RETINAL PRE-FILTERING FOR LIGHT FIELD DISPLAYS

Rafael Oliveira Romeiro

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Sistemas e Computação.

Orientadores: Daniel Ratton Figueiredo
Ricardo Guerra Marroquim

Rio de Janeiro
Setembro de 2024

RETINAL PRE-FILTERING FOR LIGHT FIELD DISPLAYS

Rafael Oliveira Romeiro

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Orientadores: Daniel Ratton Figueiredo
Ricardo Guerra Marroquim

Aprovada por: Prof. Daniel Ratton Figueiredo
Prof. Ricardo Guerra Marroquim
Prof. Claudio Esperança
Prof. Manuel Menezes de Oliveira Neto
Prof. Emilio Ashton Vital Brazil

RIO DE JANEIRO, RJ – BRASIL
SETEMBRO DE 2024

Romeiro, Rafael Oliveira

Retinal Pre-filtering for Light Field Displays/Rafael Oliveira Romeiro. – Rio de Janeiro: UFRJ/COPPE, 2024.

XV, 95 p.: il.; 29, 7cm.

Orientadores: Daniel Ratton Figueiredo
Ricardo Guerra Marroquim

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2024.

Referências Bibliográficas: p. 52 – 55.

1. Light field displays. 2. Retinal pre-filtering. 3. Anti-aliasing. I. Figueiredo, Daniel Ratton *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

À minha amada avó, Aurora,
que hoje ilumina o céu com seu
eterno brilho.

Acknowledgements

First and foremost, I would like to thank my parents, Maria Cristina and André Luiz, for all the unconditional emotional support (and financial support, when necessary) throughout the years. Mom and dad, you are the foundation of my entire existence and have shaped (for the better) my values and world views. This thesis is as much your achievement as it is mine.

To my sister, Paula, and to my brother-in-law, Rafael, thank you for your ever-present love and for always broadening my horizons. I am also immensely grateful to my newly arrived niece, Gimmaima, for bringing me new meanings for the future.

To my advisor, Ricardo Marroquim, thank you for being the most patient man on Earth. You probably achieved some level of sainthood after enduring me for so long. Without your unwavering guidance, this nearly a decade-long endeavor would not have been successfully accomplished.

To my "fake" advisor, Daniel Figueiredo, thank you for "adopting" me when I became an "orphan" in regards to UFRJ after Ricardo went to TU Delft and for promptly providing all the assistance with my (numerous) bureaucratic issues. Your pure and true enthusiasm for scientific research has always been a huge inspiration to me.

To my friend Emilio Vital Brazil, thank you for having extended me so many opportunities without me ever truly deserving them. I am deeply indebted to you. Your wisdom and expertise were invaluable in the development of this work.

I would also like to express my gratitude to professor Elmar Eisemann for welcoming me into his lab and for his insightful comments during my visit to the Netherlands.

To my therapist, Tania Steiner, thank you for keeping me emotionally afloat these past couple years and for not locking me away in an institute... yet.

Finally, I'd like to acknowledge the Brazilian research funding agency CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for the financial support received during the first half of my doctorate.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

PRÉ-FILTRAGEM RETINAL PARA DISPLAYS DE CAMPO DE LUZ

Rafael Oliveira Romeiro

Setembro/2024

Orientadores: Daniel Ratton Figueiredo
Ricardo Guerra Marroquim

Programa: Engenharia de Sistemas e Computação

Os coeficientes responsáveis por produzir o sinal emitido por um display de campo de luz são geralmente calculados de forma a aproximar a radiância sobre um conjunto de raios amostrados no espaço de campo de luz. No entanto, nem todas as informações contidas no sinal de campo de luz são de igual importância para um observador. Propomos uma pré-filtragem retinal das amostras de campo de luz que leva em consideração o processo de formação de imagem do observador para determinar os coeficientes de exibição que, em última análise, produzirão melhores imagens retinais para uma gama de distâncias focais. Demonstramos um aumento significativo na definição da imagem sem alterar a resolução do display.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

RETINAL PRE-FILTERING FOR LIGHT FIELD DISPLAYS

Rafael Oliveira Romeiro

September/2024

Advisors: Daniel Ratton Figueiredo
Ricardo Guerra Marroquim

Department: Systems Engineering and Computer Science

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.

Contents

List of Figures

Chapter 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 magnifying 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.

In Chapter 2, we define the light field signal and discuss related work on the topic of sampling and reconstruction of such signals. In Chapter 3, we cover multiple display architectures, referencing related work and indicating their differences and similarities to our work. In Chapter 4, we present a brief review of sampling and

reconstruction theory and reference related work concerning pre-filtering techniques. In Chapter 5, we apply well-established techniques to compute the display's coefficients that minimises the light field error while taking into account the display's reconstruction filter. In Chapter 6, we present a comprehensive theoretical framework to describe the retinal image produced by a light field display, formalise the computation of the display's coefficients as a minimisation problem over the retinal image error and propose an iterative solution. In Chapter 7, we demonstrate how the theoretical continuous methods can be further translated to a matrix formulation that can be effectively implemented in practice for discrete input signals. In Chapter 8, we describe an experiment we conducted using our matrix formulation to simulate a light field display and the image generated at the observer's retina, detailing every function and parameter used. In Chapter 9, we present our results using our virtual display with different display resolutions and different filtering strategies. We compare the retinal error under multiple conditions and provide analysis for execution time and method convergence. In Chapter 10, we delimit the conditions for our method to be applicable and specify the necessary assumptions in regards to the display and the observer. In Chapter 11, we gather our final conclusions and explore possible future work. In Appendix A, we describe the linear spaces and sub-spaces that emerge in our work and overview their definitions of inner product, norm, distance, orthogonality, projection and least squares. In Appendices B, C and D, we present our results for other scenes.

To summarise, the main contributions of this work are:

- a light field display pre-filtering approach that takes into account the display's reconstruction filter (Chapter 5);
- a continuous description of the retinal image reconstruction as a function of the display's reconstruction filter (Section 6.2) and its accompanying optimisation problem for light field display pre-filtering (Section 6.3);
- an iterative solution for the proposed retinal pre-filtering optimisation problem (Section 6.4);
- a numerical integration solution for practical use of our method in the context of a discrete input signal and the resulting matrix formulation (Chapter 7);
- a strategy to further reduce the retinal error by adapting the optimised focus range based on current observer focus distance (Chapter 9).

Also, this research resulted in the publication of a scientific article:

Rafael Romeiro, Elmar Eisemann, Ricardo Marroquim
"Retinal pre-filtering for light field displays"
Computers & Graphics, Volume 123, 2024, 104033, ISSN 0097-8493,
https://doi.org/10.1016/j.cag.2024.104033

Chapter 2

Light Field and Related Work

2.1 Plenoptic Function

In geometric optics, light propagation is modelled through idealised rays perpendicular to the wavefronts of the electromagnetic radiation. Taking into consideration all rays travelling from every position (3 dimensions) and in every direction (2 dimensions), the space of rays in a 3-dimensional space is 5-dimensional. The 5-dimensional function that represents the radiance of all light rays in a 3-dimensional space is called the plenoptic function. Sometimes, more dimensions are considered in the plenoptic function to differentiate the radiance at different times, polarisation angles and wavelengths. However, within the context of our work, we assume static scenes with incoherent lighting and, as it is typically handled in computer graphics, we consider separately the wavelengths of three independent colour channels that follow the response of the cone cells in the human eye (RGB channels).

Since the radiance remains constant along rays of light through empty space, if a surface can be seen free from any obstruction by an observer, then the plenoptic function at the surface is sufficient to represent all light that crosses this surface to reach the observer, regardless of how complex the plenoptic function might be behind this surface. In this case, we can reduce the 5-dimensional plenoptic function to a 4-dimensional function, which is then called a light field.

The light field function can be parameterised in a variety of ways. Most commonly, a two-plane parameterisation is followed, which is also used throughout this thesis. In this parameterisation system, two predefined parallel planes are used to respectively register the entry and exit points of each light ray (Figure2.1).

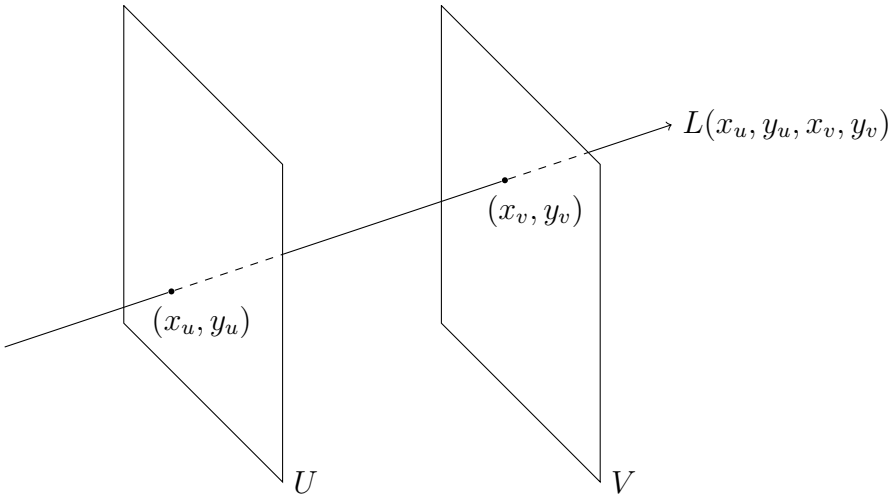


Figure 2.1: Two-plane parameterisation.

2.2 Sampling and Reconstruction

An analytical description of the scene's light field is typically not available. Therefore, instead of having a continuous representation of its function, we must resort to a discrete representation. The radiance of multiple light rays can be sampled either by ray tracing a synthetic scene or captured by a physical device (light field camera), in both cases, from different viewpoints. 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.

As previously mentioned, we follow a two-plane parameterisation. A set of entry points on one plane combined with a set of exit points on the other plane form a 4-dimensional lattice. Each 4-dimensional 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. Note that rays reconstructed this way can be used to compose a 2-dimensional render as well as to resample the entire 4-dimensional 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 Figure 2.2a. 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 Figure 2.2b. 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 Figure 2.2b. 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 Section 3.2.

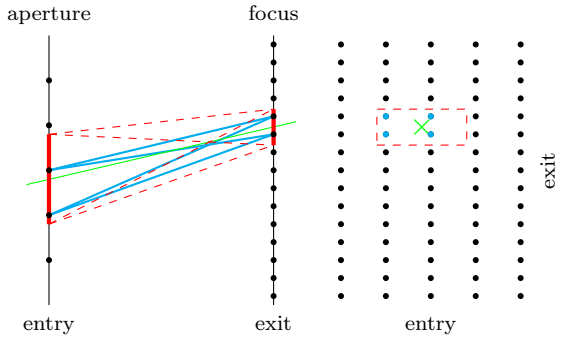
Chai et al. [4] describe the spectral support of a light field as being bounded according to the minimum and maximum depths of the scene. A reconstruction filter is then designed based on these bounds. The proposed reconstruction filter is analogous to a camera with aperture located on the sampling entry 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 Figure 2.2b. 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.

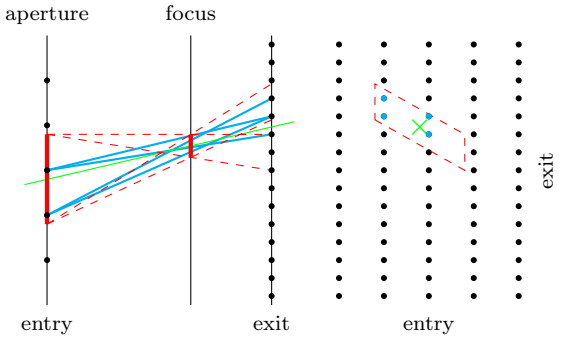
Stewart et al. [6] propose a reconstruction filter that is analogous to a combination of two cameras. One camera has a narrow aperture located on the sampling entry plane and focus located on the optimal focus depth, following Chai et al. [4]. The other camera has a wide aperture located on the sampling 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 Figure 2.2c. 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. In our simulation, for example, we use such a method with a box filter defined over two display planes independent of the original sampling planes. Moreover, it is possible to design reconstruction filters that are not separable over any axis. Nevertheless, our method is general and can handle any definition of a reconstruction filter.

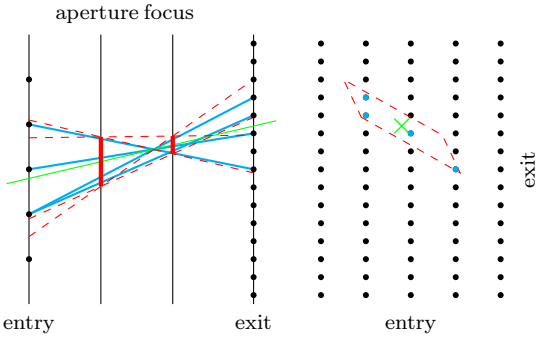
In addition, machine learning-based methods work without explicitly defining a reconstruction filter. For example, Liu et al. [7] propose a light field synthesis method based on a deep neural network. The network is first trained with light fields with multiple sub-aperture images (high angular resolution). 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.

Figure 2.2: 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).

Chapter 3

Displays and Related Work

3.1 Stereoscopic displays

Traditional 2-dimensional displays emit light equally in all directions like a Lambertian surface. Despite being limited to forming flat images, multiple techniques can be applied to traditional 2-dimensional displays to induce the brain into interpreting depth. Some of these techniques are: perspective, shading and lighting, occlusion and motion parallax (when there is movement).

Stereoscopic displays present two separate 2-dimensional images, one for the left eye and the other for the right eye. The binocular disparity present in this stereo pair of images is a major contributor to depth perception through stereopsis. This stimulates a coordinated rotation of both eyes, called vergence, as to keep the projection of an observed object centered. At the same time, each eye needs to adjust its optical power by bending its lenses in order to keep the observed object focused at varying distances, a process called accommodation. However, when binocular disparity is created using two different images, vergence follows the depth of the virtual object, while accommodation performed in each eye must still follow the depth of the images. This causes a mismatch known as the vergence-accommodation conflict. Fundamentally, this problem comes from the display not being able to reproduce the correct focus cues.

3.2 Light field displays

A light field display is a device capable of reconstructing a full 4D light field signal. Some light field displays are considered automultiscopic displays when they can generate a view-dependent image without the need for special glasses but still lack refocusing capabilities. In this work, we are interested in multifocal light field displays, that is, when the angular resolution of the display allows for multiple dif-

ferent views to enter the observer's pupil, which allows refocusing, a technique first described as integral imaging [8].

Lanman and Luebke [9] 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 2.2 and depicted in Figure 2.2b. 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.

3.3 Layered displays

Layered displays are display devices that stack two or more image layers at different depths. These layers can be combined additively or multiplicatively to produce a light field signal. However, a stack of 2D images correspond to a 3D set of coefficients, therefore the resulting 4D light field signal is restricted in degrees of freedom, thence some authors classify layered displays as compressive light field 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.

As previously stated, 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 complete 4D solution that minimises either a 4D error (Chapter 5) or a (compressive) 3D error (Chapter 6). Therefore, our method was not developed with layered displays in mind. That said, our method can be easily applied, as it is, to additive layered displays and we also believe that, with proper adaptations not provided in this work, it could be modified to contemplate multiplicative layered displays as well.

Chapter 4

Pre-filtering and Related Work

As discussed in Section 2.2, 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 analogue 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.

4.1 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 $\mathbf{c}$ (Equation 4.1). At a later stage, a continuous signal $\tilde{f}$ can be obtained by convolving the sample values $\mathbf{c}$ with a reconstruction filter φ (Equation 4.2).

$$\mathbf{c}[i] = (f * \psi)(\boldsymbol{\lambda}[i]) = \int_X f(x)\, \psi(\boldsymbol{\lambda}[i] - x)\, \mathrm{d}x \tag{4.1}$$

$$\tilde{f}(x) = (\mathbf{c} * \varphi)(x) = \sum_{i=1}^{n} \mathbf{c}[i]\, \varphi(x - \boldsymbol{\lambda}[i]) \tag{4.2}$$

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 $V(\boldsymbol{\varphi})$ (Equation 4.3).

$$V(\boldsymbol{\varphi}) = \left\{ \sum_{i=1}^{n} c_i\, \boldsymbol{\varphi}[i](\mathbf{x}) \,\middle|\, c_i \in \mathbb{R} \right\} \tag{4.3}$$

Following the sampling theorem [15], the reconstruction filter $\varphi = \text{sinc}$ 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 $\varphi = \text{sinc}$ 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 $\varphi = \text{sinc}$ regardless of the choice of pre-filter. In these cases, the pre-filter $\psi = \text{sinc}$ 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 $\left\| f - \tilde{f} \right\|$ 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 $V(\boldsymbol{\varphi})$ (Equation 4.4).

$$\tilde{f}(x) = \sum_{i=1}^{n} \langle f, \mathring{\boldsymbol{\varphi}}[i] \rangle\, \boldsymbol{\varphi}[i](\mathbf{x}) \tag{4.4}$$

Where $\mathring{\boldsymbol{\varphi}}$ is the dual basis of $\boldsymbol{\varphi}$ and can be uniquely determined by the biorthogonality condition (Equation 4.5) [16] [17].

$$\langle \mathring{\boldsymbol{\varphi}}[i], \boldsymbol{\varphi}[j] \rangle = \boldsymbol{\delta}[i - j] = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j \end{cases} \tag{4.5}$$

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., $\mathring{\varphi} = \varphi = \text{sinc}$. This coincides with the traditional anti-aliasing paradigm of sinc pre-filtering and sinc reconstruction.

4.2 Pre-filtering for displays

Zwicker et al. [18] 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 surfaces without occlusions.

Kajiya and Ullner [19] approximated the reconstruction filter of CRT displays with a Gaussian function and argued that this reconstruction filter should be taken into account during pre-filtering in order to produce the best possible images. They demonstrated this technique in the context of rendering text with each individual character being pre-filtered and stored prior to display.

Rocha et al. [20] updated this approach for modern displays, where the use of a box filter reasonably represents the reconstruction filter of LCD displays. In this case, however, the optimal (in the least-squares sense) pre-filter should also be a box function, as it is self-dual. The authors then went one step further and argued that the human optical system should also be taken into account during pre-filtering. For an ideal observer focusing on the screen, they approximated the Airy disk corresponding to the observer's theoretical PSF with a quadratic B-spline. Using the observer's PSF in combination with the display's reconstruction filter, a complete reconstruction pipeline is defined from the display's coefficients all the way to the retinal image. The proposed pre-filter is capable of producing sharper retinal images while offering a good compromise between aliasing and ringing.

Similarly, in our work we also take into account both the display's reconstruction filter and the observer's optical system for pre-filtering. In our case, however, the display emits a 4D light field instead of a 2D flat image and we model the observer's ability to focus at different distances instead of the focus being fixed on the screen. Furthermore, we don't model the observer's PSF.

Chapter 5

Light Field Display Pre-filtering

As previously stated, we don't 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 8.3, 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:

$$\boldsymbol{\lambda}_d = \left(\begin{bmatrix} x_a & y_a \\ x_b & y_b \end{bmatrix}_0, \begin{bmatrix} x_a & y_a \\ x_b & y_b \end{bmatrix}_1, ..., \begin{bmatrix} x_a & y_a \\ x_b & y_b \end{bmatrix}_{n_d} \right) \tag{5.1}$$

Where $n_d = n_a^2\, n_b^2$ is the total number of display samples and $\boldsymbol{\lambda}_d[i_d]$ 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 Equation 5.2, where $\boldsymbol{\psi}_d[i_d]$ is the display pre-filter shifted over the i_d-th display sample, i.e, $\boldsymbol{\psi}_d[i_d](\mathbf{x}_d) = \psi_d(\mathbf{x}_d - \boldsymbol{\lambda}_d[i_d])$. Likewise, a light field $\tilde{L}_d$ is reconstructed

by the display according to Equation 5.3.

$$\mathbf{c}[i_d] = (L_d * \psi_d)(\boldsymbol{\lambda}_d[i_d]) = \iint_A \iint_B L_d(\mathbf{x}_d)\, \boldsymbol{\psi}_d[i_d](\mathbf{x}_d)\, \mathrm{d}\mathbf{x}_d \tag{5.2}$$

$$\tilde{L}_d(\mathbf{x}_d) = (\mathbf{c} * \varphi_d)(\mathbf{x}_d) = \sum_{i_d=1}^{n_d} \mathbf{c}[i_d]\, \boldsymbol{\varphi}_d[i_d](\mathbf{x}_d) \tag{5.3}$$

Equation 5.3 describes the light field reconstruction process as a linear combination of base functions φ_d weighted by coefficients $\mathbf{c}$. We sought to keep this representation purposefully abstract in order to accommodate a greater variety of display architectures. The exact physical meaning of these coefficients depends on underlying structures and may be associated to forward current through an LED, opacity of a medium, exposure time, etc.

The pre-filtering method described by Equation 5.2, 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] as described in Chapter 4. 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.

Chapter 6

Retinal Pre-filtering

In the previous chapter 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.

6.1 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 [21] 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.

$$E(\mathbf{x}_r) = \frac{1}{z_r^2} \iint_P L_e\left(\begin{bmatrix}\mathbf{x}_r \\ \mathbf{x}_p\end{bmatrix}\right) \left|\cos^4\theta\right| \mathrm{d}\mathbf{x}_p \tag{6.1}$$

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 toward 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 Equation 6.1 to:

$$E(\mathbf{x}_r) = \iint_P L_e\left(\begin{bmatrix}\mathbf{x}_r \\ \mathbf{x}_p\end{bmatrix}\right) \mathrm{d}\mathbf{x}_p \tag{6.2}$$

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

$$L_e(\mathbf{x}_e) = L_d\left(\underbrace{\begin{bmatrix} z_a & 1 - z_a\zeta_f \\ z_b & 1 - z_b\zeta_f \end{bmatrix}}_{T_{de}(\zeta_f)} \mathbf{x}_e\right) \tag{6.3}$$

Where ζ_f indicates the reciprocal distance (in diopters) 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.

6.2 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 (Equation 6.4). 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 $\mathbf{x}_r$ (Figure 6.1). This space is also similar to the refocused image domain of Liu et al. [7].

$$G(\zeta_f, \mathbf{x}_r) = \iint_P L_d\left(T_{de}(\zeta_f)\begin{bmatrix}\mathbf{x}_r \\ \mathbf{x}_p\end{bmatrix}\right) \mathrm{d}\mathbf{x}_p \tag{6.4}$$

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:

$$\begin{aligned}
\tilde{G}(\zeta_f, \mathbf{x}_r) &= \iint_P \tilde{L}_d\left(T_{de}(\zeta_f)\begin{bmatrix}\mathbf{x}_r \\ \mathbf{x}_p\end{bmatrix}\right) \mathrm{d}\mathbf{x}_p \\
&= \iint_P \sum_{i_d=1}^{n_d} \mathbf{c}[i_d]\, \boldsymbol{\varphi}_d[i_d]\left(T_{de}(\zeta_f)\begin{bmatrix}\mathbf{x}_r \\ \mathbf{x}_p\end{bmatrix}\right) \mathrm{d}\mathbf{x}_p \\
&= \sum_{i_d=1}^{n_d} \mathbf{c}[i_d] \iint_P \boldsymbol{\varphi}_d[i_d]\left(T_{de}(\zeta_f)\begin{bmatrix}\mathbf{x}_r \\ \mathbf{x}_p\end{bmatrix}\right) \mathrm{d}\mathbf{x}_p \\
&= \sum_{i_d=1}^{n_d} \mathbf{c}[i_d]\, \boldsymbol{\varphi}_r[i_d](\zeta_f, \mathbf{x}_r) = \Phi\mathbf{c}
\end{aligned} \tag{6.5}$$

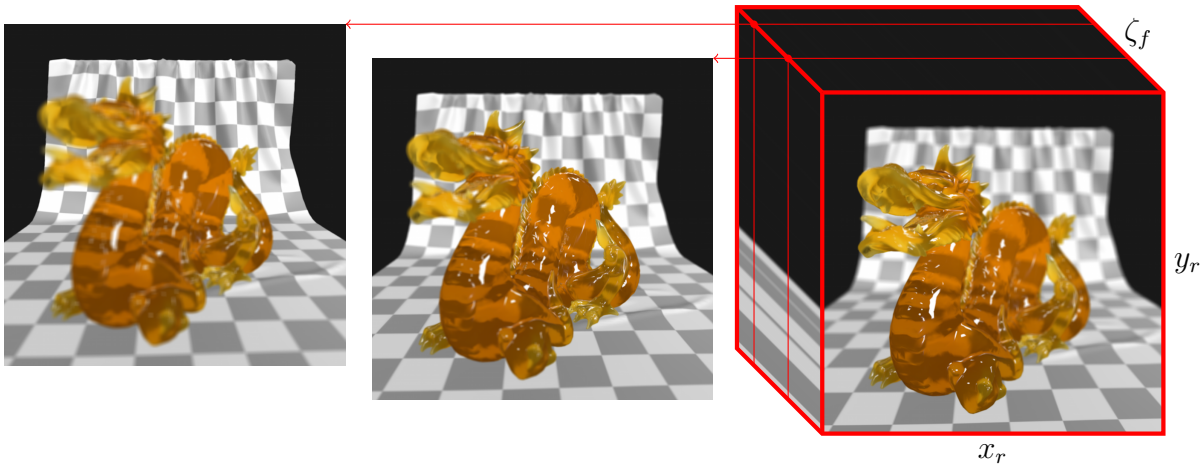


Figure 6.1: Example of a focal stack $G(\zeta_f, \mathbf{x}_r)$ where each layer defined by ζ_f is a retinal image focused at ζ_f.

$$\text{where } \boldsymbol{\varphi}_r[i_d](\zeta_f, \mathbf{x}_r) = \iint_P \boldsymbol{\varphi}_d[i_d]\left(T_{de}(\zeta_f)\begin{bmatrix}\mathbf{x}_r\\ \mathbf{x}_p\end{bmatrix}\right) \mathrm{d}\mathbf{x}_p \tag{6.6}$$

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 (Figure 6.2). 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 Φ.

The linear operator Φ 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 Φ, refer to Appendix A.

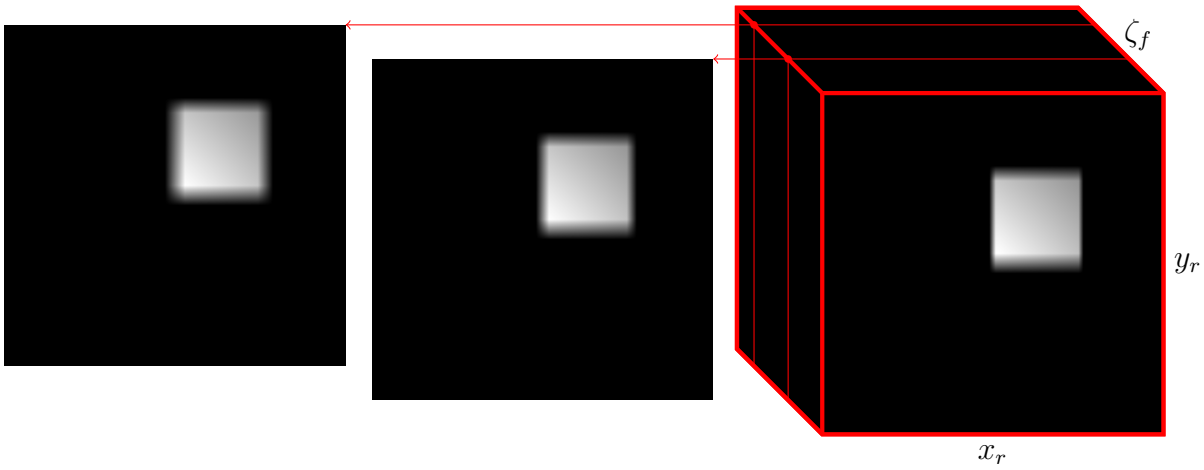


Figure 6.2: Example of an individual base function $\boldsymbol{\varphi}_r(\zeta_f, \mathbf{x}_r)$.

6.3 Optimisation problem

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

6.4 Iterative Solution

Lee and Seung [22] 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:

$$\mathbf{c} \leftarrow \mathbf{c} \otimes \frac{\Phi^\intercal G}{\Phi^\intercal\Phi\,\mathbf{c}} \tag{6.7}$$

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 $\Phi^\intercal G$ 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 Equation 6.8. The denominator $\Phi^\intercal\Phi\,\mathbf{c}$ is the result of the $n_d \times n_d$ Gram matrix $\Phi^\intercal\Phi$ multiplied by $\mathbf{c}$. Each term of $\Phi^\intercal\Phi$ corresponds to an inner product between two functions in $\boldsymbol{\varphi}_r$, as defined in Equation 6.9.

$$\begin{aligned}(\Phi^\intercal G)[i_d] &= \langle \boldsymbol{\varphi}_r[i_d], G\rangle \\ &= \int_F\int_R \boldsymbol{\varphi}_r[i_d](\zeta_f, \mathbf{x}_r)\, G(\zeta_f, \mathbf{x}_r)\, \mathrm{d}\mathbf{x}_r\, \mathrm{d}\zeta_f\end{aligned} \tag{6.8}$$

$$
\begin{aligned}
(\Phi^{\intercal}\Phi)[i_d, j_d] &= \langle \boldsymbol{\varphi}_r[i_d], \boldsymbol{\varphi}_r[j_d] \rangle \\
&= \int_F \iint_R \boldsymbol{\varphi}_r[i_d](\zeta_f, \mathbf{x}_r)\, \boldsymbol{\varphi}_r[j_d](\zeta_f, \mathbf{x}_r)\, \mathrm{d}\mathbf{x}_r\, \mathrm{d}\zeta_f
\end{aligned} \tag{6.9}
$$

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 Chapter 9.

Substituting Equation 6.4 in Equation 6.8 we have:

$$
(\Phi^{\intercal}G)[i_d] = \int_F \iint_R \iint_P L_d\left(T_{de}(\zeta_f) \begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_p \end{bmatrix}\right) \boldsymbol{\varphi}_r[i_d](\zeta_f, \mathbf{x}_r)\, \mathrm{d}\mathbf{x}_p\, \mathrm{d}\mathbf{x}_r\, \mathrm{d}\zeta_f \tag{6.10}
$$

Hence, given a target continuous light field L_d we can use Equations 6.7, 6.9 and 6.10 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.

Chapter 7

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 Chapter 5 and the retinal pre-filtering in Chapter 6. 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.

7.1 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.

$$\boldsymbol{\lambda}_s = \left(\begin{bmatrix} x_u & y_u \\ x_v & y_v \end{bmatrix}_0, \begin{bmatrix} x_u & y_u \\ x_v & y_v \end{bmatrix}_1, ..., \begin{bmatrix} x_u & y_u \\ x_v & y_v \end{bmatrix}_{n_s} \right) \tag{7.1}$$

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:

$$L_d(\mathbf{x}_d) = L_s\left(\underbrace{\frac{1}{z_b - z_a} \begin{bmatrix} z_b - z_u & z_u - z_a \\ z_b - z_v & z_v - z_a \end{bmatrix}}_{T_{sd}} \mathbf{x}_d \right) \tag{7.2}$$

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

$$L_e(\mathbf{x}_e) = L_s\left(\underbrace{\begin{bmatrix} z_u & 1 - z_u\zeta_f \\ z_v & 1 - z_v\zeta_f \end{bmatrix}}_{T_{se}(\zeta_f)} \mathbf{x}_e \right) \tag{7.3}$$

The display pre-filtering as described by Equation 5.2 can be redefined over L_s instead of L_d, becoming:

$$\mathbf{c}[i_d] = \iint_A \iint_B L_s(T_{sd}\mathbf{x}_d)\, \boldsymbol{\psi}_d[i_d](\mathbf{x}_d)\, \mathrm{d}\mathbf{x}_d \tag{7.4}$$

Similarly, the numerator of the iterative rule of the retinal pre-filtering as described by Equation 6.10 can also be redefined over L_s instead of L_d, becoming:

$$(\Phi^\intercal G)[i_d] = \iint_F \iint_R \iint_P L_s\left(T_{se}(\zeta_f) \begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_p \end{bmatrix}\right) \boldsymbol{\varphi}_r[i_d](\zeta_f, \mathbf{x}_r)\, \mathrm{d}\mathbf{x}_p\, \mathrm{d}\mathbf{x}_r\, \mathrm{d}\zeta_f \tag{7.5}$$

7.2 Numerical integration

Since there is no continuous description of L_s, only the samples $\mathbf{s}[i_s] = L_s(\boldsymbol{\lambda}_s[i_s])$ are available to any pre-filtering method. For this reason, the integrals must be approximated with a numerical approach (Monte Carlo integration).

A numerical approximation of Equation 7.4 can be defined as:

$$\mathbf{c}[i_d] \approx \frac{|A| \times |B|}{n_s} \sum_{i_s}^{n_s} \mathbf{s}[i_s]\, \boldsymbol{\psi}_d[i_d](T_{sd}^{-1}\boldsymbol{\lambda}_s[i_s]) \tag{7.6}$$

Where $|A|$ and $|B|$ are the display's finite areas over the planes A and B, respectively.

The integral in Equation 7.5, on the other hand, is particularly non-trivial. This comes from the fact that a light field sample $\mathbf{s}[i_s]$ does not correspond to a point in the integrand.

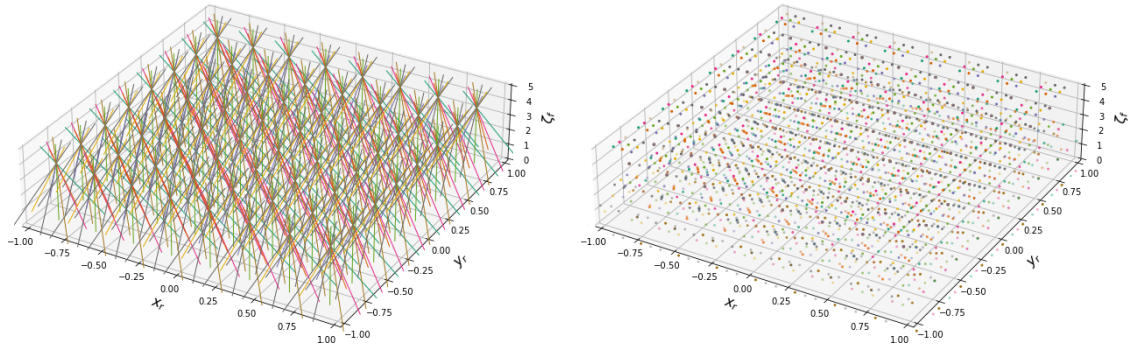
We can define the 2D retina coordinates $\boldsymbol{\xi}_r[i_s](\zeta_f)$ and 2D pupil coordinates $\boldsymbol{\xi}_p[i_s]$ corresponding to the i_s-th light field sample for when the observer is focusing at ζ_f:

$$\begin{bmatrix} \boldsymbol{\xi}_r[i_s](\zeta_f) \\ \boldsymbol{\xi}_p[i_s] \end{bmatrix} = \underbrace{\frac{1}{z_v - z_u} \begin{bmatrix} z_v\zeta_f - 1 & 1 - z_u\zeta_f \\ z_v & -z_u \end{bmatrix}}_{T_{se}^{-1}(\zeta_f)} \boldsymbol{\lambda}_s[i_s] \tag{7.7}$$

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 (Figure 7.1a).

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, $\boldsymbol{\varphi}_r[i_d]$ (Equation 6.6), 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 (Figure 7.1b).



(a) (b)

Figure 7.1: (a) Example of $(3 \times 3) \times (8 \times 8)$ light field samples in the focal stack 3D space. (b) Example of $(3 \times 3) \times (8 \times 8) \times 5$ sample coordinates of $\boldsymbol{\phi}$.

Finally, Equation 7.5 can be numerically approximated by:

$$(\Phi^{\intercal} G)[i_d] \approx \frac{|F| \times |R| \times |P|}{n_f \, n_s} \sum_{i_f}^{n_f} \sum_{i_s}^{n_s} \mathbf{s}[i_s] \, \boldsymbol{\phi}[i_f, i_d, i_s] \tag{7.8}$$

Where $|F|$ is the length of the interval of focus distances, $|R|$ is the area of the retina and $|P|$ is the area of the pupil.

The discretisation of F and its subsequent numerical integration allows us to compute each focus slice individually and later compose a desired focus interval by combining the appropriate focus samples. That is, Equation 7.8 can be rewritten as:

$$(\Phi^{\intercal} G)[i_d] \approx \frac{|F|}{n_f} \sum_{i_f}^{n_f} \underbrace{\frac{|R| \times |P|}{n_s} \sum_{i_s}^{n_s} \mathbf{s}[i_s] \, \boldsymbol{\phi}[i_f, i_d, i_s]}_{\eta[i_d, i_f]} \tag{7.9}$$

Where $\eta[i_d, i_f]$ describes what the numerator terms would be for a fixed focus distance $\boldsymbol{\zeta}_f[i_f]$ and the numerator terms $(\Phi^{\intercal} G)[i_d]$ for an interval of focus distances can be produced by averaging $\eta[i_d, i_f]$ over i_f.

It is also important to note that most light field samples lie outside any given function $\boldsymbol{\varphi}_r[i_d]$. For this reason, we divide the retina region R into n_r disjoint sub-regions $\mathbf{R}[i_r]$ such that $R = \bigsqcup_{i_r \in [1, n_r]} \mathbf{R}[i_r]$. We then stratify the light field samples into the sub-regions $\mathbf{R}[i_r]$ according to their retinal position $\boldsymbol{\xi}_r[i_s]$ for every focus distance $\boldsymbol{\zeta}_f[i_f]$. In other words, for each retinal sub-region $\mathbf{R}[i_r]$ and each focus distance $\boldsymbol{\zeta}_f[i_f]$, we compute the set of indices $\mathbf{I}_s[i_r, i_f]$ such that $i_s \in \mathbf{I}_s[i_r, i_f] \iff \boldsymbol{\xi}_r[i_s](\boldsymbol{\zeta}_f[i_f]) \in \mathbf{R}[i_r]$. Then, instead of computing an integral over the entire region R, we split the integral into multiple integrals, each defined over a sub-region $\mathbf{R}[i_r]$.

Each $\eta[i_d, i_f]$ term can then be computed as:

$$\eta[i_d, i_f] \approx |P| \sum_{i_r}^{n_r} \frac{|\mathbf{R}[i_r]|}{|\mathbf{I}_s[i_r, i_f]|} \sum_{i_s \in \mathbf{I}_s[i_r, i_f]} \mathbf{s}[i_s]\, \boldsymbol{\phi}[i_f, i_d, i_s] \tag{7.10}$$

Where $|\mathbf{R}[i_r]|$ is the area of the retinal sub-region $\mathbf{R}[i_r]$ and $|\mathbf{I}_s[i_r, i_f]|$ is the number of light field samples that reach the retina inside the sub-region $\mathbf{R}[i_r]$ when the focus distance is $\boldsymbol{\zeta}_f[i_f]$.

This stratification of the retina has the purpose of enabling us to quickly discard all light field samples associated with retinal sub-regions that do not intersect the function $\boldsymbol{\varphi}_r[i_d]$ used in the integral. This method vastly improves the efficiency of the integral computation and drastically reduces its memory footprint.

All the necessary values of $\boldsymbol{\phi}$ can be pre-computed and used temporarily to calculate the total contribution each light field sample has to each term of the numerator and stored as a matrix. The same can be done for all other operations. Once all matrices are 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.

7.3 Matrix formulation

We can now describe every operation of our methods in matrix form. For this purpose, we define five matrices, named M_1 up to M_5. Each matrix has the numerical approximation of its corresponding integral embedded into it.

When the goal is to minimise the display light field error, as described in Chapter 5, the display pre-filtering operation can be defined as an $n_d \times n_s$ matrix M_1 that encodes Equation 7.4:

$$\mathbf{c} = M_1 \mathbf{s} \tag{7.11}$$

When the goal is to minimise the retinal images error instead, as described in Chapter 6 and, more specifically, in Section 6.4, the retinal pre-filtering is defined by an iterative process using two matrices, an $n_d \times n_s$ matrix M_2 that encodes Equation 7.5 and an $n_d \times n_d$ matrix M_3 that encodes Equation 6.9:

$$\mathbf{c} \leftarrow \mathbf{c} \otimes \frac{M_2 \mathbf{s}}{M_3 \mathbf{c}} \tag{7.12}$$

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 $\mathbf{r}$ with a resolution of $n_r \times n_r$ pixels using an $n_r^2 \times n_d$ matrix M_4 computed for a given ζ_f:

$$\mathbf{r} = M_4 \, \mathbf{c} \tag{7.13}$$

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 computed for a given ζ_f:

$$\mathbf{r} = M_5 \, \mathbf{s} \tag{7.14}$$

Chapter 8

Experimental Setup

8.1 Light field sampling

To compute the sampled light field $\mathbf{s}$ for each scene, we render 20 $\times$ 20 views of 1080 $\times$ 1080 pixels each. The camera positions are evenly spaced over a 10 mm $\times$ 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 $\times$ 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 $\times$ 20) $\times$ (1080 $\times$ 1080) light rays from a light field L_s parameterised over the planes at $z_u = 0$ and $z_v = 0.265$.

8.2 Observer imaging system

We consider the observer to have an 8 mm x 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 x 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 x 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).

8.3 Light field display

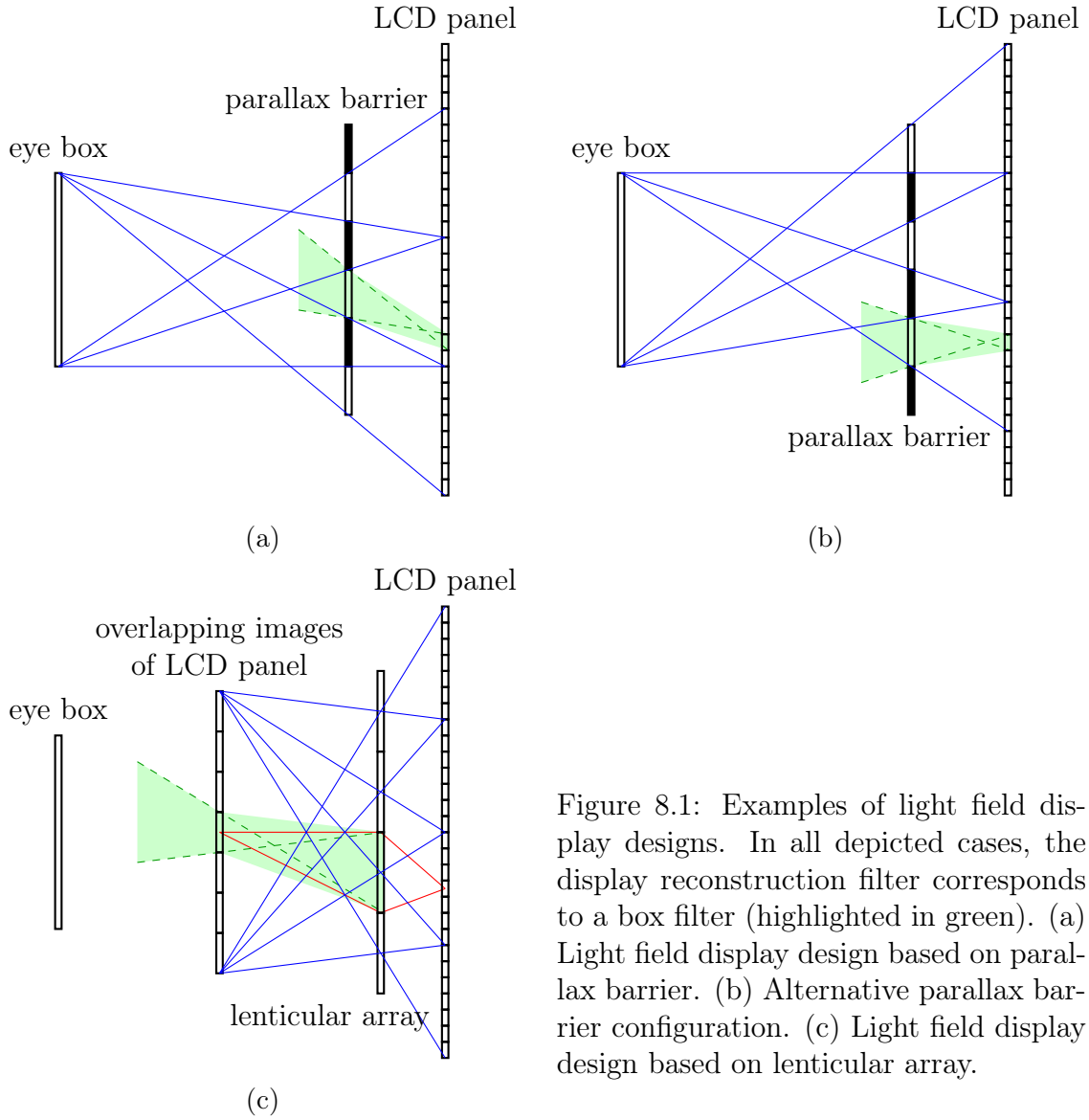


Figure 8.1: Examples of light field display designs. In all depicted cases, the display reconstruction filter corresponds to a box filter (highlighted in green). (a) Light field display design based on parallax barrier. (b) Alternative parallax barrier configuration. (c) Light field display design based on lenticular array.

In Figure 8.1, we exemplify a few possible light field display designs. Figures 8.1a and 8.1b 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. 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 can't 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 Figures 8.1a and 8.1b, can be multiplexed (through time, polarisation, etc) in order to have a better coverage of samples on the barrier plane.

Figure 8.1c 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. 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.

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

$$\varphi_d(\mathbf{x}_d) = \sqcap\left(\frac{x_a}{\mu_a}\right)\sqcap\left(\frac{y_a}{\mu_a}\right)\sqcap\left(\frac{x_b}{\mu_b}\right)\sqcap\left(\frac{y_b}{\mu_b}\right) \tag{8.1}$$

$$\text{where} \quad \sqcap(x) = \begin{cases} 0, & \text{if } |x| > \frac{1}{2} \\ \frac{1}{2}, & \text{if } |x| = \frac{1}{2} \\ 1, & \text{if } |x| < \frac{1}{2} \end{cases} \tag{8.2}$$

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).

8.4 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 Equation 8.1 is given by:

$$\psi_d(\mathbf{x}_d) = \frac{\varphi_d(\mathbf{x}_d)}{\mu_a^2\,\mu_b^2} \tag{8.3}$$

The functions ψ_d and φ_d are used over the display lattice in Equations 5.2 and 5.3 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 $[0, 1]$.

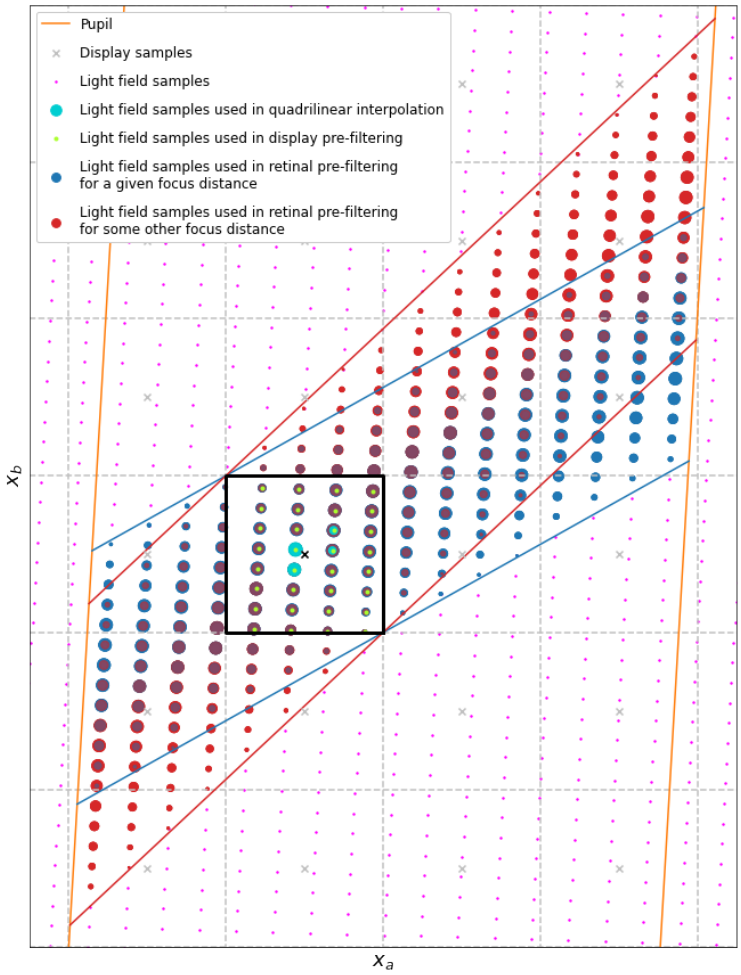


Figure 8.2: Light field samples used by different methods, drawn in display space. For each method, the size of the marks indicates the weight of the contribution of each sample.

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 Equation 6.6. The base $\boldsymbol{\varphi}_r$ is then used in Equations 6.9 and 6.10, which are part of the iterative multiplicative rule, described by Equation 6.7, 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.

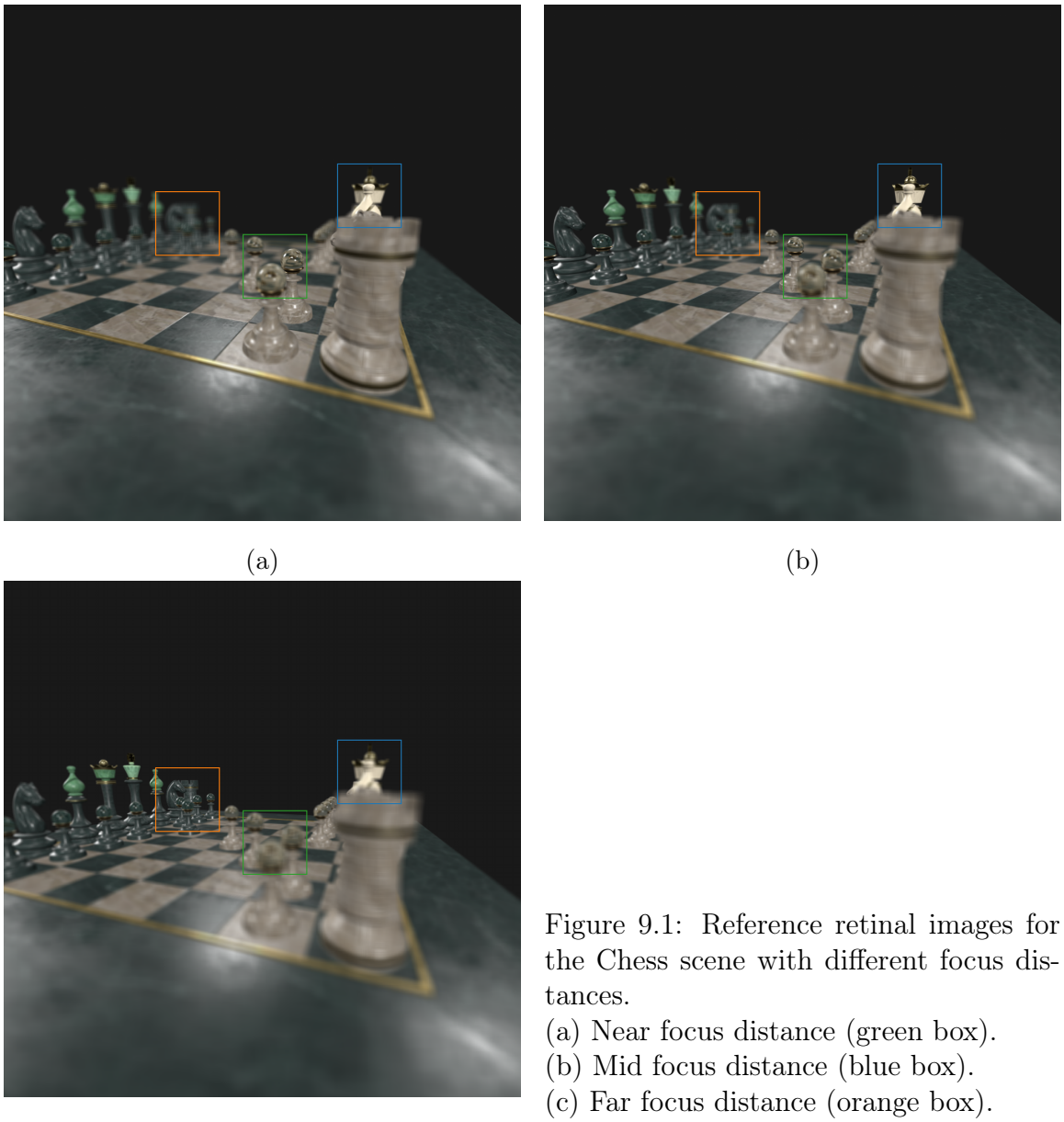
In Figure 8.2, we summarise all the pre-filtering methods we use to compute one display coefficient. The quadrilinear interpolation method (bilinear in the simplified dimensionality of the figure) uses the nearest (in sample space) 16 light field samples to the display sample (4 nearest light field samples 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.

As described in Section 7.2, we discretise F uniformly to compute the matrices of Section 7.3. In our experiment, we used 100 samples between 0 dpt and 5 dpt.

Chapter 9

Results



(a) (b)

(c)

Figure 9.1: Reference retinal images for the Chess scene with different focus distances.
(a) Near focus distance (green box).
(b) Mid focus distance (blue box).
(c) Far focus distance (orange box).

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 related inserts. Reference images for the Chess scene are shown in Figures 9.1.

We use the Chess scene for the analysis in this chapter, but similar results for the other scenes are available in Appendix B (Car), Appendix C (Dragon) and Appendix D (Sponza).

We ran our simulations on an Intel Core i5-8400 CPU @ 2.80GHz with 32GB RAM and an NVIDIA GeForce GTX 1070 GPU with 8GB 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 Figures 9.2.

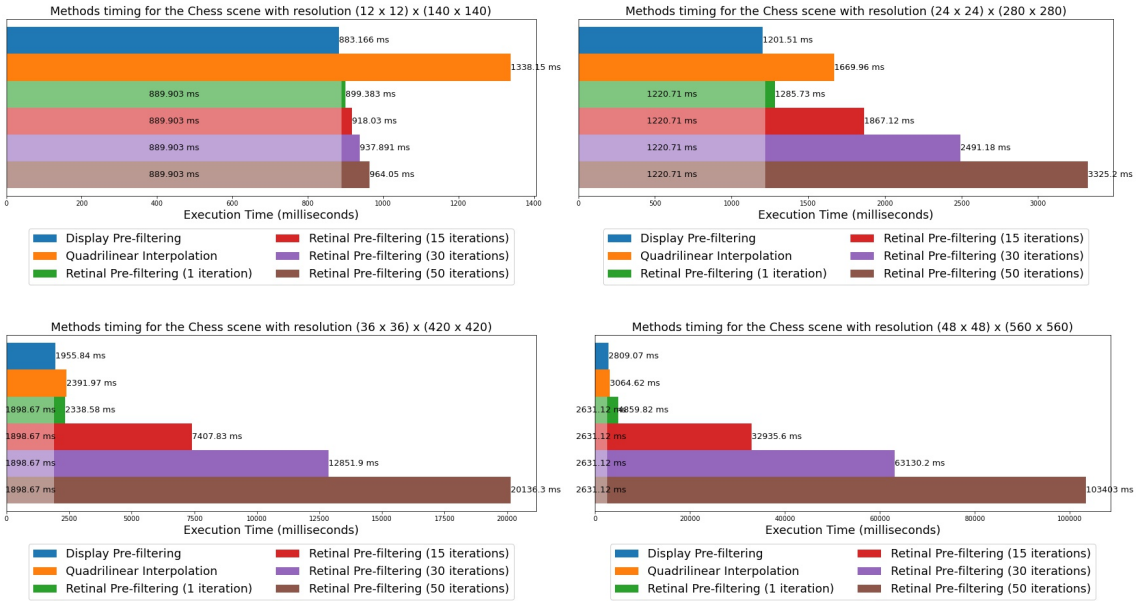


Figure 9.2: Methods timing for the Chess scene with different display resolutions.

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 Equation 7.12 only once. For every iteration, we compute 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.

Figures 9.12, 9.13, 9.14, 9.15, 9.16 and 9.17 show 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.

Figure 9.4 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.

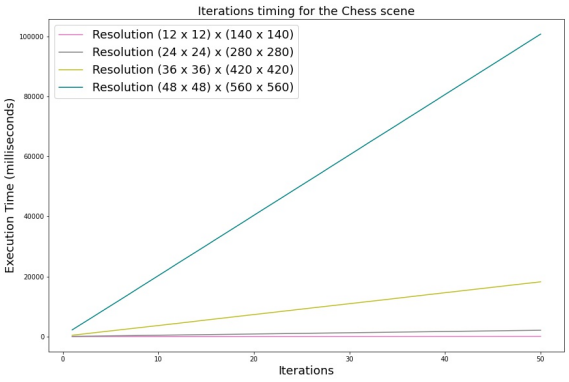


Figure 9.3: Iterations timing for the Chess scene with different display resolutions.

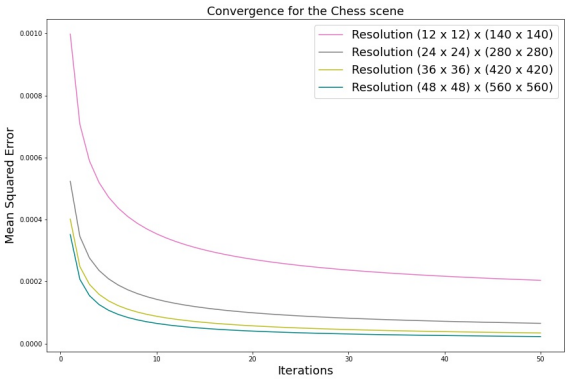
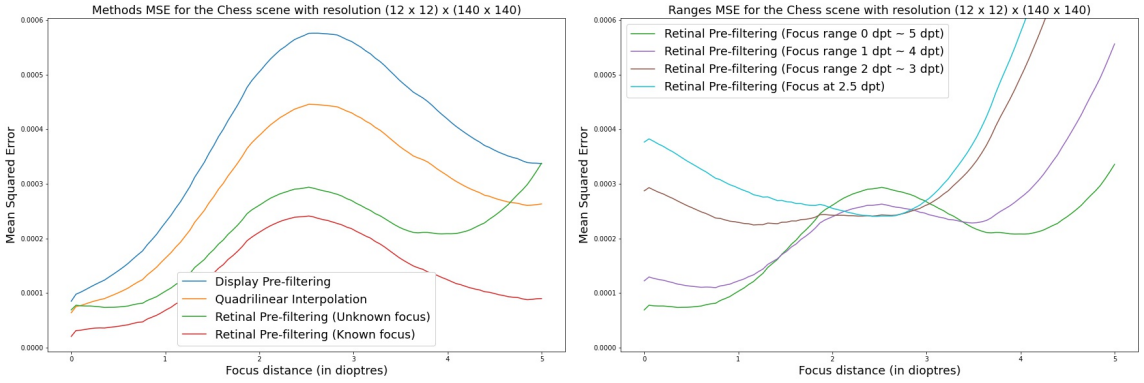


Figure 9.4: Convergence of the retinal pre-filtering for the Chess scene with different display resolutions.

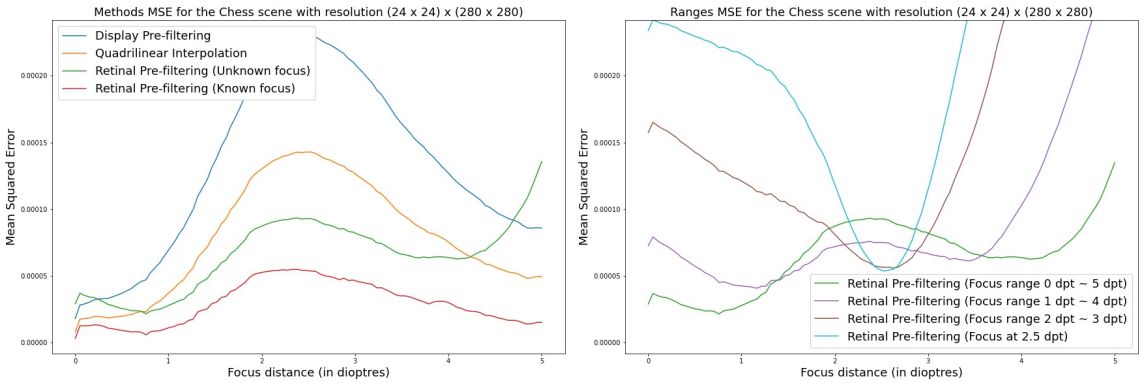
Figures 9.5, 9.6, 9.7 and 9.8 show, for each display resolution, the retinal mean squared error as the observer's focus distance varies between 0 dpt and 5 dpt (i.e.

from an infinite distance down to 20 cm in front of the pupil). The retinal pre-filtering methods in Figures 9.5b, 9.6b, 9.7b and 9.8b 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.



(a) (b)

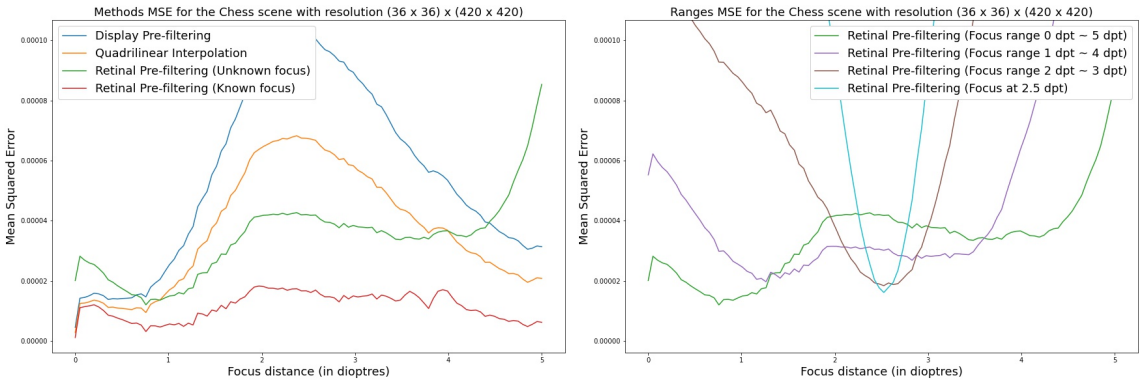
Figure 9.5: Comparison of pre-filtering methods (a) and retinal pre-filtering with different focus ranges (b) for the Chess scene with display resolution $(12 \times 12) \times (140 \times 140)$.



(a) (b)

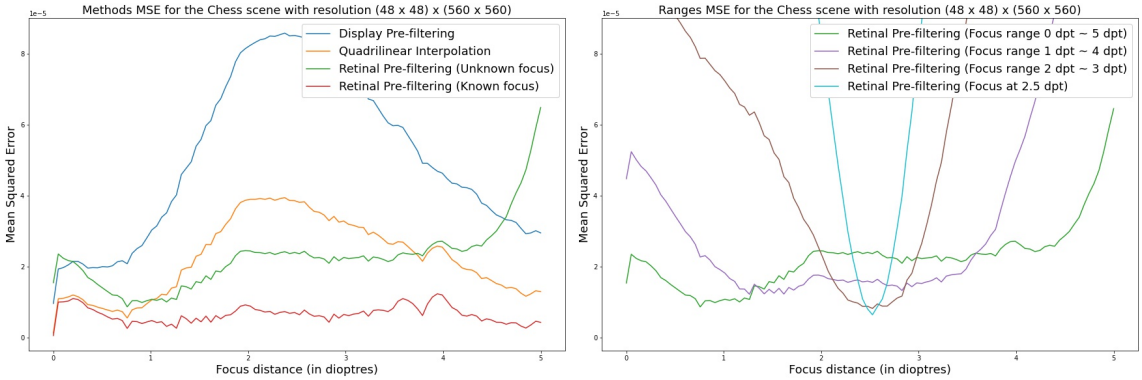
Figure 9.6: Comparison of pre-filtering methods (a) and retinal pre-filtering with different focus ranges (b) for the Chess scene with display resolution $(24 \times 24) \times (280 \times 280)$.

We then consider two scenarios for the use of retinal pre-filtering. One for when the focus distance is unknown, where 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, where the retinal pre-filtering is dynamically applied, targeting a single distance that follows the current focus distance.



(a) (b)

Figure 9.7: Comparison of pre-filtering methods (a) and retinal pre-filtering with different focus ranges (b) for the Chess scene with display resolution $(36 \times 36) \times (420 \times 420)$.



(a) (b)

Figure 9.8: Comparison of pre-filtering methods (a) and retinal pre-filtering with different focus ranges (b) for the Chess scene with display resolution $(48 \times 48) \times (560 \times 560)$.

Figures 9.5a, 9.6a, 9.7a and 9.8a compare the retinal mean squared error of the different pre-filtering methods for different display resolutions. 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 8.4. 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 optimization 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.

Figures 9.9, 9.10 and 9.11 show three inserts for the Chess scene, each showcasing a different focus distance and comparing display resolutions and pre-filtering methods.

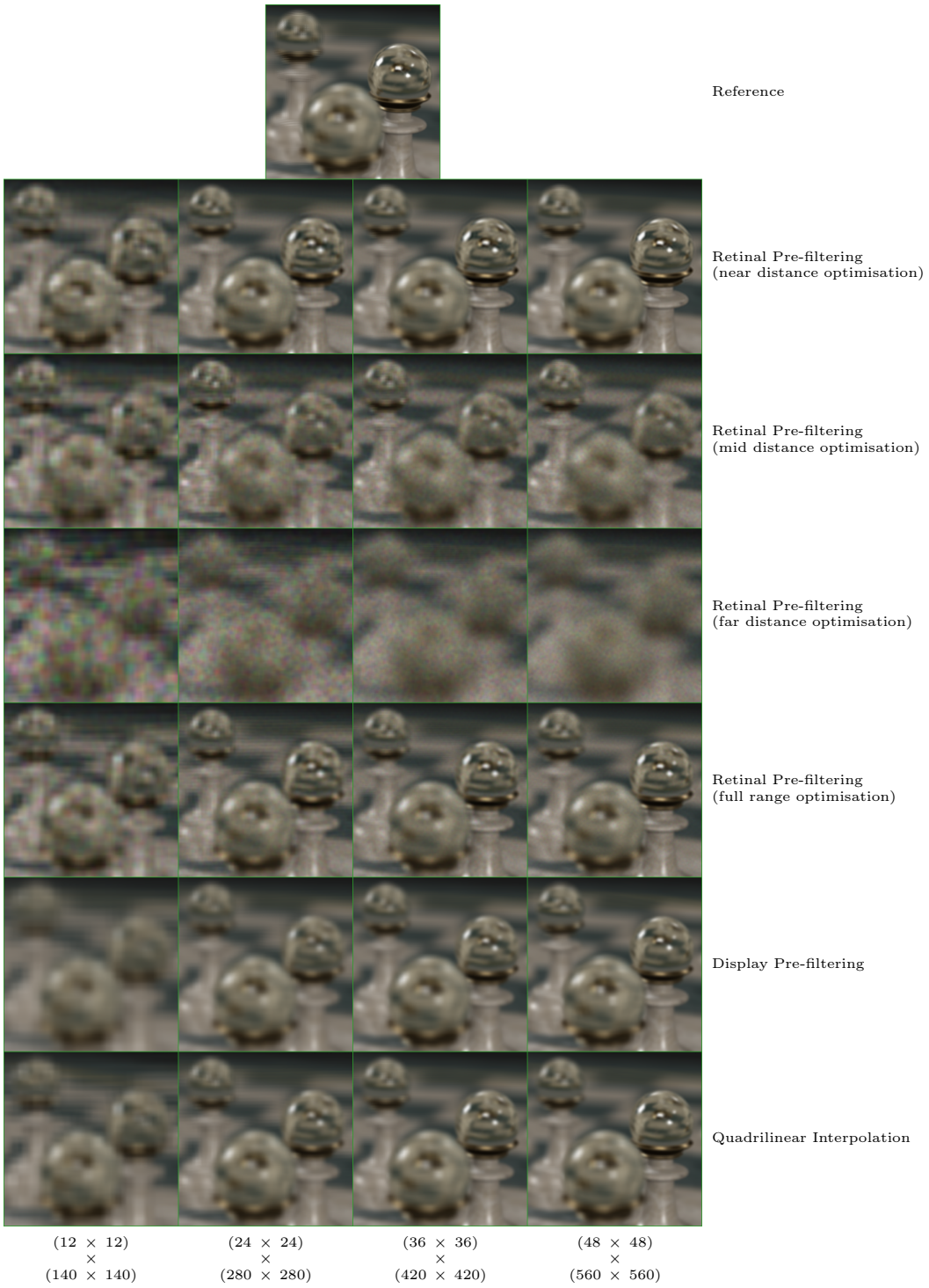


Figure 9.9: Comparison of pre-filtering methods for the near insert of the Chess scene with different display resolutions.

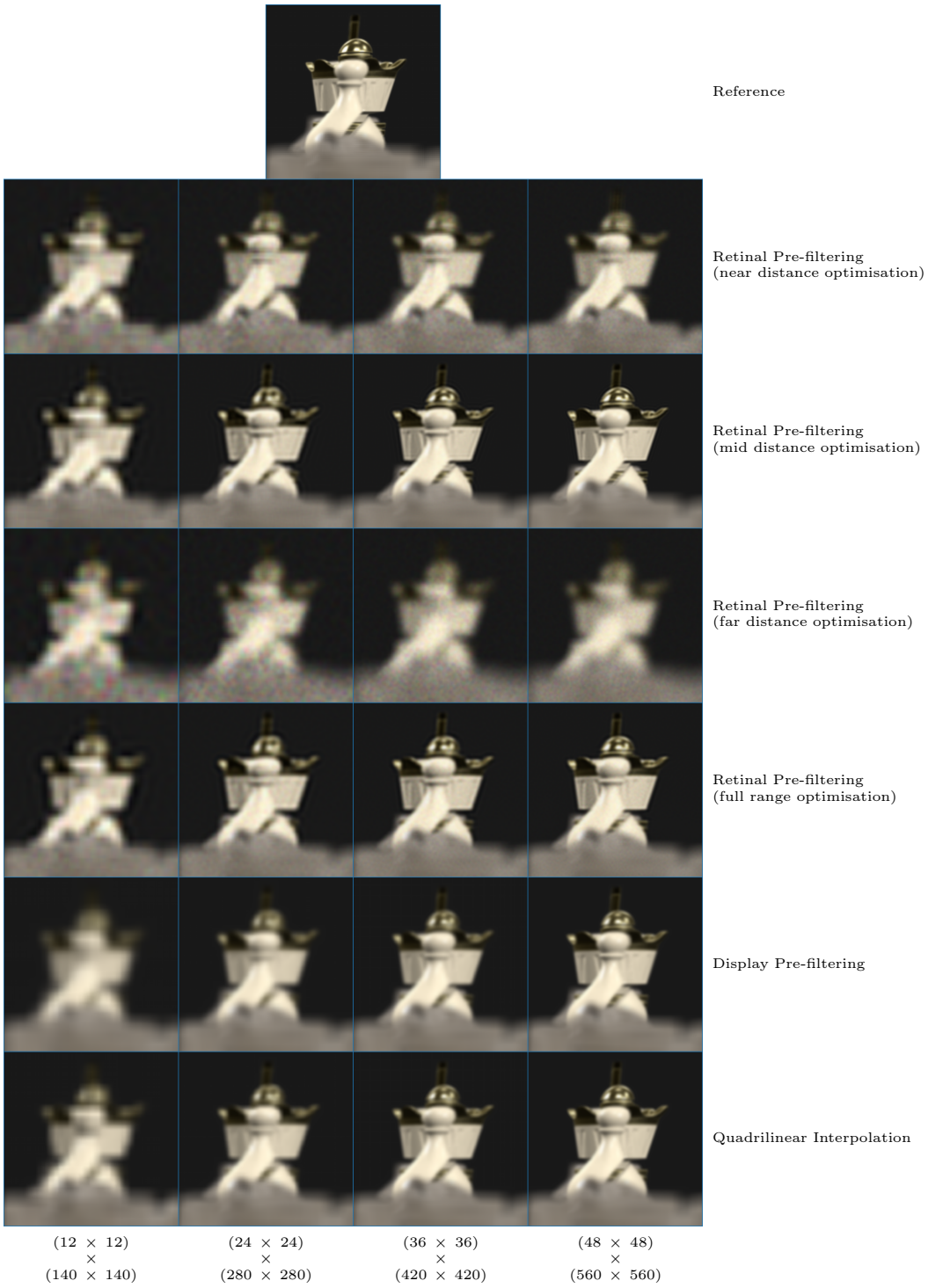


Figure 9.10: Comparison of pre-filtering methods for the mid insert of the Chess scene with different display resolutions.

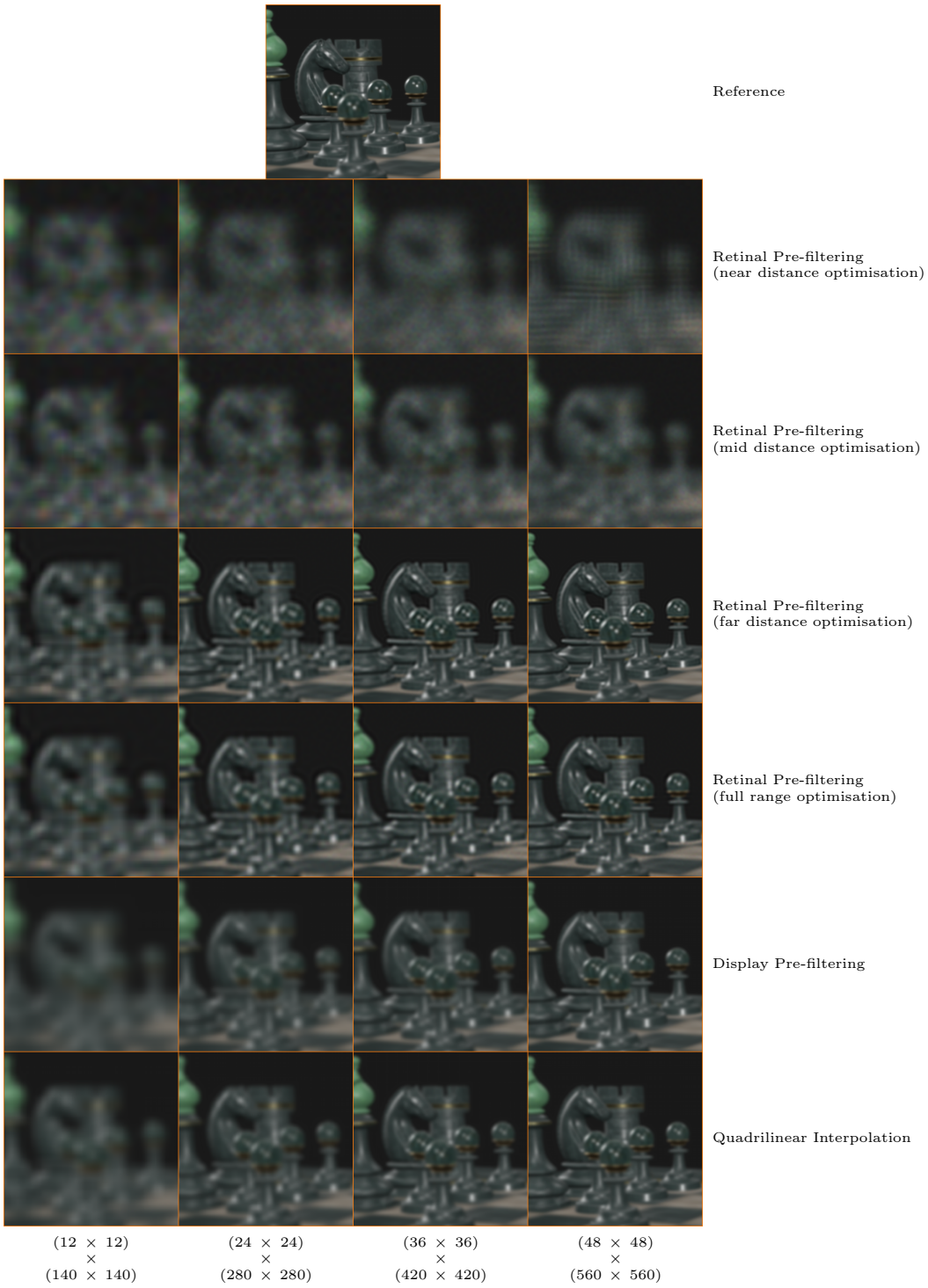


Figure 9.11: Comparison of pre-filtering methods for the far insert of the Chess scene with different display resolutions.

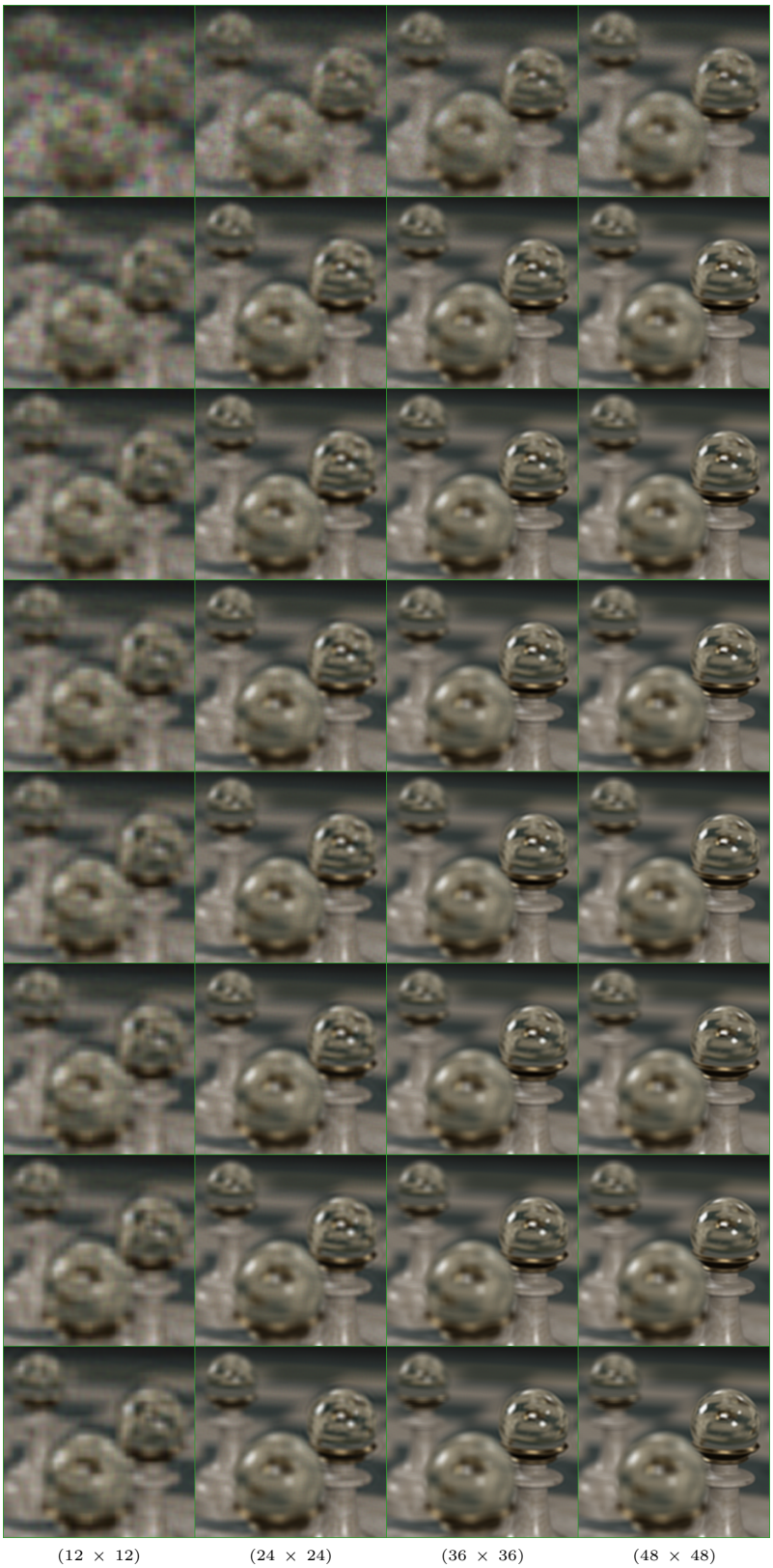


Figure 9.12: Incremental improvements of retinal pre-filtering (near distance) across multiple iterations for the near insert of the Chess scene with different display resolutions.

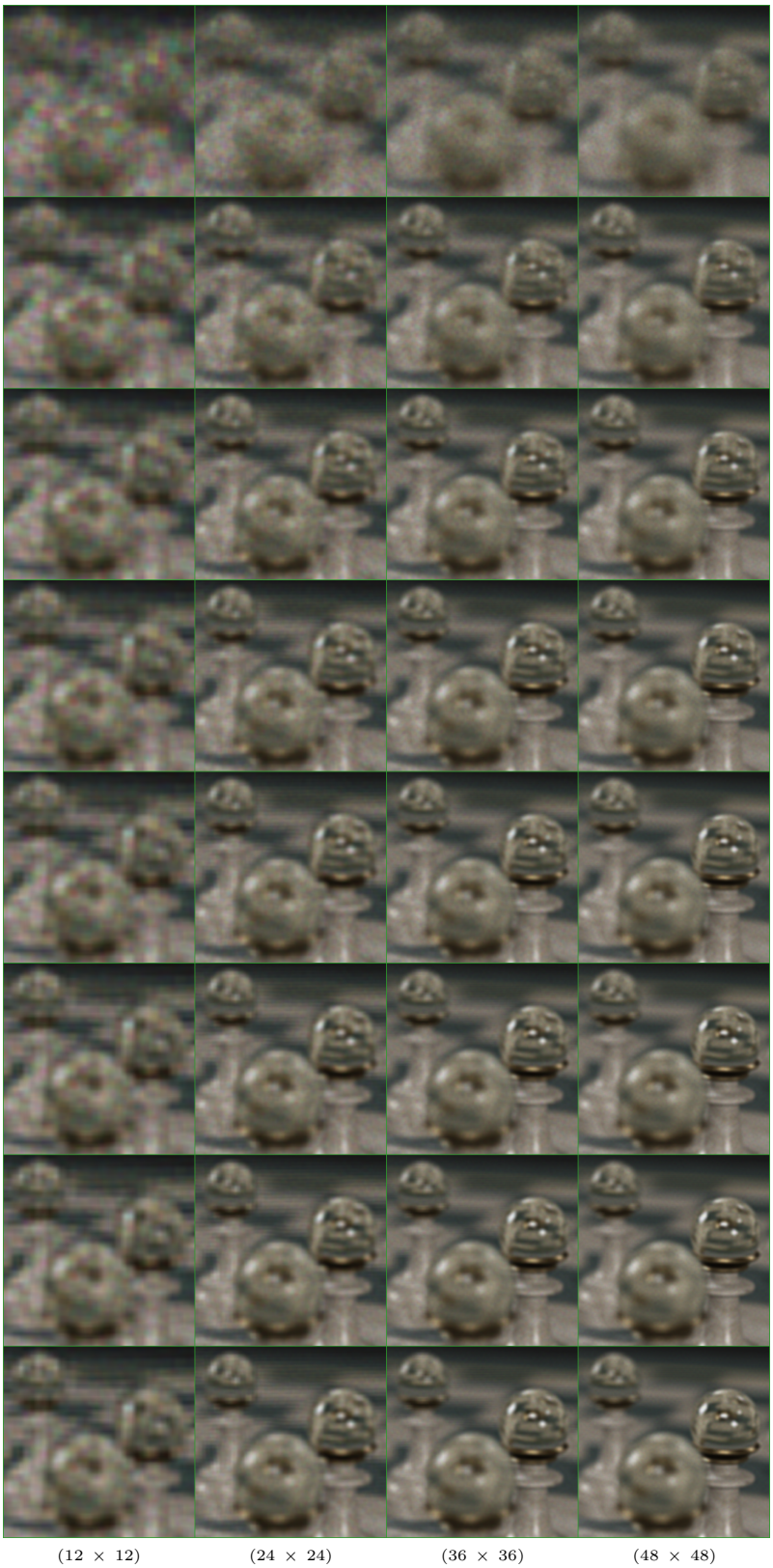


Figure 9.13: Incremental improvements of retinal pre-filtering (full range) across multiple iterations for the near insert of the Chess scene with different display resolutions.

Figure 9.14: Incremental improvements of retinal pre-filtering (mid distance) across multiple iterations for the mid insert of the Chess scene with different display resolutions.

Figure 9.15: Incremental improvements of retinal pre-filtering (full range) across multiple iterations for the mid insert of the Chess scene with different display resolutions.

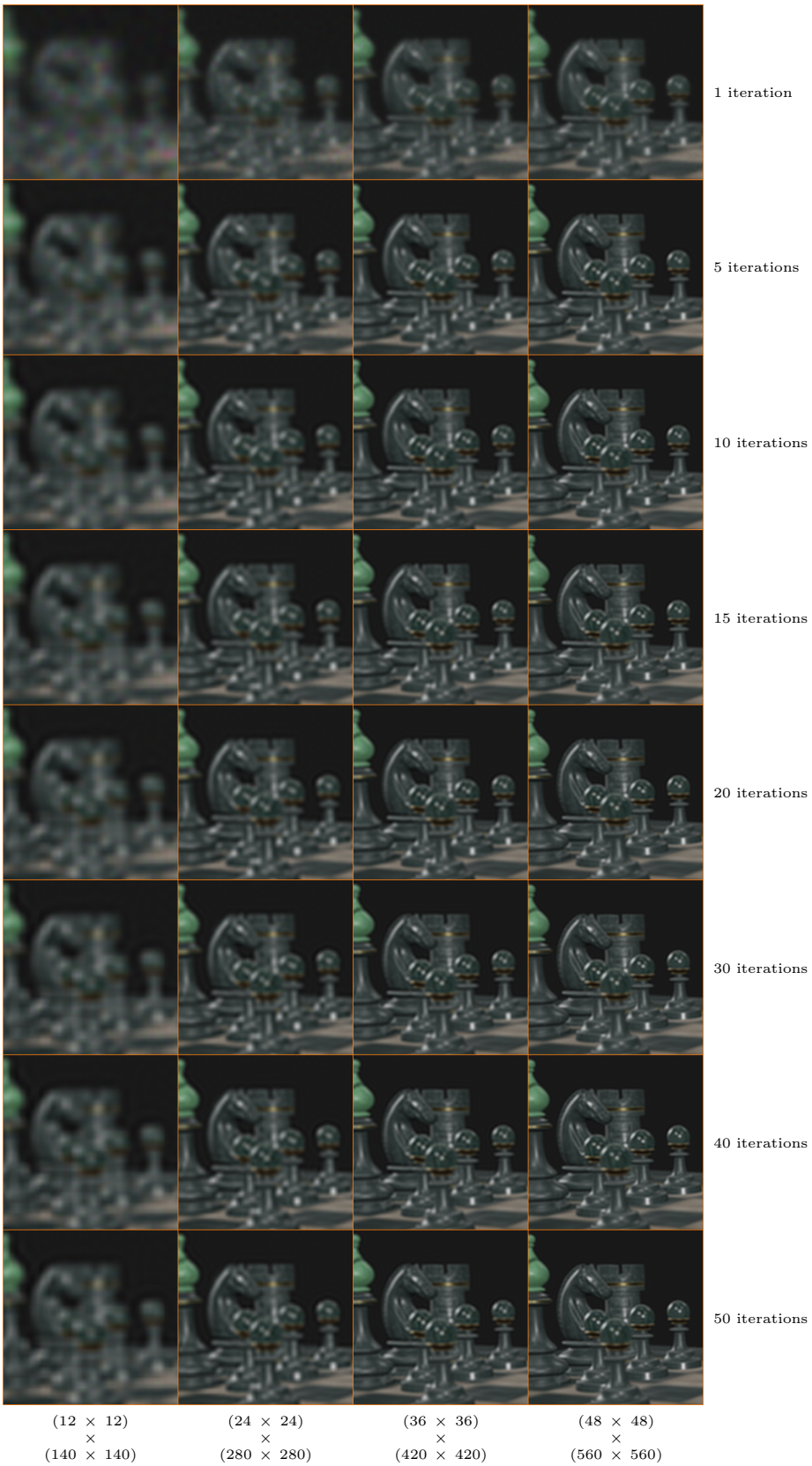


Figure 9.16: Incremental improvements of retinal pre-filtering (far distance) across multiple iterations for the far insert of the Chess scene with different display resolutions.

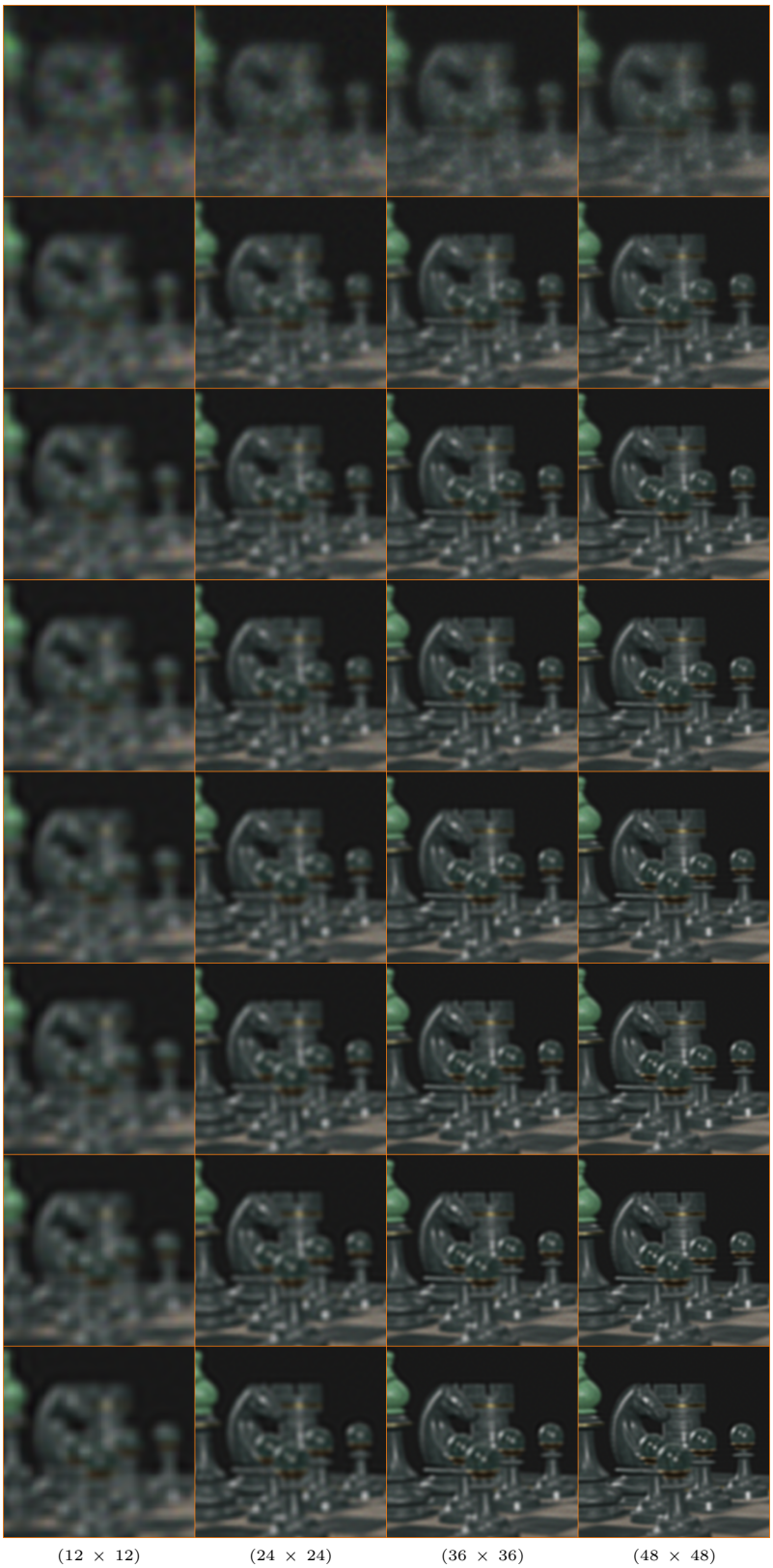


Figure 9.17: Incremental improvements of retinal pre-filtering (full range) across multiple iterations for the far insert of the Chess scene with different display resolutions.

Chapter 10

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 in 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 (Equation 5.3), while the linearity of the observer model comes from its integral transform over the incoming light field (Equation 6.1).

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.

10.1 Display limitations

Display architectures that employ lenticular arrays [8, 9] or parallax barriers [23] 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 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 is not covered in this work.

In the end, the only information needed from the display is the reconstruction function $\boldsymbol{\varphi}_d[i_d]$ associated with 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 subpixel 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 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.

10.2 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 work when our method optimises for different ranges of focus distances (Figures 9.5b, 9.6b, 9.7b and 9.8b).

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.

Chapter 11

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 (Chapter 5) or the retinal image error (Chapter 6) by leveraging information about the display and the observer. We further describe a detailed matrix formulation to allow for a practical implementation (Chapter 7).

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.

On the one hand, our work achieves the optimal result for the defined error metric. On the other hand, it should not be seen as a definitive solution but as a possible building block for many future strategies.

One important aspect is that this work can contribute to delivering better light field displays without necessarily increasing the physical resolution. This might be specifically important for near-eye displays where, given the small form factor of the display, increasing the resolution can quickly run into issues due to diffraction.

11.1 Future work

There are many possible directions for future work. Some were already pointed out in Chapter 10, but we go over them again here for the sake of completeness.

First, time performance was not our primary goal in this work; hence, our implementation was not fully optimised in this regard. We also considered many aspects to make the simulation and tests viable that might not be necessary for a concrete implementation for a given light field display. In this case, we believe many software and hardware optimisations can be employed to achieve interactive frame rates.

Given that all our results come from simulations, a natural next step would be to test our work with a physical display. This would also allow running user studies to confirm how much our theoretical contributions match human perception.

In this work, we adopted the thin lens model which is widely employed in computer graphics methods. This is true even when it comes to the specific subject of multifocal displays with an observer model, such is the case of Narain et al. [11], Mercier et al. [12] and Ebner et al. [13]. Nevertheless, a more sophisticated eye model that closely follows human physiology may lead to better results.

As was previously stated in Section 10.2, we can also increase the objective function dimensionality to account for many other variations, such as eye movement and pupil contraction. Again, note that this does not increase the dimensionalities of the display's input (coefficients) and output (light field).

Finally, it would be interesting to generalise our method to handle other display architectures where the signal is not reconstructed from a linear combination of reconstruction kernels. This is the case, for example, of tensor displays [24].

References

[1] LEVOY, M., HANRAHAN, P. "Light Field Rendering". In: *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, p. 31–42, New York, NY, USA, 1996. Association for Computing Machinery. ISBN: 0897917464. doi: 10.1145/237170.237199. Disponível em: <https://doi.org/10.1145/237170.237199>.

[2] GORTLER, S. J., GRZESZCZUK, R., SZELISKI, R., et al. "The Lumigraph". In: *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, p. 43–54, New York, NY, USA, 1996. Association for Computing Machinery. ISBN: 0897917464. doi: 10.1145/237170.237200. Disponível em: <https://doi.org/10.1145/237170.237200>.

[3] ISAKSEN, A., MCMILLAN, L., GORTLER, S. J. "Dynamically Reparameterized Light Fields". In: *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, p. 297–306, USA, 2000. ACM Press/Addison-Wesley Publishing Co. ISBN: 1581132085. doi: 10.1145/344779.344929. Disponível em: <https://doi.org/10.1145/344779.344929>.

[4] CHAI, J.-X., TONG, X., CHAN, S.-C., et al. "Plenoptic Sampling". In: *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, p. 307–318, USA, 2000. ACM Press/Addison-Wesley Publishing Co. ISBN: 1581132085. doi: 10.1145/344779.344932. Disponível em: <https://doi.org/10.1145/344779.344932>.

[5] ZHANG, C., CHEN, T. *Generalized plenoptic sampling.* Relatório Técnico AMP01-06, Carnegie Mellon University, September 2001. Disponível em: <https://www.microsoft.com/en-us/research/publication/generalized-plenoptic-sampling/>.

[6] STEWART, J., YU, J., GORTLER, S. J., et al. "A New Reconstruction Filter for Undersampled Light Fields". In: *Proceedings of the 14th Eurograph-*

ics Workshop on Rendering, EGRW '03, p. 150–156, Goslar, DEU, 2003. Eurographics Association. ISBN: 3905673037.

[7] LIU, C.-L., SHIH, K.-T., HUANG, J.-W., et al. "Light Field Synthesis by Training Deep Network in the Refocused Image Domain", *IEEE Transactions on Image Processing*, v. 29, pp. 6630–6640, 2020. doi: 10.1109/TIP.2020.2992354.

[8] LIPPMANN, G. "Épreuves réversibles donnant la sensation du relief", *J. Phys. Theor. Appl.*, v. 7, n. 1, pp. 821–825, 1908. doi: 10.1051/jphystap:019080070082100. Disponível em: <https://hal.science/jpa-00241406>.

[9] LANMAN, D., LUEBKE, D. "Near-Eye Light Field Displays", *ACM Trans. Graph.*, v. 32, n. 6, nov 2013. ISSN: 0730-0301. doi: 10.1145/2508363.2508366. Disponível em: <https://doi.org/10.1145/2508363.2508366>.

[10] WETZSTEIN, G., LANMAN, D., HEIDRICH, W., et al. "Layered 3D: Tomographic Image Synthesis for Attenuation-Based Light Field and High Dynamic Range Displays", *ACM Trans. Graph.*, v. 30, n. 4, jul 2011. ISSN: 0730-0301. doi: 10.1145/2010324.1964990. Disponível em: <https://doi.org/10.1145/2010324.1964990>.

[11] NARAIN, R., ALBERT, R. A., BULBUL, A., et al. "Optimal Presentation of Imagery with Focus Cues on Multi-Plane Displays", *ACM Trans. Graph.*, v. 34, n. 4, jul 2015. ISSN: 0730-0301. doi: 10.1145/2766909. Disponível em: <https://doi.org/10.1145/2766909>.

[12] MERCIER, O., SULAI, Y., MACKENZIE, K., et al. "Fast Gaze-Contingent Optimal Decompositions for Multifocal Displays", *ACM Trans. Graph.*, v. 36, n. 6, nov 2017. ISSN: 0730-0301. doi: 10.1145/3130800.3130846. Disponível em: <https://doi.org/10.1145/3130800.3130846>.

[13] EBNER, C., MORI, S., MOHR, P., et al. "Video See-Through Mixed Reality with Focus Cues", *IEEE Transactions on Visualization and Computer Graphics*, v. 28, n. 5, pp. 2256–2266, 2022. doi: 10.1109/TVCG.2022.3150504.

[14] LOVE, G. D., HOFFMAN, D. M., HANDS, P. J., et al. "High-speed switchable lens enables the development of a volumetric stereoscopic display", *Opt. Express*, v. 17, n. 18, pp. 15716–15725, Aug 2009. doi: 10.1364/OE.17.

015716. Disponível em: <https://opg.optica.org/oe/abstract.cfm?URI=oe-17-18-15716>.

[15] SHANNON, C. "Communication in the Presence of Noise", *Proceedings of the IRE*, v. 37, n. 1, pp. 10–21, 1949. doi: 10.1109/JRPROC.1949.232969.

[16] UNSER, M. "Sampling-50 years after Shannon", *Proceedings of the IEEE*, v. 88, n. 4, pp. 569–587, 2000. doi: 10.1109/5.843002.

[17] NEHAB, D., HOPPE, H. "A Fresh Look at Generalized Sampling", *Foundations and Trends® in Computer Graphics and Vision*, v. 8, n. 1, pp. 1–84, 2014. ISSN: 1572-2740. doi: 10.1561/0600000053. Disponível em: <http://dx.doi.org/10.1561/0600000053>.

[18] ZWICKER, M., MATUSIK, W., DURAND, F., et al. "Antialiasing for Automultiscopic 3D Displays". In: Akenine-Moeller, T., Heidrich, W. (Eds.), *Symposium on Rendering*. The Eurographics Association, 2006. ISBN: 3-905673-35-5. doi: 10.2312/EGWR/EGSR06/073-082.

[19] KAJIYA, J., ULLNER, M. "Filtering high quality text for display on raster scan devices", *SIGGRAPH Comput. Graph.*, v. 15, n. 3, pp. 7–15, ago. 1981. ISSN: 0097-8930. doi: 10.1145/965161.806784. Disponível em: <https://doi.org/10.1145/965161.806784>.

[20] ROCHA, L., OLIVEIRA, M., GASTAL, E. "Prefilters for Sharp Image Display", *Computer Graphics Forum*, v. 39, pp. 435–449, 07 2020. doi: 10.1111/cgf.13942.

[21] PHARR, M., JAKOB, W., HUMPHREYS, G. *Physically Based Rendering: From Theory to Implementation*. 3rd ed. San Francisco, CA, USA, Morgan Kaufmann Publishers Inc., 2016. ISBN: 0128006455.

[22] LEE, D., SEUNG, H. "Algorithms for Non-negative Matrix Factorization", *Adv. Neural Inform. Process. Syst.*, v. 13, 02 2001.

[23] PERLIN, K., PAXIA, S., KOLLIN, J. S. "An autostereoscopic display". In: *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, p. 319–326, USA, 2000. ACM Press/Addison-Wesley Publishing Co. ISBN: 1581132085. doi: 10.1145/344779.344933. Disponível em: <https://doi.org/10.1145/344779.344933>.

[24] WETZSTEIN, G., LANMAN, D., HIRSCH, M., et al. "Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional

Backlighting", *ACM Trans. Graph. (Proc. SIGGRAPH)*, v. 31, n. 4, pp. 1–11, 2012.

[25] BERBERIAN, S. *Introduction to Hilbert Space*. AMS Chelsea Publishing Series. USA, AMS Chelsea, 1999. ISBN: 9780821819128.

Appendix A

Linear Spaces

The focal stack, denoted by G in Chapter 6, is the function comprised of all images formed on the observer's retina as the focus distance varies. From Equation 6.4, G is the result of an integral transform, henceforth denoted by $\mathcal{T}$, applied to the light field function:

$$G(\zeta_f, \mathbf{x}_r) = \iint_P L_d\left(T_{de}(\zeta_f)\begin{bmatrix}\mathbf{x}_r\\ \mathbf{x}_p\end{bmatrix}\right) \mathrm{d}\mathbf{x}_p = (\mathcal{T}L_d)(\zeta_f, \mathbf{x}_r) \tag{A.1}$$

It follows directly from the linear properties of integrals, that $\mathcal{T}(\alpha L_1 + \beta L_2) = \alpha\,(\mathcal{T}L_1) + \beta\,(\mathcal{T}L_2)$, $\forall\alpha, \beta \in \mathbb{R}$. Therefore, the set of all possible functions G, i.e. the space of focal stacks, is a vector space over the field $\mathbb{R}$. Here we are not making any non-negativity constraint and a negative (unfeasible) focal stack would be the result of a negative (unfeasible) light field.

The function $G(\zeta_f, \mathbf{x}_r)$, with $\zeta_f \in F$ and $\mathbf{x}_r \in R$, is a real continuous 3D function defined on the 1D allowed focus range F and the 2D retina region R. Therefore, we can define the inner product of the vector space of focal stacks as:

$$\langle G_1, G_2\rangle = \int_F \iint_R G_1(\zeta_f, \mathbf{x}_r)\, G_2(\zeta_f, \mathbf{x}_r)\, \mathrm{d}\mathbf{x}_r\, \mathrm{d}\zeta_f \tag{A.2}$$

This is the standard way of defining inner product for continuous functions in functional analysis as it is the direct extension to infinite dimension of the discrete inner product for finite dimensional vectors. In other words, this is a regular inner product and has no special features particular to our case.

From the inner product, we also have a norm, given by $\|G\| = \sqrt{\langle G, G\rangle}$, which in turn induces a distance function $dist\,(G_1, G_2) = \|G_1 - G_2\|$. This distance function yields a complete metric space and, therefore, the space of focal stacks is not only a vector space but it is, in fact, a Hilbert space. For a more comprehensive description of Hilbert spaces and their properties, we refer the reader to Berberian [25].

In Equation 6.6, we defined the retinal basis function $\boldsymbol{\varphi}_r[i_d]$ as the focal stack produced from the light field comprised solely from one single display basis function $\boldsymbol{\varphi}_d[i_d]$:

$$\boldsymbol{\varphi}_r[i_d](\zeta_f, \mathbf{x}_r) = \iint_P \boldsymbol{\varphi}_d[i_d] \left(T_{de}(\zeta_f) \begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_p \end{bmatrix} \right) \mathrm{d}\mathbf{x}_p = (\mathcal{T}\boldsymbol{\varphi}_d[i_d])(\zeta_f, \mathbf{x}_r) \tag{A.3}$$

We also defined, in Equation 6.5, a linear operator Φ that encapsulates the entire process from the discrete set of display coefficients $\mathbf{c}$ to the resulting continuous 3D focal stack $\tilde{G}$ formed on the observer's retina:

$$\tilde{G}(\zeta_f, \mathbf{x}_r) = \Phi\mathbf{c} = \sum_{i_d=1}^{n_d} \mathbf{c}[i_d]\, \boldsymbol{\varphi}_r[i_d](\zeta_f, \mathbf{x}_r) \tag{A.4}$$

Every basis function $\boldsymbol{\varphi}_r$ belongs to the vector space of focal stacks. Every function $\tilde{G}$ is a linear combination of functions $\boldsymbol{\varphi}_r$ and, consequently, every function $\tilde{G}$ also belongs to the vector space of focal stacks. A trivial consequence of the linear combination embedded in Φ is that $\Phi\left(\alpha\mathbf{c}_1 + \beta\mathbf{c}_2\right) = \alpha\left(\Phi\mathbf{c}_1\right) + \beta\left(\Phi\mathbf{c}_2\right)$, $\forall\alpha, \beta \in \mathbb{R}$. Therefore, the set of all possible functions $\tilde{G}$ (which is the column space of Φ) is also a vector space over the field $\mathbb{R}$ and, more specifically, it is a linear subspace of the space of focal stacks.

Our goal is to find coefficients $\mathbf{c}$ that produce the function $\tilde{G}$ that is closest to a desired function G. The term "closest" here pertains to the distance function already discussed. The least squares problem then is based on minimising $dist^2\left(G, \tilde{G}\right)$.

$$\begin{aligned} dist^2\left(G, \tilde{G}\right) = \left\|G - \tilde{G}\right\|^2 = \langle G - \tilde{G}, G - \tilde{G}\rangle = \left(G - \tilde{G}\right)^\intercal \left(G - \tilde{G}\right) = \\ G^\intercal G - G^\intercal\tilde{G} - \tilde{G}^\intercal G + \tilde{G}^\intercal\tilde{G} = G^\intercal G - 2\tilde{G}^\intercal G + \tilde{G}^\intercal\tilde{G} \end{aligned} \tag{A.5}$$

where $G^\intercal\tilde{G} = \tilde{G}^\intercal G$ since it is a scalar.

With $\tilde{G} = \Phi\mathbf{c}$, it follows that:

$$dist^2(G, \Phi\mathbf{c}) = G^\intercal G - 2\mathbf{c}^\intercal\Phi^\intercal G + \mathbf{c}^\intercal\Phi^\intercal\Phi\mathbf{c} \tag{A.6}$$

Taking the derivative with respect to $\mathbf{c}$ and equating it to zero, we have:

$$\begin{aligned} \frac{\partial dist^2(G, \Phi\mathbf{c})}{\partial \mathbf{c}} &= 0 \\ -2\Phi^\intercal G + 2\Phi^\intercal \Phi\mathbf{c} &= 0 \\ \Phi^\intercal \Phi\mathbf{c} &= \Phi^\intercal G \end{aligned} \tag{A.7}$$

Alternatively, the closest $\tilde{G}$ is also given by the orthogonal projection of the desired G into the linear subspace of possible functions $\tilde{G}$. This orthogonal projection is characterised by the inner product of each function $\boldsymbol{\varphi}_r$ and the residual $G - \tilde{G}$ being equal to zero:

$$\langle \boldsymbol{\varphi}_r[i_d], G - \tilde{G} \rangle = 0 \quad \forall \quad i_d \in [1, n_d] \tag{A.8}$$

which can be equivalently written in matrix notation as:

$$\Phi^\intercal \left(G - \tilde{G} \right) = 0 \tag{A.9}$$

With $\tilde{G} = \Phi\mathbf{c}$, it follows that:

$$\begin{aligned} \Phi^\intercal (G - \Phi\mathbf{c}) &= 0 \\ \Phi^\intercal G - \Phi^\intercal \Phi\mathbf{c} &= 0 \\ \Phi^\intercal \Phi\mathbf{c} &= \Phi^\intercal G \end{aligned} \tag{A.10}$$

Either way we arrive at $\Phi^\intercal \Phi\mathbf{c} = \Phi^\intercal G$, which is called the normal equation. The conventional least squares solution to $\Phi\mathbf{c} \approx G$ is then given by $\mathbf{c} = (\Phi^\intercal \Phi)^{-1} \Phi^\intercal G$, with $(\Phi^\intercal \Phi)^{-1} \Phi^\intercal$ being the Moore–Penrose inverse of Φ and $\Phi(\Phi^\intercal \Phi)^{-1} \Phi^\intercal$ being the orthogonal projector from G to $\tilde{G}$.

Note that, in our case, we don't require the coefficients $\mathbf{c}$ to be unique for any given G. Therefore, we don't require $\Phi^\intercal \Phi$ to be invertible or the functions $\boldsymbol{\varphi}_r$ to be linearly independent. We also don't require the functions $\boldsymbol{\varphi}_r$ to be orthogonal or to have unitary norm. If the functions $\boldsymbol{\varphi}_r$ formed an orthonormal basis, than the solution would be greatly simplified to $\mathbf{c} = \Phi^\intercal G$ since $(\Phi^\intercal \Phi)^{-1} = \Phi^\intercal \Phi = I$ due to the orthonormality definition:

$$\langle \boldsymbol{\varphi}_r[i_d], \boldsymbol{\varphi}_r[j_d] \rangle = \boldsymbol{\delta}[i_d - j_d] = \begin{cases} 0, & \text{if } i_d \neq j_d \\ 1, & \text{if } i_d = j_d \end{cases} \tag{A.11}$$

In any case, we still have to take into account the non-negativity constraint of $\mathbf{c}$. For this reason, instead of following the conventional least squares solution, we

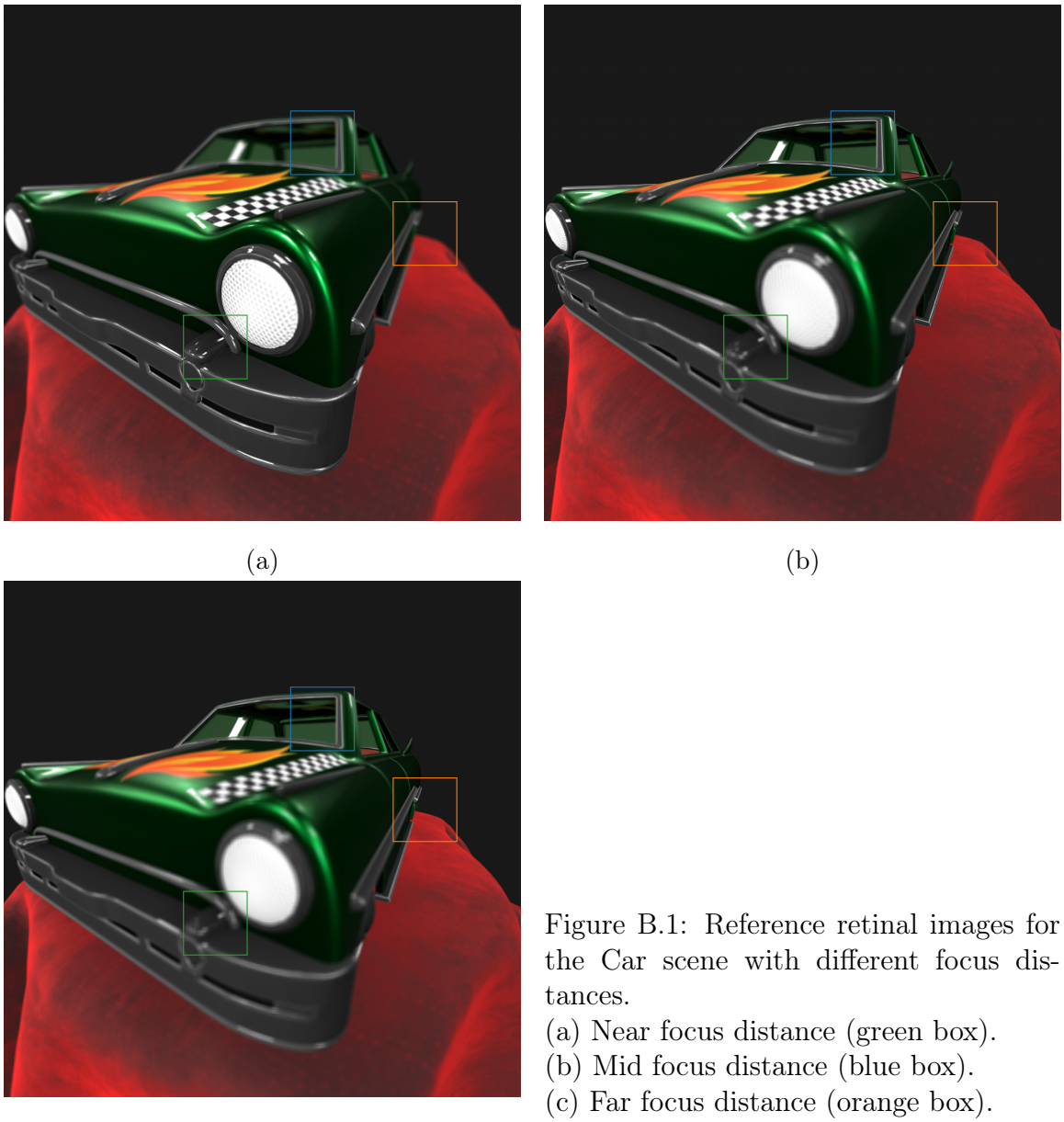
employ an iterative method as described in the primary text (Equation 6.7):

$$\mathbf{c} \leftarrow \mathbf{c} \otimes \frac{\Phi^{\intercal} G}{\Phi^{\intercal} \Phi \, \mathbf{c}} \tag{A.12}$$

Note that the normal equation $\Phi^{\intercal}\Phi\mathbf{c} = \Phi^{\intercal}G$ corresponds to a fixed point of the iterative method. This iterative method minimises the least squares error while keeping $\mathbf{c}$ always non-negative and is guaranteed to converge monotonically. We refer the reader to Lee and Seung [22] where the related proofs are provided.

Appendix B

Results - Car Scene



(a)

(b)

(c)

Figure B.1: Reference retinal images for the Car scene with different focus distances.
(a) Near focus distance (green box).
(b) Mid focus distance (blue box).
(c) Far focus distance (orange box).

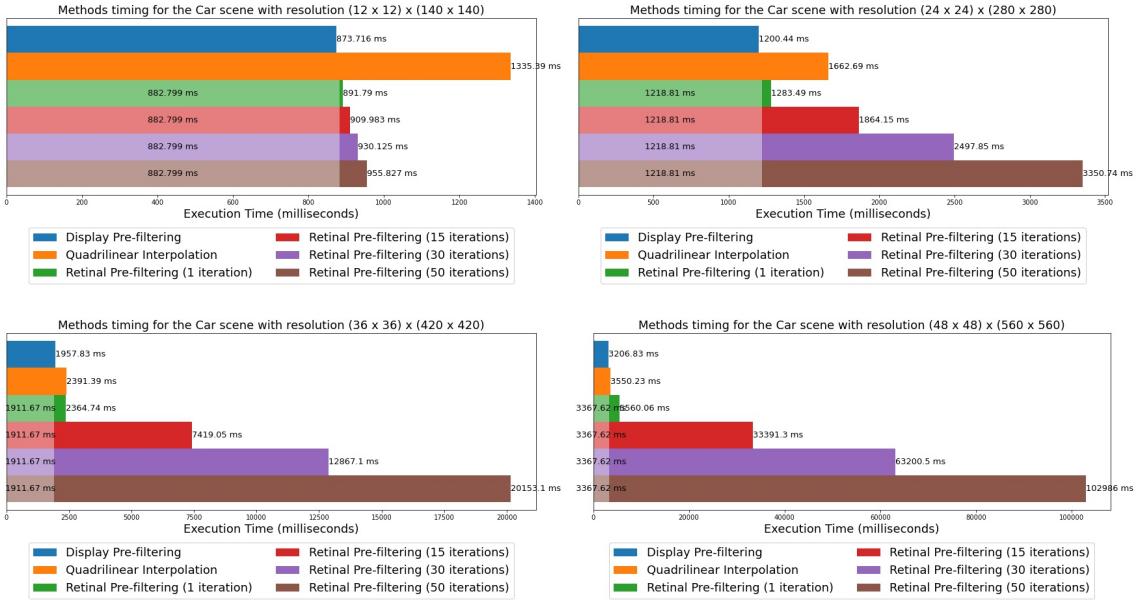


Figure B.2: Methods timing for the Car scene with different display resolutions.

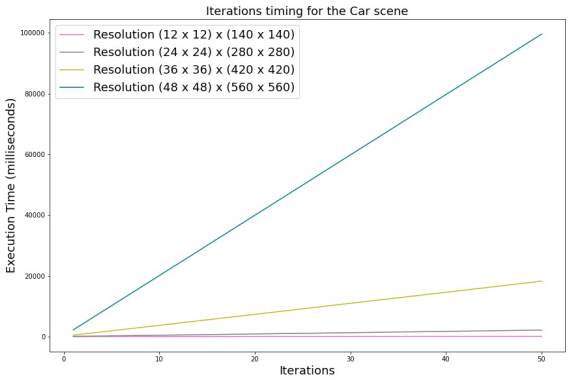


Figure B.3: Iterations timing for the Car scene with different display resolutions.

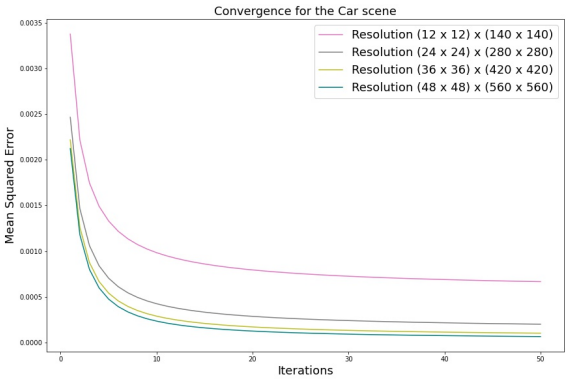
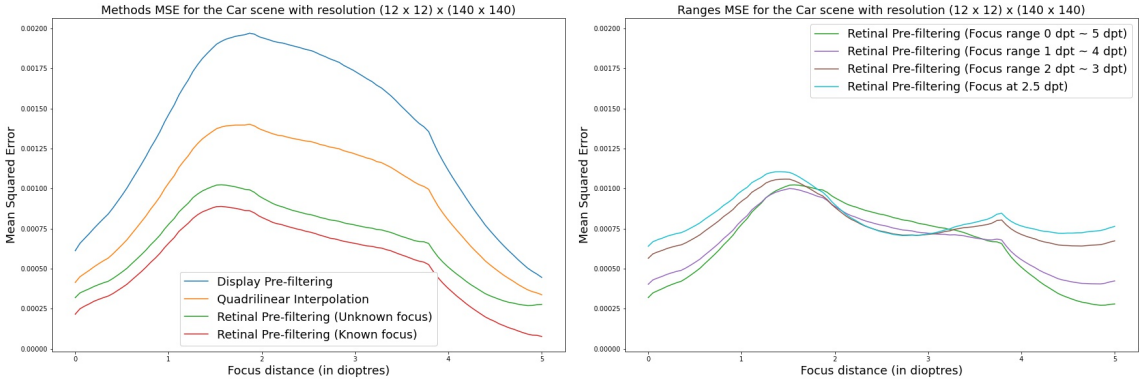
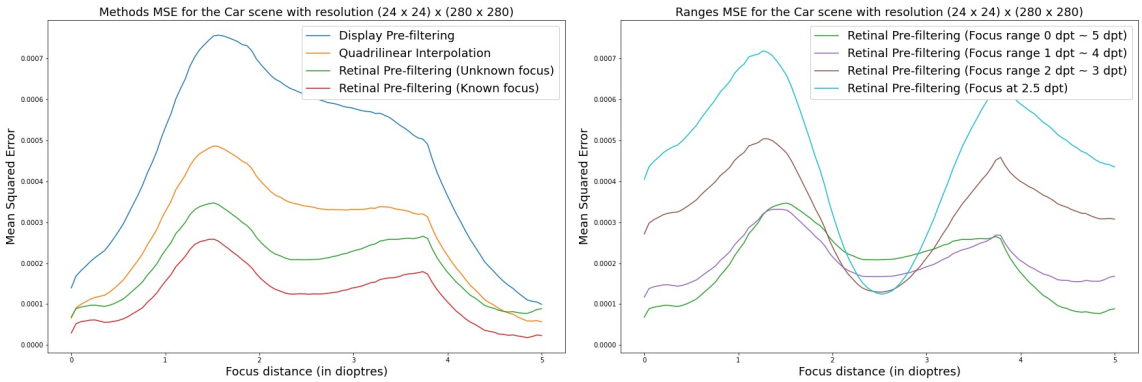


Figure B.4: Convergence of the retinal pre-filtering for the Car scene with different display resolutions.



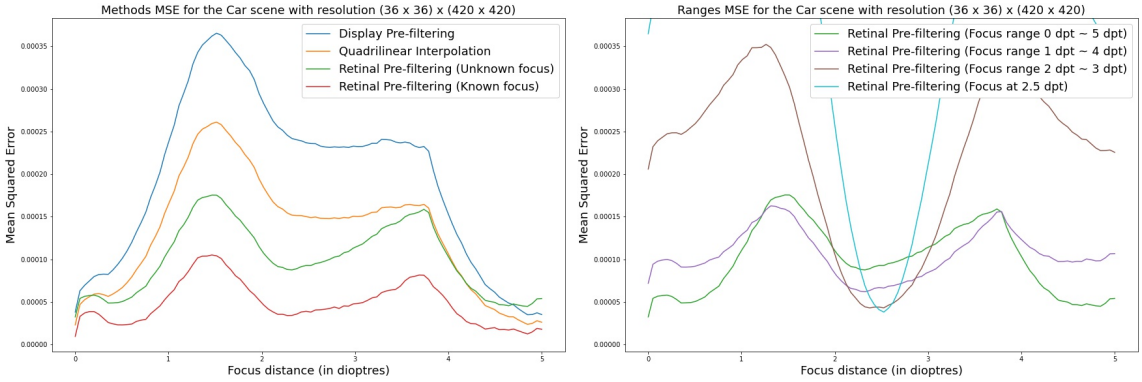
(a) (b)

Figure B.5: Comparison of pre-filtering methods (a) and retinal pre-filtering with different focus ranges (b) for the Car scene with display resolution $(12 \times 12) \times (140 \times 140)$.



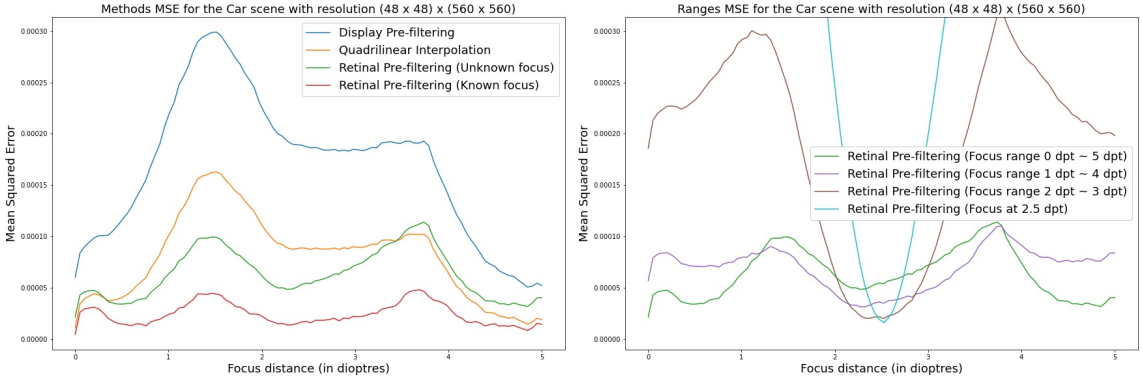
(a) (b)

Figure B.6: Comparison of pre-filtering methods (a) and retinal pre-filtering with different focus ranges (b) for the Car scene with display resolution $(24 \times 24) \times (280 \times 280)$.



(a) (b)

Figure B.7: Comparison of pre-filtering methods (a) and retinal pre-filtering with different focus ranges (b) for the Car scene with display resolution $(36 \times 36) \times (420 \times 420)$.



(a) (b)

Figure B.8: Comparison of pre-filtering methods (a) and retinal pre-filtering with different focus ranges (b) for the Car scene with display resolution $(48 \times 48) \times (560 \times 560)$.

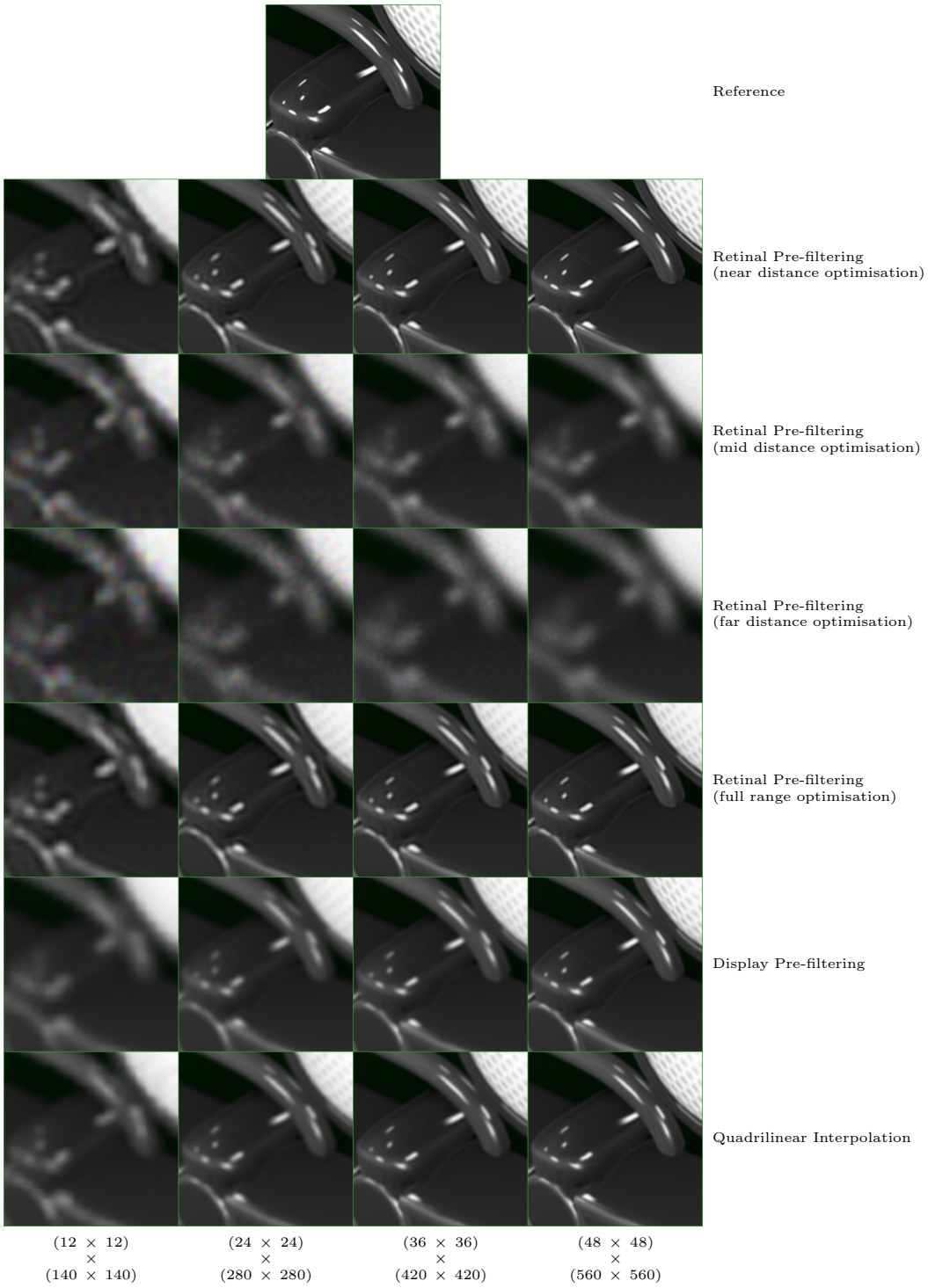


Figure B.9: Comparison of pre-filtering methods for the near insert of the Car scene with different display resolutions.

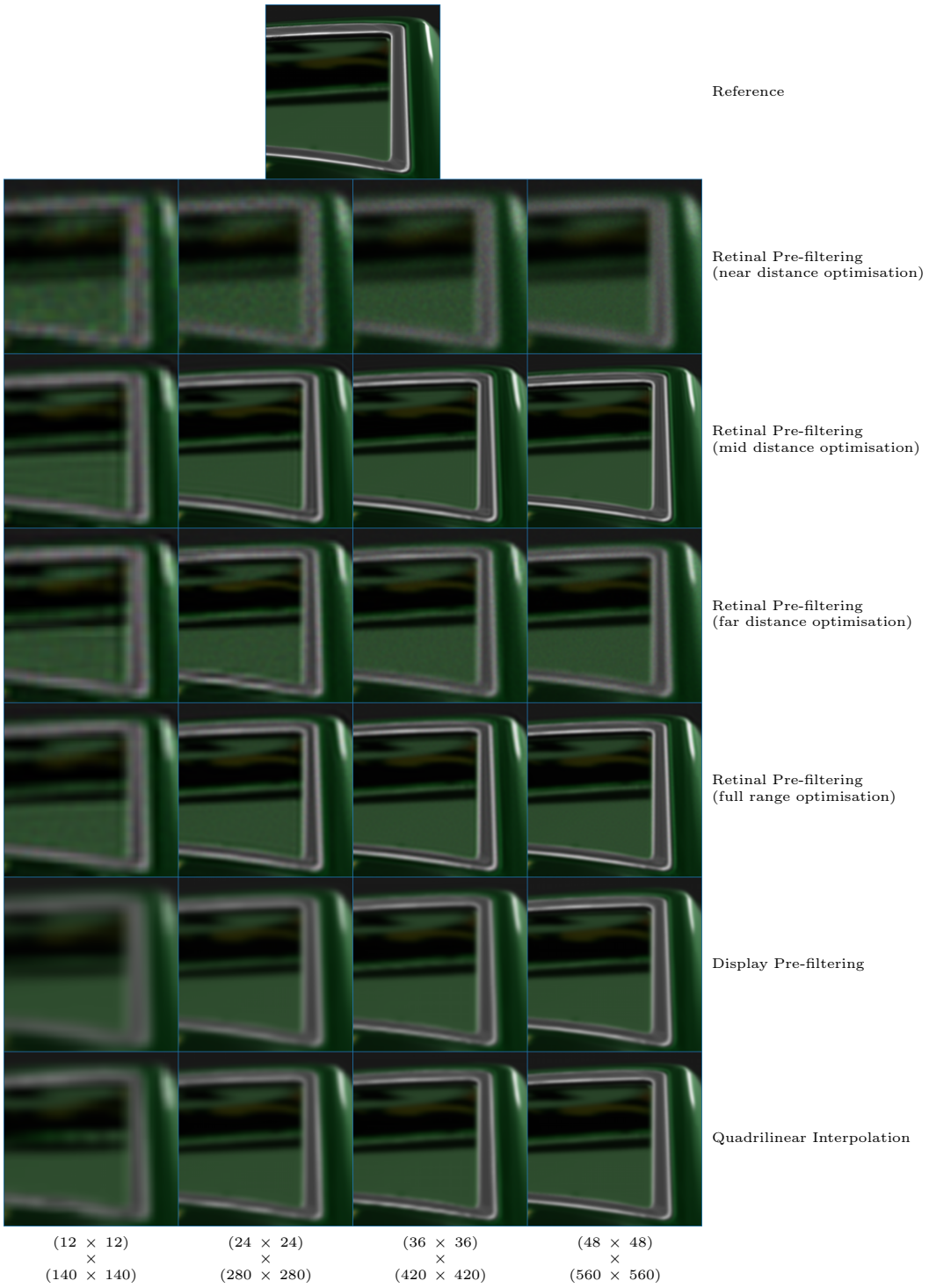


Figure B.10: Comparison of pre-filtering methods for the mid insert of the Car scene with different display resolutions.

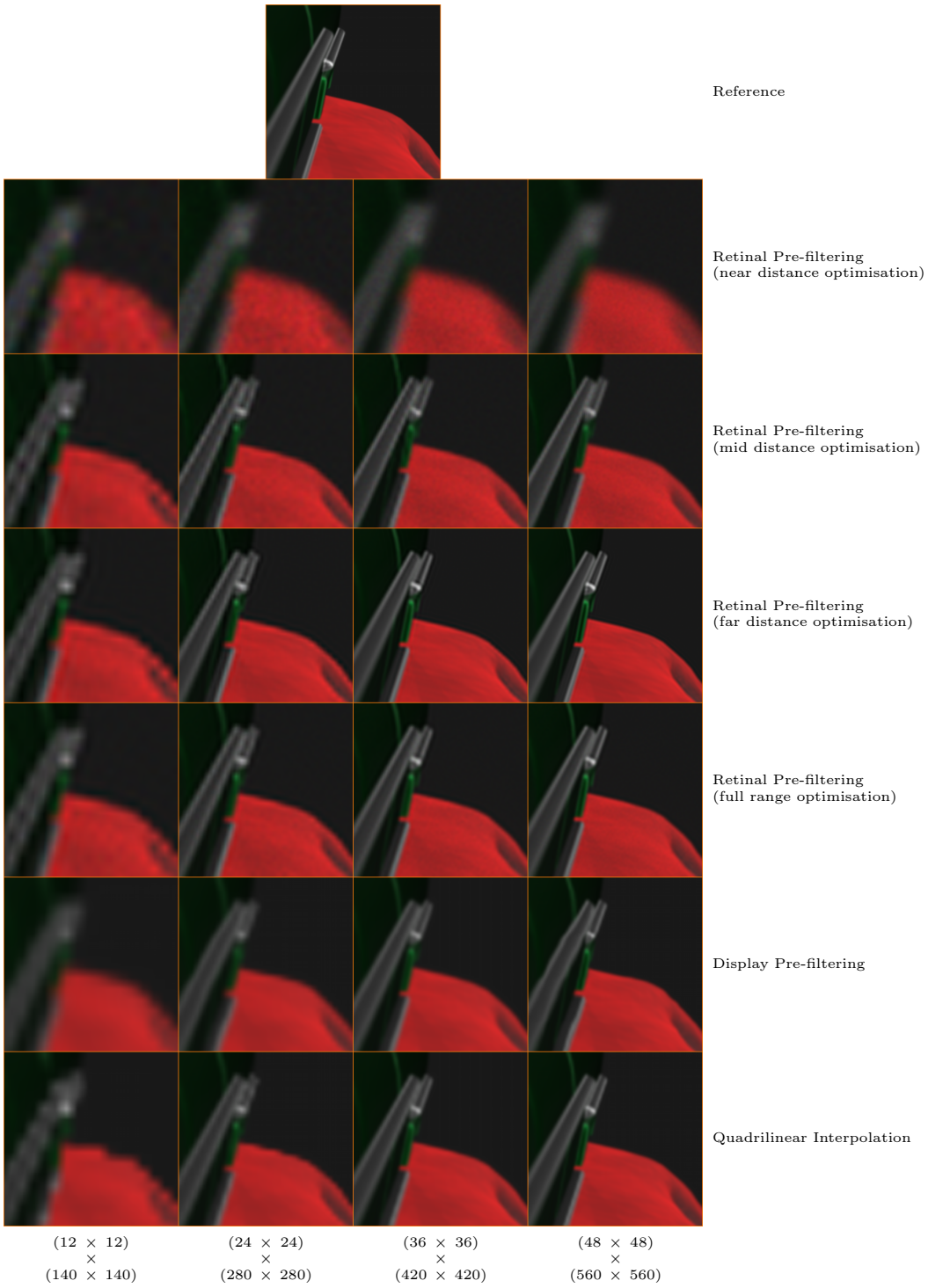


Figure B.11: Comparison of pre-filtering methods for the far insert of the Car scene with different display resolutions.

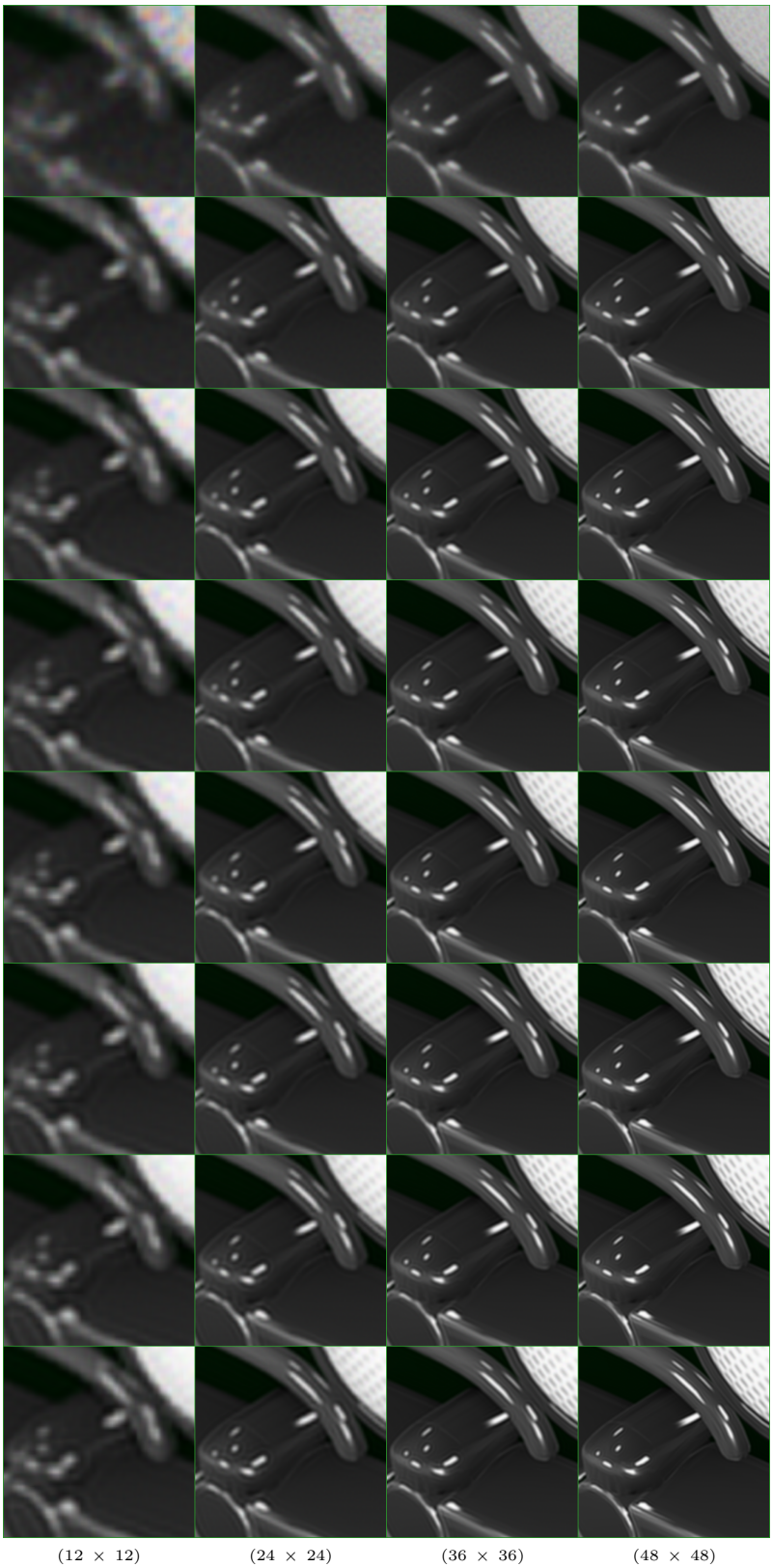


Figure B.12: Incremental improvements of retinal pre-filtering (near distance) across multiple iterations for the near insert of the Car scene with different display resolutions.

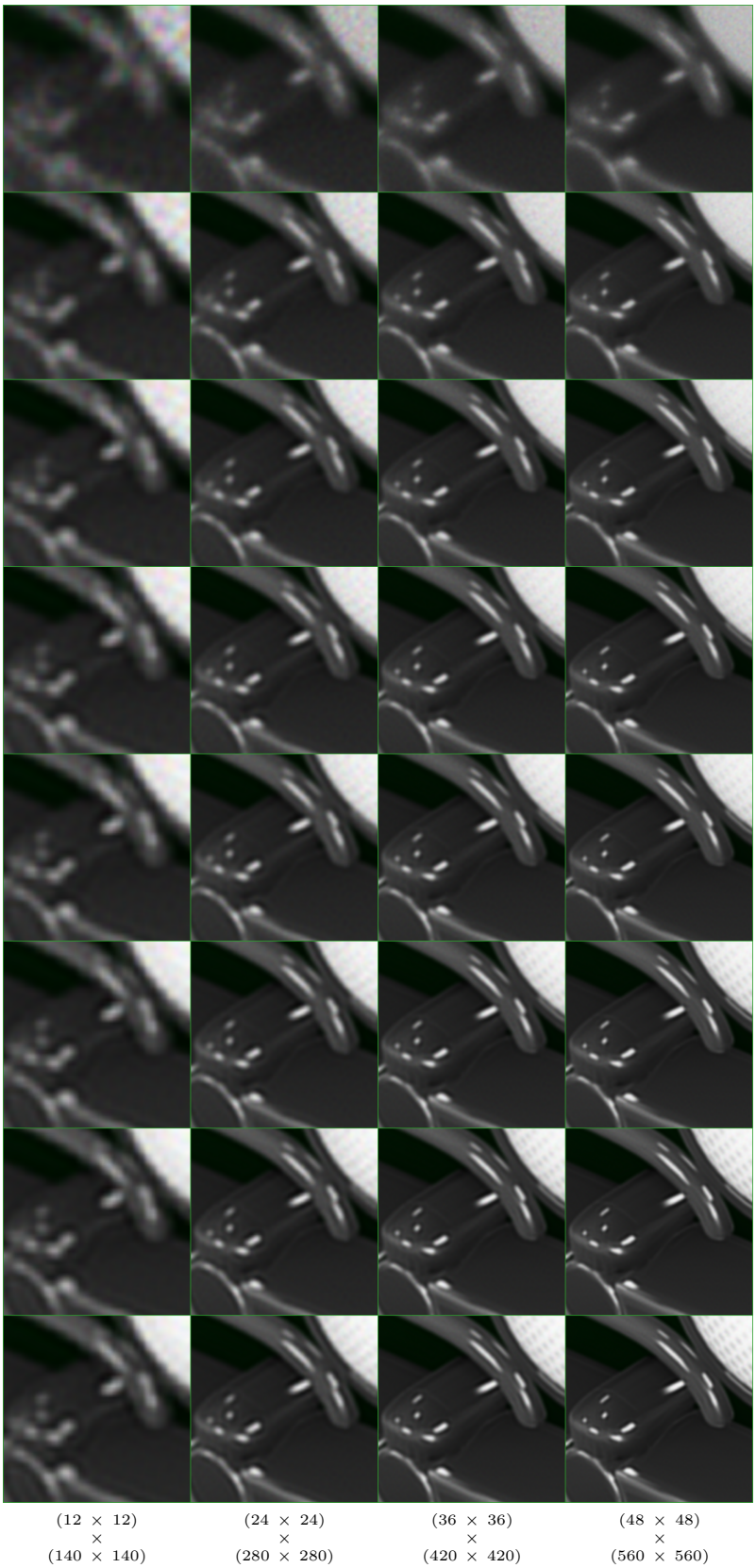


Figure B.13: Incremental improvements of retinal pre-filtering (full range) across multiple iterations for the near insert of the Car scene with different display resolutions.

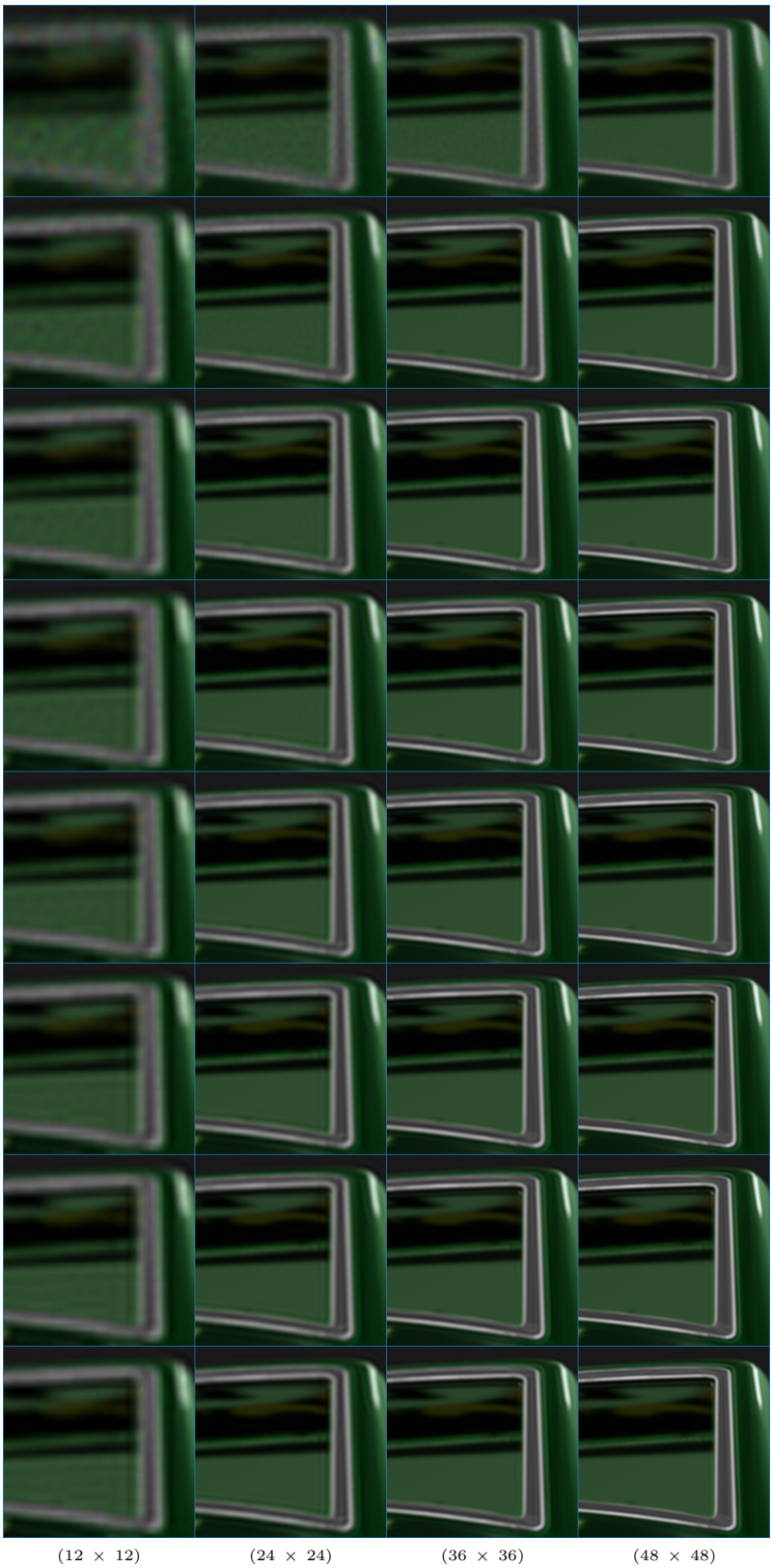


Figure B.14: Incremental improvements of retinal pre-filtering (mid distance) across multiple iterations for the mid insert of the Car scene with different display resolutions.

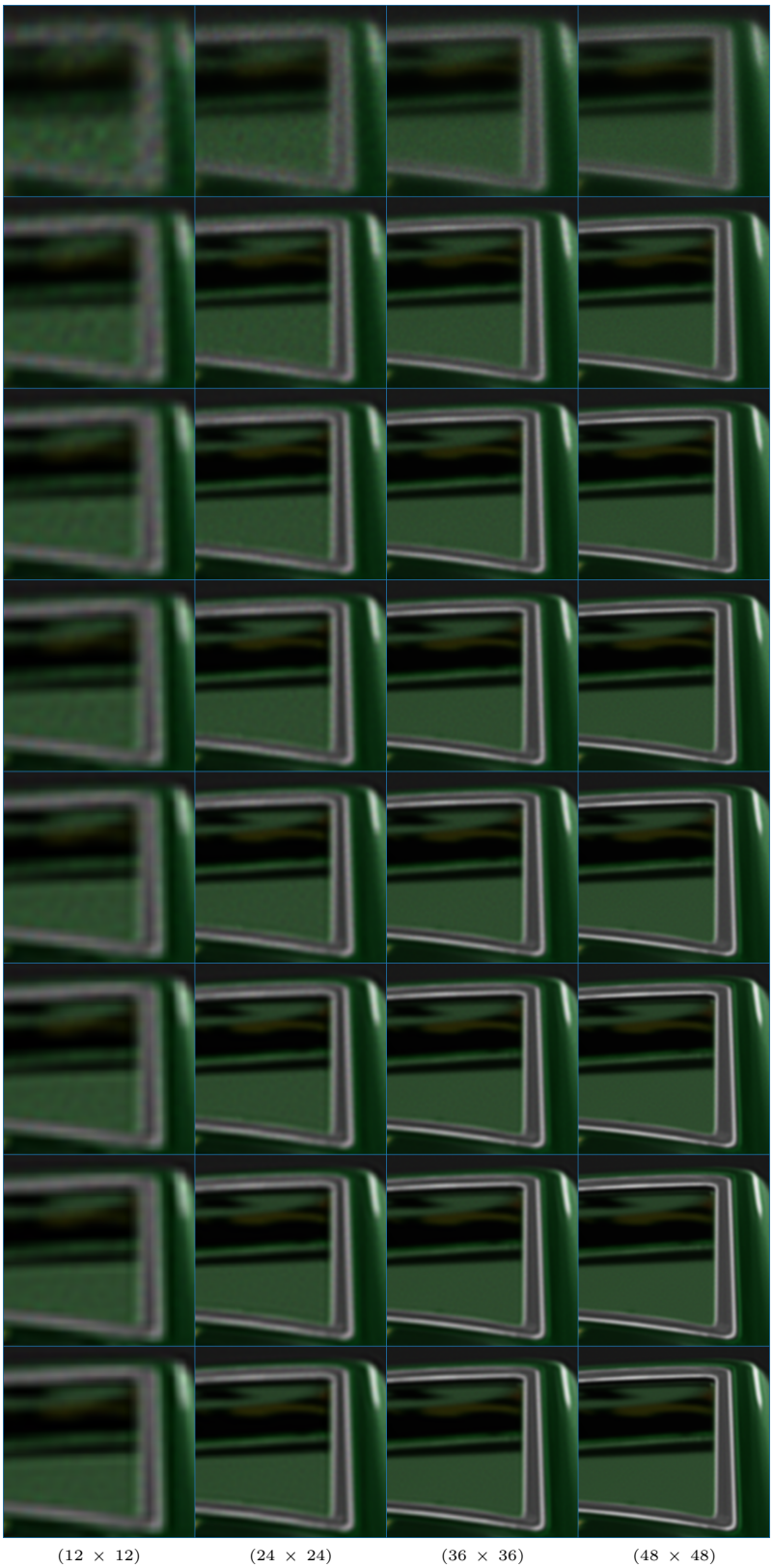


Figure B.15: Incremental improvements of retinal pre-filtering (full range) across multiple iterations for the mid insert of the Car scene with different display resolutions.

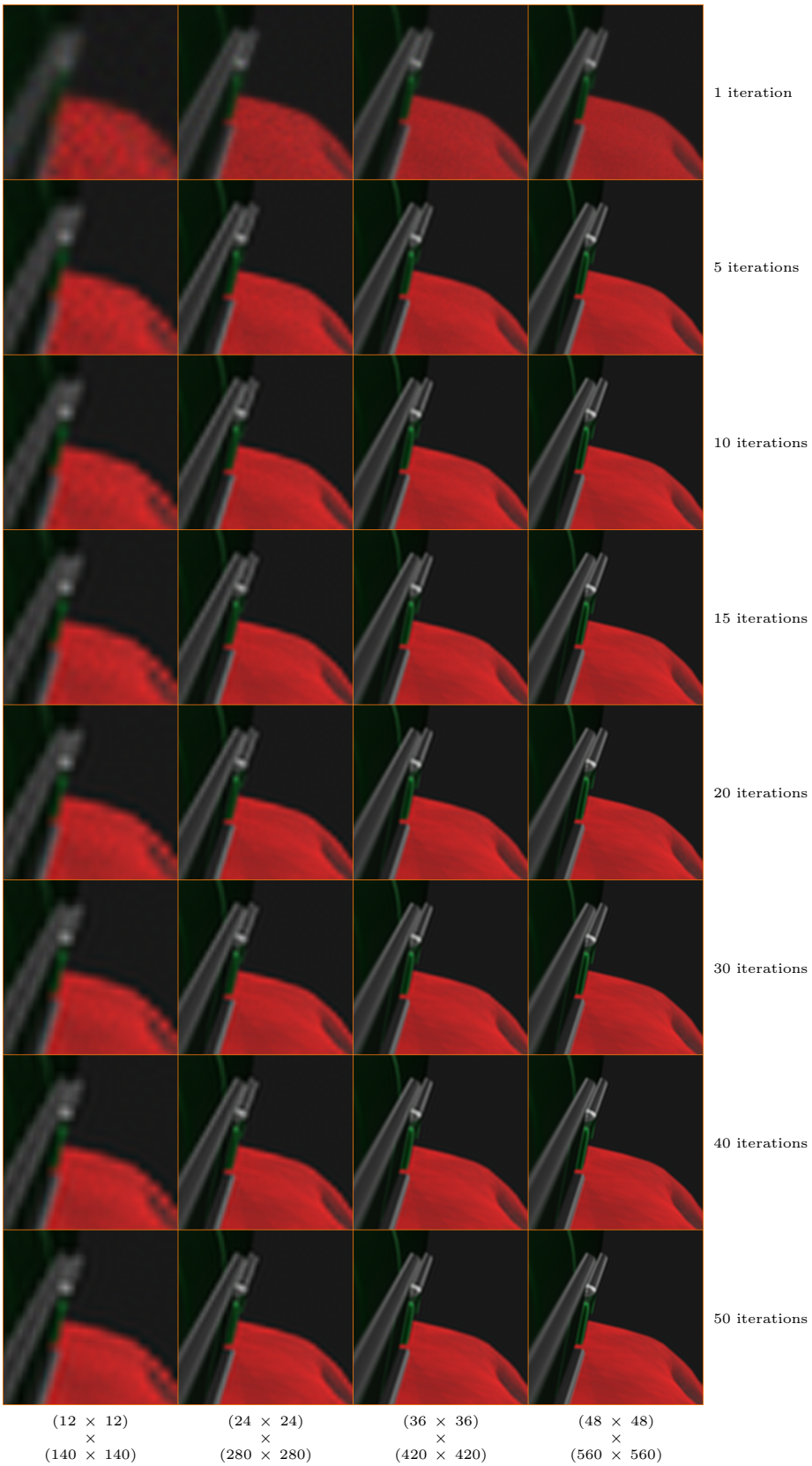
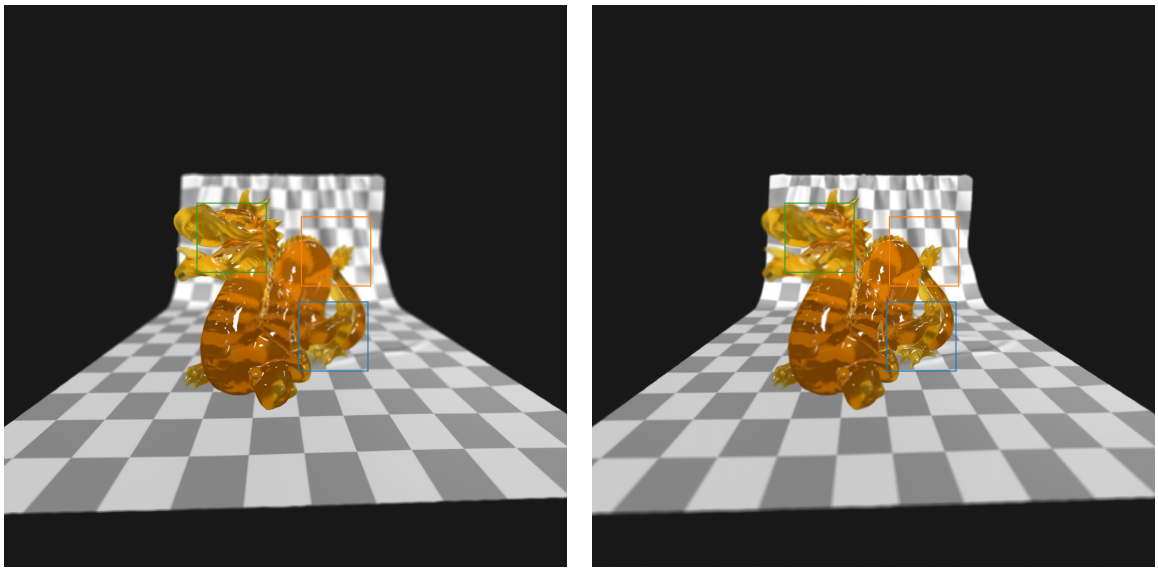


Figure B.16: Incremental improvements of retinal pre-filtering (far distance) across multiple iterations for the far insert of the Car scene with different display resolutions.

Figure B.17: Incremental improvements of retinal pre-filtering (full range) across multiple iterations for the far insert of the Car scene with different display resolutions.

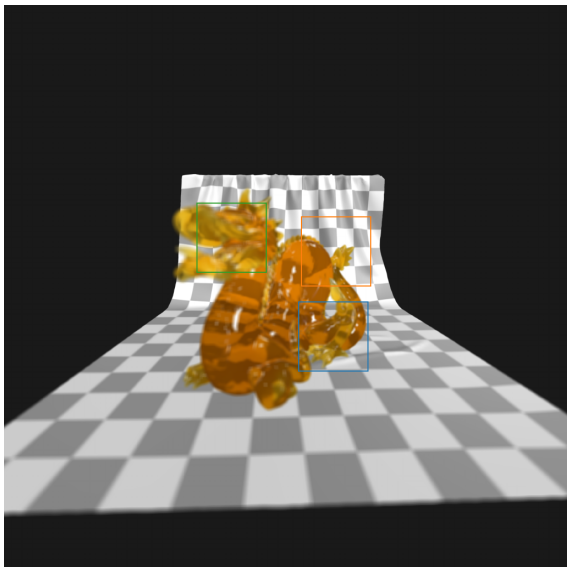
Appendix C

Results - Dragon Scene



(a)

(b)



(c)

Figure C.1: Reference retinal images for the Dragon scene with different focus distances.
(a) Near focus distance (green box).
(b) Mid focus distance (blue box).
(c) Far focus distance (orange box).

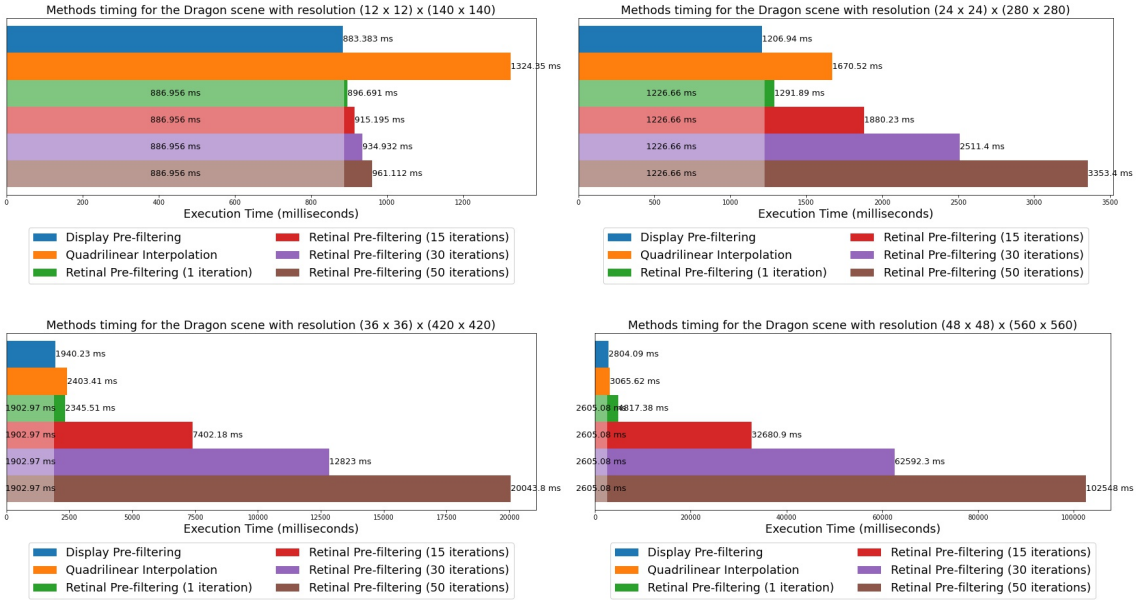


Figure C.2: Methods timing for the Dragon scene with different display resolutions.

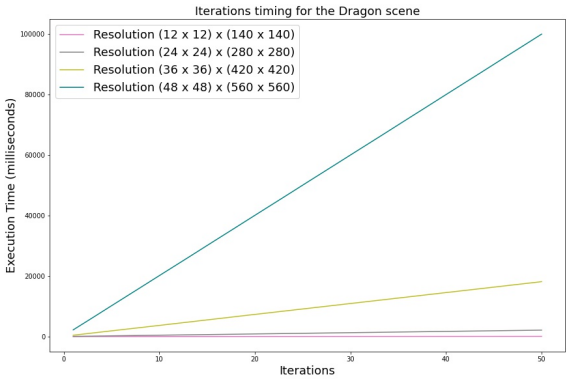


Figure C.3: Iterations timing for the Dragon scene with different display resolutions.

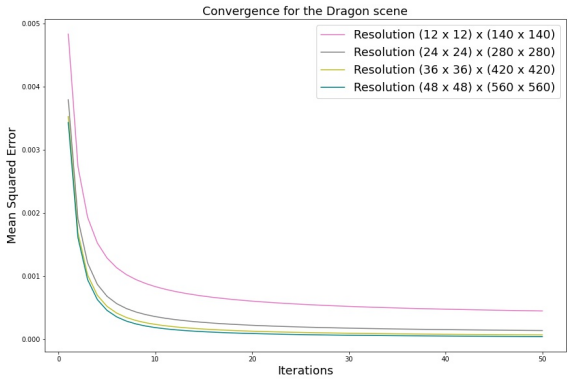
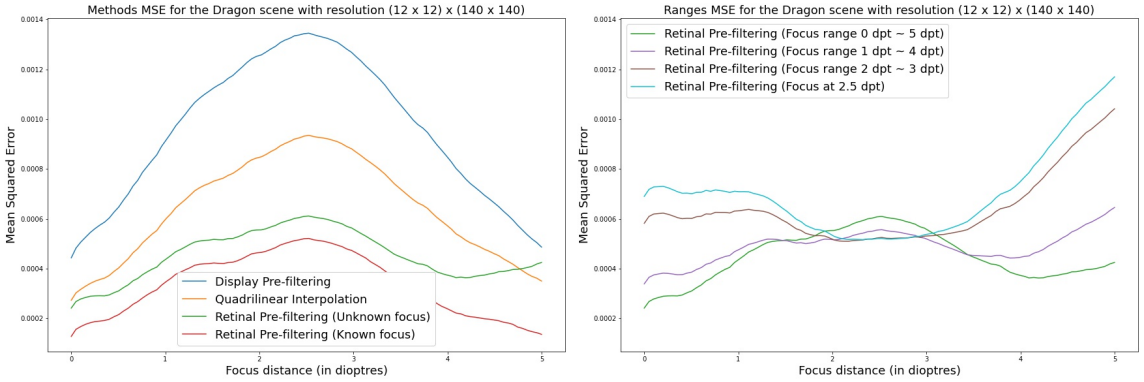
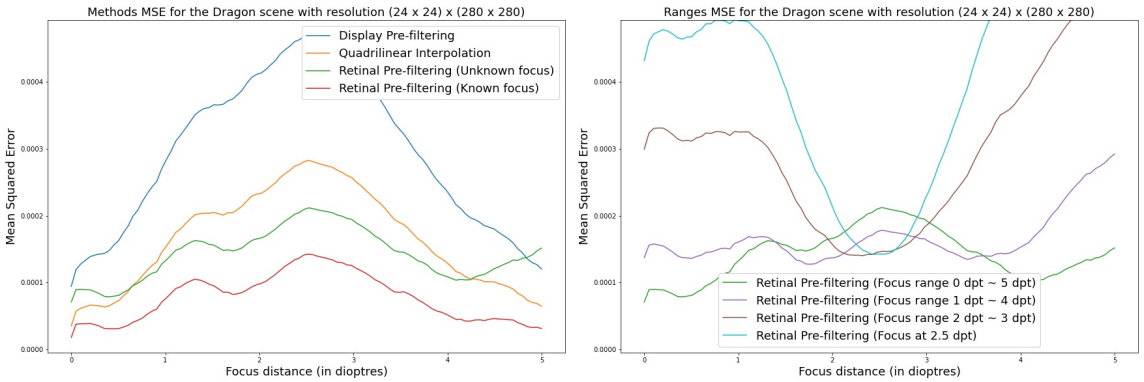


Figure C.4: Convergence of the retinal pre-filtering for the Dragon scene with different display resolutions.



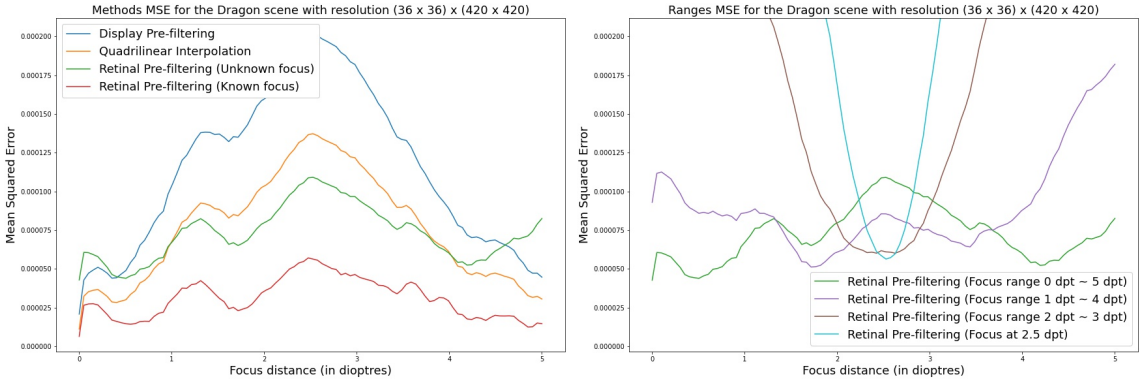
(a) (b)

Figure C.5: Comparison of pre-filtering methods (a) and retinal pre-filtering with different focus ranges (b) for the Dragon scene with display resolution $(12 \times 12) \times (140 \times 140)$.



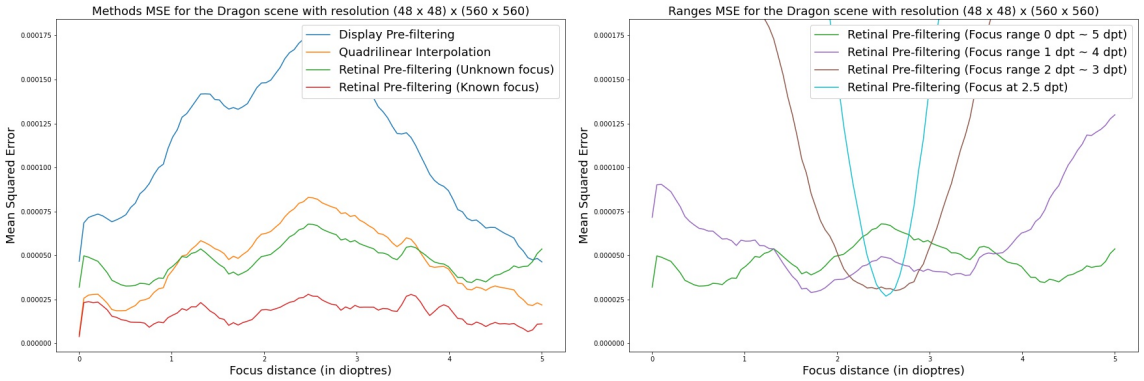
(a) (b)

Figure C.6: Comparison of pre-filtering methods (a) and retinal pre-filtering with different focus ranges (b) for the Dragon scene with display resolution $(24 \times 24) \times (280 \times 280)$.



(a) (b)

Figure C.7: Comparison of pre-filtering methods (a) and retinal pre-filtering with different focus ranges (b) for the Dragon scene with display resolution $(36 \times 36) \times (420 \times 420)$.



(a) (b)

Figure C.8: Comparison of pre-filtering methods (a) and retinal pre-filtering with different focus ranges (b) for the Dragon scene with display resolution $(48 \times 48) \times (560 \times 560)$.

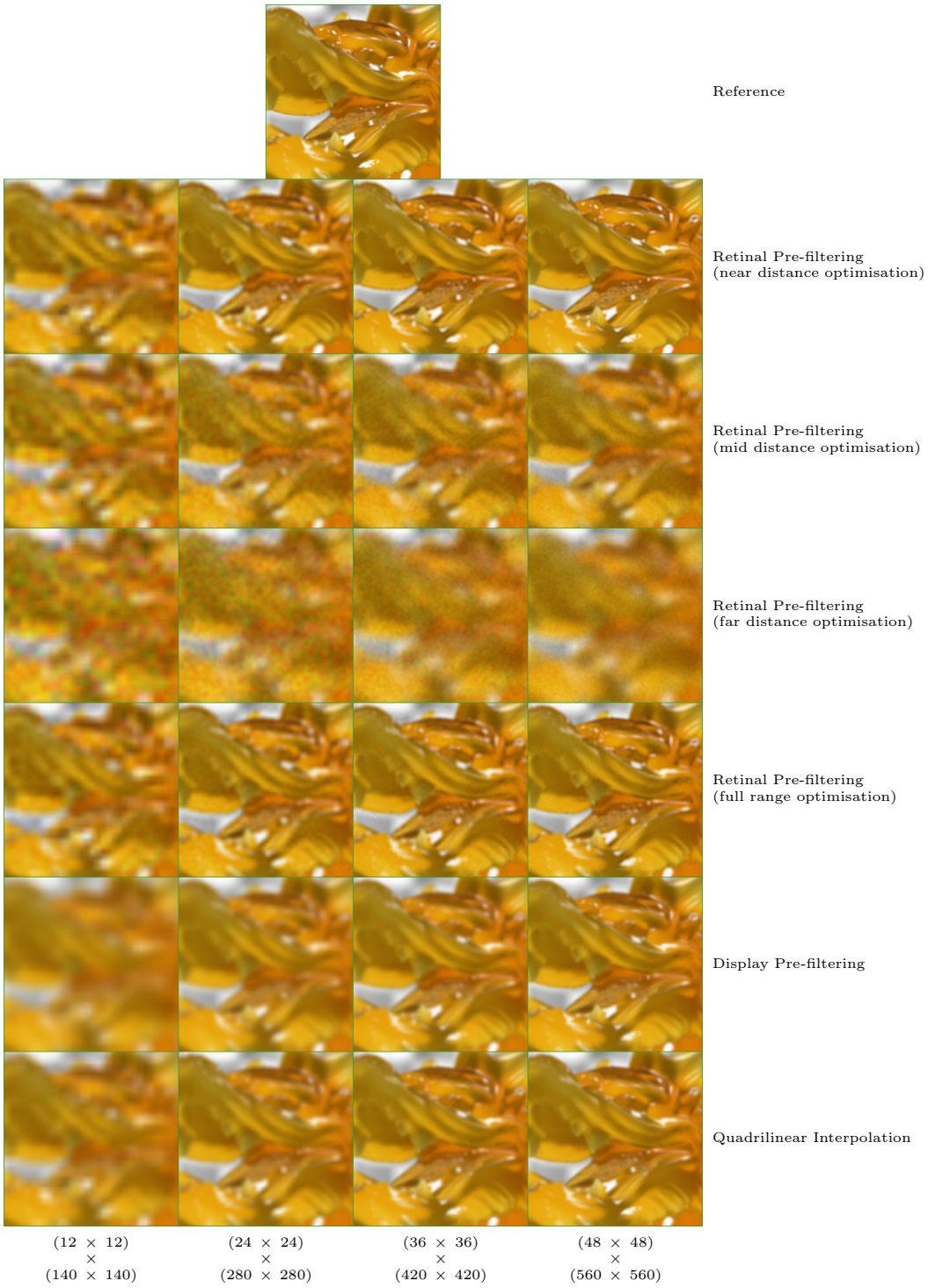


Figure C.9: Comparison of pre-filtering methods for the near insert of the Dragon scene with different display resolutions.

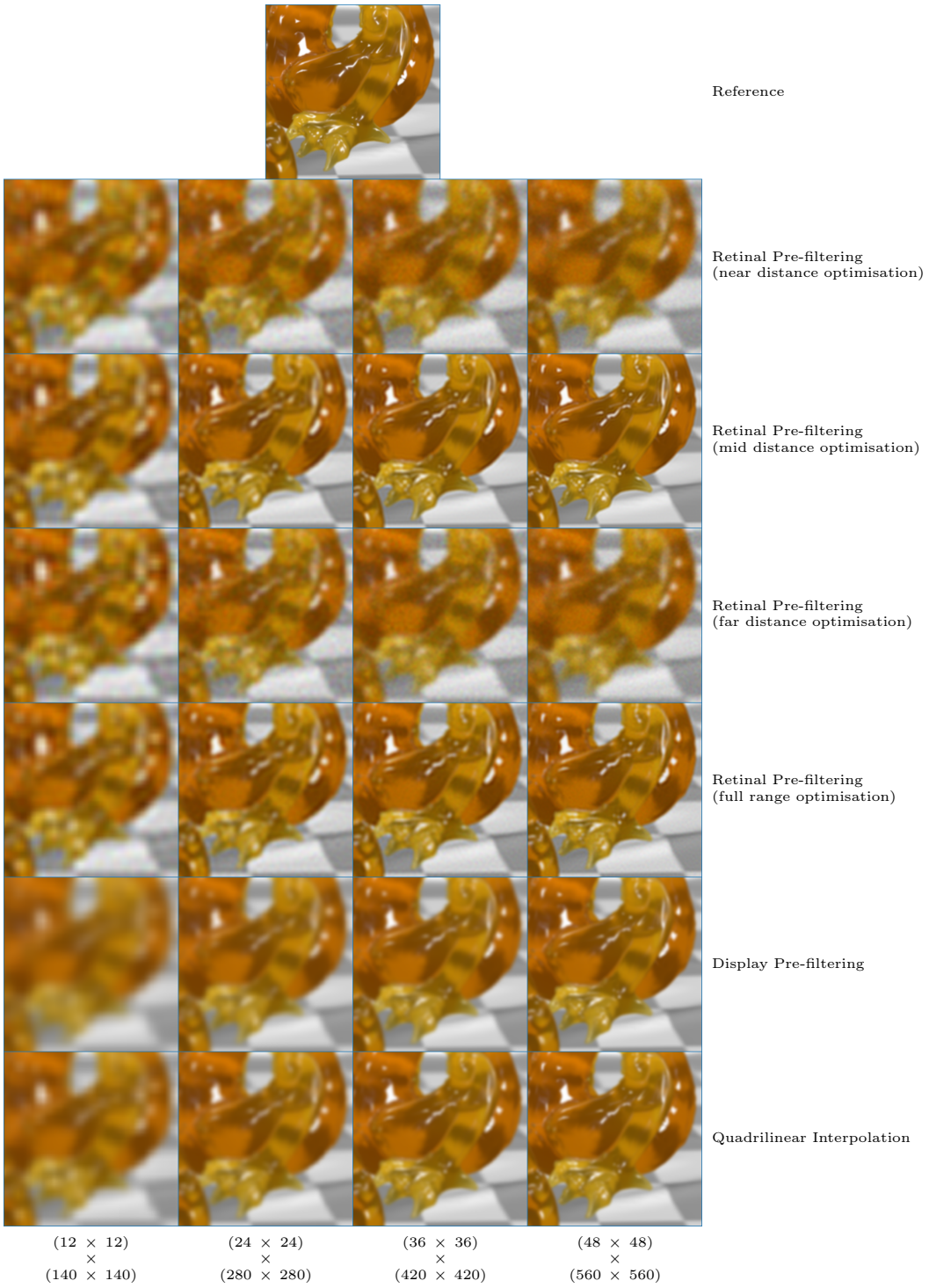


Figure C.10: Comparison of pre-filtering methods for the mid insert of the Dragon scene with different display resolutions.

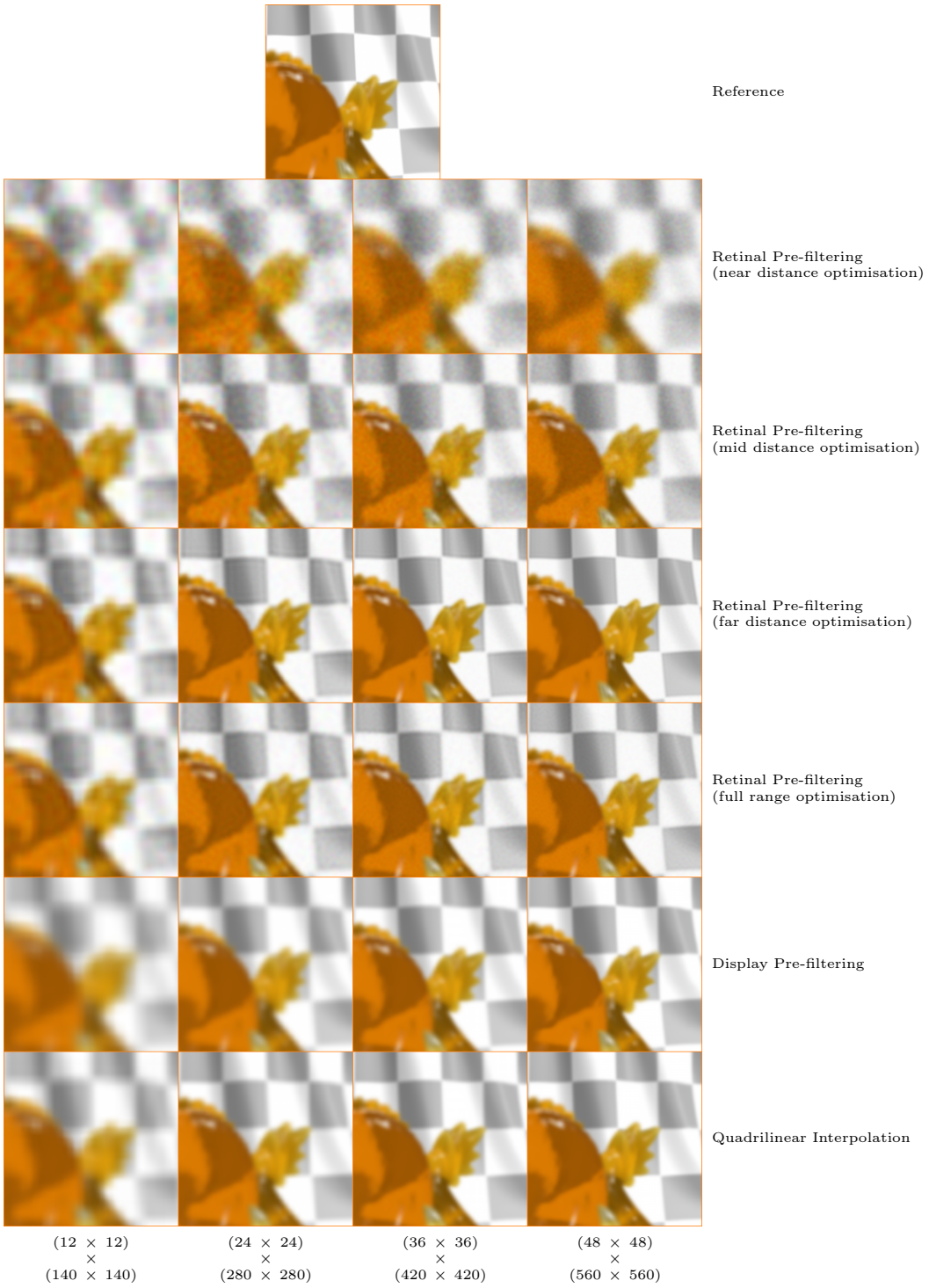


Figure C.11: Comparison of pre-filtering methods for the far insert of the Dragon scene with different display resolutions.

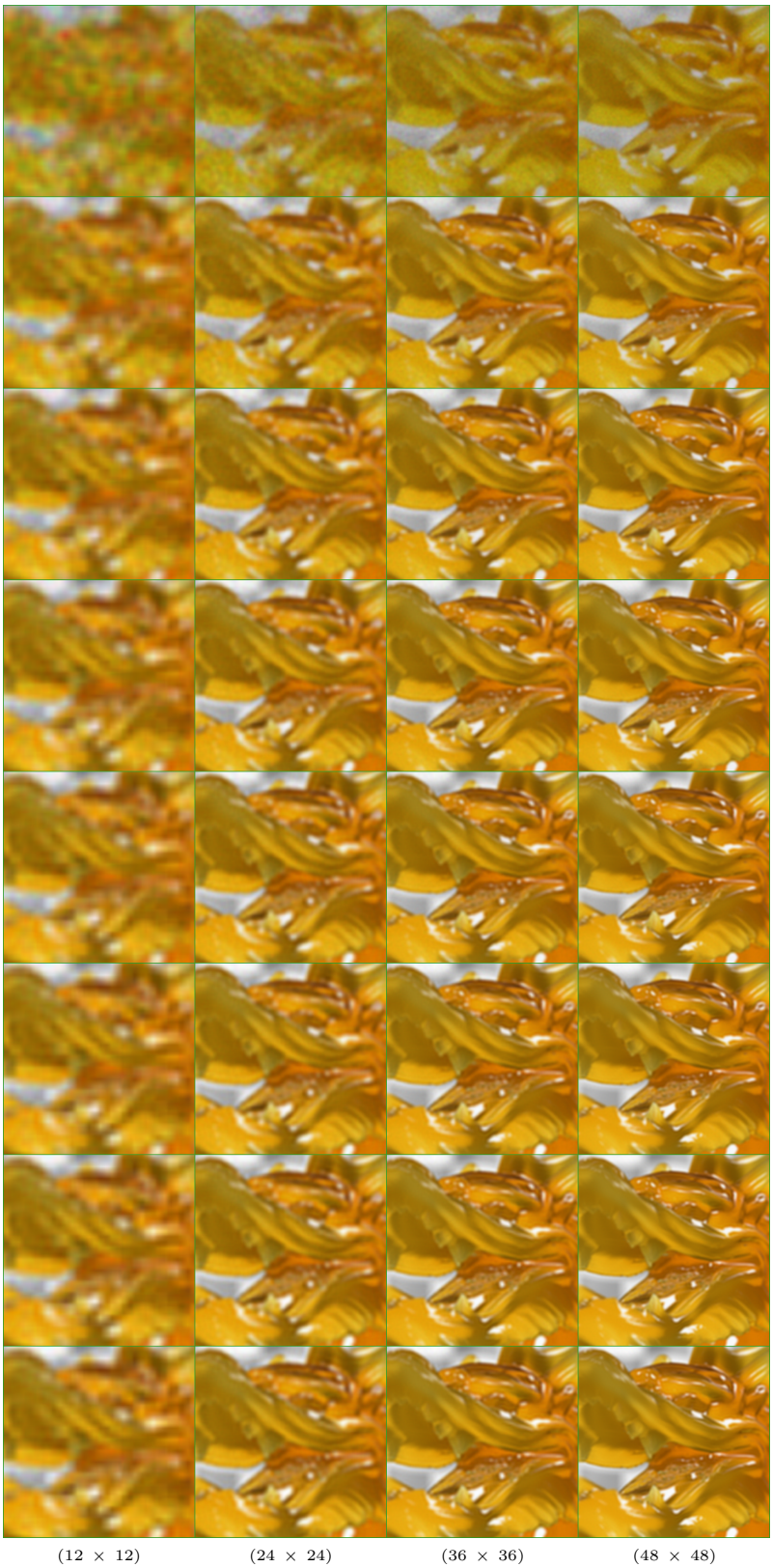


Figure C.12: Incremental improvements of retinal pre-filtering (near distance) across multiple iterations for the near insert of the Dragon scene with different display resolutions.

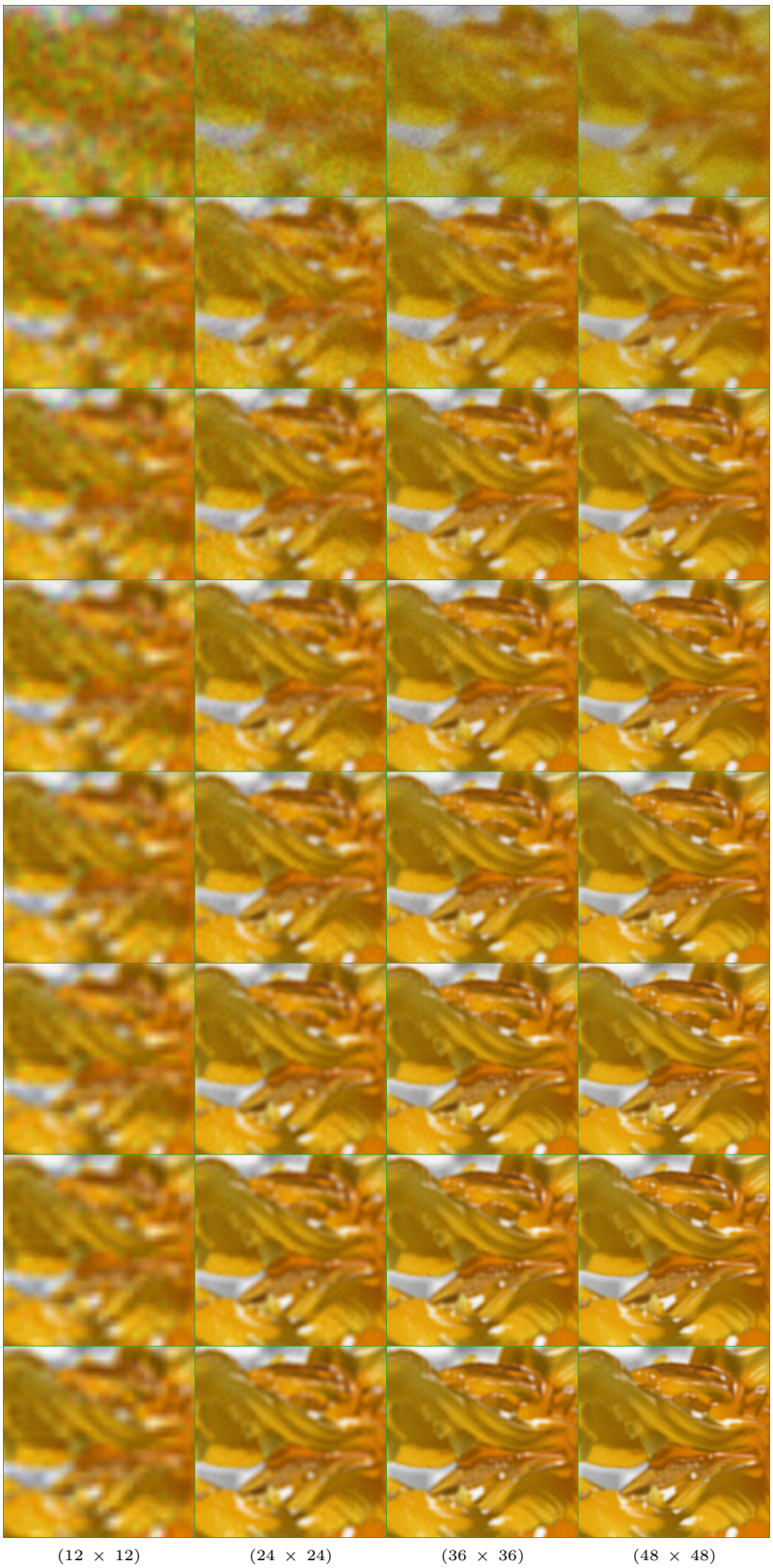


Figure C.13: Incremental improvements of retinal pre-filtering (full range) across multiple iterations for the near insert of the Dragon scene with different display resolutions.

Figure C.14: Incremental improvements of retinal pre-filtering (mid distance) across multiple iterations for the mid insert of the Dragon scene with different display resolutions.

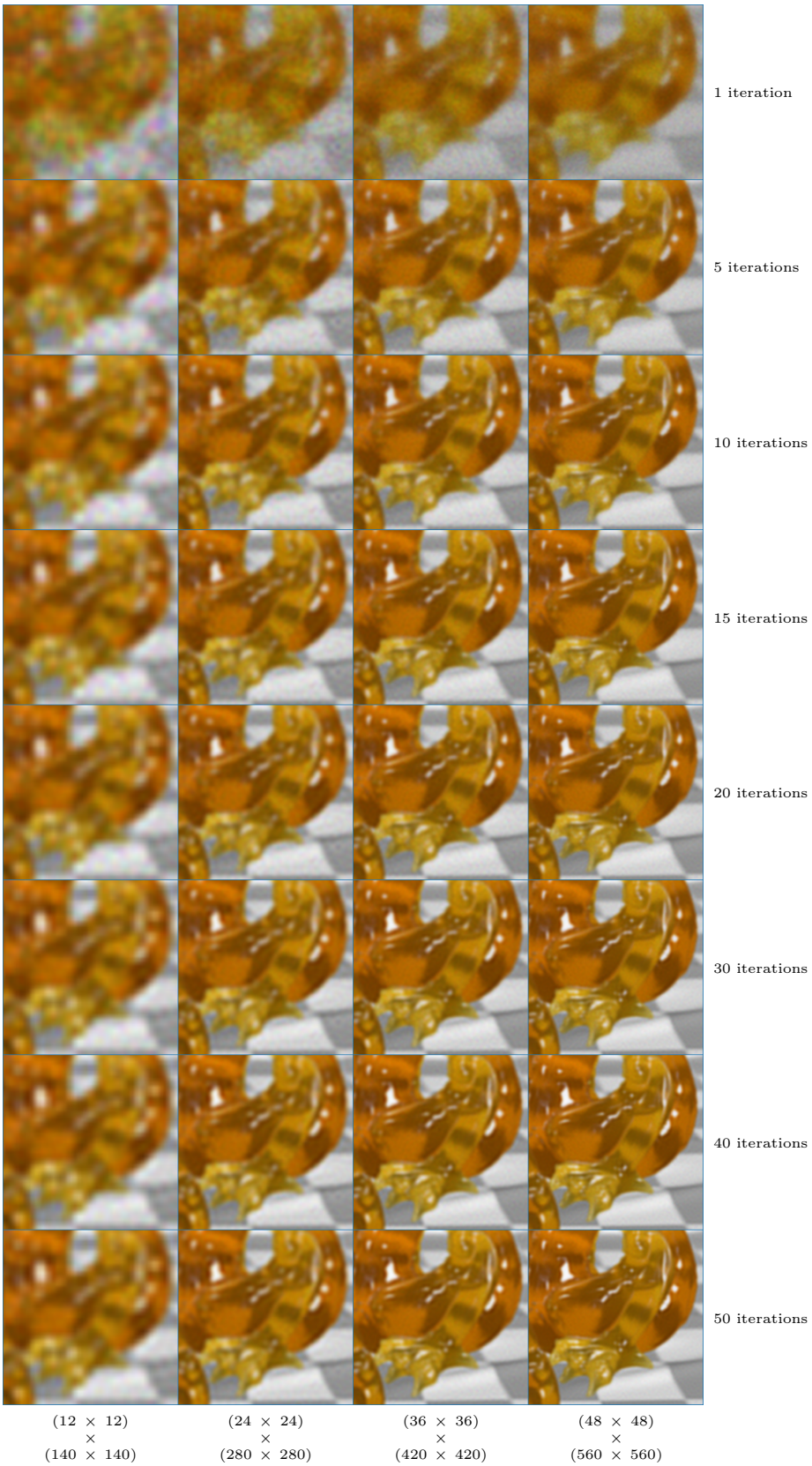


Figure C.15: Incremental improvements of retinal pre-filtering (full range) across multiple iterations for the mid insert of the Dragon scene with different display resolutions.

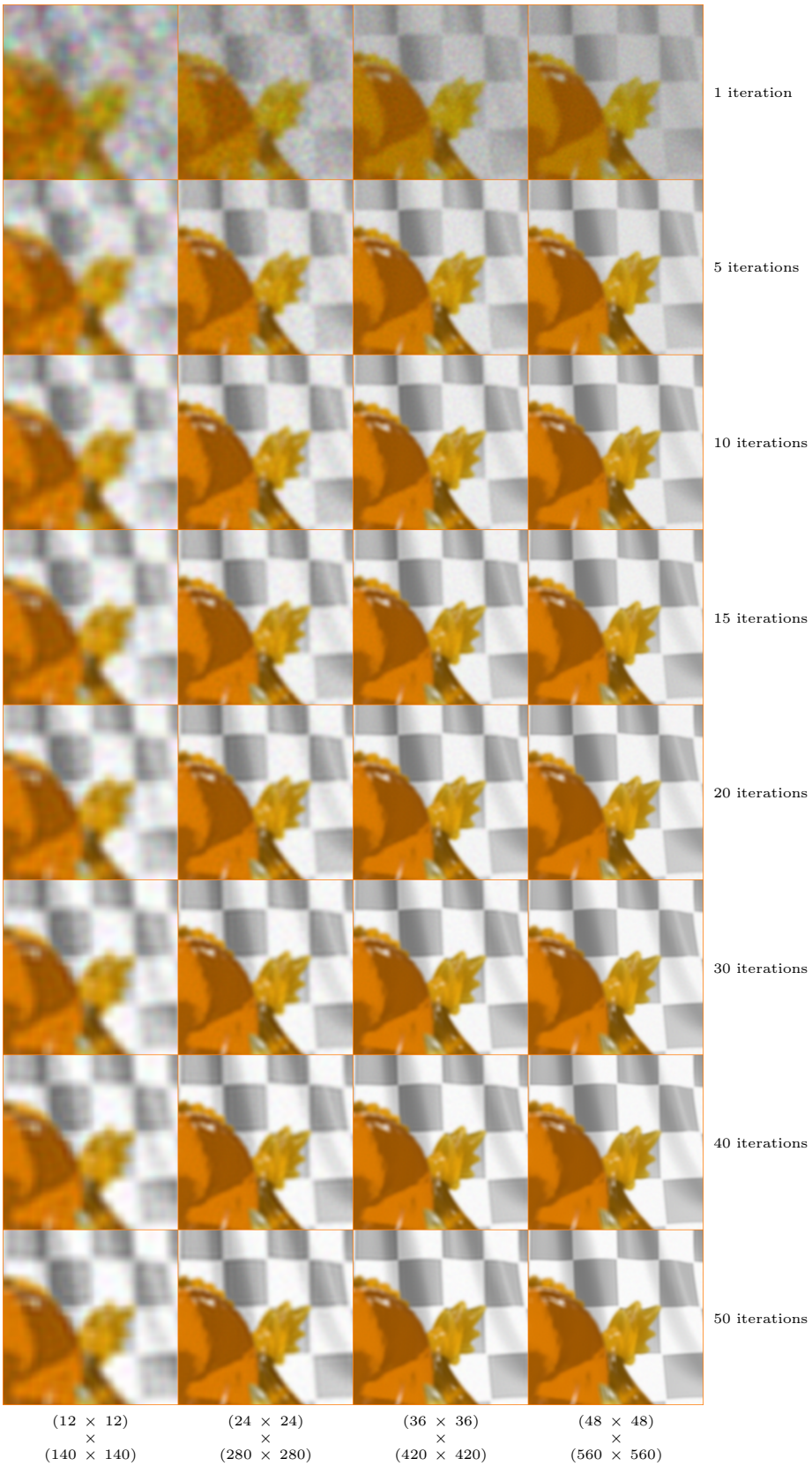


Figure C.16: Incremental improvements of retinal pre-filtering (far distance) across multiple iterations for the far insert of the Dragon scene with different display resolutions.

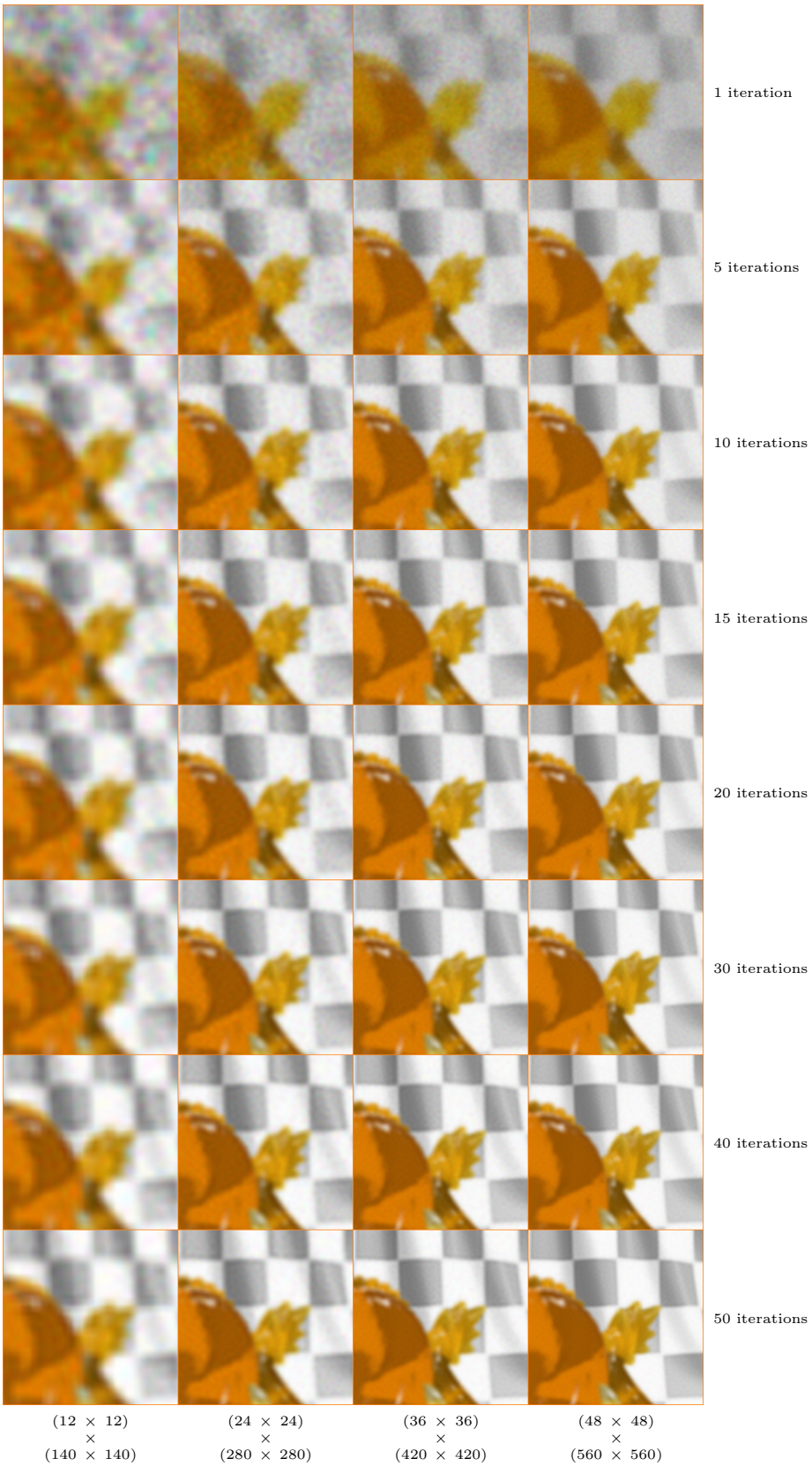


Figure C.17: Incremental improvements of retinal pre-filtering (full range) across multiple iterations for the far insert of the Dragon scene with different display resolutions.

Appendix D

Results - Sponza Scene

(a)

(b)

(c)

Figure D.1: Reference retinal images for the Sponza scene with different focus distances.
(a) Near focus distance (green box).
(b) Mid focus distance (blue box).
(c) Far focus distance (orange box).

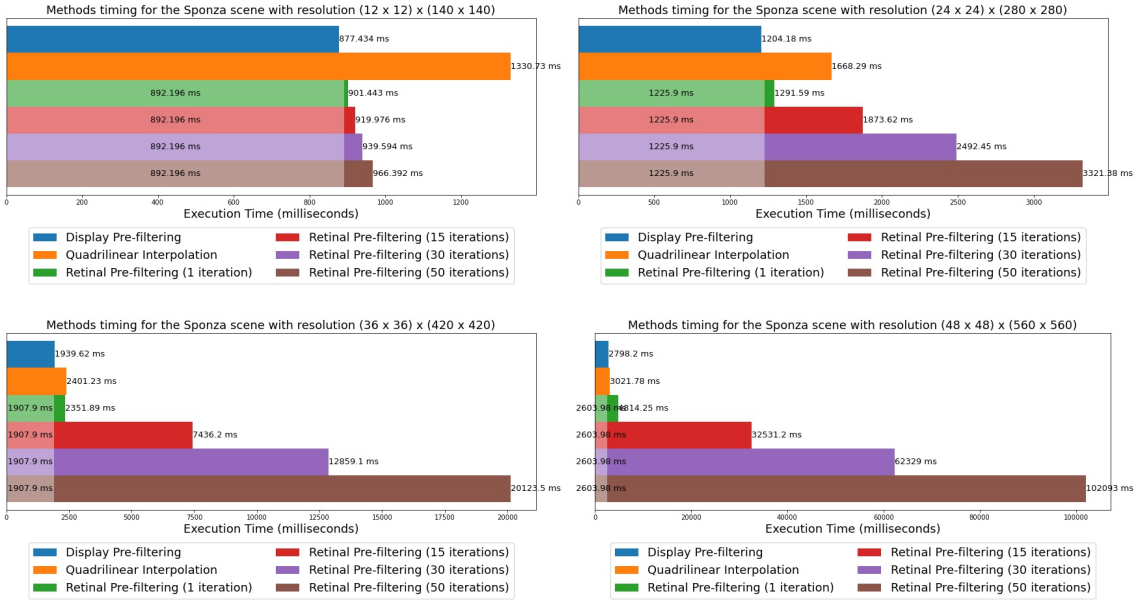


Figure D.2: Methods timing for the Sponza scene with different display resolutions.

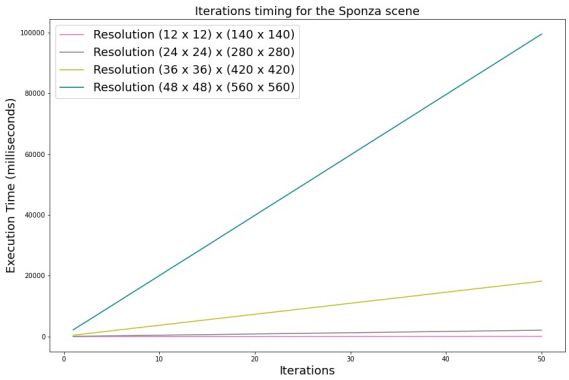


Figure D.3: Iterations timing for the Sponza scene with different display resolutions.

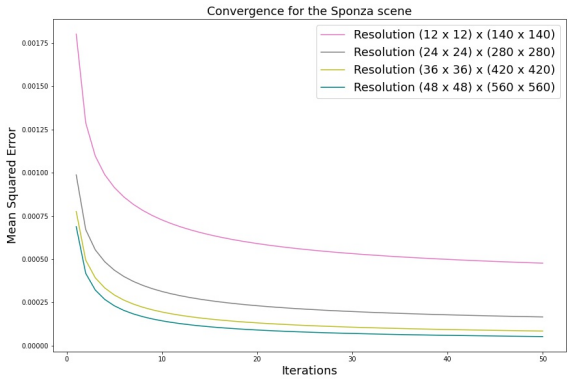
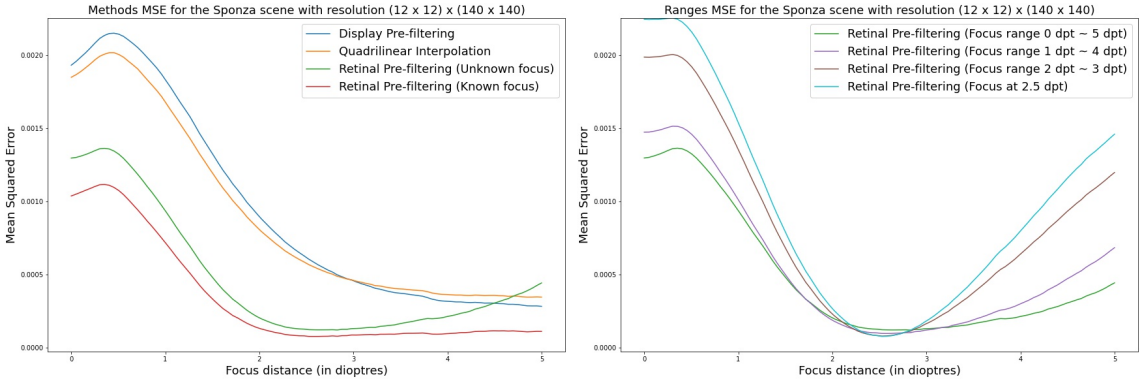
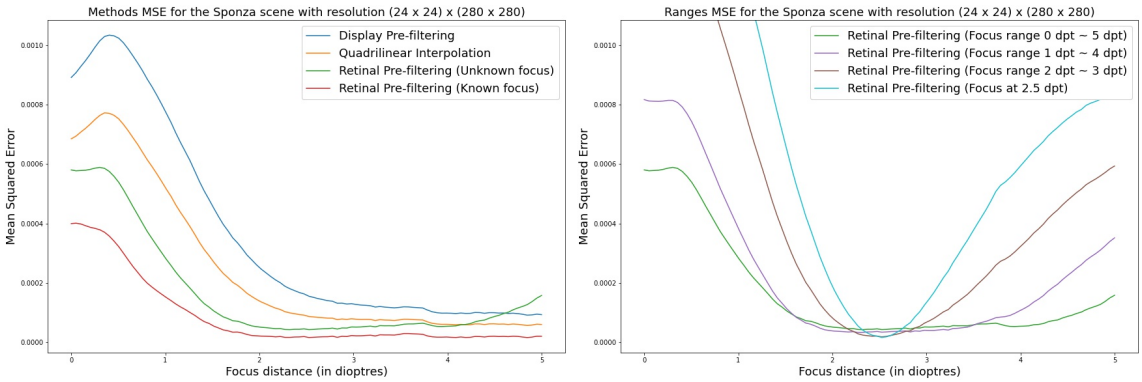


Figure D.4: Convergence of the retinal pre-filtering for the Sponza scene with different display resolutions.



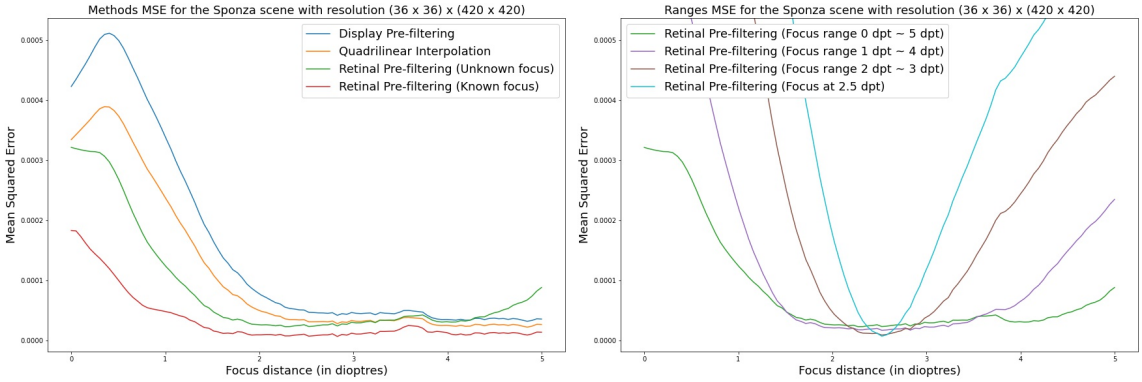
(a) (b)

Figure D.5: Comparison of pre-filtering methods (a) and retinal pre-filtering with different focus ranges (b) for the Sponza scene with display resolution $(12 \times 12) \times (140 \times 140)$.



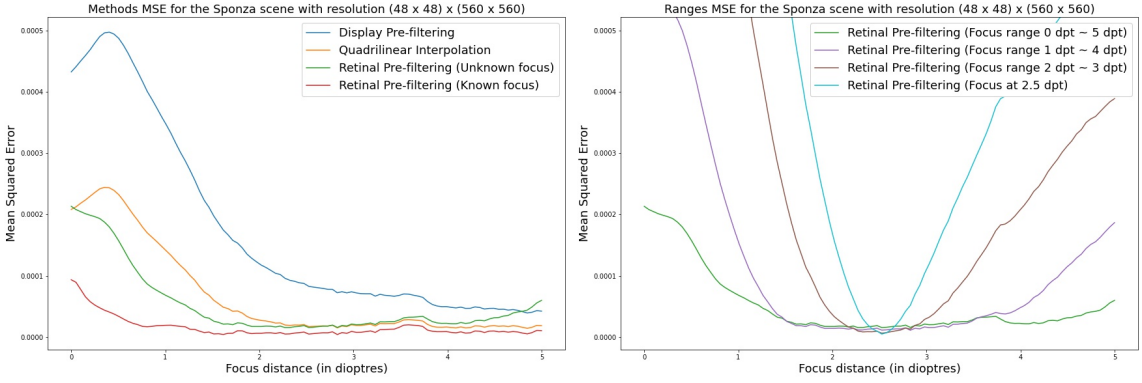
(a) (b)

Figure D.6: Comparison of pre-filtering methods (a) and retinal pre-filtering with different focus ranges (b) for the Sponza scene with display resolution $(24 \times 24) \times (280 \times 280)$.



(a) (b)

Figure D.7: Comparison of pre-filtering methods (a) and retinal pre-filtering with different focus ranges (b) for the Sponza scene with display resolution $(36 \times 36) \times (420 \times 420)$.



(a) (b)

Figure D.8: Comparison of pre-filtering methods (a) and retinal pre-filtering with different focus ranges (b) for the Sponza scene with display resolution $(48 \times 48) \times (560 \times 560)$.

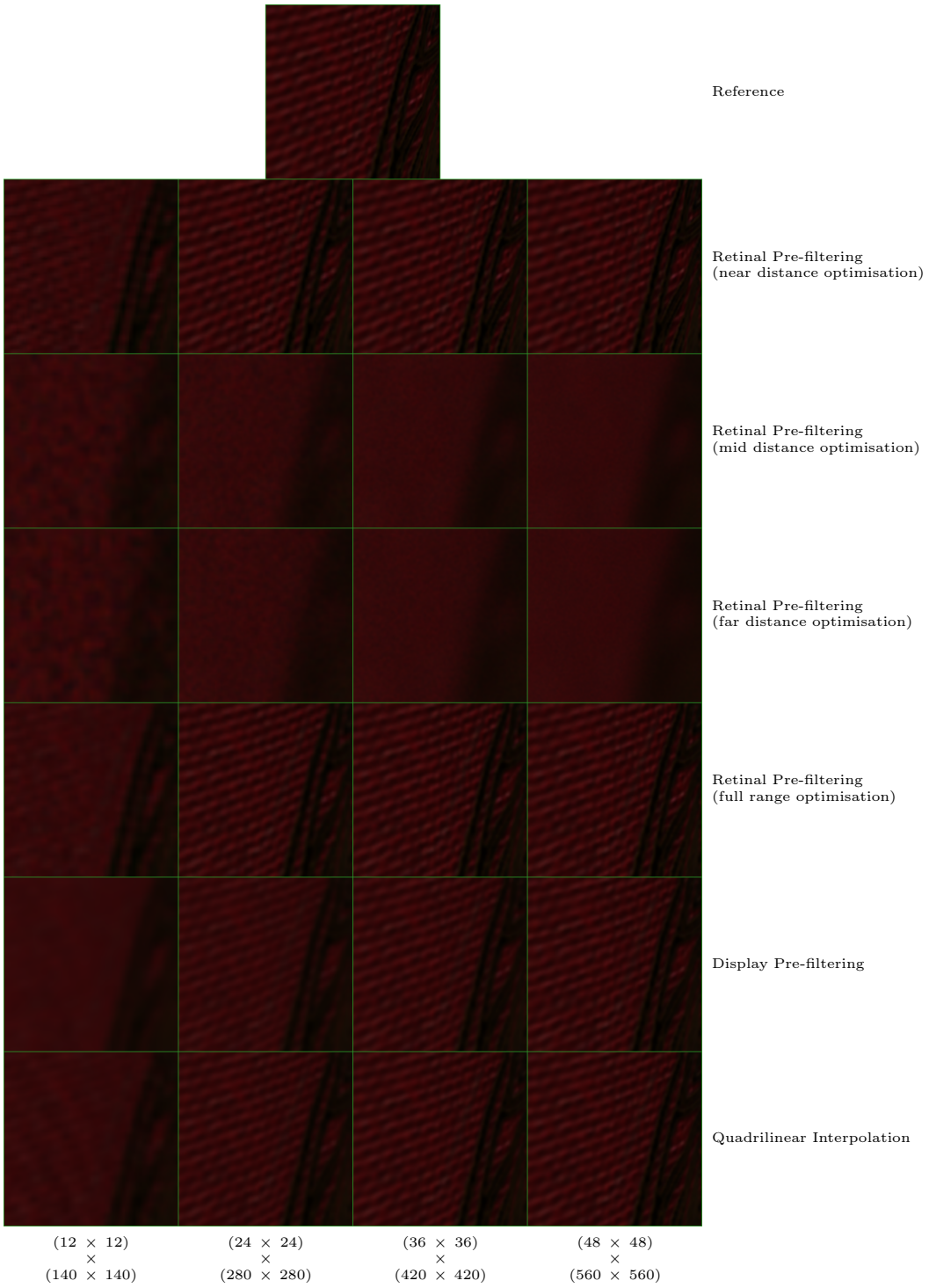


Figure D.9: Comparison of pre-filtering methods for the near insert of the Sponza scene with different display resolutions.

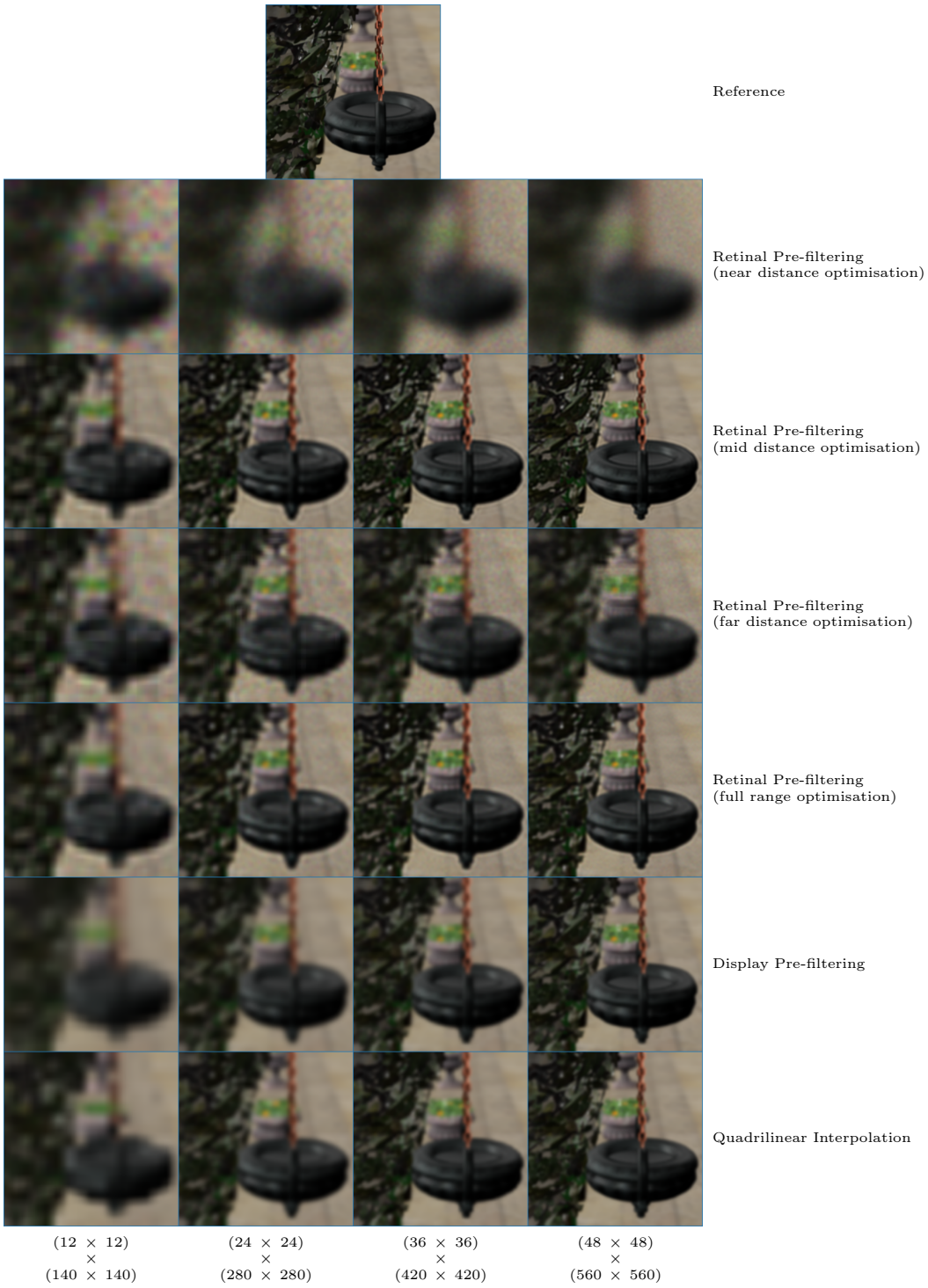


Figure D.10: Comparison of pre-filtering methods for the mid insert of the Sponza scene with different display resolutions.

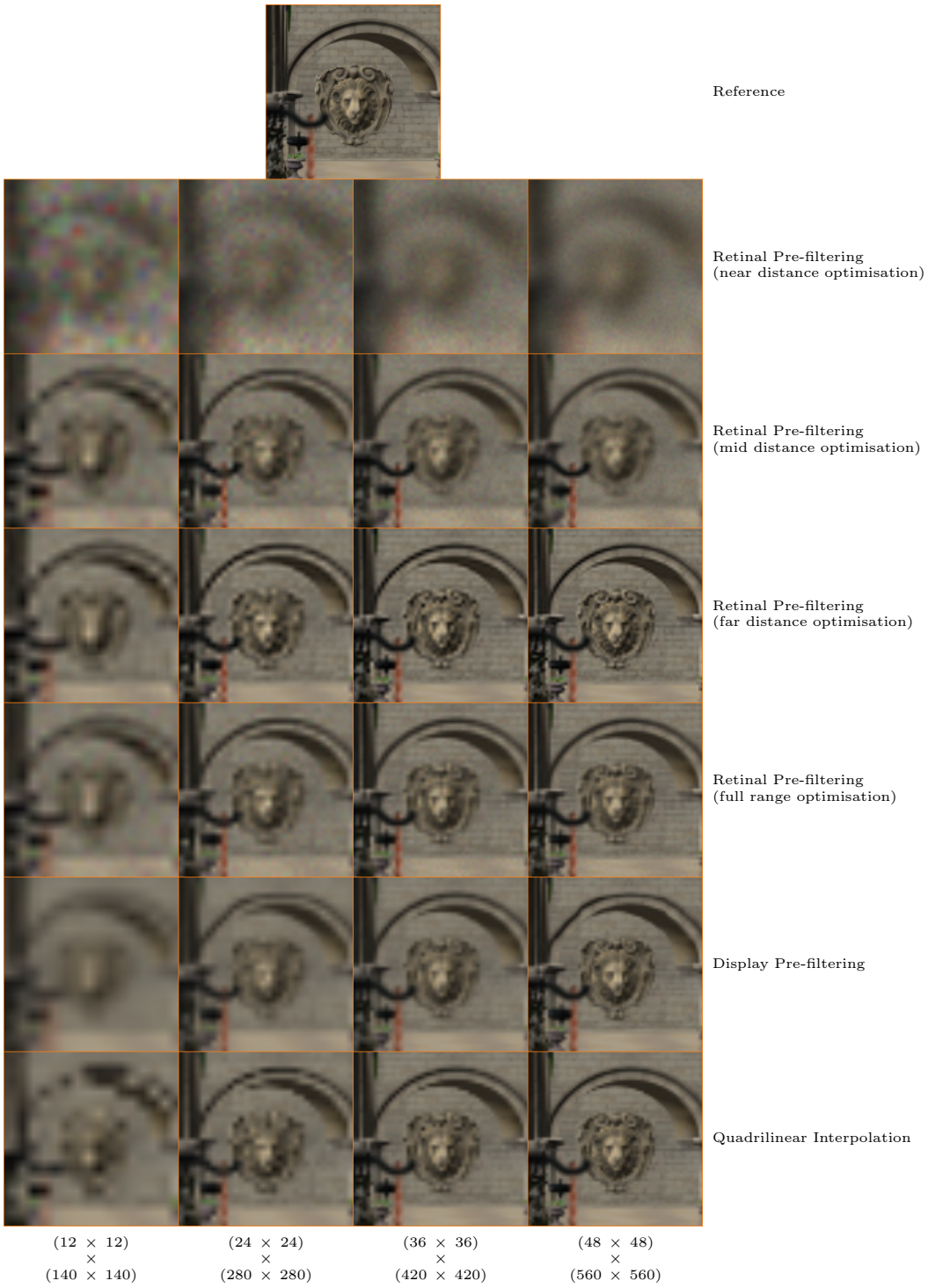


Figure D.11: Comparison of pre-filtering methods for the far insert of the Sponza scene with different display resolutions.

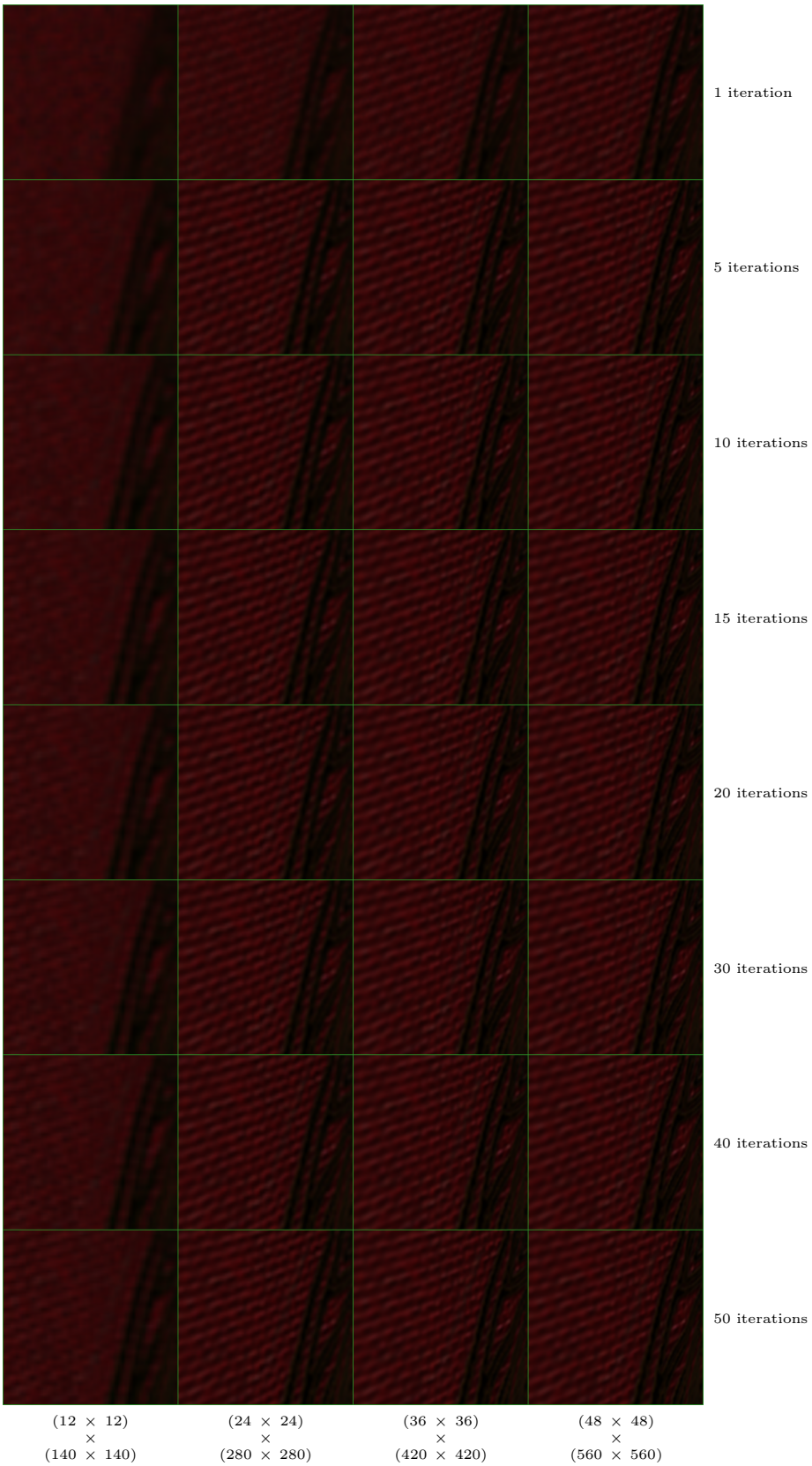


Figure D.12: Incremental improvements of retinal pre-filtering (near distance) across multiple iterations for the near insert of the Sponza scene with different display resolutions.

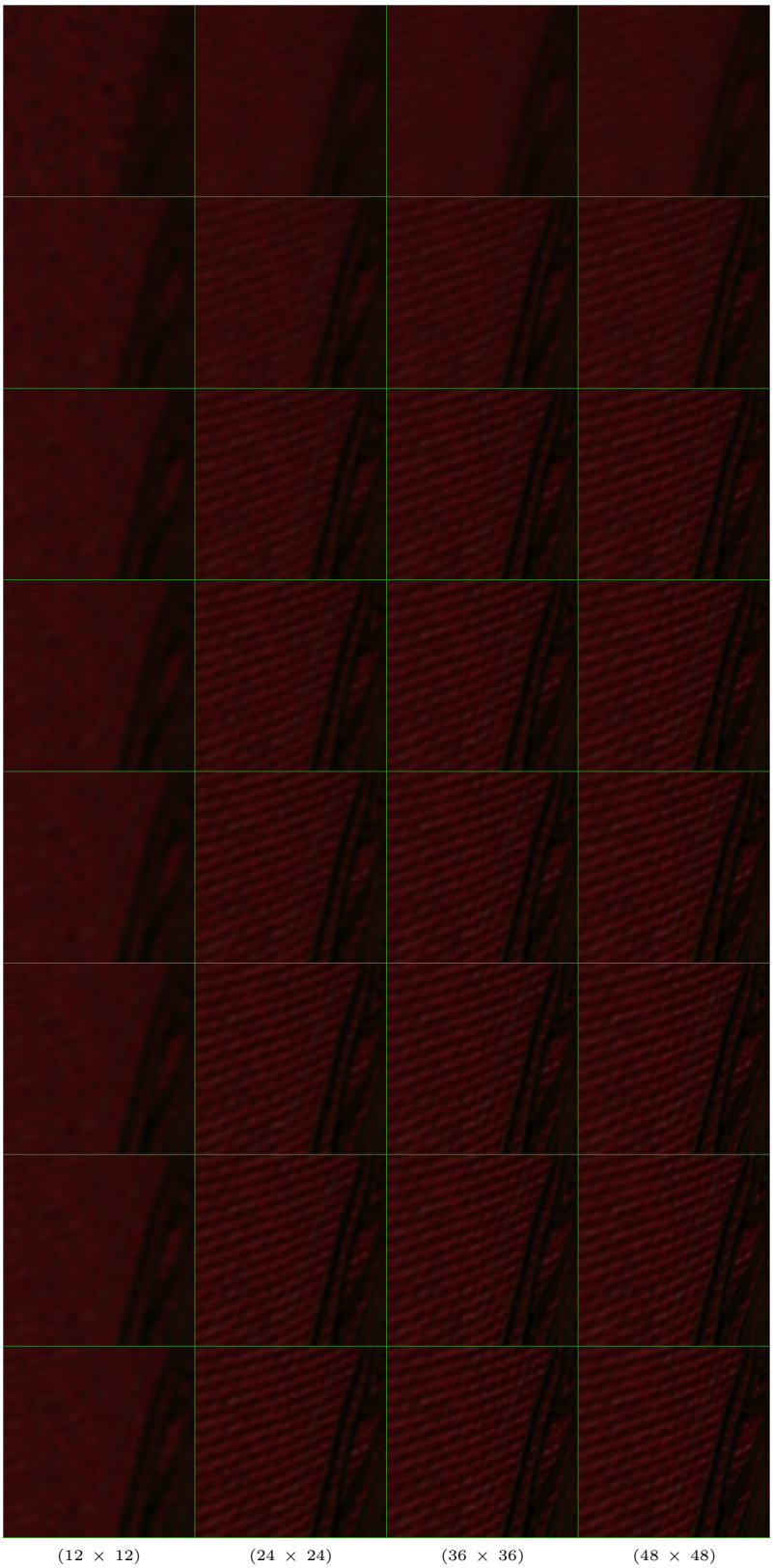


Figure D.13: Incremental improvements of retinal pre-filtering (full range) across multiple iterations for the near insert of the Sponza scene with different display resolutions.

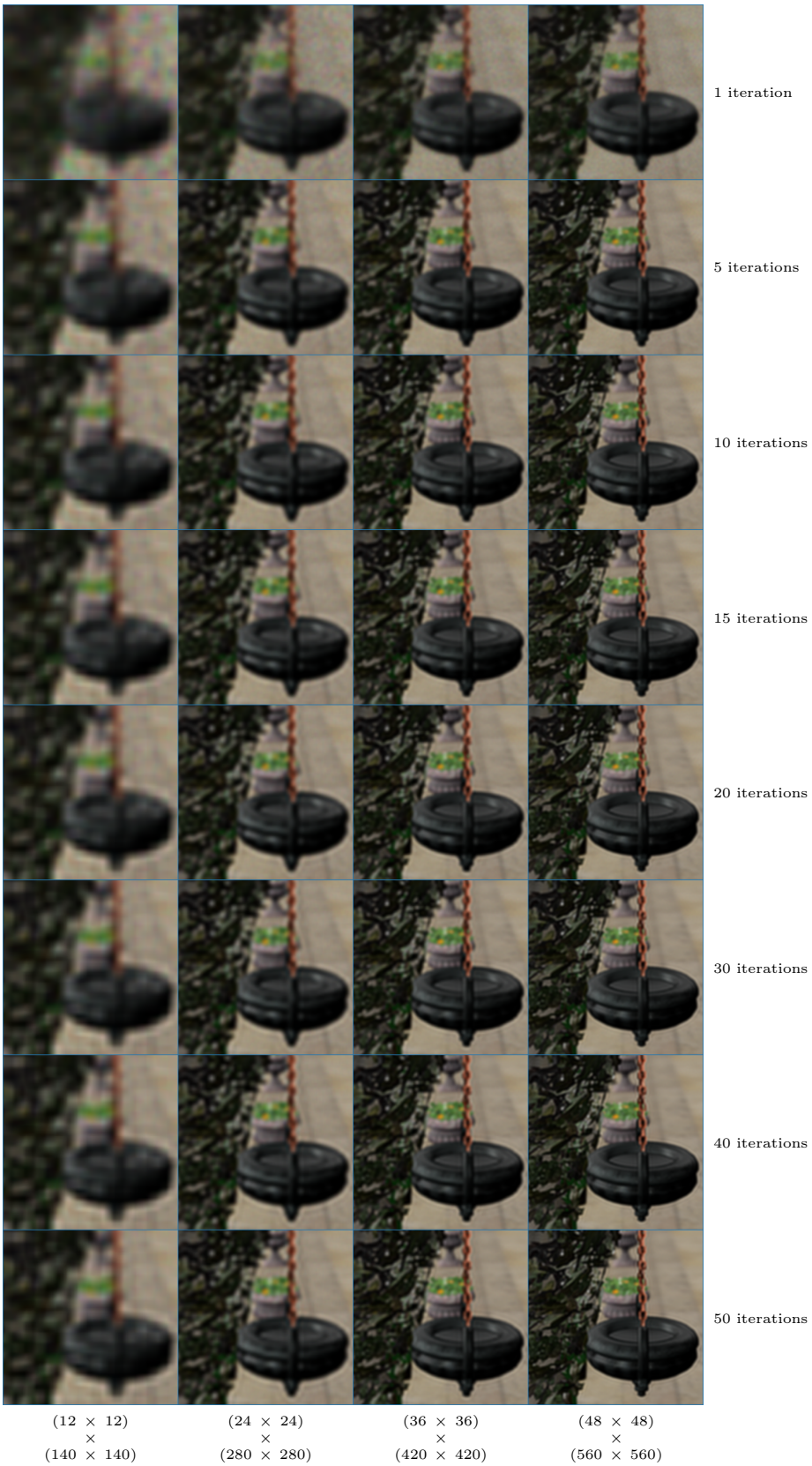


Figure D.14: Incremental improvements of retinal pre-filtering (mid distance) across multiple iterations for the mid insert of the Sponza scene with different display resolutions.

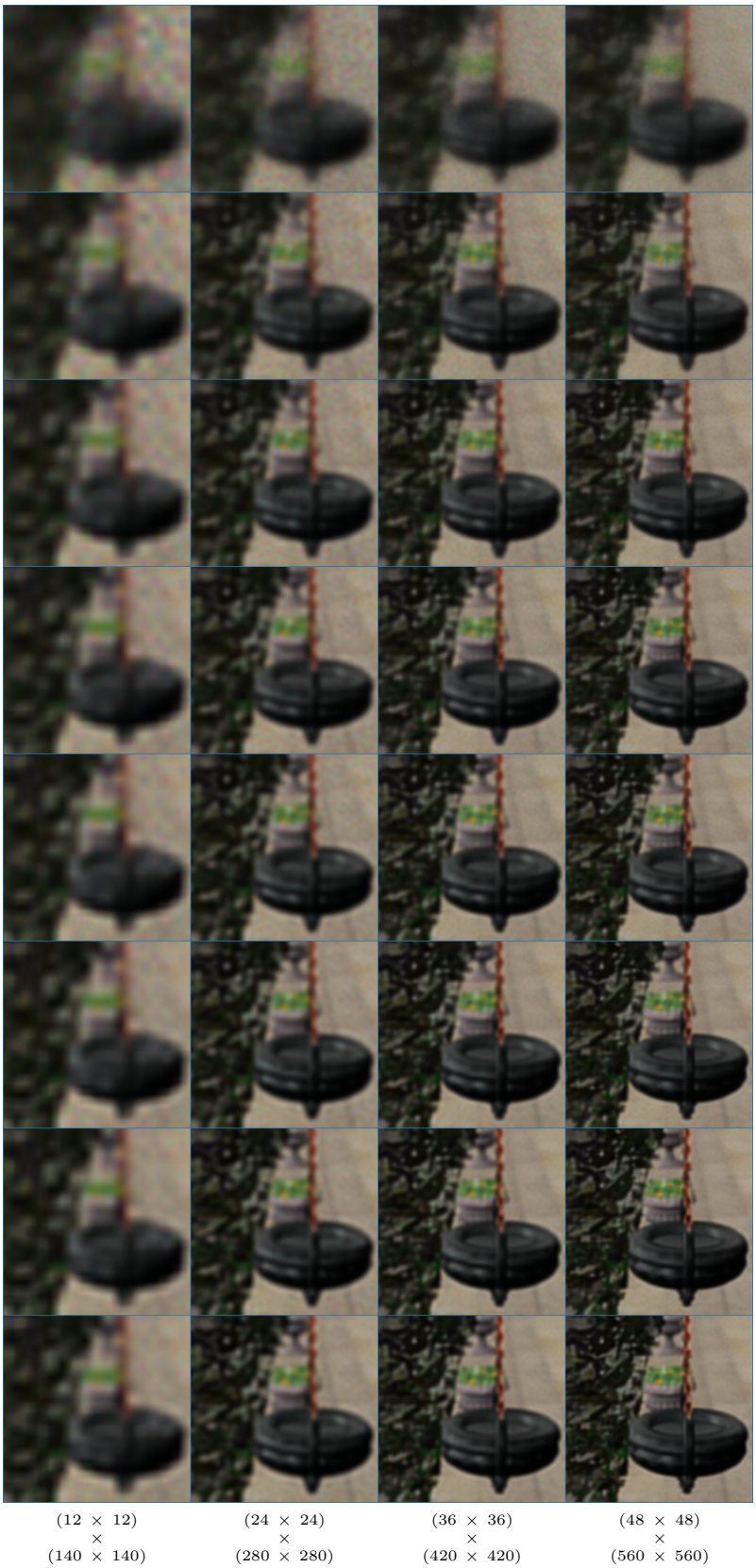


Figure D.15: Incremental improvements of retinal pre-filtering (full range) across multiple iterations for the mid insert of the Sponza scene with different display resolutions.

Figure D.16: Incremental improvements of retinal pre-filtering (far distance) across multiple iterations for the far insert of the Sponza scene with different display resolutions.

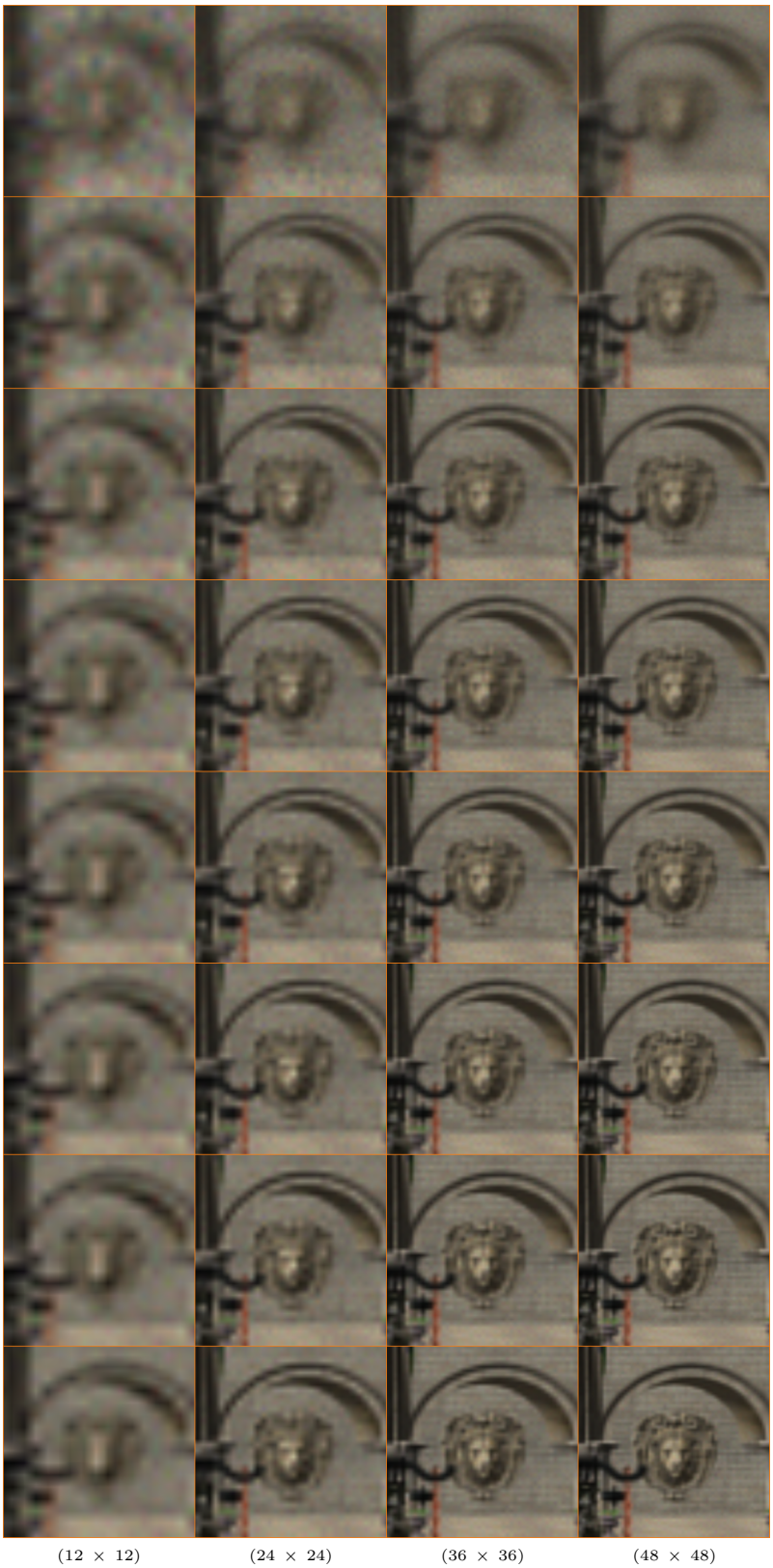


Figure D.17: Incremental improvements of retinal pre-filtering (full range) across multiple iterations for the far insert of the Sponza scene with different display resolutions.