

Display-specific light-field analysis

Robert Bregović⁽¹⁾, Péter Tamás Kovács^(1,2), Tibor Balogh⁽²⁾, Atanas Gotchev⁽¹⁾

⁽¹⁾ Department of Signal Processing, Tampere University of Technology, Tampere, Finland

⁽²⁾ Holografika, Budapest, Hungary

ABSTRACT

Present generations of 3D displays including stereoscopic and autostereoscopic displays have very limited number of views, and thus limited parallax. In contrast, the emerging light field (LF) displays support hundred(s) of views with acceptable spatial resolution thereby enabling a more realistic representation of 3D scenes. This is accomplished for the price of high data throughput, complex data acquisition and a high demand of computational power. Thus, the optimization of the content representation is of crucial importance for the performance of the whole display system. In this paper, we discuss the requirements for LF based processing of 3D content for representation on the new generation of ultra-realistic LF displays. We analyze the overall processing chain from acquisition through LF based modeling and representation to visualization on the considered displays. By analyzing the visualization capabilities of a given LF display using spatial and frequency domain analysis, we draw guidelines on how to properly acquire the required data and repurpose it based on the targeted display. We show that by taking into account the properties of the display during scene sensing and during LF processing, a good visual representation of 3D content on a given display can be achieved with a minimalistic capture setup.

Keywords: Plenoptic function, light field, multidimensional sampling, ray space, 3D displays, Voronoi cells

1. INTRODUCTION

Contemporary 3D displays are based on either stereoscopic (requiring glasses) or autostereoscopic (providing limited number of discrete views) visual technologies¹. While recreating the binocular visual cue to some extent, they largely fail to support continuous head parallax, thus providing only a limited realism in the visualized scene². Supporting continuous parallax goes through the display's capability of reconstructing a precise-enough optical replica of the continuous light field (LF), as emitted by the real-world scene³.

There are two major problems to be resolved in order to reach an ultra-realistic visualization of a 3D scene. The first problem lies in the display technology itself. A perfect 3D display has to be capable of emitting a large number of rays in multiple directions to support the varying spatial and angular visual content of the scene. As the scene LF is continuous, the display technology should allow for an adequate continuous function reconstruction out of discrete rays (samples) with high spatial and angular resolution. The most promising candidates, as of today, to achieve this are so-called projection-based LF displays. These displays are built from projection engines considered as ray generators and a custom-build holographic screen, where the continuous LF reconstruction takes place. The second problem is related with the amount of data required for driving such 3D displays. The data, constituting a LF discrete representation, has to be captured, properly processed, stored, transmitted to the display, and then used for ray generation. In order to properly design LF displays as well as content for them, one needs analysis tools based on some LF mathematical formalization which is also technology-friendly.

In an earlier work^{4,5}, we have developed analysis tools for multiview displays based on the notation of measured display passband, that is, the 'visual' bandwidth that a multiview display can show with acceptable distortions. Another concept of display bandwidth focused on autostereoscopic displays and making use of the plenoptic sampling theory has been presented in⁶. It uses a ray-space parameterization of the LF, which results in a rectangular sampling pattern suitable for straightforward analysis and data pre-processing. A distinctive feature of LF displays in comparison to autostereoscopic displays is that the generated rays appear in non-rectangular grids, which makes the overall display analysis more complex.

In this paper we aim at analyzing the current generation of projection-based LF displays in terms of the bandwidth they can generate. We apply this analysis for defining optical camera setting providing best scene visualization with

minimum data load. In comparison to the measurement-based analysis of LF displays that has been presented in ⁷, here we will look into an analytical approach making use of ray-space LF representation.

The outline of the paper is as follows. In Section 2 the description of the plenoptic function is given with the emphasis being on the 4D LF parameterization. The projection-based LF displays are explained in Section 3. The formalization of the LF in ray space (spatial and frequency domain) for LF displays is presented in Section 4. The optimization approach for deriving minimalistic camera capture setup for a given display, including a typical use case, is discussed in Section 5. Finally, concluding remarks are given in Section 6.

2. PLENOPTIC FUNCTION AND LIGHT FIELD

In this paper, we adopt the formalization of light through its geometric behavior as described by the ray-optics theory assuming that every point in space emits infinite number of directional light rays. Mathematically, this is expressed through the notion of the 7D continuous plenoptic function (PF) ⁸

$$P(\theta, \varphi, \lambda, t, V_x, V_y, V_z), \quad (1)$$

where (V_x, V_y, V_z) is a location in the 3D space, (θ, φ) are directions (angles) of observation, λ is wavelength, and t is time, see Figure 1 for illustration.

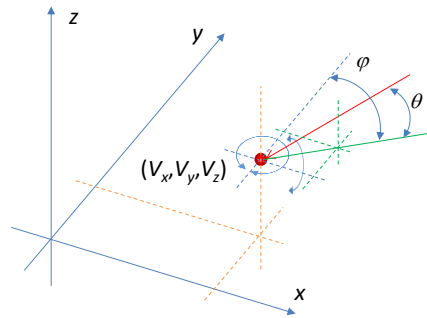


Figure 1. Illustration of the 7D PF.

Several simplifications of the 7D PF have been introduced to cope with the large amount of data required for its representation. By considering only static scenes, replacing the wavelength with RGB components (only visible wavelengths are of interest), and limiting the scene to half space one ends up with a discrete 4D function which is the most popular practical representation of the PF. This 4D function can be represented either by using a two-plane parameterization $L(x, y, s, t)$ or one plane and direction parameterization $L(x, y, \varphi, \theta)$, as illustrated in Figure 2 ^{9,10,11}.

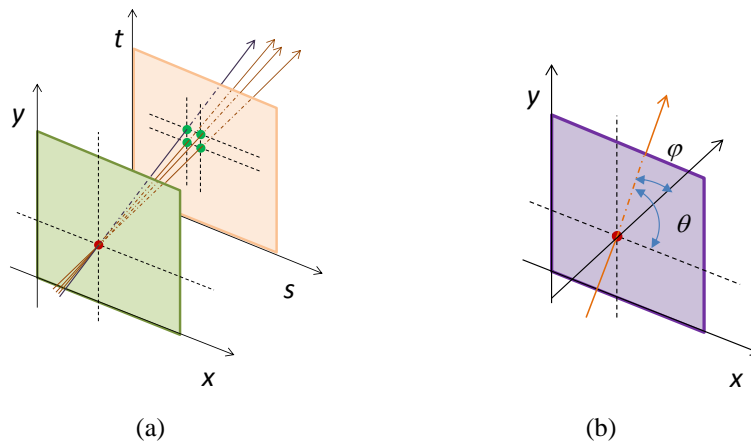


Figure 2. 4D PF parameterizations. (a) Two-plane parameterization. (b) One plane and direction parameterization.

For horizontal parallax only (HPO) 3D displays, considered hereafter, the variable y related to vertical parallax can be dropped from either representation. For illustration purposes one can drop also the variable responsible for vertical resolution (t or θ) thus getting a 2D slices of the ray space as shown in Figure 3. The relation between the planes parameterized by (x, s) and (x, φ) , is given by

$$s = l \tan \varphi \tag{2}$$

with $l = 1$ being a typical choice for l and x being the same in both representations.

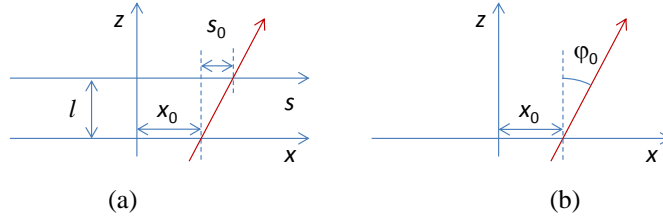


Figure 3. HPO LF parameterizations. (a) Two-plane parameterization. (b) One plane and direction parameterization.

The propagation of the LF from one plane (Plane 1) to second plane (Plane 2) that are distance d apart can be expressed as¹²

$$L_1 \left(\begin{bmatrix} x_1 \\ s_1 \end{bmatrix} \right) = L_0 \left(\begin{bmatrix} x_0 \\ s_0 \end{bmatrix} \right) = L_0 \left(\begin{bmatrix} 1 & -d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ s_1 \end{bmatrix} \right), \tag{3}$$

with $L_0(x_0, s_0)$ and $L_1(x_1, s_1)$ being the LFs on first and second plane, respectively (see Figure 4a). In essence, this means that the rays building the LF keep their intensity and direction while their position on the x -axis gets re-arranged based on the distance between planes under consideration through a linear transformation that depends on the ray direction and distance. The linear transformation, given in Eq. (3), is a shifting operation along the x -axis (also referred to as shearing in the LF context). This is illustrated in Figure 4(a)-(c). Similar relation holds for the case of (x, φ) parameterization though the transform as in (3) is not linear due to the relation between s and φ as indicated by (2).

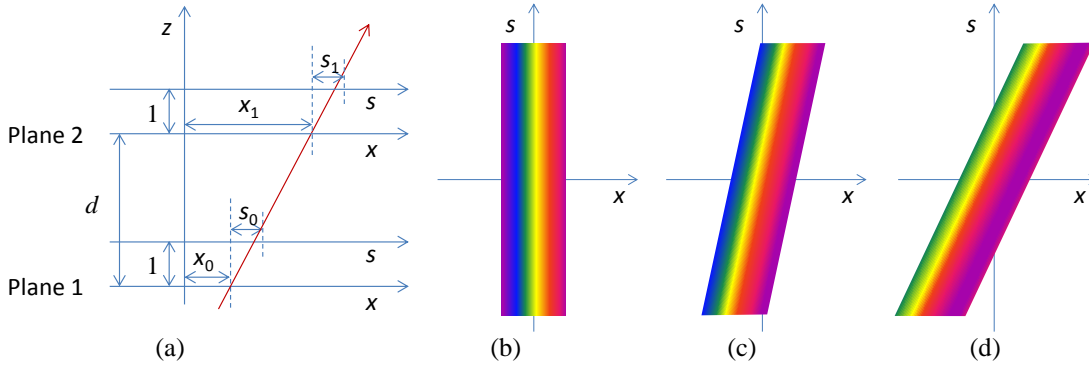


Figure 4. Light (ray) propagation – Representation on two different planes (a) and example of LF at on the surface of a flat object for $d=0$ (b), $d=d_0$ (c), and $d=2d_0$ (d).

The PF and its lower-dimensional approximations are continuous functions. In order to process light by digital means one has to have acquisition, computational, and visualization tools built upon proper sampling and reconstruction of the LF. In general, sampling of multi-dimensional functions follows the Shannon theorem assuming that the function into question belongs to the class of functions with limited frequency support. LFs originating from 3D visual scenes are not strictly bandlimited. They are formed by objects at different depths with sharp transitions at edges and occlusions depending on the viewing ray directions. Still, bandlimited sampling theory has played an important role in studying the plenoptic sampling^{13,14}. Relations have been established between the scene minimum and maximum depth and the frequency support of the corresponding LF, and anti-aliasing conditions addressing the density of cameras and their spatial resolution have been formulated. More complex LF models have been considered for reconstruction of scenes containing tilted planes or scenes with occlusions^{15,16,17}.

From a display perspective, the scenes to be reconstructed by the display are considered bandlimited. There is a twofold reason for this assumption. First, the display can reproduce a finite number of discrete rays. Second, in order to create a continuous LF out of discrete set of rays, it utilizes some kind of discrete-to-analog converter which in most cases is considered as a low-pass filter. Both factors limit the bandwidth of the LF that can be reconstructed by the display. Therefore, in order to adapt the content creation tools to the display requirements, one has to adopt the multidimensional sampling theory for bandlimited signals and ensure that the sampling configuration is adequate to the reconstruction one.

3. PRINCIPLES OF OPERATION OF PROJECTION-BASED LIGHT-FIELD DISPLAYS

Projection-based LF displays use multi-projection at their core to achieve a high number of light rays, that is, a large number of projection engines are used in parallel, which project light rays from slightly different physical positions, and are usually stacked in an equidistant linear or arc setup (linear setups are considered in this paper for simplicity). The main advantage of such distributed projection system is scalability – novel light ray directions can be added, increasing thereby the total number of pixels (rays), by introducing more projection engines, without sacrificing the resolution of the existing directions. This is in sharp contrast with 3D displays which use a single light modulator with a fixed pixel count (e.g. a flat panel) as sources of light rays, and thus have a tradeoff between directions and resolution per direction. The sum of directions covered by light rays forms the displays’ field of view (FOV), which shows the angle under which a 3D image is visible on the screen.

Beside projection engines, the second important part of a projection-based LF display is the holographic screen. This screen is located on the front side of the display and this is the plane where the desired LF gets reconstructed out of discrete rays. Light rays hit this screen from multiple angles at various positions, and the screen lets these light rays pass through without changing their direction creating a narrow angular beam. The holographic screen does not have an explicit pixel structure and taking a finite area on it, one can see it emits different light rays to different directions. This is an essential property of any glasses-free (autostereoscopic) display, as any LF that represents a non-flat scene must have direction selective light emission (see Figure 5).

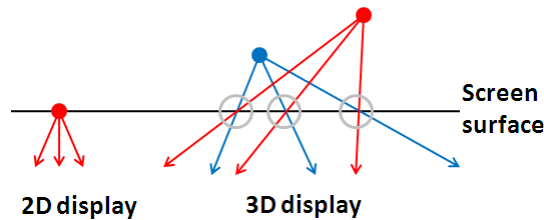


Figure 5. Comparison of light rays emitted from a 2D display and a 3D display. 2D displays emit the same color to all directions from a single pixel (left). 3D displays must emit different colors to different directions to show a 3D scene (right). Notice direction selective light emission on the “3D pixels” marked with grey circles.

This approach can be implemented in front-projected or back-projected configuration (using a reflective holographic screen in the front projected case), as it has been demonstrated in various prototype and commercial LF displays¹⁸.

The image of a single engine is projected all over the visible screen area, thus light rays hitting the screen at different horizontal positions propagate (diverge) into different directions. As a consequence, a viewer’s eye can never see such an image in its entirety from a single position, that is, a single projection engine is not projecting a “view” in the sense typically used in 3D display terminology. Rather, a single 2D image as perceived by one eye of the viewer is made up of light rays emitted from multiple sources (see Figure 6).

As LF displays use a limited number of projection engines acting as light sources, the directions emitted via a single point on the screen are discretized. The holographic screen creates a continuous LF from the adjacent light rays. Direction selectivity (angular resolution) of a LF display is a design parameter, which is scalable up to physical implementation limits – practical displays target an angular resolution of less than one degree. This imposes a certain bandwidth of the LF reconstructed by the display.

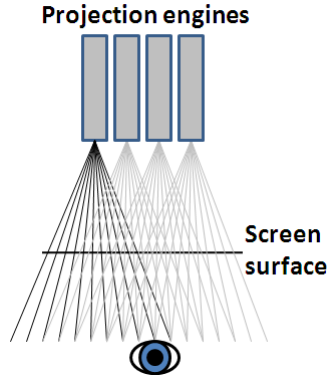


Figure 6. Light rays emitted by a single projection engine (black rays) are not seen from a single position, as they are emitted at different screen positions and different directions. The image seen by a single eye is made up of light rays originating from many sources.

LF displays today reproduce HPO: as viewer's eyes are displaced horizontally, and viewers are typically moving horizontally in front of the screen, the practical effects of missing vertical parallax are negligible. Instead, the same image is shown when observing the screen from different heights (the image follows the viewer). There is no theoretical limitation to extend this principle to vertical parallax, however the complexity and price of such systems is considered prohibitive today.

To put these principles in perspective, some examples for typical design parameters are provided as follows: the number of projection engines ranges from 32 to 128; the LF display's FOV ranges from 30 degrees up to 180 degrees; the resolution of individual projection engines varies from VGA to HD; and the total light ray count emitted by the display ranges from 10 to 80 Mpixels (Mrays). More details about the projection-based LF displays, including principles, implementation, computational background, and applications, are given in ¹⁸.

4. DISPLAY SPECIFIC LIGHT FIELD FORMALIZATION

4.1 Ray distribution of an light-field display

In this paper we assume a basic linear setup of the projection engines of an HPO projection-based LF display as illustrated in Figure 7. In the figure, N_p stands for the number of projection engines, FOV_{proj} denotes an engine's field of view, z_p is the distance between the light ray sources and the screen, and d_s is the (horizontal) size of the screen. The projection engines are equidistantly distributed with the distance between two adjacent units being

$$x_p = d_p / (N_p - 1). \quad (4)$$

Each projection engine is capable of generating N_x rays over the horizontal FOV_{proj} . However, as seen from the figure, the projection engines towards the edges of the setup contribute with a smaller number of rays due to the finite size of the screen (only the rays crossing the surface of the screen are of interest). It is worth pointing out that all these parameters are fixed by the display setup and cannot be changed by a display user.

We assume that the rays from one projection engine hit the screen plane at equidistant points thereby generating pixels of equal width as illustrated in Figure 8. Consequently, the angular distribution of the rays is nonuniform – angular distance between two rays is larger at the center of FOV_{proj} and smaller at the edges of FOV_{proj} . Nevertheless, for small FOV_{proj} , this nonuniformity can be ignored.

The position of a ray at a distance z from its generator can be evaluated by

$$x_z^{(r)} = x_0^{(r)} + z \tan(\varphi^{(r)}) \quad (5)$$

with $\varphi^{(r)}$ being the angle of the ray under consideration as illustrated in Figure 7. The ray keeps the same direction as at the origin but appears on a different place along the x coordinate as it propagates in the z direction away from origin. Consequently, the distribution of rays will be different at different planes perpendicular to the z direction. It should be

noted that although the rays from one projection engine generate a rectangular grid, this is not the case for rays originating from different projection engines as discussed in Section 3.

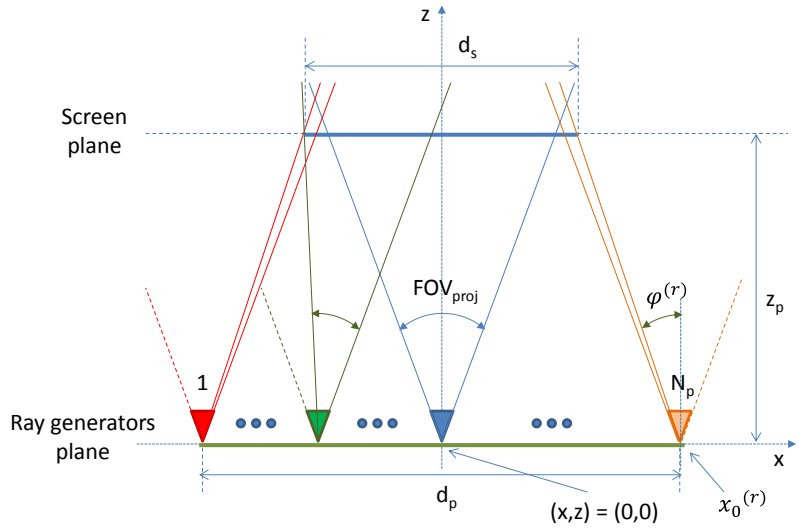


Figure 7. LF display setup.

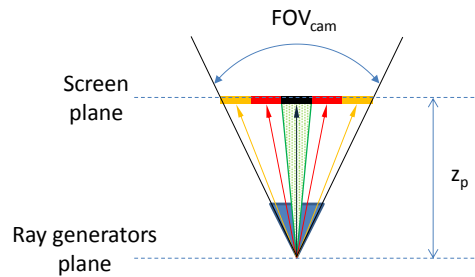


Figure 8. Example of ray distribution over FOV_{proj} in one projection engine (light ray generator).

4.2 Ray-space representation of the light field generated by the display

As discussed in Section 3, projection-based LF displays generate continuous LF from a set of discrete ray sources. Following the LF representations described in Section 2, one can opt for three coordinate systems for representing the generating rays as depicted in Figure 9.¹²

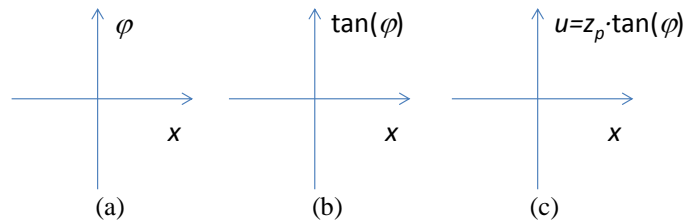


Figure 9. LF ray-space parameterizations. (a) Ray angle vs. position. (b) Two-planes with second plane being at unit distance. (c) Two-planes with second plane being at the screen distance.

First option is to have samples on the (x, φ) system (Figure 9a), thus indexing the rays through their position and direction (angle). The second option, as in Figure 9b, is to consider the $(x, \tan(\varphi))$ coordinate system. It corresponds to a classical two-plane parameterization with the second plane located at unit distance away from the first one (c.f. Eq.

(2)). The third option corresponds to the $(x, z, \tan(\varphi))$ system (Figure 9c), which is essentially a two-plane parameterization with the second plane located at the screen level (c.f. Figure 7). Although all those describe the same LF, they offer different computational features when dealing with propagation of discrete rays (samples in the corresponding spaces). The first option yields an almost rectangular initial grid of samples. The grid keeps changing quite isotropically when considering the ray propagation from the projecting engines plane to the screen plane. The second option yields a perfect rectangular grid for $z=0$, however, there is a big difference in values between the two axes – typical values of $\tan(\varphi^{(r)})$ are much smaller than one and values of x are greater than or approximately equal to one. This imposes quite anisotropic sampling grids at the planes of interest and subsequently, requires corresponding anisotropic sampling and reconstruction kernels. Similar problem appears with the third option, specifically when one tries to match the sampling grids arising from the display ray generators and from cameras aimed at LF capture. This issue will be further clarified in Section 5. Therefore, (x, φ) is our choice for a ray-space, as illustrated in Figure 10(a). As seen in the figure, for $z=0$, the ray space representation has samples on a rectangular grid. These are individual rays as generated by the projecting engines. Upon propagating, the rays form nonrectangular grids for other values of z , as illustrated in Figure 10(b). It is worth mentioning that in practice, the grid is not perfectly rectangular even for $z=0$ with the ‘amount’ of non-regularity being proportional to the FOV_{proj} . However, this non-regularity is minor for typical use scenario and as such does not influence the properties of the LF. Therefore, for considerations in this paper we can assume that the grid is rectangular for $z=0$, that is, rays are equidistantly distributed throughout the FOV of the projection engines.

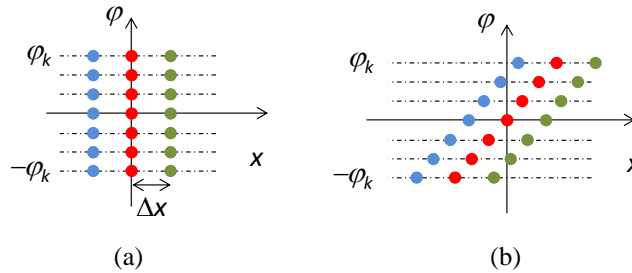


Figure 10. Ray-space example for an HPO system. (a) Ray space at $z=0$. (b) Ray space at distance z .

4.3 Frequency domain analysis of the ray-space representation

The ray space gives the relation between the angular resolution and the spatial (horizontal) resolution that a LF display can generate at a certain distance z from the projection engines. The non-rectangular ray-space sampling grid (i.e. the distribution of rays at a given plane) directly depends on the display setup. Assuming that the projection engines emit relatively regular angular-wise distribution of rays, the sampling grid at every plane is non-rectangular yet regular. Due to the regularity of the sampling grid, a sampling pattern can be identified at any ray space plane of ray propagation. This is essential for estimating the corresponding frequency domain support and its periodical replicas, following the theory of multi-dimensional signal sampling¹⁹. This in turn, defines the angular and spatial data throughput of the display.

Assume an arbitrary two-dimensional non-rectangular sampling pattern, as illustrated by green dots in Figure 11(a). The dots are at positions (x_k, φ_l) for $k, l \in \mathbb{Z}$, with \mathbb{Z} being the set of integers. The points of such regular sampling pattern can be expressed through the notion of lattice Λ , that is, as linear combinations with integer coefficients of two linearly independent vectors in \mathbb{R}^2 :

$$\Lambda = \{n_1 \mathbf{v}_1 + n_2 \mathbf{v}_2 | n_1, n_2 \in \mathbb{Z}\} \quad (6)$$

with $(\mathbf{v}_1, \mathbf{v}_2)$ being a set of basis vectors. If the basis vectors are expressed in a sampling matrix form as

$$\mathbf{V} = [\mathbf{v}_1 \quad \mathbf{v}_2] = \begin{bmatrix} v_1^{(x)} & v_2^{(x)} \\ v_1^{(y)} & v_2^{(y)} \end{bmatrix} \quad (7)$$

then the corresponding lattice is defined as $\Lambda = \text{LAT}(\mathbf{V})$. For a given sampling pattern, the sampling matrix \mathbf{V} is not unique since $\text{LAT}(\mathbf{V}) = \text{LAT}(\mathbf{E}\mathbf{V})$ where \mathbf{E} is any integer matrix with $|\det \mathbf{E}| = 1$.

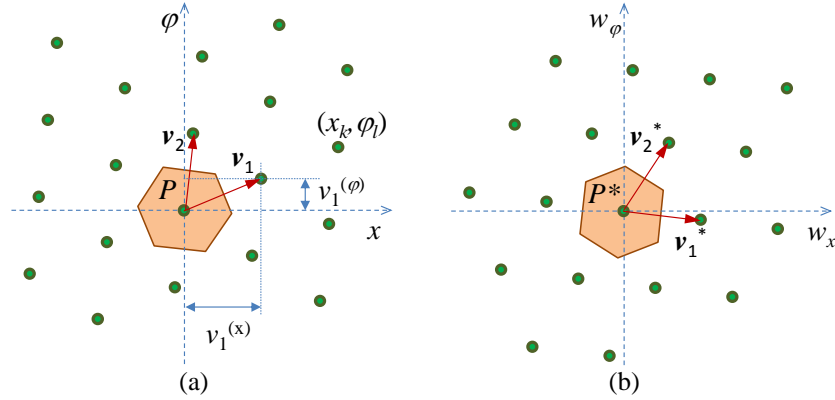


Figure 11. Non-rectangular sampling pattern (lattice) and correspond unit (Voronoi) cell. (a) Spatial domain. (b) Fourier domain.

A unit cell P of a given lattice Λ , is a set in \mathbb{R}^2 such that the union of sets centered on each lattice point covers the whole sampling space without overlapping. As with the sampling matrix, the unit cell is not unique. One of the possible unit cells is the Voronoi cell²⁰. The Voronoi cell 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 than is inside the cell than to any other lattice point. See Figure 11(a) for illustration.

The question here is what kind of continuous bandlimited function can be sampled by and correspondingly reconstructed from such a pattern. This can be evaluated in the Fourier domain. The Fourier transform of a continuous signal $f_c(x, \varphi)$ is given by

$$F_c(w_x, w_\varphi) = \iint f_c(x, \varphi) e^{-j2\pi(xw_x + \varphi w_\varphi)} dx d\varphi \quad (8)$$

with w_x and w_φ being the spatial frequencies in x and φ direction, respectively. Similarly, for a discrete signal $f(x, \varphi)$ with $(x, \varphi) \in \Lambda$, the discrete (lattice) Fourier transform is

$$F(w_x, w_\varphi) = \sum_{(x, \varphi) \in \Lambda} f(x, \varphi) e^{-j2\pi(xw_x + \varphi w_\varphi)}. \quad (9)$$

The lattice Fourier transform is periodic. For a given sampling pattern defined with lattice Λ and a sampling matrix \mathbf{V} as defined by (7), the periodicity of the Fourier transform (position of replicas in the frequency domain) is defined through the reciprocal lattice Λ^* , that can be evaluated as (see¹⁹ for more details)

$$\Lambda^* = \text{LAT}((\mathbf{V}^T)^{-1}) \quad (10)$$

and is illustrated in Figure 11(b). Due to the periodicity, $F(w_x, w_\varphi)$ is uniquely defined by its values in a unit cell of Λ^* , e.g. Voronoi cell P^* as shown in Figure 11(b). Consequently, the unit cell represents the bandwidth of the continuous signal that can be represented by and reconstructed from the given sampling pattern.

The effect of sampling a continuous signal with spectrum F_c with sampling structure given by lattice Λ in the frequency domain can be written as

$$F(w_x, w_\varphi) = \frac{1}{d(\Lambda)} \sum_{\mathbf{k} \in \Lambda^*} F_c(w_x, w_\varphi + \mathbf{k}). \quad (11)$$

As seen from this formula, the discrete spectrum contains the continuous spectrum F_c (for $\mathbf{k}=0$) and its replicas positioned on points defined by lattice Λ^* . If there is a non-zero overlapping between $F_c(w_x, w_\varphi)$ and $F_c(w_x, w_\varphi + \mathbf{k})$ for any \mathbf{k} other than $\mathbf{k}=0$, then aliasing occurred. In order to avoid aliasing, the spectrum of F_c has to be bandlimited to a unit cell P^* of Λ^* .

At the reconstructions stage, a continuous bandlimited function f_c can be reconstructed from its sampled version f based on the multidimensional sampling theorem applied to two dimensions as

$$f_c(x, \varphi) = \sum_{(\tilde{x}, \tilde{\varphi}) \in \Lambda} f(\tilde{x}, \tilde{\varphi}) h(x - \tilde{x}, \varphi - \tilde{\varphi}) \quad (12)$$

with

$$h(x, \varphi) = d(\Lambda) \int_{P^*} e^{j2\pi(xw_x + \varphi w_\varphi)} dw_x dw_\varphi \quad (13)$$

being the impulse response of an ideal lowpass (reconstruction) filter with passband P^* . The main purpose of the reconstruction filter is to remove spectrum replicas and keep only the baseband of the signal.

As mentioned before, the unit cell is not unique. This means, that signals with different frequency properties can be reconstructed from a given discrete representation on a grid described by a lattice Λ . This assumes that during the reconstruction, the reconstruction filter is tailored to the unit cell – in an ideal case it should be equal to the one given by Eq. (13), but in practice an approximation of the ideal one is to be used. As described in Section 3, in the case of projection based LF displays, the holographic screen is performing the discrete to continuous transformation. Therefore, the reconstruction filter is defined with the reconstruction properties of the holographic screen. Since the reconstruction function of the screen is approximately rectangular with some Gaussian type weights¹⁸, in this paper we adopt the use the Voronoi cell as the unit cell describing the bandwidth of the display. In particular, the properties of the Voronoi cell ensure that one get the most compact unit cell thereby threatening both directions (the ray space coordinates in 2D space) in a similar way. Furthermore, for a given sampling pattern, there is a one-to-one correspondence between the Voronoi cell in the Fourier domain, which is the available bandwidth of the display and the Voronoi cell in the spatial domain. Therefore, without loss of generality, when estimating the optimal camera setup we aim at matching the Voronoi cells of the display and camera rays in ray-space domain. This makes the overall optimization procedure faster. The frequency support can be easily estimated once the matching criterion is satisfied.

In the case under consideration, namely a projection-based LF display, we look into the ray-space representation – angle of rays vs. position. Out of this, we are able to estimate the required angular and spatial resolution for the given display. In ray space, the propagation of the rays changes the sampling grid based on the geometry of the system, thus changing the frequency support. These changes can be perceived as rotation of Voronoi cells and consequently rotation of the frequency support. This is illustrated in Figure 12. We will do the matching on the distance that corresponds to the position of the holographic screen, since that is the place at which the image is formed, that is, the screen serves as the discrete to continuous reconstruction filter.

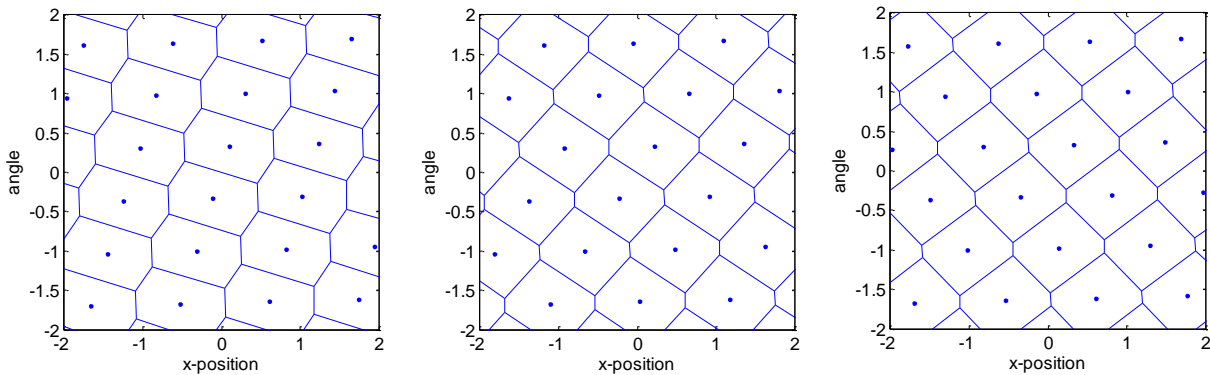


Figure 12. Example of ray-space vs. distance expressed through rotation of Voronoi cells – data for three different distances is depicted.

5. LF DISPLAY SPECIFIC CAMERA OPTIMIZATION

5.1 Display – camera (viewer) setup

In this paper we consider the whole system – from scene capture through visualization on an LF display to the observer with the goal of determining a minimalistic capture setup that provides sufficient information in terms of sampled rays for best possible representation of a scene by a given projection-based LF display. A simplified illustration of the system under consideration is shown in Figure 13. The three planes of interest are the ray-generator plane, the screen plane, and the viewing / camera plane. The position of the planes is defined through their distance from the ray generators. In the whole system, in addition to the light-ray generators, we also added the cameras capturing the LF. The rays entering the cameras behave as described by (5).

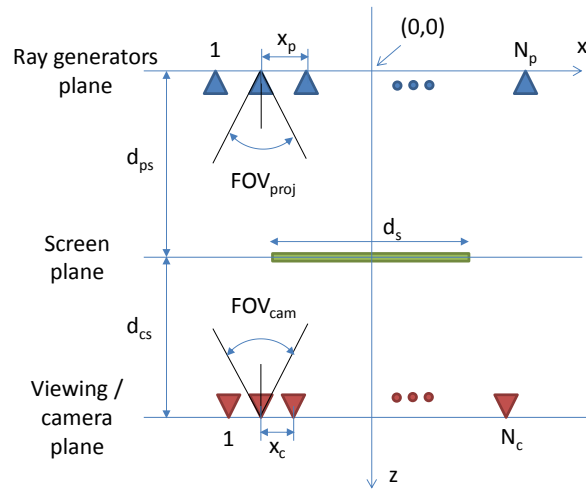


Figure 13. Display – camera (viewer) setup.

The projection engines generate a finite number of rays. These rays have to be available from the set of rays captured by the cameras. In the optimal case, for each projected ray, one needs a camera ray as in a one-to-one mapping. Since this might require a separate camera for every ray, such solution, although desirable, is not practical. Alternatively, one can aim at capturing a LF with the same bandwidth as the one supported by the display. Then, the generation of the projection rays becomes a bandlimited interpolation problem. Based on the discussion of the previous section, one can estimate the spatial and angular resolution (therefore the passband) that a given display is capable of producing. Furthermore, due to the special properties of the screen (see Section 3), the formation of the image is indirectly done on the screen (at the screen ray-space plane). Therefore we are mostly interested in the capability of the display to reconstruct the 3D content around the screen level and its bandwidth there. In practice, we aim at matching as well as possible the bandwidth of the display to the bandwidth of the cameras at the screen plane. Assuming a variable camera setup, we can optimize the camera position with the goal to minimize the overlapping between the camera and display bandwidth.

5.2 Camera optimization limitations

There are three major parts to be taken into consideration when optimizing the capture and visualization of 3D content on a projection-based LF display. All of them are influenced primarily by the display configuration. The design of the display predefines the following parameters:

- Projection engine's spatial resolutions and FOV_{proj}
- Distance between the projection engines
- Distance of the projection engines to screen plane and viewing plane

Those parameters uniquely define the bandwidth of the display as well as the geometrical configuration of rays that have to be reconstructed from the available camera rays.

In a theoretical consideration, the optical camera setup could be quite arbitrary. However, for practical purposes, the camera setup can be limited by following assumptions:

- All cameras are identical – same FOV_{cam} and spatial resolutions
- Practical (arbitrary or fixed) camera resolutions – we do not consider one-pixel cameras nor cameras with impractically large number of pixels
- FOV_{cam} equal or larger than FOV_{proj}
- Limited minimum camera-to-camera distance – cameras cannot be too close to each other (in practice this is limited by the physical size of a camera)
- Camera plane and viewer plane are the same (same distance from screen) – no scaling required between capture and visualization

The optimal camera setup is the one that would maximize the overlap between the spectra of the rays generated by the display and spectra of the rays captured by the camera setup. Accordingly, the optimization problem is to find the best camera setup for a given display by changing the camera-to-camera spacing and camera resolution taking into account aforementioned limitations.

5.3 Camera optimization criteria

Due to the correspondence between Voronoi cells evaluated for a given non-rectangular sampling grid and its spectral support, as discussed in Section 4.3, we do the similarity estimation in the ray-space domain by comparing Voronoi cell shapes as generated by the projection-engines and cameras. We used a simple three-part similarity criteria E given as:

$$E = |A_{cam} - A_{proj}| + |\Delta x_{cam} - \Delta x_{proj}| + |\Delta \varphi_{cam} - \Delta \varphi_{proj}|, \quad (14)$$

where A stands for the area of the Voronoi cell, Δx is the width of the Voronoi cell (related to the spatial resolution) and $\Delta \varphi$ is the height of the Voronoi cell (related to angular resolution). Subscripts *cam* and *proj* refer to camera and projection engines, respectively. These parameters are illustrated in Figure 14. In the optimal case E should be equal to zero. Please note that E does not have a physical meaning since it is defined as a sum of differences of different units (area, angle, and distance). It is only a similarity measure that we want to minimize, that is, the smaller the value of E , the better is the match between the camera and projection engine ray grid.

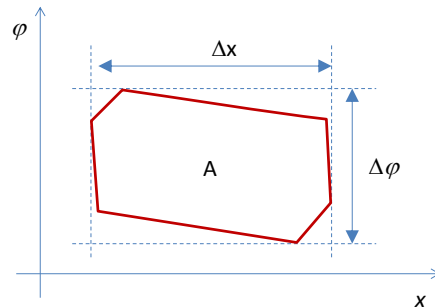


Figure 14. Similarity criteria parameters for Voronoi cells.

After determining the camera-to-camera distance, the number of required cameras is determined by the baseline that the cameras have to cover and the camera-to-camera distance, that is, more cameras are required when individual cameras are closer to each other since the required camera baseline is fixed by the display.

The overall methodology for estimating the optimal capture for visualization of 3D content on a LF display can be summarized as follows:

1. Based on the display setup, determine the display's ray space sampling grid at the screen plane
2. Calculate the corresponding Voronoi cell describing the display-based sampling grid
3. Find an optimal camera configuration by varying camera positions and resolutions and minimizing Eq. (14).

- Convert a captured scene to the one with estimated optimal camera parameters.

5.4 Example

In this section we show the evaluation of an optimal camera setup for a typical projection-based LF display with 112 projection engines generating a total of 65 MRays. For such display configuration, the ray-space representations for the three planes shown in Figure 13 are depicted in Figure 15. From the three planes under consideration, the most interesting one is the screen plane. The corresponding Voronoi cells for the screen plane are shown in Figure 15(d). As discussed earlier, for practical purposes we assume that all those cells are identical, that is, a single cell defines the angular-spatial bandwidth of the display.

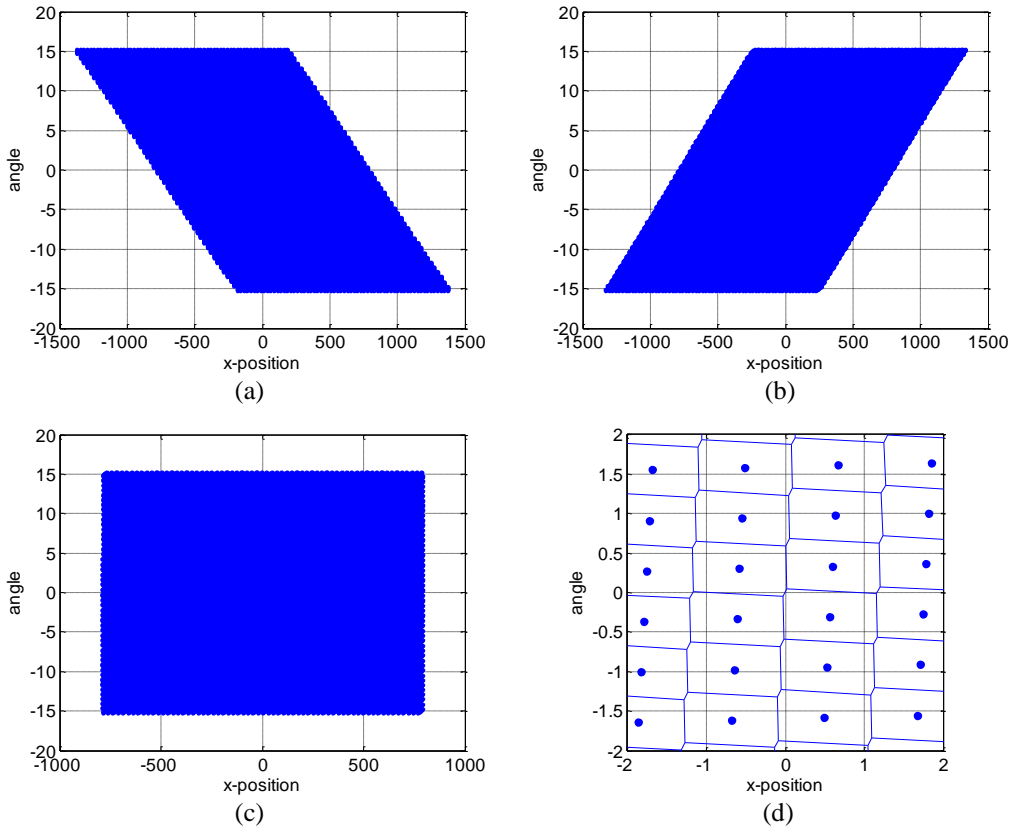


Figure 15. Ray-space representation of a projection-based LF display at different planes. (a) Ray-generators plane. (b) Viewer plane. (c) Screen plane. (d) Screen plane (zoomed in version) with corresponding Voronoi cells.

In the optimization procedure we are matching the Voronoi cells evaluated from the display's rays to Voronoi cells evaluated from the cameras' rays. The two free parameters are camera-to-camera distance x_c and the camera resolution. Since this is a very nonlinear optimization problem, we evaluated the similarity measure given by Eq. (14) on a dense grid for various values of x_c ($1 < x_c < 40$) and camera resolutions ($50\text{px} < \text{res} < 1600\text{px}$). The results of the optimization are shown in Figure 16(a). For better visualization, all values of the similarity criteria E above one have been thresholded to one since we are only interested in combinations of x_c and camera resolutions that result in small value of the similarity criteria.

Based on these results, two observations can be made. First, the similarity measure has many local minima corresponding to acceptable combinations of camera spacing and camera resolution that can be used for achieving a good capture setup for the given display. Second, optimal camera-to-camera spacing depends on camera resolution. This can be seen in Figure 16(b) that shows how the smallest value of the similarity measure changes depending on a given resolution, and Figure 16(c) that shows what the optimal camera-to-camera distance is for a given camera resolution.

The changes of the error (similarity measure) with camera-to-camera distance for a given camera resolution are shown in Figure 16(d). It is obvious that for a given camera resolution, the camera-to-camera distance has to be carefully adjusted.

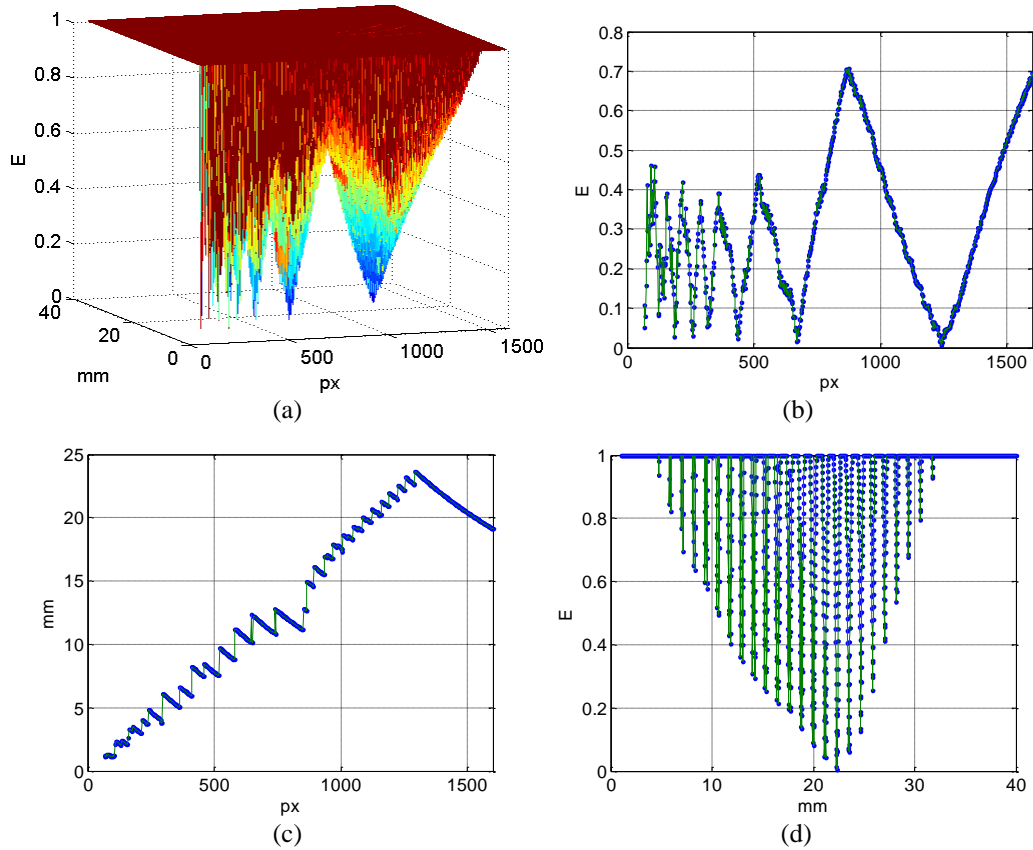


Figure 16. Camera optimization example. (a) Similarity measure with respect to camera-to-camera distance and camera resolution. (b) Smallest achievable similarity measure for a given resolution. (c) Optimal camera-to-camera spacing for a given resolution. (d) Similarity measure for fixed camera resolution (1240 px) with respect to camera-to-camera distance.

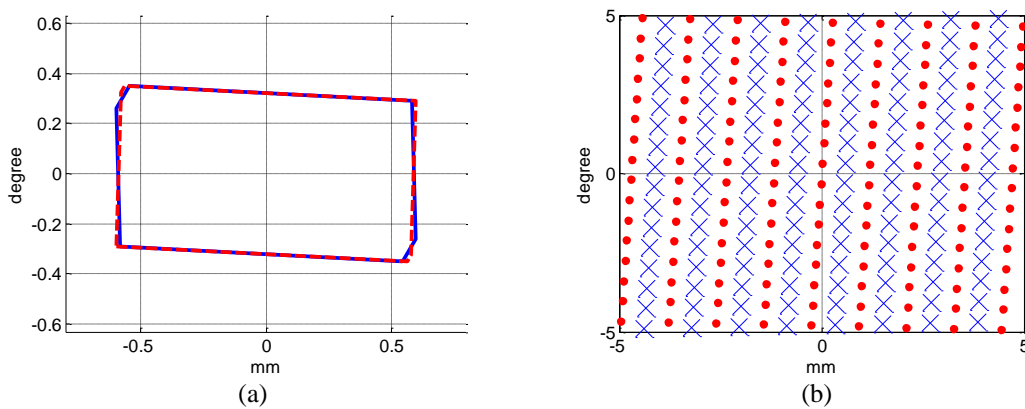


Figure 17. Optimized solution for $x_c = 22.36$ mm and 1240 pixel horizontal camera resolution – display rays (blue, solid line / crosses) and camera rays (red, dashed line / dots) at screen plane. (a) Voronoi cells. (b) Sampling grids.

The optimized solution for the camera horizontal resolution of 1240 pixels is shown in Figure 17. The display's sampling grid's Voronoi cell is almost perfectly matched by the cameras' sampling grid Voronoi cell. Consequently, the

sampling patterns are very well aligned – the sampling density along the camera rays and display rays is practically the same. This shows that a good match between the display and camera rays can be obtained making the camera to display ray interpolation efficient – using a minimalistic set of camera rays and maximizing the amount of information the display can visualize.

6. CONCLUDING REMARKS

In this paper we have presented a display-specific LF analysis, focusing on projection-based LF displays with a linear configuration of projection engines. However, the analysis is quite general and can be applied to other LF displays. At the core, the analysis only assumes that there are light generators capable of generating a finite number of rays on a regular (not necessary rectangular) grid. The configuration of such light generators put limits on the LF that a display can create, and consequently, following a similar analysis, put limits on the required scene capture setup.

Both LFs, the one reconstructed by the display and the one captured by the cameras are considered bandlimited. This allowed us to use a direct algorithm for matching the shape and size of the Voronoi cells obtained from the ray-space grid generated by the ray generators with the ones generated by the cameras. Such a match ensures that the two LF functions have the same bandwidth. There are two things that should be commented at this point. First, as discussed in various places¹⁷, a typical real-world visual scene is not bandlimited. Therefore, one cannot directly record the scene with the capture setup as determined in Step 3 of the algorithm described in Section 5.3. Instead, one has to ensure a proper anti-aliasing capture, that in practice means oversampling the scene (using more cameras than estimated) and then performing proper downsampling to the desired number of cameras. Second, the estimated resolution in Step 2 corresponds to the ideal resolution the display should be able to reconstruct based on the sampling pattern. This would assume that the holographic screen described in Section 3 has a reconstruction filter bandwidth equal to the estimated bandwidth of the display. However, in practice the reconstruction filter has a response that is more frequency restrictive than the estimated one¹⁸. This means that the display itself will smooth further the input rays. Nevertheless, this is not an issue since pre-filtering the data to the estimated bandwidth will eliminate all frequencies that cannot be properly treated by the real screen reconstruction filter.

The interpolation between the camera rays and rays available in the display has to be executed properly. Optimally, one need to do an interpolation on non-rectangular grids, with spectra of both signals limited as estimated by the ray-space analysis. Effectively, this need to be a custom resampling algorithm tailored to the capture and display setup. This can be done either on the camera plane or the screen plane. Both approaches have pros and cons and their analysis is a topic for our future work.

Finally, it should be evaluated which of the local minima (combination of camera resolution and camera-to-camera spacing) gives best visual result. The proposed similarity criterion is a quantitative criterion that does not take into account the properties of the human visual system. While we treat the two dimensions of the ray-space uniformly, subjective experiments should be conducted in order to find out whether camera resolution or camera-to-camera distance contribute more to the visual quality of the reconstructed 3D scene.

ACKNOWLEDGEMENT

This work is supported by the PROLIGHT-IAPP Marie Curie Action of the People Programme of the European Union's Seventh Framework Programme, REA grant agreement 32449.

REFERENCES

- ¹ Boev, A., Bregović, R. and Gotchev, A., "Signal processing for stereoscopic and multi-view 3D displays," *Handbook of signal processing systems*, 2nd edition, edited by Bhattacharyya, S., Deprettere, E., Leupers, R. and Takala, J., Springer, 3-47 (2013).
- ² IJsselsteijn, W., Seuntjens, P. and Meesters, L., "Human factors of 3D displays", in *3D Video Communication*, edited by Schreer, O., Kauff, P. and Sikora, T., Wiley (2005).
- ³ Pastoor, S., "3D displays", in *3D video Communication*, edited by Scheer, O., Kauff, P. and Sikora, T., Wiley, (2005).

- ⁴ Boev, A., Bregović, R. and Gotchev, A., “Visual-quality evaluation methodology for multiview displays,” *Displays*, vol. 33, 103-112 (2012).
- ⁵ Boev, A., Bregović, R. and Gotchev, A., “Measuring and modeling per-element angular visibility in multiview displays,” *Journal of the Society for Information Display*, vol. 18, 686–697 (2010).
- ⁶ Zwicker, M., Matusik, W., Durand, F. and Pfister, H., “Antialiasing for Automultiscopic 3D displays”, *Proc. of Eurographics Symposium on Rendering*, Cyprus (2006).
- ⁷ Kovács, P. T., Boev, A., Bregović, R. and Gotchev, A., “Quality measurements of 3D light-field displays,” in *Proc. 8th Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics, VPQM 2014*, Arizona, USA, 6 pages (2014).
- ⁸ Adelson, E. and Bergen, J., “The plenoptic function and the elements of early vision”, *Computational models of visual processing*, edited by Landy, M. and Movshon, J. A., MIT (1991).
- ⁹ Levoy, M. and Hanrahan, P., “Light field rendering”, *SIGGRAPH (Computer Graphics)*, New Orleans, 31-42 (1996).
- ¹⁰ Gortler, S. J., Grzeszczuk, R., Szeliski, R. and Cohen, M. F., “The lumigraph”, *SIGGRAPH (Computer Graphics)*, New Orleans, 43-54 (1996).
- ¹¹ Camahort, E. “4D light-field modeling and rendering”, PhD thesis, University of Texas at Austin (2001).
- ¹² Liang, C.-K., Shih, Y.-C. and Chen, H. H., “Light field analysis for modeling image formation”, *IEEE Trans. Image Processing*, 446-460 (2011).
- ¹³ Chai, J.-X., Tong, X., Chan, S.-C. and Shum, H.-Y., “Plenoptic sampling”, *SIGGRAPH (Computer Graphics)*, pp. 307-318, (2000).
- ¹⁴ Zhang, C. and Chen, T., “Generalized plenoptic sampling”, *Technical report AMP 01-06*, Carnegie Mellon University (2001).
- ¹⁵ Pearson, J., Brookes, M., and Dragotti, P. L., “Plenoptic layer based modelling for image based rendering”, *IEEE Trans. Image Proc.*, 3405-3419 (2013).
- ¹⁶ Criminisi, A., Kang, S. B., Swaminathan, R., Szeliski, R., and Anandan, P., “Extracting layers and analyzing their specular properties using epipolar plane image analysis”, *Comput. Vis. Image Und.* 97, 51–85 (2005).
- ¹⁷ Gilliam, C., Dragotti, P. L., and Brookes, M., “On the spectrum of the plenoptic function”, *IEEE Trans. Image Processing*, 502-516 (2014).
- ¹⁸ Balogh, T., “The HoloVizio system”, *Proc. SPIE 6055, Stereoscopic Displays and Virtual Reality Systems XIII*, San Jose, (2006)
- ¹⁹ Dubois, E., “Video sampling and interpolation”, chapter 2 in *The essential guide to video processing*, edited by Bovik, J., Academic Press (2009).
- ²⁰ Aurenhammer, F., “Voronoi Diagrams – A survey of a fundamental geometric data structure”, *ACM Computing Surveys* 23, 245-405, (1991).