

ỨNG DỤNG MẠNG NƠ-RON TRONG KỸ THUẬT TẠO THÔNG TIN PHỤ TRỢ CHO MÃ HÓA VIDEO PHÂN TÁN

Nguyễn Thị Hương Thảo, Vũ Hữu Tiến

Học viện Công nghệ Bưu chính Viễn thông

Tóm tắt: Mã hóa video phân tán (DVC) là giải pháp đầy tiềm năng cho các ứng dụng đường lên như mạng giám sát video không dây hay mạng cảm biến đa phương tiện. Trong mã hóa video phân tán, chất lượng của thông tin phụ trợ (SI) có ảnh hưởng quan trọng đến hiệu năng tốc độ bit-chất lượng khung hình sau giải mã (RD) của hệ thống. Tuy nhiên, chất lượng thông tin phụ trợ thay đổi theo chuỗi và thậm chí trong mỗi khung hình. Vì vậy, để cải thiện thông tin phụ trợ, kỹ thuật tạo thông tin phụ trợ thích ứng với nội dung chuyển động của chuỗi video được đề xuất trong bài báo này. Cụ thể, bài báo đề xuất phương pháp sử dụng mạng Nơ-ron để dự đoán kích thước cửa sổ tìm kiếm cho thuật toán ước lượng chuyển động trong thuật toán tạo thông tin phụ trợ tương ứng với đặc tính của chuỗi video. Các kết quả mô phỏng cho thấy hiệu năng của bộ mã hóa video phân tán được cải thiện đáng kể khi sử dụng phương pháp này.

Từ khóa: Video phân tán, thông tin phụ trợ.

I. GIỚI THIỆU

Trong các chuẩn mã hóa video truyền thống như H.264/AVC hay HEVC [1][2], nhiệm vụ ước lượng và bù chuyển động đầy phức tạp được thực hiện tại bộ mã hóa. Do đó trong các hệ thống mã hóa video truyền thống, bộ mã hóa rất phức tạp trong khi bộ giải mã khá đơn giản. Kiến trúc này phù hợp cho hầu hết các ứng dụng kiểu đường xuống mà ở đó dữ liệu được mã hóa một lần và được giải mã hàng triệu lần tại các bộ giải mã. Tuy nhiên, gần đây nhiều thứ đã thay đổi. Nhiều ứng dụng đường lên mới như các mạng camera không dây đã xuất hiện. Dữ liệu video giờ đây có thể được tạo ra và xử lý trên các thiết bị đa phương tiện hạn chế. Trong các ứng dụng này, bộ mã hóa được yêu cầu có độ phức tạp thấp và vì vậy chuẩn mã hóa video truyền thống giờ đây không còn phù hợp nữa. Một giải pháp thay thế đầy tiềm năng là mã hóa video phân tán (DVC).

Nền tảng của DVC là mã hóa nguồn phân tán. Mã hóa nguồn phân tán được xây dựng trên hai định lý của lý thuyết thông tin: định lý Slepian-Wolf [3] và định lý Wyner-Ziv [4]. Định lý Slepian-Wolf phát biểu rằng khi hai tín hiệu phụ thuộc thống kê với nhau được mã hóa độc lập nhưng được giải mã kết hợp thì tốc độ bit đạt được cũng tương tự như đối với hệ thống mã hóa và giải mã kết hợp. Định lý Wyner-Ziv mở rộng định lý Slepian-Wolf trong trường hợp nén có tổn thất. Dựa trên cách tiếp cận này, nhiệm vụ ước lượng chuyển động phức tạp sẽ được dịch chuyển từ bộ mã hóa sang bộ giải mã.

Dựa trên các kết quả lý thuyết này, các kiến trúc DVC thực tế đã được đề xuất trong đó nổi lên hai kiến trúc chính là kiến trúc DVC Berkeley [5] và kiến trúc DVC Stanford [6]. Tuy nhiên, bộ mã hóa DISCOVER [7] được phát triển từ kiến trúc Stanford giờ đây được coi như một bộ mã hóa DVC tham chiếu chuẩn hiện nay. Trong bộ mã hóa DISCOVER, chuỗi video được chia tách thành các khung hình chính (KF) và khung hình Wyner-Ziv (WZF). Trong khi các KF được mã hóa sử dụng giải pháp mã hóa video truyền thống như H.264/AVC Intra hay HEVC Intra. Các WZF được mã hóa theo nguyên tắc mã hóa video phân tán như sau. Đầu tiên, biến đổi DCT dựa trên khối được áp dụng cho WZF. Sau biến đổi, các hệ số DCT của toàn bộ WZF được nhóm lại với nhau, tạo thành dải các hệ số DCT. Các dải này được tách ra thành các mặt phẳng bit và các mặt phẳng bit này được đưa tới bộ mã hóa kênh để tạo ra các bit chắn lẻ. Sau đó, các bit chắn lẻ này được gửi tới bộ giải mã theo từng cụm và các bit hệ thống được loại bỏ. Tại bộ giải mã, thông tin phụ trợ, một phiên bản nhiễu của WZF, được tạo ra bằng cách sử dụng các KF đã giải mã trước đó. Bộ giải mã sẽ sử dụng các bit chắn lẻ để “sửa” các sai lỗi trong thông tin phụ trợ và cuối cùng, WZF ban đầu được tái tạo.

Rõ ràng, SI có ảnh hưởng lớn đến hiệu năng RD của DVC. Nếu SI được tạo ra càng giống với WZF, bộ mã hóa chỉ phải gửi ít bit tới bộ giải mã để sửa sai cho SI và vì vậy, tốc độ bit sẽ giảm đi với cùng một chất lượng. Đó là lý do các nghiên cứu về tạo SI chiếm một phần lớn trong các nghiên cứu về DVC. SI thường được tạo ra bằng cách nội suy hoặc ngoại suy các khung hình tham chiếu. Các phương pháp nội suy sử dụng các KF đã giải mã trước và sau trong khi các phương pháp ngoại suy chỉ sử dụng các KF đã giải mã trước. Các kết quả mô phỏng cho thấy các kỹ thuật tạo SI dựa trên nội suy thường cho SI chất lượng tốt hơn nhưng độ trễ cao hơn khi so với các kỹ thuật tạo SI dựa trên ngoại suy. Tuy nhiên, các phương pháp này thường cho kết quả không tốt khi các khung hình tham chiếu ở cách xa nhau hoặc trong các trường hợp đặc tính chuyển động của chuỗi video nhanh hoặc bất thường.

Trong nghiên cứu thời kỳ đầu [6], SI đơn giản được tạo ra bằng cách lấy trung bình các KF được giải mã trước đó. Tuy nhiên, chất lượng của SI không tốt bởi vì không tính đến thông tin chuyển động trong quá trình tạo SI. Sau đó, các phương pháp tạo SI cải tiến được đề xuất và so sánh

Tác giả liên hệ: Nguyễn Thị Hương Thảo

Email: thaonh@ptit.edu.vn

Đến tòa soạn: 9/2020, chỉnh sửa: 10/2020, chấp nhận đăng: 10/2020

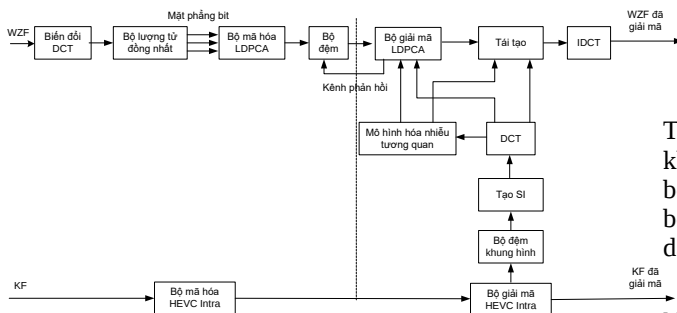
trong [8]. Năm 2005, Ascenso và các cộng sự đã phát triển một phương pháp nội suy thời gian bù chuyển động (MCTI) [9] với giải thuật làm mịn chuyển động để tạo ra SI. Phương pháp này được sử dụng rộng rãi trong các nghiên cứu về DVC và cũng được chấp nhận sử dụng trong bài báo này. Một cách tiếp cận khác là lọc SI ban đầu một số lần, thường là sau khi giải mã các mặt phẳng bit hoặc các dải hệ số [10][11]. Đề xuất trong [10] lọc các KF lân cận để lấy được vector chuyển động và đề xuất trong [11] tìm kiếm các ứng viên SI trong một cửa sổ cho trước trong SI ban đầu. Các giải pháp tạo SI dựa trên lọc liên tục thường cho SI cuối cùng có chất lượng tốt hơn nhưng có nhược điểm lớn là độ phức tạp bộ giải mã sẽ tăng lên.

Bởi vì nội dung chuyển động thay đổi theo dọc chuỗi, nó cần được tính đến trong quá trình tạo SI. Điều này có thể giúp tạo ra SI với chất lượng tốt hơn. Vì vậy, mục đích của bài báo này là đề xuất một phương pháp tạo SI thích ứng theo nội dung chuỗi video. SI ban đầu được tạo ra bằng cách sử dụng các khung hình tham chiếu trước và sau tương tự như cách tạo SI trong bộ mã hóa DISCOVER. Tuy nhiên, vùng tìm kiếm trong bước ước lượng chuyển động được điều chỉnh tùy theo nội dung chuyển động của chuỗi video. Đối với đoạn chuyển động nhanh hoặc phức tạp, rất khó để tạo ra SI chính xác. Do đó, SI cần được ước lượng chuyển động trong dải tìm kiếm rộng hơn để đạt được chất lượng tốt hơn. Ngược lại, đối với vùng chuyển động chậm, để dự đoán SI đúng và SI chỉ cần tìm kiếm trong vùng tìm kiếm nhỏ. Bài báo này đề xuất một phương pháp để lựa chọn vùng tìm kiếm phù hợp cho mỗi đoạn video dựa trên kỹ thuật học máy. Để đánh giá đề xuất này, các thử nghiệm được thực hiện trên bộ mã hóa DISCOVER với các KF được mã hóa sử dụng chuẩn HEVC Intra.

Bài báo này được tổ chức như sau: Phần II giới thiệu về kiến trúc bộ mã hóa video WZ miền biến đổi. Phần III mô tả giải pháp đề xuất và phần IV là thảo luận về hiệu năng RD của đề xuất và so sánh với các nghiên cứu liên quan. Cuối cùng, phần V giới thiệu các kết luận và hướng nghiên cứu tiếp theo.

II. KIẾN TRÚC MÃ HÓA VIDEO WZ MIỀN BIẾN ĐỔI

Phần này giới thiệu về bộ mã hóa WZ miền biến đổi với tên gọi DVC-HEVC. Khác với bộ mã hóa DISCOVER gốc [7], các KF trong bộ mã hóa này được mã hóa bằng bộ mã hóa HEVC-Intra. Vì vậy nó có tên gọi là DVC-HEVC và được mô tả trong Hình 1.



Hình 1. Kiến trúc bộ mã hóa DVC-HEVC

A. Quá trình mã hóa

Chuỗi video đầu vào được phân tách thành các KF X_{2t-1} và các WZF X_{2t} . Các KF được mã hóa bằng bộ mã hóa HEVC Intra và WZF được mã hóa theo nguyên tắc phân tán như sau:

Biến đổi DCT: Mỗi WZF được chia thành các khối không chồng nhau có kích thước 4×4 và biến đổi DCT được áp dụng cho mỗi khối để tạo thành các hệ số DCT tương ứng với các pixel trong mỗi khối.

Lượng tử hóa: Để giảm tốc độ bit, các hệ số DCT được lượng tử hóa. Các hệ số DCT được nhóm thành 16 dải hệ số b_k ($k = 0; 15$) ở đó mỗi dải hệ số gồm các hệ số có cùng vị trí trong các khối khác nhau. Các dải DCT này được lượng tử đồng nhất với bộ lượng tử 2^{M_k} ở đó M_k là số bit biểu diễn cho các hệ số DCT của dải hệ số b_k . Các ma trận lượng tử được chọn tương ứng với các mức tốc độ bit khác nhau như trong [7].

Bộ mã hóa LDPCA: Mỗi mặt phẳng bit được đưa vào bộ mã hóa LDPCA để tạo ra các bit chẵn lẻ tương ứng. Các bit chẵn lẻ này được lưu tại bộ đệm khung hình và các bit hệ thống bị loại bỏ. Các bit chẵn lẻ được truyền tới bộ giải mã theo từng cụm tùy theo yêu cầu từ bộ giải mã.

B. Quá trình giải mã

Tại bộ giải mã, các KF được giải mã bằng bộ giải mã HEVC Intra và các WZF được giải mã với sự trợ giúp của các KF này.

Tạo thông tin phụ trợ: Khối nội suy thời gian bù chuyển động tạo ra thông tin phụ trợ, một ước lượng của WZF ở bộ mã hóa, bằng cách sử dụng các KF đã giải mã trước và sau như trong [9]. Trong mô-đun này, hai khung hình tham chiếu được lọc thông thấp để giảm nhiễu cho quá trình ước lượng chuyển động. Sau đó thực hiện ước lượng chuyển động trước để có được vector chuyển động trước. Tiếp đó, ước lượng chuyển động song hướng lọc các vector chuyển động thu được ở bước trước với ràng buộc về phép chiếu tuyến tính giữa khung hình tham chiếu trước và sau. Để cải thiện độ mượt không gian, bộ lọc trung vị có trọng số sẽ được sử dụng. Cuối cùng, bù chuyển động song hướng được thực hiện để tạo ra thông tin phụ trợ, một ước lượng của WZF tương ứng.

Mô hình nhiễu tương quan: Thống kê dư thừa giữa các hệ số DCT của WZF gốc và khung hình SI tương ứng được giả định tuân theo phân bố Laplace như sau:

$$f_{X/Y}(x) = \frac{\alpha}{2} e^{-\alpha|x-y|} \quad (1)$$

ở đó $f_{X/Y}$ là hàm mật độ xác suất của X với y cho trước. α là tham số của phân bố Laplace được xác định bởi công thức sau:

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

Trong công thức (2), σ^2 là phương sai của dư thừa giữa khung hình WZF gốc và khung hình SI tương ứng. Trong bộ mã hóa được sử dụng ở bài báo này, tham số α của phân bố Laplace được ước lượng trực tuyến tại bộ giải mã ở mức dải hệ số.

Bộ giải mã LDPCA: Khi đã có các hệ số DCT của khung hình SI và nhiễu tương quan đối với một dải hệ số DCT, các mặt phẳng bit của dải hệ số b_k được giải mã LDPCA với sự hỗ trợ của các bit chẵn lẻ được gửi từ bộ mã hóa thông 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, tiêu chí dừng yêu cầu được sử dụng. Để phát hiện và sửa các lỗi còn lại trong mỗi mặt phẳng bit, tổng kiểm tra dư thừa vòng (CRC) được tính cho mỗi mặt

phẳng bit tại bộ mã hóa và gửi tới bộ giải mã. Bộ giải mã sẽ tính CRC tương ứng cho mỗi mặt phẳng bit đã giải mã LDPCA thành công và so sánh với CRC nhận được từ bộ mã hóa để đảm bảo loại bỏ xác suất lỗi cho mỗi mặt phẳng bit được giải mã. Sau khi một mặt phẳng bit được giải mã LDPCA thành công, các mặt phẳng bit còn lại của cùng một dải hệ số sẽ được xử lý tiếp theo cách tương tự.

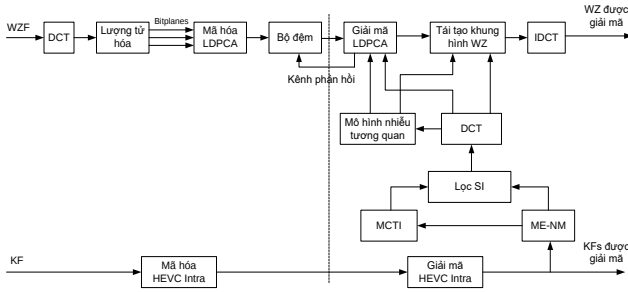
Tái tạo: Các mặt phẳng bit đã giải mã LDPCA cùng với khung hình SI và thông kê dư thừa cho mỗi dải hệ số DCT được sử dụng cùng nhau để tái tạo lại hệ số DCT ban đầu. Giá trị tái tạo cho mỗi hệ số DCT của mỗi dải hệ số DCT được cho trong [12]:

$$x' = E[x|q', y] = \frac{\int_l^u x f_{X/Y}(x|y) dx}{\int_l^u f_{X/Y}(x|y) dx} \quad (3)$$

ở đó x' là hệ số DCT tái tạo, y là hệ số DCT tương ứng của WZF và $E[\cdot]$ là toán tử kỳ vọng. l, u là các biên dưới và trên tương ứng của q' .

Biến đổi DCT ngược (IDCT): Cuối cùng, các hệ số DCT tái tạo được biến đổi DCT ngược để nhận được khung hình miền pixel.

III. PHƯƠNG PHÁP ĐỀ XUẤT



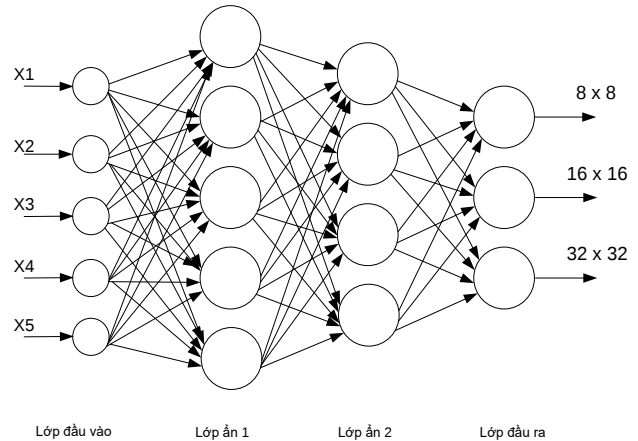
Hình 2. Kiến trúc bộ mã hóa ME-NN-MCTI-HEVC

Hình 2 mô tả bộ mã hóa video miền biến đổi dựa trên cấu trúc bộ mã hóa WZ được đề cập ở mục II. Thông tin phụ trợ được tạo ra bằng cách sử dụng các khung hình chính đã được giải mã trước đó. Chính vì vậy, các đặc tính chuyển động của các khung hình chính ảnh hưởng trực tiếp tới chất lượng của khung hình phụ trợ. Cụ thể, nếu các khung hình chính có chuyển động chậm, tức là độ tương quan giữa các khung hình lớn sẽ dẫn đến chất lượng của khung hình SI cao. Ngược lại, các khung hình chính có chuyển động nhanh dẫn đến chất lượng của SI bị suy giảm. Để đảm bảo chất lượng của khung hình SI trong những trường hợp chuyển động nhanh, phương pháp tạo khung hình SI thích ứng với nội dung video được đề xuất. Cụ thể, để tạo khung hình SI từ hai khung hình chính, thuật toán MCTI được sử dụng để nội suy khung hình SI dựa trên kỹ thuật ước lượng chuyển động và bù chuyển động song hướng. Trong kỹ thuật MCTI, ước lượng chuyển động là một trong những bước quan trọng nhằm xác định vector chuyển động. Sau đó, khối bù chuyển động song hướng dựa vào các vector chuyển động để tạo khung hình SI. Như vậy, nếu việc ước lượng chuyển động càng chính xác thì khung hình SI sẽ càng giống với khung hình WZ. Trong các bộ mã hóa video truyền thống, ước lượng chuyển động cho một khối hình được thực hiện thông qua việc tìm kiếm trên một cửa sổ tìm kiếm của khung hình tham chiếu và cửa sổ này có kích thước cố định. Kích thước của cửa sổ tìm kiếm có ảnh hưởng trực tiếp tới hiệu quả của việc giải mã. Nếu kích

thước cửa sổ lớn sẽ làm tăng thời gian tìm kiếm, trong khi kích thước nhỏ sẽ ước lượng chuyển động không chính xác. Vì vậy, trong bài báo này, phương pháp ước lượng vector chuyển động thích ứng với nội dung video dựa trên mạng Nơ-ron được đề xuất (ME-NN). Cụ thể, để phân loại mức độ chuyển động của khung hình, mạng ME-NN được sử dụng để dự đoán kích thước của cửa sổ tìm kiếm cho bộ ước lượng chuyển động dựa trên các đặc tính của khung hình. Kiến trúc mạng Nơ-ron được mô tả chi tiết trong phần dưới đây.

A. Kiến trúc mạng ME-NN

Trong nghiên cứu này, kiến trúc mạng NN được sử dụng để ước lượng kích thước của cửa sổ tìm kiếm (ME-NN) có 4 lớp bao gồm 1 lớp đầu vào, 2 lớp ẩn và 1 lớp đầu ra như mô tả trong hình 3. Lớp đầu vào nhận các giá trị X_1, X_2, X_3, X_4, X_5 tương ứng với các đặc trưng của khung hình bao gồm MAD, giá trị lớn nhất, nhỏ nhất, giá trị trung bình và phương sai của các hệ số DCT trong băng DC và ba băng của ba hệ số AC đầu tiên của các khối hình kích thước 4x4 trong khung hình. Trong các lớp ẩn, hàm kích hoạt sigmoid được sử dụng tại các nút mạng. Tại lớp đầu ra, hàm softmax được sử dụng để quyết định kích thước của cửa sổ tìm kiếm cho thuật toán MCTI.



Hình 3. Mô hình mạng ME-NN

B. Quá trình huấn luyện

Để huấn luyện mô hình, tập dữ liệu huấn luyện được trích chọn từ 10 chuỗi video Coastguard, Hall-Monitor, News, Container, Flower Garden, Mobile, Mother, Claire, Grandma và Harbour với độ phân giải 176x144 (QCIF), số lượng khung hình của mỗi chuỗi là 300 và tốc độ khung hình là 15 khung hình/giây. Đây là các chuỗi có nội dung đa dạng bao gồm chuyển động chậm, chuyển động nhanh, nội dung đơn giản, nội dung phức tạp. Để lấy ra các đặc trưng của các khung hình của chuỗi trên, mỗi khung hình sẽ được mã hóa 3 lần trên bộ mã hóa DVC với 3 cửa sổ tìm kiếm có kích thước là 8x8, 16x16 và 32x32. Một trong ba kích thước cửa sổ tìm kiếm cho hiệu năng nén cao nhất sẽ được chọn làm nhãn cho tập các tham số của khung hình đó. Mô hình ME-NN được huấn luyện trên nền tảng Google Colaboratory [12] với số epoch là 1000, batch-size bằng 10.

Mô hình mạng ME-NN sau khi huấn luyện sẽ được sử dụng trong thuật toán MCTI để tạo ra khung hình SI như sau:

Bước 1: Tính khung hình sai khác của 2 khung hình chính.

Bước 2: Tính các giá trị MAD, giá trị lớn nhất, nhỏ nhất, giá trị trung bình và phương sai của các hệ số DCT trong băng DC và ba băng của ba hệ số AC đầu tiên của các khối hình kích thước 4x4 của khung hình sai khác ở bước 1.

Bước 3: Các đặc trưng ở Bước 2 sẽ là đầu vào của mô hình ME-NN. Mô hình sẽ quyết định kích thước của cửa sổ tìm kiếm tương ứng với bộ đặc trưng đầu vào.

Bước 4: Với kích thước cửa sổ tìm kiếm ở Bước 2, thuật toán MCTI được sử dụng để tạo khung hình SI từ hai khung hình chính. Bốn giai đoạn của MCTI bao gồm: Ước lượng chuyển động hướng tiến, ước lượng chuyển động song hướng, lọc không gian và bù chuyển động song hướng.

IV. ĐIỀU KIỆN THỬ NGHIỆM VÀ PHÂN TÍCH KẾT QUẢ

A. Điều kiện thử nghiệm

Để đánh giá hiệu năng của giải pháp tạo thông tin phụ trợ đề xuất so với các phương pháp khác, bốn chuỗi video sau được sử dụng là Akiyo, Carphone, Foreman và Coastguard với các đặc điểm được mô tả trong Bảng I. Bốn cặp giá trị (tốc độ bit (Rate), độ méo (Distortion)) hay còn gọi là các điểm RD được đo tương ứng với ma trận lượng tử kích thước 4x4. Các phần tử của ma trận là các giá trị lượng tử tương ứng cho các hệ số của các giá trị DCT của các khối hình. Khi các giá trị lượng tử tăng thì tốc độ bit và chất lượng cũng tăng. Để cải thiện chất lượng video tại phía giải mã, các khung hình chính được mã hóa bằng bộ mã hóa H.265/HEVC ở chế độ Intra.

Bảng I. Điều kiện thử nghiệm

Chuỗi Video	Độ phân giải	Số khung hình	Tham số lượng tử
Carphone	176x144	300	{25,29,34,40}
Foreman			{25,29,34,40}
Akiyo			{25,29,34,40}
Coastguard			{26,30,34,38}

Tại mỗi điểm RD, tốc độ bit và giá trị PSNR được tính cho thành phần chói của mỗi khung hình. Các kết quả của phương pháp đề xuất với tên gọi NN-based-MCTI được mô phỏng trên hai bộ mã hóa DVC-HEVC và DVC-H.264 và được so sánh với một số phương pháp trước đây bao gồm: (1) Phương pháp sử dụng của sổ tìm kiếm thích ứng theo giá trị MAD-MCTI [13], (2) Phương pháp MCTI với cửa sổ tìm kiếm có kích thước cố định. Để đánh giá hiệu quả của các phương pháp, giá trị Bjontegaard được sử dụng để đo độ khác biệt giữa các đường RD của các phương pháp.

B. Kết quả

Bảng II. So sánh hiệu năng RD của phương pháp đề xuất ME-NN-MCTI-HEVC và MAD-MCTI-HEVC với phương pháp MCTI-HEVC theo BD-Rate

Chuỗi Video	MAD-MCTI-HEVC	ME-NN-MCTI-HEVC
Carphone	-21.44	-24.37
Foreman	-2.89	-7.89
Akiyo	-24.00	-29.52
Coastguard	-0.66	-11.56
Trung bình	-12.25	-18.33

Bảng III. NN-MCTI-HEVC và MAD-MCTI-HEVC với phương pháp MCTI-HEVC theo BD-PSNR

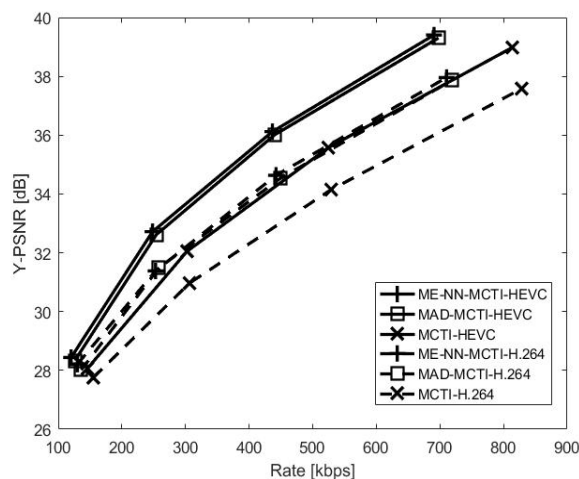
Chuỗi Video	MAD-MCTI-HEVC	ME-NN-MCTI-HEVC
Carphone	1.54	1.76
Foreman	0.19	0.53
Akiyo	1.88	2.36
Coastguard	0.04	0.72
Trung bình	0.91	1.34

Bảng IV. So sánh hiệu năng RD của phương pháp đề xuất ME-NN-MCTI-H.264 và MAD-MCTI-H.264 với phương pháp MCTI-H.264 theo BD-Rate

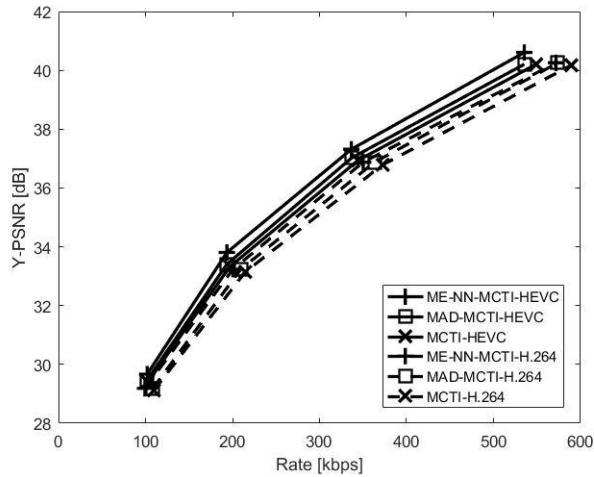
Chuỗi Video	MAD-MCTI-H.264	ME-NN-MCTI-H.264
Carphone	-20.59	-22.64
Foreman	-3.58	-7.60
Akiyo	-25.24	-27.16
Coastguard	-0.65	-2.89
Trung bình	-12.51	-15.07

Bảng V. So sánh hiệu năng RD của phương pháp đề xuất ME-NN-MCTI-H.264 và MAD-MCTI-H.264 với phương pháp MCTI-H.264 theo BD-PSNR

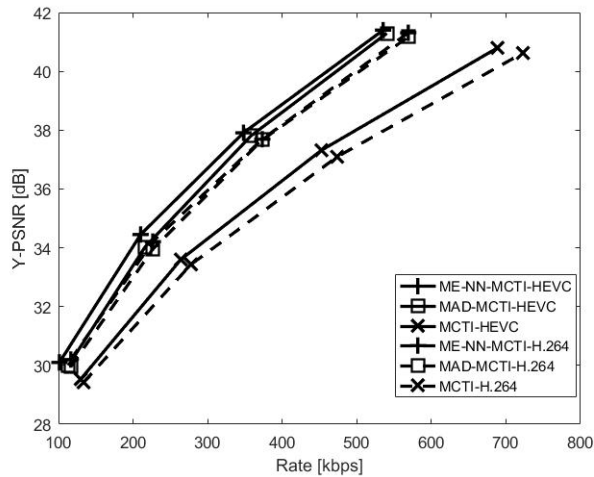
Chuỗi Video	MAD-MCTI-H.264	ME-NN-MCTI-H.264
Carphone	1.36	1.48
Foreman	0.24	0.50
Akiyo	1.96	2.13
Coastguard	0.04	0.16
Trung bình	0.9	1.07



Hình 4. So sánh hiệu năng RD của phương pháp đề xuất với chuỗi Carphone



Hình 5. So sánh hiệu năng RD của phương pháp đề xuất với chuỗi Foreman

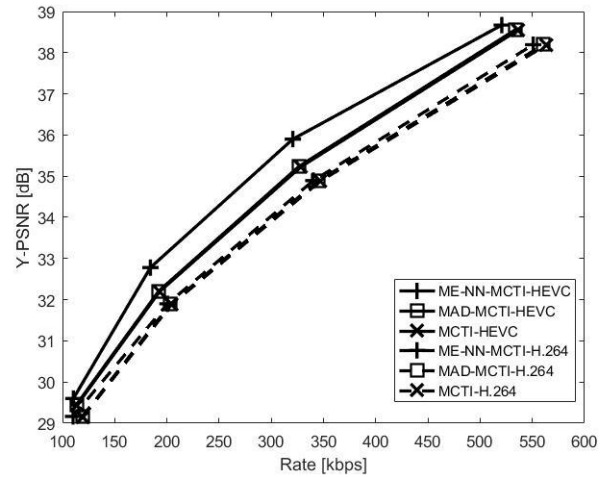


Hình 6. So sánh hiệu năng RD của phương pháp đề xuất với chuỗi Akiyo

Bảng II, III, IV và V mô tả mức độ cải thiện hiệu năng (trên phương diện tốc độ bit và chất lượng khung hình sau giải nén) của phương pháp đề xuất so với phương pháp được đề xuất trước đây MAD-MCTI [13] và phương pháp sử dụng kỹ thuật MCTI với kích thước cửa sổ tìm kiếm cố định. Kết quả mô phỏng chỉ ra rằng với bộ mã hóa sử dụng HEVC, phương pháp MAD-MCTI-HEVC tiết kiệm được trung bình 12.25% tổng số bit, trong khi phương pháp ME-NN-MCTI-HEVC tiết kiệm được 18.33% tổng số bit so với phương pháp MCTI-HEVC. Tương tự, nếu các khung hình chính được mã hóa bởi bộ mã hóa H.264 từ phương pháp MAD-MCTI-H.264 tiết kiệm 12.51% trong khi ME-NN-MCTI-H.264 tiết kiệm tới 15.07%. Xét trên phương diện chất lượng video sau giải mã, phương pháp MAD-MCTI trên HEVC và H.264 cải thiện được 0.9dB và 0.91dB trong khi ME-NN-MCTI cải thiện được 1.34dB và 1.07dB.

Hình 4, 5, 6, 7 so sánh hiệu năng RD của 3 phương pháp được triển khai trên bộ mã hóa DVC-HEVC và DVC-H.264 tại bốn giá trị lượng tử. Từ các kết quả mô phỏng ta nhận thấy phương pháp đề xuất ME-NN-MCTI có hiệu năng cao hơn so với các phương pháp trước đây trong cả hai trường hợp sử dụng HEVC và H.264 để mã hóa khung hình chính. Kết quả này có được là do phương pháp đề xuất sử dụng mạng Nơ-ron dựa trên các đặc trưng của chuỗi

video để dự đoán kích thước của cửa sổ tìm kiếm vector chuyển động trong thuật toán MCTI. Từ đó bộ giải có thể mã tiết kiệm được số bit mã hóa đồng thời tăng chất lượng của thông tin phụ trợ SI, giúp cải thiện chất lượng khung hình video sau giải mã.



Hình 7. So sánh hiệu năng RD của phương pháp đề xuất với chuỗi Coastguard

V. KẾT LUẬN

Bài báo đã đề xuất phương pháp tăng cường chất lượng khung hình phụ trợ SI trong bộ mã hóa video phân tán DVC. Cụ thể, phương pháp đề xuất đã sử dụng mạng Nơ-ron với đầu vào là các đặc trưng của chuỗi video và đầu ra là kích thước cửa sổ tìm kiếm trong thuật toán MCTI để tạo khung hình SI từ hai khung hình chính. Các kết quả mô phỏng cho thấy phương pháp đề xuất đã đạt được kết quả cao hơn một số phương pháp trước đây trên phương diện tiết kiệm được nhiều hơn số lượng bit dùng để mã hóa đồng thời chất lượng khung hình sau giải mã được tăng lên.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Học Viện Công nghệ Bưu chính Viễn thông trong đề tài mã số 07-2020-HV-ĐPT-CN và 03-HV-2020-ĐT1.

TÀI LIỆU THAM KHẢO

- [1] P.L Dragotti and M. Gastpar, "Distributed Source Coding: Theory, Algorithms and Applications," Academic Press, Feb. 2009.
- [2] X. HoangVan, J. Ascenso, and F. Pereira, "HEVC backward compatible scalability: A low encoding complexity distributed video coding based approach," Signal Process.: Image Commun., vol. 33, no. 4, pp. 51-70, Apr. 2015.
- [3] X. HoangVan, J. Ascenso, and F. Pereira., "Adaptive Scalable Video Coding: a HEVC based Framework Combining the Predictive and Distributed Paradigms", IEEE TCSVT, vol. 99, no. 00, pp. 1-14, Mar. 2016.
- [4] A. Aaron, R. Thang, and B. Girod (2002) "Wyner-Ziv Coding of Motion Video", in Proc. Asilomar Conference on Signals and Systems, Pacific Grove, CA, USA, November 2002.
- [5] J. Ascenso, C. Brites, and F. Pereira, Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding" in EURASIP Conference on Speech and Image Processing, Multimedia

Communications and Services (EC-SIPMCS), Smolenice, Slovak Republic, June 2005.

- [6] J. Ascenso, C. Brites, and F. Pereira, "Motion Compensated Refinement for Low Complexity Pixel based Distributed Video Coding," in IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS), Como, Italy, September 2005.
- [7] R. Puri and K. Ramchandran, PRISM: A new robust video coding architecture based on distributed compression principles, 40th Allerton Conf. Communication, Control and Computing, Allerton, IL, USA, 2002.
- [8] R. Puri, A. Majumdar, and K. Ramchandran, "PRISM: a video coding paradigm with motion estimation at the decoder," IEEE Transactions on Image Processing, vol. 16, no. 10, pp. 2436-2448, Oct. 2007.
- [9] T. Wiegand, G. J. Sullivan, S. Member, G. Bjøntegaard, A. Luthra, and S. Member, "Overview of the H.264/AVC Video Coding Standard," IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 7, pp. 560-576, 2003
- [10] D. Slepian, J. K. Wolf, and D. Slepian, "Noiseless Coding of Correlated Information Sources," IEEE Trans. Inf. Theory, vol. 19, no. 4, pp. 471-480, 1973, doi: 10.1109/TIT.1973.1055037
- [11] A. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder," IEEE Trans. Inf. Theory., vol. 22, no. 1, pp. 1-10, 1976
- [12] Google, Colaboratory: frequently asked questions, 2018, [Access: 6-21-2018]. [Online]. Available: <https://research.google.com/colaboratory/fag.html>.
- [13] T. Nguyen, T. Huong, T. V. Huu and S. Vu Van, "Content based side information creation for distributed video coding," 2019 6th NAFOSTED Conference on Information and Computer Science (NICS), Hanoi, Vietnam, 2019, pp. 223-227, doi: 10.1109/NICS48868.2019.9023813.

NEURAL NETWORK BASED SIDE INFORMATION GENERATION IN DISTRIBUTED VIDEO CODING

Abstract: Distributed Video Encoding (DVC) is a potential solution for uplink applications such as wireless video surveillance networks or multimedia sensor networks. In distributed video encoding, the quality of the side information (SI) has a significant effect on the bit-distortion (RD) performance. However, the quality of the side information varies in video sequence and even in every frame. Therefore, to improve the side information, the technique of creating side information adapted to the content of the video sequence is proposed in this article. Specifically, the paper proposes a method to use neural networks to predict the search window size for motion estimation algorithm in the algorithm of side information generation. The simulation results show that the performance of the distributed video codec is significantly improved using this method.

Keyword: DVC, Wyner-Ziv Coding



Nguyễn Thị Hương Thảo, Nhận bằng tốt nghiệp đại học và thạc sỹ Học viện Công nghệ Bưu chính Viễn thông vào các năm 2003 và 2010. Hiện giảng dạy và làm Nghiên cứu sinh tại Khoa Kỹ thuật Điện tử 1 - Học viện Công nghệ Bưu chính Viễn thông, Lĩnh vực nghiên cứu: Xử lý tín hiệu Video, Xử lý Ảnh, Lý thuyết thông tin.



Vũ Hữu Tiến, Tốt nghiệp đại học và cao học ngành Điện tử Viễn thông vào các năm 2002, 2004 tại Đại học Bách khoa Hà nội. Năm 2010 nhận bằng Tiến sĩ tại Đại học Chulalongkorn (Thái Lan) ngành Điện tử Viễn thông. Hiện đang công tác tại Khoa Đa phương tiện, Học viện Công nghệ Bưu chính Viễn thông.

Lĩnh vực nghiên cứu hiện nay: Xử lý tín hiệu và truyền thông đa phương tiện, Phát triển ứng dụng đa phương tiện.