Noise-based Enhancement for Foveated Rendering

TAIMOOR TARIQ, Università della Svizzera italiana, Switzerland
CARA TURSUN, Università della Svizzera italiana, Switzerland and University of Groningen, Netherlands
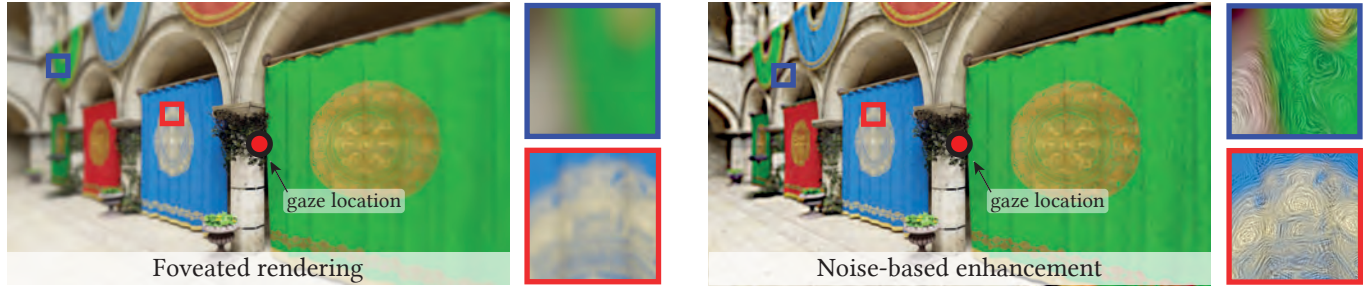PIOTR DIDYK, Università della Svizzera italiana, Switzerland

Fig. 1. Our method enhances the output of standard foveated rendering (left) by adding carefully tuned procedural noise (right). The enhancement replaces accurate rendering of spatial details, which are detectable but not resolvable by the human visual system, and allows applying more aggressive foveation during the rendering process. This images are designed to be viewed at an 80-degree wide field-of-view, with gaze at the center. Scene courtesy of Morgan McGuire and Frank Meinl, Crytek.

Human visual sensitivity to spatial details declines towards the periphery. Novel image synthesis techniques, so-called foveated rendering, exploit this observation and reduce the spatial resolution of synthesized images for the periphery, avoiding the synthesis of high-spatial-frequency details that are costly to generate but not perceived by a viewer. However, contemporary techniques do not make a clear distinction between the range of spatial frequencies that must be reproduced and those that can be omitted. For a given eccentricity, there is a range of frequencies that are detectable but not resolvable. While the accurate reproduction of these frequencies is not required, an observer can detect their absence if completely omitted. We use this observation to improve the performance of existing foveated rendering techniques. We demonstrate that this specific range of frequencies can be efficiently replaced with procedural noise whose parameters are carefully tuned to image content and human perception. Consequently, these frequencies do not have to be synthesized during rendering, allowing more aggressive foveation, and they can be replaced by noise generated in a less expensive post-processing step, leading to improved performance of the rendering system. Our main contribution is a perceptually-inspired technique for deriving the parameters of the noise required for the enhancement and its calibration. The method operates on rendering output and runs at rates exceeding 200 FPS at 4K resolution, making it suitable for integration with real-time foveated rendering systems for VR and AR devices. We validate our results and compare them to the existing contrast enhancement technique in user experiments.

Authors' addresses: Taimoor Tariq, Università della Svizzera italiana, Lugano, Switzerland, tariqt@usi.ch; Cara Tursun, Università della Svizzera italiana, Switzerland and University of Groningen, Groningen, Netherlands, tursuo@usi.ch; Piotr Didyk, Università della Svizzera italiana, Lugano, Switzerland, piotr.didyk@usi.ch.

1 INTRODUCTION

The quality of rendered images directly impacts user experience, immersion, and comfort. While today's computer graphics techniques offer endless opportunities to improve image quality, either by more accurate lighting simulation or modeling of the virtual world, additional constraints, such as frame rate or power efficiency, can significantly limit image quality. Rendering efficiency is critical in the context of new virtual and augmented reality headsets that require rendering systems to operate at high spatial and temporal resolutions. Therefore, it is essential to save computational resources whenever possible, ideally matching the quality of the images with the capabilities of human perception.

Significant computational savings can be achieved when the rendering quality is guided by the estimation of viewers' gaze location provided by an eye-tracking device. So-called foveated rendering techniques [Mohanto et al. 2021] reduce the rendering quality according to the decay in the sensitivity of the human visual system (HVS) to image distortions in the periphery. The most common approach is to reduce the shading rate or spatial resolution for image regions distant from the gaze location. This process aims to omit the costly synthesis of high-spatial frequency details irrelevant for human perception. The benefits of foveated rendering are
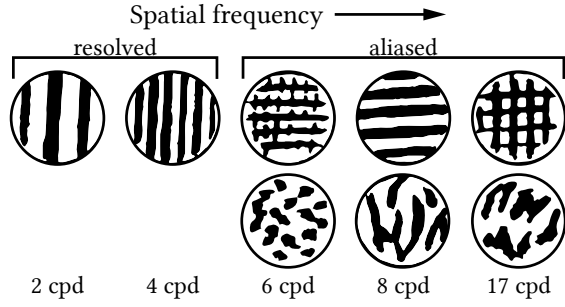
Spatial frequency ⟶

Fig. 2. Subjective appearance of gratings located at 20° retinal eccentricity in the nasal visual field. Redrawn from Thibos *et al.* [1998].
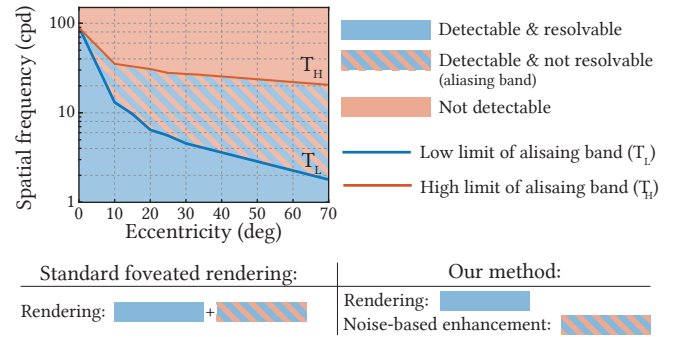

Fig. 3. Characterization of spatial acuity for peripheral vision. Resolution acuity, detection acuity, and the region of spatial aliasing are shown as a function of retinal eccentricity [Thibos et al. 1996].

already being exploited by the latest VR displays such as the Sony Play-Station VR2[1] and Varjo VR-3[2].

Unfortunately, there is no clear distinction between the spatial frequencies relevant to a human observer and the spatial frequencies that can be excluded from the rendering procedure. In fact, there is a range of spatial frequencies that are detectable by the HVS but not resolvable, i.e., the human sensitivity to its spatial localization and orientation is low [Thibos et al. 1987a]. This means that even though these frequencies do not need to be accurately synthesized, they cannot be omitted in the process of foveated rendering since this would lead to visible quality degradation (Figure 2). As a result, to provide images perceptually equivalent to full-resolution rendering, foveated rendering techniques must synthesize this frequency range.

Inspired by the existence of spatial frequencies that are detectable but not resolvable, we argue that a direct rendering of this range of spatial details is not necessary and can be replaced by an inexpensive procedural noise synthesis. Consequently, we propose a foveated rendering procedure where we first use foveated rendering to synthesize the image content that falls into the spatial frequency range which is both detectable and resolvable by a human observer. Next, we use the synthesized information to estimate the parameters of the noise, which we synthesize and combine with the previously rendered content in a post-processing enhancement step. Figure 3 shows the human visual acuity regions, and how they are addressed by different rendering approaches. The main contribution of our paper lies in the derivation of the method for estimating the noise parameters, based on the underlying content. Our procedure is based on the perceptual literature and consists of several simple image processing steps, making it efficient to implement and temporally coherent. The method is calibrated by a series of perceptual experiments to fine-tune the range of spatial frequencies added with the noise and their amplitudes. In additional evaluation experiments, we demonstrate that our noise-based enhancement step can successfully improve the fidelity of aggressive foveated rendering. We also compare our method to the recent work on contrast enhancement for foveated rendering [Patney et al. 2016]. Code for our technique can be found at https://github.com/taimoor6864/Noised-Foveation.

2 BACKGROUND AND RELATED WORK

In this section, we introduce the previous studies on the characteristics of peripheral vision, which form the basis of our work (Section 2.1). We discuss the existing applications of foveated rendering (Section 2.2) and introduce works on metamerism (Section 2.3).

2.1 Peripheral Vision

While observing our environment if we fixate on a point with our eyes, we notice that we cannot see everything at high acuity in our visual field [Rosenholtz 2016; Strasburger et al. 2011]. This is due to a decreasing limit on the visual acuity as the distance from the point of fixation increases [Aubert and Foerster 1857; Hering 1899]. Our eyes feature a central high acuity region called *fovea* that is approximately 2° wide and *peripheral vision* that covers whole hemisphere outside fovea with low acuity [Lettvin et al. 1976]. The decreasing peripheral visual acuity was systematically measured using square-wave spatial contrast gratings at different retinal eccentricities [Wertheim 1894] and those measurements were initially related to the sampling rate determined by the density of retinal rod and cone type of photoreceptor cells [Curcio et al. 1990; Fick 1898; Hirsch and Curcio 1989; Levi and Klein 1986; Østerberg 1935]. More recent studies show that the contrast sensitivity in the peripheral vision can be as high as those measured at the fovea, provided that the stimulus size is magnified depending on an estimate of ganglion cell density in the periphery [Legge and Kersten 1987; Rossi and Roorda 2010; Rovamo and Virsu 1979; Rovamo et al. 1978; Virsu et al. 1987]. These studies on the retinal anatomy have led to models of peripheral sensitivity for computing visibility across the visual field [Haun 2021; Schütt and Wichmann 2017; Watson 2018; Watson and Ahumada 2016].

The neural resolution of the photoreceptor cells is only one of the factors defining the spatial frequency limits of the human eye. The other factor is the optical system mainly represented by the cornea and the lens. The neural resolution limits were measured in isolation from optical aberrations using an interferometer for both fovea and periphery [Thibos and Walsh 1985; Williams 1985a,b]. Although both the optical quality and neural resolution decline with retinal eccentricity, the decline in neural resolution is more significant than the optical limit [Anderson et al. 1991; Green 1970; Millodot

et al. 1975]. As a result, the photoreceptors in peripheral vision may get stimulated by spatial frequencies beyond their resolution limit, which manifests itself as aliasing [Thibos et al. 1987b]. Beyond resolvable spatial frequency levels, the presence of the stimuli in the peripheral vision can still be detected; however, low-level details are not resolvable (e.g., orientation of isolated gratings) [Anderson et al. 2002]. In experiments with filtered images, this may lead to participants reporting a subjective visual change in the image without being able to define what exactly changed [Sere et al. 2000]. Thibos *et al.* [1998; 1987a] studied the perception of signals beyond neural sampling limit and identified the range of spatial frequencies where visual stimuli are detectable but not resolvable (Figure 3).

2.2 Foveated Rendering

Foveated rendering techniques reduce the rendering quality according to the loss of human visual sensitivity in the periphery. Earlier applications involved image-video compression, spatially adaptive ray tracing and tuning parameters such as the resolution and bit-depth [Browder and Chambers 1988; Glenn 1994; Kortum and Geisler 1996; Levoy and Whitaker 1990; Tong and Fisher 1984; Tsumura et al. 1996]. More recently, Guenter *et al.* [2012] demonstrated that eccentricity-dependent shading rate reduction can lead to computational savings without significant quality degradation. Stengel *et al.* [2016] made use of additional information from the geometry pass in rendering, such as depth, normal, and texture properties, to derive the local information on silhouettes, object saliency, and specular highlights from the last frame to reduce the shading rate. Patney *et al.* [2016] introduced a cost-effective method based on contrast enhancement of attenuated spatial frequencies that significantly reduces the visibility of peripheral blur in foveated rendering. Swafford *et al.* [2016] extended the HDR-VDP2 [Mantiuk et al. 2011] quality metric for peripheral vision and introduced a method for tuning ambient occlusion, tessellation, and ray-casting quality. Meng *et al.* [2018] used kernel log-polar mapping for an efficient GPU-based implementation of foveated rendering. Tursun *et al.* [2019] modeled the influence of underlying content on the visibility and proposed a model that takes both the retinal eccentricity and the content into account. Kaplanyan *et al.* [2019] showed an application of neural networks on foveated image reconstruction with very sparse sampling but with a significant computational overhead similar to other neural reconstruction techniques. For a comprehensive survey of foveated rendering methods, please refer to the work of Mohanto *et al.* [2021].

2.3 Visual Metamerism

Visual metamerism refers to the phenomena of physically different images being perceptually indistinguishable. The concept has recently attracted research interest, aimed at finding perceptually equivalent stimuli that are computationally simpler to reconstruct in the context of gaze-contingent applications. Initial studies on the neural representation of vision hypothesized that the role of early vision is to have an efficient representation by removing statistical redundancy in natural images [Barlow 1961; Simoncelli and Olshausen 2001]. Portilla and Simoncelli[2000] showed that a statistical model based on wavelet representations can be used to synthesize visually

plausible natural and artificial textures. Inspired by Portilla and Simoncelli, there have been more recent studies aimed at synthesizing visual metamers using deep convolutional neural network representations [Deza et al. 2019; Freeman and Simoncelli 2011]. Very recently, Walton *et al.* [Walton et al. 2021] exploited spatial pooling in human peripheral vision and proposed a technique for foveated graphics to synthesize visual metamers. The technique requires the full-resolution image statistics; therefore, its application is limited to tasks where the reference image is available, such as foveated compression, but it is not applicable to the real-time foveated rendering pipeline.

A comparison of our noise-based enhancement method to the most relevant works from the state-of-the-art is shown in Table 1. For foveated rendering, it is essential for a method to avoid a preliminary full-resolution rendering pass to compute the input. Among the compared methods, only Walton *et al.* [2021] requires the full-resolution input to compute the image statistics for the output; therefore, it is more appropriate for compression tasks rather than foveated rendering. In addition, a desirable property motivated by the work of Thibos *et al.* [1987a; 1996; 1987b] is enhancing the output by synthesizing novel image details at *detectable & not resolvable band* (Figure 3) that do not already exist in the input image. Among the compared methods, Kaplanyan *et al.* [2019] uses neural reconstruction trained on natural and synthesized images for this task while Walton *et al.* [2021] captures the required statistics from the full-resolution reference input. Although Patney *et al.* [2016] enhance the input image contrast, they do not add novel spatial details to their output. Regarding running times, due to the computational efficiency of the simple contrast enhancement, Patney *et al.* [2016] is the fastest among compared methods. Our method has second place with our computationally efficient real-time GPU implementation. Walton *et al.* [2021] report real-time running times for low resolutions, their application becomes challenging at high resolutions due to the costly image analysis step. Kaplanyan *et al.* [2019] is the most costly method among all compared because their network inference requires 4 GPUs to process a 1920 × 1080 input at interactive rates.

Table 1. Comparison of our method with relevant works from the literature.

Method	Requires full-res input	Spatial-Frequency synthesis	Running time
Patney *et al.* [2016]	No	No	Fastest
Kaplanyan *et al.* [2019]	No	Yes	Slow
Walton *et al.* [2021]	Yes	Yes	Moderate
Ours	No	Yes	Fast

3 METHOD OVERVIEW

Our method enhances the output of adaptive-resolution foveated rendering techniques [Mohanto et al. 2021] by synthesizing novel image details whose spatial frequencies fall into the range of detectable but not resolvable frequencies by the HVS (Figure 3). We show that these details can be synthesized using carefully controlled
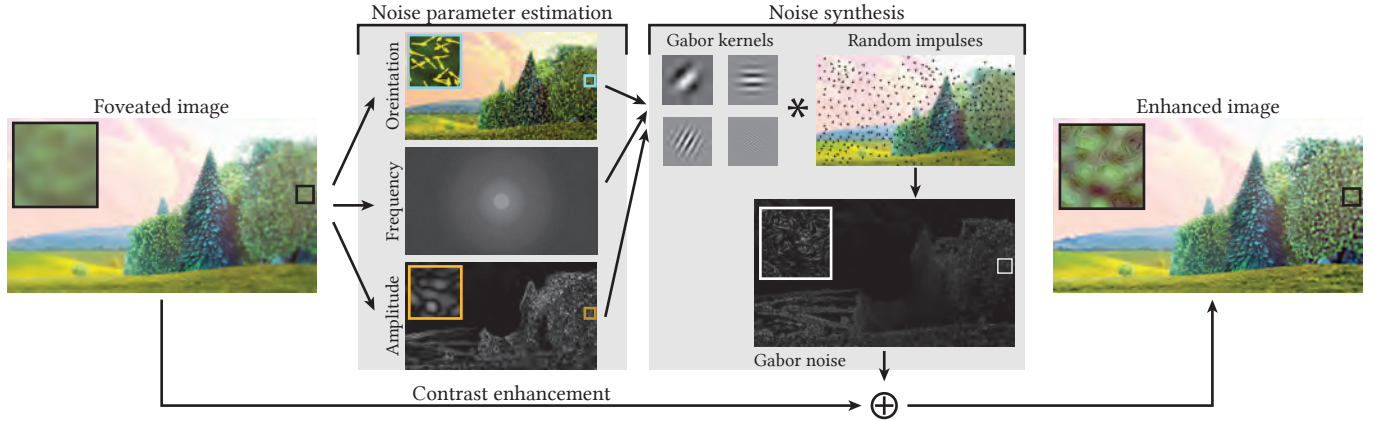
Fig. 4. The figure demonstrates our noise-based enhancement for foveated rendering. The input is a foveated image that is processed to estimate parameters for Gabor noise, i.e., orientation, frequency, and amplitude. The resulting Gabor kernels are convolved with random impulses to synthesize procedural noise. Next, the noise is added to the contrast-enhanced foveated image.

procedural noise. While several types of noise can be considered, we rely on sparse Gabor convolution which provides good frequency control and is computationally efficient for real-time performance [Lagae et al. 2010, 2009]. The main contribution of this work is the perceptually inspired method for estimating the parameters of the noise based on the underlying content. Our method relies solely on the information available in the foveated image and it does not require a full-resolution rendering pass. This makes it suitable for practical real-time foveated rendering applications.

The overview of our noise-based enhancement is depicted in Figure 4. The input to our technique is a single frame from a foveated rendering sequence. We assume that the rendering does not contain significant spatial or temporal aliasing already. Consequently, following the studies of Albert *et al.* [2017], Hoffman *et al.* [2018] and the previous work on foveated rendering [Patney et al. 2016; Tursun et al. 2019], we assume that the resolution reduction due to foveation can be modeled with a spatially-varying Gaussian filter. We denote the standard deviation of Gaussian filter as $\sigma(\mathbf{x})$, where $\mathbf{x} = [x_h \, x_v]$ is the vector of normalized horizontal and vertical image coordinates such that $x_h, x_v \in [0, 1]$. Given the input foveated image and $\sigma(\mathbf{x})$, our technique estimates the local noise parameters: amplitude, spatial frequency, and orientation. The parameters are later used to generate the additive noise for the input. In a separate step, we apply the contrast enhancement [Patney et al. 2016] to the foveated image. The technique effectively enhances the spatial frequencies that have been attenuated but not removed by the foveated rendering. In the final step, we add the generated noise to the contrast-enhanced version of the input image to introduce novel spatial details in the aliasing band.

Below, we describe the method for noise-based enhancement (Section 4) and its calibration to the data collected from perceptual experiments (Section 5). For details of the procedural noise generation and contrast enhancement methods that we use, please refer to the original papers [Lagae et al. 2009; Patney et al. 2016].

4 NOISE-BASED ENHANCEMENT

In our work, we use sparse Gabor convolution to generate the noise [Lagae et al. 2009]. The method convolves Gabor kernels with sparse, random impulses (Figure 4), which is equivalent to placing Gabor patches at random locations across the image. The noise characteristics are controlled by the properties of individual Gabor kernels, i.e., amplitude, spatial frequency, and orientation. In this section, we describe our technique for estimating these parameters from a foveated rendering input. For a given impulse location $\mathbf{x}$, our method estimates the spatial frequency, $f(\mathbf{x})$, amplitude, $K(\mathbf{x})$, and orientation, $\omega(\mathbf{x})$ of the Gabor kernel convolved with this impulse.

It is critical that these parameters are not estimated using a full-resolution image as that would require a full-resolution rendering pass and hinder any potential computational gains from foveated rendering by making it as slow as standard full-resolution rendering. Instead, we assume that there is a sufficient correlation between the low and high-spatial frequency components of an image that allows us to derive the noise parameters based on the foveated image. This assumption is motivated by the fact that the power spectrum of natural images follows a power-law [Hyvärinen et al. 2009; Ruderman 1994]. In addition, such an assumption has already been used in the context of foveated rendering to estimate missing luminance contrast [Tursun et al. 2019]. While it does not hold in all cases (Section 7), we demonstrate its successful application in our scenario.

4.1 Frequency

Our method generates noise that is detectable but not resolvable by the HVS (Figure 3). We follow the study of Thibos *et al.* [1996] which provides the range of spatial frequencies with these properties as a function of eccentricity. We denote the upper limit of the range by $T_H(e_\mathbf{x})$ and the lower limit by $T_L(e_\mathbf{x})$, where $e_\mathbf{x}$ is the retinal eccentricity at location $\mathbf{x}$ in the input image, i.e., the distance from the gaze location measured in visual degrees. While the original work provides only a few measured data points of the limits (Appendix A), we linearly interpolate the measurements to obtain

$T_H(e_\mathbf{x})$ and $T_L(e_\mathbf{x})$ for a wider range of eccentricities. The result of the interpolation is shown in Figure 3.

To respect the limits defined by Thibos et al. [1996], we define the bounds of the spatial frequency of the noise, $f(\mathbf{x})$, as $T_H(e_\mathbf{x})$ and $T_L(e_\mathbf{x})$. Additionally, $f(\mathbf{x})$ should also be bounded by the range of frequencies that are absent from or attenuated by foveated rendering. This additional lower bound on $f(\mathbf{x})$ can be computed from the spatial frequency cut-off point depending on the rendering resolution of foveated rendering or in our case, from the standard deviation of Gaussian kernel, $\sigma(\mathbf{x})$, as we use this low-pass filter to simulate the adaptive-resolution effects of foveation. We use the fact that the amplitude attenuation of a Gaussian blur is also a Gaussian in the frequency domain with standard deviation $\sigma_f = \frac{1}{2\pi\sigma}$, where σ is the standard deviation of the Gaussian in the spatial domain. The Gaussian filter has an infinite-band impulse response and it does not have a sharp frequency cut-off. In our work, we define the cut-off frequency at $3\sigma_f$, where the filter response is approximately 1% of the peak. Consequently, we compute the lower limit on the noise frequency as the maximum of the lower bound from Thibos et al. and the cut-off frequency:

$$F_L(\mathbf{x}) = \max\left(T_L(e_\mathbf{x}), \frac{3}{2\pi\sigma(\mathbf{x})}\right). \tag{1}$$

It is common to define the cut-off frequency at the 50% of the peak filter response in signal processing applications. We make this conservative selection of the cut-off point at $3\sigma_f$ to avoid adding noise to the resolvable frequency range because it would be perceived as undesired distortions to the image by observers. Apart from the frequency spectrum of the foveated image, the noise added to the image must also respect the physical limitations of the display. Consequently, we define the upper limit on the noise frequency to be the minimum of the upper bound from Thibos et al. and the limit of the display:

$$F_H(\mathbf{x}) = \min\left(T_H(e_\mathbf{x}), f_d\right), \tag{2}$$

where f_d is the highest spatial frequency the display can reproduce. According to the Nyquist–Shannon sampling theorem, this limit is equal to 0.5 cycles per pixel. Note that for brevity in our derivation, we omit conversion between different units of spatial frequency, which are either cycles per visual degree or cycles per pixel. The equations hold as long as the units are the same, and the conversion is performed based on the display pixel size and the viewing distance of an observer.

Given the theoretical frequency limits on the noise ($F_L(\mathbf{x})$ and $F_H(\mathbf{x})$), we model the local noise frequency $f(\mathbf{x})$ as a random variable $\chi_\mathbf{x}(f)$. In the early vision, the bandwidth of cortical simple cells is symmetrical in log spatial frequency domain [De Valois et al. 1982]. In order to establish a closer resemblance to this bandwidth symmetry, we define the distribution of $\chi_\mathbf{x}(f)$ as log-normal, with the mean, $\mu_n(\mathbf{x})$, at the midpoint of $F_L(\mathbf{x})$ and $F_H(\mathbf{x})$ in log domain and standard deviation, $\sigma_n(\mathbf{x})$, as the quarter of this range such that the range between $F_H(\mathbf{x})$ and $F_L(\mathbf{x})$ is spanned by $\mu_n(\mathbf{x}) \pm 2\sigma_n(\mathbf{x})$ in log frequency domain. In order to fine tune the bandwidth around $\mu_n(\mathbf{x})$, we additionally define a scaling factor s_f as a free parameter of our model for adjusting $\sigma_n(\mathbf{x})$:

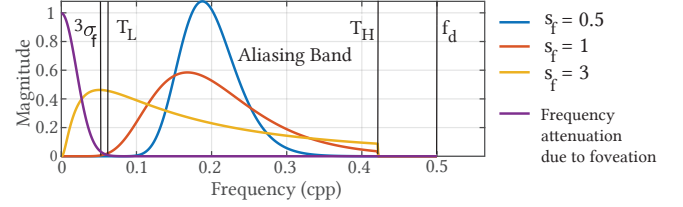$$\mu_n(\mathbf{x}) = 0.5 \cdot (\ln F_L(\mathbf{x}) + \ln F_H(\mathbf{x})), \tag{3}$$

Fig. 5. Examples of different distributions of random variable $\chi_\mathbf{x}(f)$, for Gabor kernels, with all considered limits.

$$\sigma_n(\mathbf{x}) = 0.5 \cdot s_f \cdot (\mu_n(\mathbf{x}) - \ln F_L(\mathbf{x})). \tag{4}$$

Using these mean and standard deviation parameters, we formally define the distribution of $\chi_\mathbf{x}(f)$ as:

$$\chi_\mathbf{x}(f) \sim e^{\mu_n(\mathbf{x}) + \sigma_n(\mathbf{x})Z}, \tag{5}$$

where Z is the normal distribution $\mathcal{N}(0, 1)$. Figure 5 shows all the limits considered when estimating the frequency distribution, $\chi_\mathbf{x}(f)$, and three examples of the distributions for different values of s_f.

We calibrate bandwidth scaling factor, s_f, using the participants' responses from a subjective experiment (Section 5). s_f is a scalar that is independent of location, eccentricity and underlying content as these factors are already considered by $F_L(\mathbf{x})$ and $F_H(\mathbf{x})$ in our model. Since the distribution of $\chi_\mathbf{x}(f)$ has infinite falloff, we additionally truncate it to range $(0, F_H(\mathbf{x}))$. While generating the noise, for every location of an impulse, $\mathbf{x}$, we compute the corresponding spatial frequency $f(\mathbf{x})$ of the Gabor patch by randomly sampling from the truncated log-normal distribution of $\chi_\mathbf{x}(f)$.

4.2 Amplitude

While estimating the local noise amplitude, $K(\mathbf{x})$, we aim to synthesize novel details, which resemble those in full-resolution rendering. The challenge comes from the fact that our noise-based enhancement generates spatial frequencies that are not present in the foveated input. Inspired by previous work [Tursun et al. 2019] and natural image statistics literature (Section 2), we assume that there is a sufficient correlation between low and high-frequency components of an image to infer the amplitude of higher frequencies from lower frequency information. More precisely, we propose to estimate the noise amplitude $K(\mathbf{x})$ based on the amplitude that corresponds to the highest frequency component generated by the foveated rendering around location $\mathbf{x}$.

To efficiently analyze the amplitude of different frequency bands in the input foveated image, we compute its Laplacian pyramid decomposition, L, where $L_l(\mathbf{x})$ is the value at the level l at the location $\mathbf{x}$. We treat each level l as a frequency band with cut-off frequency 2^{-l} cycles per pixel, and the central frequency

$$f^l = 2^{-(l+0.5)}, \tag{6}$$

which is the mean of the cut-off frequencies of two consecutive bands in the log domain. It is important to remember that while the Laplacian pyramid does not provide strict separation between different spatial frequency bands [Peli 1990], it can be efficiently implemented on a GPU using MIP maps.

In order to estimate the amplitude of the noise at location $\mathbf{x}$, we first calculate the highest spatial frequency present in the foveated image. As discussed in Section 4.1, the frequency attenuation due to foveation can be modeled in the frequency domain using a Gaussian function with $\sigma_f(\mathbf{x}) = \frac{1}{2\pi\sigma(\mathbf{x})}$. Given the Gaussian function modeling the frequency attenuation and the attenuation factor a, we can solve for the frequency value that was attenuated with this factor. The solution gives the frequency value:

$$f_c(\mathbf{x}) = \sqrt{-2 \cdot \sigma_f(\mathbf{x}) \cdot \ln(a)} = \sqrt{\frac{-\ln(a)}{\pi\sigma(\mathbf{x})}}. \tag{7}$$

In other words, the above formula estimates the cut-off frequency due to foveation at location $\mathbf{x}$ for a given attenuation cut-off, a, beyond which we consider the spatial frequencies to be absent or unreliable. Furthermore, by combining the above equation with Equation 6, we can compute the Laplacian level $l_a(\mathbf{x})$ that represents the highest reliable spatial frequencies:

$$l_a(\mathbf{x}) = \log_2\left(\sqrt{\frac{-\pi\sigma(\mathbf{x})}{\ln(a)}}\right) - 0.5. \tag{8}$$

We use this equation to infer the amplitude of the noise. We assume that the amplitude of the spatial frequencies omitted in the foveated rendering is directly related to the amplitude of the frequencies that are present in the foveated image. Consequently, we estimate the noise amplitude $K(\mathbf{x})$ as:

$$K(\mathbf{x}) = s_k \cdot |L_l(\mathbf{x})| \quad \text{with} \quad l = l_a(\mathbf{x}), \tag{9}$$

where s_k is a constant parameter that is calibrated in the subjective experiment (Section 5). There is a trade-off in setting the attenuation cut-off, a. Higher values make the estimation of $K(\mathbf{x})$ based on the lower frequencies. These are more reliable, but they may correlate less with the amplitude of high frequencies we seek to estimate. While this correlation is potentially better for low values of a, the content of these frequencies might be less reliable. For all our experiments, we set $a = 0.25$, which provides good trade-off. It also is important to mention that the value $l_a(\mathbf{x})$ is not guaranteed to be an integer. For non-integer values, to compute Equation 8, we perform a linear interpolation between two Laplacian pyramid levels in the log domain.

The above derivation does not account for the fact that adding noise may result in clipping the image values. This is critical for our technique which relies on precise control of the noise frequency spectrum and any untreated clipping may significantly alter it. To prevent changes to the noise frequency spectrum, we propose to further attenuate the estimated noise amplitude according to the minimum and maximum image values in a small neighborhood. More precisely, for a given location $\mathbf{x}$, we compute a minimum, $N_{\min}(\mathbf{x})$, and maximum, $N_{\max}(\mathbf{x})$, image value in a small image neighborhood. Given these values, we can estimate the remaining range of image values available for the noise as $\min(1 - N_{\max}(\mathbf{x}), N_{\min}(\mathbf{x}))$. Using this range, we scale previously estimated amplitude (Equation 9) to respect this range.

Efficient implementation of estimating minimum and maximum image values in a neighborhood is critical for the real-time performance of our technique. Therefore, we build two additional image pyramids of the foveated image to which we add the noise. One pyramid contains local minimum values of all the image pixels located below, while the other pyramid contains corresponding maximum values. Such pyramids can be efficiently constructed subsequently for each level by using the minimum and maximum values of 2×2 neighborhood from the previous level. The resulting pyramids at level l encode minimum and maximum from $2^l \times 2^l$-pixel neighborhood. It is important to note that such a pyramid decomposition is not equivalent to computing exact minimum and maximum values directly at the original resolution, but it serves as a computationally efficient approximation of it. In all our results, we use third pyramid levels, which provide a neighborhood size similar to the spatial support of the Gabor patches used for generating the noise.

4.3 Orientation

Similar to the amplitude estimation, we seek an orientation estimate, $\omega(\mathbf{x})$, that closely follows underlying content. Consequently, we compute it from the lower frequency component of the foveated image. To this end, we use the Sobel operator and convolve the foveated image with horizontal and vertical Sobel filters, which provide corresponding image gradients. Then we use them for estimating the local gradient direction. To ensure spatial and temporal smoothness of the gradients, we apply the orientation estimation to the third level of a Gaussian pyramid of the input foveated image and interpolate the results to the original resolution using bicubic interpolation. The Gaussian pyramid does not introduce any additional computational cost since it is an intermediate step when building the Laplacian pyramid for amplitude estimation and we reuse the result of that computation.

4.4 Noise Generation

The noise generation is based on the work of Lagae *et al.* [2009]. The authors propose a technique that generates Gabor noise by convolving a random set of impulses with Gabor kernels. To enable efficient implementation, the impulses are generated in cells that divide the image into a regular grid. While the size of each cell corresponds to the width of the spatial envelope of the Gabor, the number of impulses in each cell can be adjusted to control the trade-off between the quality of the noise and the computational efficiency. The density of impulses is expressed in their number per kernel, i.e., the area of the truncated Gabor kernel. The authors report that 25-50 impulses provide a good quality noise, but their analysis suggests that even lower numbers provide satisfactory results.

Our generation of noise follows the above procedure closely. We chose the size of our Gabor kernels to accommodate the lowest spatial frequency we aim to generate, which resulted in setting the width of the Gabor patch in the frequency domain to be 0.06 cycles per pixel. The size of the Gabor patch also determined the size of each cell. For all our experiments, we used 64 impulses per kernel unless mentioned otherwise. We use the same random number generator as the original implementation for generating impulses. Before generating the noise, each impulse at location $\mathbf{x}$, is assigned a Gabor kernel parameters: frequency, $f(\mathbf{x})$, amplitude $K(\mathbf{x})$, and orientation $\omega(\mathbf{x})$, estimated using our method described above. To avoid temporal instabilities, the locations of impulses throughout

the entire input image sequence are the same. This is achieved by using a constant seed value for the random number generator. Constant positions of impulses ensure that the Gabor patches do not change their position. Only their parameters change smoothly according to the underlying motion in the content.

4.5 Implementation

We implemented two versions of our enhancement method. The prototype version was implemented in MATLAB, and the real-time counterpart was implemented in OpenGL. As an input, both implementations take a foveated image together with all display and viewing parameters from accurate conversions of spatial frequencies. The output is an enhanced version of the foveated image.

Our prototype in MATLAB relies on a straightforward implementation of the method for estimating the noise parameters. The noise generation is performed using the C++ implementation provided by the authors of the original paper [Lagae et al. 2009]. Our OpenGL implementation takes advantage of the simplicity of all the operations, which can be performed in a per-pixel fashion. The method, including the noise generation, is implemented as a series of GLSL fragment shaders. We leverage the MIP maps to efficiently implement the construction of all image pyramids used in our technique.

In all our experiments, we generate the input foveated images with the fovea region radius equal to 8 visual degrees. Beyond this region, we assumed the resolution fall-off to be linear as a function of eccentricity and defined by blur rate. The foveation was simulated using spatially-varying Gaussian blur.

Our enhancement method processes only luminance information. Consequently, the input color images are first converted to linearized luminance values which are then processed by our technique. The resulting noise contains only luminance information that is added to the input color images after gamma correction.

5 CALIBRATION

Our enhancement method contains three free parameters that allow us to fine-tune the technique. The first two are the scaling factor for the amplitude of noise, s_k, and the scaling factor for the noise bandwidth, s_f. The third parameter is the parameter of the contrast enhancement method [Patney et al. 2016], f_e, which scales the amount of high-frequency content added to the input foveated image. While f_e is fixed in the original work, the authors mention that it can be adjusted. Below, we describe the subjective experiment for estimating the values of s_k, s_f, and f_e. With this experiment, we aim to find the parameters such that the result of our enhancement to closely matches the full-resolution rendering. In our experiments, we assume that their optimal values solely depend on the strength of the foveation. Figure 6 shows the effect of our calibration parameters on the spatial frequency spectrum of our enhanced image. Figure 6 (left) shows that increasing f_e increases the amount of contrast enhancement. However, as expected, the contrast enhancement has very little influence on high spatial frequencies that are lost due to foveation. Figure 6 (middle) shows that increasing s_k increases the energy of the frequencies we synthesize. Figure 6 (right) shows that increasing s_f increases the bandwidth of the distribution of
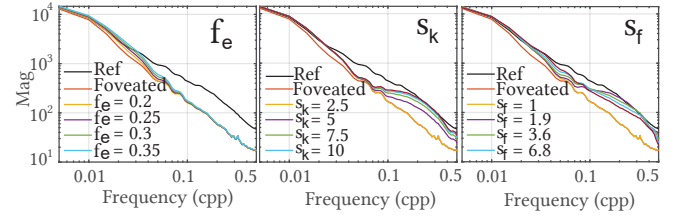

Fig. 6. Effect of calibration parameters on our synthesis (Sponza scene). Left: Effect of varying f_e on the spatial frequency spectrum of the contrast enhanced image. Middle: Effect of varying s_k on the spatial frequency spectrum of the final enhanced image for fixed f_e and s_f. Right: Effect of varying s_f on the spatial frequency spectrum of our final enhanced image for fixed f_e and s_k.
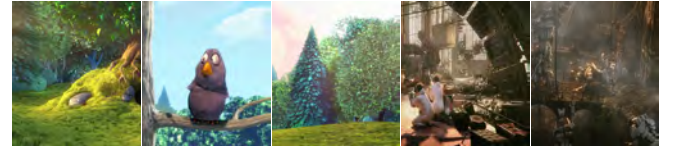

Fig. 7. Representative stimuli used in the calibration experiment.

frequencies we sample from and a wider range of frequencies are synthesized.

5.1 Experiment Setup

Hardware. The experiment was conducted using LG OLED55CX. The 55-inch screen has 4K spatial and 120Hz temporal resolution. The peak luminance of the display was set to 167.33 cd/m^2. The position of the participant's head was fixed using a chin-rest at a 71.5cm distance from the screen. This allowed covering 80° field of view and display of spatial frequencies up to 24cpd at the center of the screen. We used the Tobii Pro Spectrum 600Hz eye tracker to track the participants' gaze location. All the experiments were conducted under diffuse office illumination behind a dark curtain to prevent any reflections on the screen. The experimental setup is presented in Figure 10.

Stimuli. We selected five 4K images from two rendered video sequences: Big Buck Bunny[3] and Tears of Steel[4] (Figure 7). We chose the images such that they include both dark and bright scenes with high-contrast and high-spatial-frequency details.

Task. In each trial, the participant's task was to adjust one of the parameters of a technique such that it best matches the reference full-resolution rendering. To this end, participants were shown half of the full-resolution image on the left. A mirrored version of the reference image was foveated using the adjusted parameter value and shown on the right. The participants were able to manipulate one of the parameters of our technique using a mouse wheel, and the image on the right (test image) was updated in real-time during the experiment to reflect the result of changes in the value. Participants were asked to adjust the parameters such that the foveated image
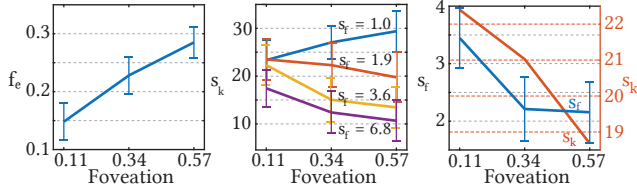
---

[3]https://peach.blender.org
[4]https://mango.blender.org

Fig. 8. Results of the calibration experiment. *Left*: contrast enhancement parameter (f_e), middle: amplitude scaling (s_k) for different values of the noise bandwidth parameter (s_f), and *Right*: the final combinations of amplitude scaling (s_k) and noise bandwidth parameter (s_f). The error bars visualize the standard error of the mean (SEM) for each measurement point.

on the right best matches the full-resolution rendering on the left, and confirm their choice by hitting the Enter key on the keyboard. During the experiment, participants were asked to keep looking at the center of the screen marked with a cross and their gaze position was constantly monitored with the eye tracker. In order to ensure the correct retinal positioning of the stimuli during the experiment, stimuli are hidden by rendering a black frame whenever participants' gaze position moved outside a small circular region with a radius $1.25°$ at the center of the screen.

Participants. University students were recruited for this experiment. Prior to the experiment, they were naive about the purpose of the experiment. All of them had normal or corrected-to-normal vision and received financial compensation for their participation. The experiment procedure was approved by the ethics committee of the hosting institution.

5.2 Experiment Procedure

Each procedure of adjusting a parameter is performed separately for a different amount of foveation. For adaptive-resolution foveated rendering techniques, the cutoff point of spatial frequencies shift to lower frequencies as the eccentricity increases as a result of lower resolution rendering in the periphery. In order to simulate this effect with our low pass Gaussian filter, we linearly increase the standard deviation of the Gaussian kernel with increasing eccentricity. We refer to the scalar representing this linear increase as *blur-rate* in arcmin-per-degree units, and we represent it with σ_{BlurRate}. We define the foveal region with $8°$ radius around the gaze position where the content is rendered in full resolution. The increase in kernel standard deviation at $8°$ eccentricity and beyond. We considered three levels of foveation for data collection: $\sigma_{\text{BlurRate}} = 0.11, 0.34, 0.57$, based on the previous analysis of Tursun *et al.* [2019], which reported percentiles of foveation detectability for different blur-rates. We included the limits so that they may encompass variation of detectability across scenes.

Contrast enhancement parameter. We first investigated the parameter of the contrast enhancement step [Patney et al. 2016], f_e, which scales the amount of added high-frequency signal to the foveated image. Here, we consider the contrast enhancement alone; therefore, the test image shown on the right did not contain additional noise. The participants could freely manipulate the parameter f_e within

the range $(0, 0.4)$. A total of five people performed the adjustment for each of the five images (Figure 8, left).

Amplitude scaling. We used the estimated f_e values to investigate the remaining parameters, amplitude and bandwidth scaling. Again, our design choice was to study one parameter at a time. Therefore, in the next step we asked participants to adjust s_k parameter within the range $(0, 45)$ for a small set of $s_f = 1, 1.9, 3.61, 6.81$. In total, 10 participants adjusted the s_k parameter for all combinations of five images, three s_f values and three foveation levels (Figure 8, middle).

Bandwidth scaling. The previous experiments provided an optimal scaling factor for contrast enhancement as well as a relation between optimal amplitude, s_k, and bandwidth scaling, s_f, for different levels of foveation, σ_{fov}. To choose the final parameters for our method across different foveation levels, we ran a final experiment where participants adjust the bandwidth scaling factor, s_f, and the amplitude, s_k, was computed by linearly interpolating the values from the previous experiment. Again, the total of 10 participants performed this experiment (Figure 8, right).

5.3 Results and Discussion

Figure 8 shows results obtained in the three experiments described above. The range of values obtained for the contrast enhancement parameter (left plot) matches the default parameter value, 0.2, reported by Patney *et al.* [2016]. The users prefer more aggressive contrast enhancement (higher f_e values) as foveation increases, which can be explained by more visible loss of spatial details due to the foveation. The participants choose higher amplitude scaling (s_k), for our noise enhancement as the bandwidth parameter (s_f) decreases (middle plot). This is explained directly by the fact that the increase of the bandwidth parameter widens the spectrum of the noise beyond the aliasing range, which makes it distinguishable from the image content in the reference image. The participants compensated for this by lowering the noise amplitude. The final parameters s_k and s_f (right plot) demonstrate that the spectrum of the noise has to be restricted (lower s_f values) more as the foveation increases. The same trend holds for corresponding amplitude scaling (s_k), even though the relative change in the parameter is less significant. We attribute this trend to the fact that for more aggressive foveation, the range of missing high-frequency details which are easily resolvable increases. Any attempt of compensating for this loss with noise makes it visible due to the ability of the HVS to resolve the details. Therefore, the participants prefer to limit the noise-based enhancement. All the final parameters of our technique for each considered foveation level are listed in Table 2. We provide a detailed statistical analysis of the user-data obtained from calibration in the supplementary.

Table 2. The optimal values of our parameters after calibration.

Foveation (σ_{BlurRate})	f_e	s_k	s_f
0.11	0.15	22.4	3.45
0.34	0.23	21.02	2.21
0.57	0.28	18.68	2.19

Fig. 9. Representative images of the stimuli used in the validation experiment. San miguel and Vokselia spawn scenes courtesy of : https://casual-effects.com/data/
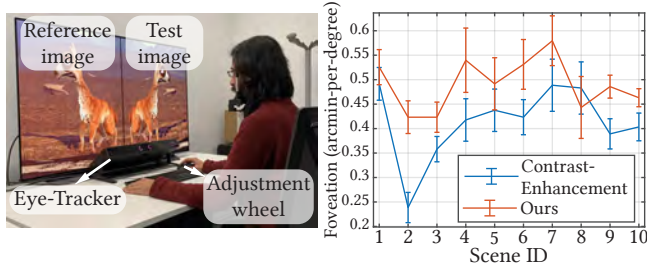

Fig. 10. The experimental setup of our subjective experiments is shown on the left. The average foveation strength (with higher value representing lower spatial frequency cutoff) selected by the participants in the validation experiment (Section 6.1) for each scene is shown on the right. Overall, our enhancement (orange line) allows using more aggressive foveation without undesired visual artifacts compared to contrast enhancement (blue line) ($p < 0.001$, t-test). The error bars visualize the standard error of the mean (SEM) for each measurement point.

6 EVALUATION

We evaluate the effectiveness of our noise-based enhancement with a subjective experiment. The main goal was to demonstrate that subjects can tolerate more aggressive foveation when our noise-based enhancement is used in a post-processing step. We compare our method to contrast enhancement [Patney et al. 2016], which to the best of our knowledge, is the only technique with similar goals (Section 2), i.e., low-cost enhancement of foveated rendering.

6.1 Subjective Experiment

Stimuli. We used 10 rendered scenes to create results for contrast enhancement and the combination of our enhancement with the contrast enhancement. We selected the stimuli manually to include scenes with different brightness levels, contrast, and luminance spatial frequencies (Figure 9). For both techniques and for all results, we used the same parameters determined in the previous calibration experiments (Section 5). We simulated adaptive-resolution foveation using spatially-varying Gaussian blur whose standard deviation parameter depends on the visual eccentricity similar to Section 5.2. In Figure 11, we provide foveation and enhancement results of four scenes for visual comparison (please refer to the supplemental materials for other scenes).

Task. The task of each participant was to select the blur-rate at which the loss of image details due to foveation became visible with respect to the reference (full-resolution) image. To this end, in each trial, we showed the original image on the left half of the display,

while the right half was used for showing the output from the evaluated techniques (i.e., our method and contrast enhancement [Patney et al. 2016]). The stimuli were randomized during the experiment and the participants were not aware of the enhancement method being shown in any trial. The participants were asked to keep looking at a cross at the center of the screen during the experiment. The eye-tracker is used for monitoring the gaze position throughout the experiment and ensuring that they adhere to the instructions (similar to our calibration experiment in Section 5). The participants were asked to adjust the blur-rate using the mouse-wheel (initially starting with no foveation) and press enter at the highest blur-rate at which they could not perceive foveation with respect to the reference image on the right.

Hardware. The hardware setup was identical to the equipment used in our calibration experiment (Section 5).

Participants. 10 university students (3 females, 7 males) with normal or corrected-to-normal vision, were recruited for the experiment. Prior to the experiment, the participants were given a written description of the experimental procedure. In addition, we explained the general concept of foveated rendering and showed a sample adaptive-resolution foveated image generated using sub-sampling and nearest-neighbor interpolation. In order to avoid introducing bias to their preferences, the sample image and the foveation method is chosen different from the techniques evaluated in the experiment. The participants received financial compensation after completing the experiment. The whole experiment took approximately 15 minutes and was completed in a single session by each participant.

Results. The average values of σ_{BlurRate} selected by the participants for each scene are shown in Figure 10. The selected σ_{BlurRate} values represent how "aggressively" high-spatial-frequency details can be removed from the reference full-resolution image by foveated rendering before the loss of high-frequency details become visible to the observers. In comparison to contrast enhancement, our technique allows for higher σ_{BlurRate} without visible artifacts ($p < 0.001$, t-test). The results demonstrate that our method reduces the amount of perceived loss of spatial details from foveated rendering and enables rendering in lower resolution in the periphery.

We provide an analysis of spectral characteristics of our outputs in Figure 12. As expected, foveation effectively removes higher-spatial frequencies which cannot be properly recovered through contrast enhancement, especially in the Aliasing Band ($f \geq T_L$). The attenuation increases with eccentricity and blur-rate, σ_{BlurRate}. This analysis confirms that our technique is effectively able to compensate for the lost high-frequencies. Interestingly, the noise sometimes over-compensates in the aliasing band; however, the participants
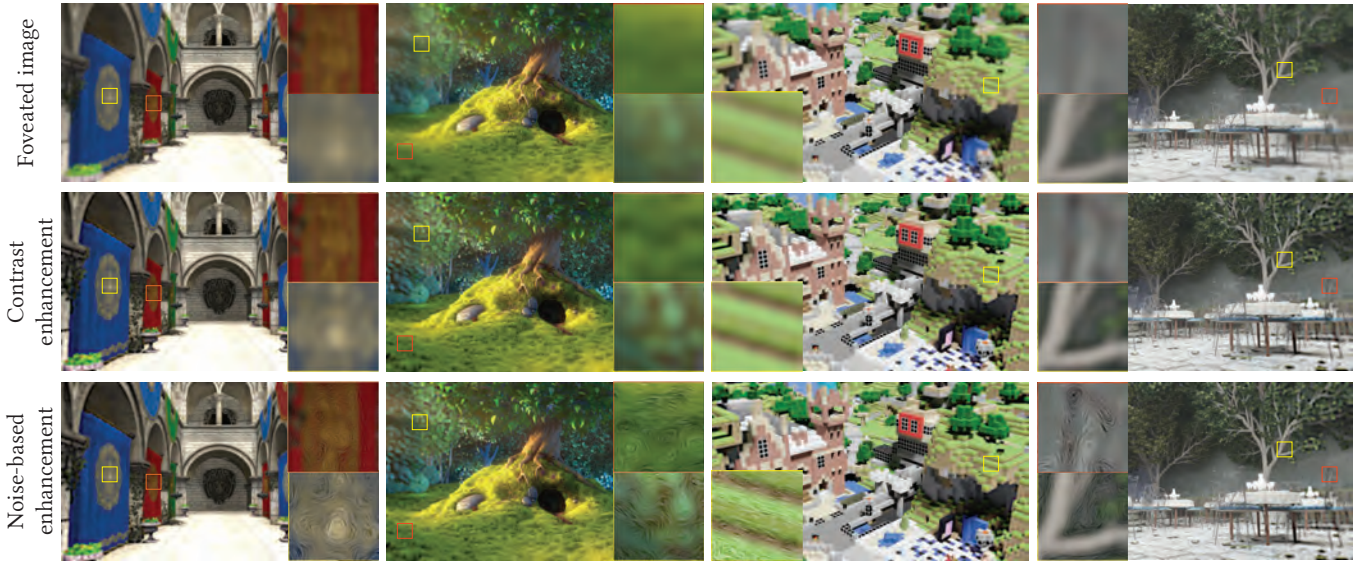
Fig. 11. The comparison of contrast enhancement and noise-based enhancement methods with input foveated images. The insets show magnified views of two image regions for visual comparison. These images are designed to be viewed at an 80 degree wide field-of-view with the gaze at the center. For more results, please refer to the supplemental materials.
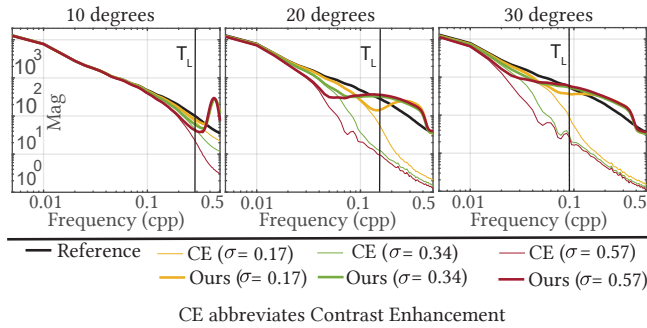

Fig. 12. Fourier analysis of our technique for three different foveations and eccentricities; averaged over 5 scenes. Our approach synthesizes lost high frequency content in the aliasing band which mitigates the perception of foveation.
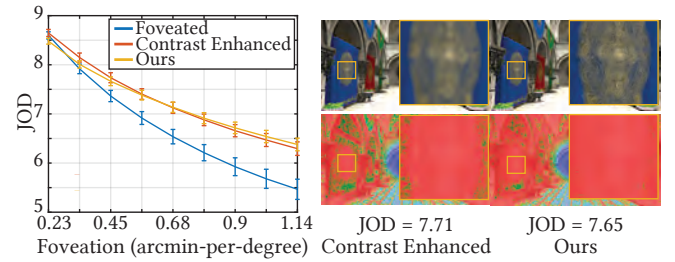

Fig. 13. Output of FovVDP for our noise-based enhancement, contrast enhancement, and standard foveated rendering results (FovVDP reference: full-resolution image). *Left*: FovVDP produces very similar JOD scores for our method and contrast enhancement, whereas standard foveation always has worse JOD scores. *Right*: FovVDP input images and error maps for contrast enhancement and our method with similar JOD scores at the same amount of foveation. ($\sigma_{BlurRate} = 0.68$).

did not report visual artifacts and we relate this to the frequency band being not resolvable by the HVS.

Additionally, we report the outputs from FovVDP [Mantiuk et al. 2021] objective quality metric for our method, contrast enhancement, and standard foveation in Figure 13 (FovVDP reference: full-resolution image). Overall, the average JOD scores obtained from FovVDP are significantly different between the standard deviation and both contrast enhancement and noise enhancement methods ($p < 0.001$, t-test). However, the visual quality difference between the contrast-enhanced and noise-enhanced images is not significantly different ($p = 0.07$, t-test). We believe that relatively close JOD scores obtained for our method and contrast enhancement is due to the greatly reduced contrast sensitivity function used in

FovVDP for the frequencies in the aliasing band. We find this observation intuitive because contrast sensitivity functions usually represent the measurements from psychovisual experiments where the visibility threshold is defined by resolvable contrast rather than detectable contrast. Hence, we believe that FovVDP and similar objective quality metrics are not suitable references for evaluating the visual quality of methods that enhance the aliasing band. This is further supported by the results of our subjective evaluation which indicates a difference between our and contrast enhancement method.
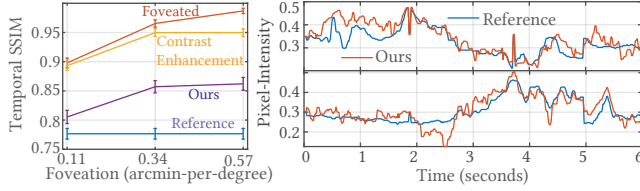
Fig. 14. The figure demonstrates the temporal coherence of our method compared to full-resolution rendering, foveated input, and contrast enhancement. The error bars visualize the standard error of the mean (SEM) for each measurement point. On the right, we demonstrate temporal slices of pixel intensities for two image locations. On the left, we measure the temporal coherence by computing SSIM metric values between consecutive frames and averaging them across the image sequence.

6.2 Temporal Coherence and Motion Preservation

Enhancing videos with additive noise requires efforts on two fronts: avoiding temporal artifacts and preserving temporal motion characteristics of the foveated input. During the derivation of our method, we took various steps to mitigate potential issues regarding temporal stability without significantly disturbing motion cues. Our method for estimating the noise parameters relies on simple image processing operations, e.g., Gaussian/Laplacian pyramids and Sobel filtering. Due to their filtering nature, these operations usually preserve the temporal coherence of the input image sequence and generate stable results as long as the input is stable. Furthermore, the frequency spectrum of our noise is defined using a smooth log-normal distribution which prevents any abrupt changes. We also fix the impulses locations for noise generation for the duration of a rendered sequence. As a result, the locations of the Gabor patches do not change, and only their parameters adapt to the underlying content. Finally, our noise is generated for less sensitive regions of our visual field, making potential temporal problems less visible.

Achieving temporal consistency alone is not enough. Noise, which does not reflect the motion in the scene, can alter motion perception. In particular, the static nature of our impulse locations could lead to the shower-door effect, which occurs when a dynamic scene is overlaid with static content. We let amplitude and orientation change according to the underlying content to limit this effect and make the noise pattern follow temporal content changes. Consider a single line moving across the screen. Even though the impulse locations are fixed, the amplitudes and orientations at each impulse adjust for underlying content such that noise appears only at the line edges. This effectively makes the noise follow the line, i.e., the noise is not static and adheres to underlying motion.

To the best of our knowledge, there is no established metric for evaluating the temporal stability of a video sequence, especially for wide-field-of-view setups. To hint at the temporal coherency of our solution, we first visualize the temporal pixel intensity variation for two locations in one of our sequences and compare it to the corresponding signal in the full-resolution rendering. Figure 14 (right) shows the results of this analysis. While the temporal variation for our technique is slightly larger than in the full-resolution rendering, the deviation remains low without creating significant temporal instabilities. Additionally, we follow the procedure used in previous

work [Kaplanyan et al. 2019], where the temporal consistency was measured using SSIM metric [Wang et al. 2004] values between consecutive frames averaged across an image sequence. Figure 14 (left) shows the results of this analysis for one sequence and different levels of foveations. For comparison, we report values for our method, full-resolution rendering, foveated input, and when only the contrast enhancement is applied. The foveated sequence has the highest average SSIM between frames because of spatial smoothing introduced by foveation. It is followed by the contrast enhancement, our method and ground truth, respectively. As our method synthesizes noise over the contrast-enhanced frames, a decrease in SSIM with respect to the contrast enhancement is expected. Still, our method delivers a high average inter-frame SSIM when compared to the reference sequence.

The above analysis gives only hints on the temporal properties of our technique. We refer to our supplemental materials for rendering sequences processed with our method.

6.3 Performance

We evaluated the performance of the noise-based enhancement method using a PC with Intel(R) Core(TM) i7-9700K CPU at 3.60GHz, 32 GB of RAM, and a single NVIDIA GeForce RTX 2080 Ti GPU. Considering the practical requirements of foveated rendering, we measured the performance for 4K resolution (3840×2160 px) input images. We considered the two main steps of our

Fig. 15. Running times

method: noise parameter estimation and noise synthesis. The parameter estimation component consists of building all the image pyramids and estimating frequency, amplitude, and orientation for each Gabor patch. The synthesis considers sparse convolution of impulses with Gabor patches. Furthermore, we consider different impulse densities for the noise generation as it provides a trade-off between the noise quality and performance (Section 4.4). The corresponding performance plots are shown in Figure 15. We can observe that the main cost of our technique is the parameter estimation for the noise. Overall, the increase in performance is approximately linear as the number of kernels increases. Note the running times are plotted in a logarithmic domain. Our initial choice of 64 impulses per kernel was motivated by the recommendation in the original paper [Lagae et al. 2009]. However, this recommendation did not take into account more flexible perceptual requirements of showing the noise in the peripheral visual field. Consequently, we experimented with lowering the number of impulses. Our initial visual inspection revealed that the noise pattern preserves its quality even for as little as 12 impulses per kernel (Figure 16, left). To evaluate that this number is still sufficient for our noise-based enhancement technique, we repeated the validation experiment (Section 6.1) for 12 impulses. The results of this experiment (Figure 16, right) demonstrate that the advantage provided by our method over the contrast-enhancement method remains unchanged when compared to the experiment for 64 impulses ($p < 0.001$, t-test). This suggests that reducing the
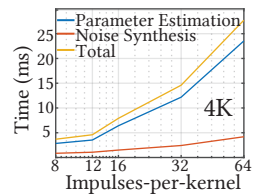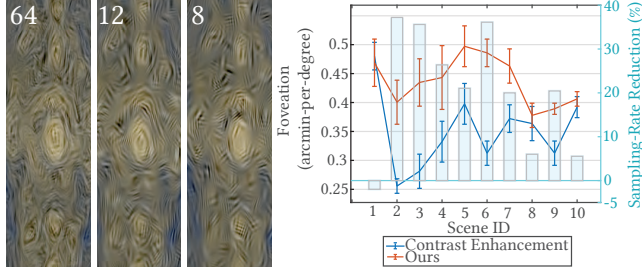
Fig. 16. Reducing the number of impulses per kernel for noise generation has little effect on the quality (left). The results of the additional experiment (right) demonstrate that also using 12 impulses per kernel allows more aggressive foveation ($p < 0.001$, t-test), and the gains are similar to those obtained for 64 impulses. The error bars visualize the standard error of the mean (SEM) for each measurement point. The bar plot shows the estimated sampling-rate reduction of our technique over contrast enhancement.
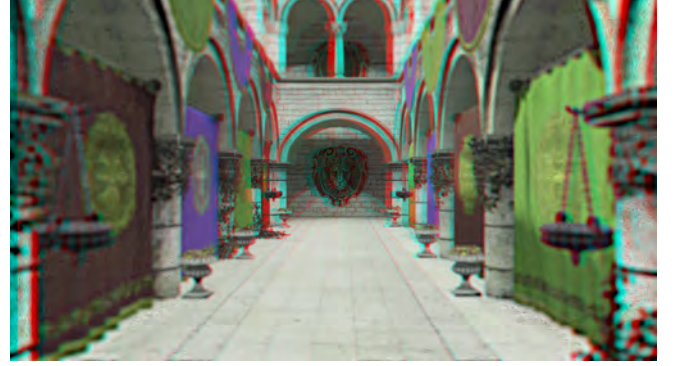

Fig. 17. The image shows our technique applied to a stereoscopic rendering of the Sponza scene. The image is presented in red-cyan anaglyph colors. This image is designed to be viewed at an 80-degree wide field-of-view, with gaze at the center.

number of impulses to 12 is a viable option for improving the performance of our method. This brings down the time required for running the entire enhancement process to 4.6 ms (217 frames per second), compared to 27.7 ms (36 FPS) for 64 impulses.

We report the scene-independent running times from an unoptimized implementation of our method on GPU. It may be challenging to quantify the actual gains from foveated rendering when using rasterization because the gains are highly dependent on the complexity of the scene, i.e., geometry and shading. Overall, the computational savings should compensate for the overhead of foveated rendering before a net gain is observed [Tursun et al. 2019]. For ray tracing, however, it is more straightforward to estimate the computational gains from our method by correlating the tolerable amount of increase in the blur-rate from our experiments with the sampling-rate reduction for ray tracing. Assuming a maximum tolerable blur to be $\sigma(x)$ at location x, the corresponding minimum sampling rate required to represent the spatial frequencies is inversely proportional to the $\sigma(x)$, according to the Nyquist theorem. Using this scheme, we aggregate the sampling-rate reduction with respect to contrast enhancement across the entire image. The gains for all the stimuli used in our experiments are reported in Figure 16 (right).

It is also important to note that for stereo rendering on VR-HMDs, our noise synthesis is performed only once, followed by a computationally efficient warping [Didyk et al. 2010]. We believe that foveated rendering applications on higher resolution (e.g., 8K) displays will benefit even more from our technique.

Our technique is designed for real-time foveated rendering. However, it is also applicable to foveated image compression. In order to compute potential gains in foveated image compression, we apply JPEG compression at a quality factor of 90 to foveated images and compare the file sizes. We measure up to 23% reduction in foveated image file size for our method compared to contrast enhancement for the set of images used in our evaluation. The blur rates used were in accordance with user experiments in Figure 16 (right).

6.4 Stereoscopic Content

Application to virtual and augmented reality devices requires stereoscopic rendering. The results presented so far consist of monoscopic images. As a simple extension to our technique for stereoscopic rendering, we propose to generate the noise only for one eye and use an image-based warping technique [Didyk et al. 2010] to generate the noise for the second eye by warping the noise according to the disparity. Such a technique creates an illusion of noise being a part of the object's surfaces and textures. Due to the random nature of our noise, we did not observe visible problems at disocclusion regions. Figure 17 shows a red-cyan anaglyph version of our enhancement applied to a stereoscopic image. Please refer to supplemental materials for more examples of our technique applied to stereoscopic content.

7 LIMITATIONS AND FUTURE WORK

Our technique estimates the parameters of the noise based on the available information in a foveated image. In particular, it propagates the statistics such as spatial frequency amplitude and orientation from lower frequencies to higher ones. This process relies on the assumption that low and high spatial frequencies share similar statistics. However, this does not always need to be the case. Some parts of the scene or textures may inherently contain only low-frequency content, for example, sky, water, or image regions with a depth-of-field effect. Despite this, our technique will still generate high-spatial frequency noise. Scene 10 in our evaluation validation experiment (Section 6.1) is a particular example of it. In future work, it would be interesting to look for ways to address this limitation. For example, the particular case of the depth-of-field effect could be addressed by exploiting the camera parameters used for rendering the scene during the estimation of the noise parameters. Similarly, information about the frequency spectra of textures could be precomputed and stored prior to the rendering. This would inform the noise-based enhancement technique not to generate noise for regions containing only low-frequency content.

The efficiency of our enhancement method is limited in very high-contrast regions where adding noise leads to clipping image values.
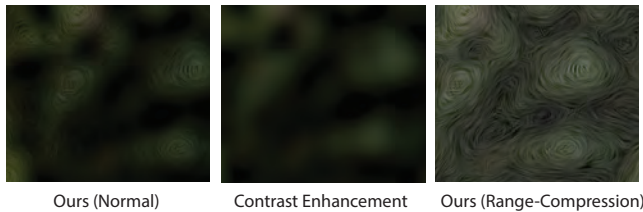
Ours (Normal)   Contrast Enhancement   Ours (Range-Compression)

Fig. 18. Saturation & Noise (*Scene*-8). *Left*: Our enhancement after contrast enhancement leaves less range for the noise to be applied.*Center*: Saturation after contrast enhancement. *Right*: Dynamic range reduction after contrast enhancement leaves more range for noise.

Our method accounts for such situations by attenuating the noise amplitude. However, such a strategy may prevent enhancement in very dark or bright image regions. Examples of such cases can be found in Scenes 1 and 8 in our validation experiment. The scenes contain dark regions that are clipped after contrast enhancement, which prevents adding noise. It would be interesting to investigate the possible trade-off between the dynamic range and our enhancement. In particular, it is possible to compress local image contrast such that the noise can be added. We provide an example of such an operation in Figure 18.

Our technique assumes that the input is a foveated rendering sequence that does not contain any spatial or temporal aliasing. Furthermore, following the studies of Albert *et al.* [2017], Hoffman *et al.* [2018] and the previous work on foveated rendering [Patney et al. 2016; Tursun et al. 2019], we simulate and model foveated rendering using spatially-varying Gaussian blur. An exciting venue for future work is to investigate the enhancement of the foveated rendering containing some form of aliasing. An analysis of the aliasing could provide additional information to the noise generation technique on the range of spatial frequencies that were not correctly synthesized.

We calibrated our technique using static images. However, we argue that the obtained parameters can be directly used for animated sequences as images represent the most challenging case in this regard. This is because motion is known to mask spatial distortions, in particular, any spatial artifacts due to foveation or our technique. In the future, it would be interesting to exploit this type of masking and design a technique that accounts for potentially higher tolerance of observers to noise in dynamic content when estimating the noise parameters. We believe that such a more accurate modeling would further increase the computational gains of the noise-based enhancement technique.

Most of the widely available virtual reality headsets provide limited spatial resolution, especially for large eccentricity [Beams et al. 2020]. Therefore, we performed all experiments using a large 4K screen, which, for our viewing conditions, can reproduce spatial frequencies up to 24 cycles per visual degree for the fovea region. In comparison, the upper limit on the spatial frequency reproduction for the HTC Vive Pro Eye headset is 5 cycles per visual degree, which is insufficient to fully exploit the benefits of our enhancement. Newer display models such as the Varjo VR-3 HMD can represent up to 15 cycles-per-degree in the periphery, which is sufficient to exploit the full benefits of our method.

In this work, we demonstrate the utility of our technique in the context of rendering. However, we believe that similarly, the method can be used to enhance foveated video compression and complement such techniques as [Walton et al. 2021]. Furthermore, we believe that there is significant potential in applying noise-based techniques for many problems in real-time foveated graphics. We have only explored temporally-coherent spatial characteristics of 2D Gabor noise for our technique. In the future, other noise variations, such as 3D Gabor noise, which can effectively control temporal characteristics of the content, or techniques that explicitly preserve motion cues during noise synthesis [Bénard et al. 2010] can pave the way for efficient rendering of temporal metamers. Considering the spectral nature of various psycho-visual models, the ability to precisely control spatio-temporal spectral properties of synthesized content presents an exciting tool to bridge human perception and graphics.

8 CONCLUSIONS

Foveated rendering enables significant computation benefits for high-resolution wide-field-of-view displays. Therefore, it becomes a crucial component and key enabler for high-quality rendering systems in virtual and augmented reality headsets. In this work, we present an efficient technique that can further reduce the cost of the foveated rendering by allowing more aggressive foveation without perceived quality loss. We achieve this goal thanks to our new noise-based enhancement step, which replaces the rendering of high-spatial frequency details with inexpensive procedural noise synthesis. This paper presents a perceptually-inspired method for estimating the parameters of the noise and calibration of the technique based on data collected in perceptual experiments. The method consists of a series of simple image processing steps applied directly to the output of the foveated rendering. Thanks to these properties, our enhancement method is suitable for direct integration into foveated rendering systems.

A DETECTION AND RESOLUTION FREQUENCY THRESHOLD MEASUREMENTS

Table 3. The measurements from Thibos *et al.* [1996], that we linearly interpolate to obtain $T_H(e_\mathbf{x})$ and $T_L(e_\mathbf{x})$ for different values of $e_\mathbf{x}$.

Eccentricity (deg)	0	5	10	15	20	25	30
T_L	60	27	10.5	8	5.5	4.8	4
T_H	60	40	26	24	23	21	20.5

REFERENCES

Rachel Albert, Anjul Patney, David Luebke, and Joohwan Kim. 2017. Latency requirements for foveated rendering in virtual reality. *ACM Trans. Appl. Percept.* 14, 4 (2017).

Roger S. Anderson, Margarita B. Zlatkova, and Shaban Demirel. 2002. What limits detection and resolution of short-wavelength sinusoidal gratings across the retina? *Vision Res.* 42, 8 (2002), 981–990.

Stephen J. Anderson, Kathy T. Mullen, and Robert F. Hess. 1991. Human peripheral spatial resolution for achromatic and chromatic stimuli: limits imposed by optical and retinal factors. *J. Physiol. (Lond.)* 442, 1 (1991), 47–64.

H. R. Aubert and C. F. R. Foerster. 1857. Beitrage zur kenntnisse der indirecten sehens [Translation: Contributions of knowledge to indirect vision]. *Graefes Archiv fur Ophthalmologie* 3 (1857).

Horace B. Barlow. 1961. Possible principles underlying the transformation of sensory messages. *Sensory communication* 1, 01 (1961).

Ryan Beams, Brendan Collins, Andrea S. Kim, and Aldo Badano. 2020. Angular dependence of the spatial resolution in virtual reality displays. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 836–841.

Pierre Bénard, Ares Lagae, Peter Vangorp, Sylvain Lefebvre, George Drettakis, and Joëlle Thollot. 2010. A dynamic noise primitive for coherent stylization. *Comput Graph Forum* 29, 4 (2010), 1497–1506.

G. Browder and W. Chambers. 1988. Eye-enslaved area-of-interest display systems. In *Proc. of Flight Simulation Technologies Conference*.

Christine A. Curcio, Kenneth R. Sloan, Robert E. Kalina, and Anita E. Hendrickson. 1990. Human photoreceptor topography. *Journal of comparative neurology* 292, 4 (1990), 497–523.

Russell L. De Valois, Duane G. Albrecht, and Lisa G. Thorell. 1982. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res.* 22, 5 (1982), 545–559.

Arturo Deza, Aditya Jonnalagadda, and Miguel P. Eckstein. 2019. Towards metamerism via foveated style transfer. In *International Conference on Learning Representations*. Openreview.net, New Orleans, USA.

Piotr Didyk, Tobias Ritschel, Elmar Eisemann, Karol Myszkowski, and Hans-Peter Seidel. 2010. Adaptive image-space stereo view synthesis. In *VMV*. 299–306.

A. Eugen Fick. 1898. Ueber stäbchensehschärfe und zapfensehschärfe. *Albrecht von Graefes Archiv für Ophthalmologie* 45, 2 (1898), 336–356.

Jeremy Freeman and Eero P. Simoncelli. 2011. Metamers of the ventral stream. *Nature neuroscience* 14, 9 (2011), 1195–1201.

William E. Glenn. 1994. Real-time display systems, present and future. In *Visual Science and Engineering: Models and Applications*. Marcel Dekker, 387–413.

Daniel G. Green. 1970. Regional variations in the visual acuity for interference fringes on the retina. *J. Physiol. (Lond.)* 207, 2 (1970), 351–356.

Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. 2012. Foveated 3D graphics. *ACM Trans. Graphics* 31, 6 (2012).

Andrew M. Haun. 2021. What is visible across the visual field? *Neuroscience of Consciousness* 2021, 1 (06 2021).

E. Hering. 1899. Concerning the limits of visual acuity. *Ber. d. math.-phys. Kl. d. K. Sachs. Gesellsch. d. Wissensch. zu Leipzig* (1899).

Joy Hirsch and Christine A. Curcio. 1989. The spatial resolution capacity of human foveal retina. *Vision Res.* 29, 9 (1989), 1095–1101.

David Hoffman, Zoe Meraz, and Eric Turner. 2018. Limits of peripheral acuity and implications for VR system design. *Journal of the Society for Information Display* 26, 8 (2018), 483–495.

Aapo Hyvärinen, Jarmo Hurri, and Patrick O. Hoyer. 2009. *Natural image statistics: A probabilistic approach to early computational vision*. Vol. 39. Springer Science & Business Media.

Anton Kaplanyan, Anton Sochenov, Thomas Leimkühler, Mikhail Okunev, Todd Goodall, and Gizem Rufo. 2019. DeepFovea: Neural reconstruction for foveated rendering and video compression using learned statistics of natural videos. *ACM Trans. Graphics* 38, 6 (2019).

Philip Kortum and Wilson S. Geisler. 1996. Implementation of a foveated image coding system for image bandwidth reduction. In *Human Vision and Electronic Imaging*, Vol. 2657. International Society for Optics and Photonics, 350–360.

Ares Lagae, Sylvain Lefebvre, Rob Cook, Tony DeRose, George Drettakis, David S. Ebert, John P. Lewis, Ken Perlin, and Matthias Zwicker. 2010. A survey of procedural noise functions. *Comput Graph Forum* 29, 8 (2010), 2579–2600.

Ares Lagae, Sylvain Lefebvre, George Drettakis, and Philip Dutré. 2009. Procedural noise using sparse Gabor convolution. *ACM Trans. Graphics* 28, 3 (2009).

Gordon E. Legge and Daniel Kersten. 1987. Contrast discrimination in peripheral vision. *J. Opt. Soc. Am. A* 4, 8 (Aug 1987), 1594–1598.

Jerome Y. Lettvin et al. 1976. On seeing sidelong. *The Sciences* 16, 4 (1976), 10–20.

Dennis M. Levi and Stanley A. Klein. 1986. Sampling in spatial vision. *Nature* 320, 6060 (1986), 360–362.

Marc Levoy and Ross Whitaker. 1990. Gaze-directed volume rendering. *SIGGRAPH* 24, 2 (feb 1990), 217–223.

Rafał Mantiuk, Kil Joong Kim, Allan G. Rempel, and Wolfgang Heidrich. 2011. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graphics* 30, 4 (2011).

Rafał K. Mantiuk, Gyorgy Denes, Alexandre Chapiro, Anton Kaplanyan, Gizem Rufo, Romain Bachy, Trisha Lian, and Anjul Patney. 2021. FovVideoVDP: A visible difference predictor for wide field-of-view video. *ACM Trans. Graphics* 40, 4 (2021).

Xiaoxu Meng, Ruofei Du, Matthias Zwicker, and Amitabh Varshney. 2018. Kernel foveated rendering. *Proc. of the ACM on Computer Graphics and Interactive Techniques* 1, 1 (2018).

Michel Millodot, Chris A. Johnson, Anne Lamont, and Herschel W. Leibowitz. 1975. Effect of dioptrics on peripheral visual acuity. *Vision Res.* 15, 12 (1975), 1357–1362.

Bipul Mohanto, A. B. M. Tariqul Islam, Enrico Gobbetti, and Oliver Staadt. 2021. An integrative view of foveated rendering. *Computers & Graphics* (2021).

G. A. Østerberg. 1935. Topography of the layer of the rods and cones in the human retina. *Acta ophthalmol* 13, 6 (1935).

Anjul Patney, Marco Salvi, Joohwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. 2016. Towards foveated rendering for gaze-tracked virtual reality. *ACM Trans. Graphics* 35, 6 (2016).

Eli Peli. 1990. Contrast in complex images. *J. Opt. Soc. Am. A, Optics and image science* 7 10 (1990), 2032–2040.

Javier Portilla and Eero P. Simoncelli. 2000. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int J Comput Vision* 40, 1 (2000), 49–70.

Ruth Rosenholtz. 2016. Capabilities and limitations of peripheral vision. *Annual Review of Vision Science* 2, 1 (2016), 437–457.

Ethan A. Rossi and Austin Roorda. 2010. The relationship between visual resolution and cone spacing in the human fovea. *Nature neuroscience* 13, 2 (2010), 156–157.

Jyrki Rovamo and Veijo Virsu. 1979. An estimation and application of the human cortical magnification factor. *Experimental brain research* 37, 3 (1979), 495–510.

Jyrki Rovamo, Veijo Virsu, and Risto Näsänen. 1978. Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision. *Nature* 271, 5640 (1978), 54–56.

Daniel L. Ruderman. 1994. The statistics of natural images. *Network: computation in neural systems* 5, 4 (1994), 517–548.

Heiko H. Schütt and Felix A. Wichmann. 2017. An image-computable psychophysical spatial vision model. *J. Vis.* 17, 12 (2017).

Boubakar Sere, Christian Marendaz, and Jeanny Herault. 2000. Nonhomogeneous resolution of images of natural scenes. *Perception* 29, 12 (2000), 1403–1412.

Eero P. Simoncelli and Bruno A. Olshausen. 2001. Natural image statistics and neural representation. *Annual review of neuroscience* 24, 1 (2001), 1193–1216.

Michael Stengel, Steve Grogorick, Martin Eisemann, and Marcus Magnor. 2016. Adaptive image-space sampling for gaze-contingent real-time rendering. In *Comput Graph Forum*, Vol. 35. Wiley Online Library, 129–139.

Hans Strasburger, Ingo Rentschler, and Martin Jüttner. 2011. Peripheral vision and pattern recognition: A review. *J. Vis.* 11, 5 (12 2011).

Nicholas T. Swafford, José A. Iglesias-Guitian, Charalampos Koniaris, Bochang Moon, Darren Cosker, and Kenny Mitchell. 2016. User, metric, and computational evaluation of foveated rendering methods. In *Proc. of the ACM Symposium on Applied Perception*. 7–14.

L. N. Thibos. 1998. Acuity perimetry and the sampling theory of visual resolution. *Optometry and vision science: official publication of the American Academy of Optometry* 75, 6 (1998), 399–406.

L. N. Thibos, F. E. Cheney, and D. J. Walsh. 1987a. Retinal limits to the detection and resolution of gratings. *J. Opt. Soc. Am. A* 4, 8 (1987), 1524–1529.

L. N. Thibos, D. L. Still, and A. Bradley. 1996. Characterization of spatial aliasing and contrast sensitivity in peripheral vision. *Vision Res.* 36, 2 (1996), 249–258.

L. N. Thibos and D. J. Walsh. 1985. Detection of high frequency gratings in the periphery. *J. Opt. Soc. Am. A* 2 (1985).

L. N. Thibos, D. J. Walsh, and F. E. Cheney. 1987b. Vision beyond the resolution limit: aliasing in the periphery. *Vision Res.* 27, 12 (1987), 2193–2197.

H. M. Tong and R. A. Fisher. 1984. *Progress report on an eye-slaved area-of-interest visual display*. Technical Report. Singer Co Silver Spring Md Link Simul. Systems Div.

Norimichi Tsumura, Chizuko Endo, Hideaki Haneishi, and Yoichi Miyake. 1996. Image compression and decompression based on gazing area. In *Human Vision and Electronic Imaging*, Vol. 2657. International Society for Optics and Photonics, 361–367.

Okan Tarhan Tursun, Elena Arabadzhiyska-Koleva, Marek Wernikowski, Radosław Mantiuk, Hans-Peter Seidel, Karol Myszkowski, and Piotr Didyk. 2019. Luminance-contrast-aware foveated rendering. *ACM Trans. Graphics* 38, 4 (2019).

Veijo Virsu, Risto Näsänen, and Kari Osmoviita. 1987. Cortical magnification and peripheral vision. *J. Opt. Soc. Am. A* 4, 8 (Aug 1987), 1568–1578.

David R. Walton, Rafael Kuffner Dos Anjos, Sebastian Friston, David Swapp, Kaan Akşit, Anthony Steed, and Tobias Ritschel. 2021. Beyond blur: real-time ventral metamers for foveated rendering. *ACM Trans. Graphics* 40, 4 (2021).

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Processing* 13, 4 (2004), 600–612.

Andrew B. Watson. 2018. The field of view, the field of resolution, and the field of contrast sensitivity. *Electronic Imaging* 2018, 14 (2018).

Andrew B. Watson and Albert J. Ahumada. 2016. The pyramid of visibility. *Electronic Imaging* 2016, 16 (2016).

T. H. Wertheim. 1894. Uber die indirekte sehscharfe. *Zeitschrift fur Psychologie* 7 (1894), 172–187.

David R. Williams. 1985a. Aliasing in human foveal vision. *Vision Res.* 25, 2 (1985), 195–205.

David R. Williams. 1985b. Visibility of interference fringes near the resolution limit. *J. Opt. Soc. Am. A* 2, 7 (1985), 1087–1093.