No-Reference Light Field Image Quality Assessment Exploiting Saliency

Kamal Lamichhane, *Graduate Student, IEEE*, Michael Neri, *Graduate Student, IEEE*, Federica Battisti, *Senior Member, IEEE*, Pradip Paudyal, and Marco Carli, *Senior Member, IEEE*

Abstract—In the near future, the broadcasting scenario will be characterized by immersive content. One of the systems for capturing the 3D content of a scene is the Light Field imaging. The huge amount of data and the specific transmission scenario impose strong constraints on services and applications. Among others, the evaluation of the quality of the received media cannot rely on the original signal but should be based only on the received data. In this direction, we propose a no-reference quality metric for light field images which is based on spatial and angular characteristics. In more details, the estimated saliency and cyclopean maps of light field images are exploited to extract the spatial features. The angular consistency features are, instead, measured with the use of the Global Luminance Distribution knowledge and the Weighted Local Binary Patterns operator on Epipolar Plane Images. The effectiveness of the proposed metric is assessed by comparing its performance with state-of-the-art quality metrics using 4 datasets: SMART, Win5-LID, VALID 10-bit, and VALID 8-bit. Furthermore, the performance is analyzed in cross-datasets, with different distortions, and for different saliency maps. The achieved results show that the performance of the proposed model outperforms state-of-the-art approaches and perform well for different distortion types and with various saliency models.

Index Terms—No-reference image quality assessment, quality of experience, immersive media, light field imaging, multimedia broadcasting.

I. INTRODUCTION

OVER the last two decades, media consumption has been changing towards immersivity. Current large-scale distribution systems, such as TV and Cinema, are being updated to meet new consumer needs [1]. The advent of the third and fourth generation of mobile networks contributed to the spread of the "TV Everywhere" paradigm thanks to improved technologies and to the definition of new coding standards. To cope with the high demand for multimedia streaming, the fifth generation of mobile communication was launched in 2021. This technology enables higher download speeds up to 10Gbps. However, to provide immersive services, including

Kamal Lamichhane, Michael Neri, and Marco Carli are with the Department of Industrial, Electronic and Mechanical Engineering, Roma Tre University, 00154 Rome, Italy.

Federica Battisti is with the Department of Information Engineering, University of Padova, 35122 Padua, Italy (e-mail: federica.battisti@unipd.it).

Pradip Paudyal is with Nepal Telecommunications Authority, Kathmandu 44600, Nepal.

Virtual Reality (VR), Augmented reality (AR), and eXtended Reality (XR), a new transmission architecture must be devised, according to CISCO white papers [2]. For this reason, the research community has started investigating on the Future Generation Wireless Network (FGWN), i.e., 6G, to satisfy new consumers demands. Its launch is scheduled in 2030 and it will allow the broadcasting of new immersive multimedia such as Light Field Images (LFIs).

Light field imaging has been proposed as a system for capturing three-dimensional (3D) scenes for next-generation broadcast multimedia services [1]. As highlighted in [3], this technology is still in its early stages, and there are several challenges to cope with. One of the main problems is defining a strategy for efficiently storing and sharing this type of content. In fact, a single LFI can be several gigabytes in size, making it difficult to handle. Therefore, new models are needed for data preprocessing, compression, content editing, rendering, and display. These operation can be perceptually relevant thus impacting the subjective quality score [4]. In particular, the transmission of LFI through noisy telecommunication channels can degrade their quality. Hence, Light Field Quality Metrics (LFQMs) are required for evaluating the impact of each transmission step.

In general, to assess the perceptual LFI quality, subjective experiments and objective quality assessment metrics are adopted. The subjective assessment is a reliable quality assessment method but it is a costly and time-consuming process. On the other hand, objective metrics are complex to design and, often, only partially matching the subjective judgment. In full-reference [5], [6], [7] or reduced-reference [8], [9] metrics the full or part of the spatial and angular features of the original LFI are generally used. However the availability of this information may be difficult, especially in a broadcasting scenario, thus motivating the development of no-reference metrics. In the literature, few efforts [10], [11], [12], [13], [14], [15], [16] have been devoted to devise no-reference metrics.

In this work, a novel no-reference metric exploiting spatial and angular characteristics of the light field is presented. Differently from state-of-the-art approaches, the proposed method combines the information from saliency and the cyclopean maps without the use of learning-based convolutional filters thus providing an explainable and robust feature extractor for light field quality estimation. The joint use of the information extracted from these maps can improve the performance of a quality model inspired by the human vision system. In [17], the authors successfully exploited these

models to evaluate the quality of stereo images. This framework has been adopted as light field images can be regarded as a set of stereo pairs.

The major contributions of the article are:

- *Analysis of the impact of distortions on the estimated saliency map:* distortions have been studied to analyze the impacts on the saliency map. The obtained results provide the rationale for using the saliency map as a spatial feature [18];
- *Design of a new angular feature:* angular irregularities of light field impair the judgement of the human eyes on the overall quality perception [19]. For this reason, the feature set Global Luminance Distribution Pattern (GLDP) has been proposed;
- *Definition of a new no-reference light field quality metric:* a new metric which exploits both spatial and angular features of light field images has been designed. The performance of the proposed model has been analyzed on different datasets, cross-datasets, distortion types, and saliency maps.

The rest of the paper is organized as follows: the related works are presented in Section II. The learning-based quality assessment model is presented in Section III while, in Section IV, the objective experiment process is reported. Results analysis and discussions are carried out in Section V. Finally, in Section VI, the conclusions are drawn.

II. RELATED WORK

Based on the availability of original or reference information, an objective quality metric for LFIs can be classified as full-reference, reduced-reference, and no-reference. Full-reference metrics exploit complete original data and measure the similarity between reference and distorted image. The reduced-reference model utilizes only partial information about the reference image, while the no-reference evaluates image quality without any information about the original data. Several efforts have been made in recent years to develop such LFIs quality metrics.

A. Full and Reduced-Reference Quality Metrics

As a first attempt, 2D image quality metrics had been used for LFI quality assessment [20], [21], [22], [23], [24]. These methods usually average the scores obtained by applying the 2D full-reference metrics to single views or sub-aperture images of the LFI. Recently, specifically designed LFI full-reference quality metrics have been proposed [5], [6], [7]. A Log-Gabor-based model is proposed in [5] where the saliency features are extracted from the reference and distorted LFIs, employing the multi-scale and single-scale Gabor wavelet. A full-reference approach is proposed in [7] based on images obtained from the focus stacking of LFIs. Regarding reduced-reference approaches, in [8] a metric based on depth map estimation of LFIs is proposed by exploiting a multi-resolution approach. Then, the Structural Similarity Index Measurement (SSIM) metric is used to analyze the distortion level of depth maps. Similarly, in [25], depth information considering multiple views is used to predict the subjective quality.

However, these techniques have limited use in a broadcasting scenario because of the required reference data.

B. No-Reference Quality Metrics

A no-reference model is presented in [12] by combining the 2D and 3D characteristics of the LFI with a Support Vector Regressor (SVR). The hue, saturation, and value components of each Sub Aperture Image (SAI) are considered as 2D features. The 3D features are obtained using sparse depth maps of horizontal and vertical Epipolar Plane Images (EPIs). Finally, the 2D and 3D features are concatenated and given as input to the SVR. However, limited-size datasets are used for performance analysis, thus prone to overfitting.

Shi et al. [10] proposed a metric that evaluates spatial and angular degradations in a LFI. The spatial degradation is measured by capturing the naturalness distribution of the cyclopean map. The angular consistency, instead, is estimated by applying the Weighted Local Binary Pattern (WLBP) operator on EPIs.

A tensor-based quality metric is proposed in [11] where the light field is regarded as a low-rank 4D tensor. The principal components of four oriented sub-aperture view stacks are obtained via Tucker decomposition. Then, the spatial quality of the LFI is measured by considering the global naturalness and local frequency properties. In the final step, the tensor angular variation index is proposed to measure the angular consistency quality by analyzing the structural similarity distribution between the first principal component and each view in the stack.

The Visualization-based Blind Light Field Image (VBLFI) model is proposed by Xiang et al. [16] which exploits LFI visualization features. In more details, the approach employs the Mean Difference Images, which are obtained by applying partial derivative to LFI, highlighting the depth and structural information. In [13], an extended version of [16] is proposed where SAIs of distorted LFI are the input data. However, only two datasets are used to evaluate the model performance, without any distortions.

In [15], a Convolutional Neural Network (CNN)-based metric is proposed. Its two novelties are the use of discriminative EPI patches for the training of a CNN and a multi-task learning. Qu et al. proposed ALAS-DADS [14], a quality metric that exploits a CNN to extract both spatial and angular features. The neural network is composed of 3 modules. The first branch extracts the spatial features from a light field by exploiting separable convolutions. Similarly, the second module captures angular consistencies. Finally, the third module fuses both spatial and angular features to predict the quality score. However, the experimental results are obtained through only two state-of-the-art datasets.

To the best of our knowledge, previous works did not take into account compression-based distortions for evaluating the performance of their model. Moreover, from the analysis of related works, we highlight the lack of a no-reference model that is validated on multiple LFI datasets. Therefore, a no-reference LFI quality metric has been designed and validated on multiple datasets. The proposed model is fully explainable and of lower computational complexity with respect to learning-based
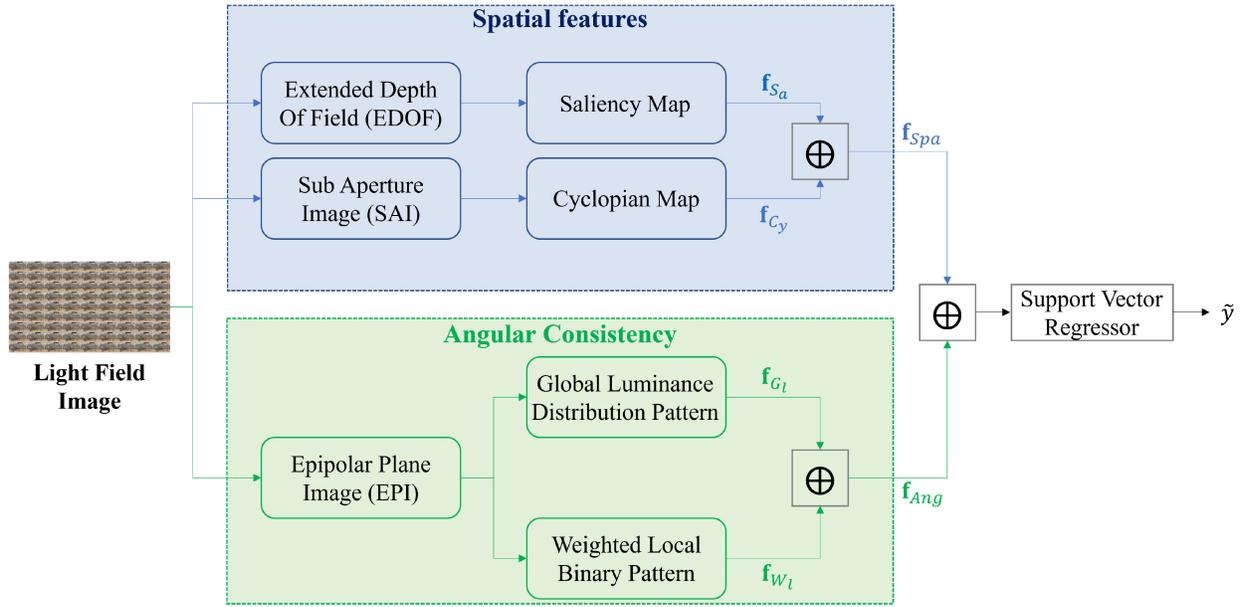
Fig. 1.   Block diagram of the proposed model. The operator $\oplus$ indicates the concatenation function.

methods. Moreover, the adoption of the estimated saliency map improves the effectiveness of prediction of the subjective quality score. In addition, the proposed approach exploits the Global Luminance Distribution knowledge in the design of the feature set GLDP. To the best of our knowledge this is the first attempt in the literature. Finally, this work demonstrates that the combination of saliency- and cyclopean-based features can lead to state-of-the-art performance without the use of learned convolutional filters.

III. PROPOSED MODEL

This section describes the proposed light field quality metric, depicted in Figure 1. It is composed of two parallel units used for extracting spatial and angular LFI characteristics. These proprieties are relevant for the subjective quality assessment of light field images [26]. Finally, the obtained features are concatenated and fed to a SVR to predict the overall quality score. We note as (u, v, s, t) the coordinate vector of a LFI. Moreover, let (U, V, S, T) be the maximum value of u, v, s, and t respectively. Here, u and v are the angular coordinates along horizontal and vertical directions of a LFI, while t and s are the spatial coordinates along vertical and horizontal directions of the SAI. The horizontal EPI are obtained by setting the vertical coordinates v and t. Similarly, vertical EPIs are extracted by fixing horizontal coordinates u and s.

A. Spatial Features

In broadcasting applications, each processing step, such as lossy compression, transmission, and rendering [10], can affect the spatial quality of the LFI. Following [26], in the proposed metric spatial LFI characteristics are considered by analyzing both saliency and cyclopean maps.

1) Saliency Map: The saliency map conveys information on perceptually important areas of the input image [27]. In the proposed metric, to identify saliency regions, we exploit

TABLE I
EMPLOYED STATISTICAL PARAMETERS

Feature	Significance
μ	Average brightness value of the pixel.
σ	Spread of the pixel values.
σ^2	Measuring how individual pixel values differ from their mean.
$\bar{\mu}_3$	Asymmetry of the probability distribution of pixel values from their mean.
$\bar{\mu}_4$	Tailedness of the probability distribution of pixel values.
$q_{0.3}$	Cut-point dividing the range of distribution of pixel values into continuous intervals with equal probabilities.
μ_3	Imbalance measure of pixel distribution.
w	Measure of randomness of the pixel values.

the Itti model [28] applied on the Extended Depth Of Field (EDOF) image, that is a light field representation in which all objects are in focus. In more detail, an EDOF image is obtained by performing a multi-focus fusion and refocusing the LFI at multiple depth planes, followed by a wavelet decomposition-based stacking [29].

Let $S_a \in \mathbb{R}^{T \times S}$ be the saliency map of the EDOF where T and S are the height and width values in pixels, respectively. From the distribution of S_a, we compute 8 statistical parameters [30], [31], [32], [33]: mean (μ), standard deviation (σ), variance (σ^2), skewness ($\bar{\mu}_3$), kurtosis ($\bar{\mu}_4$), quantile of the distribution with probability 0.3 ($q_{0.3}$), 3rd central moment (μ_3), and entropy (w). The significance of each statistical parameter is detailed in Table I.

Then, the feature vector of the saliency map $\mathbf{f}_{S_a} \in \mathbb{R}^8$ is defined as:

$$\mathbf{f}_{S_a} = \mu_{S_a} \oplus \sigma_{S_a} \oplus \sigma^2_{S_a} \oplus \bar{\mu}_{3_{S_a}} \\ \oplus \bar{\mu}_{4_{S_a}} \oplus q_{0.3_{S_a}} \oplus \mu_{3_{S_a}} \oplus w_{S_a}, \quad (1)$$
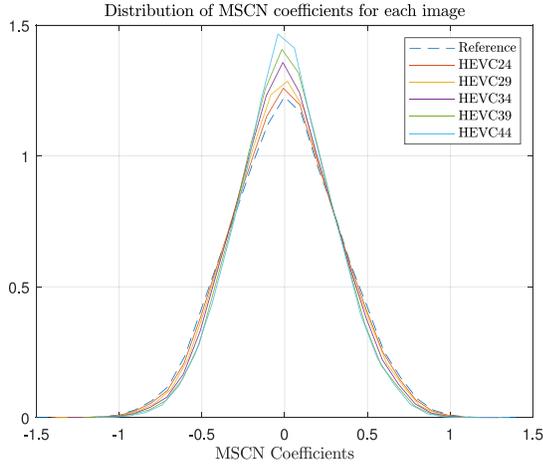
where $\oplus$ is the concatenation operator.

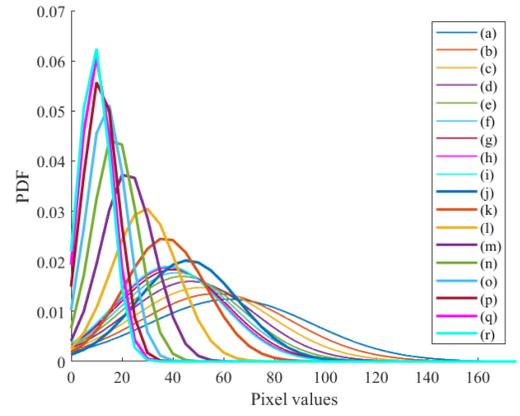Fig. 2. Distribution of MSCN coefficients for reference and HEVC distorted Bikes from Win5-LID dataset.

Fig. 3. Luminance distribution in 18 EPIs (9 horizontal and 9 vertical). (a) to (i) are related to the horizontal EPIs and (j) to (r) are related to the vertical EPIs.

2) Cyclopean Map: The cyclopean map estimates the disparity from each SAI pair, evaluating angular inconsistencies [10], [34], [35]. In more details, the cyclopean map allows to measure the naturalness, i.e., the impact of spatial alterations, by comparing adjacent sub-apertures. This phenomenon is analyzed by computing the Mean Subtracted Contrast Normalized (MSCN) coefficients [36]. In fact, when no distortion is applied, the distribution of the pixels belonging to the MSCN coefficients map shows a Gaussian trend. An example of this behavior is shown in Figure 2 where the pixel distribution of the cyclopean map and of the corresponding MSCN map are depicted. We can see that the MSCN distribution changes when a distortion is applied to a LFI [36]. To compute the MSCN-coefficients, the binocular fusion and rivalry features between adjacent SAIs in both directions are mimicked to obtain the sub-cyclopean map [10], [37], [38]. In the practical realization, stereo disparity estimation [39] and spatial activity map [10] are applied. Then, all the SAI sub-cyclopean maps are superimposed, i.e., the average of them, to extract the MSCN-coefficients.

To analyze the Gaussian trend of MSCN coefficients, the zero mean Asymmetric Generalized Gaussian Distribution (AGGD) procedure is utilized [40], [41]. The algorithm focuses on spatially asymmetric distributions of pixels in the MSCN-map [42], [43], providing four statistical metrics. More precisely, 2 shape parameters (α, η) and 2 scale coefficients along left and right directions $(\sigma_{le}, \sigma_{ri})$ are estimated. By doing so, a parametrised approximation of MSCN coefficients distribution represents the angular characteristics of the LFI.

Finally, the cyclopean-based features $\mathbf{f}_{C_y} \in \mathbb{R}^{12}$ are defined as:

$$\mathbf{f}_{C_y} = \underbrace{\alpha_{C_y} \oplus \eta_{C_y} \oplus \sigma_{le_{C_y}} \oplus \sigma_{ri_{C_y}} \oplus \bar{\mu}_{3C_y} \oplus \bar{\mu}_{4C_y}}_{\text{Original size}}$$
$$\underbrace{\oplus \alpha_{C_{y_2}} \oplus \eta_{C_{y_2}} \oplus \sigma_{le_{C_{y_2}}} \oplus \sigma_{ri_{C_{y_2}}} \oplus \bar{\mu}_{3C_{y_2}} \oplus \bar{\mu}_{4C_{y_2}}}_{\text{Downsampled by 2}}, \quad (2)$$

where $C_y \in \mathbb{R}^{T \times S}$ is the MSCN-coefficients map of the original cyclopean map and C_{y_2} is its version downsampled by a factor of 2. This design choice considers that distortions and artifacts affect the image across different scales [44].

Finally, the spatial features $\mathbf{f}_{S_{pa}} \in \mathbb{R}^{20}$ are obtained by combining both spatial and angular characteristics as follows:

$$\mathbf{f}_{S_{pa}} = \mathbf{f}_{S_a} \oplus \mathbf{f}_{C_y}. \quad (3)$$

B. Angular Consistency

Artifacts in light fields can be caused by angular inconsistencies, due to transmission, reconstruction, or rendering [45]. Using angular consistency features can help improving the performance of quality estimation. For this reason, EPIs are analyzed since they are the main sources of LFI angular consistency [10], [16]. In fact, the luminance of each pixel in a EPI is determined by the angular resolution of the light field camera [46], [47]. From this evidence, a new feature set, called Global Luminance Distribution Pattern (GLDP), is proposed to account for the angular consistency by exploiting the EPI luminance distribution. In addition, another source of angular consistency of LFI is related to the difference between SAIs where each SAI belongs to different angular coordinates. This feature can help analyze the relative relationship among pixels for measuring the change in angular consistency. We explore this feature through the Weighted Local Binary Pattern operator.

1) Global Luminance Distribution Pattern: As demonstrated in [48] and [49], different luminance distributions have an impact on the human perception thus affecting the subjective quality evaluation. Inspired by these results, we apply the same concept to LFI. More precisely, we analyze the luminance distribution of each EPI. In Figure 3 some examples are shown. It can be noticed that all the Probability Density Functions (PDFs) follow an asymmetric Gaussian distribution. We consider this occurrence with the proposed feature set GLDP, extracting statistical parameters from luminance PDFs.

First, a Gaussian low-pass filter is applied to each EPI to remove high-frequency components. Then, four statistical
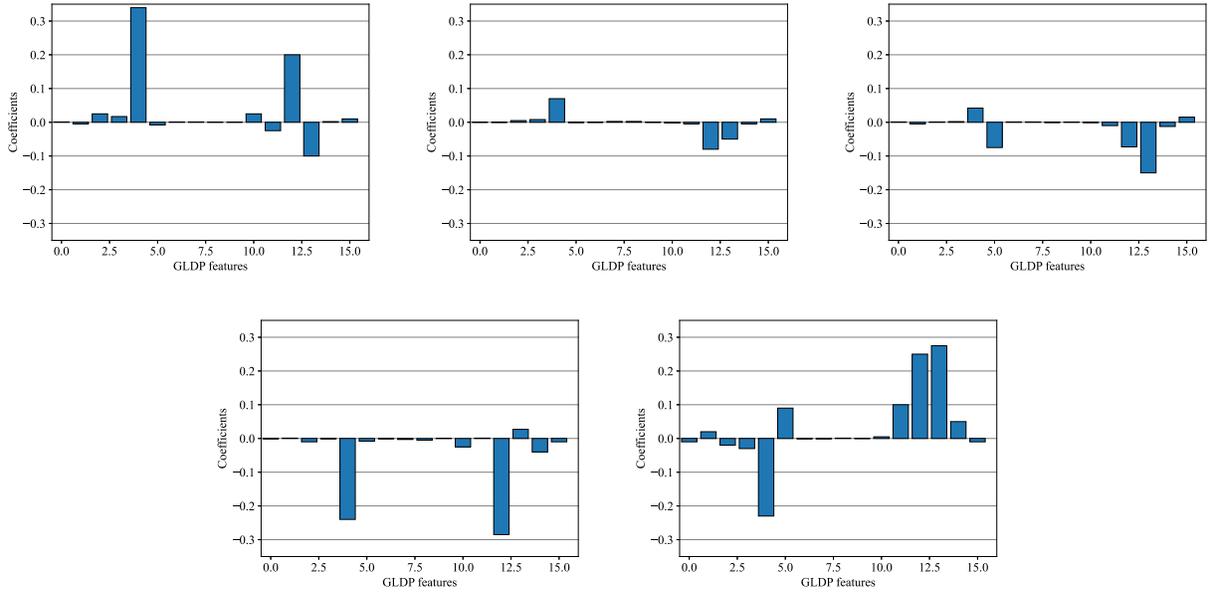
Fig. 4. GLDP coefficients from NN_bikes.bmp (with NN distortion) of Win5-LID dataset. 5 different distortions levels have been applied to the reference image: 10, 20, 30, 40, and 50, respectively.

features are extracted from each PDF: μ, w, $\bar{\mu}_3$, and $\bar{\mu}_4$ [50]. However, there are many horizontal and vertical EPIs in a light field. Hence, concatenating all the features leads to unsatisfactory space and time complexity. To cope with this issue, we evaluate the mean and the variance (μ and σ) of the features for the horizontal and vertical EPIs sets, separately.

In more details, let $\mathcal{H} = \{Hr_i \in \mathbb{R}^{U \times S}, i = 1, \ldots, V \cdot T\}$ be the set of all horizontal EPIs in grayscale with height U and width S. The GLDP features of horizontal EPIs $\mathbf{f}_{G_l}^{\mathcal{H}} \in \mathbb{R}^8$ are obtained as:

$$\begin{aligned} \mathbf{f}_{G_l}^{\mathcal{H}} = {} & \mu_{\mu_{\mathcal{H}}} \oplus \sigma_{\mu_{\mathcal{H}}} \oplus \mu_{w_{\mathcal{H}}} \oplus \sigma_{w_{\mathcal{H}}} \\ & \oplus \mu_{\bar{\mu}_{3\mathcal{H}}} \oplus \sigma_{\bar{\mu}_{3\mathcal{H}}} \oplus \mu_{\bar{\mu}_{4\mathcal{H}}} \oplus \sigma_{\bar{\mu}_{4\mathcal{H}}}. \end{aligned} \tag{4}$$

Similarly, the GLDP features of vertical EPIs $\mathbf{f}_{G_l}^{\mathcal{V}} \in \mathbb{R}^8$ are computed where $\mathcal{V} = \{Vr_i \in \mathbb{R}^{V \times T}, i = 1, \ldots, U \cdot S\}$ is the set of all vertical EPIs in grayscale with height V and width T.

The final feature vector of GLDP for both horizontal and vertical EPIs, $\mathbf{f}_{G_l} \in \mathbb{R}^{16}$, is then arranged by:

$$\mathbf{f}_{G_l} = \mathbf{f}_{G_l}^{\mathcal{H}} \oplus \mathbf{f}_{G_l}^{\mathcal{V}}. \tag{5}$$

GLDP coefficients can be inspected in Figure 4. The proposed feature set is able to highlight and quantify the impact of a distortion to a LFI.

2) Weighted Local Binary Patterns: In [51], [52] the Local Binary Pattern (LBP) operator is introduced. This operator has been successfully applied to the quality assessment of 2D images in [53] since it is able to detect the statistics of local structure primitives at the early stage of vision. More recently, it has been applied also to light field in [10]. The LBP operator uses circular neighborhoods (with radius r and number of pixels p) of different sizes. It is possible to apply a Weighted LBP operator (WLBP) to reduce the number of LBP features by adopting a weighted rotation invariant operator, which has shown effective performance with many 2D quality metrics [54], [55]. Moreover, in our implementation,

we apply three times the WLBP computation with the following values $[r, p] = \{[1, 8], [2, 16], [3, 24]\}$, following the same rationale of [44]. By doing so, the process extracts 54 WLBP coefficients.

The general expression for extracting the WLBP features of horizontal EPIs, $\mathbf{f}_{\mathcal{H}}^{r,p} \in \mathbb{R}^{54}$ is defined as:

$$\mathbf{f}_{\mathcal{H}}^{r,p} = \sum_{v=1}^{V} \sum_{t=1}^{T} \frac{w_{v,t}^{r,p} \mathbf{L}_{v,t}^{r,p}}{w_{v,t}^{r,p}}, \tag{6}$$

where $\mathbf{L}_{v,t}^{r,p}$ represents the rotation invariant uniform LBP operator and $w_{v,t}^{r,p}$ is the entropy of the corresponding EPI. To estimate the impact of distortions in the image structure across scales, we compute the WLBP coefficients at two resolutions: the original and the reduced one by a factor of 2.

Horizontal and vertical WLBP coefficients for both the resolutions are concatenated to obtain the final feature vector $\mathbf{f}_{W_l} \in \mathbb{R}^{216}$ as represented below:

$$\mathbf{f}_{W_l} = \underbrace{\mathbf{f}_{\mathcal{H}}^{r,p} \oplus \mathbf{f}_{\mathcal{V}}^{r,p}}_{\text{Original size}} \oplus \underbrace{\mathbf{f}_{\mathcal{H}*}^{r,p} \oplus \mathbf{f}_{\mathcal{V}*}^{r,p}}_{\text{Downsampled by 2}}, \tag{7}$$

where $\mathbf{f}_{\mathcal{H}*}^{r,p} \in \mathbb{R}^{54}$, $\mathbf{f}_{\mathcal{V}*}^{r,p} \in \mathbb{R}^{54}$ are the WLBP features of $Hr^* \in \mathbb{R}^{\frac{U}{2} \times \frac{S}{2}}$ and $Vr^* \in \mathbb{R}^{\frac{V}{2} \times \frac{T}{2}}$, respectively.

The features obtained from GLDP and WLBP are concatenated in order to compose the angular consistency features $\mathbf{f}_{Ang} \in \mathbb{R}^{232}$ as below:

$$\mathbf{f}_{Ang} = \mathbf{f}_{G_l} \oplus \mathbf{f}_{W_l}. \tag{8}$$

Finally, $\mathbf{f}_{S_{pa}}$ and $\mathbf{f}_{Ang}$ are concatenated to obtain the final feature vector, $\mathbf{f}_{S_{pa}} \oplus \mathbf{f}_{Ang}$, which is applied as input to the SVR [56] that outputs the estimated LFI quality.

TABLE II
LIGHT FIELD IMAGE QUALITY ASSESSMENT (LFIQA) DATASETS

Database	Year	Ref.	Dist.	Distortion source/type	Subjects	Scores	Acquisition Device
SMV [57]	2013	3	15	GN, GB, JPEG	20	DMOS	Basler's ACE 1300GC
SMART [9]	2016	16	256	HEVC-intra, JPEG, JPEG2K, SSDC	19	BT	Lytro Illum
MPI-LF	2017	14	336	DQ, OPT, LI, NNI, GN, HEVC	40	JOD	CANON EOS 5D Mark II
VALID 10-bits [34]	2018	5	100	HEVC,VP9, P3, P4, P5	23	MOS	Lytro Illum
VALID 8-bits [34]	2018	5	40	HEVC, VP9	23	MOS	Lytro Illum
Win5-LID	2018	10	220	HEVC, JPEG2K, LI, NNI, CNN1, CNN2	29	MOS	Lytro Illum
NBU-LF1.0	2019	14	210	NNI, BI, EPICNN [60], Zhang [60], VDSR [61]	22	MOS	Lytro Illum
MC [12]	2019	8	240	JPEG, JPEG2K, GB, WN, MB	22	MOS	Lytro Illum

Ref.: Number of referenced LFI. Dist.: Number of distorted LFI. MOS: mean opinion score. DMOS: Difference MOS. BT: Bradley Terry. JOD: Just Objectionable Difference. GN: Gaussian Noise. GB: Gaussian Blur. MB: Motion Blur. SSDC: Sparse Set and Disparity Coding. DQ: Quantized depth maps. OPT: Optical Flow Interpolation. LI: Linear Interpolation. NNI: Nearest Neighbour Interpolation.

IV. EXPERIMENT DESIGN

In this section we provide details on datasets, experiment setup, and metrics used for evaluating the performances of the proposed approach.

A. Dataset

Table II illustrates the available datasets, at the time of writing, and their characteristics.

In this study, 4 publicly available datasets have been used. These datasets have been selected based on the range of image contents and distortion types. The first dataset is SMART [9] which contains 16 original and 256 distorted LFIs. The distorted images are obtained by introducing 4 types and 4 levels of distortions (HEVC-intra, JPEG, JPEG2K, and Sparse Set and Disparity Coding (SSDC)). The pair-wise comparison method is adopted to collect the subjective Bradley-Terrey ratings for each LFI. A higher Bradley-Terrey score is related to a higher preference rate. Two VALID datasets are considered in this work: VALID 10-bits and VALID 8-bits [34]. Both VALID datasets rely on the same 5 original LFIs. In the VALID 10-bits dataset, 100 distorted LFIs are available, applying HEVC [62], P3 [63], P4 [64], P5 [65], and VP9 [66] compression algorithms with 4 levels of severities. In the VALID 8-bits dataset, instead, 40 distorted LFIs are available with HEVC [62] and VP9 [66] codecs. Both VALID datasets provide subjective scores in terms of Mean Opinion Score (MOS). In the Win5-LID dataset, 10 images (6 real and 4 synthetic) are available. The total number of distorted LFIs is 220 using HEVC and JPEG2K codecs. In addition, reconstruction distortions are present in the dataset such as Linear Interpolation (LI), Nearest Neighborhood Interpolation (NNI), CNN1 [67], and CNN2 [68]. CNN1 and CNN2 are learning-based reconstruction distortions which extract features from EPI by a CNN. The perceptual qualities of LFIs are measured in terms of MOS scores.

B. Implementation Details

EDOF images for each dataset are obtained by using the approach proposed in [29]. To train and test the SVR model, each dataset has been randomly split into train and test sets with a 80%:20% ratio. VALID 10-bits, VALID 8-bits, and

Win5-LID performances are measured with linear kernel function whereas SMART dataset is evaluated with a polynomial one. The choice of each kernel function is based on random hyperparameter optimization [69].

C. Performance Evaluation

To evaluate the performance of the proposed model, Spearman's Rank-order Correlation Coefficient (SRCC), Pearson's Linear Correlation Coefficient (PLCC), Kendall's Rank-order Correlation Coefficient (KRCC), and Root Mean Square Error (RMSE) have been employed. SRCC, PLCC, and KRCC measure monotonic, linear, and ordinal relationships between predicted and ground truth quality scores, respectively. Similarly, the RMSE is used to measure the distance between the predicted and the ground truth quality scores. Following [10], [11], a nonlinear activation function $g : \mathbb{R} \to \mathbb{R}$ has been applied to the model output:

$$g(x, \boldsymbol{\beta}) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp[\beta_2 (x - \beta_3)]} \right) + \beta_4 x + \beta_5, \quad (9)$$

where x is the model output and $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_5]$ are the coefficients optimized in the data nonlinear fitting step.

V. RESULTS ANALYSIS AND DISCUSSIONS

First, to demonstrate the effectiveness of using the saliency map as a spatial feature, an analysis of the impact between distortion and saliency map is reported. Then, the proposed metric is tested on LFI datasets and the results compared with state-of-the-art models. In addition, model's performance with cross-datasets are reported. Finally, an ablation study related to the features adopted in our model is provided.

A. Analysis of the Relationship Among Distortion, Saliency, and Perceived Quality

As highlighted in Section IV-A, LFIQA datasets are composed by different number of reference images (n), distortion types (z), and level of distortions (m). To study whether a distortion of the LFI causes a degradation on its saliency map, the pipeline in Figure 5 has been designed. The SMART dataset has been considered for this analysis, thus z, m, and n are 4, 4, and 16, respectively.

In more details, let $I_{ed} \in \mathbb{R}^{3 \times T \times S}$ and $I_{er} \in \mathbb{R}^{3 \times T \times S}$ be the distorted and reference EDOF images, respectively. Then, I_{de}

TABLE III
CORRELATION BETWEEN THE TYPE OF DISTORTION IN NORMALIZED IMAGE AND SALIENCY MAPS BY USING FIVE SALIENCY MAP MODELS

| Model characteristics		HEVC		JPEG		JPEG2K		SSDC	
		SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
Image normalization		0.9438	0.8875	0.9569	**0.9187**	0.9566	0.8987	0.9784	0.8885
Saliency map	Itti	0.9473	0.7213	0.9654	0.6016	0.9621	0.6568	0.9762	0.7501
	GBVS	0.9460	0.8390	0.9654	0.6991	0.9617	0.7067	0.9762	0.8836
	Geometry	**0.9874**	**0.9782**	**0.9671**	0.7949	**0.9646**	0.8735	**0.9791**	0.9125
	BMS	0.9446	0.8628	0.9645	0.7196	0.9607	0.7695	0.9735	0.8234
	EBMS	0.9464	0.9511	0.9655	0.8388	0.9623	**0.9118**	0.9769	**0.9133**
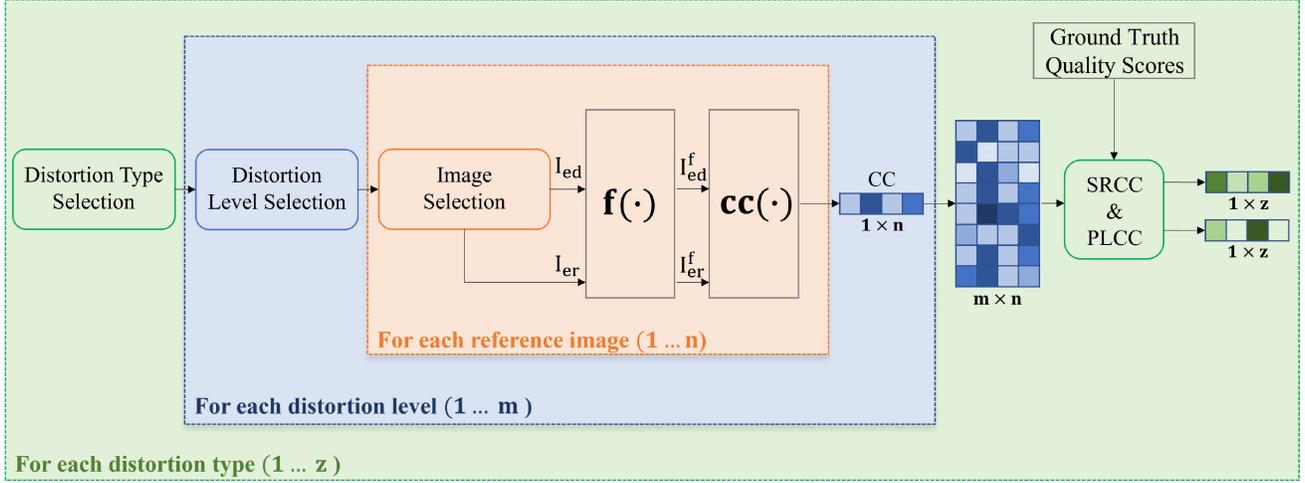
Fig. 5. Block diagram of the test module. I_{ed} and I_{er} are the distorted and reference EDOF images of corresponding LFIs, respectively; $f(\cdot)$ is the saliency extractor; I_{ed}^f and I_{re}^f are the processed I_{ed} and I_{er} from $f(\cdot)$ block.

and I_{er} undergo a saliency extraction step with the function $f(\cdot)$

$$f : \mathbb{R}^{2 \times 3 \times T \times S} \to \mathbb{R}^{2 \times T \times S}, \quad (10)$$

where the normalized image or one of the five saliency map models (Itti [28], GBVS [70], Geometry [71], BMS [72], and EBMS [73]) are computed. The image pairs are obtained as follows:

$$f(I_{ed}) = I_{ed}^f \in \mathbb{R}^{T \times S}, \quad (11)$$
$$f(I_{er}) = I_{er}^f \in \mathbb{R}^{T \times S}. \quad (12)$$

To measure the similarity between the I_{ed}^f and I_{er}^f, we compute the correlation $cc(\cdot) : \mathbb{R}^{2 \times T \times S} \to \mathbb{R}$, described in [74]. This process is iterated for n image pairs and for each distortion level m to obtain a $m \times n$ correlation matrix. A comparison between the obtained matrix and the corresponding subjective quality scores is computed by means of SRCC and PLCC metrics.

The results of the analysis are reported in Table III. The strong correlation between the distortion and the saliency map demonstrates the effectiveness of this representation for quality assessment purposes. Hence, statistics of the saliency map are employed in the proposed metric as spatial features.

B. Selection of the Saliency Model

To select the saliency model to be used in the proposed LFI quality metric, samples extracted from VALID 10-bits are randomly selected. For this purpose, saliency models

TABLE IV
INFLUENCE OF DIFFERENT SALIENCY MAPS ON THE PROPOSED
MODEL'S PERFORMANCE WITH VALID 10-BITS DATASET

Saliency map	SRCC	PLCC	RMSE	KRCC
GBVS	0.9364	0.9644	0.2444	0.8042
BMS	0.9377	0.9655	0.2427	0.8071
EBMS	0.9343	0.9624	0.2529	0.8002
Geometry	0.9345	0.9627	0.2521	0.8011
Itti	**0.9387**	**0.9658**	**0.2416**	**0.8086**

presented in Section V-A have been considered. The obtained results, shown in Table IV, highlight the superiority of the Itti approach [28], motivating its use in the proposed metric.

C. Performance Comparison With State-of-the-Art Quality Metrics

The performance of the proposed model has been compared with 2D image quality metrics:
- *Full-reference:* PSNR [20], SSIM [21], FSIM [22], NQM [23], IFC [24];
- *No-reference:* BRISQUE [36], PIQE [75], NIQE [76]).
In addition, LFI quality metrics are considered:
- *Full-reference:* LGF-LFC [5], Refocus [6], FS-based [7];
- *Reduced-reference:* Depth Map-based [77];
- *No-reference:* MC [12], VBLFI [16], Tensor-LFQ [11], BELIF [78], QMLI [79], LFQA-DSA [13], LFQA [10], Cui et al. [80], ALAS-DADS [14], *Alamgeer and Farias* [81].

TABLE V
PERFORMANCE COMPARISON OF THE PROPOSED LFQM IN TERMS OF SRCC, PLCC, RMSE
AND KRCC ON DIFFERENT DATASETS WITH DIFFERENT QUALITY METRICS

Metric Type	Metric Name	SMART				Win5-LID				VALID 10-Bits				VALID 8-bits			
		SRCC	PLCC	RMSE	KRCC	SRCC	PLCC	RMSE	KRCC	SRCC	PLCC	RMSE	KRCC	SRCC	PLCC	RMSE	KRCC
2D FR	PSNR [20]	0.5605	0.6419	1.6797	0.2831	0.8895	0.8956	0.4597	0.7145	0.8153	0.7856	0.6361	0.6069	0.6429	0.7510	0.9917	0.5714
	SSIM [21]	0.3868	0.7187	1.8913	0.2220	0.7293	0.9223	0.3994	0.5825	0.8774	0.4962	0.7097	0.7090	0.8810	0.9500	0.4690	0.7143
	FSIM [22]	0.4006	0.6792	1.7372	0.2894	0.7810	0.8163	0.5973	0.6210	0.8435	0.7704	0.6532	0.6384	0.8904	0.9120	0.6135	0.7432
	NQM [23]	0.4532	0.5328	1.6234	0.3135	0.8911	0.9299	0.3800	0.7303	0.8153	0.8804	0.4852	0.6069	0.9048	0.8684	0.7448	0.7857
	IFC [24]	0.4184	0.3869	2.3623	0.2910	0.7125	0.8495	0.5450	0.5118	0.7032	0.7632	0.5569	0.5832	0.5873	0.6835	1.0240	0.5035
2D NR	BRISQUE [36]	0.0868	0.0486	0.8450	0.0556	0.2074	0.2080	0.9140	0.1468	0.6864	0.6802	0.6489	0.4889	0.6878	0.7253	0.9421	0.5078
	PIQE [75]	0.3718	0.4641	1.8913	0.2565	0.3756	0.4592	0.9300	0.2581	0.3303	0.5231	0.7219	0.2540	0.3593	0.7120	1.0357	0.3273
	NIQE [76]	0.1727	0.2384	1.8313	0.1184	0.2755	0.5090	0.9923	0.1698	0.3032	0.2632	0.7886	0.2222	0.6347	0.4474	1.3472	0.4728
LFI FR	LGF-LFC [5]	0.8250	0.8270	1.2110	0.6270	–	–	–	–	–	–	–	–	–	–	–	–
	Refocus [6]	–	–	–	–	–	–	–	–	0.9490	0.9680	0.2430	0.8100	0.9770	0.9870	0.2170	0.8900
	FS based [7]	–	–	–	–	–	–	–	–	0.9560	0.9740	0.2190	0.8270	0.9750	0.9860	0.2270	0.8800
LFI RR	DM based [77]	0.8300	0.8300	0.1600	–	–	–	–	–	–	–	–	–	–	–	–	–
LFI NR	MC [12]	0.8920	0.9110	0.8390	0.6950	–	–	–	–	–	–	–	–	–	–	–	–
	VBLFI [16]	–	–	–	–	0.9010	0.9230	0.3510	–	–	–	–	–	–	–	–	–
	Tensor-LFQ [11]	0.8700	0.9020	0.8230	–	0.9100	0.9220	0.3780	–	0.9370	0.9640	0.2300	–	0.9290	0.9850	0.1830	–
	BELIF [78]	0.8370	0.8830	0.8350	–	0.8720	0.8910	0.4290	–	0.9190	0.9620	0.2390	–	0.9280	0.9860	0.1680	–
	QMLI [79]	–	–	–	–	0.8800	0.9040	0.4150	–	–	–	–	–	0.9290	0.9690	0.2800	–
	LFQA-DSA [13]	–	–	–	–	0.9170	0.9390	0.3470	–	–	–	–	–	–	–	–	–
	LFQA [10]	0.8800	0.9110	0.8300	0.7130	0.9030	0.9210	0.3900	0.7350	0.9230	0.9520	0.2770	0.8020	0.9290	0.9800	0.2380	0.8450
	Cui [80]	-	-	-	-	0.9116	0.9262	0.3810	-	0.9095	0.9604	0.2584	-	-	-	-	-
	ALAS-DADS [14]	0.8540	0.9344	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Alamgeer et al. [81]	0.9364	0.9294	-	-	0.9469	0.9361	-	-	0.9489	0.8888	-	-	0.9246	0.9406	-	-
	LFQIA (Ours)	0.9243	0.9193	0.7939	0.7694	0.9331	0.9558	0.3268	0.7862	0.9387	0.9658	0.2416	0.8086	0.9533	0.9892	0.1631	0.8570

TABLE VI
PROPOSED MODEL PERFORMANCE WITH DIFFERENT DISTORTIONS

Performance metrics	SMART				Win5-LID				VALID 10-bits					VALID 8-bits	
	HEVC	JPEG	JPEG2K	SSDC	HEVC	JPEG2K	LI	NNI	HEVC	P3	P4	P5	VP9	HEVC	VP9
SRCC	0.9179	0.9332	0.9487	0.8891	0.9478	0.9532	0.9001	0.8972	0.9253	0.9526	0.8628	0.9140	0.9601	0.9487	0.9509
PLCC	0.9264	0.9133	0.9490	0.8584	0.9591	0.9637	0.9331	0.9182	0.9480	0.9736	0.8874	0.9105	0.9791	0.9764	0.9782
KRCC	0.7823	0.7591	0.8464	0.7382	0.8751	0.8874	0.7971	0.7793	0.8159	0.8871	0.7266	0.7972	0.9201	0.8732	0.8786
RMSE	0.7487	0.8453	0.8576	1.1610	0.3153	0.2904	0.1722	0.1493	0.2593	0.2595	0.4390	0.1902	0.0901	0.0961	0.0613

The results, reported in Table V, show that the proposed metric outperforms state-of-the-art approaches. For both SMART and Win5-LID datasets, the proposed model ranks first among all quality metrics under consideration. Regarding the 8- and 10-bit VALID datasets, the performance of the proposed model is superior to all no-reference metrics considered. Moreover, the results are comparable with full-reference metrics. A more detailed analysis of the model performances is shown in Table VI, demonstrating its effectiveness with respect to the compression artifacts. A slight performance variation can be noticed for different encoding schemes, e.g., an higher correlation is obtained from HEVC with respect to JPEG. This difference could be due to compression methods that handle differently spatial and angular consistency.

D. Robustness Evaluation of Proposed Model With Cross-Datasets

The cross-datasets test is used to validate the model between different datasets. Win5-LID, VALID 10-bits, and VALID 8-bits datasets have been selected for the study. The achieved results are presented in Table VII, where a high correlation between VALID 10-bits and VALID 8-bits is shown. This result is obtained since VALID 10-bits and VALID 8-bits images share similar proprieties, i.e., spatial and angular consistency features. However, the proposed model's performance are low when the Win5-LID dataset is used as a training set and VALID as test. This result could be due to the angular resolution of Win5-LID which is [9 × 9] whereas VALID is [13 × 13]. In addition, the image content and the spatial and angular consistency features are different, thus changing the pixel luminance distribution.

TABLE VII
CROSS-DATASET PERFORMANCE

Training dataset	Test dataset	Performance Metric			
		SRCC	PLCC	KRCC	RMSE
Win5-LID	VALID 10-bits	0.6621	0.6876	0.5221	0.9337
	VALID 8-bits	0.6429	0.6618	0.5014	1.3200
VALID 10-bits	VALID 8-bits	0.8881	**0.9334**	**0.7306**	**0.6384**
VALID 8-bits	VALID 10-bits	**0.8848**	0.8976	0.7166	0.9397

E. Ablation Study

We study the performance of our model with 3 feature sets: $\mathbf{f}_{S_{pa}}$, $\mathbf{f}_{Ang}$, and $\mathbf{f}_{S_{pa}} \oplus \mathbf{f}_{Ang}$. This study evaluates the importance of spatial and angular consistency features in the proposed LFQM. Similarly, it shows the improvement in model performance when both spatial and angular consistency features are involved in the prediction process. The results are reported in Table VIII. It is notable the superiority of $\mathbf{f}_{Ang}$ compared to the $\mathbf{f}_{S_{pa}}$ for all datasets. The main reason is that $\mathbf{f}_{Ang}$ considers the distortion effect on higher-dimensional space of LFI at once, i.e., EPI are employed with full length of angular resolution along horizontal and vertical directions. Although $\mathbf{f}_{S_{pa}}$ has a lower number of dimensions than $\mathbf{f}_{Ang}$, it provides additional information, increasing the correlation between subjective and predicted score.

VI. CONCLUSION

In this work the use of saliency information for light field quality assessment has been exploited. As a first step, we have verified the impact of a distortion in LFI on the estimated saliency map. The achieved results on multiple LFI datasets show high correlation scores between the measure of distortion in saliency map and the subjective quality score

TABLE VIII
SIMILARITY MEASUREMENT OF THE PREDICTED QUALITY
SCORES AND GROUND TRUTH QUALITY SCORES FOR
7 TYPES OF INPUT FEATURES TO THE SVR

Dataset	Features	Performance Indices			
		SRCC	PLCC	KRCC	RMSE
SMART	$\mathbf{f}_{S_a}$	0.4572	0.5164	0.3271	1.7424
	$\mathbf{f}_{C_y}$	0.5760	0.5680	0.4243	1.6746
	$\mathbf{f}_{S_{pa}}$	0.7471	0.7892	0.5835	1.2173
	$\mathbf{f}_{G_l}$	0.5654	0.6694	0.4165	1.5133
	$\mathbf{f}_{W_l}$	0.6954	0.747	0.5122	1.3527
	$\mathbf{f}_{Ang}$	0.8927	0.8993	0.7412	0.8493
	$\mathbf{f}_{S_{pa}} \oplus \mathbf{f}_{Ang}$	**0.9243**	**0.9193**	**0.7694**	**0.7939**
Win5-LID	$\mathbf{f}_{S_a}$	0.4212	0.4787	0.3050	0.8505
	$\mathbf{f}_{C_y}$	0.5994	0.6114	0.4578	0.8119
	$\mathbf{f}_{S_{pa}}$	0.6912	0.7071	0.4960	0.7861
	$\mathbf{f}_{G_l}$	0.4992	0.6637	0.3583	0.4760
	$\mathbf{f}_{W_l}$	0.8649	0.8874	0.7132	0.3824
	$\mathbf{f}_{Ang}$	0.8944	0.9175	0.7509	0.3604
	$\mathbf{f}_{S_{pa}} \oplus \mathbf{f}_{Ang}$	**0.9331**	**0.9558**	**0.7862**	**0.3268**
VALID 10-bits	$\mathbf{f}_{S_a}$	0.4549	0.6290	0.3016	0.7318
	$\mathbf{f}_{C_y}$	0.8272	0.8786	0.6868	0.4056
	$\mathbf{f}_{S_{pa}}$	0.8992	0.9241	0.7510	0.3414
	$\mathbf{f}_{G_l}$	0.5648	0.7090	0.4191	0.6682
	$\mathbf{f}_{W_l}$	0.8636	0.8963	0.7149	0.3607
	$\mathbf{f}_{Ang}$	0.9124	0.9461	0.7791	0.2904
	$\mathbf{f}_{S_{pa}} \oplus \mathbf{f}_{Ang}$	**0.9387**	**0.9658**	**0.8086**	**0.2416**
VALID 8-bits	$\mathbf{f}_{S_a}$	0.4658	0.6371	0.3216	0.6318
	$\mathbf{f}_{C_y}$	0.8375	0.8798	0.7057	0.3546
	$\mathbf{f}_{S_{pa}}$	0.9048	0.9550	0.7857	0.2525
	$\mathbf{f}_{G_l}$	0.5759	0.7191	0.4302	0.6282
	$\mathbf{f}_{W_l}$	0.8833	0.9161	0.7359	0.2907
	$\mathbf{f}_{Ang}$	0.9379	0.9691	0.8387	0.1821
	$\mathbf{f}_{S_{pa}} \oplus \mathbf{f}_{Ang}$	**0.9533**	**0.9892**	**0.8570**	**0.1631**

of the corresponding LFI. Then, we have exploited the spatial and angular consistency features of LFIs based on a machine learning approach. The model performance has been evaluated on LFI quality datasets: SMART, WIN5LID, VALID 10-bits, and VALID 8-bits. The experimental results show that the proposed model outperforms state-of-the-art no-reference quality metrics. Moreover, the results show that our model guarantees good results in cross-dataset evaluation with different distortions types and saliency map models, yielding good generalization capabilities. Finally, it is possible that learning-based approaches, with saliency and cyclopean maps as input, can reach better performances that hand-crafted features. Clearly, an ad-hoc modeling study has to be carried out to account for computational complexity and performance. Hence, the use of learning-based features extracted from saliency and cyclopean maps is set as future work.

REFERENCES

[1] H. P. S. Rufino and R. Prasad, "6G networks for next generation of digital TV beyond 2030," *Wireless Pers. Commun.*, vol. 121, pp. 1363–1378, Sep. 2021.

[2] "Cisco Annual Internet Report 2018–2023." Accessed: Oct. 19, 2022. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[3] E. C. Strinati et al., "6G: The next Frontier: From holographic messaging to artificial intelligence using subterahertz and visible light communication," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 42–50, Sep. 2019.

[4] H. Huang, H. Zeng, Y. Tian, J. Chen, J. Zhu, and K.-K. Ma, "Light field image quality assessment: An overview," in *Proc. IEEE Conf. Multimedia Inf. Process. Retrieval (MIPR)*, 2020, pp. 348–353.

[5] Y. Tian, H. Zeng, J. Hou, J. Chen, and K.-K. Ma, "Light field image quality assessment via the light field coherence," *IEEE Trans. Image Process.*, vol. 29, pp. 7945–7956, 2020.

[6] C. Meng, P. An, X. Huang, and C. Yang, "Objective quality assessment for light field based on refocus characteristic," in *Proc. Int. Conf. Image Graph.*, 2019, pp. 193–204.

[7] C. Meng, P. An, X. Huang, C. Yang, L. Shen, and B. Wang, "Objective quality assessment of lenslet light field image based on focus stack," *IEEE Trans. Multimedia*, vol. 24, pp. 3193–3207, 2021.

[8] G. Arru, F. Battisti, and M. Carli, "Evaluating the effectiveness of image quality metrics in a light field scenario," *Electron. Imag.*, vol. 2019, no. 11, pp. 1–6, 2019.

[9] P. Paudyal, R. Olsson, M. Sjöström, F. Battisti, and M. Carli, "SMART: A light field image quality dataset," in *Proc. Int. Conf. Multimedia Syst.*, 2016, pp. 1–6.

[10] L. Shi, W. Zhou, Z. Chen, and J. Zhang, "No-reference light field image quality assessment based on spatial-angular measurement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 4114–4128, Nov. 2020.

[11] W. Zhou, L. Shi, Z. Chen, and J. Zhang, "Tensor oriented no-reference light field image quality assessment," *IEEE Trans. Image Process.*, vol. 29, pp. 4070–4084, 2020.

[12] L. Shan, P. An, C. Meng, X. Huang, C. Yang, and L. Shen, "A no-reference image quality assessment metric by multiple characteristics of light field images," *IEEE Access*, vol. 7, pp. 127217–127229, 2019.

[13] J. Xiang, G. Jiang, M. Yu, Y. Bai, and Z. Zhu, "No-reference light field image quality assessment based on depth, structural and angular information," *Signal Process.*, vol. 184, Jul. 2021, Art. no. 108063.

[14] Q. Qu, X. Chen, V. Chung, and Z. Chen, "Light field image quality assessment with auxiliary learning based on depthwise and angle-wise separable convolutions," *IEEE Trans. Broadcast.*, vol. 67, no. 4, pp. 837–850, Dec. 2021.

[15] P. Zhao, X. Chen, V. Chung, and H. Li, "Low-complexity deep no-reference light field image quality assessment with discriminative EPI patches focused," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, 2021, pp. 1–6.

[16] J. Xiang, M. Yu, H. Chen, H. Xu, Y. Song, and G. Jiang, "VBLFI: Visualization-based blind light field image quality assessment," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, 2020, pp. 1–6.

[17] J. Yang et al., "Quality assessment metric of stereo images considering cyclopean integration and visual saliency," *Inf. Sci.*, vol. 373, pp. 251–268, Dec. 2016.

[18] K. Lamichhane, F. Battisti, P. Paudyal, and M. Carli, "Exploiting saliency in quality assessment for light field images," in *Proc. Picture Coding Symp. (PCS)*, 2021, pp. 1–5.

[19] C. Meng, P. An, X. Huang, C. Yang, and D. Liu, "Full reference light field image quality evaluation based on angular-spatial characteristic," *IEEE Signal Process. Lett.*, vol. 27, pp. 525–529, 2020.

[20] J. Korhonen and J. You, "Peak signal-to-noise ratio revisited: Is simple beautiful?" in *Proc. 4th Int. Workshop Qual. Multimedia Exp. (QoMEX)*, 2012, pp. 37–38.

[21] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[22] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[23] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 636–650, Apr. 2000.

[24] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.

[25] P. Paudyal, F. Battisti, and M. Carli, "Reduced reference quality assessment of light field images," *IEEE Trans. Broadcast.*, vol. 65, no. 1, pp. 152–165, Mar. 2019.

[26] G. Wu et al., "Light field image processing: An overview," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 926–954, Oct. 2017.

[27] T. Kadir and M. Brady, "Saliency, scale and image description," *Int. J. Comput. Vis.*, vol. 45, no. 2, pp. 83–105, 2001.

[28] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[29] M. Brizzi, F. Battisti, and A. Neri, "Light field depth-of-field expansion and enhancement based on multifocus fusion," in *Proc. 8th Eur. Workshop Visual Inf. Process. (EUVIP)*, 2019, pp. 146–151.

[30] D. Li, T. Jiang, W. Lin, and M. Jiang, "Which has better visual quality: The clear blue sky or a blurry animal?" *IEEE Trans. Multimedia*, vol. 21, no. 5, pp. 1221–1234, May 2019.

[31] X. Chen, Q. Zhang, M. Lin, G. Yang, and C. He, "No-reference color image quality assessment: From entropy to perceptual quality," *EURASIP J. Image Video Process.*, vol. 2019, no. 1, pp. 1–14, 2019.

[32] P. Mazumdar and F. Battisti, "A content-based approach for saliency estimation in 360 images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2019, pp. 3197–3201.

[33] S. Rajkumar and G. Malathi, "A comparative analysis on image quality assessment for real time satellite images," *Indian J. Sci. Technol.*, vol. 9, no. 34, pp. 1–11, 2016.

[34] I. Viola and T. Ebrahimi, "VALID: Visual quality assessment for light field images dataset," in *Proc. Int. Conf. Qual. Multimedia Exp. (QoMEX)*, 2018, pp. 1–3.

[35] M. F. Bradshaw, K. M. Elliott, S. J. Watt, P. B. Hibbard, I. R. Davies, and P. J. Simpson, "Binocular cues and the control of prehension," *Spatial Vis.*, vol. 17, nos. 1–2, pp. 95–110, 2004.

[36] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[37] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks, "The zone of comfort: Predicting visual discomfort with stereo displays," *J. Vis.*, vol. 11, no. 8, p. 11, 2011.

[38] B. A. Steinman and R. P. Garzia, *Foundations of Binocular Vision: A Clinical Perspective*. New York, NY, USA: McGraw-Hill Educ., 2000.

[39] V. Mohandas and R. Paily, "Stereo disparity estimation algorithm for blind assisting system," *CSI Trans. ICT*, vol. 1, no. 1, pp. 3–8, 2013.

[40] A. Tesei and C. S. Regazzoni, "The asymmetric generalized Gaussian function: A new HOS-based model for generic noise PDFs," in *Proc. 8th Workshop Stat. Signal Array Process.*, 1996, pp. 1–8.

[41] N. Lasmar, Y. Stitou, and Y. Berthoumieu, "Multiscale skewed heavy tailed model for texture analysis," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, 2009, pp. 2281–2284.

[42] T. Kato, S. Omachi, and H. Aso, "Asymmetric Gaussian and its application to pattern recognition," in *Proc. Joint IAPR Int. Workshops Stat. Techn. Pattern Recognit. (SPR) Struct. Syn. Pattern Recognit. (SSPR)*, 2002, pp. 405–413.

[43] X. Zhao, J. Sun, and W. Gao, "A novel rate estimation model for mode decision of H.264/AVC," in *Proc. Vis. Commun. Image Process.*, vol. 7257, 2009, Art. no. 72570R.

[44] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden, "Pyramid methods in image processing," *RCA Eng.*, vol. 29, no. 6, pp. 33–41, 1984.

[45] V. K. Adhikarla et al., "Towards a quality metric for dense light fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 3720–3729.

[46] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *Int. J. Comput. Vis.*, vol. 1, no. 1, pp. 7–55, 1987.

[47] Bob Fisher. "Epipolar Geometry." 1997. [Online]. Available: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/EPSRC_SSAZ/node18.html

[48] D. L. Ruderman, "The statistics of natural images," *Netw. Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.

[49] A. R. Buffett, "Visual perception of luminance contrast: Implications for the assessment of subjective image quality," *Displays*, vol. 7, no. 1, pp. 30–36, 1986.

[50] V. Kumar and P. Gupta, "Importance of statistical measures in digital image processing," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 8, pp. 56–62, 2012.

[51] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[52] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[53] M. Zhang, C. Muramatsu, X. Zhou, T. Hara, and H. Fujita, "Blind image quality assessment using the joint statistics of generalized local binary pattern," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 207–210, Feb. 2015.

[54] F. Rezaie, M. S. Helfroush, and H. Danyali, "No-reference image quality assessment using local binary pattern in the wavelet domain," *Multimedia Tools Appl.*, vol. 77, no. 2, pp. 2529–2541, 2018.

[55] A. Satpathy, X. Jiang, and H. Eng, "LBP-based edge-texture features for object recognition," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 1953–1964, May 2014.

[56] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 1999.

[57] R. Tamboli, K. K. Vupparaboina, J. Ready, S. Jana, and S. Channappayya, "A subjective evaluation of true 3D images," in *Proc. IEEE Int. Conf. 3D Imag. (IC3D)*, 2014, pp. 1–8.

[58] L. Shi, S. Zhao, W. Zhou, and Z. Chen, "Perceptual evaluation of light field image," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, 2018, pp. 41–45.

[59] Z. Huang, M. Yu, G. Jiang, K. Chen, Z. Peng, and F. Chen, "Reconstruction distortion oriented light field image dataset for visual communication," in *Proc. Int. Symp. Netw. Comput. Commun. (ISNCC)*, 2019, pp. 1–5.

[60] S. Zhang, H. Sheng, D. Yang, J. Zhang, and Z. Xiong, "Micro-lens-based matching for scene recovery in lenslet cameras," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1060–1075, Mar. 2018.

[61] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 1646–1654.

[62] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[63] S. Zhao and Z. Chen, "Light field image coding via linear approximation prior," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2017, pp. 4562–4566.

[64] W. Ahmad, R Olsson, and M. Sjöström, "Interpreting plenoptic images as multi-view sequences for improved compression," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2017, pp. 4557–4561.

[65] I. Tabus, P. Helin, and P. Astola, "Lossy compression of lenslet images from plenoptic cameras combining sparse predictive coding and JPEG 2000," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2017, pp. 4567–4571.

[66] D. Mukherjee et al., "The latest open-source video codec VP9—An overview and preliminary results," in *Proc. Picture Coding Symp. (PCS)*, 2013, pp. 390–393.

[67] G. Wu, M. Zhao, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field reconstruction using deep convolutional network on EPI," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 1638–1646.

[68] N. K. Kalantari, T. Wang, and R. Ramamoorthi, "Learning-based view synthesis for light field cameras," *Trans. Graph.*, vol. 35, no. 6, pp. 1–10, 2016.

[69] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, no. 2, pp. 281–305, 2012.

[70] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 545–552.

[71] P. Mazumdar, K. Lamichhane, M. Carli, and F. Battisti, "A feature integrated saliency estimation model for omnidirectional immersive images," *Electronics*, vol. 8, no. 12, p. 1538, 2019.

[72] J. Zhang and S. Sclaroff, "Saliency detection: A Boolean map approach," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 153–160.

[73] J. Zhang and S. Sclaroff, "Exploiting surroundedness for saliency detection: A Boolean map approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 5, pp. 889–902, May 2016.

[74] Z. Bylinskii, T. Judd, A. Oliva, A. Torralba, and F. Durand, "What do different evaluation metrics tell us about saliency models?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 3, pp. 740–757, Mar. 2019.

[75] N. Venkatanath, D. Praneeth, B. M. Chandrasekhar, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features," in *Proc. 21st Nat. Conf. Commun. (NCC)*, 2015, pp. 1–6.

[76] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[77] P. Paudyal, F. Battisti, M. Sjöström, R. Olsson, and M. Carli, "Towards the perceptual quality evaluation of compressed light field images," *IEEE Trans. Broadcast.*, vol. 63, no. 3, pp. 507–522, Sep. 2017.

[78] L. Shi, S. Zhao, and Z. Chen, "Belif: Blind quality evaluator of light field image with tensor structure variation index," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2019, pp. 3781–3785.

[79] Z. Luo, W. Zhou, L. Shi, and Z. Chen, "No-reference light field image quality assessment based on micro-lens image," in *Proc. Picture Coding Symp. (PCS)*, 2019, pp. 1–5.

[80] Y. Cui, M. Yu, Z. Jiang, Z. Peng, and F. Chen, "Blind light field image quality assessment by analyzing angular-spatial characteristics," *Digit. Signal Process.*, vol. 117, Oct. 2021, Art. no. 103138.

[81] S. Alamgeer and M. C. Q. Farias, "A two-stream CNN based visual quality assessment method for light field images," *Multimedia Tools Appl.*, vol. 82, pp. 5743–5762, Jul. 2022.

Kamal Lamichhane (Graduate Student Member, IEEE) received the B.Sc. degree in electronics and communication engineering from the Pokhara Engineering College of Pokhara Univiersity, Nepal, in 2003, the M.Sc. degree in information and communication engineering from the Institute of Engineering, Pulchowk Campus, Nepal, in 2010, and the master's degree in business adminstration from Indira Gandhi National Open University in 2014. He is currently pursuing the Ph.D. degree in applied electronics with Roma Tre University, Rome, Italy. He was a Research Fellow from La Sapienza University, Rome, in 2018.

Pradip Paudyal received the M.Sc. degree in information and communication engineering from Tribhuvan University, Kathmandu, Nepal, in 2010, and the Ph.D. degree from Università degli Studi Roma TRE, Italy, in 2017. He is currently the Deputy Director of Nepal Telecommunications Authority. His current research fields include multimedia communication and quality assessment.

Michael Neri (Graduate Student Member, IEEE) received the Laurea (B.Sc.) degree in information engineering in 2019, and the Laurea Magistrale (M.Sc.) degree from the University of Padova in 2021. He is currently pursuing the Ph.D. degree in applied electronics with the Department of Industrial, Electronic, and Mechanical Engineering, Roma Tre University. His main research interests are in the area of computer vision, deep learning, and audio processing.

Federica Battisti (Senior Member, IEEE) is an Associate Professor with the Department of Information Engineering, University of Padova. Her research interests include multimedia quality assessment and security. She serves as an Associate Editor for the IEEE TRANSACTIONS ON MULTIMEDIA, *EURASIP Journal on Image and Video Processing*, and *Signal Processing: Image Communication* (Elsevier).

Marco Carli (Senior Member, IEEE) received the Laurea degree in telecommunication engineering from the Università degli Studi di Roma "La Sapienza," Rome, Italy, and the Ph.D. degree from the Tampere University of Technology, Tampere, Finland. He is an Associate Professor with the Università degli Studi Roma Tre, Rome. He was a Visiting Researcher with the Image Processing Laboratory, University of California at Santa Barbara, Santa Barbara, CA, USA, from 2000 to 2004. He has been an Associate Editor of *EURASIP Journal on Image and Video Processing* since 2011. He is an Associate Editor of IEEE TRANSACTIONS ON MULTIMEDIA and an Area Editor of *Signal Processing: Image Communication* (Elsevier).