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

Previous generations of 3D displays (e.g. stereoscopic displays, autostereoscopic displays) had 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, for the price of high data throughput, complex data acquisition (sensing) 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 sensing (acquisition) through LF based modeling and representation up 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 (scene sensing) 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. The overall process is demonstrated on the state of the art displays produced by Holografika.

Keywords: Plenoptic function, light field, non-uniform sampling, ray space, 3D displays, Voronoi cells

1. INTRODUCTION

3D displays are the next big step in immersive technology. Although, the quality of the current generation off-the-shelf 3D displays is still not good enough for a perfect visual merge between the physical and artificial world, the 3D display technology reached a point at which 3D displays become part of our everyday life (e.g. projection of movies in cinemas and living rooms, visualization of research data, streamlining production and design). Today's mass-produced 3D displays are developed around two approaches [1]. First approach is based on 3D displays with glasses. These displays create a 3D visualization of a scene by a combination of a screen that sends alternately (spatially or time multiplexed) images to be received by the left and right eye. The separation of images at the viewer side is achieved by glasses – typically shutter or polarization glasses. In this way each eye sees a slightly different, properly prepared (pre-processed) image thereby enabling the illusion of 3D vision. The second approach is based on autostereoscopic displays. In such displays, each pixel on the screen is visible only from a specific observation points in front of the display. This is typically achieved by adding in front of the pixels (TFT matrix) either a parallax barrier or a lenticular sheet. In this way, each eye sees a different set of pixels (and consequently a different image), again enabling a 3D illusion of the scene assuming that the data presented on the display is properly prepared.

Unfortunately, both of those approaches have limitations. The glasses-based displays require a viewer to wear glasses (source of discomfort) and are capable of reproducing only two views of a scene (identical two views are visible independently of the position of the observer). The autostereoscopic displays reproduce typically from two to 20+ views (enabling different perspective of a scene), but a viewer has to position himself / herself precisely with respect to the screen (each eye has to be in the center of a view). Otherwise, due to the crosstalk between the adjacent views (view overlapping) the 3D illusion is lost and, moreover, the views seen by the viewer can become very disturbing. Furthermore, due to the low number of views generated by an autostereoscopic display (mostly limited by the maximum size of a TFT matrix that can be produced by today's technology), the realism of the scene (e.g. moving around the scene) is still lacking [2].

The ultimate aim of a 3D display is to recreate a 'perfect' 3D visualization of a scene. Here, 'perfect' implies that a human observer cannot differentiate a scene visualized by a display from the real one. In practice, this means that a display should be capable of reconstructing the continuous light field (LF), as emitted by the scene, from each

observation point and not just from two observation points like in the case of displays with glasses or several observation points like in the case of autostereoscopic displays. A continuous reconstruction of the LF describing the scene would enable an observer to see an uninterrupted visualization of the scene when moving in the front of the display (continuous parallax) [3]. This would be similar to observing the real scene.

There are two major problems that have 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, that is, it must have large number of multiview pixels (pixels with directional visibility – different information represented by the pixel is seen depending on observation direction). The most promising candidates, as of today, to achieve this are so-called projection-based LF displays. These displays are built from projection engines (ray generators) and a custom-build holographic screen. Such displays, although not perfect, are capable of achieving acceptable spatial resolution and a good enough motion parallax (angular resolution). The second problem lies in the large amount of data required for driving 'perfect' 3D displays. The large amount of data describing the scene has to be captured, properly processed, stored, transmitted to the display, and finally visualized on the display thereby making the overall process computationally very intensive. Each of those two problems must be solved, either separately or simultaneously, before displays capable of ultra-realistic representation of 3D scenes will become a mainstream technology.

Since today's LF displays are not perfect, it is important to understand the capabilities and limitations of such displays. Knowing display's limitations, the 3D content that one wants to visualize can be tailored to a given display, thereby achieving the best possible visualization of an arbitrary content on a given display.

In an earlier work [4,5] we have shown how to estimate the passband of a multiview display, that is, the 'visual' bandwidth (amount of information) that a multiview display can show without or with acceptable distortion. In contrast to [6] that presented a technique for estimating the display passband for autostereoscopic displays, the proposed methods in [4,5] are measurement based thereby taking into account various nonlinear distortions that different parts of the display can introduce but that cannot be easily modeled. As it will be seen in this paper, in comparison to autostereoscopic displays where the ray-space parameterization of the LF is uniform (consequence of the way how the rays are generated) and therefore it is quite straightforward to analyze such displays (as well as pre-process data to be visualized on such displays), in projection-based LF displays the distribution is nonuniform. This makes the overall display analysis more complex.

In this paper we will analyze the current generation of projection-based LF displays and discuss what kind of data can be represented on such displays and what the necessary capture setup looks like for getting the best possible visualization of the scene on a given projection-based LF display. In comparison to the measurement-based analysis of LF displays that has been presented in [7], here we will look into an analytical approach, namely, we will analyze projection-based LF displays in ray-space. We will express the 3D visualization capability of a display in terms of angular and spatial resolution through its effective display bandwidth. This will give a better understanding about the content that can be represented on the display and we will be able to estimate the minimalistic capture setup (number and position of cameras with a given properties) that is required for recording that content.

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 is discussed in Section 5 with Section 6 showing a typical use case. Finally, concluding remarks are given in Section 7.

2. PLENOPTIC FUNCTION AND LIGHT FIELD

Every point on an object surface emits light into surrounding space, either being an active source of light or merely reflecting incoming light. The light propagation through space can be described, in the most complete form, by wave-optics theory or, in a simpler form, by ray-optics theory. In this paper we will formalize the light through its geometric behavior as described by the ray-optics theory. In the ray-optics theory, it is assumed that every point in space emits infinite number of directional rays. Mathematically, this is expressed through the notion of the plenoptic function (PF) [8]. PF is a 7D continuous function that describes the amount of light through every point in space in every direction and can be denoted as

$$P(\theta, \varphi, \lambda, t, V_x, V_y, V_z), \tag{1}$$

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 also Figure 1 for illustration). Implicitly, the PF describes the underlying LF.

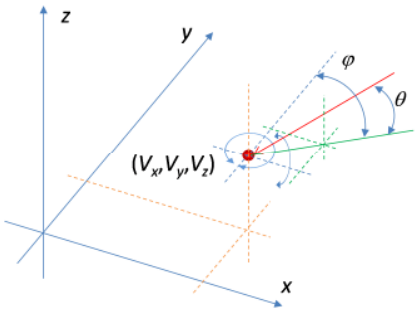
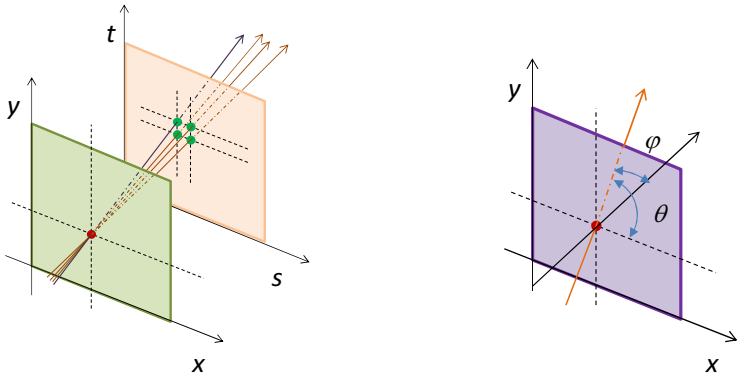


Figure 1. Illustration of the 7D PF.

Although the continuous 7D PF is the most complete (formal) description of a LF, it is not really useful in practice due to the enormous amount of data required for its description that cannot be efficiently captured, processed, or reproduced. Therefore, several simplifications of the 7D PF have been introduced. 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].



(a) (b)

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

Due to practical (technical) limitations, most of the today's 3D displays support horizontal parallax only (HPO) and ignore vertical parallax. The benefit of an HPO display over a full-parallax one is in the considerably reduced amount of data (rays) that has to be recorded and reproduced. Our discussion in this paper assumes HPO displays, however it can be easily extended to full-parallax displays. Ignoring the vertical parallax allows omitting the parameters y and θ in the (x, y, φ, θ) parameterization (y and t in the (x, y, s, t) parameterization) as all rays behave in the same way along y – axis. Consequently, the two parameterizations from Figure 2 are simplified to their 2D versions illustrated in Figure 3. The relation between those two, (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.

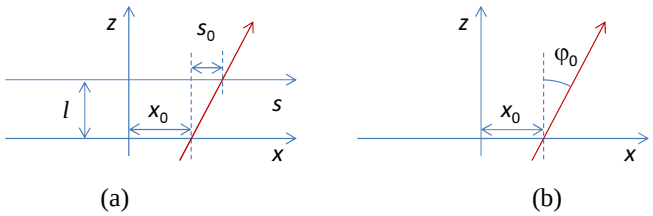


Figure 3. HPO LF parameterizations. (a) Two-plane parameterization. (b) One plane and direction parameterization.

The propagation of the LF [12] from one plane (Plane 1) to second plane (Plane 2) that are distance d apart can be expressed as

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

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 4). 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 (3), is a shifting operation along the x-axis (also referred to as shearing in the LF context). This is illustrated in Figure 5. 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).

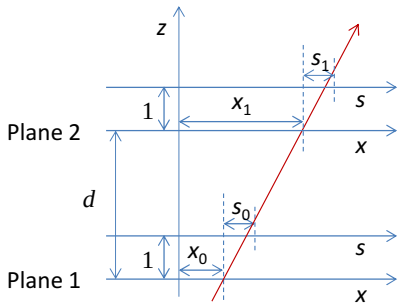


Figure 4. Light (ray) propagation – Representation on two different planes.

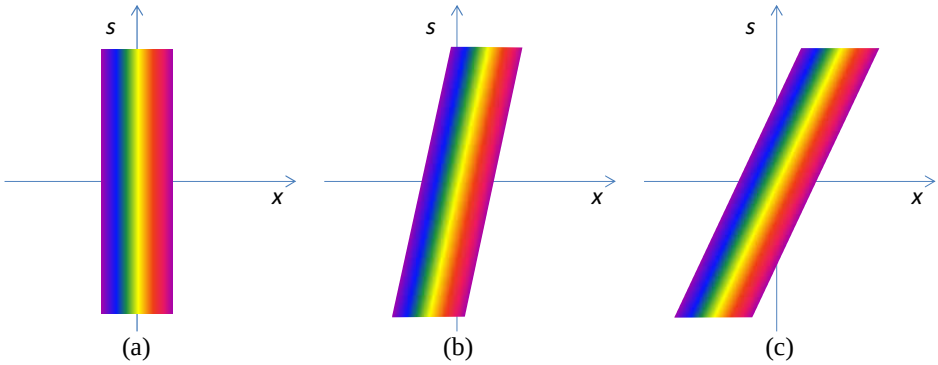


Figure 5. LF at different planes (positions). (a) For d=0 (on the surface of a flat object). (b) For d=d_0. (c) For d=2d_0.

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. This is discussed in the following section.

2.1 Plenoptic sampling

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 band-limited. They are formed by objects at different depths with sharp transitions at edges and occlusions depending on the viewing ray directions. Still, band-limited sampling theory has played an important role in studying the plenoptic sampling. Strictly speaking, a band-limited LF would arise from a scene with no occlusions (all points in the scene are visible from all observation points), which also satisfies the Lambertian property (a point in space has same color and intensity from all observation angles). It has been shown that for such scenes sampling intervals along axes in the used LF parameterization can be directly related with and interpreted in terms of camera spatial resolution, camera-to-camera spacing, and the number of depth planes resolved. A detail analysis of sampling and reconstructing such scenes can be found in [13,14]. Here, we will emphasize only few points related to plenoptic sampling that are needed in the context of this paper.

Consider a Lambertian scene without occlusions as shown in Figure 6(a). The continuous frequency support of such scene, when using the two-plane parameterization, is given by blue cones in Figure 6(b). It should be noted that the support is limited only by the minimal, z_{min}, and maximal, z_{max}, distance of the object from the camera plane. In the extreme case, for objects at infinity or at the focal plane, the spectral support is limited to the green cones. Sampling leads to replication of this spectrum determined by the sampling frequencies along *x* and *s*. Therefore, before sampling, the spectrum should be limited in order to avoid aliasing. The frequency support (only base spectrum) in the discrete domain is shown in Figure 6(c). Assuming the two planes are co-located with the camera plane and focal plane respectively, the repetition of the LF spectrum in the horizontal direction depends on the density of cameras and in the vertical direction it depends on the horizontal camera resolution. In order to reconstruct the continuous spectrum out of the discrete one, the sampling density must be such that the spectral replicas do not overlap with the base spectrum.

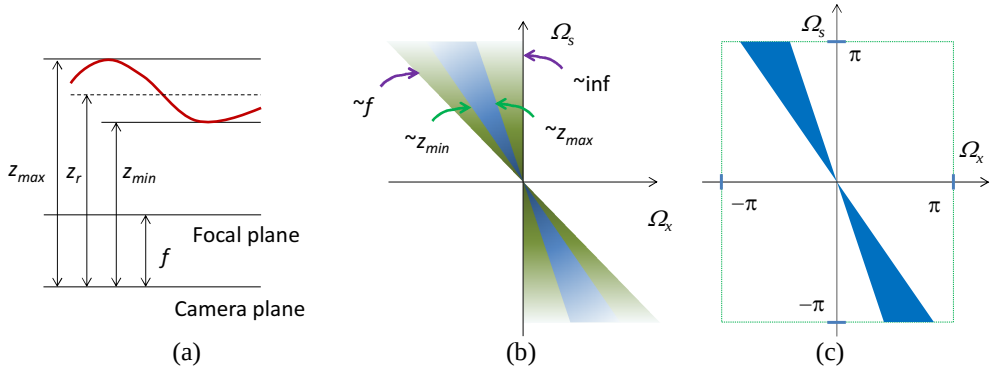


Figure 6. Example of a Lambertian scene without occlusion. (a) Scene. (b) Continuous frequency support (two-plane parameterization). (c) Frequency support in the discrete domain (two-plane parameterization).

Another factor that has to be taken into account when sampling a scene, is the technique that will be used for reconstruction. Depending on the additional knowledge about the scene (e.g. depth values of points in the scene), different number of images is required as illustrated in Figure 7 [13]. The more information about depth we possess, the fewer images are needed for scene reconstruction.

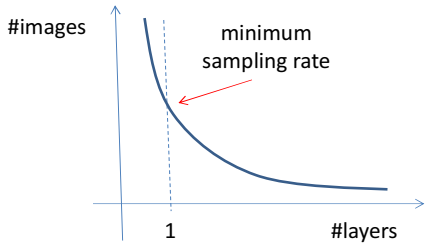


Figure 7. Plenoptic sampling – minimum sampling rate.

If only one reconstruction plane is used (e.g. due to simplicity or lack of information related to depth), the optimal reconstruction plane z_r, as shown in Figure 6(a), can be estimated from the minimum and maximum depths in the scene as

$$z_r = \frac{2}{\frac{1}{z_{min}} + \frac{1}{z_{max}}}. \quad (4)$$

When more layers for rendering are used, then every layer can be treated in a similar manner. More complex models for scene description are required for reconstruction of scenes with occlusions [15,16,17].

From a display perspective, the scenes to be reconstructed by the display are considered band limited. There is a twofold reason for this assumption. Fist, 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 band limited 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 8).

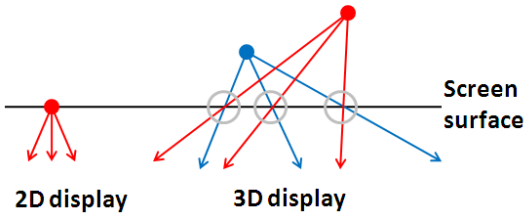


Figure 8. Comparison of light rays emitted from a 2D display and a 3D display. 2D displays emit the same color to all directions from a single pixel (left). 3D displays must emit different colors to different directions to show a 3D scene (right). Notice direction selective light emission on the "3D 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 displays [18].

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 9).

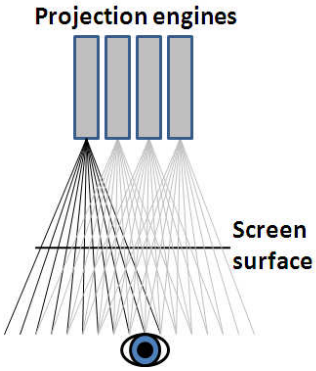


Figure 9. Light rays emitted by a single projection engine (black rays) are not seen from a single position, as they are emitted at different screen positions and different directions. The image seen by a single eye is made up of light rays originating from many sources.

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.

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

For determining the capabilities of projection-based LF displays in visualizing 3D content, in the following sections we will perform a frequency domain analysis of such displays.

4.1 Ray distribution of an light-field display

An HPO projection-based LF display consists of several projection engines (light ray generators) and a special holographic screen. In this paper we assume a basic linear setup of the projection engines as illustrated in Figure 10. 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

$$x_p = d_p/(N_p - 1). \tag{5}$$

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.

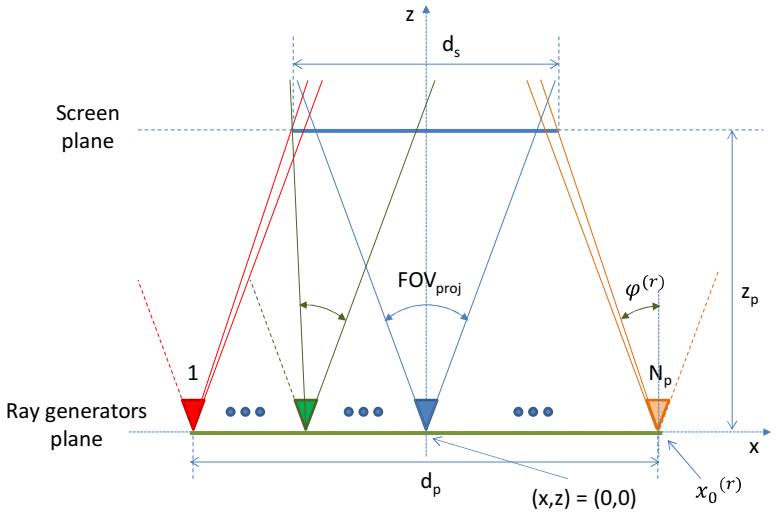


Figure 10. LF display setup.

For the analysis in this paper 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 11. 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 nonunifomity can be ignored.

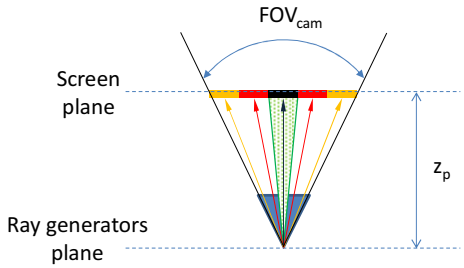


Figure 11. Example of ray distribution over FOV_{proj} in one projection engine (light ray generator).

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

$$x_z^{(r)} = x_0^{(r)} + z\,tan(\varphi^{(r)}) \tag{6}$$

with $\varphi^{(r)}$ being the angle of the ray under consideration as illustrated in Figure 10. 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 uniform grid, this is not the case for rays originating from different projection engines as discussed in Section 3.

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 12. [12]

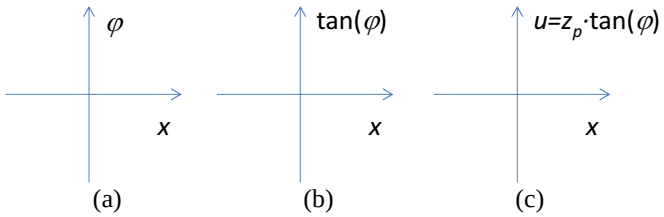


Figure 12. LF ray-space parameterizations. (a) Ray angle vs. position. (b) Two-planes with second plane being at unit distance. (c) Two-planes with second plane being at the screen distance.

First option is to have samples on the (x, φ) system (Figure 12a), thus indexing the rays through their position and direction (angle). The second option, as in Figure 12b, 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 $(x, z_p\, tan(\varphi))$ system (Figure 12c), which is essentially a two-plane parameterization with the second plane located at the screen level (c.f. Figure 10). 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 uniform 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 13(a). As seen in the figure, for $z = 0$, the ray space representation has samples on a uniform grid. These are individual rays as generated by the projecting engines. Upon propagating, the rays form nonuniform grids for other values of z, as illustrated in Figure 13(b). It is worth mentioning that in practice, the grid is not perfectly uniform even for $z = 0$ with the 'amount' of nonuniformity being proportional to the FOV_{proj}. However, this nonuniformity 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 uniform for $z = 0$, that is, rays are equidistantly distributed throughout the FOV of the projection engines.

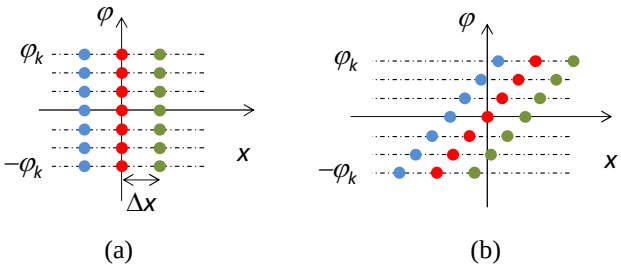


Figure 13. Ray-space example for an HPO system. (a) Ray space at $z = 0$. (b) Ray space at distance z.

As discussed in Section 2, the PF describing the LF is a continuous function. However, a projection-based LF display can only generate a finite number of rays. This implies that the discrete LF generated by projection-based LF displays represents a sampled version of the underlying continuous PF. Due to the nonuniform structure of the samples in the ray space, the first problem to be solved (analyzed) is to determine which kind of PF can be properly represented by the available rays (LF samples) as generated by a LF display. For this we will analyze the behavior of rays along z direction (distance from projection engines) in the frequency domain.

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 nonuniform ray-space sampling grid (distribution of

rays at a given plane) directly depends on the display setup. Assuming that the projection engines emit relatively uniform angular-wise distribution of rays, the sampling grid at every plane is nonuniform but regular. Due to the regularity of the sampling grid, a sampling pattern can be identified. We know from the nonuniform sampling theory [19] that for such grid we can estimate the corresponding frequency domain support (characteristic), which in turn, defines the angular and spatial data throughput of the display. This will be elaborated in more detail next.

Let us assume an arbitrary two-dimensional nonuniform regular sampling pattern, as shown by green dots in Figure 14(a). The dots are at positions (x_k, y_l) for $k,l \in \mathbb{Z}$, with $\mathbb{Z}$ being the set of integers. In the sampling theory, 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 $\mathbb{R}^2$:

$$\Lambda = \{n_1\boldsymbol{v}_1 + n_2\boldsymbol{v}_2 | n_1, n_2 \in \mathbb{Z}\} \tag{7}$$

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

$$\boldsymbol{V} = [\boldsymbol{v}_1 \quad \boldsymbol{v}_2] = \begin{bmatrix} {v_1}^{(x)} & {v_2}^{(x)} \\ {v_1}^{(y)} & {v_2}^{(y)} \end{bmatrix} \tag{8}$$

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

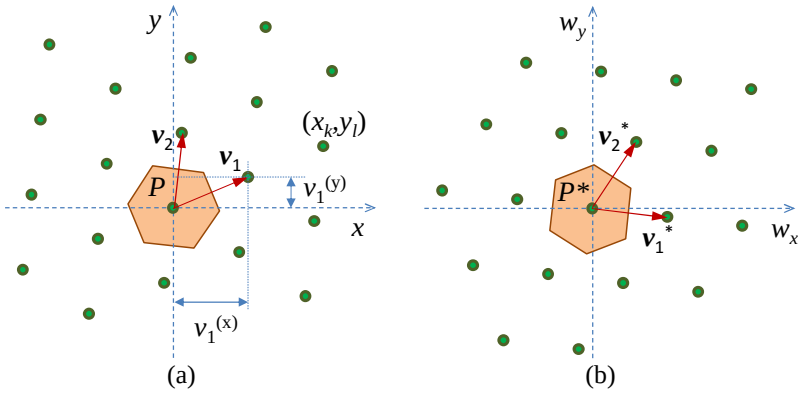


Figure 14. Nonuniform sampling pattern (lattice) and correspond unit (Voronoi) cell. (a) Spatial domain. (b) Fourier domain.

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 cell [20]. 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 14(a) for illustration.

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

$$F_c(w_x, w_y) = \iint f_c(x, y)e^{-j2\pi(xw_x+yw_y)}dxdy \tag{9}$$

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

$$F(w_x, w_y) = \sum_{(x,y)\in\Lambda} f(x, y)e^{-j2\pi(xw_x+yw_y)} \tag{10}$$

The lattice Fourier transform is periodic. For a given sampling pattern defined with lattice Λ and a sampling matrix $\boldsymbol{V}$ as defined by (8), 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 (see [19] for more details)

$$\Lambda^* = \mathrm{LAT}((\boldsymbol{V}^T)^{-1}) \tag{11}$$

and is illustrated in Figure 14(b). Due to periodicity, $F(w_x, w_y)$ is uniquely defined by its values in a unit cell of Λ^*, e.g. Voronoi cell as shown in Figure 14(b). Consequently, the unit cell represents the bandwidth of the continuous signal that can be represented by (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

$$F(w_x, w_y) = \frac{1}{d(\Lambda)} \sum_{\boldsymbol{k} \in \Lambda^*} F_c(w_x, w_y + \boldsymbol{k}). \tag{12}$$

As seen from this formula, the discrete spectrum contains the continuous spectrum F_c (for $\boldsymbol{k}$=0) and its replicas positioned on points defined by lattice Λ^*. If there is a non-zero overlapping between $F_c(w_x, w_y)$ and $F_c(w_x, w_y + \boldsymbol{k})$ for any $\boldsymbol{k}$ other than $\boldsymbol{k}$=0, than 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

$$f_c(x, y) = \sum_{(\tilde{x}, \tilde{y}) \in \Lambda} f(\tilde{x}, \tilde{y}) h(x - \tilde{x}, y - \tilde{y}) \tag{13}$$

with

$$h(x, y) = d(\Lambda) \int_{P^*} e^{j2\pi(xw_x + yw_y)} dw_x dw_y \tag{14}$$

being the impulse response of an ideal lowpass (reconstruction) filter with passband P^*. The main purpose of the reconstruction filter is to remove or 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 (14), but in practice an approximation of the ideal one has 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 weighs [18], in this paper we will use the Voronoi cell as the unit cell describing the bandwidth of the display. The properties of the Voronoi cell, as mentioned earlier, will ensure that we get the most compact unit cell thereby threating both directions (variables in 2D space) in a similar way.

Finally, as discussed above, for a given sampling pattern, there is a one-to-one correspondence between the Voronoi cell in the Frequency domain (that in turn is the available bandwidth of the display) and Voronoi cell in the spatial domain. Therefore, without loss of generality, when estimating the optimal camera setup we will compare (match) Voronoi cells of the display and camera rays in spatial domain. This makes the overall optimization procedure, described in the following sections, 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 will be 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 15. 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.

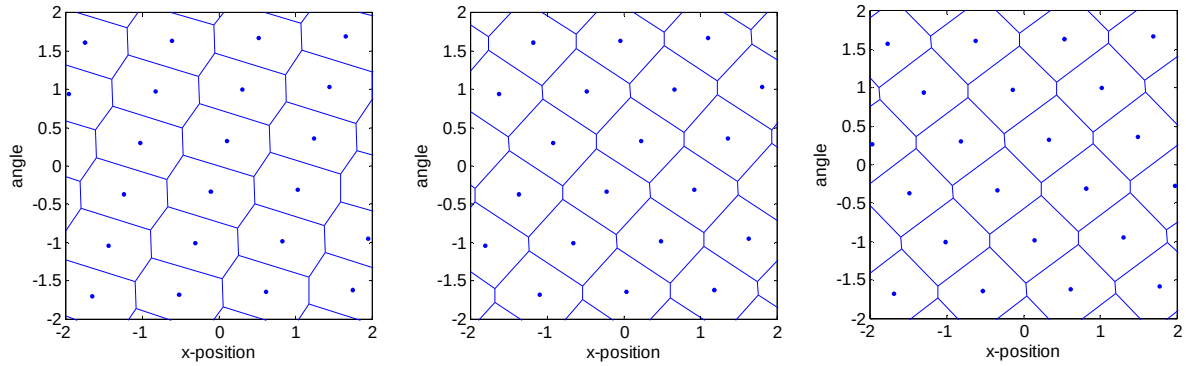


Figure 15. Example of ray-space vs. distance expressed through rotation of Voronoi cells – data for three different distances is depicted.

After determining the frequency support of the display, we can have a look at behavior of other parts of the system under consideration, that is, the desired camera setup, in order to optimize the overall system, from capture to display.

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 will provide enough data (information) for best possible representation of a scene on a given projection-based LF display. A simplified illustration of the system under consideration is shown in Figure 16. The three planes of interest are the ray-generators plane (projection engine 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 – devices that will capture the LF. From the ray distribution point of view, rays entering the cameras behave in the same way as rays living the ray generators – both of those sets of rays follow the expression given by (6).

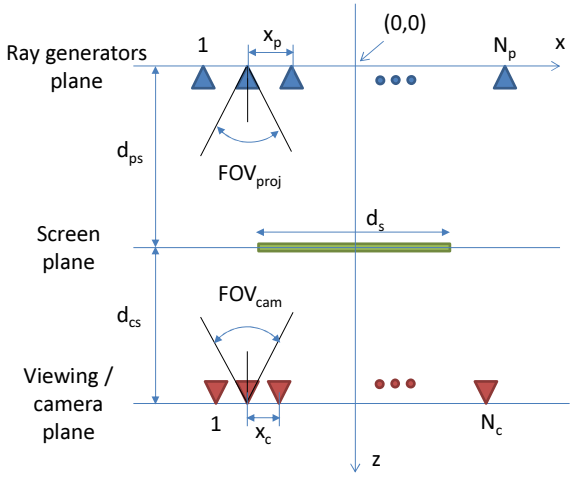


Figure 16. Display – camera (viewer) setup.

The projection engines generate a finite number of rays. These rays have to be reconstructed from rays captured by the cameras that correspond to a discrete version of the LF generated by the scene under consideration. The ray reconstruction itself is an interpolation problem. In the optimal case, for each projected ray, we would like to have a camera ray thereby making the interpolation a simple one to one mapping. Since this might require a separate camera for every ray, such solution, although desirable, is not practical. Therefore, we have to look into solutions where we can use

a finite (as small as possible) number of cameras of a given resolution and still being able to perform good reconstruction of rays that can be generated by the projection engines.

Following the discussion of the previous section, we know how to estimate the spatial and angular resolution (passband) that a given display is capable to produce. We can also apply the same bandwidth estimation for an arbitrary camera setup. Furthermore, due to the special properties of the screen (see Section 3), the formation of the image is indirectly done on the screen (screen plane). Therefore we are mostly interested in the capability of the display to reconstruct the 3D content around the screen level. In practice, this means that we want to match 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. This will be shown in the following section.

5.2 Camera optimization limitations

As seen in the previous section, 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 (display design). The design of the display predefines 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 will uniquely define the bandwidth of the display as well as the geometrical configuration of rays that have to be reconstructed from 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 want one pixel cameras nor cameras with impractically large number of pixels (non-existing sensor size)
- 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 nonuniform regular sampling grid and its spectral (frequency) support, as discussed in Section 4.3, we do the similarity estimation in the spatial (ray-space) domain, that is, we are comparing Voronoi cells evaluated from the grids generated by the projection-engines and cameras. If two sampling grids have similar shaped Voronoi cells (spatial offset between the cells is unimportant), then those two will have similar frequency support. In this paper we used a simple three-part similarity criteria E given as:

$$E = \left|A_{cam} - A_{proj}\right| + \left|\Delta x_{cam} - \Delta x_{proj}\right| + \left|\Delta\varphi_{cam} - \Delta\varphi_{proj}\right|, \tag{15}$$

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 17. 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.

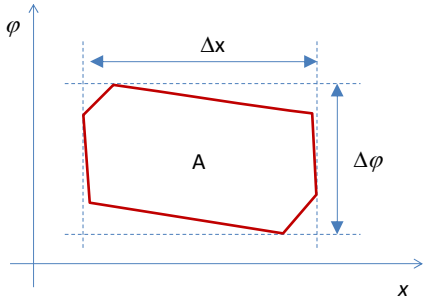


Figure 17. Similarity criteria parameters for Voronoi cells.

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 (15).
4. Convert a captured scene to the one with estimated optimal camera parameters.

This will be illustrated, in the following section, by means of an example for a typical LF projection-based display. Before that, there are two things that should be pointed out at this point. First, as discussed in various places [17], a typical real scene is not bandlimited. Therefore we can not only record the scene with the capture setup as determined in Step 3 of the algorithm. Instead, we have to ensure proper anti-aliasing capture, that in practice means oversampling the scene (using more cameras than estimated) and then do proper downsampling to the desired number of cameras. Second, the estimated resolution in Step 2 corresponds to ideal resolution the display should be able to reconstruct based on the sampling pattern. This would assume that the holographic screen (see Section 3) has a reconstruction filter equal to the estimated bandwidth of the display. However, in practice the reconstruction filter has a more rectangular response that is more restrictive than the estimated one [18]. This means that the display itself will smooth further the data we provide. 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.

6. EXAMPLE

In this section we will 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 16 are depicted in Figure 18. 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 19. 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.

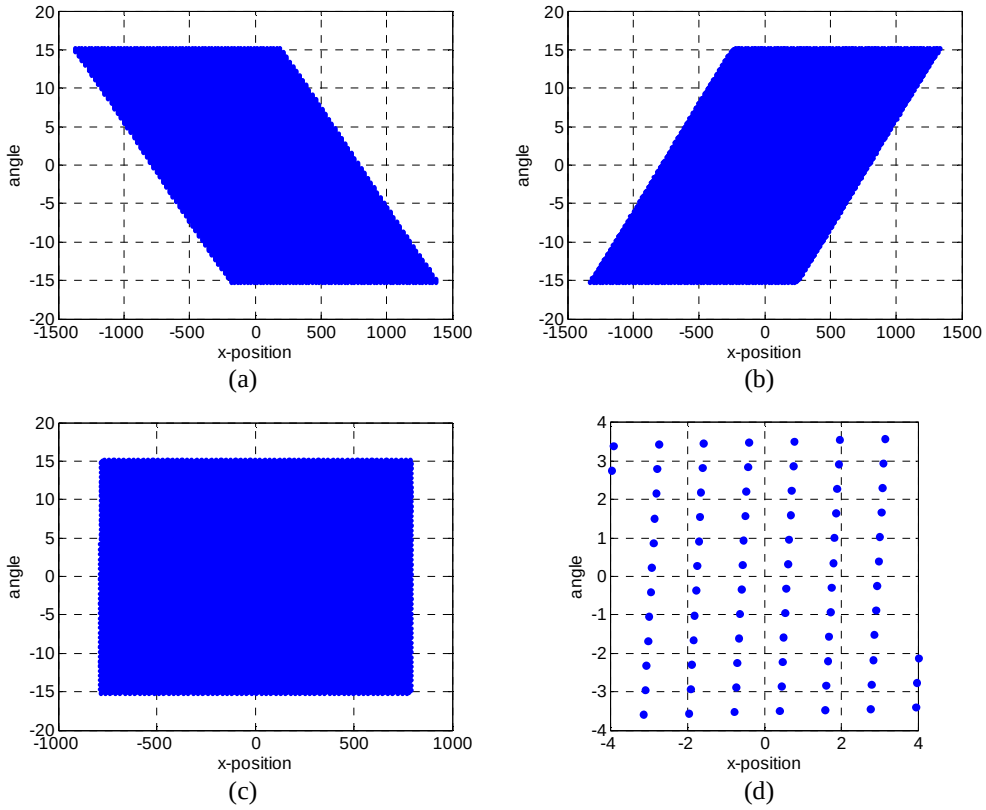


Figure 18. Ray-space representation of a projection-based LF display at different planes. (a) Ray-generators plane. (b) Viewer plane. (c) Screen plane. (d) Screen plane – zoomed in version.

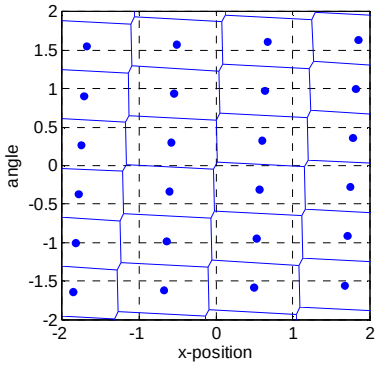


Figure 19. Voronoi cells for a projection-based LF display in ray space at the screen plane.

In the optimization procedure we are matching the Voronoi cells evaluated from the display's rays (spectrum supported by the display) to Voronoi cells evaluated from the Cameras' rays (spectrum captured by cameras). 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 (15) on a dense grid for various values of x_c ($1 < x_c < 40$) and camera resolutions (50px < res < 1600px). The results of the optimization are shown in Figure 20. 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, there are many good combinations of camera spacing and camera resolution that can be used for achieving a good capture setup for the given display (the similarity measure has many local minima). Second, optimal camera-to-camera spacing depends on camera resolution. This can be seen in Figure 21(a) that shows how the smallest value of the similarity measure changes depending on a given resolution, and Figure 21(b) 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 22. It is obvious that for a given camera resolution, the camera-to-camera distance has to be carefully adjusted.

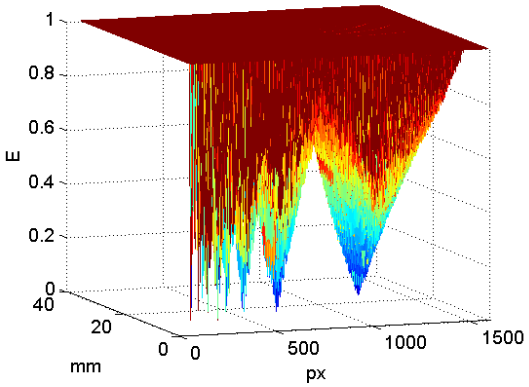


Figure 20. Camera optimization – similarity measure with respect to camera-to-camera distance and camera resolution.

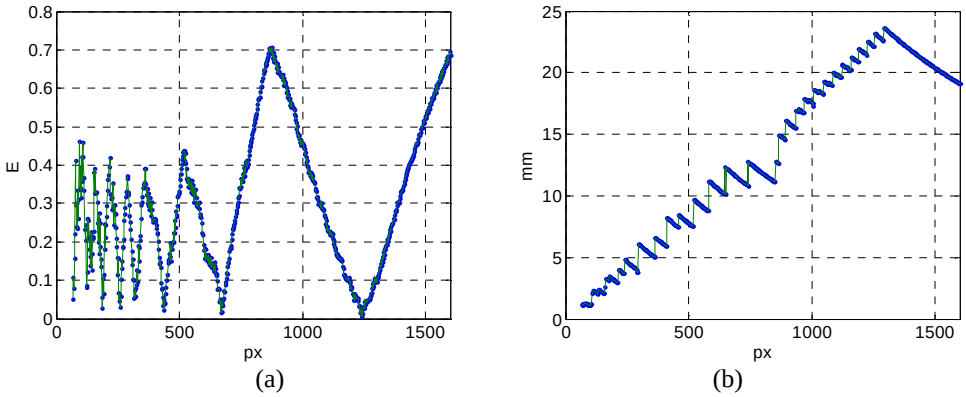


Figure 21. Camera optimization – compromise between camera-to-camera spacing and camera resolution. (a) Smallest achievable similarity measure for a given resolution. (b) Optimal camera-to-camera spacing for a given resolution.

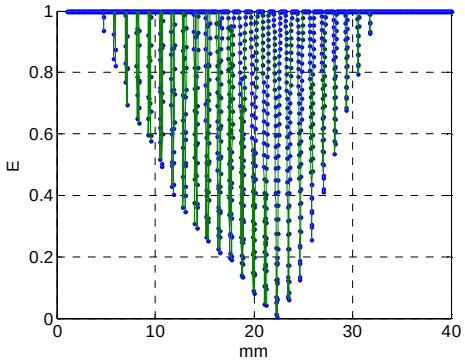


Figure 22. Similarity measure for fixed camera resolution (1240 px) with respect to camera-to-camera distance.

The optimized solution for camera horizontal resolution of 1240 pixels is shown in Figure 23. The display's sampling grid's Voronoi cell is almost perfectly matched with 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 similar. 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.

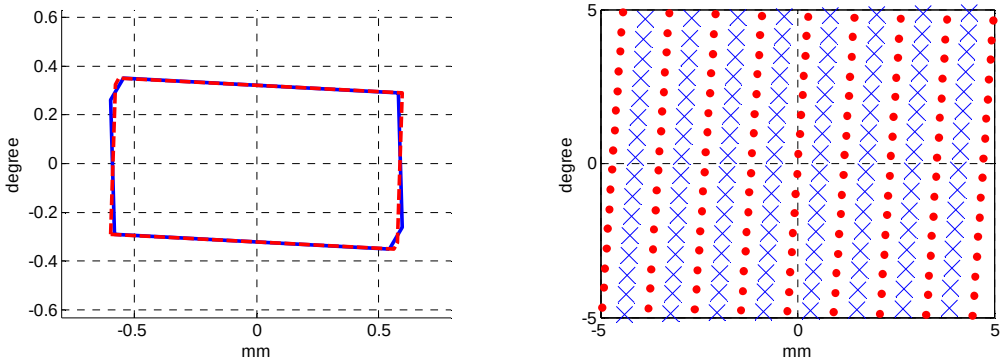


Figure 23. Optimized solution for x_c = 22.36 mm and 1240 pixel horizontal camera resolution – display rays (blue, solid line / crosses) and camera rays (red, dashed line / dots) at screen plane. (a) Voronoi cells. (b) Sampling grids.

7. CONCLUDING REMARKS

In this paper we put emphasize on projection-based LF displays with a linear configuration of projection engines – the main reason for this being the existence of those. Nevertheless, similar concepts can be applied to any other display that recreates the LF. At the core, the analysis in this paper only assumes that there are light generators capable of generating a finite number of rays on a regular (not necessary uniform) grid. The configuration of such light generators will put limits on the LF that a display can create, and consequently, following a similar analysis as presented in this paper, put limits on the required scene capture setup.

We used a very simple algorithm for determining similarities between the Voronoi cells obtained from the ray-space grid generated by the display (ray generators) and cameras. It is obvious that this might result in some mismatches. We plan to look into more robust algorithms in the future that will ensure more reliable similarity comparison but at the same time still keep the complexity of the similarity measure as low as possible.

The final step, after determining the required minimalistic camera setup for a given display is to capture a scene and process the captured data (evaluate the display rays out of the captured camera rays). Two things should be taken under consideration when doing this. First, it should be ensured, by some means, that the desired scene can be 'correctly'

captured by the estimated camera setup (e.g. with the desired level of detail and without aliasing). In practice this would require proper anti-aliasing filters during capture or capturing the scene with a higher sampling rate and then properly downsampling it. Second, the interpolation between the camera rays and rays available in the display has to be executed properly. Optimally, we need to do a nonuniform to nonuniform interpolation with spectra of both signals limited as estimated by the ray-space analysis – effectively, this would 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 cons and pros. We will discuss this in more detail in the 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 (or more generally a human itself). This will have to be evaluated by subjective experiments that should show what is more important – camera resolution or camera-to-camera distance.

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., Seuntiens, 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).