

PHÂN TÍCH HIỆU SUẤT CỦA DỮ LIỆU ẢNH LIGHT FIELD VỚI CÁC CHUẨN MÃ HÓA VIDEO VÀ HÌNH ẢNH

Nguyễn Cảnh Châu

Khoa Đa phương tiện, Học Viện Công Nghệ Bưu Chính Viễn Thông

Tóm tắt—Gần đây, công nghệ hình ảnh Light Field (LF) đã thu hút nhiều sự quan tâm do các ứng dụng tiềm năng của nó trong một số lượng lớn các lĩnh vực bao gồm Thực tế ảo, Thực tế tăng cường (VR/AR), Hội nghị từ xa và E-learning. Dữ liệu LF là tập dữ liệu đa chiều, cung cấp thông tin hình ảnh phong phú như kết xuất cảnh với những thay đổi về độ sâu trường ảnh, điểm nhìn và độ dài tiêu cự mà người xem có thể thay đổi. Tuy nhiên, dữ liệu của LF cũng gặp một vấn đề quan trọng, đó là lượng dữ liệu được lưu trữ quá lớn. Do đó, nén dữ liệu LF là một trong những vấn đề trọng tâm đối với nghiên cứu LF. Bài báo này đưa ra những phân tích theo ba phần chính. Thứ nhất, bài báo trình bày một cách toàn diện luồng mã hóa hình ảnh LF, có thể được sử dụng làm chuẩn để đánh giá nén LF trong tương lai. Thứ hai, bài báo mô tả ngắn gọn các tiêu chuẩn mã hóa hình ảnh và video hiện có, được áp dụng để nén dữ liệu LF, bao gồm ba tiêu chuẩn mã hóa video nổi tiếng, H.264/AVC, H.265/HEVC, Google/VP9, và chuẩn mã hóa hình ảnh, JPEG-2000. Cuối cùng, hiệu quả nén và độ phức tạp của các tiêu chuẩn này sẽ được so sánh và thảo luận chi tiết.

Từ khóa—Dữ liệu Light Field, H.264/AVC, H.265/HEVC, VP9, JPEG-2000.

I. GIỚI THIỆU CHUNG

Dữ liệu Light Field (LF) đề cập đến việc thu thập thông tin về góc đi của tia sáng trên cảm biến hình ảnh, bên cạnh các thông tin về không gian và cường độ được ghi lại theo cách truyền thống. Việc nắm bắt thông tin góc cạnh của LF, nó giúp dữ liệu LF có nhiều thông tin phong phú. Nghiên cứu đầu tiên về lĩnh vực này đó là hệ thống kết xuất dựa trên hình ảnh (IBR) do Levoy và Hanrahan đề xuất vào năm 1996 [1]. Trong hai thập kỷ qua, đã có một sự cải tiến rất lớn trong các kỹ thuật LF, ví dụ: kỹ thuật thu nhận, kết xuất và lấy mẫu. Đặc biệt, nhiều máy ảnh đã được phát triển để thu dữ liệu LF, ví dụ, Lytro LF, Illum [2] và Raytrix [3]. Những máy ảnh này cung cấp các tính năng tuyệt vời cho dữ liệu LF như thay đổi góc nhìn và góc nhìn, lấy nét lại kỹ thuật số, trích xuất dữ liệu ba chiều (3-D) và ước tính độ sâu [4].

Dữ liệu không lồ liên quan đến công nghệ LF mang lại những tính năng tuyệt vời được liệt kê ở trên nhưng cũng đi kèm với những bất lợi về mặt xử lý tín hiệu và triển khai phần mềm. Đặc biệt, loại dữ liệu này đưa ra các vấn đề về lấy mẫu với khoảng cách mẫu không đều và nhu cầu lấy mẫu theo hướng không gian cũng như dữ liệu với nhiều thông tin tạo ra một thách thức đối với các tác vụ tiền xử lý cổ điển. Tuy nhiên, với thông tin phong phú, dữ liệu LF có một tương lai đầy hứa hẹn. Dữ liệu LF có tiềm năng tăng cường sản xuất hậu kỳ cho quá trình xử lý phim, cho phép trải nghiệm nhập vai mới trong thực tế hỗn hợp, hội nghị từ xa và cải thiện nội dung hình ảnh trong thực tế ảo và tăng cường (VR-AR) [5]. Ngoài ra, màn hình LF cho cảm nhận chân thực nhất khi hiển thị dữ liệu 3D [6].

Để đạt được mục tiêu này, nhiều kỹ thuật khác nhau đã được đề xuất để kiểm soát phương pháp đánh giá và hiệu suất nén. Các kỹ thuật được đề xuất chủ yếu tập trung vào nén đa điểm và đa tiêu điểm từ ảnh nhỏ của khẩu độ phụ thu được từ dữ liệu LF [7]. Một cải tiến gần đây là giải pháp mã hóa H.265/HEVC dựa trên dự đoán tương đồng với LF [8]. Ý tưởng chính là sử dụng dự đoán đa hướng của các khối trong cửa sổ tìm kiếm với thông tin tương tự từ hình ảnh LF. Một ý tưởng thú vị khác là của Dong Liu và cộng sự, để nén tất cả các góc nhìn cấu thành của hình ảnh LF dưới dạng một chuỗi video giả. Cách tiếp cận này đã đạt được mức độ mã hóa đáng kể so với các lược đồ mã hóa dựa trên hình ảnh [9]. Nghiên cứu tại đề xuất [10] đã đưa ra một số kết quả so sánh với các chuẩn mã hóa hình ảnh và video với dữ liệu LF, tại đề xuất này, các kết quả so sánh về hiệu quả nén của dữ liệu LF giữa một số giải pháp mã hóa bao gồm H.265/HEVC, H.264/AVC, JPEG2000 và JPEG. Tuy nhiên, đánh giá này không đề cập đến giải pháp mã hóa gần đây nhất của Google, phần mềm mã hóa video VP9 [12] và độ phức tạp tính toán liên quan đến mỗi bộ mã hóa.

Vì vậy, bài báo này đề xuất một phân tích toàn diện về hiệu quả nén và độ phức tạp với các tiêu chuẩn mã hóa hình ảnh và video gần đây, bao gồm H.265/HEVC [11], Google/VP9 [12], H.264/AVC [13] và JPEG-2000 [14]. Tập dữ liệu hình ảnh LF do T. Ebrahimi và cộng sự cung cấp trong [15].

Phần còn lại của bài viết này được tổ chức như sau. Phần 2 mô tả ngắn gọn về cơ sở lý thuyết cơ bản về hình ảnh LF, trong khi các chi tiết về mã hóa LF với các tiêu chuẩn hình ảnh/video được liệt kê trong Phần 3. Phần 4 chủ yếu phân tích kết quả thử nghiệm cho từng trường hợp, và Phần 5 đưa ra một số kết luận và công việc trong tương lai.

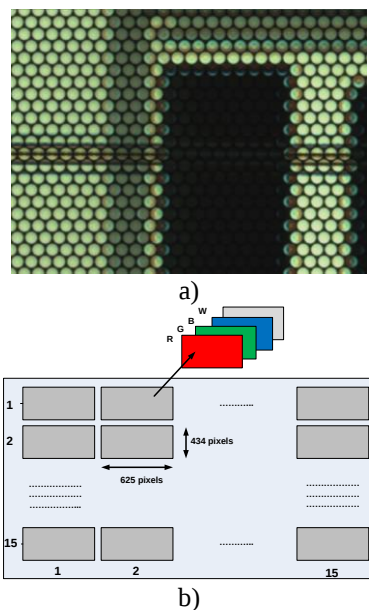
Tác giả liên hệ: Nguyễn Cảnh Châu,

Email: chaunc@ptit.edu.vn

Đến tòa soạn: 10/2021, chỉnh sửa: 11/2021, chấp nhận đăng: 12/2021.

II. CƠ SỞ LÝ THUYẾT

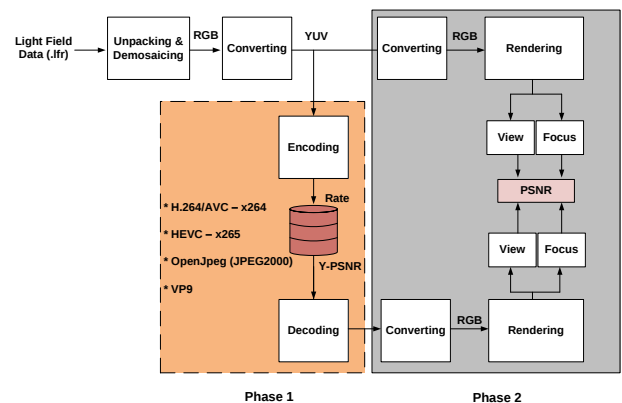
Hình ảnh LF được coi là dữ liệu đầu ra từ cảm biến gốc của máy ảnh [7]. Thông thường, đầu ra cảm biến gốc không hiển thị được trên hầu hết các thiết bị trình chiếu. Để chuyển đổi và xem dữ liệu LF, công cụ LF toolbox (LFT) [16], được sử dụng. Dữ liệu LF có thể được ghi lại bằng nhiều máy ảnh có cấu trúc khác nhau như Raytrix, Lytro Illum, hệ thống mảng máy ảnh Stanford [17], v.v., do đó, mỗi máy ảnh cung cấp dữ liệu khác nhau. Dữ liệu LF được sử dụng trong bài báo này là dữ liệu trong [15] được chụp bởi máy ảnh Lytro Illum, nó hiển thị với độ phân giải 7728×5368 mẫu và với thứ tự màu “GRBG” trong bộ lọc Bayer như trong Hình 1 (a). Sau khi giải nén và chuyển đổi bằng LFT, hình ảnh LF đầy đủ có thể được hiển thị dưới dạng hình ảnh 2D 7728×5368 , hiển thị từng hình ảnh lenslet như trong Hình 1.a. Để sử dụng nhiều chế độ xem và tiêu cự thay đổi của hình ảnh LF, nó phải chuyển đổi hình ảnh 2D thành mảng 4D cung cấp nhiều kích thước là $15 \times 15 \times 434 \times 625$ trong đó 15×15 thể hiện số lượng ống kính nhỏ, 434×625 đại diện cho độ phân giải của mỗi ống kính nhỏ và 4 liên quan đến không gian màu của RGB và thành phần trọng số hình ảnh như được hiển thị trong Hình 1 (b)



Hình 1. a) Hình ảnh OF đầy đủ với thấu kính lớn, b) Cấu trúc dữ liệu OF do LFT tạo ra

Quá trình hiển thị một chế độ xem cụ thể của dữ liệu LF được thể hiện trong Hình 2. Có hai giai đoạn để có được một chế độ xem cụ thể và tiêu điểm của một hình ảnh 2D thông thường sau khi tính toán tỷ lệ biến dạng điểm (Rate Distortion – RD). Tuy nhiên, trong bài báo này, điểm RD chính trong giai đoạn đầu tiên chủ yếu được xem xét để phân tích hiệu quả nén và độ phức tạp của ảnh LF gốc. Theo kích thước lớn của hình ảnh LF, đầu vào LFR ở giai đoạn một được chọn để giải nén và giải mã sau khi cung cấp đầu ra LFR dưới dạng RGB và sau đó chuyển nó thành không gian màu YUV bằng cách sử dụng biểu diễn số nguyên 8 bit được tương thích với phần mềm mã hóa có sẵn, đó là H.264/AVC, H.265/HEVC, Google/VP9 và JPEG-2000. Tại đây, thành phần Y của hình ảnh đã được giải mã được sử dụng để tính RD chính và kích thước của

luồng được mã hóa sẽ trở thành tỷ lệ cho phân tích hiệu suất RD sau đây.



Hình 2. Khung đánh giá việc mã hóa dữ liệu LF

III. MÃ HÓA LIGHT FIELD VỚI CÁC CHUẨN NÉN HÌNH ẢNH VÀ VIDEO

Dữ liệu LF chứa một tập hợp thông tin phong phú và có thể được trình bày dưới dạng ảnh tĩnh hoặc ảnh chuyển động. Do đó, cả hai tiêu chuẩn mã hóa hình ảnh và video đều có thể được áp dụng cho dữ liệu LF như Google/VP9, H.265/HEVC, JPEG2000, và H.264/AVC. Trong phần này, việc sử dụng các tiêu chuẩn trên sẽ được mô tả chi tiết.

A. Google/VP9

VP9 là codec video mã nguồn mở thế hệ tiếp theo của Google. Nó cạnh tranh trực tiếp với bộ giải mã video hiện đại H.265/HEVC. Google/VP9 có một số công nghệ chung với H.265/HEVC như: cây tứ phân (64×64) trong kích thước khối dự đoán, Biến đổi Cosine rời rạc (DCT) trong các loại biến đổi và các chế độ dự đoán như inter và intra. Tuy nhiên, có một vài điểm khác biệt đáng chú ý, tức là sự phân rã khối của VP9 tương tự như cây tứ phân nhưng có 4 chế độ phân vùng, đó là: không có phân vùng, phân vùng ngang, phân vùng dọc và phân chia so với hai chế độ phân vùng (không có và phân chia) của HEVC [12]. Ngoài ra, các loại biến đổi của Google/VP9 không chỉ là DCT, mà còn là Biến đổi hình sin rời rạc bất đối xứng (ADST) được cập nhật [12].

Để đánh giá, chúng tôi sử dụng codec hỗ trợ mã hóa và giải mã của Google/VP9 (libvpx-vp9 phiên bản 1.6.1) có trong FFMPEG phiên bản 3.3.2. Cấu hình mặc định của Google/VP9 là đặt “-b: v 0” cùng với tốc độ không đổi “-crf” và bộ thông số “-deadline”. Cấu hình này cung cấp sự cân bằng tốt giữa chất lượng và tỷ lệ. Có thể tìm thêm chi tiết về các cấu hình trong Bảng I, II và III.

B. H.265/HEVC

Mã hóa video hiệu quả cao (HEVC), được gọi là H.265/HEVC là tiêu chuẩn nén video mới nhất. Như đã báo cáo về đánh giá mã hóa video [11], H.265/HEVC đã đạt được mức tăng nén đáng kể khi so sánh với tiêu chuẩn H.264/AVC trước đây. Điều này đến từ một số lượng lớn các cải tiến trong cả cấu trúc mã hóa, cụ thể là phân chia cây mã hóa với tối đa 64×64 , chế độ mã hóa Hợp nhất, Độ lệch thích ứng mẫu, v.v. [11].

Dựa trên hiệu suất vượt trội này, việc đánh giá nén trên dữ liệu LF là rất có ý nghĩa. Tương tự như H.264/AVC,

FFMPEG phiên bản 3.3.2 được sử dụng để mã hóa và giải mã với codec x265 tương tự như x264 với cài đặt mặc định như trong Bảng I, Bảng II và Bảng III.

C. H.264/AVC

H.264/AVC là định dạng được sử dụng phổ biến để nén và phân phối nội dung video được phát triển bởi Nhóm chuyên gia mã hóa video ITU-T (VCEG) cùng với Nhóm chuyên gia hình ảnh chuyển động (MPEG). Nó là một tiêu chuẩn nén video theo khối dựa trên bù chuyển động [13]. Có hai lớp chính trong H.264/AVC, một là lớp mã hóa video (VCL) chỉ định bù chuyển động, mã hóa biến đổi và mã hóa entropy, trong khi lớp kia là lớp mạng trừu tượng (NAL) bao gồm các lát được mã hóa vào đối tượng mạng trong mạng. Các mục tiêu chính của H.264/AVC là nâng cao hiệu quả mã hóa và tính linh hoạt để sử dụng hiệu quả trên nhiều loại video khác nhau.

Một đơn vị mã hóa cơ bản trong H.264/AVC là Macroblock được mã hóa ở chế độ trong khung (Intra) hoặc liên khung (Inter). Trong bài báo này, chế độ trong khung H.264/AVC Intra được lựa chọn để tính hiệu suất nén vì dữ liệu ảnh LF được kiểm thử không tồn tại mối tương quan thời gian. Để có sự so sánh công bằng, phần mềm được sử dụng để mã hóa và giải mã là FFMPEG phiên bản 3.3.2 [19] hỗ trợ mã hóa x264, triển khai video tương thích H.264/AVC. Nó là một bộ mã hóa chất lượng rất cao và tạo ra các luồng bit chất lượng đáng kể. Tham số lượng tử hóa được đặt ở 16, 24, 31 và 39 tương ứng với tỷ lệ nén 'r' của OpenJPEG trong Bảng II, cấu hình của cài đặt trước là trung bình cho chất lượng cân bằng và mã hóa tốc độ bit cũng như cài đặt mặc định dòng lệnh dưới dạng Bảng I

D. JPEG-2000

JPEG-2000 là một trong những tiêu chuẩn mã hóa hình ảnh mới nhất [14], được tạo ra bởi Nhóm chuyên gia nhiếp ảnh nhằm cải thiện những điểm yếu của tiêu chuẩn trước đó JPEG [14] như: nén tốc độ bit kém, xử lý hình ảnh lớn, kiến trúc nền đơn, v.v. JPEG-2000 không chỉ mang lại hiệu quả nén cao hơn mà còn tạo ra một hình ảnh biểu diễn mới với một bộ tính năng phong phú so với JPEG, tức là khả năng mở rộng và truy cập ngẫu nhiên. Biến đổi Wavelet là một trong những lý do chính hỗ trợ cho các tính năng linh hoạt này. Nội dung này các bạn có thể đọc mô tả đầy đủ hơn về JPEG-2000 trong [14].

Việc thực hiện mã hóa và giải mã JPEG-2000 được sử dụng trong bài báo này là OpenJPEG phiên bản 2.1.2 [18]. Đây là bộ mã hóa JPEG 2000 mã nguồn mở và phần mềm được cấu hình mặc định cho tất cả các cài đặt. Tỷ lệ biến dạng (RD) được đặt bằng tỷ lệ nén (CR) 'r' theo 4 tỷ lệ, cụ thể là 10: 1, 20: 1, 40: 1 và 100: 1. Các dòng lệnh mã hóa và giải mã được trình bày chi tiết trong Bảng I trong khi Bảng II trình bày các tỷ lệ nén tương ứng giữa các bộ mã hóa.

Bảng I. Một số câu lệnh sử dụng để mã hóa

Bộ mã hóa	Câu lệnh
OpenJPEG	JPEG-2000 Encode opj_compress -i input.bmp -o output.jp2 -r selected ratio
	JPEG-2000 Decode opj_decompress -i output.jp2 -o output.bmp

JPEG		
FFMPG-EG	HEVC (x265) encode	ffmpeg -f rawvideo -vcodec rawvideo -s 7728x5368 -r 1 -pix_fmt yuv420p -i input.yuv -c:v libx265 -preset medium -crf selected rate output.mp4
	HEVC (x265) decode	ffmpeg -i input.mp4 -pix_fmt yuv420p -s 7728x5368 -r 1 output.yuv
	H.264/AVC (x264) encode	ffmpeg -f rawvideo -vcodec rawvideo -s 7728x5368 -r 1 -pix_fmt yuv420p -i input.yuv -c:v libx264 -preset medium -crf selected rate output.mp4
	H.264/AVC (x264) decode	ffmpeg -i input.mp4 -pix_fmt yuv420p -s 7728x5368 -r 1 output.yuv
	VP9 (libvpx-vp9) encode	ffmpeg -f rawvideo -vcodec rawvideo -s 7728x5368 -r 1 -pix_fmt yuv420p -i input.yuv -c:v libvpx-vp9 -b:v 0 -deadline good -crf selected rate output.mp4
	VP9 (libvpx-vp9) decode	ffmpeg -i input.mp4 -pix_fmt yuv420p -s 7728x5368 -r 1 output.yuv

Bảng II. Các tham số lượng tử (qps) sử dụng trong các bộ mã hóa

Bộ mã hóa	R1	R2	R3	R4
JPEG-2000	10:1	20:1	40:1	100:1
x264	16	24	31	39
x265	16	24	31	39
VP9	13	30	41	53

Bảng III. Các cấu hình cho các bộ mã hóa

Bộ mã hóa	x264	x265	VP9
Encoder	Libx264	Libx265 v2.4	Libvpx-vp9 v1.6.1
Profile	High	Main	-
Level	6.2	8.5	-
8x8 Transform	On	-	-
Rate Control	On	On	On
Internal Bit Depth	8	8	8

IV. ĐÁNH GIÁ VÀ KIỂM THỬ

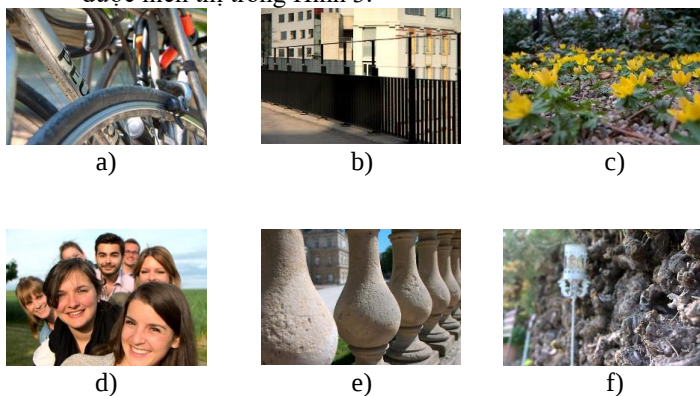
Trong phần này, các trường hợp thử nghiệm khác nhau được cung cấp và chúng tôi kiểm tra hiệu quả nén và độ phức tạp với các bộ mã hóa khác nhau.

A. Phương pháp đánh giá

Trong bài báo này, phương pháp thử nghiệm sử dụng các tiêu chuẩn mã hóa khác nhau để tính toán tốc độ bit chính và chất lượng của hình ảnh LF gốc. Hình ảnh LF với độ phân giải 7728×5368 sẽ được coi là mẫu đầu vào cho bài kiểm tra hiệu suất. Thông thường, độ chói (thành phần Y) của hình ảnh là một giá trị hiệu quả cho quá trình nén hơn là giá trị màu [10]. Do đó, hình ảnh LF gốc sau khi được giải mã trong Hình 2, được chuyển đổi sang định dạng YUV với giá trị Y được lựa chọn để xử lý tiếp theo. Mỗi tiêu chuẩn mã hóa sẽ được cài đặt mặc định (như trong Bảng 2) và được sử dụng để nén thành phần Y của 6 hình

ảnh LF đã được giải mã như trong Hình 3.

Để hiển thị nội dung, hình thu nhỏ của các hình ảnh LF được hiển thị trong Hình 3.



Hình 3. Các hình ảnh mẫu của LF: (a) Bikes, (b) Black Fence, (c) Flowers, (d) Friends 1, (e) Stone Pillars Outside, (f) Vine Wood

B. Đánh giá hiệu năng nén

Đối với nội dung video, HEVC (x265) và Google/VP9 (libvpx-vp9) được cho là sẽ hoạt động tốt hơn các bộ mã hóa khác như H.264/AVC (x264) và JPEG-2000. Trong đánh giá này, hiệu suất RD được so sánh và trình bày trong Hình 4 và khả năng tiết kiệm bitrate theo Bjøntegaard Delta (BD) -Rate [20] so với JPEG được tính trong Bảng IV.

Bảng IV. Phần trăm (%) bitrate tiết kiệm so sánh với jpeg-2000

Dữ liệu LF	H.264/AVC	H.265/HEVC	VP9
Bikes	0.79	-26.95	-25.53
Black Fence	2.85	-23.19	-20.78
Flowers	-4.71	-28.90	-28.61
Friends 1	9.02	-18.17	-17.98
Stone Pillars Outside	1.14	-25.73	-22.77
Vine Wood	0.14	-22.02	-20.64
Trung bình	1.53	-24.16	-22.71

Từ đánh giá tỷ lệ BD-Rate của tất cả các tiêu chuẩn với JPEG-2000, như thể hiện trong Bảng IV và Hình 4, có thể rút ra một số kết luận:

- Khi nén hình ảnh LF, chi khai thác mối tương quan không gian, hiệu quả nén của tiêu chuẩn mã hóa video H.264/AVC gần với tiêu chuẩn mã hóa hình ảnh JPEG, đáng chú ý là chỉ tăng khoảng 1,5% tỷ lệ BD-Rate trong khi cung cấp chất lượng xem tương tự.
- H.265/HEVC (x265) vẫn là lựa chọn tốt nhất cho hiệu quả nén trong mọi tình huống với tốc độ bit tiết kiệm được khoảng 24% so với JPEG-2000. Điều này chủ yếu đến từ một số lượng lớn các chế độ dự đoán Intra được sử dụng trong HEVC [11].
- Kết quả cho Google/VP9 (libvpx-vp9) chỉ ra rằng hiệu suất nén gần bằng H.265/HEVC (x265), đặc biệt là ở chế độ nội bộ Intra (nén một khung hình).

- Ngoài ra còn có một quan sát thú vị về hiệu suất H.265/HEVC (x265) và Google/VP9 (libvpx-vp9). Cả hai bộ mã hóa đều có một số so sánh trong [21, 22] mà lựa chọn ra bộ mã hóa nào hoạt động tốt hơn. Kết quả cho thấy H.265/HEVC vượt trội hơn Google/VP9 về hiệu quả nén [21, 22], nhưng kết quả trong Hình 4 và Bảng IV đưa ra một góc nhìn mới. Hiệu quả nén của Google/VP9 (libvpx-vp9) có kết quả ấn tượng khi nó rất gần với chất lượng và tỷ lệ H.265/HEVC (x265) trong tất cả các thử nghiệm, đặc biệt là trong Bikes, Flowers và Friends_1.

C. Đánh giá thời gian xử lý

Độ phức tạp của nén là một phần thiết yếu trong đánh giá hiệu suất. Đối với đánh giá này, các tiêu chuẩn mã hóa được thử nghiệm trên cùng một PC với bộ xử lý Intel core i7-7700HQ (2,8 GHz), RAM 8GB và HĐH Windows 10-Home. Thử nghiệm được chạy trên tất cả các tỷ lệ nén và các thông số lượng tử hóa được liệt kê trong Bảng I và II. Các kết quả hiển thị trong Bảng V và Bảng VI lần lượt cho R1 (CR 10: 1) và R4 (CR 100: 10). Để tránh ảnh hưởng của quá trình xử lý đa luồng trong quá trình thử nghiệm, kết quả của 5 lần lặp lại của cùng một cài đặt nén được tính trung bình.

Dựa trên kết quả hiển thị trong Bảng V với CR ở mức 10: 1, bộ mã hóa nhanh nhất là H.264/AVC (x264) nhanh hơn 7 lần so với bộ mã hóa Google/VP9 (libvpx-vp9) và nhanh hơn 3 lần so với H.265/HEVC (x265) và JPEG-2000. Nhưng kết quả ấn tượng nhất được thể hiện trong Bảng VI với CR ở tỷ lệ 100: 1, trong đó codec H.264/AVC (x264) vẫn là bộ mã hóa nhanh nhất so với bộ mã hóa chậm nhất Google/VP9 (libvpx-vp9). Thời gian mã hóa H.264 (x264) và H.265/HEVC (x265) giảm đi hai lần so với kết quả thu được đối với CR ở tỷ lệ 10: 1, trong khi thời gian mã hóa Google/VP9 (libvpx-vp9) được tăng lên khoảng 1,3 lần và JPEG-2000 hầu như vẫn giữ nguyên.

Bảng V. Thời gian xử lý giữa các chuẩn mã hóa tại chỉ số nén 10:1

Dữ liệu LF	Bộ mã hóa			
	X265	X264	VP9	JPEG 2000
Bikes	8.63	3.31	22.45	9.78
Black Fence	8.61	3.35	22.36	10.25
Flowers	8.92	3.27	22.65	9.47
Friends 1	9.15	3.55	22.41	9.69
Stone Pillars Outside	9.15	3.55	22.41	9.69
Vine Wood	9.42	3.59	23.15	10.40
Trung bình	8.98	3.43	22.57	9.88

Từ các kết quả về độ phức tạp, một số điểm được nhận thấy dưới đây:

- H.265/HEVC (x265) là bộ mã hóa cân bằng với chất lượng cao về hiệu quả nén cùng với thời gian mã hóa

- hợp lý ở tất cả các giá trị của thông số lượng tử hóa. H.264/AVC (x264) và JPEG-2000 có cùng hiệu suất về mặt hiệu quả, tuy nhiên H.264/AVC (x264) là bộ mã hóa nhanh nhất trong số các tiêu chuẩn này, đáng chú ý là nhanh hơn khoảng 20 lần so với Google/VP9 (libvpx-vp9) bộ mã hóa ở CR 100: 1
- Bộ mã hóa chậm nhất là Google/VP9 (libvpx-vp9) cho tất cả các thử nghiệm. Đây có thể là một sự đánh đổi phức tạp hơn hiệu quả để có được hiệu suất chất lượng gần với H.265/HEVC (x265).

Bảng VI. Thời gian xử lý giữa các chuẩn mã hóa tại chỉ số nén 100:1

Dữ liệu LF	Bộ mã hóa			
	X265	X264	VP9	JPEG 2000
Bikes	4.27	1.07	28.4	9.23
Black Fence	4.55	1.16	30.23	9.55
Flowers	4.12	1.01	25.42	8.82
Friends 1	4.37	1.16	27.94	9.33
Stone Pillars Outside	4.35	1.12	26.34	9.05
Vine Wood	4.55	1.14	32.20	9.59
Trung bình	4.37	1.11	28.42	9.26

V. KẾT LUẬN

Bài báo này đề xuất một đánh giá hiệu suất toàn diện cho mã hóa hình ảnh LF với các tiêu chuẩn mã hóa hình ảnh và video tiên tiến gần đây. Như đã trình bày, hiệu suất nén của H.265/HEVC (x265) và Google/VP9 (libvpx-vp9) vượt trội hơn đáng kể so với JPEG-2000 trong khi hiệu suất nén của H.264/AVC (x264) gần bằng JPEG-2000. Ngoài ra, để đạt được hiệu suất nén gần với HEVC (x265), Google/VP9 (libvpx-vp9) cần số lượng tính toán lớn hơn. Những kết quả thử nghiệm này là những bước khởi đầu hỗ trợ cho việc nghiên cứu LF trong tương lai. Theo đó, công việc trong tương lai có thể tập trung vào phân tích sâu về hiệu quả nén và độ phức tạp với HEVC (HM) và Google/VP9 trong hình ảnh khẩu độ phụ của LF.

TÀI LIỆU THAM KHẢO

- [1] Marc Levoy and Pat Hanrahan, "Light field rendering," in Proc. SIGGRAPH, 1996, pp. 31-42.
- [2] Lytro camera, <https://www.lytro.com/>
- [3] Raytrix, <https://www.raytrix.de/>
- [4] Ivo Ihrke, John Restrepo, and Loïs Mignard-Debise, "Principles of Light Field Imaging", IEEE Signal Processing Magazine, 2016
- [5] Touradj Ebrahimi, Siegfried Foessel, Fernando Pereira, Peter Schelkens, "JPEG Pleno: Toward an Efficient Representation of Visual Reality," IEEE Multimedia, Oct-Dec 2016.
- [6] M. Lamboij and W. Ijsselstein, "Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review", Journal of Imaging Science and Technology, 53(3), 030201-030201-14, 2009.

- [7] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, Pat Hanrahan, "Light Field Photography with a Hand-held Plenoptic Camera," Stanford Tech Report CTSR, 2005.
- [8] C. Conti, P. Nunes and L. D. Soares, "HEVC-based light field image coding with bi-predicted self-similarity compensation," 2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Seattle, WA, 2016, pp. 1-4.
- [9] Dong Liu, Lizhi Wang, Li Li, Zhiwei Xiong, Feng Wu, Wenjun Zeng, "Pseudo-Sequence-Based Light Field Image Compression", Multimedia & Expo Workshops (ICMEW), IEEE International Conference, July 2016
- [10] G. Alves, F. Pereira and E. A. B. da Silva, "Light field imaging coding: Performance assessment methodology and standards benchmarking," 2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Seattle, WA, 2016, pp. 1-6.
- [11] G. J. Sullivan, J. R. Ohm, W. J. Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.
- [12] D. Mukherjee et al., "A Technical Overview of VP9 – The Latest Open-Source Video Codec," SMPTE 2013 Annual Technical Conference & Exhibition, Hollywood, CA, USA, , pp. 1-17, 2013.
- [13] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC video coding standard," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560-576, July 2003.
- [14] M. Rabbani, R. Joshi, "An overview of the JPEG2000 still image compression standard", Signal Processing: Image Communication 17, 3-48, 2002.
- [15] M. Řeřábek and T. Ebrahimi, "New Light Field Image Dataset," 8th International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, 2016.
- [16] D. Dansereau, "Light Field Toolbox v0.4," <http://www.mathworks.com/matlabcentral/fileexchange/49683-light-field-toolbox-v0-4>.
- [17] Stanford Multi-Camera Array, <http://graphics.stanford.edu/projects/array/>
- [18] OpenJPEG software, <http://www.openjpeg.org/>
- [19] FFMPEG software, <https://www.ffmpeg.org/>
- [20] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves", Doc. ITU-T SG16 VCEG-M33, Austin, TX, USA, April 2001.
- [21] M. Rerabek, T. Ebrahimi, "Comparison Of Compression Efficiency Between Hvc/H.265 And Vp9 Based On Subjective Assessments", Multimedia Signal Processing Group (MMSPG), Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland, 2014.
- [22] D. Grois, D. Marpe, A. Mulayoff, B. Itzhaky and O. Hadar, "Performance comparison of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC encoders," 2013 Picture Coding Symposium (PCS), San Jose, CA, 2013, pp. 394-397.

ANALYZE THE PERFORMANCE OF LIGHT FIELD IMAGE WITH VIDEO AND IMAGE STANDARDS

Abstract – Recently, Light Field (LF) imaging technology has attracted much interest due to its potential applications in a large number of fields including Virtual Reality, Augmented Reality (VR) /AR), Teleconference and E-learning. LF data is a multidimensional dataset that provides rich visual information such as scene rendering with variable depth of field, point of view, and focal length changes that the viewer can change. However, LF data also has an important problem, that is, the amount of data stored is too large. Therefore, compression of LF data is one of

the central issues for LF research. This paper presents the analysis in three main parts. First, the paper comprehensively presents the LF image encoding stream, which can be used as a benchmark for evaluating LF compression in the future. Second, the paper briefly describes the existing image and video coding standards, which are applied to LF data compression, including three well-known video coding standards, H.264/AVC, H.265/HEVC, Google/VP9, and image coding standard, JPEG-2000. Finally, the compression efficiency and complexity of these standards will be compared and discussed in detail.

Keywords—Light Field Data, H.264/AVC, H.265/HEVC, VP9, JPEG-2000.



Nguyễn Cảnh Châu tốt nghiệp đại học tổng hợp, khoa Toán cơ tin và Vật lý vô tuyến (nay là Đại học Khoa học tự nhiên – Đại học quốc gia Hà Nội, đang làm nghiên cứu sinh tại Học viện Công nghệ Bưu chính Viễn thông. Hiện đang công tác tại Khoa Đa phương tiện, Học viện Công nghệ Bưu chính Viễn thông, nghiên cứu chuyên sâu các lĩnh vực liên quan tới xử lý ảnh, video, đồ họa máy tính.