A No-Reference Image Quality Assessment Metric by Multiple Characteristics of Light Field Images

LIANG SHAN , PING AN , (Member, IEEE), CHUNLI MENG, XINPENG HUANG , CHAO YANG, AND LIQUAN SHEN , (Member, IEEE)

Key Laboratory for Advanced Display and System Application, Ministry of Education, Shanghai University, Shanghai 200072, China
Shanghai Institute for Advanced Communication and Data Science, School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China

Corresponding author: Ping An (anping@shu.edu.cn)

ABSTRACT Evaluation of light field image (LFI), especially micro-lens camera light field (LF), is a new and challenging work. The development of image quality assessment (IQA) metric of LFIs relies on the subjective quality assessment database. In this paper, we establish a perceptual quality assessment dataset consisting of 240 distorted images from 8 source images with five distortion types. Furthermore, a no-reference IQA metric is proposed by combining 2D and 3D characteristics of LFI with the Support Vector Regression (SVR) model. The performance of the proposed metric is demonstrated by comparing with some classical full reference IQA metrics both on the presented dataset and a third-party dataset. The experiment results show that our method has a better performance than others.

INDEX TERMS Light field images, subjective quality assessment, objective quality assessment, light field characteristics, SVR.

I. INTRODUCTION

With the development of Computational Photography [1], the capability of a camera can be greatly improved. It benefits from the combination of digital image capturing and processing techniques. Meanwhile, traditional photography theory seems dwarfed due to the limitation of optical processing. As a significant branch of Computational Photography, Light Field (LF) imaging, also known as plenoptic imaging, holoscopic imaging and integral imaging [2], shows a huge advantage in refocusing, 3-dimension (3D) imaging, virtual reality (VR) and many other technologies in computer vision. LF describes the distribution of light rays in free space, including their position, angle, and radiance [3]. The concept of LF was introduced as early as 1936 by Arun [4], and in 1991, a function called seven-dimensional plenoptic function used to describe the LF imaging was proposed by Edward and Bergen [5], the function contains three-dimensional coordinates of each spot in a light, the direction, the wavelength of light, and time parameter, to describe the spatially distributed geometric light. To simplify the unnecessary parts of the function, in 1996, Marc and Hanrahan [6] only used

four dimensions to express the plenoptic function: $L = (x, y, u, v)$. Where x, y and u, v are two parallel planes, (x, y) and (u, v) are two points on two planes respectively, L is a light that passes through the above two points. Technically, all types of images captured by the devices that can obtain the LF information based on the function can be called LF images (LFIs). These devices can be multiple camera gantry, single camera slide system or micro-lens array camera, etc. And the micro-lens array camera with the advantages of portability and refocusing from one shot has been becoming a significant development direction of LF cameras. And after Lytro [7] announced their first consumer micro-lens array camera, the study of the post-processing of LF data entered a new era.

The LF data obtained by the micro-lens array camera is quite different from the ordinary image. Therefore, there are wealth post-processing applications for such LFIs, including refocusing, depth extraction, super-resolution, 3D display, etc. None of these can bypass the storage and transmission of LFIs. Since the LFI contains a huge amount of information, a common LFI occupies a large storage space (the amount of original image data captured by a Lytro Illum LF camera is $5368 \times 7728 \times 24b/8b = 124451712$ Bytes). In the transmission process, the problem of transmission efficiency has

to be considered, which makes compression coding of light field images an important research topic. At present, the compression coding for the characteristics of the light field image has a certain research basis and is constantly developing. However, the current methods for verifying the experimental results of compression coding of LFIs are directly applied to the image quality assessment (IQA) metrics of ordinary 2D images, and there is no objective quality evaluation standard based on the characteristics of LFIs. In the past a few years, researchers have proposed some subjective IQA methods for evaluating the performance of coding algorithms. But still, more methods for evaluating more types of light field distortions are lacking. At the same time, although the current LFIs dataset is numerous, most of them are used in the compression coding, deep extraction, 3D display, etc. There is no comprehensive dataset for IQA research. So currently, the number of objective LF IQA methods that relying on subjective evaluation results is small. The above factors make the research of LFI quality evaluation have a broad prospect.

The IQA system is necessary to verify the performance of the image processing system. In the field of LFI, whether it is the verification of post-processing algorithms such as compression coding and super-resolution, or the optimization of LF camera imaging, IQA has an important guiding role as a standard. Consequently, it is the main work of this paper to re-establish an LF IQA dataset containing more abundant distortion factors and use this as the basis for proposing an objective IQA method. To propose an objective LF IQA method is our ultimate goal. Therefore, the context of our overall research method is: firstly, establish the subjective IQA dataset of LFIs, and in the process of establishing the dataset, the display form of the LFI should be considered in the display and test. At the same time, the existing objective image quality evaluation algorithm is applied to the established dataset to compare the analysis algorithm for further improvement. Finally, based on the established dataset, the LFI features are fully utilized to study the objective IQA algorithm. The above steps are the main links in the current LFIs quality evaluation.

As mentioned above, the micro-lens array LF camera is currently a mainstream direction for studying LFIs. A considerable number of datasets are based on LFIs taken by micro-lens array cameras, and we are no exception. However, the LFI captured by the micro-lens array camera cannot be directly displayed, and a series of processing is required to extract the sub-aperture image (SAI). The extracted SAIs array conforms to the 4-dimensional plenoptic function, so this process is called the 4D LF decoding [8], [9].

During the process of the 4D LF decoding, the most noteworthy is the resulting lenslet image and the final 4D LFI during the decoding process. The former is an important research content of the LFI correlation image algorithm, and the SAI obtained by the latter is the main form of the external representation of the LFI. We use Fig. 1 to finally show the two types. Fig. 1 (a) is a lenslet image, the magnified part can be seen as a lens diagram showing a hexagonal arrangement; Fig. 1 (b) is a model diagram of the 4D LF structure obtained
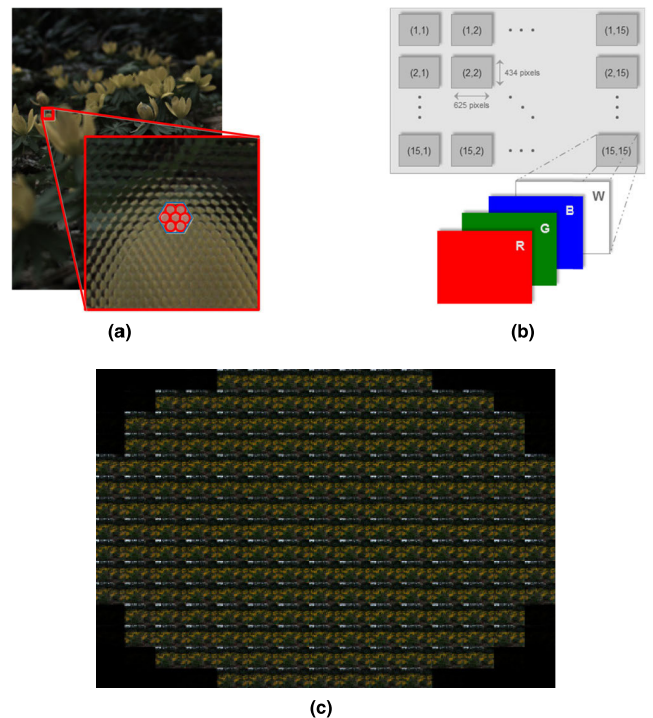

FIGURE 1. LFI: (a) Lenslet Image, (b) 4D LF structure, (c) Sub-aperture images array.

by the second generation Lytro camera. Fig. 1 (c) is the extracted SAIs arrays.

The rest of the paper is structured as follows: Section 2 describes the previous works related to LF IQA. Section 3 presents a subjective IQA scheme of LFIs and build an LF IQA dataset. In Section 4, the proposed objective metric of LFIs is introduced in detail. Section 5, the conclusion is presented.

II. RELATED WORK

Although the research on the LF IQA is in the initial stage, there are many research results in the field of subjective and objective IQA in the past two years due to the application requirements of LF processing. In the following, we will introduce them separately.

A. STUDIES ON SUBJECTIVE IQA OF LFIs

A team from the EPFL [10] in Lausanne has done a lot of work on the quality assessment methods for compressed LFIs. In [11], the authors compared two different subjective methods (common method and interactive method that allow the subject in the test to select any viewing SAI.) to evaluate the LF contents and complete the LFI encoding method to compare different compression performances. The final dataset they constructed is called the Visual Quality Assessment Light Field Image Dataset (VALID) [12]. A group of researchers from the Technical University of Rome also conducted in-depth research. In [13], they used the all-in-focus view to evaluate the quality of the LFI and compared the different 2D objective quality metrics. They used the central

TABLE 1. Comparison of existing IQA datasets of LFIs.

Dataset	Year	Content	Distortion	Number
SMART	2017	16（Lytro Illum）	JPEG，JPEG2000 HEVC intra，SSDC	256
VALID	2018	5（Lytro Illum）	8bit/10bit HEVC 8bit/10bit VP9 3 kinds of LF coding	140
Win5-LID	2018	6（Lytro Illum） ＋4（synthesis）	HEVC，JPEG2000 Linear Interpolation Nearest Interpolation	200

view of all SAIs as an all-in-focus view, providing all depth information for the LFI. Finally, an LFI visual quality dataset called SMART [14] is obtained, which contains four quality loss of compression coding distortion. In [15], researchers used an electric linear platform with a common digital camera to capture LF scenes and build the dataset, which greatly inspired the conclusions of the performance of certain 2D and 3D metrics. Recently, there was coming a new dataset with five windowed 5 degrees of freedom LFIs Database (Win5-LID) [16] from the University of Science and Technology of China.

The main features of above datasets (based on micro-lens array camera) are listed in the TABLE 1.

In order to establish a dataset that accurately reflects the true quality of the LF contents, we also refer to the work of the predecessors. The three datasets shown in TABLE 1 have their advantages, but there are some shortcomings for our follow-up research: SMART dataset distortion type is insufficient, and the evaluation scheme of the SMART dataset focuses on the process of subjective scores. It is not a traditional scoring evaluation. Instead, the reference image and the distorted image are listed on both sides of the screen to make the tester choose good and bad, and finally pass a set of statistics. The calculation system calculates the quality score of each image, but a big problem is that the display material used in its evaluation scheme is only a central SAI, which falls into the evaluation category of traditional 2D images.

The researchers of the VALID dataset did a lot of work in the early stage. As described in [11], they compared two different test viewing methods, the common method and the interactive method that allows the subject in the test to select any SAI. The ordinary evaluation method, the material of the evaluation is a pseudo-sequence composed of sub-aperture images, and this method is finally adopted by us. But the problem of the VALID dataset is it has insufficient samples.

Win5 -LID dataset directly used the composite image of the LFI to make a 3D image as the evaluation material and then passed the 3D display and the evaluation of 3D glasses,

which takes into account the 3D characteristics of the LFI, is a big improvement. The pseudo-sequence evaluation scheme used by the VALID dataset attempts to bring a 3D experience through the change of the parallax during sequence playback, but there is still a gap between the real 3D display and the advantage is more in the LF 2D features, and the ability to fuse more SAIs to maximize the quality of the complete LFI. However, this dataset contains a synthetic LFI, not the type studied in this paper.

B. STUDIES ON OBJECTIVE IQA OF LFIs

Most studies on Objective LF IQA have still limited to verification of the study, which means verifying the traditional objective IQA method after subjective experiments. Pure objective IQA algorithms for LFIs have only had a few in the past year. Some researchers like to use objective performance to compare different forms of LFIs [17]. Also as shown in [18], they proposed a 3D full reference metric for keyframes extracted from the LF video content. In 2018, researchers at the University of Cagliari in Italy began with the subjective IQA of LFI and then the objective IQA method, and proposed an objective IQA metric based on LF rendering [19]. In 2019, there is also an LF IQA method based on depth map distortion [20]. In [21], the researchers began the objective quality assessment method from the features of the LFI structure such as edge similarity and multi-order derivative features.

Although there are not many methods, the ideas are all based on the unique features of the LFI. Among them, the depth information of the LFI is of particular interest to us.

III. SUBJECTIVE QUALITY DATASET

Our research idea is to establish a subjective IQA dataset that accurately reflects the LFI quality loss and contains sufficient distortion types and sample LFIs. The preliminary evaluation of the subjective IQA scheme was conducted in the previous study [22], and the results showed that our scheme
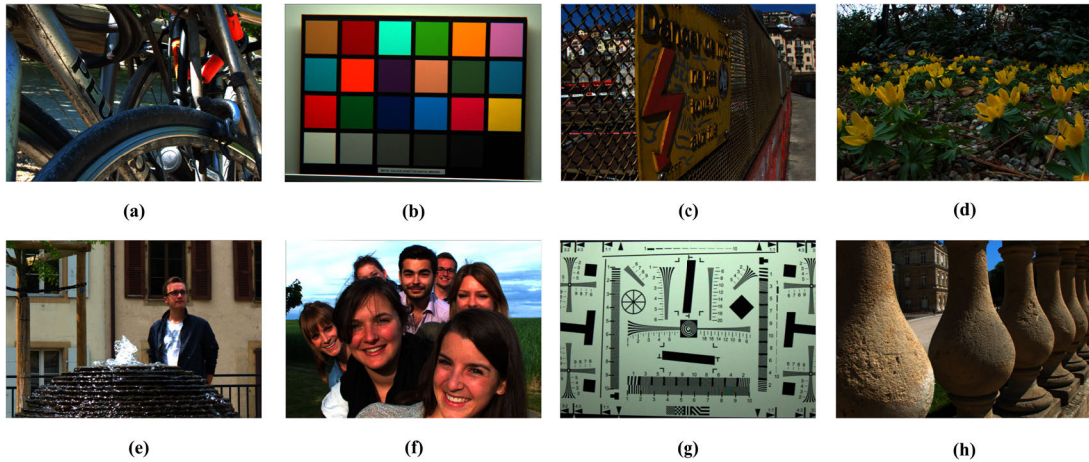
FIGURE 2. Contents of source LFIs used in the database: (a) Bike, (b) Color Chart 1, (c) Danger de Mort, (d) Flowers, (e) Fountain & Vincent 2, (f) Friends, (g) ISO Chart 12, (h) Stone Pillars Outside.
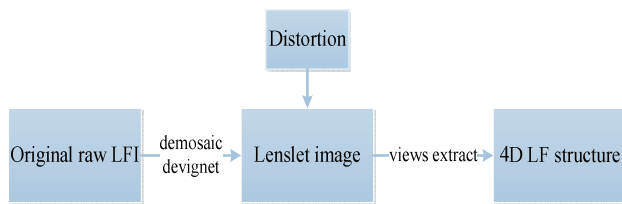

FIGURE 3. Flowchart of 4D LF decoding process.

was feasible. However, the scheme in [22] still has shortcomings such as low sample data, so the scheme is improved and the experience is summarized in [23]. Below is our complete subjective IQA method.

A. DATA PREPARING

All of the source LFIs we selected are from EPFL LFIs database [10]. 8 LFI that covered a wide range of scenes (including people, objects, plants, single and complex colorfulness or textures) were selected, as shown in Fig. 6. In these images, except for Fig. 2 (b) and (g), the rest LFIs clearly show the difference between the far and near depths of field, because such images can better reflect the unique characteristics of the LFI in terms of 3D and focus.

Each original LFI was decoded to obtain lenslet image using the function *LFDecodeLensletImageSimple* in the MATLAB toolbox LFToolBox0.4 [24]. The distortion was added to the lenslet image (Fig. 1 (a)) for considering most of LFIs processing methods are directed to the lenslet image, especially various compression methods. The distortion added in this step can also be reflected in the final SAI. And the subsequent process of SAIs extraction and other processes will still produce a certain distortion, and adding distortion in this intermediate link can also play the role of control variables. The specific operation process is shown in Fig. 3.

In order to fully reflect the quality of the LFI, all the SAIs with good decoding effect should be used as the
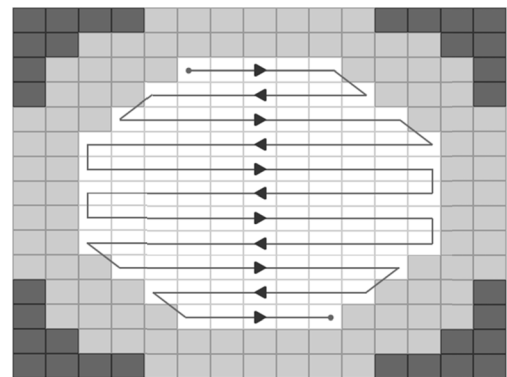

FIGURE 4. Views order in pseudo sequence video.

display material. At the same time, the distortion of the edge SAIs should be avoided, so the edge and corner SAIs should be discarded. We used the white area in Fig. 4 as the final selected area. A pseudo-sequence method using video instead of an image is used to display the LFI. Each selected SAI is used as a frame of video to form a pseudo-sequence video so that each of the all-in-focus SAI is fully displayed. This form of the display can also reflect the difference in parallax between the SAIs. For the continuity of parallax, the sequence of SAIs sequence combinations in Fig. 4 was employed. According to [25], when the frame rate is greater than 15 frames, the subjective perception of the human vision will not be much different, and it is at a relatively high level, and after the frame rate is less than 15 frames, the subjective feeling of the human vision will drop sharply. Therefore, the frame rate of the final synthesis into a lossless video sequence is 15 frames per sec, and the pseudo-sequence playback time per segment is about 6.5 sec.

We chose Gaussian blur, JPEG2000, JPEG, motion blur, and white noise these five types of distortion to add into the lenslet image. The 5 types of distortion can reflect a broad range of random noise, image impairments, smoothing and structured distortions [26]. These distortions cover all
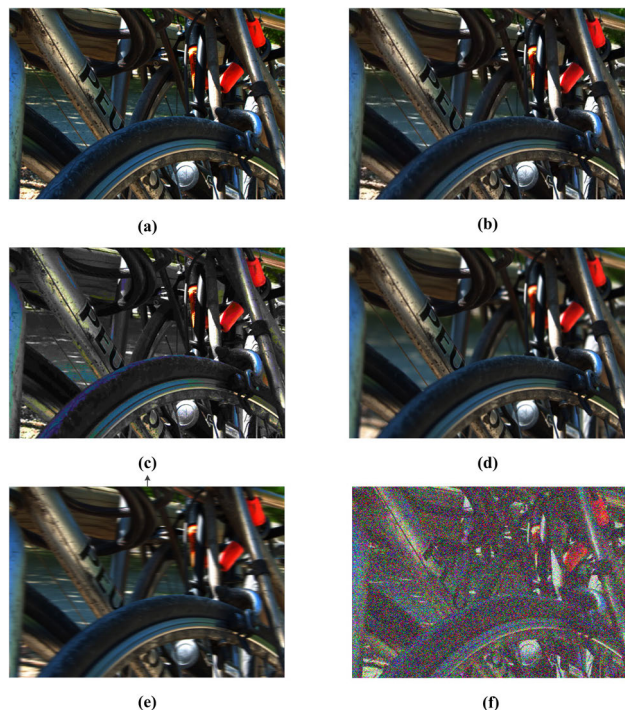
FIGURE 5. The source LFI and its corrupted versions: (a) Source LFI, (b) JPEG, (c) JPEG 2000, (d) Gaussian blur, (e) Motion blur, (f) White noise.

processes of LF. Each type of distortion has six levels, which covered from a small degree to high levels of impairment. After adding distortion, each reference image can generate 30 distorted LFIs. Finally, a total of 240 distorted LFIs was derived from 8 sources LFIs. Fig. 5 shows a central view of the source reference image *Bike* and its associated distortion types after 4D LF decoding.

Here's how to add various types of distortion:

- JPEG2000 compression: The MATLAB *imwrite* command was used to create JPEG2000 compressed LFIs with a bitrate ranging from 0.04 bits per pixel (bpp) to 2.40 bpp.
- JPEG compression: The MATLAB *imwrite* command was used to create JPEG compressed LFIs with bitrates from 0.132 bpp to 2.34 bpp.
- Gaussian blur: Each color channel (R, G, and B) used the same Gaussian kernel, and the standard deviation σ was used to generate Gaussian blur LFIs. In this paper, the values of σ were set as 0.5, 2, 4, 5, 10, and 20.
- White noise: The standard deviation σ was added to the RGB components of each LFI, and the values of σ were set as 0.05, 0.1, 0.3, 0.5, 1, and 2.
- Motion blur: The MATLAB *fspecial* and *imfilter* commands were employed to create Motion blurred LFIs. Setting the motion angle to 0, and the motion length parameters were set as 10, 20, 60, 100, 150, and 200.

B. TESTING ENVIRONMENT AND EXPERIMENT

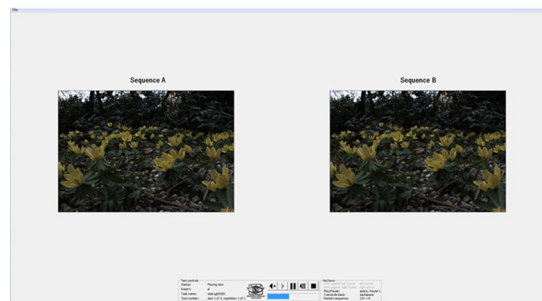To create a natural viewing environment, the distance between the tester and the monitor was not specified but

FIGURE 6. MSU perceptual video quality tool.

determined by the test subject himself, and the principle is to make them feel comfortable. To avoid external interference, a quiet test environment is required, and the influence of external illumination is also avoided. The entire subjective IQW testing was performed in a laboratory environment with a 21-inch display. The display resolution is 1920×1080 pixels. 20 subjects (including 15 males and 5 females, aged 20 to 36 years old) participated in the experiment.

Each tester must go through a few minutes of training before the test: watching the distortion performance of an LFI according to the distorted level and understanding the scoring basis of the evaluation. Then the formal test is carried out. The Double Stimulus Continuous Quality Scale (DSCQS) method is adopted according to the ITU-R BT.500 [27] for the subjective evaluation. The reference sequences and the distortion sequence are randomly arranged on both sides of the screen and played simultaneously. The test platform used here is a perceptual video quality evaluation tool developed by Moscow State University (MSU) [28]. Fig. 6 shows the playback interface when testing image Flowers.

Each pseudo-sequence was played twice, and after two plays, the subject simultaneously scores the sequence displayed on both sides. The subjects were asked to rate the quality of both test videos on a discrete scale from 5 (excellent) to 0 (bad). If the subject has not seen it, he or she can repeat it. Each type of distorted LFIs is tested separately, but images of different distortion levels within each distortion type are randomly played. After all distortion types have played for about 10 to 15 minutes, there will be two minutes break before testing the next distortion type. The complete test takes more than an hour. The data from the subjective IQA testing should be analyzed and excluded the abnormal data. We followed the ITU-R BT.500 [27] guidelines about the observers screening procedures to exclude the outliers. After hypothesis testing elucidated in [27], we excluded all the scores of one subject. The scores of the remaining 19 subjects were used for the final calculation. Eventually, the MOS of each LFI was calculated by Eq. (1):

$$MOS_j = \frac{1}{N} \sum_{i=1}^{N} m_{ij}, \qquad (1)$$

where N is the number of participants and m_{ij} is the score for stimulus j by participant i.
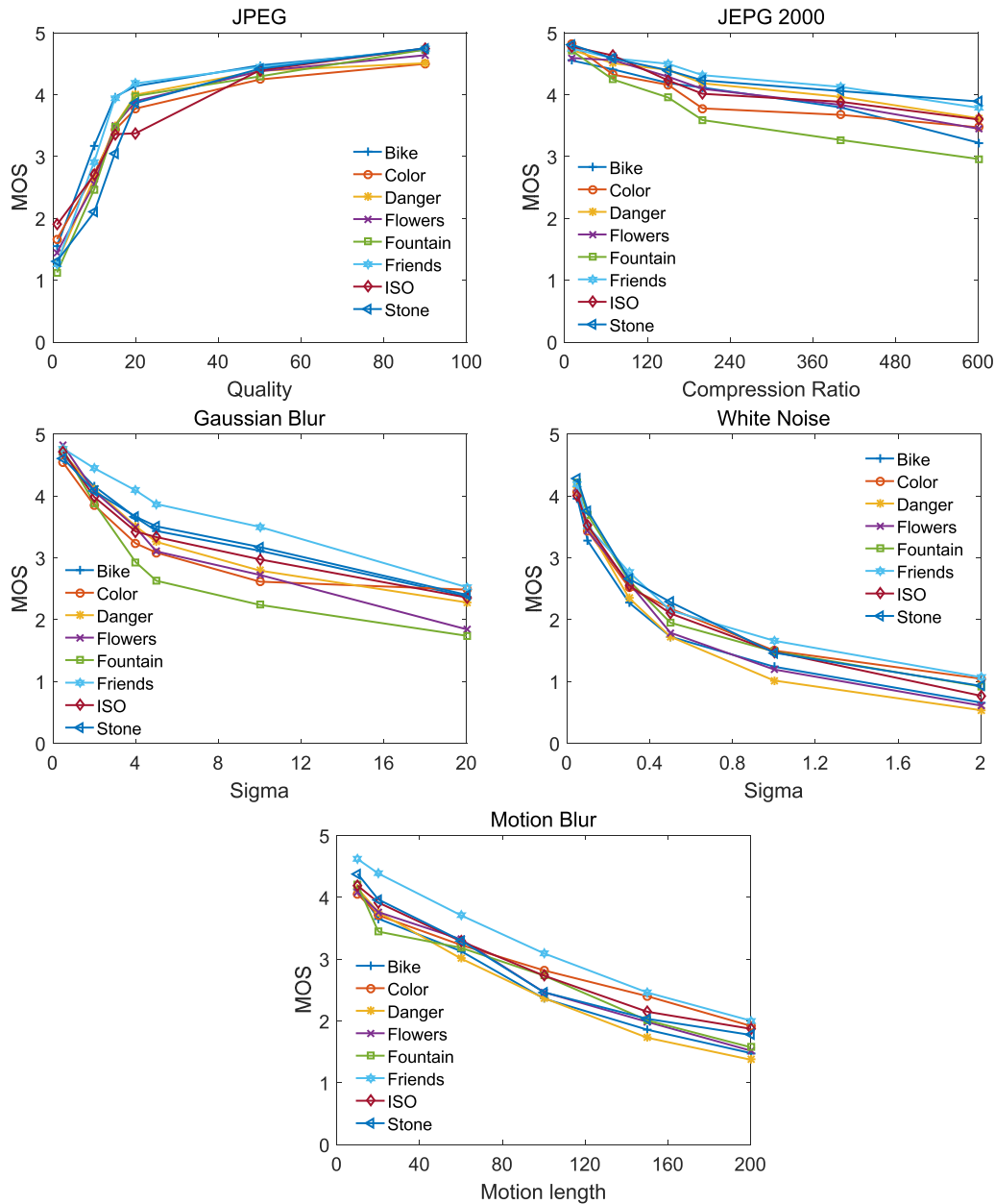
FIGURE 7. MOS vs. distortion: Classify the MOS distribution of all images by distortion type.

C. CRESULTS AND ANALYSIS

The MOS values of all distorted LFIs of the dataset are shown in Fig. 7 in the form of the line chart.

Some important information about the performance of the distortion can be seen in Fig. 7. For both compression distortions, JPEG 2000 has a higher MOS value due to better compression, even higher than all other distortions, and the resulting visual damage needs to be extremely high. The compression ratio can be significantly reflected. For JPEG, when the compression ratio is increased to a certain extent (the quality parameter is less than 15), severe color distortion is generated, which is also the reason why the lines in the JPEG line graph rise significantly before the parameter 15.

In addition to JPEG 2000, Gaussian blur also has a higher MOS score. As previously mentioned, the test pseudo-sequence video consists of 97 SAIs of the 4D LF structure. There is parallax between these SAIs, so the pseudo-sequences we make will look somewhat unstable. Gaussian blur can reduce jitter, which creates an illusion that the more ambiguous the sequence looks more stable, which is also reflected by the tester during the test, so we can give the conclusion that Gaussian blur will affect the 3D characteristics of the LFI. For white noise distortion, the MOS curve is relatively smoother and grading significantly, because the test object can easily distinguish the noise difference, and the visual quality damage of the white noise image to the LFI
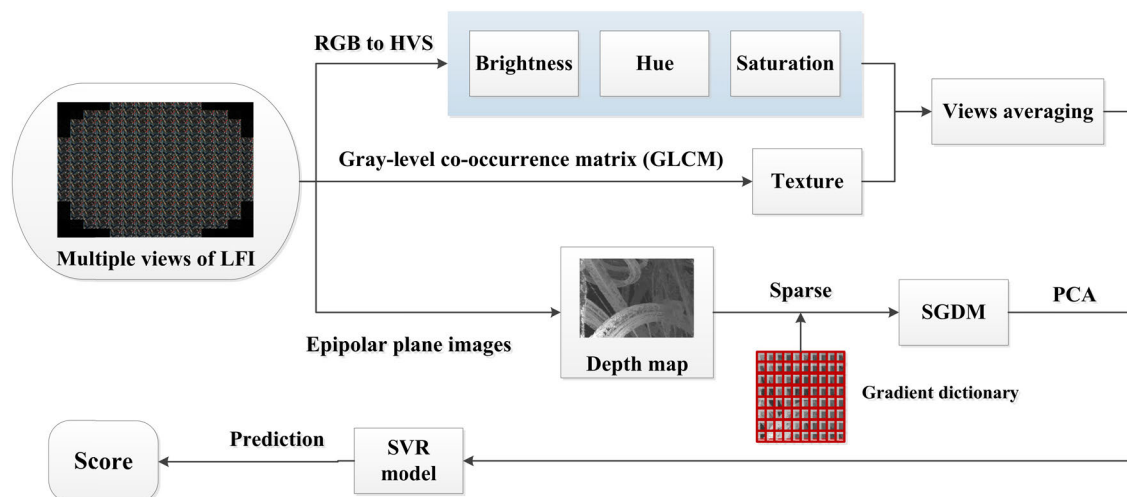
FIGURE 8. The framework of the proposed metric.

is most obvious. Finally, motion blur, which is specifically added to simulate distortion caused by camera or subject shake, also has a relatively uniform MOS score distribution.

Then we analyze each angle of the LFI. From the perspective of a particular LFI with a smaller depth of field, simpler textures and colors exhibit less visual quality than blur or noise distortion, while compression distortion is the opposite. For a large depth of field image with obvious 3D characteristics, the blur distortion will have a relatively large effect, while the compression distortion, especially the better compression method, will have less impact.

IV. PROPOSED OBJECTIVE QUALITY METRIC

A. FRAMEWORK

The acquired LF content can be presented to the user in a variety of ways, such as LF displays, helmets, refocused images, and multi-views arrays. Different ways of visualizing the content of the LF are different, and the degree of attention to the features of the LF is also different. For example, visualization applications for LF refocusing will focus more on the depth of field of the LF content. Therefore, it is especially important to select and extract the appropriate and reasonable visual features of the LFI for the objective IQA of LFIs. In view of the fact that LFI combines the features of 2D and 3D characteristics, the following features are what we consider to extract:

- 2D characteristic: Inspired by the human visual system, brightness, hue, and saturation as color information features are also an important part of the perceived quality of the scene, so these three basic image features are included in the primary consideration. In addition to the basic features, image texture, as an important indicator of spatial perception information, plays an important role in the interaction of LF content and structure, and is also an important factor we consider. The 4D LF structure can extract dense SAIs. In order to better reflect

this feature, the three basic image features and texture features need to be processed by multi-view mean.

- 3D characteristics: In the analysis part of subjective quality evaluation, we know that the larger the depth of field of the LFI, the stronger the 3D visual perception. Depth of field is an important indicator of the 3D characteristics of the LF. Not only that, but the depth of field is also an important manifestation of the refocusing characteristics of the LFI. In the processing steps of virtual view reconstruction and LF 3D synthesis, the extraction of LF depth information is an important step. So here we choose depth map as the only feature to represent 3D features of LF. The depth map of LFI is obtained by using the Epipolar Plane Images (EPI) method.

After extracting these two main features group, they are combined as a complete LF feature represent an image, and then input into the SVR model together with the subjective IQA score for training and learning, and the final prediction model is obtained. The overall framework of the proposed method is shown in Fig. 8.

As can be seen from the framework, the depth information representing the 3D feature of the LF needs to be sparse because it is too large. We trained a gradient dictionary and employed it to obtain the sparse representation of depth map of each LFI and called it the sparse gradient depth matrix (SGDM). After summing-up and principle component analysis (PCA) we can get a set of one-dimensional data which represent the depth feature. Then all of the features are sent into the SVR model to predict the final objective assessment score.

B. FEATURES EXTRACTION

1) BASIC VISUAL FEATURES

We converted the RGB values of each SAI to the HVS coordinates. Three matrixes represent brightness hue and saturation can be obtained by the processing. We used M to present each

matrix. The final values of brightness, hue, and saturation of LFI can be calculated by:

$$V_x = \frac{1}{kl} \sum_{k=1}^{K} \sum_{l=1}^{L} mean(M_x(k, l)), \qquad (2)$$

where k and l correspond to k-th and l-th individual SAI, K and L correspond to the amount of SAIs in each horizontal and vertical direction, respectively. ($K = L = 15$). x refers to brightness, hue or saturation. $mean(M_x(k, l))$ refers to averaging SAIs. Finally, the V corresponding to different x was combined to get the basic visual features $[V_1, V_2, V_3]$.

There is a phenomenon that the larger the value kl, the better the experiment result. That is when we used all of the SAIs of an LFI we can get the best effect. On the contrary, the experiment result was the worst with only one SAI.

2) LF TEXTURE

To better describe the 2D structural characteristics of the LFI, in addition to the previous basic features, the texture feature is also an important feature of our choice. It is a kind of feature that describes the image quality which is very suitable for SVM learning. Here we choose the classical texture feature extraction method which is simple and easy to implement, the gray-level co-occurrence matrix (GLCM).

The GLCM of an image can reflect the comprehensive information about the direction, adjacent interval, variation amplitude, etc. which is the basis for analyzing the local pattern of the image and its arrangement rules. The essence is to count the number of simultaneous occurrences of pixels (x+Dx,y+Dx) whose distance is d and whose gradation is j from the pixel whose gradation is i (the position is (x, y)) in the image. The number p (i, j, d, θ) can be calculated by:

$$\begin{aligned} p(i, j, d, \theta) &= [(x, y), (x + Dx, y + Dx)|f(x, y) \\ &= i, f(x + Dx, y + Dy) = i], \quad (3) \end{aligned}$$

where x, y $= 1, 2, \ldots, N - 1$ is the pixel coordinate in the image, i, j $= 1, 2, \ldots, L - 1$ is the gray level, Dx, Dy is the position offset, the d is used to generate the step size of the GLCM, θ generation direction can take four directions of 0°, 45°, 90°, and 135° to generate four different directions of the co-occurrence matrix. To make its eigenvalues unaffected by the range of regions, it is also necessary to normalize this GLCM.

The GLCM of each SAI of an LFI is obtained by the MATLAB function graycomatrix. Here, take the GLCM in different directions (0, 45, 90, 135 degrees), and the GLCM in each direction is calculated by 10 loops and normalized (calculation contrast, inverse gap, entropy, autocorrelation), and then the mean and variance are taken as the characteristics of the final extraction. Of course, the same processing of SAIs averaging like Eq. (2) needs to be executed. An 80-dimensional feature of an LFI called [G] should be gotten after the extraction.
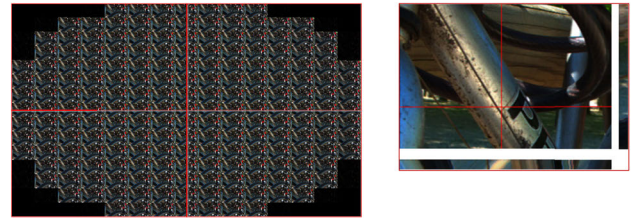

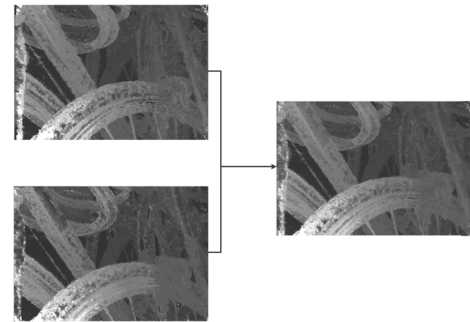**FIGURE 9.** The horizontal and vertical EPIs.

FIGURE 10. Depth map of LFI.

3) DEPTH MAP

We used the method from [29] to extract the depth map feature of LFI. All SAIs of an LFI are utilized to estimate the depth map according to EPI image analysis. As the previous section mentioned, 4D LF contained both spatial and angular information about LF. In the function $L(x, y, u, v)$, (x, y) can be regarded as a set of cameras and (u, v) is the focal plane. Through collecting the SAI pixels at the fixed spatial coordinate x and an angular coordinate u (or y and x), horizontal Epipolar Plane Image (or vertical Epipolar Plane Image) can be produced [30]. As shown in Fig. 9.

The EPI can reflect the 3D information of the scene of the LFI because the slope of the line is closely related to the depth of the 3D space contained in the scene of the LFI [31], [32]. The depth of the scene can be calculated by the following:

$$D_{(i)} = f / (1 - tan \propto_{(i)}), \qquad (4)$$

where i is an enumeration value, which represents horizontal or vertical when i is 0 or 1, respectively. $D_{(i)}$ and f represent scene depth and camera focal length, and $\propto_{(i)}$ indicates the slope of a certain sample in the EPI. Here, the calculation of the slope $\propto_{(i)}$ of the specific oblique line in the EPI also needs to be optimized. The optimal slope with the lowest cost can be found by calculating the decision model considering the limitation of the angular resolution of the LFI. After all points and angles are calculated, two depth maps can be generated from the horizontal and vertical EPIs, respectively. Since the calculated horizontal and vertical disparity and the two depth maps have the same weight, the depth map is achieved by weighting the average of the two depth maps and limiting the depth value between 0 and 255. As shown in Fig. 10, at the left column, the top row is the depth map obtained
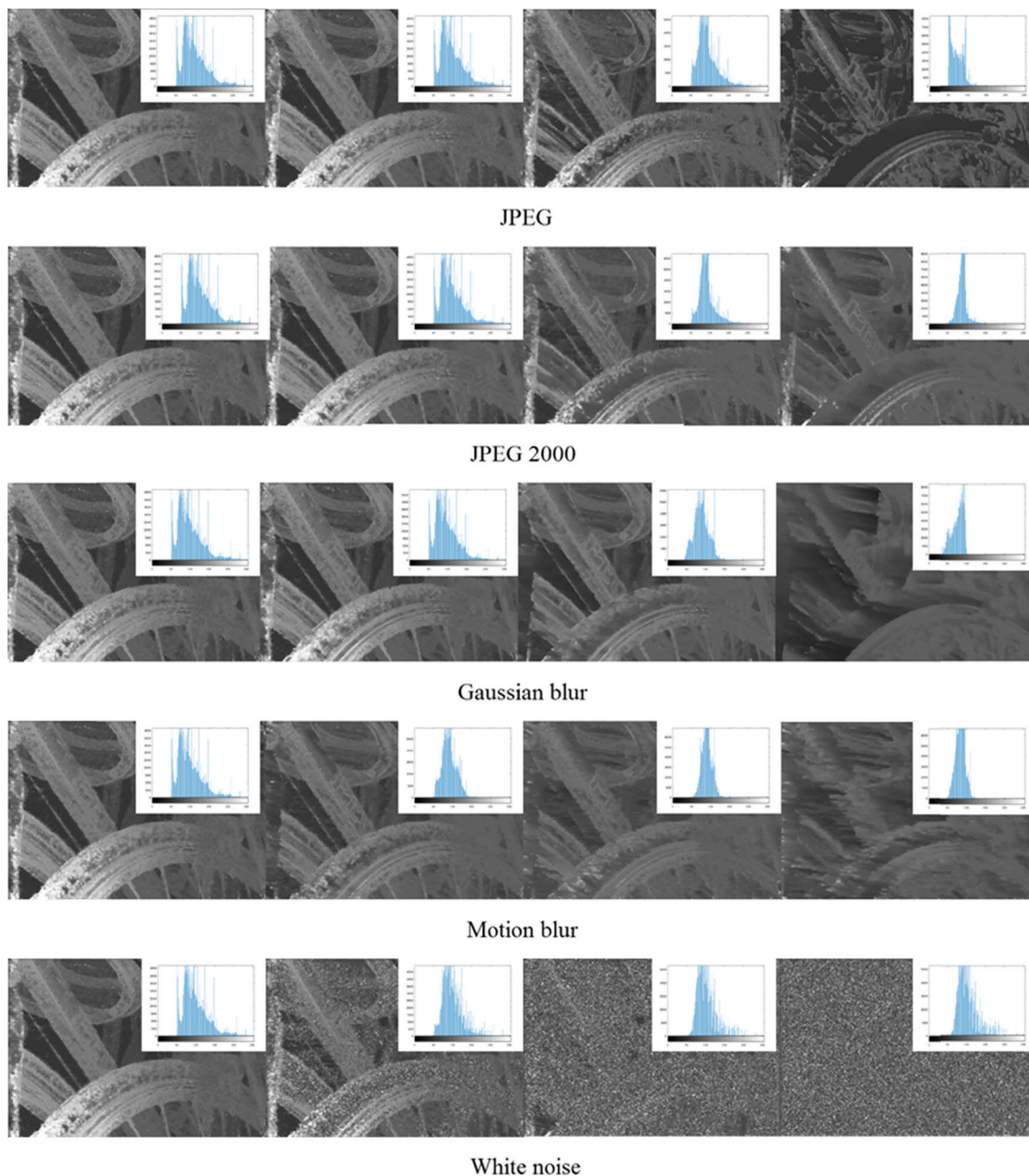
FIGURE 11. Depth map of LFI- *Bike* and their grayscale histogram.

by the horizontal EPI, and the lower row is the depth map obtained by the vertical EPI.

Fig. 11. shows the final depth map of the partially distorted image derived from the LFI-Bike and the gray histogram corresponding to the depth. We selected four distortion level LFIs for each type of distortion to display. As can be seen from the figure, the histogram shows the variation of each distortion. For example, the histogram distribution of each distortion converges to a certain point as the distortion deepens. The change of the histogram branch and the convergence position in different distortions are different.

This indicates the feasibility of the LFI depth map to represent image quality. However, the mere histogram as the information representing the depth map is also inaccurate. The data volume of the complete depth map is too large. The direct use of the depth map data will increase the computational complexity, and the final training result may also be over-fitting. So, the depth map needs to sparse representing to get the representative depth map. Here we choose the K-SVD method [33], through a training gradient dictionary to sparse the depth map. The following is a specific processing method on the depth map of the LFI.

We chose 10 depth maps of different LFIs from our dataset randomly to train the depth gradient dictionary. Then the gray-scale depth images and the absolute values are used to build 10 gradient magnitude maps [34]. The gradient magnitude can be calculated by:

$$GM = [G_x^2(i,j) + G_y^2(i,j)]^{1/2}, \qquad (5)$$

where $G_x(i,j)$ and $G_y(i,j)$ are directional derivatives in the horizontal x and vertical y directions in sample area (i, j). Then we divided each gradient magnitude map into nonoverlapped local patched of 8*8 and select 2500 patches and build a matrix denoted as $N = [n_1, n_1, \cdots, n_{2500}] \in \mathbb{R}^{64 \times 2500}$, where n is the column vector vectorized by each patch. At this time, the sparse coefficient matrix $C \in \mathbb{R}^{64 \times 2500}$ can be calculated by the dictionary learning formula represented by Eq. (6):

$$N = UC, \qquad (6)$$

$$min \|N\text{-}UC\|_2 \quad s.t \ \forall k \in [1, 2500], \ \|C_k\|_0 \leq \tau, \qquad (7)$$

where $\|\cdot\|_2$ is the l_2 norm, $\|\cdot\|_0$ is the l_0 norm meaning the number of nonzero elements of a vector. This means when the error is the minimum, each atoms column after column and coefficients have been updated. Then the dictionary $DIC = [U_1, U_2, \ldots U_{10}] \in \mathbb{R}^{64 \times 2500}$ can be gotten. Where U is a dictionary obtained by Eq. (6), and 10 dictionaries obtained from 10 gradient amplitude maps are used to sparse the depth map to form the final depth gradient dictionary.

The following is using the dictionary to sparse each depth map. The depth map of size 625×434 is divided into 4212 patches of 8×8 size and vectorized into column vectors to build matrix Z. Using the trained dictionary DIC, the sparse representation matrix can be obtained by followed:

$$SGDM = argmin \frac{1}{2} \|Z - (DIC)(SGDM)\|_2 + \lambda \|SGDM\|_1, \qquad (8)$$

where $\|\cdot\|_1$ is the l_1 norm and λ is positive parameter balancing the importance of error term and constraint term. The *SGDM* still have so many dimensions that may cause high redundancy and high correlation between columns. After summing up the values of each column of *SGDM* and principal component analysis (PCA), the linear feature represented depth map can finally be obtained. We named it [*DM*].

C. EXPERIMENTS AND RESULTS

We took 80% of the dataset for training and the rest 20% for predicting. LIBSVM-3.22 was employed for the SVR model [35], [36] and the Radial Basis Function (RBF) k(A, A_n) as a kernel function.

$$f(x) = \sum_{n}^{m} W_n \times k(A, A_n) + b, |f(x) - MOS| \qquad (9)$$

$$= \begin{cases} 0, & |f(x) - MOS| \in \varepsilon \\ |f(x) - MOS| - \varepsilon, & otherwise, \end{cases} \qquad (10)$$
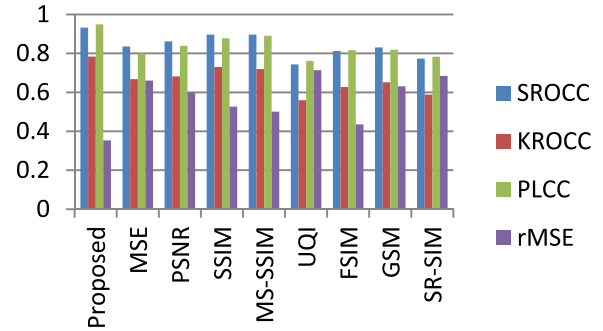

FIGURE 12. Histogram of performance of IQA metrics.

where $f(x)$ is the objective score. When the error of $f(x)$ and *MOS* is less than a fixed minimum value, the needed model can be obtained. We applied the grid search method to do the parameters optimizing. By feeding the below sample features into the trained SVR regression model then looped the model program 1000 times to take the median, the predicted objective score can be obtained. The reason we chose to loop the model 1000 times is that the median value of 1000 is sufficient to eliminate the instability. Higher looping times do not ensure better prediction results, and it will cause the system to run longer and increase computational complexity.

$$F = [V_1, V_2, V_3, G, DM], \qquad (11)$$

$$Objective \ Score = SVR\_model \{F\}, \qquad (12)$$

The performance of the proposed method was evaluated by four wildly used performance criteria Pearson linear correlation coefficient (PLCC), Kendall's rank correlation coefficient (KRCC), Spearman rank correlation coefficient (SRCC) and Root mean-squared error (RMSE). The lower value of RMSE and the higher value of SCRCC, KRCC, and PLCC indicate a better performance of the testing metric. Light field quality assessment usually involves resolution in terms of both spatial and angular dimensions, spatial quality, and angular consistency assessment [3]. For effect comparison, the proposed LFI IQA dataset is also employed as testing platform for eight classical full reference IQA metrics including MSE [37], PSNR [38], SSIM [39], MS-SSIM [40], UQI [41], FSIM [42], GSM [43] and SR-SIM [44]. The Objective experiment results are given in Fig. 12 and TABLE 2.

In the above table, the best performance of the proposed method and the best performance of other methods are bolder. In general, the full reference image quality evaluation method is better than the no-reference method. The method we proposed is a no-reference IQA metric, but it can be seen from the table that even if compared with some of the best performing full reference objective IQA metrics, the proposed method still exceeds many indicators (PLCC, SROCC, KROCC are closer to 1, and RMSE values are lower).

From the results, we can see the proposed metric is much better than other methods on each indicator. We sort the performance and obtain results that Proposed > MS-SSIM > SSIM > PSNR > MSE > GSM > FSIM > SR-SIM > UQI.

TABLE 2. Data of performance of IQA metrics.

	SROCC	KROCC	PLCC	rMSE
Proposed	**0.933**	**0.7838**	**0.9494**	**0.3535**
MSE	0.8353	0.6682	0.7997	0.6608
PSNR	0.8618	0.6816	0.8393	0.5984
SSIM	**0.8967**	**0.7306**	0.8779	0.5269
MS-SSIM	0.8961	0.7201	**0.8904**	0.5009
UQI	0.7436	0.5597	0.7609	0.7141
FSIM	0.8128	0.6275	0.8165	**0.4354**
GSM	0.8307	0.6516	0.8191	0.6314
SR-SIM	0.7736	0.5886	0.783	0.6846

TABLE 3. Data of performance of IQA metrics (SMART dataset).

	SROCC	KROCC	PLCC	rMSE
Proposed	**0.8917**	**0.6986**	**0.9106**	**0.8392**
PSNR	0.7733	0.5546	0.7545	1.8525
SSIM	0.5783	0.4256	0.5323	2.3895
MS-SSIM	0.8578	0.6627	0.8798	1.3418
FSIM	**0.8775**	**0.6839**	**0.8933**	**1.2689**
GSM	0.8600	0.6564	0.8698	1.3929
VSI	0.8733	0.6716	0.8566	1.4564
VIF	0.8657	0.6706	0.8766	1.3585

Obviously, some of the mainstream classical full reference IQA metrics have a certain degree of reliability. PSNR and SSIM used for evaluating the quality of LFI are feasible. But for further improvement, the using of the characteristics of the LF is extremely necessary. Our method also proves this point. The combining of 2D features and 3D features is a significant characteristic of LF itself. This points the direction for us to continue exploring ways to combine more LF features such as multi-views disparity and so on.

In addition to the datasets we created, we conducted an objective experiment on the proposed method on the SMART dataset, as we mentioned in Section 2. The SMART dataset only collects compression coding distortion (SSDC, HEVC Intra, JPEG, JPEG2000), which cannot enough reflect the loss of LFIs quality. But the sample data quantity (256 distorted LFIs) of the dataset is suitable for the proposed method. So we conducted an objective experiment based on this dataset. Since the dataset does not give specific experimental data of the existing classical objective IQA metrics, we applied the full-reference objective IQA metrics mentioned in the paper of the dataset (PSNR, SSIM, MS-SSIM, FSIM, GSM, VSI [45], and VIF [46]) as a comparative test to compare the results of the proposed method. We still experimented the experiments several rounds and chose the best and worst results. In the following TABLE 3, we show the performance of the method and the best performance of other metrics.

Fig. 13 shows a visual comparison of the proposed method with other objective metrics. Since the RMSE indicator of the proposed method is significantly better than the other methods (the value is much lower), only the first three indicators are listed.

The above experimental results reflect that the performance of the proposed method is much better than some existing classical objective IQA methods. These results prove that the LF combination features can accurately reflect the LFI quality. We have achieved good results only when the
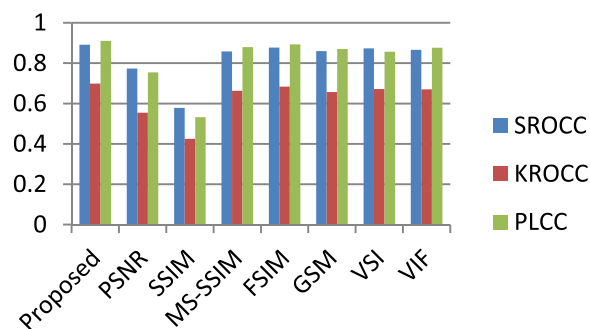

FIGURE 13. Histogram of performance of IQA metrics (SMART dataset).

combination of the 2D and 3D features of the LFI is employed. When combining more LFI features, such as refocusing, the angle parallax features, etc., in theory, there will be further improved. Of course, the 2D and 3D characteristics of the LFI are still the most important features, which is the main consideration when we make feature selection.

V. CONCLUSION

The micro-lens array LF camera is the most widely used LF photographing device because of its portability and operability. Various types of post-processing of LFIs based on micro-lens array cameras have developed rapidly in recent years, but the lack of LF IQA standards to evaluate the quality of LFIs has gradually become a vital issue limiting the development of LF. This paper deeply studies the work of predecessors and gives a complete research plan for IQA of LFIs.

We propose a subjective IQA scheme of LFIs. Using the representation of the distortion and the pseudo-sequence of the LF SAIS in the 4D decoding process, the subject was called to conduct a subjective test. Through subjective experiments, we established a new LF IQA dataset containing a total of 240 samples of 5 types of distortion. Then we propose a no-reference objective IQA method based on SVR model to evaluate the quality of our dataset. Combining the 2D and 3D features of the LFI, more kinds of features were extracted, and finally, multiple cycles of regression experiments were

carried out. We also applied the proposed model to a third-party dataset for related experiments. The results of these subjective and objective experiments show that the quality of LFIs can be well evaluated through the proposed method, and the performance can be greatly improved compared to other metrics.

The IQA of LFIs is a relatively new research field, which has the advantage of great development space but also has the disadvantage of less experience. Although this paper gives a relatively systematic set of research programs, we should also see some shortcomings. For example, the number of sampled subjective quality assessment dataset is still not enough. And the objective quality evaluation method we proposed can be more open in terms of LF characteristics to achieve better results.

REFERENCES

[1] R. Ramesh, and J. Tumblin, *Computational Photography: Mastering New Techniques for Lenses, Lighting, and Sensors*. Natick, MA, USA: AK Peters, 2009.

[2] C. Conti, L. D. Soares, and P. Nunes, "HEVC-based 3D holoscopic video coding using self-similarity compensated prediction," *Signal Process. Image Commun.*, vol. 42, pp. 59–78, Mar. 2016.

[3] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field image processing: An overview," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 926–954, Oct. 2017.

[4] G. Arun, "The light field," *J. Math. Phys.*, vol. 18, pp. 51–151, Apr. 1939.

[5] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*. Cambridge, MA, USA: MIT Press, 1991, pp. 3–20.

[6] L. Marc and P. Hanrahan, "Light field rendering," in *Proc. 23rd Annu. Conf. Comput. Graph. Interact. Techn.*, 1996, pp. 31–42.

[7] *Lytro*. Accessed: Jun. 5, 2018. [Online]. Available: https://www.lytro.com/

[8] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, Jun. 2013, pp. 1027–1034.

[9] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Linear volumetric focus for light field cameras," *ACM Trans. Graph.* vol. 34, no. 2, pp. 1–15, 2015.

[10] *Rerabek, Martin, and Touradj Ebrahimi*. New Light Field Image Dataset. CONF. 2016.

[11] I. Viola, M. Řerábek, and T. Ebrahimi, "Comparison and evaluation of light field image coding approaches," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 1092–1106, Oct. 2017.

[12] I. Viola and T. Ebrahimi, "Valid: Visual quality assessment for light field images dataset," in *Proc. 10th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, Cagliari, Italy, May/Jun. 2018, pp. 1–3.

[13] 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.

[14] P. Paudyal, R. Olsson, M. Sjöström, F. Battisti, and M. Carli, "SMART: A light field image quality dataset," in *Proc. 7th Int. Conf. Multimedia Syst.*, 2016, p. 49.

[15] V. K. Adhikarla, M. Vinkler, D. Sumin, R. K. Mantiuk, K. Myszkowski, H.-P. Seidel, and P. Didyk, "Towards a quality metric for dense light fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, 3720–3729.

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

[17] R. J. S. Monteiro, N. M. M. Rodrigues, S. M. M. Faria, and P. J. L. Nunes, "Light field image coding: Objective performance assessment of Lenslet and 4D LF data representations," *Proc. SPIE*, vol. 10752, Sep. 2018, Art. no. 107520D.

[18] R. R. Tamboli, P. A. Kara, A. Cserkaszky, A. Barsi, M. G. Martini, B. Appina, S. S. Channappayya, and S. Jana, "3D objective quality assessment of light field video frames," in *Proc. 3DTV-Conf., True Vis.-Capture, Transmiss. Display 3D Video (3DTV-CON)*, Helsinki, Finland, Jun. 2018, pp. 1–4.

[19] C. Perra, "Assessing the quality of experience in viewing rendered decompressed light fields," *Multimedia Tools Appl.*, vol. 77, no. 16, pp. 21771–21790, 2018.

[20] 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.

[21] Y. Tian, H. Zeng, L. Xing, J. Chen, J. Zhu, and K.-K. Ma, "A multi-order derivative feature-based quality assessment model for light field image," *J. Vis. Commun. Image Represent.*, vol. 57, pp. 212–217, Nov. 2018.

[22] L. Shan, "Subjective evaluation of light field images for quality assessment database," in *International Forum on Digital TV and Wireless Multimedia Communications*. Singapore: Springer, 2017, pp. 267–276.

[23] L. Shan, D. Liu, P. An, and X. Huang, "Research on subjective quality assessment of light field images," in *Proc. VII Int. Conf. Netw., Commun. Comput.*, 2018, pp. 278–282.

[24] D. Dansereau. *Light-Field Toolbox for MATLAB*. Accessed: Mar. 13, 2018. [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/49683-light-field-toolbox-v0-4

[25] Y.-F. Ou, T. Liu, Z. Zhao, Z. Ma, and Y. Wang, "Modeling the impact of frame rate on perceptual quality of video," in *Proc. 15th IEEE Int. Conf. Image Process.*, San Diego, CA, USA, Oct. 2008, pp. 689–692.

[26] G. Yue, C. Hou, and K. Gu, "Subjective quality assessment of animation images," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, St. Petersburg, FL, USA, Dec. 2017, pp. 1–4.

[27] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, BT, Recommendation ITU-R, 2002.

[28] Moscow State University. *MSU Perceptual Video Quality Tool*. Accessed: Oct. 25, 2018. [Online]. Available: http://www.compression.ru/video/quality_measure/perceptual_video_quality_tool_en.html

[29] X. Huang, P. An, L. Shan, R. Ma, and L. Shen, "View synthesis for light field coding using depth estimation," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, San Diego, CA, USA, Jul. 2018, pp. 1–6.

[30] 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.

[31] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sydney, NSW, Australia, Dec. 2013, pp. 673–680.

[32] H.-G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.-W. Tai, and I. S. Kweon, "Accurate depth map estimation from a lenslet light field camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1547–1555.

[33] P. Irofti and B. Dumitrescu, "Overcomplete dictionary design: The impact of the sparse representation algorithm," in *Proc. 20th Int. Conf. Control Syst. Comput. Sci.*, Bucharest, Romania, May 2015, pp. 901–908.

[34] J. Yang, P. An, J. Ma, K. Li, and L. Shen, "No-reference stereo image quality assessment by learning gradient dictionary-based color visual characteristics," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Florence, Italy, May 2018, pp. 1–5.

[35] M. Narwaria and W. Lin, "SVD-based quality metric for image and video using machine learning," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 347–364, Apr. 2012.

[36] Y. Zhang, X. Jin, and Q. Dai, "A SVR based quality metric for depth quality assessment," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Montreal, QC, Canada, May 2016, pp. 2567–2570.

[37] M. Gaubatz and S. Hemami. (2016). *MeTriX MuX Visual Quality Assessment Package*. [Online]. Available: http://foulard,ece.cornell.Edu/gaubatz/metrix_mux

[38] 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.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[39] Z. Wang, 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.

[40] Z.-S. Xiao, "A Multi-scale Structure SIMilarity metric for image fusion qulity assessment," in *Proc. Int. Conf. Wavelet Anal. Pattern Recognit.*, Guilin, China, Jul. 2011, pp. 69–72.

[41] D. Li, M. Hao, J. Zhang, B. Hu, and Q. Lu, "A universal hypercomplex color image quality index," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, Graz, Austria, May 2012, pp. 985–990.

[42] 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.

IEEE *Access*

[43] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.* vol. 21, no. 4, pp. 1500–1512, Apr. 2012.

[44] L. Zhang and H. Li, "SR-SIM: A fast and high performance IQA index based on spectral residual," in *Proc. 19th IEEE Int. Conf. Image Process.*, Orlando, FL, USA, Sep./Oct. 2012, pp. 1473–1476.

[45] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270–4281, Aug. 2014.

[46] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Montreal, QC, Canada, 2004, p. III-709.

XINPENG HUANG received the B.S. and M.E. degrees from the Zhengzhou University of Light Industry, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree with the School of Communication and Information Engineering, Shanghai University. His research interests include light field coding and processing, multi-view image/video coding, and processing.

LIANG SHAN received the B.E. degree from Anhui University, Hefei, China, in 2016. He is currently pursuing the M.E. degree with the School of Communication and Information Engineering, Shanghai University. His research interests include light field image assessment and processing, and multi-view image/video processing.

PING AN received the B.E. and M.E. degrees from the Hefei University of Technology, Hefei, China, in 1990 and 1993, respectively, and the Ph.D. degree from Shanghai University, Shanghai, China, in 2002, where she joined in 1993. From 2011 to 2012, she was a Visiting Professor with the Communication Systems Group, Technical University of Berlin, Berlin, Germany. She is currently a Professor with the Video Processing Group, School of Communication and In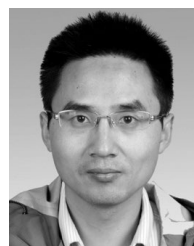formation Engineering, Shanghai University. She has finished more than 15 projects supported by the National Natural Science Foundation of China, the National Science and Technology Ministry, and the Science and Technology Commission of Shanghai Municipality. Her research interests include image and video processing, with a focus on 3D video processing and computer vision. He was a recipient of the Second Prize of the Shanghai Municipal Science and Technology Progress Award, in 2011, and the Second Prize in Natural Sciences of the Ministry of Education, in 2016.

CHAO YANG received the B.S. and Ph.D. degrees from the School of Communication and Information Engineering, Shanghai University, Shanghai, China, in 2012 and 2017, respectively. He is currently a Postdoctoral Fellow with the Department of Electrical Engineering, University of Southern California, Los Angeles, USA. His research interests include image and video processing.

CHUNLI MENG received the B.E. degree from the Jiangxi University of Science and Technology, Ganzhou, China, in 2015, and the M.E. degree from the University of Shanghai for Science and Technology, Shanghai, China, in 2018. She is currently pursuing the Ph.D. degree with Shanghai University, Shanghai, China. Her current research interests include light field imaging and quality assessment.

LIQUAN SHEN received the B.S. degree in automation control from Henan Polytechnic University, Henan, China, in 2001, and the M.E. and Ph.D. degrees in communication and information systems from Shanghai University, Shanghai, China, in 2005 and 2008, respectively. From 2013 to 2014, he was a Visiting Professor with the Department of Electrical and Computer Engineering, University of Florida, Gainesville. Since 2008, he has been with the Faculty of the School of Communication and Information Engineering, Shanghai University, where he is currently a Professor. He has authored or coauthored more than 100 refereed technical articles in international journals and conferences in the field of video coding and image processing. He holds ten patents in the areas of image/video coding and communications. His current research interests include high efficiency video coding, perceptual coding, video codec optimization, 3DTV, and video quality assessment.

• • •