Light Field Image Quality Assessment Using Contourlet Transform

Hailiang Huang[1], Huanqiang Zeng[1,2,*], Jing Chen[1], Canhui Cai[2], and Kai-Kuang Ma[3]
[1]School of Information Science and Engineering, Huaqiao University, Xiamen, China 361021
[2]School of Engineering, Huaqiao University, Quanzhou, China 362021
[3]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798
[*]Corresponding author, Email: zeng0043@hqu.edu.cn

2021 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB) | 978-1-6654-4908-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/BMSB53066.2021.9547190

***Abstract*—In this paper, a new full-reference *image quality assessment* (IQA) method for performing perceptual quality evaluation on the *light field* (LF) image is proposed, called the *contourlet transform-based model* (CTM). The LF image consists of a set of *sub-aperture images* (SAIs) that have similar image content but with small angular deviations due to different viewing angles. Hence, the abundant image details can be extracted from the SAIs. To fulfil this goal, the *contourlet* transform is used to extract the multi-scale spatial features of the reference and distorted SAIs, respectively. Based on our proposed IQA metric, the degree of similarity can be computed based on the above-measured quantities for arriving at the final IQA score of the distorted LF image under evaluation. Experimental simulation results obtained from the dense light fields datasets clearly show that the proposed CTM algorithm is more in line with the quality assessment of the LF images perceived by the *human visual system* (HVS) when compared with that of using other state-of-the-art IQA algorithms.**

***Index Terms*—light field image, image quality assessment, sub-aperture image, contourlet transform**

I. Introduction

With the growth of immersive experience highly demanded in virtual reality [1] and augmented reality [2], the *light field* (LF) image [3], [4] has been widely received as a special type of new media for conveying visual information. It has been deployed in many computer vision and computer graphics applications, such as multi-perspective imaging [5], *three-dimensional* (3-D) reconstruction [6], depth estimation [7], to name a few.

Unlike the traditional *two-dimensional* (2-D) images, the LF images consist of a set of *sub-aperture images* (SAIs), which have the small deviations occurred in the content. Due to the different viewing angles of the LF sensor, these SAIs can capture more image details from different angles. By leveraging on the spatial information provided from these, our goal is to develop a full-reference *image quality assessment* (IQA) model, called the *contourlet transform-based model* (CTM), for objectively measuring the quality of the processed LF images. Ideally, these measurements should be consistent with the judgement made by our *human visual system* (HVS).

It is well-known that some IQA models have been designed to evaluate various kinds of images, such as natural images [8]–[10], screen content images [11]–[13], and underwater images [14], to name a few. Since the LF images have different characteristics toward these images, they are not suitable for these IQA models. For that, some efforts have been devoted to the IQA models designed for LF images. Considered that the viewing angles could be exploited to reflect image details, Tian *et al.* [15] proposed a full-reference model, MDFM, by extracting multiple orders of derivatives from the SAIs as image features. Fang *et al.* [16] proposed a full-reference model to measure the local and global image quality based on the extracted gradient features. Paudyal *et al.* [17] proposed a reduced-reference IQA model, called the LF-IQM. In their work, the quality of the depth map and that of the LF image are measured, followed by computing their correlation to quantify the degree of distortion. Shi *et al.* [18] proposed a no-reference IQA model to measure the degradation of the LF image by jointly considering the spatial quality and angular consistency.

In our work, the *contourlet* transformation [19], performed as the multi-scale framework, was used for the development of a full-reference model. That is motivated by the fact that multi-scale is capable of providing both local details (using smaller scales) and global contour information (using larger scales). Here the contourlet coefficients from different scales are computed to depict the spatial features. Then the similarities between these multi-scale features are measured, followed by further combining them to compute the final quality scores of the distorted LF images. Extensive simulation experiments have been conducted on the dense light fields [20] dataset. The simulation results show that our proposed CTM model is clearly superior to existing state-of-the-art IQA models in the perceptual quality evaluation of the LF images.

The rest of this paper is organized as follows. Section II describes our developed full-reference IQA model, CTM, for conducting the IQA of LF images in detail. Section III presents experimental results of the proposed CTM and compares

This work was supported in part by the National Natural Science Foundation of China under the grants 61871434, 61802136, and 61976098, in part by the Natural Science Foundation for Outstanding Young Scholars of Fujian Province under the grant 2019J06017, in part by the Natural Science Foundation of Fujian Province under the grant 2018J01090, in part by the Fujian-100 Talented People Program, in part by the Key Science and Technology Project of Xiamen City under the grant 3502ZCQ20191005, and in part by the High-Level Talent Project Foundation of Huaqiao University under the grants 14BS201, 14BS204 and 16BS108.

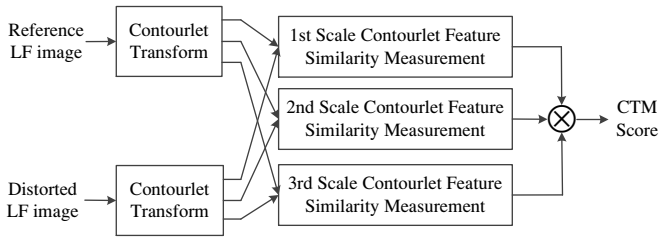

Fig. 1. The framework of the proposed *image quality assessment* (IQA) model, called the *contourlet transform-based model* (CTM), is developed for the evaluation of the *light field* (LF) image quality.

its performance with that of existing state-of-the-art models. Section IV concludes this paper.

II. Proposed Light Field Image Quality Assessment Model

In this paper, a *full*-reference LF IQA model, called the *contourlet transform-based model* (CTM), is proposed by exploiting the multi-scale spatial features extracted from the SAIs, as shown in Fig. 1. The extracted features are consistent with the perception of our HVS on comparing the reference and distorted LF images. The proposed CTM mainly consists of three stages: (1) contourlet transform (2) contourlet feature extraction; (3) contourlet-feature pooling. More details are described in the following sub-sections.

A. Contourlet Transform

According to the psychophysical findings, the operation of receptive fields in the mammalian primary visual cortex is mainly local, directional, and bandpass, and suffered specific stimuli at different scales [21]. This implies that the HVS processes visual signals via different scales, the multi-resolution and multi-directional representation of the image can match the HVS perception. Moreover, the contourlet transform [19] consists of two parts: *Laplacian pyramid* (LP) representation and *directional filter bank* (DFB) decomposition, which essentially perform such multi-resolution and multi-directional representation of the image. Furthermore, it can explore the image details from coarse level to fine level, which captures the basic spatial features. Motivated by these observations, the contourlet features extracted over the SAIs are expected to reflect the quality degradation of the LF images.

As shown in Fig. 2 (a), the contourlet transform use a combined iterated non-separable filter bank which is composed of the LP and DFB. It is performed on each SAI to extract the contourlet coefficients, and the extracted features are used to represent the spatial information of the LF images. Specifically, the LP [22] is firstly conducted on the input image a_0 to obtain a low-frequency subband a_1 and a high-frequency subband b_1. Then the above low-frequency subband is taken as the new input to iteratively perform the same decomposition. The final result of these LP decompositions is a low-frequency subband a_S and S high-frequency subbands b_s, where $s = 1, 2, \cdots, S$, representing the fine scale to coarse scale. That is, each LP decomposition generates a new high-frequency subband at the coarser scale.

Followed by the high-frequency subbands b_s are decomposed by the DFB [23] to generate the multi-directional subbands H_s^o, where the o denotes the direction index. Fig. 2 (b) depicts the composition of the DFB, it is generated by the cascade of the 2-D dual-channel filter bank highlighted in the blue box. The dual-channel filter bank consists of fan filters [24] and quincunx sampling, where the black regions represent the ideal frequency output of the fan filters, and the Q denotes the quincunx sampling matrix. They divide the 2-D spectrum into the horizontal and vertical directions, then these filtering results are reordered by the resampling matrixes R. From Fig. 2 (b), these resampling matrixes are added before and after the dual-channel filter banks to achieve the cascade structure, and realize the frequency division in 2^k directions, where k is the total number of the decomposition levels. Additionally, the contourlet decomposition process in each scale is highlighted by the dashed box shown in Fig. 2 (a), they make up the multi-directional representation of S scales.

B. Contourlet Feature Extraction

It is worthwhile to mention that these multi-directional subbands can achieve the multi-resolution and multi-directional representation of the image. Moreover, the human eyes are sensitive to the image details contained in the high-frequency components [25]. Based on the above information, the multi-directional subband coefficients of each scale are extracted to form the spatial feature map. That is,

$$C_{\xi,s}(x,y) = \bigcup_{o=1}^{2^k} H_{\xi,s}^o, \tag{1}$$

where $C_{\xi,s}(x,y)$ represent the spatial feature maps, $\xi := \{r, d\}$ represent the reference and distorted LF SAI, respectively. (x,y) denotes the coordinate of each pixel in the SAIs. $s = 1, 2, 3$ is the decomposition scale index, $o = 1, 2, \cdots, 2^k$ denotes the direction index, 2^k is the total number of directions decomposed by each scale, here $k = 4$. The symbol "$\bigcup$" denotes the union operator, which means the multi-directional subbands are spliced and combined into a large matrix.

To demonstrate, Fig. 3 shows the extracted spatial feature maps of a chosen SAI and its distorted version represented in three different scales. From which one can be observed that these feature maps from different scales can capture the distortion of the LF images. Additionally, it is worthwhile to mention that the size of these feature maps varies from the scales, that is because the length and width of the high-frequency subbands are halved along with the LP decomposition. Hence the resolution of these feature maps is different by scale, e.g., the resolution of the original image in Fig. 3 (a) is 720×960, the same as that of the 1st-scale feature map in Fig. 3 (b), and the resolution of 2nd-scale feature map in Fig. 3 (c) is halved to 360×480. Comparing these feature maps with different resolution, the feature map from the 1st

978-1-6654-4908-3/21/$31.00 ©2021 IEEE

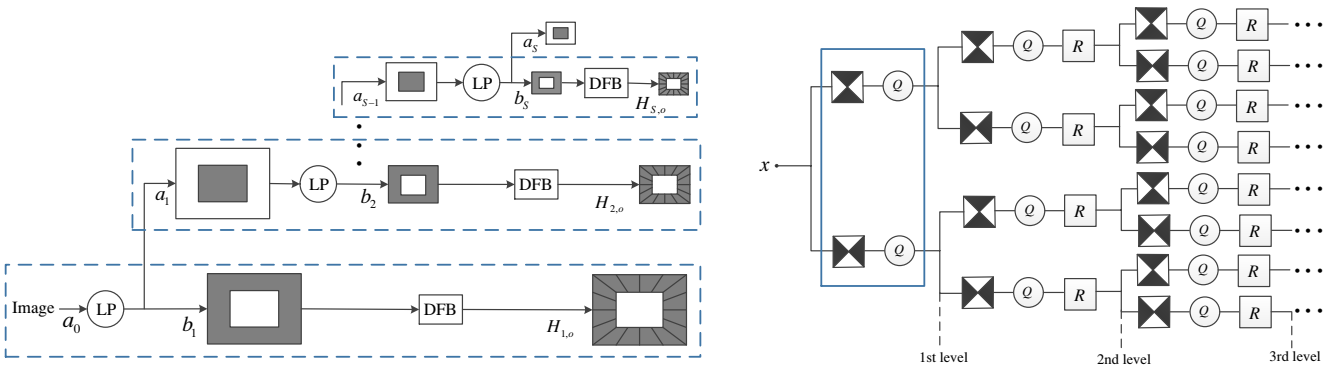

Fig. 2. An illustration of the *contourlet* transform [19]: (a) a schematic diagram of the contourlet filter bank, where the *Laplacian pyramid* (LP) is used to perform multi-scale decomposition, then each high-frequency subband is decomposed by the *directional filter bank* (DFB) to generate multi-directional subbands. The content in each dashed box represents the decomposition process of a scale. (b) The composition of the DFB, which is generated by the cascade of the 2-D dual-channel filter bank highlighted in the blue box. Where the black regions represent the ideal frequency output of the fan filters, the Q and R denote the quincunx sampling matrix and resampling matrix, respectively.

scale contains more details of the object, while the larger-scale maps reflect the coarser information. Moreover, these feature maps consist of multi-directional subbands, which can reflect the image features by different directions.

C. Contourlet-feature Pooling

The contourlet feature similarity map is obtained by the contourlet feature maps from the reference and distorted SAIs at scale $s = 1, 2, 3$, defined as:

$$SC_s(x,y) = \frac{2C_{r,s}(x,y) \cdot C_{d,s}(x,y) + c}{C_{r,s}^2(x,y) + C_{d,s}^2(x,y) + c}, \tag{2}$$

where c is a positive constant to ensure the stability of the value. Following the procedure in [9], [12], we make a test on a subset of the dense light fields dataset (i.e., 4 randomly selected reference LF images and their corresponding 96 distorted LF images). Based on the standard in the best effect of the algorithm, c is taken as 10.

As highlighted above, these contourlet features can effectively capture the key information of the images. Hence the larger values of each pixel in these feature maps can reflect the higher visual perception. The multi-scale weighted maps are generated by the larger $C_{\xi,s}(x,y)$:

$$w_s(x,y) = \max\{C_{r,s}(x,y), C_{d,s}(x,y)\}, \tag{3}$$

Therefore, the weighted contourlet feature scores can be calculated:

$$wscore_s = \frac{\sum\limits_{(x,y)} w_s(x,y) \cdot SC_s(x,y)}{\sum\limits_{(x,y)} w_s(x,y)}, \tag{4}$$

The distorted SAI quality assessment values can be computed by combining the quality scores of each scale:

$$CScore = \prod_{s=1}^{3} wscore_s, \tag{5}$$

where the symbol "$\prod$" denotes the multiplication operator.

Since the quality assessment of each SAI carries out independently, the quality score of each LF image is taken by the average of the above $CScore$. Then the natural logarithmic transformation suggested in [26] is employed on these quality scores to increase the linear relationship, that is,

$$CTM = \ln\left(\frac{1}{M}\sum_{m=1}^{M} CScore(m)\right), \tag{6}$$

where $m = 1, 2, \cdots, M$, M is the total number of the SAIs in each LF image.

III. Experimental Results and Analysis

A. LF IQA Dataset and Evaluation Criteria

To prove the effectiveness of the proposed CTM, the simulation test is conducted on the dense light fields dataset [20]. It provides 14 reference images, including 9 synthetic scenes and 5 real scenes, each scene has 101 SAIs. Reference images are affected by 6 types of distortion, including *HEVC compression* (HEVC) distortion, *linear* (LN) distortion, *nearest neighbor interpolation* (NN) distortion, *optical flow estimation* (OPT) distortion, *quantization depth maps* (DQ) distortion, and *gaussian blur* (Gauss). Each distortion is affected in 6 different levels to generate 336 distorted LF images in total.

To make a fair performance comparison of each model, the predicted scores are mapped to the same scale space as the subjective quality scores. This mapping process is conducted by the nonlinear logistic regression equation suggested in the *video quality experts group* (VQEG) HDTV test [30]. After fitting, the three commonly-used performance evaluation criteria, *Pearson linear correlation coefficient* (PLCC), *Spearman rank correlation coefficient* (SROCC), and *Root mean square error* (RMSE), are used to evaluate the performance of these IQA models. Among them, the PLCC and SROCC with the higher values mean the higher correlation between the subjective and objective quality scores. Moreover, the lower the RMSE value means the better the performance.

978-1-6654-4908-3/21/$31.00 ©2021 IEEE

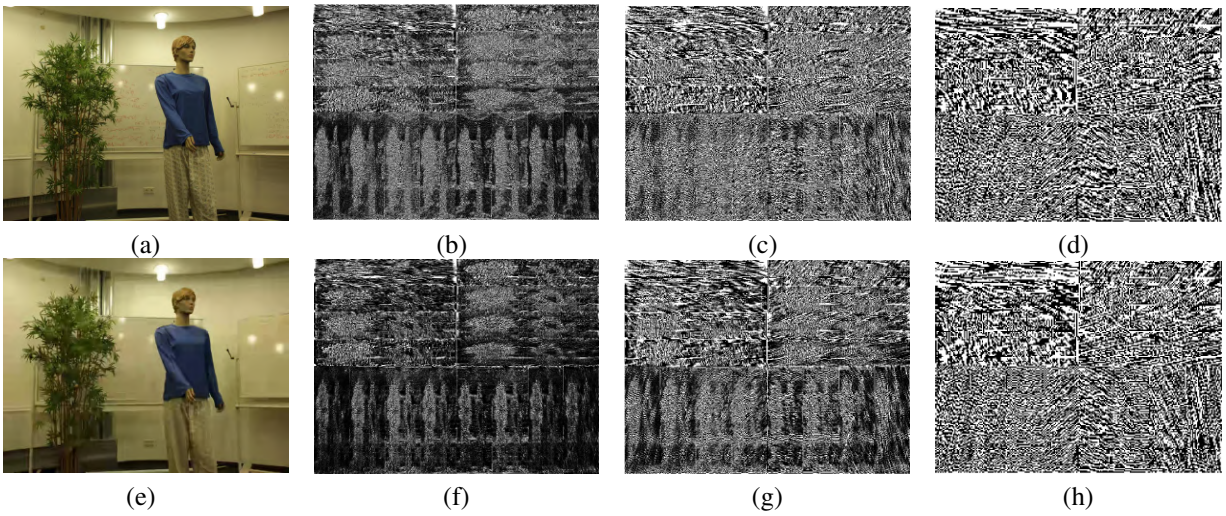

Fig. 3. An illustration of the contourlet feature map. [Row 1]: (a) an original SAI of an LF image selected from the dense light fields dataset [20], and the corresponding contourlet feature maps arrive at: (b) 1st scale; (c) 2nd scale; (d) 3rd scale. [Row 2]: (e) is the distorted version of (a), with respect to (b), (c), and (d), the (f), (g), and (h) are the corresponding distorted contourlet feature maps.

TABLE I
PERFORMANCE COMPARISONS OF THE PROPOSED AND DIFFERENT IQA MODELS ON THE DENSE LIGHT FIELDS DATASET [20].

Criteria	Distortions	PSNR	SSIM [8]	IWSSIM [27]	FSIM [9]	VIF [10]	VSI [28]	VSNR [29]	ESIM [11]	GFM [12]	Fang [16]	MDFM [15]	CTM [Ours]
PLCC	DQ	**0.9017**	0.3155	0.6460	0.5616	0.7217	0.4650	0.7403	**0.8145**	0.6837	-	**0.7933**	0.7227
	OPT	**0.7635**	0.3808	0.6679	0.5957	0.6265	0.5712	0.5541	0.5795	0.4905	-	**0.8057**	**0.7257**
	GAUSS	0.7998	0.8441	0.8224	0.8225	**0.9122**	0.8095	0.8378	**0.9266**	0.8256	-	0.8300	**0.9499**
	HEVC	0.9229	**0.9741**	0.9453	0.9391	**0.9672**	0.9406	0.9559	0.9606	**0.9879**	-	0.8790	0.8769
	LINEAR	0.8205	0.7415	0.8404	0.7454	**0.9221**	0.7460	0.8127	**0.9142**	0.8696	-	0.8653	**0.9420**
	NN	**0.9146**	0.7437	0.8801	0.8229	**0.9043**	0.8236	0.7824	0.8044	0.8393	-	0.8627	**0.9400**
	Overall	0.6250	0.6837	0.7319	0.6971	0.7860	0.6716	0.6835	0.7771	0.7421	**0.7942**	**0.8123**	**0.8435**
SROCC	DQ	**0.8761**	0.4286	0.7628	0.7598	0.7067	0.7756	**0.8332**	0.8242	0.7457	-	**0.8454**	0.7540
	OPT	**0.7976**	0.5428	0.7335	0.6859	0.6455	0.6734	0.6992	0.6351	0.6370	-	**0.7844**	**0.7483**
	GAUSS	0.8005	0.8615	0.9087	0.9045	**0.9205**	0.8954	0.8673	**0.9537**	0.8252	-	0.8355	**0.9420**
	HEVC	0.9195	0.9733	0.9831	**0.9853**	0.9818	0.9524	0.9702	**0.9849**	**0.9867**	-	0.9800	0.8563
	LINEAR	**0.9490**	0.8096	0.9144	0.8698	0.9181	0.9018	0.8706	**0.9362**	0.8728	-	0.8822	**0.9389**
	NN	**0.9262**	0.7711	**0.9256**	0.8662	0.9050	0.8907	0.9049	0.8224	0.8457	-	0.8733	**0.9422**
	Overall	0.7325	0.6972	0.8122	0.7735	0.7844	0.7666	0.7578	0.7902	0.7664	**0.8064**	**0.8346**	**0.8423**
RMSE	DQ	**0.7693**	1.6883	1.3582	1.4721	1.2316	1.5752	1.1962	**1.0323**	1.2984	-	**1.0832**	1.1951
	OPT	**1.1664**	1.6700	1.3441	1.4507	1.4076	1.4824	1.5034	1.4718	1.5738	-	**1.0697**	**1.2426**
	GAUSS	1.3423	1.1990	1.2723	1.2721	**0.9163**	1.3131	1.2209	**0.8079**	1.2618	-	1.2474	**0.6716**
	HEVC	0.9796	**0.5756**	0.8303	0.8742	**0.6461**	0.8635	0.7484	0.7075	**0.3951**	-	1.2132	1.0655
	LINEAR	1.2334	1.4476	1.1694	1.4383	**0.8347**	1.4368	1.2573	**0.8746**	1.0680	-	1.0816	**0.7164**
	NN	**0.6782**	1.1212	0.7961	0.9529	**0.7160**	0.9511	1.0445	1.1776	0.9117	-	0.8482	**0.5722**
	Overall	1.5935	1.4898	1.3910	1.4636	1.2620	1.5125	1.4900	1.2847	1.3704	**1.2300**	**1.1907**	**1.0963**

B. Performance Comparison and Analysis

To verify the performance, the proposed CTM make the comparison with multiple classic and state-of-the-art IQA models (i.e., PSNR, SSIM [8], IWSSIM [27], FSIM [9], VIF [10], VSI [28], VSNR [29], ESIM [11], GFM [12], Fang [16], MDFM [15]) on the dense light fields dataset. All experiments are conducted under the same test process and environment to ensure the fairness of the experiment. Except for Fang [16], the source codes of the above IQA models are from the authors or websites.

Table I records the overall and individual performance of the proposed CTM and various IQA models on the dense light fields dataset. For a more intuitive display, the first three excellent results of each evaluation standard are shown in bold, with the colors in red, blue, and black respectively. Compared with a variety of classic and state-of-the-art IQA models, CTM achieves the highest PLCC, SROCC values, and the lowest RMSE value in overall performance. Moreover, the result of CTM is more consistent with the subjective quality scores than that of the LF IQA models (i.e., Fang and MDFM). That is due

978-1-6654-4908-3/21/$31.00 ©2021 IEEE

to the fact that it applies the multi-scale representation to LF images, which realize a more comprehensive measurement of the distorted LF images. Furthermore, from Table I, CTM can evaluate the quality degradation by most kinds of distortion.

IV. Conclusions

In this paper, a full-reference LF IQA model is proposed, called the *contourlet transform-based model* (CTM). The contourlet transform is performed on the LF SAIs to extract the multi-scale spatial features. Then the similarity of these reference and distorted features are computed to effectively measure the quality degradation. Experimental results show that the proposed CTM has a better performance than the various classic and state-of-the-art IQA methods in evaluating the perceptual quality of LF images.

References

[1] J. Yu, "A light-field journey to virtual reality," *IEEE MultiMedia*, vol. 24, no. 2, pp. 104–112, Apr. 2017.

[2] I. Schillebeeckx and R. Pless, "Using chromo-coded light fields for augmented reality," *2016 IEEE Virtual Reality*, pp. 281–282, 2016.

[3] G. Wu, B. Masia, A. Jarabo, Y. Zhang, and L. Wang, "Light field image processing: an overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 926–954, Oct. 2017.

[4] H. Huang, H. Zeng, Y. Tian, J. Chen, J. Zhu, and K.-K. Ma, "Light field image quality assessment: an overview," *2020 IEEE Conference on Multimedia Information Processing and Retrieval*, vol. 15, pp. 348–353, 2020.

[5] S. Fujigaki, K. Kodama, and T. Hamamoto, "Multi-view imaging system using paraboloidal mirror arrays for efficient acquisition of dynamic light fields," *2019 IEEE International Conference on Image Processing*, pp. 3532–3536, 2019.

[6] J. Peng, Z. Xiong, Y. Zhang, D. Liu, and F. Wu, "LF-fusion: dense and accurate 3D reconstruction from light field images," *2017 IEEE Visual Communications and Image Processing*, pp. 1–4, 2017.

[7] J. Chen, J. Hou, Y. Ni, and L. Chau, "Accurate light field depth estimation with superpixel regularization over partially occluded regions," *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 4889–4900, Oct. 2018.

[8] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

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

[10] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[11] Z. Ni, L. Ma, H. Zeng, J. Chen, C. Cai, and K.-K. Ma, "ESIM: Edge similarity for screen content image quality assessment," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4818–4831, Oct. 2017.

[12] Z. Ni, H. Zeng, L. Ma, J. Hou, J. Chen, and K.-K. Ma, "A gabor feature-based quality assessment model for the screen content images," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4516–4528, Sep. 2018.

[13] Y. Fu, H. Zeng, L. Ma, Z. Ni, J. Zhu, and K.-K. Ma, "Screen content image quality assessment using multi-scale difference of gaussian," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 9, pp. 2428–2432, Sep. 2018.

[14] P. M. Hmue and S. Pumrin, "Image enhancement and quality assessment methods in turbid water: a review article," *2019 IEEE International Conference on Consumer Electronics*, pp. 59–63, 2019.

[15] Y. Tian, H. Q. Zeng, L. Xing, J. Chen, J. Q. Zhu, and K.-K. Ma, "A multi-order derivative feature-based quality assessment model for light field image," *Journal of Visual Communications and Image Representation*, vol. 57, pp. 212–217, Nov. 2018.

[16] Y. Fang, K. Wei, J. Hou, W. Wen, and N. Imamoglu, "Light field image quality assessment by local and global features of epipolar plane image," *2018 IEEE Fourth International Conference on Multimedia Big Data*, pp. 1–6, 2018.

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

[18] L. Shi, W. Zhou, Z. Chen, and J. Zhang, "No-reference light field image quality assessment based on spatial-angular measurement," *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.

[19] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.

[20] V. Adhikarla, M. Vinkler, D. Sumin, R. Mantiuk, K. Myszkowski, H. Seidel, and P. Didyk, "Towards a quality metric for dense light fields," *2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[21] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *Journal of Physiology*, 1962.

[22] P. Burt and E. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, 1983.

[23] R. H. Bamberger and M. J. T. Smith, "A filter bank for the directional decomposition of images: theory and design," *IEEE Transactions on Signal Processing*, vol. 40, no. 4, pp. 882–893, Apr. 1992.

[24] M. Vetterli, "Multidimensional subband coding: Some theory and algorithms," *Signal Process*, vol. 6, no. 2, pp. 97–112, Feb. 1984.

[25] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[26] X. Min, J. Zhou, G. Zhai, P. Le Callet, X. Yang, and X. Guan, "A metric for light field reconstruction, compression, and display quality evaluation," *IEEE Transactions on Image Processing*, vol. 29, pp. 3790–3804, 2020.

[27] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, May 2011.

[28] L. Zhang, Y. Shen, and H. Li, "VSI: a visual saliency-induced index for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4270–4281, Oct. 2014.

[29] D. Chandler and S. Hemami, "VSNR: a wavelet-based visual signal-to nosie ratio for natural image," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.

[30] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," *http://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx*, Aug. 2015.

978-1-6654-4908-3/21/$31.00 ©2021 IEEE