

# CÁC CÔNG CỤ MÃ HÓA CỦA CHUẨN H.266/VVC VỚI DỮ LIỆU ẢNH LIGHT FIELD

Phí Công Huy, Nguyễn Cảnh Châu

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

**Abstract**—Các tiêu chuẩn mã hóa video đã được phát triển để truyền và lưu trữ video kỹ thuật số một cách hiệu quả, đáp ứng nhiều yêu cầu như bitrate, chất lượng hình ảnh, độ trễ, khả năng truy cập ngẫu nhiên, độ phức tạp, và nhiều yếu tố khác. Bên cạnh đó, dữ liệu kỹ thuật số cũng được khai thác để đáp ứng nhu cầu của người dùng trong trải nghiệm xem, bao gồm cả hình ảnh trường sáng (Light Field), điểm đám mây (Point Cloud), và hình ảnh phản chiếu (Holography). Hình ảnh trường sáng hay còn gọi là dữ liệu ảnh Light Field, là một phương pháp thu thập dữ liệu plenoptic cho phép xử lý hình ảnh đa dạng như trích xuất 3D, thay đổi điểm nhìn, và lấy nét kỹ thuật số (lấy nét sau khi chụp). Hơn nữa, Light Field cung cấp khả năng ghi lại thông tin phong phú về một cảnh chụp, ví dụ như cấu trúc, thông tin hình học, v.v. Do đó, việc nén hiệu quả đóng vai trò quan trọng đối với Light Field do lượng dữ liệu lớn cần thiết để biểu diễn. Các tiêu chuẩn mã hóa video tiên tiến như H.264/AVC, H.265/HEVC, và H.266/VVC đã xem xét các đặc điểm của video thông thường để cải thiện hiệu suất nén. Hiện nay, tiêu chuẩn mã hóa video mới nhất - H.266/VVC cung cấp hiệu suất nén cải thiện so với các tiêu chuẩn khác, nhưng đi kèm với độ phức tạp tính toán rất cao, do việc thêm vào một số công cụ mã hóa mới (ví dụ: chế độ phân vùng tam giác (TPM), đơn vị cây mã hóa (CTU) lớn với 128x128 điểm ảnh, v.v.). Tuy nhiên, chưa có một phân tích đầy đủ nào về các công cụ mã hóa cho bộ mã hóa này để tận dụng tối đa hiệu suất hiệu quả của nó, đặc biệt đối với dữ liệu Light Field. Do đó, bài báo này trình bày một nghiên cứu toàn diện về các công cụ mã hóa của tiêu chuẩn mã hóa video - H.266/VVC cho dữ liệu Light Field. Các công cụ mã hóa của H.266/VVC (ví dụ: SMVD, MIP, IBC, v.v.) được đánh giá trên các đoạn video giả lập, từ đó thống kê kết quả về hiệu suất nén và độ phức tạp.

**Keywords**— Mã hóa dữ liệu Light Field, chất lượng ảnh Light Field, độ phức tạp, Chuẩn mã hóa H.266/VVC.

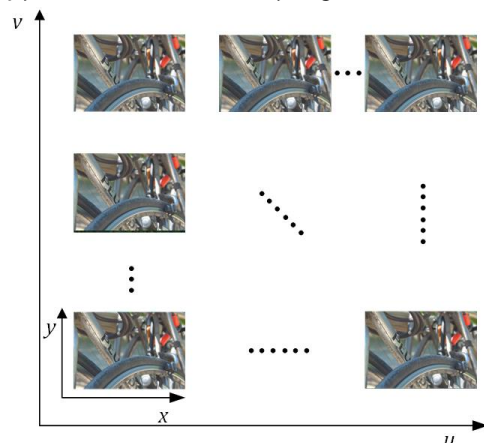
## I. GIỚI THIỆU CHUNG

Dữ liệu hình ảnh trường sáng (LF) đã tiến xa và được áp dụng rộng rãi trong nhiều lĩnh vực như y học (như kính soi tai [1]), kỹ thuật viễn thám (như kính hiển vi [2]), điều

khiển robot dựa trên thị giác [3], và đo vận tốc [4]. Điều này bắt nguồn từ cấu trúc đặc biệt của máy ảnh LF. Máy ảnh LF không chỉ ghi lại cường độ ánh sáng mà còn ghi lại hướng ánh sáng đến cảm biến của máy, thông qua các thấu kính nhỏ được thêm vào ống kính chính trước cảm biến máy ảnh [5, 6]. Do đó, dữ liệu ảnh LF có cấu trúc phức tạp và kích thước lớn, đồng thời đòi hỏi phân tích và khám phá kỹ lưỡng.

Dữ liệu ảnh LF biểu diễn tập hợp các tia sáng truyền theo mọi hướng góc tại mọi điểm trong không gian 3D [6], và do đó chứa một lượng lớn thông tin về mỗi tia sáng bao gồm vị trí  $x, y, z$ , góc  $\Theta, \Phi$ , bước sóng  $\gamma$ , và thời gian  $t$ . Dữ liệu ảnh LF có kích thước lớn và thường được mô tả bằng hàm đa sơ đồ 7D như sau, LF  $(x, y, z, \Theta, \Phi, \gamma, t)$  [6]. Tuy nhiên, tính phức tạp của hàm chức năng này mang lại nhiều thách thức trong việc chụp và xử lý hình ảnh trong thực tế, do đó, hàm chức năng này thường được giảm xuống thành hàm chức năng 4D.

Dữ liệu hình ảnh bao gồm các thông tin như không gian  $(x, y, z)$ , thời gian  $(t)$ , vị trí  $(\theta, \phi)$  và bước sóng của ánh sáng  $(\lambda)$ . Nó cung cấp biểu diễn ảnh LF. Trong Hình 1, tham số  $P_{LF}(x, y, z, \theta, \phi, \lambda, t)$  như Hình 1. Để đơn giản hóa tham số của LF gốc, chức năng LF đã được giảm xuống còn 4 tham số hay còn gọi là 4D-LF,  $P_{LF} = L(u, v, x, y)$ , với  $(x, y)$  là điểm nhìn còn  $(u, v)$  là góc nhìn.



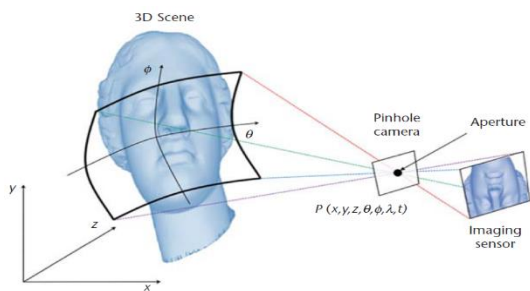
Hình 1. Cấu trúc ảnh Light Field theo khẩu độ (Sub-Aperture Image -SAI)

Một hình ảnh LF gốc bao gồm các hình ảnh cực nhỏ (MI) và một tập hợp các hình ảnh khẩu độ kèm theo (SAI) thu được bằng cách sắp xếp lại các pixel cùng vị trí từ mỗi MI như Hình 2.

Tác giả liên hệ: Phí Công Huy,

Email: [huy@ptit.edu.vn](mailto:huy@ptit.edu.vn)

Đến tòa soạn: 10/2023, chỉnh sửa: 11/2023, chấp nhận đăng: 12/2023.



Hình 2. Các loại thông tin thể hiện cho dữ liệu Light Field

Dữ liệu hình ảnh LF, do cấu trúc của nó, có dung lượng khoảng 50 MB. Gần đây, vấn đề mã hóa LF đã thu hút sự chú ý để giải quyết vấn đề này. Nhu cầu lưu trữ được giải quyết bằng các phương pháp xử lý khác nhau, bao gồm nén dữ liệu LF và tái tạo chế độ xem. Năm 2017, nhóm nghiên cứu chuẩn hóa JPEG đã đưa ra mục tiêu nghiên cứu về nén 4D LF [2] và chuẩn hóa JPEG Pleno đã được giới thiệu, có dữ liệu ảnh LF [3].

Tiêu chuẩn mã hóa video mới nhất, H.266/VVC [7], là một tiêu chuẩn mã hóa video mới xuất hiện, được công bố chung bởi ITU-T và ISO/IEC. Dự án mã hóa video này đã khởi đầu từ năm 2015 trong giai đoạn khám phá; trong giai đoạn phát triển, nhóm cộng tác chung đã bắt đầu soạn thảo phiên bản đầu tiên của codec này vào năm 2018. Vào tháng 7 năm 2020, codec VVC được công bố là tiêu chuẩn nén video thế hệ kế tiếp. Tiêu chuẩn video này được thiết kế để đáp ứng nhu cầu truyền thông đa phương tiện hiện đại bằng cách cải thiện hiệu suất nén, cho phép truyền thông và lưu trữ hiệu quả nội dung video lớn.

Để tăng hiệu suất mã hóa, H.266/VVC đã thêm vào nhiều công cụ mã hóa mới so với tiêu chuẩn trước đó H.265/HEVC [8], như độ phân giải vectơ chuyển động thích ứng (AMVR), chế độ phân chia tam giác (TPM), dự đoán nội suy dựa trên ma trận (MIP), phân chia khối linh hoạt với nhiều loại cây (MTT), dự đoán song hướng với trọng số CU (BCW), biến đổi phân khối con (SBT), đơn vị cây mã hóa 128x128 điểm ảnh lớn (CTU), và nhiều cải tiến nhỏ trong thiết kế mã hóa video hỗn hợp cổ điển [7]. Dựa trên sự cải tiến của các công cụ mã hóa mới, giảm tỷ lệ bit được báo cáo là khoảng 50% so với H.265/HEVC, đặc biệt, H.266/VVC có thể được sử dụng cho video siêu nét, ví dụ 4K hoặc 8K, cũng như có phạm vi động rộng và không gian màu rộng [7]. Đối với tính đa dạng, tiêu chuẩn mã hóa mới bao gồm ba khía cạnh quan trọng, được giải quyết cho các ứng dụng cụ thể, tức là mã hóa nội dung màn hình cho nội dung mới nổi (ví dụ: chơi game, chia sẻ màn hình), tái mẫu ảnh tham chiếu cho truyền straming thích ứng với việc chuyển đổi độ phân giải và các phần hình ảnh con độc lập cho việc truyền streaming của video 360 độ [7].

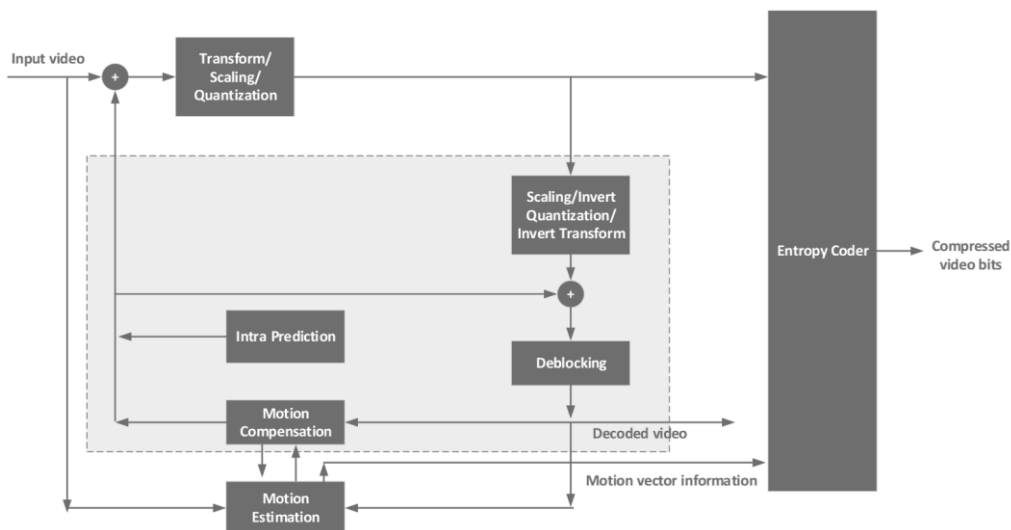
Một chủ đề nghiên cứu đang thu hút sự chú ý trong ngữ cảnh này là việc nén hình ảnh LF dựa trên tiêu chuẩn H.266/VVC. Một số nghiên cứu đã đề xuất và đánh giá hiệu suất nén LF của codec H.266/VVC. Ví dụ, Avramelos và đồng nghiệp [9] đã thực hiện nghiên cứu đầu tiên về việc nén hình ảnh LF sử dụng H.266/VVC bằng cách so sánh hiệu suất mã hóa của ba thế hệ tiêu chuẩn nén video mới nhất: H.264/AVC [10], H.265/HEVC [8] và H.266/VVC [7]. Nghiên cứu gần đây của Amirpour và đồng nghiệp [11]

tập trung vào phần mềm mã hóa bao gồm HM, VP9 [12], AV1 [13], XVC [14], x265 [8], và VTM [7]. Nghiên cứu này đã so sánh hiệu suất của các trình mã hóa video trong việc nén hình ảnh LF. Tuy nhiên, Umabayashi và đồng nghiệp [15] đã sử dụng hình ảnh đa tập trung và các quan điểm tham chiếu để nghiên cứu nén LF 4D với codec VVC. Về phương pháp mô hình mạng, công việc của Bakir và đồng nghiệp [16] đã sử dụng khả năng mở rộng thời gian của VVC để chọn các quan điểm tham chiếu và bị loại bỏ cho hình ảnh LF với mô hình Mạng Sinh học đối lập kép (GAN) hai bộ phận. Mặt khác, mã hóa VVC nội suy với độ phức tạp thấp cho học tập hai chế độ nhẹ là chủ đề của Pakdaman và đồng nghiệp [17].

Dựa trên tài liệu nghiên cứu, có thông tin cho thấy hiệu suất nén xuất sắc của H.266/VVC đồng thời đi kèm với độ phức tạp tính toán đáng kể cao hơn nhiều lần so với H.265/HEVC. Ví dụ, trong các cấu hình Low-Delay (LD) và All-Intra (AI), H.266/VVC đòi hỏi độ phức tạp tính toán lần lượt là 5 và 31 lần so với mã hóa H.265/HEVC [18]. Mức độ phức tạp tính toán gia tăng này đặt ra một thách thức đáng kể đối với việc triển khai mã hóa thời gian thực sử dụng H.266/VVC, đặc biệt trong bối cảnh truyền phát video. Để đạt được sự cân bằng hợp lý giữa độ phức tạp tính toán và hiệu suất nén trong codec VVC, việc sử dụng các công cụ mã hóa xuất phát có vai trò quan trọng. Tuy nhiên, hiện tại còn thiếu tài liệu nghiên cứu phân tích khía cạnh này. Vì lý do đó, bài báo này mang lại phân tích toàn diện đầu tiên về các công cụ mã hóa của H.266/VVC áp dụng vào hình ảnh LF. Các đóng góp của bài báo này có thể được tóm tắt như sau:

- ***Phân tích toàn diện về các công cụ mã hóa phạm vi rộng:*** Các công cụ mã hóa mới của H.266/VVC được phân loại và nghiên cứu kỹ lưỡng với trạng thái bật hoặc tắt. Cụ thể, chúng được chia thành các nhóm như: nhóm phân chia khối (ví dụ: Quadtree kết hợp MTT, v.v.), nhóm dự đoán chuyển động giữa các khung hình (ví dụ: Dự đoán Vector Chuyển Động Thời Gian Dựa Trên Phân Khối (SBTMVP), Hiệu Chênh Vector Chuyển Động Đối Xứng (SMVD), v.v.), nhóm dự đoán trong cùng khung hình (ví dụ: Nhiều Đường Tham Chiếu (MRL), Dự đoán Trong Các Khung Hình Dựa Trên Ma Trận (MIP), v.v.), và nhóm biến đổi và lượng tử hóa (ví dụ: Lượng tử hóa Phụ Thuộc (DQ), Lựa Chọn Nhiều Biến Đổi (MTS), v.v.). Kết quả bao gồm độ phức tạp tính toán và hiệu suất nén.
- ***Phân tích khả năng nén dựa trên nội dung LF:*** Nghiên cứu tập trung vào nhiều điều kiện nội dung LF khác nhau, chẳng hạn như điều kiện hẹp và rộng (liên quan đến sâu độ trường ảnh), điều kiện phức tạp và đơn giản (liên quan đến mức độ phức tạp của đối tượng trong hình ảnh LF), và điều kiện tiền cảnh và hậu cảnh (liên quan đến điểm lấy nét), được lựa chọn từ tập dữ liệu EPFL để cung cấp một phân tích kỹ lưỡng về độ phức tạp tính toán và hiệu suất nén của H.266/VVC.

Cấu trúc của bài báo được phân chia như sau. Phần 2 trình bày bối cảnh và công trình liên quan đến các bộ mã hóa video cũng như tổng quan về các công cụ mã hóa trên



Hình 3. Cấu trúc mã hóa ở cấp cao

H.266/VVC. Phần 3 trình bày phân tích về các công cụ mã hóa này kèm theo phương pháp quan sát và đánh giá. Sau đó, phần 4 triển khai đánh giá thực nghiệm. Cuối cùng, phần 5 tóm tắt công trình và tiến hành thảo luận.

## II. NHỮNG NGHIÊN CỨU LIÊN QUAN

Trong phần này, chúng tôi sẽ tiến hành một tóm tắt và mô tả ngắn gọn về các nghiên cứu liên quan, đặc biệt tập trung vào các công cụ mã hóa đổi mới của mỗi tiêu chuẩn mã hóa video, cũng như các công trình liên quan đến việc nén hình ảnh LF bằng cách sử dụng H.266/VVC.

### A. Các chuẩn mã hóa video

Trong suốt ba thập kỷ vừa qua, các tiêu chuẩn mã hóa video đã phát triển trong Nhóm Chuyên gia Mã hóa Video (ITU-T VCEG) của Tổ chức Tiêu chuẩn Viễn thông Quốc tế (ITU-T). H.261, ra đời vào năm 1990, đã định rõ cách mã hóa và giải mã video kèm âm thanh ở tốc độ  $p \times 64$  kbit/s, trong đó  $p$  nằm từ 1 đến 30. Hơn nữa, H.261 cũng xác định hai định dạng hình ảnh, gọi là Định dạng Trung gian Phổ biến (CIF) và Định dạng Trung gian Phổ biến Gấp Đôi (QCIF). Codec này quan trọng đối với hội nghị video và hệ thống truyền thông, tuy nhiên, nó phù hợp hơn cho Frame Relay hoặc Internet TCP/IP vì tối ưu hóa cho tốc độ dữ liệu thấp và video ít chuyển động.

Sau phiên bản đầu tiên của H.261, ITU đã bổ sung thêm tính năng và đưa vào nhiều thành viên mới vào gia đình mã hóa video H.26x, bao gồm H.263, H.264/AVC, H.265/HEVC và bản mới nhất là H.266/VVC. Mỗi thành viên trong gia đình H.26x mang lại ưu điểm và hạn chế riêng, phù hợp với yêu cầu cụ thể của người dùng và dữ liệu. Trong thực tế, H.264/AVC vẫn là codec phổ biến nhờ tính đơn giản của nó. Cấu trúc mã hóa tiêu chuẩn này giống với các tiêu chuẩn video lớn trước đó, dựa trên mã hóa biến đổi được bù trừ chuyển động.

Tiêu chuẩn H.264/AVC thực hiện nén hình ảnh từ một chuỗi bằng cách chia thành một hoặc nhiều "slice". Mỗi slice bao gồm các "macroblock" có kích thước cố định ( $16 \times 16$  mẫu). Các công cụ mã hóa bao gồm: dự đoán không gian nội (dựa trên các khối), dự đoán thời gian ngoại, tính năng mã hóa xen kẽ, lượng tử hoá véc-tơ, bộ lọc giảm nhiễu, v.v. Công cụ mới của H.264/AVC so với tiêu chuẩn

trước đó là bộ lọc giảm nhiễu linh hoạt và dự đoán nội khung. Bộ lọc giảm nhiễu linh hoạt được sử dụng để giảm các hiện tượng artefact khối trong vòng lặp dự đoán. Đối với việc dự đoán khối cần mã hóa, chế độ nội của tiêu chuẩn H.264/AVC sử dụng tín hiệu hình ảnh của các macroblock truyền đi trong cùng một hình ảnh. Phép biến đổi số nguyên của Biến đổi Cosin Rời rạc (DCT) được áp dụng để thay thế các tiêu chuẩn trước đó. Điều quan trọng là tiêu chuẩn này có khả năng lưu trữ nhiều khung video hơn so với tiêu chuẩn trước đó (Hình 3).

Để cải thiện hiệu suất mã hóa, tiêu chuẩn H.265/HEVC đã được hoàn thiện vào năm 2013 với cấu trúc tương tự như tiêu chuẩn tiền nhiệm nhưng có thêm một số tính năng nổi bật. Tiêu chuẩn này giới thiệu biểu diễn khối dựa trên cây phân tầng (Quadtree) và Đơn vị Cây Mã hóa (CTU) với kích thước lên đến  $64 \times 16$  mẫu. CTU có thể được chia thành các "CUs" với cấu trúc cây phân tầng để có biểu diễn khối linh hoạt hơn. Ngoài ra, Tiên đoán Vector Chuyển động Nâng cao (AMVP) được sử dụng trong việc truyền tín hiệu vector chuyển động (MV), kèm theo chế độ hợp nhất cho việc mã hóa vector chuyển động. Điều quan trọng là tiêu chuẩn H.265/HEVC sử dụng bộ lọc 7-tap hoặc 8-tap để nội suy vị trí mẫu phân số thay vì bộ lọc 6-tap cho vị trí mẫu nửa. Hơn nữa, dự đoán hình ảnh nội của tiêu chuẩn này hỗ trợ 33 chế độ hướng so với 8 chế độ trong tiêu chuẩn H.264/AVC, bao gồm chế độ dự đoán phẳng và DC.

Tiêu chuẩn mã hóa video mới nhất là H.266/VVC, được công bố vào tháng 7 năm 2020, đã đem lại nhiều cải tiến quan trọng về hiệu suất mã hóa. Đầu tiên, tiêu chuẩn này cập nhật cấu trúc khối linh hoạt và lớn hơn bằng cách tăng kích thước Đơn vị Mã hóa Cây (CTU) lên  $128 \times 128$  mẫu luma. Điều này cho phép tiêu chuẩn xử lý độ phân giải cao và các đặc tính video HDR một cách hiệu quả. Ngoài ra, nó áp dụng các kỹ thuật mã hóa được thừa hưởng từ tiêu chuẩn mở rộng mã hóa nội dung màn hình H.265/HEVC, như Intra Block Copy (IBC), block-level differential pulse code modulation (BDPCM) và Adaptive Color Transform (ACT). Tiêu chuẩn cũng giới thiệu mã hóa đa tầng, cho phép mở rộng thời gian, không gian và chất lượng, cùng với mã hóa nhiều góc nhìn. Đặc biệt, dự đoán bù trừ chuyển

Bảng 1 Tổng quan về định nghĩa dữ liệu LF

Loại lấy nét LF	Mô tả	Số lượng mẫu
CWF	Lớp phức tạp với FOV rộng và lấy nét vào <b>Tiền cảnh</b>	10
CWB	Lớp phức tạp với FOV rộng và lấy nét vào <b>Hậu cảnh</b>	10
CNF	Lớp phức tạp với FOV hẹp và lấy nét vào <b>Tiền cảnh</b>	10
CNB	Lớp phức tạp với FOV hẹp và lấy nét vào <b>Hậu cảnh</b>	10
SWF	Lớp đơn giản với FOV rộng và lấy nét vào <b>Tiền cảnh</b>	8
SWB	Lớp đơn giản với FOV rộng và lấy nét vào <b>Hậu cảnh</b>	8
SNF	Lớp đơn giản với FOV hẹp và lấy nét vào <b>Tiền cảnh</b>	10
SNB	Lớp đơn giản với FOV hẹp và lấy nét vào <b>Hậu cảnh</b>	10

Bảng 2. Cấu hình thay đổi điểm nét cho dữ liệu LF

Tên LF	LF slope	Nét hậu cảnh	Nét tiền cảnh	Nét trung tâm	Tên LF	LF slope	Nét hậu cảnh	Nét tiền cảnh	Nét trung tâm
House_Lake	SW	-	-0.3	0	Backlight_1	CW	0.4	-0.4	-
Overexposed_Sky		0.3	-0.3	-	Bench_in_Paris		0.4	-0.4	-
Paved_Road		0.2	-0.3	-	Black_Fence		-	-0.4	0
Red_White_Building		-	-0.3	0	Fountain_Vince nt_2		0.6	-0.3	-
Reeds		-	-0.3	0	Friends_3		0.6	-0.3	-
Semi_reflecting_Structure_1		-	-0.3	0	Palais_du_Luxe mbourg		0.3	-0.6	-
Semi_reflecting_Structure_2		-	-0.3	0	Parc_du_Luxe mbourg		0.3	-0.6	-
Wheat_Silos		0.3	-0.6	-	Pillars		-	-0.6	0
Chain_link_Fence_1	SN	1.6	-	0	Rolex_Learning _Center	CN	0.6	-0.3	-
					Swans_1		0.4	-0.3	-
					Bikes		0.5	-0.6	-
Color_Chart_3		-	-0.3	0	Books		1	-0.7	-
Fountain_1		0.3	-0.3	-	Bridge		1	-0.7	-
Geometric_Sculpture		-	-0.5	0	Danger_de_Mo rt		0.6	-0.3	-
ISO_Chart_16		-	-0.3	0	Mirabelle_Prune e_Tree		0.6	-0.7	-
Perforated_Metal_1		1.6	-0.5	-	Poppies		0.5	-0.7	-
Perforated_Metal_3		0.3	-0.3	-	Rusty_Fence		0.4	-0.5	-
Sewer_Drain		0.5	-0.3	-	Stone_Pillars_I nside		0.7	N/A	0
Wall_Decoration		0.7	-0.7	-	Stone_Pillars_ Outside		0.4	-0.4	-
Water_Drops		0.3	-0.3	-	University		0.3	-0.4	-

động được tính toán một cách hiệu quả để bao quanh biên độ hình ảnh, hỗ trợ các ứng dụng thực tế ảo và tăng cường cho video 360°. Hơn nữa, đối với truyền phát trực tuyến, tiêu chuẩn này cung cấp khả năng truyền phát trực tuyến với độ trễ cực thấp và điều chỉnh linh hoạt thông qua việc xử lý Gradual Decoding Refresh (GDR) tích hợp sẵn và tận dụng lợi ích của Reference Picture Resampling (RPR).

**B. Chuẩn mã hóa video H.266/VVC và dữ liệu Light Field**

Hiệu suất nén video H.266/VVC, được công bố vào tháng 7 năm 2020, vượt trội so với các tiêu chuẩn khác [22]. Điều này khiến nó trở thành một codec phổ biến để sử dụng cho các tập dữ liệu hiện đại như video có độ phân giải cao (ví dụ: 4K, 16K), video sống động (ví dụ: video 360 độ), Cloud Point và LF Image. Mặt khác, nghiên cứu

về việc nén hình ảnh LF dựa trên H.266/VVC vẫn còn rất sớm.

Các nghiên cứu ban đầu đã so sánh phương pháp nén LF [9, 11, 23] để đánh giá hiệu suất nén giữa nhiều tiêu chuẩn. Kết quả cho thấy H.266/VVC là tiêu chuẩn mã hóa tốt hơn đáng kể so với các tiêu chuẩn và phần mềm khác. Đặc biệt, việc bổ sung và cải tiến các công cụ hỗ trợ mã hóa của H.266/VVC đã làm tăng đáng kể sự hiệu quả và tính linh hoạt của nó. Nếu H.266/VVC được chuẩn bị để đáp ứng nhu cầu công nghệ mới như nén video 360 độ, thực tế ảo và xử lý hình ảnh Point Cloud, thì các công cụ này sẽ hữu ích.

Đối với chế độ mã hóa liên khung của chuẩn mã hóa video, độ tương tự giữa các SAI là một tham số quan trọng trong hiệu suất nén. Từ nghiên cứu của Liu et al. [32], chuỗi giả 4D-LF được tạo bằng cách tổ chức SAI từ cấu trúc mảng ống kính. Tuy nhiên, thứ tự mã hóa và quản lý khung tham chiếu được triển khai một cách thô sơ theo cách không thích ứng với các kịch bản cụ thể. Li và cộng sự [33] trình bày giải pháp khai thác triệt để thông tin giữa các góc nhìn khác nhau. Vì vậy có thể thấy dữ liệu ảnh Light Field và nội dung của các SAI sẽ là một trong những yếu tố có thể ảnh hưởng tới hiệu suất nén và các tham số nén.

### III. DỮ LIỆU LIGHT FIELD VÀ CÁC CÔNG CỤ MÃ HÓA H.266/VVC

#### A. Dữ liệu ảnh Light Field

Bộ dữ liệu ảnh Light Field (LF) được lựa chọn từ nguồn dữ liệu EPFL [34]. Nó bao gồm các loại khác nhau về Chiều Sâu Trường (DOF) và Trường Nhìn (FOV) được tạo ra để hiển thị cho người quan sát. Bảng 1 tổng quan về dữ liệu này và giải thích lý do lựa chọn, cùng với các cấu hình thay đổi điểm nét tại Bảng 2. Dữ liệu này được chia thành hai phần dựa trên độ phức tạp của FOV LF, cụ thể là FOV lớp phức tạp và FOV lớp đơn giản (đơn giản hay phức tạp dựa trên số lượng vật thể trong cảnh mà chúng ta có thể tập trung và xác định một cách dễ dàng; nếu có hơn hai vật thể thì được coi là lớp phức tạp). Mỗi loại FOV LF chứa 2 loại DOF LF: DOF hẹp và DOF rộng (hẹp và rộng tương ứng với phạm vi chiều sâu được tập trung trong cảnh). Để thay đổi tiêu cự của ảnh LF, chúng tôi điều chỉnh thông số đường nghiêng của ảnh LF bằng cách sử dụng Matlab LF toolbox v.0.4 [35].

Dựa vào định nghĩa bộ dữ liệu LF trong Bảng 1, bộ dữ liệu có thể được phân loại theo ba điều kiện phức tạp cảnh: độ phức tạp (đơn giản và phức tạp), độ sâu (hẹp và rộng), và tiêu cự (phần trước và phần sau). Trong nghiên cứu này, bộ dữ liệu LF được chọn từ nhiều nội dung hình ảnh LF khác nhau để thực hiện đánh giá nén toàn diện. Bộ dữ liệu LF này được tạo ra bằng phương pháp không can thiệp sử dụng chuỗi video giả tạo (PVS) dựa trên ảnh SAI và phương pháp hình ảnh được tiêu cự lại (RI). Kỹ thuật PVS sử dụng 193 ảnh SAI của hình ảnh LF với thứ tự quét xoắn ốc với tốc độ 30 khung hình mỗi giây (fps). Ngoài ra, các PVS này cũng được sử dụng với các tiêu điểm tiêu cự khác nhau, và các phạm vi độ sâu dựa trên các kỹ thuật RI.

Một lệnh chạy thử nghiệm đã được viết trên Matlab kết hợp với công cụ trình chiếu MPV player, các ảnh SAI sẽ được thể hiện song song với các độ nét và độ lấy nét khác nhau. Từ đó, kết quả của người quan sát sẽ được thu thập thông qua điểm số bình phương thấp nhất (Mean Opinion Square – MOS). Điểm MOS này được sử dụng để xác định

các thay đổi về nội dung của ảnh LF có ảnh hưởng ra sao tới người dùng, dựa trên đó, đưa ra phân tích sự thay đổi của công cụ mã hóa với bộ mã hóa H.266/VVC.

#### B. Các công cụ mã hóa trên bộ mã hóa H.266/VVC

- Nhóm Dự đoán hình ảnh liên khung (inter-picture)
  - Vectơ chuyển động theo thời gian dựa trên khối con (SBTMVP): TMVP ở chế độ hợp nhất kế thừa một tập hợp thông tin chuyển động được phân vùng theo thời gian từ CU. Trong VVC, phương pháp SBTMVP cho phép kế thừa thông tin chuyển động từ hình ảnh được chỉnh màu theo đơn vị khối con  $8 \times 8$ .
  - MVD đối xứng (SMVD): Khi chuyển động của khối hiện tại nằm trên một quỹ đạo chuyển động không đổi giữa quá khứ tạm thời và hình ảnh tham chiếu tạm thời trong tương lai theo thứ tự hiển thị, các MV tương ứng và hình ảnh tham chiếu của các chỉ số có xu hướng đối. Điều này được SMVD sử dụng để báo hiệu chỉ mục hình ảnh tham chiếu và lưu các bit cho MVD.
  - Chế độ phân vùng hình học (GPM): là một biến thể của chế độ hợp nhất trong VVC, cho phép bù chuyển động trên các phân vùng không hình chữ nhật của khối. Khi chế độ này được sử dụng, CU được chia thành hai khu vực hình học nằm trên đường thẳng. Hai chỉ số hợp nhất, một chỉ số cho mỗi khu vực, cũng được hiển thị.
  - Dự đoán nội bộ/liên ảnh kết hợp (CIIP): Khi CU được mã hóa ở chế độ hợp nhất trong VVC, một tín hiệu bổ sung được truyền đạt để xác định liệu một chế độ CIIP có được áp dụng cho CU hiện tại hay không. Khi CU có chiều rộng và chiều cao nhỏ hơn 128 mẫu luma, chế độ CIIP có thể được áp dụng.
- Nhóm dự đoán hình ảnh liên khung (inter-picture)
  - Nhiều đường tham chiếu (MRL): để dự đoán trong ảnh, vùng lân cận không gian gần nhất các mẫu được tái tạo sẽ có nhiều đường tham chiếu hơn. Trong chế độ này, có thể sử dụng mẫu từ hai dòng - một dòng tham chiếu hai dòng cách xa và một dòng tham chiếu ba dòng cách đó - thay vì sử dụng dòng gần nhất của mẫu làm dòng tham chiếu cho dự đoán trong ảnh.
  - Dự đoán không gian ảnh dựa trên ma trận (MIP): một chế độ dự đoán mới đã được thêm vào VVC. Nó ban đầu được đề xuất như một dự đoán dựa trên mạng nơ-ron, nhưng sau đó nó được đơn giản hóa để sử dụng phép nhân ma trận và lập chỉ mục bảng ma trận.
  - Phân vùng phụ nội bộ (ISP): chế độ chia một Luma CU theo chiều dọc hoặc ngang thành hai hoặc bốn phân vùng con tùy thuộc vào kích thước khối. Chế độ này mã hóa tất cả các tiêu đề con,

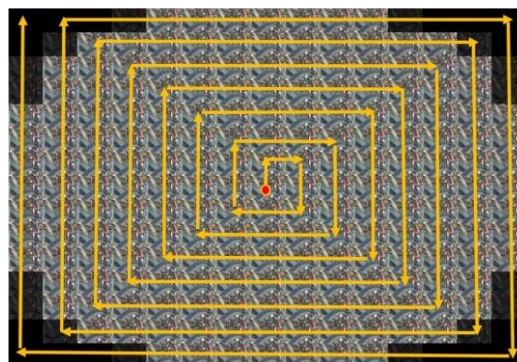
nhưng dự đoán và biến đổi được làm riêng. ISP có thể có kích thước khối tối thiểu là  $4 \times 8$  hoặc  $8 \times 4$  và kích thước khối tối đa là  $64 \times 64$ .

- Nhóm biến đổi và lượng tử hóa
  - Lựa chọn nhiều biến đổi (MTS): được sử dụng để mã hóa dư cho cả mã hóa giữa ảnh và nội ảnh của các khối. Nó cung cấp khả năng lựa chọn trong số một tập hợp các phép biến đổi hình sin được tăng cường, bao gồm Biến đổi DCT loại II, DST loại VII và DCT loại VIII cho CU nhỏ hơn hoặc bằng 32.
  - Lượng tử hóa phụ thuộc (DQ): đây là một phương pháp trong đó tập hợp các giá trị tái tạo có sẵn cho một hệ số biến đổi phụ thuộc vào các giá trị tái tạo đã được chọn cho các hệ số biến đổi đứng trước nó theo thứ tự quét. Hiệu quả chính của phương pháp này là có thể giảm độ méo trung bình giữa vectơ đầu vào và vectơ tái cấu trúc gần nhất. Vectơ đầu vào được đưa vào không gian vectơ M chiều, bao gồm tất cả các hệ số biến đổi trong một TB.
  - Giải mã dư thừa sắc độ chung (JCCR): được sử dụng để giảm thêm sự dư thừa của hai sắc độ tín hiệu dư của các thành phần khi chúng tương tự nhau. Một trong ba chế độ JCCR với các kết hợp trọng số khác nhau của một mã đơn dư sắc độ có thể được áp dụng một cách chọn lọc ở mức CU, thay vì báo hiệu phần dư cho hai thành phần sắc độ khác nhau.
- Nhóm mã hóa nội dung màn hình
  - Intra-picture block copy (IBC): sử dụng các mẫu lặp lại bên trong một bức tranh. Nó có thể được coi là một cơ bản dạng dự đoán bù chuyển động với vectơ khối (số nguyên MV) tham chiếu đến các khu vực được mã hóa trước đó của cùng một hình ảnh (thay vì các hình ảnh tham chiếu được mã hóa trước đó). IBC trong VVC đơn giản hơn về bộ đệm mẫu tham chiếu so với các phần mở rộng HEVC SCC.
  - Chế độ Bảng màu (PM): Các giá trị mẫu trong CU được hiển thị bằng các giá trị màu đại diện trong chế độ bảng màu (PM). Tập hợp màu này có tên là bảng màu. Đối với CU được mã hóa trong chế độ bảng màu, bảng màu đầu tiên được hiển thị và sau đó một chỉ số bảng màu được hiển thị cho mỗi mẫu CU. Trong VVC, đối với các lát có cây mã hóa luma và chroma khác nhau, bảng màu được sử dụng cho cả luma (thành phần Y) và chroma (thành phần Cb và Cr). Bảng màu luma mục nhập chỉ chứa giá trị Y, trong khi bảng màu chroma mục nhập chỉ chứa giá trị Cb và Cr.

## IV. KẾT QUẢ VÀ KIỂM THỬ

### A. Phương pháp kiểm thử

Chuỗi giả 4D-LF với thứ tự quét xoắn ốc được tạo ra từ những hình ảnh LF này. Chỉ có 193 khung hình được chọn chủ yếu từ trung tâm, ngoại trừ các SAI từ các góc xa hơn, để giảm hiệu ứng làm mờ nét ảnh. Các chuỗi giả thử nghiệm được chọn ở độ phân giải  $624 \times 432$ , tốc độ khung hình 30 Hz, định dạng yuv và kiểu nội dung LF phức tạp và đơn giản. Chuỗi dữ liệu LF sẽ được sử dụng để kiểm thử như hình 4.



Hình 4. Phương pháp quét xoắn ốc cho chuỗi giả 4D-LF

Dựa trên sự tương quan cao của các SAI và các loại LF, nghĩa là các lớp phức tạp và đơn giản, các công cụ mã hóa của VVC codec được đánh giá bật và tắt tương ứng để xác định các công cụ mã hóa không được sử dụng. Các công cụ mã hóa VVC codec được đánh giá cao nhất vì chúng đòi hỏi lượng bitrate lớn nhưng thu được ít màn biểu diễn do sự tương quan cao của SAI và các loại LF, nghĩa là các lớp phức tạp và đơn giản. Nhưng hình ảnh LF có mối tương quan cao giữa tất cả các SAI, vì vậy thử nghiệm đầu tiên cần được xem xét như sau:

- Mỗi thử nghiệm chạy năm lần liên tục với các cấu hình giống nhau (lưu ý: cần đóng tất cả phần mềm nền) để xác định độ phức tạp về thời gian.
- Mỗi công cụ mã hóa cần chạy hai lần, một cho TẮT và một cho BẬT, để xác định công bằng

### B. Kết quả so sánh với các công cụ hỗ trợ mã hóa

Hiệu suất thông qua BD-PSNR của các công cụ mã hóa được thể hiện cho dữ liệu ảnh LF trong Bảng 3, 4 và Hình 6. Kết quả BD-PSNR cao hơn cho thấy hiệu suất cao hơn. Bộ mã hóa VVC thường được mã hóa nếu tất cả các công cụ hỗ trợ đều được bật. Do đó, tác giả sẽ thử nghiệm so sánh điều kiện khi các công cụ mã hóa được tắt. Bảng 3 cho thấy các công cụ nào có chỉ số BD-PSNR tốt sẽ làm giảm hiệu năng của bộ mã hóa VVC khi chúng được bật. Tuy nhiên, Bảng 3 cho thấy phần lớn các công cụ mã hóa đều có chỉ số âm, ngoại trừ hai công cụ có giá trị dương là CIIP và MIP. Công cụ MTS có giá trị cao nhất là 0,04 dB, nhưng chỉ được sử dụng cho một dạng dữ liệu LF. Hiệu suất thay đổi của các giá trị này không đáng kể và rất nhỏ. Do đó, có thể thấy rằng hiệu năng mã hóa của chính nó



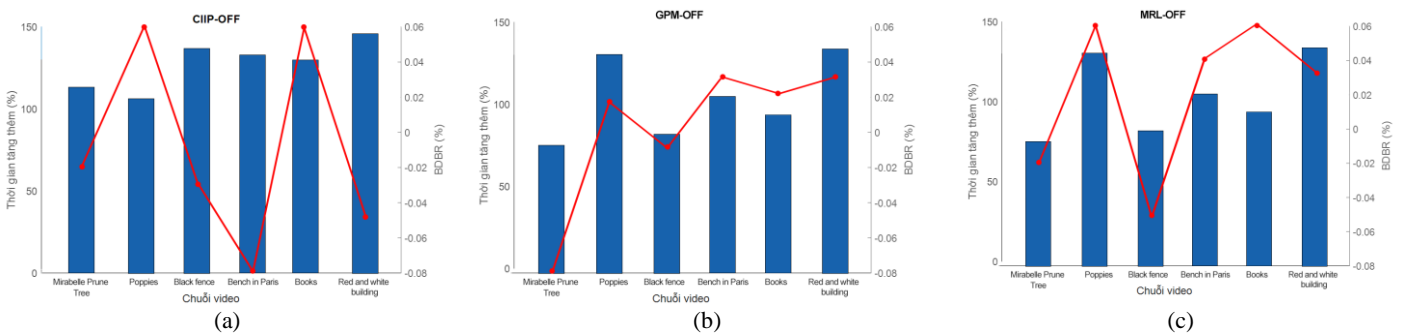
Hình 5. Các dữ liệu LF sử dụng kiểm thử: (a) Mirabelle Prune Tree, (b) Bench in Paris, (c) Poppies, và (d) Black fence, (e) Books, (f) Red and white building

Bảng 3. Kết quả so sánh BD-PSNR của các công cụ hỗ trợ mã hóa khi tắt và bật

Dữ liệu LF	Loại LF	BD-PSNR với các công cụ hỗ trợ mã hóa tắt (OFF)											
		GPM	MRL	SBTMVP	SMVD	CIIP	DQ	JCCR	MTS	PM	IBC (on)	ISP	MIP
Mirabelle Prune Tree	CN	N/A	0.02	-0.01	0.01	0.03	-0.01	-0.11	N/A	-0.02	N/A	-0.02	N/A
Poppies	CN	-0.01	-0.02	0.05	-0.01	0.02	-0.06	-0.11	N/A	-0.01	0.02	-0.02	0.01
Black fence	CW	N/A	N/A	N/A	N/A	0.04	-0.01	-0.12	N/A	-0.02	N/A	0.01	N/A
Bench in Paris	CW	-0.01	-0.03	0.06	-0.02	0.02	-0.04	N/A	-0.03	-0.01	N/A	N/A	0.01
Books	CN	-0.01	-0.02	N/A	-0.04	0.01	-0.06	-0.08	N/A	-0.01	0.02	N/A	0.01
Red and white building	SW	N/A	N/A	0.01	N/A	N/A	N/A	-0.06	0.04	0.01	-0.01	N/A	0.02

Bảng 4. Kết quả so sánh BD-Rate của các công cụ hỗ trợ mã hóa khi tắt và bật

Dữ liệu LF	Loại LF	BD-Rate với các công cụ hỗ trợ mã hóa tắt (OFF)											
		GPM	MRL	SBTMVP	SMVD	CIIP	DQ	JCCR	MTS	PM	IBC (on)	ISP	MIP
Mirabelle Prune Tree	CN	-0.08	-0.02	-0.03	N/A	0.01	N/A	0.22	-0.14	-0.06	0.04	0.02	-0.09
Poppies	CN	0.02	0.06	N/A	-0.04	0.06	0.23	0.49	-0.01	0.03	-0.05	0.02	-0.05
Black fence	CW	-0.01	-0.05	N/A	N/A	-0.03	-0.02	0.56	N/A	0.11	-0.02	0.03	0.05
Bench in Paris	CW	0.03	0.04	0.02	-0.07	-0.08	0.15	N/A	0.12	0.06	-0.01	N/A	0.06
Books	CN	0.02	0.06	N/A	-0.04	0.06	0.23	0.49	-0.01	0.03	-0.05	0.02	-0.05
Red and white building	SW	0.03	0.03	0.02	-0.07	-0.05	0.15	N/A	0.12	0.06	-0.01	N/A	0.06



không bị ảnh hưởng bởi việc tắt các công cụ hỗ trợ mã hóa của chuẩn VVC.

Dựa trên kết quả cho thấy, khi tắt mở độc lập các công cụ mã hóa của bộ mã hóa H.266/VVC thì hiệu năng không thay đổi đáng kể. Tuy nhiên, một số công cụ mã hóa như CIIP và MIP đã có sự thay đổi, từ đó chứng minh rằng nếu kết hợp các công cụ mã hóa theo cặp hoặc theo nhóm sẽ có khả năng cải thiện và thay đổi hiệu suất của bộ mã hóa tiên tiến này, đặc biệt cho bộ dữ liệu ảnh LF.

## V. KẾT LUẬN

Một nghiên cứu toàn diện về các công cụ hỗ trợ mã hóa trong bộ mã hóa H.266/VVC được cung cấp trong bài báo này. Kết quả cho thấy hiệu năng và sự tiết kiệm bitrate của các công cụ hỗ trợ mã hóa không thay đổi nhiều khi tắt mã hóa. Điều này cho thấy rằng sự phức tạp của bộ mã hóa nằm ở một nơi khác và cần phải được nghiên cứu rõ hơn và rộng hơn. Đặc biệt, một điều đáng chú ý là nội dung của dữ liệu ảnh LF ảnh hưởng đáng kể đến các công cụ mã hóa.

## TÀI LIỆU THAM KHẢO

1. N. Bedard, T. Shope, A. Hoberman, M. A. Haralam, N. Shaikh, J. Kovačević, N. Balram, and I. Tošić, "Light field otoscope design for 3D in vivo imaging of the middle ear," *Biomedical optics express*, vol. 8, pp. 260-272, 2017.
2. H. Li, C. Guo, and S. Jia, "High-resolution light-field microscopy," *Frontiers in Optics*, vol. FW6D. 3, 2017.
3. D. Tsai, D. G. Dansereau, T. Peynot, and P. Corke, "Image-Based Visual Servoing With Light Field Cameras," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 912-919, 2017.
4. K. Lynch, T. Fahringer, and B. Thurow, "Three-dimensional particle image velocimetry using a plenoptic camera," in *American Institute of Aeronautics and Astronautics (AIAA)*, 2012.
5. R. Ng, et al., "Light field photography with a hand-held plenoptic camera," in *Computer Science Technical Report*, vol. 2, no. 11, pp. 111, 2005.
6. G. Wu et al, "Light Field Image Processing: An Overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 926-954, 2017.
7. B. Benjamin, C. Jianle, L. Shan, W. Ye-Kui, "Versatile Video Coding (Draft 10)," 19th Meeting: by teleconference, Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, 2020
8. 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
9. V. Avramelos, J. D. Praeter, G. V. Wallendael, and P. Lambert, "Light Field image compression using Versatile Video Coding," in *IEEE 9th International Conference on Consumer Electronics (ICCE)*, Berlin, 2019
10. T. Wiegand, G. J. Sullivan, G. Bjøntegaard and A. Luthra, "Overview of the H.264/AVC Video Coding Standard", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, issue 7, pp. 560-576, 2003
11. H. Amirpour, A. M. G. Pinheiro, M. Pereira, M. Ghanbari, "Performance comparison of video encoders in light field image compression", in *Proc. IS&T Int'l. Symp. on Electronic Imaging: 3D Imaging and Applications*, pp 60-1 - 60-7, 2021, <https://doi.org/10.2352/ISSN.2470-1173.2021.18.3DIA-060>
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. J. Han et al., "A Technical Overview of AV1," in *Proceedings of the IEEE*, vol. 109, no. 9, pp. 1435-1462, Sept. 2021, doi: 10.1109/JPROC.2021.3058584.
14. J. Samuelsson, "The XVC Video Code: A Revolutionary Software-Defined Video Compression Format," in *SMPTE Motion Imaging Journal*, vol. 128, no. 10, pp. 1-8, Nov.-Dec. 2019, doi: 10.5594/JMI.2019.2937737.
15. S. Umebayashi, K. Kodama and T. Hamamoto, "A Study on 4D Light Field Compression Using Multi-focus Images and Reference Views," *2021 International Conference on Visual Communications and Image Processing (VCIP)*, pp. 1-5, 2021, doi: 10.1109/VCIP53242.2021.9675378.
16. N. Bakir, W. Hamidouche, S. A. Fezza, K. Samrouth and O. Déforges, "Light Field Image Coding Using Dual Discriminator Generative Adversarial Network and VVC Temporal Scalability," *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1-6, 2020, doi: 10.1109/ICME46284.2020.9102880.
17. F. Pakdaman, M. A. Adelimanesh, and M. R. Hashemi. "BLINC: Lightweight Bimodal Learning for Low-Complexity VVC Intra Coding." *arXiv preprint arXiv:2201.07823*, 2022.
18. F. Pakdaman, M.A. Adelimanesh, M. Gabbouj, M.R. Hashemi, "Complexity Analysis Of Next- Generation VVC Encoding and Decoding". in: *IEEE International Conference on Image Processing (ICIP)*. pp. 3134–3138, 2020
19. F. Bossen, J. Boyce, K. Suehring, X. Li, V. Seregin, "JVET common test conditions and software reference configurations for SDR video." *Jt. Video Expert. Team ITU-T SG 16 WP 3 ISO/IEC JTC 1/SC 29/WG 11, 14th Meet*, 2019.
20. S. Vetrivel, K. Suba, and G. Athisha, "An Overview of H.26x Series and Its Applications," *International Journal of Engineering Science and Technology*, vol. 2, no. 9, pp. 4622-4631, 2010.
21. D. Marpe, T. Wiegand, and G.J. Sullivan, "The H.264/MPEG4 advanced video coding standard and its applications", *IEEE Communications Magazine*, vol. 44, no. 8, pp. 134–143, 2006.
22. J. Ostermann, et al. "Video coding with H.264/AVC: tools, performance, and complexity." *IEEE Circuits and Systems Magazine*, vol. 4, pp. 7-28, 2004.
23. High Efficiency Video Coding, *Rec. ITU-T H.265 and ISO/IEC 23008-2*, Jan. 2013.
24. B. Bross, J. Chen, J. -R. Ohm, G. J. Sullivan and Y. -K. Wang, "Developments in International Video Coding Standardization After AVC, With an Overview of Versatile Video Coding (VVC)," in *Proceedings of the IEEE*, vol. 109, no. 9, pp. 1463-1493, Sept. 2021, doi: 10.1109/JPROC.2020.3043399.
25. H. P. Cong, S. Perry, V. T. Anh, and X. H. Van, "Joint Exploration Model based Light Field Image Coding: A



- Comparative Study”, In: 2017 4th NAFOSTED Conference on Information and Computer Science (NICS 2017), Hanoi, Vietnam, Nov. 2017.
26. T. Nguyen, D. Marpe, “Future video coding technologies: a performance evaluation of av1, jem, vp9, and hm”, In: 2018 picture coding symposium (PCS), pp 31–35, 2018.
  27. A. S. Panayides, M. S. Pattichis, M. Pantziaris, A. G. Constantinides, C. S. Pattichis, “The battle of the video codecs in the healthcare domain - a comparative performance evaluation study leveraging VVC and AV1”, IEEE Access, vol. 8, pp. 11469–11481, 2020.
  28. F. Zhang, A.V. Katsenou, M. Afonso, G. Dimitrov, D.R. Bull, “Comparing vvc, hevc and av1 using objective and subjective assessments”, arXiv:2003.10282 (eess. IV), 2020.
  29. D. Barina, M. Solony, T. Chlubna, D. Dlabaja, O. Klima, P. Zemcik, “Comparison of light field compression methods”, Multimedia Tools and Applications, vol. 81, pp. 2517 – 2528, 2022.
  30. V. V. Duong, T. N. Huu, L. Jealin, and J. Byeungwoo, “Light Field image compression using Versatile Video Coding Intra Prediction”, Proceedings of the Korean Society of Broadcast Engineers Conference, pp. 222 – 224, 2019.
  31. T. N. Huu, V. V. Duong, M. Xu, and J. Byeungwoo, “Analysis of Affine Motion Compensation for Light Field Image Compression”, Proceedings of the Korean Society of Broadcast Engineers Conference, pp. 216 – 217, 2019.
  32. D. Liu et al, "Pseudo-sequence-based light field image compression," IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Seattle, WA, 2016.
  33. L. Li et al, "Pseudo Sequence Based 2-D Hierarchical Coding Structure for Light-Field Image Compression," 2017 Data Compression Conference (DCC), Snowbird, UT, 2017.
  34. M. Řeřábek, T. Ebrahimi, “New Light Field Image Dataset”. In Proceedings of the 8th International Conference on Quality of Multimedia Experience, Lisbon, Portugal, 6 - Jun. 2016.
  35. D. Dansereau, “Light Field Toolbox for Matlab,” Feb. 2015. Available online: <http://www.mathworks.com/matlabcentral/fileexchange/49683-light-field-toolbox-v0-4> (accessed on 12/01/2021).



**Phí Công Huy** tốt nghiệp đại học Oxford Brookes, UK; lấy bằng thạc sĩ tại Monash, Úc, và đang làm nghiên cứu sinh trong chương trình hợp tác giữa đại học quốc gia Hà nội và đại học công nghệ Sydney. 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, kiểm soát chất lượng dữ liệu ảnh Light Field.

Email: huygc@ptit.edu.vn



**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 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.

Email: chaunc@ptit.edu.vn

## VERSATILE VIDEO CODING BASED LIGHT FIELD IMAGE COMPRESSION: A COMPREHENSIVE STUDY ON CODING TOOLS

**Abstract:** Video coding standards have been developed to efficiently transmit and store digital video with a variety of requirements on bitrate, picture quality, delay, random accessibility, complexity, and so on. Meanwhile, the digital data is also explored to meet the user’s need for the viewing experience, including the Light Field, Point Cloud, and Holography. Light Field (LF) imaging is a plenoptic data collection method enabling a wide variety of image post-processing such as 3D extraction, viewpoint change, and digital refocusing. Moreover, LF provides the capability to capture rich information about a scene e.g., texture, geometric information, etc. Therefore, efficient compression plays a significant role in Light Field imaging due to the huge amount of data needed for their representation. The state-of-the-art video coding standards have taken into account the characteristics of common videos to improve the compression efficiency such as H.264/AVC, H.265/HEVC, and H.266/VVC. Currently, the latest video coding standard – H.266/VVC provides outstanding compression performance comparing to other standards, but it comes at price of extremely high computational complexity due to several innovation coding tools added (e.g., triangle partition mode (TPM), the large  $128 \times 128$  pixels Coding Tree Unit (CTU), etc.). However, there is not yet have an extensive analysis of coding tools for this codec to take the most advantage of its performance efficiency, especially for LF images. Hence, this paper presents a comprehensive study on coding tools of the up-to-date video coding standard – H.266/VVC for LF images. Most of the coding tools from H.266/VVC (e.g., SMVD, MIP, IBC, etc) are evaluated on 4D-LF pseudo-sequences, then the compression and complexity performance are reported.

**Keywords:** Light Field Coding, Light Field Imaging, Video coding standards, Coding tools.