

Optimization of light field display-camera configuration based on display properties in spectral domain

Robert Bregović,^{1,*} Péter Tamás Kovács,^{1,2} and Atanas Gotchev¹

¹Department of Signal Processing, Tampere University of Technology, P.O. Box 553, FIN-33101 Tampere, Finland

²Holografika, P.O. Box 100, H-1704 Budapest, Hungary

*robert.bregovic@tut.fi

Abstract: The visualization capability of a light field display is uniquely determined by its angular and spatial resolution referred to as display passband. In this paper we use a multidimensional sampling model for describing the display-camera channel. Based on the model, for a given display passband, we propose a methodology for determining the optimal distribution of ray generators in a projection-based light field display. We also discuss the required camera setup that can provide data with the necessary amount of details for such display that maximizes the visual quality and minimizes the amount of data.

©2016 Optical Society of America

OCIS codes: (100.6890) Three-dimensional image processing; (120.2040) Displays; (110.6880) Three-dimensional image acquisition; (100.2000) Digital image processing; (100.0100) Image processing.

References and links

1. A. Boev, R. Bregović, and A. Gotchev, "Signal processing for stereoscopic and multi-view 3D displays," in *Handbook of Signal Processing Systems, 2nd edition*, S. Bhattacharyya, E. Depretere, R. Leupers, and J. Takala, eds. (Springer, 2013).
2. W.A. IJsselstein, P.J.H. Seuntiëns, and L.M.J. Meesters, "Human factors of 3D displays," in *3D Video Communication*, O. Schreer, P. Kauff, and T. Sikora, eds. (Wiley, 2005).
3. T. Balogh, "The HoloVizio system," *Proc. SPIE* 6055, 12 pages (2006).
4. J.H. Lee, J. Park, D. Nam, S.Y. Choi, D.S. Park, and C.Y. Kim, "Optimal projector configuration design for 300-Mpixels multi-projection 3D display," *Opt. Express* **21**(22), 26820–26835 (2013).
5. E. Adelson and J. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*, M. Landy and J.A. Movshon, eds. (MIT, 1991).
6. A. Stern, Y. Yitzhaky, and B. Javidi, "Perceivable light fields: Matching the requirements between the human visual system and autostereoscopic 3-D displays," *Proceedings of the IEEE* **102**(10), 1571-1587 (2014).
7. S.A. Benton and V.M. Bove, *Holographic Imaging*, (Wiley, 2008).
8. R. Bregović, P.T. Kovács, T. Balogh, and A. Gotchev, "Display-specific light-field analysis," *Proc. SPIE* 9117, 15 pages (2014).
9. S.J. Gortler, R. Grzeszczuk, R. Szeliski, and M.F. Cohen, "The lumigraph," *SIGGRAPH (Computer Graphics)*, 43-54 (1996).
10. M. Levoy and P. Hanrahan, "Light field rendering," *SIGGRAPH (Computer Graphics)*, 31-42 (1996).
11. C.K. Liang, Y.C. Shih, and H.H. Chen, "Light field analysis for modeling image formation," *IEEE Trans. Image Processing* **20**(2), 446-460 (2011).
12. C. Zhang and T. Chen, "Spectral analysis for sampling image-based rendering data," *IEEE Trans Circuits and Systems for Video Technology* **13**(11), 1038-1050 (2003).
13. E. Dubois, "The sampling and reconstruction of time-varying imagery with application in video systems," *Proc. IEEE* **73**, 502–522 (1985).
14. E. Dubois, "Video sampling and interpolation," in *The Essential Guide to Video Processing*, J. Bovik, ed. (Academic Press, 2009).
15. P.Q. Nguyen and D. Stehlé, "Low-dimensional lattice basis reduction revisited," *ACM Trans. Algorithms* **5**(4), 46 pages (2009).
16. F. Aurenhammer, "Voronoi diagrams – A survey of a fundamental geometric data structure," *ACM Computing Surveys* **23**, 245-405 (1991).

17. E.B. Tadmor and R.E. Miller, *Modeling Materials: Continuum, Atomistic and Multiscale Techniques*, (Cambridge University, 2011).
 18. X. Cao, Z. Geng, and T. Li, "Dictionary-based light field acquisition using sparse camera array," *Opt. Express* **22**(20), 24081-24095 (2014).
 19. M. Zwicker, W. Matusik, F. Durand, and H. Pfister, "Antialiasing for automultiscopic 3D displays", *Proc. of Eurographics Symposium on Rendering*, 10 pages (2006).
 20. A. Boev, R. Bregović, and A. Gotchev, "Methodology for design of antialiasing filters for autostereoscopic displays," *IET Signal Processing* **5**, 333-343 (2011).
 21. N. Holliman, N. Dodgson, G. Favalora, and L. Pockett, "Three-dimensional displays: A review and application analysis," *IEEE Trans. Broadcasting*, **57**(2), 362-371 (2011).
 22. Y. Takaki and N. Nago, "Multi-projection of lenticular displays to construct a 256-view super multi-view display," *Opt. Express* **18**(9), 8824-8835 (2010).
 23. Y. Takaki, Y. Urano, S. Kashiwada, H. Ando, and K. Nakamura "Super multi-view windshield display for long-distance image information presentation," *Opt. Express* **19**(2), 704-716 (2011).
 24. G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, "Tensor displays: Compressive light field synthesis using multilayer displays with directional backlighting," *ACM Trans. Graph.* **31**(4), 11 pages (2012).
 25. D. Kane, P. Guan, and M. Banks, "The limits of human stereopsis in space and time", *Journal of Neuroscience* **34**(4), 1397-1408 (2014).
-

1. Introduction

Most of the commercially available, stereoscopic as well as autostereoscopic, 3D displays concentrate on reproducing the binocular visual cue for single or multiple observers thereby giving the illusion of 3D [1]. However, a typical consumer display is not capable or has difficulty in reproducing other cues important for 3D vision, most notable one being the continuous head parallax [2]. There are two practical ways for achieving the illusion of continuous head parallax. First way is by performing user's eye tracking and rendering parallax-correct views depending on user's position. This can be achieved by either using a head mounted display (e.g. Oculus Rift, Samsung Gear VR, Zeiss VR) or a custom built display with eye-tracking capabilities (e.g. zSpace). Disadvantage of those lies in the fact that, typically, only one user is supported. Second way of achieving a reasonably convincing continuous parallax is by using so called light field (LF) displays [1,3,4].

A LF display strives to reproduce the underlying plenoptic function describing the scene that is visualized [5]. It can be observed by multiple users simultaneously without a need of user tracking or glasses. In order to support continuous parallax, a large and dense set of light rays have to be generated to reconstruct the underlying LF function. In today's LF displays this is achieved by using projection-based systems [3,4]. There are two major drawbacks of such LF displays. First, only a finite number of light rays can be generated in practice. Based on the properties of the human visual system (HVS) it is possible to estimate the optimal (required) number of rays needed to achieve a level of detail that is sufficient for a human observer [6,7]. Unfortunately, achieving that level of detail is impractical with today's technology. Second, due to the multiple sources of rays, it is very difficult to achieve the desired uniform density of rays (position wise as well as intensity wise) on the screen surface [4]. Both drawbacks reduce the perceived resolution of the display. Therefore it is important to optimize the display setup and properly preprocess data sent to the display in order to mitigate the aforementioned two drawbacks as much as possible.

We have shown earlier [8] that by performing a frequency domain analysis of a typical LF display it is possible to determine the throughput of the display in terms of its spatial and angular resolution. This enables one to calculate the optimal amount of data that has to be captured and sent to the display to maximally utilize its visual capability. Moreover, it gives a user a good idea what to expect from the display in terms of visual quality. In this paper, we build on some of the ideas presented in [8] in order to achieve a deeper understanding of the relations of various hardware and software parts building a LF display. We present an analysis assuming a desired ray-sampling pattern at the screen plane that will define display specifications. For such display, we estimate the throughput of the display in terms of its

angular-spatial bandwidth. Having the display specifications, we develop a methodology for determining the optimal distribution of ray generators that will result in the desired display properties as well as a camera setup that can provide data with required amount of details. This is achieved by developing an optimization / estimation method for determining the required display / camera parameters.

Outline of this paper is as follows: Section 2 introduces the LF concepts and notations followed by the description of the principle of operation and properties in spatial and frequency domain of projection based LF displays. The proposed display-camera system optimization is introduced in Section 3, with several examples given in Section 4. Finally, concluding remarks are given in Section 5.

2. Light field displays

2.1. Light field basics

In the most general case, by using ray-optics assumptions, the propagation of light in space can be described by a 7D continuous plenoptic function $R(\theta, \phi, \lambda, \tau, A_x, A_y, A_z)$, where (A_x, A_y, A_z) is a location in the 3D space, (θ, ϕ) are directions (angles) of observation, λ is wavelength, and τ is time [5]. For practical reasons, the continuous plenoptic function is typically simplified to its 4D version, which describes the static and monochromatic light ray propagation in half space. This 4D approximation of the plenoptic function is referred to as LF [10]. In this approximation, the LF ray positions are indexed either by their Cartesian coordinates on two parallel planes, the so-called two-plane parameterization $L(x, y, s, t)$ or by their one plane and direction coordinates $L(x, y, \phi, \theta)$ [9,10].

In this paper, without loss of generality and in line with today's display technology, we concentrate on the so-called horizontal parallax only (HPO) case, ignoring the vertical parallax and subsequently dropping variables t or θ in the aforementioned parameterization. Furthermore, we assume that the relation between planes parameterized by (x, s) and (x, ϕ) , is given by $s = \tan \phi$ with x being the same in both representations. In this parameterization, the origin of the s axis is relative to the given x coordinate.

The position of two parallel planes x and s can be chosen depending on the application. Two such positions, where the distances between parameterizing planes are taken as unit, are given in Fig 1. According to the figure, the propagation of light rays through space can be mathematically expressed as [8,11]

$$L_2 \begin{pmatrix} x_2 \\ s_2 \end{pmatrix} = L_1 \begin{pmatrix} x_1 \\ s_1 \end{pmatrix} = L_1 \begin{pmatrix} 1 & -d \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_2 \\ s_2 \end{pmatrix} \quad (1)$$

$$L_2 \begin{pmatrix} x_2 \\ \phi_2 \end{pmatrix} = L_1 \begin{pmatrix} x_1 \\ \phi_1 \end{pmatrix} = L_1 \begin{pmatrix} x_2 - d \tan \phi_2 \\ \phi_2 \end{pmatrix} \quad (2)$$

with L_1 and L_2 referring to LFs on plane position 1 and plane position 2, respectively, and d being the distance between the plane positions along the z axis. As can be seen from Eq. (2), when considering propagation of light rays in plane and direction representation, the relation between parameters on both planes is not strictly linear. However, for small angles, this nonlinearity can be ignored. More detailed evaluation on light ray propagation can be found in [8,11].

The continuous LF function has to be sampled in a way, which allows its reconstruction from samples. The plenoptic sampling theory, that considers the LF as a multidimensional bandlimited function has been developed in [12]. In general, it states that LF frequency support depends on the min and max depth of the visual scene, and sampling along x and s creates the usual replication of the baseband, which should be taken into account when designing the end-to-end LF camera to display system. While the sampling physically occurs

at the LF acquisition (sensing) stage, it is the LF display, which recreates the LF originating from a visual 3D scene. In the sampling theory formalism, an LF display can be considered as a discrete-to-continuous (D/C) converter that converts a sampled LF into its continuous version, thereby achieving a continuous visualization of a 3D scene, with continuous parallax being part of it. Consequently, we can consider a LF display as a multidimensional sampling system and as such apply multidimensional sampling theory when analyzing LF displays.

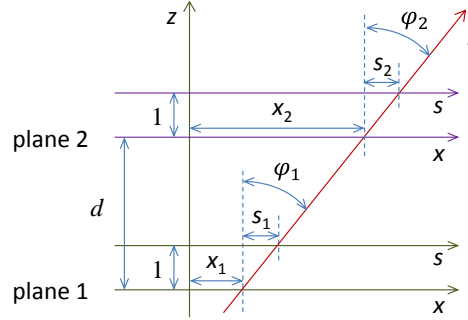


Fig. 1. Light ray (r) propagation through space (representation on two planes).

2.2. Light field display as sampling-reconstruction system

In our general model, we consider the LF display being composed by a set of ray generators and a continuous LF reconstruction optical module. The ray generators act as discrete sources of light rays and the module is the D/C converter that converts the set of samples (rays) into its continuous representation that is observed by a viewer. While different display settings can fall into this general model, we specifically concentrate on a LF display consisting of a set of projection engines and a special screen, dubbed as holographic screen as illustrated in Fig. 2(a). Each light ray generated by a ray generator, hits this screen from a different angle at a different position, and the screen converts (diffuses) each light ray into an angular beam around the main direction of the ray. The span of the beam after diffusion is anisotropic with narrow horizontal angle δ_x and wide vertical angle δ_y , as illustrated in Fig. 2(b) [3]. The screen does not have an explicit pixel structure. A finite area on it emits different light rays to different directions. The properties of such screen are described in more detail in [3]. In this paper we will assume that the screen is a perfect D/C converter. In practice it introduces some low-frequency selectivity that additionally smooths the reconstructed LF. However, this can be ignored for the purposes of our work.

From the observer viewpoint, a point (object) in space is reconstructed by the interaction of rays originating from different sources (i.e. coming from different directions). This is illustrated in Fig. 2(a) for two observers and several points in space. Each ray can be traced from its origin (ray generator) to the screen surface and it is uniquely described by its starting position and angle or its starting position and the place it hits the screen surface. This is reminiscent to the two-plane LF parameterization discussed in the previous section.

The overall throughput of the display is directly related to the number of light rays the display can generate. Denser set of rays produces finer spatial and angular details. Technology limitations prevent us from achieving the resolution power of the HVS [6, 7]. Therefore, it is important to take these limits into account when building the display and/or processing the visual data to be represented on it. Frequency domain analysis of the sampled and reconstructed light field is the proper tool for doing this.

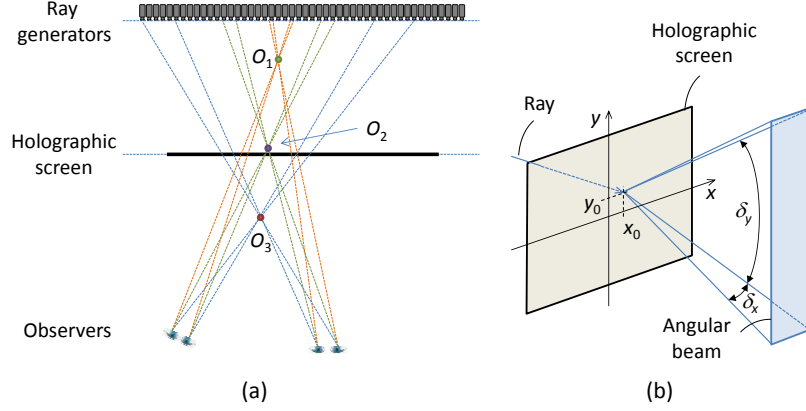


Fig. 2. Projection based light field display – principle of operation. (a) Different ray-generators participate in forming a point in space (O) depending on the place of the point and the position of the observer. (b) The holographic screen acts as diffusor with a narrow spread in horizontal and a wide spread in vertical direction.

2.3. Spatial and frequency domain analysis of light field displays

A typical LF display under consideration is illustrated in Fig. 3. It consists of N_p projection engines uniformly distributed on the ray generators (RG) plane (p - plane) over distance d_p thereby making the distance between engines $x_p = d_p / (N_p - 1)$. Each projection engine generates N_x rays over its field of view FOV_p . We assume that the rays hit a certain plane (screen plane, s - plane) parallel to the RG plane at equidistant points. As a consequence, the angular distribution of rays inside the FOV is not uniform. Nevertheless, for small angles we can assume that this is uniform and approximate the angular resolution at the RG plane as $\alpha_p = FOV_p / N_x$.

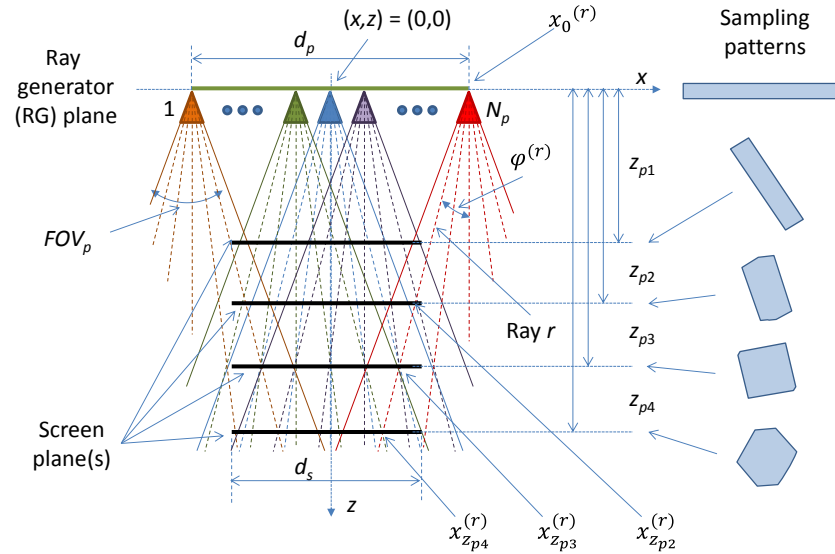


Fig. 3. Ray propagation in a light field display – different sampling patterns are illustrated for different positions of the screen plane.

The 'trajectory' of a ray can be uniquely defined by its origin $x_0^{(r)}$ at the RG plane and its direction determined by angle $\varphi^{(r)}$. The position of the ray at a distance z from the display is given as

$$\begin{bmatrix} x_z^{(r)} \\ \varphi_z^{(r)} \end{bmatrix} = \begin{bmatrix} x_0^{(r)} + z \tan(\varphi^{(r)}) \\ \varphi^{(r)} \end{bmatrix} \quad (3)$$

which is according to the propagation of rays in (x, φ) LF parameterization – see Eq. (2).

The screen of the display is where rays recombine to reconstruct the desired continuous LF function to be observed by a viewer. In Fig. 3, several positions for the screen are illustrated with thick black lines. As seen in the figure, the ray (r) crosses those 'screens' at different horizontal positions ($x_{z_{p2}}^{(r)}, x_{z_{p3}}^{(r)}, x_{z_{p4}}^{(r)}$) and, due to a finite width of the screen d_s , it even does not contribute to the screen at distance z_{p1} . In practice this means that the ray would contribute to a different part of the screen depending on the screen position. Moreover, at different screen positions, it intersects with different rays originating from different ray generators, that is, depending on the screen position, a different combination of rays will be responsible for forming a multiview pixel at that position. As a consequence, the uniform distribution of rays we had at the RG plane is lost.

Rays are indexed (parameterized) by their spatial position and direction (x, φ) and thus represented as samples in the corresponding ray space. This parameterization has been selected among several possibilities (e.g. φ vs. x , $\tan \varphi$ vs. x , $z \tan \varphi$ vs. x) since both ray-space axes can be allocated with measurable (quantifiable) units (position can be expressed in mm and angle in degrees) that are easy to understand by a user. Consequently, at the screen plane, the display can be quantified by its spatial resolution (e.g. number of pixels per mm or per screen size) and its angular resolution (e.g. number of rays per degree or FOV of the display FOV_{disp}).

For the need of frequency analysis, each ray is considered as a sample, positioned in the 2D ray-space plane for fixed z (in the case of full parallax, this turns into a 4D plane). Since the position of the ray is changing along z , as given by Eq. (3), for a set of ray generators, different sampling patterns are obtained at different distances from the screen. This is illustrated by means of an example in Fig. 4 (see also Fig. 3). The figures on the top row for $z=0, z_{p2}, z_{p4}$ show how the whole LF that the display is capable of generating is sheared along the x -axis as the screen plane moves away from the RG plane. For better visualization, one set of rays is marked in blue. The figures in bottom row show zoomed in versions of the LF at different distances from the RG plane. One can observe that for every distance, the sampling pattern is regular although not rectangular. The fact that the sampling patterns are regular, enables us to utilize the multi-dimensional sampling theory [13,14].

Any regular 2D pattern can be uniquely described through a notion of sampling lattice Λ . The elements of the lattice are calculated as a linear combination of two linearly independent vectors

$$\Lambda(\mathbf{V}) = \{n_1 \mathbf{v}_1 + n_2 \mathbf{v}_2 \mid n_1, n_2 \in \mathbb{Z}\} \quad (4)$$

with $\mathbf{v}_k = \begin{bmatrix} v_k^{(x)} & v_k^{(\varphi)} \end{bmatrix}^T$ for $k=1,2$ referred to as basis vectors and T being the transpose operator. The vectors building the lattice can be expressed in matrix form as

$$\mathbf{V}(\mathbf{v}_1, \mathbf{v}_2) = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} v_1^{(x)} & v_2^{(x)} \\ v_1^{(\varphi)} & v_2^{(\varphi)} \end{bmatrix} \quad (5)$$

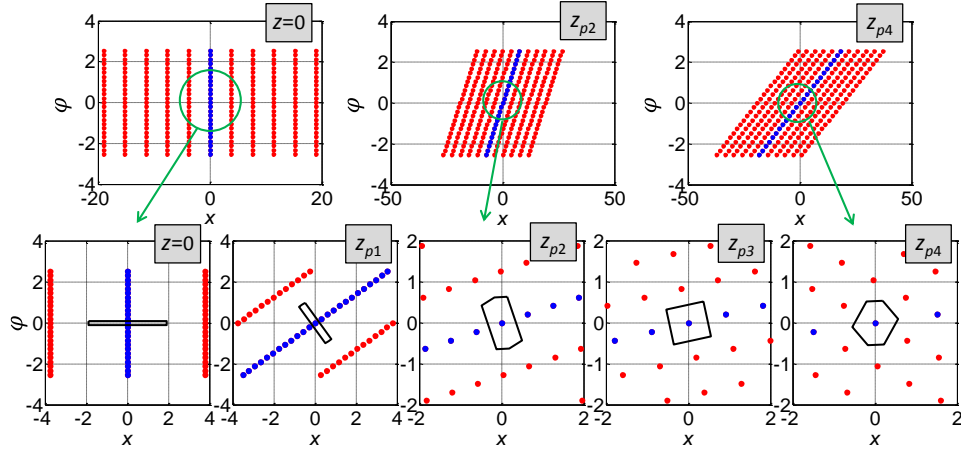


Fig. 4. Light field display – ray space spatial sampling patterns at different distances from the RG plane (c.f. Fig. 3)

with \mathbf{V} being referred to as the sampling matrix. It is important to point out that the sampling matrix is not unique for a given sampling pattern since $\Lambda(\mathbf{V}) = \Lambda(\mathbf{E}\mathbf{V})$ where \mathbf{E} is any integer matrix with $|\det \mathbf{E}| = 1$. Consequently, there are multiple basis vectors describing the same lattice. In practice the set of basic vectors with minimum length (norm) is preferred. Therefore, given a set of basis vectors $(\mathbf{v}_1, \mathbf{v}_2)$, one should find a pair of vectors $(\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2)$ such that $\Lambda(\mathbf{V}) = \Lambda(\tilde{\mathbf{V}})$. Here, tilde denotes the sampling matrix with minimized basis vectors (length $\|\mathbf{v}_1\| + \|\mathbf{v}_2\|$ is minimized) – see Fig. 5(a). The problem of finding such vectors is known in literature as the lattice basis reduction problem [15]. The solution applicable to our 2D case can be obtained using the following Lagrange's algorithm applied to a pair of basis vectors $(\mathbf{v}_1, \mathbf{v}_2)$:

$$\begin{aligned}
 &\text{do} \\
 &\quad \text{if } \|\mathbf{v}_1\| > \|\mathbf{v}_2\| \text{ then swap } \mathbf{v}_1 \text{ and } \mathbf{v}_2 \\
 &\quad \mathbf{v}_2 = \mathbf{v}_2 - \left\lfloor \frac{\langle \mathbf{v}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2} \right\rfloor \mathbf{v}_1 \\
 &\quad \text{until } \|\mathbf{v}_1\| \leq \|\mathbf{v}_2\|
 \end{aligned} \tag{6}$$

The pair of vectors $(\mathbf{v}_1, \mathbf{v}_2)$ resulting from the algorithm are with the smallest norm for the given sampling pattern, that is, $(\mathbf{v}_1, \mathbf{v}_2) \equiv (\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2)$.

For a regular grid described with a lattice Λ , one can also define a unit cell P that is a set in \mathbb{R}^2 such that the union of all cells centered on each lattice point covers the whole sampling space without overlapping or leaving empty space. Similar to the basis vectors, the unit cell is not unique, as illustrated in Fig. 6. The figure illustrates three possibilities out of an infinite set of valid unit cells describing the same lattice. The shapes become even stranger if the underlying sampling pattern is not rectangular.

In this paper we use the Voronoi cell as the unit cell representing a given sampling pattern [16]. As illustrated in Fig. 5(b), the Voronoi cell, denoted by P (green shaded area in the

figure), is a set in \mathbb{R}^2 such that all elements of the set are closer, based on Euclidean distance, to the one lattice point that is inside the cell than to any other lattice point – this makes it the most compact unit cell. In the literature, Voronoi cells are also known as Wigner-Seitz cell – e.g. in solid-state physics [17 p.122]. By using the minimum length basis vectors, the construction of the Voronoi cell is straightforward and is illustrated in Fig. 5(b) (in Fig. 6 the Voronoi cell is the one shown by the leftmost example).

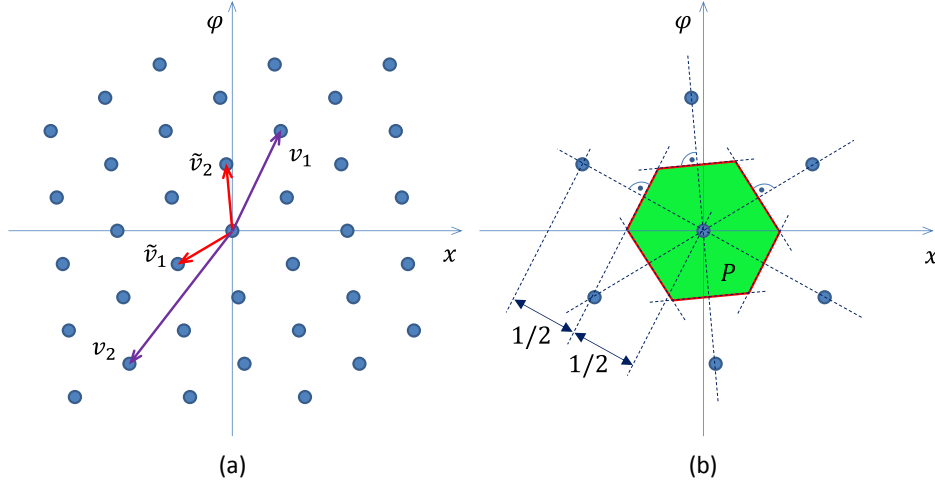


Fig. 5. (a) Example of two sets of basis vectors, (v_1, v_2) and $(\tilde{v}_1, \tilde{v}_2)$, describing the same lattice. (b) Illustration of Voronoi cell construction.

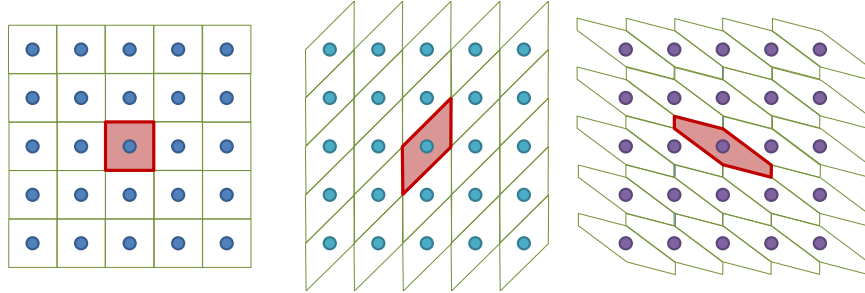


Fig. 6. Some possible unit cells P (shaded area) for a given lattice Λ (points).

The samples in ray space forming regular sampling patterns at different depths z represent a bandlimited function. In frequency domain, it has periodic structure with multiple replicas of the baseband. The periodicity and at the same time the baseband frequency support is defined through the reciprocal lattice Λ^* , that can be evaluated as [8,14]

$$\Lambda^*(V) = \Lambda\left((V^T)^{-1}\right). \quad (7)$$

There are many possible unit cells for a given lattice Λ^* . Each possible unit cell describes a set of bandlimited functions that can be represented by the sampling pattern and can be reconstructed from a given discrete representation assuming that the reconstruction filter has the shape of the selected unit cell. Furthermore, this also means that an arbitrary continuous

function has to be pre-filtered with a filter aimed at removing all frequency content outside of the selected unit cell in order to prevent aliasing errors during sampling. This can be achieved either by using (if possible) a proper continuous-domain filter before sampling the function or first oversampling the continuous function and then performing filtering and down sampling in the discrete domain. It should be pointed out that in the case under consideration, it might not be possible to perform pre-filtering in the continuous domain since this would require an optical filter in spatial and angular direction. Therefore, in this paper we assume that we oversample the continuous function at the sampling stage and perform all filtering in the discrete domain. If the scene is captured by sparse cameras, the dense (oversampled) LF can be reconstructed by compressive sensing approaches, e.g. [18].

The most compact (isotropic) unit cell for a given sampling pattern is, as in the spatial domain, a Voronoi cell, denoted in this paper as P^* . The importance of this unit cell is twofold. First, it will represent frequency support that treats equally both directions (spatial and angular direction in ray space representation) – this is beneficial from the HVS viewpoint. Second, the screen in the display that will perform the D/C conversion has for practical reasons a 'low-pass' type characteristics (typically it is rectangular with Gaussian type weights [3]) that has to be matched to available ray distribution or vice versa. As such, the Voronoi cell will be the most convenient unit cell to match the screen reconstruction filter.

The Voronoi cell of a sampling pattern can be considered equivalently in spatial or frequency domain. Given its isotropic behavior, it is precisely the quantity, which characterizes the properties of the reconstructed bandlimited function. Therefore, the estimation of the optimal display and camera setup can be done by comparing Voronoi cells formed on the screen plane. From one side, there is the sampling pattern of the rays generated by the display; from another side there is the sampling pattern of the rays as captured by cameras. Both sampling patterns and the respective bandlimited LF are compared for similarity through their Voronoi cells in ray-space domain at the screen plane. This makes the overall optimization procedure computationally less demanding and thus faster. The frequency bandwidth of the system can be easily estimated once the optimal configuration is determined.

3. Light field display–camera configuration optimization

In an ideal case, one would require that a display perfectly reconstructs the underlying plenoptic function or at least up to the level of detail supported by the HVS. With limited resources, one can target the best possible continuous LF approximation out of a given discrete set of rays. In such a case, it is important to determine the optimal display and camera setup that maximizes the visual capabilities of the display.

We tackle the problem in two steps. First, we evaluate the optimal setup of ray generators for a given or desired density of rays at the screen plane. Second, we estimate the bandwidth for such system from the perspective of the scene, that is, what kind of capture setup and pre-processing is required to sense enough data for the given display setup. It should be pointed out that step two can be applied to an arbitrary display setup, as long as the basic setup parameters (ray generators, distances, screen properties, etc.) are available. The complete display-camera setup considered in this paper, with all adjustable parameters, is illustrated in Fig. 7 and is discussed in more detail in the following two sections. To streamline the text in the rest of this paper, we use the notations for various sampling patterns emerging from the display setup as in Fig. 7. Subscripts p , s , and c are used to denote the parameters related to the RG, screen, and camera/viewer plane, respectively, with z increasing in the direction of the observer and $z=0$ being relative to the parameter's origin, e.g. for parameters originating on the screen plane, $z=0$ is on the screen plane. Practical angles are denoted by α in contrast to 'theoretical' angles φ used in the LF parameterization. The estimated and optimized parameters are denoted by hat and bar, respectively, e.g. $\hat{\alpha}_p$ and \bar{x}_p . Finally, tilde is used to denote parameters after the lattice basis reduction operation.

The proposed optimization technique can be extended for other display-camera configurations, than the one shown in Fig. 7, as long as those configurations result in regular sampling patterns in the angular-spatial domain and as such can be described by sampling matrices as illustrated for cases under consideration next.

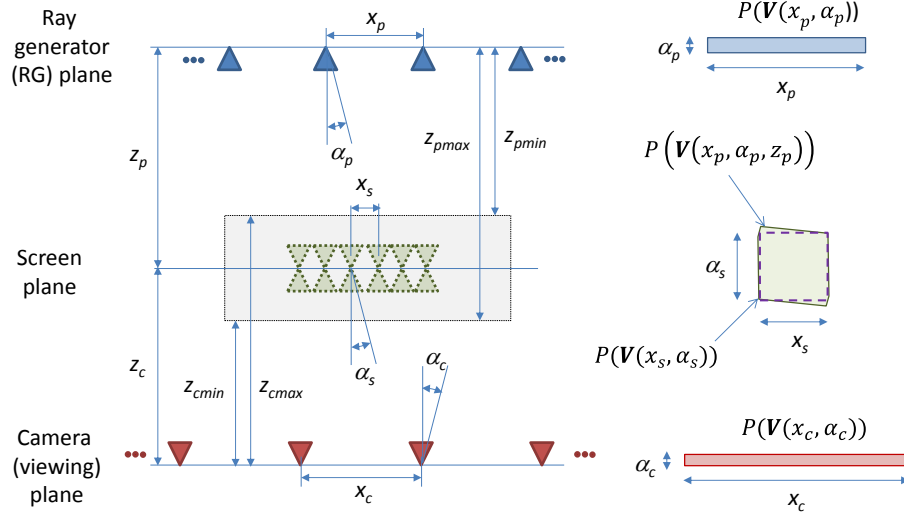


Fig. 7. Light field display-camera setup together with notations for expected sampling patterns.

3.1. Light field display configuration optimization

For the purpose of stating the problem under consideration, we start from the center of Fig. 7, namely, the screen plane. We require that the display should be able to reproduce a LF with a desired bandwidth or, equivalently, a LF with a given density at the screen plane – the density being defined by the spatial and angular resolution. This determines the values (x_s, α_s) in the ray space representation, and in turn, it determines the desired sampling pattern at the screen plane. We assume that the pattern is rectangular – this is realistic assumption due to the properties of the screen and the requirements that both directions (spatial and angular) should be treated in a similar manner. The sampling pattern is uniquely defined through the following sampling matrix:

$$\mathbf{V}(x_s, \alpha_s) = \begin{bmatrix} x_s & 0 \\ 0 & \alpha_s \end{bmatrix}. \quad (8)$$

Having the sampling pattern at the screen plane, the problem is to determine optimal parameters of the ray generators (x_p, α_p) and distance between the RG plane and the screen plane z_p for which the sampling pattern mapped to the screen plane will match the desired one, that is, grids described by sampling lattices $\Lambda(\mathbf{V}(x_p, \alpha_p, z_p))$ and $\Lambda(\mathbf{V}(x_s, \alpha_s))$ should match. With reference to Eq. (7), this will ensure the same Fourier domain bandwidth of the desired LF. Mismatches between the lattices $\Lambda(\mathbf{V}(x_p, \alpha_p, z_p))$ and $\Lambda(\mathbf{V}(x_s, \alpha_s))$ will manifest themselves either as aliasing effects in the reconstructed continuous LF or as inefficient utilization of the display bandwidth. The targeted lattice matching is done by an optimization technique presented below aimed at mitigating the aforementioned two problems.

The ray generators' sampling matrix mapped to screen plane is defined as

$$\mathbf{V}(x_p, \alpha_p, z_p) = \begin{bmatrix} x_p & z_p \tan \alpha_p \\ 0 & \alpha_p \end{bmatrix}. \quad (9)$$

Two sampling patterns will match in the case they have identical unit cells or will be similar if the difference between the unit cells that can be expressed as $\|P(\mathbf{V}(x_p, \alpha_p, z_p)) - P(\mathbf{V}(x_s, \alpha_s))\|$ is small. Here we assume that the similarity criterion is defined via the difference in the area (size) of the unit cells and the difference in the shape of the unit cell. Furthermore, based on the sampling theory [14] a unit cell is uniquely described by its (lattice basis reduced) sampling matrix. Consequently, the similarity measure of two sampling grids (and correspondingly the underlying unit cells) can be expressed through the similarity between the corresponding sampling matrices. In summary, the problem under consideration is, for given (x_s, α_s) , to find (x_p, α_p, z_p) that minimizes δ_p

$$\delta_p = \|\tilde{\mathbf{V}}(x_p, \alpha_p, z_p) - \mathbf{V}(x_s, \alpha_s)\| \quad (10)$$

with $\mathbf{V}(x_s, \alpha_s)$ being the desired sampling matrix at the screen plane and $\tilde{\mathbf{V}}(x_p, \alpha_p, z_p)$ being the lattice basis reduced sampling matrix of the ray generators $\mathbf{V}(x_p, \alpha_p)$ mapped to the screen plane. It should be pointed out that $\mathbf{V}(x_s, \alpha_s) = \tilde{\mathbf{V}}(x_s, \alpha_s)$. Furthermore, when implementing Eq. (10), it should be kept in mind that the reduced matrix is unique up to the sign and sequence of basis vectors, that is, $\Lambda(\mathbf{V}(\mathbf{v}_1, \mathbf{v}_2)) \equiv \Lambda(\mathbf{V}(\pm \mathbf{v}_1, \pm \mathbf{v}_2)) \equiv \Lambda(\mathbf{V}(\pm \mathbf{v}_2, \pm \mathbf{v}_1))$.

The lattice basis reduced sampling matrix of the ray generators mapped to the screen plane can be expressed as

$$\tilde{\mathbf{V}}(x_p, \alpha_p, z_p) = \begin{bmatrix} x_1 & x_2 \\ \alpha_1 & \alpha_2 \end{bmatrix} = \begin{bmatrix} x_s + \Delta x_1 & \Delta x_2 \\ \Delta \alpha_1 & \alpha_s + \Delta \alpha_2 \end{bmatrix} = \begin{bmatrix} x_s & 0 \\ 0 & \alpha_s \end{bmatrix} + \begin{bmatrix} \Delta x_1 & \Delta x_2 \\ \Delta \alpha_1 & \Delta \alpha_2 \end{bmatrix}. \quad (11)$$

In this notation, minimizing the difference between $\mathbf{V}(x_s, \alpha_s)$ and $\tilde{\mathbf{V}}(x_p, \alpha_p, z_p)$ corresponds to minimizing

$$\delta_p = \|\tilde{\mathbf{V}}(x_p, \alpha_p, z_p) - \mathbf{V}(x_s, \alpha_s)\| = \left\| \begin{bmatrix} \Delta x_1 & \Delta x_2 \\ \Delta \alpha_1 & \Delta \alpha_2 \end{bmatrix} \right\| \quad (12)$$

with Δx_k and $\Delta \alpha_k$ (for $k=1,2$) depending on the unknowns (x_p, α_p, z_p) , and in the ideal case should be zero (in practice they can never be zero but one can attempt making them small enough). The minimization of the measure δ_p in Eq. (12) is illustrated in Fig. 8(a). The lattice basis reduction procedure $\mathbf{V}(x_p, \alpha_p, z_p) \rightarrow \tilde{\mathbf{V}}(x_p, \alpha_p, z_p)$ is an iterative procedure with no analytical solution and there is no analytical relation between Δx_k and $\Delta \alpha_k$ (for $k=1,2$) and unknowns (x_p, α_p, z_p) .

The above problem can be tackled by fixing one of the parameters (x_p, α_p, z_p) and finding a solution for the other two that achieves the smallest δ_p . Unfortunately, the optimization problem is not convex and has multiple local minima, which complicates finding the global minimum. However, since there are only three unknowns, a good practical approach is to do a grid search over a reasonable range of the unknown variables. This is a time consuming yet a reliable way to obtain the global minima. We will illustrate this by an example in Section 4.

Practical limitations of a projection-based light field display are related with the physical size and resolution of the ray generators, the number of generated rays, and other screen

properties – see [3] for more details. These limitations translate to a finite number of ray sources with high angular and lower spatial density at the RG plane, i.e. small α_p , and larger x_p . Desired spatial resolution at the screen plane is higher, which can be achieved by reducing the angular resolution. This leads to practical limitations expressed as

$$\alpha_p \ll \alpha_s \text{ and } x_s \ll x_p, \quad (13)$$

where the difference between parameters at the RG and screen planes is at least one order of magnitude. This is illustrated in Fig. 9(a) and it is the starting point for considering the effect of shearing when moving from RG to screen plane.

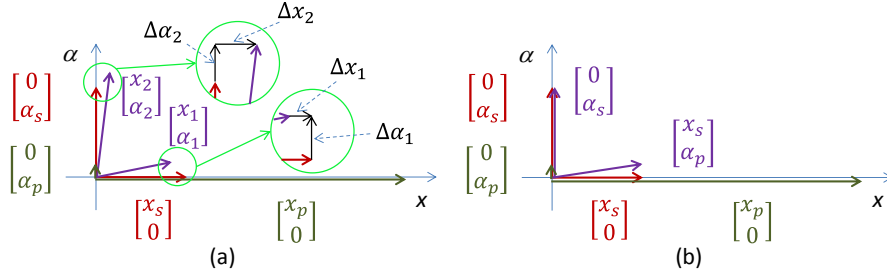


Fig. 8. Display optimization – matching $V(x_s, \alpha_s)$ and $\tilde{V}(x_p, \alpha_p, z_p)$. (a) General optimization solution obtained by minimizing Eq. (12). (b) Expected approximation, according to Eq. (14), for the setup under consideration.

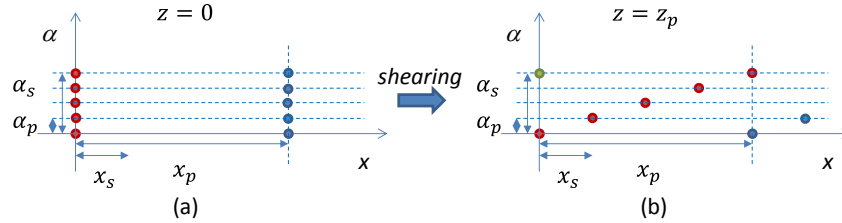


Fig. 9. Display optimization – shearing of $\Lambda(V(x_p, \alpha_p))$ to $\Lambda(V(x_p, \alpha_p, z_p))$ to match $\Lambda(V(x_s, \alpha_s))$. (a) Sampling grid on RG plane. (b) Sheared sampling grid on the screen plane.

The sampling grid at the RG plane is described by $V(x_p, \alpha_p)$. After shearing that grid by distance z_p it turns into the sampling grid at the screen plane described by sampling matrix $V(x_p, \alpha_p, z_p)$ as given in Eq. (9). The question is: Which sampling points in the original grid contribute to the basis vectors after shearing and lattice basis reduction? The approach for finding a good candidate can be graphically visualized as shown in Fig. 9. The original pattern corresponding to $V(x_p, \alpha_p)$ in Fig. 9(a) is sheared to position $z=z_p$ in Fig. 9(b). The best approximation of the pattern $V(x_s, \alpha_s)$ is achieved when (see also Fig. 8(b) for illustration)

$$\tilde{V}(x_p, \alpha_p, z_p) = \begin{bmatrix} x_s & 0 \\ \alpha_p & \alpha_s \end{bmatrix} = \begin{bmatrix} z_p \tan \alpha_p & 0 \\ \alpha_p & \frac{x_p}{x_s} \alpha_p \end{bmatrix}. \quad (14)$$

This leads to the following estimates of two out of three unknown parameters (x_p, α_p, z_p) :

$$\hat{z}_p = \frac{x_s}{\tan \alpha_p} \Leftrightarrow \hat{\alpha}_p = \tan^{-1} \frac{x_s}{z_p} \quad (15)$$

$$\hat{x}_p = x_s \frac{\alpha_s}{\alpha_p} \Leftrightarrow \hat{\alpha}_p = \alpha_s \frac{x_s}{x_p} \quad (16)$$

Since we still have two equations and three unknowns, we need to select one of them by some other means. In the case under consideration, a good selection for α_p is

$$\hat{\alpha}_p \approx \alpha_s / L \text{ for } L \in \mathbb{N}. \quad (17)$$

The reason for this selection lies in the fact that the sampling grid on the screen plane is a sheared version of the sampling grid at the RG plane with shearing performed only in the horizontal direction according to Eq. (3). Under these circumstances, the selection of α_p according to Eq. (17) will ensure that there exist a point in the sheared grid that approximately matches the desired sampling vector $[0 \ \alpha_s]^T$ thereby minimizing $\Delta\alpha_2$ and Δx_2 . This is illustrated in Fig. 9(b).

The estimated parameters, as illustrated in Section 4 by means of examples, will be very close to the optimal ones, e.g. the optimal value of \tilde{x}_p will be in the range $\hat{x}_p \pm x_s / 2$. Based on this, we can formulate the optimization technique as follows:

1. Select a value $\hat{\alpha}_p$ according to available hardware resources and Eq. (17).
2. Use estimation formulas given by Eqs. (15) and (16) to get \hat{x}_p and \hat{z}_p .
3. Refine the result by applying iterative search / general purpose optimization in range $\hat{x}_p \pm x_s / 2$ thereby obtaining an optimal set of parameters $(\bar{x}_p, \bar{\alpha}_p, \bar{z}_p)$.

The evaluated sampling density in ray space $(\bar{x}_p, \bar{\alpha}_p)$ determines the spatial and angular resolution of the display. This technique will be illustrated by means of examples in Section 4.

3.2. Camera setup optimization

The camera setup should provide the rays required by the display for proper recreation of the LF of the scene. While the display is a band-limited device, the 3D visual scene is not (except for simple scenes with low-frequency spatial content, limited depth, and no occlusions) This means that when discussing the optimization of the camera setup, we have two problems to consider. First, how to estimate the optimal camera setup in terms of minimal amount of data that will provide the information needed for rendering all rays generated by the display. Second, how to ensure an alias-free capture of the scene to be recreated by the display.

Both problems are directly related to the display parameters and the corresponding display bandwidth they determine. The ultimate goal is to match that bandwidth with an optimal camera setup which allows rendering all rays needed by the display in an anti-aliased pass band manner. The solution goes through matching the sampling patterns of the display and cameras at the screen plane. With reference to Fig. 7, the optimization problem is formulated as follows: for a given display sampling pattern described by (x_p, α_p) , find (x_c, α_c, z_c) that minimizes

$$\delta_c = \|\tilde{V}(x_p, \alpha_p, z_p) - \tilde{V}(x_c, \alpha_c, -z_c)\| \quad (18)$$

with $\tilde{V}(x_p, \alpha_p, z_p)$ and $\tilde{V}(x_c, \alpha_c, -z_c)$ being the lattice basis reduced sampling matrices of the ray generators and cameras mapped to the screen plane, respectively – the argument $(-z_c)$ indicates that the camera sampling matrix $V(x_c, \alpha_c)$, is mapped to the screen plane with the

minus being there due to the orientation of the z axis. Two comments regarding the above optimization criteria. First, we are doing the matching on the screen plane since this is the place where the D/C conversion takes place – sampling criteria must be satisfied at that plane. Second, in order to speed up the optimization, instead of $\tilde{V}(x_p, \alpha_p, z_p)$ we could also use $V(x_s, \alpha_s)$, assuming that the display sampling grid at the screen plane approximates well enough the desired one. This is perfectly fine in practice since it is expected that practical limitations (e.g. anti-aliasing filter, screen's D/C conversion) will affect the overall visual performance much more than mismatch between the desired and obtained display properties.

By applying an iterative optimization as described in the previous section, we can determine the optimal camera setup in terms of camera parameters $(\bar{x}_c, \bar{\alpha}_c, \bar{z}_c)$. In comparison to display optimization, there are additional restrictions that have to be taken into account, e.g. reasonable distance of a viewer from the screen, practical camera resolutions, $FOV_c \geq FOV_p$, and camera-to-camera distance that cannot be too small.

After determining the minimal camera sampling pattern $\Lambda(V(\bar{x}_c, \bar{\alpha}_c))$ for a given \bar{z}_c , we map the optimized display unit cell in the frequency domain at the screen plane $P^*(V(\bar{x}_p, \bar{\alpha}_p, \bar{z}_p))$ to the camera plane where it turns into $S_{\bar{z}_c}(P^*(V(\bar{x}_p, \bar{\alpha}_p, \bar{z}_p)))$ where the $S_z(P^*)$ is the mapping (shearing) operator. Since P^* is a convex set with points being the vertices of the unit cell in frequency domain that can be defined as

$$P^* = \{(\omega_k, \psi_k) \in \mathbb{R}^2\}_{k=1,2,\dots,K} \quad (19)$$

the mapping operator $S_z(P^*)$ maps each of those points such that

$$\begin{aligned} S_z(P) &= \{F_z(\omega, \psi) \mid (\omega, \psi) \in P\} \\ F_z(\omega, \psi) &= (\omega_k, \psi_k - z \tan \omega_k). \end{aligned} \quad (20)$$

The obtained shape of the cell determines the bandwidth of the display as observed from the camera plane. This is illustrated in Fig. 10. As can be seen, the shape is very different from the unit cell on the screen plane. Comparing this with the plenoptic sampling theory, it is obvious that, with respect to the overall plenoptic function, the display can only reproduce a finite amount of data concentrated around a particular distance from the viewer / camera plane. The limitation applies to angular and spatial coordinates. It should be pointed out that area (bandwidth) wise, $S_{\bar{z}_c}(P^*(V(\bar{x}_p, \bar{\alpha}_p, \bar{z}_p)))$ and $P^*(V(\bar{x}_c, \bar{\alpha}_c))$ are of approximately same size (depending on how good minima have been found), but they cover different set of frequencies. In practice this means that for proper preparation (sensing) of content to be shown on the display, one has to do the following steps:

1. Capture the scene with sampling rate (large number of cameras) that will ensure proper anti-alias capture. This depends on the scene. However, the smallest bandwidth that has to be captured is marked by $P^*(V(x_c^{BIG}, \alpha_c^{BIG}))$.
2. Filter the captured content with filter having the passband determined by $S_{\bar{z}_c}(P^*(V(\bar{x}_p, \bar{\alpha}_p, \bar{z}_p)))$.
3. Down-sample the filtered signal to $P^*(V(\bar{x}_c, \bar{\alpha}_c))$.

This sensing procedure will result in properly pre-processed minimal amount of data that at the same time maximally utilizes the visualization capabilities of the display.

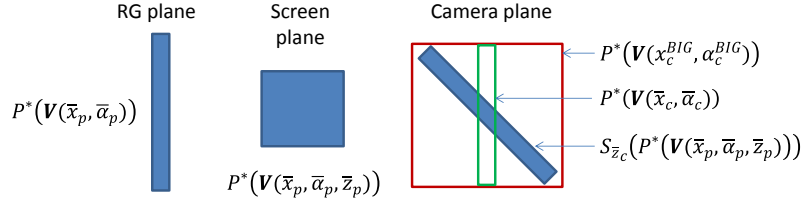


Fig. 10. Illustrations of display/camera sampling unit cell in frequency domain at different planes.

4. Examples

We illustrate the proposed optimization procedure on a ‘realistic’ display with reasonable quality as it can be built today, illustrate the optimization approach for optimal capture setups, and finally show / discuss what would be the setup of a display matching the requirements of the HVS.

4.1. Display optimization examples

First we illustrate the proposed display configuration optimization on a display having desired spatial and angular resolution at the screen plane such that $x_s = 1$ mm and $\alpha_s = 1^\circ$. We fix the angular resolution of the ray generators at the RG plane to $\alpha_p = 0.0391^\circ$ (this resolution corresponds to a spatial resolution of 1024px over FOV of 40 degrees). For fixed α_p , we evaluate the matching error δ_p on the screen plane for various values of x_p ($10\text{ mm} \leq x_p \leq 40\text{ mm}$) and z_p ($600\text{ mm} \leq z_p \leq 1800\text{ mm}$). The results of the optimization are shown in Fig. 11. In the figure, the left column shows the overall optimization range and the right column shows a zoomed-in range around the minimal value. Top row shows the result of overall optimization whereas middle and bottom row show the best solutions for a given z_p and x_p . As it can be seen, there is a dominant minimum at $(\bar{x}_p, \bar{z}_p) = (26.01\text{ mm}, 1465.90\text{ mm})$ with an error value of $\delta_p = 0.04249$. The Voronoi unit cell $P(V(\bar{x}_p, \bar{\alpha}_p, \bar{z}_p))$ of such optimized display is shown in Fig. 12. As it can be seen, the match with the desired $P(V(x_s, \alpha_s))$ one is almost perfect. The downside of such grid-based search is in the need to evaluate many combinations of (x_p, z_p) not knowing which one will result in optimal solution. This can be considerably speeded-up by using the estimation approach described in Section 3.1. Following Eqs. (15) and (16), the estimated values for the problem under consideration are $(\hat{x}_p, \hat{z}_p) = (25.58\text{ mm}, 1465.36\text{ mm})$. As can be seen they are very close to the ones above obtained by the grid search. By performing single gradient-based optimization from the estimate, we end up with $(\bar{x}_p, \bar{z}_p) = (26.00\text{ mm}, 1465.34\text{ mm})$ with an error value of $\delta_p = 0.04249$. This is almost identical to the one obtained by the grid-based search and is obtained with a small amount of computational resources – fraction of a second instead of 10-15 min needed by the grid-based approach. Since almost identical result is obtained with both approaches, we can conclude that our proposed estimation method is correct and useful.

By using the fast estimation method, we can easily calculate optimal display setups for various screen parameters. First, Fig. 13 shows display optimization results for $x_s = 1$ mm and $\alpha_s = 1^\circ$ for various values of α_p . It is seen that for a good approximation we need small values of α_p . However, very small values of α_p require impractically large values of z_p and x_p . Therefore, in practice, a compromise between those has to be made. For illustration, the unit cells for optimal solutions for several values of α_p are shown in Fig. 14.

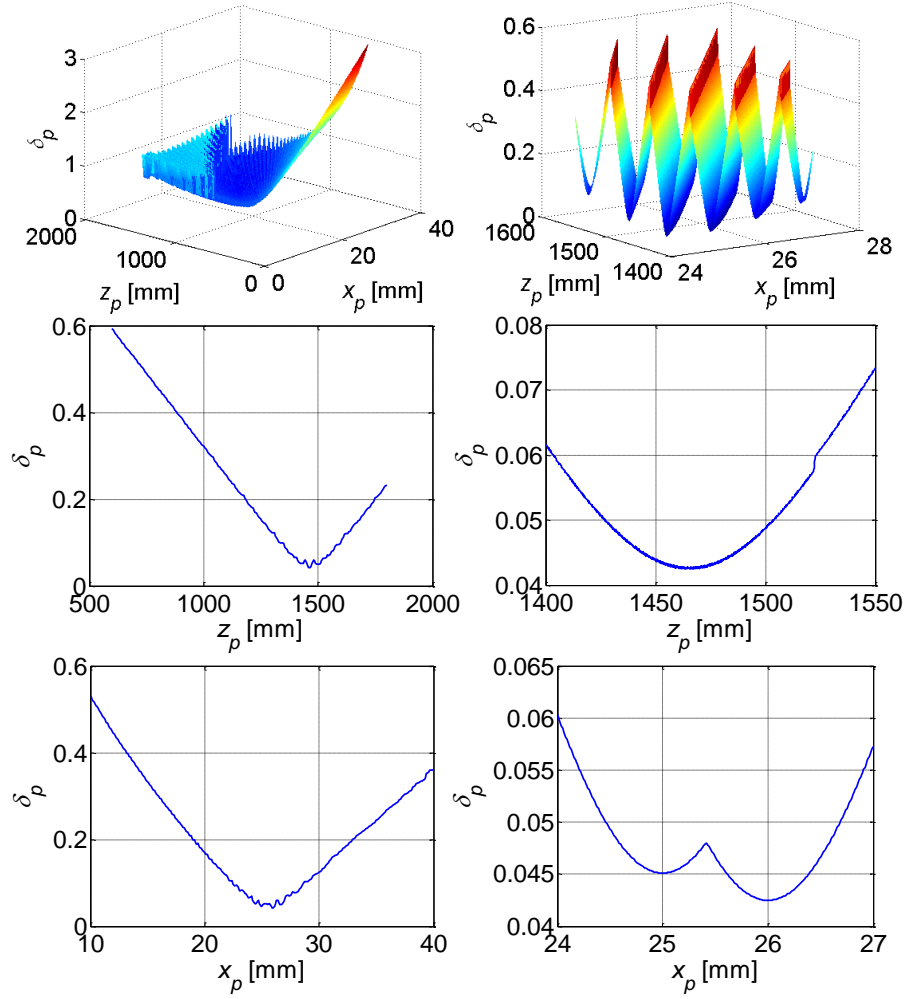


Fig. 11. Display optimization example based on grid search for $x_s=1$ mm, $\alpha_s=1^\circ$, and $\alpha_p=0.0391^\circ$ – figures in the right column are zoomed in version of figures in the left column.

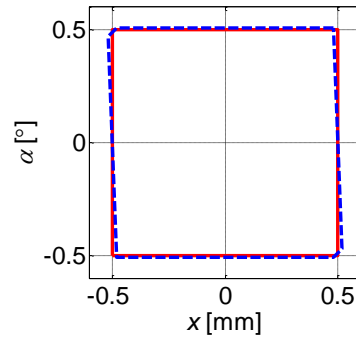


Fig. 12. Unit cells at screen plane for the optimized display setup solution for $x_s=1$ mm, $\alpha_s=1^\circ$, and $\alpha_p=0.0391^\circ$ – $P(V(\bar{x}_p, \bar{\alpha}_p, \bar{z}_p))$ dashed/blue and $P(V(x_s, \alpha_s))$ solid/red.

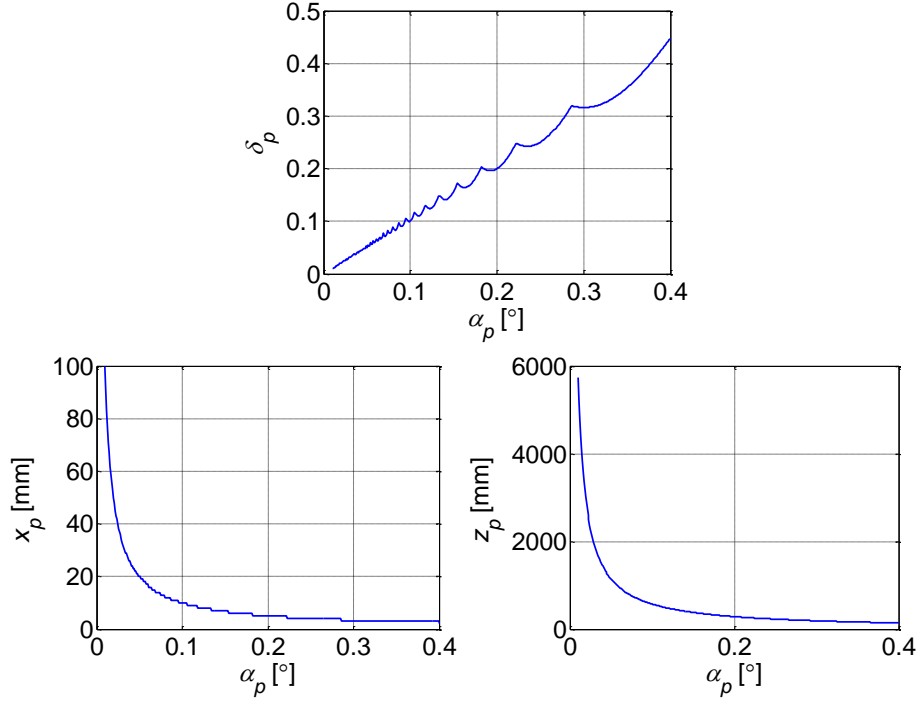


Fig. 13. Display optimization example for $x_s = 1$ mm and $\alpha_s = 1^\circ$ for various values of α_p .

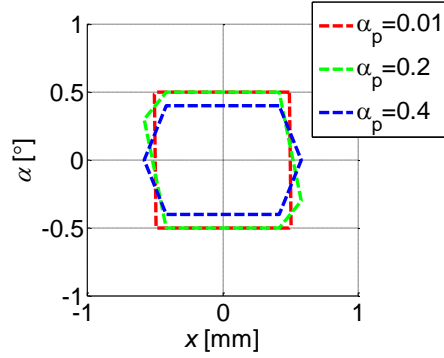


Fig. 14. Normalized ray-generator unit cells at the screen plane, $P(\bar{V}(\bar{x}_p, \bar{\alpha}_p, \bar{z}_p))$ for various values of α_p .

Next, we investigate the influence of different values of (x_s, α_s) on (z_p, x_p) for fixed $\alpha_p = 0.0391^\circ$. As seen in Fig. 15 a similar reconstruction error δ_p can be achieved independently of the choice for x_s and α_s . Furthermore, distance z_p is influenced only on the desired x_s and finally, x_p has to be increased if either x_s or α_s is increased. These figures give us a good understanding about the relation between involved parameters and can help us in making proper selection decisions.

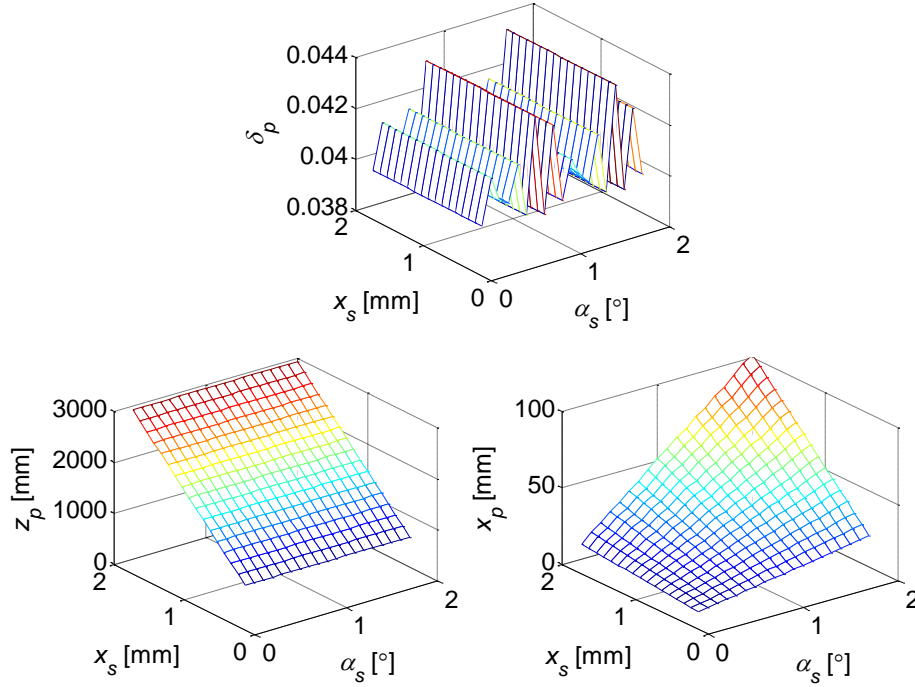


Fig. 15. Display optimization example for $\alpha_p = 0.0391^\circ$ and various values of x_s and α_s .

4.2. Camera optimization examples

For an optimized display as described in the previous section, the display bandwidth is uniquely defined by $P(\mathbf{V}(\bar{x}_p, \bar{\alpha}_p, \bar{z}_p))$ with a good approximation being described by $P(\mathbf{V}(x_s, \alpha_s))$. Content captured by any means has to be pre-filtered to this bandwidth. The question here is what is the optimal camera/viewer setup, that is, what are the optimal parameters (x_c, α_c, z_c) that would support the display bandwidth in the best possible way. In comparison to display optimization where it was logical to fix parameter α_p , here it is more convenient to fix the screen to viewer distance z_c since the viewer distance is typically ‘fixed’ / ‘selected’ by the user preferences / general recommendation for ‘TV’ watching. For a fixed distance $z_c = 2000$ mm, the result of optimization are shown in Fig. 16. Matching is performed again at the screen plane. There is dominant minimum at $(\bar{x}_c, \bar{\alpha}_c) = (35.72 \text{ mm}, 0.0284^\circ)$ with an error value of $\delta_c = 0.06832$. For comparison purpose, optimized ray generators and camera unit cells are shown in Fig. 17.

The grid search can be made faster by a better initial estimation. This can be done by assuming that the unit cell at the screen distance is ideal, that is, it is defined by (x_s, α_s) . By following the approach presented in Section 3.2, we obtain $(\hat{x}_c, \hat{\alpha}_c) = (34.91 \text{ mm}, 0.0286^\circ)$. This is very close to the aforementioned optimal solution. Due to a high nonlinearity (see Fig. 16, middle row, right), one cannot use gradient based optimization but can perform a grid search only in the vicinity of the estimated values. Since this drastically limits the search space, it can be performed much faster than the full grid search.

The sampling pattern in the spatial domain can be converted to the frequency domain by using Eq. (7). By converting the frequency domain unit cell belonging to optimized display pattern from the screen plane to the camera plane, we obtain the bandwidth of the display –

shown in blue in Fig. 18. As discussed before, one should sample the scene with wide enough bandwidth to avoid aliasing, then pre-filter and then downsample. After downsampling, one obtains the maximum amount of data required by the display – display cannot show more and as such there is no point to provide more. It should be pointed out that this is in line with similar analysis performed for autostereoscopic displays [19,20].

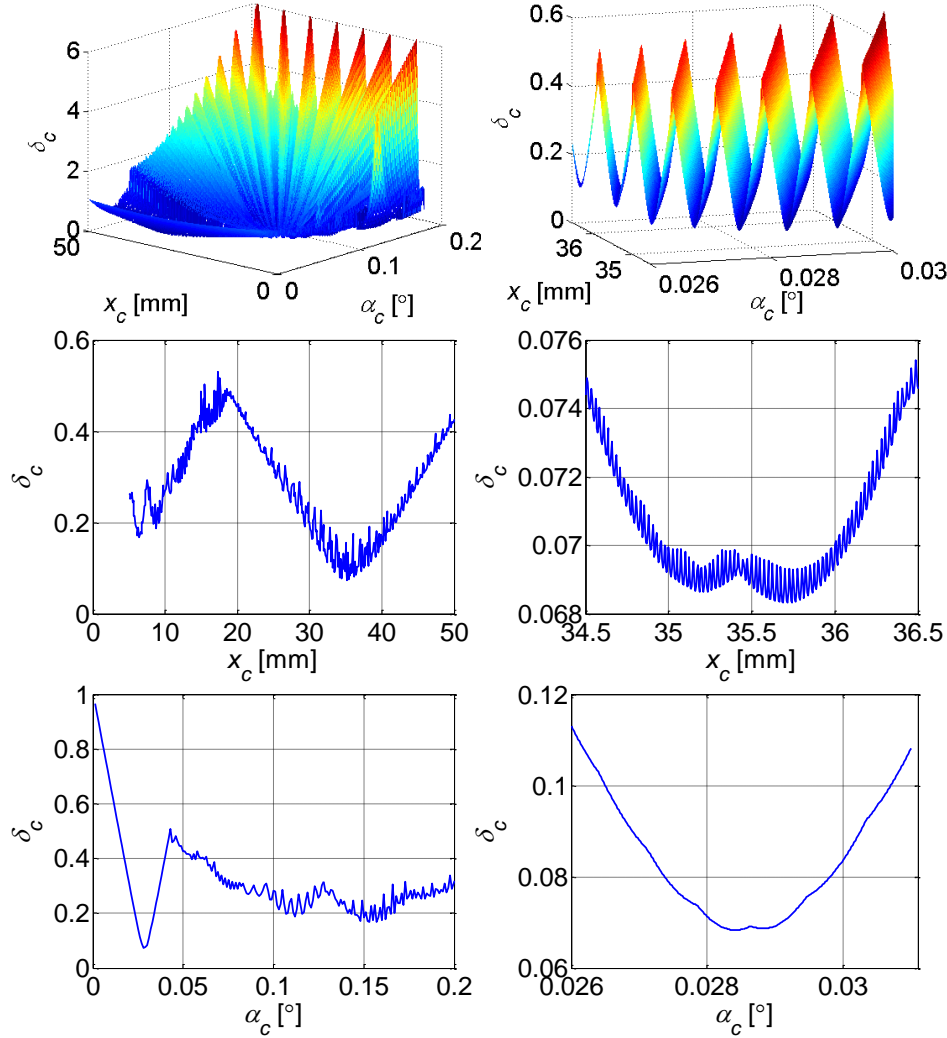


Fig. 16. Camera optimization example based on grid search for optimal $P(V(\bar{x}_p, \bar{\alpha}_p))$ optimized for $x_s=1$ mm, $\alpha_s=1^\circ$, and $\alpha_p=0.0391^\circ$ with $z_c=2000$ mm – figures in the right column are zoomed in version of figures in the left column.

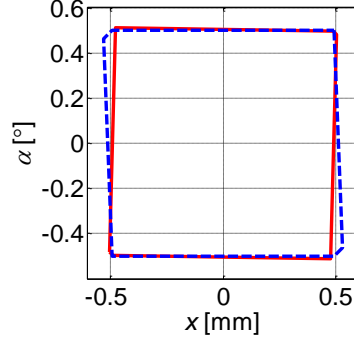


Fig. 17. Unit cells at screen plane for the optimized display and camera setup solution for $x_s = 1$ mm, $\alpha_s = 1^\circ$, $\alpha_p = 0.0391^\circ$ and $z_c = 2000$ mm – $P(V(\bar{x}_p, \bar{\alpha}_p, \bar{z}_p))$ dashed/blue and $P(V(\bar{x}_c, \bar{\alpha}_c, -\bar{z}_c))$ solid/red.

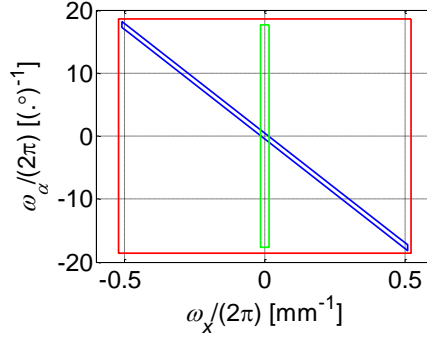


Fig. 18. Unit cells (bandwidths) at the camera (viewer) plane – $S_{z_c}(P^*(V(\bar{x}_p, \bar{\alpha}_p, \bar{z}_p)))$ blue, $P^*(V(\bar{x}_c, \bar{\alpha}_c))$ green, and $P^*(V(\bar{x}_c^{BIG}, \bar{\alpha}_c^{BIG}))$ red.

4.3. Consideration related to an ‘ideal’ HPO 3D display

An ideal display should deliver the resolution required by the HVS. For estimating the required display angular-spatial resolution, in this section, we follow the discussion presented in [7, p 219-220]. It is assumed that an eye at distance z_c from the display can differentiate spatial changes equal to $1/60^\circ$ – this is equivalent to resolution of 30cpd (cycles per degree). This maps to

$$x_s = z_c \tan 1/60^\circ. \quad (21)$$

Furthermore, the angular deviation that the eye can distinguish depends on the pupil size d_p and can be estimated as [7]

$$\alpha_s = \tan^{-1}(d_p / z_c). \quad (22)$$

Assuming that average pupil size, as reported in the literature, is $d_p = 3$ mm and the viewing distance is fixed at $z_c = 2000$ mm, we end up with required display resolution of $(x_s, \alpha_s) = (0.58 \text{ mm}, 0.086^\circ)$. This means that an ‘ideal’ HPO display with 60 degree FOV, for

the assumed fixed distance, is required to reproduce at least $2 \cdot 10^9$ rays per square meter of the screen surface.

Following the proposed display optimization, one can determine that for fixed $\alpha_p = 0.0313^\circ$, the optimal parameters of the ray generators should be $(\bar{x}_p, \bar{z}_p) = (1.74 \text{ mm}, 1062.86 \text{ mm})$ with the matching error being $\delta_p = 0.0322$. By mapping this values to the camera plane (c.f. Fig. 19), we can determine the necessary sampling rates as discussed in the previous section.

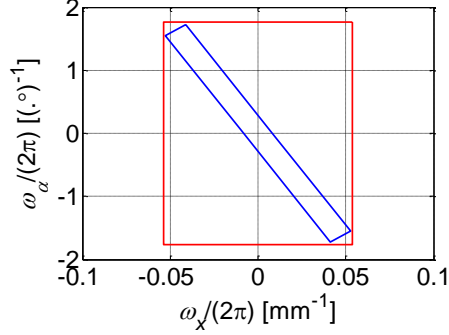


Fig. 19. Unit cells (bandwidths) at the camera (viewer) plane – $S_{z_c} \left(P^* \left(V(\bar{x}_p, \bar{\alpha}_p, \bar{z}_p) \right) \right)$ blue and $P^* \left(V(x_c^{BLG}, \alpha_c^{BLG}) \right)$ red.

5. Concluding remarks

In this paper we presented a sampling model of the LF display-camera channel. We have shown that, from the sampling theory viewpoint, we can start with the required properties of the display specified in the ray space at the screen plane and then calculate the display setup fulfilling those requirements. Having the display setup, we can estimate the minimal set of data that the display needs to maximally utilize its visualization capabilities together with filter bandwidth for data pre-filtering aimed at alias-free reproduction.

Several points should be emphasized beyond the scope of this paper. First, we did not discuss all practical (hardware) aspects of implementing such displays, e.g. additional limitations to the design might be enforced by the available components and space like overall size, available ray sources, etc. Nevertheless, the same methodology presented in the paper still applies. Second, we assumed ideal D/C properties of the (holographic) screen. In practice, the screen will introduce additional smoothing that will further band limit the content the display can reproduce. Third, it should be always kept in mind that while the display is a bandlimiting device, a typical visual scene is not bandlimited. This means that special care has to be taken when sensing a scene and preparing the content for its optimal anti-aliasing filtering prior of its visualization on the LF display.

The discussion in the paper concentrated on projection-based displays employing diffusion-based holographic screen for reconstructing the continuous light field. This specific setting allows to clearly demonstrate the importance of ray sampling patterns for characterizing the display bandwidth and to directly relate it with the ray acquisition setting. However, the proposed approach can be used with any type of display system that attempts recreating a continuous light field and has an underlying (not necessary uniform) sampling of the input light field in the angular-spatial domain. Examples include autostereoscopic [21] and super multi-view displays [22, 23]. Further work and a more comprehensive analysis is required for displays having a non-uniform density of input light rays or /and ones that are

capable of changing the density of light rays based on content (e.g. tensor displays [24]) and it will be a topic of further research.

The estimates of 'ideal' projection-based LF display parameters as obtained in Section 4.3 were based on geometrical assumptions about the resolution power of the human eye. They show that a projection-based LF display matching the sampling density of the HVS is possible given that the individual optical modules are spaced 1,74 mm apart, while each module delivers rays with 0.03 rad angular step. Such a display is still difficult to produce and can be attempted in the future. Any other display designs with lower resolutions shall greatly benefit from the solution presented in this paper for delivering alias-free imagery. While the resolution power of the HVS estimated by geometrical assumptions is quite high, further studies are needed to characterize the perceptual threshold of continuous parallax, in the fashion how the window of visibility in disparity domain has been estimated [25]. Such perceptual characterization of continuous parallax would be more instructive when specifying the desired LF display bandwidth. A similar problem does exist with LF content creation. To cope with the anti-aliasing requirements, a very dense set of cameras is required for capturing content to be further processed for the specific display. Future development of intermediate view generation out of a set of sparsely captured views and employing signal processing sparsification approaches is of great interest.

Acknowledgments

The research leading to these results has received funding from the PROLIGHT-IAPP Marie Curie Action of the People programme of the European Union's Seventh Framework Programme, REA grant agreement 32449 and from the Academy of Finland, grant No. 137012: High-Resolution Digital Holography: A Modern Signal Processing Approach.