ số DCT được biến đổi ngược để tạo ra dư thừa EL.
- Cuối cùng, dư thừa EL kết hợp cùng với khung hình BL giải mã để tái tạo lại WZF. Các WZF và KF lớp EL được kết hợp lại để tạo ra chuỗi video EL.

### 3.3.3 Phương pháp tạo thông tin phụ trợ dựa trên kỹ thuật SVM

Trước khi đi vào chi tiết của quá trình lựa chọn SI dựa trên kỹ thuật SVM, phần đầu sẽ mô tả kiến trúc tạo SI tổng thể và tiếp đó là phương pháp tạo các





Hình 3.8: Kiến trúc tạo SI

ứng viên SI.

Kiến trúc tạo SI gồm hai mô-đun chính là mô-đun tạo ba ứng viên SI và mô-đun lựa chọn ứng viên SI được mô tả trong Hình 3.8. Như được mô tả trong Hình 3.8, quá trình tạo SI gồm ba nhánh để tạo ra ba ứng viên SI và một quá trình lựa chọn một trong ba ứng viên.

### A. Quá trình tạo ứng viên SI

Ba SI ứng viên tương ứng với với các loại tương quan thời gian và tương quan liên lớp được tạo ra [25] và có thể được tổng kết như sau:

#### 1) Tạo SI dựa trên trường chuyển động BL, $SI_{BLC}$ :

Do mức tương quan cao giữa trường vector chuyển động (MVF) của lớp BL và lớp EL, BL MVF được sử dụng trực tiếp để tạo ra thông tin chuyển động EL và sau đó  $SI_{BLC}$  được tạo ra bằng cách thực hiện bù chuyển động (MC) dựa trên các KF lớp EL giải mã.

- Sao chép BL MV: Các khung hình BL được mã hóa với bộ mã hóa HEVC Inter và mỗi đơn vị mã hóa BL (BL CU) có thể chứa một hoặc nhiều PU, các PU này có thể được mã hóa ở chế độ Inter hoặc Intra. Vì vậy, tùy thuộc vào chế độ dự đoán được lựa chọn cho mỗi PU của BL, sẽ có hai trường hợp khi tạo ra trường vector chuyển động cho SI: i) Nếu BL PU liên quan được mã hóa liên ảnh, thông tin chuyển động của PU cùng vị trí

BL,  $mv_B^f, mv_B^b$ , đơn giản được sao chép tới SIB của EL hiện thời; ii) Nếu BL PU được mã hóa nội ảnh, PU được giải mã BL tương ứng sẽ được sử dụng làm  $SI_{BLC}$  và quá trình tạo SI dựa trên trường chuyển động BL sẽ kết thúc.

- Bù chuyển động: Sử dụng thông tin chuyển động từ bước trước, SI ứng viên  $SI_{BLC}$  được tạo ra bằng cách bù chuyển động dựa trên các khung hình EL trước và sau đã giải mã,  $\hat{X}_E^f, \hat{X}_E^b$ .

Bởi vì thông tin chuyển động được sử dụng ở đây đơn giản là trường chuyển động BL, nhánh SI này có xu hướng hiệu quả khi trường chuyển động BL có tương quan cao với trường chuyển động EL.

## 2) Tạo SI dựa trên lọc trường chuyển động BL, $SI_{ILC}$ :

Bởi vì việc sử dụng trực tiếp BL MVF có thể tạo ra SI không tin cậy, BL MVF sau khi sao chép vào EL như đã mô tả sẽ được lọc sử dụng tham chiếu BL và EL. Nhánh này sẽ sử dụng các thông tin chuyển động liên quan đến cả hai loại khung hình giải mã BL và EL hay nói cách khác việc tính SI ứng viên này có xem xét đến tương quan giữa các lớp (IL). Hơn nữa, bởi vì chất lượng của khung hình bù chuyển động có thể thấp nếu MVF được ước lượng không chính xác nên khung hình bù chuyển động sẽ được kết hợp tuyến tính mức pixel với khung hình BL đã giải mã để tạo nên ứng viên cuối cùng. Nhánh tạo  $SI_{ILC}$  làm việc như sau:

- Lọc BL MVF: Trước tiên, BL MVF được sao chép tới EL như đã mô tả trong nhánh tạo  $SI_{BLC}$ . Các MV này sau đó được lọc sử dụng các tham chiếu BL và EL. Trong trường hợp này, bắt đầu từ vị trí phù hợp liên quan đến vector chuyển động BL,  $mv_B^f, mv_B^b$  trong khung hình tham chiếu EL, quá trình lọc ME được áp dụng để tìm kiếm một vector chuyển động tốt hơn theo khía cạnh tối thiểu hóa tổng các sai khác bình phương (SSD) giữa khối BL cùng vị trí và khối EL được bù chuyển động, từ đó thu được  $mv_R^f, mv_R^b$ .

- Bù chuyển động: Hai vector chuyển động  $mv_R^f, mv_R^b$  được sử dụng cùng với tham chiếu EL để tạo ra hai khối bù chuyển động tương ứng. Các khối bù chuyển động được lấy trung bình để tạo ra  $P_{mc}$ .
- Kết hợp tuyến tính mức pixel: Cuối cùng, khung hình bù chuyển động  $P_{mc}$  được kết hợp tuyến tính ở mức pixel với khối BL đã giải mã để tạo ra  $SI_{ILC}$  như sau:

$$SI_{ILC}(x, y) = w(x, y) \times P_{mc}(x, y) + (1 - w(x, y)) \times \widehat{X}_B(x, y) \quad (3.7)$$

ở đó  $(x, y)$  là vị trí pixel trong khi  $w(x, y)$  là trọng số xác định mức đóng góp của  $P_{mc}$  và  $\widehat{X}_B$  đối với ứng viên  $SI_{ILC}$ .  $w(x, y)$  được tính như trong [26].

### 3) Tạo SI dựa trên lọc trường chuyển động EL, $SI_{ELC}$ :

Do sự vắng mặt của dữ liệu gốc nên trường vector chuyển động BL và quá trình lọc MV có thể không phải lúc nào cũng chính xác, vì vậy hai khung hình tham chiếu EL trước và sau được sử dụng để tạo ra EL MVF. Nhánh tạo SI này gồm các bước sau:

- Ước lượng EL MV: Thực hiện hai quá trình ước lượng chuyển động để tạo ra hai MVF: i) Ước lượng chuyển động trước: ME được áp dụng để tạo ra trường chuyển động trước, trường này xác định chuyển động từ khung hình tham chiếu EL trước đến khung hình tham chiếu EL sau ( $\widehat{X}_E^f \rightarrow \widehat{X}_E^b$ ); ii) Ước lượng chuyển động sau: ME được sử dụng để tạo ra trường chuyển động sau, trường này xác định chuyển động từ khung hình tham chiếu EL sau tới khung hình tham chiếu EL trước ( $\widehat{X}_E^b \rightarrow \widehat{X}_E^f$ ). Vector chuyển động liên quan tới mỗi MVF sẽ được coi là ứng viên cho mỗi khối không chồng lấn trong khung hình nội suy. Đối với mỗi khối trong khung hình nội suy, tất cả các vector ứng viên được xem xét và vector chuyển động nào giao với khung hình nội suy ở gần tâm của khối đang xem xét nhất sẽ được lựa chọn. Các khối nào không có vector chuyển động đi qua được coi là có MV bằng 0.

- Lọc EL MV: Mỗi SI MVF ban đầu được lọc bằng bộ lọc trung vị có trọng số để tạo ra hai trường MVF được làm mịn.
- Bù chuyển động: Sử dụng hai trường MVF được làm mịn ở trên để bù chuyển động song hướng cho hai khung hình chính EL đã giải mã  $\widehat{X}_E^f, \widehat{X}_E^b$  để tạo ra hai khối nội suy SI. Hai khối bù chuyển động này được lấy trung bình để tạo ra ứng viên SI thứ ba  $SI_{ELC}$ .

Bởi vì giải pháp này chỉ sử dụng thông tin EL đã giải mã để tạo ra ứng viên SI nên nó sẽ được lựa chọn khi chất lượng khung hình EL tốt hơn nhiều so với khung hình BL.

## B. Quá trình lựa chọn SI

Do sự vắng mặt của khung hình gốc, việc lựa chọn ứng viên SI trở nên khó khăn. Để lựa chọn một trong ba ứng viên SI một cách hiệu quả, quá trình lựa chọn SI được coi là một bài toán phân loại. Trong ngữ cảnh này, SVM [24] được lựa chọn để giải bài toán phân loại. Như được mô tả trong Hình 3.8, với ba khối ứng viên SI đầu vào, bộ phân loại SVM cần phải chọn ra một trong ba khối ứng viên đó. Tiêu chí để lựa chọn ở đây là phải chọn được khối SI ứng viên nào giống với khối tương ứng trong khung hình WZ gốc nhất. Việc lựa chọn ấy được quyết định dựa vào thuật toán SVM và mục tiêu cuối cùng là khi xây dựng lại hoàn chỉnh, SIF phải có chất lượng tốt.

Để quá trình chọn SI hiệu quả nhất cũng như để giảm thiểu chi phí tính toán, việc lựa chọn và định nghĩa tập các vector hỗ trợ (SV) cần phải được thực hiện một cách cẩn thận.

### 1) Xác định đặc trưng

Để lựa chọn một tập các đặc trưng SV cần phải xem xét các đặc điểm của mỗi nhánh tạo ứng viên SI. Các đặc trưng SV này phải mô tả đúng nhất chất lượng của mỗi nhánh ứng viên SI. Trong ngữ cảnh này, các đặc trưng sau được lựa chọn:

- Đặc trưng đánh giá trường MV: Ba đặc trưng SV đánh giá chất lượng của

trường vector chuyển động được tính toán sử dụng tổng các sai khác bình phương (SSD) giữa các khung hình trước và sau được bù chuyển động.  $SSD_{BLMV}$ ,  $SSD_{ILMV}$ ,  $SSD_{ELMV}$  liên quan đến BL MVF, BL MVF lọc và EL MVF.

$$SSD_{BLMV} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (\hat{X}_E^f(x, y, mv_B^f) - \hat{X}_E^f(x, y, mv_B^b))^2 \quad (3.8)$$

$$SSD_{ELMV} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (\hat{X}_E^f(x, y, mv_E^f) - \hat{X}_E^b(x, y, mv_E^b))^2 \quad (3.9)$$

$$SSD_{ILMV} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (\hat{X}_E^f(x, y, mv_R^f) - \hat{X}_E^b(x, y, mv_R^b))^2 \quad (3.10)$$

ở đó,  $(x, y)$  là vị trí pixel trong mỗi khối mã hóa với kích thước  $N \times N$ .

- Đặc trưng sai khác SI: Tính toán ba SSD liên quan đến sự sai khác giữa mỗi cặp SI ứng viên,  $SSD_{SIBE}$ ,  $SSD_{SIBI}$ ,  $SSD_{SIEI}$ .

$$SSD_{SIBE} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (SI_{BLC}(x, y) - SI_{ELC}(x, y))^2 \quad (3.11)$$

$$SSD_{SIEI} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (SI_{ELC}(x, y) - SI_{ILC}(x, y))^2 \quad (3.12)$$

$$SSD_{SIBI} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (SI_{BLC}(x, y) - SI_{ILC}(x, y))^2 \quad (3.13)$$

$SSD_{SIBE}$  đo là sự sai khác giữa  $SI_{BLC}$  và  $SI_{ELC}$ ,  $SSD_{SIBI}$  đo là sự sai khác giữa  $SI_{BLC}$  và  $SI_{ILC}$  và  $SSD_{SIEI}$  đo sự sai khác giữa  $SI_{ELC}$  và  $SI_{ILC}$ .

- Đặc trưng tương quan thời gian: Trong nhánh tạo  $SI_{BLC}$  và  $SI_{ELC}$ , các ứng viên SIB được tạo ra bằng cách sử dụng các KF của lớp EL trong khi  $SI_{ILC}$  được tạo ra sử dụng cả khung hình BL và EL. Do đó, tương quan thời gian đóng vai trò quan trọng đối với  $SI_{BLC}$  và  $SI_{ELC}$  hơn là  $SI_{ILC}$ . Vì vậy, tương quan thời gian (TC) là một đặc trưng tốt để phân biệt hai ứng

viên  $SI_{BLC}$  và  $SI_{ELC}$  với ứng viên  $SI_{ILC}$ . Đặc trưng tương quan thời gian  $SSD_{TC}$  được định nghĩa là sự sai khác giữa khung hình BL giải mã và các khối bù chuyển động của nó trong các khung hình BL tham chiếu,  $\hat{X}_B^f, \hat{X}_B^b$  như sau:

$$SSD_{TC} = \frac{SSD_{TC\_1} + SSD_{TC\_2}}{2} \quad (3.14)$$

$$SSD_{TC\_1} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (\hat{X}_B(x, y) - \hat{X}_E^f(x, y, mv_B^f))^2 \quad (3.15)$$

$$SSD_{TC\_2} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (\hat{X}_B(x, y) - \hat{X}_E^b(x, y, mv_B^b))^2 \quad (3.16)$$

$SSD_{TC}$  có xu hướng cao khi tương quan thời gian thấp. Trong trường hợp này,  $SI_{BLC}$  và  $SI_{ELC}$  sẽ thấp vì chúng phụ thuộc chủ yếu vào tương quan thời gian. Vì vậy, khi  $SSD_{TC}$  cao thì  $SI_{BLC}$  và  $SI_{ELC}$  sẽ không được lựa chọn.

- Đặc trưng tương quan giữa các lớp: Do  $SI_{ILC}$  được tạo ra sử dụng tương quan giữa các lớp, nên tương quan giữa các lớp cũng có ảnh hưởng quan trọng đến chất lượng ứng viên SI. Vì vậy, đặc trưng tương quan giữa các lớp được sử dụng trong quá trình huấn luyện và phân loại. Đặc trưng này là sự sai khác giữa các khối trong các KF lớp EL bù chuyển động và các khối trong các khung hình BL giải mã bù chuyển động ở cùng thời điểm với khung hình EL như sau:

$$SSD_{ILC} = \frac{SSD_{ILC\_1} + SSD_{ILC\_2}}{2} \quad (3.17)$$

$$SSD_{ILC\_1} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (\hat{X}_B^f(x, y, mv_B^f) - \hat{X}_E^f(x, y, mv_B^f))^2 \quad (3.18)$$

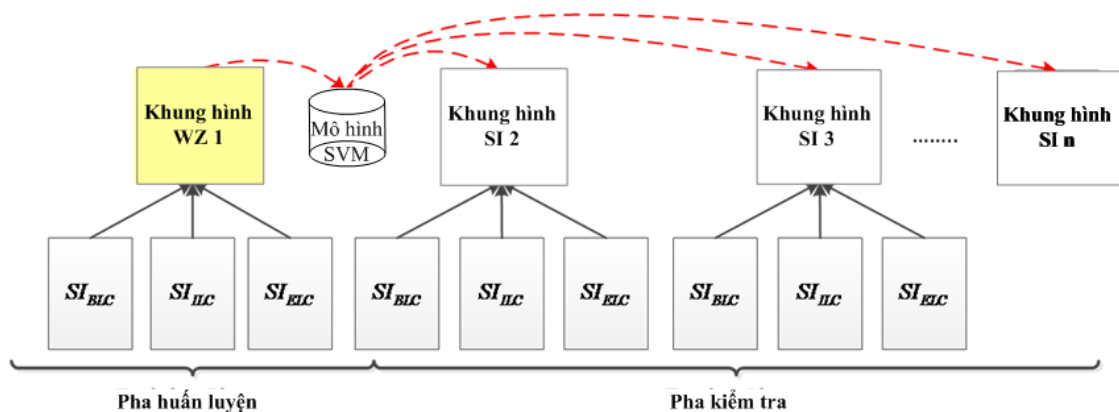
$$SSD_{ILC\_2} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (\hat{X}_B^b(x, y, mv_B^b) - \hat{X}_E^b(x, y, mv_B^b))^2 \quad (3.19)$$

Nhưng  $SSD_{ILC}$  đối lập với  $SSD_{TC}$ , điều đó có nghĩa rằng nếu  $SSD_{ILC}$  nhỏ thì  $SI_{ILC}$  được lựa chọn.

## 2) Huấn luyện mô hình SVM

Bài toán phân loại ở đây là với ba đầu vào là ba khối SI ứng viên  $SI_{BLC}$ ,  $SI_{ELC}$  và  $SI_{ILC}$  cần phải chọn đầu ra là một trong ba ứng viên đó.

Mô hình SVM này cần được huấn luyện trước khi sử dụng cho giai đoạn phân loại. Bởi vì nội dung video rất đa dạng, nên cần sử dụng SVM trực tuyến mà ở đó mô hình huấn luyện được tạo ra một cách thích ứng cho mỗi chuỗi video tùy thuộc vào nội dung thông tin của nó. Để làm được điều này, các đặc trưng SVM được mô tả như ở trên được trích xuất từ khung hình SI đầu tiên của mỗi chuỗi video thử nghiệm. Thông thường, giai đoạn huấn luyện được thực hiện định kỳ sau một khoảng thời gian nào đó. Tuy nhiên, để cân bằng giữa độ chính xác mô hình và thời gian xử lý để tạo SI nên trong đóng góp này, huấn luyện SVM được đề xuất thực hiện chỉ một lần, cụ thể là cho khung hình SI đầu tiên của mỗi chuỗi thử nghiệm. Quá trình huấn luyện đề xuất được minh họa trong Hình 3.9 được thực hiện ở mức khối như sau:



Hình 3.9: Quá trình huấn luyện trực tuyến SVM đề xuất cho mỗi chuỗi video

- **Bước 1.** Tạo ba SI ứng viên: Đối với khung hình WZ đầu tiên tại bộ giải mã, ba khối ứng viên SI tương ứng với mỗi khối của khung hình WZ được tạo ra trong kiến trúc tạo SI.

- **Bước 2.** Trích xuất đặc trưng: Với ba SI ứng viên này, tám đặc trưng SV tương ứng được trích xuất.
- **Bước 3.** Tạo nhãn tham chiếu: Khung hình WZ đầu tiên được giải mã và mỗi khối trong khung hình này được so sánh với ba khối SI ứng viên. Khối SI ứng viên nào có sai khác ít nhất với khối tương ứng trong khung hình WZ giải mã sẽ được coi là nhãn tham chiếu.
- **Bước 4.** Huấn luyện mô hình SVM: Tám đặc trưng SV cùng với nhãn tương ứng được đưa vào huấn luyện cho ba mô hình SVM tương ứng với ba cặp SI ứng viên bằng cách sử dụng phần mềm *SVM<sup>light</sup>* [30] với một số tham số tối ưu được thiết lập như sau:
  - *Hàm Kernel*: RBF với hệ số  $\gamma$  được chọn trong khoảng [0,01 100].
  - *Mức độ phạt đối với lỗi*: Hằng số C được lựa chọn trong khoảng [0,01 1000].

Độ chính xác thu được khi huấn luyện cho ba mô hình được mô tả trong Bảng 3.4.

Bảng 3.4: Độ chính xác của các mô hình huấn luyện

	<b>Độ chính xác (%)</b>	<b>Recall (%)</b>	<b>F1-score (%)</b>
<b>Mô hình 1</b>	97	95	96
<b>Mô hình 2</b>	93	97	95
<b>Mô hình 3</b>	91	89	90

### 3) Trích xuất các đặc trưng SV

Đối với các khung hình tiếp theo bắt đầu từ khung hình thứ 2, ba SI ứng viên lại được tạo ra và từ đó trích xuất ra tám đặc trưng SV tương ứng theo từng khối:

$$F = \left\{ \begin{array}{l} SSD_{BLMV}, SSD_{ELMV}, SSD_{ILMV}, SSD_{TC} \\ SSD_{ILC}, SSD_{SIBE}, SSD_{SIBI}, SSD_{SIEI} \end{array} \right\} \quad (3.20)$$



#### 4) Phân loại SVM

Ở bước này, việc phân loại SVM được thực hiện và đầu ra là ứng viên SI được lựa chọn. Thuật toán SVM ban đầu [24] được thiết kế để thực hiện phân loại nhị phân cho hai đầu vào trong khi bài toán ở đây là lựa chọn một trong số ba SI đầu vào. Vì vậy cần phải sử dụng thuật toán SVM nhiều lớp. Có nhiều cách để thực hiện thuật toán SVM nhiều lớp nhưng trong đề xuất này sử dụng cách tiếp cận SVM nhiều lớp dựa trên so sánh một - một (one versus one) vì phương pháp này có độ chính xác cao hơn so với một số cách tiếp cận khác ví dụ một - tất cả (one versus all) [27].

Trong đề xuất này sử dụng ba bộ phân loại SVM nhị phân cho ba cặp SI ứng viên với hai bước sau:

- Phân loại SVM nhị phân cho mỗi cặp SI: Trong bước này, ba quá trình phân loại SVM nhị phân được thực hiện cho mỗi cặp SI ứng viên:  $SI_{BLC}$  và  $SI_{ELC}$ ;  $SI_{BLC}$  và  $SI_{ILC}$ ;  $SI_{ILC}$  và  $SI_{ELC}$ .
- Tìm ra SI tốt nhất: Sau khi lựa chọn nhãn đầu ra cho ba bộ phân loại SVM nhị phân ở bước trước, ta được ba nhãn tương ứng với ba bộ phân loại. Nhãn nào được lựa chọn nhiều nhất sẽ được coi là SI cuối cùng. Trong trường hợp cả ba SI đều xuất hiện ở đầu ra ba bộ phân loại nhị phân thì nhãn tương ứng với đầu ra của bộ phân loại hai đầu vào  $SI_{BLC}$  và  $SI_{ILC}$  sẽ được chọn là SI cuối cùng bởi vì các kết quả thực nghiệm cho thấy giải pháp này cho chất lượng SI cao nhất.

### 3.3.4 Đánh giá hiệu năng

Trong phần này sẽ đánh giá chất lượng của SI được tạo ra bằng phương pháp đề xuất và so sánh với một số phương pháp tạo SI khác.

#### A. Điều kiện thử nghiệm

Đánh giá hiệu năng được thực hiện trên bốn chuỗi video thử nghiệm *Race-Horses*, *BlowingBubbles*, *Basketball Pass*, và *BQSquare* với sự đa dạng của các

Bảng 3.5: Đặc điểm của các chuỗi video thử nghiệm

Chuỗi video	Độ phân giải không gian	Độ phân giải thời gian	Số lượng khung hình
BlowingBubbles	416 × 240	50	497
BasketballPass		50	497
RaceHorses		30	297
BQSquare		60	599
Tham số lượng tử	$QP_B = 34; QP_E = QP_B - \{4; 6; 8; 10\}$		



Hình 3.10: Các khung hình đầu tiên của các chuỗi video thử nghiệm

đặc tính chuyển động và cấu trúc bề mặt. Đặc điểm chi tiết của các chuỗi video này được mô tả ở Bảng 3.5 và các khung hình đầu tiên của các chuỗi được minh họa ở Hình 3.10.

Phần mềm *SVM<sup>light</sup>* được sử dụng để lựa chọn các SI ứng viên. Phần mềm tham chiếu HEVC HM phiên bản 14.0 được sử dụng để mã hóa BL và phần mềm tham chiếu SHVC SHM phiên bản 6.0 được sử dụng để mã hóa các KF lớp EL.

### B. Đánh giá chất lượng tạo SI

Các giải pháp tạo SI được sử dụng để so sánh với phương pháp đề xuất *SI<sub>SVM</sub>* gồm *SI<sub>BLC</sub>*, *SI<sub>ELC</sub>*; *SI<sub>ILC</sub>* và *SI<sub>SPIC</sub>* [61]. Trong đó *SI<sub>BLC</sub>*, *SI<sub>ELC</sub>*; *SI<sub>ILC</sub>* là các SI được tạo ra dựa trên trường chuyển động lớp cơ sở, SI được tạo ra dựa trên lọc trường chuyển động lớp tăng cường và SI được tạo ra dựa trên lọc trường chuyển động lớp cơ sở tương ứng. *SI<sub>SPIC</sub>* là SI được tạo ra bằng giải pháp được đề cập trong [61].

Bảng 3.6: Chất lượng SI tạo bằng kỹ thuật SVM so với các phương pháp khác [dB]

Chuỗi video	$QP_B; QP_E$	$SI_{BLC}$	$SI_{ELC}$	$SI_{ILC}$	$SI_{SPIC}$	$SI_{SVM}$
<b>RaceHorses</b>	{34;30}	28,53	28,86	34,51	33,78	34,48
	{34;28}	28,65	29,02	34,83	34,08	34,90
	{34;26}	28,72	29,12	35,04	34,29	35,20
	{34;24}	28,77	29,18	35,17	34,42	35,41
<b>Trung bình</b>		<b>28,67</b>	<b>29,05</b>	<b>34,89</b>	<b>34,14</b>	<b>35,00</b>
<b>BlowingBubbles</b>	{34;30}	31,79	31,70	33,68	33,56	34,16
	{34;28}	32,45	32,38	34,33	34,36	35,25
	{34;26}	32,93	32,87	34,72	34,94	36,09
	{34;24}	33,28	33,23	34,94	35,37	36,73
<b>Trung bình</b>		<b>32,61</b>	<b>32,55</b>	<b>34,42</b>	<b>34,56</b>	<b>35,56</b>
<b>BasketballPass</b>	{34;30}	30,29	29,61	35,69	35,16	39,54
	{34;28}	30,45	29,81	35,97	35,54	36,31
	{34;26}	30,57	29,97	36,12	35,83	36,52
	{34;24}	30,67	30,11	36,21	36,06	36,77
<b>Trung bình</b>		<b>30,50</b>	<b>29,88</b>	<b>36,00</b>	<b>35,65</b>	<b>36,29</b>
<b>BQSquare</b>	{34;30}	33,4	31,22	33,19	32,45	33,78
	{34;28}	34,26	31,73	33,62	32,72	34,51
	{34;26}	34,93	32,12	33,85	32,92	35,11
	{34;24}	35,45	32,39	34,00	33,05	35,69
<b>Trung bình</b>		<b>34,51</b>	<b>31,87</b>	<b>33,67</b>	<b>32,79</b>	<b>34,77</b>
<b>Trung bình tổng cộng</b>		<b>31,57</b>	<b>30,83</b>	<b>34,74</b>	<b>34,28</b>	<b>35,41</b>
<b>Mức cải thiện PSNR</b>		<b>3,83</b>	<b>4,57</b>	<b>0,66</b>	<b>1,12</b>	

Bảng 3.6 mô tả các kết quả về chất lượng SI của phương pháp tạo SI đề xuất

và mức cải thiện PSNR [dB] của phương pháp tạo mới này so với các phương pháp tạo SI có liên quan.

Từ các kết quả ở Bảng 3.6, có thể rút ra một số kết luận như sau:

- Giải pháp tạo SI đề xuất đạt kết quả tốt hơn so với các phương pháp có liên quan. Điều này đến từ thực tế rằng phương pháp lựa chọn SI dựa trên SVM có thể lựa chọn một cách hiệu quả một SI từ một số các SI ứng viên.
- Giải pháp tạo SI đề xuất có thể đạt được SI với chất lượng tốt hơn, đáng chú ý là mức cải thiện trung bình lên đến 4,57 dB khi so sánh với  $SI_{ELC}$ .
- Chất lượng của các SI ứng viên có thể thay đổi tùy vào nội dung chuỗi video. Trong các chuỗi chuyển động nhanh và trung bình như *BasketballPass* and *BlowingBubbles*,  $SI_{ELC}$  vượt trội so với hai ứng viên còn lại. Còn đối với chuỗi chuyển động chậm như *BQSquare*,  $SI_{BLC}$  lại chiếm ưu thế. Điều đó cho thấy vai trò của mỗi ứng viên trong các ngữ cảnh khác nhau là khác nhau và cần một cơ chế lựa chọn phù hợp với nội dung chuỗi video.
- Thêm vào đó, Bảng 3.6 cũng cho thấy rằng phương pháp đề xuất đạt mức cải thiện cao trong trường hợp chuỗi video chuyển động chậm (ví dụ chuỗi *BQSquare*). Mức cải thiện chủ yếu đến từ việc sử dụng thông tin chuyển động BL và độ chính xác của giải pháp lựa chọn SI dựa trên học máy.

### 3.4 Kết luận chương

Trong chương này, Luận án đã đề xuất:

- Đề xuất 3: Phương pháp tạo thông tin phụ trợ dựa trên kỹ thuật kết hợp ảnh cho mã hóa video phân tán liên lớp không gian và thời gian. Trong đề xuất này, thông tin phụ trợ được tạo ra bằng cách khai thác thông tin tương quan thời gian và thông tin giữa các lớp. Các SI ứng viên này được

kết hợp lại với nhau để tạo nên thông tin phụ trợ cuối cùng có chất lượng tốt hơn.

- Đề xuất 4: Luận án cũng đã đề xuất một phương pháp tạo thông tin phụ trợ dựa trên kỹ thuật SVM cho mã hóa video phân tán liên lớp thời gian và chất lượng. Trong đề xuất này, ba ứng viên SI được tạo ra bằng cách khai thác các tương quan khác nhau bao gồm tương quan thời gian và tương quan giữa các lớp. Ba ứng viên này có chất lượng thay đổi phụ thuộc vào nội dung chuỗi video và việc lựa chọn ứng viên trong ngữ cảnh cụ thể được quyết định bởi thuật toán học máy SVM. Các kết quả mô phỏng cho thấy chất lượng của SI tạo ra bằng phương pháp đề xuất luôn tốt hơn so với các ứng viên đơn lẻ. Đó là nhờ sử dụng thuật toán SVM huấn luyện trực tuyến một cách hiệu quả. Cụ thể, phương pháp đề xuất có thể tăng chất lượng SI lên đến 4,57 dB nếu so với phương pháp tạo SI chỉ sử dụng thông tin sẵn có của lớp tăng cường.

Với các kết quả đạt được, có thể thấy tiềm năng của mã hóa video phân tán trong các bộ mã hóa video liên lớp. Việc sử dụng mã hóa video phân tán ở lớp tăng cường làm giảm sự phức tạp của bộ mã hóa trong khi vẫn duy trì được hiệu năng nén tương đương so với các bộ mã hóa video liên lớp truyền thống. Đây là một lợi thế quan trọng khi áp dụng cho các ứng dụng mới mà ở đó bộ mã hóa cần có sự phức tạp thấp.

#### **Các bài báo đã công bố:**

[4] (2017) N. T. H. Thao, V. Van San, and V. H. Tien, *Fusion Based Side Information Creation Method for Distributed Scalable Video Coding*, Tạp chí Khoa học và Công nghệ, vol. 121, pp. 48–53, 2017, ISSN 2354-1083.

[5](2017) X. Hoangvan and T. N. T. Huong, *An online SVM based side information creation for efficient distributed scalable video coding*, in International Conference on Advanced Technologies for Communications, 2017, vol. 2017-October, pp. 225–228, doi: 10.1109/ATC.2017.8167622.

# Kết luận

Trong mã hóa video phân tán, thách thức lớn đặt ra là làm thế nào đạt được hiệu năng nén gần với các hệ thống mã hóa video dự đoán tốt nhất hiện nay trong khi vẫn duy trì được các ưu điểm ban đầu là độ phức tạp mã hóa thấp và khả năng chống nhiễu tốt. Luận án này tập trung vào các kỹ thuật nhằm cải tiến hiệu năng nén của mã hóa video phân tán và mã hóa video phân tán liên lớp. Các kết quả nghiên cứu chính của luận án bao gồm các nội dung sau đây:

1. Cải tiến hiệu năng nén cho mã hóa video phân tán tại phía mã hóa với mô-đun thay đổi kích thước nhóm ảnh. Không sử dụng kích thước GOP cố định như các bộ mã hóa khác, bộ mã hóa DVC đề xuất lựa chọn kích thước  $GOP=2$  hay  $GOP=4$  tùy thuộc vào nội dung của từng phân đoạn video gồm 5 khung hình video. Việc quyết định lựa chọn kích thước GOP nào được dựa trên giải thuật cây quyết định. Các kết quả mô phỏng cho thấy bộ mã hóa DVC đề xuất có thể tiết kiệm tốc độ bit so với khi sử dụng kích thước GOP cố định cho toàn bộ chuỗi video.
2. Cải tiến hiệu năng nén cho mã hóa video phân tán tại phía giải mã. Luận án đã đề xuất phương pháp tạo thông tin phụ trợ mới dựa trên lọc liên tục trong quá trình giải mã và phương pháp ước lượng tham số  $\alpha$  cho mô hình nhiễu tương quan dựa trên mạng nơ-ron. Nhờ việc lọc thông tin phụ trợ trong suốt quá trình giải mã, chất lượng SI cuối cùng được cải thiện. Đối với phương pháp ước lượng tham số  $\alpha$  dựa trên mạng nơ-ron, các kết quả

cho thấy  $\alpha$  được ước lượng một cách chính xác hơn so với phương pháp trước đó. Khi kết hợp hai mô-đun vào bộ mã hóa, các kết quả mô phỏng cho thấy hiệu năng nén tổng thể của bộ mã hóa DVC đề xuất đạt được kết quả tốt hơn khi so với một số bộ mã hóa video liên quan.

3. Cải tiến hiệu năng nén cho mã hóa video phân tán liên lớp bằng cách cải tiến chất lượng của thông tin phụ trợ. Có hai phương pháp tạo thông tin phụ trợ được đề xuất cho DSVC. Phương pháp thứ nhất đề xuất sử dụng kỹ thuật kết hợp ảnh để kết hợp hai SI ứng viên cho mã hóa video phân tán liên lớp thời gian và không gian. Phương pháp thứ hai đề xuất sử dụng thuật toán SVM huấn luyện trực tuyến để lựa chọn một trong số ba SI ứng viên cho mã hóa video phân tán liên lớp thời gian và chất lượng. Các kết quả mô phỏng cho thấy cả hai phương pháp tạo SI cho DSVC đều có chất lượng tốt hơn so với các ứng viên riêng rẽ cũng như tốt hơn so với một số phương pháp trước đó.

### **Hướng nghiên cứu tiếp theo**

Kiến trúc DVC đã cho thấy khả năng ứng dụng trong một dải rộng, từ mã hóa video đơn góc nhìn cho đến đa góc nhìn, từ khả năng chống lỗi mạnh mẽ cho đến mã hóa video liên lớp, và vì vậy, cần phải có nghiên cứu đầy đủ và sự khai thác triệt để sơ đồ mã hóa video này. Để cải thiện hơn nữa hiệu năng của các codec video này sao cho đạt được hiệu năng nén tương đương với các chuẩn mã hóa video mới nhất và các ứng dụng khác thì các nghiên cứu tiếp theo là vô cùng cần thiết.

Các nghiên cứu tiếp theo có thể tập trung vào khai thác hơn nữa khả năng chống lỗi mạnh mẽ của kiến trúc này và các kịch bản sử dụng nguyên tắc DVC trong hệ thống mã hóa video đa góc nhìn. Đây là những lợi thế rất lớn của DVC nhưng chưa được khai thác nhiều trong các hệ thống mã hóa video. Nghiên cứu sinh cũng mong muốn được phát triển các bộ mã hóa video phân tán đề xuất trên các hệ thống nhúng trong tương lai.

# Các công trình khoa học

[1] (2018) **T. N. Thi Huong**, H. Phi Cong, T. V. Huu, and X. Hoang Van, *Artificial Intelligence Based Adaptive GOP Size Selection for Effective Wyner-Ziv Video Coding*, International Conference on Advanced Technologies for Communications, vol. 2018-October, pp. 120–124, 2018, doi: 10.1109/ATC.2018.8587573.

[2] (2019) T. V. Huu, **T. Nguyen Thi Huong**, M. N. Ngoc, and X. Hoangvan, *Improving performance of distributed video coding by consecutively refining of side information and correlation noise model*, Proceedings - 2019 19th International Symp Symposium on Communications and Information Technologies (ISCIT 2019), pp. 502–506, 2019, doi: 10.1109/ISCIT.2019.8905187.

[3] (2020) Tien Vu Huu, **Thao Nguyen Thi Huong**, Xiem Hoang Van, San Vu Van, *Improving TDWZ Correlation Noise Estimation: A Deep Learning based Approach*, REV Journal on Electronics and Communications: Article scheduled for publication in Vol. 10, No. 1–2, January–June, pp. 11-20, 2020, ISSN 1859 – 378X.

[4] (2017) **N. T. H. Thao**, V. Van San, and V. H. Tien, *Fusion Based Side Information Creation Method for Distributed Scalable Video Coding*, Tạp chí Khoa học và Công nghệ, vol. 121, pp. 48–53, 2017, ISSN 2354-1083.

[5] (2017) X. Hoangvan and **T. N. T. Huong**, *An online SVM based side information creation for efficient distributed scalable video coding*, in Interna-



tional Conference on Advanced Technologies for Communications, 2017, vol. 2017-Octob, pp. 225–228, doi: 10.1109/ATC.2017.8167622.

# Bibliography

- [1] A. Aaron, S. Rane, and B. Girod (2004), “Wyner-Ziv video coding with hash-based motion compensation at the receiver”, *International Conference on Image Processing, ICIP '04., Singapore, pp. 3097–3100.*
- [2] A. Aaron, R. Zhang, and B. Girod (2002), “Coding of Motion Video”, *Asilomar Conf. Signals, Syst. Comput. Pacific Grove, CA, pp. 240–244.*
- [3] A. Aaron et al. (2004), “Transform-domain Wyner-Ziv codec for video”, *Vis. Commun. Image Process. 2004, vol. 5308, p. 520.*
- [4] A. Abou-Elailah et al. (2012), “Successive refinement of side information using adaptive search area for long duration GOPs in distributed video coding”, *19th Int. Conf. Telecommun. ICT 2012, no. Ict.*
- [5] I. Ahmad, Z. Ahmad, and I. Abou-Faycal (2009), “Content adaptive GOP size control with feedback channel suppression in distributed video coding”, *Proceedings - International Conference on Image Processing, ICIP, pp. 1397–1400.*
- [6] I. Ahmad, Z. Ahmad, and I. Abou-Faycal (2009), “Delay-efficient GOP size control algorithm in Wyner-Ziv video coding”, *IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2009, pp. 403–407.*

- 
- [7] J. M. Boyce et al. (2016), “Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 20–34.
- [8] J. Q. Pedro et al. (2007), “Studying Error Resilience Performance for a Feedback Channel Based Transform Domain Wyner-Ziv Video Codec”, *Picture Coding Symposium, Lisbon, Portugal, 2007*, no. November.
- [9] X. Artigas et al. (2007), “The DISCOVER codec: architecture, techniques and evaluation”, *Proc. Pict. Coding Symp.*, pp. 6–9.
- [10] J. Ascenso, C. Brites, and F. Pereira (2005), “Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding”, *5th EURASIP Conf. Speech Image Process. Multimed. Commun. Serv.*, no. Dvc, pp. 1–6.
- [11] J. Ascenso, C. Brites, and F. Pereira (2006), “Content Adaptive Wyner-ZIV Video Coding Driven by Motion Activity”, *International Conference on Image Processing*, pp. 605–608.
- [12] B.G.Haskell (1997), “Digital video: an introduction to MPEG-2”, *Springer*.
- [13] G. Bjontegaard (2001), “Calculation of average PSNR differences between RD curves”, *Doc. VCEG-M33, 13th ITU-T VCEG Meeting, Austin, TX, USA*.
- [14] C. Brites, J. Ascens, and F. Pereira (2006), “Studying temporal correlation noise modeling for pixel based Wyner-Ziv video coding”, *International Conference on Image Processing, Atlanta, GA.*, pp. 273–276.
- [15] C. Brites, J. Ascenso, and F. Pereira (2006), “Improving transform domain Wyner-Ziv video coding performance”, *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2, pp. 525–528.

- 
- [16] C. Brites, J. Ascenso, and F. Pereira (2006), “Modeling correlation noise statistics at decoder for pixel based Wyner-Ziv video coding”, *25th PCS Proc. Pict. Coding Symp.*
- [17] C. Brites and F. Pereira (2005), “Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding”, *EURASIP Conf. Speech Image Process. Multimed. Commun. Serv. Smolenice, Slovak Repub., no. January.*
- [18] C. Brites and F. Pereira (2008), “Correlation noise modeling for efficient pixel and transform domain Wyner-Ziv video coding”, *IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 9, pp. 1177–1190.*
- [19] G. Cote et al. (1998), “H. 263+: Video coding at low bit rates”, *Circuits and Systems for Video Technology, IEEE Transactions on.*
- [20] B. Dash et al. (2018), “Multi-resolution extreme learning machine-based side information estimation in distributed video coding”, *Multimed. Tools Appl., vol. 77, no. 20, pp. 27301–27335.*
- [21] K. DinhQuoc, X. HoangVan, and B. Jeon (2011), “An iterative algorithm for efficient adaptive GOP size in transform domain Wyner-Ziv video coding”, *Ho YS. Adv. Image Video Technol. PSIVT 2011. Lect. Notes Comput. Sci. vol 7088. Springer, Berlin, Heidelb., vol. 7088 LNCS, no. PART 2, pp. 347–358.*
- [22] B. Girod et al. (2005), “Distributed video coding”, *Proc. IEEE, vol. 93, no. 1, pp. 71–83.*
- [23] C. Guillemot and F. Pereira (2007), “Distributed Monoview and Multiview Video Coding”, *IEEE Signal Process. Mag., vol. 24, no. 5, pp. 67–76.*
- [24] S. R. Gunn (1998), “Support vector machines for classification and regression”, *Technical Report, University of Southampton, Southampton, UK.*

- [25] X. HoangVan, J. Ascenso, and F. Pereira (2017), “Adaptive Scalable Video Coding: An HEVC-Based Framework Combining the Predictive and Distributed Paradigms”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 8, pp. 1761–1776.
- [26] X. Hoangvan, J. Ascenso, and F. Pereira (2015), “Improving enhancement layer merge mode for HEVC scalable extension”, *Picture Coding Symposium, PCS 2015 - with 2015 Packet Video Workshop, PV 2015 - Proceedings*, pp. 15–19.
- [27] Chih-Wei Hsu and Chih-Jen Lin (2002), “A comparison of methods for multiclass support vector machines”, *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425,
- [28] X. Huang and S. Forchhammer (2012), “Cross-band noise model refinement for transform domain Wyner-Ziv video coding”, *Signal Process. Image Commun.*, vol. 27, no. 1, pp. 16–30.
- [29] N. Imran, B. C. Seet, and A. C. M. Fong (2015), “Distributed video coding for wireless video sensor networks: a review of the state - of - the - art architectures”, *Springerplus*, vol. 4.
- [30] T. Joachims (2008), “Support Vector Machine: SVM-light”, Available: <http://svmlight.joachims.org/>.
- [31] D. Kubasov, J. Nayak, and C. Guillemot (2007), “Optimal reconstruction in Wyner-Ziv video coding with multiple side information”, *2007 IEEE 9Th Int. Work. Multimed. Signal Process. MMSP 2007 - Proc.*, pp. 183–186.
- [32] D. Kubasov et al. (2006), “Mesh-Based Motion-Compensated Interpolation for Side Information Extraction in Distributed Video Coding”, *International Conference on Image Processing, Atlanta, GA, 2006*, pp. 261–264.
- [33] M. Liou (1991), “Overview of the p x 64 kbit/s video coding standard”, *Communications of the ACM*, vol. 34, no. 4, pp. 59–63.

- 
- [34] H. Van Luong, L. L. Raket, and S. Forchhammer (2014), “Re-estimation of motion and reconstruction for distributed video coding”, *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 2804–2819.
- [35] R. Martins et al. (2009), “Refining side information for improved transform domain wyner-ziv video coding”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 9, pp. 1327–1341.
- [36] R. Martins et al. (2010), “Statistical motion learning for improved transform domain Wyner-Ziv video coding”, *IET Image Process.*, vol. 4, no. 1, pp. 28–41.
- [37] T. Maugey et al. (2010), “Using an exponential power model for Wyner Ziv video coding”, *IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, pp. 2338–2341.
- [38] S. Milani, J. Wang, and K. Ramchandran (2007), “Achieving H.264-like compression efficiency with distributed video coding”, *Proc. SPIE 6508, Vis. Commun. Image Process.*, vol. 6508, pp. 1013–1024.
- [39] W. Miled et al. (2009), “Image interpolation with dense disparity estimation in multiview distributed video coding”, *Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC), Como*, pp. 1–6.
- [40] M. Ouaret, F. Dufaux, and T. Ebrahimi (2009), “Error-resilient scalable compression based on distributed video coding”, *Signal Process. Image Commun.*, vol. 24, no. 6, pp. 437–451.
- [41] J. Park et al. (2009), “Wyner-Ziv video coding with region adaptive quantization and progressive channel noise modeling”, *IEEE Int. Symp. Broadband Multimed. Syst. Broadcast. BMSB 2009*.
- [42] F. Pereira and J. Ascenso (2007), “Adaptive Hash-Based Side Information Exploitation for Efficient Wyner-Ziv Video Coding”, *International Conference on Image Processing, San Antonio, TX*, vol. 9, pp. 29–32.

- [43] R. Puri, A. Majumdar, and K. Ramchandran (2007), “PRISM: A video coding paradigm with motion estimation at the decoder”, *IEEE Trans. Image Process.*, vol. 16, no. 10, pp. 2436–2448.
- [44] R. Puri and K. Ramchandran (2002), “PRISM: A new robust video coding architecture based on distributed compression principles”, *Proceedings of the Allerton Conference on Communication Control and Computing*.
- [45] H. Qin and B. Song (2012), “Adaptive Correlation Noise Model for DC Coefficients in Wyner-Ziv Video Coding”, *ETRI J.*, vol. 34, pp. 190–198.
- [46] J. Ross Quinlan (1994), “C4.5: Programs for Machine Learning”, *Morgan Kaufmann Publishers, Mach Learn 16, PP. 235–240*.
- [47] I. E. Richardson (2004), “H.264 and MPEG-4 video compression: video coding for next-generation multimedia”, *Wiley.com*.
- [48] K. Sakomizu, T. Nishi, and T. Onoye (2012), “A hierarchical motion smoothing for distributed scalable video coding”, *Picture Coding Symposium*, pp. 209–212.
- [49] M. Salmistraro et al. (2014), “A robust fusion method for multiview distributed video coding”, *EURASIP J. Adv. Signal Process.*, no. December.
- [50] H. Schwarz, D. Marpe, and T. Wiegand (2007), “Overview of the Scalable Video Coding Extension of the H . 264 / AVC Standard”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120.
- [51] A. Sehgal, A. Jagmohan, and N. Ahuja (2004), “Scalable video coding using Wyner-Ziv codes”, *Picture Coding Symposium, San Francisco, CA., 2004*, vol. 6, no. 1.
- [52] A. Sehgal, A. Jagmohan, and N. Ahuja (2004), “Wyner – Ziv Coding of Video: An Error-Resilient Compression Framework”, *IEEE Trans. Multimed.*, vol. 6, no. 2, pp. 249–258.
- [53] D. Slepian and J. K. Wolf (1973), “Noiseless Coding of Correlated Information Sources”, *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480.

- 
- [54] G. J. Sullivan et al. (2012), “Overview of the High Efficiency Video Coding”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668.
- [55] M. Tagliasacchi (2004), “A distributed-source-coding based robust spatio-temporal scalable video codec”, *Proc. Picture Coding Symposium*.
- [56] M. Tagliasacchi, M. Abhik, and K. Ramchandran (2004), “A distributed-source-coding based robust spatio-temporal scalable video code”, *Proc. Pict. Coding Symp.*
- [57] M. Tagliasacchi et al. (2006), “Intra Mode Decision Based on Spatio-Temporal Cues in Pixel Domain Wyner-ZIV Video Coding”, *IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, pp. 57–60.
- [58] B. Tian and W. Xiong (2018), “A Side Information Generation method using Deep Learning for Distributed Video Coding”, *J. Phys. Conf. Ser.*, vol. 1087, no. 6.
- [59] D. C. Tsai, C. M. Lee, and W. N. Lie (2006), “Dynamic key block decision with spatio-temporal analysis for Wyner-Ziv video coding”, *Proc. - Int. Conf. Image Process. ICIP*, vol. 6, no. Dvc, pp. 425–428.
- [60] X. Hoang Van, J. Ascenso, and F. Pereira (2014), “Correlation modeling for a distributed scalable video codec based on the HEVC standard”, *IEEE 16th International Workshop on Multimedia Signal Processing (MMSP)*, Jakarta, pp. 1–6.
- [61] X. Hoang Van, J. Ascenso, and F. Pereira (2015), “HEVC backward compatible scalability: A low encoding complexity distributed video coding based approach”, *Signal Process. Image Commun.*, vol. 33, pp. 51–70.
- [62] K. R. Vijayanagar and J. Kim (2011), “Dynamic GOP size control for low-delay distributed video coding”, *18th IEEE International Conference on Image Processing, Brussels.*, pp. 157-160.



- 
- [63] D. J. Le Gall W. B. Pannebaker and W. B. Pennebaker (1995), “MPEG1: Video compression standard”, *COMPCON Spring '91 Digest of Papers*.
- [64] H. Wang, N. M. Cheung, and A. Ortega (2006), “A framework for adaptive scalable video coding using Wyner-Ziv techniques”, *URASIP J. Appl. Signal Processing*, vol. 2006, pp. 1–18.
- [65] T. Wiegand et al. (2003), “Overview of the H.264/AVC Video Coding Standard”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576.
- [66] A. Wyner and J.Ziv (1976), “The Rate-Distortion Function for Source Coding with Side Information at the Decoder”, *IEEE Trans. Inf. Theory.*, vol. 22, no. 1, pp. 1–10.
- [67] Hoangvan Xiem, Ascenso Joao, and Pereira Fernando (2014), “Optimal reconstruction for a HEVC backward compatible distributed scalable video codec”, *IEEE Visual Communications and Image Processing Conference, VCIP 2014*, pp. 193–196.
- [68] Q. Xu and Z. Xiong (2006), “Layered Wyner–Ziv Video Coding”, *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3791–3803.