Optics and Lasers in Engineering 165 (2023) 107539

Contents lists available at ScienceDirect

Optics and Lasers in Engineering

journal homepage: www.elsevier.com/locate/optlaseng

Consistency analysis of focal stack-based light field reconstruction

Yang Liu [a,b,c,1], Runnan Zhang [a,b,c,1], Shijie Feng [a,b,c], Chao Zuo [a,b,c], Qian Chen [a,c], Zewei Cai [a,b,c,*]

[a] Smart Computational Imaging Laboratory (SCILab), School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[b] Smart Computational Imaging Research Institute (SCIRI), Nanjing University of Science and Technology, Nanjing 210019, China
[c] Jiangsu Key Laboratory of Spectral Imaging and Intelligent Sense, Nanjing 210094, China

ARTICLE INFO

Keywords:
Light field reconstruction
Computational imaging
Focal stack
Transport of intensity
Object-image sapce consistency

ABSTRACT

Light field camera with a microlens array can realize spatio-angular joint sampling of light ray field at the cost of a trade-off between spatial and angular resolutions. Alternatively, focal stack-based light field reconstruction can computationally retrieve full-pixel-resolution light fields in object space by virtue of the transport-of-intensity property in an image sequence recorded at different focal depths in image space. However, traditional camera imaging generally involves a nonlinear mapping between object and image spaces. The inconsistency of image-space recording and object-space reconstruction will reduce the accuracy of reconstructed light fields. In this work, we focus on analyzing and addressing the problem caused by the object-image space inconsistency for high-resolution, high-accuracy focal stack-based light field reconstruction. With a pre-calibrated light field camera as a reference, light field reconstructions in object and image spaces are experimentally compared and discussed from different aspects, such as digital refocusing, viewpoint switching, angular resolution, and depth range and sampling rate. All experimental results demonstrate that the light field reconstruction accuracy can be significantly improved when satisfying the object-image space consistency, which can serve as a mechanism for the realization and application of high-quality computational light field imaging and measurement in the situation of nonlinear recording and reconstruction.

Introduction

Due to the human retina's well-known "what you see is what you get" property, traditional two-dimensional imaging sensors provide a significant optical difficulty that only records two-dimensional spatial intensity information. The emerging of computational photography, typically light field imaging, breaks through this limitation. Light field is a representation of full four-dimensional (4D) radiance of all rays with spatial and angular information in free space. Light field camera embedded with a microlens array can simultaneously record the 4D spatio-angular information of light rays [1,2], enabling many new development potentials for computational imaging such as viewpoint switching, digital refocusing, panoramic focusing, depth-of-field expansion, and depth/three-dimensional (3D) sensing [3–5].

However, such spatial multiplexing technique using a microlens array suffers from a trade-off between spatial and angular resolutions. In general, the spatial resolution is one to two order magnitudes lower than the used pixel amount of the sensor. Although higher spatial (angular) resolution can be resolved from light field images because of spatio-angular sampling redundancy [6,7], the multiplexed space of the sensor still strictly limits the super-resolution capability. Alternatively, scanning light field imaging with high spatial resolution can be achieved by dithering a microlens array [8,9] or using a galvanometer [10] at the expense of requiring additional motion devices and making the system architecture more complicated. Another spatio-angular sampling modality is to record the spatial information of light fields angle to angle through a camera array [11,12] or single-camera motion [13], but the angular resolution is limited by the array size or the motion number.

In contrast to direct spatio-angle sampling, one can use other optical components, such as coded aperture [14,15], attenuation mask [16,17], diffuser [18,19], and graphene layers [20], instead of the microlens array, as modulators to pre-modulate the incident light field for computational light field imaging. These techniques can increase the spatio-angular sampling redundancy in the spatial or frequency domain to demodulate light fields with high spatio-angular resolution. In addition to physically encoded devices, the incident light field can be digitally modulated in terms of light field propagation, the core of which is the inverse reconstruction of Fourier slice photography [21]. With a traditional camera to capture images at different focal depths, the full-pixel-resolution light field information can be retrieved according to the

* Corresponding author.
E-mail addresses: zuochao@njust.edu.cn (C. Zuo), zeweilinc@foxmail.com (Z. Cai).
[1] These authors contributed equally.

https://doi.org/10.1016/j.optlaseng.2023.107539
Received 20 December 2022; Received in revised form 31 January 2023; Accepted 18 February 2023
0143-8166/© 2023 Elsevier Ltd. All rights reserved.

transport-of-intensity property in a focal stack (FS). Levin and Durand [22] in 2010 and Alonso et al. [23] in 2016 deconvoluted a dense FS using different defocused blur kernels. In 2013, Orth and Crozier [24] proposed light field moment imaging to deconvolve the first-order moment of light fields from a defocused image pair. These methods can demodulate partial information of the 4D light field from 3D FS, achieving viewpoint synthesis and rendering. In 2014, Park et al. [25] proposed computed tomography-like light field reconstruction (LFR) by back-projecting the multifocal intensity images. However, the incomplete reconstruction restricted by the depth-of-field of cameras may suffer from artifacts. Subsequently, Mousnier et al. [26] and Chen et al. [27] overcame the defocus noise problem in LFR by recognizing the focused information in FS. Recently, researchers regarded focal stack-based light field reconstruction (FS-LFR) as an inverse problem and proposed different iterative algorithms to obtain optimized light fields [28–32].

Although the works mentioned above can successfully recover the light field information from FS, there is still a problem in FS-LFR that has not been discussed and analyzed yet. In the process of acquiring FSs in image space and reconstructing light fields in object space, the inconsistency of image-space recording and object-space reconstruction will reduce the LFR accuracy due to the nonlinear mapping between the image and object spaces of the camera. In this work, for the first time, we investigate the object-image space consistency problem for high-accuracy FS-LFR. We used a pre-calibrated light field camera to establish the precise metric mapping between image and object spaces [37], based on which light fields in the two spaces were reconstructed using the same FS. The reconstructed light fields were compared with the reference signal provided by the light field camera to discuss and analyze the reconstruction performance in digital refocusing, viewpoint switching, angular resolution, depth range, and sampling rate. Experimental results demonstrate that the object-image space inconsistency significantly reduces the FS-LFR accuracy, while the reconstruction error is more sensitive to the edges of objects, viewpoint, and depth range. It verifies the necessity of object-image space consistency for high-quality FS-LFR, which can be used as a mechanism to realize and apply computational light field imaging when the recording and reconstruction lie in nonlinear mapping spaces.

Principle and method

2.1. Light field characterization

The term "Light Field" was first introduced by Gershun in 1939 [38] to describe the radiation propagation of light rays in space. In 1991, Adelson and Bergen [39] proposed a seven-dimensional plenoptic function to characterize the distribution of light rays in visual space: $L(x, y, z, \theta, \omega, \lambda, t)$, which represents that a light ray with a wavelength λ passes through a position (x, y, z) at angles (θ, ω) at a moment t. In 1996, Levoy and Hanrahan [40] proposed the light field rendering theory in which the attenuation of light propagation in free space was neglected so that the plenoptic function was reduced to 4D. As shown in Fig. 1(a), each light ray in space can be represented by a straight line intersecting two parallel planes, i.e., $L(x, y, u, v)$, where L denotes the radiance and (x, y), (u, v) denote the spatial and angular coordinates, respectively. In 1998, Camahort et al. [41] proposed another light field representation, as shown in Fig. 1(b), where a light ray in space crosses a plane (x, y) at angles (θ, ω), denoted as $L(x, y, \theta, \omega)$.

2.2. Light field reconstruction based on focus stack

In the process of FS-LFR, the sensor records a series of intensity images to obtain an FS by changing the relative distance between the camera and the measured object (e.g., object distance, image distance, and focal length), as shown in Fig. 2(a). The FS recorded using a traditional camera reflects the transport-of-intensity property in light field propagation, which can be regarded as a digital modulator for incident light fields. In the spatial domain, the intensity transmission of light fields is equivalent to light field projection along different directions, as shown on the right side of Fig. 2(b). Light field projection carries rich information about the light field structure, based on which the light field can be reconstructed from FS. Levin et al. [42] pointed out that in the spatial frequency domain, the energy of light field intensity transmission is mainly concentrated in a 3D subset of the 4D light field spectrum, i.e., a dimensionality gap exists. The light field information can be (partially) recovered from FS by giving an appropriate light field prior assumption or establishing an appropriate inverse reconstruction model. The latter for FS-LFR is discussed in detail below.

As visualization in the light field coordinate system, light field propagation is equivalent to a shear operation on light fields, namely resampling spatial coordinates using angular coordinates and a shear factor. In object space, the shear factor is the light field propagation distance Δz. By spatially shearing and angularly integrating the light field, an intensity image on a specific focal plane can be obtained, based on which a light field forward imaging model can be established as

$$i_{\Delta z}(x, y) = \iint L_{\text{obj}}(x + \Delta z\theta, y + \Delta z\omega, \theta, w)d\theta d\omega \tag{1}$$

where $i_{\Delta z}$ represents the integrated image in object space. In fact, images are acquired in image space using a camera. In this case, the light field forward imaging mode can be expressed as

$$i_{\alpha_m}(x, y) = \iint L_{\text{im}}\left(x + \left(1 - \frac{1}{\alpha}\right)u, y + \left(1 - \frac{1}{\alpha}\right)v, u, v\right)dudv \tag{2}$$

where α denotes the shear factor in image space, which is the ratio of the distance from the sensor plane to the main lens and the focal length. FS can be obtained by equivalently varying the shear factor: $\{i_{\alpha_m}\}, m = 1, 2, \cdots, M$, where M is the number of images in FS.

Once FS is obtained, LFR becomes an inverse solution of the forward imaging model. In practice, the propagation distance Δz is known, for example, using a translation stage, and the object-space light field can be reconstructed from FS by constructing an inverse problem as follows:

$$\underset{L_{\text{obj}}}{arg\,\min} \sum_m ||i_{\alpha_m}(x, y) - \sum_\theta \sum_\omega L_{\text{obj}}\left(x + \Delta z_m\theta, y + \Delta z_m\omega, \theta, \omega\right)||^2 \tag{3}$$

The left side of Fig. 2(b) illustrates the sampling result of the reconstructed light field computationally optimized from the light field projection, where a point corresponds to a light ray in Fig. 2(a). The inverse

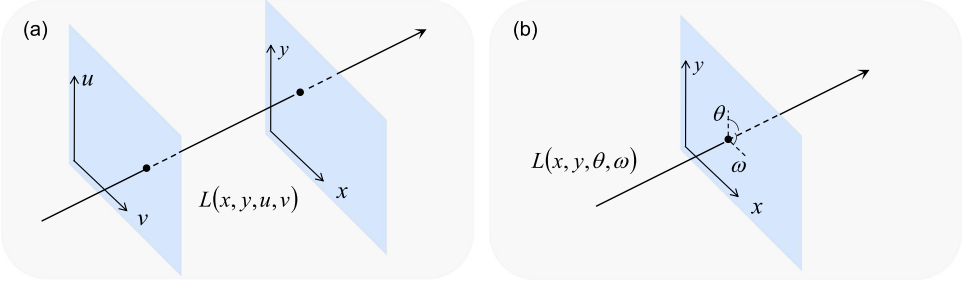

Fig. 1. Light field parameterization: (a) two-plane model; (b) position-angle model.

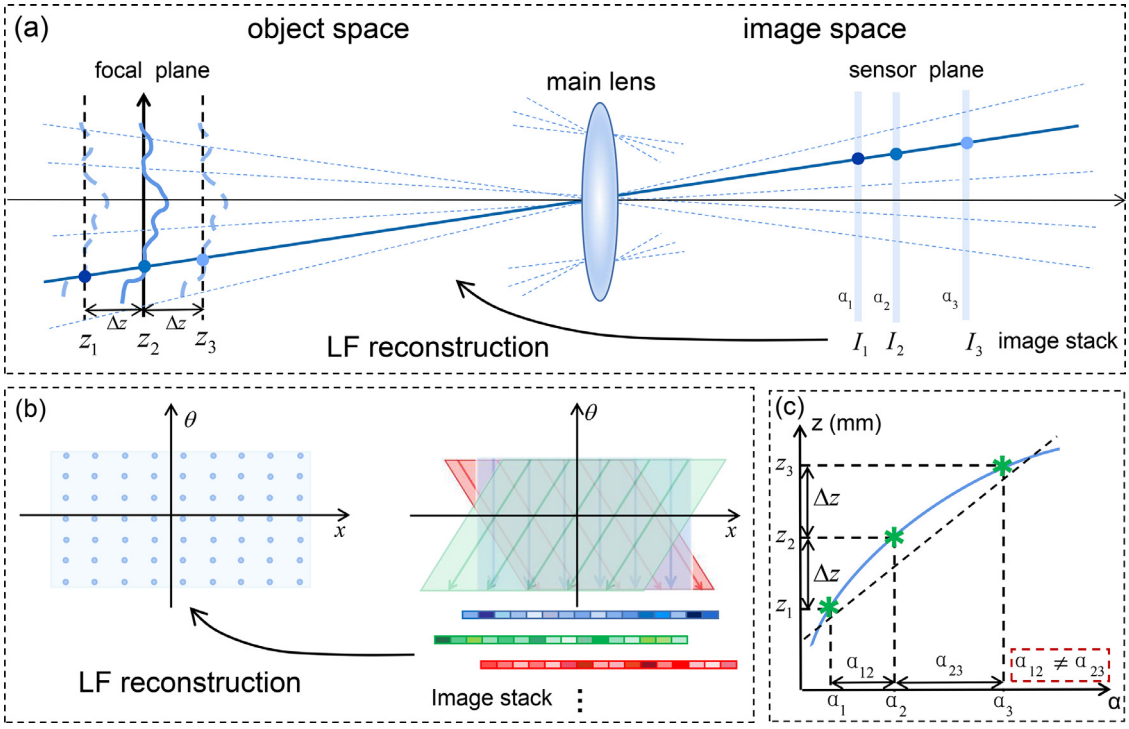

Fig. 2. Schematic diagram of FS-LFR: (a) light field in image space propagates to different image planes to form an image stack, which is used to reconstruct the light field in object space by virtue of the transport-of-intensity property in FS; (b) light field coordinate system associated with (a), the image stack is equivalent to the projection of the light field along different directions, and LFR is equivalent to correspondingly recover a point set in the coordinate system, where a point corresponds to a light ray in (a); (c) nonlinear metric mapping curve of depths between object and image spaces.

model in Eq. (3) involves two light field shear factors, Δz and α, which correspond to the depth dimensions of light field propagation in object and image spaces, respectively. The relationship of depth mapping is nonlinear [33–37], as shown by the mapping curve in Fig. 2(c). Cai *et al.* [37] established an accurate depth metric mapping of a light field camera between object and image spaces: $Z_f = (m_1\alpha + m_2)/(\alpha + m_3)$, where Z_f is the object-space depth in the light field coordinate system, and $\{m_1, m_2, m_3\}$ are the mapping coefficients. It can be seen that the same propagation distance in object space actually corresponds to different depth distances in image space. This inconsistency between the image-space recording and object-space reconstruction will inevitably introduce LFR error.

In this paper, we discuss and analyze the object-image space consistency problem in FS-LFR in detail. To this end, based on the previous work, we precisely establish the object-image space mapping relationship of a light field camera. Accordingly, the propagation distance Δz in object space can be converted to the shear factor α in image space so that the inverse reconstruction model can be modified as follows:

$$\arg\min_{L_{\text{im}}} \sum_m \left\| i_{\alpha_m}(x,y) - \sum_u \sum_v L_{\text{im}}\left(x + \left(1 - \frac{1}{\alpha_m}\right)u, y + \left(1 - \frac{1}{\alpha_m}\right)v, u, v\right) \right\|^2 \tag{4}$$

In Eq. (4), both FS and LFR belong to the image space of the camera and thus satisfy the consistency condition, which will result in high-quality FS-LFR theoretically.

Results and analysis

In this section, we experimentally validate and analyze the effect of object-image space consistency on FS-LFR. A light field camera (Lytro Illum) was used to capture light field images as reference light fields L_{ref}. The light field camera was pre-calibrated to determine the depth metric mapping coefficients between object and image spaces. It should be noted that the implementation of FS-LFR does not depend on whether the light field camera is calibrated or not. The calibrated depth range was 200 mm. The simulated FS was the image sequence calculated by the forward imaging model in Eq. (2) using shear factors corresponding to equal propagation distances in the depth range. Then, light fields in object and image spaces were reconstructed according to the inverse models in Eqs. (3) and (4), obtaining L_{obj} and L_{im}, respectively. By quantitatively comparing with L_{ref}, the object-image space consistency analysis of FS-LFR was performed.

In the experiment, an FS consisted of 15 refocused images based on L_{ref} decoded with spatial and angular resolutions of 434×625 and 9×9 [33]. In the process of iterative LFR, the residual distribution curves of L_{im} and L_{obj} are shown in Fig. 3. It can be seen that both reconstructions can converge after a certain number of iterations, but the overall error

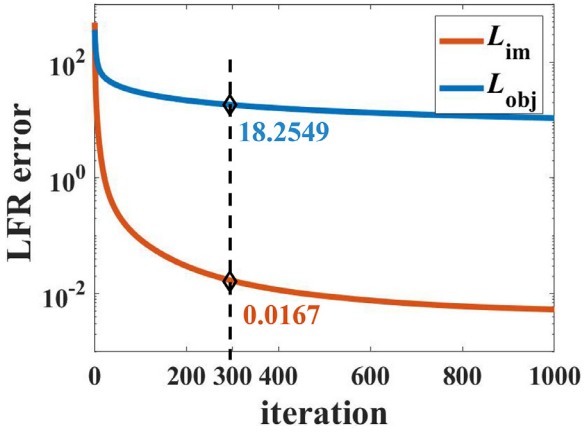

Fig. 3. Residual distribution curves of iterative LFR in object and image spaces, respectively.

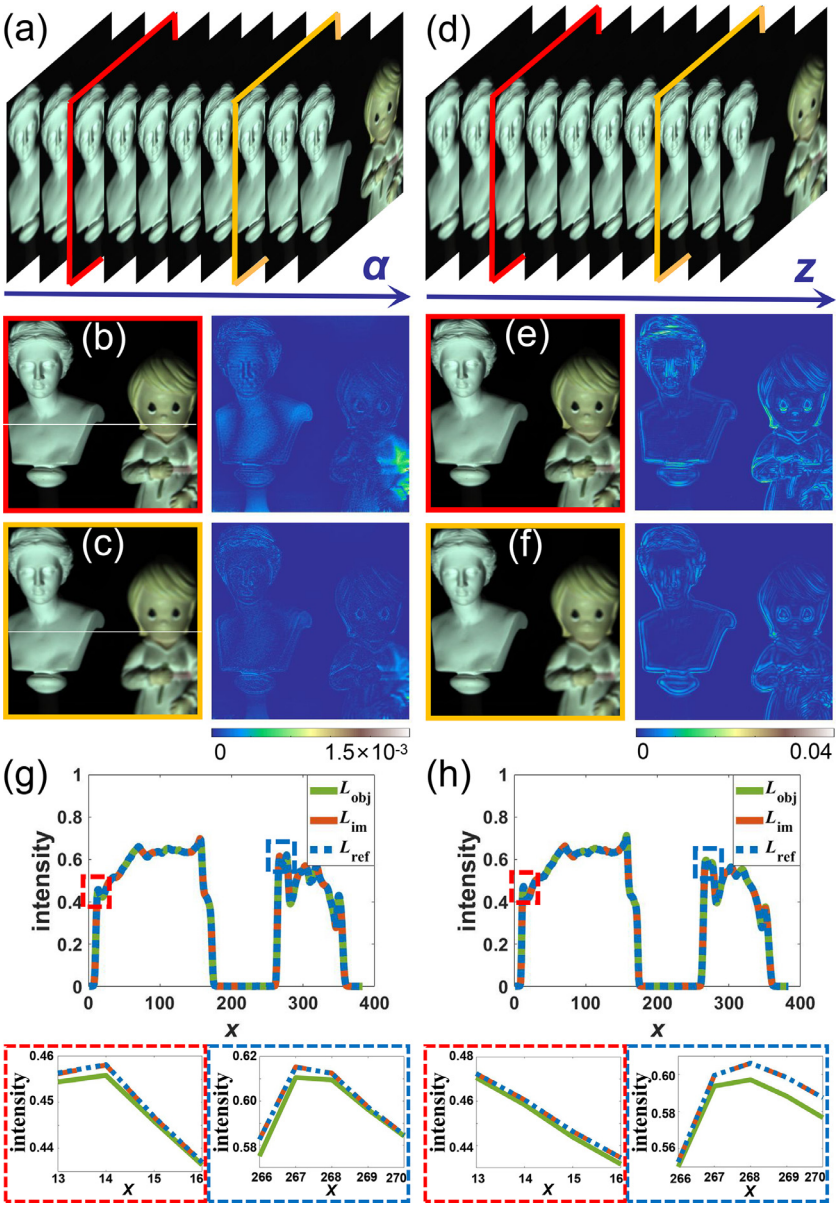

Fig. 4. Digital refocusing of reconstructed light fields in object and image spaces: (a) refocused images of L_{im}; (b) – (c) front views related to the red and orange wireframes marked in (a) and their error maps, respectively; (d) refocused images of L_{obj}; (e) – (f) front views related to the red and orange wireframes marked in (d) and their error maps, respectively; (g) – (h) cross-sectional distribution curves of the refocused images of L_{im}, L_{obj}, and L_{ref} at the same focal depths, related to the white lines marked in (b) and (c), along with the enlarged local views marked by the red and blue wireframes in (g) and (h), respectively.

of L_{im} is always less than that of L_{obj}. Specifically, the error of L_{im} drops to 0.0167 after 300 iterations, while the error of L_{obj} is still as high as 10.8670 after 1000 iterations. Intuitively, FS-LFR exhibits the overall high-quality performance on efficiency and accuracy when satisfying the object-image space consistency. Comprehensively considering the efficiency and accuracy of FS-LFR, we chose L_{im} and L_{obj}, which were optimized through 300 iterations and 1000 iterations respectively, for the following comparison and analysis from different aspects in detail.

3.1. Digital refocusing

The reconstructed L_{im} and L_{obj} were first used to perform digital refocusing, which involved partial 3D structure of light fields, as shown in Fig. 4(a) and 4(d), respectively. Fig. 4(b) and 4(c) show two front views of the refocused images marked by the red and orange wireframes in Fig. 4(a), along with the corresponding error maps compared with the original FS. Fig. 4(e) and 4(f) show similar results related to Fig. 4(d). In comparison, the refocused error of L_{im} is significantly less than that of L_{obj}, which is consistent with the result shown in Fig. 3. In particular, the refocusing error of L_{obj} is more sensitive to the edges of objects. Fig. 4(g) and 4(h) show the cross-sectional distribution curves of the refocused images of L_{im}, L_{obj}, and L_{ref} at the same focal depths, corresponding to the white lines marked in Fig. 4(b) and 4(c), respectively. The local views related to the red and blue wireframes in Fig. 4(g) and 4(h) are enlarged and shown below. It can be seen that there is a higher degree of anastomosis between L_{im} and L_{ref}, while L_{obj} has a degree of overall offset.

3.2. Viewpoint switching

In this subsection, multi-view imaging of L_{im} and L_{obj} was exhibited and compared with that of L_{ref}. Fig. 5(a) and 5(d) show the sketch maps of viewpoint switching related to L_{im} and L_{obj}, respectively (see Visualization 1). The central and boundary views of reconstructed light fields were chosen for comparison. Fig. 5(b) and 5(c) show the enlarged central and boundary sub-aperture images of L_{im} related to the red and orange wireframes marked in Fig. 5(a), along with the error maps compared with that of L_{ref}, respectively. Similarly, Fig. 5(e) and 5(f) show the results related to the central and boundary views marked in Fig. 5(d). Compared with digital refocusing in Fig. 4, the multi-view error becomes

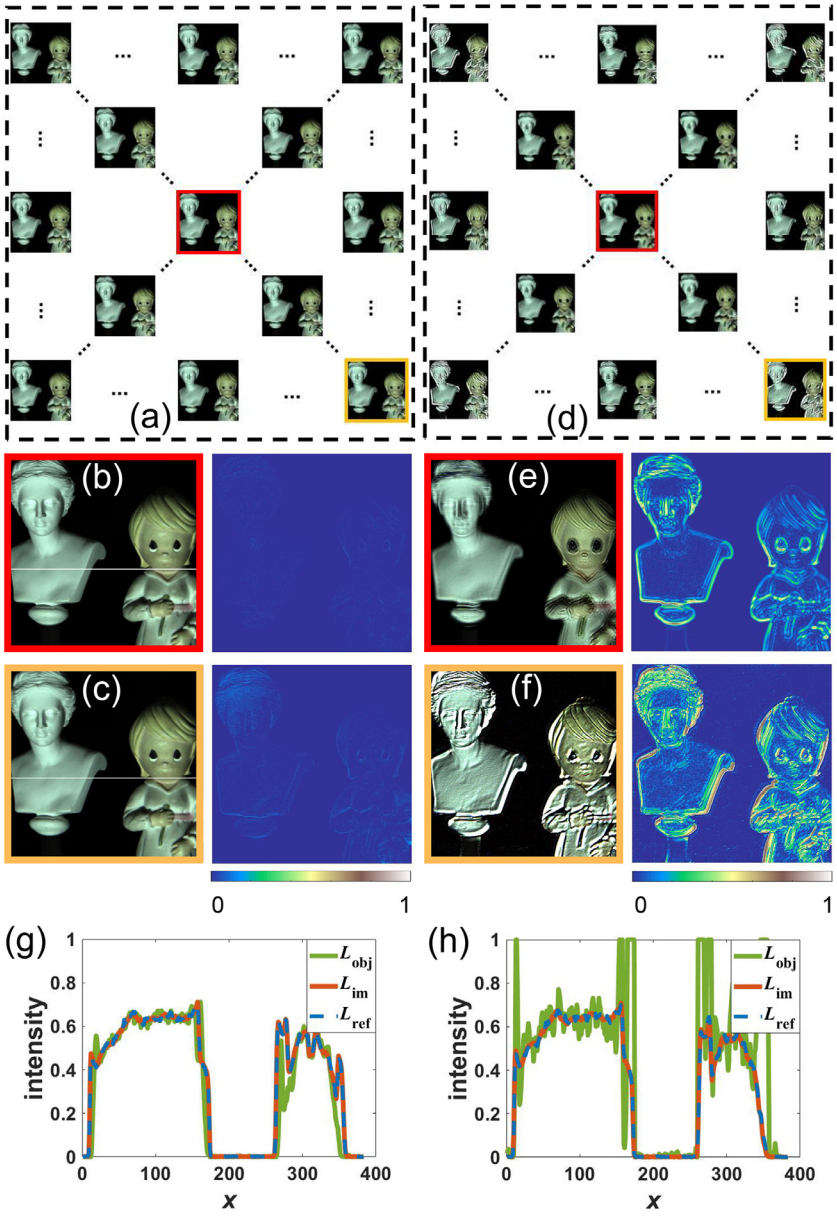

Fig. 5. Viewpoint switching of reconstructed light fields in object and image spaces: (a) sketch map of multi-view imaging of L_{im}; (b) – (c) enlarged central and boundary sub-aperture images related to the red and orange wireframes marked in (a), along with the corresponding reconstruction error maps, respectively; (d) sketch map of multi-view imaging of L_{obj}; (e) – (f) enlarged central and boundary sub-aperture images related to the red and orange wireframes marked in (d), along with the corresponding reconstruction error maps, respectively; (g) – (h) cross-sectional distribution curves of the central and boundary sub-aperture images of L_{im}, L_{obj}, and L_{ref}, related to the white lines marked in (b) and (c), respectively.

more considerable because the entire 4D structure of light fields is considered in this situation. In addition, it can be seen that the reconstruction accuracy from the central view is higher than that from the boundary view for both L_{im} and L_{obj}. This may be related to the lens distortion since the aberration at the edge part of the lens is generally larger than that at the central part. In practice, taking imaging distortion into account in light field metric mapping can improve the precision of light field depth/3D measurement [43]. In this work, we do not analyze this effect in depth but leave it for future work.

Comparing respectively Fig. 5(b) and 5(c) with Fig. 5(e) and 5(f), it can be seen that the accuracy of L_{im} is higher than that of L_{obj} at different viewpoints. As with the refocused images in Fig. 4, the multi-view image error of L_{obj} is more sensitive to the edges of objects. Fig. 5(g) and 5(h) show the cross-sectional distribution curves for the central and boundary sub-aperture images of L_{im}, L_{obj}, and L_{ref}, related to the white lines marked in Fig. 5(b) and 5(c), respectively. Similarly, the distribution curves for L_{im} and L_{ref} largely coincide with each other in both central and boundary views. In the central view, the distribution curve for L_{obj} shows a fluctuation effect to some extent, while in the boundary view, the fluctuation effect becomes more significant.

3.3. Angular resolution

The above comparison experiments of FS-LFR were based on a constant angular resolution (i.e., 9 × 9). In this subsection, we discussed the effect of angular resolution on FS-LFR. As a computational light field imaging technique, FS-LFR with additionally selected angular resolutions of 5 × 5 and 7 × 7 were performed in object and image spaces, respectively. In the experiment, the angular sampling rate was kept constant, so different angular resolutions were equivalent to using the main lens with different aperture sizes for FS-LFR. Fig. 6(a) and 6(c) show the error maps from different views of L_{im} and L_{obj} corresponding to the three angular resolutions, respectively. Fig. 6(b) and 6(d) plot the error distribution curves of L_{im} and L_{obj} for different views, corresponding to the red lines marked in Fig. 6(a) and 6(c), respectively. It can be seen that the reconstruction error of L_{im} is much lower than that of L_{obj}, and the latter is still sensitive to the edges of the measured objects, regardless of the angular resolution chosen for FS-LFR. Furthermore, the errors of both L_{im} and L_{obj} increase with the angular resolution. From the perspective of numerical computation, the attempt to recover more information using the same detected signals is associated with greater

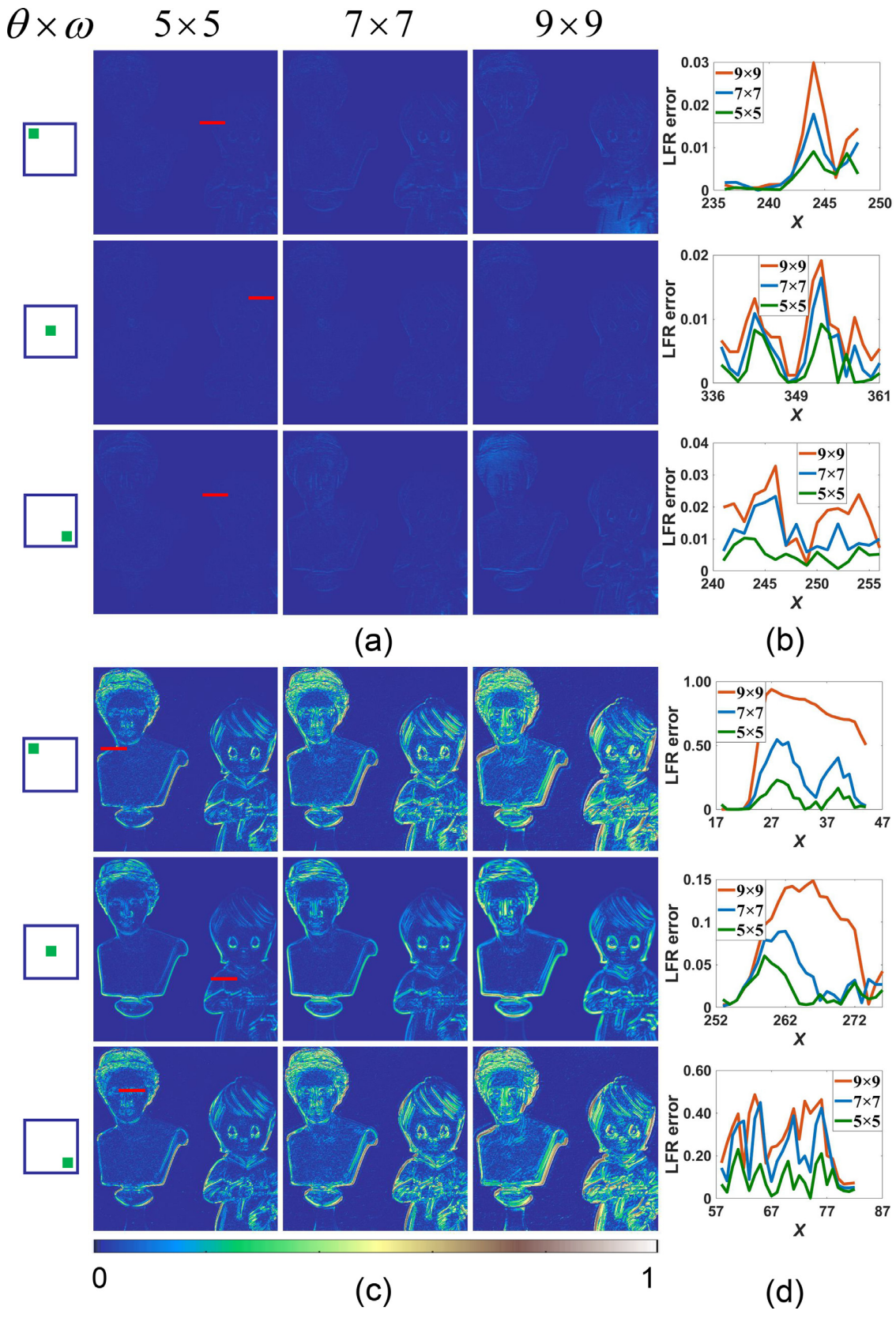

Fig. 6. FS-LFR with different angular resolutions: (a) error maps of L_{im} at different views; (b) error distribution curves related to the red lines marked in (a); (c) error maps of L_{obj} at different views; (d) error distribution curves related to the red lines marked in (c).

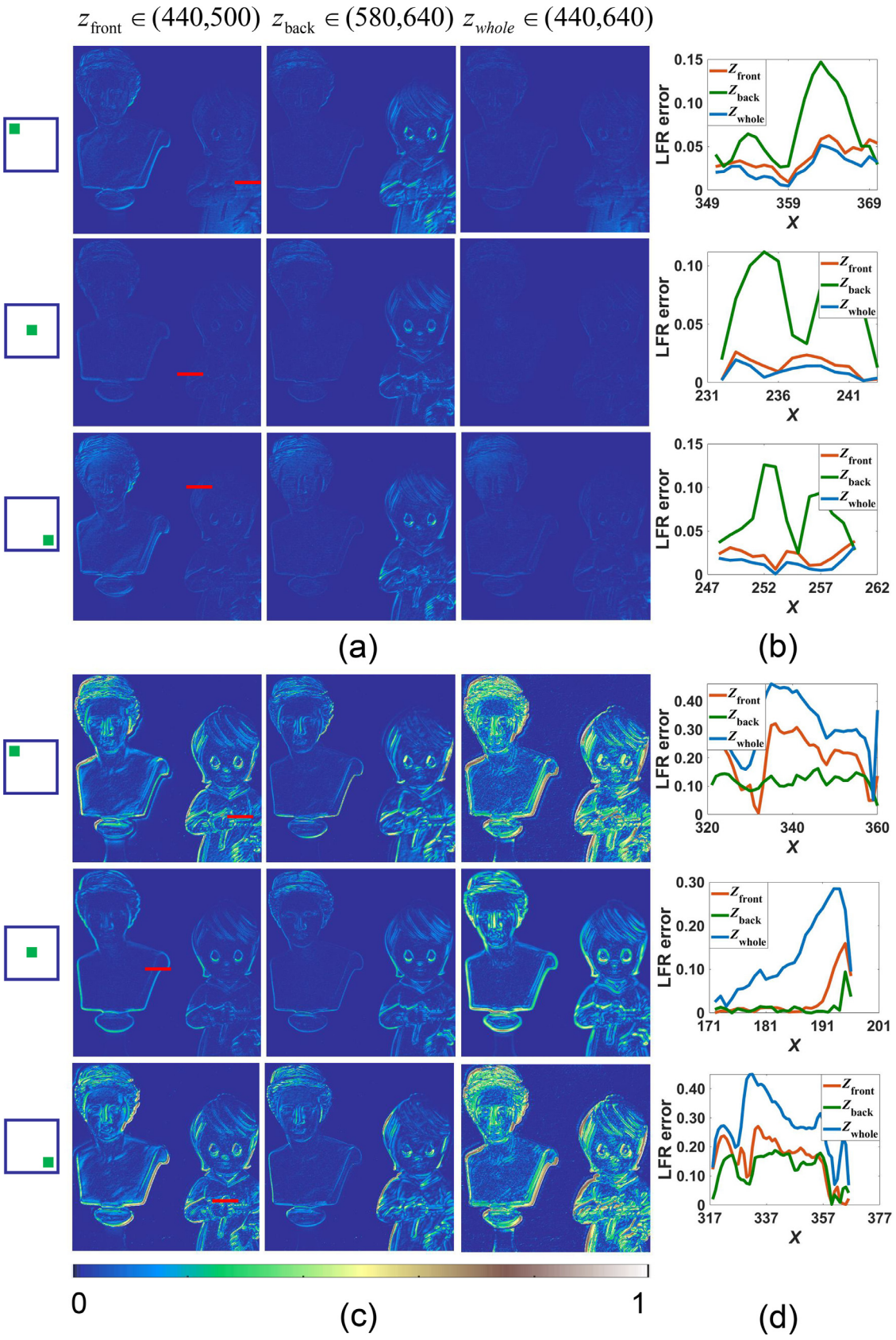

Fig. 7. FS-LFR at different depth ranges: (a) error maps of L_{im} at different views; (b) error distribution curves related to the red lines marked in (a); (c) error maps of L_{obj} at different views; (d) error distribution curves related to the red lines marked in (c).

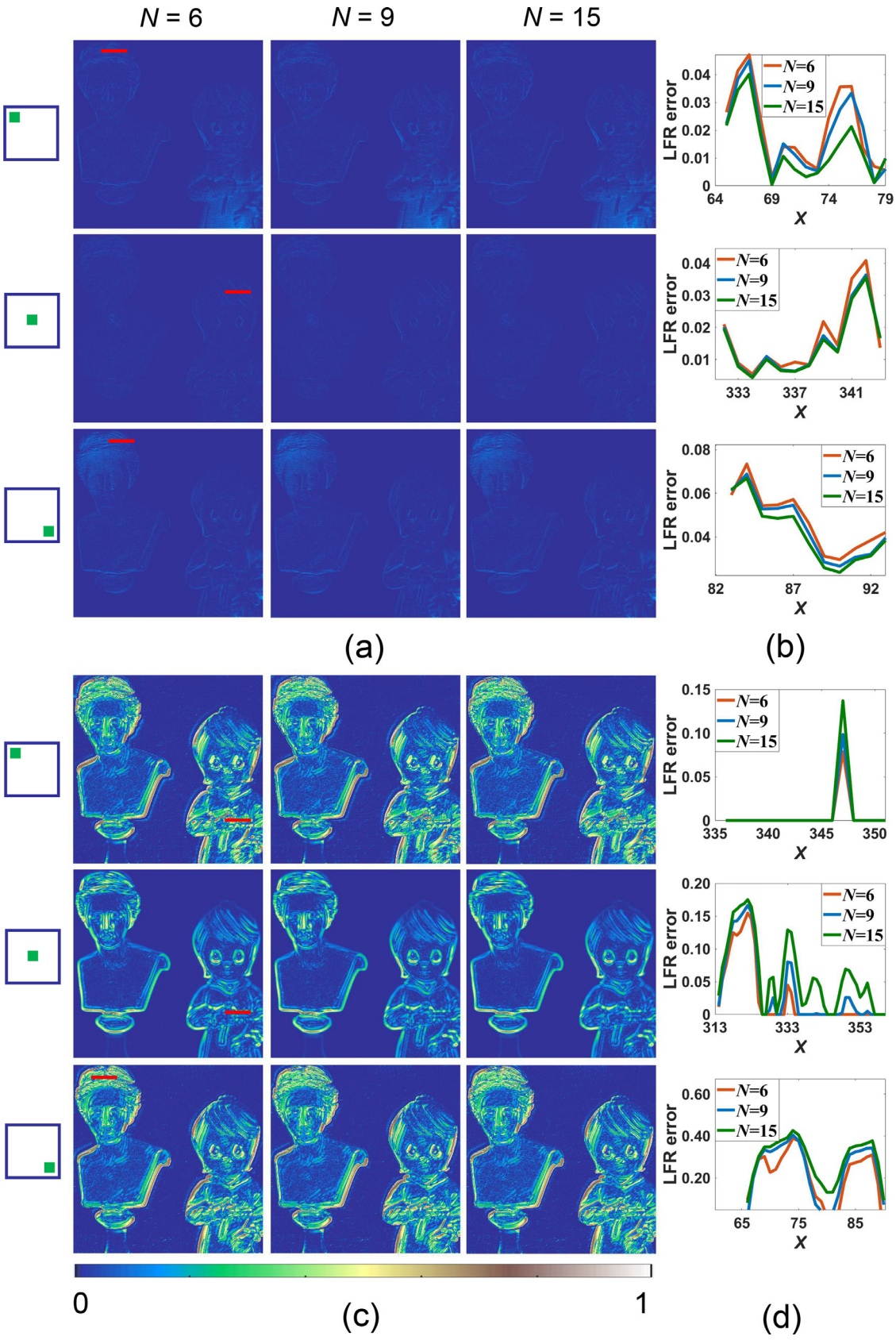

Fig. 8. FS-LFR at different depth sampling rates: (a) error maps of L_{im} at different views; (b) error distribution curves related to the red lines marked in (a); (c) error maps of L_{obj} at different views; (d) error distribution curves related to the red lines marked in (c).

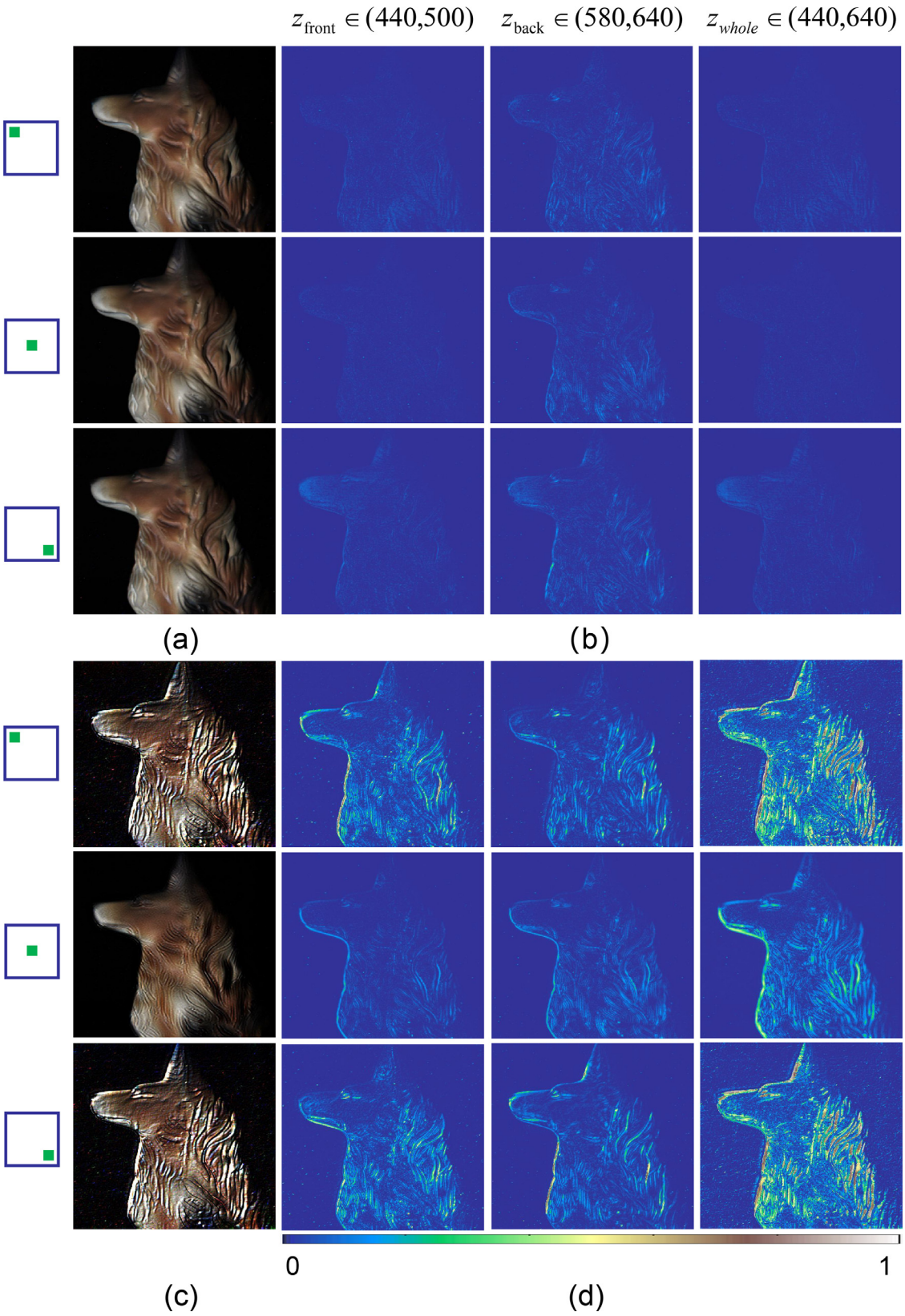

Fig. 9. FS-LFR at different viewpoints and depth ranges: (a) different viewpoint maps of L_{im} in z_{whole}; (b) error maps of L_{im} at different views; (c) different viewpoint maps of L_{obj} in z_{whole}; (d) error maps of L_{obj} at different views.

Table 1
LFR error at different depth ranges.

(θ, ω)		(–4, –4)			(0, 0)			(4, 4)		
		RMSE	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE	PSNR	SSIM
L_{im}	z_{front}	0.0102	39.7912	0.8673	0.0089	40.6937	0.8662	0.0115	38.7947	0.8273
	z_{back}	0.0141	36.9885	0.8587	0.0177	35.0519	0.7885	0.0175	35.1547	0.7874
	z_{whole}	0.0061	44.2559	0.9123	0.0048	46.3601	0.9210	0.0080	41.9536	0.8694
L_{obj}	z_{front}	0.0443	27.0645	0.6751	0.0360	28.8680	0.7036	0.0441	27.1125	0.6073
	z_{back}	0.0285	30.9158	0.7196	0.0369	28.6530	0.6861	0.0327	29.7087	0.6934
	z_{whole}	0.0872	21.1899	0.5418	0.1025	19.7819	0.5454	0.0879	21.1190	0.5419

Table 2
LFR error at different depth sampling rates.

(θ, ω)		(-4,-4)			(0,0)			(4,4)		
		RMSE	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE	PSNR	SSIM
L_{im}	$N = 6$	0.0064	43.8939	0.9067	0.0049	46.2201	0.9204	0.0082	41.7278	0.8641
	$N = 9$	0.0062	44.1068	0.9112	0.0048	46.3351	0.9210	0.0080	41.8983	0.8686
	$N = 15$	0.0061	44.2559	0.9123	0.0048	46.3601	0.9210	0.0080	41.9536	0.8694
L_{obj}	$N = 6$	0.0855	21.3592	0.5512	0.1050	19.5751	0.5449	0.0870	21.2139	0.5462
	$N = 9$	0.0861	21.2970	0.5502	0.1038	19.6747	0.5462	0.0874	21.1735	0.5486
	$N = 15$	0.0872	21.1899	0.5418	0.1025	19.7819	0.5454	0.0879	21.1190	0.5419

computational uncertainty. From the perspective of light field imaging mechanism, a smaller angular resolution corresponds to a lower angular frequency sampling rate when the angular sampling rate is constant. Thus, the effect of the dimensionality gap between 3D FS and 4D light field on FS-LFR is reduced to some extent. However, high-accuracy FS-LFR using small angular resolution may sacrifice the sensitivity of reconstructed light fields to depth sensing.

3.4. Depth range and sampling rate

According to the Fourier slice photography, the spectrum of a refocused image is a slice of 4D light field spectrum, so FS is equivalent to the plane sampling in light field spectrum space. In this subsection, we analyzed the effect of the spectrum sampling rate on FS-LFR. The primary considerations are the depth range of FS and the depth sampling rate. The former is associated with the frequency coverage of the light field spectrum, and the latter corresponds to the plane sampling rate in a specific frequency range.

For the depth range, we divided the calibration space of the light field camera into three parts:$z_{\text{front}} \in (440, 500)$, $z_{\text{back}} \in (580, 640)$, and $z_{\text{whole}} \in (440, 640)$. Keeping the same depth sampling rate, L_{im} and L_{obj} were reconstructed in the three depth ranges and then compared with L_{ref}. Fig. 7(a) shows the error maps of L_{im} at different views (see Visualization 2), where the error distribution curves related to the marked red lines are shown in Fig. 7(b). Fig. 7(c) and 7(d) show the error maps and distribution curves of L_{obj} at different views, respectively. Similarly, no matter which depth range is used, the error of L_{im} is smaller than that of L_{obj}. Intuitively, it can be seen that the increase in the depth range improves the accuracy of L_{im}. However, the result related to L_{obj} shows an opposite trend, i.e., the accuracy decreases with increasing the depth range. Table 1 lists the relative values of root-mean-square error (RMSE), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) corresponding to the error maps in Fig. 7. These values quantitatively reflect the observed phenomena in Fig. 7. Specifically, the increase of the depth range respectively improves the average RMSE, PSNR, and SSIM by –52.69%, 17.07%, and 8.21% for L_{im}, but respectively impair the relative values by 149.53%, –27.94%, and –20.24% for L_{obj}. Consequently, increasing the frequency coverage with the detection depth range can effectively improve the FS-LFR accuracy when the object-image space consistency is satisfied. On the contrary, if the condition is not met, the reconstruction error becomes significantly larger with increasing the depth range.

For the depth sampling rate, we reconstructed light fields in object and image spaces using FS containing 6, 9, and 15 images, respectively, within the same depth range z_{whole}. Fig. 8(a) and 8(c) show the error maps of L_{im} and L_{obj} at different views, respectively, while the corresponding error distribution curves related to the red lines marked in Fig. 8(a) and 8(c) are shown in Fig. 8(b) and 8(d), respectively. Table 2 lists the RMSE, PSNR, and SSIM values corresponding to the error maps in Fig. 8. Similar to FS-LFR at different depth ranges, the accuracy of L_{im} is higher than that of L_{obj} at different depth sampling rates. It can be seen from Fig. 8 and Table 2 that a higher depth sampling rate results in a lesser error of L_{im} and a larger error of L_{obj}. However, the effect of the depth sampling rate on FS-LFR is small compared to that of the depth range.

3.5. Other scenes

In this section, we demonstrate the influencing factors by measuring another scene. It can be seen that the viewpoint and depth range are two main factors for FS-LFR (see also Discussion). We reconstructed L_{im} and L_{obj} in the three depth ranges: $z_{\text{front}} \in (440, 500)$, $z_{\text{back}} \in (580, 640)$, and $z_{\text{whole}} \in (440, 640)$. Fig. 9(a) shows different viewpoint maps of L_{im} in z_{whole}. The error maps of L_{im} at different viewpoints and depth ranges are shown in Fig. 9(b). Fig. 9(c) and (d) show the viewpoint and error maps of L_{obj}. Similar to the comparison results in Fig. 7, the error of L_{im} is smaller than that of L_{obj} in any depth range. Furthermore, when satisfying the object-image space consistency, the increase of the depth range can effectively improve the LFR accuracy.

Discussion

The comparison and analysis of FS-LFR between object and image spaces from four aspects, involving digital refocusing, viewpoint switching, angular resolution, and depth range and sampling rate, can be summarized as follows:

(1) FS-LFR satisfying the object-image space consistency boasts a higher accuracy and less error in multi-view and refocused images regardless of angular resolution, depth range, and depth sampling rate.
(2) Among the influencing factors on FS-LFR in comparison and analysis, two aspects have nothing to do with the object-image space consistency. First, the FS-LFR error in the central view is generally smaller than in other views, which may be associated with

the imaging distortion of the main lens. Second, the FS-LFR error decreases with the angular resolution, which in turn may reduce the sensitivity of light field depth sensing.

(3) When against the object-image space consistency, the FS-LFR error for refocused and multi-view images is sensitive to the edges of objects. Even for the central sub-aperture image, the reconstruction error is still more significant at the edges of objects.

(4) Increasing the depth range and sampling rate of FS can effectively improve the LFR accuracy when satisfying the object-image space consistency. On the contrary, increasing the depth range of FS will significantly reduce the LFR accuracy in the case of inconsistency. In comparison, the effect of the depth sampling rate on FS-LFR is weaker.

Conclusion

In this paper, we investigate the object-image space consistency of FS-LFR on the base of the pre-calibrated metric mapping of depths between object and image spaces. In experiments, we quantitatively compared and analyzed the reconstruction performances of FS-LFR in the two spaces in terms of four aspects: digital refocusing, viewpoint switching, angular resolution, and depth range and sampling rate. The results validate that the object-image space consistency needs to be met for high-accuracy FS-LFR. In practice, when using a traditional camera to capture an image sequence for high-quality computational light field imaging and measurement, the images can be transformed into the space where light field information are recovered through pre-calibrating the camera, which will be studied and realized in our future work.

Author statement

Y.L. and R.Z. contributed equally to this work. Z.C. and R.Z initiated the project and developed the theory and method. Y.L., R.Z and Z.C wrote code for the experiment and simulations. Y.L. and R.Z. designed the experiments. Y.L., R.Z and S.F. prepared and performed the experiments. Y.L. analysed the data. Y.L. and R.Z. prepared the supplementary movies. Y.L., R.Z. and Z.C. wrote the manuscript with contributions from all authors. Q. Chen, C. Z, and Z.C. provided overall supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (62175109, 62061136005, 61905115, 62105151, U21B2033), National Major Scientific Instrument Development Project (62227818), Leading Technology of Jiangsu Basic Research Plan (BK20192003), Youth Foundation of Jiangsu Province (BK20190445, BK20210338), Biomedical Competition Foundation of Jiangsu Province (BE2022847), Key National Industrial Technology Cooperation Foundation of Jiangsu Province (BZ2022039), Fundamental Research Funds for the Central Universities (30920032101), and Open Research Fund of Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense (JSGP202105, JSGP202201).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.optlaseng.2023.107539.

References

[1] Ng R, Levoy M, Brédif M, Duval G, Horowitz M, Hanrahan P. Light field photography with a hand-held plenoptic camera. Stanford University. Stanford Technical Report CTSR; 2005. 2005-02.
[2] Lumsdaine A, Georgiev T. The focused plenoptic camera. In: 2009 IEEE International Conference on Computational Photography. IEEE; 2009. p. 1–8.
[3] Levoy M. Light fields and computational imaging. Computer (Long Beach Calif) 2006;39:46–55.
[4] Ihrke I, Restrepo J, Mignard-Debise L. Principles of light field imaging: briefly revisiting 25 years of research. IEEE Signal Process Mag 2016;33:59–69.
[5] Wu G, Masia B, Jarabo A, Zhang Y, Wang L, Dai Q, Chai T, Liu Y. Light field image processing: an overview. IEEE J Sel Top Signal Process 2017;11:926–54.
[6] Georgiev T, Chunev G, Lumsdaine A. Superresolution with the focused plenoptic camera. In: Proc, 7873. SPIE; 2011. p. 232–44.
[7] Wanner S, Goldluecke B. Spatial and angular variational super-resolution of 4D light fields. In: European Conference on Computer Vision. Springer; 2012. p. 608–21.
[8] Lim Y-T, Park J-H, Kwon K-C, Kim N. Resolutionenhanced integral imaging microscopy that uses lens array shifting. Opt Express 2009;17:19253–63.
[9] Mukati MU, Gunturk BK. Light field super resolution through controlled micro-shifts of light field sensor. Signal Process Image Commun 2018;67:71–8.
[10] Wu J, Lu Z, Jiang D, Guo Y, Qiao H, Zhang Y, Zhu T, Cai Y, Zhang X, Zhanghao K. Iterative tomography with digital adaptive optics permits hour-long intravital observation of 3D subcellular dynamics at millisecond scale. Cell 2021;184:3318–32.
[11] Wilburn BS, Smulski M, Lee H-HK, Horowitz MA. Light field video camera. In: Media Processors 2002, 4674. SPIE; 2001. p. 29–36.
[12] Wilburn B, Joshi N, Vaish V, Talvala E-V, Antunez E, Barth A, Adams A, Horowitz M, Levoy M. High performance imaging using large camera arrays. In: ACM SIGGRAPH 2005 Papers. ACM; 2005. p. 765–76.
[13] Unger J, Wenger A, Hawkins T, Gardner A, Debevec P. Capturing and rendering with incident light fields. Eurographics Workshop on Rendering 2003:141–9.
[14] Levin A, Fergus R, Durand F, Freeman WT. Image and depth from a conventional camera with a coded aperture. ACM Trans Graphics (TOG) 2007;26:70.
[15] Liang C-K, Lin T-H, Wong B-Y, Liu C, Chen HH. Programmable aperture photography: multiplexed light field acquisition. In: ACM siggraph 2008 papers. ACM; 2008. p. 1–10.
[16] Veeraraghavan A, Raskar R, Agrawal A, Mohan A, Tumblin J. Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. ACM Trans Graphics (TOG) 2007;26:69.
[17] Marwah K, Wetzstein G, Bando Y, Raskar R. Compressive light field photography using overcomplete dictionaries and optimized projections. ACM Trans Graphics (TOG) 2013;32:1–12.
[18] Cai Z, Chen J, Pedrini G, Osten W, Liu X, Peng X. Lensless light-field imaging through diffuser encoding. Light: Sci Appl 2020;9:1–9.
[19] Antipa N, Necula S, Ng R, Waller L. Single-shot diffuser-encoded light field imaging. In: 2016 IEEE International Conference on Computational Photography IEEE; 2016. p. 1–11.
[20] Lien M-B, Liu C-H, Chun IY, Ravishankar S, Nien H, Zhou M, Fessler JA, Zhong Z, Norris TB. Ranging and light field imaging with transparent photodetectors. Nat Photonics 2020;14:143–8.
[21] Ng R. Fourier slice photography. ACM Trans Graphics(TOG) 2005;24:735–44.
[22] Levin A, Durand F. Linear view synthesis using a dimensionality gap light field prior. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE; 2010. p. 1831–8.
[23] Alonso JR, Fernández A, Ferrari JA. Reconstruction of perspective shifts and refocusing of a three-dimensional scene from a multi-focus image stack. Appl Opt 2016;55:2380–6.
[24] Orth A, Crozier KB. Light field moment imaging. Opt Lett 2013;38:2666–8.
[25] Park J-H, Lee S-K, Jo N-Y, Kim H-J, Kim Y-S, Lim H- G. Light ray field capture using focal plane sweeping and its optical reconstruction using 3D displays. Opt Express 2014;22:25444–54.
[26] Mousnier A, Vural E, Guillemot C. Partial light field tomographic reconstruction from a fixed-camera focal stack. ArXiv Preprint ArXiv 2015 1503.01903.
[27] Chen N, Ren Z, Li D, Lam EY, Situ G. Analysis of the noise in backprojection light field acquisition and its optimization. Appl Opt 2017;56:F20–6.
[28] Pérez F, Pérez A, Rodríguez M, Magdaleno E. Lightfield recovery from its focal stack. J Math Imaging Vis 2016;56:573–90.
[29] Liu C, Qiu J, Jiang M. Light field reconstruction from projection modeling of focal stack. Opt Express 2017;25:11377–88.
[30] Blocker CJ, Chun Y, Fessler JA. Low-rank plus sparse tensor models for light-field reconstruction from focal stack data. In: 2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP). IEEE; 2018. p. 1–5.
[31] Gao S, Qu G, Sjöström M, Liu Y. A TV regularisation sparse light field reconstruction model based on guided-filtering. Signal Process Image Commun 2022;109:116852.
[32] Yin X, Wang G, Li W, Liao Q. Iteratively reconstructing 4D light fields from focal stacks. Appl Opt 2016;55:8457–63.
[33] Dansereau DG, Pizarro O, Williams SB. Decoding, calibration and rectification for lenselet-based plenoptic cameras. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2013. p. 1027–34.
[34] Bok Y, Jeon H-G, Kweon IS. Geometric calibration of micro-lens-based light field cameras using line features. IEEE Trans Pattern Anal Mach Intel 2017;39:287–300.
[35] Zhang Q, Zhang C, Ling J, Wang Q, Yu J. A generic multi-projection-center model and calibration method for light field cameras. IEEE Trans Pattern Anal Mach Intel 2019;41:2539–52.
[36] Li C, Zhang X, Tu D. Metric three-dimensional reconstruction model from a light field and its calibration. Opt Eng 2017;56:013105.

[37] Cai Z, Liu X, Pedrini G, Osten W, Peng X. Unfocused plenoptic metric modeling and calibration. Opt Express 2019;27:20177–98.
[38] Gershun A. The light field. J Math Phys 1939;18:51–151.
[39] Bergen JR, Adelson EH. The plenoptic function and the elements of early vision. Comput Models Visual Process 1991;1:8.
[40] Levoy M, Hanrahan P. In: Light field rendering. 23rd Annual Conference on Computer Graphics and Interactive Techniques. ACM; 1996. p. 31–42.
[41] Camahort E, Lerios A, Fussell D. Uniformly sampled light fields. In: Eurographics workshop on rendering techniques. Springer; 1998. p. 117–30.
[42] Levin A, Hasinoff SW, Green P, Durand F, Freeman WT. 4D frequency analysis of computational cameras for depth of field extension. ACM Trans Graphics (TOG) 2009;28:1–14.
[43] Cai Z, Liu X, Pedrini G, Osten W, Peng X. Accurate depth estimation in structured light fields. Opt Express 2019;27:13532–46.