

NGHIÊN CỨU CÁC PHƯƠNG PHÁP ĐÁNH GIÁ KHÁCH QUAN VỚI DỮ LIỆU ẢNH LIGHT FIELD

Phí Công Huy^{*,#}, Nguyễn Cảnh Châu^{*}

^{*}Khoa Đa phương tiện, Học Viện Công Nghệ Bưu Chính Viễn Thông

[#]Khoa Điện tử Viễn thông, Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội

Tóm tắt—Hình ảnh Light Field (LF) là một công nghệ mới cho phép chụp và mô phỏng thông tin hình ảnh ở nhiều khoảng cách tiêu cự và các vị trí của ảnh. Bằng cách thu thập bức xạ của các tia sáng từ các hướng khác nhau, LF có thể cung cấp độ sâu và góc nhìn để đưa cho người dùng trải nghiệm hình ảnh phong phú hơn mà nhiếp ảnh truyền thống không thể có. Vì vậy, việc đánh giá chất lượng hình ảnh LF là một nhu cầu thiết yếu trong nhiều lĩnh vực. Nhiều nghiên cứu đánh giá chất lượng ảnh LF đã được giới thiệu gần đây, nhưng chưa có một báo cáo tổng quát để so sánh các phương pháp này. Trong bài báo này, chúng tôi đưa ra một nghiên cứu tổng quát về phương pháp đánh giá chất lượng ảnh khách quan dựa trên dữ liệu LF. Dữ liệu ảnh LF được sử dụng để đánh giá bao gồm 228 hình ảnh từ nguồn EPFL với hai mức lấy nét khác nhau (lấy nét tiền cảnh và lấy nét hậu cảnh). Kết quả được so sánh với điểm số từ phương pháp đánh giá chủ quan, nhằm so sánh sự tương quan giữa các phương pháp đánh giá khách quan với MOS. Một tập hợp các phương pháp đánh giá khách quan phổ biến, như: PSNR, SSIM, MS-SSIM, IW-PSNR, IW-SSIM, IW-MSE, GMSD, VMAF, IFC, JQS, VIF, và FSIM, được kiểm thử và đưa ra kết luận cụ thể trong bài báo này.

Keywords—Mã hóa dữ liệu Light Field, chất lượng ảnh Light Field, phương pháp đánh giá khách quan.

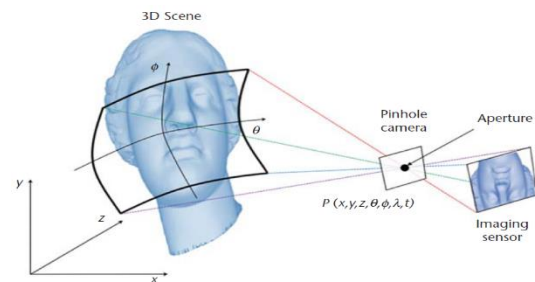
I. GIỚI THIỆU CHUNG

Với nhu cầu ngày càng tăng đối với các ứng dụng dựa trên hình ảnh, việc đánh giá chất lượng hình ảnh hiệu quả và đáng tin cậy ngày càng trở nên quan trọng. Đo lường chất lượng hình ảnh có tầm quan trọng cơ bản đối với nhiều ứng dụng xử lý hình ảnh, đó là mục tiêu của phương pháp đánh giá chất lượng hình ảnh (IQA) mong muốn đồng nhất với thị giác của mắt người với các phương pháp IQA. Nhiều phương pháp IQA đã được đề xuất trong những năm qua để thực hiện mục tiêu này. Phần lớn các đề xuất tập trung vào các phương pháp đánh giá chất lượng hình ảnh tham chiếu đầy đủ (FR-IQA) và 9 phương pháp đánh giá chất lượng thường được sử dụng (bao gồm cả sai số trung bình bình phương (MSE), chỉ số tương tự về cấu trúc

(SSIM), chỉ số tương tự về cấu trúc đa thang đo (MS-MQI)).

Biểu diễn ảnh LF được thể hiện thông qua dữ liệu hình ảnh bao gồm các thông tin [1] như không gian (x, y, z) , thời gian (t) , vị trí (θ, ϕ) bước sóng của ánh sáng (λ) . Nó được tổng hợp chung trong 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.

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.



Hình 1. Các loại thông tin thể hiện cho dữ liệu Light Field

Dữ liệu hình ảnh LF có dung lượng khá lớn bởi cấu trúc của nó, một bức ảnh LF có dung lượng khoảng 50MB. Để giải quyết vấn đề này, vấn đề mã hóa LF đã thu hút được sự chú ý gần đây. Các kỹ thuật xử lý khác nhau chẳng hạn như nén dữ liệu LF và các phương pháp tái tạo chế độ xem đều được sử dụng nhằm giải quyết nhu cầu lưu trữ. Nhóm nghiên cứu chuẩn hóa JPEG đã đưa ra mục tiêu nghiên cứu về nén 4D LF vào năm 2017 [2] và chuẩn hóa JPEG Pleno đã được giới thiệu, trong đó có dữ liệu ảnh LF [3].

Trong khi các phương pháp đánh giá chất lượng khách quan cho dữ liệu LF vẫn đang phát triển [4-7], điều quan trọng là phải kiểm tra hiệu năng của các phương pháp đánh giá, từ đó có thể thiết lập thước đo tốt nhất để sử dụng rộng rãi cho dữ liệu LF. Nhiều phương pháp dự đoán chất lượng hình ảnh và video đã được đề xuất gần đây để áp dụng đánh giá chất lượng LF. Độ chính xác của các phương pháp này thường được đánh giá chỉ trên một tập dữ liệu duy nhất với một thang điểm chủ quan cụ thể như trong các đề xuất tại [8-10]. Tuy nhiên, hiệu suất của các chỉ số có thể khác nhau qua nhiều tập dữ liệu với các kiểu biến dạng và cách tính

Tác giả liên hệ: Phí Công Huy,

Email: huyphc@ptit.edu.vn

Đến tòa soạn: 16/9/2022, chỉnh sửa: 22/10/2022, chấp nhận đăng: 27/10/2022.

điểm khác nhau.

Trong bài báo này, mục tiêu là nghiên cứu các phương pháp đánh giá khách quan phổ biến nhất hiện nay dựa trên bộ dữ liệu thực nghiệm chất lượng chủ quan ảnh LF, từ đó đánh giá sự tương quan (hiệu quả) của các phương pháp đánh giá khách quan trên bộ dữ liệu đó.

Cấu trúc của bài báo này như sau. Phần 2 tóm lược cơ sở lý thuyết về các phương pháp đánh giá khách quan và chủ quan. Phần 3 mô tả các kỹ thuật sẽ kiểm thử và cách tạo ra dữ liệu LF. Tiếp theo, Phần 4 mô tả các kết quả và kiểm thử, cuối cùng Phần 5 đưa ra một số nhận xét kết luận và các công việc trong tương lai.

II. CƠ SỞ LÝ THUYẾT

Để có được phương pháp đánh giá khách quan chính xác, cần có một tập dữ liệu tin cậy và ổn định, bộ dữ liệu này được tổng hợp và tạo ra từ các phương pháp đánh giá chủ quan. Phương pháp này là một trong những phương pháp cần sự tham gia của con người và là phương pháp đáng tin cậy nhất để đánh giá chất lượng hình ảnh, vì người quan sát là người dùng cuối cùng trong hầu hết các ứng dụng đa phương tiện. Trong thử nghiệm chủ quan, một nhóm người được yêu cầu đưa ra ý kiến của họ về chất lượng của từng hình ảnh. Để thực hiện kiểm tra chất lượng hình ảnh chủ quan, một số tiêu chuẩn quốc tế được đề xuất [11-16] cung cấp kết quả đáng tin cậy. Các chuẩn quốc tế này được mô tả tổng quát như sau: ITU-R BT.500-11 [11] đề xuất các phương pháp khác nhau để đánh giá chất lượng chủ quan của hình ảnh truyền hình. Đây là một tiêu chuẩn được sử dụng rộng rãi, chứa thông tin về cách đánh giá, và hướng dẫn cách thực hiện các thí nghiệm chủ quan. ITU-T P.910 [12, 13] đề xuất phương pháp tiêu chuẩn để đánh giá chất lượng video kỹ thuật số với tốc độ truyền dưới 1,5 Mb/s/giây. ITU-R BT.814-1 [14] được đề xuất để đặt độ sáng và độ tương phản của thiết bị hiển thị. ITU-R BT.1129-2 [15] được đề xuất để đánh giá chất lượng của chuỗi video độ nét tiêu chuẩn (SD).

Dựa trên các chuẩn được giới thiệu, các kỹ thuật đánh giá chủ quan để tạo ra dữ liệu có thể kể đến: Kỹ thuật mô phỏng đơn/đôi, kỹ thuật dự đoán tương đồng, và kỹ thuật tính toán điểm khác biệt DMOS hoặc Z-score. Đối với kỹ thuật mô phỏng đơn, các hình ảnh thử nghiệm được hiển thị trên màn hình trong một khoảng thời gian cố định, sau đó, chúng sẽ biến mất khỏi màn hình và người quan sát sẽ được yêu cầu đánh giá chất lượng của chúng trên thang điểm được định nghĩa như sau: xuất sắc, tốt, công bằng, kém, hoặc xấu. Tất cả các hình ảnh thử nghiệm được hiển thị ngẫu nhiên. Còn đối với kỹ thuật mô phỏng đôi, nó tương tự như kỹ thuật mô phỏng đơn. Tuy nhiên, trong phương pháp này, cả hình ảnh thử nghiệm và hình ảnh tham chiếu đều được hiển thị trong một khoảng thời gian cố định. Sau đó, hình ảnh sẽ biến mất khỏi màn hình và người quan sát sẽ được yêu cầu đánh giá chất lượng của hình ảnh thử nghiệm theo thang đo đã mô tả trước đó. Ngoài ra, còn một số kỹ thuật sắp xếp thứ tự theo cặp chọn lọc, phương pháp này sử dụng hai hình ảnh của cùng một cảnh được hiển thị cho người quan sát. Sau đó, họ được yêu cầu chọn hình ảnh có chất lượng cao hơn. Người quan sát luôn được yêu cầu chọn một hình ảnh ngay cả khi cả hai hình ảnh không có sự khác biệt. Không có giới hạn thời gian cho người quan sát

để đưa ra quyết định. Hạn chế của phương pháp này là nó đòi hỏi nhiều thử nghiệm hơn để so sánh từng cặp điều kiện [16]. Trong [17, 18], hai phương pháp để giảm số lượng thử nghiệm được đề xuất.

Đối với kỹ thuật dự đoán tương đồng người quan sát được yêu cầu chọn một hình ảnh ngay cả khi họ không thấy sự khác biệt giữa cặp hình ảnh. Tuy nhiên, trong việc đánh giá sự giống nhau theo cặp, những người quan sát không chỉ được yêu cầu chọn hình ảnh có chất lượng cao hơn mà còn chỉ ra mức độ khác biệt giữa chúng trên thang đo. Người quan sát có thể bị ảnh hưởng để sử dụng các kết quả đánh giá thô như: xuất sắc, tốt, khá, v.v. cho điểm chất lượng. Tuy nhiên, những kết quả đánh giá này là không đáng tin cậy. Một lý do cho điều này là các nhà quan sát có khả năng chỉ định các thang chất lượng khác nhau cho từng cảnh và thậm chí các loại biến dạng [19 - 21]. Ở đây, đề tài giới thiệu ngắn gọn hai phương pháp tính điểm được sử dụng trong IQA chủ quan.

Từ những kỹ thuật trên, 2 thuật toán thống kê điểm từ người dùng được sử dụng đó là DMOS hoặc Z-score.

Kỹ thuật DMOS thay vì áp dụng trực tiếp kết quả xếp hạng, các chỉ số IQA hiện đại sử dụng sự khác biệt về chất lượng giữa các hình ảnh. DMOS được định nghĩa là sự khác biệt giữa điểm chất lượng thô của hình ảnh tham chiếu và thử nghiệm. DMOS được tính bằng công thức sau:

$$d_{i,j} = r_{i,ref}(i) - r_{i,j} \quad (1)$$

Trong đó, $r_{i,j}$ là điểm gốc cho đối tượng i-th và hình ảnh j-th. Đồng thời, $r_{i,ref}(i)$ biểu thị điểm gốc được cho bởi đối tượng i-th đối với ảnh tham khảo tương ứng j-th

Còn đối với kỹ thuật Z-score, để dễ dàng so sánh ý kiến của mỗi người quan sát về chất lượng hình ảnh, một phép biến đổi tuyến tính làm cho giá trị trung bình và phương sai bằng nhau cho tất cả người quan sát được sử dụng. Kết quả của phép biến đổi như vậy được gọi là điểm Z và nó có thể được tính bằng công thức sau

$$z_{i,j} = \frac{d_{i,j} - \bar{d}_i}{\sigma_i} \quad (2)$$

Trong đó DMOS là $\bar{d}_i$ và độ lệch chuẩn σ_i được tính toán thông qua tất cả các hình ảnh đã được đánh giá bởi đối tượng i-th

Các phương pháp đánh giá chất lượng chủ quan cung cấp các phép đo chính xác và đáng tin cậy về chất lượng của các tín hiệu hình ảnh. Tuy nhiên, các phương pháp này có những hạn chế khác nhau làm hạn chế các ứng dụng như: tốn thời gian và tốn kém, điều này là do thực tế là các kết quả chủ quan thu được thông qua các thí nghiệm với nhiều người quan sát; không thể được tích hợp vào các ứng dụng thời gian thực như nén hình ảnh và hệ thống truyền; kết quả của đánh giá phụ thuộc nhiều vào điều kiện thể chất và trạng thái cảm xúc của đối tượng. Hơn nữa, các yếu tố khác như thiết bị hiển thị và điều kiện ánh sáng ảnh hưởng đến kết quả của các thí nghiệm đó.

Do đó, cần thiết kể các mô hình toán học có khả năng dự đoán chất lượng cảm nhận của các tín hiệu hình ảnh một cách phù hợp với các đánh giá chủ quan, các mô hình toán học này được coi là phương pháp đánh giá khách quan. Các phương pháp đánh giá khách quan có thể chia ra làm 3 thể loại: Đánh giá chất lượng hình ảnh không tham chiếu (NR-

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 Tiền cảnh	10
CWB	Lớp phức tạp với FOV rộng và lấy nét vào Hậu cảnh	10
CNF	Lớp phức tạp với FOV hẹp và lấy nét vào Tiền cảnh	10
CNB	Lớp phức tạp với FOV hẹp và lấy nét vào Hậu cảnh	10
SWF	Lớp đơn giản với FOV rộng và lấy nét vào Tiền cảnh	8
SWB	Lớp đơn giản với FOV rộng và lấy nét vào Hậu cảnh	8
SNF	Lớp đơn giản với FOV hẹp và lấy nét vào Tiền cảnh	10
SNB	Lớp đơn giản với FOV hẹp và lấy nét vào Hậu cảnh	10

IQA), đánh giá chất lượng hình ảnh tham chiếu một phần (RR-IQA), và đánh giá chất lượng hình ảnh với tham chiếu đầy đủ (FR-IQA).

Trong nhiều ứng dụng thực tế, chẳng hạn như hệ thống liên lạc bằng hình ảnh, hình ảnh tham chiếu không có sẵn và việc đánh giá chất lượng chỉ dựa trên hình ảnh thử nghiệm. NR-IQA là một nhiệm vụ khó khăn hơn so với RR-IQA và FRIQA các phương pháp. Tuy nhiên, con người thường có thể đánh giá hiệu quả chất lượng của hình ảnh thử nghiệm mà không cần sử dụng bất kỳ hình ảnh tham chiếu nào. Điều này có lẽ là do bộ não của chúng ta lưu giữ rất nhiều thông tin về việc một hình ảnh nên hoặc không nên trông như thế nào trong thế giới thực [22]. Một số phương pháp NR-IQA có thể được giới thiệu trong [23-28].

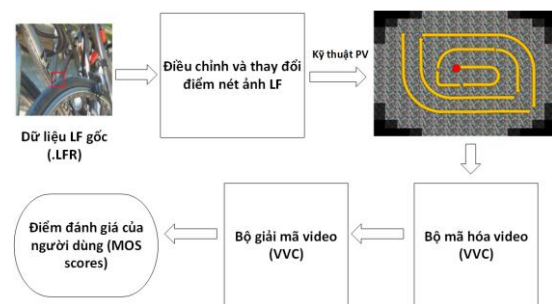
Trong RR-IQA, không thể truy cập đầy đủ hình ảnh tham chiếu. Thay vào đó, một số tính năng được trích xuất từ hình ảnh tham chiếu. Các tính năng này được sử dụng bởi phương pháp đánh giá chất lượng như là thông tin phụ để đánh giá chất lượng của hình ảnh thử nghiệm. Các phương pháp RR-IQA có thể được sử dụng trong một số ứng dụng. Ví dụ, chúng có thể được sử dụng để theo dõi mức độ suy giảm chất lượng hình ảnh của dữ liệu hình ảnh và video được truyền qua mạng giao tiếp hình ảnh thời gian thực.

Trong các thước đo FR-IQA, hình ảnh tham chiếu sẽ có chất lượng hoàn hảo cho quá trình dự đoán. Phạm vi ứng dụng của các thông tin dữ liệu này bao gồm nén hình ảnh [30], watermarking [31, 32], v.v. Trong các phần tiếp theo, đề tài sẽ mô tả toàn diện sáu phương pháp FR-IQA. Các phương pháp này là những phương pháp phổ biến và đề xuất rộng rãi trong các tài liệu, và đã được các nhà nghiên cứu đánh giá có hiệu quả tốt. Sáu chỉ số FR-IQA được mô tả trong các phần sau đây bao gồm sai số trung bình bình phương (MSE), chỉ số tương tự về cấu trúc (SSIM) [33], chỉ số tương tự về cấu trúc đa cấp độ (MS-SSIM) [34], độ trung thực của thông tin hình ảnh (VIF) [35], biến dạng rõ nhất (MAD) [36], và chỉ số tương tự tính năng (FSIM) [37]. Điều quan trọng cần lưu ý là tất cả sáu chỉ số đánh giá chất lượng này đều được thiết kế và đánh giá cho hình ảnh ở thang màu xám.

III. CÁC PHƯƠNG PHÁP ĐÁNH GIÁ KHÁCH QUAN VÀ BỘ DỮ LIỆU ẢNH LIGHT FIELD

A. Xây dựng bộ dữ liệu LF

Dữ liệu LF trong bài báo này được thu thập bằng phương pháp đánh giá chủ quan thông qua phương pháp bị động, bằng kỹ thuật tạo video giả lập (PV) [38], được chọn từ 101 SAI của hình ảnh LF thông qua cách thức quét xoắn ốc, với 30 khung hình/giây (fps). Đồng thời, kỹ thuật thay đổi điểm nét của ảnh (RI) [39] cũng được sử dụng kết hợp. Thông thường, phương pháp thể hiện hình ảnh chủ động có thể cung cấp QoE tốt hơn bằng cách trao quyền kiểm soát các điểm lấy nét cho người đánh giá. Tuy nhiên, việc không kiểm soát được số lượng điểm nhìn mà mỗi người đánh giá lựa chọn là một hạn chế của phương pháp này. Trong tập dữ liệu LF của bài báo này, người dùng không thể kiểm soát các điểm lấy nét, nhưng chúng tôi cung cấp nhiều điểm lấy nét khác nhau (từ hậu cảnh đến tiền cảnh), các điều kiện độ sâu khác nhau (từ hẹp tới rộng) và độ phức tạp khác nhau của đối tượng trong mỗi ảnh LF, nhằm thu thập tối đa các đánh giá chủ quan của người dùng. Quy trình tạo ra dữ liệu ảnh LF để phục vụ đánh giá chất lượng ảnh có thể tham khảo dựa trên Hình 2.



Hình 2. Quy trình tạo dữ liệu LF để phục vụ đánh giá chất lượng ảnh

Dữ liệu LF để đánh giá được lựa chọn từ EPFL [40] gồm 38 thể loại nội dung khác nhau, với các loại độ sâu (DOF), và khung nhìn (FOV) khác nhau, được mã hóa và giải mã bằng bộ mã hóa VVC [41]. Các định nghĩa về bộ dữ liệu được mô tả trong Bảng 1.

B. Các phương pháp đánh giá khách quan

Trong bài báo này các phương pháp đánh giá khách quan được giới thiệu và mô tả tổng quan, với mục tiêu tính toán và xác định được phương thức đánh giá của từng phương pháp. Trong giới hạn của bài báo, một số phương pháp tiêu biểu sẽ được mô tả chi tiết như: PSNR, SSIM, VIF, FSIM, và MS-SSIM.

Phương pháp đánh giá khách quan PSNR được sử dụng phổ biến nhất để đo chất lượng của việc tái tạo lại các tín hiệu từ các bộ nén dữ liệu có tổn thất (ví dụ: để nén hình ảnh). Tín hiệu trong trường hợp này là dữ liệu gốc và nhiễu là lỗi do nén. Công thức tính PSNR được mô tả như sau

$$PSNR = 10 \log_{10} \left(\frac{255^2}{MSE} \right) \quad (3)$$

Với

$$MSE(x, y) = 1/MN \sum_{i=1}^M \sum_{j=1}^N (x_{ij} - y_{ij})^2 \quad (4)$$

Ở đây, x_{ij} và y_{ij} là vector các giá trị quan sát được và giá trị dự báo, và M và N là tổng số mẫu.

Còn đối với phương pháp đánh giá khách quan SSIM, nó là một phương pháp để đo mức độ giống nhau giữa hai hình ảnh thông qua dự đoán chất lượng hình ảnh gốc và hình ảnh nhiễu. Phương pháp này được coi là phù hợp với hệ thống thị giác của con người. Phương pháp đánh giá khách quan SSIM được tính toán dựa trên các đặc tính của hình ảnh như: độ sáng, độ tương phản, và độ tương đồng. Công thức SSIM được tính như sau:

$$SSIM(x, y) = l_{(x,y)} c_{(x,y)} s_{(x,y)} \quad (5)$$

Với

$$\begin{aligned} l_{(x,y)} &= 2\mu_x\mu_y + C_1/\mu_x^2 + \mu_y^2 + C_1 \\ c_{(x,y)} &= 2\sigma_x\sigma_y + C_2/\sigma_x^2 + \sigma_y^2 + C_2 \\ s_{(x,y)} &= \sigma_{xy} + C_3/\sigma_x\sigma_y + C_3 \end{aligned} \quad (6)$$

Phương pháp đánh giá mở rộng của SSIM đó là MS-SSIM được tính toán như sau

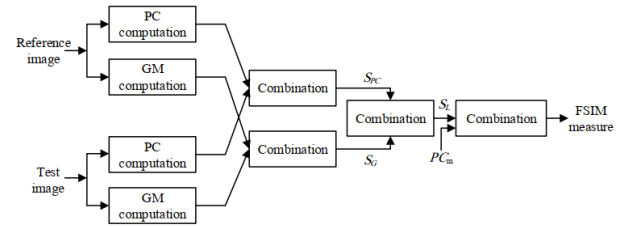
$$MSSIM(I_{ref}, I_{tst}) = \frac{1}{M_w} \sum_{i=1}^{M_w} SSIM(I_{ref}^i, I_{tst}^i) \quad (7)$$

trong đó M_w là tổng số cửa sổ cục bộ, và I_{ref}^i và I_{tst}^i là nội dung hình ảnh tại cửa sổ cục bộ thứ i.

Cũng phát triển và cải tiến từ phương pháp SSIM, tập trung vào các đặc trưng của ảnh tham chiếu và ảnh dự đoán, FSIM được tính theo công thức sau

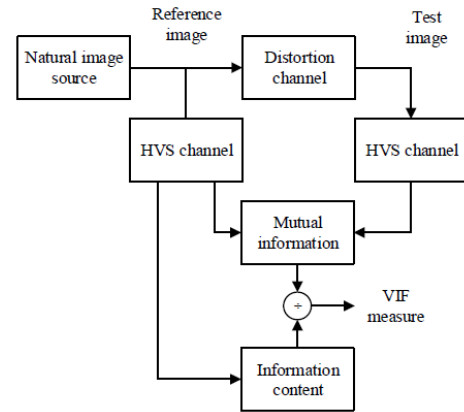
$$FSIM = \frac{\sum_{x \in \Omega} S_L(x) PC_m(x)}{\sum_{x \in \Omega} PC_m(x)} \quad (8)$$

Trong đó Ω là toàn bộ miền không gian ảnh. Sơ đồ khối của chỉ số FSIM được trình bày trong Hình 3



Hình 3. Mô hình thuật toán của phương pháp FSIM

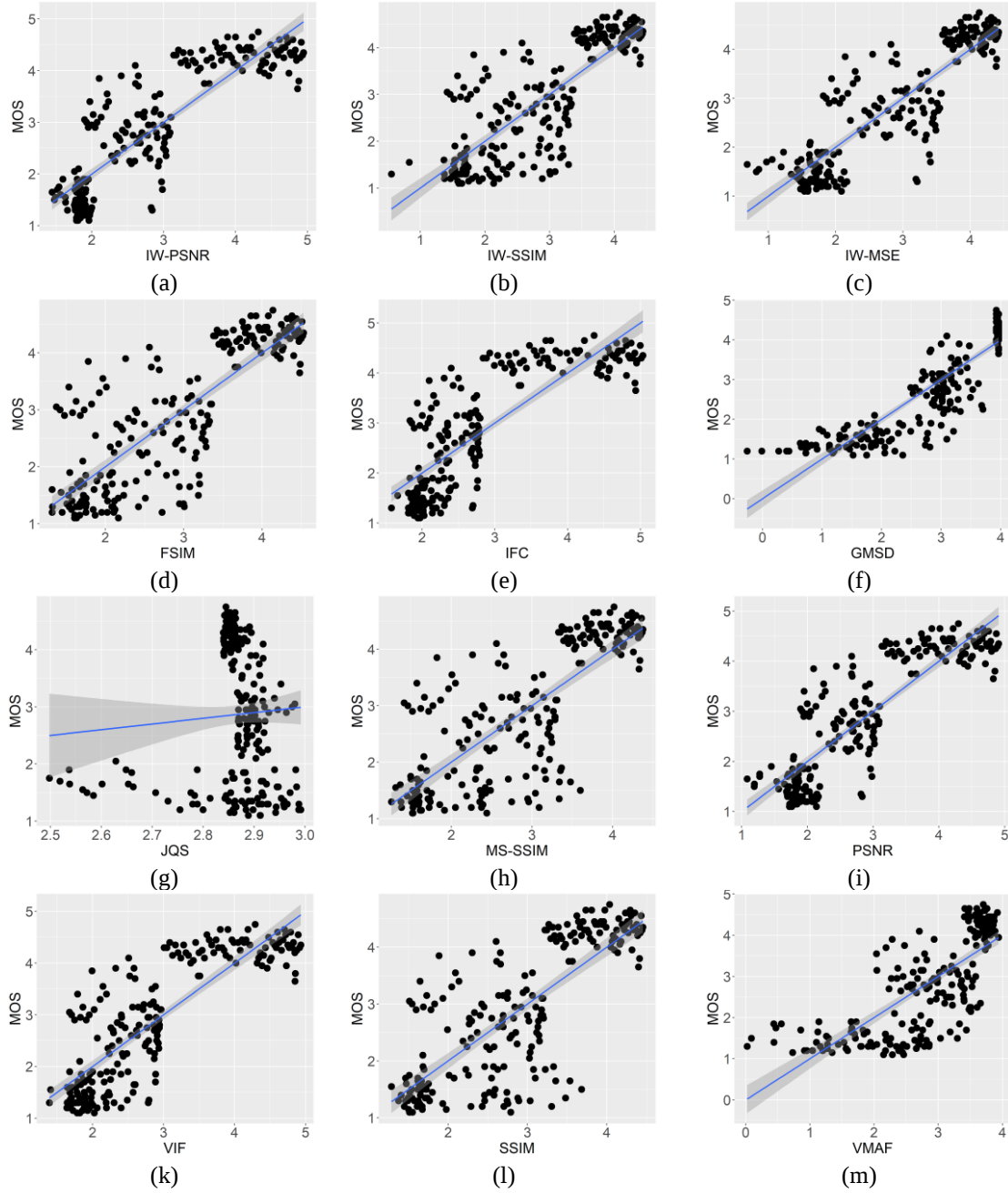
Khác biệt với các thuật toán của SSIM và PSNR, thuật toán VIF lập mô hình ảnh tự nhiên trong miền wavelet bằng cách sử dụng hỗn hợp thang đo Gaussian (GSM). Hình ảnh và video được lấy từ môi trường tự nhiên bằng cách sử dụng các thiết bị chụp chất lượng cao hoạt động trong quang phổ hình ảnh được phân loại là cảnh tự nhiên. Thuật toán VIF bao gồm ba thành phần: mô hình nguồn, mô hình biến dạng và mô hình HVS



Hình 4. Mô hình thuật toán của phương pháp VIF

BẢNG 2. SỰ TƯƠNG QUAN GIỮA CÁC PHƯƠNG PHÁP ĐÁNH GIÁ KHÁCH QUAN VÀ MOS

Metric	PSNR	SSIM	MSSSIM	FSIM	IWSSIM	IWPSNR	IFC	VIF	GMSD	JQS	VMAF	IWMSE
Kendall	0.63	0.56	0.56	0.59	0.58	0.64	0.63	0.62	0.71	-0.22	0.58	0.64
Spearman	0.85	0.76	0.77	0.80	0.79	0.86	0.83	0.83	0.89	-0.28	0.81	0.86
Pearson	0.86	0.77	0.77	0.82	0.80	0.87	0.82	0.83	0.89	0.07	0.76	0.87
MAE	0.48	0.58	0.58	0.53	0.55	0.48	0.58	0.54	0.44	1.06	0.65	0.45
RMSE	0.60	0.76	0.76	0.69	0.71	0.60	0.69	0.66	0.55	1.20	0.78	0.59
R2	0.75	0.60	0.60	0.67	0.65	0.75	0.67	0.70	0.79	0.00	0.58	0.76



Hình 5. Biểu đồ sự tương đồng giữa MOS và các phương pháp đánh giá khách quan. Các điểm chấm đen là dữ liệu của LF, đường kẻ xanh là đường fit line của mô hình các phương pháp đánh giá

Công thức tính VIF như sau

$$VIF = \frac{\sum_{j \in \text{subbands}} I(C^j; F^j | Z^j)}{\sum_{j \in \text{subbands}} I(C^j; E^j | Z^j)} \quad (9)$$

trong đó j là chỉ số phụ, và $I(C^j; F^j | Z^j)$ và $I(C^j; E^j | Z^j)$ là thông tin lẫn nhau tương ứng của tham số phụ i -th.

Các thuật toán gần đây như VMAF, IW-PSNR, IW-SSIM, IW-MSE, và GMSD được tính toán và mô tả tương ứng theo [42 - 45].

IV. KẾT QUẢ VÀ KIỂM THỬ

A. Quy trình đánh giá

Quy trình đánh giá được áp dụng theo tiêu chuẩn đánh giá thực nghiệm chủ quan từ dữ liệu ảnh LF, từ đó các điểm số chủ quan của người dùng (MOS) sẽ được tính toán và kết hợp với các phương pháp khách quan để đưa ra kết quả của sự tương quan với mỗi phương pháp. Để có được kết quả MOS, bộ dữ liệu được lấy từ Bảng 1, và thiết lập cho người dùng đánh giá giữa 2 ảnh (1 bên là ảnh gốc, 1 bên là ảnh đã qua xử lý), và đưa ra điểm số của mình. Số lượng

người tham gia đánh giá gồm 30 người có đủ các độ tuổi (20 – 40 tuổi) và chuyên môn khác nhau.

Sau khi có kết quả của các chỉ số MOS, để có được kết quả đánh giá trực quan và chính xác nhất, các phương pháp thống kê sau đây sẽ được sử dụng nhằm thể hiện hiệu năng của các mô hình đánh giá khách quan cho dữ liệu ảnh LF. Các phương pháp thống kê gồm có, hệ số tương quan của Kendall (KRCC), hệ số tương quan của Spearman (SRCC) [46], hệ số tương quan tuyến tính Pearson (PLCC) [47], sai số tuyệt đối trung bình (MAE), lỗi bình phương trung bình (RMSE) [48] và R bình phương (R2) [49]. KRCC và SRCC được sử dụng để đánh giá và dự đoán tính đơn điệu, trong khi PLCC được sử dụng để đánh giá độ chính xác của mô hình dự đoán, MAE và RMSE được sử dụng để đánh giá tính ổn định của mô hình dự đoán và R2 được sử dụng để đánh giá chất lượng tương quan của mô hình dự đoán.

Chỉ số cho các phương pháp thống kê được xác định như sau: hiệu năng được coi là tốt nếu kết quả KRCC, SRCC, PLCC, R2 có chỉ số cao và ngược lại, đối với MAE, RMSE có chỉ số thấp, phạm vi giá trị của các phương pháp thống kê được đo từ 0 đến 1.

Phương pháp thống kê được áp dụng trên 12 mô hình đánh giá khách quan bao gồm: PSNR, SSIM, MS-SSIM, FSIM, IW-SSIM, IW-PSNR, IFC, VIF, GMSD, JQS, VMAF, IW-MSE.

B. Đánh giá hiệu năng tổng thể dựa trên sự tương quan

Dựa trên Bảng 2 và Hình 5, ta thấy rằng trong tất cả các phương pháp đánh giá khách quan phổ biến hiện nay, sự tương quan cao nhất thuộc về 3 phương pháp, đó là: PSNR, IW-PSNR, và IW-MSE. Chỉ số tương đồng của 3 phương pháp này tương ứng theo thứ tự sau 75% có độ tương quan cao của phương pháp PSNR và IW-PSNR, và cao hơn chút là 76% so với MOS. Đáng chú ý, chỉ số thấp nhất lại thuộc về phương pháp đánh giá JQS và VMAF. Phương pháp JQS thể hiện không có sự tương quan nào với MOS, điều này có thể giải thích bởi phương pháp JQS chuyên thực hiện để đánh giá ảnh tĩnh, không phù hợp cho đánh giá video. Tuy nhiên, với phương pháp VMAF, là phương pháp được đánh giá tốt và có mối liên hệ với thị giác con người, lại không thể hiện sự tương đồng cao với dữ liệu LF. Với đặc thù của dữ liệu LF, có thể giả thiết rằng VMAF chỉ phù hợp với chất lượng ảnh có độ phân giải cao và nét, còn với dữ liệu ảnh thay đổi đa dạng về độ lấy nét sẽ không phù hợp. Thêm vào đó, phương pháp đánh giá VMAF được xây dựng trên mô hình huấn luyện và kết hợp của các phương pháp đánh giá khác, vì vậy đây cũng là một phần lý do làm hiệu suất tương quan của phương pháp này không cao khi so sánh với các phương pháp độc lập.

Ngoài ra, hiệu năng của 2 phương pháp đánh giá SSIM và PSNR cho kết quả thú vị, đó là với dữ liệu LF thay đổi điểm nét, thì hiệu năng của PSNR đều cho kết quả tốt hơn so với SSIM, mặc dù SSIM luôn được đánh giá là phương pháp thể hiện tốt đối với thị giác con người. Điều này có thể giải thích rằng, các đặc trưng ảnh hưởng tới thị giác mắt người trong phương pháp SSIM được thiết kế để đánh giá cho các dữ liệu của ảnh tự nhiên thông thường, chứ không phục vụ được cho dữ liệu ảnh LF. Đối với phương pháp đánh giá PSNR, kỹ thuật của nó đơn giản chỉ trừ hai ảnh để tìm điểm khác nhau, nên không bị ảnh hưởng bởi dữ liệu thiết kế cho ảnh tự nhiên thông thường. Do vậy, PSNR có thể là điểm nghiên cứu khởi đầu để phát triển các mô hình

đánh giá dữ liệu ảnh LF thay vì dựa trên phương pháp SSIM như trước đây.

Hình 5 thể hiện trực quan cho thấy sự phân bố của dữ liệu tại các phương pháp PSNR, IW-MSE rất ổn định và tuyến tính, tuy nhiên ngược lại với JQS và MS-SSIM lại có sự phân bố không đồng đều so với MOS. Hai phương pháp PSNR và IW-MSE cho thấy sự tương quan cao với chỉ số MOS, cũng một phần lý do là sự tính toán đơn giản của hai phương pháp này, đảm bảo độ chính xác cao nhất và không bị ảnh hưởng bởi các đặc trưng khác. Qua đây, có thể phân nào kết luận được phương pháp nào là lựa chọn tối ưu khi làm việc đánh giá chất lượng hình ảnh, đặc biệt là dữ liệu LF

V. KẾT LUẬN

Bài báo này trình bày một nghiên cứu tổng quát về các phương pháp đánh giá khách quan phổ biến hiện nay dựa trên dữ liệu ảnh LF, đặc biệt khả năng thay đổi điểm nét của bộ dữ liệu này. Với phương pháp xây dựng bộ dữ liệu LF đặc thù, nhằm thu thập kết quả đánh giá chủ quan đầy đủ nhất từ người dùng, 228 dữ liệu với các điểm lấy nét và các điều kiện khác nhau đã được sử dụng. Độ chính xác của các phương pháp đánh giá khách quan được so sánh với điểm số chủ quan ứng với mỗi phương pháp. Kết quả cho thấy phương pháp đánh giá khách quan phổ biến nhất là PSNR hoặc IW-MSE là những phương pháp có hiệu năng và sự chính xác ấn tượng nhất (độ tương quan 76%), đây cũng là lý do phương pháp này vẫn được dùng thông dụng trong lĩnh vực đánh giá ảnh và video. Ngoài ra, đề tài cũng chỉ ra phương pháp đánh giá VMAF không phù hợp so với dữ liệu LF, phương pháp này chỉ có thể phù hợp với các dữ liệu có độ phân giải cao hoặc nội dung đặc biệt phù hợp với streaming

Dựa trên kết quả nghiên cứu này, chúng tôi có thể phát triển thêm với các mô hình đánh giá khách quan khác nhau, nhằm đưa ra một phương pháp đánh giá khách quan tốt nhất sử dụng cho bộ dữ liệu LF, đồng thời có thể xây dựng được một mô hình dự đoán chính xác hơn trong tương lai.

VI. TÀI LIỆU THAM KHẢO

- [1] T. Ebrahimi, F. Pereira, P. Schelkens, "JPEG Pleno: Toward an Efficient Representation of Visual Reality", Industry and Standards, IEEE Multimedia, 2016
- [2] ISO/IEC JTC1/SC29/WG1, "JPEG Pleno Call for Proposals on Light Field Coding," wG1N74014, in: 74th JPEG Meeting, Geneva, Switzerland, 2017.
- [3] P. Schelkens, et al, "JPEG Pleno: a standard framework for representing and signaling plenoptic modalities", in A.G. Tescher (Ed.), Proceedings of SPIE Optics Photonics: Applications of Digital Image Processing XLI, in: Applications of Digital Image Processing XLI, vol. 10752, SPIE, United States, p. 107521P, 2018.
- [4] L. Shi, W. Zhou, Z. Chen, J. Zhang, "No-reference light field image quality assessment based on spatial-angular measurement," IEEE Trans. Circuits Syst. Video Technol. Vol. 1, 2019 <http://dx.doi.org/10.1109/tcsvt.2019.2955011>.
- [5] Z. Luo, W. Zhou, L. Shi, Z. Chen, "No-reference light field image quality assessment based on micro-lens image," in 2019 Picture Coding Symposium, PCS, pp. 1–5, 2019
- [6] Y. Tian, H. Zeng, J. Hou, J. Chen, K. Ma, "Light field image quality assessment via the light field coherence," IEEE Trans. Image Process. Vol. 29, pp. 7945–7956, 2020

- [7] W. Zhou, L. Shi, Z. Chen, J. Zhang, Tensor oriented no-reference light field image quality assessment, *IEEE Trans. Image Process.* Vol. 29, pp. 4070–4084, 2020.
- [8] V. Kiran Adhikarla, M. Vinkler, D. Sumin, R.K. Mantiuk, K. Myszkowski, H.-P. Seidel, P. Didyk, "Towards a quality metric for dense light fields," in: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2017
- [9] C. Perra, P. Astola, E.A. da Silva, H. Khanmohammad, C. Pagliari, P. Schelkens, I. Tabus, "Performance analysis of JPEG pleno light field coding," in: *Applications of Digital Image Processing XLII*, SPIE, 2019.
- [10] H. Amirpour, A.M. Pinheiro, M. Pereira, M. Ghanbari, "Reliability of the most common objective metrics for light field quality assessment", in: *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, IEEE, pp. 2402–2406, 2019
- [11] ITU-R Recommendation BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," ITU, Geneva, Switzerland, 2002.
- [12] ITU-R Recommendation BT.710-4, "Subjective assessment methods for image quality in high-definition television," ITU, Geneva, Switzerland, 1998.
- [13] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," ITU, Geneva, Switzerland, 2008.
- [14] ITU-R Recommendation BT.814-1, "Specification and alignment procedures for setting of brightness and contrast of displays," ITU, 1994.
- [15] ITU-R Recommendation BT.1129-2, "Subjective assessment of standard definition digital television (SDTV) systems," ITU, 1998.
- [16] ITU-R Recommendation BT.1361, "Worldwide unified colorimetry and related characteristics of future television and imaging systems," ITU, 1998.
- [17] ITU-R Recommendation BT.815-1, "Specification of a signal for measurement of the contrast ratio of displays," ITU, 1994.
- [18] R. Mantiuk, A. Tomaszewska, and R. Mantiuk, "Comparison of four subjective methods for image quality assessment," *Computer Graphics Forum*, vol. 31, pp. 2478-2491, 2012.
- [19] H. Gulliksen and L. R. Tucker, "A general procedure for obtaining paired comparisons from multiple rank orders," *Psychometrika*, vol. 26, pp. 173-183, June 1961.
- [20] D. A. Silverstein and J. E. Farrell, "Efficient method for paired comparison," *J. Electron. Imag.*, vol. 10, pp. 394-398, April 2001.
- [21] A. M. van Dijk, J.-B. Martens, and A. B. Watson, "Quality assessment of coded images using numerical category scaling," *Proc. SPIE*, vol. 2451, pp. 90-101, Feb. 1995.
- [22] Z. Wang and A. C. Bovik, *Modern image quality assessment. Synthesis Lectures on Image, Video & Multimedia Processing*, Morgan & Claypool Publishers, 2006.
- [23] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using nature scene statistics: JPEG2000," *IEEE Trans. Image Processing*, vol. 14, pp. 1918-1927, Nov. 2005.
- [24] L. Liang, S. Wang, J. Chen, S. Ma, D. Zhao, and W. Gao, "No-reference perceptual image quality metric using gradient profiles for JPEG2000," *Signal Processing: Image Communication*, vol. 25, pp. 502-516, Aug. 2010.
- [25] T. Brando and M. P. Queluz, "No-reference image quality assessment based on DCT domain statistics," *Signal Processing*, vol. 88, pp. 822-833, April 2008.
- [26] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," presented at *IEEE Int. Conf. on Image Processing*, Sept. 2002.
- [27] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Processing*, vol. 18, pp. 717-728, April 2009.
- [28] Z. Wang and A. C. Bovik, "Reduced-and no-reference image quality assessment: the natural scene statistic model approach," in *IEEE Signal Processing Magazine*, vol. 28, Nov. 2011, pp. 29-40.
- [29] A. Rehman and Z. Wang, "Reduced-reference image quality assessment by structural similarity estimation," *IEEE Trans. Image Processing*, vol. 21, pp. 3378-3389, Aug. 2012
- [30] L. Ma, S. Li, and K. N. Ngan, "Visual horizontal effect for image quality assessment," *IEEE Signal Processing Letters*, vol. 17, pp. 627-630, July 2010.
- [31] F. Zhang, W. Liu, W. Lin, and K. N. Ngan, "Spread spectrum image watermarking based on perceptual quality metric," *IEEE Trans. Image Processing*, vol. 20, pp. 3207-3218, Nov. 2011.
- [32] Y. Niu, M. Kyan, L. Ma, A. Beghdadi, and S. Krishnan, "A visual saliency modulated just noticeable distortion profile for image watermarking," in *European Signal Processing Conf.*, 2011.
- [33] Z. Wang, A. C. Bovik, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, pp. 600-612, April 2004
- [34] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," presented at *IEEE Asilomar Conf. on Signals, Systems, and Computers*, Nov. 2003
- [35] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, pp. 011006:1-011006:21; Jan. 2010
- [36] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Processing*, vol. 15, pp. 430-444, Feb. 2006
- [37] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Processing*, vol. 20, pp. 2378-2386, Aug. 2011
- [38] F. Battisti, M. Carli, and P. L. Callet, "A study on the impact of visualization techniques on light field perception," in *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 2155–2159, 2018
- [39] P. Paudyal, F. Battisti, M. Sjöström, R. Olsson, and M. Carli, "Towards the perceptual quality evaluation of compressed light eld images," *IEEE Trans. Broadcast.*, vol. 63, no. 3, pp. 507522, Sep. 2017

- [40] M. Řeřábek and T. Ebrahimi, "New Light Field Image Dataset," 8th International Conference on Quality of Multimedia Experience, Lisbon, Portugal, 2016
- [41] Fraunhofer Heinrich Hertz Institute, "VVC Overview," [Online]. Available: <https://www.hhi.fraunhofer.de/en/departments/vca/technologies-and-solutions/h266-vvc.html>
- [42] Z. Wang, Q. Li, "Information content weighting for perceptual image quality assessment," IEEE Trans. Image Process., vol. 20, no. 5, p. 1185–1198, 2010
- [43] Z. Wang, Q. Li, "Information content weighting for perceptual image quality assessment," IEEE Trans. Image Process., vol. 20, no. 5, p. 1185–1198, 2010
- [44] W. Xue, L. Zhang, X. Mou, A.C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," IEEE Trans. Image Process., vol. 23, no. 2, p. 684–695, 2013
- [45] R. Rassool, "VMAF reproducibility: validating a perceptual practical video quality metric," IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, July 2017
- [46] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans. Image Process., 15(11):3440–3451, 2006
- [47] "Pearson's Correlation Coefficient". In: Kirch W. (eds) Encyclopedia of Public Health. Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-5614-7_2569, 2008.
- [48] Wang, Weijie & Lu, Yanmin. "Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE)", in Assessing Rounding Model. IOP Conference Series: Materials Science and Engineering. 324. 012049. 10.1088/1757-899X/324/1/012049, 2018.
- [49] Figueiredo, Dalson & Júnior, Silva, & Rocha, Enivaldo. "What is R2 all about?." Leviathan-Cadernos de Pesquisa Política. Vol. 3. pp. 60-68, 2011.

A COMPREHENSIVE STUDY ON OBJECTIVE ASSESSMENT METHODS WITH LIGHT FIELD IMAGE.

Abstract: Light Field (LF) imaging is a new technology that enables informative imaging to be captured and simulated at multiple focal distances and positions of images. By collecting the radiation of light rays from different directions, LF can provide depth and angle of view to give users a richer visual experience that is not possible with traditional photos. Therefore, evaluating LF image quality is a necessity in many fields. Many studies evaluating LF image quality have been introduced recently, but there is no comprehensive report comparing these methods. In this paper, we present a comprehensive study on the objective assessment methods based on LF data. The LF image data is used to review includes 228 images from an EPFL source with two different focus levels (pre-focus and background focus). The results are compared with points from the critical assessment method, in order to compare the similarity between the objective assessment methods and the MOS. A set of common objective assessment methods, such as: PSNR, SSIM, MS-SSIM, IW-PSNR, IW-SSIM, IW-MSE, GMSD, VMAF, IFC, JQS, VIF and FSIM, has been developed. tested and the conclusions can be made in this paper.

Keywords: Light Field image, Objective quality assessment, Light Field coding.



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.



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.