Angular-spatial analysis of factors affecting the performance of light field reconstruction

Xinjue Hu 1 | Lin Zhang 2,3

1 Beijing University of Posts and Telecommunication, Beijing, China (Email: huxinjue@bupt.edu.cn)
2 Research Center of Networks and Communications, Peng Cheng Laboratory, Shenzhen, China
3 Beijing Information Science and Technology University, Beijing, China

Correspondence
Lin Zhang, Beijing Information Science and Technology University
Email: zhl@bistu.edu.cn

Funding information
BUPT Excellent Ph.D Students Foundation, Grant/Award Number: CX2019 102; National Key Research and Development Program of China, Grant/Award Number: 2019YFB1803103; The Verification Platform of Multi-tier Coverage Communication Network for Oceans, Grant/Award Number: LZC0020

Abstract
As a new VR multimedia format, Light Field (LF) has received more and more attention. LF can support users to move freely within a certain range during the experience. However, the limitation of LF acquisition technology has become the main obstacle to its large-scale application. Since the current hardware technology cannot produce a lens small enough to meet the needs of LF capture, the intensity of the collected LF image is usually insufficient. Although there are many software LF reconstruction algorithms have been proposed to make captured result denser, they all face the problem of poor generalization in the application process. In this work, it is attempted to locate the factors that affect the performance of LF reconstruction from both the angular and spatial domains. The analysis results show that it is affected not only by the edge and texture of the LF content in the spatial domain, but also by the adjacent disparity and the reflection characteristics in the angular domain. An indicator to quantify the impact of these factors on reconstruction performance is also proposed, which will be helpful to design a new generalized adaptive LF reconstruction algorithm in the future.

1 | INTRODUCTION

The difference between traditional media and VR media is that the latter can provide users with an immersive experience. That means users can freely choose the direction and position to observe the scene when viewing the VR media. The ideal VR media need to be able to support the user’s viewing behavior on 6 Degrees of Freedom (6DoF). As shown in Figure 1(a), the users’ viewing angle has 3 Degrees of Freedom (3DoF), usually defined as pitch, yaw and roll, represented by \( \alpha \), \( \beta \) and \( \gamma \). For the 6DoF, an additional 3 degrees of freedom (\( x \), \( y \) and \( z \)) are required to determine the viewing position of the user, as illustrated in (b).

The 360° video and the 3-Dimension (3D) model are the most well-known VR media format for now. The former is captured by the panoramic camera, but only supports the 3DoF viewing experience. The user’s viewing position is limited by the center of the camera. Although the 3D model (e.g. point cloud and mesh model) can allow users to explore freely within the virtual scene, how to accurately model any scene in real-time is still a challenge that cannot be solved currently. Because of the limitations of both 360° videos and 3D models, the Light Field (LF) is seen as a new and promising VR format for a true 6DoF immersive experience. Furthermore, LF has advantages over the other VR format, including the lower computational requirement for rendering, and scene-complexity independence of the cost of interactive viewing [1, 2].

The LF image records a scene with a set of images from sub-aperture lenses of different viewpoints, so it can provide more structure and texture detail of the scene. As shown in Figure 2, by switching between existing viewpoints or synthesizing new viewpoints, the user can move freely in both the vertical and horizontal directions. At the same time, the equally spaced view acquisition form is very beneficial for the extraction of the depth information for the scene, thus can support the positional movement of users in the forward–backward direction. Each
FIGURE 1 3DoF and 6DoF for immersive experience

FIGURE 2 LF viewpoints images matrix

sub-aperture image stores a scene image captured by the LF camera from different perspectives. These sub-aperture images are the same with traditional images, and they represent the spatial domain of the LF. Therefore, the spatial domain resolution of the LF is the pixel resolution of each sub-aperture image. At the same time, each sub-aperture image corresponds to a different viewpoint of the shooting angle, and all the viewpoints constitute the angular domain of the LF. The denser the angular domain, the more scene details are captured and recorded.

The insufficient angular resolution of LF images is a common problem encountered in acquisition, mainly because the production technology of the acquisition lens cannot meet the actual needs. LF reconstruction is a solution to this problem, which can reconstruct high angular resolution LF images with low angular resolution input data. Although many reconstruction methods have been proposed, they all face performance generalization problems in actual use [9-15]. It is found that the performance of the reconstruction algorithm on different datasets varies greatly, and the reconstruction results of some datasets are seriously distorted. We analyze the LF data in both the angular domain and the spatial domain, and locate the main factors that interfere with the performance of LF reconstruction. At last, we establish indicators to quantify the impact of these factors. The contributions of this work are:

• We analyze the factors that affect the performance of LF reconstruction in the spatial domain and the angular domain, which can greatly help solve the problem of unstable performance of the reconstruction algorithm.
• In the spatial domain, the performance of LF reconstruction is found to be poor in the edges and texture regions of the image. In the angular domain, the adjacent disparity and the reflection characteristics are also regarded as the main factors that interfere with LF reconstruction.
• An indicator is proposed to sufficiently evaluate the influence of interference factors both in the spatial and angular domain.

The rest of this paper is organized as follows. Section 2 discusses related work. Section 3 shows the performance test results of various reconstruction schemes on different LF images. Section 4 presents the analysis and conclusion of the test result in both the spatial and angular domain and proposes corresponding indicators to quantify the impact of these factors on reconstructing performance. Section 5 designs a comprehensive index to evaluate the complexity of LF scenes. Finally, Section 6 concludes the paper and discusses future work.

2 | RELATED WORK

We first present an overview of the basic concepts of LF image processing. As shown in Figure 3(a), LF can be represented by a 4D plenoptic function \( L(u, v, x, y) \), which characterizes the light rays between two parallel planes, called the boldmath \( uv \) camera plane and the \( xy \) image plane. The coordinate of \((u, v)\) corresponds to the location of the viewpoint in the microlens matrix, and the coordinate of \((x, y)\) corresponds to a pixel in the sub-aperture image of single viewpoint. The \( xy \) plane contains the spatial information of the LF, while the \( uv \) plane forms the angular domain of the LF. \( L_{u,v}(x,y) \) represents the 2D spatial slice of LF, which is actually the 2D sub-aperture image of each viewpoint. Similarly, \( L_{x,y}(u,v) \) represents, respectively, the angular slice of LF. Usually, camera array or lenslet [3, 4] are used to collect LF, as illustrated in Figure 3(b). Figure 2 shows an LF image used in this work. The camera array contains \( 17 \times 17 \)
sub-aperture lenses with each having a $1024 \times 1024$ spatial resolution. The scene is recorded with a set of cameras from different viewpoints. The image captured by each separate camera is called a sub-aperture image, and all the sub-aperture images together form an image matrix according to their corresponding camera positions. For example, $(u_0, v_0)$ determines the position of the sub-aperture image shown by the red box, and $(u_1, v_1)$ indicates the position of the pixel in the image, shown by the blue dot in the figure. For more details, readers can refer to LF tutorials such as [5].

Since a single lens needs to occupy a certain area, it is impossible to deploy sub-aperture lenses infinitely densely on the camera plane. Therefore, the angular resolution of the LF that can be acquired is limited by the lens production process. In order to improve the angular resolution of LF through post-processing algorithms, many methods of LF reconstruction have been proposed. The LF reconstruction algorithm can be divided into two kinds: the one with depth information reference and the one without depth reference.

The first kind of schemes needs to collect the depth map of the scene when collecting the LF image or estimate it from the LF's Epiolar Plane Image (EPI), and achieve the upsampling of the LF image with the help of the depth information. Bilevich et al. [6] propose a three-pass algorithm for light field interpolation in EPI. [7] introduces the self-similarity based LF image compression scheme, compressing light field image by the commonly used image compression standards. Rizkallah et al. [8] proposed a light field compression scheme based on CNN, which uses entropy coding with graph transform. LF reconstruction is also helpful for 3D modeling in Virtual Reality (VR). Vagharshakyan et al. [9] utilize sparse representation of EPI with a directionally sensitive transform; However, all these LF reconstruction schemes suffer from a certain degree of performance degradation when dealing with some LF images containing complex scenes. By using locally linear embedding techniques, Lucas et al. [10] propose a new way to improve the performance of 3D modeling. With the help of the high-precision model, any virtual viewpoints can be rendered freely.

David et al. [11] cast the problem of upsampling camera arrays as a directional super-resolution in 4D space. Nima et al. [12] introduces disparity and depth estimation using a learning-based approach and proposes an intermediate synthesis technique to reconstruct the LF image. Suren et al. [13] use the shearer transform as the sought sparsifying transform and develop an effective reconstruction of the LF represented by EPIs. Jason et al. [14] combine band-limited filtering with wide-aperture reconstruction which is essentially directional edge-preserving filtering. Bishop et al. [15] recover superresolution for Lambertian scenes in the spatial or angular domain on the base of Lambertian assumption.

The second kind of schemes directly reconstructs the sub-aperture image matrix of LF. Shi et al. [16] considered LF reconstruction as a sparsity optimization problem in the continuous Fourier domain. Sparse Fast Fourier Transform (SFFT) is used to reconstruct the full LF images. Irene Viola et al. [17] introduce graph theory to model the LF images. They regard each sub-aperture image as a vertex of a graph, with weighted edges represent the correlation. The determination of the graph model is defined as an optimization problem, and the gradient descent algorithm is used to search for the optimal solution. Graph Convolution Networks (GCN) [18] can efficiently and accurately calculate the graph model of LF, which can be used to reconstruct the high angular resolution LF image. Xu et al. [19] present a reconstruction algorithm that can restore the 4D light field from a portable light field camera without the need for calibration images. They proposed a 4D demosaicing algorithm based on kernel regression to enhance the quality of the reconstructed sub-aperture images.

Because the second kind of schemes is more in line with the actual application requirements, it does not need other information to assist except the collected LF image. Therefore, this work mainly analyzes the factors that affect the performance generalization of the second kind of LF reconstruction schemes. To our best knowledge, this is the only exploration work on this issue at present. But there are many works about quality metrics specially designed for LF images. In [20] Paudyal et al. present a SMART light field image quality dataset, which consists of source images, compressed images, and annotated subjective quality scores to help test new processing tools or even assess the effectiveness of objective quality metrics. Murgia and Giusto [21] present an open-access database of light field images and corresponding views to help fairly compare the results in light field imaging applications. Yuan Yan [22] explores the effects of light field photography on image quality by quantitatively evaluating some basic criteria for an imaging system. Kera et al. [23] mainly focuses on visual content quality, aiming to measure the perceptual difference between light field reconstruction and different angular resolutions via a series of subjective image quality assessments.

## 3 PERFORMANCE TEST FOR LF RECONSTRUCTION

In this work, three typical reconstruction schemes are selected for the test. They are LF reconstruction based on SFFT [16], Graph-Based Learning (GBL) reconstruction method [17] and GCN-based LF reconstruction [18], respectively. These three reconstruction algorithms all directly perform signal processing and recovery on the LF sub-aperture matrix, instead of simply interpolating between viewpoints to achieve super-resolution in the angular domain. They can reconstruct the ground truth of LF images more accurately, and have been used in other applications [24–26].

All tested reconstruction methods are used to process six different groups of LF images (Stanford Crystal-large angular extent(Crystal-L), Stanford Crystal-small angular extent(Crystal-S), Bunny, Knight [24], Tower, Town) [27] and test their reconstruction performance, as illustrated in Figure 4. Among them, Crystal-L and Crystal-S capture the same scene, but the camera array used by Crystal-L has larger spacing between adjacent sub-aperture lenses, that is, the adjacent disparity of Crystal-L is larger.

The reconstruction quality is measured by average Peak Signal to Noise Ratio (PSNR) and average Structural Similarity...
The reconstructed PSNR is the average PSNR of all sub-aperture images in the $uv$ plane, which is defined in Equation (1):

$$
\text{PSNR} = \sum_{(m,n) \in (u,v)} 10 \log_{10} \left( \frac{255^2}{\text{MSE}(m,n)} \right),
$$

$$
\text{MSE}(m,n) = \sum_{(i,j) \in (x,y)} \text{dist}(m,n,i,j),
$$

$$
\text{dist}(m,n,i,j) = \| L_{m,n}(i,j)_{\text{rec}} - L_{m,n}(i,j)_{\text{raw}} \|.
$$

In addition to objective quality metrics, we also test for subjective quality metric VMAF [28]. VMAF has been trained and tested on a large amount of users’ data on Netflix servers, which can reflect the subjective feeling that the image quality brings to the user. VMAF scores range from 0 to 100, with 100 indicating the best subjective experience. The SSIM and VMAF are also the results after averaging on the $uv$ plane, and the calculation method is similar to Equation (1). All LF images are down-sampled before reconstruction, raw data without preprocessing are used as the ground truth when evaluating the reconstructed performance. In Figure 4, each row represents the reconstruction quality results of the same method on different LF images. The upper left corner of the table also shows the spatial resolution and angular resolution of all LF images.

As shown in Figure 4, there is a significant difference between the reconstruction results of different LF images, and this difference appears in the results of all reconstruction methods. For example, the reconstruction result of Bunny and Team are better than other LF images. Another valuable test result is that the output quality of Crystal-S is better than that of Crystal-L even when the collected scenes are identical. Next, we will analyze the test results from two aspects: spatial domain and angular domain. We will focus on the causes of this poor performance generalization and establish indicators to evaluate them quantitatively.

### 4 | INTERFERENCE FACTORS OF LF RECONSTRUCTION

#### 4.1 | Analysis in spatial domain

The spatial domain of LF represents the imaging result of the scene when the viewpoint is fixed. Like the traditional image, the high frequency component of LF spatial domain corresponds to the edge and texture of the object, while the low frequency component corresponds to the smooth region. In order to determine the specific regions in the spatial domain that bring adverse effects to LF reconstruction, we make a more fine-grained quality analysis of the reconstruction results. As shown in Figure 5, we define the $16 \times 16$ pixels block around target
pixel $P$ as $U_R$. By calculating the MSE between reconstructed output and ground truth of all pixels in $U_R$, we can evaluate the reconstruction quality of the local area around $P$. For the pixels at the edge of the image, we only use the remaining pixels in $U_R$ to calculate MSE. We will not consider the points that do not exist outside the boundary.

Figure 6 presents the visualized result of fine-grained quality analysis on Bunny, Tower, Knight and Crystal-L using the GCN-based reconstruction method. The darker the pixel, the worse the reconstruction quality in the local area. In Bunny, the background is quite pure and the quality of those areas is significantly higher than in other areas. We also notice that for those pixels in the edge and texture regions, the resulting quality of their surrounding regions is very poor. These phenomena are also obvious in the other three LF images. Therefore, we think that the high frequency components such as edge and texture in LF spatial domain are the main factors to interfere with the reconstruction performance.

We consider calculating the high frequency ratio of all LF images in spatial domain to verify our analysis conclusion. Two dimensional Discrete Cosine Transform (DCT) is used to analyze the frequency component distribution for spatial slice $L_{xx}(x, y)$. Since the signal energy is mainly concentrated in the low-frequency region, we reorder the spectrum according to the energy level, and select the proportion of the last 90% of the frequency point energy to the total energy as the high-frequency ratio $R_{H−I_{spa}}$ as shown in Equation (2).

$$R_{H−I_{spa}} = \frac{E_{H−I_{spa}}}{E_{all−I_{spa}}}.$$  

(2)

Table 1 gives the $I_{spa}$ and the PSNR of its corresponding reconstructed results on four groups of LF images: Crystal-L, Knight, Crystal-S and Bunny. All results of SFFT, GBL and GCN are given. The last column is the $I_{spa}$ of each group of LF images.

In Table 1, it can be clearly observed that the lower $I_{spa}$ is, the less spatial high frequency components of LF image are, and the better reconstructed results are. Therefore, the high-frequency components such as edge and texture are one of the factors of interference reconstruction. At the same time, we also notice that $I_{spa}$ cannot explain the reason why Knight is better than Crystal-L, but worse than Crystal-S. So we believe that there are other interference factors besides the spatial domain.

### 4.2 Analysis in angular domain

We mentioned before in Section 3 that the quality of reconstruction results is different due to the difference of adjacent disparity between Crystal-S and Crystal-L. This shows that in addition to the interference factors inside the sub-aperture viewpoint, there should also be some factors between the viewpoints. We notice that the results in Figure 6 seem to be related to the location of objects in the scene. For example, the quality of the tower area in the upper left corner of the Tower is significantly better than that of other objects.

We further use the depth map of each LF image to compare with its fine-grained quality analysis result. We use the GCN reconstruction algorithm for Tower to reconstruct the LF matrix from $5 \times 5$ to $9 \times 9$. Figure 7 presents the comparison result of Tower. The left is the reconstructed result; we transform it into a color temperature map for easy observation. The higher the temperature, the better the reconstruction quality, and vice
From Figure 7, we notice that the objects with better quality are all concentrated near the same depth plane. Apart from the very far background area, whether it is closer or farther away from the plane, the reconstruction quality of the target is declining. This means that the depth plane of the object is related to its reconstruction quality. In order to verify that, we count the reconstructed PSNR of all pixels according to the depth plane; the result is shown in Figure 8. The horizontal axis represents different depth planes, and the vertical axis represents the average value of the reconstructed PSNR of all pixels on the same plane. The depth value is normalized to (0,1) and divided into 100 planes.

There is an obvious peak near the depth value of 0.51 in Figure 8. When moving away from the peak plane to both sides, the quality of their result gradually decreases. However, when the depth value is close to 1, the resulting quality begins to pick up. This is mainly because the depth plane close to depth value 1 represents the background area of the scene, which is consistent with the results observed in Figure 7. In order to determine the special physical significance of the depth plane near the depth value 5.1, we investigated the acquisition camera array of the LF image Tower. We find that the camera array of Tower is not arranged in parallel, but in a horizontal arc surface around the center of the scene, as shown in Figure 9.

Therefore, the depth plane near 0.51 is likely to correspond to the rotation center of the horizontal arc surface of the camera array. By analyzing the sub-aperture image matrix of Tower, we find that the displacement of the object near 0.51 is the smallest in different viewpoints, and it is determined as the horizontal rotation center of the camera array. To sum up, we confirm that the reconstruction quality of each object in the scene is related to their depth distance from the rotation center. As shown in Figure 9, the further away from the center of rotation, the larger the shift of objects in different viewpoints, the more unfavorable the reconstruction algorithm is to synthesize new viewpoints. At the same time, the closer the depth value is to 1, the closer the depth plane is to infinity. In this case, the camera array can be approximately seen as a parallel arrangement, and the change of the background in different viewpoints is very small, so the reconstruction quality is picking up.

Through the analysis, we know that the depth distance will increase the difference between viewpoints, and then lead to reconstruction distortion. In addition, the larger interval between adjacent viewpoints will also lead to the larger disparity. Shooting scenes with camera array can be regarded as discrete sampling in LF angular domain. When the array is more sparse, the more discrete the collected signal is, the less easy it is to reconstruct the LF image. We believe that this is the reason for the difference in reconstruction performance between Crystal-L and Crystal-S. In the LF image data set used in this work, only the crystal ball scene was collected with different adjacent disparity (Crystal-L and Crystal-S) during capturing. The rest of the scenes only captured the data of one adjacent disparity, and it is difficult to directly compare them with different disparity reconstruction results. So if we need to design a verification experiment, we...
need to intentionally subsample the existing LF images to generate subsets of the same scene with different adjacent disparity.

As shown in Figure 10, we design two kinds of sampling schemes: large interval and small interval. They all downsample the angular resolution of the LF matrix from 17 × 17 to 9 × 9. The red small box represents the viewpoint to be sampled, and the white one represents the viewpoint to be discarded. On the left is the small interval sample, on the right is the large interval sample. The angular resolution of these two schemes’ output is the same, but their adjacent disparity is different.

Because the original angular resolution of Tower and Town is too small (only 9 × 9), it is not conducive to down-sampling. So we choose Crystal-L, Knight and Bunny three groups of LF images for experiment. After two kinds of subsampling for all LF images, we use the GCN-based method to test the reconstruction performance for their output, the result is shown in Table 2.

Table 2 verifies that even in the case of the same angular resolution, the adjacent disparity difference of LF image will affect the quality of reconstruction.

The adjacent disparity and the depth distance from the rotation center are both the reasons for the great differences between LF viewpoints, which are actually the reflection of the high frequency component of LF image in the angular domain. While the region which tends to be invariable in different viewpoints represents the low frequency angular component. Therefore, similar to the methods in the spatial domain, we design indicators in the angular domain to evaluate the impact of these factors. We analyzed the frequency components for angular slice $I_{\theta_1 \theta_2, \phi}(u, v)$. By reformulating the LF image and using DCT processing, we can get the energy distribution of frequency points in the angular domain, as shown in Equations (4) and (5).

$$R_{\theta_1 \theta_2 \rightarrow \theta_1 \theta_2} = \frac{E_{\theta_1 \theta_2 \rightarrow \theta_1 \theta_2}}{E_{\theta_1 \theta_2 \rightarrow \theta_1 \theta_2}}$$

$$I_{\theta_1 \theta_2} = \frac{C_{\theta_1 \theta_2}}{\log_2 \left( \frac{R_{\theta_1 \theta_2 \rightarrow \theta_1 \theta_2}}{R_{\theta_1 \theta_2 \rightarrow \theta_1 \theta_2}} \right)}$$

where $C_{\theta_1 \theta_2}$ is also a scaling parameter and set to 50 because there are fewer high frequency components in the angular domain. The comparison result between reconstructed quality and $I_{\theta_1 \theta_2}$ is shown in Table 3. The Tower and Town are reconstructed from 5 × 5 to 9 × 9. The Crystal-L, Knight, Crystal-S and Bunny are reconstructed from 9 × 9 to 17 × 17. Overall, the lower $I_{\theta_1 \theta_2}$ is, the less angular high frequency components of LF image are, and the better reconstructed results are.

From Tables 1 and 3, we find that the reconstruction quality of LF image is affected by both spatial and angular domain factors. For example, although the $I_{\phi_0}$ of Crystal-L and Crystal-S are similar, the $I_{\theta_1 \theta_2}$ difference between them is obvious, which shows that the main reason for the performance difference between them is the angular domain factors. The huge difference between Knight and Tower can also be reflected by $I_{\theta_1 \theta_2}$. But for why Town is better than Tower, $I_{\theta_1 \theta_2}$ is not a good explanation, which shows that the dominant factors are in the spatial domain.

### 5 | A COMPREHENSIVE INDEX TO EVALUATE THE COMPLEXITY OF LF SCENES

As analyzed in the previous section, we believe that factors in the spatial domain and angular domain together affect the performance of LF reconstruction. We want to build a comprehensive index by the weighted sum of $I_{\phi_0}$ and $I_{\theta_1 \theta_2}$ to describe the influence of all possible factors in the LF scene on reconstruction. In essence, this is a quantitative evaluation of LF Scene Complexity (LFSC). As follows, we define LFSC as:

$$LFSC = \alpha I_{\phi_0} + \beta I_{\theta_1 \theta_2}$$

Table 4 shows the LFSC results of different $\alpha$ and $\beta$, and compare them with the actual reconstruction quality.

When $\alpha = 0.6$ and $\beta = 0.4$, the quality of reconstruction results decreases monotonically with LFSC. In the LF image dataset we used, LFSC can quantify and reflect the influence of
TABLE 4 Different $\alpha$ and $\beta$ for LFSC and the reconstruction quality

| Quality indicator | Crystal-L | Knight | Crystal-S | Tower |
|-------------------|-----------|--------|-----------|-------|
| PSNR (dB)         | 24.6533   | 32.2792| 34.7149   | 39.6411|
| SSIM (0–1)        | 0.7871    | 0.9432 | 0.9139    | 0.9520 |
| VMAF (0–100)      | 58.7755   | 86.1330| 87.0080   | 81.0729|
| $\alpha$          | $\beta$   | LFSC   |           |       |
| 0.2               | 0.8       | 10.5673| 8.0531    | 7.6317 | 6.6826 |
| 0.4               | 0.6       | 9.7488 | 7.5647    | 7.2367 | 6.5325 |
| 0.6               | 0.4       | 8.8045 | 6.9767    | 6.7699 | 6.3824 |
| 0.8               | 0.2       | 7.6000 | 6.1835    | 6.1714 | 6.2323 |
| PSNR (dB)         |           |        |           |       |
| SSIM (0–1)        |           |        |           |       |
| VMAF (0–100)      |           |        |           |       |
| $\alpha$          | $\beta$   | LFSC   |           |       |
| 0.2               | 0.8       | 6.7411 | 5.9384    |       |
| 0.4               | 0.6       | 6.5499 | 5.6803    |       |
| 0.6               | 0.4       | 6.3588 | 5.3624    |       |
| 0.8               | 0.2       | 6.1676 | 4.9260    |       |

There are also some limitations in our work. In Section 5, our design of LFSC is relatively simple, and we did not further explore the mechanism for working together of spatial domain factors and angular domain factors. This is what we need to focus on in our future work. At the same time, how to apply the results of this work to modify various application technologies of LF is also our future focus. We hope that it can at least help to design a new LF reconstruction method which can adjust automatically according to the input LF image.

ACKNOWLEDGEMENTS

The work is supported by the BUPT Excellent Ph.D. Students Foundation (CX2019 102). This work is also supported by the project “The Verification Platform of Multi-tier Coverage Communication Network for oceans (LZC0020)” and National Key Research and Development Program of China (2019YFB1803103).

ORCID

Xinjue Hu https://orcid.org/0000-0001-8304-9720

REFERENCES

1. Levoy, M., Hanrahan, P.: Light field rendering. In: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, pp. 31–42. ACM, New York (1996)
2. Gortler, S.J., et al.: The lumigraph. In: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, pp. 43–54. ACM, New York (1996)
3. Venkataraman, K., et al.: Picam: An ultra-thin high performance monolithic camera array. ACM Trans. Graphics (TOG) 32(6), 166 (2013)
4. Liu, X., et al.: Camera array based light field microscopy. Biomed. Opt. Express 6(9), 3179–3189 (2015)
5. Wu, G., et al.: Light field image processing: An overview. IEEE J. Sel. Top. Signal Process. 11(7), 926–954 (2017)
6. Bilevich, L., et al.: High-quality light field interpolation in epipolar plane image domain by DCT-based shearing. AIP Conf. Proc. 1631, 77 (2014)
7. Conti, C., et al.: HEVC-based 3D holoscopic video coding using self-similarity compensated prediction. Signal Process. Image Commun. 42, 59–78 (2016)
8. Rizkallah, M., et al.: Graph-based transforms for predictive light field compression based on super-pixels. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, Piscataway (2018)
9. Vagharshakyan, S., et al.: Light field reconstruction using shearlet transform. IEEE Trans. Pattern Anal. Mach. Intell. PP(99), 133–147 (2018)
10. Lucas, L.F., et al.: Locally linear embedding-based prediction for 3D holoscopic image coding using HEVC. Signal Process. Conf. 11–15 (2014)
11. Schell, D.C., et al.: Directional super-resolution by means of coded sampling and guided upsampling. In: 2015 IEEE International Conference on Computational Photography (ICCP), pp. 1–10. IEEE, Piscataway (2015)
12. Kalantar, N.K., et al.: Learning-based view synthesis for light field cameras. ACM Trans. Graphics (TOG) 35(6), 193 (2016)
13. Vagharshakyan, S., et al.: Light field reconstruction using shearlet transform. IEEE Trans. Pattern Anal. Mach. Intell. 40(1), 133–147 (2017)
14. Stewart, J., et al.: A new reconstruction filter for undersampled light fields. ACM International Conference Proceeding Series (2003)
15. Bishop, T.E., et al.: Light field supersresolution. In: 2009 IEEE International Conference on Computational Photography (ICCP), pp. 1–9. IEEE, Piscataway (2009)
16. Shi, L., et al.: Light field reconstruction using sparsity in the continuous fourier domain. ACM Trans. Graphics (TOG) 34(1), 12 (2014)
17. Viola, I., et al.: A graph learning approach for light field image compression. In: Applications of Digital Image Processing XLI, p. 12. SPIE, Washington, DC (2018)
18. Niepert, M., et al.: Learning convolutional neural networks for graphs. In: Proceedings of the 33rd International Conference on International Conference on Machine Learning, pp. 2014–2023. JMLR.org, New York (2016)
19. Xu, S., et al.: Multi-view image restoration from plenoptic raw images. In: Asian Conference on Computer Vision, pp. 3–15. Springer, Cham (2014)
20. Paudyal, P., et al.: Toward the perceptual quality evaluation of compressed light field images. IEEE Trans. Broadcast. 63(3), 507–522 (2017)
21. Murgia, F., Giusto, D.: A database for evaluating the quality of experience in light field applications. In: 24th Telecommunications Forum (TELFOR). IEEE, Piscataway (2017)
22. Fu, Q., et al.: Image quality evaluation of light field photography. Proc. SPIE Int. Soc. Opt. Eng. 7867(4), 357–366 (2011)
23. Kara, P.A., et al.: The effect of light field reconstruction and angular resolution reduction on the quality of experience. In: International Conference on Signal-Image Technology & Internet-Based Systems. IEEE, Piscataway (2017)
24. Hu, X., et al.: An adaptive two-layer light field compression scheme using GNN-based reconstruction. ACM Trans. Multimedia Comput. Commun. Appl. (TOMM) 16(2c), 1–23 (2020)
25. Hawary, F., et al.: Scalable light field compression scheme using sparse reconstruction and restoration. In: 2017 IEEE International Conference on Image Processing (ICIP), pp. 3250–3254. IEEE, Piscataway (2017)
26. Hu, X., et al.: Adaptive two-layer light field compression scheme based on sparse reconstruction. In: Proceedings of the 10th ACM Multimedia Systems Conference, pp. 74–85. ACM, New York (2019)
27. Honauer, K., et al.: A dataset and evaluation methodology for depth estimation on 4D light fields. In: 13th Asian Conference on Computer Vision, pp. 19–34. Springer, Cham (2017)
28. Rassool, R.: VMAF reproducibility: Validating a perceptual practical video quality metric. In: IEEE International Symposium on Broadband Multimedia Systems and Broadcasting. IEEE, Piscataway (2017)

How to cite this article: Hu X, Zhang L. Angular-spatial analysis of factors affecting the performance of light field reconstruction. IET Image Process. 2021;1–9. https://doi.org/10.1049/ipr.12203