IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 4 August 2022, accepted 11 September 2022, date of publication 21 September 2022, date of current version 29 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3208261

RESEARCH ARTICLE

Refocusing Metric of Light Field Image Using Region-Adaptive Multi-Scale Focus Measure

CHUN ZHAO AND BYEUNGWOO JEON, (Senior Member, IEEE)
Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, South Korea

Corresponding author: Byeungwoo Jeon (bjeon@skku.edu)

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT under Grant NRF-2020R1A2C2007673.

ABSTRACT Compared with conventional photography, the newly emerging light field image capturing technique has dramatically extended potential capabilities of post processing. Among the new capabilities, refocusing is of the most interest. In this paper, we first investigate a region-adaptive multi-scale focus measure (RA-MSFM) that is able to more robustly and accurately measure focus of light field images. It is especially superior when measuring focus in flat areas where previous methods struggle. Following we design a novel refocusing measure metric which employs the RA-MSFM as core technique. Using the metric, refocusing capability of a given light field image as a whole can be measured in a single number by combining focus score maps of each refocused image in the focal stack. The focus score maps are generated using the proposed RA-MSFM. In RA-MSFM, different multi-scale factor is adaptively selected depending on different regions such as texture-rich or flat areas using a multi-layer perceptron network. Different from most light field image metrics that assess image quality, our metric targets to assess refocusing capability. Our experiments have shown that not only does the proposed refocusing metric have high correlation with subjective evaluations given in the form of mean opinion scores, but it also produces all-in-focus images having 0.7 $\sim$ 4.6dB higher PSNRs compared to previous state-of-the-art methods. The proposed refocusing metric can be used to measure refocusing loss in practical application such as compression, tone mapping, denoising, and smoothing.

INDEX TERMS Refocusing measure, light field images, multi-scale focus measure, all-in-focus, subjective experiment, refocusing capability.

I. INTRODUCTION

Light field photography has drawn much attention from academia, consumers, and industries due to its wide range of potential applications such as in photography, astronomy, microscopy, robotics, and medical imaging among others [1]. The light field (LF) camera allows effective reverse raytracing from already recorded image so that the image itself can be adjusted in post processing [2]. That is, focus, exposure, viewing angle, and depth of field can be adjusted after the picture is taken [3]. The availability of depth information over an entire scene also facilitates users in adjusting other aspects of the image like controlling the depth-of-field [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Senthil Kumar.

Recently, various rendering and display techniques for LF data have been reported, for example, techniques related to light field displays and head-mounted displays [5]. With the development of 5G, head-mounted displays of LFs such as in AR/VR are becoming popular [6]. In addition, conventional displays can be used to simulate light field applications taking advantage of techniques such as refocusing or viewing angle change [7]. It is also easy to see many commercial smart phones equipped with 2 or 3 cameras which are able to capture multi-view images. This would make it possible to refocus an image even after the time of its capture. In these applications, the in-focus region of interest (ROI) can be arbitrarily designated by user, thus, the ROI could be placed at any position in the whole image. It is thus very desirable to measure how successful these refocusing operations are.

TABLE 1. Existing focus measure operators and their performance.

Method	Theoretical Principle	Operators	Performances
Spatial domain	Assumes that focused images present more sharp edges than blurred ones [8]. This family takes advantage of image statistics as texture descriptors.	Gradient-based operators, e.g. Thresholded absolute gradient [11][17]	Sensitive to window size; Good response to noise.
		Laplacian-based operators, e.g. SML (sum-modified-Laplacian) [10][18]	Best overall performance in normal imaging conditions; Most sensitive to noise.
		Statistics-based operators, e.g. Gray-level variance, Histogram range [12][13]	The highest robustness to noise; Better autofocus performance.
		RDF (ring difference filter) [15]	Robust to noise
Frequency domain	Uses coefficients of the discrete wavelet transform to describe the frequency and spatial content of images.	Wavelet-based operators, e.g. Sum/Variance of wavelet coefficients [8]; DCT-based operators, e.g. DCT energy ratio.	Performs well in small window size
Others	Conditional operator consisting of multiply operands.	Miscellaneous operators, e.g. Image contrast [14], Image curvature, Local binary patterns-based, Steerable filters-based.	Performance varies in terms of contrast, noise, window size.

In this regard, one likes to measure how much in focus a pixel is in a refocused image. The general refocusing capability of an LF image can be thought as the overall degree of focusing in the images refocused at arbitrary points. There has been research on focus measurement, however, it is noted that most focus measures [10], [11], [12], [13], [14] were only developed with general 2D images in mind, and there are only a few studies [15], [16], [18] that have looked at refocusing of light field images. The motivation and contribution of this paper is to design a measure of general refocusing capability of light field images.

Focus measures (FM) are widely used for many problems of image processing and computer vision, such as for depth from focus [15], [16], [20], autofocus [18]. The existing FM operators for conventional 2D images can be divided into two main categories: spatial domain FM and frequency domain FM [8]. The operators and their advantages are listed in Table 1. There are roughly four families of spatial domain focus measures: Gradient-based, Laplacian-based, statistics-based, and the rest. The analysis and ranking results of FM operators [9] shows that the modified Laplacian (LAP2) [10], tenengrad of gradient variance (GRA7) [11], eigenvalue based (STA2) [3], and image contrast (MIS3) [14] methods achieve the overall high performance among the four families. Noise robustness tests in [9] indicate that STA2 and GRA7 methods handle noise very well.

However, it should also be noted that there are a few FM methods that have been developed for LF images. Surh [15] proposed a FM employing a ring difference filter (RDF) which maintains high robustness and confidence by utilizing a relatively large window of neighboring pixels and placing a ring gap space to ignore certain regions in that window. The RDF focus measure is especially useful in depth estimation for LF images. Rizkallah [17] proposed a metric to decide whether a certain pixel in a focal stack is in focus or not by thresholding of pixel gradient. This metric is used to evaluate compression loss by counting the number of in-focus pixels. It is a very simple metric which judges whether a pixel is in-focus or not by hard thresholding on gradient value, but performs not so well in flat in-focus area. Chantara [18] proposed a FM based on the Summation of the Modified Laplacian, which is sensitive to noise. This FM is used to select the focus area of an LF focal stack.

These existing FM schemes work well in high frequency areas such as rich textures or high sharpness areas, but encounter difficulties when dealing with in-focus flat areas [9]. A multi-scale concept [21], [22], [23] is seen promising in improving the accuracy in those flat areas. Thus, in this paper, we propose a region-adaptive multi-scale focus measure (RA-MSFM) which plays an important role in our refocusing measure methodology. That is, for each refocused image in a focal stack, the RA-MSFM is employed to generate a pixel-wise focus map of the focal stack. An overall refocusing capability score of a whole LF image is calculated by combining the pixel-wise focus maps of the focal stack.

In this paper, we design a new refocusing measure for LF images. Our contribution of this paper lies in introducing 1) an assessment metric for LF refocusing capability, different from existing metrics mostly targets for image quality; 2) the RA-MSFM method performs higher accuracy in both texture-rich and flat area; 3) an appropriate focal stack range and step size are determined by analyzing in a mathematical way instead of simply predefining their values as most research did.

To evaluate the proposed RA-MSFM method, we take two approaches: rendering all-in-focus images and comparing with other state-of-the-art approaches; carrying out subjective experiments on the focus level of the ROI to analyze correlation. To evaluate the proposed refocusing measure metric for whole LF image, another subjective experiment is carried out on in-focus pixels coverage. The proposed refocusing measure can be used in many image processing tasks which may result in refocusing loss such as compression, tone curve mapping, noise reduction, blurring, smoothing, etc. Compression is a method which minimizes data size of image even with image quality degradation, thus the consequential information loss may reduce refocusing capability. Tone curve mapping extends the dynamic range of one region while

suppressing in another region where the suppressed signal may cause refocusing capability loss. The noise reduction process may also distort the pixel values where the distortion affects the refocusing capability of the pixel. Image smoothing and blurring are achieved by convolving the image with a low pass filter, thus refocusing capability is expected to be reduced due to the loss of high frequency information.

The rest of this paper is organized as follows. Section II mainly describes the proposed focal-stack oriented refocusing measure methodology of which the RA-MSFM is a core part. Section III shows the overall structure, flowchart, and the procedure optimization. Section IV presents experiments and analysis, including the all-in-focus rendering experiment and two subjective experiments. The all-in-focus rendering experiment and one of the subjective experiments on focus level are used to evaluate RA-MSFM. The other subjective experiment on in-focus pixel coverage is used to evaluate the refocusing metric. Section V concludes the paper.

II. PROPOSED REFOCUSING MEASURE METHODOLOGY

In this Section, we present the proposed RA-MSFM and a new refocusing measure metric. The proposed RA-MSFM not only works well with high frequency rich textures, but also achieves good performance in the in-focus flat areas meanwhile being robust to noise. Using the proposed refocusing metric, we can compute a single value representing overall degree of focusing on a refocused image generated by post processing of LF image.

A. PROPOSED REGION-ADAPTIVE MULTI-SCALE FOCUS MEASURE (RA-MSFM)

A focus map is a collection of focus levels of pixels in an image. The focus level indicates how much in-focus a given pixel is. A higher focus level indicates that the pixel has the better focus. By conventional methods [9], [18], the focus level of a pixel $I(x, y)$, denoted by $F(x, y)$, can be measured by a selected focus measure operator FM.

$$F(x, y) = FM(I(x, y)) \tag{1}$$

The FM operators generally consider spatial gradient, Laplacian, or pixel variance, thus, their accuracies depend much on the image content.

For example, they work well in areas full of textures and edges [9], however, they tend to fail in in-focus flat area or out-of-focus texture-rich area. This is an evident limitation of the existing FM operators. Besides, light field images may have low spatial resolution and suffer from significant camera noise in images [7], [15]. In this respect, we investigate a new FM method achieving limitation-break and noise-resistance. To have the noise-resistance, a noise-robust FM operator is designed. To break the limitation, we introduce a region-adaptive multi-scale method for which different multi-scale factor is adaptively selected depending on different regions such as texture-rich or flat areas.

1) SELECTION OF A FOCUS MEASURE OPERATOR

To generate a noise-robust pixel-wise focus map, we experimentally evaluate the five well-known FM operators [9], GRA7, LAP2, MIS3, STA2, and RDF for their comparative noise-robustness since they have been reported as showing the best performance with relatively low computational complexity [9]. The test result of the five operators is shown in Fig. 1(a). To simulate realistic camera noise, we also applied gaussian noise to the original light field images in the experiment. A refocused image in a focal stack is referred by its focal stack index k, $k = 1, \ldots, K$. K is the total number of images in the focal stack. The horizontal axis in Fig. 1 indicates the focal stack image index, and the vertical axis shows the focus level of a pixel at the center of its window (shown in yellow in Fig. 1). Note that the more in-focus a pixel is, the higher its focus level is. The focus level of a pixel at (x, y) are different between the kth and $(k+1)$th images in the focal stack. In an ideal case, with increasing focal stack index, the focus level should increase and will decrease after a pick so that showing a hill-shape. In comparison of the focus level curves generated by the five operators in Fig. 1(a), RDF, GRA7, and STA2 are seen to show better resistance to noise than the others since their focus level curves are more similar to the desired smooth hill shape than the others. Meanwhile, LAP2 was found to be the most sensitive to noise since its rises and falls are out of order. As for the accuracy, RDF and STA2 show better performance than GRA7 since the peak of the GRA7's curve does not match with the most in-focus image index marked by the black dashed line. Focus measure performance test on in-focus flat area is shown in Fig. 1(b). It is noted that none of the five operators show their focus level curve having a smooth hill shape for the in-focus area. To compare complexity, we analyzed the average computation time of the five focus operators to see that STA2 consumes the most time. In the overall tradeoff between noise-robustness, accuracy, and complexity, RDF was found to be better than GRA7 and STA2. As such, we select RDF to generate the pixel-wise focus map.

2) REGION-ADAPTIVE MULTI-SCALE FM ARCHITECTURE

An image varies a lot in terms of its spatial contents so that its processing is better to be adaptive especially when it comes to focus measure regarding whether a pixel is in textured or flat area. To deal with this issue for achieving accurate focus map, we investigate a region-adaptive multi-scale focus measure (RA-MSFM). An appropriate degree of down scaling highly depends on local regions, therefore, one single fixed scale-down factor cannot cover all cases. For example, the regions of an image can be roughly divided into 4 types: in-focus flat area (p1), out-of-focus flat area (p2), out-of-focus texture-rich area (p3), and in-focus texture-rich area (p4), as shown in Fig. 2. In this work, an adaptive selection scheme is designed using a multi-layer perceptron network which selects an appropriate scale-down factor. For this, we down-scale the original image (having resolution

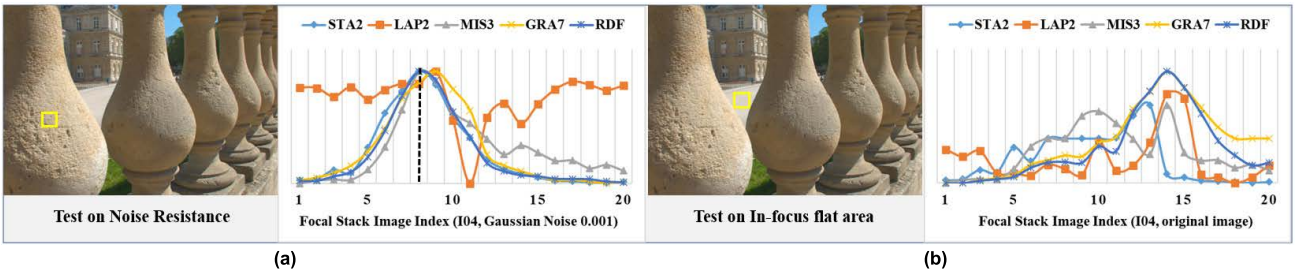

FIGURE 1. Performance comparison of different focus operators. (a) Test on noise resistance; (b) Test in in-focus flat area. The horizontal axis indicates the focal stack image index, and the vertical axis shows the focus level. In comparison of computation time, STA2 takes the longest computation time (STA2 > MIS3 > RDF > GRA7 > LAP2). The pixel under focus measurement is at the center of a 5 × 5 window shown in yellow.

624 × 432) using three factors (1/4, 1/16, 1/64) and train the network to classify the scale-down factor appropriate for each pixel. The ground truth scale-down factor is set manually according to local region structure, which is described later in this Section II-A.

a: STRUCTURE OF THE CLASSIFICATION NETWORK

The adaptive selection scheme is shown in Fig. 3. A fully connected network is employed to predict an appropriate scale-down factor for a given pixel p. Input information to the network are multi-scale blocks centered at pixel p which are extracted from the down-scaled images. In order to reduce redundancy and dimensionality, intended features are extracted from the original, 1/4, 1/16, and 1/64 scale-downed blocks at the feature layer of the network. The intended features not only include general image information such as luminance histogram, variance, and color histogram, but also include custom features that are specially designed for judging proper type such as texture frequency, gradient histogram, and detail information. To improve convergence speed, all the feature variables are normalized to range 0∼1. By the fully connected layer, probabilities of four scale-down factors are generated at the output layer. The final predicted label is the scale-down factor having the highest probability. The loss, which measures the model accuracy in training, is computed in terms of the absolute difference of probabilities between the predicted and the true labels. With the proposed RA-MSFM as a selected focus measure operator FM in (1), the focus level of a pixel in an image is re-written as:

$$F(x,y) = MSFM(I(x,y)) = R_{\hat{t}}(x,y)$$
$$\text{where } \hat{t} = \underset{t}{\operatorname{argmax}}\, Pr(t), \quad t \in (t0, t1, t2, t3) \qquad (2)$$

where $\hat{t}$ is the classified label for a pixel $I(x,y)$ and $Pr(t)$ is the probability of the label t provided as network output. $R_{t0}(x,y)$, $R_{t1}(x,y)$, $R_{t2}(x,y)$, and $R_{t3}(x,y)$ are the focus levels corresponding to the pixel $I(x,y)$, respectively in the original, 1/4, 1/16, and 1/64 scaled-down resolution images.

b: PREPARATION OF GROUND TRUTH SCALE-DOWN FACTORS FOR TRAINING

The scale-down factor should be selected considering structures of local regions. For training, the ground truth scale-down factor of each pixel is prepared among four scale factors (1, 1/4, 1/16, 1/64) by selecting the one that reflects its actual focus level the best. The focus level is calculated as in (1) using a selected focus measure operator FM, for example, RDF [15]. A high FM value of a pixel indicates that the pixel is in-focus, on the other hand, a low value indicates an out-of-focus pixel. Accordingly, for a pixel located at the in-focus region like p1 and p4, its ground truth scale-down factor is selected as the one giving the maximum focus level. For a pixel located at the out-of-focus regions (like p2 and p3), the one giving the minimum focus level is selected as the ground truth. Fig. 2 illustrates the ground truth of the scale-down factor (indicated by bolded rectangle) for the pixels p1∼p4.

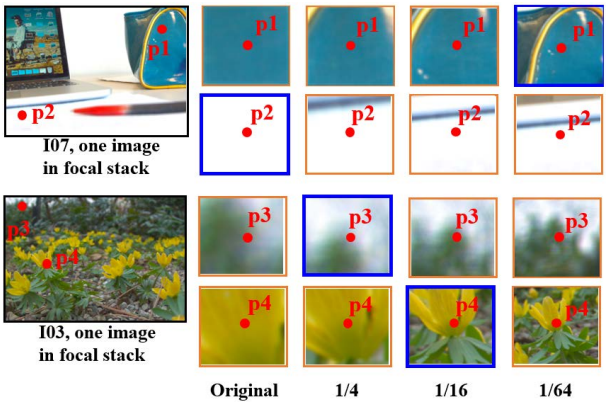

FIGURE 2. Image texture type and its ground truth label (p1: in-focus flat area; p2: out-of-focus flat area; p3: out-of-focus texture-rich area; p4: in-focus texture-rich area). The bolded rectangle indicates the selected ground truth scale-down factor of each selected pixel.

A focus map showing focus level of each pixel is generated with our RA-MSFM. Fig. 4 compares the focus maps generated by three well-known methods and the proposed RA-MSFM. Unlikely the existing methods, the proposed RA-MSFM is shown to work well even in flat areas by selecting an appropriate scale-down factor to show high accuracy focus level in in-focus flat area.

3) GENERATION OF FOCUS SCORE MAP

A focal stack is a collection of the same image but focused on multiple planes, thus, refocusing can be understood in a

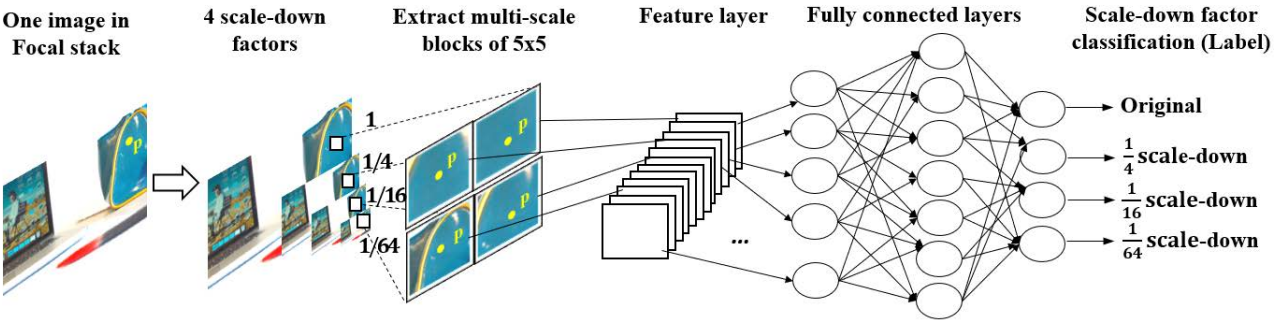

FIGURE 3. Proposed region-adaptive multi-scale focus measure (RA-MSFM) scheme. (4 scale-down factors: 1, 1/4, 1/16, 1/64; multi-scale blocks of 5 × 5: extract 5 × 5 pixels window with its center at a pixel p from 4 types of scale-downed images).

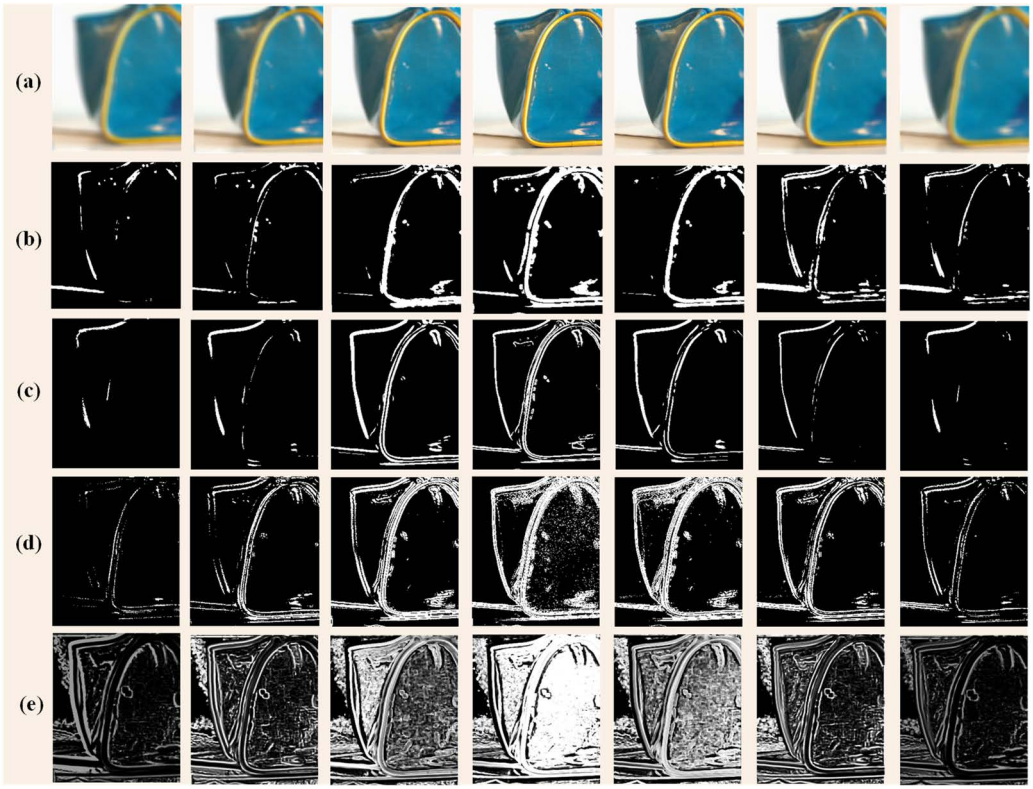

FIGURE 4. Comparison of the generated focus maps. (a) Original focal stack images; Generated focus maps by (b) Chantara's method (Laplacian based) [18]; (c) Rizkallah's method (Gradient based) [17]; (d) Surh's method (Ring Difference Filter) [15]; (e) the proposed method.

simple term as selecting an appropriate image (or rendering an image by selecting appropriate parts of images) in the focal stack corresponding to a given desired focus plane. Since multiple refocused images can be rendered from one single light field image, availability of possible in-focus areas of the images in the focal stack that can cover the whole image as much as possible can provide a very good indication of the overall refocusing capability of a light field image.

We compute the focus score $S_k(x, y)$ by normalizing $F_k(x, y)$ in (2) by F_{LF}^{max} as below.

$$S_k(x, y) = \frac{F_k(x, y)}{F_{LF}^{max}}$$
$$\text{where } F_{LF}^{max} = \max_{k=1\ldots k} \max_{\substack{0\le x<W \\ 0\le y<H}} (F_k(x, y)) \tag{3}$$

$F_k(x, y)$ represents the focus level of a pixel (x, y) in the kth image in the focal stack. F_{LF}^{max} is the maximum focus level among all K refocused images in the focal stack for a given LF image. H and W are respectively width and height of the refocused image. In computing the maximum focus level, 1% of the highest values are excluded as outliers to minimize the noise effect.

B. PROPOSED REFOCUSING CAPABILITY MEASURE METRIC

In this section, we propose a refocusing capability measure for the given 4D light field image and its rendered 2D refocused images.

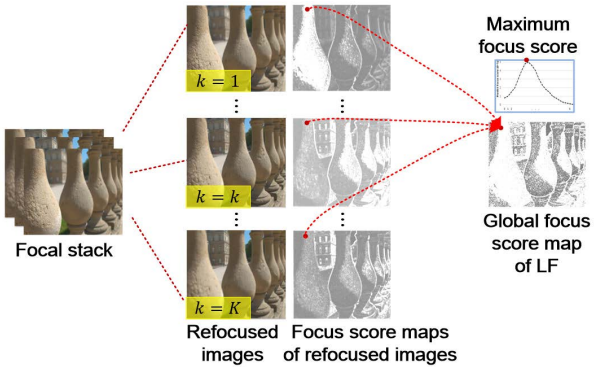

FIGURE 5. Focus score map generation for light field image (LF).

1) REFOCUSING CAPABILITY OF ONE REFOCUSED IMAGE

Refocusing capability of a refocused image is proposed to be measured by the refocusing pixel coverage (RPC) which is expressed as a percentage of the in-focus pixels in the given refocused image. RPC of the kth refocused image in the focal stack is defined as:

$$RPC(k) = \frac{1}{HW}\sum_{x=0}^{W-1}\sum_{y=0}^{H-1} b_k(x,y)$$
$$b_k(x,y) = \begin{cases} 0, & S_k(x,y) \le THD \\ 1, & S_k(x,y) > THD \end{cases} \quad (4)$$

where $b_k(x,y)$ represents whether the given pixel at (x,y) is in-focus or not: $b_k = 1$ indicates an in-focus pixel while $b_k = 0$ indicates an out-of-focus (that is, blurred) pixel. A pixel is thought perceptually in-focus if its focus score is higher than a certain threshold *THD*. *THD* is determined by experiment.

2) OVERALL REFOCUSING CAPABILITY OF LIGHT FIELD IMAGE

It is possible to measure the refocusing capability of LF image as a whole by integrating the refocusing capabilities of all the possible 2D refocused images rendered from the given light field image. From the 1st to the kth image in the focal stack, the focus score increases to a peak (indicating highest in-focus), and then decreases (indicating out-of-focus) like an inverted valley. We find a maximum focus score $S_{LF}(x,y)$ at (x,y) among all the images in the focal stack and generate a global focus score map for 4D light field (LF) image as shown in Fig. 5. The global focus score map of an LF image reveals the best possible focus for each pixel.

$$S_{LF}(x,y) = \max_{k=1,\ldots,K}(S_k(x,y)) \quad (5)$$

The overall refocusing pixel coverage (RPC) of a light field image, RPC_{LF}, is defined as below.

$$RPC_{LF} = \frac{1}{HW}\sum_{x=0}^{W-1}\sum_{y=0}^{H-1} b_{LF}(x,y)$$
$$b_{LF}(x,y) = \begin{cases} 0, & S_{LF}(x,y) \le THD \\ 1, & S_{LF}(x,y) > THD \end{cases} \quad (6)$$

III. OVERALL STRUCTURE AND PROCEDURE OPTIMIZATION

In this Section, first, the overall structure of refocusing metric is presented. Following, we optimize the number of images in a focal stack by setting appropriate α values through mathematical analysis.

A. OVERALL FRAMEWORK

In our refocusing measure methodology, we first generate a focal stack using a 4D light field image. In contrast to the general focal stack with a user-defined fixed step size [17], [33], we define the range and the sample step size through mathematical analysis. Then, we generate a pixel-wise focus map of each image in the focal stack using RA-MSFM as in (2). Furthermore, we normalize the focus maps of each refocused image in the focal stack using (3). Finally, a single score value representing the refocus capability of a given LF image as a whole is calculated using (6) by combining the focus score maps of all refocused images in the focal stack as in (5). The main structure of this approach is illustrated in Fig. 6. It includes five submodules: focal stack generation from 4D light field input data; RA-MSFM for each image in the focal stack to generate focus maps; normalization and combination of the focus maps in the focal stack to have a global focus map; calculation of the overall LF refocusing score. Our method has three advantages. First, we can compute a single value representing overall degree of focusing in refocused image generated from LF image through post-processing. Second, the proposed RA-MSFM works well in both high frequency rich textures and in-focus flat areas. Third, our generated focal stack is superior to existing methods [17], [33], [34], [41] since the range and the step size used for the focal stack generation are not set simply as in other methods [17], [33], but determined based on the proposed rigorous mathematical analysis.

B. FOCAL STACK GENERATION WITH APPROPRIATE α VALUES

Light field images are captured using a plenoptic cameras which typically have microlens arrays placed in front of an image sensor to record incoming light rays from many directions [24]. This architecture allows differentiating as many directions as there are pixels behind each microlens. The light field can be used to digitally reconstruct an image corresponding to a different camera focus, which we call the refocused image [25], [26]. The light ray of 4D light field can be parameterized by two parallel planes uv and xy, known as the directional and spatial dimensions. The camera aperture is positioned along the uv plane, while xy indicates the sensor plane [19]. As illustrated in Fig. 7, the sensor plane is located at a distance F from the aperture plane, and the light ray $L(x,y,u,v)$ reaches at the position x on the sensor plane. For the refocus plane RP_k, its distance from the aperture plane is $F' = \alpha F$ where $\alpha = F'/F$ is defined as the relative depth [27]. The dashed light ray converging on a position x

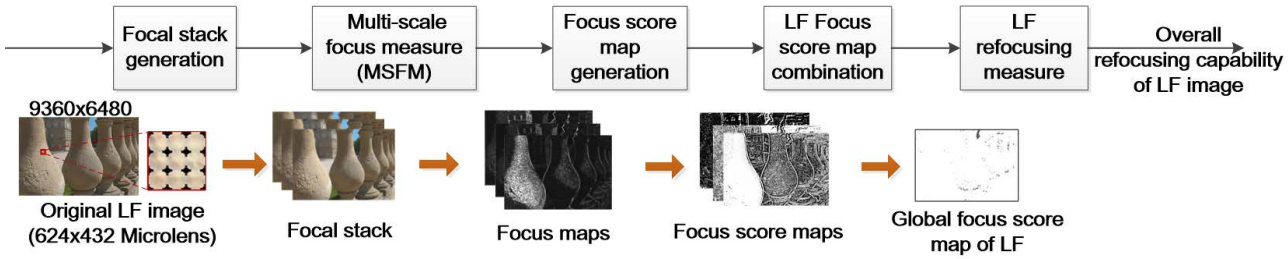

FIGURE 6. Framework of the proposed refocusing measure methodology.

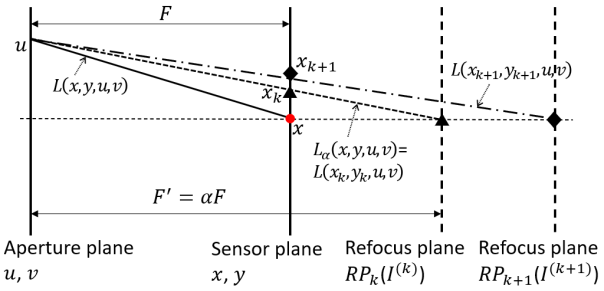

FIGURE 7. Light field image refocusing.

on the refocus plane RP_k is denoted by $L_\alpha(x, y, u, v)$. Since it also reaches at x_k in the sensor plane, $L_\alpha(x, y, u, v) = L(x_k, y_k, u, v)$. By similar triangles in Fig. 7, $x_k = \frac{x}{\alpha} + u\left(1 - \frac{1}{\alpha}\right)$, thus

$$L_\alpha(x, y, u, v) = L\left(\frac{x}{\alpha} + u\left(1 - \frac{1}{\alpha}\right), \frac{y}{\alpha} + v\left(1 - \frac{1}{\alpha}\right), u, v\right) \tag{7}$$

The refocused image $I_\alpha(x, y)$ is rendered by integrating all directional rays at a specific position on the sensor plane:

$$I_\alpha(x, y) = \iint L\left(\frac{x}{\alpha} + u\left(1 - \frac{1}{\alpha}\right), \frac{y}{\alpha} + v\left(1 - \frac{1}{\alpha}\right), u, v\right) dudv \tag{8}$$

By scaling and normalizing $\frac{x}{\alpha} \to x$, $\frac{y}{\alpha} \to y$, we have,

$$I_a(\alpha x, \alpha y) = \iint L\left(x + u\left(1 - \frac{1}{\alpha}\right), y + v\left(1 - \frac{1}{\alpha}\right), u, v\right) dudv \tag{9}$$

The image $I_a(\alpha x, \alpha y)$ is a scaled version of $I_a(x, y)$ by a factor α. However, we can ignore the scaling factor in digital processing since we can just scale the pitch size of a virtual sensor plane by exactly α times so the resolution of a rendered image matches exactly with spatial resolution of light field. The explanation of ignoring dilated factor α is first presented by Ng *et al.* [27] and is adopted in the following research [28], [29], [30] for digital refocusing. The digital refocusing realization at RP_k is given by rendering equation in discrete domain,

$$I_\alpha(x, y) = \sum_u \sum_v L^{(u,v)}\left(x + u\left(1 - \frac{1}{\alpha}\right), y + v\left(1 - \frac{1}{\alpha}\right)\right) \tag{10}$$

Here, $L^{(u,v)}(x, y) = L(x, y, u, v)$ denotes a 2D view from the point (u, v). (10) shows the digital refocusing is realized by shifting a factor $u(1 - 1/\alpha)$, $v(1 - 1/\alpha)$. If we denote the shift offset as Δx, Δy, then, $\Delta x = u(1 - 1/\alpha)$, $\Delta y = v(1 - 1/\alpha)$.

Regarding the refocusing parameter α, an α value less than 1 indicates its refocus plane being close to the aperture plane (that is, $F' < F$), and α value larger than 1 means a refocus plane far from the aperture plane (that is, $F' > F$). In general, α can assume any real value, and it is related to one refocused image in the focal stack, that is, one refocused plane. The number of images in the focal stack, K, is determined by α. If K is of a high value, a large and redundant focal stack demands huge hardware consumption as shown in Fig. 8(a); if the value of K is low, the generated focal stack cannot cover all the refocusing ranges as shown in Fig. 8(b), thus the measured refocusing capability will be smaller than its real refocusing capability. Therefore, it is necessary to decide an appropriate K value. The K value depends on the range $\alpha_{max} - \alpha_{min}$ and step size $\Delta a = \alpha_{k+1} - \alpha_k$.

There have been some studies [31], [32], [33], [34] on setting the α value range and step size. One work [31] investigated setting these values based on the image content depth, but since content always varies, the depth of each image will also vary, so it is difficult to give a consistent definition in this way; another work [32] also investigated this based on the plenoptic camera's focal length and microlens diameter, but this approach is limited to only typical cameras; another work [33] just used a fixed range and a step size was found experimentally. In this paper, we study the problem again mathematically and propose a method to define the range and sample step size of parameter α.

1) SETTING FOCAL STACK RANGE

Suppose a plenoptic camera has $M \times N$ microlens each of which has $P \times P$ aperture views. It is $M \times N$ in the spatial

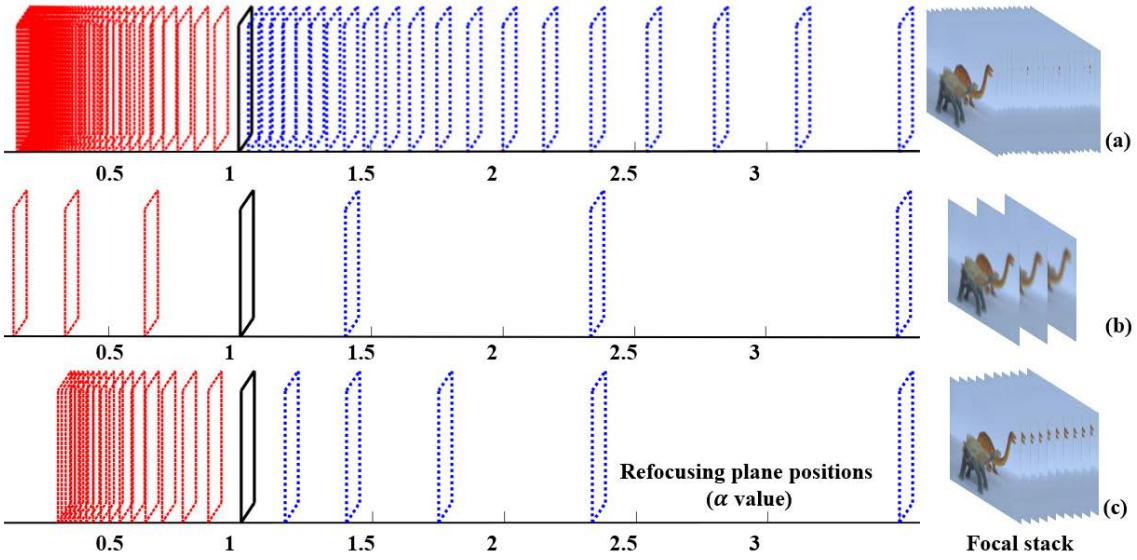

FIGURE 8. Refocusing plane positions (Original focus plane $\alpha = 1$ ($F = F'$); left side planes: $\alpha < 1$ ($F > F'$) background direction; right side planes: $\alpha > 1$ ($F < F'$) foreground direction). (a) Number of images in focal stack is large and redundant due to a small $\Delta\alpha$; (b) Number of images in focal stack is small due to a large $\Delta\alpha$; (c) Number of images in a focal stack is appropriate with the proposed $\Delta\alpha$ setting.

xy dimension, and $P \times P$ aperture views in the directional uv dimension. A sub-aperture image $L^{(u,v)}(x, y)$ extracted from an LF image has a set of $P \times P$ views, where P is usually an odd number. The possible ranges for u, v, x, y in each sub-aperture image $L^{(u,v)}(x, y)$ are:

$$u(\textit{and also } v) \in \left\{-\frac{P-1}{2}, \ldots, 0, \ldots, \frac{P-1}{2}\right\},$$
$$x \in \{0, 1, \ldots, M-1\}, y \in \{0, 1, \ldots, N-1\}$$

In the refocusing plane, the directional shifts $\Delta x, \Delta y$ have the following constraint due to the limitation of $|\Delta x| \leq M-1, |\Delta y| \leq N-1$. We analyze α value in both two cases of $0 < \alpha < 1$ and $\alpha > 1$. For $0 < \alpha < 1$, using the constraint of $|\Delta x| \leq M-1$,

$$\left|u\left(1-\frac{1}{\alpha}\right)\right| \leq M-1, \text{ since } |u|_{max} = \frac{P-1}{2}$$
$$\left|1-\frac{1}{\alpha}\right| \leq \frac{2(M-1)}{P-1} \tag{11}$$
$$\alpha \geq \frac{P-1}{2(M-1)+P-1} \tag{12}$$

Similarly, using the constraint on $|\Delta y|$ above,

$$\alpha \geq \frac{P-1}{2(N-1)+P-1} \tag{13}$$

So, the minimal α value depends on M, N, P. For the case of $\alpha > 1$, we always have $0 < 1 - \frac{1}{\alpha} < 1$, thus,

$$|\Delta x| = \left|u\left(1-\frac{1}{\alpha}\right)\right| < |u|_{max} \tag{14}$$

Since $|\Delta x|$ can be close to $|u|_{max}$ but cannot be $|u|_{max}$, to reduce redundancy, we set $|u|_{max} - 1$ as the maximum $|\Delta x|$ value, thus, $|\Delta x| \leq |u|_{max} - 1$,

$$\left(1-\frac{1}{\alpha}\right)|u|_{max} \leq |u|_{max} - 1$$
$$\alpha \leq |u|_{max}$$
$$\alpha \leq \frac{P-1}{2} \tag{15}$$

So, the maximal α value depends on P.

2) SETTING FOCAL STACK STEP SIZE

In Fig. 7, the refocus planes RP_k and RP_{k+1} respectively correspond to $I^{(k)}$ and $I^{(k+1)}$. The kth and $(k+1)$th refocused images, $I^{(k)}(x, y) = I_{\alpha_k}(x, y)$ and $I^{(k+1)}(x, y) = I_{\alpha_{k+1}}(x, y)$ are rendered using the refocus parameters, α_k and α_{k+1}. The step size between the two refocused images $I^{(k)}$ and $I^{(k+1)}$ is defined as $\Delta\alpha = |\alpha_k - \alpha_{k+1}|$.

In order to cover as wide refocusing ranges as possible with the minimal number of images in the focal stack, we should define a proper value for $\Delta\alpha$. The light ray $L^{(u,v)}(x_k, y_k)$ and $L^{(u,v)}(x_{k+1}, y_{k+1})$ converge to a position x of $I^{(k)}$ and $I^{(k+1)}$, respectively.

$$x_k = x + u\left(1-\frac{1}{\alpha_k}\right)$$
$$x_{k+1} = x + u\left(1-\frac{1}{\alpha_{k+1}}\right) \tag{16}$$

and $\Delta d = |x_k - x_{k+1}|$ is the distance between the two light rays. If $\Delta d > 1$, refocusing possibility is compromised due to omission of some light rays. If $\Delta d \leq 1$, all the refocusing ranges can be covered, therefore, $\Delta d = |x_k - x_{k+1}| \leq 1$ is an essential constraint as shown below.

$$|u|\left|\frac{1}{\alpha_{k+1}} - \frac{1}{\alpha_k}\right| \leq 1 \tag{17}$$

To ensure all the u values meet the constraint, it is required that $|u| = |u|_{max} = (p-1)/2$. If Δd is too small, there will be many images in a focal stack which causes redundancy and high computational cost, thus, $\Delta d = 1$ is selected as an appropriate value.

$$|u|_{max} \left| \frac{1}{\alpha_{k+1}} - \frac{1}{\alpha_k} \right| = 1 \tag{18}$$

In case of $0 < \alpha < 1$, then $\alpha_{k+1} < \alpha_k$, thus,

$$\alpha_{k+1} = \frac{\left(\frac{p-1}{2}\right)\alpha_k}{\left(\frac{p-1}{2}\right) + \alpha_k} \tag{19}$$

In the same way, when $\alpha > 1$, then, $\alpha_{k+1} > \alpha_k$, thus,

$$\alpha_{k+1} = \frac{\left(\frac{p-1}{2}\right)\alpha_k}{\left(\frac{p-1}{2}\right) - \alpha_k} \tag{20}$$

The refocusing plane positions using the step size decided in this paper are shown in Fig. 8(c). With the proposed α setting, the generated focal stack can cover all the refocusing range while having only the minimum number of images.

IV. EXPERIMENT AND ANALYSIS

In this Section, we evaluate the performance of 1) the proposed RA-MSFM, and 2) the proposed refocusing capability metric RPC. To demonstrate advantages of the proposed RA-MSFM, we render an all-in-focus (AIF) image using the proposed RA-MSFM. For performance comparison, AIF images using other state of art methods [15], [17], [18] are also generated. A higher PSNR value between the rendered and the ground truth (GT) AIF images is an indication of achieving better performance.

We also carry out a subjective evaluation experiment (subjective test I) by collecting subjective Mean Opinion Score (MOS) values following the recommendation in [39]. Subsequently we compare the correlation coefficients achieved by the proposed objective focus level and the subjective MOS value to verify whether the proposed measure agrees well with the subjective judgements.

To evaluate our refocusing capability metric, we also conduct another subjective experiment (subjective test II) for the refocusing pixel coverage (RPC) of each image in a focus stack.

A. OBJECTIVE EVALUATION BY RENDERING ALL-IN-FOCUS IMAGE

To evaluate the proposed focus measure, we render AIF images since the focus measure is an essential technique in their generation [35], [36]. The AIF image is generated by finding the highest focus score pixel by pixel from the focal stack images [37]. k_{max} is the index of the focal stack image that has the highest focus score.

$$\begin{aligned} AIF\,(x,y) &= I^{(k_{max})}\,(x,y) \\ \text{where } k_{max} &= \underset{k=1,\ldots,K}{\text{argmax}}\,(S_k\,(x,y)) \end{aligned} \tag{21}$$

Fig. 9 shows the all-in-focus images generated by the different methods [15], [17], [18]. To ensure a fair experiment, the window size is set to 5×5 for all the focus measure operators tested. The AIF images rendered using our RA-MSFM are seen better than the images generated by existing methods (take note of the sky, wall, and face). While the AIF images rendered by the proposed method is very clean like the ground truth AIF images, the other existing methods show unpleasing artifacts in the flat areas.

TABLE 2. PSNR and RMSE comparison.

	DB	Chantara's [18]	Rizkallah's [17]	Surh's [15]	Proposed
PSNR [dB]	I01	30.54	26.93	30.15	**31.14**
	I02	29.07	26.67	28.82	**29.69**
	I03	30.14	26.75	28.96	**30.52**
	I04	32.76	29.16	32.00	**32.94**
	I05	36.12	31.81	35.29	**36.94**
	I06	37.08	32.77	36.14	**37.63**
	I07	30.39	25.78	29.15	**30.70**
	I08	35.74	30.98	34.12	**35.95**
	I09	28.59	25.80	28.09	**30.02**
	I10	33.19	27.79	32.37	**34.20**
	I11	29.27	24.89	28.17	**29.37**
	I12	28.09	25.08	28.01	**29.48**
RSME	I01	3.88	5.20	4.12	**3.69**
	I02	4.38	5.49	4.59	**4.28**
	I03	3.96	5.30	4.27	**3.81**
	I04	3.55	4.67	3.78	**3.53**
	I05	2.58	3.83	2.91	**2.49**
	I06	2.31	3.64	2.69	**2.15**
	I07	3.51	5.98	4.23	**3.64**
	I08	2.44	3.97	2.89	**2.36**
	I09	4.88	6.48	5.14	**4.34**
	I10	3.16	5.44	3.64	**2.96**
	I11	3.82	6.52	4.14	**3.77**
	I12	5.17	7.78	5.61	**4.42**

(best one in bold)

Table 2 illustrates the correlations between the reference ground truth image and the rendered AIF images. Test datasets I01 $\sim$ I12 are shown in Fig. 10(a). Peak signal to noise ratio (PSNR) and root mean squared error (RMSE) are used to evaluate the correlation between the images. In checking similarity between the reference and the rendered images, a high PSNR indicates that the rendered AIF image is very similar to the reference AIF image. So, in this evaluation of all-in-focus images, a high PSNR is seen to indicate a high-quality focus measure method. Since the RMSE quantifies difference between the reference and the rendered image, a smaller RMSE indicates better performance of the given focus measure. The experiment results reveal that the proposed method has the highest PSNR and the lowest RSME. The rendered AIF image using the proposed method has average 0.7dB, 4.6dB, and 1.5dB higher PSNR than Chantara's [18], Rizkallah's [17], and Surh's [15] methods, respectively. The RSME results also show that the proposed method produces 4.6%, 35.3%, and 13.8% smaller error than the existing methods, respectively.

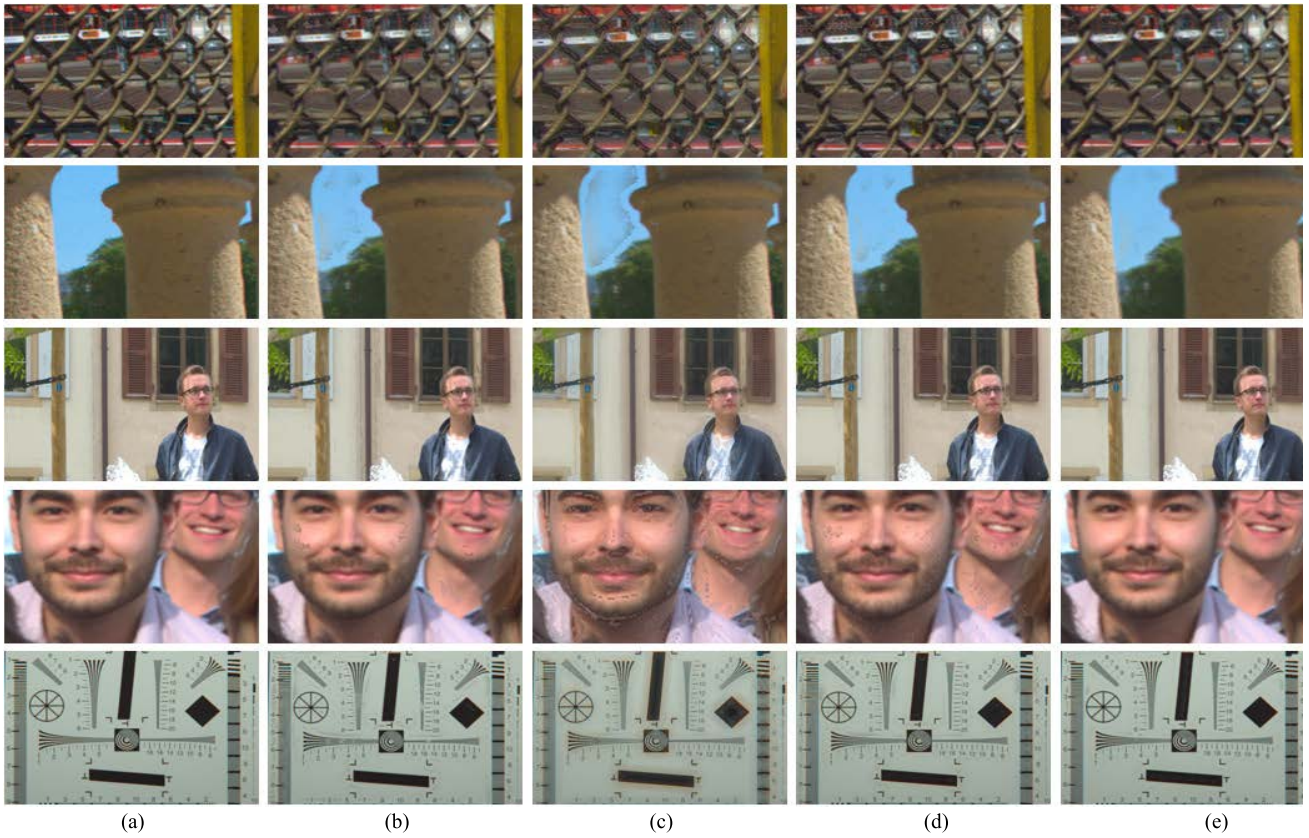

FIGURE 9. Various all-in-focus images generated by different focus measure methods. (a) Reference GT AIF images; (b) Chantara's method [18] (Laplacian based); (c) Rizkallah's method [17] (Gradient based); (d) Surh J 's method [15] (Ring Difference Filter); (e) The proposed method.

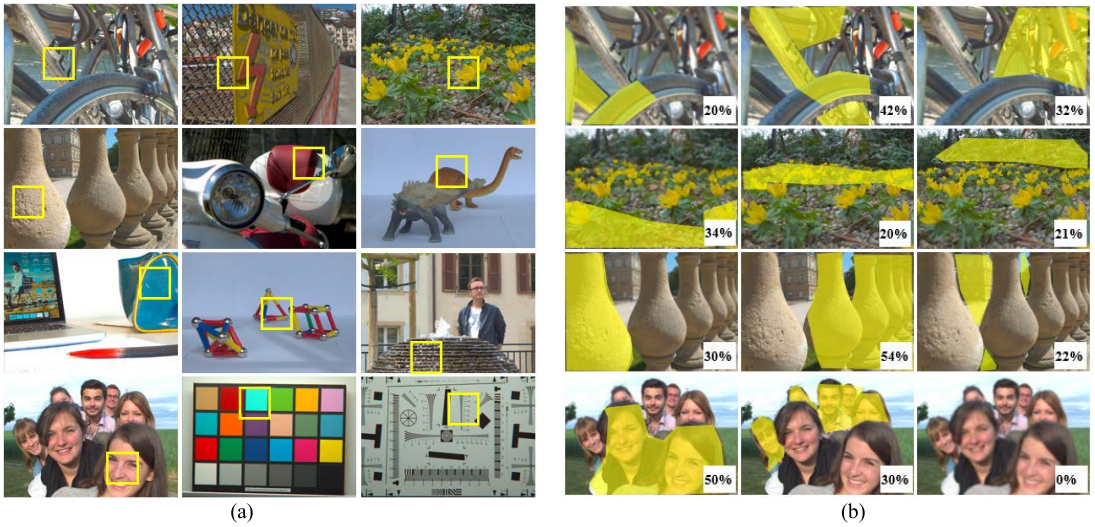

FIGURE 10. Subjective experiments. (a) Subjective test I: focus measure on marked ROI area (marked area is target area which is used to subjectively score focus capability level as 0 ∼ 1) (I01 ∼ I12); (b) Subjective test II: refocusing metric (marked area is refocusing pixel coverage (RPC) of sample images in focal stack) (I01, I03, I04, I10).

B. SUBJECTIVE EVALUATION

1) EXPERIMENT SETTING

Using the popular LF image dataset [38] which covers all the general categories, we perform two subjective evaluations of our RA-MSFM and the refocusing measure metric RPC. Our experimental environment is arranged according to the recommendations specified in ITU-R BT.500-12 [39].

TABLE 3. Experiment configuration for subjective test.

Configuration	Values
Method	Single stimulus (SS)
Evaluation score scales	0~5
Number of light field images	12
Rendered image resolution	624x432
Viewing distance	Three times the image height
Ambient light	Dark room

A single-stimulus (SS) method is used in these tests under the configuration in Table 3. For each light field image in the test, we generate a focal stack that includes 21 refocused images with different focus depths. In the subjective test I, for certain regions of interest (ROI) marked with yellow boxes in Fig. 10(a), the focus capability level of each ROI is scored from $0 \sim 1$.

Among the 21 rendered images in the focal stack, the best in-focus one is scored as 1, while the most blurred (that is, out-of-focus) one is given 0. In the subjective test II, the refocusing pixel coverage range is marked and the ratio of the focused area to the whole image is scored as $0 \sim 100\%$ as shown in Fig. 10(b). Total 15 subjects are asked to mark all the in-focus pixels for each image in the focal stack and the percentage of focused area is calculated as the score. For this, we develop a tool using MATLAB code to experiment the focus measure MOS. The experiment consists of training and rating stages. The subjects preview some examples at the training stage to understand how to score properly. When tallying the result, one outlier score is removed from computing an average MOS value.

2) EXPERIMENT ANALYSIS

After the subjective test, for performance comparison with three existing methods, we analyze the correlations between the objective focus score generated by the proposed method and the subjective MOS score employing four commonly used statistical indexes: Spearman rank order correlation coefficient (SROCC), Kendall rank order correlation coefficient (KROCC), Pearson Linear Correlation Coefficient (PLCC), and Absolute prediction error root mean square error (RMSE). SROCC, KROCC, and PLCC indicate correlation, thus, a higher value means higher correlation. RMSE, meanwhile, measures error, thus, a smaller value indicates the measured focus score is closer to the subjective score.

For noise-resistance comparison of different focus measure methods in the subjective evaluation I, we test the proposed MSFM, Chantara's [18], Rizkallah's [17], and Surh's [15] focus measure methods. Fig. 11 illustrates the focus score for the focal stacks I01, I07, and I10 corrupted by Gaussian noise of $\sigma = 0.001$. The black line shows subjective focus score MOS value (scaled to $0 \sim 1$) given by human. The measured focus score for the focal stack generated with the various methods show that the proposed method is closest to the reference subjective result.

TABLE 4. Subjective I & objective comparision on FM of each image.

Co-relation		Chantara's [18]	RizkallahM's [17]	SurhJ's [15]	Proposed
	I01	0.39	0.61	0.67	**0.90**
	I02	0.30	0.65	0.82	**0.87**
K	I03	0.24	0.76	0.74	**0.77**
R	I04	0.25	0.62	**0.91**	0.88
O	I05	0.36	0.72	0.78	**0.80**
C	I06	0.38	0.76	0.73	**0.94**
C	I07	0.26	0.63	**0.95**	**0.95**
	I08	0.20	0.75	0.92	**0.98**
	I09	0.35	0.81	**0.94**	0.92
	I10	0.19	0.52	0.94	**0.97**
	I11	0.84	0.83	0.63	**0.93**
	I12	0.56	0.79	0.74	**0.92**
	I01	0.64	0.84	0.81	**0.98**
	I02	0.09	0.61	0.89	**0.98**
P	I03	0.38	0.6	0.78	**0.80**
L	I04	0.59	0.72	0.87	**0.97**
C	I05	0.48	0.58	0.82	**0.93**
C	I06	0.48	0.62	0.93	**0.99**
	I07	0.46	0.62	0.89	**0.99**
	I08	0.10	0.54	0.91	**0.98**
	I09	0.42	0.64	0.87	**0.98**
	I10	0.26	0.64	0.88	**0.99**
	I11	0.92	0.88	0.69	**0.95**
	I12	0.65	0.73	0.88	**0.98**

(best one in bold)

Table 4 shows the correlations between the objective and subjective focus score for the datasets I01 $\sim$ I12 with Gaussian noise of $\sigma = 0.001$. The Laplacian-based Chantara's method is found to be most sensitive to noise.

TABLE 5. Subjective I & objective correlation of focus meansures.

Correlation		Chantara's [18]	Rizkallah M's [17]	Surh J's [15]	Proposed
Original	SROCC	0.93	0.85	0.91	**0.97**
	KROCC	0.90	0.80	0.85	**0.90**
	RMSE	0.07	0.12	0.08	**0.05**
	PLCC	0.91	0.82	0.89	**0.96**
Gaussian noise σ=0.001	SROCC	0.40	0.66	0.78	**0.92**
	KROCC	0.36	0.70	0.81	**0.90**
	RMSE	0.24	0.18	0.13	**0.05**
	PLCC	0.46	0.67	0.85	**0.96**
Gaussian noise σ=0.02	SROCC	0.21	0.55	0.64	**0.82**
	KROCC	0.15	0.53	0.60	**0.70**
	RMSE	0.28	0.24	0.19	**0.13**
	PLCC	0.23	0.50	0.68	**0.81**

(best one in bold)

Table 5 shows the average performance over 12 LF images for the original and its noise-added ones. In the case of the original images, the proposed method is seen to achieve 2% $\sim$ 4% higher correlation than Chantara's, 5% $\sim$ 8% higher than Surh's, and 10% $\sim$ 15% higher than Rizkallah's. In the case of the noisy images, the Laplacian-based Chantara's method shows the worst performance. The comparison data in Table 4, Table 5, and the measured focus level for the ROI in Fig. 11 show that the proposed method is superior to the

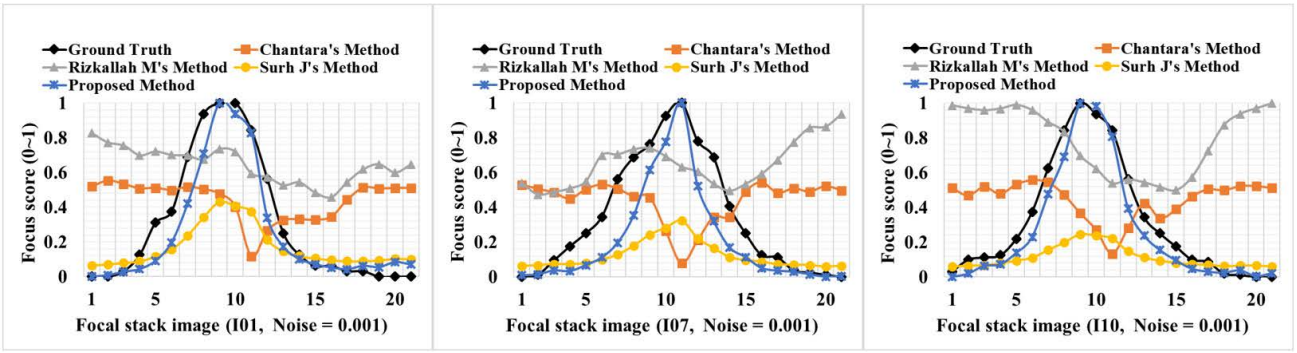

FIGURE 11. Subjective evaluation I on MSFM methods (with noisy images).

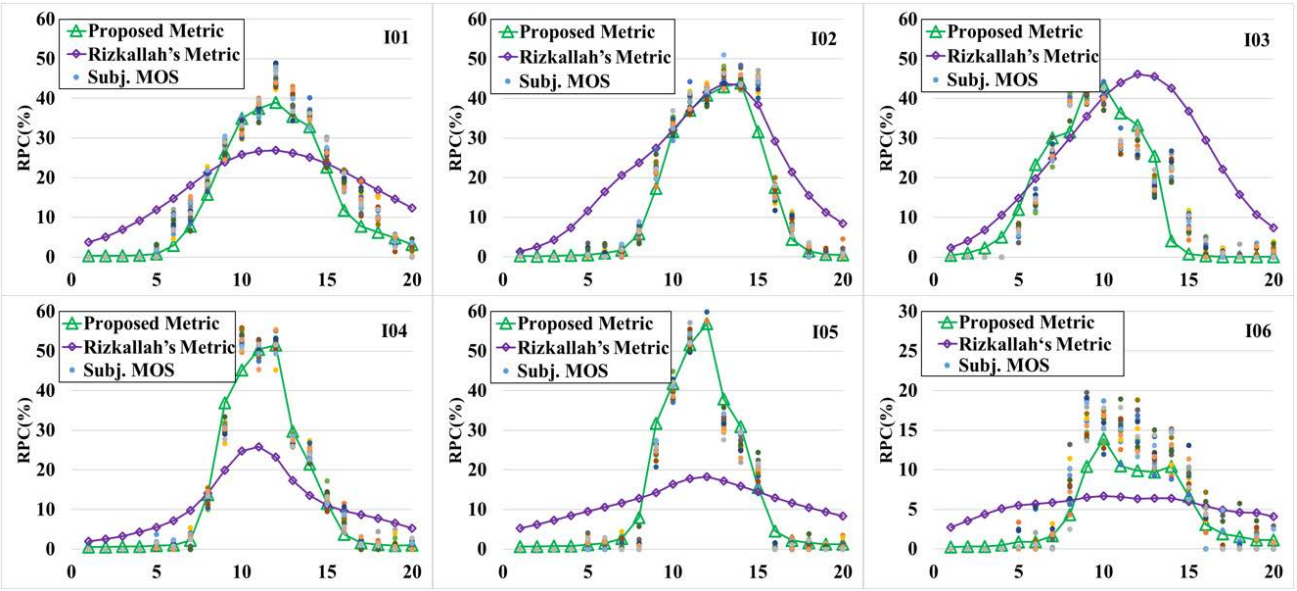

FIGURE 12. Subjective evaluation II on refocusing capability metric (with original images).

existing methods not only for the original images but also with noisy images.

Apart from the subjective test I of focus score for a given local area, the subjective test II is for refocusing capability metric regarding refocusing pixel coverage. In the subjective test II, subjects are asked to mark all the in-focus pixels for each image in the focal stack.

The subjective refocusing capability is measured in terms of the in-focus pixel percentage over the whole image. The comparative subjective and objective refocusing capability evaluation is shown in Fig. 12 using I01 $\sim$ I12 data sets. The proposed metric RPC gets closer to the subjective score than the state-of-the-art Rizkallah's metric [17]. The correlation analysis results with PLCC and RMSE are shown in Table 6 which shows also that the proposed metric achieves 12% higher correlation and 5% lower error than the other state-of-the-art methods. The experiment result shows that the proposed RA-MSFM exhibits high accuracy in measuring focus level.

TABLE 6. Subjective II & objective correlation of refocusing metrics.

DB	PLCC		RMSE	
	Proposed	Rizkallah M's [17]	Proposed	Rizkallah M's [17]
I01	**0.98**	0.90	**3.01**	7.00
I02	**0.99**	0.91	**2.69**	8.00
I03	**0.94**	0.80	**5.08**	12.28
I04	**0.99**	0.93	**2.15**	9.73
I05	**0.99**	0.73	**2.80**	13.37
I06	**0.98**	0.67	**2.05**	4.57
I07	**0.97**	0.78	**2.24**	6.03
I08	**0.98**	0.85	**2.01**	5.19
I09	**0.92**	0.87	**4.51**	9.81
I10	**0.96**	0.84	**3.02**	6.92
I11	**0.97**	0.89	**2.87**	8.03
I12	**0.95**	0.90	**3.19**	7.48

(best one in bold)

An accurate focus level is essential in improving many practical applications especially in auto focusing (AF) [40], shape-from-focus (SFF) [9], digital refocusing [42], [43], etc.

In AF, the focus measure is used to determine the position of the best focused image. Auto-focus augmented reality (AR) eyeglasses is one of the key AF applications [41]. Depth information is estimated in SFF with an accurate focusing measure since the local focus variation can be used as a depth cue [9]. SFF applications include robot manipulation and control [44], 3D model reconstruction [45] and manufacturing [46]. Digital refocusing enables users to choose the focus and depth-of-field for the image after capture [42]. A common interactive method is to allow the user to point-and-click on the image to choose the location of the refocus plane in practical AR/VR application [42], [43].

V. CONCLUSION

This paper presented a focal stack-based refocusing capability measure for light field images. We first addressed determination of the range and the sample step size for the focal stack generation. Second, we proposed the region-adaptive multi-scale focus measure (RA-MSFM) to evaluate focus level of each refocused image in the focal stack, then introduced a refocusing capability measure which is applicable to a light field image as a whole or single refocused images in a focal stack.

To carry out objective and subjective performance evaluations of the proposed scheme, we experimented the proposed RA-MSFM and the proposed refocusing metric.

By rendering all-in-focus images in an objective experiment using our focus measure method, we compared the proposed method with three well-known state-of-the-art focus measure methods introduced in literatures.

The all-in-focus image generated using our RA-MSFM method was markedly superior compared to the existing approaches, achieving PSNR 0.7dB, 4.6dB, and 1.5dB higher than those achieved by the respective methods of Chantara, Rizkallah, and Surh. Additionally, we undertook the subjective experiment I on focus measure method and the subjective experiment II on refocusing capability assessment for widely-used well-known light field images. Our experimental results indicate that there is significant correlation between the proposed refocusing metric and subjective scores.

The proposed RA-MSFM model can potentially be useful in many applications including all-in-focus image generation, depth map estimation, and other potential practical applications such as auto focusing AR eyeglasses, digital refocusing on user assigned area, and 3D model reconstruction with depth information, etc. The proposed refocusing metric can be used to measure refocusing capability loss after some image processing such as image compression, denoise, smoothing, tone or inverse tone mapping, etc. The proposed metric in this paper is based on a focal stack generated from $P \times P$ sub-aperture images (SAI). However, if P is small, the generated focal stack by shift and add may be not fine enough. Handling high accuracy focal stack generation from a few SAIs will be one of our future research directions. Another potential future extension of this work is the design of a perceptual focus measure which is more closely related to how the human visual system evaluates images.

REFERENCES

[1] P. Paudyal, F. Battisti, M. Sjostrom, 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.

[2] I. Viola, M. Rerabek, T. Bruylants, P. Schelkens, F. Pereira, and T. Ebrahimi, "Objective and subjective evaluation of light field image compression algorithms," in *Proc. Picture Coding Symp. (PCS)*, 2016, pp. 1–5.

[3] R. Lütolf, "A multiresolution representation for light field acquisition and processing," Ph.D. dissertation, Dept. Comput. Sci., SWISS Federal Inst. Technol., ETH Zurich, Switzerland, 2004.

[4] I. Viola, M. Øeøábek, and T. Ebrahimi, "A new approach to subjectively assess quality of plenoptic content," in *Proc. SPIE*, vol. 9971, Sep. 2016, Art. no. 99710X.

[5] Q. Zhao, F. Dai, J. Lv, Y. Ma, and Y. Zhang, "Panoramic light field from hand-held video and its sampling for real-time rendering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 4, pp. 1011–1021, Apr. 2020.

[6] Z. Cheng, Z. Xiong, and D. Liu, "Light field super-resolution by jointly exploiting internal and external similarities," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 8, pp. 2604–2616, Aug. 2020.

[7] 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.

[8] G. Yang and B. J. Nelson, "Wavelet-based autofocusing and unsupervised segmentation of microscopic images," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, vol. 3, Oct. 2003, pp. 2143–2148.

[9] S. Pertuz, D. Puig, and M. A. Garcia, "Analysis of focus measure operators for shape-from-focus," *Pattern Recognit.*, vol. 46, no. 5, pp. 1415–1432, May 2013.

[10] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 8, pp. 824–831, Aug. 1994.

[11] J. L. Pech-Pacheco, G. Cristóbal, J. Chamorro-Martínez, and J. Fernández-Valdivia, "Diatom autofocusing in brightfield microscopy: A comparative study," in *Proc. 15th Int. Conf. Pattern Recognit.*, vol. 3, Sep. 2000, pp. 314–317.

[12] Y. Sun, S. Duthaler, and B. J. Nelson, "Autofocusing in computer microscopy: Selecting the optimal focus algorithm," *Microsc. Res. Technique*, vol. 65, no. 3, pp. 139–149, Oct. 2004.

[13] C.-Y. Wee and R. Paramesran, "Image sharpness measure using eigenvalues," in *Proc. 9th Int. Conf. Signal Process.*, Oct. 2008, pp. 840–843.

[14] H. Nanda and R. Cutler, "Practical calibrations for a real-time digital omnidirectional camera," in *Proc. CVPR Tech. Sketch*, 2001, vol. 20, no. 2, pp. 1–4.

[15] J. Surh, H.-G. Jeon, Y. Park, S. Im, H. Ha, and I. S. Kweon, "Noise robust depth from focus using a ring difference filter," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6328–6337.

[16] H.-G. Jeon, J. Surh, S. Im, and I. S. Kweon, "Ring difference filter for fast and noise robust depth from focus," *IEEE Trans. Image Process.*, vol. 29, pp. 1045–1060, 2020.

[17] M. Rizkallah, T. Maugey, C. Yaacoub, and C. Guillemot, "Impact of light field compression on focus stack and extended focus images," in *Proc. 24th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2016, pp. 898–902.

[18] W. Chantara and M. Jeon, "All-in-focused image combination in the frequency domain using light field images," *Appl. Sci.*, vol. 9, no. 18, p. 3752, Sep. 2019.

[19] A. Levin and F. Durand, "Linear view synthesis using a dimensionality gap light field prior," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1831–1838.

[20] Y. Zhang, W. Dai, M. Xu, J. Zou, X. Zhang, and H. Xiong, "Depth estimation from light field using graph-based structure-aware analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 4269–4283, Nov. 2020.

[21] R. Yan and L. Shao, "Blind image blur estimation via deep learning," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1910–1921, Apr. 2016.

[22] R. Huang, W. Feng, M. Fan, L. Wan, and J. Sun, "Multiscale blur detection by learning discriminative deep features," *Neurocomputing*, vol. 285, pp. 154–166, Apr. 2018.

[23] Y. Pang, H. Zhu, X. Li, and X. Li, "Classifying discriminative features for blur detection," *IEEE Trans. Cybern.*, vol. 46, no. 10, pp. 2220–2227, Oct. 2016.

[24] Y. Chen, X. Jin, and Q. Dai, "Distance measurement based on light field geometry and ray tracing," *Opt. Exp.*, vol. 25, no. 1, pp. 59–76, 2017.
[25] S. Pertuz, E. Pulido-Herrera, and J.-K. Kamarainen, "Focus model for metric depth estimation in standard plenoptic cameras," *ISPRS J. Photogramm. Remote Sens.*, vol. 144, pp. 38–47, Oct. 2018.
[26] C. Hahne, A. Aggoun, S. Haxha, V. Velisavljevic, and J. C. J. Fernández, "Light field geometry of a standard plenoptic camera," *Opt. Exp.*, vol. 22, no. 22, pp. 26659–26673, 2014.
[27] R. Ng, "Digital light field photography," Ph.D. dissertation, Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, 2006.
[28] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Comput. Sci. Tech. Rep.*, vol. 2, no. 11, pp. 1–11, 2005.
[29] W. Zhou, E. Zhou, Y. Yan, L. Lin, and A. Lumsdaine, "Learning depth cues from focal stack for light field depth estimation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 1074–1078.
[30] J. Fiss, B. Curless, and R. Szeliski, "Refocusing plenoptic images using depth-adaptive splatting," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, May 2014, pp. 1–9.
[31] R. Ng, "Fourier slice photography," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 735–744, Jul. 2005.
[32] C. Hahne, A. Aggoun, V. Velisavljevic, S. Fiebig, and M. Pesch, "Refocusing distance of a standard plenoptic camera," *Opt. Exp.*, vol. 24, no. 19, pp. 21521–21540, 2016.
[33] O. Frigo, "Epipolar plane diffusion: An efficient approach for light field editing," in *Proc. Brit. Mach. Vis. Conf.*, 2017, pp. 1–13.
[34] K. Wu, Y. Yang, M. Yu, and Q. Liu, "Block-wise focal stack image representation for end-to-end applications," *Opt. Exp.*, vol. 28, no. 26, pp. 40024–40043, 2020.
[35] W. Zhang and W.-K. Cham, "Single-image refocusing and defocusing," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 873–882, Feb. 2012.
[36] M. Levoy and P. Hanrahan, "Method and system for light field rendering," U.S. Patent 6 097 394, Aug. 1, 2000.
[37] K. Takahashi, A. Kubota, and T. Naemura, "Focus measurement and all in-focus image synthesis for light-field rendering," *Syst. Comput. Jpn.*, vol. 37, no. 1, pp. 1–12, Jan. 2006.
[38] I. Viola and T. Ebrahimi, "VALID: Visual quality assessment for light field images dataset," in *Proc. 10th Int. Conf. Quality Multimedia Exper. (QoMEX)*, May 2018, pp. 1–3.
[39] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU-R Rec. BT. 500-12, 2019.
[40] H. P. Hariharan, T. Lange, and T. Herfet, "Low complexity light field compression based on pseudo-temporal circular sequencing," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, Jun. 2017, pp. 1–5.
[41] P. Chakravarthula, D. Dunn, K. Aksit, and H. Fuchs, "FocusAR: Auto-focus augmented reality eyeglasses for both real world and virtual imagery," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 11, pp. 2906–2916, Nov. 2018.
[42] M. Alain, W. Aenchbacher, and A. Smolic, "Interactive light field tilt-shift refocus with generalized shift-and-sum," in *Proc. Eur. Light Field Imag. Workshop*, 2019, pp. 1–5.
[43] C. Chang, K. Bang, G. Wetzstein, B. Lee, and L. Gao, "Toward the next-generation VR/AR optics: A review of holographic near-eye displays from a human-centric perspective," *Optica*, vol. 7, no. 11, pp. 1563–1578, Nov. 2020.
[44] I. R. Nourbakhsh, D. Andre, C. Tomasi, and M. R. Genesereth, "Mobile robot obstacle avoidance via depth from focus," *Robot. Auto. Syst.*, vol. 22, no. 2, pp. 151–158, Nov. 1997.
[45] S. Salokhiddinov and S. Lee, "Depth from focus for 3D reconstruction by iteratively building uniformly focused image set," in *Proc. ACM SIGGRAPH Posters*, 2018, pp. 1–2.
[46] Z. Jiang, D. Xu, M. Tan, and H. Xie, "An improved focus measure for MEMS assembly," in *Proc. IEEE Int. Conf. Mechatronics Autom.*, vol. 2, 2005, pp. 1118–1122.

CHUN ZHAO received the B.S. and M.S. degrees from the Department of Electronics Science and Technology, North University of China, Shanxi, China, in 2005 and 2008, respectively, and the M.S. degree, in 2008, through the Exchange Student Program. She is currently pursuing the Ph.D. degree at the Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, South Korea. From September 2008 to September 2014, she was with the Research and Design Center, Samsung Electronics, South Korea, where she worked for image/video enhancement algorithm development and system on chip (SOC) design, implement algorithm based on FPGA/Chip, and RTL design. Since 2015, she has been a Senior Engineer in visual display business at Samsung Electronics, South Korea, where she worked for practical algorithm development on LED/OLED/WOLED/QLED/QD-OLED displays by analyzing panel's characteristic. Her research interests include multimedia signal processing, panel color calibration, machine learning, and light field refocusing representation.

BYEUNGWOO JEON (Senior Member, IEEE) received the B.S. *(magna cum laude)* and M.S. degrees from the Department of Electronics Engineering, Seoul National University, Seoul, South Korea, in 1985 and 1987, respectively, and the Ph.D. degree from the School of Electrical Engineering, Purdue University, West Lafayette, USA, in 1992. From 1993 to 1997, he was with the Signal Processing Laboratory, Samsung Electronics, South Korea, where he worked for research and development of video compression algorithms, design of digital broadcasting satellite receivers, and other MPEG-related research for multimedia applications. Since September 1997, he has been with Sungkyunkwan University (SKKU), South Korea, where he is currently a Professor. His research interests include multimedia signal processing, video compression, statistical pattern recognition, and remote sensing.