Retinal pre-filtering for light field displays

Romeiro, Rafael; Eisemann, Elmar; Marroquim, Ricardo

DOI
10.1016/j.cag.2024.104033

Publication date
2024

Document Version
Final published version

Published in
Computers and Graphics (Pergamon)

Computers & Graphics 123 (2024) 104033

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Technical Section

Retinal pre-filtering for light field displays

Rafael Romeiro [a], Elmar Eisemann [b], Ricardo Marroquim [b,*]

[a] Computer Graphics Lab, Federal University of Rio de Janeiro, Brazil
[b] Computer Graphics and Visualization, TU Delft, Netherlands

ARTICLE INFO

Keywords:
Light field displays
Retinal pre-filtering
Anti-aliasing

ABSTRACT

The display coefficients that produce the signal emitted by a light field display are usually calculated to approximate the radiance over a set of sampled rays in the light field space. However, not all information contained in the light field signal is of equal importance to an observer. We propose a retinal pre-filtering of the light field samples that takes into account the image formation process of the observer to determine display coefficients that will ultimately produce better retinal images for a range of focus distances. We demonstrate a significant increase in image definition without changing the display resolution.

1. Introduction

Typically, head-mounted displays (HMDs) employ a single display screen divided into two parts, one seen by the left eye and the other by the right eye, together with a pair of magnifier lenses that bring the display panel to a comfortable accommodation distance for the user's eyes. The binocular disparity present in this stereo pair of images provides appropriate vergence cues for depth perception. The accommodation distance, however, remains fixed to that of the flat virtual image of the display panel. The natural correlation of vergence and accommodation is thereby lost, which results in a variety of detrimental effects such as eye fatigue, visual discomfort and headaches.

To avoid the vergence–accommodation conflict, light field displays have been proposed to support correct focus cues. Nonetheless, the finite resolution of a light field display still imposes limitations on how accurately a light field signal can be reproduced. The coefficients that determine the signal reproduced on the display are usually calculated in an effort to minimise a reproduction error.

We believe, however, that the image formed on the observer's retina is of greater importance than the light field signal itself. Therefore, display resources should not be employed to mitigate errors in the light field signal reproduction that are not actually perceived by the observer. Rather, they should be applied where it makes the most difference to the quality of the resulting retinal image. With this in mind, we propose a method to optimally compute the light field display coefficients that minimise the retinal error instead of the light field error. Our method optimises the coefficients to minimise the error inside any given focus range. When the observer's focus is known, a narrower range can be used that dynamically moves to match the focus distance.

We present a comprehensive theoretical framework to compute the display's coefficients that minimises the light field error (Section 3) or the retinal image error (Section 4). The theoretical continuous methods are further translated to a matrix formulation (Section 6.3) that can be implemented in practice for discrete input signals. With this matrix formulation, we conducted an experiment (Section 7) that simulates a light field display and the image generated at the observer's retina. Finally, we present results using our virtual display with different filtering strategies (Section 8). We show that with increasing knowledge about the display and the observer, the results improve (see Fig. 1).

To summarise, the main contributions of this paper are:

- a light field display pre-filtering approach that takes into account the display's reconstruction filter (Section 3.2);
- a continuous description of the retinal image reconstruction as a function of the display's reconstruction filter (Section 4.2) and its accompanying optimisation problem for light field display pre-filtering (Section 4.3);
- an iterative solution for the proposed retinal pre-filtering optimisation problem (Section 5);
- a numerical integration solution for practical use of our method in the context of a discrete input signal (Sections Section 6.2) and the resulting matrix formulation (Section 6.3);
- a strategy to further reduce the retinal error by adapting the optimised focus range based on current observer focus distance (Section 8).

* Corresponding author.
E-mail address: r.marroquim@tudelft.nl (R. Marroquim).

https://doi.org/10.1016/j.cag.2024.104033
Received 31 January 2024; Received in revised form 31 July 2024; Accepted 1 August 2024
Available online 6 August 2024

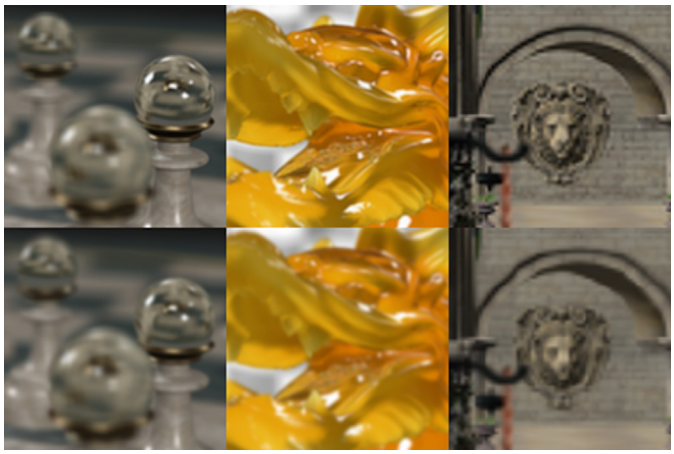


Fig. 1. Our proposed retinal pre-filtering (top row) and display pre-filtering that do not consider the retinal image formation (bottom row).

2. Related work

2.1. Light field sampling and reconstruction

An analytical description of the scene's light field is typically not available. Therefore, the radiance of multiple light rays must first be sampled to produce a discrete representation of the light field, either by ray tracing a synthetic scene or captured by a physical device (light field camera). A continuous representation of the light field can then be obtained by applying a reconstruction filter to the light field samples. The choice of initial sampling scheme and reconstruction filter are relevant factors for how well the original signal can be reconstructed.

Light field sampling is done predominantly following a two-plane parameterisation. In this parameterisation system, two predefined parallel planes are used as entry and exit planes for the light rays to be sampled. A set of entry points on one plane combined with a set of exit points on the other plane form a 4D lattice. Each 4D point in this lattice corresponds to a sampled ray of the light field.

To produce new light rays different from those already sampled, the existing samples can be combined in various ways using reconstruction filters. The shape and values of the reconstruction filter will determine what samples will contribute most, if at all, to a new light ray being produced. The properties of a reconstruction filter can be intuited by reinterpreting it as a camera-like ray detector that combines multiple light rays through its aperture and sensor to arrive at a single value that approximates the radiance of the desired ray. Do note that rays reconstructed this way can be used to compose a 2D render as well as to resample the entire 4D light field signal.

A separable reconstruction filter aligned to the original sampling axes, such as the quadrilinear interpolation filter proposed by Levoy and Hanrahan [1], combines sample rays whose entry and exit points are close to those of the new ray being produced. For this reason, this reconstruction filter is analogous to a camera with aperture located on the sampling entry plane and focus located on the sampling exit plane, as depicted in Fig. 2(a). This filter does not exploit any inherent property of the light field function or of the scene. Nonetheless, due to separability, this filter can be applied to samples very efficiently.

Gortler et al. [2] also propose a quadrilinear interpolation filter. This filter, however, is dynamically sheared for each reconstructed ray to better capture a different depth. It is analogous to a camera with an aperture located on the sampling entry plane and focus located on a plane at a chosen depth, as depicted in Fig. 2(b). For each ray being reconstructed, the depth to be prioritised is chosen based on the scene's geometry. This not only requires knowledge of the scene geometry but also requires the assignment of a single depth in the scene for each light ray. This depth per ray association is only suitable for scenes comprised of non-transparent Lambertian surfaces as, otherwise, the necessary spatial coherence in the light field would not be present.

Isaksen et al. [3] propose a similar approach but conflates the light field reconstruction with rendering, as in this case, the reconstruction filter is tailored to produce a specific rendered image. Again, this reconstruction filter is analogous to a camera with aperture located on the sampling entry plane and focus located on a plane at a chosen depth, as depicted in Fig. 2(b). The focus distance, however, no longer comes from geometry but is arbitrarily chosen based on visualisation interest. The authors further describe the use case of their filter for autostereoscopic displays, which we will cover in the following Section 2.2.

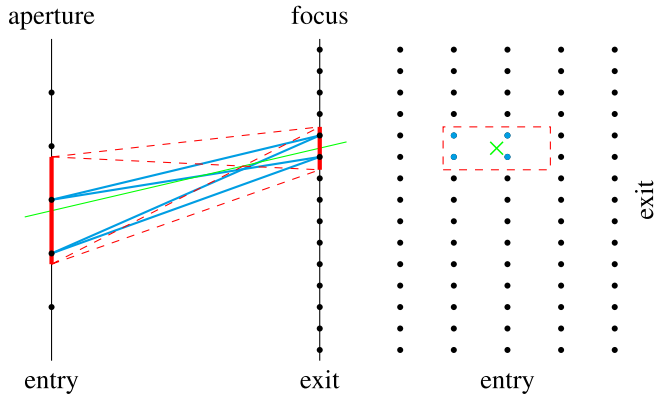
Chai et al. [4] describe the spectral support of a light field as being bounded according to the minimum and maximum depths of the scene. A reconstruction filter is then designed based on these bounds. The proposed reconstruction filter is analogous to a camera with aperture located on the sampling entry plane and focus located on an optimal constant depth that is calculated solely from the minimum and maximum depths of the scene. Once more, such a filter also corresponds to Fig. 2(b). Here, the choice of focus distance is equivalent in photography to adjusting the depth of field of a fixed-focus camera to properly cover the depth range of the scene, i.e., to keep its nearest and the farthest objects under acceptably sharp focus. Given the sampling rate over the focus plane, it is possible to determine the minimum sampling rate over the aperture plane to avoid aliasing. In other words, given the resolution of the reconstruction camera, it is possible to determine the optimal number of pictures to avoid aliasing. This, however, is only optimal under the assumption that the sampling lattice must remain rectangular with respect to the filter axes. Furthermore, the spectral bounds outlined based on minimum and maximum depths only hold for strictly Lambertian scenes without occlusions.

Zhang and Chen [5] generalise the work of Chai et al. [4] by allowing full freedom to the 4D sampling lattice. A lower total sampling density can then be achieved using a non-rectangular lattice. The authors also suggest measures to mitigate aliasing in non-Lambertian or occluded scenes.

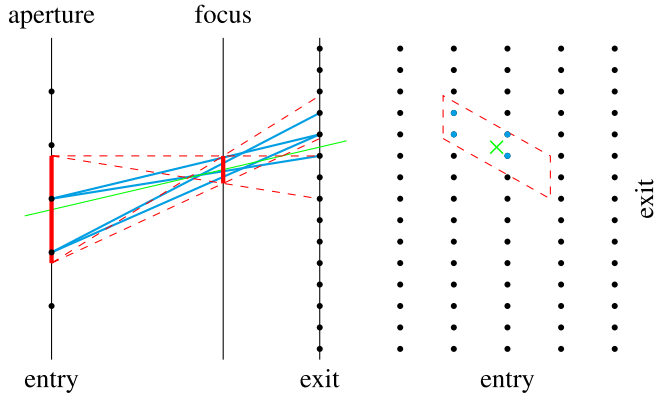
Stewart et al. [6] propose a reconstruction filter that is analogous to a combination of two cameras. One camera has a narrow aperture located on the sampling entry plane and focus located on the optimal focus depth, following Chai et al. [4]. The other camera has a wide aperture located on the sampling entry plane and focus located according to visualisation interest, following Isaksen et al. [3]. This combination aims to eliminate aliasing using the filter from Chai et al. [4] while reintroducing high frequencies at a particular depth through the filter from Isaksen et al. [3].

All the reconstruction filters described so far have kept the camera aperture on the sampling entry plane. However, this does not need to be the case. Just like with the focus plane, the aperture can also be corrected for a plane other than the sampling entry plane, as depicted in Fig. 2(c). This camera-like composition of the reconstruction filter is separable over the aperture and focus axes and each individual axial component can follow different functions such as box filter, linear interpolation, sinc filter, etc. Moreover, it is possible to design reconstruction filters that are not separable over any axis.

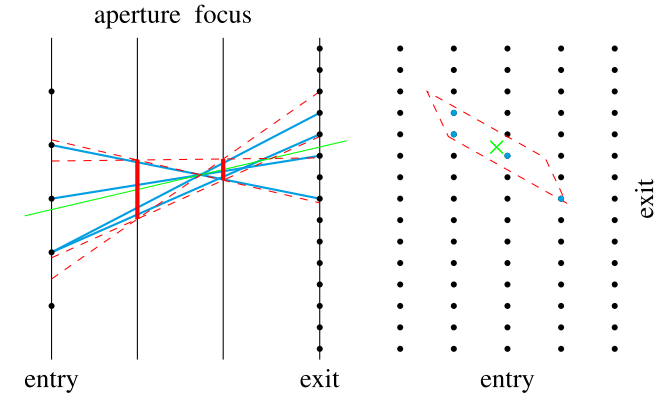
Liu et al. [7] propose a light field synthesis method based on a deep neural network. The network is first trained with light fields with multiple sub-aperture images (high angular resolution). Once trained, the network is capable of producing new in-between views for an input light field with few sub-aperture images (low angular resolution). The loss function used in the training process takes into account not only the view-wise error (error in the original light field domain) but also the refocused image error (error in the retinal image domain). This regularisation factor allows the synthesised light fields to produce refocused images more effectively.



(a) Reconstruction filter separable over the sampling axes.



(b) Reconstruction filter with focus correction.



(c) Reconstruction filter with aperture and focus correction.

Fig. 2. A new ray (green) can be reconstructed by combining all the sampled rays (blue) that lie inside some region (red) around the new ray. The diagrams on the left show the sampling planes (entry and exit), the planes (aperture and focus) over which the reconstruction filter is defined (red) and depict rays as lines (green and blue). On the right side, we represent the corresponding sampling space (entry and exit axes) where rays are points (green, blue and black), and the reconstruction filter support is a parallelogram (red).

2.2. Light field displays

We define a light field display as a device capable of emitting a 4D signal reconstructed from a 4D set of coefficients. In this case, the sample density and reconstruction filter used are defined physically in its construction. What remains to be specified during its operation are the values of the coefficients.

Lanman and Luebke [8] demonstrate a near-eye light field display architecture based on a microlens array that could drive correct focus cues within a viewer's natural accommodation range despite the display being in close proximity to the eye. This architecture trades a part of the spatial resolution of a traditional display that emits light isotropically for angular resolution, with each microlens acting as a view-dependent pixel. Each individual image behind a microlens encodes angular information while the resulting spatial resolution is reduced to that of the microlens array, which is significantly lower than that of the original display. Furthermore, the microlenses may also introduce optical aberrations and fundamentally limit resolution due to diffraction. To determine coefficient values, a direct solution provided by ray tracing was implemented, where from each pixel a ray is cast passing through the optical centre of its associated microlens.

Isaksen et al. [3] describe a method to determine coefficient values from an already sampled light field. In this case, instead of casting a new ray into a virtual scene for each display sample, the radiance of the ray passing through the centre of a pixel and the optical centre of its associated microlens is calculated by combining the values of previously sampled rays. This combination is done using a reconstruction filter as described in Section 2.1 and depicted in Fig. 2(b). Note that the authors do not relate the filter aperture with the display microlenses or the filter focus plane with the focused image of the display panel through the microlenses. Instead, they propose the use of an arbitrarily sized aperture located on the sampling entry plane and focus located according to visualisation interest.

Zwicker et al. [9] characterise a light field display bandwidth from the Nyquist limits associated with the display sample density. This corresponds to a spectral support shaped as a 4D box. In order to avoid aliasing, a pre-filter of the same shape must then be applied to the continuous light field before the samples for display are taken. The continuous light field can be obtained from previous samples using a band-limited reconstruction filter such as the one proposed by Stewart et al. [6]. This entire process can be summarised in a single digital resampling filter that combines the reconstruction filter with the display pre-filter. However, without taking into account the display reconstruction filter, this method only prevents the pre-aliasing that may occur during the display resampling and not the post-aliasing that may occur during the lattermost reconstruction on the display. In general, post-aliasing is unavoidable, as to do so would require a band-limited reconstruction filter, which is physically unfeasible. Furthermore, even the pre-aliasing is only preventable for a very limited category of scenes, namely, Lambertian without occlusions.

2.3. Layered displays

Wetzstein et al. [10] develop a tomographic technique for multifocal displays by stacking light-attenuating layers. This is a compressive multi-layer architecture where the layers are combined multiplicatively. The coefficients for each layer are calculated to minimise the error over a discrete set of rays. The compressive nature of this model makes it possible to achieve higher resolutions at the cost of the degrees of freedom of its content (lower-rank representation). Due to the interaction between layers, this model is fundamentally limited by diffraction.

Narain et al. [11] propose a multifocal display architecture that combines layers additively and, therefore, is not limited by diffraction. The set of retinal images that would be seen by the viewer for a range of different accommodation distances called a focal stack, encodes all the information necessary to describe the scene from the viewer's point of view. Using a model of image formation in the eye, it is possible to predict the focal stack for both the scene as viewed directly and when viewing the display. The contents of each layer can be computed to minimise the error between those two focal stacks.

Mercier et al. [12] present a more efficient method for decomposing the scene into layers for multifocal displays. The proposed numerical

method is provably stable and reaches interactive performance on GPU implementations. If eye tracking is available, this method can also correct for misalignments due to eye movement after the scene decomposition.

Ebner et al. [13] propose a combination of the multifocal and varifocal approaches. Traditionally, the varifocal display architecture changes the position of the flat virtual image of a display panel according to the viewer's accommodation distance while applying the corresponding focus blur to the panel image [14]. This allows for high resolution and high contrast but requires eye tracking and is limited by the tracker's speed and precision. The multifocal display architecture does not require an eye tracker as it can simultaneously reconstruct correct focus cues within a working volume delimited by the display layers. However, the quality of each individual focus reconstruction decreases with the working volume size. By adding an eye tracker to an adjustable multifocal display, the working volume of a multifocal display can be dynamically repositioned according to the viewer's accommodation distance as a varifocal display would do. This hybrid approach uses the small working volume of the multifocal display to cover the lack of accuracy of the eye tracker while allowing a bigger working volume without compromising the reconstruction quality.

Layered displays reconstruct a 4D light field signal from a 3D set of coefficients (compressive space). The optimisation in such devices searches for a (compressive) 3D solution that minimises either a 4D error [10] or a (compressive) 3D error [11–13]. Our work aims to find a 4D solution that minimises either a 4D error (Section 3) or a (compressive) 3D error (Section 4).

3. Display pre-filtering

As discussed in Section 2, multiple reconstruction filters have been proposed to produce a continuous representation of a light field from a set of samples. These filters were designed for the purpose of light field resampling or image synthesis and are based on scene content or general features of light fields.

Similarly, in the context of light field displays, a continuous-space analog signal is produced from a set of display coefficients. However, the reconstruction filter employed in this case is one imposed by the physical properties of the display and the assumption of an ideal band-limited reconstruction is improper. Even though previous works have proposed anti-aliasing pre-filtering methods for light field displays, to the best of our knowledge, they were limited to evaluating the display's capabilities through the Nyquist rate criterion associated with the display resolution and have completely neglected the display reconstruction filter. The actual display reconstruction filter (and its accompanying limitations) should influence the pre-filtering process. In this section, we make a brief review of sampling and reconstruction theory and demonstrate its application for light field displays.

3.1. Signal reconstruction and anti-aliasing

Assuming a traditional sampling and reconstruction pipeline, a continuous signal $f(x)$, $x \in X$, is first pre-filtered with an analysis filter ψ and then sampled over an array of n sample positions $\boldsymbol{\lambda}$, resulting in an array of n sample values $\mathbf{c}$ (Eq. (1)). At a later stage, a continuous signal $\tilde{f}$ can be obtained by convolving the sample values $\mathbf{c}$ with a reconstruction filter φ (Eq. (2)).

$$\mathbf{c}[i] = (f * \psi)(\boldsymbol{\lambda}[i]) = \int_X f(x)\,\psi(\boldsymbol{\lambda}[i] - x)\,\mathrm{d}x \tag{1}$$

$$\tilde{f}(x) = (\mathbf{c} * \varphi)(x) = \sum_{i=1}^{n} \mathbf{c}[i]\,\varphi(x - \boldsymbol{\lambda}[i]) \tag{2}$$

The reconstruction filter φ placed over the sample positions $\boldsymbol{\lambda}$ forms a shift-invariant array of functions $\boldsymbol{\varphi}$, where $\boldsymbol{\varphi}[i](\mathbf{x}) = \varphi(\mathbf{x} - \boldsymbol{\lambda}[i])$. $\boldsymbol{\varphi}$ spans the reconstruction space $V(\boldsymbol{\varphi})$ (Eq. (3)).

$$V(\boldsymbol{\varphi}) = \left\{ \sum_{i=1}^{n} c_i\,\boldsymbol{\varphi}[i](\mathbf{x}) \,\middle|\, c_i \in \mathbb{R} \right\} \tag{3}$$

Following the sampling theorem [15], the reconstruction filter $\varphi = \mathrm{sinc}$ is capable of perfectly reconstructing band-limited signals. The reconstruction space $V(\mathbf{sinc})$ is the subspace of band-limited signals in compliance with the sampling rate of $\boldsymbol{\lambda}$.

If f is band-limited and lies within $V(\mathbf{sinc})$, f can be perfectly reconstructed with $\varphi = \mathrm{sinc}$ from the samples in $\mathbf{c}$ without requiring a pre-filter ψ. When f is not band-limited, it lies outside $V(\mathbf{sinc})$ and cannot be perfectly reconstructed with $\varphi = \mathrm{sinc}$ regardless of the choice of pre-filter. In these cases, the pre-filter $\psi = \mathrm{sinc}$ can be used to create a band-limited approximation of f before sampling, which in turn can be perfectly reconstructed from the obtained samples. The use of sinc as a pre-filter is coupled to the use of sinc as the reconstruction filter.

The traditional approach to sampling and reconstruction can be reinterpreted as a minimisation problem over the $\mathcal{L}_2$ norm of the residual $\|f - \tilde{f}\|$ conditioned to $\tilde{f} \in V(\boldsymbol{\varphi})$. For that effect, the residual must be orthogonal to $V(\boldsymbol{\varphi})$ and $\tilde{f}$ must then be the orthogonal projection of f into $V(\boldsymbol{\varphi})$ (Eq. (4)).

$$\tilde{f}(x) = \sum_{i=1}^{n} \langle f, \mathring{\boldsymbol{\varphi}}[i] \rangle\, \boldsymbol{\varphi}[i](\mathbf{x}) \tag{4}$$

Where $\mathring{\boldsymbol{\varphi}}$ is the dual basis of $\boldsymbol{\varphi}$ and can be uniquely determined by the biorthogonality condition (Eq. (5)) [16,17].

$$\langle \mathring{\boldsymbol{\varphi}}[i], \boldsymbol{\varphi}[j] \rangle = \delta[i - j] = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j \end{cases} \tag{5}$$

As a consequence, $\mathring{\boldsymbol{\varphi}}$ inherits the shift-invariance of $\boldsymbol{\varphi}$ and the orthogonal projection of f into $V(\boldsymbol{\varphi})$ becomes equivalent to the convolution of f with a filter $\mathring{\varphi}$. In the traditional sampling approach, this corresponds to the use of the pre-filter $\psi = \mathring{\varphi}$. Moreover, sinc is an orthogonal kernel, i.e., $\mathring{\varphi} = \varphi = \mathrm{sinc}$. This coincides with the traditional anti-aliasing paradigm of sinc pre-filtering and sinc reconstruction.

3.2. Optimal display pre-filtering for any reconstruction filter

However, as previously stated, we do not believe that sinc is an appropriate representation of a reconstruction filter for a physical implementation of a light field display. The sinc filter, whether used as a pre-filter or reconstruction filter, is an ideal low-pass filter. In reality, the use of ideal low-pass filters is impractical or even impossible, and it is generally desirable for ψ and φ to be compactly supported. The sinc filter not only has a slow-decaying infinite support, but its negative lobes would lead to some physically unfeasible light field reconstructions.

Instead of assuming a band-limited reconstruction with sinc filter, in this work, we will refer to the display reconstruction filter as φ_d and its associated optimal pre-filter as ψ_d. We derive our formulations based on φ_d and later, in Section 7.3, we support our choice of φ_d as the box filter for our simulations.

We use a two-plane light field parameterisation, with the display space $A \times B$ being defined by the absolute positions $\mathbf{x}_a = \begin{bmatrix} x_a & y_a \end{bmatrix}$ and $\mathbf{x}_b = \begin{bmatrix} x_b & y_b \end{bmatrix}$ on the planes A and B, respectively. We can then specify a point in display space by $\mathbf{x}_d = \begin{bmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{bmatrix} = \begin{bmatrix} x_a & y_a \\ x_b & y_b \end{bmatrix}$.

We assume the display samples are arranged according to a regular 4D lattice with $n_a \times n_a \times n_b \times n_b$ samples, with samples spaced by μ_a on plane A and by μ_b on plane B. We collapse the 4 dimensions of the

display lattice into a single index i_d so that the display sample array is given by:

$$\lambda_d = \left(\begin{bmatrix} x_a & y_a \\ x_b & y_b \end{bmatrix}_0, \begin{bmatrix} x_a & y_a \\ x_b & y_b \end{bmatrix}_1, \dots, \begin{bmatrix} x_a & y_a \\ x_b & y_b \end{bmatrix}_{n_d} \right) \tag{6}$$

Where $n_d = n_a^2 n_b^2$ is the total number of display samples and $\lambda_d[i_d]$ are the 4D coordinates of the i_d-th display sample.

Hereafter, we refer as display pre-filtering of a light field L_d the process described by Eq. (7), where $\boldsymbol{\psi}_d[i_d]$ is the display pre-filter shifted over the i_d-th display sample, i.e, $\boldsymbol{\psi}_d[i_d](\mathbf{x}_d) = \psi_d(\mathbf{x}_d - \lambda_d[i_d])$. Likewise, a light field $\tilde{L}_d$ is reconstructed by the display according to Eq. (8).

$$\mathbf{c}\left[i_d\right] = \left(L_d * \varphi_d\right)\left(\lambda_d\left[i_d\right]\right) = \iint_A \iint_B L_d(\mathbf{x}_d)\, \boldsymbol{\psi}_d\left[i_d\right](\mathbf{x}_d)\, \mathrm{d}\mathbf{x}_d \tag{7}$$

$$\tilde{L}_d(\mathbf{x}_d) = \left(\mathbf{c} * \psi_d\right)(\mathbf{x}_d) = \sum_{i_d=1}^{n_d} \mathbf{c}[i_d]\, \boldsymbol{\varphi}_d[i_d](\mathbf{x}_d) \tag{8}$$

The pre-filtering method described by Eq. (7), where an optimal pre-filter ψ_d conditioned to the reconstruction filter φ_d is used, is a straightforward application of well established techniques in the field of signal sampling [16,17]. However, to the best of our knowledge, this is the first time this concern has been raised within the context of light field display pre-filtering.

4. Retinal pre-filtering

In the previous section we were concerned with approximating $\tilde{L}_d$ compared to L_d directly (that is, minimising the error in the 4D light field domain) while taking into account the reconstruction filter of the display. In this section, we are concerned with approximating what an observer sees when exposed to $\tilde{L}_d$ compared to what they would have seen if exposed to L_d (that is, minimising the error in the 3D refocusable retinal image domain). Now, in addition to taking into account the display reconstruction filter, we will also consider the image formation process of the observer.

4.1. Observer model

The retina image of an observer is proportional to the irradiance function $E(\mathbf{x}_r)$ over the retina plane. The irradiance can be defined from the incident radiance on the retina [18] which is given by the light field L_e in the eye space $R \times P$, with R denoting the retina region and P denoting the pupil region.

$$E(\mathbf{x}_r) = \frac{1}{z_r^2} \iint_P L_e\left(\begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_p \end{bmatrix} \right) |\cos^4\theta|\, \mathrm{d}\mathbf{x}_p \tag{9}$$

where z_r is the position of the retina plane relative to the pupil plane and θ is the angle between the light ray and the optical axis.

We neglect the $|\cos^4\theta|$ term as its effects (such as vignetting) are only noticeable towards the periphery of the image, where the flat approximation of the retina is already less representative. Since the retina will always be mapped to the plane at focus distance (with appropriate scaling), we can assume, without loss of generality, that $z_r = 1$.

This simplifies Eq. (9) to:

$$E(\mathbf{x}_r) = \iint_P L_e\left(\begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_p \end{bmatrix} \right) \mathrm{d}\mathbf{x}_p \tag{10}$$

Approximating the eye lens by the Gaussian thin lens formula, the relationship between L_e and L_d can be expressed as follows:

$$L_e(\mathbf{x}_e) = L_d\left(\underbrace{\begin{bmatrix} z_a & 1 - z_a\zeta_f \\ z_b & 1 - z_b\zeta_f \end{bmatrix}}_{T_{de}(\zeta_f)} \mathbf{x}_e \right) \tag{11}$$

Where ζ_f indicates the reciprocal distance (in dioptres) from the pupil at which the observer is focusing while z_a and z_b are the distances (in meters) from the pupil of the two planes, A and B, over which the display space is parameterised.

Notice that from the same light field L_d, each value of ζ_f will produce a different L_e and, consequently, a different retinal image E.

4.2. Retinal reconstruction

Given a light field L_d, we define a 3D function G comprised of all 2D retinal images E produced by continuously varying ζ_f (Eq. (12)). This function G is a continuous counterpart for the discrete focal stack of Narain et al. [11] and Mercier et al. [12], both in terms of the focus distance ζ_f and the retina position $\mathbf{x}_r$. This space is also similar to the refocused image domain of Liu et al. [7].

$$G(\zeta_f, \mathbf{x}_r) = \iint_P L_d\left(T_{de}(\zeta_f) \begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_p \end{bmatrix} \right) \mathrm{d}\mathbf{x}_p \tag{12}$$

Analogously, we can define $\tilde{G}(\zeta_f, \mathbf{x}_r)$ for when a reconstructed light field $\tilde{L}_d$ is observed instead of L_d:

$$\begin{aligned} \tilde{G}(\zeta_f, \mathbf{x}_r) &= \iint_P \tilde{L}_d\left(T_{de}(\zeta_f) \begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_p \end{bmatrix} \right) \mathrm{d}\mathbf{x}_p \\ &= \iint_P \sum_{i_d=1}^{n_d} \mathbf{c}[i_d]\, \boldsymbol{\varphi}_d[i_d]\left(T_{de}(\zeta_f) \begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_p \end{bmatrix} \right) \mathrm{d}\mathbf{x}_p \\ &= \sum_{i_d=1}^{n_d} \mathbf{c}[i_d] \iint_P \boldsymbol{\varphi}_d[i_d]\left(T_{de}(\zeta_f) \begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_p \end{bmatrix} \right) \mathrm{d}\mathbf{x}_p \\ &= \sum_{i_d=1}^{n_d} \mathbf{c}[i_d]\, \boldsymbol{\varphi}_r[i_d](\zeta_f, \mathbf{x}_r) = \boldsymbol{\Phi}\mathbf{c} \end{aligned} \tag{13}$$

$$\text{where } \boldsymbol{\varphi}_r[i_d](\zeta_f, \mathbf{x}_r) = \iint_P \boldsymbol{\varphi}_d[i_d]\left(T_{de}(\zeta_f) \begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_p \end{bmatrix} \right) \mathrm{d}\mathbf{x}_p \tag{14}$$

Now $\boldsymbol{\varphi}_r$ forms a base for the retinal reconstruction, like $\boldsymbol{\varphi}_d$ formed a base for the display reconstruction. The functions in $\boldsymbol{\varphi}_r$ correspond to the retinal image of each display element individually. However, due to the fact that the pupil region P is finite, the functions in $\boldsymbol{\varphi}_r$ are not shift-invariant as the functions in $\boldsymbol{\varphi}_d$. Therefore the reconstruction procedure can no longer be represented by a convolution and instead will be represented by the linear operator $\boldsymbol{\Phi}$.

The linear operator $\boldsymbol{\Phi}$ encapsulates the entire process from the discrete set of display coefficients $\mathbf{c}$ to the continuous 3D focal stack $\tilde{G}$ formed on the observer's retina. For more details on the linear properties of $\boldsymbol{\Phi}$, refer to Appendix 1 of the supplementary material.

4.3. Optimisation problem

Our retinal pre-filtering method is then an optimisation problem that aims to minimise $\left\|G - \tilde{G}\right\| = \|G - \boldsymbol{\Phi}\mathbf{c}\|$ for coefficients $\mathbf{c}[i_d] \in [0, 1]$. Typically, the least squares solution to $\boldsymbol{\Phi}\mathbf{c} \approx G$ is given by the normal equation $\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\,\mathbf{c} = \boldsymbol{\Phi}^\top G$. However, calculating $\mathbf{c} = (\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\,\boldsymbol{\Phi}^\top G$ leads to values in $\mathbf{c}$ outside the allowed interval and inverting $\boldsymbol{\Phi}^\top\boldsymbol{\Phi}$ may present numerical instability or even be impossible.

Multiple strategies have been devised to solve bounded-variable least squares (BVLS) problems such as this. In the following section, we describe the algorithm we developed for our implementation.

5. Iterative solution

Lee and Seung [19] proposed a pair of alternating multiplicative update rules for non-negative matrix factorisation that are a good compromise between speed and ease of implementation. The iterative nature of this method allows it to be more adaptable to available execution times and it is not sensitive to the choice of step size as gradient-based methods.

Matrix factorisation can be interpreted as a generalisation of least squares. Therefore, to solve our least squares problem, we adapted this solution to a single iterative multiplicative rule:

$$\mathbf{c} \leftarrow \mathbf{c} \otimes \frac{\boldsymbol{\Phi}^\intercal G}{\boldsymbol{\Phi}^\intercal \boldsymbol{\Phi}\,\mathbf{c}} \tag{15}$$

where $\otimes$ and —— are the Kronecker product and division, respectively.

The coefficients in $\mathbf{c}$ can be initialised with random values in the $]0,1]$ interval and clamped back into $]0,1]$ after each iteration. The numerator $\boldsymbol{\Phi}^\intercal G$ is a discrete sequence of n_d terms, where each term is the inner product between a function in $\boldsymbol{\varphi}_r$ and G, as defined in Eq. (16). The denominator $\boldsymbol{\Phi}^\intercal\boldsymbol{\Phi}\,\mathbf{c}$ is the result of the $n_d \times n_d$ Gram matrix $\boldsymbol{\Phi}^\intercal\boldsymbol{\Phi}$ multiplied by $\mathbf{c}$. Each term of $\boldsymbol{\Phi}^\intercal\boldsymbol{\Phi}$ corresponds to an inner product between two functions in $\boldsymbol{\varphi}_r$, as defined in Eq. (17).

$$\begin{aligned}(\boldsymbol{\Phi}^\intercal G)[i_d] &= \langle \boldsymbol{\varphi}_r[i_d], G\rangle \\ &= \int_F \iint_R \boldsymbol{\varphi}_r[i_d](\zeta_f,\mathbf{x}_r)\,G(\zeta_f,\mathbf{x}_r)\,\mathrm{d}\mathbf{x}_r\,\mathrm{d}\zeta_f\end{aligned} \tag{16}$$

$$\begin{aligned}(\boldsymbol{\Phi}^\intercal\boldsymbol{\Phi})[i_d,j_d] &= \langle \boldsymbol{\varphi}_r[i_d], \boldsymbol{\varphi}_r[j_d]\rangle \\ &= \int_F \iint_R \boldsymbol{\varphi}_r[i_d](\zeta_f,\mathbf{x}_r)\,\boldsymbol{\varphi}_r[j_d](\zeta_f,\mathbf{x}_r)\,\mathrm{d}\mathbf{x}_r\,\mathrm{d}\zeta_f\end{aligned} \tag{17}$$

Note that the discrete inner product from the original method is translated into the inner product for functions, which, in our case, is an integral over the focus stack domain.

Here F delimits the region of G that is considered relevant in terms of depth of field (restriction over ζ_f). In other words, F defines the allowed range of focus distances for the observer, and this range will later be manipulated to produce different results in Section 8.

Substituting Eq. (12) in Eq. (16) we have:

$$(\boldsymbol{\Phi}^\intercal G)[i_d] = \int_F \iint_R \iint_P L_d\left(T_{de}(\zeta_f)\begin{bmatrix}\mathbf{x}_r\\ \mathbf{x}_p\end{bmatrix}\right)\boldsymbol{\varphi}_r[i_d](\zeta_f,\mathbf{x}_r)\,\mathrm{d}\mathbf{x}_p\,\mathrm{d}\mathbf{x}_r\,\mathrm{d}\zeta_f \tag{18}$$

Hence, given a target continuous light field L_d we can use Eqs. (15), (17) and (18) to determine discrete display coefficients $\mathbf{c}$. When used as display values, these coefficients produce the light field $\tilde{L}_d$ that, among all the possible light fields the display can produce, is the one that induces the retinal images closest to those induced when observing L_d directly.

6. Discrete input signal

So far, we have used a target continuous light field L_d as the input signal for both the display pre-filtering in Section 3 and the retinal pre-filtering in Section 4. We have not discussed any properties of the scene since our method does not imply any restrictions on L_d. Nevertheless, there is usually no analytical description of the scene light field so a sampled input light field $\mathbf{s}$ is required.

6.1. Sampled light field

Similarly to the display space, we assume the sampling space $U \times V$ is defined by the absolute positions $\mathbf{x}_u$ and $\mathbf{x}_v$ on the planes U and V located at z_u and z_v, respectively. We also assume $\mathbf{s}$ to be an array of n_s samples and $\boldsymbol{\lambda}_s$ to be their corresponding 4D coordinates in sample space.

$$\boldsymbol{\lambda}_s = \left(\begin{bmatrix}x_u & y_u\\ x_v & y_v\end{bmatrix}_0, \begin{bmatrix}x_u & y_u\\ x_v & y_v\end{bmatrix}_1, \ldots, \begin{bmatrix}x_u & y_u\\ x_v & y_v\end{bmatrix}_{n_s}\right) \tag{19}$$

Note that the sampling parameterisation does not need to follow the display or eye spaces. It can be chosen in a way that is most suitable for the content of the scene. A reparameterisation to display space can be performed as follows:

$$L_d(\mathbf{x}_d) = L_s\Bigg(\underbrace{\frac{1}{z_b - z_a}\begin{bmatrix}z_b - z_u & z_u - z_a\\ z_b - z_v & z_v - z_a\end{bmatrix}}_{T_{sd}}\mathbf{x}_d\Bigg) \tag{20}$$

Likewise, a reparameterisation to eye space can be performed as:

$$L_e(\mathbf{x}_e) = L_s\Bigg(\underbrace{\begin{bmatrix}z_u & 1 - z_u\zeta_f\\ z_v & 1 - z_v\zeta_f\end{bmatrix}}_{T_{se}(\zeta_f)}\mathbf{x}_e\Bigg) \tag{21}$$

The display pre-filtering as described by Eq. (7) can be redefined over L_s instead of L_d, becoming:

$$\mathbf{c}[i_d] = \iint_A \iint_B L_s(T_{sd}\mathbf{x}_d)\,\boldsymbol{\psi}_d[i_d](\mathbf{x}_d)\,\mathrm{d}\mathbf{x}_d \tag{22}$$

Similarly, the numerator of the iterative rule of the retinal pre-filtering as described by Eq. (18) can also be redefined over L_s instead of L_d, becoming:

$$(\boldsymbol{\Phi}^\intercal G)[i_d] = \int_F \iint_R \iint_P L_s\left(T_{se}(\zeta_f)\begin{bmatrix}\mathbf{x}_r\\ \mathbf{x}_p\end{bmatrix}\right)\boldsymbol{\varphi}_r[i_d](\zeta_f,\mathbf{x}_r)\,\mathrm{d}\mathbf{x}_p\,\mathrm{d}\mathbf{x}_r\,\mathrm{d}\zeta_f \tag{23}$$

6.2. Numerically integrating the focal stack

Since there is no continuous description of L_s, only the samples $\mathbf{s}[i_s] = L_s(\boldsymbol{\lambda}_s[i_s])$, $i_s \in [1, n_s]$ are available to any pre-filtering method. Still, the numerical integration of Eq. (22) is quite trivial. On the other hand, the integral in Eq. (23) becomes particularly non-trivial.

We can define the 2D retina coordinates $\boldsymbol{\xi}_r[i_s](\zeta_f)$ and 2D pupil coordinates $\boldsymbol{\xi}_p[i_s]$ corresponding to the i_s-th light field sample for when the observer is focusing at ζ_f:

$$\begin{bmatrix}\boldsymbol{\xi}_r[i_s](\zeta_f)\\ \boldsymbol{\xi}_p[i_s]\end{bmatrix} = \underbrace{\frac{1}{z_v - z_u}\begin{bmatrix}z_v\zeta_f - 1 & 1 - z_u\zeta_f\\ z_v & -z_u\end{bmatrix}}_{T_{se}^{-1}(\zeta_f)}\boldsymbol{\lambda}_s[i_s] \tag{24}$$

Each individual light field sample is a point $\boldsymbol{\lambda}_s[i_s]$ in the 4D $U \times V$ sampling space but corresponds to a line described by $\boldsymbol{\xi}_r[i_s](\zeta_f)$ in the 3D $F \times R$ space of G (Fig. 3).

We can then define $\boldsymbol{\zeta}_f$ as a discrete sequence of n_f samples of ζ_f covering F uniformly and define $\boldsymbol{\phi}[i_f, i_d, i_s]$ as the value of the retinal reconstruction function of the i_d-th display sample, $\boldsymbol{\varphi}_r[i_d]$ (Eq. (14)), evaluated at the focus distance $\boldsymbol{\zeta}_f[i_f]$ and at the retina coordinates $\boldsymbol{\xi}_r[i_s](\boldsymbol{\zeta}_f[i_f])$. In this way, we have effectively covered the focal stack domain with 3D sample points (Fig. 4) with values $\mathbf{s}[i_s]\boldsymbol{\phi}[i_f, i_d, i_s]$ that can then be used to numerically integrate Eq. (23).

The values of $\boldsymbol{\phi}$ can be pre-computed and used temporarily to calculate the total contribution each light field sample have to each term of the numerator and stored it as a matrix. Once this matrix is complete, $\boldsymbol{\phi}$ is not needed to calculate the coefficients in $\mathbf{c}$, and therefore n_f can be chosen as high as necessary for accuracy without compromising the subsequent optimisation stage.

6.3. Matrix formulation

We can now describe every operation of our methods in matrix form. When the goal is to minimise the display light field error, as described in Section 3, the display pre-filtering operation can be defined as an $n_d \times n_s$ matrix M_1 that encodes Eq. (22):

$$\mathbf{c} = M_1\mathbf{s} \tag{25}$$

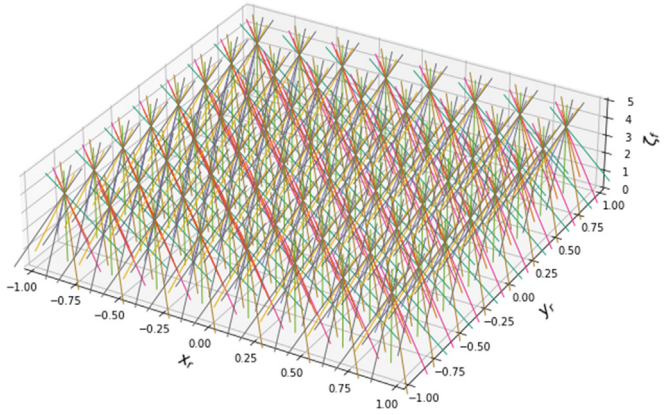


Fig. 3. Example of (3 × 3) × (8 × 8) light field samples in the 3D space of G.

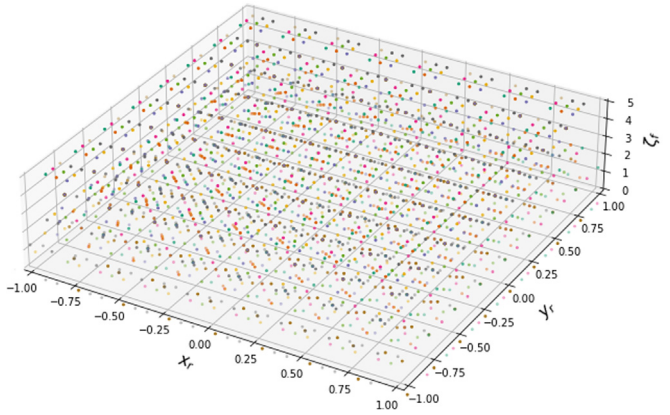


Fig. 4. Example of (3 × 3) × (8 × 8) × 5 sample coordinates of $\boldsymbol{\phi}$.

When the goal is to minimise the retinal images error instead, as described in Sections 4 and 5, the retinal pre-filtering is defined by an iterative process using two matrices, an $n_d \times n_s$ matrix M_2 that encodes Eq. (23) and an $n_d \times n_d$ matrix M_3 that encodes Eq. (17):

$$\mathbf{c} \leftarrow \mathbf{c} \otimes \frac{M_2 \mathbf{s}}{M_3 \mathbf{c}} \tag{26}$$

Subsequently, we simulate resulting retinal images $\mathbf{r}$ with a resolution of $n_r \times n_r$ pixels either from $\mathbf{s}$ or from $\mathbf{c}$ for some focus distance ζ_f. This retina resolution is only used in the simulation of the retinal images to illustrate our results. The computation of $\mathbf{c}$ is not based on any discretisation of retinal space.

Regardless of how $\mathbf{c}$ was produced, we simulate what an observer would see through a display given those coefficients when focusing at ζ_f. We simulate the resulting retinal image $\mathbf{r}(\zeta_f)$ with a resolution of $n_r \times n_r$ pixels using an $n_r^2 \times n_d$ matrix $M_4(\zeta_f)$:

$$\mathbf{r}(\zeta_f) = M_4(\zeta_f)\,\mathbf{c} \tag{27}$$

To assess the quality of our methods, we compare our simulated results to an appropriate reference image. A reference image is generated by simulating the observer's view of the scene directly without any display, which is done using an $n_r^2 \times n_s$ matrix $M_5(\zeta_f)$:

$$\mathbf{r}(\zeta_f) = M_5(\zeta_f)\,\mathbf{s} \tag{28}$$

An exact derivation for each matrix is given in Appendix 2 of the supplementary material.

7. Experimental setup

7.1. Light field sampling

To compute the sampled light field $\mathbf{s}$ for each scene, we render 20 × 20 views of 1080 × 1080 pixels each. The camera positions are evenly spaced over a 10 mm × 10 mm region in a plane coinciding with the observer's pupil.

On each camera position, we perform a sheared perspective projection (as that used to generate stereo pair images) where each frustum is skewed so that they share a 540 mm × 540 mm region in a plane at a distance of 265 mm from the pupil. This translates to a sample array $\mathbf{s}$ comprised of (20 × 20) × (1080 × 1080) light rays from a light field L_s parameterised over the planes at $z_u = 0$ and $z_v = 0.265$.

7.2. Observer imaging system

We consider the observer to have an 8 mm × 8 mm square pupil and a 90° field of view. We chose to use a square pupil to keep linear operations separable. As previously stated, we place the retina plane at a distance of 1 mm from the pupil to simplify the formulas. This decision, however, does not incur a loss of generality. At 1 mm distance, the 90° field of view corresponds to a 2 mm × 2 mm retina region. Although this size and position do not correspond to the actual size and position of the retina, the results are identical had physically plausible dimensions been used. We divide the retina region into 1024 × 1024 square retina pixels. This retina resolution is used to simulate retinal images and is not part of the pre-filtering process.

We also assume that the observer's focal distance ζ_f can vary from 5 dioptres (200 mm) up to 0 dioptres (at infinity).

7.3. Light field display

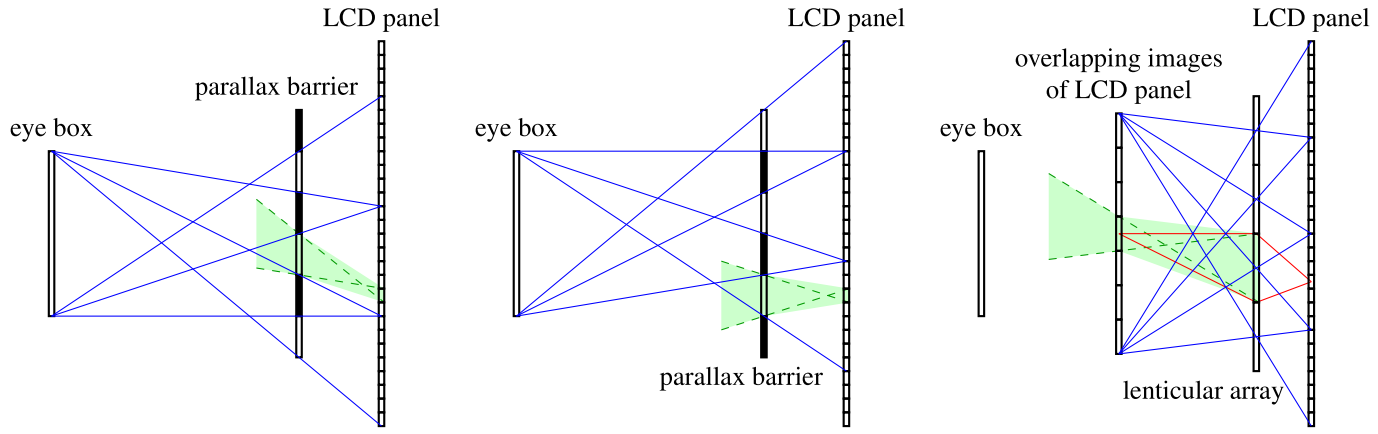
In Fig. 5, we exemplify a few possible light field display designs. Figs. 5(a) and 5(b) are based on parallax barrier, and the associated reconstruction filter is a box filter defined over the slits of the parallax barrier and the individual pixels on the LCD panel. Fig. 5(c) is based on lenticular array, and the associated reconstruction filter is a box filter defined over the individual lenslets and the converging images of individual pixels.

Formally, we define the display reconstruction filter φ_d for our simulations as being the 4D box filter:

$$\varphi_d(\mathbf{x}_d) = \sqcap\left(\frac{x_a}{\mu_a}\right)\sqcap\left(\frac{y_a}{\mu_a}\right)\sqcap\left(\frac{x_b}{\mu_b}\right)\sqcap\left(\frac{y_b}{\mu_b}\right) \tag{29}$$

$$\text{where} \quad \sqcap(x) = \begin{cases} 0, & \text{if } |x| > \frac{1}{2} \\ \frac{1}{2}, & \text{if } |x| = \frac{1}{2} \\ 1, & \text{if } |x| < \frac{1}{2} \end{cases} \tag{30}$$

Regardless of the underlying display design, we define the display reconstruction space for our simulations over a 24 mm × 24 mm region at a distance of 8 mm from the pupil, together with a 280 mm × 280 mm region at a distance of 136 mm from the pupil. These two regions are divided according to 4 resolution options: 12 × 12, 24 × 24, 36 × 36 or 48 × 48 for the first region and 140 × 140, 280 × 280, 420 × 420 or 560 × 560 for the second region. Therefore, the display light field L_d is parameterised over the planes at $z_a = 0.008$ and $z_b = 0.136$ and sampled under 4 resolutions: (12 × 12) × (140 × 140), (24 × 24) × (280 × 280), (36 × 36) × (420 × 420) and (48 × 48) × (560 × 560).



(a) Light field display design based on parallax barrier.

(b) Alternative parallax barrier configuration.

(c) Light field display design based on lenticular array.

Fig. 5. Examples of light field display designs. In all depicted cases, the display reconstruction filter corresponds to a box filter (highlighted in green). Designs (a) and (b) are based on parallax barrier. Each slit on the barrier separates a segment of the LCD panel that can be seen from the eye box (visibility traced in blue). Since a pixel cannot be seen through more than one slit, each visible combination of slit and pixel is an independent sample of the light field, which is uniformly reconstructed over the surface of the associated slit and pixel. The parallax barrier configuration used, such as in (a) and (b), can be multiplexed (through time, polarisation, etc.) in order to have a better coverage of samples on the barrier plane. Design (c) is based on lenticular array. Each lenslet forms an image of a different segment of the LCD panel (delimited by the lines in blue) that overlaps with the image formed by all the other lenslets at the same region. The overlapping pixel images in this region become view-dependent virtual pixels. Each combination of virtual pixel and lenslet is an independent sample of the light field which is uniformly reconstructed over the surface of the associated virtual pixel and lenslet.

7.4. Pre-filtering methods

Like the sinc filter, the box filter is also an orthogonal kernel. Therefore, the optimal pre-filter for a box filter reconstruction is also a box filter. More precisely, the optimal pre-filter ψ_d associated with the display reconstruction filter φ_d defined in Eq. (29) is given by:

$$\psi_d(\mathbf{x}_d) = \frac{\varphi_d(\mathbf{x}_d)}{\mu_a^2\,\mu_b^2} \tag{31}$$

The functions ψ_d and φ_d are used over the display lattice in Eqs. (7) and (8) to perform the display pre-filtering and reconstruction, respectively. Notice that a light field L_d with values in the range $[0,1]$ is guaranteed to produce coefficients $\mathbf{c}$ also in the range $[0,1]$, which in turn also guarantees to reconstruct a light field $\tilde{L}_d$ with values in the range $[0,1]$.

In the context of retinal pre-filtering, the base for retinal reconstruction $\boldsymbol{\varphi}_r$ is computed from the base for display reconstruction $\boldsymbol{\varphi}_d$ using Eq. (14). The base $\boldsymbol{\varphi}_r$ is then used in Eqs. (17) and (18), which are part of the iterative multiplicative rule, described by Eq. (15), to determine the contribution that each display coefficient receives from each light field sample. Keep in mind that changing the focus range F changes the region of ζ_f over which the error between G and $\tilde{G}$ is measured and, therefore, produces different sample weight distributions for the same display coefficient.

In Fig. 6, we summarise all the pre-filtering methods we use to compute one display coefficient. The quadrilinear interpolation method (bilinear in the simplified dimensionality of the figure) uses the nearest (in sample space) 16 light field samples to the display sample (4 nearest light field samples in the figure) weighted according to their distances. The display pre-filtering method uses all light field samples that lie inside ψ_d, which in our simulation is a 4D box (2D box in the figure) and gives the same weight to all the light field samples involved. Then, we show the retinal pre-filtering method being done for two different settings of F, each comprised of a single fixed focus distance instead of a range of focus distances. If the entire range between both focus distances is of concern, then all the indicated samples (blue and red) would be used with weights calculated following the continuous variation of ζ_f in this interval.

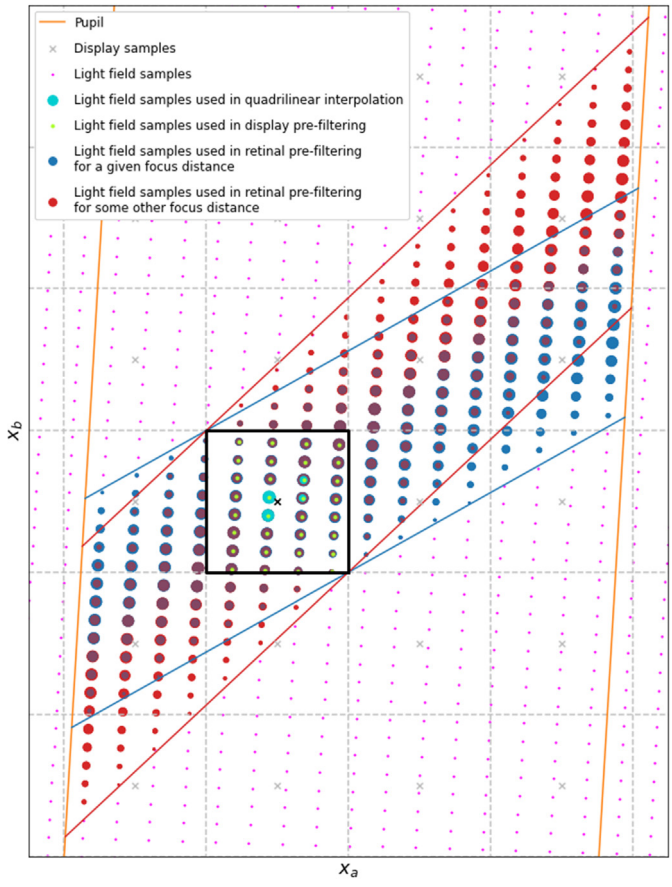
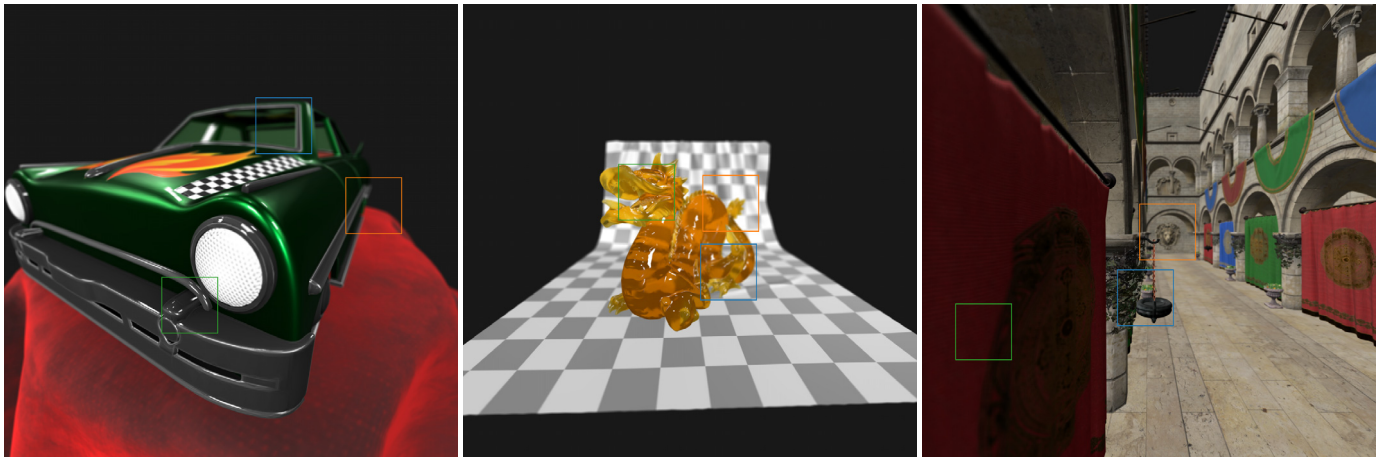


Fig. 6. Light field samples used by different methods, drawn in display space. For each method, the size of the marks indicates the weight of the contribution of each sample.

(a) Near focus distance (green box). (b) Mid focus distance (blue box). (c) Far focus distance (orange box).

Fig. 7. Reference retinal images for the Chess scene with different focus distances.



(a) Near focus distance (green box) for the Car scene. (b) Near focus distance (green box) for the Dragon scene. (c) Mid focus distance (blue box) for the Sponza scene.

Fig. 8. Reference retinal images for other scenes.

As described in Section 6.2, we discretise F uniformly to compute the matrices of Section 6.3. In our experiment, we used 100 samples between 0 dpt and 5 dpt.

8. Results

We tested our methods with four different scenes: Chess, Car, Dragon and Sponza. For each simulation configuration (scene, display resolution and pre-filtering method), we produce results simulating 100 focus distances between 0 dpt and 5 dpt. Among these, we select three focus distances (near, medium and far) per scene to showcase with inserts. Reference images for the scenes are shown in Figs. 7 and 8.

We used the chess scene for the error, time and convergence analysis, but similar plots for the other scenes are available in the supplementary material.

We ran our simulations on an Intel Core i5-8400 CPU @ 2.80 GHz with 32 GB RAM and an NVIDIA GeForce GTX 1070 GPU with 8 GB VRAM. Our simulator was implemented with Python using the CuPy library for GPU acceleration.

Our focus in this work is not on time performance, and there is definitely room for improvement in this regard. Nevertheless, to provide some insight into the relative behaviour of the methods under different display resolutions, we show some timings in Figs. 9 and 10.

Please note that all the methods measured here involve processing approximately 500M light field samples, regardless of display resolution. We use a significantly high number of light field samples so we can measure the achievable quality of each method without the concern of the samples quantity being a limiting factor. In the context of an interactive application, far fewer samples would be used.

The average time per iteration we found running our retinal pre-filtering method is 1.48 ms, 42.29 ms, 364.18 ms and 1999.11 ms corresponding to the display resolutions $(12 \times 12) \times (140 \times 140)$, $(24 \times 24) \times (280 \times 280)$, $(36 \times 36) \times (420 \times 420)$ and $(48 \times 48) \times (560 \times 560)$, respectively. There is no significant variation for different scenes or iteration numbers.

Note, however, that given a display configuration, we can precompute the matrices M_2 and M_3. For every frame, we need to compute the numerator $M_2\mathbf{s}$ of Eq. (26) only once. For every iteration, we compute

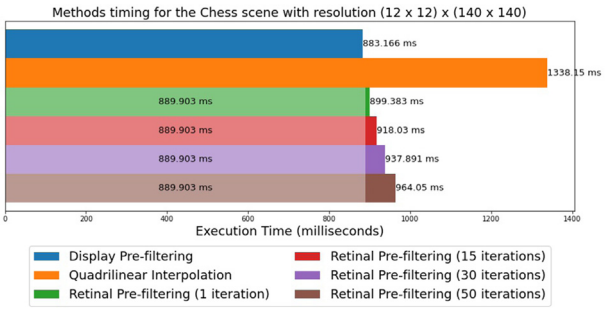


Fig. 9. Time comparison of pre-filtering methods under lower resolution. The highlighted segment of the retinal pre-filtering timings corresponds to the computation of the numerator.

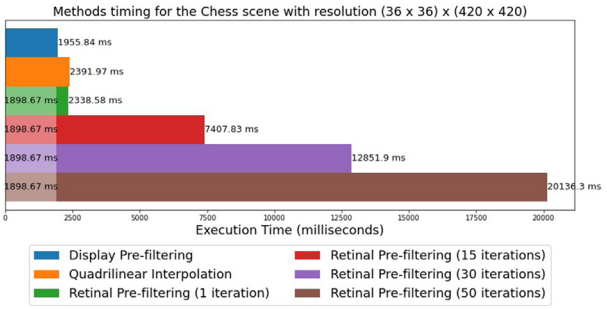


Fig. 10. Time comparison of pre-filtering methods under higher resolution. The highlighted segment of the retinal pre-filtering timings corresponds to the computation of the numerator.

all the remaining operations, among which the most expensive is the multiplication of the coefficients $\mathbf{c}$ by the $n_d \times n_d$ matrix M_3 in the denominator.

The time complexity to compute the numerator is linear with the number of display coefficients n_d and linear with the number of light field samples n_s. The two non-retinal methods also follow this same asymptotic behaviour. On the other hand, the iteration time is completely independent of n_s but grows quadratically with n_d. Consequently, the iteration time steeply increases with the display resolution.

Fig. 11 shows that the perceived difference between the iterations quickly diminishes. Hence, one could adjust the number of iterations for a particular requirements scenario. Also, in this work we initialise $\mathbf{c}$ with random values prior to the iterations. It is possible that a good initial guess could be provided requiring less iterations in total to achieve similar results. An example would be to employ progressive rendering strategies to improve the result across frames such as feeding the coefficients $\mathbf{c}$ of the previous frame as the initial state of the next frame in order to exploit spatial and temporal coherence.

Fig. 12 shows the convergence for the retinal pre-filtering method for all used display resolutions. All the following results involving retinal pre-filtering were achieved using 50 iterations. As with the light field samples, here we use a significantly high number of iterations so we can measure the achievable quality of our method without extensive considerations about convergence.

Fig. 13 shows the retinal mean squared error as the observer's focus distance varies between 0 dpt and 5 dpt (i.e. from an infinite distance down to 20 cm in front of the pupil). The retinal pre-filtering methods in this plot target four different options of focus distance range. As the target range narrows, the error inside the range decreases while the error outside the range increases. If the focus distance of the observer is known by, for example, using a tracking device, the retinal pre-filtering can target that exact focus distance, which would result in the smallest possible error. Otherwise, the retinal pre-filtering would need to target a range of possible focus distances.

From now on, we will consider two scenarios for the use of retinal pre-filtering. One for when the focus distance is unknown, and a single static retinal pre-filtering targeting the entire range of focus distances from 0 dpt to 5 dpt is used. The second is for when the focus distance is known, and the retinal pre-filtering is dynamically applied, targeting a single distance that follows the current focus distance.

Figs. 14 and 15 compare the retinal mean squared error of the different pre-filtering methods for a lower resolution display and a higher resolution display, respectively. The four methods being compared are the quadrilinear interpolation, the display pre-filtering, the retinal pre-filtering for the unknown focus distance scenario and the retinal pre-filtering for the known focus distance scenario. All methods have been previously described in Section 7.4. Note, however, that only the retinal pre-filtering with known focus changes the content of the display as the observer changes its focal distance.

The retinal pre-filtering with unknown focus produces, on average, lower retinal errors than the display pre-filtering and quadrilinear interpolation. The error is, however, higher on the extremes of the target range due to the retinal image coherence present when the focus distance varies. This coherence incentivises the optimisation to allocate more resources to reduce the error for focus distances in the middle of the target range, as this also benefits other distances that are still within the range. Reducing the error associated with focus distances closer to the range limits yields less benefit as a portion is wasted outside the range.

Fig. 16 shows three inserts for the Chess scene, each showcasing a different focus distance and comparing the four display resolutions and the four pre-filtering methods. Fig. 17 shows inserts for the remaining three scenes. Some more results can be seen in Fig. 1 and in the supplementary material.

9. Practical discussion

In this work, we proposed two different pre-filtering methods. One takes into account characteristics of the display and the other, in addition to the display, also incorporates characteristics of the observer. The display and the observer are represented by abstract models that approximate their behaviour in reality.

The fundamental property that is required on both models for the operation of the proposed methods is linearity. The linearity of the display model comes from the linear combination of the display elements (Eq. (8)) while the linearity of the observer model comes from its integral transform over the incoming light field (Eq. (9)).

Linearity allows us to describe the reconstruction of the 4D light field as a combination of 4D basis functions $\boldsymbol{\varphi}_d$ and the reconstruction of the 3D focal stack as a combination of 3D basis functions $\boldsymbol{\varphi}_r$. In both cases, the combination is physically performed and carried out according to the coefficients $\mathbf{c}$, which is the digital signal transmitted to the display.

9.1. Display limitations

Display architectures that employ lenticular arrays [8,20] or parallax barriers [21] to direct the emitted light along different directions trivially satisfy the linearity condition as their 4D signal is composed of non-overlapping elements. Even when incorporating time-multiplexing, the result is still linear.

Some display architectures explore compressive techniques where the emitted 4D signal has restrained degrees of freedom, layered displays being the most prominent. Layered displays where the layers are combined additively [11–13] also satisfy the linearity condition. On

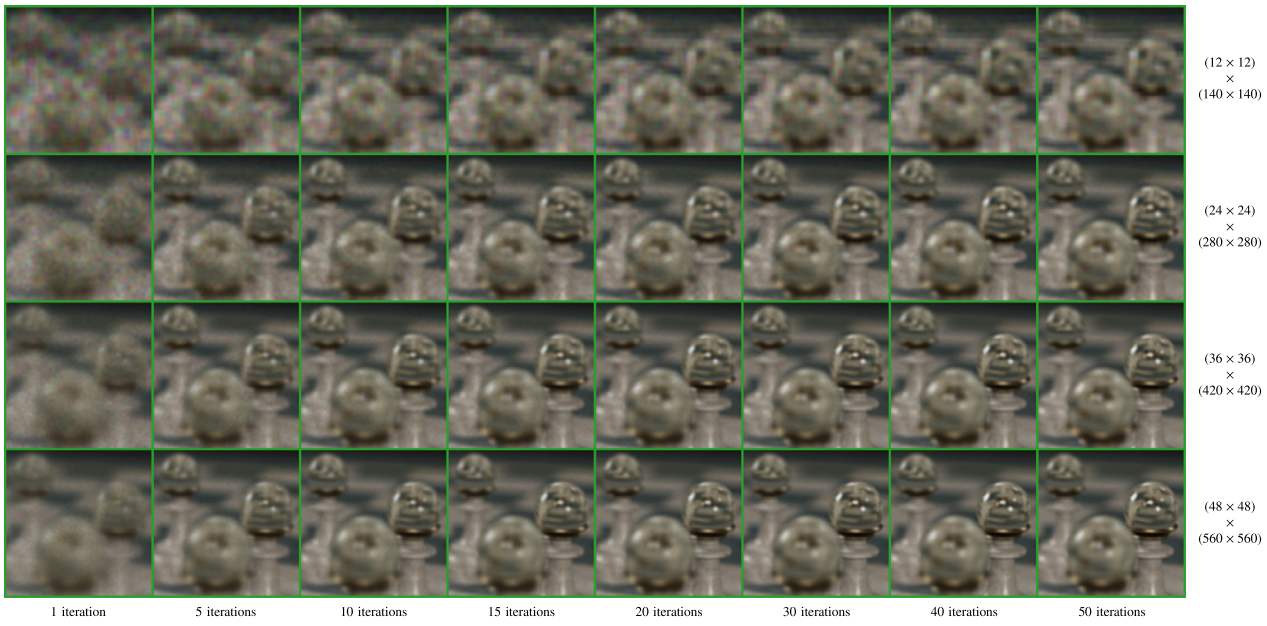


Fig. 11. Incremental improvements of retinal pre-filtering across multiple iterations for different display resolutions.

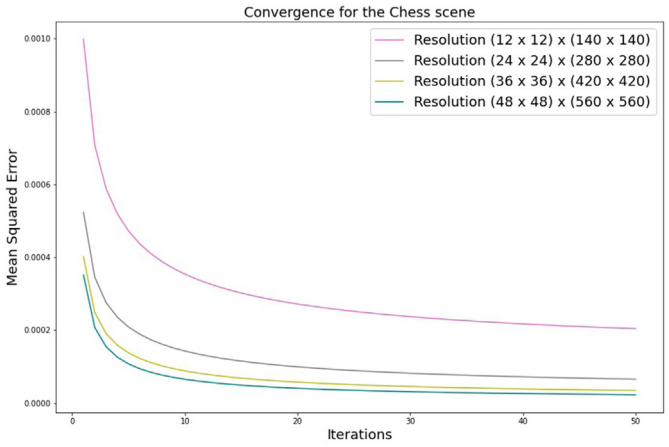


Fig. 12. Convergence of the retinal pre-filtering method throughout iterations of the iterative multiplicative rule for different display resolutions.

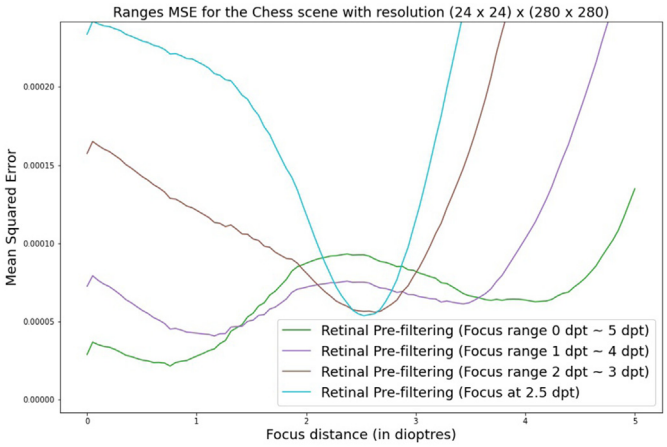


Fig. 13. Comparison of the retinal pre-filtering method for different focus distance ranges.

the other hand, when the layers are combined multiplicatively [10], the resulting 4D signal is not linear in regards to the input coefficients and our methods cannot be readily applied as they are. It is still possible to pursue a retinal pre-filtering method for those cases, but their mathematical modelling are not covered in this paper.

In the end, the only information needed from the display is the reconstruction function $\boldsymbol{\varphi}_d[i_d]$ associated to each display coefficient $\mathbf{c}[i_d]$. In many cases, $\boldsymbol{\varphi}_d$ is the same function shifted through the display 4D lattice. Even if each $\boldsymbol{\varphi}_d[i_d]$ is entirely different, our methods can still be used.

The vast majority of light field displays utilises LCD panels, often times combined with some optical elements. The exact shape of the sub-pixel structure, spreading from diffraction and optical distortions from lenses among many other factors could be taken into consideration to determine the display reconstruction filter. At first glance, it may seem that this makes our method more difficult to use. However, all these intricacies will be present in the display, whether they are modelled or not.

By making the reconstruction filter explicit and adaptable in our model instead of imposing an implicit (and unfeasible) sinc reconstruction filter, our methods are more general than previous methods and, as such, they are actually better equipped to deal with a wider range of display designs. The used approximation of the reconstruction filter can still be as simple as desired. We believe the box filter to be a good compromise and it is the filter we used for our simulations.

Alternatively, if we treat the display as a black box, each $\boldsymbol{\varphi}_d$ can be individually measured activating the display coefficients one at a time during a calibration stage. If this calibration is done with a camera that mimics the observer's eye, then even $\boldsymbol{\varphi}_r$ could be measured directly, and no further observer model would be needed.

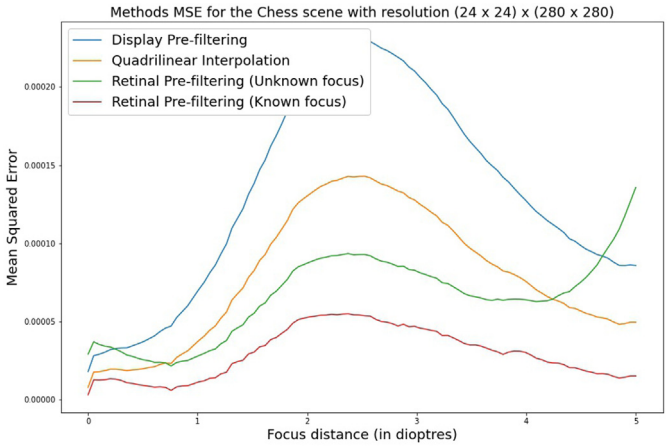


Fig. 14. Comparison of pre-filtering methods under lower resolution.

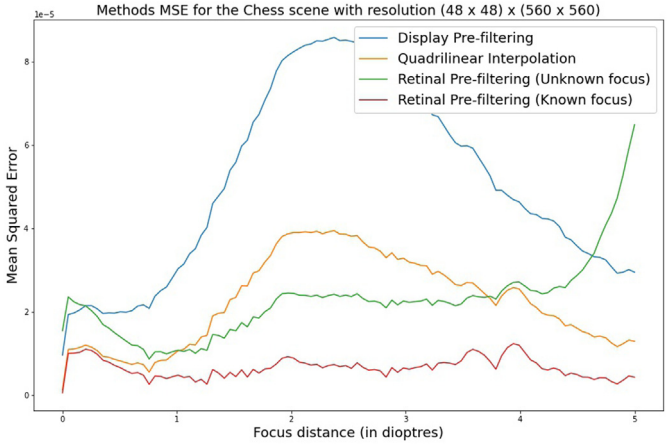


Fig. 15. Comparison of pre-filtering methods under higher resolution.

9.2. Observer limitations

Our retinal pre-filtering method imposes strong restrictions on the observer. We assume the exact pupil diameter and the distance from the eye to the display as known. Also, we assume the eye orientation is always perpendicular to the display. However, it is possible to incorporate all these variations into our method in the future.

Our method currently incorporates the variation in the observer's focus which increases the dimensionality of the optimised error space from what would be the error of a single 2D retinal image to the error of a 3D focal stack. Likewise, the variation of pupil diameter, eye position and eye orientation could each be added as new dimensions of the optimised error space. Note, however, that all of these new considerations do not change the size of the input light field samples $\mathbf{s}$, the size of the output display coefficient $\mathbf{c}$, the size or shape of any of the matrices or the execution time of our method. Only the precomputed values within the matrices would change.

The more restrictive the assumptions towards the observer, the higher the potential for improvement in the results when the assumptions align with reality. As more variation and freedom are given to the observer state, the benefits of the optimisation would be diluted as the method attempts to satisfy multiple scenarios simultaneously. This is true not only for how many different variation types are considered but also for the range within which they can vary. This behaviour has already been demonstrated in this paper when our method optimises for different ranges of focus distances (Fig. 13).

It is also possible to dynamically adapt the targeted observer state and ranges either by eye tracking or adjusting according to the existing contents of the scene and desired depth of field. For example, an application that only involves manipulating objects at arm's length could forego optimising focus on the horizon. If tracking is employed, the variation ranges should be chosen as to cover the accuracy and update speed of the tracking system.

Finally, a more sophisticated eye model could be employed instead of a thin lens approximation. Since, at this point, we only presented simulation results, it is still not clear whether the thin lens model sufficiently captures the fundamental properties of the human imaging process to drive the optimisation or if a more sophisticated eye model would lead to significant practical improvements.

10. Conclusions

In this work, we presented a framework to compute a light field display's coefficients that makes no assumption on scene properties such as being Lambertian or free of occlusions. Our methods target either minimising the light field error or the retinal image error by leveraging information about the display and the observer. We further describe a matrix formulation to allow for a practical implementation.

We present results for our retinal pre-filtering in two scenarios: when the observer focus is known and when it is unknown. Through simulations, we show that when the focus is known, we always achieve better qualitative and quantitative results. When the focus is unknown, our optimisation strategy is still able to significantly improve the results on average.

The natural future work would be to confirm the benefit our method brings in practice by running user studies with a physical display. Furthermore, for such a practical test, it would be interesting to incorporate the observer variations described in Section 9.2. Finally, even though we did not focus on time performance in this work, many software and hardware optimisation strategies could be incorporated to accelerate the method to be used in an interactive application.

CRediT authorship contribution statement

Rafael Romeiro: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Elmar Eisemann:** Writing – review & editing, Supervision, Funding acquisition. **Ricardo Marroquim:** Writing – review & editing, Writing – original draft, Supervision, Investigation, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the Brazilian funding agency CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) under the Process No. 160966/2015-9. The authors would like to thank the Khronos Group for the glTF Sample Models made publicly available and Don McCurdy for the glTF Viewer application.

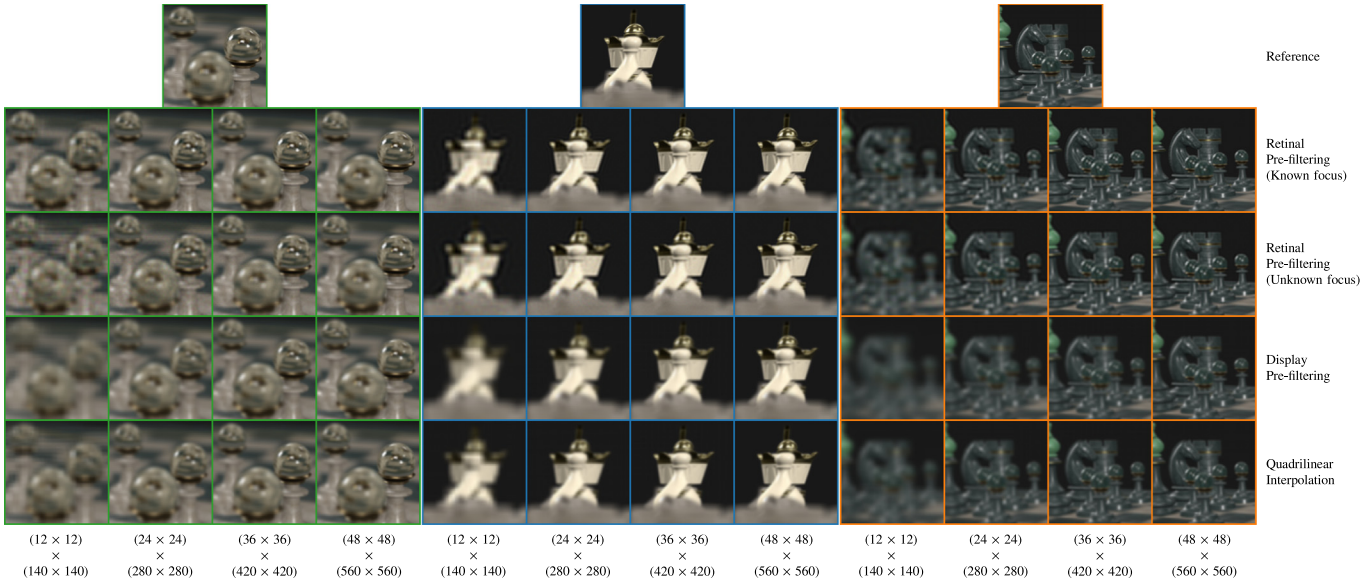


Fig. 16. Comparison of pre-filtering methods for the Chess scene with different display resolutions.

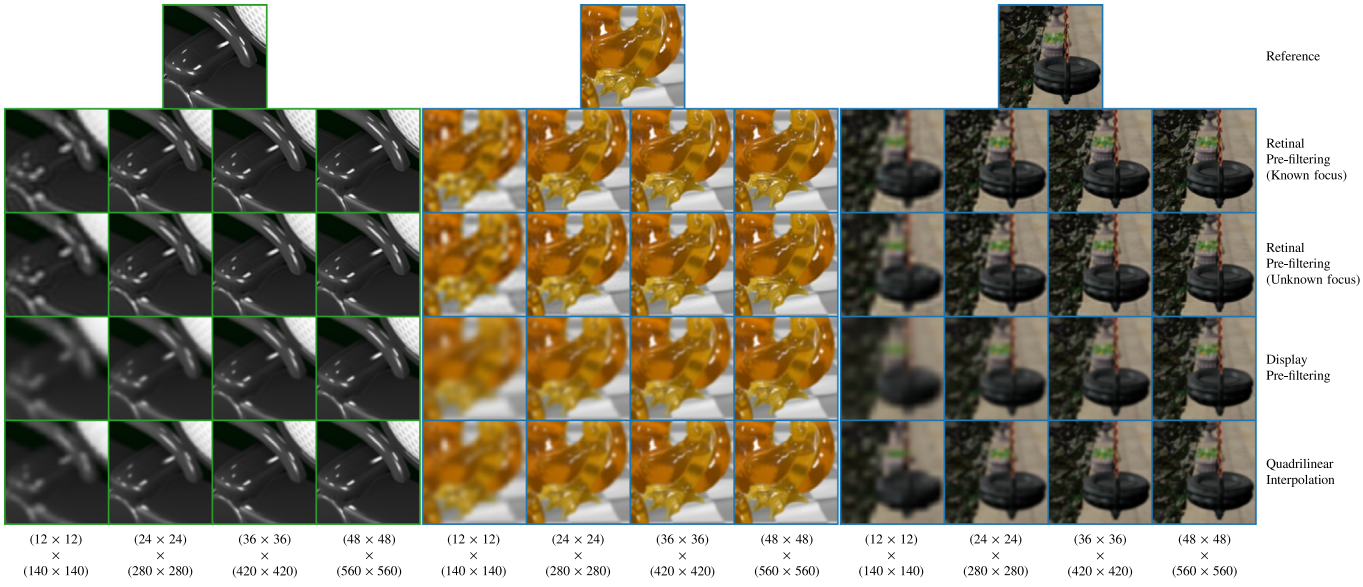


Fig. 17. Comparison of pre-filtering methods for the Car, Dragon and Sponza scenes with different display resolutions.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cag.2024.104033.

References

[1] Levoy M, Hanrahan P. Light field rendering. In: Proceedings of the 23rd annual conference on computer graphics and interactive techniques. New York, NY, USA: Association for Computing Machinery; 1996, p. 31–42. http://dx.doi.org/10.1145/237170.237199.

[2] Gortler SJ, Grzeszczuk R, Szeliski R, Cohen MF. The lumigraph. In: Proceedings of the 23rd annual conference on computer graphics and interactive techniques. New York, NY, USA: Association for Computing Machinery; 1996, p. 43–54. http://dx.doi.org/10.1145/237170.237200.

[3] Isaksen A, McMillan L, Gortler SJ. Dynamically reparameterized light fields. In: Proceedings of the 27th annual conference on computer graphics and interactive techniques. USA: ACM Press/Addison-Wesley Publishing Co.; 2000, p. 297–306. http://dx.doi.org/10.1145/344779.344929.

[4] Chai JX, Tong X, Chan SC, Shum HY. Plenoptic sampling. In: Proceedings of the 27th annual conference on computer graphics and interactive techniques. USA: ACM Press/Addison-Wesley Publishing Co.; 2000, p. 307–18. http://dx.doi.org/10.1145/344779.344932.

[5] Zhang C, Chen T. Generalized plenoptic sampling. Tech. rep. AMP01-06; 200, Carnegie Mellon University; 2001, URL: https://www.microsoft.com/en-us/research/publication/generalized-plenoptic-sampling/.

[6] Stewart J, Yu J, Gortler SJ, McMillan L. A new reconstruction filter for undersampled light fields. In: Proceedings of the 14th eurographics workshop on rendering. Goslar, DEU: Eurographics Association; 2003, p. 150–6.

[7] Liu CL, Shih KT, Huang JW, Chen HH. Light field synthesis by training deep network in the refocused image domain. IEEE Trans Image Process 2020;29:6630–40. http://dx.doi.org/10.1109/TIP.2020.2992354.
[8] Lanman D, Luebke D. Near-eye light field displays. ACM Trans Graph 2013;32(6). http://dx.doi.org/10.1145/2508363.2508366.
[9] Zwicker M, Matusik W, Durand F, Pfister H. Antialiasing for Automultiscopic 3D Displays. In: Akenine-Moeller T, Heidrich W, editors. Symposium on rendering. The Eurographics Association; 2006, http://dx.doi.org/10.2312/EGWR/EGSR06/073-082.
[10] Wetzstein G, Lanman D, Heidrich W, Raskar R. Layered 3D: Tomographic image synthesis for attenuation-based light field and high dynamic range displays. ACM Trans Graph 2011;30(4). http://dx.doi.org/10.1145/2010324.1964990.
[11] Narain R, Albert RA, Bulbul A, Ward GJ, Banks MS, O'Brien JF. Optimal presentation of imagery with focus cues on multi-plane displays. ACM Trans Graph 2015;34(4). http://dx.doi.org/10.1145/2766909.
[12] Mercier O, Sulai Y, Mackenzie K, Zannoli M, Hillis J, Nowrouzezahrai D, et al. Fast gaze-contingent optimal decompositions for multifocal displays. ACM Trans Graph 2017;36(6). http://dx.doi.org/10.1145/3130800.3130846.
[13] Ebner C, Mori S, Mohr P, Peng Y, Schmalstieg D, Wetzstein G, et al. Video see-through mixed reality with focus cues. IEEE Trans Vis Comput Graphics 2022;28(5):2256–66. http://dx.doi.org/10.1109/TVCG.2022.3150504.
[14] Love GD, Hoffman DM, Hands PJ, Gao J, Kirby AK, Banks MS. High-speed switchable lens enables the development of a volumetric stereoscopic display. Opt Express 2009;17(18):15716–25. http://dx.doi.org/10.1364/OE.17.015716, URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-17-18-15716.
[15] Shannon C. Communication in the presence of noise. Proc. IRE 1949;37(1):10–21. http://dx.doi.org/10.1109/JRPROC.1949.232969.
[16] Unser M. Sampling-50 years after Shannon. Proc IEEE 2000;88(4):569–87. http://dx.doi.org/10.1109/5.843002.
[17] Nehab D, Hoppe H. A fresh look at generalized sampling. Found Trends Comput Graph Vis 2014;8(1):1–84. http://dx.doi.org/10.1561/0600000053.
[18] Pharr M, Jakob W, Humphreys G. Physically based rendering: From theory to implementation. 3rd ed.. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 2016.
[19] Lee D, Seung H. Algorithms for non-negative matrix factorization. Adv Neural Inform Process Syst 2001;13.
[20] Lippmann G. Épreuves réversibles donnant la sensation du relief. J Phys Theor Appl 1908;7(1):821–5. http://dx.doi.org/10.1051/jphystap:019080070082100, URL: https://hal.science/jpa-00241406.
[21] Perlin K, Paxia S, Kollin JS. An autostereoscopic display. In: Proceedings of the 27th annual conference on computer graphics and interactive techniques. USA: ACM Press/Addison-Wesley Publishing Co.; 2000, p. 319–26. http://dx.doi.org/10.1145/344779.344933.