Densely-sampled Light Field Reconstruction

Suren Vagharshakyan[1][0000−0003−1687−2391], Robert Bregovic[1][0000−0002−3878−7588], and Atanas Gotchev[1][0000−0003−2320−1000]

Tampere University, FI
Firstname.Lastname@tuni.fi

Abstract. In this chapter, we motivate the use of densely-sampled light field as the representation which can bring the required density of light rays for the correct recreation of 3D visual cues such as continuous parallax and focus cues and can serve as an intermediary between light field sensing and light field display. We consider the problem of reconstructing such representation from a sparse set of camera views and approach it in a sparsification framework. More specifically, we demonstrate that the light field is well structured in the set of so-called epipolar images and can be sparsely represented there by a dictionary of directional and multi-scale atoms called shearlets. We present the corresponding regularization method, along with its main algorithm and speed-accelerating modifications. Eventually, we illustrate its applicability for the cases of holographic stereograms and light field compression.

Keywords: Light field · Sparsification · Shearlet transform.

1 Introduction

Human observers interact with the visual world through light. It is light what is sensed by photoreceptors and converted into neural impulses to be further processed by the brain. A number of visual cues, such as stereopsis, focus, motion and head parallax help the human to perceive, understand and navigate through the three-dimensional world. These cues all depend on the way how light is presented and sensed. And it is light what contemporary displays emit in their attempt to recreate the visual world. The aim in designing 3D displays, such as multi-view, light field, holographic, or head-mounted has been to generate visual cues as realistically as possible. Generally speaking, this is achievable by generating and controlling a high amount of directional light rays to meet the visual acuity of the human visual system. This brings the question how to formalize and represent light in such a way so to effectively drive 3D displays and how to sense visual scenes in order to generate that display-driven representation.

In this book chapter, we advocate the use of densely-sampled light field: an overcomplete, yet discrete representation of light in terms of ray optics. The densely-sampled light field can play the role of intermediary between the light as captured by (multiple) cameras and light as recreated by 3D displays. It comes within the framework of the plenoptic function, and as a result of a particular effective light field parameterization. Therefore, in Section 2, we overview the mathematical formalization of the plenoptic function, and its various aspects of parameterization, approximation and sampling. By

discussing the spectral support of 4D light field in different cases, we come up with the corresponding sampling conditions. We overview also recent methods for spatial and angular light field reconstruction which employ various approaches: from depth estimation, through machine learning to sparsity.

Section 3 presents our take on the problem of densely-sampled light field reconstruction. We motivate a sparsification-based approach and discuss the corresponding transform and regularization method. Eventually, we present a few applications to illustrate the applicability and performance of the proposed method.

The presented chapter originates from a doctoral thesis with the same title [56].

2 Basics of Light Field Processing

2.1 Light field modeling and parameterization

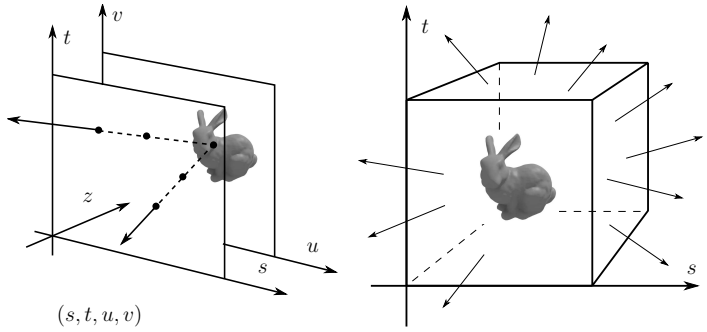
Plenoptic function, as a concept to describe the space of all possible light rays, was first presented by E. Adelson and J. Bergen in [1]. The idea arised from observation that information about the scene, can be modeled as a dense array of light rays of various intensities. Plenoptic function was introduced as tool for efficient mathematical parameterization of light. In such setup, all light rays are parameterized by their location (V_x, V_y, V_z) and direction (θ, ϕ). By adding wavelength λ of light and time instance t, the plenoptic $7D$ function describing light intensity in a given space is defined in the form

$$P = P(V_x, V_y, V_z, \theta, \phi, \lambda, t)$$

This function can be simplified, as shown by L. McMillan and G. Bishop in [45], by considering still light field (fixing time) and replacing the intensity (wavelength) by an RGB representations. Moreover, in practice we tend to discretize positions and angles, which in turn, results in a discrete 5D light field function.

Two plane parameterization M. Levoy and P. Hanrahan remarked in [41], that the aforementioned 5D light field function can be further reduced to only 4 dimensions assuming that the medium through which light rays propagate is completely transparent. Since in this case the ray intensity is constant, a ray can be simply parameterized using corresponding intersection points with two planes - see Figure 1 (b) for illustration. This is referred to as the two-plane light field parameterization $L(s, t, u, v)$ with the (s, t) and the (u, v) plane being also referred to as the camera and the image plane, respectively.

A special case of two-plane parameterization is the so called Lumigraph presented by S. J. Gortler et al. in [24]. Lumigraph enables a convenient description for scene or object that is placed inside a virtual cube. Each of six cube faces is parameterized using a two-plane parameterization with the planes itself being the sides of the cube as illustrated in Figure 1 (a). Therefore, having 4D Lumigraph description of a scene, an arbitrary view can be formed by selecting required samples directly from the Lumigraph thereby avoiding complex calculations.



(a) (b)

Fig. 1: (a) Two-plane parametrization, where an arbitrary ray is parameterized using intersection points with two parallel planes. (b) Light radiance information described by considering radiance over the rays intersecting the sides of the cube.

A two-plane parameterization is a simple and efficient tool especially useful in the problems of analysis and synthesis of views (continuous lightf field) from a given discrete light field. This is achieved by first considering discrete subdivision in each of the s, t, u, v dimensions and second associating each discrete sample (i, j, p, q) to coefficient $x_{i,j,p,q}$ with reconstruction kernel $B_{i,j,p,q}(s, t, u, v)$. Consequently, reconstructed continuous light field $\tilde{L}$ is obtained as follows

$$\tilde{L}(s, t, u, v) = \sum_i \sum_j \sum_p \sum_q x_{i,j,p,q} B_{i,j,p,q}(s, t, u, v).$$

A comparision of the reconstruction quality for different basis functions B has been performed in [24]. It has shown that the use of as quadralinear kernel is beneficially in terms of computational efficiency vs. compromise in quality due to lack of band-limited property of light field function.

Two-plane parameterization is a convenient way to efficiently represent a light field acquired with an array of cameras. Example of such an acquisition system is presented in [60], where a single camera is moving on a plane using gantry, or in [61] where instead of a gantry an array of cameras is used. Recently, a Light field camera (plenoptic camera) capturing system is introduced in [46], [21]. The main difference of the light field camera from the conventional camera is the additional layer of a microlens array in front of the sensor. Obtained data from a light field camera can be interpreted as uniformly sampled two-plane parameterized light field over a small baseline.

Epipolar plane images Epipolar constraint term comes from early stage analysis of stereo images [27], where it is shown, that search of matching features between stereo pair of images from two dimension problem can be reduced to one dimension, if locations of the cameras are available.

In [6] epipolar constraint is generalized for the case of a light field captured by a dense set of cameras that are strictly over a straight-line (t constant). Such special case of the light field is also referred to as a Horizontal parallax only light field. Stacking those images into cube (s, u, v) and slicing the cube along u the obtained 2D image (s, v) is a so-called epipolar-plane images (EPI). Independent analysis / processing performend on each EPIs can be combined into a three-dimensional representation of the whole scene. More details covering the definition and morphological properties of EPI will be presented in Section 3.

Alternative parametrizations The considered visual acquisition system usually one of the main motivation for introducing new parameterizations of the light field function. Two notable examples are the spherical and cylindrical parameterizations that are introduced for efficient parameterization of multiple captured images from the same location.

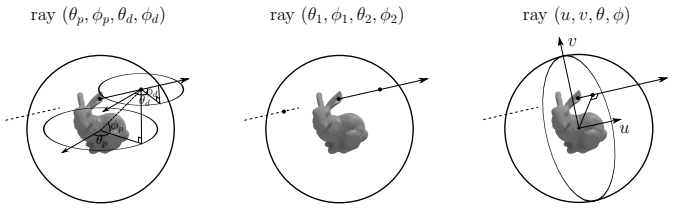


Fig. 2: Identical ray in different spherical light field parameterizations.

In concentric mosaic parameterization, introduced by He and Shum in [52], each ray is described by only three parameters: radius, rotation angle, and vertical elevation, thereby reducing the plenoptic function to three dimensions. The acquisition system consists of a camera moving over planar concentric circles. Similar to view synthesis in case of panoramic images, novel views are rendered by combining the appropriate captured rays. Novel view synthesis works only when the corresponding viewpoint is located inside the planar circular region with the quality of the reconstruction increasing with the number of concentric circles.

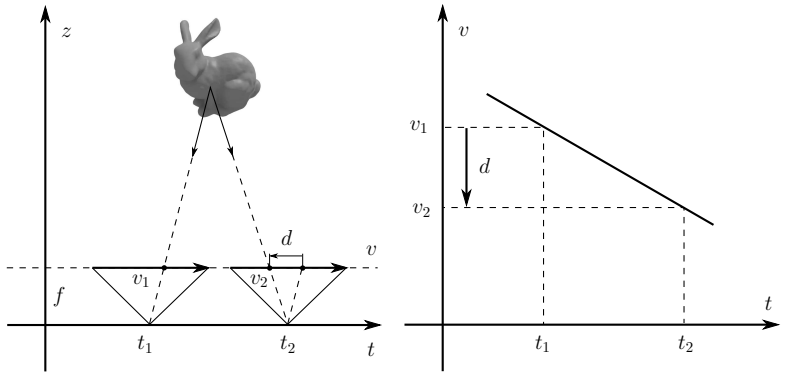
Similar to the Lumigraph, spherical parameterizations assume a finite size scene, such that a unit sphere ecapsulates the whole scene. As shown by Ihm in [31], in a spherical parameterization the position of the light ray emanated from a scene is parameterized using an intersection point on the positional sphere (θ_p, ϕ_p) used as a convex hull of the scene and the direction of the ray is identified by the intersection point with the directional sphere (θ_d, ϕ_d). This is illustrated in Figure 2 (a). Thus, the two-sphere or spherical 4D light field parameterization is defined as $l^{\text{sphere}}(\theta_p, \phi_p, \theta_d, \phi_d)$ function.

Alternative sphere-sphere parameterization (2SP) and sphere-plane parameterization (SPP) are presented in [10]. In 2SP parameterization each light ray is parameter-

ized by its two intersection points with the same sphere, as illustrated in Figure 2 (b). In SPP parameterization, light ray is parameterized by its angle and 2D coordinate of the intersection point of the ray and orthogonal plane, as illustrated in Figure 2 (c).

The spherical LF parametrizations are easily applicable to synthetic data, though, they can also be used when recording real scenes, one example of a sensing setup being the Stanford spherical gantry [41]. More recently Debevec et. all in [47] also presented a spherical LF capturing system. In the proposed new capturing system, two cameras are rotated over the sphere surface in the space, which allows capturing spherical LF of the outside environment. Captured LF data provides information to efficiently generate novel views located within the recorded spherical volume.

2.2 Light field sampling and reconstruction



(a) (b)

Fig. 3: (a) Two plane parameterization of a light field for fixed values of s and u axis. (b) Radiance of same 3D point observed from different position of camera.

The two-plane parameterization is instrumental when defining light field sampling. Consider the camera plane (s,t) plane and the image plane (u,v), as illustrated in Figure 3 (a). For the purposes of sampling, it is suitable to define the (u,v) plane relatively to (s,t) coordinates [16]. A discrete uniform grid is considered on the camera plane (s,t), where each point on the grid represents a pin-hole camera location. Pixels corresponding to the camera form a uniform grid on the image plane (u,v). Each pixel value is formed by the weighted sum of the light radiance arriving at a certain angle to the camera plane. Thus, an arbitrary ray intersecting both planes uniquely determines the quadruple $q = (u,v,s,t)$.

A simplified spectral analysis of the 4D LF function can be carried out if assuming occlusion-free scenes with Lambertian reflectance. The former assumptions implies that the same 3D point can be observed from any location of the camera plane and latter

implies that the radiance of a point is constant in all directions. Then, the observed 4D LF function has a distinct form, which can be easily explained using the EPI notion. Recall that EPI is formed by fixing the parameters s, u and varying parameters t, v for the LF function, such as $E_{(s,u)}(t,v) = l(u,v,s,t)$ is conventionally called horizontal EPI. Identically, vertical EPI is defined as $E_{(v,u)}(t,v) = l(u,v,s,t)$. An EPI example is presented in Figure 3. In EPI, any scene point is represented by a corresponding line with an intensity proportional to the light radiance from the point in different directions. For Lambertian scenes, this line has constant intensity. The disparity d between two images located at (s,t_1) and (s,t_2) of the same observed 3D point is

$$d = v_2 - v_1 = (t_1 - t_2)f/z,$$

where $z = z(q)$ represents the scene depth, i.e. the distance of the surface point corresponding to the ray q from the camera plane.

Assuming $t_1 = 0$ to be the origin of the axis t,

$$l(q) = l\left(u + \frac{f}{z(q)}s, v + \frac{f}{z(q)}t, 0, 0\right).$$

For a simplified case of constant-depth plane $z(q) = z_0$, it can be shown that:

$$L(\Omega_u,\Omega_v,\Omega_s,\Omega_t) = 4\pi^2 L^{'}(\Omega_u,\Omega_v)\delta\left(\Omega_s - f\Omega_u/z_0\right)\delta\left(\Omega_t - f\Omega_v/z_0\right) \quad (1)$$

where $L^{'}(\Omega_u,\Omega_v)$ is the Fourier transform of $l^{'}(u,v) = l(u,v,0,0)$ and δ is the Dirac delta function. Thus, for the constant-depth plane scene, the 4D function L support on the 2D plane (Ω_v,Ω_t) is bounded by the line $\Omega_t = \Omega_v f/z_0$, as shown in Figure 4 (a). The same is true for (Ω_u,Ω_s) plane and the corresponding line $\Omega_s = \Omega_u f/z_0$.

Let's assume an uniform 4D lattice $\Delta q = (\Delta u,\Delta v,\Delta s,\Delta t)$, and a sampling function $p(q) = \mathrm{III}_{\Delta q}(q)$, where $\mathrm{III}_T(t)$ is the Dirac comb function, [44], and $l_s(q) = l(q)p(q)$.

The Fourier transform of the sampled LF L_s at angular frequency $\Omega_q = (\Omega_u,\Omega_v,\Omega_s,\Omega_t)$ is

$$L_s(\Omega_q) = L(\Omega_q) * \mathrm{III}_{2\pi/\Delta q}(\Omega_q).$$

The convolution with the Dirac comb function implies that L_s consists of periodical replicas of L at a 4D uniform lattice defined as

$$\left\{\frac{2\pi}{\Delta u}m_1, \frac{2\pi}{\Delta v}m_2, \frac{2\pi}{\Delta s}l_1, \frac{2\pi}{\Delta t}l_2\right\}_{m_1,m_2,l_1,l_2\in\mathbb{Z}},$$

as illustrated in Figure 4 (b) for the case of constant depth.

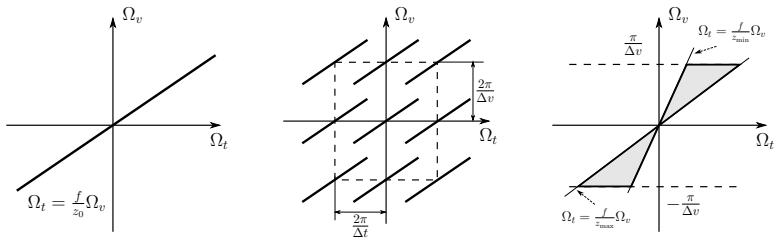
For the case of multiple depth planes, instead of single line, there will be multiple lines on the Fourier plane. It has been proved that all of them are confined between the lines $\Omega_t = f\Omega_v/z_{\max}, \Omega_t = f\Omega_v/z_{\min}$ corresponding to the minimum and maximum depth $[z_{\min}, z_{\max}]$, c.f. Figure 4 (c) [16]. This bow-tie support shape forms the baseband of the light field. Its periodical replicas would be generated during sampling and have to be filtered out during reconstruction. To avoid aliasing, the sampling intervals have

to be chosen with respect to the min and max depth of the scene. This is specifically important for the distances between cameras on the camera plane, i.e. along the t and s axes. It has be shown that

$$\Delta t_{\max} = \frac{1}{K_{f_v} f \left(z_{\min}^{-1} - z_{\max}^{-1}\right)}.$$

where $K_{f_v} = \min\left(B_v^s, 1/(2\Delta v), 1/(2\delta v)\right)$ is the maximum frequency in axis Ω_v. K_{f_v} depends on the complexity of texture information represented with the highest scene texture frequency B_v^s and on the rendering camera resolution δv. If textural complexity is ignored and full resolution images are rendered then the maximal frequency is $K_{f_v} = 1/(2\Delta v)$ [16].

Figure 5 illustrates cases of different reconstruction filters [16]. Figure 5 (a) presents direct interpolation when the constant-depth plane is at the infinity. Figure 5 (b) depicts an optimal filter support of a constant-depth plane rendering at z_{opt}, where $z_{\text{opt}}^{-1} = (z_{\min}^{-1} + z_{\max}^{-1})/2$. For the case of the optimal filter, camera spacing can be increased such that replicas are placed compactly as shown in Figure 5 (c).

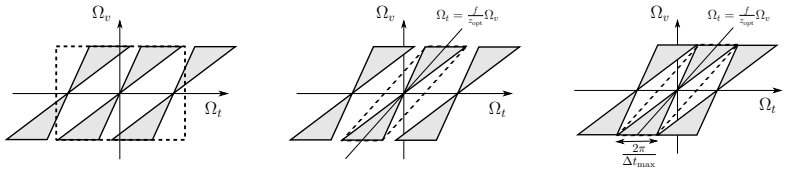


(a) (b) (c)

Fig. 4: The Fourier transform support on (Ω_t, Ω_v) plane, (a) continuous light field with a constant depth, (b) sampled light field with a constant depth, (c) depth varies between $z_{\min}$ and $z_{\max}$.

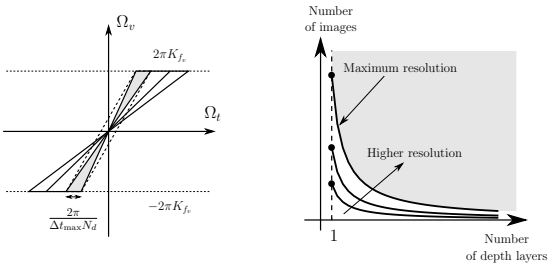
Further optimization of the sampling rate and corresponding reconstruction filters can be achieved by considering depth layering. Narrower bands corresponding to dominant depth layers can be specified and then each depth layer can be processed by the corresponding optimal filter, as illustrated in Figure 6 (a). Thus, for N_d number of layers, the minimum sampling rate is decreased and corresponding maximum distance of camera spacing is increased. $\Delta t_{\max,N_d} = \Delta t_{\max} N_d$.

The fundamental relation between the number of depth layers N_d and number of camera images N_i has been derived in [16] in the form $N_d\sqrt{N_i} = K_{f_v}$, resulting in the so-called *optimal sampling curve*, c.f. Figure 6 (b). The curve particularly suggests that the number of required images in classical plenoptic sampling is still high and more advanced methods are required for achieving similar reconstruction quality with smaller number of images.



(a) (b) (c)

Fig. 5: (a) Direct reconstruction filter with implicit assumption of infinite depth. (b) Filtering using z_{opt}. (c) Optimal packing in frequency domain is achieved in case of critical camera spacing distance $\Delta t_{\max}$.



(a) (b)

Fig. 6: (a) Uniform multi-layer depth decomposition represented in the frequencies domain. (b) Minimum sampling curve for different rendering resolutions. Any point in the highlighted region represents redundancy for rendering in joint image and geometric space.

More recent works have attempted addressing the problem of the high number of images needed for LF sampling and reconstruction. Alternatively to depth layering, the concept of surface light field has been proposed in [62]. The approach makes a connection between surface geometric modeling and the corresponding radiated light field. A scene surface is modeled by a simplified base mesh K_0, which is projected onto the more complex geometric surface M Figure 7. The latter gives rise to a surface light field $L(u, \omega)$, which represents the radiance of a light ray in direction ω at a point u on K_0. $L(u, \omega)$ is considered piecewise-linear and composed of LF primitives, called *lumispheres*, parameterized by the ray direction ω). Lumispheres are recovered from given images by least-squares approximation. The model is applicable to non-Labertian scenes with complex radiance functions, however the reconstruction quality heavily depends on the accuracy of the approximated scene geometry.

The relation between the two-plane parameterized light field and surface light field along with the corresponding spectral analysis has been developed in [15]. In summary, the work has aimed at approximating the non-bandlimited Fourier transform of

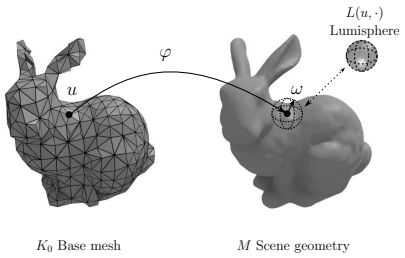


Fig. 7: Surface light field as presented in [62].

the surface light field and to modeling occlusions in the spectral domain. Further, more accurate LF bandwidth estimations have been developed for the cases of essential bandwidth [18], finite field of view [23] and finite scene width [22].

2.3 General methods for light field reconstruction

In this section, we present an overview of recent approaches and methods which deal with light field reconstruction. The majority of methods uses depth in one or another form and therefore the accurate depth estimation from light field is also discussed. More recent methods employ also modern machine learning approaches. Continuous light field reconstruction through sparsification is discussed as well.

Depth based methods Wanner et al. [58] have proposed a method for disparity estimation directly from the light field. A structure tensor is applied on epipolar plane images for fast local disparity estimation. Then, globally consistent disparity maps with sub-pixel precision are obtained from local estimates through convex regularization. They are used in a variational inverse problem aimed at spatial and angular super-resolution. The method has been developed for processing data from plenoptic cameras, therefore relatively small disparity range has been considered.

Conventional disparity estimation methods employ a three-step framework comprising cost volume construction based on hypotheses, cost volume filtering (i.e. regularization using aggregation in the spatial domain), and label selection (i.e. selection of most probable hypothesis, typically winner-takes-all) [30]. A similar framework has been used in [32] for an accurate disparity estimation from light fields acquired by plenoptic cameras (i.e. relatively small disparity range between sub-aperture images). To handle such data, an accurate sub-pixel displacement algorithm using 2D Fourier transform for the cost volume construction. Suitable depth hypothesis layers are formed and used to find per-pixel disparities. The latter are enhanced through a discrete multi-label optimization based on graph cuts and iterative quadratic fitting. The final disparity map is with sub-pixel precision and can be used for LF angular or spatial super-resolution.

Alternative methods for disparity estimation from plenoptic data have been presented in [53], [65].

The case of wide field of view horizontal parallax LF capture has been addressed in [34]. Conventional disparity estimation methods are inefficient for the high amount of data in such imagery. Therefore, the proposed technique utilizes a fine-to-coarse refinement technique with the aim to obtain accurate disparity maps from sufficiently densely sampled light fields and avoid explicit global regularization. A novel sparse representation for a set of adjacent EPIs has been presented, comprising a set of distinct lines, obtained by considering densely sampled LF. This representation is obtained at edges on the high-resolution image first and then further proceeded to successive coarse EPI resolutions to obtain disparity estimation on smooth spatial areas, where edges are not well-defined. The proposed technique implies EPI constraints between images and is especially efficient for processing high spatio-angular LF datasets.

Machine learning methods The machine learning instrumentation has proven quite effective for solving the problem of light field reconstruction. Kalantari at. al. [33] have proposed a learning-based approach aimed at synthesizing intermediate views from sub-sampled plenoptic images, e.g. such captured by the Lytro Illum camera. The work utilizes two neural networks: one for disparity estimation and another for view synthesis using the estimated disparity. Both networks have been trained simultaneously by minimizing the error between synthesized and ground truth views. The disparity-estimating CNN, which consists of four convolutions layers with decreasing kernel sizes followed by a rectified linear unit, has generated high quality disparity maps. Even though, a subsequent color prediction CNN is required to model the complex relationship between the final image and warped images around occlusions. The method has demonstrated superior results when compared with [58], [32], especially around occlusion boundaries.

In [64], the LF angular super-resolution has been formulated as a problem of EPI high-frequency details reconstruction. The given low resolution EPI is considered as a subsampled version of the densely sampled EPI. The former undergoes a covolution with a smoothing kernel (e.g. Gaussian kernel) with the aim to extract low-frequency features. The result is processed by a CNN, which acts as a high-frequency reconstruction operator. It is designed as a residual neural network with three convolution layers with decreasing kernel sizes together with a rectified linear unit. This network is used only to predict angular domain detail information from blurred and upsampled EPI. Further, the spatial detail of the EPI is recovered through a non-blind deblur operation based on the method from [35]. The whole densely sampled light field is reconstructed by applying the proposed "blur - restoration - deblur" framework for every EPI in both horizontal and vertical directions.

The so-proposed method has demonstrated good reconstruction results for to 5 pixels disparity between adjacent views. For higher disparities, the deblurring kernel has proven inefficient. This limitation, has been addressed in [63], by proposing a depth-assisted rendering technique for multiview imagery with large disparity range [57]. The method uses a roughly discretized disparity map, obtained using the method in [30]. For each discrete disparity region, appropriate shearing is applied on corresponding EPI region, to get a disparity range small enough to be processed by the original "blur

- restoration - deblur" method. The final result is formed by blending together multiple super-resolved EPIs regions.

A two-step method for disparity estimation by EPI analysis has been presented in [25]. For a given 4D LF, hyperplane orientations are predicted for the central image using a CNN applied on horizontal and vertical EPIs. The predicted orientations (i.e. disparities) are then refined by a generalized total variation regularization procedure based on the method in [7]. The approach has been improved in [26] by designing a neural network working on 3D subsets of the 4D LF (using 2D spatial and one angular dimension). This allows to effectively suppress artifacts in spatial domain.

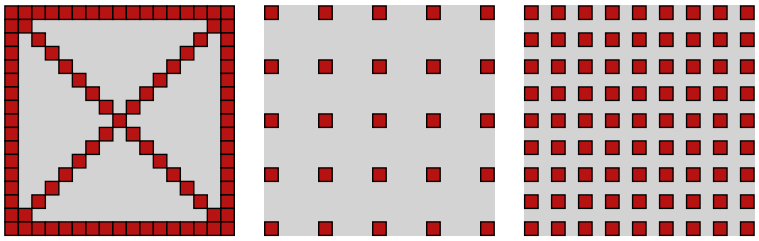
An end-to-end neural network architecture for disparity estimation from 4D light field has been proposed in [51]. The input data consists of views containing horizontal, vertical and two diagonal images in stucks, always containing the central view. The designed network has a multi-stream structure, such that every 1D image stack subset is processed through three convolution layers in order to get sets of features describing the corresponding image stack. The feature sets are concatenated together and processed together by additional convolution layers followed by a rectified linear unit. The work discusses also the optimal number of input views. The proposed method has demostrated high-quality results for the HCI 4D Light Field Benchmark [29].

Extraction of non-Lambertian scene properties from LF has been attempted in [4]. The authors have proposed an encoder-decoder network aimed at decomposing the LF data into disparity, diffuse and specular components. The encoder part reduced the multidimensional structure of the light field. Further, multiple decoders extract the targeted intrinsic components. The encoder is applied on each epipolar-plane independently. It contains 18 residual blocks, which are gradually decreasing the input epipolar volume to in spatial domain and increasing its feature domain. The encoder features are further processed by the multiple decoder pathways. The auto-encoder path reconstructs the original input data, while the three other decoders generates disparity, diffuse and specular components, respectively. All decoders are constructed of residual blocks with transpose convolution layers. A dichromatic reflection model is considered, such that the final radiance is formed by the sum of the diffusion and specular information.

Light field reconstruction by sparsification in Fourier domain Shi et al. [50] have proposed for the first time to cast the LF reconstruction as a problem of sparsification in a transform domain. The authors have motivated the choice to seek LF sparsity continuous Fourier domain rather than in discrete Fourier domain.

A signal of length N is k-sparse in the continuous Fourier domain, if it can be represented as a combination of $k < N$ frequencies, not necessarily located at integer coordinates (hence, continuous). Therefore, the signal reconstruction requires estimating both the frequencies and the corresponding transform coefficients. Consider the reconstruction of a two dimensional signal $\{x[u,v], \forall u, v = 0, \ldots, N-1\}$ from a set of measurements $x_S = \{x[u,v], \forall (u,v) \in S\}$. The sparsifying solution is obtained by solving the following minimization problem

$$\underset{a_l, \omega_{u_l}, \omega_{v_l}}{\arg\min} \sum_{(u,v)\in S} \left\| x(u,v) - \frac{1}{N} \sum_{l=0}^{k} a_l \exp\left(2\pi i \frac{u\omega_{u_l} + v\omega_{v_l}}{N}\right) \right\|_2^2,$$



(a) (b) (c)

Fig. 8: Sampling pattern where every rectangle represents one view from a LF consisting of 17×17 views. (a) box and two diagonals pattern consisting of 93 views used for method [50]. (b), (c) uniformly decimated setup consisting of 5×5 and 9×9 views respectively.

which can be represented more compactly in matrix form

$$\arg\min_{a,\omega} \|x_S - A_\omega a\|_2^2 ,$$

where $a = \{a_l\}_{l=0}^k , \omega = \{(\omega_{u_l}, \omega_{v_l})\}_{l=0}^k$.

The problem is solved by alternating minimization: for fixed k frequency locations ω, the corresponding optimal coefficients a are estimated as $a = A_\omega^\dagger x_S$, while the optimal frequency locations ω are found by

$$\omega^* = \arg\min_{\omega} \left\| x_s - A_\omega A_\omega^\dagger x_S \right\|_2^2 .$$

The functional is minimized by gradient descent, where the gradient is approximated by evaluating the error function over 8 directions around every frequency position and updating it in most descending direction.

In [50], the sampling set S is composed from a set of 1D discrete sampling lines. These lines are used in a voting scheme based on the Fourier slice theorem aimed at obtaining reliable initial estimates for the frequency positions.

Using the proposed sparsification method, the 4D light field $L(x, y, u, v)$ is reconstructed at all angular locations (u, v) from the given sampling set S, illustrated by the red squares in 8, by independently reconstructing each $\hat{L}_{\omega_x,\omega_y}(u, v)$ 2D slice for fixed spatial frequencies.

The proposed method has shown prominent reconstruction quality especially for light fields representing non-Lambertian scenes.

3 Light field reconstruction trough sparse modelling

3.1 Problem formulation

Densely sampled light field (DSLF) refers to a regular light field representation consisting of full-parallax camera views, where the maximum disparity between adjacent

views is one pixel at most. This is an attractive representation since it allows generating arbitrary rays by simple (quad-)linear interpolation [43].

Direct capture of a densely-sampled light field over a large baseline at full parallax requires a high number of densely positioned cameras. Such setting is inefficient and in many cases unfeasible. In reality, light fields have been captured by an array of cameras, either unstructured or uniformly located on a line or 2D grid. This discrepancy between a desired representation and physically limited capture settings sets the fundamental problem of reconstructing, computationally, the densely sampling LF from multiperspective images, considered as LF samples on coarser grids. For the sake of simplicity, these grids are usually assumed regular, which reflects the case of horizontally and vertically aligned (rectified) cameras. As discussed in the previous section, rectified views when put together, form LF slices referred to as EPIs, which implicitly represents the scene geometry.

Hereafter, we set the DSLF reconstruction problem as problem of reconstructing densely sampled EPI (DSEPI) from decimated (in angular dimension or in camera plane) LF samples. The concept of densely sampled EPI is inspired by [43], where the limit of necessary sampling in angular dimension is formulated in terms of depth.

For the sake of simplicity, in most derivations, we consider the horizontal parallax case, where EPIs are formed after stacking all images together and taking 2D slices along the horizontal and camera motion dimensions for a fixed vertical coordinate. Rows in a particular EPI represents horizontal lines from different perspective views. While put in the context of DSEPI, these rows are separated by blank areas of the missing intermediate views. An illustration of this sampling is presented in Figure 10 (a),(c). Reconstructing DSEPIs for all vertical coordinates gives the fully-reconstructed DSLF.

We formulate the problem DSEPI reconstruction in terms of signal reconstruction with sparsity constrains. More specifically, we consider the reconstruction in some suitable transform domain, where the LF in sparse. Based on the structural properties of DSEPI, we adopt shearlet frames as the sparsification transform employing their directional sensitivity properties and present the main reconstruction method and its accelerations. While it is initially formulated to handle coarse sampling in angular domain, the approach is flexible enough to address also the problem of LF super-resolution is spatial domain.

3.2 Sparse representation

Various image processing problems, such as denoising, debluring, inpainting, superresolution can be formalized by a system of linear equations

$$y = Ax + \epsilon, \tag{2}$$

where A is the process, modeled in some metric space, which acts on the input (unknown) image x resulting in the available (distorted, undersampled, blurred) image y, additionally contaminated by noise ϵ. The latter is usually modeled as an independent and identically distributed Gaussian with zero mean and standard deviation of σ, i.e. ($\epsilon \sim \mathcal{N}(0, \sigma^2 I)$).

The corresponding inverse problem of finding x can be formulated in least squares (LS) sense

$$\bar{x} = \arg\min_x \|y - Ax\|_2^2 .$$

Typical image processing problems result in an underdetermined system of linear equations, which implies an infinite set of solutions for x satisfying $\|y - Ax\|_2^2 = 0$. Such cases are considered ill-posed and require additional regularization to determine the desirable solution. In the general case, a minimization of a cost function, composed of a fidelity term f and a penalty (regularization) term s is attempted:

$$\arg\min_x f(x) + \lambda s(x). \tag{3}$$

The fidelity term $f(x)$ ensures the consistency between the solution and the measurements. The penalty term or regularizer $s(x)$ guarantees the prior model of the signal. In line with the least squares formulation, the fidelity term can defined as $f(x) = \frac{1}{2\sigma^2}||y - Ax||_2^2$. In order to solve Equation 3 given the model Equation 2, state of the art approaches have been employing the Alternating direction method of multipliers (ADMM) [59]. A more recent version, referred to as Plug-and-Play (P&P ADMM) has proposed to avoid explicitly presenting s by introducing a regularization procedure in the form of denoising thresholding [17].

Consider a dictionary given by a matrix D, the sparse representation of a signal x refers to the coefficients α, such that

$$\arg\min \|\alpha\|_0 , \text{ subject to } x = D\alpha, \tag{4}$$

where $\|\alpha\|_0 = \#(\alpha_k \neq 0)$ denotes the pseudo-norm l_0 which equals the number of non-zero coefficients. The entries of the matrix D can be the analysis functions of a fixed transform $D = \{\phi_n\}_{n\in\Gamma}$, such that $(Dx)[n] = \langle \phi_n, x\rangle$. Finding the sparse representation is an NP-Hard problem. A sufficiently sparse α can be found by replacing the l_0 pseudo-norm with the l_1 norm, leading to what is known as Basis Pursuit (BP) algorithm [11].

$$\min_\alpha \|\alpha\|_1 , \text{ subject to } x = D\alpha \tag{5}$$

It has been proved that the problems 4 and 5 are equivalent in the case of *sufficient sparsity* quantified as

$$\|x\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu(D)}\right),$$

where $\mu(D) = \max_{k\neq l} \frac{\langle \phi_k \phi_l\rangle}{\|\phi_k\|_2 \|\phi_l\|_2}$ represents the *mutual coherence* of the dictionary [12].

The design of the sparsifying dictionary depends on the set of considered signals and the application at hand. For signals representing natural images formed by conventional digital cameras, various dictionaries have been proposed ranging from the widely-used discrete cosine and wavelet transforms to more dedicated, usually directional transforms such as curvelets, ridgelets, bandlets, etc. [44]. Dictionaries can be also learned through approaches such as K-SVD algorithm [2] and sparse coding [38].

Consider a transform in the form of a tight frame, which is a generalization of a basis. The frame is defined by its analysis $\alpha = \Phi x$ and synthesis $x = \Psi\alpha$ transforms, such that in the general case $\Psi\Phi = I$ and $\Phi\Psi \neq I$. The regularization term in the minimizer in Equation 3 can be formulated in terms of sparse transform coefficients $s(\alpha) = ||\alpha||_p$, where $p = 0, 1$. The regularization is implemented in the form of a thresholding operator $\mathcal{T}_t(\cdot)$, acting on the transform coeffiecinets and yelding a denoised version of the signal $\mathcal{D}(x, \sigma) = \Psi\mathcal{T}_{t(\sigma)}(\Phi x)$ [44].

Following the approach proposed in [8], the solution can be found by iterations

$$x_{k+1} = \Psi\mathcal{T}_\lambda\Phi(x_k + A(y - x_k)) \tag{6}$$

It has been shown that the convergence of Equation 6 is equivalent to solving the minimization problem

$$\arg\min_\alpha \frac{1}{2}\left\|A\Psi\alpha - y\right\|_2^2 + \frac{\varkappa}{2}\left\|(I - \Phi\Psi)\alpha\right\|_2^2 + \lambda\left\|\alpha\right\|_1. \tag{7}$$

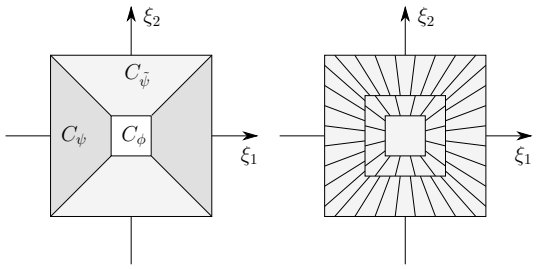
with $\varkappa = 1$, referred to as balanced approach [8], [49]. Its solution has been eventually derived in the following iterative form [9]

$$\begin{aligned} a) \quad & x_k = \Psi\alpha_k \\ b) \quad & \eta_k = \Phi\left(x_k + \frac{1}{\varkappa}(y - Ax_k)\right) \\ c) \quad & \omega_{k+1} = \mathcal{T}_{\gamma\lambda}\left(\alpha_k + \gamma\varkappa(\eta_k - \alpha_k)\right) \end{aligned} \tag{8}$$

3.3 Shearlet frame

Suitable sparsifying dictionaries (transforms) have been studied predominantly for the case of natural images. For such images, dictionaries have been requred to optimally approximate curvilinear singularities of the underlying 2D functions (images). The optimal approximation, in this case, is defined using the decay rate of l_2 error of the best N-term approximation. More specifically, the development of such systems has been specified for *cartoon-like functions* consisting of C^2 functions being compactly supported on the unit square, except for a closed C^2 discontinuity curve. Examples of developed systems include *curvelets* [14], and *contourlets* [19]. What is common for such systems is their ability to handle directional properties in images.

The construction of discrete shearlet frame has followed a similar approach by controlling the orientation of the system's atoms trough a shear operator [36]. It is precisely this property, which makes the shearlet frame particularly interesting for epipolar-plane image representation, since EPI structure is formed by shearing rather than rotation or other curve motion. The compactly-supported shearlet system is of special interest since it contains atoms which are compacty supported in both spatil and Fourier domains [37]. Though it is not a Parseval frame, it is still applicable for approximating (sparsifying) *cartoon-like functions*. The theoretical framework of the universal shearlet system has been developed in [20]. It includes a parameterized systems family, which, for varying parameter value, can describe the wavelet system, the parabolic shearlet system and the



(a) (b)

Fig. 9: (a) Outlined regions are corresponding to frequency plane separation for shearlet transform design. Two cone-adapted regions are corresponding to $C_\psi, C_{\tilde{\psi}}$ set of filters and central rectangle region corresponds to C_ϕ low pass filter. (b) Frequency plane tilling obtained by whole shearlet transform using two scales of decomposition $J = 2$.

ridglet system [13]. Departing from the nonseparable shearlet transform described in [42], hereafter we present a modified version, which has been purposefully designed for efficient representation of functions having singularities along straight lines, in contract to the image-inspired case of parabolic curves approximated by ridges.

The cone-adapted discrete shearlet system SH is defined as a set of 2D functions formed by shearing S, translation and parabolic scaling A transforms applied on generator functions: a scaling function ϕ and two shearlets $\psi, \widetilde{\psi} \in L^2(\mathbb{R}^2)$. For $c = (c_1, c_2) \in \mathbb{R}^2_+$, the system is defined as

$$\mathrm{SH}(\phi, \psi, \widetilde{\psi}; c) = \Phi(\phi; c_1) \cup \Psi(\psi; c) \cup \widetilde{\Psi}(\widetilde{\psi}; c), \tag{9}$$

. With reference to Figure 9 (a), the subset $\Psi(\psi; c)$ corresponds to the cone-shaped region C_ψ, the subset $\widetilde{\Psi}(\widetilde{\psi}; c)$ corresponds to the region $C_{\tilde{\psi}}$ and $\Phi(\phi; c_1)$ - to the central part C_ϕ. This division of the frequency plane is achieved using the following definitions

$$\begin{aligned}
&\Phi(\phi; c_1) = \{\phi_m = \phi(\cdot - c_1 m) : m \in \mathbb{Z}^2\} \\
&\Psi(\psi; c) = \{\psi_{j,k,m} = 2^{3/4j}\psi(S_k A_{2^j} \cdot - M_c m) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\} \\
&\widetilde{\Psi}(\widetilde{\psi}; c) = \{\widetilde{\psi}_{j,k,m} = 2^{3/4j}\widetilde{\psi}(S_k^{\mathsf{T}} \widetilde{A}_{2^j} \cdot - \widetilde{M}_c m) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\}
\end{aligned}$$

where A and $\widetilde{A}$ are scaling matrices, S_k is shearing matrix, $M_c = \mathrm{diag}(c_1, c_2)$ and $\widetilde{M}_c = \mathrm{diag}(c_2, c_1)$ are (translation) sampling matrices, as follows

$$A = \begin{pmatrix} 2^j & 0 \\ 0 & 2^{j/2} \end{pmatrix}, \widetilde{A} = \begin{pmatrix} 2^{j/2} & 0 \\ 0 & 2^j \end{pmatrix}, S_k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}.$$

This construction is suitable for images with parabolic singularities. By modifying the scaling matrix to become $A = \mathrm{diag}(2^j, 2^{-1})$, the shearlet system can be tuned to handle images with line singularities, where the new scaling matrix would guide the required number of shears in each scale of the frequency plane tilling.

The above proposed transform is continuous however it has to handle discrete signals. A natural assumption to start with is to consider a sufficiently large $J > 0$, for which a continuous 2D signal function f is represented by the discrete signal f^d and the scaling function ϕ

$$f(x_1, x_2) = \sum_{(k_1,k_2)\in\mathbb{Z}^2} 2^J f^d[k_1, k_2]\phi(2^J x_1 - k_1, 2^J x_2 - k_2).$$

In a further assumption, $\phi(x_1, x_2) = \phi^1(x_1)\phi^1(x_2)$. Then, the 1D scaling and wavelet functions $\psi^1(x)$, $\phi^1(x)$ are represented by two-scale equations

$$\phi^1(x) = \sum_{k\in\mathbb{Z}} h[k]\sqrt{2}\phi^1(2x - k) \quad \text{and} \quad \psi^1(x) = \sum_{k\in\mathbb{Z}} g[k]\sqrt{2}\phi^1(2x - k).$$

The Fourier coefficients of the trigonometric polynomial H_j and G_j

$$H_0 \equiv 1, \quad H_j(\xi) = \prod_{i=0}^{j-1} H(2^i\xi), \quad G_j(\xi) = G(2^{j-1}\xi)H_{j-1}(\xi), \quad j = 0, \ldots, J \tag{10}$$

are denoted by g_j and h_j.

For better performance, it has been suggested to select a 2D nonseparable wavelet function $\psi(x_1, x_2)$ corresponding to the scaling function $\phi(x_1, x_2)$ such as

$$\hat{\psi}(\xi_1, \xi_2) = P(\xi_1/2, \xi_2)\hat{\psi}^1(\xi_1)\hat{\phi}^1(\xi_2),$$

where $P(\xi_1, \xi_2)$ is trigonometric polynomial representing 2D fan filter with wedge-shaped essential support [42]. By appropriate selection of the sampling grid M_c, the coefficients of the shearlet transform corresponding to the system elements $\{\psi_{j,0,m}\}_{m\in\mathbb{Z}^2}$ can be calculated by applying a digital filter $p_j * \left(g_{J-j} \otimes h_{J-j/2}\right)$ on the discrete signal f^d, where p_j are the Fourier coefficients of a scaled 2D fan filter $P(2^{J-j-1}\xi_1, 2^{J-j/2}\xi_2)$.

A poper discretization of the whole system [42], [55] eventually leads to the following digital implementation:

$$\Psi^d_{j,k} = S^d_{k2^{-(j+1)}} \left(p_j * (g_{J-j} \otimes h_{J+1})\right), j = 0, \ldots, J-1, |k| \le 2^j + 1.$$

This set of transform filters corresponds to the cone-shaped region C_ψ of the frequency plane highlighted in Figure 9 (a). The region $C_{\tilde{\psi}}$ is covered by the filters $\hat{\tilde{\psi}}^d_{j,k}(\xi_1, \xi_2) = \hat{\psi}^d_{j,k}(\xi_2, \xi_1)$. The central region C_Φ is dealt with a single filter $\phi^d = h_J \otimes h_J$.

The constructed discrete shearlet system is not orthogonal, therefore dual frame elements are required for the synthesis transform. Using auxiliary notation

$$\hat{\Psi}^d = |\hat{\phi}^d|^2 + \sum_{j=0}^{J-1} \sum_{|k|\le 2^j+1} \left(|\hat{\psi}^d_{j,k}|^2 + |\hat{\tilde{\psi}}^d_{j,k}|^2\right)$$

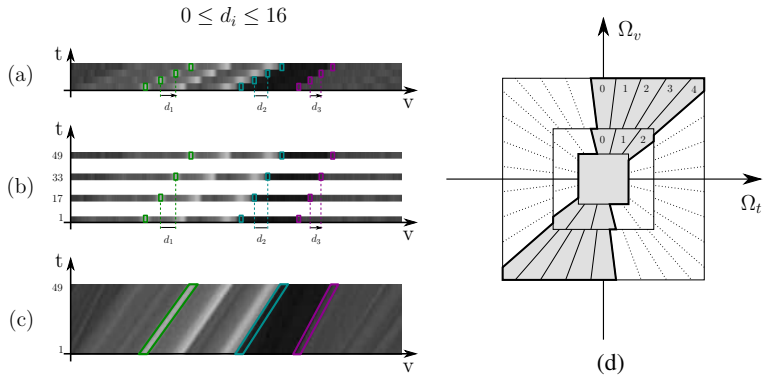


Fig. 10: (a) Subsampled densely sampled epipolar-plane image, assuming that disparities between consecutive rows are no more than 16px. (b) Subsampled data can be interpreted as every 16-th row if desirable densely sampled light field. (c) Corresponding densely sampled light field with disparities don't exceeding 1px. (d) Highlighted shearlet transform atoms used in EPI reconstruction algorithm. Selected atoms correspond to EPI structure. Each disparity layers ($k = 0, 1, ...4$) represented with one transform atom in each scale ($j = 0, 1$).

the dual elements are defined as follows

$$\hat{\varphi}^d = \frac{\hat{\phi}^d}{\hat{\Psi}^d},\ \hat{\gamma}^d_{j,k} = \frac{\hat{\psi}^d}{\hat{\Psi}^d},\ \hat{\tilde{\gamma}}^d_{j,k} = \frac{\hat{\tilde{\psi}}^d}{\hat{\Psi}^d}.$$

Finally, the analysis operator corresponding to the construction shearlet frame is given by

$$S(f^d_J) = \left\{ s_{j,k} = f^d_J * \bar{\psi}^d_{j,k}, \tilde{s}_{j,k} = f^d_J * \bar{\tilde{\psi}}^d_{j,k}, s_0 = f^d_J * \bar{\phi}^d \right\}$$

and the synthesis operator uses the dual elements

$$S^*\left(\{s_{j,k}, s_0\}\right) = \sum_{j=0}^{J-1} \sum_{|k| \le 2^j+1} \left(s_{j,k} * \gamma^d_{j,k} + \tilde{s}_{j,k} * \tilde{\gamma}^d_{j,k} \right) + s_0 * \phi^d.$$

3.4 Epipolar-plane image reconstruction

Main method The epipolar-plane image reconstruction can be formulated as a sparse regularization problem utilizing the shearlet frame [55]. The input signal is a subsampled EPI y with respect to the desired DSEPI x: $y = Mx$, where M is the masking or subsampling matrix. The input EPI has disparities between adjacent views in the range $[d_{\min}, d_{\max}]$ pixels. A pre-shearing operation is applied to guarantee positive disparities $[0, d_{\text{range}}]$, with $d_{\text{range}} = d_{\max} - d_{\min}$. The subsampled EPI is organized to get the size of the target DSEPI, i.e. the k-th row (view) takes kd_{range}-th row of the densely sampled epipolar-plane image. This organization enforces the densely sampled condition, where

the disparities are in the range $[0, 1]$px. An example of desirable densely sampled EPI is given in Figure 10 (c) where every 16-th row is taken from input EPI at Figure 10 (a).

The sparsifying shearlet transform is implemented at $J = \lceil \log_2(d_{\text{range}}) \rceil$ scales. This number guarantees an alias-free central low-pass region. The participating shearlet atoms are selected according to the desired disparity range $[0, 1]$ pixels, c.f. Figure 10 (b) and Figure 10 (d).

The proposed algorithm is a version of the algorithm in Equation 8 and employs an iterative scheme involving the analysis transform S and its synthesis (dual) counterpart S^*[55]

$$x_{k+1} = S^* \left(\mathcal{T}_{\lambda_k} \left(S(x_k + \alpha_k(y - Mx_k)) \right) \right), \tag{11}$$

where $\left(\mathcal{T}_\lambda x\right)(k) = \begin{cases} x(k), & |x(k)| \geq \lambda \\ 0, & |x(k)| < \lambda \end{cases}$ is a hard thresholding operator taking linearly decaying thresholding values λ_n in the range $[\lambda_{\max}, \lambda_{\min}]$. A large value of the parameter α_k provides additional convergence acceleration. This result is partially related to the sparsity of the measurements matrix M. Typically, the number of available samples is significantly smaller than the number of reconstructed samples. Therefore, significant amplification is required to increase the influence of available samples at every thresholding iteration. Nevertheless, an unlimited increase of the parameter α_k diverges the series x_k. The factor can be also made adaptive [5], [55].

Full parallax processing The presented reconstruction algorithm assumes EPI formed by an 1D parallax. Full parallax light field data can be processed in consequent manner, i.e. reconstructing the vertical DSEPIs after obtaining all horizontal DSEPIs. This direct approach assumes the same disparity range in both directions. However, the number of shearlet scales, and hence the computational cost, is directly determined by the disparity range. This motivates processing full-parallax EPIs in a hierarchical reconstruction (HR) order [55]. Consider the example in Figure 11: there is a 5×5 array of input images and the targeted DSLF contains an 17×17 array of images. The reconstruction in performed in three steps by alternating the reconstruction directions, thus making use of the already twice-decreased maximal disparity. For large disparity range, the number of alternating steps can be increased in the same fashion, building a hierarchy where disparities of the subsequent step are reduced twice by the current step.

Performance This method has demonstrated a superior performance against the state of the art and specifically against view interpolation methods relying on depth and hence requiring multi-view depth estimation [55].

While the main method has been developed to handle uniformly-sampled LFs (i.e. rectified views from equidistantly spaced cameras), non-uniform sampling can be handle equally successfully [55]. It is worth mentioning that the required input parameter d_{range} has no direct interpretation in the case of non-uniform sampling and the number of shearlet scales has to be determined by the maximum disparity between adjacent views in order to increase the performance, c.f. Figure 12 (b).

The shearlet atoms are directly related with properties of LF imagery and the imposed sparsity allows avoiding any depth estimation, which might be required for view

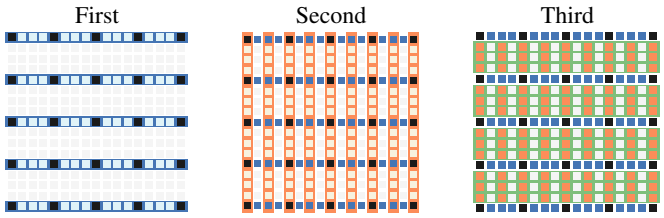


Fig. 11: Proposed fast processing order illustrated for 17 x 17 array of images. Reconstruction is divided into three steps (blue, orange, green) to decrease the disparity range in the successive steps.

interpolation otherwise. The LF views are reconstructed as a weighted combination of atoms which can handle cases corresponding to non-Lambertian scenes, which are challenging for depth estimation. Figure 13 illustrates the performance of the proposed method for a semi-transparent scene and against a well-known disparity estimation algorithm, referred to as semi-global block matching (SGBM) [28].

3.5 Acceleration methods

The basic reconstruction algorithm in Equation 11 is applied per EPI of a given LF. However, there are correlations between neighboring EPIs as well as between different color components in the same EPI, which can be further explored to achieve an accelerated processing [54].

Colorization Colorization uses a grayscale image, which guides the reconstruction of another image where the color information is available in isolated regions only [40], [39]. The missing color information is recovered and propagated using the local structure of the guiding image [39]. Colorization is attractive and computationally efficient approach for DSEPI reconstruction, considering the luminance channel as the appropriate guidance map for the two chrominance channels [54].

Let E denotes the given luminance channel of DSEPI, which is fully reconstructed. The unknown color image x is modelled as a linear function of the known guidance map at every pixel within a small spatial window w,

$$x[i] \approx aE[i] + b, \forall i \in w.$$

Finding the unknowns is performed through a cost function minimization

$$\min_{x,a,b} J(x,a,b) = \sum_j \left(\sum_{i \in w_j} (x[i] - a_j E[i] - b_j)^2 + \varepsilon a_j^2 \right),$$

where a regularization coefficient ε ensures the numerical stability. The minimization problem can be reformulated in terms of *matting Laplacian* Λ matrix such as

$$\min_{a,b} J(x,a,b) = x^\intercal \Lambda x,$$

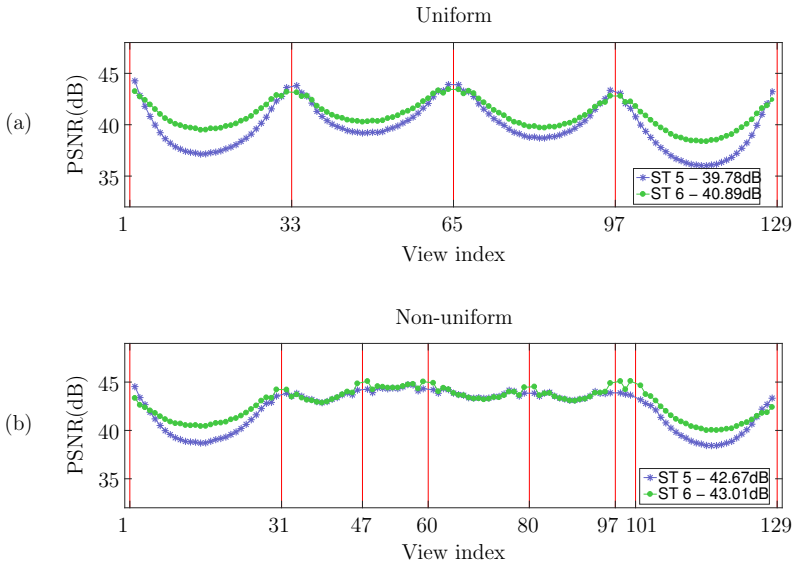


Fig. 12: Comparison of the DSLF reconstruction performance between uniform (a) and non-uniform (b) sampling. Used *ST 5* and *ST 6* methods correspond to the used $J = 5$ and $J = 6$ number of scales in the shearlet transform.

where the symmetric matrix Λ depends on E and w only [39]. The proper choice of the local window w plays a crucial role for the algorithm performance. Figure 14 (a) illustrates possible windowing for DSEPI reconstruction, where the window shape has been motivated in [54]. With structure and local windowing information represented in Λ, the colorization problem can be formulated as constrained quadratic minimization,

$$\min_x x^\intercal \Lambda x, \quad \text{subject to } Mx = y, \tag{12}$$

where M is the diagonal measuring matrix and y contains the available color information. The problem can be solved by e.g. the conjugate gradient method $(\Lambda + \lambda M)x = \lambda My$ with sufficiently high λ. Figure 15 illustrates the approach in terms of input guidence map and color and output reconstructed (colorized) DSEPI. The quality of the colorization is mainly dependent on the accuracy of the guidance map. Therefore in order to provide an overall high quality accelerated reconstruction it is required to efficiently distribute the processing resources between reconstructing the luminance channel Y and colorizing the RGB color channels. As seen in Figure 16, same or better quality of reconstruction can be achieved for less time. Alternatively, the three decorelated channels YUV can be reconstructed independently using the shearlet-based sparsification, with higher priority assigned to the Y channel reconstruction. As a rule of thumb, it should be processed with twice more iterations than the U and V channels to get decent reconstruction results.

Fig. 13: Semi-transparent DSEPI reconstruction using the proposed method and SGBM [28].

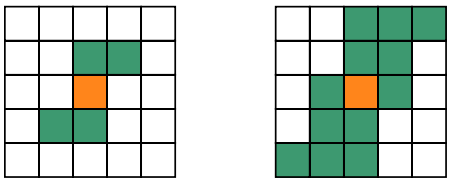


Fig. 14: (a) Proposed w window (green) for modelling guidance map. (b) Neighbourhood (green) for forming matting Laplacian matrix entry with respect to reference pixel (orange).

Decorrelation Transform Another acceleration can be considered by exploiting spatial correlation between neighboring EPIs of the same light field. The winning idea is to use one-dimensional wavelet transform along the vertical direction (i.e. between EPIs), as illustrated in Figure 17 [54]. In this way, the imagery is decomposed in coarse (low-pass) component and detail components. The coarse DSEPI approximation is reconstructed first: per-EPI and then by inverse 1D wavelet transform to get a global low-pass approximation which serves as an initial estimate for the sparsity-based reconstruction. This approach strongly depends on the spatial image structure. For scenes containing objects with vertically uniform color, the wavelet based processing allows to

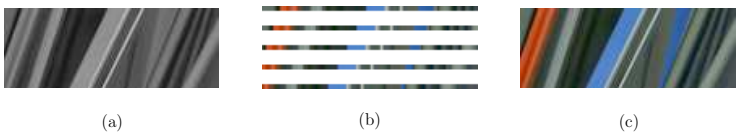


Fig. 15: (a) Grayscale DSEPI obtained by luminance channel reconstruction using shearlet transform used as a guide for colorization. (b) Color information of the DSEPI is available only from input coarse set of views. (c) Colorization result obtained by solving problem Equation 12.

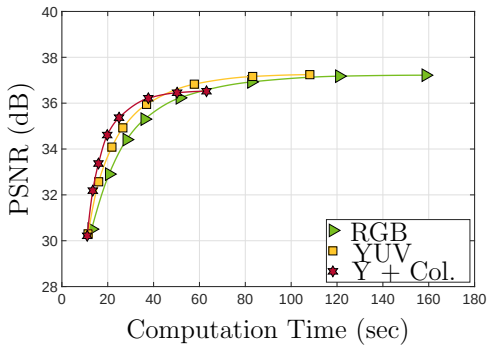


Fig. 16: Average performance over multiple datasets of the colorization technique (Y+Col.) compared with reference reconstruction method applied on every channel independent (RGB) and efficiently reconstruction in YUV color space (YUV).

significantly decrease the computation time while for scenes with more complex structures in vertical direction, the method shows no significant acceleration [54].

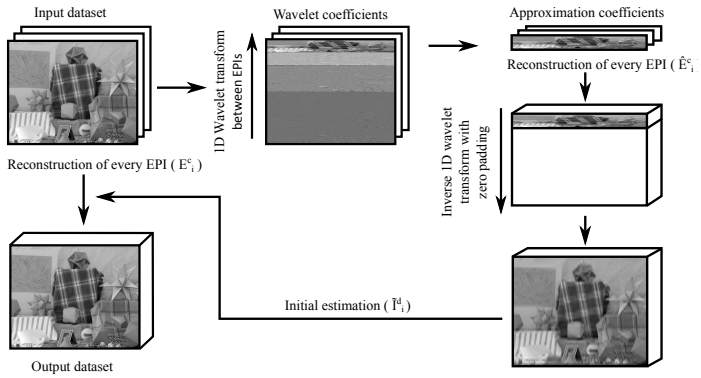


Fig. 17: Reconstruction flowchart using wavelet transform approximation (lowpass) coefficients as an initial estimation for original set reconstruction.

4 Applications

Densely sampled light field has a number of applications, where a dense set of views or rays is required. Here, we present a few applications, to illustrate its use.

4.1 Holographic stereograms

Holographic stereograms are printed digital holograms, where small holographic elements (so-called hogels) act as multi-view pixels when illuminated with proper light [48]. In holographic stereograms, each ray of the LF is considered source of a windowed plane wave with the corresponding amplitude and the entire LF is formed as a superposition of plane waves. For convenience, the two planes of the LF parameterization are located at the camera plane and the hologram plane. Thus, all rays intersecting a point on the hologram plane form a hogel. A fringe pattern corresponding to a hogel on the holographic stereogram is calculated using the superposition of the plane waves generated by the rays in the hogel. The resolution of each hogel is directly related to the angular resolution of the given LF and high angular resolution is needed for an accurate calculation of the fringe pattern corresponding to each hogel. All fringe patterns together form the entire holographic stereogram.

The method described in Section 3 has been used to generate synthetic holographic stereograms and to compare with depth-based view synthesis methods [48]. The presented results have demonstrated the importance of using DSLF for the holographic stereogram calculation and the efficiency of the proposed shearlet based algorithm for obtaining DSLF, which performed better compared to the depth-based approaches [48].

4.2 Light field compression

Typically, the LF compression problem is interpreted as compression of the corresponding sub-aperture views. By using an enhanced inter-view prediction scheme, significant improvement in compression can be achieved. DSLF reconstruction can be used for predicting views [3]. In this approach, the given LF is uniformly decimated in the angular domain first to form a set of key views. This is aimed at decreasing the number of images which go to the compression engine. The key views are converted further into a pseudo video sequence and compressed using high-efficiency video coding (HEVC) encoder. At the decoding side, the full LF is obtained by decoding the key views and then using them for DSLF reconstruction. Apparently, the interplay between view decimation and encoding parameters is the key performance factor [3]. As an anchor method, the direct encoding of the full set of the image from LF as a pseudo video sequence has been considered. The obtained results have demonstrated the efficiency of the proposed compression scheme especially in low bit-rates compared to the anchor method. Since the reconstruction method based on the shearlet transform relies only on keys views, in a low bit-rate scenario, the bit budget allows to achieve high quality for the key views and, as a consequence, high quality for the reconstructed DSLF. On the other hand, the anchor achieves effective compression in high bit-rates due to its various prediction modes for highly correlated views and handling residual information in an effective manner.

References

1. Adelson, E.H., Bergen, J.R.: The plenoptic function and the elements of early vision. In: Computational Models of Visual Processing. pp. 3–20 (1991)

2. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing **54**(11), 4311–4322 (12 2006)
3. Ahmad, W., Vagharshakyan, S., Sjöström, M., Gotchev, A., Bregovic, R., Olsson, R.: Shearlet transform based prediction scheme for light field compression. In: 2018 Data Compression Conference. pp. 396–396 (Mar 2018)
4. Alperovich, A., Johannsen, O., Strecke, M., Goldluecke, B.: Light field intrinsics with a deep encoder-decoder network. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Jun 2018)
5. Blumensath, T., Davies, M.E.: Normalized iterative hard thresholding: Guaranteed stability and performance. IEEE Journal of Selected Topics in Signal Processing **4**(2), 298–309 (Apr 2010)
6. Bolles, R.C., Baker, H.H., Marimont, D.H.: Epipolar-plane image analysis: An approach to determining structure from motion. International Journal of Computer Vision **1**(1), 7–55 (Mar 1987)
7. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. SIAM Journal on Imaging Sciences **3**(3), 492–526 (Jan 2010)
8. Cai, J.F., Chan, R.H., Shen, Z.: A framelet-based image inpainting algorithm. Applied and Computational Harmonic Analysis **24**(2), 131–149 (Mar 2008)
9. Cai, J.F., Shen, Z.: Framelet based deconvolution. Journal of Computational Mathematics **28**(3), 289–308 (2010)
10. Camahort, E., Lerios, A., Fussell, D.: Uniformly sampled light fields. In: Rendering Techniques '98. pp. 117–130 (1998)
11. Candès, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory **52**(2), 489–509 (Feb 2006)
12. Candes, E.J., Tao, T.: Decoding by linear programming. IEEE Transactions on Information Theory **51**(12), 4203–4215 (Dec 2005)
13. Candès, E.J., Donoho, D.L.: Ridgelets: A key to higher-dimensional intermittency? Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences **357**(1760), 2495–2509 (1999)
14. Candès, E.J., Donoho, D.L.: New tight frames of curvelets and optimal representations of objects with piecewise c^2 singularities. Communications on Pure and Applied Mathematics **57**(2), 219–266 (2004)
15. Cha Zhang, Tsuhan Chen: Spectral analysis for sampling image-based rendering data. IEEE Transactions on Circuits and Systems for Video Technology **13**(11), 1038–1050 (Nov 2003)
16. Chai, J.X., Tong, X., Chan, S.C., Shum, H.Y.: Plenoptic sampling. In: 27th Annual Conference on Computer Graphics and Interactive Techniques. pp. 307–318. SIGGRAPH '00 (2000)
17. Chan, S.H., Wang, X., Elgendy, O.A.: Plug-and-play admm for image restoration: Fixed-point convergence and applications. IEEE Transactions on Computational Imaging **3**(1), 84–98 (Mar 2017)
18. Do, M.N., Marchand-Maillet, D., Vetterli, M.: On the bandwidth of the plenoptic function. IEEE Transactions on Image Processing **21**(2), 708–717 (Feb 2012)
19. Do, M.N., Vetterli, M.: The contourlet transform: an efficient directional multiresolution image representation. IEEE Transactions on Image Processing **14**(12), 2091–2106 (Dec 2005)
20. Genzel, M., Kutyniok, G.: Asymptotic analysis of inpainting via universal shearlet systems. SIAM Journal on Imaging Sciences **7**(4), 2301–2339 (2014)
21. Georgiev, T., Intwala, C.: Light field camera design for integral view photography. Adobe Technical Report (2006)

22. Gilliam, C., Dragotti, P., Brookes, M.: On the spectrum of the plenoptic function. IEEE Transactions on Image Processing **23**(2), 502–516 (Feb 2014)
23. Gilliam, C., Dragotti, P.L., Brookes, M.: A closed-form expression for the bandwidth of the plenoptic function under finite field of view constraints. In: IEEE International Conference on Image Processing (ICIP). pp. 3965–3968 (Sep 2010)
24. Gortler, S.J., Grzeszczuk, R., Szeliski, R., Cohen, M.F.: The lumigraph. In: 23rd Annual Conference on Computer Graphics and Interactive Techniques. pp. 43–54. SIGGRAPH '96 (1996)
25. Heber, S., Pock, T.: Convolutional networks for shape from light field. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3746–3754 (Jun 2016)
26. Heber, S., Yu, W., Pock, T.: Neural EPI-Volume networks for shape from light field. In: IEEE International Conference on Computer Vision (ICCV). pp. 2271–2279 (Oct 2017)
27. Heyden, A., Pollefeys, M.: Multiple view geometry. Emerging topics in computer vision pp. 45–107 (2005)
28. Hirschmuller, H.: Stereo processing by semiglobal matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(2), 328–341 (Feb 2008)
29. Honauer, K., Johannsen, O., Kondermann, D., Goldluecke, B.: A dataset and evaluation methodology for depth estimation on 4D light fields. In: Asian Conference on Computer Vision (2016)
30. Hosni, A., Rhemann, C., Bleyer, M., Rother, C., Gelautz, M.: Fast cost-volume filtering for visual correspondence and beyond. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(2), 504–511 (Feb 2013)
31. Insung Ihm, Sanghoon Park, Rae Kyoung Lee: Rendering of spherical light fields. In: Fifth Pacific Conference on Computer Graphics and Applications. pp. 59–68 (Oct 1997)
32. Jeon, H., Park, J., Choe, G., Park, J., Bok, Y., Tai, Y., Kweon, I.S.: Accurate depth map estimation from a lenslet light field camera. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1547–1555 (Jun 2015)
33. Kalantari, N.K., Wang, T.C., Ramamoorthi, R.: Learning-based view synthesis for light field cameras. ACM Transactions on Graphics (TOG) **35**(6), 1–10 (Nov 2016)
34. Kim, C., Zimmer, H., Pritch, Y., Sorkine-Hornung, A., Gross, M.: Scene reconstruction from high spatio-angular resolution light fields. ACM Transactions on Graphics (TOG) **32**(4), 1–12 (Jul 2013)
35. Krishnan, D., Fergus, R.: Fast image deconvolution using hyper-laplacian priors. In: Advances in Neural Information Processing Systems 22, pp. 1033–1041 (2009)
36. Kutyniok, G., Lemvig, J., Lim, W.Q.: Shearlets: Multiscale analysis for multivariate data (2012)
37. Kutyniok, G., Lim, W.Q.: Compactly supported shearlets are optimally sparse. Journal of Approximation Theory **163**(11), 1564–1589 (2011)
38. Lee, H., Battle, A., Raina, R., Ng, A.Y.: Efficient sparse coding algorithms. In: Proceedings of the 19th International Conference on Neural Information Processing Systems. pp. 801–808. NIPS'06, MIT Press (2006)
39. Levin, A., Lischinski, D., Weiss, Y.: A closed-form solution to natural image matting. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(2), 228–242 (Feb 2008)
40. Levin, A., Lischinski, D., Weiss, Y.: Colorization using optimization. ACM Transactions on Graphics (TOG) **23**(3), 689–694 (Aug 2004)
41. Levoy, M., Hanrahan, P.: Light field rendering. In: 23rd Annual Conference on Computer Graphics and Interactive Techniques. pp. 31–42 (1996)
42. Lim, W.Q.: Nonseparable shearlet transform. IEEE Transactions on Image Processing **22**(5), 2056–2065 (May 2013)
43. Lin, Z., Shum, H.Y.: A geometric analysis of light field rendering. International Journal of Computer Vision **58**(2), 121–138 (Jul 2004)

44. Mallat, S.: A Wavelet Tour of Signal Processing: The Sparse Way. Academic Press (2008)
45. McMillan, L., Bishop, G.: Plenoptic modeling: An image-based rendering system. In: 22nd Annual Conference on Computer Graphics and Interactive Techniques. pp. 39–46. SIGGRAPH '95 (1995)
46. Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., Hanrahan, P., et al.: Light field photography with a hand-held plenoptic camera. Computer Science Technical Report (CSTR) **2**(11), 1–11 (2005)
47. Overbeck, R.S., Erickson, D., Evangelakos, D., Pharr, M., Debevec, P.: A system for acquiring, processing, and rendering panoramic light field stills for virtual reality. ACM Transactions on Graphics (TOG) **37**, 1–15 (Dec 2018)
48. Sahin, E., Vagharshakyan, S., Mäkinen, J., Bregovic, R., Gotchev, A.: Shearlet-domain light field reconstruction for holographic stereogram generation. In: IEEE International Conference on Image Processing (ICIP). pp. 1479–1483 (2016)
49. Shen, Z., Toh, K., Yun, S.: An accelerated proximal gradient algorithm for frame-based image restoration via the balanced approach. SIAM Journal on Imaging Sciences **4**(2), 573–596 (Jan 2011)
50. Shi, L., Hassanieh, H., Davis, A., Katabi, D., Durand, F.: Light field reconstruction using sparsity in the continuous fourier domain. ACM Transactions on Graphics (TOG) **34**(1), 1–13 (Dec 2014)
51. Shin, C., Jeon, H., Yoon, Y., Kweon, I.S., Kim, S.J.: EPINET: A fully-convolutional neural network using epipolar geometry for depth from light field images. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4748–4757 (Jun 2018)
52. Shum, H.Y., He, L.W.: Rendering with concentric mosaics. In: 26th Annual Conference on Computer Graphics and Interactive Techniques. pp. 299–306. SIGGRAPH '99 (1999)
53. Tao, M.W., Hadap, S., Malik, J., Ramamoorthi, R.: Depth from combining defocus and correspondence using light-field cameras. In: IEEE International Conference on Computer Vision (ICCV). pp. 673–680 (Dec 2013)
54. Vagharshakyan, S., Bregovic, R., Gotchev, A.: Accelerated shearlet-domain light field reconstruction. IEEE Journal of Selected Topics in Signal Processing **11**(7), 1082–1091 (Oct 2017)
55. Vagharshakyan, S., Bregovic, R., Gotchev, A.: Light field reconstruction using shearlet transform. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(1), 133–147 (Jan 2018)
56. Vagharshakyan, S.: Densely-sampled light field reconstruction. Ph.D. thesis, Tampere University (2020)
57. Vaish, V., Adams, A.: The (new) stanford light field archive. http://lightfield.stanford.edu (2008)
58. Wanner, S., Goldluecke, B.: Variational light field analysis for disparity estimation and super-resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence **36**(3), 606–619 (Mar 2014)
59. Wen, Z., Goldfarb, D., Yin, W.: Alternating direction augmented lagrangian methods for semidefinite programming. Mathematical Programming Computation **2**(3-4), 203–230 (2010)
60. Wilburn, B., Joshi, N., Vaish, V., Talvala, E.V., Antunez, E., Barth, A., Adams, A., Horowitz, M., Levoy, M.: High performance imaging using large camera arrays. ACM Transactions on Graphics (TOG) **24**(3), 765–776 (Jul 2005)
61. Wilburn, B.S., Smulski, M., Lee, H.H.K., Horowitz, M.A.: Light field video camera. In: Media Processors 2002. vol. 4674, pp. 29–37 (2001)
62. Wood, D.N., Azuma, D.I., Aldinger, K., Curless, B., Duchamp, T., Salesin, D.H., Stuetzle, W.: Surface light fields for 3D photography. In: 27th Annual Conference on Computer Graphics and Interactive Techniques. pp. 287–296. SIGGRAPH '00 (2000)

63. Wu, G., Liu, Y., Fang, L., Dai, Q., Chai, T.: Light field reconstruction using convolutional network on epi and extended applications. IEEE Transactions on Pattern Analysis and Machine Intelligence (2018)
64. Wu, G., Zhao, M., Wang, L., Dai, Q., Chai, T., Liu, Y.: Light field reconstruction using deep convolutional network on epi. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1638–1646 (Jul 2017)
65. Yu, Z., Guo, X., Ling, H., Lumsdaine, A., Yu, J.: Line assisted light field triangulation and stereo matching. In: IEEE International Conference on Computer Vision (ICCV). pp. 2792–2799 (Dec 2013)