

## Research Article

# A Novel Light Field Image Compression Method Using EPI Restoration Neural Network

Jinghuai Liu <sup>1</sup>, Qian Zhang <sup>1</sup>, Ang Shen <sup>2</sup>, Ying Gao <sup>1</sup>, Jiaqi Hou <sup>1</sup>, Bin Wang <sup>1</sup>, and Tao Yan <sup>3</sup>

<sup>1</sup>College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 200234, China

<sup>2</sup>Distribution Grid Dispatching and Control Center, State Grid Qingdao Power Supply Company, Qingdao, Shandong 266001, China

<sup>3</sup>School of Mechanical, Electrical & Information Engineering, Putian University, Putian, Fujian 351100, China

Correspondence should be addressed to Qian Zhang; [qianzhang@shnu.edu.cn](mailto:qianzhang@shnu.edu.cn)

Received 12 March 2022; Accepted 23 May 2022; Published 13 June 2022

Academic Editor: Wen Si

Copyright © 2022 Jinghuai Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Different from traditional images, light field images record not only spatial information but also angle information. Due to the large volume of light field data brings great difficulties to storage and compression, light field compression technology has attracted much attention. The epipolar plane image (EPI) contains a lot of low rank information, which is suitable for recovering the complete EPI from a part of EPI. In this paper, a light field image coding framework based on EPI restoration neural network has been proposed. Compared with previous algorithms, the proposed algorithm further takes advantage of the inherent similarity in light field images, and the proposed framework has higher performance and robustness. Experimental results show that the proposed method has superior performance compared to the state-of-the-art both in quantitatively and qualitatively.

## 1. Introduction

With the development of new technology, the three-dimensional world can be more abundant and immersive sampling. Light field imaging technology is one of the main ways to collect three-dimensional scene information. Objects and scenes in the real world can be represented by the light field, which is widely used in refocusing, view synthesis, depth estimation, three-dimensional reconstruction, and so on. The all-optical function  $I = P(x, y, z, \theta, \phi, \omega, t)$  was proposed by Adelson and Bergen [1], which was used to describe the position, direction, time, and wavelength information of light in three-dimensional space. Nevertheless, it was difficult to record and process this kind of high-dimensional information. Levoy and Hanrahan [2] further simplified the full-light model and proposed a four-dimensional light field model  $L(u, v, s, t)$ , where  $(u, v)$  was used to describe the angle information of the light field and  $(s, t)$  was used to describe the spatial information of

each angle of the light field. The four-dimensional light field model is shown in Figure 1.

The light field images contain not only angular information of light distribution but also spatial information, which means that the light field contains a large amount of high-dimensional redundant data. Raw data can range from hundreds of megabytes to thousands of gigabytes depending on light field sampling and capture equipment. For example, the Lytro Illum (microlens array light field camera) can collect light from 225 directions, generating about 1.82 GB of light field data [3]. In order to store, transmit, and render light field images, a more efficient encoding scheme is required.

The mainstream light field image coding methods are closely related to video coding methods. For example, a sub-aperture image stream compression scheme is proposed in [4], which takes a subaperture image as a frame in a video and encodes it with standard video compression methods. Yet, only using video compression methods is not enough;

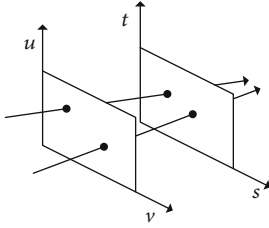


FIGURE 1: Four-dimensional light field model.

it does not take advantage of the relationship between nonadjacent subaperture images. Monteiro et al. [5] optimized the intrablock prediction model of HEVC, not only using the translation prediction model but also adding the projection and bilinear transformation model. Jia et al. [6] combine deep learning algorithms with HEVC. Unsampling subaperture images are predicted by generation adversarial networks from sampled subaperture images. However, this deep learning-based framework relies heavily on the previous training data. Once the content of light field data changes, it will face a very time-consuming retraining process.

In this paper, we consider that there is a large amount of low-rank information in EPI and propose an efficient light field image compression method. Wu et al. [7] proposed an algorithm framework for recovering a complete EPI from a sampled EPI. We get inspiration from this and propose a framework of light field image coding algorithm. Therefore, we integrate this structure into the light field image coding framework. Firstly, the optical field subaperture image is sampled and encoded with VVC. Then, in the decoding stage, the sampled subaperture image is transformed into EPI form. Finally, a recovery algorithm module is used to restore a complete EPI. The main contributions of this paper are listed as follows:

- (i) The paper presents a novel framework for the compression of light field images by using the relationship between subaperture images and EPI
- (ii) The EPI is sampled and then restored to facilitate the compression of the low-rank part
- (iii) The proposed compression scheme presents a generalized solution for light field images

The remainder of this paper is organized as follows: In Section 2, several relevant solutions about light field image compression are described. Section 3 introduces the proposed light field image coding framework. In Section 4, the simulated experimental results and test conditions are presented. Ultimately, this paper is concluded in Section 5.

## 2. Related Work

Several coding schemes of light field images described in the literature all take advantage of the redundancy of light field images. These schemes depend on different image represen-

tation and coding techniques, and their basic methods can be divided into three categories: transform-based coding, pseudo video sequence coding, and prediction-based coding.

Some low-frequency signal coding schemes essentially rely on the use of the transform, such as Discrete Cosine Transform (DCT) [8] or Discrete Wavelet Transform (DWT) [9]. After several micro images are stacked together, 3D-DCT is used to remove the spatial redundancy between adjacent micro images. The resulting transformation coefficients are then quantified and entropy encoding. In [9], light field images are decomposed into view images, and 3D-DWT is performed on a group of view images. 2D-DWT is used to transform the low-frequency part, and then, arithmetic coding is carried out, while the other high-frequency coefficients are simply quantized and arithmetic coded. Since the light field image can be regarded as a four-dimensional form, the 4D-DWT is used to compress the light field images in [10]. In [11], a light field coding method based on 4D-DCT is proposed, which introduces four-dimensional quantization and block traversal. In [12], each subview image is regarded as the node of the graph, and the similarity between adjacent views is regarded as the edge of the graph. At the coding end, the optimal graph representation is obtained by minimizing the sparsity and connectivity of the graph. According to the graph representation, the reference viewpoint is selected, and the light field image is compressed by graph transform.

Recently, HEVC is used to process light field data in many solutions [13]. HEVC specializes in using internal prediction tools to eliminate redundancy in traditional two-dimensional images. Nevertheless, it does not carry out specific processing according to the characteristics of the light field image, so many researchers put forward corresponding improvement measures. The method based on pseudo video sequence was studied in [14], mainly changing the order of subaperture images to deal with the redundant interference between subaperture images. Various scanning sequences, such as rotation sequence [4], raster scanning sequence [14], horizontal zigzag, and U-shaped mixed scanning sequence [15], are studied in the process of pseudo video sequence generation. Liu et al. improved on HEVC [16]; the prediction unit (PU) is divided into three different categories, and different prediction models based on Gaussian process regression are used for each category. In [17], a new light field multiview coding prediction structure is designed, which extends the interview prediction to a bidirectional parallel structure, and analyzes the relationship between the prediction structure and the coding performance. Because of the high similarity between subaperture images, Helin et al. [18] proposed a method of lossless compression with correction for low-frequency images.

With the improvement of hardware performance, deep learning technology has developed rapidly, attracting more and more attention from researchers and has been applied in many fields. For example, Coherence Constrained Graph LSTM (CCG-LSTM) with Spatio-Temporal Context Coherence (STCC) and Global Context Coherence (GCC) are proposed to recognize group activities [19]. Skeleton-joint Coattention Recurrent Neural Networks are proposed to generate future motions based on the observed human

motions [20]. Expansion-Squeeze-Excitation Fusion Networks (ESE-FN) are proposed to recognize elderly activity [21]. Chen et al. [22] proposed a prediction scheme based on HEVC, in which a group of the sparse subaperture images is encoded in a base layer. Other subaperture images in the decoder are reconstructed by using interpolation. The reconstructed image is then used to predict the entire light field image, and a prediction residual is transmitted. In [23], the author uses a convolutional neural network to predict all the views from the four corners, while the author in [24] uses a DIBR (depth-image-based rendering) method, by estimating the depth and using the depth-reference view to predict the other views. In [25], firstly, a set of selected subaperture images are taken as reference views and then encoded into video sequences by HEVC and transmitted to the decoder. Finally, the unselected image is reconstructed from the decoded subset of the selected image by using the low-frequency sparsity of the angular continuous Fourier domain.

Monteiro et al. [5] proposed a two-stage block high-order prediction model by using eight degrees of freedom geometric transformation. In [26], a minimum mean square error estimation method based on scalable kernel is proposed to accelerate the prediction process of light field image coding. Based on the inherent nonlocal spatial redundancy of low-frequency images, a coding method combined with local linear embedding is proposed in [27]. Monteiro et al. [28] proposed a method based on locally linear embedding and a compensation prediction method. In [24], a sparse predictor is used to predict multiple reference points. The disparity compensation wavelet coding technology is applied in [29]. Huang et al. [3] studied content consistency and structure consistency to achieve efficient light field compression at a low bit rate.

As mentioned above, according to the characteristics of light field images, the above methods combine 4D-DCT, linear approximation, convolutional neural network, and other methods with traditional video coding algorithms to achieve the purpose of compression of light field images. However, there are some problems with this approach. For example, the linear relationship between nonadjacent subaperture images is not properly handled, the coding complexity is high, and the data need to be trained in advance and time-consuming. In this paper, the relationship between subaperture images can be linked by EPI, which can be restored from a partial EPI. Therefore, we can compress the light field image by the characteristic of EPI.

### 3. Proposed Coding Solution

As EPI contains both angle information and spatial information of light field, the light field image compression method proposed in this paper focuses on processing of EPI. As shown in Figure 2, EPI consists of pixels of the same number of rows of each subaperture image. We consider sampling the EPI in coding and then recovering the original EPI in decoding, so as to reduce the amount of data transferred. Since the parallax between adjacent subaperture images of dense light field is less than 1 pixel, the sampling of EPI

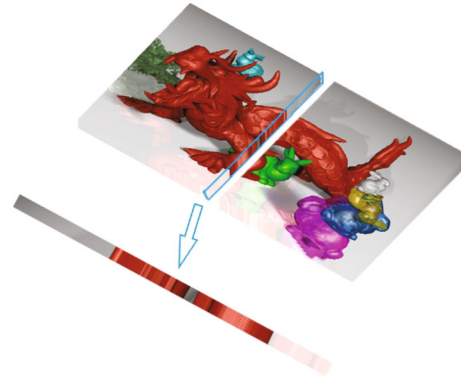


FIGURE 2: EPI generation method. Stacking a row of subaperture images, the cross section is the epipolar plane image.

meets the Nyquist sampling rate, and the original EPI can be restored according to the all-optical function.

Due to the characteristics of light field image, it contains not only spatial domain information but also angular domain information. The light field EPI connects the spatial information of the light field with the angle information in its own way. VVC is optimized on the basis of HEVC and obtains the maximum coding gain as far as possible by improving every link of coding. VVC can save about 50% bit-rate compared with HEVC at the same quality. Therefore, this paper starts with EPI, builds an image coding network based on EPI enhancement, and combines it with the existing video coding standard VVC to obtain a more efficient compression coding scheme for light field images.

Wu et al. [7] proposed a light field image reconstruction network. Its main idea is to add blur before recovering sampled images using CNN and remove blur after upsampling, which can make reconstructed images have better quality than those without the process of “blur-restoration-deblurring.” Based on this framework, this paper constructs a more suitable frame for light field image coding. Due to different problems to be solved, Wu et al.’s problem is to reconstruct the light field with high angular resolution from the light field with low angular resolution, and they do not know the true value of the light field with high angular resolution to be reconstructed in advance. However, in this paper, the original high-resolution light field is sampled and reduced to the light field with low angular resolution. This process is then reconstructed into the original high-resolution light field, the values of which are known. Therefore, in this paper, two-dimensional kernel function is used to blur the original high-resolution EPI of the light field, instead of the one-dimensional function used after sampling. In this way, the blur can be carried out in two-dimensional as a whole, instead of in one-dimensional direction, so that the reconstructed light field image quality is higher.

The whole workflow of the light field image compression coding framework proposed in this paper is shown in Figure 3. Firstly, the original light field image is transformed into EPI form; EPI is blurred by two-dimensional kernel function. Secondly, EPI is downsampled. Finally, VVC

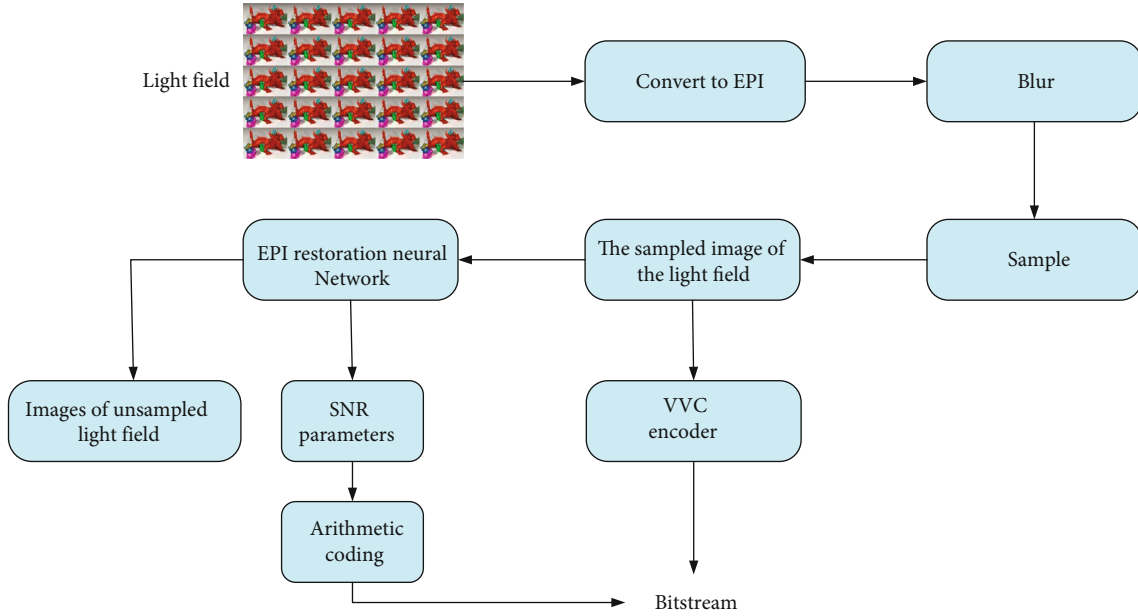


FIGURE 3: The framework of light field image coding. Firstly, the light field images are selected by the select module. The selected subaperture images are converted into EPIs through the rearrange module. Restore the EPIs through the restoration module. Finally, the selected subaperture image and the residual value are transmitted to the VVC encoder.

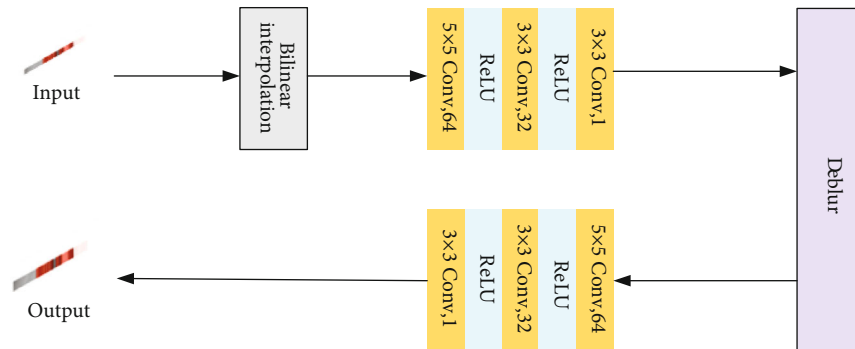


FIGURE 4: Network framework of recovery model.

encoder is used to encode the selected sampled light field image, and EPI recovery network is used to recover the sampled light field image for the noncoded part.

The overall framework of the EPI recovery model is shown in Figure 4. The input is the EPI after blur processing and then sampling. First, the pixels to be restored are initialized by bilinear interpolation, and low-frequency details are recovered by three-layer convolutional network. Then, a deblur module is used to recover the high frequency detail features. Finally, the noise generated in the deblur process is removed through three-layer convolutional network, and the restored EPI is obtained.

Bilinear interpolation is used to initialize the pixels to be restored, which is convenient for subsequent convolutional neural network processing. The three-layer convolutional network is used to recover low frequency details of EPI. The first layer contains 64 convolution kernels with a size of  $1 \times 5 \times 5$  for feature extraction. The second layer

contains 32 convolution kernels with a size of  $64 \times 3 \times 3$  for nonlinear mapping. The last layer contains a convolution kernel with a size of  $32 \times 3 \times 3$  for detail reconstruction. The activation function between the first and second layers is ReLU. Using ReLU is more efficient than other activation functions, does not have the problem of gradient disappearance, and prevents over fitting. The EPI is filled with zeros before each convolution to keep the input and output sizes the same. In the deblur module, Wiener filtering is used to deblur. Use arithmetic encoding SNR parameter. Finally, a three-layer convolutional network is used to deal with the noise generated in the process of deblur. The first layer contains 64 convolution kernels of size  $1 \times 5 \times 5$ . The second layer contains 32 convolution kernels of size  $64 \times 3 \times 3$ . The last layer contains a convolution kernel of size  $32 \times 3 \times 3$ .

In order to ensure the high quality signal recovery and perceived quality of CNN, this paper adopts the mean

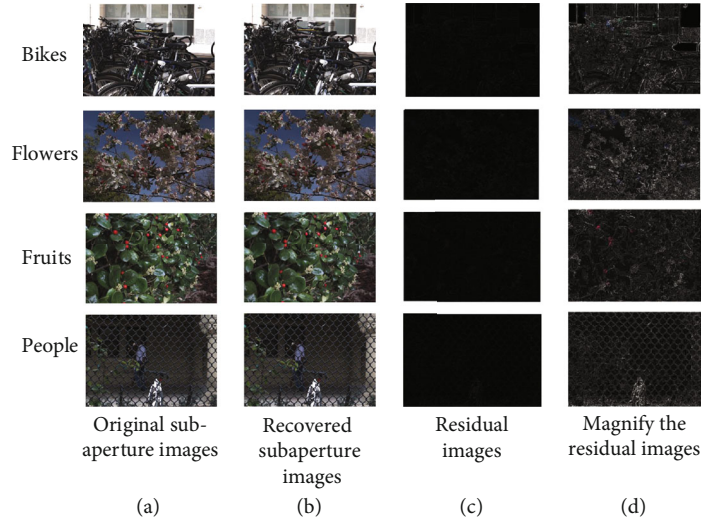


FIGURE 5: Experimental comparison of subaperture image restoration: (a) original subaperture image; (b) repaired subaperture image; (c) residual images; (d) magnify the residual images.

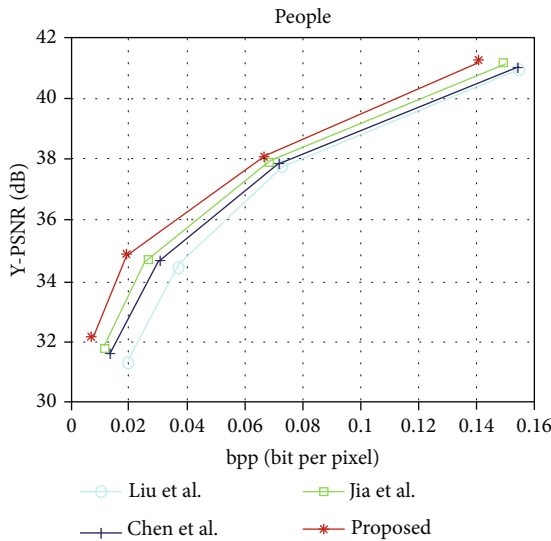


FIGURE 6: Rate-distortion curves of "People."

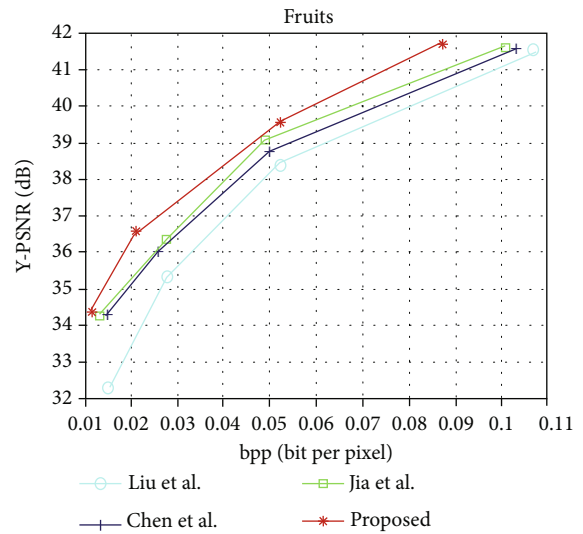


FIGURE 7: Rate-distortion curves of "Fruits."

square error loss function in the model. The formula of loss function is defined as follows:

$$L = \frac{1}{n} \sum_{i=1}^n \|E^i - E' i\|^2, \quad (1)$$

where  $n$  represents the number of trained EPI,  $E$  represents the original EPI, and  $E'$  represents the EPI for detail recovery through convolutional neural network.

#### 4. Experimental Results

In order to verify the performance of the proposed light field image coding framework, the coding efficiency of light field image is tested. A set of comparative experiments are designed to verify the performance of the proposed

method. The light field image compression coding algorithm proposed in this paper is compared with the method proposed by Liu et al. [13], Chen et al. [22], and Jia et al. [6]. During the experiment, the standard deviation of Gaussian blur kernel was set as 0.83, and the selected subaperture image after blur was constructed into pseudo video sequence according to sinuous array. VVC coding software VTM-12.2 was used to compress and encode pseudo video sequence, and the coding method was set as "Random Access." In the initialization parameter setting of convolutional neural network, the total number of iterations of training is set as, and the learning rate is set as 0.01. Each iteration, the learning rate attenuates to one-tenth of the original. The batch size is set to 64, and the momentum parameter is set to 0.9. The Gaussian distribution function with mean value 0 and standard deviation is

used to initialize the convolution kernel of the convolutional neural network.

In this paper, “People,” “Bikes,” “Fruits,” and “Flowers” in Stanford light field data set are selected for the experiment. In order to facilitate the observation of experimental results, this paper presents the original image of one of the subaperture images in each light field before the experiment, the restored result image after the experiment, and the residual image of the two images. The experimental results are shown in Figure 5. The original subaperture image is on the far left, followed by the restored subaperture image. The residual image is obtained by subtracting the pixels of the original subaperture image and the restored subaperture image. It can be seen from the figure that most areas of the residual image are black, indicating that the residual value of the image is very small, and the difference between the restored subaperture image and the original subaperture image is very small; that is, the image restoration effect of the EPI restoration model is good. In order to facilitate the observation of the difference between the original subaperture image and the restored subaperture image, the pixel value of the residual image is enlarged to obtain the right-most residual enlarged image. It can be seen that the difference is mainly concentrated in the edge part of the object in the subaperture image.

Peak signal-to-noise ratio (PSNR) is the average value of all subaperture images, and its formula is defined as follows:

$$\text{PSNR} = \frac{1}{7 \times 7} \sum_{u=1}^7 \sum_{v=1}^7 \text{PSNR}[u][v]. \quad (2)$$

The peak signal-to-noise ratio of each subaperture image is calculated as follows:

$$\text{PSNR}[u][v] = 10 * \log_{10} \left[ \frac{(2^n - 1)^2}{\text{MSE}} \right], \quad (3)$$

$$\text{MSE} = \frac{1}{s \times t} \sum_i^s \sum_j^t (\text{SAI}_r[i][j] - \text{SAI}_o[i][j])^2,$$

where  $H$  and  $W$  are the spatial resolution of each subaperture image,  $\text{SAI}_r$  and  $\text{SAI}_o$  represent reconstructed subaperture image and uncompressed subaperture image, respectively, and  $n$  indicates the bit depth.

The light field image compression coding algorithm is proposed in this paper and the rate-distortion curves of Liu et al. [13], Chen et al. [22], and Jia et al. [6], under the light field data set “People,” “Fruits,” “Bikes,” and “Flowers.” The rate-distortion curves of “People,” “Fruits,” “Bikes,” and “Flowers” are shown in Figures 6–9, respectively. It can be seen from the figure that the performance of the light field image compression and coding algorithm proposed in this paper is superior to that proposed by Liu et al., Chen et al., and Jia et al. Table 1 shows the rate-distortion performance in terms of BD-PSNR rate for each light field image. As can be seen from Table 1, the average BD-PSNR and average BD-BR of the light field image compression coding algo-

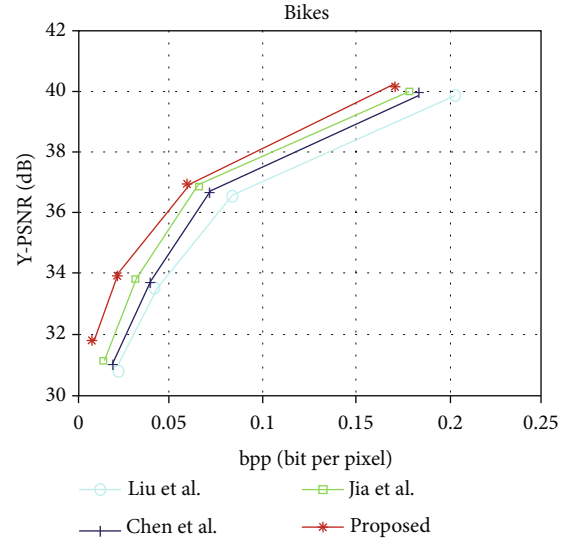


FIGURE 8: Rate-distortion curves of “Bikes.”

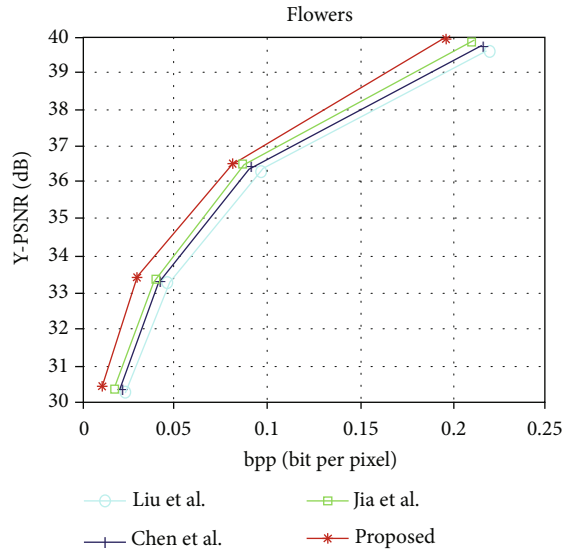


FIGURE 9: Rate-distortion curves of “Flowers.”

rithm proposed in this paper are 2.02 dB and -39.47%, respectively. Compared with the light field image compression and coding algorithm proposed by Chen et al., the average BD-PSNR increases by 0.64 dB, and the average BD-BR decreases by 15.58%. Compared with the light field image compression algorithm proposed by Jia et al., the average BD-PSNR is improved by 1.08 dB, and the average BD-BR is reduced by 24.87%. Experimental results show that the proposed image compression coding algorithm can ensure the image quality while reducing the bit rate.

In order to prove the improvement of the proposed module over the method proposed by Wu et al., an ablation experiment was added in this paper. One group directly uses the network built by Wu et al. on the basis of VVC coding; the other group uses the network structure improved in this paper. We also conducted experiments on “People,” “Fruits,”

TABLE 1: BD-PSNR (dB)/BD-BR (%): performance calculated with respect to the anchor method proposed by Liu et al.

Sequence	Chen et al. [22]		Jia et al. [6]		Proposed	
	BD-PSNR	BD-BR	BD-PSNR	BD-BR	BD-PSNR	BD-BR
Bikes	0.61	-14.50	1.26	-29.01	2.11	-47.35
Flowers	0.35	-8.13	0.62	-13.97	1.61	-31.22
Fruits	0.86	-25.39	1.10	-31.41	1.93	-38.41
People	0.75	-14.29	1.35	-25.10	2.44	-40.88
Average	0.64	-15.58	1.08	-24.87	2.02	-39.47

TABLE 2: BD-PSNR (dB)/BD-BR (%): performance calculated with respect to the anchor method proposed by Liu et al.

Sequence	Wu et al. [7]		Proposed	
	BD-PSNR	BD-BR	BD-PSNR	BD-BR
Bikes	0.73	-15.60	2.11	-47.35
Flowers	0.43	-10.67	1.61	-31.22
Fruits	0.94	-25.22	1.93	-38.41
People	0.87	-19.52	2.44	-40.88
Average	0.74	-17.75	2.02	-39.47

“Bikes,” and “Flowers” light field data sets. The experimental results are shown in Table 2. The improved light field image coding performance is better than the method proposed by Wu et al.

## 5. Conclusion

In this paper, a new light field image coding algorithm framework is constructed. In order to ensure that the quality of the original light field EPI is recovered from the sampled light field EPI, the light field EPI is blurred first and then sampled. The sampled EPI was transformed into pseudo video sequence, which was encoded by the VVC reference software VTM-12.2. The unselected light field EPI is reconstructed by EPI recovery network model. In order to test the performance of the light field image compression, compared with the light field image compression and coding algorithm proposed by Liu et al., Chen et al., and Jia et al., the bit rate is relatively reduced, and the image quality is relatively higher, which confirms the superior performance of the light field image compression.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

There is no conflict of interest regarding the publication of this paper.

## Acknowledgments

This study was supported by the Natural Science Foundation of Fujian Province (No. 2019J01816), the Putian Science and Technology Bureau (2021G2001-8), and the New Century Excellent Talents in Fujian Province University (Yantao, 2018JY7RC (PU)).

## References

- [1] E. H. Adelson and J. R. Bergen, “The plenoptic function and the elements of early vision,” in *Computational Models of Visual Processing*, pp. 3–20, MIT Press, Cambridge, MA, USA, 1991.
- [2] M. Levoy and P. Hanrahan, “Light field rendering,” in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, pp. 31–42, New York, 1996.
- [3] X. Huang, P. An, Y. Chen, D. Liu, and L. Shen, “Low Bitrate Light Field Compression With Geometry and Content Consistency,” *IEEE Transactions on Multimedia*, vol. 24, pp. 152–165, 2022.
- [4] F. Dai, J. Zhang, Y. Ma, and Y. Zhang, “Lenselet image compression scheme based on subaperture images streaming,” in *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 4733–4737, Quebec City, QC, Canada, 2015.
- [5] R. Monteiro, P. Nunes, N. M. M. Rodrigues, and S. M. M. Faria, “Light field image coding using high-order intrablock prediction,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 1120–1131, 2017.
- [6] C. Jia, X. Zhang, S. Wang, S. Wang, and S. Ma, “Light field image compression using generative adversarial network-based view synthesis,” *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 177–189, 2019.
- [7] G. Wu, Y. Liu, L. Fang, Q. Dai, and T. Chai, “Light field reconstruction using convolutional network on epi and extended applications,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, pp. 1681–1694, 2019.
- [8] A. Aggoun, “A 3D DCT compression algorithm for omnidirectional integral images,” in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, Toulouse, France, 2006.
- [9] A. Aggoun, “Compression of 3D integral images using 3D wavelet transform,” *Journal of Display Technology*, vol. 7, no. 11, pp. 586–592, 2011.
- [10] Y. Li, M. Sjostrom, R. Olsson, and U. Jennehag, “Scalable coding of plenoptic images by using a sparse set and disparities,” *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 80–91, 2016.

- [11] N. Liyanage, C. Wijenayake, C. U. Edussooriya, A. Madanayake, R. Cintra, and E. Ambikairajah, "Low-complexity real-time light field compression using 4-D approximate DCT," in *2020 IEEE international symposium on circuits and systems (ISCAS)*, pp. 1–5, Seville, Spain, 2020.
- [12] I. Viola, H. P. Maretic, P. Frossard, and T. Ebrahimi, "A graph learning approach for light field image compression," in *Applications of Digital Image Processing XLI (Vol. 10752, p. 107520E)*, International Society for Optics and Photonics, 2018.
- [13] D. Liu, L. Wang, L. Li, Z. Xiong, F. Wu, and W. Zeng, "Pseudo-sequence-based light field image compression," in *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 1–4, Seattle, WA, 2016.
- [14] C. Perra and P. Assuncao, "High efficiency coding of light field images based on tiling and pseudo-temporal data arrangement," in *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, Seattle, WA, USA, 2016.
- [15] S. Zhao, Z. Chen, K. Yang, and H. Huang, "Light field image coding with hybrid scan order," in *2016 Visual Communications and Image Processing (VCIP)*, pp. 1–4, Chengdu, China, 2016.
- [16] D. Liu, P. An, R. Ma, W. Zhan, X. Huang, and A. A. Yahya, "Content-based light field image compression method with Gaussian process regression," *IEEE Transactions on Multimedia*, vol. 22, no. 4, pp. 846–859, 2020.
- [17] G. Wang, W. Xiang, M. Pickering, and C. W. Chen, "Light field multi-view video coding with two-directional parallel inter-view prediction," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5104–5117, 2016.
- [18] P. Helin, P. Astola, B. Rao, and I. Tabus, "Sparse modelling and predictive coding of subaperture images for lossless plenoptic image compression," in *2016 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, pp. 1–4, Hamburg, Germany, 2016.
- [19] J. Tang, X. Shu, R. Yan, and L. Zhang, "Coherence constrained graph LSTM for group activity recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 2, pp. 636–647, 2022.
- [20] X. Shu, L. Zhang, G. -J. Qi, W. Liu, and J. Tang, "Spatiotemporal co-attention recurrent neural networks for human-skeleton motion prediction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 6, pp. 3300–3315, 2022.
- [21] X. Shu, J. Yang, R. Yan, and Y. Song, "Expansion-squeeze-excitation fusion network for elderly activity recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, 2022.
- [22] J. Chen, J. Hou, and L. P. Chau, "Light field compression with disparity-guided sparse coding based on structural key views," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 314–324, 2018.
- [23] X. Jiang, M. Le Pendu, and C. Guillemot, "Light fields compression using depth image based view synthesis," in *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, Hong Kong, China, 2017.
- [24] P. H. Tabus and P. Astola, "Lossy compression of lenslet images from plenoptic cameras combining sparse predictive coding and JPEG 2000," in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 4567–4571, Beijing, China, 2017.
- [25] F. Hawary, C. Guillemot, D. Thoreau, and G. Boisson, "Scalable light field compression scheme using sparse reconstruction and restoration," in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3250–3254, Beijing, China, September. 2017.
- [26] Z. You, P. An, and D. Liu, "Scalable kernel-based minimum mean square error estimate for light-field image compression," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, 14 pages, 2018.
- [27] L. F. Lucas, C. Conti, P. Nunes et al., "Locally linear embedding-based prediction for 3D holoscopic image coding using HEVC," in *2014 Proceedings of the 22nd European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, 2014.
- [28] R. Monteiro, L. Lucas, C. Conti et al., "Light field HEVC-based image coding using locally linear embedding and self-similarity compensated prediction," in *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 1–4, Seattle, WA, USA, 2016.
- [29] T. H. Tran, Y. Baroud, Z. Wang, S. Simon, and D. Taubman, "Light-field image compression based on variational disparity estimation and motion-compensated wavelet decomposition," in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3260–3264, Beijing, China, 2017.