EURASIP Journal on Image and Video Processing

REVIEW Open Access

Learning-based light field imaging: an overview

Saeed Mahmoudpour[1,2*], Carla Pagliari[3] and Peter Schelkens[1,2]

*Correspondence:
Saeed.Mahmoudpour@vub.be

[1] Department of Electronics and Informatics (ETRO), Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1050 Brussel, Belgium
[2] imec, Kapeldreef 75, B-3001 Leuven, Belgium
[3] PGEE/PGED/IME, Instituto Militar de Engenharia, Rio de Janeiro, Brazil

Abstract

Conventional photography can only provide a two-dimensional image of the scene, whereas emerging imaging modalities such as light field enable the representation of higher dimensional visual information by capturing light rays from different directions. Light fields provide immersive experiences, a sense of presence in the scene, and can enhance different vision tasks. Hence, research into light field processing methods has become increasingly popular. It does, however, come at the cost of higher data volume and computational complexity. With the growing deployment of machine-learning and deep architectures in image processing applications, a paradigm shift toward learning-based approaches has also been observed in the design of light field processing methods. Various learning-based approaches are developed to process the high volume of light field data efficiently for different vision tasks while improving performance. Taking into account the diversity of light field vision tasks and the deployed learning-based frameworks, it is necessary to survey the scattered learning-based works in the domain to gain insight into the current trends and challenges. This paper aims to review the existing learning-based solutions for light field imaging and to summarize the most promising frameworks. Moreover, evaluation methods and available light field datasets are highlighted. Lastly, the review concludes with a brief outlook for future research directions.

Keywords: Light fields, Depth estimation, Image reconstruction, Compression, Machine learning, Deep learning

1 Introduction

Light field imaging is one of the most promising three-dimensional (3D) imaging technologies that can deliver a photo-realistic representation of the viewing environment and bring viewers rich and immersive visual experiences. Light fields offer additional angular information, compared to conventional two-dimensional (2D) imaging, by gathering light rays from multiple directions. A light field consists of light rays in 3D space flowing through every point and in every direction [1].

The light field market is rapidly expanding, with an increased focus on glasses-free displays. Pioneering 160° horizontal viewing angle displays [2] discretize light field data into many spaced views, providing 3D experience for many spectators at the same time. Light fields can be captured by an array of monocular cameras (rather than a single camera), by a plenoptic camera [3], or be computer-generated. The addition of angular information to the already existing spatial 2D information sets the scene for 6 degrees of freedom (DoF) experiences in emerging immersive media applications. The 6 DoF immersive media will likely become a key asset in future broadcasting services [4] and gaming industry, providing the viewer with a sense of depth and presence in the scene.

Like traditional 2D images, light fields undergo various processing stages from acquisition to visualization; however, their inherent high dimensionality, information redundancy, and inter-view dependency impose new challenges and opportunities in the development of light field processing algorithms. Considering the added angular dimension in light fields, processing tasks such as compression and super-resolution are potentially more complex in light fields than in 2D images. Moreover, with the implicitly recorded depth information in light fields, new processing challenges such as depth estimation, view synthesis, and 3D reconstruction have been introduced that require innovative solutions to exploit the inherent inter-view correlation between multiple views.

As plenoptic cameras capture light fields at low spatial resolutions with narrow baselines, spatial super-resolution methods are often used as an intermediate processing stage to enhance resolution. A camera array enables capturing light fields with a higher spatial resolution and a wide field of view (FoV); however, the larger distance between neighboring cameras results in sparse views. It is therefore necessary to use methods of angular super-resolution and view synthesis to reconstruct dense light fields. As light fields are much larger than 2D images, compression algorithms are vital for efficient data storage and transmission. Light fields have inter-view information redundancy, allowing data compression to be more efficient by sending only selected views at the encoder and reconstructing the missing views at the decoder. Thus, light field compression methods often include a view synthesis/reconstruction stage to improve compression efficiency. Depth estimation is a fundamental light field processing task that is critical for depth sensing and it is serving as an intermediate step in many other vision tasks including view synthesis, spatial super-resolution, and compression.

As learning-based approaches offer advantages for solving complex tasks, light field processing algorithms are increasingly relying on learning frameworks to enhance efficiency. Considering the diversity of light field processing tasks, content characteristics, and learning-based methods, a systematic summary of existing research is essential to understanding the main applications and architectures of light field processing, as well as identifying limitations and providing a roadmap for future research.

In recent years, several reviews on light field processing have been published. A comprehensive review paper by Wu et al. [5] in 2017 covered the theory and different processing tasks in light fields, as well as their applications in computer vision. However, most of the methods discussed in that paper were based on traditional image processing techniques, and since then, there have been significant advances in using learning-based methods. Two detailed surveys on light field compression, published in 2020 [6, 7], discussed different coding schemes as well as future research directions and standardization efforts.

In this review paper, a comprehensive overview of the learning-based solutions—emerged as a promising paradigm in recent years—for light field processing is provided. Unlike previous surveys that covered more high-level reviews of light field technology or focused on a specific light field processing task, this paper focuses specifically on the applications of learning-based techniques to different stages of light field processing, such as depth estimation, reconstruction, rendering, compression, and quality assessment. We describe the most popular learning-based architectures for light field processing, summarize the available benchmarking datasets and evaluation methods, and discuss the current challenges and future perspectives. By reviewing the state-of-the-art methods and highlighting the key challenges and opportunities in this field, we aim to understand the current status quo and to identify the open challenges and future directions for research in this field, offering a timely and valuable reference for researchers.

The fundamentals of light field imaging, as well as light field acquisition/generation and visualization, which are necessary stages to deliver light field data to the viewer, are introduced in Sect. . Section outlines a number of light field processing tasks and elaborates on the main learning-based frameworks deployed in the literature for each task. The existing light field datasets and quality assessment methods to evaluate light field processing algorithms are described in Sect. . Section discusses the current challenges and future research directions, and finally, Sect. concludes the paper.

2 Light field imaging background

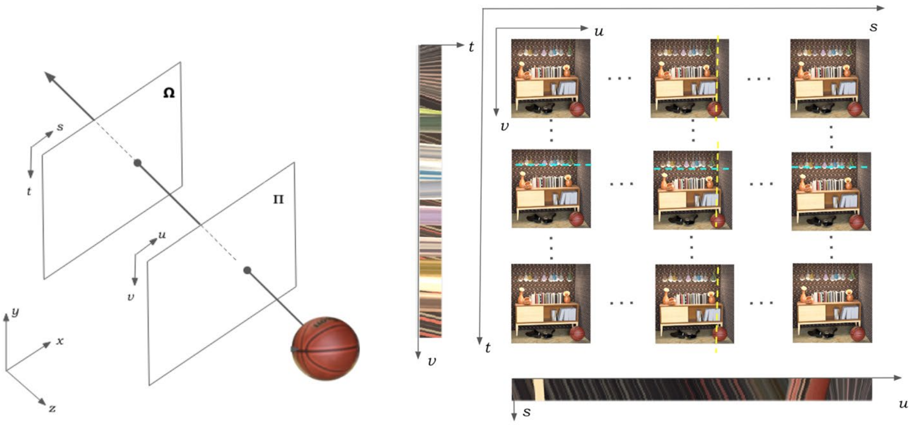
When acquiring a 3D scene using a light field camera, the scene representation is transformed into 2D encoded light field data, with the captured light rays recorded in 2D planes. This process can be exchangeable when a recorded 3D object on 2D planes is reconstructed on a light field display, for example. Light field data need to be processed to provide an immersive 3D experience with full parallax and wide FoV to the viewer. Many challenges need to be overcome before this can be achieved, such as compressing the huge amount of data, rendering the plenoptic scene at different focus planes, and exhibiting the light field to single users wearing headsets or to multiple users in light field displays. Therefore, acquisition/generation and visualization of light fields are key components to meet the functionalities given various use case contexts, such as XR-based, industrial, and medical imaging [5, 8] applications.

2.1 Light field fundamentals

To render 3D scenes that look as realistic as possible, a camera system needs to capture information from a huge number of viewpoints. A light field consists of light rays in 3D space flowing through every point and in every direction [1]. Light is electromagnetic radiation moving along rays in space. A grayscale snapshot of a 3D scene is produced when a stationary person, or a pinhole camera, captures the intensity of light from a single viewpoint, at a single time, averaged over the wavelengths of the visible spectrum, parametrized as $P(\theta, \phi)$. A color snapshot of the same scene is produced when the intensity of light is given as a function of wavelength and parametrized as $P(\theta, \phi, \lambda)$. The intensity of light captured from a single viewpoint, over time, as a function of wavelength can be parametrized as $P(\theta, \phi, \lambda, \tau)$, thus corresponding to a sequence of snapshots (or a movie). The intensity of light captured from any viewpoint, over time, as a function of wavelength can be parametrized as $P(\theta, \phi, \lambda, \tau, V_x, V_y, V_z)$. This 7D structure is known as the plenoptic function describing a 7D ray space, with each point in this space corresponding to a single light ray [1].

Capturing the complete plenoptic function for a scene requires densely placing, ideally a huge number of cameras to scan every point and in every direction, which is naturally performed by the human visual system (HVS) [1]. A plenoptic image modality can be represented as a light field, a point cloud, or a hologram, which are sampled representations of the plenoptic function in the form of, respectively, a vector function that represents the radiance of a discretized set of light rays, a collection of points with position and attribute information, or a complex wavefront. The first plenoptic image modality can be computer-generated or acquired by several capturing devices in the form of light field images. The classical work on photometry [9] considered the light field as a vector field in 3D space. A plenoptic function [1] parameterizes each light ray of a point in space with its 3D location, direction of arrival, and time-varying power spectral density. Thus, light field data carry both spatial and angular information about the light reaching the sensor, providing different viewpoints of the same 3D scene.

Considering fixed lighting conditions and a static scene, it is possible to discard the wavelength (λ) and the time (τ) information [1] from the 7D plenoptic function which becomes a 5D function with 3D position (V_x, V_y, V_z) and 2D direction (θ, ϕ) inputs. This 5D function is used to represent scenes as Neural Radiance Fields (NeRF) [10], detailed in section . Furthermore, a four-dimensional (4D) light field parametrization was simultaneously proposed in [11] and in [12], considering that the intensity of light rays remains constant along a straight line. This constraint allows the representation of each ray of light using the two angles (θ and ϕ) as the propagation direction from viewpoints V_x and V_y. Instead of using the angles and the rays direction to the eye (or the camera), one can parameterize the plenoptic function in terms of spatial coordinates (u, v) of plane Ω (image or focal plane), which is parallel to plane Π (camera plane) that gives the angular distribution of the light rays indexed by (s, t), as depicted in Fig. 1a. This 4D parameterization is known as the two-plane parametrization [11] with $P(u, v, s, t)$ defining the 4D light field representation in terms of the spatial and view coordinates. Different ways of parameterizing light fields can be found in [13].



(a) (b)

Fig. 1 **a** Two-plane parametrization; **b** light field represented as 2D arrays of 2D images and horizontal and vertical EPIs

With the two-plane parametrization, any 4D light field can be conveniently represented by an array of 2D images, indexed by (u, v), at different points of view, indexed by (s, t), as shown in Fig. 1b. This representation enables the compression of light fields using standard codecs designed for 2D sequences, also allowing the employment of 2D native algorithms to process light field data.

The two-plane parameterization also allows for another way to represent light fields. To extract geometric information about the scene, Bolles et al. [14] introduced a technique for building a 3D description of a static scene from a dense sequence of images. The geometry that relates the cameras, points in 3D, and the corresponding observations is referred to as the epipolar geometry (of a stereo pair). When the image planes of cameras are parallel to each other and the baseline (distance between the cameras optical centers) [15], the camera centers are at the same height, and the focal lengths are equal, then the epipolar lines [15] fall along the horizontal scan lines of the images. Therefore, a spatio-temporal (or spatial-view, or spatial–angular, or space-view) representation, named epipolar plane image (EPI) [14] (Fig. 1b), can be constructed. The EPI is a 2D spatial–angular slice of the light field containing patterns of oriented lines, making it possible to estimate depth maps by analyzing the disparity of each line in the EPIs [14]. Light fields represented as 2D arrays of 2D images and EPIs are widely used as input data by light field imaging methods.

2.2 Light field acquisition

A light field could be acquired by a single camera, with special lenses (lenslet-based cameras), or by an array of monocular cameras distributed on a planar or spherical surface to simultaneously capture light field samples from different points of view [11]. The former type of camera is equipped with an array of microlenses positioned between the main lens and the camera sensor. This lens arrangement multiplexes the light rays from the 3D scene, creating microimages [16]. The pixels with the same coordinates relative to each microlens are grouped, forming a sub-aperture image (SAI). Several plenoptic cameras are available in the consumer market, all supplied by one of the two leading manufacturers of this type of imaging technology, namely Raytrix [17] and Lytro (Lytro ceased operations in late March 2018) [18]. This acquisition method provides a dense angular sampling of viewpoints using a single camera. Alternatively, light fields can be acquired using arrays of sensors (cameras), where the pixels given by the spatial coordinates (u, v) of the 4D light field are determined by the cameras, and the view (angular) coordinates (s, t) are given by the number of cameras and their distribution (e.g., planar, spherical) [5]. Light fields can also be computer-generated with accurate disparity maps per view, providing benchmarks for depth and disparity estimation algorithms. Light field acquisition is also possible with a hand-held camera [19], achieving good scene coverage via the proposed “Simultaneous Localization and Mapping” technique. The same work presents a rendering algorithm tailored to the unstructured dense captured data. Acquisition systems designed to capture an unstructured set of limited viewpoints become serviceable due to efficient learning-based algorithms that are able to efficiently generate the missing viewpoints by learning priors on scene geometry, for example.

In [20], LiDAR (Light Detection And Ranging) sensors and an array of multiple cameras are used by autonomous robotics systems for capturing light fields. Salient surfaces given by LiDAR sensors minimize human intervention during the rendering process.

2.3 Light field visualization

The main goal of developing light field visualization techniques is to provide a 3D experience, which is essential to the successful launch of related use cases. Use cases can refer to static light fields, corresponding to a single time sample, where spatial and angular information of a 3D scene are simultaneously captured, and/or to non-static light fields corresponding to multiple time samples. Also, the use cases of light field visualization may be either passive (no viewer interaction) or active (e.g., the viewer will be able to rotate the scene or an object) [21]. In [22], the impact of visualization techniques on light field quality of experience is presented. The results indicate that the perceptual quality of light field images is highly dependent on the rendering methods.

Time-sensitive use cases, such as medical imaging, demand challenging visualization technology to provide view-related interactions for example. For passive use cases, visualization can be accomplished using the simplified 4D representation of a light field. To provide immersive experiences, the captured (or CGI-generated) 3D scenes need to be presented with a large number of viewing angles. The capture of densely sampled views in the real world is a very challenging task, frustrating the experience with 3D light field displays. A way to get around this problem is to employ view interpolation techniques to synthesize (render) dense virtual views for 3D light field display [23–27].

During the acquisition/generation of light fields, two essential properties are defined: the angular (given by the baseline) and the spatial resolution. The former defines the maximum distance between the change of perspective within a light field, while the latter affects visual realism. When recovering 3D from (2D) images, there are cues in the image providing 3D information, such as shading, texture, focus, motion, and perspective [28]. However, existing 2D displays fail to deliver a fully immersive visual experience as critical perceptual cues such as vergence and accommodation, associated with realistic 3D perception, are not provided. In contrast, specific 3D displays would be able to replicate the rays of light (light field), including the directional and color components, to simultaneous viewers in a content-related interaction use case, for example.

The challenge of light field displays is to convey/recreate the light rays that come from the scene to the viewer's eyes. Pixels in 2D displays convey the same color light in all directions, despite the viewer's point of view, while 3D displays convey unique color rays in each direction in a bundle of rays (3D pixels). To reproduce these directional light rays as accurately as possible, the display needs acuity (spatial resolution), several possible viewing angles of the light rays, and their associated depth.

Light field displays cannot reproduce an infinite number of light rays while having to be able to address the vergence–accommodation conflict [29]. Considering the two-plane parametrization (Sect.), the spatial resolution (2D image size) and the angular resolution (2D array view spacing: horizontal and vertical), the display resolution and FoV will establish a volume where the scene objects are exhibited. So, the observer will have to be at a 'valid' point of view to 'see' the object [30, 31]. The visualization techniques and output devices may allow an extended depth of field, refocusing, and 3D views.

Horizontal-parallax-only and vertical-parallax-only displays limit the practical application of 3D display technology and the assessment of the discomfort issue [30–33].

3 Learning-based light field processing

The previous section discussed two key components of the light field imaging pipeline, acquisition/generation, and display/visualization. Light fields go through several intermediate processing stages between the initial acquisition and the final visualization. These processing steps differ from the conventional 2D image processing steps as light fields require handling rays in 3D space to represent a 3D scene. Conventional algorithms are facing their limitation, as realistic applications employing light fields may demand higher accuracy and large-scale computational infrastructure. Learning-based approaches are very promising to improve image processing tasks and support high data-demanding applications as in the light field imaging areas. This section summarizes the most prominent light field processing tasks studied in the literature and highlights the learning-based imaging techniques deployed for each processing task.

3.1 Depth estimation

Depth estimation targets measuring the distance of each pixel relative to the camera, thus inferring 3D information from the images. Depth information can be obtained using active sensing, such as structured light projection, time-of-flight (ToF) measurement, and LiDAR to name a few. Passive disparity/depth techniques are detailed in [34]. Light field imaging enables capturing a scene from multiple viewpoints so depth information is implicitly encoded in the light field representation and can be acquired by computing the inter-view pixel disparity information, as is done in stereo-matching methods. Scene reconstruction and image-based rendering by depth maps involve the ill-posed problem of finding homologous points within the views, which is mitigated by the number of viewpoints of the 3D scene given by a light field. Accurate depth estimation from light field images remains a challenging task especially when it comes to occlusions and photo-consistency constraints of non-Lambertian surfaces. Several works proposed depth/disparity estimation methods using light fields [35–40], handling difficult problems such as specular surfaces and occlusions.

In recent years, learning-based approaches for depth estimation have gained significant traction due to their remarkable performance gains. One of the first learning-based methods for estimating depth was proposed by Johannsen et al. [41], where a light field sparse coding and a disparity-based dictionary were employed. Dictionary atoms represent patches of epipolar plane images (EPIs) constructed from the center view and transformed to the 4D domain using a generative model. Three main strategies have been identified in the literature for learning-based light field depth estimation: (1) autoencoder architectures; (2) stereo matching and refinement; and (3) end-to-end feature extraction and disparity regression. Figure 2 illustrates a generic framework of the three strategies.

3.1.1 Autoencoders

The first depth estimation method based on deep learning is proposed by Heber and Pock [51] in which the horizontal and vertical EPIs were inserted into a five-block

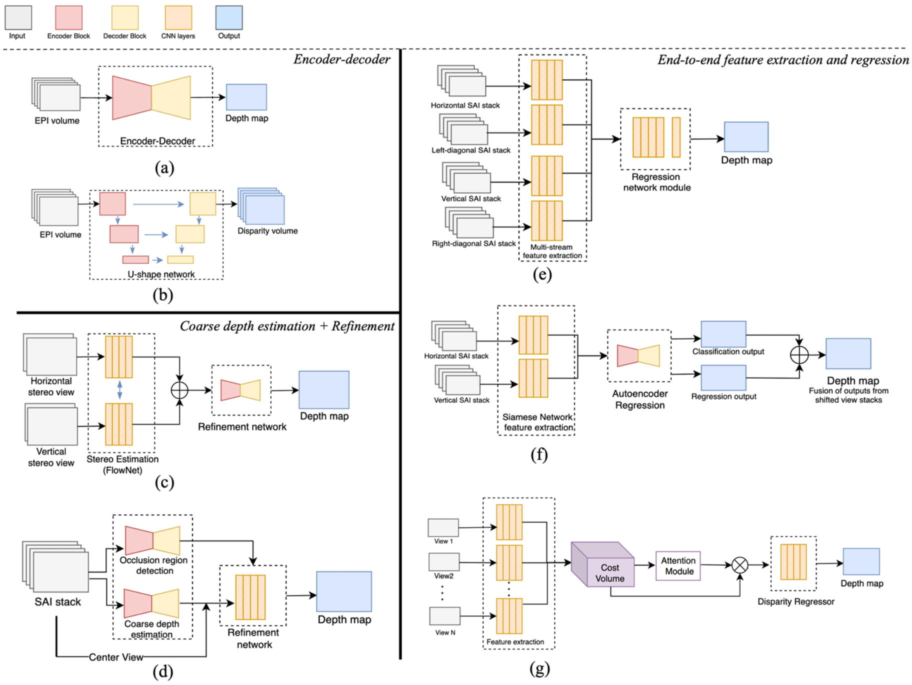


Fig. 2 Examples of architectures for the three existing learning-based depth estimation frameworks. **a** Autoencoders [42], **b** U-shaped encoder–decoder [43, 44], **c** stereo-matching and depth refinement networks [45, 46], **d** residual learning for coarse estimation and occlusion-aware refinement [47], **e** EPINET: multi-stream feature extraction and disparity regression [48], **f** Siamese network for feature extraction and U-Net architecture for disparity regression [49], and **g** end-to-end depth estimation incorporating cost volume and an attention map [50]

convolutional neural network (CNN) architecture to estimate the 2D hyperplane orientation. The slope of the line formed by the pixel shifts across the different views in EPIs is used to obtain the magnitude of disparity. Finally, a 4D regularization term was adopted to compensate for errors in texture-less regions. This method is computationally expensive due to the sliding window processing of EPIs. Later on, the authors proposed a U-shaped encoder–decoder CNN to address this problem and analyze the entire EPI [43]. The encoder part shrinks the size of the input image and represents a low-dimensional form of the image using a set of features. On the decoder side, the features are expanded to obtain the disparity information. The encoder–decoder network is symmetric and includes some skip connections (called pinholes) to preserve high-frequency information. Further, the authors proposed an extension to the U-shape CNN in [44] by learning 3D filters and inserting EPI volumes as inputs.

Following the success of the deep autoencoders for latent feature extraction and depth estimation, Alperovich et al. [42] developed an autoencoder that encodes horizontal and vertical EPI stacks simultaneously using six stages of residual blocks. Then, the compressed representation is expanded using three decoder pathways to address the disparity, diffusion, and specularity estimation problems.

3.1.2 Stereo matching and refinement

While autoencoder-based approaches typically estimate line orientation in EPIs for depth estimation, another group of approaches is based on computing the disparity of matching pixels—using conventional or new stereo-matching methods—on sub-aperture images (SAIs). After a coarse stereo matching depth estimation, these techniques often include a refinement stage to compensate for the large errors, especially in occluded areas (schemes (c) and (d) in Fig. 2). The methods in [45, 46] are based on estimating the disparity between a set of anchor views using a fine-tuned encoder–decoder network designed for optical flow learning (called FlowNet 2.0). Next, the initial depth estimates are combined using an occlusion-aware mask, and finally, a residual learning network is applied to correct the warping errors and the depth map. Rogge et al. [52] performed a coarse depth estimation based on stereo matching and used a belief propagation for regularization. Next, a residual learning network was deployed to refine the depth map. In the same context, an encoder–decoder network was designed by Guo et al. [47] to estimate coarse depths by concatenating SAIs. An occlusion region detection is then performed to obtain a guidance map for estimating depth separately on the occluded and non-occluded areas. In the final stage, the coarse depth map is fed into a refinement network to smooth depth and correct errors.

3.1.3 End-to-end feature extraction and disparity regression

Recently, more depth estimation methods have been proposed that include end-to-end feature extraction followed by disparity regression/classification (schemes (e-f-g) in Fig. 2). Epinet [48] is a fully end-to-end convolutional network for depth estimation. With Epinet's multi-stream architecture, features are extracted in local patches of SAIs horizontally, vertically, and diagonally. In the next stage, the features from several streams are merged by passing them through eight convolutional blocks, and a per-pixel depth value is finally obtained using a regression block. As with stereo matching, Epinet attempts to find correspondence between sub-aperture views, but the goal is to achieve a latent representation instead of a coarse depth estimation and use the features to predict disparity. Leistner et al. [49] deployed a Siamese neural network for feature extraction from vertical and horizontal SAI stacks. A U-Net architecture then merges the information and generates classification and regression outputs. Based on the classification label, a coarse subpixel disparity value is derived, and it is refined using an offset derived from the regression output. To train the network for wide baseline light fields, the authors applied a virtual shift to the SAI stacks to generate views with different disparity ranges. A two-stream CNN architecture is proposed in [53] that receives horizontal and vertical EPIs and performs a multi-scale feature extraction in four convolutional stages. After concatenating the feature sets from the two streams, the final disparity value is calculated using multi-label regression. Zhu et al. [54] employed a hybrid approach that combines focal stacks, center view, and EPIs to extract rich feature sets. A pixel-wise disparity classification is then performed by combining the features from the three pass-ways and feeding them into fully connected and softmax layers.

To better estimate the displacements when computing the multi-view feature maps, several methods deploy the concept of cost volume [55] where shifts are applied to the input views and the features are merged into a cost volume (Fig. 2g). Tsai et al. [50] proposed to manually shift SAIs at different disparity levels and pass them through residual blocks and a spatial pyramid pooling module for feature extraction. The resulting features are concatenated into a cost volume. During the learning process, an attention map is used in conjunction with the cost volume to determine the importance of different views, and finally, the pixel disparity is determined by solving a regression equation. A multi-scale feature extraction method is proposed in [56] that uses a cost volume with a low memory footprint. SAI feature maps are shifted according to the center view for several disparity levels and then concatenated to construct a 4D cost volume.

The three major techniques for depth estimation, summarized in Fig. 2, come with some drawbacks. Auto-encoder-based methods are generally early-stage works in the domain with significant artifacts, especially in occluded regions, and they are not competitive with state-of-the-art approaches. Moreover, as EPIs are 2D slices in both spatial–angular dimensions, they are only limited to specific horizontal and vertical coordinates for depth estimation, so they do not use the full 4D light field information for accurate depth estimation. The application is also limited in wide baseline and more complex scenes, where finding the relation between the slope of the EPI line and depth is more challenging.

Methods based on coarse depth estimation and refinement can better exploit the 4D nature of the light field information by using SAIs at different dimensions and better handling wide-baseline scenarios. However, stereo-matching approaches based on optical flow use heavyweight models and are often vulnerable to errors on non-Lambertian surfaces and occluded regions.

End-to-end methods like Epinet use SAIs stacked at multiple directions and can achieve better performance than the other two strategies, but they are still computationally expensive and show degraded performance for wide-baseline scenarios. Moreover, the performance is highly bound to the training set. Therefore, depth estimation for the wide baseline scenario, with an acceptable trade-off between accuracy and computation, is still an open research problem.

3.2 Light field reconstruction

Light field acquisition is often limited by the underlying hardware constraints of cameras and the captured data can suffer from low spatial and/or angular resolution. Plenoptic cameras enable recording dense light fields though with lower spatial resolution and narrow baseline, while bulky camera rigs allow larger baseline and spatial resolution but with a sparse set of views. To enable higher spatial and angular resolutions, the development of light field reconstruction/super-resolution (SR) methods has gained significant attention. Based on the available literature on light field reconstruction, existing works on spatial SR, view synthesis (angular SR), and reconstruction (both spatial and angular SR) are discussed in separate subsections. Finally, the concept of neural scene representation is introduced and reviewed.

3.2.1 Spatial super-resolution

Two popular strategies were identified in the literature for spatial SR of sub-aperture views as depicted in Fig. 3a and b:

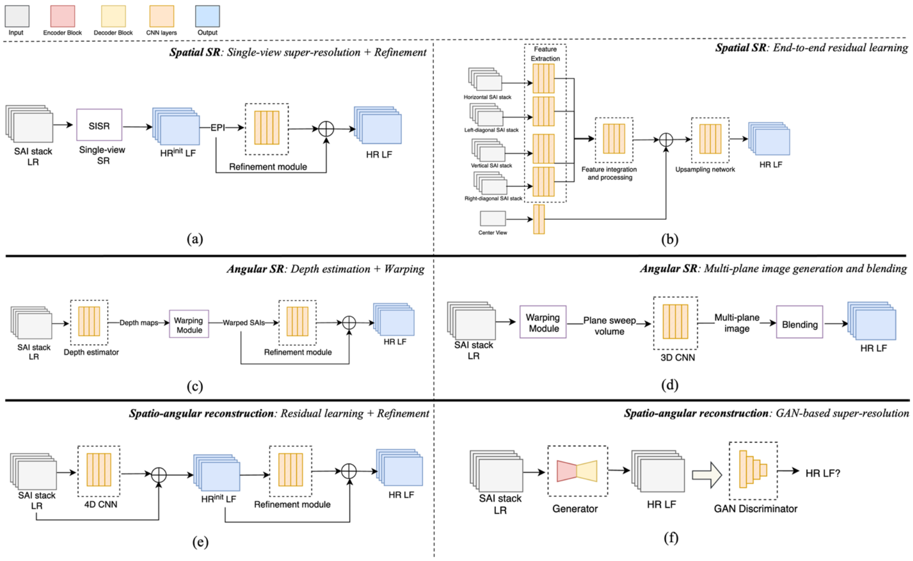


Fig. 3 Examples of architectures for the spatial, angular, and spatio-angular super-resolution (SR) frameworks. **a** Single-view SR using single image super-resolution (SISR) network and inter-view enhancement [57], **b** end-to-end residual learning [58], **c** warping and residual learning for refinement [59], **d** multi-plane image generation [60], **e** residual learning using 4D CNNs and refinement [61], **f** GAN-based method [62]

3.2.1.1 Single-view super-resolution and refinement Following the success of the learning-based single image SR (SISR) methods, these techniques were deployed to independently super-resolve SAIs of light fields. However, SISR cannot exploit the correlation between views, and thus, the inter-view consistency is not preserved. Therefore, the quality of the reconstructed views is often improved via an additional learning-based inter-view enhancement stage. In one of the early works, Fan et al. [63] developed a two-stage method that first super resolves each view separately using a SISR method (called VDSR) and then performs a patch-level view alignment using residual learning to compensate for the inter-view misalignments. Cheng et al. [64] employed a similar approach to super-resolve the SAIs individually and warp them to the center view. Afterward, the warped view stack was inserted into an enhancement residual block to obtain high-resolution (HR) light fields. Yuan et al. [57] used an EPI enhancement network to preserve the geometrical consistency of views. After the initial SR stage of SAIs, the EPIs are extracted and passed through feature extraction and reconstruction steps to achieve the final super-resolved light fields. Farrugia et al. [65] used sparse representation in which warping is applied to align the sub-aperture view with the center view, and a low-rank approximation was used to reduce the light field dimension. Finally, a super-resolved light field is recovered and further enhanced by inverse warping and inpainting methods.

3.2.1.2 End-to-end residual learning Inspired by the residual learning techniques deployed in SISR algorithms, several methods have been developed to restore high-frequency information of light field views using end-to-end deep residual networks. The recovered information is finally combined with bicubic-upsampled light fields to achieve an HR light field. A residual network is proposed in [58] that exploits stacks of light field views in four angular directions for feature extraction and residual information learning. Yeung et al. [66] deployed a residual scheme with spatial–angular separable (SAS) convolutions for more computationally efficient feature extraction. Jin et al. [67] proposed an all-to-one strategy in which per-view feature extraction is performed in the first step, and then the features are combined across views to construct an intermediate HR light field. Next, regularization is applied in the enhancement stage to improve the cross-view structural consistency. More recent work by Wang et al. [68] used the concept of deformable convolution to compute residual information for light field SR.

Apart from the two main schemes described above, there are several other innovative methods worth mentioning. LFNet [69] is an end-to-end deep learning approach for light field SR which replaced the conventional warping/registration approach for alignment with a multi-scale contextual information extraction scheme. The spatial relation between views is exploited by fusing the contextual data, and a bidirectional recurrent CNN was deployed to super-resolve horizontal and vertical image stacks. Finally, a stacked generalization is used to linearly combine the horizontal and vertical image stacks.

Some techniques adopt a hybrid capturing system where a single HR image is recorded using a standard 2D camera, and low-resolution (LR) light field views are recorded by a plenoptic camera. The high-frequency information provided by the HR image is used as a reference and propagated to the adjacent LR views for super-resolution. Zheng et al. [70, 71] deploy a patch matching scheme to find the corresponding patches between an HR reference and LR views and then adopt a view synthesis network to build an HR light field. In another hybrid approach, Jin et al. [72] proposed two parallel pipelines to make intermediate super-resolved light fields. One model is based on deep feature extraction from multiple views, and the other transforms the HR components to LR views using a warping network. The two intermediate HR light fields are subsequently combined using the learned attention maps to build the final super-resolved light field.

3.2.2 Angular super-resolution

Light field views captured using camera arrays are often sparsely sampled, and significant efforts have been made to synthesize the intermediate views between the captured SAIs to increase the angular resolution and obtain dense light fields. Real-world scenery often introduces occlusions, specularities, and non-Lambertian surfaces that challenge the view synthesis task. Three main image-based schemes have been devised in the literature for light field view synthesis including: (a) EPI super-resolution, (b) depth estimation and warping, and (c) multi-plane image rendering and blending. In addition to the three main approaches, we provide an introduction to neural scene representation, which has recently become a hot topic.

3.2.2.1 EPI super-resolution Angular SR can be done using similar approaches deployed for the spatial SR (as in Fig. 3a, b) while using EPIs instead of SAIs as inputs to the networks. Several view synthesis approaches follow a similar scheme of residual learning deployed in spatial SR in which EPIs were used as inputs to the networks for angular SR and synthesizing novel views. Wu et al. [73, 74] applied a blur kernel to extract low-frequency information of light fields and then performed a bicubic-upsampling on EPIs. Next, high-frequency information is restored using a residual network and an HR EPI is generated after an EPI deblurring step. Guo et al. [75] used five convolutional stages, each with three residual blocks to restore high-frequency information on EPIs. The work in [76] first transformed the EPI to the shearlet domain and then used an encoder–decoder generative adversarial network (GAN) to estimate the residual information and reconstruct the shearlet coefficients. Finally, the coefficient is transformed back to the image domain. Wang et al. [77] deployed 3D convolutions operated on EPI stacks to apply an upsampling followed by a high-frequency detail reconstruction based on residual learning.

The method in [78] manually shears the EPIs for certain shift values and upsamples them. By evaluating the similarity between the sheared EPI and the original EPI, an evaluation CNN determines whether the EPI was correctly sheared. The evaluation CNN is composed of an encoder–decoder structure that delivers a similarity score map (instead of the disparity) for fusing a set of sheared EPIs. Finally, a pyramid decomposition-reconstruction framework was used to reconstruct the high-resolution EPI. Liu et al. [79] constructed EPI images in four (vertical, horizontal, and two diagonal) directions to better explore the rich LF information for angular reconstruction. The four EPIs were fed into a deep learning framework that includes a feature extraction step followed by residual learning and coarse reconstruction. Further enhancement of the reconstructed EPIs was carried out based on an additional residual network. Using a learned filter bank, Fang et al. [80] proposed a sparse regularization scheme to obtain an intermediate reconstructed light field that was further enhanced through an encoder-like deep architecture for inpainting EPIs on the occluded and non-Lambertian regions.

3.2.2.2 Depth estimation and warping Several methods break down the angular SR task into depth estimation and warping components, as shown in the flow diagram of Fig. 3c. An additional enhancement step is often added at the end of the pipeline for quality enhancement of the reconstruction. Kalantari et al. [81] used two sequential CNN architectures for disparity estimation and color prediction of warped views. Sparse light fields are subjected to feature extraction and a learning-based disparity estimation process. The disparity data are then used to warp images, which are then fed into the second CNN for dense light field reconstruction. Gao et al. [82] deployed a learning-based optical flow estimation method to compute the disparity between the sub-aperture views, and an intermediate dense reconstruction is obtained using warping. In the next step, EPIs are extracted and inserted into an interactive enhancement process that uses a mask-accelerated shearlet transform to correct the errors on the EPIs.

The authors in [59] proposed depth estimation and blending networks for view synthesis in which novel views are initially generated through a warping module and by using the estimated depth map. In the next step, the blending module further refines the novel views by exploiting the spatio-temporal relationship using a residual network. Shi et al. [83] developed an end-to-end learning architecture that first computes disparity using an optical flow estimation network and then uses the disparity-guided warped images to obtain two reconstructions in pixel and feature domains. The feature-domain reconstruction uses a VGG network for feature extraction and a multi-scale feature warping to construct the views. Finally, the two reconstructions are combined to achieve the final reconstructed views. Meng et al. [84] adopted two dense neural networks for disparity estimation and warping. A disparity map is first obtained by computing the pixel shift between two corner views. This intermediate disparity map and the input views were then inserted into a second residual network to estimate a confidence map for warping. Finally, a refinement network based on alternating spatial–angular convolutional filters was adopted that exploits the 4D light field and cross-view dependencies. Liu et al. [85] used a similar depth estimation, warping, and refinement strategy but proposed a new loss function that computes the error of refocused views instead of SAI or EPI. The image quality is optimized in the refocused domain by minimizing the loss in the pixel and frequency domains.

3.2.2.3 Multi-plane image generation In addition to the previously reviewed articles that use EPI and SAIs for view synthesis, an innovative scheme has recently been developed based on rendering LFs in the form of multiplane images (MPI) (Fig. 3d). Sampled light field views are first projected to several depth planes using warping to form plane sweep volumes (PSVs). Next, a 3D CNN network is adopted to convert PSVs to MPI, and each layer is characterized by color, depth, and transparency information. The novel views are reconstructed by blending renders from multiple nearby MPIs [60, 86, 87].

3.2.3 Spatio-angular reconstruction

Multiple methods have been developed to deliver a joint spatial and angular super-resolution for an input light field. Early learning-based solutions for spatio-angular light field reconstruction are based on sparse representation and compressive sensing theory. In 2013, Marwah et al. [88] developed an over-complete dictionary for sparse representation of the light fields and their reconstruction. The idea is that a light field can be modeled using a set of atoms due to information redundancy. Later on, Farrugia et al. [89] deployed dictionaries of LR and HR patches in which the patch volumes were projected to sparse subspaces with the lower dimension using principle component analysis. As a final step, a linear mapping function is learned from LR to HR subspace.

Deep learning methods have become dominant for the simultaneous spatial and angular reconstruction of light fields in recent years. The first deep learning methods often deployed separate architectures for spatial and angular SR. In [90, 91], a deep learning-based approach is used to first increase the spatial resolution of light field views, which was followed by a second CNN for novel view synthesis. Gul et al. [92] deployed a reconstruction scheme with two successive CNNs to perform an angular followed by a spatial SR on stacks of lenslet images. However, using such a pixel-level reconstruction strategy can lead to jagged and lattice artifacts near the edges, which can significantly affect the quality of the results.

Gupta et al. [93] deployed an autoencoder and 4D CNNs in two branches to reconstruct light fields. The autoencoder branch consists of a series of fully connected layers followed by a 4D convolution to compress the information and exploit the parallax. The second branch has five 4D CNN blocks. The reconstructions from the two branches are combined to obtain the final spatial–angular super-resolved light field. The use of compressed sensing for light field reconstruction is further expanded using deep learning; In [94], 4D tensors were generated from patches of light field views, and a deep learning scheme embedding 3D convolutions was adopted to build a sparse representation of light fields. Finally, reconstruction is performed by passing the features through a set of fully connected layers and reshaping them into the 4D tensor.

Modern deep learning architectures aim to exploit rich features from light fields for an integrated spatial and angular SR [95–97] (see Fig. 3e). A U-shaped encoder–decoder reconstruction scheme is proposed in [98] that embeds convolutional long short-term memory (LSTM) for simultaneous spatial and angular SR. Meng et al. [61] used 4D CNN layers to exploit both the spatial characteristics as well as the cross-view relationships. The method consists of an intermediate 4D light field reconstruction step based on residual learning followed by an enhancement step to refine the spatial information. Moreover, a perceptual loss function based on VGG19 network features was added to the angular loss function for better reconstruction. The authors further extended their work in [99] by using GANs for light field reconstruction (Fig. 3f). A GAN architecture is proposed in [62] for light field reconstruction, wherein the generator module, EPIs are fed to an encoder–decoder network for angular SR and then go through a spatial upscaling block based on the residual block. The discriminator module aims to distinguish the reconstructed light field from the ground-truth ones by comparing pixels as well as the high-frequency information. Chandramouli et al. [100] deployed a generative model based on a variational autoencoder that encodes light fields to a latent space and learns to generate them from the latent code. The model includes an additional CNN to extract features from central views to be used as auxiliary information in the encoder.

The reconstruction techniques summarized in Fig. 3 have some limitations inherent to their design. Early reconstruction techniques, such as [81], are slow and trained based on a predefined sampling pattern of the views, which reduces their applicability to different scenarios. Techniques based on single-view super-resolution (Fig. 3a) suffer from geometrical inconsistency as each view is super-resolved independently without exploiting the information between the SAIs. Angular SR methods based on depth estimation and warping (Fig. 3c), such as [59], are more suitable for wide-baseline applications, but the quality of view synthesis heavily depends on the accuracy of the estimated depth information, which is naturally error-prone, inducing artifacts such as tearing and ghosting in occluded regions and depth errors in non-Lambertian surfaces. MPI-based methods (Fig. 3(d)), like [86], are memory-intensive and complex to train, taking several days on multiple GPUs. Moreover, the number of depth planes depends on the maximum disparity, which can increase the model size significantly with the depth budget. A major drawback of this method is that the MPI network may assign high opacity to the wrong layers, resulting in blurry output rendering. Other advanced techniques that use 4D CNNs (Fig. 3e) showed higher reconstruction quality but at higher computational costs due to the 4D convolution operation. GAN-based methods (Fig. 3f) also have potential issues, such as the need for large training datasets and the challenge of avoiding problems such as mode collapse.

3.2.4 Neural scene representation

Apart from the previous works that follow an image-based rendering scheme (i.e., using neighboring views for novel view synthesis), the use of neural networks for scene representation and rendering has gained significant attention in recent years [101]. A 3D representation of a scene may be generated by explicit methods such as meshes [102], voxels [103], or point clouds [104]. Alternatively, 3D scene representation networks utilize differentiable ray marching algorithms along with continuous functions that map coordinates to features. Henzler et al. [105] and Sitzmann et al. [103] used deep-learning techniques and volumetric ray-marching to synthesize novel views from a continuous differentiable density field. However, these approaches were computationally expensive as they required many samples to query the volume. Neural radiance field (NeRF) [10] is a recent technique with significant influence in the computer vision community, and new NeRF variants are constantly emerging in the literature [106–108].

Using NeRF, volumetric representations of scenes can be generated as the weights of a non-convolutional multi-layer perceptron (MLP) with few samples. The weights are obtained by training on images with the known poses. The network produces a volume density and a view-dependent emitted radiance by inputting 5D coordinates (spatial and viewing directions). Through the use of volume rendering techniques, the output colors and densities of a view can be simulated by querying 5D coordinates along camera rays. A coarse sampling MLP in NeRF learns to estimate the density at a particular spatial location to estimate a coarse shape for an object. The coarsely sampled data are then used in a second fine network to obtain a denser sampling along the viewing ray. Although NeRF has shown promising results, it still suffers from long training time and quality issues. As a result, methods have been proposed to improve the quality and speed, despite the fact that faster methods often compromise quality for speed.

Methods focused on improving NeRF speed often rely on using different encodings [109] or deploying representations where the continuous NeRF representation is obtained by interpolating values stored in variants of spatial discretizations, such as voxels [110], 4D tensors [111], octrees [112], etc. Fridovich et al. [110] used a sparse voxel grid to store opacity and spherical harmonic coefficients, which are then interpolated to model the full plenoptic function. The method is improved by removing empty voxels and using a coarse-to-fine optimization. TensoRF [111] leverages the power of 4D tensors to represent the radiance field of a scene. This representation factorizes the scene into a combination of several low-rank tensor components that are more compact and efficient. This also enables the use of powerful tensor decomposition methods for modeling radiance fields. Another representation called a sparse neural radiance grid (SNeRG), is proposed in [113]. It stores a trained NeRF model as a sparse 3D voxel grid data structure, where each SNeRG voxel contains opacity, diffuse color, and a learned feature vector that captures view-dependent effects. Yu et al. [112] achieved faster rendering by an octree representation where the continuous plenoptic function is pre-sampled on a sparse voxel-based octree volume. However, this representation has a higher memory footprint than the original NeRF and may not be suitable for large-scale scenes.

NeX [114] is a scene representation that models view-dependent effects by performing basis expansion on the pixel representation of a multiplane image (MPI). Unlike traditional MPI, which stores static *RGBα* values, NeX represents each color as a function of the viewing angle and approximates this function using a linear combination of learnable spherical basis functions. Attal et al. [115] proposed a memory-efficient representation that uses a ray-space embedding network to transform 4D ray-space coordinates into a latent space that can be interpolated. The embedding network enables a nonlinear, many-to-one mapping from different ray coordinates to shared latent features, resulting in better feature embedding and a reduced representation size. To handle sparse input, a spatial subdivision with a voxel grid of local light fields is deployed, which enhances quality at the cost of increased rendering time. KiloNeRF [116] speeds up NeRF rendering by shrinking a single large MLP into thousands of tiny MLPs, each representing parts of the scene. Each small network encodes features for a single cell on a regular grid that covers the scene. This allows one to take advantage of the spatial structure of the scene and skip computations for empty or occluded regions. A more recent approach is proposed in [117] that represents scenes using 3D Gaussians and shows promising real-time rendering at reasonable quality. This representation can maintain the characteristics of continuous volumetric radiance fields and skip the computation in empty space.

Müller et al. [109] proposed using a new encoding approach based on a hash table that allows a more compact network architecture. The model consists of a small neural network that maps the input features to a low-dimensional hash code and a hash table that stores a trainable feature vector for each hash bucket. The feature vectors are updated by stochastic gradient descent. The method can handle hash collisions by exploiting the multiresolution structure of the network, which allows it to learn different embeddings for similar inputs at different levels of granularity. This results in a simple and efficient architecture that can be easily parallelized on modern GPUs.

Several research papers have proposed quality improvements to the NeRF framework. Mip-NeRF [118] is a multi-scale extension of NeRF that uses conical frustums for anti-aliased rendering. Mip-NeRF can handle different resolutions of scene content more robustly than NeRF. Unlike NeRF, which uses rays to sample the scene and trains separate neural networks for each scale, Mip-NeRF uses cones to sample the scene and trains a single neural network that can model multiple scales. This way, Mip-NeRF can encode the position and size of each cone segment and render the scene at a variable scale. This reduces the aliasing and blurring artifacts that NeRF suffers from when the scene content varies in resolution across training or testing images.

The original NeRF framework uses 3D Cartesian coordinates to model scenes that represent all-angle captures of objects with transparent backgrounds. However, for front-facing scenes where all images have similar orientations, NeRF uses projective coordinates instead. When rendering scenes that are unbounded in all directions, a different parameterization is required. This idea was explored by NeRF++ [119], which used an additional network to model distant objects. The method divides the scene into two inner and outer volumes modeled using two NeRFs. The outer volume network learns a spherical representation of the scene that can handle large variations in depth and viewing angle. The outer network is combined with the inner volume network, which models the foreground objects, to produce high-quality novel views of complex scenes. Mip-NeRF 360 [120] is an extension of Mip-NeRF for synthesizing novel views of unbounded scenes. It leverages a nonlinear scene parameterization, online distillation, and a distortion-based regularizer to achieve high-quality results. The method consists of a NeRF MLP to predict color and opacity, and a second proposal MLP that predicts density and weights.

The latest NeRF methods, such as Mip-NeRF 360 [120], can produce high-quality visual results, but they require a long training time. On the other hand, methods that use explicit volumetric representations can achieve fast training, but they have some limitations. First, they cannot match the visual quality of the implicit NeRF methods, and second, they cannot use gradient-based optimization directly on their methods, because they rely on complex data structures or lack prior knowledge. Therefore, they still need to convert a trained implicit model (e.g., NeRF) to their final representation that supports real-time rendering. This conversion step adds to their training time and complexity. Thus, real-time high-quality rendering is still an open research question.

3.3 Compression

As mentioned earlier, the original 7D plenoptic function can be reduced to a 4D one [11], yet representing a huge amount of raw data compared to 2D images. Therefore, compression schemes are essential to enable efficient light field storage and transmission.

An efficient codec should be able to explore not only the spatial and angular redundancies independently (as two-dimensional data), but also the combined spatial–angular redundancy (4D data). As the light field imaging modality carries a huge amount of data, efficient compression algorithms are paramount, and light field coding has been extensively researched in the last few years [6, 7]. A wide range of light field compression schemes have been proposed, from well-known, off-the-shelf standard codecs, such as H.264 [121], HEVC [122], and VVC [123] to those specially designed for light field data [124–126]. These compression methods can be based on the 2D images (views) (a.k.a as the SAIs in the case of lenslet-based cameras), on microimages, on EPIs [6], or on other alternative representation [10]. The joint photographic experts group (JPEG) committee (ISO/IEC JTC1/SC29/WG1) developed a plenoptic coding standard [124, 127–130] triggered by this new technology. In [124], two independent coding modes are defined: one exploiting the redundancy using a 4D prediction process, the other exploiting the redundancy in 4D light field data by utilizing a 4D transform technique [125, 131].

Several works propose learning-based compression methods to code light field data, including light field videos. Three main strategies have been identified in the literature for learning-based light field compression: 1) Light field compression using learning-based view synthesis on the decoder side; and 2) Light field compression using learning-based view synthesis on the encoder and decoder sides, and 3) End-to-end light field compression architecture; The learning-based view synthesis methods used in the first two compression schemes closely follow the frameworks discussed in Section 3.2.2 (Angular SR). Figure 4 illustrates the generic frameworks of these three strategies.

3.3.1 Learning-based view synthesis on the decoder side

The key idea of this compression architecture is bitrate saving by sparsely encoding the views, with the remaining views being reconstructed from the encoded views on the decoder side. This type of strategy is highly dependent on the quality of available decoded views and on the learning-based reconstruction method. The method in [132] employs a sampling strategy to select key views to be encoded by the multiview extension of HEVC [133]. A disparity-based view warping method is used to synthesize the missing views at the decoder side, with the reconstruction accuracy improved by a CNN.

A sparse light field compression method is presented in [134]. An initial set of sparse views is encoded using the joint exploration model (JEM), while another set of sparse views is estimated using a linear approximation. Deep learning is used to reconstruct the remaining missing views. Another deep learning-based light field compression

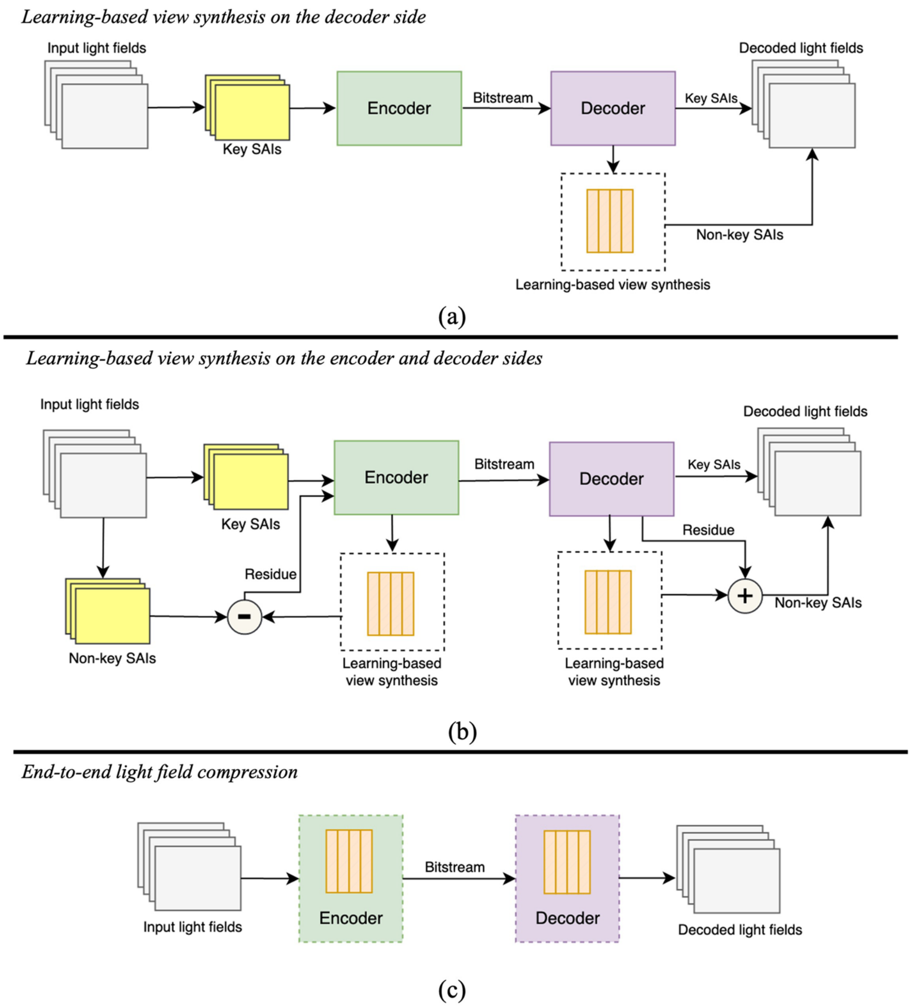


Fig. 4 Light field compression architectures using: **a** learning-based view synthesis on the decoder side; **b** learning-based view synthesis on the encoder and decoder sides, and **c** end-to-end scheme

method is proposed in [135] in which the missing views are reconstructed from the encoded neighboring views and a multi-view joint enhancement network is introduced to improve the coding performance. The light field compression scheme proposed in [136] super resolves the EPI via CNNs. Twenty-five percent of views of the light field are compressed on the encoder side, and these encoded views are used to reconstruct the entire light field by taking advantage of the special structure of EPI on the decoder end. As low-resolution EPIs are generated from the 25% selected views, they are super-resolved using a deep residual network. The high-resolution EPIs and the decoded views are used to reconstruct the whole light field. In [136], compression distortions suppression and super-resolution of the EPIs are performed by the same network, thus increasing the network's learning burden. In addition, compression distortions produced by the encoded view are hard to remove in the form of EPI. In [137], the learning burden is reduced by assigning quality enhancement and super-resolution to two different networks. The training data production process is also different from the previous scheme, taking the decoded sparse views as input images and the uncompressed views in the same position as their labels to train the decoder-side quality enhancement CNN.

3.3.2 Learning-based view synthesis on the encoder and decoder sides

Hou et al. [138] proposed a bi-level view compensation method to exploit angular redundancies, separating the light field data into key and non-key views. Aiming to exploit the 4D redundancy inherent to light field geometry, a CNN-based angular SR approach is used to synthesize super-resolved non-key views using the small number of key views. The residue between the key and non-key views is calculated and those super-resolved non-key views are arranged as a pseudo-sequence to be encoded by the HEVC video standard [122]. The efficient HEVC video encoder performs block-wise motion compensation to further exploit the inter-view (angular) correlation. The key views are also encoded by HEVC for the reconstruction of non-key views to obtain the whole light field. On the decoder side, both residual and key views are decoded and the CNN-based angular SR approach is applied to the decoded key view to be combined with the decoded residues to reconstruct the non-key views. This hybrid scheme uses a specific approach to explore the geometrical light field inter-view redundancy and the block-based motion estimation to further reduce redundancy is very efficient. A similar design is proposed in [139] using a GAN-based approach instead of a CNN. Another light field compression method using GAN-based view synthesis is proposed in [140]. Unlike the scheme in [139], it divides the generative model into disparity estimation and non-key view prediction components. Also, a perceptual quality-based loss function is introduced to supervise the GAN and preserve the quality of the synthesized views.

For the compression of densely-sampled light fields, Su et al. [141] used a CNN-based view synthesis as an initial prediction stage to exploit inter-view correlation. This compression solution employs the concept of super-rays, which is a grouping of rays within and across views taking into account disparity information. Super-rays are computed on synthesized residues, produced in the first prediction stage, to increase the correlation with the residues to be encoded. Then, neighboring super-rays are merged into a larger super-ray according to a rate-distortion cost. A 4D shape adaptive discrete cosine transform is applied per super-ray on the prediction residues in both the spatial and angular dimensions, followed by quantization and entropy coding on the transformed coefficients.

3.3.3 End-to-end light field compression architecture

Schemes of end-to-end learning for light field compression are gaining traction. In the recent [142], a dynamic adaptive light field video transmission scheme is proposed. The work introduces a description scheduling algorithm for unstable network conditions, which is capable of decoding the light field video with the highest possible quality even if partial data cannot be received completely and/or timely. A multiple description coding (MDC) based solution, employing the HEVC [122] video coding standard, was designed to transport the light field video compressed by a graph neural network (GNN) model. The proposed scheme separates the light field data with multi-level descriptions, with a GNN-based light field compression method as the basic encoder, enabling each description to be independently decoded aiming to increase robustness against eventual packet loss.

A hybrid coding scheme that combines a learning-based compression approach with a traditional video coding scheme is presented in [143]. An end-to-end trained compression scheme is used as a base layer in this hybrid compression solution to provide high gains at low/mid bitrates. To circumvent quality saturation at high bitrates, the proposed method uses standard-based HEVC coding in its enhancement layer.

An end-to-end spatial–angular-decorrelated network light field image compression is detailed in [144]. This method decouples the angular and spatial information by dilation convolution and stride convolution in spatial–angular interaction. Feature fusion is employed to jointly compress spatial and angular information.

The framework proposed in [145] uses neural representations to represent light fields aiming at efficiently randomly accessing any compressed view. As neural representations map positional information into color values, a multi-layer perceptron (MLP) is trained for each light field to map positional information to color information. The random access is facilitated by this direct mapping. The results show that the proposed method outperforms HEVC inter-coding in terms of compression efficiency. A more recent work [146] learns an MLP-based NeRF from the light field input views. The rate-distortion results show that the NeRF scene representation efficiently compresses light fields, outperforming coding standards and other learning-based methods.

No computational complexity is reported, nor subjective assessment results are provided by the methods reviewed in this subsection. The methods used densely angular sampled light fields, easing inter-view correlation.

Among the three techniques summarized in Fig. 4, end-to-end schemes are gaining more attention due to their effectiveness in image compression. This technique also enables computer vision tasks to be performed in the compressed domain without full decoding, offering greater flexibility and speed in machine applications. The performance of the first 2 techniques is highly dependent on the quality of the view synthesis task, which is defective under occlusion and when encoding non-Lambertian surfaces. The quality can be improved by sending the residues, as proposed in the second technique (Fig. 4b), but this does not significantly mitigate the problem, especially in the case of wide baselines and sparse light fields. Despite its exorbitant inference time, the NeRF technique proved its efficiency in compressing light field data. View synthesis drawbacks can be circumvented by neural representations that achieve a level of detail that is challenging for traditional methods.

3.4 Other light field imaging applications

Compared to conventional 2D images, light fields enable the acquisition of depth data, focus cues, and parallax, thereby improving accuracy for fundamental computer vision tasks. In addition to the applications described earlier in this section, learning-based light field solutions have been developed for other vision tasks such as saliency detection [147, 148], face recognition [149–151], image classification [152], low-light imaging [153] and light field microscopy [154–157]. Light field microscopy is able to capture 3D spatial information in a single camera frame, allowing almost instantaneous 3D imaging. This high-speed capability promoted the realistic applications of light field microscopy in biology and neurobiology providing the visualization of cardiovascular dynamics, the reduction of image reconstruction artifacts, and the recording of brain neuronal activity, for example [158].

Saliency object detection (SOD) aims to mimic the HVS to detect objects or areas that attract human attention. A comprehensive review of light field SOD methods is provided in [159] where the authors benchmark several learning-based models and compare them with RGB-D models of saliency detection. Moreover, the most notable architectures and existing light field datasets for SOD are extensively discussed and summarized. Encoder–decoder two-stream networks are commonly used for light field SOD that combine features from all-in-focused or center images in one stream with a focal stack or multi-view features in the other stream.

Face recognition using light fields can take advantage of both inter- and intra-view information for better accuracy. VGG network features were inserted into LSTM cell architectures in [150] to improve face recognition by exploiting spatio-angular information. A capsule network is developed in [151] that uses a pose matrix to capture viewpoint shifts and share knowledge across different object parts and locations. Moreover, the authors introduced two datasets of light field faces in the wild (LFFW) and light field face constrained (LFFC) for benchmarking the light field learning-based models.

light field microscopy makes use of deep learning models to improve the speed and quality of the image reconstructions, allowing the formation of volumetric dynamics in real-time. Encoder–decoder networks can be trained to transform raw light field microscopy data into 3D image stacks. Wang et al. [155] deployed a view-channel-depth (VCD) network with U-net architecture to convert the 2D light field views into 3D depth data. Wagner et al. [156] designed a deep architecture with 2D and 3D residual blocks for 3D reconstruction of light field microscopy data. A convolutional sparse coding (CSC) model is deployed in [157] for fast 3D localization of neurons in tissues from light fields. The network uses EPIs as inputs and generates sparse codes, representing depth data, as outputs.

4 Datasets and quality assessment

Light field datasets with diverse content characteristics are essential for the development and benchmarking of light field processing systems. Real-world light field data can be captured by plenoptic cameras such as lenslets, by an array of single-lens or plenoptic cameras, or using a single-lens moving camera, capturing the scene from different viewpoints. Synthetic light field data can be generated using computer simulation and view rendering. Content characteristics such as angular resolution and view sparsity, spatial resolution, scene complexity, specularity, and transparency can challenge light field imaging algorithms. Shafiee et al. [160] provided a detailed comparison of 33 light field datasets ranging from content-only datasets to specific task-based and quality assessment datasets. Although many datasets have been introduced in the literature, only a few of them have been widely adopted by the research community to evaluate light field imaging algorithms. Table 1 summarizes all the public datasets with their characteristics that have been used in the literature for benchmarking the learning-based compression, depth estimation, and reconstruction algorithms.

Table 1 Characteristics of the light field datasets used to benchmark the light field imaging systems

Dataset	Capturing device	No. scenes (u×v×s×t)	Year
Kalantari et al. [161]	Lytro lenslet camera	100 (625×434×15×15)	2016
EPFL [162]	Lytro lenslet camera	118	2016
4D light field (HCI) [163]	Synthetic	24 (512×512×9×9)	2016
Fraunhofer IIS [164]	Sony Alpha 7RII camera array	9 (7952×5304×21×99) and (7952×5304×21×101)	2017
Stanford Multiview [165]	Single Lytro lenslet camera 3 Lytro lenslet cameras	4211 3042	2019
INRIA [45]	Synthetic	53 Sparse and 39 Dense (512×512×9×9)	2019
WLF [56]	Synthetic	381 (1920×1080×9×9) and (512×512×9×9)	2021
Spaces [86]	GoPro Hero4	100 (5-10 camera rig positions, spatial res. 2k and 800×480)	2019
Real Forward-Facing [10]	Smartphone captures	8 (spatial res. 1008 × 756) 20-62 views	2021
Shiny [114]	Smartphone captures	8 (spatial res. 1008 × 567) 35-307 views	2021
NeRF 360 [120]	Smartphone captures	9 indoor/outdoor scenes 125-311 views	2022

The first public datasets in the domain such as Kalantari et al. [161] and EPFL [162] consist of realistic light fields captured using plenoptic Lytro Illum cameras. The 4D light field also known as HCI [163] includes 24 densely sampled synthetic light fields with accurate disparity information. In the Fraunhofer IIS dataset [164], Sony Alpha 7RII camera rigs were used to acquire density-sampled high-resolution light field images with wide baseline. The Stanford multiview light field dataset [165] consists of single Lytro shots as well as triple shots of three Lytro Illum cameras mounted together. More than 7000 light fields are collected in this dataset covering challenges such as non-Lambertian surfaces, occlusions, and specularity. INRIA [45] and WLF [56] are other recent synthetic datasets deployed in assessing light field algorithms for depth estimation and reconstruction.

The aforementioned datasets contain light fields with regular grids of sampled views. The dataset named Spaces [86] contains 100 scenes captured by 16 cameras with an arbitrary camera rig structure and 10 cm spacing between the cameras. The Spaces dataset was used for view synthesis using the MPI approach discussed in section 3.2.2. Table 1 also summarizes widely used datasets (Real Forward Facing of NeRF [10], Shiny [114], and NeRF 360 [120]) that represent smartphone captures of scenes on an irregular grid and have been used for the purpose of reconstruction using neural scene representation. Other related datasets used for NeRF benchmarking include BlendedMVS [166], Synthetic-NSVF [167], Tanks &Temples [168], and DeepVoxels [103].

Light field imaging algorithms are typically evaluated using quantitative methods by comparing generated data to a ground-truth. Peak signal-to-noise ratio (PSNR) and Structural SIMilarity index (SSIM) are the most widely used objective metrics for evaluating imaging algorithms. However, these two metrics tend to be less accurate when dealing with light field data. A comprehensive benchmark of 24 objective metrics on three public light field datasets revealed that PSNR and SSIM are only ranked 13th and 14th, respectively, in terms of their consistency with human quality preferences [169]. Moreover, CNN-induced artifacts and GAN reconstruction errors are different from conventional artifacts, and the existing metrics perform poorly on these emerging artifacts. Therefore, more reliable objective metrics are required to benchmark the imaging

Table 2 Subjectively annotated light field datasets

Dataset	Stimulus		Visualization		Display		Distortion types
	Single	Double	Passive	Active	2D	3D	
Tian et al. (2020) [172]	✓			✓		✓	JPEG, Gaussian noise, Gaussian blur
SMART (2017) [173]		✓	✓		✓		JPEG, JPEG2000, HEVC, SSDC
MPI-Lightfield (2017) [174]		✓		✓		✓	3D-HEVC, Display crosstalk, Quantized depth map Interpolation (linear, nearest neighbor, optical flow)
SHU (2017) [175]		✓	✓		✓		JPEG, JPEG2000, Gaussian blur, white noise
VALID (2018) [176]		✓	✓	✓	✓		HEVC, VP9, and 3 light field codecs
Win5-LID (2018) [177]		✓		✓		✓	HEVC, JPEG2000, Interpolation (linear, nearest neighbor, and 2 CNN methods)
NBU-LF1.0 (2019) [178]		✓	✓	✓		✓	Interpolation (nearest neighbor, bicubic, CNN-based) disparity-based reconstruction, spatial super-resolution (VDSR)
LFDD (2020) [179]		✓	✓		✓		Gaussian and impulse noise, Pincushion distortion, JPEG, JPEG 2000, HEVC, VP9, AV1, H.264, BPG

algorithms. Due to the diversity of light field acquisition procedures, distortions, and rendering processes, light field quality assessment remains a challenging task. To this end, subjectively annotated light field datasets are required to be used as ground-truth for developing new metrics. A summary of publicly available light field datasets with subjective scores can be found in Table 2. These datasets are characterized by the stimulus type for the subjective experiment (double or single stimulus), light field visualization method (passive in the form of pseudo videos or active), display type used in the subjective experiment, and the distortion types available in the dataset.

Given the multi-dimensional nature of light field data and the presence of distortions in both spatial and angular dimensions, a recent focus has been on developing more accurate objective algorithms that extract features from both spatial and angular domains for light field quality assessment. Metrics can be classified into three categories based on the availability of the reference image: full-reference (FR), reduced-reference (RR), and no-reference (NR). Table 3 summarizes the proposed metrics with their specifications. Learning-based approaches—especially for developing NR metrics—are gaining more attention due to their success in improving accuracy, and more deep-learning-based approaches have emerged recently. Although significant work has been done in this domain and more reliable metrics have been proposed, current learning-based light field algorithms are still only evaluated using conventional PSNR and SSIM methods. Therefore, a shift toward using more recent metrics is desirable. In this context, the IEEE established a new standard called 'IEEE P3333.1.4' [170], which defines metrics and provides recommended practices for light field quality assessment. A standardization activity, namely 'JPEG Pleno Quality Assessment' [171], was recently initiated within the JPEG committee aiming to explore the most promising subjective quality assessment practices as well as the objective methodologies for plenoptic modalities in the context of multiple use cases. The first phase of this effort will address the light field modality.

Table 3 Summary of the objective quality assessment methods for light fields

Authors	Type	Feature extraction	Learning method
Shi et al. [180], 2019	NR	Features based on cyclopean image naturalness, Gradient direction distribution, and weighted local binary pattern (LBP).	Support vector regression (SVR)
Tian et al. [181], 2020	FR	Using radial symmetry transform and Depth features	N/A
Tian et al. [182], 2020	FR	Multi-scale log-Gabor features	N/A
Min et al. [183], 2020	FR	Global and local spatial quality based on view structure matching, and near-edge MSE, multi-view quality for angular distortions	N/A
Meng et al. [184], 2020	FR	Texture features using Difference of Gaussian (DoG), and SSIM on refocused images for angular quality assessment	N/A
Zhou et al. [185], 2020	NR	Regarding LFs as 4D tensors to exploit global naturalness and local frequency properties	SVR
Liu et al. [186], 2021	NR	Making pseudo-reference quality assessment. Singular value decomposition (SVD) and DCT applied to micropixel blocks	SVR
Xiang et al. [187], 2021	NR	4D DCT features	PCA and SVR
Qu et al. [188], 2021	NR	Using separable convolutions to extract spatial and angular features	Deep auxiliary learning
Zhao et al. [189], 2021	NR	Three deep residual blocks for feature extraction from EPI	Deep learning
Pan et al. [190], 2021	NR	Spatial feature extraction (sharpness, slice distribution) on tensor slices. Angular features extraction using SVD	SVR
Meng et al. [191], 2021	FR	Gradient and phase congruency of key refocused views + chrominance features. Visual saliency map for pooling	N/A
Huang et al. [192], 2022	NR	Contourlet transform for spatial, and 3D-Gabor filter for geometry feature extraction	N/A
Alamgeer et al. [193], 2022	NR	Two-stream CNN and Atrous blocks for feature extraction	Deep learning
Alamgeer et al. [194], 2022	NR	Two-stream Long Short-Term Memory (LSTM) blocks for feature extraction from EPI and micro-lens images	Deep learning (LSTM)
Alamgeer et al. [195], 2022	NR	Use GANs to generate distortion maps and feature extraction using isometric mapping of the GAN-generated maps	GANs and random forest regressor
Zhang et al. [196], 2022	NR	Two-stream CNN feature extraction on spatio-angular patches	Deep learning
Zhang et al. [197], 2023	FR	Hierarchical discrepancy feature extraction using CNNs	Deep learning
Zhang et al. [198], 2023	NR	Two-stream CNNs for features extraction from Pseudo video blocks. Saliency- and variance-guided pooling	Deep learning
Ma et al. [199], 2023	NR	Spatial–angular feature extraction using self-attention	Deep learning (Swin Transformer)
Ma et al. [200], 2023	NR	Feature extraction using LBP, and NSS features on spatial-domain MSCN coefficient and in curvelet domain	SVR
Qu et al. [201], 2023	NR	Angular-wise attention using three new attention kernels: anglewise self-, anglewise grid-, and anglewise central-attention	Deep learning (Tranformers)

Table 3 (continued)

Authors	Type	Feature extraction	Learning method
Lamichhane [202], 2023	NR	Spatial feature extraction on saliency and cyclopean maps. Angular features using global luminance distribution and the weighted LBP on EPIs	SVR
Xiang et al. [203], 2023	RR	4D wavelet transform	SVR
Chai et al. [204], 2023	NR	SVD, LBP, 3D and 2D Log-Gabor for spatio-angular features	SVR

5 Discussion, challenges and perspectives

Research in light field imaging has been active in recent years, and machine-learning frameworks have been widely deployed in this domain to enhance various light field processing stages. While parallel to light fields, other plenoptic modalities like point cloud and holography have also been developed, there is no indication yet as to which will gain more dominance in the future. The three modalities offer unique advantages and they are interchangeable so for instance, one could obtain light fields from point clouds or vice versa. Even though point clouds and holographic content processing and compression have advanced significantly in recent years, these content types may eventually need to be converted to light field views for visualization on the display. This assumption of course depends on the types of displays to be introduced to the market in the future. Light fields provide more comprehensive information when it comes to capturing scenes. The Lidar can be used to create 3D point clouds by measuring precise distance to objects. However, light fields capture not only the 3D objects but also the entire scene information, which can be essential in many applications like autonomous driving, that require accurate 3D recreation of the vehicle surroundings.

The widely known plenoptic modeling methods proposed in [11] and [12] require the acquisition system (e.g., the camera grid) to be densely and regularly sampled, or the target viewing ray being a linear combination of the source views [205]. However, the number of required samples differs for scenes with different spatial complexity, occlusions, depth range, and non-Lambertian surfaces. Recent advances in using deep learning for spatio-angular reconstruction and the emergence of the NeRF-based approaches have demonstrated the potential of 3D rendering from a limited number of views, but they also face several challenges, as discussed in this paper. Some of these challenges are the long training times required for the models, the aliasing artifacts that occur at different resolutions, and the quality degradation that happens when dealing with scenes that are not bounded by a finite volume.

The reduction of the number of views might be efficient for use cases that do not demand real-time interaction, such as industrial visualization (e.g., prototype review) and digital signage (e.g., advertising). Faster radiance field methods, discussed in section , can achieve interactive rendering times on GPU depending on the complexity of the scene. However, they fall short of achieving true real-time rendering at high resolutions. More recent methods, such as 3D Gaussian splatting [117], show improvements in the quality–speed trade-off, whereas the quality achieved might not yet be optimal for applications that require high-fidelity 3D reconstructions. Enhancing the quality–speed trade-off could enable new use cases such as real-time telepresence and robotic tasks with fewer views for reconstruction.

The advent of different neural scene representations of light fields will bring discussions about which representation is most useful in the light field domain. In contrast to an explicit representation based on multiple SAIs (or micro-lens images), an implicit neural representation encodes light fields as parameters of an MLP. Therefore, the evolution of these two representations in the future may change the direction of research in light field processing, coding, and quality assessment. In light field compression, for example, image/video coding methods can be applied to encode the explicit representations, but implicit representation might require network data coding or a more compact neural representation with fewer parameters to achieve better storage and streaming performance [206].

Advances in deep learning frameworks are expected to significantly improve the performance of light field processing algorithms and solve the existing challenges. For example, the current depth estimation methods are not flawless and always come with the cost of artifacts, especially for occluded regions and in the presence of specularities. The imperfect depth data directly impacts the performance of the light field processing algorithms (such as view synthesis) that use depth information as an intermediate step. Therefore, better depth estimation or depth-free approaches are critical. Moreover, the success of the light field reconstruction algorithms is highly dependent on the number of available views and the baseline, and many algorithms used on narrow baseline will fail when deployed in wide baseline light fields.

When it comes to light field coding, advanced methods are needed to efficiently exploit the huge amount of redundant information about the light rays in the same scene that conveys angular and spatial information. End-to-end learning-based approaches have proved their effectiveness for 2D image coding. However, several challenges arise when coding 4D data including: (a) How to efficiently exploit the inherent 4D geometry using learning-based approaches? (b) What is the quality impact of the artifacts produced by learning-based compression methods which are different, in nature, from the ones produced by other light field codecs? and (c) How to deal with the increased complexity of the learning-based codecs that may result in impractical decoding runtimes. The JPEG committee has started the JPEG Pleno light field learning-based coding activity, targeting the creation of a learning-based coding standard to provide competitive compression efficiency compared to state-of-the-art light field coding solutions. To this end, two workshops have been promoted to discuss challenges and current solutions in learning-based coding solutions for light field data, to explore relevant use cases and requirements, and to provide a forum for researchers to discuss the latest findings in this area [207].

The evaluation of light field imaging systems has several shortfalls related to the content and assessment approaches available. The current datasets used in evaluations are often captured by Lytro cameras, which provide a very narrow baseline for comparison. Some synthetic or naturalistic light field datasets presented a slightly wider baseline, and there are very few wide baseline datasets that are used for evaluations in the literature. The development of advanced light field imaging systems for real-world applications requires additional datasets with diverse content properties. Among these properties are spatial resolution, scene complexity, wide color gamut, wide baselines and parallax, sparse and dense view sampling, specularity, and transparency of objects. It is also necessary for deep learning methods to be trained on large-scale datasets, however, the existing data are often small and deep frameworks are mostly trained using limited data. It is therefore essential to provide more comprehensive light field datasets from both the quantitative and content diversity perspectives.

The assessment of plenoptic image quality also faces various challenges because of the variety of quality aspects and complexity of the content when compared to the assessment of 2D images. In the context of the JPEG Pleno standardization process with regard to light field coding, a variety of subjective visual quality assessment procedures have been designed and significant knowledge has been built regarding challenges and good practices. For further improvement in this area, JPEG has begun developing a light field quality assessment standard, defining a framework with subjective quality assessment protocols and objective quality assessment procedures for lossy decoding of light field data within the context of multiple use cases and requirements outlined in [208]. The IEEE is also developing a standard called "P3333.1.4—Recommended Practice for the Quality Assessment of light field Imaging" that targets to establish methods of quality assessment of light field imaging based on psychophysical studies [209].

6 Conclusions

This paper provided an overview of the prominent learning-based paradigms for the most popular light field processing tasks. Depth estimation, compression, super-resolution and reconstruction are among the most important processing tasks while other vision tasks such as light field microscopy, saliency estimation, face recognition, refocusing and relighting have also been studied in the literature.

This review demonstrates the broad integration of learning-based frameworks for light field processing which is expected to be further expedited with the advances in light field capturing and visualization devices and the establishment of larger light field datasets. However, researchers still face many challenges, and there are many more to come. The advent of NeRF and its variants has paved the way for more innovative and efficient light field representation models requiring specific light field processing techniques and posing their own challenges. Despite the advances in developing super-resolution and view synthesis methods to improve the spatial and angular resolution of the light field, the existing content is still limited in FoV and DoF and far from offering a real 6 DoF experience where users can freely explore a scene from different viewports. Therefore, significant efforts are expected to increase in the coming years towards capturing larger datasets with wider baselines. Larger datasets will lead to further computational complexity and lower processing efficiency, bringing new challenges for learning-based solutions and deep learning models, in particular.

A number of deep learning frameworks have been developed for light field microscopy reconstruction, and 3D imaging using light field is expected to gain more attention in biology and neurobiology.

The existing image-based representation of light fields and implicit data-driven representations pose their own coding challenges, and further advancements in light field compression can broaden the application of these representations. Learning-based image compression solutions have gained momentum recently and this is expected to make a significant impact on the development of light field compression techniques.

Acknowledgements
Not applicable

Author contributions
All authors have made contributions to this manuscript. SM and CP: review and analysis, original draft writing. PS: review and substantively draft revising.

Funding
The research was funded by Research Foundation—Flanders (FWO) (G0B3521N).

Availability of data and materials
All data generated or analyzed during this study are included in this published article.

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 5 September 2023 Accepted: 19 May 2024
Published online: 30 May 2024

References

1. E.H. Adelson, J.R. Bergen, The plenoptic function and the elements of early vision. M. Landy, J. A. Movshon, (eds) *Computational Models of Visual Processing* (1991)
2. L. Liu, X. Sang, X. Yu, X. Gao, Y. Wang, X. Pei, X. Xie, B. Fu, H. Dong, B. Yan, 3d light-field display with an increased viewing angle and optimized viewpoint distribution based on a ladder compound lenticular lens unit. Opt. Express **29**(21), 34035–34050 (2021). https://doi.org/10.1364/OE.439805
3. E.H. Adelson, J.Y.A. Wang, Single lens stereo with a plenoptic camera. IEEE Trans. Pattern Anal. Mach. Intell. **14**(2), 99–106 (1992). https://doi.org/10.1109/34.121783
4. Y. Sawahata, Y. Miyashita, K. Komine, Estimating angular resolutions required in light-field broadcasting. IEEE Trans. Broadcast. **67**(2), 473–490 (2021). https://doi.org/10.1109/TBC.2020.3047218
5. G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, Y. Liu, Light field image processing: an overview. IEEE J. Select. Topics Signal Process. **11**(7), 926–954 (2017). https://doi.org/10.1109/JSTSP.2017.2747126
6. C. Conti, L.D. Soares, P. Nunes, Dense light field coding: a survey. IEEE Access **8**, 49244–49284 (2020). https://doi.org/10.1109/ACCESS.2020.2977767
7. C. Brites, J. Ascenso, F. Pereira, Lenslet light field image coding: classifying, reviewing and evaluating. IEEE Transactions on Circuits and Systems for Video Technology, 1–1 (2020)
8. R. Tao, W. Guo, T. Zhang, An overview on theory and algorithm of light field imaging technology. In: Y. Jiang, X. Ma, X. Li, M. Pu, X. Feng, B. Kippelen (eds.) 9th International Symposium on advanced optical manufacturing and testing technologies: optoelectronic materials and devices for sensing and imaging, vol. 10843, p. 108431. SPIE, China (2019). https://doi.org/10.1117/12.2514826. International Society for Optics and Photonics
9. A. Gershun, The light field. J. Math. Phys. **18**(1–4), 51–151 (1939). https://doi.org/10.1002/sapm193918151
10. B. Mildenhall, P.P. Srinivasan, M. Tancik, J.T. Barron, R. Ramamoorthi, R. Ng, NeRF: Representing scenes as neural radiance fields for view synthesis. cite arxiv:2003.08934 Comment: ECCV 2020 (oral). Project page with videos and code: http://tancik.com/nerf (2020)
11. M. Levoy, P. Hanrahan, Light field rendering. In Proceedings of the 23rd Annual Conference on computer graphics and interactive techniques, pp. 31–42. ACM, New York, NY, USA (1996)
12. S.J. Gortler, R. Grzeszczuk, R. Szeliski, M.F. Cohen, The lumigraph. In Proceedings of the 23rd Annual Conference on computer graphics and interactive techniques. SIGGRAPH '96, pp. 43–54. Association for computing machinery, New York, NY, USA (1996). https://doi.org/10.1145/237170.237200
13. D.G. Dansereau, 4D light field processing and its application to computer vision. PRISM (2003). https://doi.org/10.11575/PRISM/10182. https://prism.ucalgary.ca/handle/1880/42305
14. R.C. Bolles, H.H. Baker, D.H. Marimont, Epipolar-plane image analysis: an approach to determining structure from motion. Int. J. Comput. Vis. **1**(1), 7–55 (1987)
15. R. Hartley, A. Zisserman, *Multiple View Geo. Comput. Vis.*, 2nd edn. (Cambridge University Press, New York, NY, USA, 2003)
16. R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, P. Hanrahan, Light field photography with a hand-held plenoptic camera. Research report CSTR 2005-02, Stanford university (April 2005). https://hal.archives-ouvertes.fr/hal-02551481
17. Raytrix. http://www.raytrix.de/
18. Light Field Forum. http://lightfield-forum.com/
19. A. Davis, M. Levoy, F. Durand, Unstructured light fields. Comput. Graphics Forum (2012). https://doi.org/10.1111/j.1467-8659.2012.03009.x

20. A. Bajpayee, A.H. Techet, H. Singh, real-time light field processing for autonomous robotics. In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4218–4225 (2018). https://doi.org/10.1109/IROS.2018.8594477
21. P.A. Kara, A. Simon, The good news, the bad news, and the ugly truth: a review on the 3d interaction of light field displays. Multimodal technologies and interaction **7**(5) (2023). https://doi.org/10.3390/mti7050045
22. P. Paudyal, F. Battisti, P. Le Callet, J. Gutiérrez, M. Carli, Perceptual quality of light field images and impact of visualization techniques. IEEE Trans. Broad. **67**(2), 395–408 (2021). https://doi.org/10.1109/TBC.2020.3034445
23. S.C. Chan, H.Y. Shum, A spectral analysis for light field rendering. In Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101), vol. 2, pp. 25–282 (2000). https://doi.org/10.1109/ICIP.2000.899215
24. Z. Lin, H.-Y. Shum, On the number of samples needed in light field rendering with constant-depth assumption. In Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662), vol. 1, pp. 588–5951 (2000). https://doi.org/10.1109/CVPR.2000.855873
25. Z. Lin, H.-Y. Shum, H. Shum, A geometric analysis of light field rendering. Int. J. Comput. Vis. **58**, 121 (2004)
26. X. Yu, R. Wang, J. Yu, Real-time depth of field rendering via dynamic light field generation and filtering. Computer Graphics Forum **29**(7), 2099–2107 (2010). https://doi.org/10.1111/j.1467-8659.2010.01797.x. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2010.01797.x
27. Y.J. Jeong, Light-field rendering in the view interpolation region without dense light-field reconstruction. J. Phys. Conf. Series **1098**(1), 012015 (2018). https://doi.org/10.1088/1742-6596/1098/1/012015
28. R. Rideaux, A.E. Welchman, Proscription supports robust perceptual integration by suppression in human visual cortex. Nat. Commun. (2018). https://doi.org/10.1038/s41467-018-03400-y
29. H. Hiura, K. Komine, J. Arai, T. Mishina, Measurement of static convergence and accommodation responses to images of integral photography and binocular stereoscopy. Opt. Express **25**(4), 3454–3468 (2017). https://doi.org/10.1364/OE.25.003454
30. P. Kovács, R. Bregovic, A. Boev, A. Barsi, A. Gotchev, Quantifying spatial and angular resolution of light field 3d displays. IEEE J. Selected Topics Signal Process. (2017). https://doi.org/10.1109/JSTSP.2017.2738606
31. A. Cserkaszky, P.A. Kara, R.R. Tamboli, A. Barsi, M.G. Martini, T. Balogh, Light-field capture and display systems: limitations, challenges, and potentials. In Optical Engineering + Applications (2018)
32. X. Gao, X. Sang, S. Xing, X. Yu, B. Yan, B. Liu, P. Wang, Full-parallax 3D light field display with uniform view density along the horizontal and vertical direction. Optics Commun. **467**, 125765 (2020). https://doi.org/10.1016/j.optcom.2020.125765
33. S. Shen, S. Xing, X. Sang, B. Yan, Y. Chen, Virtual stereo content rendering technology review for light-field display. Displays (2022). https://doi.org/10.1016/j.displa.2022.102320
34. M. Poggi, F. Tosi, K. Batsos, P. Mordohai, S. Mattoccia, On the synergies between machine learning and binocular stereo for depth estimation from images: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **44**(9), 5314–5334 (2022). https://doi.org/10.1109/TPAMI.2021.3070917
35. S. Wanner, B. Goldluecke, Globally consistent depth labeling of 4d light fields. In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 41–48 (2012). https://doi.org/10.1109/CVPR.2012.6247656
36. M. Diebold, B. Goldluecke, Epipolar plane image refocusing for improved depth estimation and occlusion handling. In M. Bronstein, J. Favre, K. Hormann (eds.) Vision, Modeling and Visualization. The Eurographics Association, Switzerland (2013). https://doi.org/10.2312/PE.VMV.VMV13.145-152
37. S. Wanner, B. Goldluecke, Variational light field analysis for disparity estimation and super-resolution. IEEE Trans. Pattern Anal. Mach. Intell. **36**(3), 606–619 (2014). https://doi.org/10.1109/TPAMI.2013.147
38. T.-C. Wang, A.A. Efros, R. Ramamoorthi, Occlusion-aware depth estimation using light-field cameras. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). ICCV '15, pp. 3487–3495. IEEE Computer Society, USA (2015). https://doi.org/10.1109/ICCV.2015.398
39. Y. Zhang, H. Lv, Y. Liu, H. Wang, X. Wang, Q. Huang, X. Xiang, Q. Dai, Light-field depth estimation via epipolar plane image analysis and locally linear embedding. IEEE Trans. Circuits Syst. Video Technol. **27**(4), 739–747 (2017). https://doi.org/10.1109/TCSVT.2016.2555778
40. J. Chen, J. Hou, Y. Ni, L.-P. Chau, Accurate light field depth estimation with superpixel regularization over partially occluded regions. IEEE Trans. Image Process. **27**(10), 4889–4900 (2018). https://doi.org/10.1109/TIP.2018.2839524
41. O. Johannsen, A. Sulc, B. Goldluecke, What sparse light field coding reveals about scene structure. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3262–3270 (2016)
42. A. Alperovich, O. Johannsen, M. Strecke, B. Goldluecke, Light field intrinsics with a deep encoder-decoder network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9145–9154 (2018)
43. S. Heber, W. Yu, T. Pock, U-shaped networks for shape from light field. In BMVC, vol. 3, p. 5 (2016)
44. S. Heber, W. Yu, T. Pock, Neural epi-volume networks for shape from light field. In Proceedings of the IEEE International Conference on Computer Vision, pp. 2252–2260 (2017)
45. J. Shi, X. Jiang, C. Guillemot, A framework for learning depth from a flexible subset of dense and sparse light field views. IEEE Trans. Image Process. **28**(12), 5867–5880 (2019)
46. X. Jiang, J. Shi, C. Guillemot, A learning based depth estimation framework for 4d densely and sparsely sampled light fields. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2257–2261 (2019). IEEE
47. C. Guo, J. Jin, J. Hou, J. Chen, Accurate light field depth estimation via an occlusion-aware network. In 2020 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6 (2020). IEEE
48. C. Shin, H.-G. Jeon, Y. Yoon, I.S. Kweon, S.J. Kim, Epinet: A fully-convolutional neural network using epipolar geometry for depth from light field images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4748–4757 (2018)
49. T. Leistner, H. Schilling, R. Mackowiak, S. Gumhold, C. Rother, Learning to think outside the box: Wide-baseline light field depth estimation with epi-shift. In 2019 International Conference on 3D Vision (3DV), pp. 249–257 (2019). IEEE

50. Y.-J. Tsai, Y.-L. Liu, M. Ouhyoung, Y.-Y. Chuang, Attention-based view selection networks for light-field disparity estimation. In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 12095–12103 (2020)
51. S. Heber, T. Pock, Convolutional networks for shape from light field. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3746–3754 (2016)
52. S. Rogge, I. Schiopu, A. Munteanu, Depth estimation for light-field images using stereo matching and convolutional neural networks. Sensors **20**(21), 6188 (2020)
53. M. Feng, Y. Wang, J. Liu, L. Zhang, H.F. Zaki, A. Mian, Benchmark data set and method for depth estimation from light field images. IEEE Trans. Image Process. **27**(7), 3586–3598 (2018)
54. W. Zhou, X. Wei, Y. Yan, W. Wang, L. Lin, A hybrid learning of multimodal cues for light field depth estimation. Digital Signal Process. **95**, 102585 (2019)
55. J. Zbontar, Y. LeCun et al., Stereo matching by training a convolutional neural network to compare image patches. J. Mach. Learn. Res. **17**(1), 2287–2318 (2016)
56. Y. Li, Q. Wang, L. Zhang, G. Lafruit, A lightweight depth estimation network for wide-baseline light fields. IEEE Trans. Image Process. **30**, 2288–2300 (2021)
57. Y. Yuan, Z. Cao, L. Su, Light-field image superresolution using a combined deep cnn based on epi. IEEE Signal Process. Lett. **25**(9), 1359–1363 (2018). https://doi.org/10.1109/LSP.2018.2856619
58. S. Zhang, Y. Lin, H. Sheng, Residual networks for light field image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11046–11055 (2019)
59. J. Jin, J. Hou, H. Yuan, S. Kwong, Learning light field angular super-resolution via a geometry-aware network. In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 11141–11148 (2020)
60. K.-E. Lin, Z. Xu, B. Mildenhall, P.P. Srinivasan, Y. Hold-Geoffroy, S. DiVerdi, Q. Sun, K. Sunkavalli, R. Ramamoorthi, Deep multi depth panoramas for view synthesis. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16, pp. 328–344 (2020). Springer
61. N. Meng, H.K.-H. So, X. Sun, E. Lam, High-dimensional dense residual convolutional neural network for light field reconstruction. IEEE transactions on pattern analysis and machine intelligence (2019)
62. M. Zhu, A. Alperovich, O. Johannsen, A. Sulc, B. Goldlücke, An epipolar volume autoencoder with adversarial loss for deep light field super-resolution. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops : CVPRW 2019 : Proceedings : 16-20 June 2019, Long Beach, California, pp. 1853–1861. IEEE, Piscataway, NJ (2019). https://doi.org/10.1109/CVPRW.2019.00236
63. H. Fan, D. Liu, Z. Xiong, F. Wu, Two-stage convolutional neural network for light field super-resolution. In 2017 IEEE International Conference on Image Processing (ICIP), pp. 1167–1171 (2017). https://doi.org/10.1109/ICIP.2017.8296465
64. Z. Cheng, Z. Xiong, D. Liu, Light field super-resolution by jointly exploiting internal and external similarities. IEEE Trans. Circuits Syst. Video Technol. **30**(8), 2604–2616 (2020). https://doi.org/10.1109/TCSVT.2019.2921660
65. R.A. Farrugia, C. Guillemot, Light field super-resolution using a low-rank prior and deep convolutional neural networks. IEEE Trans. Pattern Anal. Mach. Intell. **42**(5), 1162–1175 (2020). https://doi.org/10.1109/TPAMI.2019.2893666
66. H.W.F. Yeung, J. Hou, X. Chen, J. Chen, Z. Chen, Y.Y. Chung, Light field spatial super-resolution using deep efficient spatial-angular separable convolution. IEEE Trans. Image Process. **28**(5), 2319–2330 (2018)
67. J. Jin, J. Hou, J. Chen, S. Kwong, Light field spatial super-resolution via deep combinatorial geometry embedding and structural consistency regularization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2260–2269 (2020)
68. Y. Wang, J. Yang, L. Wang, X. Ying, T. Wu, W. An, Y. Guo, Light field image super-resolution using deformable convolution. IEEE Trans. Image Process. **30**, 1057–1071 (2020)
69. Y. Wang, F. Liu, K. Zhang, G. Hou, Z. Sun, T. Tan, Lfnet: A novel bidirectional recurrent convolutional neural network for light-field image super-resolution. IEEE Trans. Image Process. **27**(9), 4274–4286 (2018). https://doi.org/10.1109/TIP.2018.2834819
70. H. Zheng, M. Ji, L. Han, Z. Xu, H. Wang, Y. Liu, L. Fang, Learning cross-scale correspondence and patch-based synthesis for reference-based super-resolution. In BMVC, vol. 1, p. 2 (2017)
71. H. Zheng, M. Ji, H. Wang, Y. Liu, L. Fang, Crossnet: An end-to-end reference-based super resolution network using cross-scale warping. In Proceedings of the European Conference on Computer Vision (ECCV), pp. 88–104 (2018)
72. J. Jin, J. Hou, J. Chen, S. Kwong, J. Yu, Light field super-resolution via attention-guided fusion of hybrid lenses. In Proceedings of the 28th ACM International Conference on Multimedia, pp. 193–201 (2020)
73. G. Wu, M. Zhao, L. Wang, Q. Dai, T. Chai, Y. Liu, Light field reconstruction using deep convolutional network on epi. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6319–6327 (2017)
74. G. Wu, Y. Liu, L. Fang, Q. Dai, T. Chai, Light field reconstruction using convolutional network on epi and extended applications. IEEE Trans. Pattern Anal. Mach. Intell. **41**(7), 1681–1694 (2018)
75. M. Guo, H. Zhu, G. Zhou, Q. Wang, Dense light field reconstruction from sparse sampling using residual network. In Asian Conference on Computer Vision, pp. 50–65 (2018). Springer
76. Y. Gao, R. Bregović, A. Gotchev, Self-supervised light field reconstruction using shearlet transform and cycle consistency. IEEE Signal Process. Lett. **27**, 1425–1429 (2020)
77. Y. Wang, F. Liu, Z. Wang, G. Hou, Z. Sun, T. Tan, End-to-end view synthesis for light field imaging with pseudo 4dcnn. In Proceedings of the European Conference on Computer Vision (ECCV), pp. 333–348 (2018)
78. G. Wu, Y. Liu, Q. Dai, T. Chai, Learning sheared epi structure for light field reconstruction. IEEE Trans. Image Process. **28**(7), 3261–3273 (2019)
79. D. Liu, Y. Huang, Q. Wu, R. Ma, P. An, Multi-angular epipolar geometry based light field angular reconstruction network. IEEE Trans. Comput. Imaging **6**, 1507–1522 (2020)
80. L. Fang, W. Zhong, L. Ye, R. Li, Q. Zhang, Light field reconstruction with a hybrid sparse regularization-pseudo 4dcnn framework. IEEE Access **8**, 171009–171020 (2020)
81. N.K. Kalantari, T.-C. Wang, R. Ramamoorthi, Learning-based view synthesis for light field cameras. ACM Trans. Graphics (TOG) **35**(6), 1–10 (2016)

82. Y. Gao, R. Bregovic, A. Gotchev, R. Koch, Mast: Mask-accelerated shearlet transform for densely-sampled light field reconstruction. In 2019 IEEE International Conference on Multimedia and Expo (ICME), pp. 187–192 (2019). IEEE
83. J. Shi, X. Jiang, C. Guillemot, Learning fused pixel and feature-based view reconstructions for light fields. In= Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2555–2564 (2020)
84. N. Meng, K. Li, J. Liu, E.Y. Lam, Light field view synthesis via aperture disparity and warping confidence map. IEEE Trans. Image Process. **30**, 3908–3921 (2021)
85. C.-L. Liu, K.-T. Shih, J.-W. Huang, H.H. Chen, Light field synthesis by training deep network in the refocused image domain. IEEE Trans. Image Process. **29**, 6630–6640 (2020)
86. J. Flynn, M. Broxton, P. Debevec, M. DuVall, G. Fyffe, R. Overbeck, N. Snavely, R. Tucker, Deepview: View synthesis with learned gradient descent. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2367–2376 (2019)
87. B. Mildenhall, P.P. Srinivasan, R. Ortiz-Cayon, N.K. Kalantari, R. Ramamoorthi, R. Ng, A. Kar, Local light field fusion: practical view synthesis with prescriptive sampling guidelines. ACM Trans. Gr. (TOG) **38**(4), 1–14 (2019)
88. K. Marwah, G. Wetzstein, Y. Bando, R. Raskar, Compressive light field photography using overcomplete dictionaries and optimized projections. ACM Trans. Gr. (TOG) **32**(4), 1–12 (2013)
89. R.A. Farrugia, C. Galea, C. Guillemot, Super resolution of light field images using linear subspace projection of patch-volumes. IEEE J. Selected Topics Signal Process. **11**(7), 1058–1071 (2017)
90. Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, I. So Kweon, Learning a deep convolutional network for light-field image super-resolution. In Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 24–32 (2015)
91. Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, I.S. Kweon, Light-field image super-resolution using convolutional neural network. IEEE Signal Process. Lett. **24**(6), 848–852 (2017)
92. M.S.K. Gul, B.K. Gunturk, Spatial and angular resolution enhancement of light fields using convolutional neural networks. IEEE Trans. Image Process. **27**(5), 2146–2159 (2018)
93. M. Gupta, A. Jauhari, K. Kulkarni, S. Jayasuriya, A. Molnar, P. Turaga, Compressive light field reconstructions using deep learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 11–20 (2017)
94. L. Wei, Y. Wang, Y. Liu, Tensor-based light field compressed sensing and epipolar plane images reconstruction via deep learning. IEEE Access **8**, 134898–134910 (2020)
95. K. Ko, Y.J. Koh, S. Chang, C.-S. Kim, Light field super-resolution via adaptive feature remixing. IEEE Trans. Image Process. **30**, 4114–4128 (2021)
96. G. Wu, Y. Wang, Y. Liu, L. Fang, T. Chai, Spatial-angular attention network for light field reconstruction. IEEE Trans. Image Process. **30**, 8999–9013 (2021)
97. Y. Chen, S. Zhang, S. Chang, Y. Lin, Light field reconstruction using efficient pseudo 4d epipolar-aware structure. IEEE Trans. Comput. Imaging **8**, 397–410 (2022)
98. H. Zhu, M. Guo, H. Li, Q. Wang, A. Robles-Kelly, Revisiting spatio-angular trade-off in light field cameras and extended applications in super-resolution. IEEE Trans. Vis. Comput. Gr. **27**(6), 3019–3033 (2019)
99. N. Meng, Z. Ge, T. Zeng, E.Y. Lam, Lightgan: a deep generative model for light field reconstruction. IEEE Access **8**, 116052–116063 (2020)
100. P. Chandramouli, K.V. Gandikota, A. Gorlitz, A. Kolb, M. Moeller, A generative model for generic light field reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence (2020)
101. M. Suhail, C. Esteves, L. Sigal, A. Makadia, Light field neural rendering. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8269–8279 (2022)
102. J. Thies, M. Zollhöfer, M. Nießner, Deferred neural rendering: image synthesis using neural textures. ACM Trans. Gr. (TOG) **38**(4), 1–12 (2019)
103. V. Sitzmann, J. Thies, F. Heide, M. Nießner, G. Wetzstein, M. Zollhofer, Deepvoxels: Learning persistent 3d feature embeddings. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2437–2446 (2019)
104. M. Wu, Y. Wang, Q. Hu, J. Yu, Multi-view neural human rendering. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1682–1691 (2020)
105. P. Henzler, N.J. Mitra, T. Ritschel, Escaping plato's cave: 3d shape from adversarial rendering. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9984–9993 (2019)
106. R. Martin-Brualla, N. Radwan, M.S. Sajjadi, J.T. Barron, A. Dosovitskiy, D. Duckworth, Nerf in the wild: Neural radiance fields for unconstrained photo collections. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7210–7219 (2021)
107. A. Pumarola, E. Corona, G. Pons-Moll, F. Moreno-Noguer, D-nerf: Neural radiance fields for dynamic scenes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10318–10327 (2021)
108. K. Deng, A. Liu, J.-Y. Zhu, D. Ramanan, Depth-supervised nerf: Fewer views and faster training for free. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12882–12891 (2022)
109. T. Müller, A. Evans, C. Schied, A. Keller, Instant neural graphics primitives with a multiresolution hash encoding. ACM Trans. Gr. (ToG) **41**(4), 1–15 (2022)
110. S. Fridovich-Keil, A. Yu, M. Tancik, Q. Chen, B. Recht, A. Kanazawa, Plenoxels: Radiance fields without neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5501–5510 (2022)
111. A. Chen, Z. Xu, A. Geiger, J. Yu, H. Su, Tensorf: Tensorial radiance fields. In European Conference on Computer Vision, pp. 333–350 (2022). Springer
112. A. Yu, R. Li, M. Tancik, H. Li, R. Ng, A. Kanazawa, Plenoctrees for real-time rendering of neural radiance fields. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5752–5761 (2021)
113. P. Hedman, P.P. Srinivasan, B. Mildenhall, J.T. Barron, P. Debevec, Baking neural radiance fields for real-time view synthesis. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5875–5884 (2021)

114. S. Wizadwongsa, P. Phongthawee, J. Yenphraphai, S. Suwajanakorn, Nex: Real-time view synthesis with neural basis expansion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8534–8543 (2021)
115. B. Attal, J.-B. Huang, M. Zollhöfer, J. Kopf, C. Kim, Learning neural light fields with ray-space embedding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 19819–19829 (2022)
116. C. Reiser, S. Peng, Y. Liao, A. Geiger, Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 14335–14345 (2021)
117. B. Kerbl, G. Kopanas, T. Leimkühler, G. Drettakis, 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics **42**(4) (2023)
118. J.T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, P.P. Srinivasan, Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5855–5864 (2021)
119. K. Zhang, G. Riegler, N. Snavely, V. Koltun, Nerf++: Analyzing and improving neural radiance fields. arXiv preprint arXiv:2010.07492 (2020)
120. J.T. Barron, B. Mildenhall, D. Verbin, P.P. Srinivasan, P. Hedman, Mip-nerf 360: unbounded anti-aliased neural radiance fields. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5470–5479 (2022)
121. *ITU-T & ISO/IEC.*: Advanced Video Coding for Generic Audiovisual Services, Rec. ITU-T H.264 and ISO/IEC 14496-10 Information technology - Coding of audio-visual objects - Part 10: Advanced Video Coding (MPEG-4 AVC) (2014)
122. *ITU-T & ISO/IEC.*: High Efficiency Video Coding, Rec. ITU-T H.265 and ISO/IEC 23008-2 (2013)
123. Working Draft 4 of Versatile Video Coding. Doc. Joint Video Experts Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC 1/SC29/WG11 N18274, 13th Meeting, Marrakech, Morocco (2019)
124. Information technology – Plenoptic image coding system (JPEG Pleno) – part 2: light field coding. ISO/IEC 21794-2:2021 (2021). https://www.iso.org/standard/74532.html
125. G. De Oliveira Alves, M.B. De Carvalho, C.L. Pagliari, P.G. Freitas, I. Seidel, M.P. Pereira, C.F.S. Vieira, V. Testoni, F. Pereira, E.A.B. Da Silva, The JPEG pleno light field coding standard 4d-transform mode: how to design an efficient 4D-native codec. IEEE Access **8**, 170807–170829 (2020). https://doi.org/10.1109/ACCESS.2020.3024844
126. M.L. Pendu, C. Ozcinar, A. Smolic, Hierarchical Fourier Disparity Layer Transmission For Light Field Streaming. In 2020 IEEE International Conference on Image Processing (ICIP), pp. 2606–2610 (2020). https://doi.org/10.1109/ICIP40778.2020.9190719
127. Information technology – Plenoptic image coding system (JPEG Pleno) – part 1: framework. ISO/IEC 21794-1:2020 (2020). https://www.iso.org/standard/74531.html
128. Information technology – Plenoptic image coding system (JPEG Pleno) – part 2: Light field coding – amendment 1: profiles and levels for JPEG Pleno light field coding system. ISO/IEC 21794-2:2021/AMD 1:2021 (2021). https://www.iso.org/standard/80897.html
129. Information technology – plenoptic image coding system (JPEG Pleno) – part 3: conformance testing. ISO/IEC 21794-2:2021 (2021). https://www.iso.org/standard/74533.html
130. Information technology – plenoptic image coding system (JPEG Pleno) – part 4: reference software. ISO/IEC 21794-4:2022 (2022). https://www.iso.org/standard/74534.html
131. S. Foessel, J. Ascenso, L.A. Silva Cruz, T. Ebrahimi, P.-A. Lemieux, C. Pagliari, A.M.G. Pinheiro, J. Sneyers, F. Temmermanns, Jpeg status and progress report 2022. SMPTE Motion Imaging J. **131**(8), 111–119 (2022). https://doi.org/10.5594/JMI.2022.3190917
132. B. Wang, W. Xiang, E. Wang, Q. Peng, P. Gao, X. Wu, Learning-based high-efficiency compression framework for light field videos. Multimedia Tools Appl. **81**(6), 7527–7560 (2022). https://doi.org/10.1007/s11042-022-11955-8
133. G. Tech, Y. Chen, K. Müller, J.-R. Ohm, A. Vetro, Y.-K. Wang, Overview of the multiview and 3d extensions of high efficiency video coding. IEEE Trans. Circuits Syst. Video Technol. **26**(1), 35–49 (2016). https://doi.org/10.1109/TCSVT.2015.2477935
134. N. Bakir, W. Hamidouche, O. Déforges, K. Samrouth, M. Khalil, Light field image compression based on convolutional neural networks and linear approximation. In 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 1128–1132 (2018). https://doi.org/10.1109/ICIP.2018.8451597
135. Z. Zhao, S. Wang, C. Jia, X. Zhang, S. Ma, J. Yang, Light field image compression based on deep learning. In 2018 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6 (2018). https://doi.org/10.1109/ICME.2018.8486546
136. J. Zhao, P. An, X. Huang, L. Shan, R. Ma, Light Field Image Sparse Coding via CNN-Based EPI Super-Resolution. In 2018 IEEE Visual Communications and Image Processing (VCIP), pp. 1–4 (2018). https://doi.org/10.1109/VCIP.2018.8698714
137. J. Zhao, P. An, X. Huang, C. Yang, L. Shen, Light field image compression via CNN-based EPI super-resolution and decoder-side quality enhancement. IEEE Access **7**, 135982–135998 (2019). https://doi.org/10.1109/ACCESS.2019.2930644
138. J. Hou, J. Chen, L.-P. Chau, Light field image compression based on bi-level view compensation with rate-distortion optimization. IEEE Trans. Circuits Syst. Video Technol. **29**(2), 517–530 (2019). https://doi.org/10.1109/TCSVT.2018.2802943
139. C. Jia, X. Zhang, S. Wang, S. Wang, S. Ma, Light field image compression using generative adversarial network-based view synthesis. IEEE J. Emerg. Selected Topics Circuits Syst. **9**(1), 177–189 (2019). https://doi.org/10.1109/JETCAS.2018.2886642
140. D. Liu, X. Huang, W. Zhan, L. Ai, X. Zheng, S. Cheng, View synthesis-based light field image compression using a generative adversarial network. Inf. Sci. **545**, 118–131 (2021). https://doi.org/10.1016/j.ins.2020.07.073
141. X. Su, M. Rizkallah, T. Mauzev, C. Guillemot, Rate-distortion optimized super-ray merging for light field compression. In 2018 26th European Signal Processing Conference (EUSIPCO), pp. 1850–1854 (2018). https://doi.org/10.23919/EUSIPCO.2018.8553485

142. X. Hu, Y. Pan, Y. Wang, L. Zhang, S. Shirmohammadi, Multiple description coding for best-effort delivery of light field video using gnn-based compression. IEEE Transactions on Multimedia, 1–1 (2021) https://doi.org/10.1109/TMM.2021.3129918
143. M. Stepanov, G. Valenzise, F. Dufaux, Hybrid learning-based and hevc-based coding of light fields. In 2020 IEEE International Conference on Image Processing (ICIP), pp. 3344–3348 (2020). https://doi.org/10.1109/ICIP40778.2020.9190971
144. K. Tong, X. Jin, C. Wang, F. Jiang, Sadn: Learned light field image compression with spatial-angular decorrelation. In ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1870–1874 (2022). https://doi.org/10.1109/ICASSP43922.2022.9747377
145. H. Amirpour, C. Guillemot, C. Timmerer, FuRA: Fully Random Access Light Field Image Compression. In European Workshop on Video and Image Processing. Proceedings EUVIP Conf., pp. 1–6, Lisbon, Portugal (2022). https://hal.science/hal-03758898
146. J. Shi, C. Guillemot, Light Field Compression via Compact Neural Scene Representation. In ICASSP 2023 - IEEE International Conference on Acoustics, Speech, and Signal Processing, Rhodes Island, Greece, pp. 1–5 (2023). https://inria.hal.science/hal-04017645
147. M. Zhang, W. Ji, Y. Piao, J. Li, Y. Zhang, S. Xu, H. Lu, Lfnet: light field fusion network for salient object detection. IEEE Trans. Image Process. **29**, 6276–6287 (2020)
148. Y. Piao, Z. Rong, M. Zhang, X. Li, H. Lu, Deep light-field-driven saliency detection from a single view. In IJCAI, pp. 904–911 (2019)
149. A. Sepas-Moghaddam, M.A. Haque, P.L. Correia, K. Nasrollahi, T.B. Moeslund, F. Pereira, A double-deep spatio-angular learning framework for light field-based face recognition. IEEE Trans. Circuits Syst. Video Technol. **30**(12), 4496–4512 (2019)
150. A. Sepas-Moghaddam, A. Etemad, F. Pereira, P.L. Correia, Long short-term memory with gate and state level fusion for light field-based face recognition. IEEE Trans. Inf. Forens. Sec. **16**, 1365–1379 (2020)
151. A. Sepas-Moghaddam, A. Etemad, F. Pereira, P.L. Correia, Capsfield: light field-based face and expression recognition in the wild using capsule routing. IEEE Trans. Image Process. **30**, 2627–2642 (2021)
152. Z. Lu, H.W. Yeung, Q. Qu, Y.Y. Chung, X. Chen, Z. Chen, Improved image classification with 4d light-field and interleaved convolutional neural network. Multimed. Tools Appl. **78**(20), 29211–29227 (2019)
153. M. Lamba, K.K. Rachavarapu, K. Mitra, Harnessing multi-view perspective of light fields for low-light imaging. IEEE Trans. Image Process. **30**, 1501–1513 (2020)
154. K. Wang, Deep-learning-enhanced light-field microscopy. Nat. Methods **18**(5), 459–460 (2021)
155. Z. Wang, L. Zhu, H. Zhang, G. Li, C. Yi, Y. Li, Y. Yang, Y. Ding, M. Zhen, S. Gao et al., Real-time volumetric reconstruction of biological dynamics with light-field microscopy and deep learning. Nat. Methods **18**(5), 551–556 (2021)
156. N. Wagner, F. Beuttenmueller, N. Norlin, J. Gierten, J.C. Boffi, J. Wittbrodt, M. Weigert, L. Hufnagel, R. Prevedel, A. Kreshuk, Deep learning-enhanced light-field imaging with continuous validation. Nat. Methods **18**(5), 557–563 (2021)
157. P. Song, H.V. Jadan, C.L. Howe, P. Quicke, A.J. Foust, P.L. Dragotti, Model-inspired deep learning for light-field microscopy with application to neuron localization. In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8087–8091 (2021). IEEE
158. N. Wagner, F. Beuttenmueller, N. Norlin, J. Gierten, J. Wittbrodt, M. Weigert, L. Hufnagel, R. Prevedel, A. Kreshuk, Deep learning-enhanced light-field imaging with continuous validation. bioRxiv (2020) https://doi.org/10.1101/2020.07.30.228924. https://www.biorxiv.org/content/early/2020/07/31/2020.07.30.228924.full.pdf
159. K. Fu, Y. Jiang, G.-P. Ji, T. Zhou, Q. Zhao, D.-P. Fan, Light field salient object detection: A review and benchmark. Computational Visual Media, 1–26 (2022)
160. E. Shafiee, M.G. Martini, Datasets for the quality assessment of light field imaging: comparison and future directions. IEEE Access **11**, 15014–15029 (2023)
161. N.K. Kalantari, T.-C. Wang, R. Ramamoorthi, Learning-based view synthesis for light field cameras code. https://cseweb.ucsd.edu/~viscomp/projects/LF/papers/SIGASIA16/. [Online; accessed 10-August-2021] (2016)
162. EPFL Light Field Image Dataset. https://www.epfl.ch/labs/mmspg/downloads/epfl-light-field-image-dataset/ . [Online; accessed 10-August-2021]
163. HCI Light Field Dataset. https://lightfield-analysis.uni-konstanz.de/ . [Online; accessed 10-August-2021]
164. M. Ziegler, R. Veld, J. Keinert, F. Zilly, Acquisition system for dense lightfield of large scenes. In 2017 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), pp. 1–4 (2017). IEEE
165. Stanford Light Field Archives. http://lightfields.stanford.edu/ . [Online; accessed 10-August-2021]
166. Y. Yao, Z. Luo, S. Li, J. Zhang, Y. Ren, L. Zhou, T. Fang, L.: Quan,Blendedmvs: A large-scale dataset for generalized multi-view stereo networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1790–1799 (2020)
167. L. Liu, J. Gu, K. Zaw Lin, T.-S. Chua, C. Theobalt, Neural sparse voxel fields. Adv. Neural Inf. Process. Syst. **33**, 15651–15663 (2020)
168. A. Knapitsch, J. Park, Q.-Y. Zhou, V. Koltun, Tanks and temples: benchmarking large-scale scene reconstruction. ACM Trans. Gr. (ToG) **36**(4), 1–13 (2017)
169. S. Mahmoudpour, P. Schelkens, On the performance of objective quality metrics for lightfields. Signal Process. Image Commun. **93**, 116179 (2021)
170. M. Maria, Ieee standard on the quality assessment of light field imaging. In IEEE SA, pp. 20–55 (2022). IEEE
171. C. Perra, S. Mahmoudpour, C. Pagliari, Jpeg pleno light field: Current standard and future directions. In Optics, Photonics and Digital Technologies for Imaging Applications VII, vol. 12138, pp. 153–156 (2022). SPIE
172. R.R. Tamboli, B. Appina, S. Channappayya, S. Jana, Super-multiview content with high angular resolution: 3d quality assessment on horizontal-parallax lightfield display. Signal Process. Image Commun. **47**, 42–55 (2016)
173. P. Paudyal, F. Battisti, M. Sjostrom, R. Olsson, M. Carli, Towards the perceptual quality evaluation of compressed light field images. IEEE Trans. Broadcast. **63**(3), 507–522 (2017). https://doi.org/10.1109/tbc.2017.2704430

174. V. Kiran Adhikarla, M. Vinkler, D. Sumin, R.K. Mantiuk, K. Myszkowski, H.-P. Seidel, P. Didyk, Towards a quality metric for dense light fields. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
175. L. Shan, P. An, D. Liu, R. Ma, Subjective evaluation of light field images for quality assessment database. In Digital TV and Wireless Multimedia Communication: 14th International Forum, IFTC 2017, Shanghai, China, November 8-9, 2017, Revised Selected Papers 14, pp. 267–276 (2018). Springer
176. I. Viola, T. Ebrahimi, VALID: Visual Quality Assessment for Light Field Images Dataset. In Tenth International Conference on Quality of Multimedia Experience (QoMEX). IEEE, Italy (2018). https://doi.org/10.1109/qomex.2018.8463388
177. L. Shi, S. Zhao, W. Zhou, Z. Chen, Perceptual evaluation of light field image. In 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 41–45 (2018). IEEE
178. Z. Huang, M. Yu, G. Jiang, K. Chen, Z. Peng, F. Chen, Reconstruction distortion oriented light field image dataset for visual communication. In 2019 International Symposium on Networks, Computers and Communications (ISNCC), pp. 1–5 (2019). IEEE
179. A. Zizien, K. Fliegel, Lfdd: Light field image dataset for performance evaluation of objective quality metrics. In Applications of Digital Image Processing XLIII, vol. 11510, pp. 671–683 (2020). SPIE
180. L. Shi, W. Zhou, Z. Chen, J. Zhang, No-reference light field image quality assessment based on spatial-angular measurement. IEEE Trans. Circuits Syst. Video Technol. **30**(11), 4114–4128 (2019)
181. Y. Tian, H. Zeng, J. Hou, J. Chen, J. Zhu, K.-K. Ma, A light field image quality assessment model based on symmetry and depth features. IEEE Trans. Circuits Syst. Video Technol. **31**(5), 2046–2050 (2020)
182. Y. Tian, H. Zeng, J. Hou, J. Chen, K.-K. Ma, Light field image quality assessment via the light field coherence. IEEE Trans. Image Process. **29**, 7945–7956 (2020)
183. X. Min, J. Zhou, G. Zhai, P. Le Callet, X. Yang, X. Guan, A metric for light field reconstruction, compression, and display quality evaluation. IEEE Trans. Image Process. **29**, 3790–3804 (2020)
184. C. Meng, P. An, X. Huang, C. Yang, D. Liu, Full reference light field image quality evaluation based on angular-spatial characteristic. IEEE Signal Process. Lett. **27**, 525–529 (2020)
185. W. Zhou, L. Shi, Z. Chen, J. Zhang, Tensor oriented no-reference light field image quality assessment. IEEE Trans. Image Process. **29**, 4070–4084 (2020)
186. Y. Liu, G. Jiang, Z. Jiang, Z. Pan, M. Yu, Y.-S. Ho, Pseudoreference subaperture images and microlens image-based blind light field image quality measurement. IEEE Trans. Inst. Meas. **70**, 1–15 (2021)
187. J. Xiang, G. Jiang, M. Yu, Z. Jiang, Y.-S. Ho, No-reference light field image quality assessment using four-dimensional sparse transform. IEEE Transactions on Multimedia (2021)
188. Q. Qu, X. Chen, V. Chung, Z. Chen, Light field image quality assessment with auxiliary learning based on depthwise and anglewise separable convolutions. IEEE Trans. Broadcast. **67**(4), 837–850 (2021)
189. P. Zhao, X. Chen, V. Chung, H. Li, Delfiqe-a low-complexity deep learning-based light field image quality evaluator. IEEE Trans. Instrum. Meas. **70**, 1–11 (2021)
190. Z. Pan, M. Yu, G. Jiang, H. Xu, Y.-S. Ho, Combining tensor slice and singular value for blind light field image quality assessment. IEEE J. Selected Topics Signal Process. **15**(3), 672–687 (2021)
191. C. Meng, P. An, X. Huang, C. Yang, L. Shen, B. Wang, Objective quality assessment of lenslet light field image based on focus stack. IEEE Trans. Multimed. **24**, 3193–3207 (2021)
192. H. Huang, H. Zeng, J. Hou, J. Chen, J. Zhu, K.-K. Ma, A spatial and geometry feature-based quality assessment model for the light field images. IEEE Trans. Image Process. **31**, 3765–3779 (2022)
193. S. Alamgeer, M.C. Farias, Light field image quality assessment with dense atrous convolutions. In 2022 IEEE International Conference on Image Processing (ICIP), pp. 2441–2445 (2022). IEEE
194. S. Alamgeer, M.C. Farias, No-reference light field image quality assessment method based on a long-short term memory neural network. In 2022 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), pp. 1–6 (2022). IEEE
195. S. Alamgeer, M.C. Farias, Blind visual quality assessment of light field images based on distortion maps. Front. Signal Process. **2**, 815058 (2022)
196. Z. Zhang, S. Tian, W. Zou, L. Morin, L. Zhang, Deeblif: Deep blind light field image quality assessment by extracting angular and spatial information. In 2022 IEEE International Conference on Image Processing (ICIP), pp. 2266–2270 (2022). IEEE
197. Z. Zhang, S. Tian, W. Zou, L. Morin, L. Zhang, Eddmf: An efficient deep discrepancy measuring framework for full-reference light field image quality assessment. IEEE Trans. Image Process. **32**, 6426–6440 (2023)
198. Z. Zhang, S. Tian, W. Zou, L. Morin, L. Zhang, Pvblif: A pseudo video-based blind quality assessment metric for light field image. IEEE Journal of Selected Topics in Signal Processing (2023)
199. J. Ma, X. Zhang, J. Wang, Blind light field image quality assessment based on deep meta-learning. Optics Lett. **48**(23), 6184–6187 (2023)
200. J. Ma, X. Zhang, C. Jin, P. An, G. Xu, Light field image quality assessment using natural scene statistics and texture degradation. IEEE Transactions on Circuits and Systems for Video Technology (2023)
201. Q. Qu, X. Chen, Y.Y. Chung, W. Cai, Lfacon: introducing anglewise attention to no-reference quality assessment in light field space. IEEE Trans. Vis. Comput. Gr. **29**(5), 2239–2248 (2023)
202. K. Lamichhane, M. Neri, F. Battisti, P. Paudyal, M. Carli, No-reference light field image quality assessment exploiting saliency. IEEE Transactions on Broadcasting (2023)
203. J. Xiang, P. Chen, Y. Dang, R. Liang, G. Jiang, Pseudo light field image and 4d wavelet-transform-based reduced-reference light field image quality assessment. IEEE Transactions on Multimedia (2023)
204. X. Chai, F. Shao, Q. Jiang, X. Wang, L. Xu, Y.-S. Ho, Blind quality evaluator of light field images by group-based representations and multiple plane-oriented perceptual characteristics. IEEE Transactions on Multimedia (2023)
205. J.-X. Chai, X. Tong, S.-C. Chan, H.-Y. Shum, Plenoptic sampling. In Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '00, pp. 307–318. ACM Press/Addison-Wesley Publishing Co., USA (2000). https://doi.org/10.1145/344779.344932

206. H. Zhu, H. Wang, Z. Chen, Minl: Micro-images based neural representation for light fields. arXiv preprint arXiv:2209.08277 (2022)
207. ISO/IEC JTC1/SC29/WG1: JPEG Pleno Workshop on Learning-Based Light Field Coding Proceedings (2022). https://jpeg.org/jpegpleno/documentation.html
208. ISO/IEC JTC 1/SC29/WG1N100306:Information technology - Use Cases and Requirements for Light Field Quality Assessment v5.0. ISO/IEC 21794-1:2020 (2022)
209. IEEE Recommended Practice for the Quality Assessment of Light Field Imaging (P3333.1.4). IEEE (2022). https://standards.ieee.org/ieee/3333.1.4/10873/

Publisher's Note