



GIỚI THIỆU CHẾ ĐỘ DỰ ĐOÁN TRONG ẢNH MỚI DỰA TRÊN QUÁ TRÌNH DỰ ĐOÁN LIÊN LỚP DÀNH CHO KHẢ NĂNG MỞ RỘNG CỦA CHUẨN HEVC

Nguyễn Tăng Khả Duy¹

¹ Khoa Công nghệ, Trường Đại học Cần Thơ

Thông tin chung:

Ngày nhận: 26/11/2014

Ngày chấp nhận: 26/02/2015

Title:

A new intra prediction mode based on inter-layer prediction for scalable extension to HEVC

Từ khóa:

Dự đoán trong ảnh, dự đoán liên lớp, HEVC, SHVC, TextureRL

Keywords:

Intra prediction, inter-layer prediction, HEVC, SHVC, TextureRL

ABSTRACT

The current draft scalable extension to HEVC offers two approaches, RefIdx and TextureRL, for performing inter-layer prediction. In the framework of TextureRL, this paper first presents a mode-dependent pixel-based weighted intra prediction scheme for coding the enhancement layer (EL). The scheme first decomposes the EL intra prediction and the collocated base layer reconstructed block into their respective DC and AC components and then computes a weighted sum of both components to form a better prediction signal using a pixel-based weighting scheme. The experimental results show an average BD-rate savings of 1,0% for the AI-2x configuration and 0,5% for AI-1.5x over the SHM-1.0 anchor.

TÓM TẮT

Hiện nay, dự thảo mở rộng dành cho HEVC (gọi là SHVC) để mã hóa đa phân lớp đang có hai hướng tiếp cận đó là RefIdx và TextureRL. Trong khuôn khổ của TextureRL, nghiên cứu này sẽ trình bày một chế độ dự đoán trong ảnh mới để mã hóa các khối ở lớp mở rộng của SHVC. Chế độ này đầu tiên sẽ phân giải khối đã được dự đoán bằng chế độ dự đoán trong ảnh ở lớp mở rộng và khối có cùng vị trí ở lớp cơ sở tương ứng ra thành các thành phần DC (giá trị trung bình) và AC (giá trị khác biệt); bốn thành phần này sẽ được cộng lại theo một tỷ lệ phụ thuộc vào vị trí của từng điểm ảnh và chế độ dự đoán trong ảnh ban đầu. Các hệ số tỷ lệ này được tính toán bằng phương pháp bình phương nhỏ nhất dựa trên các dữ liệu huấn luyện chọn lọc. Kết quả cho thấy video thu được có BD-rate nhỏ hơn 1,0% và 0,5% so với các cấu hình so sánh chuẩn là AI-2x và AI-1.5x trong tập chương trình gốc SHM-1.0.

1 GIỚI THIỆU

Cách thức mã hóa liên lớp¹ của HEVC (gọi là SHVC) (Sullivan *et al.*, 2012; Chen *et al.*, 2013b) cũng giống như của H.264/SVC (Schwarz *et al.*,

2007), đó là mã hóa theo lớp. Lớp cơ sở (base layer) chứa các thông tin cơ bản nhất để đảm bảo chất lượng tối thiểu giải mã được, còn các lớp nâng cao (enhancement layer) được tạo ra nhằm cải thiện chất lượng (về tỷ lệ khung hình, độ phân giải, và tái thiết độ nét) khi kết hợp với lớp cơ sở. Bởi vì tín hiệu video ngõ vào cho các phân lớp này có cùng nội dung nên sự tương quan giữa các lớp này là rất lớn. Do đó, nếu ta có thể khai thác được sự tương quan này thì số lượng bit cần thiết để mã hóa

¹ Đây là khả năng cung cấp chế độ phân cấp chất lượng tức thời và cố định ở mức đối tượng. Sự khác nhau chia làm ba loại là: phân cấp không gian (spatial scalability), phân cấp thời gian (temporal scalability), và phân cấp về độ nét (fidelity scalability).

các lớp nâng cao có thể được giảm đáng kể. Ví dụ, trong chuẩn nén H.264/SVC, số bit cần để mã hóa các lớp nâng cao được giảm bằng thuật toán dự đoán liên lớp (inter-layer prediction – ILP) với ba chế độ cơ bản là dự đoán kết cấu liên lớp (texture), dự đoán chuyển động liên lớp (motion), và dự đoán tín hiệu sai biệt (residual).

Trong mô hình thử nghiệm hiện tại của SHVC, các cơ chế ILP có thể được thực hiện bằng hai cách tiếp cận khác nhau, đó là RefIdx và TextureRL (Chen *et al.*, 2013b). Trong cách tiếp cận RefIdx, tất cả ảnh đều được tái tạo lại dựa vào lớp cơ sở, các ảnh này sẽ được lưu trữ trong danh sách đối chiếu ảnh (reference picture list). Sau đó, các chế độ dự đoán ảnh liên lớp có thể được thực hiện như chế độ dự đoán liên ảnh (inter prediction) trong cùng một lớp. Lý do đạt được quá trình dự đoán này là do trong cơ chế RefIdx, nếu thiết bị đầu cuối yêu cầu tái tạo lại ảnh ở một lớp nâng cao nào đó thì thiết bị chỉ cần thực hiện phép tái tạo ảnh tại lớp đó do các ảnh tái lập ở lớp khác đã được đặt vào danh sách đối chiếu ảnh tại lớp nâng cao cần tái lập. Cách tiếp cận này mang lại những lợi ích lớn cho quá trình thiết kế thiết bị mã hóa và giải mã do việc thay đổi cấu trúc là ít nhất trên các hệ thống mã hóa tích hợp theo chuẩn HEVC. Ngược lại, cách tiếp cận TextureRL yêu cầu sự thay đổi lớn ở mức độ thấp (ví dụ như thay đổi cấu trúc ở Coding Unit hoặc Prediction Unit). Sự thay đổi này nhằm mục tiêu tìm ra sự giống nhau tối đa ở cấp độ thấp để số lượng bit mã hóa cuối cùng là thấp nhất. Cách tiếp cận này còn cung cấp sự linh hoạt tối đa cho người thiết kế do việc dự đoán có thể được thực hiện ở bất cứ cấp độ nào. Tuy nhiên, cần lưu ý rằng cấp độ thiết kế càng thấp càng mất nhiều thời gian cũng như tài nguyên để thiết kế và việc tương thích với các hệ thống theo chuẩn HEVC là càng thấp. Một số nghiên cứu chỉ ra rằng cách tiếp cận RefIdx đã có thể đạt được hầu hết các ưu điểm được cung cấp bởi TextureRL (Xiu *et al.*, 2013; Zan, 2013), làm lựa chọn mã hóa dữ liệu dựa trên cách tiếp cận TextureRL ít được chú ý tới. Tuy nhiên, chúng tôi tin rằng sự linh hoạt của các TextureRL đã không được sử dụng một cách tối ưu để cho kết quả tốt nhất.

Nhằm khai thác tiềm năng của phương pháp tiếp cận TextureRL trong việc hình thành một chế độ dự đoán liên lớp tốt hơn, Lainema và Ugur (2013) đề xuất phương pháp dự đoán hiệu chỉnh trong ảnh dựa trên giá trị trung bình (Intra DC Correction – IDCC) để dự đoán ảnh ở các lớp nâng cao bằng cách thay thế giá trị trung bình của khối đang được dự đoán tại lớp nâng cao bằng giá trị

trung bình của khối tương ứng tại lớp cơ sở. Tuy nhiên, các thuật toán IDCC chỉ dừng lại ở mức đơn giản là thay thế giá trị trung bình mà chưa quan tâm đến từng vị trí điểm ảnh. Điều này là chưa phù hợp bởi đối với chuẩn HEVC, các khối có kích thước thay đổi khá lớn từ 4x4 đến 64x64. Một nghiên cứu khác (Kim, 2013) giới thiệu một phương pháp dự đoán trong ảnh tỷ lệ với vị trí của từng pixel (Weighted Intra Prediction - WIP). Phương pháp này kết hợp giá trị dự đoán được tại từng điểm ảnh trong một khối ở lớp nâng cao (khối này được dự đoán bằng phương pháp dự đoán trong ảnh) với giá trị tại các điểm ảnh trong cùng một khối tương ứng ở lớp cơ sở; hệ số tỷ lệ này phụ thuộc vào vị trí của điểm ảnh so với điểm ảnh gốc tại vị trí (0,0) của khối đang được dự đoán. Vấn đề của phương pháp này là nó chỉ sử dụng duy nhất một giao thức tỷ lệ mà chưa quan tâm đến sự khác biệt do các khối ở lớp nâng cao có thể được tạo nên bằng các chế độ dự đoán trong ảnh khác nhau (chú ý rằng, hiện tại HEVC có đến 35 chế độ dự đoán trong ảnh).

Chúng tôi cố gắng kết hợp các ưu điểm của hai đề xuất trên để tạo ra một chế độ dự đoán mới có thể cải thiện hơn nữa về hiệu quả dự đoán. Với phương pháp đề xuất này, các thành phần giá trị trung bình (kể từ đây sẽ được gọi là giá trị DC) ở cấp độ dự đoán khối và sự khác biệt (chúng tôi gọi là AC) ở cấp độ dự đoán điểm ảnh của từng khối ở lớp nâng cao và lớp cơ sở sẽ được trích ra và sau đó cả bốn thành phần này sẽ được cộng lại theo một tỷ lệ nhất định. Tương ứng với cách thức hoạt động của thuật toán được đề xuất, chúng tôi tạm gọi thuật toán này có tên là Chế độ dự đoán trong ảnh liên lớp dựa trên hệ số tỷ lệ ở cấp độ điểm ảnh (Pixel-based Weighted Intra-Inter-Layer Prediction – PWILLP).

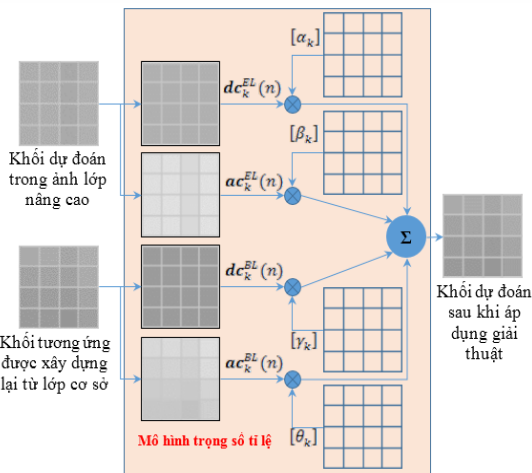
Các hệ số tỷ lệ này được xác định thông qua một quá trình huấn luyện dựa trên thuật toán bình phương nhỏ nhất (least-square). Kết quả sơ bộ cho thấy chế độ dự đoán này giảm BD-rate từ 0,5-1,0% đối với SHM 1.0.

Phần còn lại của bài viết này được tổ chức như sau: Phần II trình bày thuật toán đề xuất của chúng tôi, bao gồm các khái niệm quan trọng và các mô hình toán học. Phần III trình bày các bảng trọng số (weighting scheme) cho các thành phần khác nhau, và làm thế nào các trọng số này nên được điều chỉnh để đáp ứng với sự thay đổi trong mã hóa các thông số. Phần IV cung cấp kết quả thử nghiệm và so sánh với các nghiên cứu trước. Cuối cùng, mục V kết luận việc này với một bản tóm tắt các kết quả của chúng tôi.

2 TỔNG QUAN VỀ GIẢI THUẬT ĐỀ XUẤT

2.1 Nguyên lý hoạt động

Ý tưởng của giải thuật đề xuất về cơ bản là một sự kết hợp của hai nghiên cứu trước (Lainema và Ugur, 2013; Kim, 2013). Như mô tả trong Hình. 1, đầu tiên, các thông tin kết cấu bề mặt của khối được tạo nên từ chế độ dự đoán trong ảnh ở lớp nâng cao và khối được tái tạo ở lớp cơ được phân tách ra thành các thành phần DC và AC; với thành phần DC tìm được bằng cách lấy giá trị trung bình của tất cả điểm ảnh trong khối, còn thành phần AC của mỗi điểm ảnh tìm được bằng cách trừ giá trị hiện tại của điểm ảnh cho giá trị DC vừa tìm được. Các thành phần DC và AC ở cả hai phân lớp (lớp nâng cao và cơ sở) sau đó được nhân cho một hệ số tương ứng tham chiếu từ bảng giá trị tìm được thông qua thuật toán giá trị trung bình tối thiểu. Chú ý rằng, các bảng giá trị tham chiếu này là riêng biệt cho từng chế độ dự đoán trong ảnh ở lớp nâng cao và cũng riêng biệt cho từng khối khác nhau về kích thước. Sau cùng, khối dự đoán mới được tạo ra bằng cách cộng tất cả các thành phần vừa tìm được lại với nhau tương ứng cho từng điểm ảnh.



Hình 1: Sơ đồ trình bày nguyên tắc hoạt động của giải thuật đề xuất

2.2 Phương pháp tổng bình phương nhỏ nhất

Rõ ràng, cách mà các thành phần khác nhau được gán các trọng số khác nhau có ảnh hưởng quan trọng đến kết quả của việc dự đoán. Nghiên cứu này được tiến hành với hy vọng rằng chúng ta có thể tìm thấy một tập hợp các bảng trọng số cho dự đoán tốt nhất mà kết quả là có thể được giảm thiểu sự sai biệt giữa khối dự đoán và khối ảnh gốc. Vấn đề này có thể được giải quyết bằng cách

sử dụng phương pháp tính toán giá trị bình phương nhỏ Least-Squares (LS).

Để tiện cho việc theo dõi các công thức tính toán được trình bày trong mục này, chúng tôi áp dụng các ký hiệu sau: **chữ đậm viết thường** đại diện cho vector, **CHỮ ĐẬM VIẾT IN** biểu thị ma trận, và *chữ nghiêng viết thường* đặc trưng cho các giá trị vô hướng. Cụ thể, $\mathbf{a}_k = [a_k(1) \ a_k(2) \ \dots \ a_k(n)]^T$ và $\mathbf{b}_k = [b_k(1) \ b_k(2) \ \dots \ b_k(n)]^T$ đại diện cho các giá trị dự đoán tại điểm ảnh k được tạo nên từ khối cần dự đoán tại lớp nâng cao và khối tương ứng tại lớp cơ sở. Tương tự, $\mathbf{ac}_k = [ac_k(1) \ ac_k(2) \ \dots \ ac_k(n)]^T$ và $\mathbf{dc}_k = [dc_k(1) \ dc_k(2) \ \dots \ dc_k(n)]^T$ biểu thị tương ứng các giá trị tương ứng từ AC và DC tại điểm ảnh k . Vì vậy, chúng tôi có

$$\mathbf{a}_k = \frac{(\mathbf{a}_k - \mathbf{dc}_k^{EL})}{\mathbf{ac}_k^{EL}} + \mathbf{dc}_k^{EL} \tag{1}$$

$$\mathbf{b}_k = \frac{(\mathbf{b}_k - \mathbf{dc}_k^{BL})}{\mathbf{ac}_k^{BL}} + \mathbf{dc}_k^{BL} \tag{2}$$

Với những giá trị trên, gọi giá trị dự đoán cuối cùng là $\mathbf{p}_k = [p_k(1) \ p_k(2) \ \dots \ p_k(n)]^T$ tại điểm ảnh k có thể được viết như sau:

$$\mathbf{p}_k = \mathbf{X}_k \cdot \mathbf{w}_k \tag{3}$$

Với $\mathbf{X}_k = [\mathbf{ac}_k^{EL} \ \mathbf{dc}_k^{EL} \ \mathbf{ac}_k^{BL} \ \mathbf{dc}_k^{BL}]$ và $\mathbf{w}_k = [\alpha_k \ \beta_k \ \gamma_k \ \theta_k]^T$ là một véc-tơ trọng số mà các phần tử là các giá trị trọng số liên kết với bốn thành phần để dự đoán tại điểm ảnh k . Cụ thể, các véc-tơ trọng số đại diện cho mô hình trọng số tương ứng với một điểm ảnh duy nhất, các véc-tơ này mô tả cách các mẫu tương ứng từ các thành phần khác nhau đóng góp vào dự toán cường độ một điểm ảnh hiện tại.

Với tham chiếu đến các ký hiệu trên, chúng tôi tiếp tục biểu diễn $\mathbf{o}_k = [o_k(1) \ o_k(2) \ \dots \ o_k(n)]^T$ tương ứng với các điểm ảnh gốc tại vị trí k , có cường độ là giá trị được ước tính. Vấn đề xác định các véc-tơ trọng số tối ưu \mathbf{w}^*_k theo phương pháp bình phương nhỏ nhất sau đó có thể được xây dựng như sau:

$$\mathbf{w}^*_k = \underset{\mathbf{w}_k}{\operatorname{argmin}} (\mathbf{o}_k - \mathbf{X}_k \cdot \mathbf{w}_k)^2 \tag{4}$$

Trong công thức (4), mục tiêu là tìm ra các véc-tơ trọng số tối ưu \mathbf{w}^*_k sao cho sự khác biệt giữa \mathbf{p}_k và \mathbf{o}_k là tối thiểu. Từ mô hình hồi quy tuyến tính, trọng số tối ưu có dạng toán học như sau:

$$\mathbf{w}^*_k = (\mathbf{X}_k^T \cdot \mathbf{X}_k)^{-1} \cdot \mathbf{X}_k^T \cdot \mathbf{o}_k \tag{5}$$

Bằng cách thay đổi các chỉ số k và lặp đi lặp lại quá trình đó, chúng ta có thể có được các véc-tơ trọng lượng cho các vị trí điểm ảnh khác

nhau và do đó các chức năng trọng cho tất cả bốn thành phần.

2.3 Quá trình huấn luyện dữ liệu

Để thu thập dữ liệu cho việc tính toán hàm trọng số tối ưu, quá trình huấn luyện dữ liệu được sử dụng. Trước tiên, để đảm bảo các dữ liệu thu thập được là phù hợp, thuật toán đề xuất sẽ được áp dụng để tạo ra một chế độ dự đoán mới PWILP và chế độ này phải cạnh tranh với tất cả các chế độ thông thường thông qua quá trình chọn lọc Rate-Distortion Optimization (RDO) tại lớp nâng cao để tìm ra chế độ có RDO tốt nhất. Sau đó, những khối mã hoá bằng thuật toán đề xuất sẽ được sử dụng để tính toán nhằm tìm ra hàm trọng số tỷ lệ tối ưu dựa trên công thức (5).

Dựa trên công thức (4), ta có thể thấy rằng mỗi lần lặp lại của quá trình huấn luyện dữ liệu, chúng ta cần phải gán một giá trị khởi tạo cho hàm trọng số tỷ lệ. Cụ thể, cho quá trình lặp lại đầu tiên, giá trị trọng số tỷ lệ tương ứng với giá trị trung bình của kết cấu bề mặt giữa khối được dự đoán bằng chế độ dự đoán trong ảnh ở lớp nâng cao và khối được tái tạo từ lớp cơ sở. Giá trị khởi tạo này dùng để xác định được khối cần dự đoán p_k cho quá trình so sánh RDO. Hàm trọng số tỷ lệ tối ưu sau đó được tìm trực tiếp dựa trên công thức (5) với X_k được tìm dựa trên các khối có RDO tốt nhất, tìm được sau quá trình chọn lọc RDO. Quá trình tính toán tìm trọng số tối ưu được nhúng trực tiếp lên mô hình mẫu SHM1.0 (Chen *et al.*, 2013a). Sau đó, các hàm trọng số tỷ lệ tìm được của quá trình lặp lại thứ nhất được gán trở lại thành giá trị ban đầu cho quá trình lặp lại thứ hai. Đồng thời, các hàm trọng số tỷ lệ này là chung nhất cho tất cả các video trong quá trình huấn luyện (và kết quả của quá trình huấn luyện dữ liệu – tức là hàm trọng số tỷ lệ tối ưu – cũng được dùng chung nhất cho tất cả các video trong quá trình thí nghiệm). Quá trình này được lặp lại cho tất cả các video trong quá trình huấn luyện. Cuối cùng, điều kiện để thoát khỏi quá trình huấn luyện khi các hàm trọng số tỷ lệ được tối ưu; tức là, khi sai số bình phương là nhỏ nhất trong các quá trình lặp lại. Cụ thể, quá trình huấn luyện được dừng lại dựa trên hai điều kiện: (1) Các hàm trọng số tỷ lệ là ổn định (tức là, cả giá trị này không thay đổi quá lớn so với quá trình lặp lại trước đó); và (2) Giá trị sai số tuyệt đối khi được tính toán bằng công thức bình phương nhỏ nhất giữa quá trình lặp lại hiện tại và quá trình

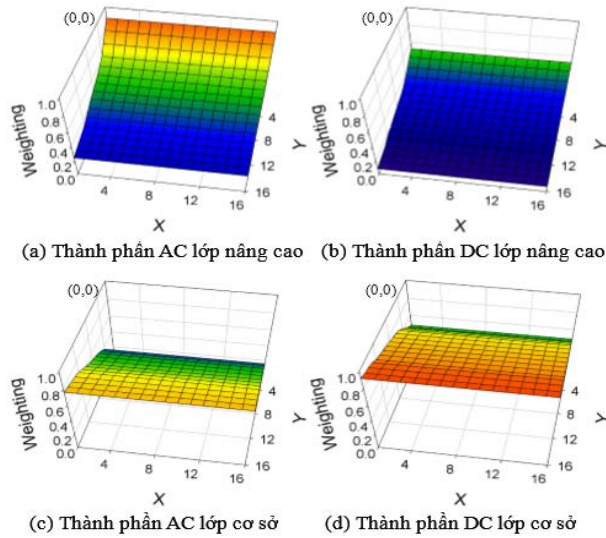
lặp lại liền trước đó nhỏ hơn 1%. Một vấn đề cần lưu ý đó là các video được sử dụng cho quá trình huấn luyện là khác với các video dùng trong thí nghiệm để xác định hiệu suất. Do đó, các hàm trọng số tỷ lệ trong nghiên cứu này nên được hiểu như là các hàm trọng số tỷ lệ tìm được từ quá trình huấn luyện với các video cụ thể (sẽ được trình bày chi tiết trong Mục 4.1 – Điều kiện và mô hình thí nghiệm).

3 PHÂN TÍCH MÔ HÌNH TRỌNG SỐ

Phần này sẽ phân tích cách tính toán các trọng số tỷ lệ dành cho thành phần DC và AC ở các phân lớp khác nhau. Giá trị cho các hàm trọng số này đặc trưng cho tỷ lệ mà mỗi thành phần sẽ góp phần vào khối dự đoán sau cùng ở cấp độ điểm ảnh. Hàm trọng số tỷ lệ của mỗi cặp thành phần (DC hoặc AC) ở lớp nâng cao và lớp cơ sở có tổng bằng 1 và giá trị tại mỗi điểm ảnh của khối dự đoán sau cùng sẽ phải nằm trong khoảng từ 0 – 255 tương ứng với giá trị 8 bit tại mỗi điểm ảnh trong mô hình thí nghiệm. Sự phân tích sâu của hàm trọng số tỷ lệ (weighting functions) sẽ tương ứng với (1) Các chế độ dự đoán trong ảnh tại lớp nâng cao, (2) Giá trị lượng tử hóa (Quantization value - QP) của lớp cơ sở và nâng cao, và (3) Kích thước của khối dự đoán.

3.1 Ảnh hưởng của các chế độ dự đoán trong ảnh

Tiêu mục này khảo sát ảnh hưởng của chế độ dự đoán trong ảnh lên hàm trọng số tỷ lệ. Hình 2 cho thấy dạng sóng của hàm trọng số tỷ lệ của chế độ dự đoán trong ảnh dọc (vertical mode) ở lớp nâng cao cho một khối kích thước 16x16. Có thể thấy rằng, dạng sóng tương tác với các thành phần từ cùng một lớp có dạng sóng giống nhau và khác nhau một cách đáng kể về mặt biên độ. Hơn nữa, ta có thể nhận thấy giá trị trọng số tỷ lệ dành cho thành phần DC của lớp nâng cao nhỏ hơn nhiều so với lớp cơ sở, điều này giải thích được cho việc thay thế thành phần DC của lớp nâng cao bằng thành phần DC của lớp cơ sở trong giải thuật IDCC. Chúng ta cũng có thể thấy rằng các trọng số tỷ lệ có dạng sóng phụ thuộc vào chế độ dự đoán trong ảnh ở lớp nâng cao (trong trường hợp này, chế độ dự đoán tạo ra khối dự đoán ở lớp nâng cao là chế độ dọc; do đó, dạng sóng cũng có dạng chạy dọc xuống). Như vậy, có thể kết luận rằng giải thuật đề xuất phụ thuộc vào chế độ dự đoán trong ảnh tại lớp nâng cao.

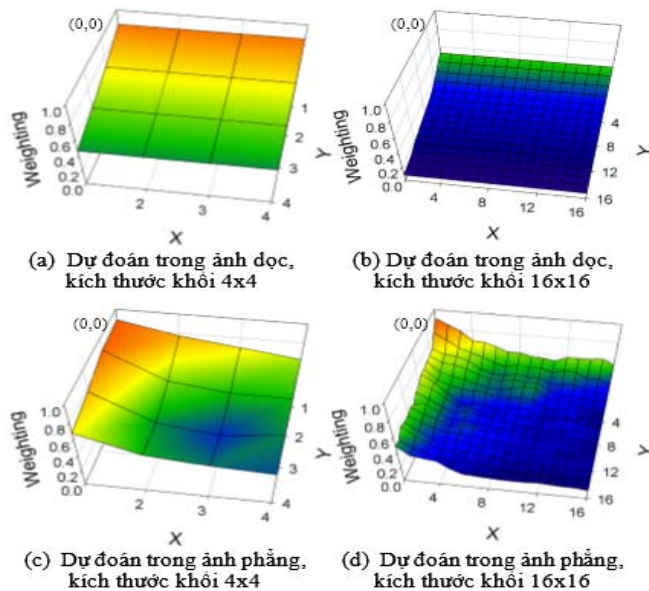


Hình 2: Dạng sóng của hàm trọng số tỷ lệ của thành phần AC và DC ở lớp nâng cao và lớp cơ sở; dạng sóng thu được từ chế độ dự đoán trong ảnh dọc (vertical intra prediction mode), kích thước khối 16x16

3.2 Ảnh hưởng của kích thước khối

Phần này sẽ tìm hiểu về tác động của kích thước của khối dự đoán lên trọng số tỷ lệ. Ảnh hưởng của kích thước khối một phần nào đó có thể dự đoán được (hiển nhiên, giá trị trọng số tại lớp nâng cao sẽ cao hơn cho các kích thước khối nhỏ hơn); mặc dù, các thông tin về kết cấu bề mặt trong lớp cơ sở có thể thay đổi kết quả dự đoán trên. Trong trường hợp này, Hình 3 mô tả dạng sóng của hàm trọng số tỷ lệ của thành phần DC ở lớp nâng

cao với kích thước khối lần lượt là 4x4 và 16x16. Đó là dạng sóng của hàm trọng số tỷ lệ tìm được tương ứng với chế độ dự đoán trong ảnh dọc và phẳng. Như đã thảo luận, kết quả mô phỏng cho thấy các giá trị trọng số cho thành phần DC ở lớp nâng cao là cao hơn cho các khối có kích thước nhỏ hơn. Điều này hoàn toàn có giải thích được vì các chế độ dự đoán trong ảnh phụ thuộc vào các điểm ảnh lân cận; do đó, với kích thước khối càng nhỏ thì độ tương quan càng lớn và kết quả dự đoán có kết quả chính xác cao hơn.



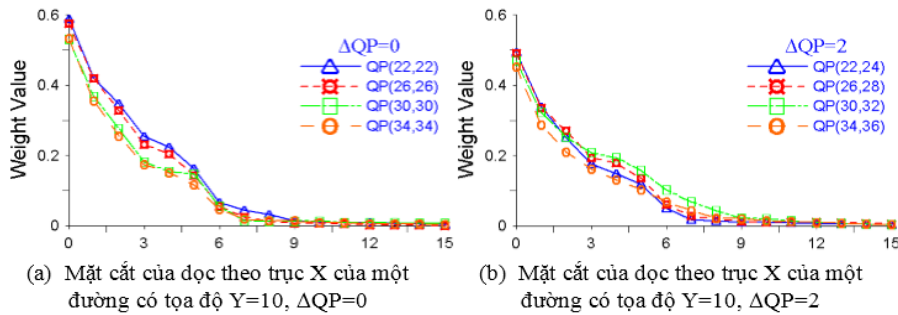
Hình 3: Dạng sóng của hàm trọng số tỷ lệ của thành phần DC ở lớp nâng cao của chế độ dự đoán trong ảnh dọc và phẳng (vertical and planar intra prediction mode) với các kích thước khối khác nhau

3.3 Ảnh hưởng của giá trị lượng tử hóa

Phần này sẽ tìm hiểu ảnh hưởng của thiết lập thông số lượng tử hóa (quantization parameter – QP) lên hàm trọng số tỷ lệ. Hình 4 mô tả dạng sóng của hàm trọng số tỷ lệ của thành phần DC ở lớp nâng cao dọc theo mặt cắt trục Y tại điểm có giá trị bằng 10 (của một hàng có 16 giá trị) cho các thiết lập giá trị lượng tử khác nhau. Cụ thể, giá trị lượng tử hóa của lớp cơ sở gồm có 4 giá trị là 22, 26, 30, và 34; và deltaQP (ΔQP) là sự khác biệt trong giá trị lượng tử hóa giữa lớp nâng cao và lớp cơ sở. Trong mô hình thí nghiệm hiện tại, có hai nhóm ΔQP lần lượt là $\Delta QP=0$ và $\Delta QP=2$ tương ứng với các giá trị lượng tử ở lớp nâng cao bằng hoặc lớn hơn 2 đơn vị so với lớp cơ sở. Lưu ý, kết quả thể hiện trong Hình 4 tương ứng với chế độ dự đoán trong ảnh ngang (horizontal intra prediction mode).

Từ kết quả ta có thể thấy được hàm trọng số tỷ lệ giảm dần khi di chuyển dọc theo trục X (trong đó X là trục nằm ngang) bởi vì bản chất của chế độ

dự đoán trong ảnh ngang và đây cũng là kết quả mà chúng ta đã thấy trong phần phân tích ảnh hưởng của chế độ dự đoán trong ảnh. Một quan sát thú vị hơn từ các giá trị trọng số tỷ lệ của các thiết lập trị số lượng tử hóa trong cùng một tập đó là mặc dù trị số lượng tử hóa tối thiểu và tối đa là khác nhau rất lớn (trị số lượng tử hóa 22 cho giá trị tối thiểu và 34 cho giá trị tối đa), thì sự khác biệt về biên độ trong dạng sóng của các thiết lập này khác nhau không đáng kể. Từ nhiều thí nghiệm với tất cả các khả năng tổ hợp của các thiết lập trên trị số lượng tử hóa, chúng tôi nhận thấy ảnh hưởng của thiết lập trị số lượng tử hóa lên các hàm trọng số tỷ lệ về phương diện tiết kiệm số lượng dữ liệu mã hóa là không khác nhau nhiều. Vì thế, từ những quan sát này cho ta một khả năng thống nhất các hàm trọng số tỷ lệ cho các thiết lập của trị số lượng tử cho tất cả các video được dùng cho thí nghiệm với các điều kiện kiểm tra được đặc tả trong các điều kiện thử nghiệm thông thường.



Hình 4: Các đường cong chỉ ra dạng sóng của trọng số tỷ lệ của thành phần DC ở lớp mở rộng của chế độ dự đoán trong ảnh ngang (horizontal intra prediction mode) với các thiết lập giá trị lượng tử hóa khác nhau và được chia thành hai nhóm: (a) Giá trị lượng tử hóa ở cả hai lớp giống nhau; (b) Giá trị lượng tử hóa ở lớp nâng cao cao hơn lớp cơ sở 2 đơn vị

Các kết quả nghiên cứu có thể được tóm tắt như sau. Đầu tiên, các hàm trọng số tỷ lệ là phụ thuộc chế độ dự đoán trong ảnh ở lớp nâng cao; và sự tách riêng các thành phần AC và DC cho nhiều ưu điểm khi tổng hợp chúng lại theo các hàm trọng số tỷ lệ. Thứ hai, lớp nâng cao có trọng số tỷ lệ cao hơn cho các khối dự đoán có kích thước nhỏ hơn. Thứ ba, mặc dù các thiết lập trị số lượng tử hóa có thể rất khác biệt trong cùng một nhóm nhưng lại có dạng sóng của các hàm trọng số tỷ lệ khá giống nhau; điều này có thể cho phép đơn giản hóa giải thuật đề xuất bằng cách thống nhất các hàm trọng số tỷ lệ cho các trị số lượng tử hóa khác nhau.

4 THÍ NGHIỆM

4.1 Điều kiện và mô hình thí nghiệm

Mô hình thí nghiệm được xây dựng dựa trên mô hình mẫu SHM1.0 [Chen *et al.*, 2013a] và tiến

hành thử nghiệm rộng rãi, chủ yếu là trên các điều kiện dự đoán trong ảnh (All Intra – AI) và các điều kiện thử nghiệm phổ biến để khảo sát hiệu suất của thuật toán được đề xuất; sau đó, so sánh giải thuật này với các đề xuất trước đó. Trong mô hình hiện tại, chỉ các thí nghiệm bắt buộc được thực hiện (tức là không thử nghiệm trên các tập mở rộng). Cụ thể, chỉ khảo sát trường hợp khi video ở lớp cơ sở được mã hóa với chuẩn HEVC (mà không khảo sát trường hợp lớp cơ sở được mã hóa với chuẩn H.264). Hơn nữa, tỷ lệ độ phân giải giữa lớp nâng cao và lớp cơ sở được giới hạn theo tỷ lệ 2x và 1.5x (tức là lớp nâng cao có độ phân giải gấp 2 hoặc gấp 1.5 lần lớp cơ sở). Bảng 1 giới thiệu về các video được sử dụng trong bài thí nghiệm. Chú ý rằng các giá trị trọng số tỷ lệ được sử dụng trong giải thuật mà bài báo đề xuất thu được từ quá trình huấn luyện với các video sử dụng trong quá trình

này khác với các video dùng trong thí nghiệm. Điều này là do nếu các video dùng trong quá trình huấn luyện và thí nghiệm là cùng một nhóm các

video thì hàm trọng số tỷ lệ thu được chỉ mang tính cục bộ. Các video được sử dụng trong quá trình huấn luyện dữ liệu được trình bày trong Bảng 2.

Bảng 1: Các Video dùng trong thí nghiệm

Phân lớp Video	Tên Video	Độ phân giải lớp cơ sở	Độ phân giải lớp nâng cao	Cấu hình dự đoán trong ảnh toàn phần
A	Traffic	1280x800	2560x1600	Tỷ lệ khung ảnh 2x
	PeopleOnStreet	1280x800	2560x1600	Tỷ lệ khung ảnh 2x
B	Kimono	960x540	1920x1080	Tỷ lệ khung ảnh 2x
		1280x720	1920x1080	Tỷ lệ khung ảnh 1.5x
	ParkScene	960x540	1920x1080	Tỷ lệ khung ảnh 2x
		1280x720	1920x1080	Tỷ lệ khung ảnh 1.5x
	Cactus	960x540	1920x1080	Tỷ lệ khung ảnh 2x
		1280x720	1920x1080	Tỷ lệ khung ảnh 1.5x
	BasketballDrive	960x540	1920x1080	Tỷ lệ khung ảnh 2x
		1280x720	1920x1080	Tỷ lệ khung ảnh 1.5x
	BQTerrace	960x540	1920x1080	Tỷ lệ khung ảnh 2x
		1280x720	1920x1080	Tỷ lệ khung ảnh 1.5x

Bảng 2: Các Video dùng trong quá trình huấn luyện dữ liệu

Phân lớp Video	Tên Video	Độ phân giải lớp cơ sở	Độ phân giải lớp nâng cao	Cấu hình dự đoán trong ảnh toàn phần
A	Nebuta	1280x800	2560x1600	Tỷ lệ khung ảnh 2x
	SteamLocomotive	1280x800	2560x1600	Tỷ lệ khung ảnh 2x
B	Blue Sky	960x540	1920x1080	Tỷ lệ khung ảnh 2x
		1280x720	1920x1080	Tỷ lệ khung ảnh 1.5x
	Tennis	960x540	1920x1080	Tỷ lệ khung ảnh 2x
		1280x720	1920x1080	Tỷ lệ khung ảnh 1.5x
	Riverbed	960x540	1920x1080	Tỷ lệ khung ảnh 2x
		1280x720	1920x1080	Tỷ lệ khung ảnh 1.5x
	Troy	960x540	1920x1080	Tỷ lệ khung ảnh 2x
		1280x720	1920x1080	Tỷ lệ khung ảnh 1.5x
	Station	960x540	1920x1080	Tỷ lệ khung ảnh 2x
		1280x720	1920x1080	Tỷ lệ khung ảnh 1.5x
	Pedestrian Area	960x540	1920x1080	Tỷ lệ khung ảnh 2x
		1280x720	1920x1080	Tỷ lệ khung ảnh 1.5x
	Rush Hour	960x540	1920x1080	Tỷ lệ khung ảnh 2x
		1280x720	1920x1080	Tỷ lệ khung ảnh 1.5x
	Tractor	960x540	1920x1080	Tỷ lệ khung ảnh 2x
		1280x720	1920x1080	Tỷ lệ khung ảnh 1.5x

4.2 Hiệu suất của thuật toán đề xuất

Từ Bảng 3 ta có thể thấy được giải thuật đề xuất cung cấp một hiệu suất nén tốt hơn, được đánh giá bởi công thức BD-rate [Bjontegaard, 2001]. Công thức BD-rate là công thức đánh giá chất lượng video một cách trực quan dựa trên sự sai biệt tỷ lệ bit theo phần trăm cho các tín hiệu có cùng tỷ

số tín hiệu cực đại trên nhiễu (Peak Signal Noise Rate – PSNR). Do các tín hiệu có cùng tỷ số tín hiệu cực đại trên nhiễu nên công thức BD-rate phản ánh được các cải thiện về tỷ lệ bit trong nghiên cứu mà không làm thay đổi chất lượng của các video sau khi được mã hóa và giải mã. Chú ý, dấu “-” (trừ) trong các bảng chỉ ra sự cải thiện về

tỷ lệ bit. Rõ ràng, hiệu suất mã hóa đạt được rất khác nhau giữa các video khác nhau. Ví dụ, hiệu suất mã hóa là nhỏ nhất với video *ParkScene*, chỉ tăng 0.3% và 0.1% lần lượt cho các tỷ lệ khung là 2x và 1.5x. Còn đối với video *BasketballDrive*, tỷ lệ này là khá cao dao động trong khoảng từ 2.4% và 1.5% cho các tỷ lệ khung 2x và 1.5x. Điều này có thể được giải thích bởi đặc điểm của các video này. Video *ParkScene* chứa nhiều chi tiết hơn trên một khung ảnh; trong khi, video *BasketballDrive* lại có nhiều vùng đồng nhất hơn. Sự cải thiện trong tỷ lệ bit của giải thuật đề xuất đi kèm với sự tăng đáng kể thời gian mã hóa; cụ thể, thời gian này

tăng lần lượt là 201,8% và 199,6% tổng thời gian để mã hóa tất cả các video dùng trong thí nghiệm khi so sánh với mô hình mẫu SHM1.0. Tất cả các thí nghiệm (trên mô hình mẫu SHM1.0 và mô hình với thuật toán đề xuất) đều được thực hiện trên máy cấu hình Core i7 tốc độ 3.0 Ghz, sử dụng Hệ điều hành Window 7; cấu hình này cũng được sử dụng để đo đặc tốc độ xử lý trong quá trình giải mã. Mặc dù, thời gian giải mã khi áp dụng mô hình thí nghiệm với thuật toán đề xuất có tăng nhưng gần như không đáng kể so với khi áp dụng mô hình mẫu.

Bảng 3: Hiệu suất mã hóa của thuật toán đề xuất

Phân lớp Video	Tên Video	Cấu hình dự đoán trong ảnh, tỷ lệ khung 2x			Cấu hình dự đoán trong ảnh, tỷ lệ khung 1.5x		
		Y	U	V	Y	U	V
A	Traffic	-0,7%	-0,6%	-0,6%	N/A		
	PeopleOnStreet	-0,6%	-0,7%	-0,4%	N/A		
	Kimono	-0,4%	-0,4%	-0,5%	-0,2%	-0,3%	-0,3%
B	ParkScene	-0,7%	-0,7%	-0,7%	-0,1%	-0,2%	-0,3%
	Cactus	-0,7%	-0,7%	-0,7%	-0,3%	-0,6%	-0,6%
	BasketballDrive	-2,4%	-2,7%	-2,7%	-1,5%	-1,6%	-1,3%
	BQTerrace	-1,3%	-1,6%	-1,8%	-0,4%	-0,6%	-0,7%
Hiệu suất so với SHM 1.0		-1,0%	-1,1%	-1,0%	-0,5%	-0,7%	-0,6%
Thời gian mã hóa [%]					201,8%		
Thời gian giải mã [%]					100,5%		

Bảng 4 so sánh hiệu suất mã hóa của giải thuật đề xuất với các giải thuật IDCC và WIP. Có thể thấy rằng cả hai giải thuật trước đó có hiệu suất mã hóa nhỏ hơn nhiều so với giải thuật đề xuất. Cụ thể, tỷ lệ BD-rate của giải thuật IDCC dao động trong khoảng từ 0,0% đến 0,5% cho tất cả các trường hợp thí nghiệm, với tỷ lệ BD-rate trung bình không quá 0,2%. Giải thuật WIP cũng cho thấy các xu hướng tương tự với tỷ lệ BD-rate trung bình

khoảng 0,3%. Trong tất cả các giải thuật trong bảng so sánh thì giải thuật đề xuất tuy có sự cải thiện đáng kể nhất về tỷ lệ bit nhưng cũng có thời gian giải mã và mã hóa cao nhất (chủ yếu do độ phức tạp của thuật toán khi phải phân tích các khối ở hai phân lớp ra thành các thành phần AC và DC, sau đó tổng hợp các thành phần này lại với các trọng số tỷ lệ tối ưu tìm được từ quá trình huấn luyện dữ liệu).

Bảng 4: So sánh hiệu suất mã hóa giữa thuật toán đề xuất và các giải thuật đề xuất trước đó

Các thuật toán	Cấu hình dự đoán trong ảnh toàn phần, tỷ lệ khung 2x			Cấu hình dự đoán trong ảnh toàn phần, tỷ lệ khung 1.5x			
	IDCC	WIP	PWILP	IDCC	WIP	PWILP	
Phân lớp A	-0,1%	-0,2%	-0,7%	N/A			
Phân lớp B	-0,2%	-0,3%	-1,1%	-0,1%	-0,1%	-0,5%	
Cấu hình tối thiểu	0,0%	0,0%	-0,3%	0,0%	0,0%	-0,1%	
Cấu hình tối đa	-0,5%	-0,7%	-2,4%	-0,1%	-0,3%	-1,5%	
Hiệu suất so với SHM 1.0		-0,2%	-0,3%	-1,0%	-0,1%	-0,1%	-0,5%
Thời gian mã hóa [%]		104,7%	192,1%	201,8%	105,3%	189,8%	199,6%
Thời gian giải mã [%]		98,5%	100,1%	100,5%	100,7%	100,4%	102,2%

5 KẾT LUẬN

Trong bài báo này, chúng tôi đã giới thiệu một giải thuật mới để kết hợp khối đã được dự đoán bằng kỹ thuật dự đoán trong ảnh tại lớp nâng cao

với khối tương ứng được xây dựng lại từ lớp cơ sở với mục đích cải thiện kết quả dự đoán khối tại lớp nâng cao trong khuôn khổ của mô hình TextureRL. Kết quả của giải thuật này cho một kết quả nên cao hơn so với các giải thuật khác trong cùng mô hình

TextureRL. Giải thuật đề xuất cung cấp một tỷ lệ giảm BD-rate từ 0,5 đến 1,0% so với SHM1.0 và sự cải thiện lên đến 0,7 đến 0,8% so với các giải thuật đề xuất trước đó. Ngoài ra, các yếu tố mã hóa có thể ảnh hưởng đến giải thuật đề xuất (ví dụ như chế độ dự đoán trong ảnh ở lớp nâng cao, kích thước khối dự đoán, và các thiết lập của hệ số lượng tử hóa) cũng được phân tích một cách kỹ càng. Với những phân tích này, người đọc có thể hiểu sâu hơn về nguyên tắc hoạt động và tạo điều kiện thuận lợi cho những nghiên cứu tiếp theo dựa trên giải thuật mà chúng tôi đề xuất.

TÀI LIỆU THAM KHẢO

1. Bjontegaard G, 2001, "Calculation of Average PSNR Differences between RD Curves". ITU-T SG16/Q6 Doc. VCEG-M33, Apr. 2001.
2. Chen J., Boyce J., Ye Y., Hannuksela M. M., 2013a, "SHVC Test Model 1 (SHM 1)", JCTVC-L1007, Geneva, Switzerland, 14-23 Jan, 2013.
3. Chen J., Boyce J., Ye Y., Hannuksela M. M., 2013b, "SHVC Working Draft 2", JCTVC-M1008, Incheon, Korea, 18-26 Apr, 2013.
4. Kim C. K., 2013, "Non-CE1: Weighted intra prediction", JCTVC-M0117, Incheon, Korea, 18-26 Apr, 2013.
5. Lainema L., Ugur K., 2013, "SCE1: Results of test 1.1 on intra DC correction", JCTVC-M0031, Incheon, Korea, 18-26 Apr, 2013.
6. Li X., Boyce J., Onno P., 2013, "Common Test Conditions and Software Reference Configurations for the Scalable Test Model". JCTVC-L1009, Jan. 2013.
7. Schwarz H., Marpe H., and Wiegand H., 2007, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard", IEEE Transactions on Circuits and Systems for Video Technology. 17:1103–20.
8. Sullivan G., Ohm J., Han W.J., Wiegand T., 2012, "Overview of the High Efficiency Video Coding (HEVC) Standard", IEEE Transactions on Circuits and Systems for Video Technology. 22:1649–68.
9. Xiu X., He Y., Ye Y., Hannuksela M.M., 2013, "TEA2: Interlayer reference picture placement", JCTVC-L0051, Geneva, Switzerland, 14-23 Jan, 2013.
10. Zan J., 2013. "TE2: Results of test 3.1.3 on CU based Intra-BL signaling", JCTVC-L0165, Geneva, Switzerland, 14-23 Jan, 2013.