

Towards Motion Metamers for Foveated Rendering

TAIMOOR TARIQ, Università della Svizzera italiana (USI), Switzerland

PIOTR DIDYK, Università della Svizzera italiana (USI), Switzerland



Fig. 1. Loss of spatial details in foveated rendering can inhibit motion perception, leading to underestimation of speed. We introduce a concept of motion metamers, i.e., sequences that are structurally different from the reference, but induce the same spatial and motion perception. As a first step towards motion metamers, we design a real-time technique that synthesizes spatio-temporal motion energy to compensate for the loss in motion perception in foveated rendering.

Foveated rendering takes advantage of the reduced spatial sensitivity in peripheral vision to greatly reduce rendering cost without noticeable spatial quality degradation. Due to its benefits, it has emerged as a key enabler for real-time high-quality virtual and augmented realities. Interestingly though, a large body of work advocates that a key role of peripheral vision may be motion detection, yet foveated rendering lowers the image quality in these regions, which may impact our ability to detect and quantify motion. The problem is critical for immersive simulations where the ability to detect and quantify movement drives actions and decisions. In this work, we diverge from the contemporary approach towards the goal of foveated graphics, and demonstrate that a loss of high-frequency spatial details in the periphery inhibits motion perception, leading to underestimating motion cues such as velocity. Furthermore, inspired by an interesting visual illusion, we design a perceptually motivated real-time technique that synthesizes controlled spatio-temporal motion energy to offset the loss in motion perception. Finally, we perform user experiments demonstrating our method's effectiveness in recovering motion cues without introducing objectionable quality degradation.

CCS Concepts: • Computing methodologies → Perception; Virtual reality; Rendering; Image manipulation.

Additional Key Words and Phrases: foveated rendering, visual perception

Authors' addresses: Taimoor Tariq, Università della Svizzera italiana (USI), Lugano, Switzerland, tariqt@usi.ch; Piotr Didyk, Università della Svizzera italiana (USI), Lugano, Switzerland, piotr.didyk@usi.ch

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
0730-0301/2024/7-ART65 \$15.00
<https://doi.org/10.1145/3528223.3530101>

ACM Reference Format:

Taimoor Tariq and Piotr Didyk. 2024. Towards Motion Metamers for Foveated Rendering. *ACM Trans. Graph.* 41, 4, Article 65 (July 2024), 10 pages. <https://doi.org/10.1145/3528223.3530101>

1 INTRODUCTION

Foveated rendering techniques [Mohanty et al. 2021] exploit decay in the sensitivity of the human visual system (HVS) in peripheral vision, and reduce rendering quality according to the viewers' gaze location. The most common approach for applying this idea is to limit the shading rate or spatial resolution for image regions distant from the gaze location. Due to its significant computational benefits, foveated rendering has emerged as a key enabler of high-quality real-time rendering on virtual reality (VR) head-mounted displays (HMDs) and has been adopted by latest VR displays such as the Sony Play-Station VR2¹ and Varjo VR-3².

Foveated rendering is a particular case of *visual metamerism*, an interesting perceptual phenomenon in which observers perceive physically different images as perceptually indistinguishable [Freeman and Simoncelli 2011]. The phenomenon has been of great interest to cognitive psychologists and graphics researchers. From a psychological perspective, understanding visual metamerism is vital for understanding the mechanism by which our brain forms a representation of the visual world [Broderick et al. 2023; Freeman and Simoncelli 2011; Portilla and Simoncelli 2000]. From a graphics perspective, a significant body of recent work has been aimed at new techniques, i.e., foveated methods, to synthesize visual metamerism of

¹<https://mixed-news.com/en/psvr-2-foveated-rendering-technical-marvel/>

²<https://developer.varjo.com/docs/native/foveated-rendering-api>

rendered scenes at a lower computational cost than traditional rendering pipelines [Guenter et al. 2012; Krajancich et al. 2023; Patney et al. 2016].

While previous work on foveated rendering focuses on synthesizing images that appear indistinguishable from original rendering or do not lead to objectionable quality loss, they often ignore other processes and tasks of the HVS. This work focuses on motion perception, an essential aspect of human perception and VR. While foveated rendering tries to save computational resources when generating content for periphery, peripheral vision plays a critical role in detecting and perceiving movements [Exner 1886; Finlay 1982]. Therefore, properly replicating motion cues in the periphery is crucial for realizing the dream of realistic VR, i.e., the sensory equivalence between simulated immersive environments and the real world. A potential loss of motion cues in the periphery due to foveated rendering may affect the accurate simulation of immersive environments and lead to unforeseen consequences, especially in environments where information about movement in the scene drives a user’s decision-making process.

In this work, we first provide ground motivation for our work. We conduct a small perceptual experiment investigating velocity perception in peripheral vision, demonstrating that a loss of spatial details in the periphery (the typical result of foveated rendering) can inhibit velocity perception, making global camera motion appear slower than it physically is. Then, we pose a simple question: just like how spatial metamers aim to replicate spatial statistics in images seeking for an equivalence in how an image is encoded by our brain [Freeman and Simoncelli 2011], can we synthesize and control motion energy via a metameristic process, i.e., by synthesizing content that is structurally different from a reference, but evokes similar motion perception? We take the first step toward answering the above question, drawing inspiration from visual illusions such as the “Double-Drift Illusion” and the Fourier theory of human motion perception. We design a perceptually motivated real-time technique that synthesizes controlled spatio-temporal motion energy and superimposing it over the underlying foveated video. Through our validation experiments, we show that our technique can offset the inhibition of velocity perception due to loss in peripheral resolution, without objectionable quality degradation, resulting in a visual metamer that can reproduce motion cues more faithfully than standard foveated rendering.

2 BACKGROUND AND RELATED WORK

In this section, we discuss relevant literature and recent advances in foveated rendering and human visual perception.

2.1 Foveated Rendering

Foveated rendering aims to progressively reduce the quality of rendering away from the user gaze point. This technique is perceptually justified by the observation that our resolution acuity decreases in the periphery of our vision [Thibos et al. 1996]. Guenter et al. [2012] demonstrated that gaze-contingent shading rate reduction can lead to significant rendering time reductions without noticeable quality degradation. Patney et al. [2016] took this a step further, demonstrating that additional eccentricity dependent contrast enhancement

can lead to more computational benefits. Around the same time, Stengel et al. [2016] utilized other information from the rendering pipeline to improve the technique further. Meng et al. [2018] employed kernel log-polar mapping to optimize the technique further for GPUs. Tursun et al. [2019] demonstrated that accounting for the masking effects of underlying content can further improve the computational benefits. Kaplanyan et al. [2019] utilized the power of deep learning for foveated image reconstruction with very sparse sampling, but incurring a significant computational overhead. Tariq et al. [2022] proposed a technique for real-time enhancement of foveated rendering by adding perceptually controlled Gabor noise. Krajancich et al. [2023] demonstrated that higher level cognitive functions, such as attention to a certain task in the fovea may enable more aggressive foveated rendering without noticeable quality loss. Surace et al. [2023] recently showed that rather than only reducing the shading-rate, one can also reduce the geometric complexity of the scene in the periphery. Lisboa et al. [2023a] recently showed that foveation can be increased as motion increases.

2.2 Peripheral Vision and Spatial Metamerism

Spatial metamerism refers to the phenomena of physically different images being perceptually indistinguishable. The key idea is that due to the loss of spatial resolution in peripheral vision, we need to preserve some local statistics rather than per-pixel equivalence. Barlow [1961] and Simoncelli et al. [2001] hypothesized that the early visual system may be aiming to encode visual stimulus efficiently. Portilla and Simoncelli [2000] later showed that a computational model based on wavelet representations can be used to synthesize visually plausible textures. Inspired by the initial work, there have been more recent studies aimed at synthesizing spatial metamers using deep learning [Deza et al. 2019]. Walton et al. [2021] leveraged previous work on spatial pooling in human peripheral vision by Freeman et al. [2011], and proposed a technique for spatial metamer synthesis that preserves local mean and variance. Inspired by earlier studies on spatial frequency resolvability in peripheral vision [Thibos and Walsh 1985], Tariq et al. [2022] proposed a technique for no-reference real-time spatial metamer synthesis by inexpensive Gabor-noise synthesis.

2.3 Motion Perception

An important leap in vision science was the application of fourier analysis to demonstrate that suitably tuned spatio-temporal filters can be used to model human motion perception [Adelson and Bergen 1985; Burr et al. 1986; van Santen and Sperling 1985; Watson and Ahumada 1985]. These models were built on the pioneering work of Werner et al. [1961]. The modelling of motion as spatio-temporal energy has found successful application in video quality assesment techniques [Watson 1998], and also modelling the critical display artifacts such as judder and flicker [Watson 2013]. Watson et al. [2013] theorized that the visibility of temporal artifacts can be estimated by analyzing the spectrum of a video in relation to the pyramid of visibility, which defines a closed pyramidal structure in the spatio-temporal frequency space.

In regards to peripheral vision, it has been stated that the periphery functions as an organ for movement detection [Exner 1886].

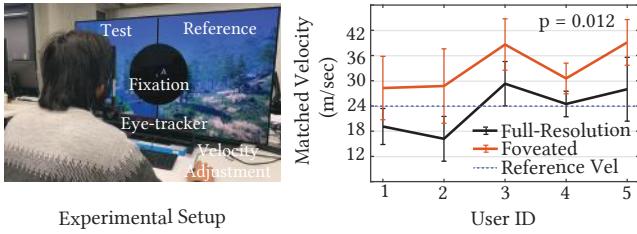


Fig. 2. Foveated Motion Perception: In a full reference task to match the velocity of foveated/full-resolution scenes a reference moving at a fixed velocity (dotted blue line). People tend to assign the foveated scene a significantly higher velocity compared to the full-resolution scene for equivalence (foveated was perceived slower). This serves as evidence for potential loss in motion perception/cues due to peripheral blur. The reported p-value is for a t-test between the two groups (full-resolution and foveated). The error bars represent Standard Error (SE).

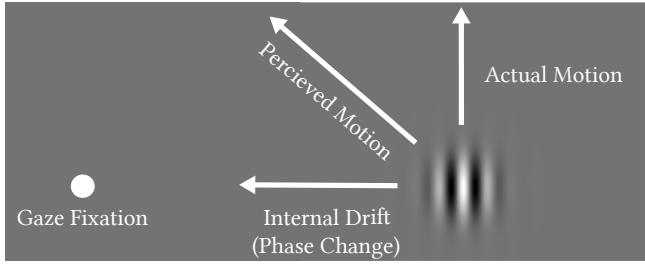


Fig. 3. Double-Drift Illusion: In peripheral vision, the visual system integrates the actual motion and the internal drift due to phase change; resulting in a perceived motion that is different from the actual motion of the Gabor stimulus.

Armstrong (quoted in [Sekuler 1875]) demonstrated that the periphery is able to detect motion in stimuli moving at 50 deg/s, while the foveal system is not. However, in terms of lower bounds, Klein *et al.* [1941] showed that the displacement threshold for moving stimuli is much less for foveal viewing compared to the far periphery. Hartman *et al.* [1979] showed that the critical flicker frequency in the mid-periphery may be higher compared to the fovea, but it reduces sharply in the far periphery. Very recent work by Krajancich *et al.* [2021], shows that this realization presents potential for reducing refresh-rates in the far periphery.

3 MOTIVATION

In the section, we discuss the motivation behind our work, showing that a loss of spatial resolution in peripheral vision may inhibit motion perception and we can take inspiration from an interesting visual illusion to mitigate this loss.

3.1 Peripheral Blur and Velocity Perception

Foveated rendering removes and attenuates spatio-temporal frequencies in the periphery. According to the spatio-temporal energy interpretation of motion [Adelson and Bergen 1985] and the fact that peripheral vision is categorically sensitive to motion, this may lead to a loss in perceived motion energy. There is also evidence

in literature to support this hypothesis, showing that even foveal reduction in resolution may lead to inhibition of motion cues [Stone and Thompson 1992]. To investigate, we ran a small controlled study (setup shown in Figure 2) to understand whether a loss in peripheral resolution inhibits the sensation of velocity. Users with their gaze fixated at the center of a display were shown two videos, playing concurrently on either halves of the screen. On the right was a full-quality reference video with a linear camera motion and fixed velocity (24 m/sec), empirically selected such that the angular movement of objects was near half of the maximum smooth pursuit velocity (30 deg/s). Depending on the trial, on the left was either a foveated version of the scene, or the same full-quality version. To focus on the peripheral vision, the content was removed from a circle with 15 visual degree radius around the central gaze location. The initial scene camera position between the left and right halves was misaligned, to aid in objective motion perception. The task was to use the method of adjustment to tune the velocity of the test scene on the left, such that users believe it is moving at the same speed as the scene on the right. At each trial, the video on the left started with a random initial velocity. The users were not informed about the purpose of the experiment, and nature of the manipulations to the scenes. Each user performed a total of 20 randomized trials (10 for both full-resolution and foveated test cases).

The study results (Figure 2) indicate that to achieve perceived velocity equivalence with the full-resolution reference, users consistently and significantly overshoot for the foveated test scene, as opposed to when the full-quality version is shown on the left. This means that even at the same physical camera velocities, users perceive the foveated test stimulus to be moving slower compared to the full-resolution one. The reported p-value is for a t-test between the two groups (full-resolution and foveated), and the error bars represent standard error (SE) per-participant.

Now an interesting question arises; not unlike how spatial metamers aim to preserve spatial statistics in images [Freeman and Simoncelli 2011], can we efficiently compensate for this loss in peripheral motion cues by "hallucinating" spatio-temporal motion energy? Interestingly, we find a potential hint towards the answer in a fascinating visual illusion called the double-drift illusion.

3.2 Double-Drift Illusion and Phase-based Motion

In the double-drift illusion³ illustrated in Figure 3 (refer to the supplementary material for the video), a Gabor patch physically moves in one direction, while continuously undergoing a phase drift in another direction. If an observer gazes at the Gabor patch directly, they perceive it moving in the actual motion direction. However, if the Gabor patch is in the periphery, the phase shift strongly alters the perceived motion path, and it now appears to be moving diagonally (Figure 3). This illusion demonstrates a significant impact of internal drift (phase change) in modulating perceived motion in the periphery, encouraging us that an approach which can synthesize spatial energy, and then modulate it using local phase changes can be effective for controlling directional motion energy.

In computer graphics, the use of phase changes to induce motion energy is not a foreign concept. Wadhwa *et al.* [2013] demonstrated

³<https://cavlab.net/Demos/DoubleDrift/>

that enforcing controlled phase change on steerable pyramid works remarkably well in magnifying underlying motion, or even synthesizing an illusion of motion without any explicit movement [Freeman et al. 1991]. However, such an approach is not applicable to our problem, as the high spatial frequencies we wish to temporally modulate are removed due to foveated rendering and phase modulations on low frequencies would cause objectionable artifacts.

Therefore, we require to design a technique that offers us the ability to synthesize high spatial frequencies with precise spectral control (as we do not want the hallucinated content to be resolvable), and the ability to modulate the phase as a source of motion energy. Furthermore, the technique needs to be efficient enough to maintain real-time performance. The recent work by Tariq *et al.* [2022] is an ideal base. The framework can synthesize/control spatial frequencies, and as it is based on superimposed Gabor patches, we have good control over phase change to introduce motion energy. Furthermore, the technique is efficient enough to maintain real-time foveated rendering rates.

4 TOWARDS MOTION METAMERS

Based on the motivation discussed in the previous section, we aim to synthesize directional motion energy using controlled phase modulations in procedural Gabor noise, and superimpose the energy over the foveated render to synthesize motion cues. Gabor noise is synthesized by the additive interactions of Gabor patches centered at randomly sampled locations. Placing a Gabor patch at a certain location can be thought of as its convolution with an impulse (i) centered at that location. [Lagae et al. 2009]. A Gabor patch corresponding to the impulse i can be defined as a product of a sinusoidal function and a Gaussian envelope.

$$g_i(x, y) = K_i e^{-\pi a_i^2(x^2+y^2)} \cos(2\pi f_s^i(x \cos \theta_i + y \sin \theta_i) + \phi_i) \quad (1)$$

where K_i is the amplitude, a_i controls the size, f_s^i is the spatial frequency, θ_i is the orientation, and ϕ_i is the phase. The noise value (N in Eq. 2) at a point (x, y) is the additive contribution of Gabor patches at different locations:

$$N(x, y) = \sum_{i \in I} w_i \cdot g_i(x - x_i, y - y_i) \quad (2)$$

where I is the set of impulses, which defines the number of different Gabor patches, (x_i, y_i) is the spatial location of an impulse i in the set I , and $w_i \in \{-1, 1\}$ are randomly sampled binary weights on the amplitude, to aid in the requirement that the synthesis should not alter the average brightness of the superimposed output (the synthesized noise should have an expected mean of zero). Each constituent g_i has its own set of parameters, corresponding to its location (x_i, y_i) . We observe that local coherent phase modulations can introduce percepts of global motion, as also indicated by the double-drift illusion. We aim to estimate the parameters of each Gabor patch (g_i) in a way that we can synthesize the required motion energy in a controllable way, such that users do not perceive any artifacts. To achieve this, we rely on models of spatio-temporal resolvability in human peripheral vision.

Our method (shown in Figure. 4) takes a foveated render as an input and firstly computes the Laplacian and Gaussian pyramids

to encode spatial details and different frequency bands. We do not rely on a full-resolution reference, which makes the technique ideal for the practical foveated rendering pipeline. Next, the underlying motion flow is computed based on the geometry information and transformation matrices as a byproduct of the rendering process. The pyramid and motion flow are thereafter analyzed by our perceptual model, which returns the per-impulse (i) parameters for each Gabor patch g_i . The parameters are used for synthesizing motion energy, which is superimposed over the foveated image to return a frame of our motion metamer as the final output. In the following subsections, we outline the procedure for estimating Gabor patch parameters for the motion metamer synthesis.

4.1 Orientation

Based on intuition from the double-drift illusion and our initial studies, we observed that an ideal strategy is to orient the noise such that the phase change is always aligned with local motion. We observed that directionally coherent local phase modulations integrate into the perception of global motion. To this end, we compute per-frame motion flow, and orient the Gabor patch according to the local motion direction.

4.2 Spatio-Temporal Frequency

It is important to limit our synthesized motion energy to spatio-temporal frequencies that are not resolvable in peripheral vision, so that the synthesized content is not perceptually objectionable. To this end, the contrast sensitivity function [Peli 2001] offers insights into the visibility of contrast as a function of factors such as eccentricity, spatial frequency, and temporal frequency. A simplified contrast sensitivity for a Gabor patch corresponding to the impulse i , centered at location x_i (concise notation for (x_i, y_i) in Eq. 2) can be defined as:

$$S(i) = S_{csf}(f_s(i), f_t(i), e(x_i)) \quad (3)$$

The function value $S(i)$ is the inverse of the contrast threshold (the minimum contrast that was measured to be resolvable). S_{csf} is a contrast sensitivity function, $f_s(i)$ is the spatial frequency of the Gabor patch in cycles-per-degree, $f_t(i)$ is the temporal frequency in hertz, $e(x_i)$ is the eccentricity (angular distance from the eye fixation location in degrees) of the location where the Gabor patch is centered. $S(i) < 1$ implies that a Gabor patch will not be resolvable, regardless of its contrast and amplitude. Given $e(x_i)$, our goal is to estimate $f_s(i)$ and $f_t(i)$ such that $S(i) = 1$. The temporal frequency is usually defined as the product of the spatial frequency and velocity ($f_t(i) = f_s(i)v(x_i)$). We have explicit control over the temporal frequency (through controlled phase modulation, as shown in Eq. 8). Therefore, we can introduce a scaling factor α on the local motion flow velocity of the foveated input $v(x_i)$ (degrees-per-second), that offers us additional parametric control over the spatio-temporal band we synthesize.

$$f_t(i) = \alpha f_s(i) \cdot v(x_i) \quad (4)$$

If $\alpha=1$, the temporal frequency will be exactly corresponding to $v(x_i)$. $\alpha < 1$ will make the synthesis more conservative, synthesizing lower temporal frequencies. Mantiuk *et al.* [2022] provide a unified

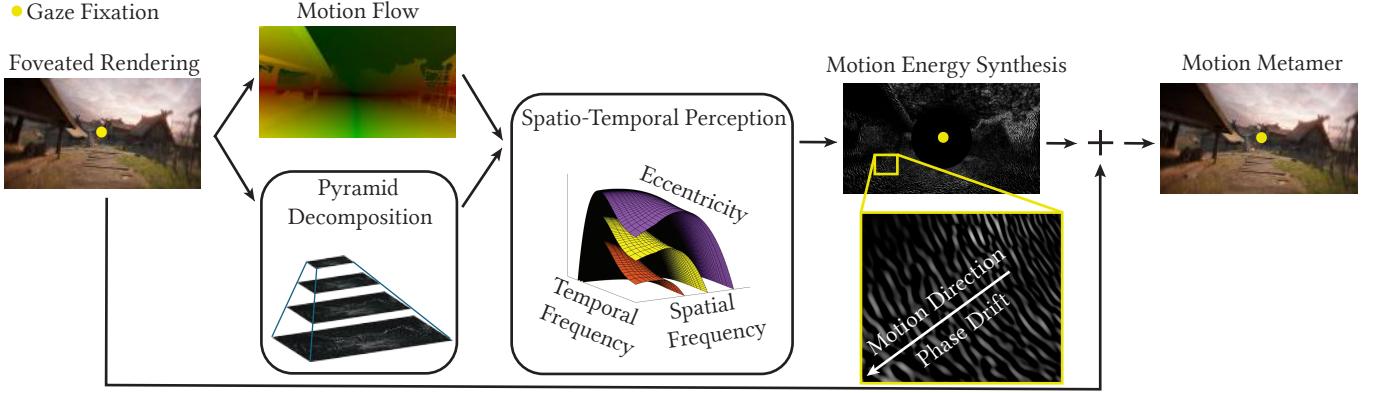


Fig. 4. Method Overview: Our method takes a foveated rendering frame as input. We first do a multi-scale decomposition of the input frame (Gaussian and Laplacian pyramid), and estimate motion flow. The obtained pyramids and motion flow are thereafter processed by our technique, which is centered around on the spatio-temporal frequency perception characteristics of the human visual system. Thereafter, we synthesize perceptually controlled directional motion energy using procedural Gabor noise, and super-impose it over the foveated frame.

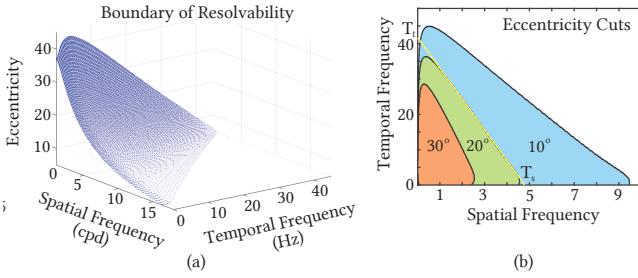


Fig. 5. Spatio-Temporal Resolvability. (a): The iso-sensitivity ($S_{csf}=1$) surface of stelaCSF [Mantiuk et al. 2022] defines a control boundary of spatio-temporal resolvability i.e., we want to limit our synthesis to the boundary and above. (b): The iso-eccentricity planar cuts of the surface.

framework for spatio-temporal contrast sensitivity i.e., stelaCSF. Using the stelaCSF, we can compute the iso-sensitivity surface ($S_{csf} = 1$) that represents limits/boundaries for spatio-temporal resolvability as a function of eccentricity (Figure 5 (a)) as:

$$B = \{(f_s, f_t, e) | S_{csf}(f_s, f_t, e) = 1\} \quad (5)$$

This boundary (B) represents spatio-temporal frequency and eccentricity triplets for which the contrast sensitivity function value is unity ($S_{csf} = 1$). The triplets above this boundary are not resolvable, and below this boundary are resolvable.

We aim to enforce that given the eccentricity $e(x_i)$, the spatio-temporal frequency pair $(f_s(i)$ and $f_t(i))$ lies on the boundary of visibility (B). As the spatial and temporal frequencies are not independent quantities (see Eq. 4), this problem cannot be solved via a trivial search. Hence, we rely on the observation that the iso-eccentricity curves (The projection of B for fixed eccentricity) in Figure 5 (b) are approximately linear for the most part, except very low spatial, and high temporal frequencies. Therefore, we model the boundary as a linear function.

$$f_t(i) = -\frac{T_t(e(x_i))}{T_s(e(x_i))} f_s(i) + T_t(e(x_i)) \quad (6)$$

where T_s and T_t are the eccentricity ($e(x_i)$) dependent intercepts of the linear model, derived from the contrast sensitivity function, as shown in Figure 5 (b). The practical validity of this approximation is grounded in the fact that the approximation we have made will (in the worst case) just act conservatively i.e., it will never push the synthesis inside the region of visibility. Given the local input velocity v , the display refresh-rate in frames-per-second R , and Eq. 4; to lie on the boundary of visibility, the spatial frequency can be computed by substituting Eq. 4 into Eq. 6, as:

$$f_s(i) = \frac{T_s(e(x_i))}{1 + \frac{T_s(e(x_i))}{T_t(e(x_i))} \cdot \alpha v(x_i)} \quad (7)$$

Note that Eq. 7 will perceptually adjust the synthesized spatial frequency based on the scaling of the velocity (α), offering us control over the synthesized spatio-temporal band. The adaption of the synthesized motion energy frequency with foveation and speed can be seen in Figure 10.

4.3 Phase-change rate

To compensate for the lost motion energy, we need to introduce a phase-change proportional to the underlying local input velocity v , which is provided by the motion flow. To this end, given the required temporal frequency $f_t(i)$ in hertz, we enforce the phase-change rate ϕ_v in radians-per-second as:

$$\phi_v(i) = 2\pi f_t(i) \quad (8)$$

which corresponds to a per-frame (k) phase increment (in radians) of each g_i (see Eq.1) as:

$$\phi_i^{k+1} = \phi_i^k + \frac{\phi_v(i)}{R} \quad (9)$$

where R is the refresh-rate of the display in hertz. An important consideration is motion aliasing, that will make the Gabor patches

appear to move in the opposite direction of the phase change, as shown in Figure 6(a). This phenomena is analogous to how a fast moving ceiling fan appears to rotate in the opposite direction of actual angular rotation of the blades, or how a slow camera shutter speed inverts the angular motion of helicopter blades. To avoid motion aliasing (respecting the Nyquist-rate), we need to ensure that:

$$|\phi_v(i)| < \pi R \quad (10)$$

This constraint introduces a lower bound on the frequency computation in Eq. 7 as:

$$\hat{f}_s(i) = \max \left\{ f_s(i), T_s(e(\mathbf{x}_i)) - \frac{R \cdot T_s(e(\mathbf{x}_i))}{2T_t(e(\mathbf{x}_i))} \right\} \quad (11)$$

ensuring that we don't synthesize spatial frequencies that require a temporal frequency greater than $\frac{R}{2}$ to be outside the visibility zone.

4.4 Amplitude

There are a few considerations that have to be made while estimating amplitude in a motion adaptive way. Firstly, the method needs to respect that the synthesized spatial frequencies adhere to natural image statistics in static content. This has been appropriately addressed by Tariq *et al.* [2022].

Secondly, as we are superimposing the synthesized energy onto the foveated video, it is important to account for the masking by background contrast [Legge and Foley 1980] and its variation with motion. Movement in the foveated scene will decrease contrast masking on our motion energy in a velocity dependent way, potentially increasing the visibility. Therefore we need to appropriately adjust the Gabor patch amplitude with velocity. The masking effect is strongest when the foreground and background contrast have similar spatial frequencies [Zeng *et al.* 2000]. For a Gabor patch with spatial frequency $f_s(i)$, the strongest masking signal will be contrast in the Laplacian pyramid level with the octave closest to $f_s(i)$. Furthermore, the higher the background contrast, the stronger the masking effect. For the required masking level $L(i)$ (nearest lower pyramid octave to $f_s(i)$), and its corresponding band central spatial frequency $f_L(i)$, we can estimate Michelson contrast as the ratio between the appropriate Laplacian (p_{lap}) and Gaussian (p_{gauss}) pyramid levels [Mantiuk *et al.* 2021].

$$C(\mathbf{x}_i, f_L(i)) = \frac{p_{lap}(\mathbf{x}_i, L(i))}{p_{gauss}(\mathbf{x}_i, L(i) + 1) + \epsilon} \quad (12)$$

The Gaussian pyramid level is one higher to account for local adaptation [Mantiuk *et al.* 2021]. Now, this masking contrast can be perceptually normalized by the CSF as:

$$C_n(\mathbf{x}_i, f_L(i)) = S_{csf}(f_L(i), f_t(i)) \cdot C(\mathbf{x}_i, f_L(i)) \quad (13)$$

We model the masking as the threshold elevation of contrast due to the background [Tursun *et al.* 2019]:

$$a_c(\mathbf{x}_i) = 1 + \frac{1}{|N_{hood}|} \sum_{p \in N_{hood}} |C_n(p, f_L(i))|^{\beta} \quad (14)$$

Where the N_{hood} is the 3x3 neighborhood around \mathbf{x}_i , and the parameter β controls the slope of the masking transducer. This masking

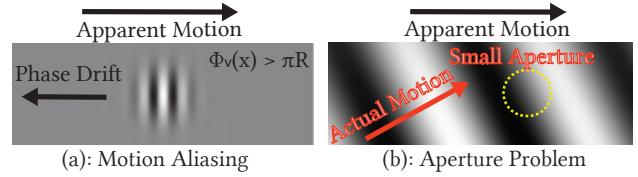


Fig. 6. Motion Considerations (a): The Nyquist-Rate must be respected so that motion aliasing is avoided. (b): The aperture (Gaussian window) size must be adjusted such that the aperture problem is avoided.

$a_c(\mathbf{x}_i)$ will appropriately decrease the amplitude when the velocity increases, accounting for the decrease in masking.

Lastly, the factors that account for natural statistics (see Section 4.2 by Tariq *et al.* [2022]) and motion dependent masking can be combined to form our amplitude estimation model:

$$K(i) = \gamma \cdot |l(\mathbf{x}_i, L_a(\mathbf{x}_i))| \cdot a_c(\mathbf{x}_i) \quad (15)$$

which can be parametrically controlled using a scaling factor γ . The adaption of the synthesized motion energy amplitude with foveation and speed can be seen in Figure 10.

4.5 Gabor Patch Size

In our phase based motion synthesis, the sinusoidal function undergoes an internal drift through the Gaussian aperture window. It is important that Gaussian aperture is not too small, as this is known to be potentially detrimental to perceiving the motion direction correctly due to the aperture problem [Kane *et al.* 2011], as shown in Figure 6(b). This issue can be even more problematic as eccentricity increases, as we are synthesizing lower frequencies. Thankfully, the receptive pooling field size in human peripheral vision increases linearly with eccentricity [Freeman and Simoncelli 2011]. This implies that without perceptual degradation, we have the freedom to increase the size of Gabor patch as eccentricity increases. We increase the Gaussian aperture radius linearly with eccentricity such that the Gabor patch captures at-least two cycles of the sinusoid.

5 REAL-TIME IMPLEMENTATION

We implemented our technique on the Universal Render Pipeline (URP) in Unity 3D. The implementation was centered around four main shaders; the motion flow shader, parameter estimation shader, the phase update shader, and the noise synthesis shader. The input foveated render is processed with a standard display photometry and geometry model. We assume constant pixels per visual degree throughout the field-of-view. The values are linearized, and luminance and color channels are separated using the YUV colorspace transformation. The processing is applied only to the linearized Y channel, with the color channels added back (same color as the input foveated render) at the end. The motion flow is computed using the full-resolution geometry the camera model-view-projection matrices i.e., the interpolated movement of vertices projected onto the image plane. The parameter estimation shader employs the motion flow and pyramid decomposition to estimate the four Gabor patch parameters (orientation, amplitude, frequency, phase) for each pixel location, returning them in the form of an RGBA render texture.

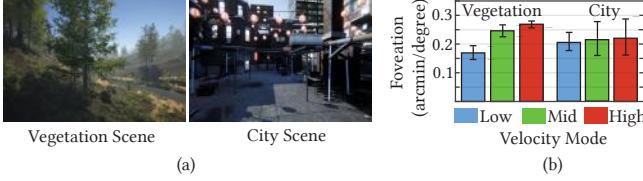


Fig. 7. (a): The scenes used in our evaluation (b): The measured foveation thresholds for the two scenes for different velocity settings

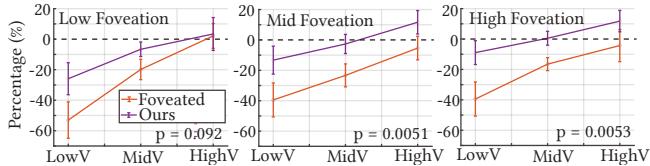


Fig. 8. Velocity User-Study Results: Percentage difference in velocity perception with respect to the full-quality reference for foveated and our technique. 0% means no difference with the reference velocity. <0% means perceived slower compared to the reference. >0% means perceived faster than the reference. The p-values for pairwise t-test's between the foveated and ours are reported. The error bars represent standard error (SE).

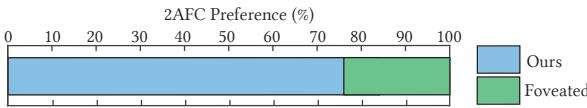


Fig. 9. 2AFC Quality Study: The average preference of our method over the foveated for a full-reference 2AFC quality study run over 7 participants. The study was run on both scenes with the "Mid" and "High" foveations across all three velocity settings

The phase update shader stores the current phase for each Gabor patch and updates it according to the parameter estimation shader's returned phase values. The noise synthesis shader inputs the parameter texture (returned from the parameter estimation shader) and synthesizes the per-frame motion energy by generating impulses and sampling the parameter textures at the impulse locations. Each Gabor patch (corresponding to its unique impulse location) uses the sampled parameters throughout its support. The support size is capped to 3σ of the Gaussian envelope, therefore the computation in Eq. 2 is simplified to only considering the contribution of Gabor patches that influence a particular point (the point is within their support). The whole pipeline (including the rendering of the underlying scenes) runs in real-time, with frame-rates of about 75-90fps for different 4K resolution scenes we tested on, on a machine with an NVIDIA GTX 4090.

For our validation, we set $\alpha = 1$, $\beta = 1$, and $\gamma = 3$ as the model parameters. The standard parameter values of α and β were chosen to represent the effectiveness and accuracy of the standard base perceptual model we developed, as the model becomes non-parametric with the specified values. The value of γ was chosen empirically.

6 VALIDATION

Through a user-study, we validate the effectiveness of our method to compensate for the loss in motion perception due to loss of spatial details in periphery.

6.1 Subjective Experiment

We follow a similar procedure from the earlier experiment (Section 3) and test the participants' ability to match the perceived speed between a test video and a reference. As the reference, we use full-resolution rendering, while the two test versions include standard foveated rendering and the same foveated rendering with super-imposed motion energy.

6.1.1 Hardware. Most readily available VR-HMDs provide limited spatial resolution [Beams et al. 2020]. To ensure high-quality images in the far periphery, we used a large 4K LG OLED55CX 55-inch 120Hz screen. The position of the participants was fixed using a chin-rest 71.5 cm from the screen, resulting in an 80-degree field-of-view and the possibility of replicating spatial frequencies up to 24 cpd at the center of the screen. We tracked the participants' gaze using the Tobii Pro Spectrum 600Hz eye tracker. To maintain frame-rate, we fixed the the display refresh-rate to 60 Hz . The peak luminance was set to 170 cd/m². All trials were conducted under similar low ambient lighting conditions.

6.1.2 Stimuli. We use high-quality 4K rendering of two scenes, VEGETATION and CITY (Figure 7a). We tested different types of camera trajectories. VEGETATION scene was rendered using a linear camera path, while CITY scene used a circular path around a fixed axis. Each scene was rendered with three different camera velocities representing low, medium, and high motion cases. To choose the corresponding velocities, we analyzed screen-space motion flow in rendered animations and adjusted the camera speed in the high-velocity case such that the motion flow magnitudes reached approximately 30-35 degrees per second. We chose this as a case in which subjects are usually able to follow objects using smooth pursuit eye movements [Daly 2001]. Higher velocities would make such a task problematic. The medium and low velocity cases were set to be 2/3 and 1/2 of the high speed, respectively.

Literature suggests that foveated rendering can be successfully simulated using Gaussian blur [Albert et al. 2017; Hoffman et al. 2018]. Consequently, in our experiments, we simulate the process of shading-rate reduction and anti-aliasing with eccentricity-dependent Gaussian blur. To focus on the peripheral vision, the content was removed from a circle with 15 visual degree radius around the gaze location. A linear increase of blur with respect to the eccentricity beyond this region was used to model foveation. We define the amount of foveation using the rate with which the blur increases with eccentricity. We test different blur rates to evaluate our method on various foveations. For our experiments, we consider the lowest foveation rate to be close to the visibility threshold. To measure, we conducted a small, informal experiment where three participants naive to the subject of the experiment adjusted the blurring rate to a just-noticeable one. Since the just-noticeable foveation depends on the content [Tursun et al. 2019] and the velocity [Lisboa et al. 2023b], we perform the measurements for each scene and camera velocity

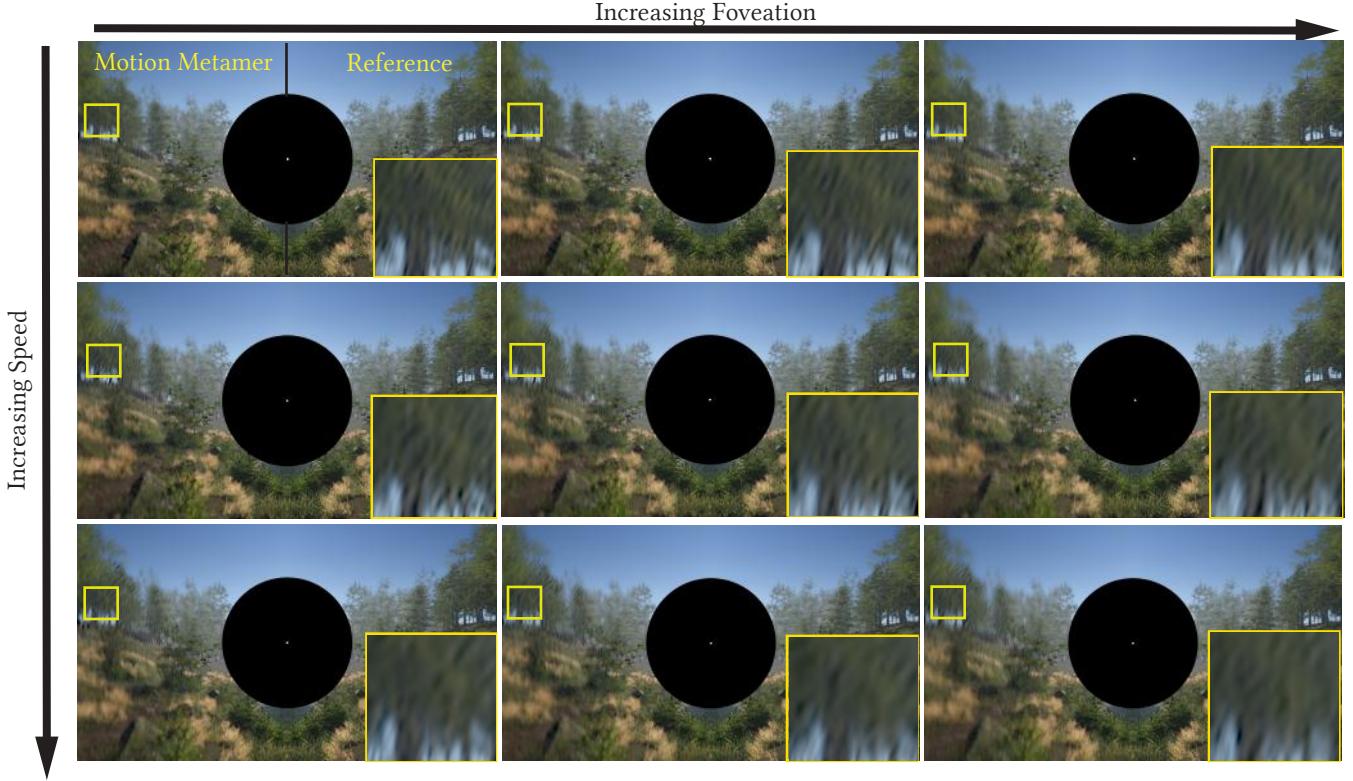


Fig. 10. Motion Metamer Frames: Sample frames from our technique (left half image), and the corresponding reference (right half image). The perceptual adaption of motion energy characteristics with foveation and speed can be seen in the snippets.

(Figure 7b). We refer to these rates as the Low Foveation case. The rates of two other considered foveations, i.e., Mid and High, were set as 25% and 50% higher than the measured just-noticeable rates. For a preview of the stimuli, we refer to the supplementary.

6.1.3 Task. The setup for the experiment is shown in Figure 2a. In each trial, users were shown two sequences played side by side. The sequence on the right was the reference scene in full resolution, with the camera moving at a constant, low, medium, or high speed. The sequence on the left presented a foveated version of the scene on the right, either with or without our method applied. At the beginning of each trial, the camera velocity of the left scene was set to random, and the task of participants was to adjust it to match the velocity of the reference sequence. Users were instructed to do this by scrolling the mouse wheel up or down (to adjust the camera velocity of the scene on the left) and confirming the choice with a keyboard key. Viewers were instructed to focus on a cross displayed at the center of the screen to ensure proper viewing. A blank screen was shown when the eye tracker detected the users' gaze location moving away. We repeat the trials (in randomized order) for all combinations of the three different foveation levels (Low, Mid, and High) and the three different velocity settings (Low, Mid, and High). We divided the experiment into three sessions to mitigate fatigue. Each of them took an average of 15–20 minutes. 14 participants (all university students) with normal or normal-corrected vision took

part in the study, 9 male and 5 female with ages in the range of 19–25.

6.2 Results and Discussion

The results of our main study on velocity perception are shown in Figure 8. The plots show the percentage loss in velocity perception with-respect-to the full-quality reference velocity. The p-values for pairwise t-test's between the foveated and ours are reported. The error bars represent standard error (SE). As it can be seen, we were able to show strong statistical significance for the "Mid" and "High" foveation cases, and not so much so for the "Low" foveation. Firstly, we report that there is a consistent loss in velocity perception due to loss of spatial details in peripheral vision. Interestingly, the perceived loss in motion cues decreases as the test velocity increases. We believe that this may be because of the fact that due to motion blur, users lose acuity to high frequency details in the reference as velocity increases, decreasing the perceived spatio-temporal mismatch between the videos. Less perceived mismatch implies that less velocity compensation is required.

Secondly, we report that as per our hypothesis, our synthesis of spatio-temporal motion energy helps significantly in mitigating the loss in peripheral motion cues. In fact, users actually may even perceive the super-imposed content to be faster than the reference itself for high velocities, demonstrating the perceptual strength

of the technique. We believe that the α , β , and γ parameters can be calibrated better in the future to mitigate the overshoot. For example, if α is decreased, it will decrease the perceived motion energy, mitigating the overshoot.

Lastly, an extremely interesting and thought provoking observation is that a loss in motion cues may be observed even at foveation levels which lie at the threshold of detection (Low Foveation). This is interesting and warrants further investigation, as it may reveal that even when there are no spatially visible differences between stimuli, they may elicit different motion perception. This observation may be explained by the hypothesis that there is a separation of spatial and motion processing channels in the human visual system [Billock and Harding 1996], and spatial equivalence may not always imply an equivalence of motion perception. Consider a simple thought experiment to get more context, assume you have a high-quality reference video, and you process each frame separately with two different spatial metamer synthesis techniques (e.g. [Freeman and Simoncelli 2011; Tariq et al. 2022; Walton et al. 2021]) to get two spatial metamer videos. Now spatially, each frame in either of the two videos will be perceptually indistinguishable from the corresponding frame in the high-quality reference. However, will the two metamer videos evoke the same perception of velocity and motion as the high-quality reference? We believe that there need to be dedicated studies investigating the relationship between spatial metamerism/equivalence and the accurate representation of perceptual cues such as motion and depth. Such a study will improve our understanding of the relationship between perceptual cues related to different aspects of perceived realism (depth, motion, spatial, dynamic range etc), and it might give us a more formalized insight into why preserving perceived spatial fidelity does not automatically imply the accurate reproduction of other aspects of our visual perception.

6.3 Quality and Temporal Consistency

We took care in the design of our technique to ensure that the synthesized motion energy does not cause objectionable artifacts. The participants of our velocity user-study also did not report observing any "unnatural" things during their trials, except for the fact that some stimulus was perceived to be "blurry" (likely the foveation). To validate this aspect even further, we performed a small 2AFC experiment. With the same setup as our main study, during each trial, the participants had to choose which option "is perceptually closer to the reference in terms of quality and resolution". The participants could switch between the two options on the left half of the screen (foveated and ours) using the SPACE key, with a 500ms black screen during the switch (so that the temporal change between the stimuli does not bias them). The choices were presented in random order per-trial. The experiment included both scenes (Vegetation and City), for the "Mid" and "High" foveations over all three velocity settings (Low, Mid, and High). This amounts to a total of 12 trials per participant.

The results of our study are reported in Figure 9, showing that a majority of judgments were made in the favor of our technique.

7 LIMITATIONS AND FUTURE WORK

We venture into a topic that is largely unexplored, so naturally there is a lot of potential for interesting future work. Firstly, we believe that more sophisticated calibration (α , β , γ) can bring our framework significantly closer to the synthesis of an idealized motion metamer i.e., perceived exactly the same speed as a full-quality reference, irrespective of content and motion magnitude.

Secondly, as we believe that the peripheral vision system plays the central role in motion perception, we limited our analysis to understanding the characteristics of motion perception in peripheral vision (we blacked out the foveal region). A natural extension would be to study how the fovea and periphery interact together to produce a coherent representation of perceived motion. An approach to study this could be to separately repeat our experiments for only foveal region, and the full field-of-view scenes (no fovea blacking). This could give us a better picture of the foveal and peripheral interaction for motion perception. The experiment could also be repeated for different peripheral regions radius's to also distinguish between the motion perception characteristics of the near and far periphery.

Thirdly, there needs to be more investigation into the interplay between spatial and temporal statistics in visual metamerism. An approach could be to study different spatial metamer synthesis techniques (e.g. [Freeman and Simoncelli 2011; Tariq et al. 2022; Walton et al. 2021]) and try to model the differences in velocity perception among them, which could provide important clues for a more formalized theory of motion metamers. Furthermore, a comprehensive content and viewing condition dependent model that can quantify the loss in velocity perception and a function of peripheral blur would be greatly beneficial to the community. Also, a study on the effects of loss in motion cues on downstream task performance would be very interesting. Lastly, there were some very subtle banding discontinuities sometimes in the periphery which are perceptually invisible in peripheral vision, so we did not introduce additional rendering cost to blend them.

It is also important to clarify that though we tested our technique on two completely orthogonal types of global camera movements (linear and rotation), we did not validate for more complicated motion. As our method parameters are adaptive to local spatio-temporal statistics of the content, we do not foresee any bottlenecks in its efficacy for application to local motion (though we did not validate for local movement settings). However, an aspect we do not cater for during our analysis is different types of acceleration. It would be an extremely interesting future study to see if metamers that preserve the perception of velocity also induce necessary acceleration perception cues.

8 CONCLUSIONS

Contemporary work on foveated rendering has collectively been pushing towards a singular goal of reducing quality in the periphery such that the result is spatially indistinguishable from full-quality synthesis. However, visual perception goes beyond that. It incorporates elements such as space, time, depth, motion, and more complicated high-level processes that all interact and help us perceive the surrounding world. For foveated rendering to become a key enabler of realistic and comfortable VR/AR experiences, it is necessary to

consider this broad spectrum of cues governing human perception. This work focuses on a small part of this endeavor yet an essential element of visual reality, i.e., motion perception. We demonstrate that foveated rendering may inhibit motion perception, potentially making immersive realities appear slower than they actually are. This problem is critical for immersive simulations where motion perception drives the decision-making process. Drawing inspiration from a fascinating visual illusion, we design a perceptually motivated real-time technique that synthesizes controlled motion energy to offset the loss in motion perception without introducing objectionable quality degradation.

ACKNOWLEDGMENTS

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 804226 PERDY).

REFERENCES

- Edward H. Adelson and James R. Bergen. 1985. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America. A, Optics and image science* 2 2 (1985), 284–99.
- 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).
- Horace B. Barlow. 1961. Possible principles underlying the transformation of sensory messages. *Sensory communication* 1, 01 (1961).
- Ryan Beams, Brendan Collins, Andrea Seung Kim, and Aldo Badano. 2020. Angular Dependence of the Spatial Resolution in Virtual Reality Displays. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2020), 836–841.
- Vincent A. Billock and Thomas H. Hardig. 1996. Evidence of spatial and temporal channels in the correlational structure of human spatiotemporal contrast sensitivity. *The Journal of Physiology* 490 (1996).
- William F. Broderick, Gizem Rufo, Jonathan A. Winawer, and Eero P. Simoncelli. 2023. Foveated metamers of the early visual system. *bioRxiv* (2023).
- Petr Kellnhofer, Brooke Krajancich, and Gordon Wetzstein. 2021. A perceptual model for eccentricity-dependent spatio-temporal flicker fusion and its applications to foveated graphics. *ACM Transactions on Graphics (TOG)* 40 (2021), 1 – 11.
- D C Burr, John Ross, and Maria Concetta Morrone. 1986. Seeing objects in motion. *Proceedings of the Royal Society of London. Series B. Biological Sciences* 227 (1986), 249 – 265.
- Scott Daly. 2001. Engineering observations from spatiotemporal and spatiotemporal visual models. In *Vision Models and Applications to Image and Video Processing*. Springer, 179–200.
- 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.
- Bernhard Josef Lachenmayer E. Hartmann and Hans Brettl. 1979. The peripheral critical flicker frequency. *Vision Research* 19 (1979), 1019–1023.
- Sigm. Exner. 1886. Ein Versuch über die Netzhautperipherie als Organ zur Wahrnehmung von Bewegungen. *Archiv für die gesamte Physiologie des Menschen und der Tiere* 38 (1886), 217–218.
- David Finlay. 1982. Motion Perception in the Peripheral Visual Field. *Perception* 11 (1982), 457 – 462.
- Jeremy Freeman and Eero P. Simoncelli. 2011. Metamers of the ventral stream. *Nature neuroscience* 14, 9 (2011), 1195–1201.
- William T. Freeman, Edward H. Adelson, and David J. Heeger. 1991. Motion without movement. *SIGGRAPH* (1991).
- Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. 2012. Foveated 3D graphics. *ACM Transactions on Graphics (TOG)* 31, 6 (2012).
- 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.
- David Kane, Peter J. Bex, and Steven C. Dakin. 2011. Quantifying "the aperture problem" for judgments of motion direction in natural scenes. *Journal of vision* 11 3 (2011).
- 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 Transactions on Graphics (TOG)* 38, 6 (2019).
- G.S. Klein. 1941. The relation between motion and form acuity in para-foveal and peripheral vision and related phenomena. *Archives of Psychology* (1941).
- Brooke Krajancich, Petr Kellnhofer, and Gordon Wetzstein. 2023. Towards Attention-aware Foveated Rendering. *ACM Transactions on Graphics (TOG)* 42 (2023), 1 – 10.
- Ares Lagae, Sylvain Lefebvre, George Drettakis, and Philip Dutré. 2009. Procedural noise using sparse Gabor convolution. *ACM Transactions on Graphics (TOG)* 28, 3 (2009).
- Gordon E. Legge and John M. Foley. 1980. Contrast masking in human vision. *Journal of the Optical Society of America* 70 12 (1980), 1458–71.
- Thállys Lisboa, Horácio Macêdo, Thiago Porcino, Eder Oliveira, Daniela Trevisan, and Esteban Clua. 2023b. Is Foveated Rendering Perception Affected by Users' Motion?. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 1104–1112.
- Thállys Lisboa, Horácio Macêdo, Thiago Malheiros Porcino, Eder de Oliveira, Daniela Trevisan, and Esteban Walter Gonzalez Clua. 2023a. Is Foveated Rendering Perception Affected by Users' Motion? *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2023), 1104–1112.
- Rafal Mantiuk, Malika Ashraf, and Alexandre Chapiro. 2022. stelaCSF: a unified model of contrast sensitivity as the function of spatio-temporal frequency, eccentricity, luminance and area. *ACM Transactions on Graphics (TOG)* 41 (2022), 145:1–145:16.
- Rafal 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 Transactions on Graphics (TOG)* 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).
- Bipul Mohanto, A. B. M. Tariqul Islam, Enrico Gobbetti, and Oliver Staadt. 2021. An integrative view of foveated rendering. *Computers & Graphics* (2021).
- Fredo Durand, Neal Wadhwa, Michael Rubinstein, and William T. Freeman. 2013. Phase-based video motion processing. *ACM Transactions on Graphics (TOG)* 32 (2013), 1 – 10.
- Anjul Patney, Marco Salvi, Joohwan Kim, Anton Kaplanyan, Chris Wyman, Nir Bentj, David Luebke, and Aaron Lefohn. 2016. Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics (TOG)* 35, 6 (2016).
- Eli Peli. 2001. Contrast sensitivity function and image discrimination. *Journal of the Optical Society of America. A, Optics, image science, and vision* 18 2 (2001), 283–93.
- 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.
- R Sekuler. 1875. Visual motion perception. *Handbook of Perception* (1875).
- 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.
- Leland S. Stone and Peter Thompson. 1992. Human speed perception is contrast dependent. *Vision Research* 32 (1992), 1535–1549.
- Luca Surace, Cara Tursun, Ufuk Celikcan, and Piotr Didyk. 2023. Gaze-Contingent Perceptual Level of Detail Prediction. *Eurographics Symposium on Rendering (EGSR)* (2023).
- Taimoor Tariq, Cara Tursun, and Piotr Didyk. 2022. Noise-based enhancement for foveated rendering. *ACM Transactions on Graphics (TOG)* 41 (2022), 1 – 14.
- 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).
- 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 Transactions on Graphics (TOG)* 38, 4 (2019).
- Jan P. H. van Santen and George Sperling. 1985. Elaborated Reichardt detectors. *Journal of the Optical Society of America. A, Optics and image science* 2 2 (1985), 300–21.
- 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 Transactions on Graphics (TOG)* 40, 4 (2021).
- Andrew B. Watson. 1998. Toward a perceptual video-quality metric. In *Electronic imaging*.
- Andrew B. Watson. 2013. High Frame Rates and Human Vision: A View through the Window of Visibility. *Smpte Motion Imaging Journal* 122 (2013), 18–32.
- Andrew B. Watson and Albert J. Ahumada. 1985. Model of human visual-motion sensing. *Journal of the Optical Society of America. A, Optics and image science* 2 2 (1985), 322–41.
- Reichardt Werner. 1961. Autocorrelation, a principle for the evaluation of sensory information by the central nervous system.
- Wenjun Kevin Zeng, Scott J. Daly, and Shawmin Lei. 2000. Point-wise extended visual masking for JPEG-2000 image compression. *International Conference on Image Processing (ICIP)* 1 (2000), 657–660 vol.1.