applied sciences

MDPI

Article

End-to-End Light Field Image Compression with Multi-Domain Feature Learning

Kangsheng Ye, Yi Li *, Ge Li, Dengchao Jin and Bo Zhao

School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China; yks1001@tju.edu.cn (K.Y.); lige97@tju.edu.cn (G.L.); jdc3159761141@tju.edu.cn (D.J.); bzzb@tju.edu.cn (B.Z.)
* Correspondence: annie@tju.edu.cn

Abstract: Recently, end-to-end light field image compression methods have been explored to improve compression efficiency. However, these methods have difficulty in efficiently utilizing multi-domain features and their correlation, resulting in limited improvement in compression performance. To address this problem, a novel multi-domain feature learning-based light field image compression network (MFLFIC-Net) is proposed to improve compression efficiency. Specifically, an EPI-based angle completion module (E-ACM) is developed to obtain a complete angle feature by fully exploring the angle information with a large disparity contained in the epipolar plane image (EPI) domain. Furthermore, in order to effectively reduce redundant information in the light field image, a spatial-angle joint transform module (SAJTM) is proposed to reduce redundancy by modeling the intrinsic correlation between spatial and angle features. Experimental results demonstrate that MFLFIC-Net achieves superior performance on MS-SSIM and PSNR metrics compared to public state-of-the-art methods.

Keywords: light field image; end-to-end compression; multi-domain feature learning

check for updates

Citation: Ye, K.; Li, Y.; Li, G.; Jin, D.; Zhao, B. End-to-End Light Field Image Compression with Multi-Domain Feature Learning. *Appl. Sci.* **2024**, *14*, 2271. https://doi.org/10.3390/app14062271

Academic Editor: João M. F. Rodrigues

Received: 6 February 2024
Revised: 25 February 2024
Accepted: 26 February 2024
Published: 8 March 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the development of light field imaging technology, light field images have been widely used in many fields, including depth estimation [1], view synthesis [2], and free viewpoint rendering [3]. However, the dense viewpoints and high-resolution characteristics of light field images pose huge challenges to transmission. Consequently, many compression methods have been researched to improve the compression efficiency of light field images.

Traditionally, many light field image compression methods based on standard video codecs have been studied, which perform specific processing on light field images and compress them using existing video codecs [4,5]. Specifically, the sub-aperture images (SAIs) of the light field are first organized into pseudo video sequences with a specific order. Subsequently, existing video codecs are applied to compress these pseudo video sequences. Finally, the reconstructed pseudo video sequences are rearranged to recover the original light field SAIs. Although these methods can effectively reduce the redundancy of light field images, the effect becomes limited when the number of SAIs increases. Considering the similarity among views in light field images, some researchers have studied compression frameworks based on view synthesis, which encode sparse views and synthesizes the remaining views based on the disparity relationship at the decoder [6–8]. Since existing standard video codecs are not specifically tailored for light field images, it is difficult for the above methods to achieve optimal coding efficiency.

Recently, deep learning technology has developed rapidly and achieved impressive success in various tasks, such as quality enhancement [9,10], content generation [11,12], and object detection [13,14]. Inspired by this, researchers have explored many deep learning-based video coding methods [15–17]. For light field image coding, some end-to-end coding

frameworks have also been proposed, and preliminary results have been achieved [18–21]. Specifically, the first step involves the extraction of light field image features. Subsequently, a nonlinear transform is applied to light field image features to obtain a compact latent representation for efficient entropy coding. Therefore, the coding performance of end-to-end light field image compression methods mainly depends on the complete extraction of light field image features and the efficient compression of image features.

In order to completely extract light field image features, several methods have been proposed, such as data structure adaptive 3D convolution (DSA-3D) [18], space-angle decoupling mechanism [19], and disparity-aware model [20]. However, the aforementioned methods mainly extract spatial feature and angle feature within a limited disparity range and ignore the exploitation of the large disparity angle information in the EPI domain. As a result, this problem generally hinders the efficient compression of light field images with large disparity.

Meanwhile, some technologies have been designed for the efficient compression of light field image features by reducing unnecessary redundancy. For example, Jia et al. [21] introduced generative adversarial networks (GANs) into light field image compression, which utilize decoded sparse sampling SAIs to generate remaining light field SAIs at specific locations, thus effectively reducing the transmission of redundant information. Tong et al. [19] decoupled the spatial feature and angle feature in the light field image and reduced the redundancy of the spatial and angle features, respectively. However, these methods usually focus on reducing the redundancy within a single domain and ignoring correlations among multi-domain features, which leads to the insufficient reduction of redundancy.

To address the above problems, a novel multi-domain feature learning-based light field image compression network (MFLFIC-Net) is proposed in this paper. In the MFLFIC-Net, the complete angle feature is obtained by exploring the large disparity angle information in the EPI domain. In addition, the correlation between the spatial feature and complete angle feature is explored to effectively reduce redundancy, thereby further improving the compression efficiency. The contributions are summarized as follows:

(1) A novel multi-domain feature learning-based light field image compression network is proposed to improve compression efficiency by effectively utilizing multi-domain features and their correlation to obtain a complete angle feature and reduce the redundancy among multi-domain features.
(2) An EPI-based angle completion module is developed to obtain a complete angle feature by fully exploring the large disparity angle information contained in the EPI domain.
(3) A spatial-angle joint transform module is proposed to reduce redundancy by modeling the intrinsic correlation between spatial and complete angle features.
(4) Experimental results on the EPFL dataset demonstrate that MFLFIC-Net achieves superior performance on MS-SSIM and PSNR metrics compared to public state-of-the-art methods.

The remaining sections of this paper are organized as follows. Section 2 briefly reviews the related works on light field image compression. Section 3 provides an introduction to the details of the proposed method. The experimental setup and experimental results are introduced in Section 4. Finally, the paper is summarized in Section 5.

2. Related Works

2.1. Traditional Light Field Image Compression

Traditional methods typically explore compression techniques suitable for light field images based on existing video codecs [22–24]. For example, Liu et al. [22] arranged the light field SAIs into pseudo-video sequences and compressed sequences using the standard video codec. Dai et al. [25] arranged all SAIs into two scan orders, i.e., the spiral order and the zigzag order, and encoded them with the video codec, thereby effectively reducing spatial redundancy in light field images. Different from the fixed arrangement order,

Porto et al. [23] designed a context-adaptive coding scheme that calculates the sum of absolute differences (SADs) among SAIs, and then dynamically arranges SAIs according to the minimum SAD, thus effectively reducing inter-frame prediction residual. To further improve the prediction accuracy, Liu et al. [24] divided the prediction unit (PU) into three different categories based on the content attributes, and designed a Gaussian process regression-based prediction method to improve compression efficiency. However, the effect becomes limited when the number of SAIs increases, making it difficult to further improve compression efficiency.

The recent standard Joint Photographic Experts Group Pleno (JPEG Pleno) only encodes selected SAIs and uses disparity information at the decoder to reconstruct unselected SAIs, which could significantly reduce the encoding bitrate and provide a new direction. Inspired by JPEG Pleno, Huang et al. [6] transmitted half of the SAIs and their corresponding disparity maps, which improves the reconstruction quality with the cost of more bit consumption. To further reduce bit consumption, Huang et al. [7] proposed a low-bitrate compression method with geometry and content consistency that achieves low-bitrate light field image compression while ensuring reconstruction quality. However, existing video codecs are not tailored for the unique characteristics of light field images, making it challenging to attain optimal compression efficiency.

2.2. Learning-Based Light Field Image Compression

With the rapid development of deep learning, deep learning technology has been widely applied to various image processing tasks and has achieved significant success. Inspired by this, some researchers have devoted to studying learning-based light field image compression methods [26–28]. The learning-based compression methods can be divided into two categories: using deep learning tools to enhance the traditional light field image compression framework and using deep learning frameworks to perform end-to-end compression of light field images.

In terms of deep learning tools, these methods have achieved impressive coding performance improvements by embedding deep learning tools into traditional light field image compression frameworks. Bakir et al. [26] proposed a light field image compression method based on view synthesis, using Versatile Video Coding (VVC) to compress the sparsely sampled SAIs, and introduced the efficient dual discriminator GAN model to generate the remaining SAIs at the decoding side. Zhao et al. [27] introduced a multi-view joint enhancement network to suppress compression artifacts, thereby improving coding performance. To further reduce bit consumption, Van et al. [28] proposed a downsampling-based light field image encoding framework that downsamples the light field image at the encoder and introduces a super-resolution network at the decoder. However, these methods only enhance a single module using deep learning tools, lacking the capability to achieve comprehensive optimization of the compression framework.

In recent years, end-to-end image compression methods have been extensively studied and achieved impressive performance. As a key component, entropy coding plays a vital role in the encoding and decoding stages of data. Researchers continue to optimize the use of entropy coding to improve compression efficiency and image quality while reducing data transmission and storage costs. For example, Ballé et al. [29] proposed a hyperprior entropy model based on side information, which assists in the parameter generation of the entropy model by capturing the hidden information of the latent representation. On the basis, Minnen et al. [30] proposed an autoregressive context model that accurately capture the correlation between pixels in an image by fully utilizing contextual information, thereby achieving more effective image compression. These end-to-end methods usually adopt a compression autoencoder framework, which is overall optimized by the rate-distortion (RD) function. Inspired by this, some researchers have devoted themselves to developing end-to-end light field image compression methods. For instance, Singh et al. [20] proposed a novel learning-based disparity assist model to compress four-dimensional light field images, which ensures the structural integrity of reconstructed light field images by

introducing disparity information. Zhong et al. [18] introduced a data structure adaptive 3D-convolutional autoencoder, which effectively remove artifacts generated during the compression process and achieve better compression efficiency. Tong et al. [19] proposed a spatial-angle decoupling network that reduces redundancy in spatial and angle features, respectively, thereby improving the compression efficiency. However, the above light field image compression methods cannot effectively exploit multi-domain features and their correlation, which prevents further improvement in compression efficiency.

3. Method

3.1. Architecture of MFLFIC-Net

To enhance the compression efficiency of light field images, a MFLFIC-Net is proposed, which exploits multi-domain features and their correlation to obtain a complete angle feature and reduce the redundancy among multi-domain features. Specifically, the large disparity angle information contained in the EPI domain is explored to obtain a complete angle feature, which contains angle information on the complete disparity range. In addition, the intrinsic correlation between spatial and complete angle features is further exploited to reduce redundant information in features.

As shown in Figure 1, the proposed network consists of three parts: an encoder, an entropy coding, and a decoder. The encoder is employed to generate a latent representation from the original image. In the entropy coding phase, the latent representation is compressed into a bitstream for transmission. Finally, the decoder, which consists of a spatial-angle joint transform module and reconstruction module, is utilized to reconstruct the image from the reconstructed latent representation. In detail, the feature extractors are first used to map the original light field image x to the feature space for obtaining spatial feature f_S, angle feature f_A, and EPI feature f_E. Subsequently, the E-ACM is designed to obtain the complete angle feature f_{A^*}. Then, the SAJTM is designed to effectively reduce the redundant information between f_S and f_{A^*}, and obtain the compact spatial feature $f_S^{c_2}$ and the compact angle feature $f_{A^*}^{c_2}$. After that, $f_S^{c_2}$ and $f_{A^*}^{c_2}$ are channel-concatenated and pass through a convolution layer to obtain the latent representation y. Then, y is entropy encoded and decoded by the entropy coding. Finally, the reconstructed latent representation $\hat{y}$ passes through the decoder to obtain the reconstructed light field image $\hat{x}$.

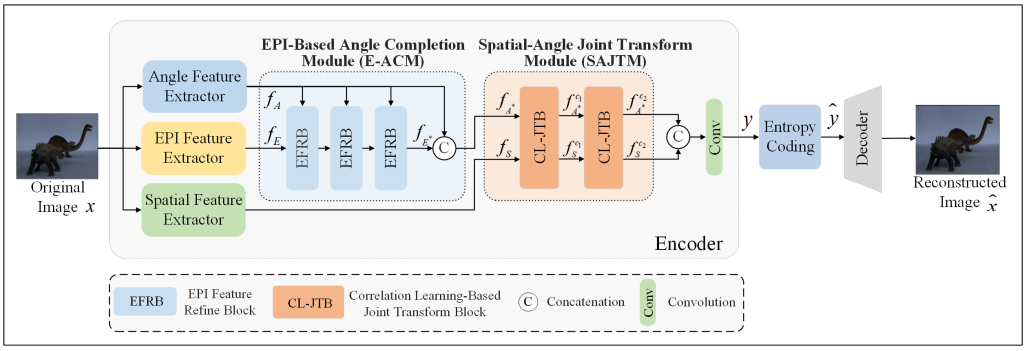


Figure 1. Architecture of MFLFIC-Net. The original image x first passes through the encoder, which consists of a series of feature extractors, E-ACM, and SAJTM, to obtain a compact latent representation y. Specifically, the E-ACM consists of three EPI feature refine blocks (EFRBs), and the SAJTM consists of two correlation learning-based joint transform blocks (CL-JTBs). After obtaining the reconstructed latent representation $\hat{y}$ from entropy coding, the reconstructed light field image $\hat{x}$ is finally generated through the decoder.

3.2. EPI-Based Angle Completion Module

The angle domain feature obtained by feature extractor only contains angle information within a limited disparity range. To improve the coding performance of large disparity light field images, an E-ACM is designed by fully exploiting the large disparity angle information contained in the EPI domain. This module takes the angle feature f_A and the

EPI domain feature f_E as inputs, and obtains a complete angle feature f_{A^*}, containing angle information on the complete disparity range.

As shown in Figure 1, the E-ACM consists of three EPI feature refine blocks (EFRBs) and a concatenation operation. Specifically, the EFRB is designed to obtain the complementary feature f_{E^*} from f_E. Then, f_{E^*} and the input f_A are concatenated in the channel dimension to obtain the complete angle feature f_{A^*}. The specific calculation process is as follows:

$$f_{A^*} = f_{E^*} \oplus f_A, \tag{1}$$

$$f_{E^*} = E(f_A, f_E), \tag{2}$$

where $E(\cdot)$ consists of three EPI feature refine blocks, and $\oplus$ denotes the channel-wise concatenation.

The architecture of the proposed EFRB is shown in Figure 2. Taking an EFRB as an example, the input angle domain feature f_A^{in} and the EPI domain feature f_E^{in} are first fed into two residual blocks to generate features $f_A^{'}$ and $f_E^{'}$. Then, a mutual attention mechanism is applied to model the similarity relationship between $f_A^{'}$ and $f_E^{'}$. Subsequently, nonlinear transform is performed after channel-wise concatenation to obtain the common feature f_{EA}. The above process can be formulated as

$$f_E^{'} = R(f_E^{in}), \tag{3}$$

$$f_A^{'} = R(f_A^{in}), \tag{4}$$

$$f_{EA} = h_L(F_L(f_A^{'}, f_E^{'}) \oplus f_E^{'}), \tag{5}$$

where $F_L(\cdot)$ denotes the mutual attention function [31]. $h_L(\cdot)$ consists of two convolution layers and an activation layer, and $\oplus$ denotes the channel-wise concatenation. $R(\cdot)$ is composed of two residual blocks. Finally, the complementary feature $f_{E^*}^{out}$ with large disparity angle information is extracted from $f_E^{'}$. The specific calculation process is as follows:

$$f_{E^*}^{out} = g_L(f_E^{'} - f_{EA}), \tag{6}$$

where $g_L(\cdot)$ consists of a convolution layer and an activation layer.

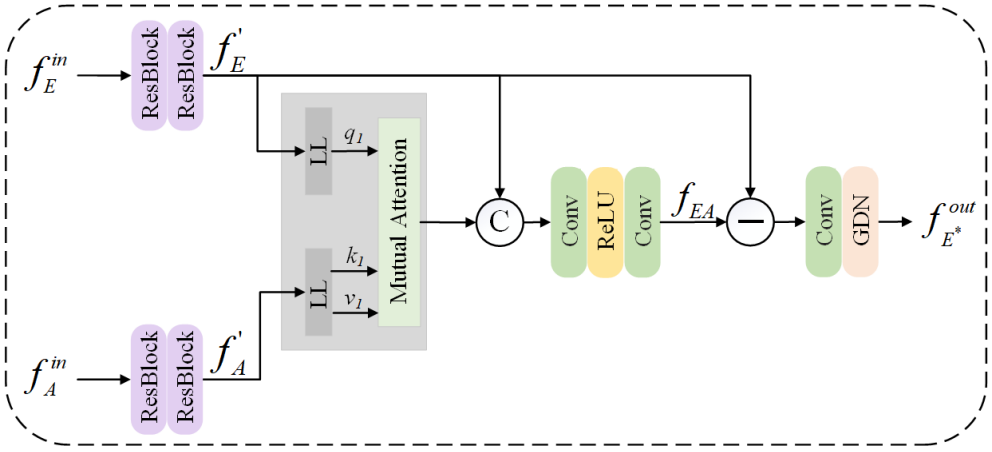


Figure 2. Illustration of EPI feature refine block.

As described above, f_{A^*} is obtained after f_A and f_E pass through E-ACM. Benefiting from the proposed E-ACM, the large disparity angle information contained in the EPI domain is fully explored, thereby improving the compression efficiency of large disparity light field images.

3.3. Spatial-Angle Joint Transform Module

To effectively reduce redundant information in light field images, a SAJTM is proposed to reduce redundancy by modeling the intrinsic correlation between spatial and complete angle features. The proposed SAJTM takes spatial feature f_S and complete angle feature f_{A^*} as inputs, and the outputs are the compact spatial feature $f_S^{c_2}$ and compact angle feature $f_{A^*}^{c_2}$.

As shown in Figure 1, the proposed SAJTM is composed of two correlation learning-based joint transform blocks (CL-JTBs). Two CL-JTBs are sequentially connected to reduce redundant information in f_S and f_{A^*}. Taking a CL-JTB as an example, its architecture is illustrated in Figure 3. To obtain a more compact spatial feature, the inputs f_S^{in} and $f_{A^*}^{in}$ are first processed through a residual block, two convolution layers, and an activation layer to generate spatial feature $f_S^{r_1}$ and angle feature $f_{A^*}^{r_1}$ with the same size, respectively. Subsequently, the spatial predictive feature f_p^s, which is the predicted redundant information that can be obtained through prediction, is obtained by modeling the intrinsic correlation between $f_S^{r_1}$ and $f_{A^*}^{r_1}$. Finally, f_p^s is subtracted from $f_S^{r_1}$ to obtain the more compact spatial feature f_S^{out}. The above process can be expressed as

$$f_p^s = F_c(R_S(f_S^{in}) \oplus R_A(f_{A^*}^{in})), \tag{7}$$

$$f_S^{out} = f_S^{r_1} - f_p^s, \tag{8}$$

where $R_S(\cdot)$ and $R_A(\cdot)$ both consist of a residual block layer, an activation layer, and two convolution layers. $F_c(\cdot)$ is composed of an activation layer and a convolution layer, and $\oplus$ denotes the channel-wise concatenation.

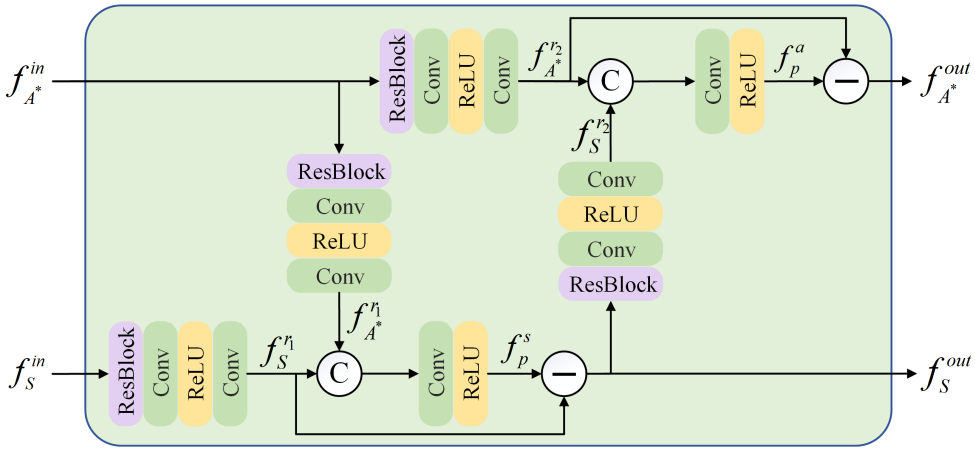


Figure 3. Illustration of correlation learning-based joint transform block.

The process of obtaining the compact angle feature is similar to the above description. The specific process can be expressed as

$$f_p^a = F_c(R_S(f_S^{out}) \oplus R_A(f_{A^*}^{in})), \tag{9}$$

$$f_{A^*}^{out} = f_{A^*}^{r_2} - f_p^a. \tag{10}$$

As mentioned above, after the input spatial feature f_S and complete angle feature f_{A^*} pass through two CL-JTBs, compact spatial feature $f_S^{c_2}$ and compact angle feature $f_{A^*}^{c_2}$ are obtained. In this way,the proposed SAJTM effectively reduces redundant information by utilizing the inherent correlation among features, thus improving compression efficiency.

3.4. Implementation Details

We adopt the same rate-distortion loss function L as [19] to train MFLFIC-Net:

$$L = \lambda D + R, \tag{11}$$

where R represents the bits consumed for encoding, and D denotes the reconstruction distortion computed by the mean square error (MSE). Following [19], the hyperparameter λ is configured with values from the set {0.0001, 0.0003, 0.0006, 0.001, 0.002, 0.004}, ensuring that the rates of all experimental results can be in the same range for fair comparison. In addition, we use the same variable code rate training strategy as SADN [19], which is divided into three phases. In the first stage, the initial network is trained for 30 epochs to obtain a single rate point model of 0.002. In the second stage, the single rate point model of 0.002 is fine-tuned using noisy quantization into a variable rate model that can be tested at multiple code rates. In the third stage, the coding performance is further improved by changing the noise quantization of the entropy model to a straight-through estimator quantization, finally obtaining a different variable code rate model.

4. Experiments

4.1. Experimental Settings

The model is trained on the PINet dataset [19], which is a large-scale natural dataset consisting of 7549 images taken by the light field camera named Lytro Illum. These images contain 178 categories, such as cats, camels, fans, bottles, etc. Each image in the PINet dataset is randomly cropped into patches of size 832 × 832 for training. The MFLFIC-Net is implemented using PyTorch and optimized with the Adam optimizer [32]. The training of the model requires approximately 21 days on a PC equipped with 4 NVIDIA GeForce RTX 3090 GPUs. In addition, it is worth noting that model training is divided into three stages, and the initial learning rate for each stage is different. The initial learning rates for the three stages are set to 1×10^{-4}, 1×10^{-5}, and 1×10^{-6}, respectively. To ensure the fairness and impartiality of the experimental comparison, all methods are tested on the commonly used EPFL dataset [33] with 12 light field images using an NVIDIA GeForce GTX 1080Ti GPU.

4.2. Comparison Results

A quantitative evaluation is performed to demonstrate the effectiveness of the proposed MFLFIC-Net, which compares MFLFIC-Net with an end-to-end light field image compression method SADN [19] and some traditional light field image compression methods, including GPR [24], HEVC reference software HM 16.9 [34], and VVC reference software VTM 10.0 [35]. It is worth mentioning that these comparison methods were selected because they have been widely used in previous related research and have achieved remarkable results. In addition, this selection covers the main methods and strategies in this field to ensure the reliability of the experimental results. Since our experimental settings are consistent with SADN [19], the experimental results of GPR [24] are obtained from the results of SADN [19]. For a fair comparison, the end-to-end light field compression method SADN [19] is retrained on the same experiment equipment as the proposed method. To objectively evaluate the compression performance, the BD-PSNR and BD-rate are utilized as the objective metrics, where HM 16.9 [34] is set as the baseline to calculate the BD-PSNR and BD-rate. As shown in Table 1, MFLFIC-Net achieves a 2.149 dB BD-PSNR increase and a 65.585% BD-rate reduction. Compared with the SADN [19], MFLFIC-Net further obtains a 0.354 dB BD-PSNR increase and a 21.329% BD-rate reduction. In addition, compared with VTM 10.0 [35], MFLFIC-Net further obtains a 1.128 dB BD-PSNR increase and a 30.028% BD-rate reduction. As shown in Table 1, the experimental results demonstrate that MFLFIC-Net achieves the highest rate-distortion performance, which verifies its effectiveness.

Furthermore, the rate-distortion curves of the MS-SSIM and PSNR metrics are shown in Figure 4. In terms of PSNR metrics, MFLFIC-Net outperforms the other methods, demonstrating the effectiveness of MFLFIC-Net. Regarding the MS-SSIM metric, as shown in Figure 4, the MFLFIC-Net also exhibits optimal performance when compared to the other methods. In addition, since the images in the test dataset do not overlap with the training set and the images are different, the test experimental results are able to verify the generalization ability of the method. Experimental results show that the proposed method has better performance than the comparative methods, which further proves that our method has stronger generalization ability.

Table 1. Comparison results of BD-rate and BD-PSNR on the EPFL dataset when using HM 16.9 as baseline.

Methods	BD-PSNR (dB)	BD-rate (%)
GPR	−2.253	138.095
VTM 10.0	1.021	−35.557
SADN	1.795	−44.256
Proposed	**2.149**	**−65.585**

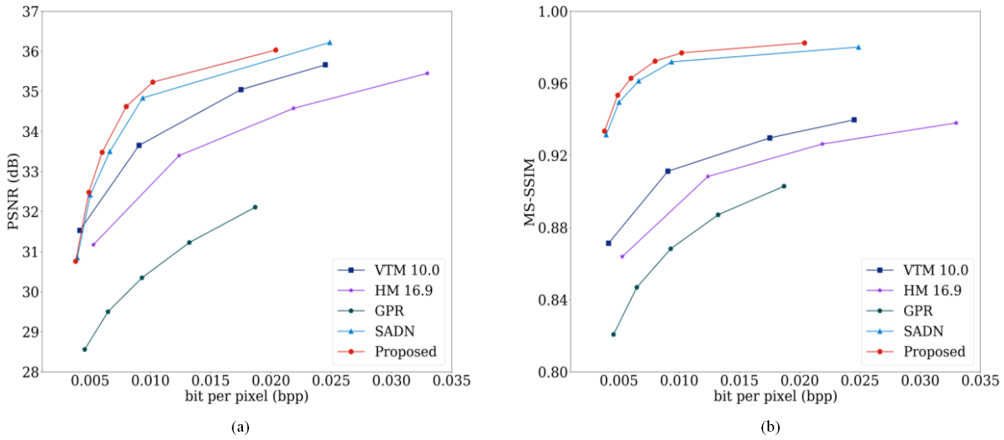


Figure 4. Rate-distortion curves . (**a**) Comparison of PSNR on the EPFL dataset. (**b**) Comparison of MS-SSIM on the EPFL dataset.

To visually verify the efficiency of the proposed MFLFIC-Net, the reconstruction results generated by SADN [19] and the proposed MFLFIC-Net are shown in Figure 5. It is worth noting that, in the visual comparison, MFLFIC-Net is only compared with SADN [19]. This is because SADN [19] significantly outperforms the other compared methods. From the visualization results, it can clearly observe that the reconstruction texture of the proposed method is clearer than SADN [19]. This is because the proposed E-ACM provide angle information with a large disparity range, which is beneficial in reconstructing light field images. At the same time, the proposed MFLFIC-Net has a lower bit consumption, which indicates that SAJTM effectively reduce redundancy and improve performance.

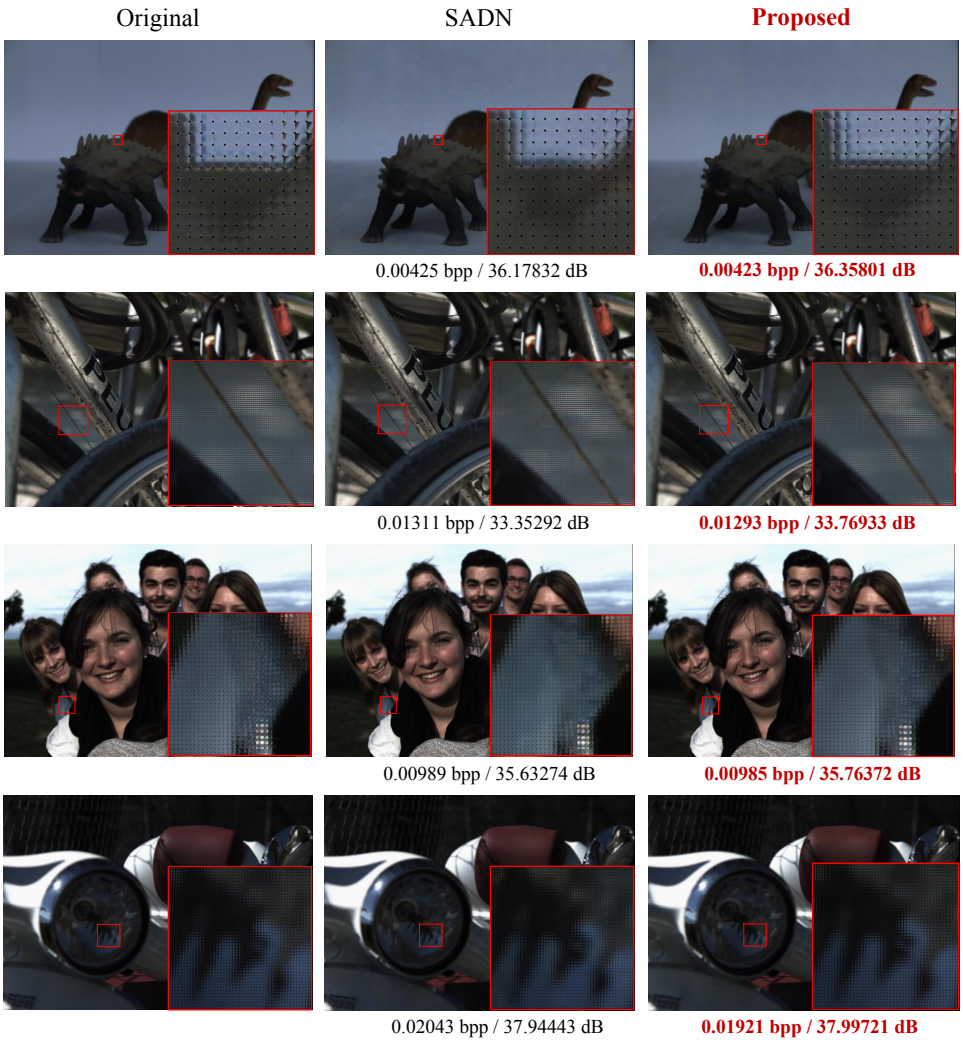


Figure 5. Visual comparison of reconstruction results generated by SADN and the proposed MFLFIC-Net. The reconstruction quality is measured by bpp/PSNR (dB).

4.3. Ablation Study

In order to verify the effectiveness of the SAJTM and E-ACM, they are successively removed from the proposed MFLFIC-Net. First, the SAJTM is replaced by the interactive module of SADN [19], which is denoted as "w/o SAJTM". Then, E-ACM is further removed from the MFLFIC-Net, which means that only utilize spatial features and angle features within a limited disparity range, denoted as "w/o E-ACM and w/o SAJTM". To ensure the fairness of the experiment, the results of the ablation experiment are tested by the third-stage training model. The comparison results are shown in Table 2 and Figure 6.

Table 2. Ablation results of BD-PSNR and BD-rate on the EPFL dataset.

Methods	BD-PSNR (dB)	BD-rate (%)
w/o E-ACM and w/o SAJTM	1.541	−48.167
w/o SAJTM	1.918	−54.970
Proposed	**2.149**	**−65.585**

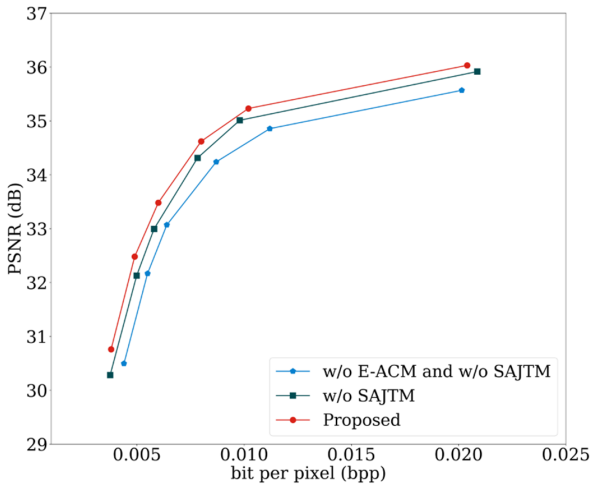


Figure 6. Ablation study on the EPFL dataset.

The SAJTM is proposed to effectively reduce redundancy between spatial and complete angle features. As shown in Table 2 and Figure 6, with the replacement of the SAJTM, the proposed network achieves a 0.231 dB BD-PSNR reduction and a 10.615% BD-rate increase. This is because the proposed SAJTM can effectively utilize the correlation between features to reduce redundant information, thereby achieving further improvement in compression performance. Meanwhile, the E-ACM is designed to obtain a complete angle feature, containing angle information on the complete disparity range. As shown in Table 2 and Figure 6, the "w/o E-ACM and w/o SAJTM" achieves a 0.377 dB BD-PSNR reduction and a 6.803% BD-rate increase compared with the "w/o SAJTM". This is because the proposed E-ACM fully utilizes the large disparity angle information contained in the EPI domain.

4.4. Complexity Analysis

In order to assess the computational complexity of MFLFIC-Net, MFLFIC-Net and the end-to-end method SADN [19] are tested to record their encoding and decoding times. The encoding and decoding times for SADN [19] are 449.441 s and 1107.301 s, respectively, while those for the proposed MFLFIC-Net are 996.422 s and 1482.784 s. Compared with SADN [19], MFLFIC-Net has a higher complexity. The increase in time can be attributed to the computational complexity of the mutual attention mechanism applied in the network, which is acceptable compared to the improvement in compression efficiency. In addition, it can be observed that the increase in encoding time is larger than the increase in decoding time. The reason is that, compared with the decoding time, the increase in the encoding time is more reliant on exploring the large disparity angle information contained in the EPI domain.

5. Conclusions

This paper proposes a novel multi-domain feature learning-based light field image compression network. In particular, an EPI-based angle completion module is developed to obtain a complete angle feature by fully exploring the large disparity angle information contained in the EPI domain. In addition, a spatial-angle joint transform module is proposed to reduce redundancy in features by modeling the intrinsic correlation between spatial and complete angle features. Experimental results demonstrate that MFLFIC-Net achieves superior performance on PSNR and MS-SSIM metrics compared to public state-of-the-art methods. In future work, it will be worthwhile to research the optimization of the algorithm structure and explore more lightweight networks for lower computational complexity, such as simpler and more effective feature extraction methods. At the same time, it is necessary to study possible trade-offs, such as maintaining the stability and accuracy of compression

performance while reducing network complexity. In addition, there is a need to balance the relationship between compression ratio and image quality to ensure that complexity is reduced while still maintaining sufficiently high compression efficiency.

Author Contributions: Conceptualization, K.Y. and Y.L.; methodology, K.Y.; software, K.Y. and G.L.; investigation, D.J. and B.Z.; resources, Y.L.; data curation, G.L.; writing—original draft preparation, K.Y.; writing—review and editing, K.Y., Y.L., G.L., D.J. and B.Z.; supervision, Y.L., D.J. and B.Z.; project administration, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Science and Technology Major Project of Tibetan Autonomous Region of China under Grant XZ202201ZD0006G03.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu, F.; Hon, G. Depth estimation from a hierarchical baseline stereo with a developed light field camera. *Appl. Sci.* **2024**, *14*, 550. [CrossRef]
2. Lei, J.; Liu, B.; Peng, B.; Cao, X.; Ling, N. Deep gradual-conversion and cycle network for single-view synthesis. *IEEE Trans. Emerge. Top. Comput.* **2023**, *7*, 1665–1675. [CrossRef]
3. Ai, X.; Wang, Y. The cube surface light field for interactive free-viewpoint rendering. *Appl. Sci.* **2022**, *12*, 7212. [CrossRef]
4. Amirpour, H.; Guillemot, C.; Ghanbari, M. Advanced scalability for light field image coding. *IEEE Trans. Image Process.* **2022**, *31*, 7435–7448. [CrossRef] [PubMed]
5. Gu, J.; Guo, B.; Wen, J. High efficiency light field compression via virtual reference and hierarchical MV-HEVC. In Proceedings of the International Conference on Multimedia and Expo, Shanghai, China, 8–12 July 2019.
6. Huang, X.; An, P.; Shan, L.; Ma, R. LF-CAE: View synthesis for light field coding using depth estimation. In Proceedings of the International Conference on Multimedia and Expo, San Diego, CA, USA, 23–27 July 2018.
7. Huang, X.; An, P.; Chen, D.; Liu, D.; Shen, L. Low bitrate light field compression with geometry and content consistency. *IEEE Trans. Multimed.* **2022**, *24*, 152–165. [CrossRef]
8. Liu, D.; Huang, X.; Zhan, W. View synthesis-based light field image compression using a generative adversarial network. *Inf. Sci.* **2021**, *545*, 118–131. [CrossRef]
9. Peng, B.; Zhang, X.; Lei, J.; Zhang, Z.; Ling, N.; Huang, Q. LVE-S2D: Low-light video enhancement from static to dynamic. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 8342–8352. [CrossRef]
10. Shi, X.; Lin, J.; Jiang, D.; Nian, C.; Yin, J. Recurrent network with enhanced alignment and attention-guided aggregation for compressed video quality enhancement. In Proceedings of the International Conference on Visual Communications and Image Processing, Suzhou, China, 13–16 December 2022.
11. Lei, J.; Song, J.; Peng, B.; Li, W.; Pan, Z.; Huang, Q. C2FNet: A coarse-to-fine network for multi-view 3D point cloud generation. *IEEE Trans. Image Process.* **2022**, *31*, 6707–6718. [CrossRef] [PubMed]
12. Shen, X.; Li, X.; Elhoseiny, M. MoStGAN-V: Video generation with temporal motion styles. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023.
13. Jung, H.K.; Choi, G.S. Improved yolov5: Efficient object detection using drone images under various conditions. *Appl. Sci.* **2022**, *12*, 7255. [CrossRef]
14. Yu, C.; Peng, B.; Huang, Q. PIPC-3Ddet: Harnessing perspective information and proposal correlation for 3D point cloud object detection. *IEEE Trans. Circuits Syst. Video Technol.* 2023, *accepted.* [CrossRef]
15. Peng, B.; Chang, R.; Pan, Z. Deep in-loop filtering via multi-domain correlation learning and partition constraint for multiview video coding. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *33*, 1911–1921. [CrossRef]
16. Li, H.; Wei, G.; Wang, T. Reducing video coding complexity based on CNN-CBAM in HEVC. *Appl. Sci.* **2023**, *13*, 10135. [CrossRef]
17. Zhang, J.; Hou, Y.; Pan, Z. SWGNet: Step-wise reference frame generation network for multiview video coding. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *accepted.* [CrossRef]
18. Hu, Y.; Yang, W.; Liu, J. 3D-CNN autoencoder for plenoptic image compression. In Proceedings of the International Conference on Visual Communications and Image Processing, Macau, China, 1–4 December 2020.
19. Tong, K.; Jin, X.; Wang, C.; Jiang, F. SADN: Learned light field image compression with spatial-angular decorrelation. In Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Singapore, 23–27 May 2022.
20. Singh, M.; Rameshan, R.M. Learning-based practical light field image compression using a disparity-aware model. In Proceedings of the Picture Coding Symposium, Speech and Signal Processing, Bristol, UK, 29 June–2 July 2021.

21. Jia, C.; Zhang, X; Wang, S.; Wang, S.; Ma, S. Light field image compression using generative adversarial network-based view synthesis. *J. Emerg. Sel. Top. Power Electron.* **2018**, *9*, 177–189. [CrossRef]
22. Liu, D.; Wang, L.; Li, L.; Xiong, Z.; Wu, F. Pseudo-sequence-based light field image compression. In Proceedings of the International Conference on Multimedia & Expo Workshops, Seattle, WA, USA, 11–15 July 2016.
23. Conceição, R.; Porto, M.; Zatt, B.; Agostini, L. LF-CAE: Context-adaptive encoding for lenslet light fields using HEVC. In Proceedings of the International Conference on Image Processing, Athens, Greece, 7–10 October 2018.
24. Liu, D.; An, P; Ma, R.; Zhan, W.; Huang, X. Content-based light field image compression method with gaussian process regression. *IEEE Trans. Multimed.* **2020**, *22*, 846–859. [CrossRef]
25. Dai, F.; Zhang, J.; Ma, Y.; Zhang, Y. Lenselet image compression scheme based on subaperture images streaming. In Proceedings of the International Conference on Image Processing, Quebec, QC, Canada, 27–30 September 2015.
26. Bakir, N.; Hamidouche, W.; Fezza, S.A.; Samrouth, K. Light field image coding using dual discriminator generative adversarial network and VVC temporal scalability. In Proceedings of the International Conference on Multimedia and Expo, London, UK, 6–10 July 2020.
27. Zhao, Z.; Wang, S.; Jia, C.; Zhang, X.; Ma, S.; Yang, J. Light field image compression based on deep learning. In Proceedings of the International Conference on Multimedia and Expo, San Diego, CA, USA, 23–27 July 2018.
28. Van, V.; Huu, T.N.; Yim, J.; Jeon, B. Downsampling based light field video coding with restoration network using joint spatio-angular and epipolar information. In Proceedings of the International Conference on Image Processing, Bordeaux, France, 16–19 October 2022.
29. Ballé, J.; Minnen, D.; Singh, S. Variational image compression with a scale hyperprior. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
30. Minnen, D.; Ballé, J.; Toderici, G.D. Joint autoregressive and hierarchical priors for learned image compression. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 4–6 December 2018.
31. Cheng, Z.; Sun, H.; Takeuchi, M. Learned image compression with discretized gaussian mixture likelihoods and attention modules. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
32. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2017**, arXiv:1412.6980.
33. Rerabek, M.; Ebrahimi, T. New light field image dataset. In Proceedings of the International Conference on Quality of Multimedia Experience, Lisbon, Portugal, 6–8 June 2016.
34. HEVC Official Test Model. Available online: https://vcgit.hhi.fraunhofer.de/jvet/HM/-/tags (accessed on 27 January 2024).
35. VVC Official Test Model. Available online: https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM (accessed on 27 January 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.