

Perception-driven Accelerated Rendering

Weier, M.; Stengel, M.; Roth, T.; Didyk, P.; Eisemann, E.; Eisemann, M.; Grogorick, S.; Hinkenjann, A.; Kruijff, E.; Magnor, M.

DOI

[10.1111/cgf.13150](https://doi.org/10.1111/cgf.13150)

Publication date

2017

Published in

Computer Graphics Forum (online)

Citation (APA)

Weier, M., Stengel, M., Roth, T., Didyk, P., Eisemann, E., Eisemann, M., Grogorick, S., Hinkenjann, A., Kruijff, E., Magnor, M., Myszkowski, K., & Slusallek, P. (2017). Perception-driven Accelerated Rendering. *Computer Graphics Forum (online)*, 36(2), 611-643. <https://doi.org/10.1111/cgf.13150>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Perception-driven Accelerated Rendering

M. Weier^{1,2} M. Stengel^{3,4} T. Roth^{1,5} P. Didyk^{2,6} E. Eisemann³ M. Eisemann⁷
S. Grogorick⁴ A. Hinkenjann¹ E. Kruijff¹ M. Magnor⁴ K. Myszkowski⁶ P. Slusallek^{2,8,9}

¹Bonn-Rhein-Sieg University of Applied Sciences, ²Saarland University, ³TU Delft, ⁴TU Braunschweig, ⁵Brunel University London, ⁶MPI Informatik Saarbrücken, ⁷TH Köln, ⁸Intel Visual Computing Institute, ⁹German Research Center for Artificial Intelligence (DFKI)

Abstract

Advances in computer graphics enable us to create digital images of astonishing complexity and realism. However, processing resources are still a limiting factor. Hence, many costly but desirable aspects of realism are often not accounted for, including global illumination, accurate depth of field and motion blur, spectral effects, etc. especially in real-time rendering. At the same time, there is a strong trend towards more pixels per display due to larger displays, higher pixel densities or larger fields of view. Further observable trends in current display technology include more bits per pixel (high dynamic range, wider color gamut/fidelity), increasing refresh rates (better motion depiction), and an increasing number of displayed views per pixel (stereo, multi-view, all the way to holographic or lightfield displays). These developments cause significant unsolved technical challenges due to aspects such as limited compute power and bandwidth. Fortunately, the human visual system has certain limitations, which mean that providing the highest possible visual quality is not always necessary. In this report, we present the key research and models that exploit the limitations of perception to tackle visual quality and workload alike. Moreover, we present the open problems and promising future research targeting the question of how we can minimize the effort to compute and display only the necessary pixels while still offering a user full visual experience.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation

1. Introduction

The dream of presenting a computer-generated scene in a convincing and compelling way that cannot be distinguished from the real world in real-time remains one of the key challenges for computer graphics. Displaying realistic-looking graphics at high refresh rates is computationally intense, particularly for high pixel densities, a wide field-of-view and stereo rendering.

In recent years, we have seen strong trends towards displays with increasing size, resolution and dynamic range, leading to higher pixel densities and the ability to display high dynamic range (HDR) content. While a resolution of 1080p has been a widespread standard in TVs as well as in desktop displays for several years, we are currently observing a shift towards 4k displays with up to 80 inches, increasing the field of view (FOV) of observers at typical viewing distances while maintaining or increasing pixel density. Moreover, display resolutions for handheld devices such as smartphones constantly increase, currently supporting pixel densities up to 800 pixels-per-inch (PPI). This enables clear readability of text and playback of high-resolution videos. In addition, we are seeing increasing refresh rates to allow for better motion depiction and an increasing number of displayed views per pixel (stereo, multi-view, all the way to holographic or lightfield displays). Large, high-

resolution, projection-based displays and high-resolution tiled display walls have become well-established installations in research institutions around the world. Furthermore, mass production has introduced a range of high-quality head-mounted displays (HMDs) with a wide FOV available on a commodity level, gaining interest from researchers in the fields of Virtual Reality (VR) and Augmented Reality (AR). High pixel densities and high refresh rates are crucial for immersion and interaction, especially in the context of VR. Thus, we consider this domain to be well-suited for demonstrating the new challenges posed in the field of computer graphics and image synthesis. The aforementioned advances in display technologies will further tighten the requirements on rendering techniques to produce realistic images. Already today, the achieved realism is strongly limited by hardware capabilities, while often many desirable but costly aspects of reality are not even considered, including real-time global illumination, accurate depth of field, motion blur or spectral effects. It is obvious that this divergence will cause significant issues on the technical side (e.g., limited compute power and bandwidth), which are yet unsolved.

One way to tackle these challenges is by taking perceptual aspects into account when rendering images, a field we refer to as *perception-driven rendering*. Perception-driven rendering is based

on a close understanding of the human visual system (HVS) to improve the quality, speed and comprehensibility of images. Although the HVS seems to process complete images at a very high quality not bound to a fixed frame rate, it has several limitations. Visual input passes through optics, where it is sampled and then filtered on the retina before it is compressed and transmitted over the optical nerves. It is processed on various levels, including the brain's ability to access and use memory. One particular limitation is visual acuity, where the acuity of human vision is at its maximum only inside a very small central region of the FOV. This limitation means that rendering at a much lower geometric or spatial quality in the peripheral regions might go unnoticed. Thus, perception-driven rendering techniques can be created that exploit essential properties of the HVS and thus optimize rendering pipelines for increased performance by omitting details in the peripheral visual area. Also, visual quality within the area of sharp vision could be improved by reallocating some of the computational resources from peripheral regions. One of the key challenges when accelerating rendering techniques in such a way is to reduce the computational effort by computing only a subset of an image's pixels or by reducing the underlying scene description's complexity, while still maintaining a high-quality visual experience. Further benefits of perception-driven rendering approaches may include the reduction of nausea and motion sickness, particularly relevant for VR systems.

In this state-of-the-art report, we provide an overview of relevant properties of the HVS and describe how different rendering techniques tap into specific perceptual mechanisms through a structural approach. We give a comprehensive overview of the main research areas, classify the different techniques by the underlying mechanisms and present open problems as well as promising future research directions.

2. Scope and Structure

This paper comprises perception-driven rendering methods that accelerate realistic rendering. We refer to realistic rendering as methods that synthesize views from virtual 3D descriptions in order to resemble real-world scenes and objects. Hence, we do not discuss non-photorealistic rendering and information visualization. Stereo and low-latency rendering are also not discussed specifically. We consider methods from these areas not to be perception-driven, but rather perception-targeting, as they may support the implementation of actual perception-driven methods, but do not exploit perceptual limitations themselves. While several exciting new technologies have been introduced recently, e.g., new developments for HMDs, most of the methods discussed here are applicable in a more general way and not restricted to specific devices.

Performance and quality can often be subject to a trade-off. In our understanding acceleration can either mean improving the performance of a method or improving image quality while maintaining performance. Critical to a system's performance and its impact on perception is not only a method's bandwidth, i.e., the number of frames per second but also its latency. The latter constitutes how long it takes for input to be translated to visual output. With this focus, we contribute to the field by giving an overview of novel methods that have not been targeted by other related state-of-the-art papers so far. Masia et al. [MWDG13] present a survey on com-

putational displays. They leverage known aspects of the HVS to provide apparent enhancements beyond the physical limits of the display. However, an in-depth discussion of rendering methods is missing. Even though the work by McNamara et al. [MMBH10] discusses similar subjects, it does not focus on the specific aspects of accelerated rendering. Moreover, enhancements in technologies like HMDs paved the way for a lot of new developments in the last years, which are not covered by their report. The work by Corsini et al. [CLL*13] is entirely focused on geometric processing. They discuss different perception-oriented metrics for static and dynamic triangle meshes. Although adapting the geometric complexity of a 3D scene is one approach used in perception-driven accelerated rendering, other aspects are equally important, e.g., temporal coherence [SYM*12]. We cover a broader set of techniques and discuss them in the context of perception-driven rendering.

In this paper, we focus on a set of three research questions, targeting perception-driven accelerated rendering. We have derived these questions directly from the process of developing a rendering technique based on exploiting perceptual mechanisms and limitations. With the premise that the general concept of a rendering method is already present, the following steps have to be taken. First, focus has to be put on getting a clear understanding of what limitations and properties of the human visual system are relevant for implementing this idea. Second, one or several specific properties/limitations have to be chosen and resembling models have to be built. Third, an understanding is required of all physiological and technical steps involved in implementing the idea as an accelerated rendering technique. As such, it describes and demonstrates how limitations of the human visual system can be used to accelerate rendering. Based on this three-step process, we propose the following research questions, which structure this report:

- **RQ1:** What are the limitations of human visual perception?
- **RQ2:** How are these limitations modeled?
- **RQ3:** How are limitations used in current state-of-the-art methods to accelerate rendering?

Based on these research questions, we derive the following structure of the paper. We discuss the most important physiological and perceptual aspects of human vision in Sec. 3. A coarse model of the HVS is introduced to approach the most important parts of the visual processing pipeline from a higher level. To answer **RQ1**, we present relevant limitations of human visual perception and derive methods to describe them. Moreover, we explain limitations that have not been exploited so far but could be subject to further research. In the course of answering **RQ2**, we present different models that are commonly used within rendering systems to describe the visual system at its early processing stages (Sec. 4). Afterwards, we move to higher-level cognitive aspects such as visual attention and gaze prediction (Sec. 5). Such selective treatment of visual signal processing in the brain mirrors the current state of knowledge in visual perception. Early vision modeling is relatively robust, while intermediate processing remains less explored and cognitive aspects are mostly limited to visual saliency, the perceptual property that makes some visual stimuli stand out from their neighbors. Existing models for the latter generally focus on salient image feature modeling (often akin to early vision understanding), while volitional, task-driven aspects are notoriously more difficult

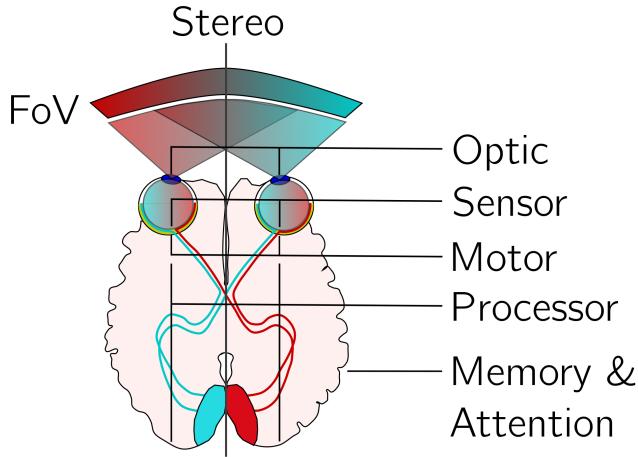


Figure 1: Model of the HVS. High-level model of the basic components responsible for human perception.

to handle. Answering **RQ3**, the main part of our work presents current state-of-the-art methods that exploit known limitations of the HVS. Techniques are divided into those entirely relying on perceptual or attentional models (Sec. 6) and those that include active measurements (Sec. 7), primarily gaze-tracking.

3. The Human Visual System and its Limitations

In this section, we describe the main principles of the HVS. Following the main approach taken in this paper, we highlight limitations that can be applied to optimize rendering techniques (**RQ1**). A model of the basic functions of the HVS is illustrated in Fig. 1. The HVS model provides an overview of the different stages a visual stimulus passes before it is processed and perceived by the user. Detailed information on the HVS from a psychophysical point of view can be found in the book by Adler et al. [AKLA11]. Following Fig. 1 from top to bottom, light enters our eyes constituting two data streams that enable us to process stereoscopic imaging over a FOV that encompasses zones with different stimuli sensitivities. The optical system passes the stimuli to the *retina* (the “sensor”), which is connected to the visual pathways. This connection transports signals from the eye to the visual cortex, where different parts of the brain are involved in processing and interpreting the signals until a final mental representation of the environment is produced: the image perceived by the user. In this process, memory as well as attention play a key role. Below, we go through each component of Fig. 1, discuss the findings regarding the physiological and perceptual properties of the HVS and describe their limitations.

3.1. Optics

The HVS is characterized by several unique *optical* qualities that are a result of both the position and shape of the eyes. With binocular vision and both eyes looking straight ahead, humans have a horizontal FOV of almost 190° . If eyeball rotation is included, the horizontal FOV extends to 290° [HR95, p.32]. While the human eye will receive visual stimuli over the full extent of the FOV, the way stimuli are processed in different parts of the visual field is

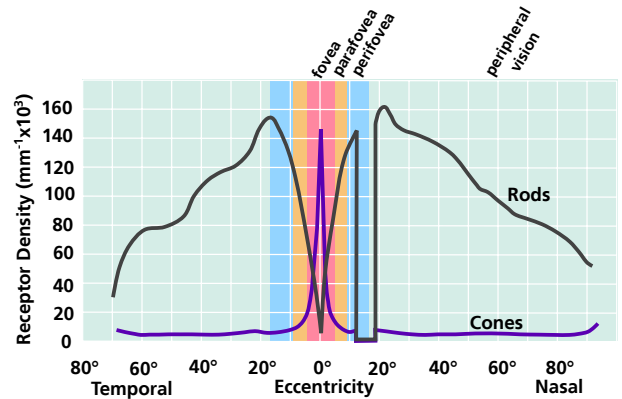


Figure 2: Retinal photoreceptor distribution. Image adapted from Goldstein [Gol10b, p. 51]

highly affected by the spatially varying properties of the retina. There are striking differences between central vision in comparison to the near and far periphery [CSKH90].

The distance between the pupils, the inter-pupillary distance (IPD), results in two streams of visual stimuli from slightly different view points, which are combined in the brain by a process called stereopsis and enable perception of depth also referred to as *stereo vision* [Pal99, Chapter 5.3]. Depth perception is additionally enabled by visual cues such as parallax, occlusion, color saturation and object size [CV95, HCOB10].

An effect that is relevant to the spatial acuity of the HVS is caused by the eye’s optics. It is known from sampling theory that *aliasing* occurs if a signal contains frequencies higher than the observer’s Nyquist frequency [Sha49]. In human vision, this under-sampling effect occurs for spatial frequencies higher than approx. 60 cycles per degree (cpd) [Wan95, p.24]. A cpd is a unit to describe spatial frequency. It is defined as one period in the alternating pattern of black and white spaces (sinusoidal grating pattern) at the projected size of 1 degree. However, the eye’s optics in the cornea and lens act as a low pass filter with a cutoff frequency around 60 cpd. This way, the signal that cannot be properly sampled and reconstructed is effectively removed through optical prefiltering.

The pupil is an additional important factor. With its adjustable diameter of 2 to 8 mm [Gol10a], it serves as an aperture. This adjustment mostly affects sharpness of the image, as only about one magnitude of light intensity difference (1 log unit) can be controlled by the pupil. The eye’s adaptation to differences in brightness sensation (dark and light adaptation) mostly takes place on the retina.

3.2. Sensor

Light that enters through the eye is projected onto the photosensitive layer of the eye, the retina. This layer consists of two types of photoreceptors; $6 \cdot 10^6$ *cones* and approximately 20 times as many *rods* [Gol10b, p. 28]. Rods consist of only one type of light-sensitive pigment and are responsible for the brightness sensation in lower-light conditions (scotopic vision) by providing monochromatic feedback. Cones are divided into three types for different wavelengths, namely L-cones (long wavelengths), M-cones

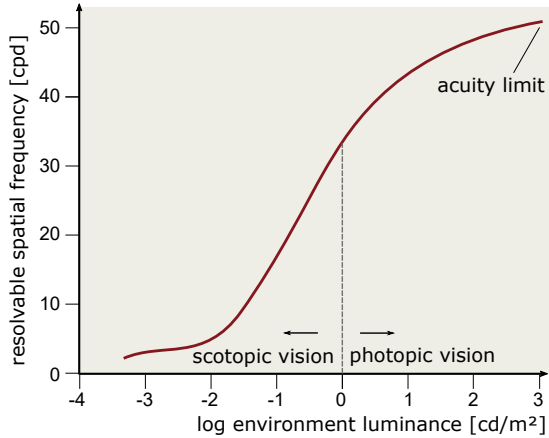


Figure 3: Adaptation-dependent acuity. Spatial acuity increases non-linearly from scotopic to photopic vision. Image adapted from Ferwerda et al. [FPSG96].

(medium wavelengths) and S-cones (short wavelengths). They are responsible for detailed color sensation (photopic vision). Photoreceptors of different types follow the distribution pattern shown in (Fig. 2). The central area of the retina, the *fovea* (approx. 5.2° around the central optical axis), consists entirely of cones. Cone density drops significantly with increasing eccentricities (the angular distance to the optical axis) [CSKH90] past the *parafovea* (approx. 5.2° to 9°) and *perifovea* (approx. 9° to 17°). These inner parts constitute *central vision*, while areas further away are referred to as *peripheral vision*. The highest density of *rods* is approximately $15 - 20^\circ$ around the *fovea*. Their density drops almost linearly. Just as the rods and cones have different densities across the retina, they have different spatial sampling distributions and follow a Poisson-Disc Distribution pattern [Wan95, ch. 3] [Yel83]. The density of cones is related to *visual acuity*, the “keenness of sight”. Visual acuity of the eye drops significantly outside the small foveal region, where humans are able to generate a sharp image (acuity is already reduced by 75% at an eccentricity of 6°). Visual acuity can be expressed as minimum angle of resolution (MAR). Normal vision corresponds to 1 MAR, a measure describing that a feature size of 0.5 minutes of arc is still visible [AKLA11, p. 627]. This minimal feature size relates to a spatial frequency of a sinusoidal grating pattern of alternating black and white spaces at 60 cpd. Models to describe the spatial acuity of the eye can be found in Sec. 4.1.1.

There are further factors influencing this keenness of sight. Visual acuity also depends on the contrast of the stimuli. The acuity limit is usually measured using a high contrast image or letter under photopic luminance conditions, which corresponds to typical daylight and display use cases. Moreover, the reduction of acuity depends on the overall lighting. The reduction of perceivable spatial detail under dimmed light is visualized in Fig. 3. The highest perceivable spatial frequency of a sinusoidal grating pattern reduces from ~ 60 cpd at photopic levels down to 2 cpd for scotopic vision, Fig. 3. In addition, contrast perception is affected [BSA91]. The eye’s sensitivity to contrast can be described by a contrast sensitivity function (CSF) for the spatial and temporal domain [Wan95, p. 33]. The CSF is defined as the reciprocal of the smallest vis-

ible contrast expressed as a function of temporal and spatial frequencies (Fig. 4 and Fig. 6). The measurements are usually performed using sinusoidal grating patterns at different contrast levels. The region under the curve is commonly called *the window of visibility* [AKLA11, pp. 613–621]. The resolvable acuity limit of (60 cpd) corresponds to the lowest contrast sensitivity value. Very high (>60 cpd) and very low frequencies (<0.1 cpd) cannot be perceived at all. While the upper limit can be explained by the cone spacing and optical filtering, the lower limit cannot be directly derived from the eye’s physiology [AKLA11, pp. 613–621]. Con-

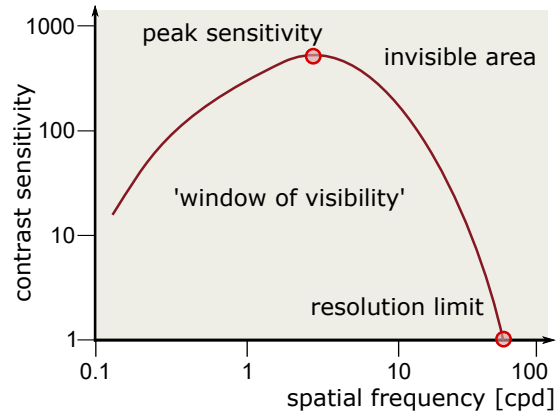


Figure 4: Spatial contrast sensitivity function (CSF). The CSF denotes the threshold contrast required for a given spatial frequency of sinusoidal pattern to be perceivable (visually detectable). All sinusoidal patterns with a contrast higher than the threshold are placed in the window of visibility under the CSF curve. Image adapted from Snowden [STT12, p. 115]

trast sensitivity depends on the number of neural cells responding to the respective grating pattern [RVN78]. From the fovea to the periphery, sensitivity decreases significantly at all frequencies. The decrease is fastest for high frequencies [RVN78]. Commonly used models for the CSF are presented in Sec. 4.1.2.

The varying distributions of rods and cones also affect the *sensitivity to colors* in different parts of the visual field [NKOE83]. The fovea is tuned to chromatic red/green stimuli, whereas those stimuli are significantly less salient in the periphery. In contrast, only about 9% of our cones are responsible for the perception of blue, but they are more spread outside the fovea. This leads to a relatively higher sensitivity to blue colors in the periphery. Hence, contrast sensitivity also depends on the chromaticity of the stimulus. Blue/yellow and achromatic stimuli result in a less-pronounced decrease in terms of contrast threshold [Mul85]. The sensitivity to red-green color variations decreases more steeply toward the periphery than the sensitivity to luminance or blue-yellow colors. Besides the different densities of the cones, neural processes are also of importance in this context [HPG09]. Information on perceptually-driven color models is given in Sec. 4.1.3.

Retinal photoreceptors have the ability to adapt to stark changes in light intensity. While adaptation to bright lighting can occur very rapidly, adapting to low-lighting conditions takes considerably longer [Ade82, Bak49]. Adaptation influences the performance of

the HVS, such as color perception, spatio-temporal contrast sensitivity and the amount of perceivable detail [LSC04]. It enables humans to perceive visual information robustly over seven orders of magnitude of brightness intensities. However, we are not able to see equally well at all intensity levels: Adaptation comes at the expense of reduced acuity at lower light levels. During daytime, contrast sensitivity is lower but visual acuity and color vision excel. Commonly used models to describe adaptation are presented in Sec. 4.1.4. Similar to the drop of acuity with eccentricity that can be observed in stereopsis, depth perception is significantly reduced in the periphery [PR98].

3.3. Motor

Our eyes are constantly in motion. Six external muscles (extraocular muscles) allow precise and fast changes of the horizontal and vertical orientation of the eye. The primary goal of moving the eyes is to move the projection of the object of interest (OOI) onto both foveae, so that the focused object is perceived with high detail. This mechanism allows exploration and scanning of the environment, shifting attention from one object to another. In addition, the extraocular muscles allow the eye's lens to adjust and to set the OOI into focus. The most important types of motion are *saccades*, the *vestibular-ocular reflex*, *smooth pursuit eye motion (SPEM)*, and coupled *vergence-accommodation* motion. An excellent survey on the properties and effects of human eye motion from a psychophysical point of view is provided by Kowler [Kow11].

During head movements occurring during activities such as walking, the HVS uses acceleration information from the vestibular system as well as information on the amount of head rotation and retinal velocity information (optic flow) to keep the orientation of the eyes aligned to the OOI. This *vestibular-ocular reflex* happens quickly with a latency of 7-15 milliseconds and is robust for fast head movements [AKLA11].

Humans constantly scan their environment. The most important mechanisms in this context are *saccades* and *fixations*. Saccades denote the motion when rapidly jumping from one OOI to another. During a saccade, peak angular speeds of up to $900^\circ/s$ [FR84] can be reached. At the same time, there is a dramatic decline in visual sensitivity, which is referred to as *saccadic suppression* [VRWM78, WDW99, RMGB01]. As a result, during saccadic eye movements, accurate visual information cannot be acquired. In contrast, fixations describe the state and duration in which visual information is perceived while our gaze rests on an OOI. Fixation durations typically vary between 100 milliseconds and 1.5 seconds [WDW99]. It is assumed that the duration of a fixation corresponds to the relative importance and visual complexity of an area in the visual field. If more information needs to be processed, fixations typically take longer. When viewing a typical natural scene, the HVS triggers around 2 to 3 saccades per second and the average fixation time is about 250 milliseconds [KFSW09]. The spacing between fixations is, on average, around 7° viewing angle. Maintaining fixations at larger eccentricities ($> 30^\circ$) is uncomfortable and usually results in a head rotation towards the target, followed by a fixation at a lower, more comfortable eccentricity. While consciously fixating on an object, the eye still performs tiny but important movements known as *tremor motion*. This unconscious motion

refreshes the retinal image. Tests have shown that the perceived image fades away if tremor motion is inhibited [AKLA11].

The unconsciously triggered tracking reflex when a moving object attracts our gaze is called smooth pursuit eye motion (SPEM). This motion enables the observer to track slow-moving targets so that the object is fixated onto the fovea. Interestingly, small eye movements up to $2.5^\circ/s$ have hardly any effect on visual acuity [AKLA11]. Researchers have found that the peak velocity for SPEM is around $100^\circ/s$ [WDW99]. However, the success rate depends on the speed of the target and decreases significantly for angular velocities $> 30^\circ/s$.

Stereopsis is highly entangled by *vergence* and *accommodation*. In order to fixate an object, both eyes are required to simultaneously rotate into opposite directions (vergence). To focus on an object, the eye adapts its lens and uses the pupil as an aperture. The mechanical ability to compress and relax the lens is called accommodation [How12]. When the ciliary muscles at the front of the eye tighten, the curvature of the lens and, correspondingly, its focusing power is increased. Accommodation describes the natural counterpart of adjusting a camera lens so that an object in the scene is set into focus. Importantly, this process happens unconsciously and without any effort in less than a second at photopic illumination levels [Gol10a, p. 289]. Typically, stereoscopic displays affect vergence by providing binocular disparity cues as a separate image for each eye. Yet, as the images are shown on the screen, the eyes focus on the screen's distance. This results in a conflict, generally known as the *vergence-accommodation conflict* [Gol10a, p. 1040]. Accommodation and vergence motions are coupled with the fixation process for binocular vision so that both eyes' gaze aims at the same point at a distance.

3.4. Processor

Retinal stimuli processing is followed by neural information processing in the *visual cortex* of the brain. Corresponding to the drop in the density of rods and cones, over 30% of the primary visual cortex are responsible for the central 5° of the visual field, while the periphery is under-represented [JH91]. Cognitive processing of images and perceptual differences between central and peripheral vision have been targeted by perception research. Thorpe et al. [TGFTB01] have shown that peripheral vision provides a rich source of information, crucial to the perception and recognition of contrast features, objects and animals. Gilchrist et al. [TGTT11] point out that the influence of color changes in the periphery is higher than that of orientation changes. Furthermore, the HVS makes extensive use of contextual information from peripheral vision, facilitating object search in natural scenes [KKK*14]. Thereby, preprocessing of visual stimuli probably occurs. There is evidence that basic visual features (such as object size, color and orientation) are pre-processed before actual attention is placed on the object by moving it into central vision [WB97]. Hence, humans may be aware of certain aspects of the scene content (shapeless bundles of basic features) in the periphery but have to pay attention on the shape to recognize its form and all its features.

Besides the process of stereopsis, the ability to interpret depth cues in the visual input to improve stereo vision and the sense of

spatial localization is highly entangled in the visual cortex. Depth cues can be static (e.g., occlusion, perspective foreshortening, texture and shading gradients, shadows and aerial perspective) or dynamic (e.g. motion parallax). Cues can also be obtained from memory, such as for relative sizes of familiar objects [Gol10b, ch. 10] [Pal99, ch. 5.5] (Sec. 3.5). Moreover, depth cues are dependent on the object's distance to the eye and dominant cues may prevail or compromise 3D scene interpretation [DRE*11]. A phenomenon that can only be observed with peripheral vision is known as *crowding*. Objects become more difficult to recognize (rather than to detect) if distracting stimuli surround them. Crowding is studied by using well-defined stimuli such as letters or sine wave patterns [Bou70, TGTT11]. The effect of crowding can also be observed for more complex content such as faces [MM05] and complex stimuli in natural images [RR07, DP08, BNR09].

Finally, vision is affected by cross-modal effects. In particular VR systems often provide non-visual cues such as audio, vibration or even smell. These effects have been studied in psychological experiments on various interplays between cues [SS01, Pai05, SS03, WP04]. When sensory channels are substituted or combined, some implications occur: sensory channels are no longer seen as separate channels but may affect each other through integration of sensory signals inside multimodal association areas in the brain [Sut02, p. 36–64] [Pai05] [LN07].

3.5. Memory and Attention

The processing of visual information is highly dependent on knowledge and patterns, stored in the memory. How such knowledge is stored, is still being discussed. According to Smith and Kosslyn [SK13], a representation is a physical state that stands for an object, event or concept, and must be constructed intentionally and carry information. Representations may encode information in different forms, including those similar to images or feature records, but also amodal symbols, and statistical patterns in neural networks. These representations are connected: Different formats work together to represent and simulate objects [SK13]. Moreover, the brain preserves certain properties of the retinal image over time, despite motion and varying occlusions [Yan95, LE13]. As noted before, there is also evidence that basic visual features such as color, size and orientation are parsed and pre-processed before directing the central gaze to them.

While attention is still not fully understood, research indicates that it has three components: orienting to sensory events, detecting signals for focused processing and maintaining a vigilant or alert state [PB71]. Attention is important for processing visual stimuli and search behavior [TG80]. It involves the selection of information for further processing and inhibiting other information from receiving further processing [SK07, p. 115]. Attention can occur in information-processing tasks in various ways [WC97]: *selective attention* is the choosing of which events or stimuli to process; *focused attention* is the effort in maintaining processing of these elements while avoiding distraction from other events or stimuli; *divided attention* is the ability to process more than one event or stimulus at a given point in time.

The focus of attention also affects perception on a cognitive

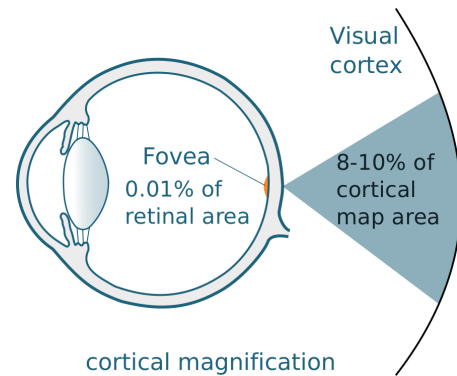


Figure 5: Cortical magnification. The cortical magnification maps the small area of the fovea to a much larger area on the visual cortex. Image adapted from Goldstein [Gol10a, p. 82]

level. A critical perceptual effect for certain tasks is the effect of *cognitive tunneling* (or *visual tunneling*) and *inattentive blindness*. Observers tend to focus attention on information from specific areas or objects. However, a strong cognitive focus on specific objects leads to an exclusion/loss of information for parts in the periphery of highly-attended regions.

Several studies conducted by Thomas et al. [TW01] were concerned with the detection of perceptual differences and showed the effects of visual tunneling during terrain rendering. Further studies showed the effects of visual tunneling as a dramatic decrease of the size of the visual field and the loss of information in the user's peripheral vision [TW06, WA09, LMS10].

4. Perceptual Models

Perceptual models are commonly used in computer graphics to approximate functions and properties of the HVS by mathematical descriptions. These models are used to steer perceptual rendering algorithms and to judge the perceptual quality of images. In the following, we present a selection of models that have been derived from the HVS (RQ2) and are relevant for computer-graphics applications. These models can be either low-level, only describing certain aspects of the HVS, or higher-level, describing an entire system.

4.1. Low-level Models

Low-level models target very specific aspects of the HVS. In the following, we will describe the most commonly used low-level models to describe spatial acuity and the temporal resolution along with brightness, contrast and color sensitivities of the HVS.

4.1.1. Spatial Sensitivity (Acuity)

One important aspect in perception-driven rendering is how the acuity of the eye can be modeled. Acuity models are often used to adapt image resolution and rendering quality based on the user's gaze (Sec. 7.2). Strasburger et al. [SRJ11] provide a historical summary and survey describing how the visual acuity drops for peripheral vision.

Green [Gre70] shows that acuity differs from what may be expected from physical cone spacing at eccentricities $> 2^\circ$. Weymouth et al. [Wey58] show that visual acuity decreases roughly linearly with eccentricity for the first 20 – 30 degrees. Visual performance decreases more rapidly for higher eccentricities [FWK63]. According to Weymouth [Wey58] and Strasburger et al. [SRJ11, ch. 3], when measured in terms of an MAR rather than acuity (its reciprocal), a linear model matches both anatomical data (e.g., receptor density) and performance results on many low-level vision tasks. Nonetheless, the slope of the MAR function is user-dependent and cannot be precisely predicted [SRJ11, p. 19]. Acuity is also affected by eye adaptation in very bright and dark areas and by eye motion and cognitive factors [Gol10a, p. 60]. Thus, any linear model clearly remains an approximation. A motivation for the often-used linear model has been provided through the concept of *cortical magnification* by Whitteridge and Daniel [DW61], and Cowey and Rolls [CR74]. The cortical magnification factor (CMF) M represents a mapping from the visual angle to a cortical diameter in millimeters (Fig. 5). The magnification factor M is largest for those areas corresponding to the fovea and decreases with eccentricity for peripheral areas.

Resulting to the linear CMF, the M -scaling hypothesis claims that performance degradation with eccentricity can be canceled out by applying spatial scaling to stimuli. For example, to compensate for the loss in acuity when trying to read letters in the periphery, those letters just have to be enlarged in accordance to the linear CMF to be equally readable again. This method has been successfully demonstrated by Cowey and Rolls [CR74] and motivated researchers to unify fovea and periphery [SRJ11, ch. 3.1]. The M -scaling concept is widely used to model the eccentricity falloff in gaze-contingent rendering systems, presented in Sec. 7.2. Strong supporters of the concept claimed that "a picture can be made equally visible at any eccentricity by scaling its size by the magnification factor" [RVN78, p. 56]. However, other researchers have pointed out difficulties of the M -scaling concept [WM78, BKM05]: First, the linear CMF model only approximates the complexity of the HVS, as peripheral vision is not a scaled-down version of foveal vision [BKM05]. Second, several studies exist in which the CMF concept is less convincing or clearly fails, such as stereo acuity, two-point separation in the far periphery or contrast sensitivity for scotopic vision [SRJ11, ch. 3.4]. In addition, due to variations in the measurements for different visual tasks and to inter-individual differences, it is still an open question whether M -scaling is also applicable to near-foveal regions [SRJ11, p. 10].

4.1.2. Brightness and Contrast Sensitivity

Sensitivity to spatial contrast of the HVS is expressed by the contrast sensitivity function (CSF) [Sch56, AL73, AET96]. It can be described by spatial parameters and the eye's adaptation (Fig. 4). However, for a complete model, the CSF also depends on eccentricity, temporal effects and retinal velocity, making it a function of an even higher order. Hence, various CSF models with various degrees of freedom exist in the literature.

Mannos and Sakrison [MS74] propose a practical model for achromatic and chromatic sensitivity:

$$csf_{mannos}(f) = 2.6 \cdot (0.0192 + 0.144f) e^{-(0.114f)^{1.1}}$$

where f is the spatial frequency of the visual stimulus in cpd. A more elaborate mathematical model used in practice was developed by Barten [Bar99]. It takes spatial contrast sensitivity for foveal and peripheral vision into account, along with an extension to the temporal domain. Another approach to describe the CSF was introduced by Gervais et al. [GHR84]. The authors fit splines to psychophysical data to derive the CSF. Daly et al. [Dal98] use precomputed CSF data for a specific illumination level and support spatial frequency and retinal velocities. Mantiuk et al. provide achromatic and chromatic CSF models fitted to visibility measurements over an extended range of adaptation luminance [MKRH11]. The latter is broadly used for image quality assessment [VDSLD14, ASG15] and tone mapping operator (TMO) [YZWH12, HČA*12, PM13, EMU15]. More details on the CSF are given by Johnson and Fairchild [JF02] and by Lukac [Luk12, p.17].

4.1.3. Color Sensitivity

The different sensitivities, distribution and densities of cone types highly affect human color perception, the sensation of visible light with wavelengths λ ranging from 390 to 750 nanometers. The most commonly used color model in computer graphics is the additive RGB model. This model additively combines the three primary colors red, green and blue (r,g,b) that relate to the rods on the retina. However, this model is not inspired by perception. The color-space depends on the underlying device. This problem has been targeted by a later standardization by the Commission Internationale de l'Éclairage (CIE). CIE-XYZ uses special reference colors for r, g and b. Assigning these references became possible by determining CIE-XYZ, a device independent and non-negative color-space.

A visible stimulus S depends on wavelength-dependent illumination $I(\lambda)$ and object reflectance $R(\lambda)$ [Luk12]. When a stimulus is observed, the (L)arge, (M)edium and (S)mall-cones respond by integrating energy over all wavelengths:

$$(L, M, S) = \int_{390}^{750} l(\lambda), m(\lambda), s(\lambda) I(\lambda) R(\lambda) d\lambda,$$

where $l(\lambda), m(\lambda)$, and $s(\lambda)$ describe the spectral sensitivities of the L-, M-, and S-cones. The CIE-XYZ was the first attempt to encompass the retinal response of the HVS with the goal to cover all perceivable colors with positive coordinates. CIE-XYZ is often used in practice, but the color space is perceptually irregular, as is RGB, HSV or many of the standard color spaces, and does not consider the different sensitivities of the L, M and S cones. In these spaces, e.g., the distance for perceptually equally-different green colors are smaller than for red or blue. Perceptual color spaces (CIE-LAB, CIE-LUV, CIEDE2000) are almost linear, and can be converted from CIE-XYZ values. Within these color spaces, perceptual differences between any two colors are directly related to the Euclidean distance. As the response of the cones cannot be measured directly, the LMS color space was designed to directly relate to the spectral responses of the LMS-cone types.

Several linear transformations from CIE-XYZ to LMS space have been proposed based on empirical measurements, such as the Hunt model, LLAB, CIECAM97s and CIECAM2000. A detailed presentation of perceptual color models is given

by Fairchild [Fai05], Gonzales et al. [GW07], and Lukac et al. [Luk12]. Hering [Her20] developed the idea of color opponency in 1892. He found that certain colors cannot be mixed, e.g., there is no reddish green. This was later empirically validated [HJ57] and proved beneficial in several image processing tasks [BBS09]. Interestingly, both CIE-LAB and CIE-LUV provide color opponency in their color channels. Color highly affects the ability of the HVS to perceive contrast. An early approach to account for the eye's sensitivity to different contrasts per wavelength and color used in computer graphics was introduced by Mitchell [Mit87]. Contrast is detected using separate, perceptually inspired thresholds for the RGB colors. Other approaches in computer graphics such as Bolin et al. [BM95, BM98], use a conversion to transform the CIE-XYZ color space to a LMS space. The LMS values are used to detect the regions that have strong perceivable contrasts. More information on these approaches can be found in Sec. 6.2.

4.1.4. Adaptation Models

This process of reducing sensitivity of the HVS as light intensity increases is known as *light adaptation*. Conversely, *dark adaptation* describes the change of vision from brightness to darkness. Adaptation allows for perceiving the environment over a high dynamic range (HDR) exceeding 24 *f-stops* [MMS15], where the illuminance reaching the sensor (retina) is doubled between two *f-stops*. This adaptation means that the HVS can perceive visual stimuli with an illuminance of more than the 2^{24} -fold of the minimum perceivable illuminance. The actual perceivable dynamic range depends, however, on the peak brightness of a scene, which is very limited on common low-dynamic range (LDR) displays. A TMO provides models to approximate the appearance of high-dynamic-range images on low-dynamic-range display devices or prints. In the following, the most influential, perceptually-motivated operators are briefly described. Detailed information can be found in the survey papers by Reinhard et al. [RWP*10], Eilertsen et al. [EUWM13], Fairchild [Fai15], and Mantiuk et al. [MMS15]. The TMO presented by Ferwerda et al. [FPSG96] globally simulates eye adaptation over time and modulates visual acuity and color perception accordingly. The model is tuned by psychophysical measurements.

Durand and Dorsey [DD00] simulate adaptation over time. Their TMO includes time-dependent chromatic, global adaptation, color shift for scotopic conditions, flare rendering for light sources, and sensitivity loss simulation for mesopic and scotopic conditions. Pattanaik et al. [PTYG00] use exponential smoothing filtering for global temporal adaptation simulation, whereas different models are used for simulating cone and rod response. Like Ferwerda et al. [FPSG96], Pattanaik et al. also utilize psychophysical measurements for calibration. Ledda et al. [LSC04] propose a simple physiological model of eye adaptation that approximates the local photoreceptor response with a temporally adjustable sigmoid curve. The time-dependent parameter is computed with respect to the characteristics of rods and cones, which results in a simulation of photopic, scotopic and mesopic vision conditions as well as receptor bleaching and regeneration. The luminance difference between succeeding frames determines the adaptation rate. Krawczyk et al. [KMS05] model temporal adaptation using an exponential decay function. In addition, local contrast and optical aberrations

are calculated by taking the pupil size into account, resulting in naturally-looking scenes. Benoit et al. [BAHLC09] present the Retina model TMO including a local, biologically-inspired retina model that enables spatiotemporal filtering with local cellular interactions and temporal stability. Mantiuk et al. [MM13] propose a real-time gaze-dependent tone mapping operator (GDTMO) by including eye tracking into the rendering pipeline. The operator simulates eye adaptation based on the fixation location. Temporal eye adaptation is controlled by the luminance of the gaze point area. Recently, a complex GDTMO has been proposed by Jacobs et al. [EJGAC*15]. This operator simulates gaze-dependent global adaptation over time as well as a variety of secondary perception effects such as bleaching, mesopic hue shift, desaturation and temporal noise under very dark illuminance conditions. The modeled *Purkinje effect* shifts the luminous efficiency function from 555 nm at photopic light levels to 507 nm at scotopic conditions, resulting in a decreasing red-green perception [WS13, p.112]. In the intermediate luminance range between 0.01 and 3 cd/m² (mesopic vision), both rods and cones are active, which is most difficult to model [WM14], but it is practically important as it overlaps with luminance ranges of typical displays. Spatial acuity drops linearly with log-luminance for mesopic and scotopic conditions. The TMO by Jacobs et al. [EJGAC*15] simulates acuity loss by Gaussian blur with adaptive kernel sizes. Another model for the brightness adaptation is presented by Vangorp et al. [VMGM15]. Their local adaptation model is based on psychophysical measurements over a high dynamic range display and predicts how the adaptation signal is integrated in the retina.

Different areas on the retina need different times to adapt to new lighting situations due to previous visual stimuli (simultaneous and successive contrast) [ARH78]. When viewed simultaneously or in quick succession, different objects having the same color seem to have different colors when viewed e.g., in front of a different colored background. Retinal photoreceptors take some time to refresh, which leads to bleaching processes [GAMS05]. Hence, the image stays locally "imprinted" in the visual system for some time, resulting in perceivable *afterimages*. A corresponding model for real-time applications is provided by Ritschel and Eisemann [RE12]. It has been refined to model color transitions when the afterimage disappears [MSR*13]. After-image-like effects can also be used to increase the perceived brightness of a light [ZC01]. Similarly, perceived brightness, as well as perceived color can be altered by flickering; an effect called *apparent brightness*. Recently, it has been used to improve perceived color saturation of images beyond the display capabilities [MCNV14].

4.1.5. Visual Masking

Another phenomenon affecting sensitivity is visual masking, which happens when the perception of one stimulus, the target, is affected by the presentation of another stimulus, called a mask. A good overview and survey on physiological findings for visual masking can be found in the work by Legge and Foley [LF80], Breitmeyer and Ogmen [BO00] as well as Enns and Lollo [EDL00]. The effects of visual masking occur spatially and temporally (backward masking). There are several methods that try to model visual masking. Spatial visual masking is typically modeled as a background over which the target pattern is superimposed. A survey on im-

age processing with information on visual masking can be found in the work from Beghdadi et al. [BLB13]. The non-linearity of visual masking as a function of contrast level is captured by a transducer function. Effectively this function models a hypothetical response of the HVS to the input contrast, which is scaled in perceptually uniform just noticeable difference (JND) units. By computing a difference between the original and distorted signals expressed by means of the transducer in the JND units, the visibility of distortion (the difference over 1 JND) as well as its magnitude can immediately be derived. A simpler approach is to directly scale the input contrast by the corresponding threshold value as derived in the CSF that is elevated proportionally to masking signal contrast. In this case, non-linearities captured by the transducer function are ignored. Visual masking is widely used for image compression [Wat93, HK97, ZDL02]. Wavelet-based approaches like JPEG 2000 employ models for self-masking and spatial masking within pixel neighborhoods. An overview of those models is given by [ZDL02].

Since visual masking happens not only spatially but also temporally, similar masking models are widely used in video compression [AKF13]. Visual masking is commonly modeled in image and video quality metrics as well. For example the Daly's Visible Differences Predictor (VDP) [Dal93] employs the simpler threshold elevation approach, while the Sarnoff Visual Discrimination Metric (VDM) [Lub95] is based on a transducer. Both approaches will be discussed in more detail in Sec. 4.2.1 and are used from methods in Sec. 6.1 and Sec. 6.2. Employing transducers is also common in computer graphics applications in the image contrast [FSPG97] as well as disparity [DRE*11] domains.

4.1.6. Temporal Resolution

Besides spatial contrast sensitivity models, temporal changes may have a strong effect on the visibility of a pattern. "The *critical flicker frequency* (CFF, also flicker fusion frequency) describes the fastest rate that a stimulus can flicker and just be perceived as a flickering rather than stable" [AKLA11, p.700]. Fig. 6 shows the estimated adaptation-dependent temporal sensitivity for different retinal illuminance values at photopic levels with an achromatic flickering stimulus. The retinal adaptation levels are measured in Trolands $T = L \frac{cd}{m^2} \cdot p_a \text{mm}^2$, taking the size of the pupil area p_a and the luminance L into account. In Fig. 6, temporal frequency along the x-axis is plotted against the modulation ratio of the flickering stimulus. The modulation ratio represents the extent that the sinusoidally modulated light deviates from its average direct current component [AKLA11, p. 705]. It can be seen that at low frequencies modulation sensitivity is approximately equal for all adaptation levels. For higher flicker frequencies, modulation sensitivity strongly depends on retinal illuminance values. The temporal CSF shown in Fig. 6 does not show the complete picture. Many other properties result in deviation from the shown CSF behavior. Kelly [Kel61] compared chromatic flickering with achromatic stimuli and explore spatial contrast sensitivity in combination with temporal contrast sensitivity resulting in a spatio-temporal CSF [Kel79]. The mathematical model can be used to describe how a feature moving across the visual field also affects the perceptibility of detail, which leads to motion blur. If a light flickers faster than the speed the HVS can resolve, we perceive the flashing light

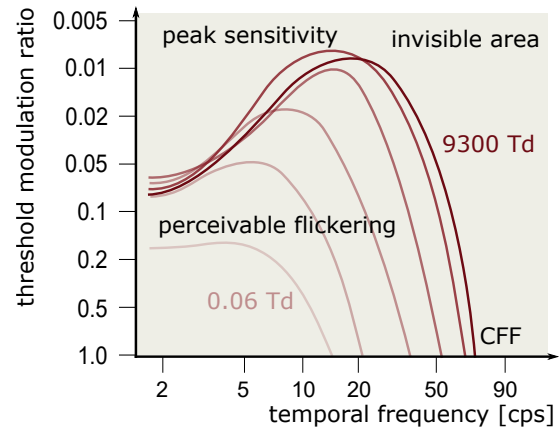


Figure 6: Temporal contrast sensitivity function (CSF) for different retinal adaptation levels. Each curve represents the threshold modulation ratio (percentage deviation of average value) of a just-detectable flicker stimulus for a given adaptation level (in Trolands) plotted against flicker frequency (in cycles per second, cps). Low levels of retinal illuminance result in a low-pass CSF, whereas higher levels reshape the CSF into a more band-pass curve. Image adapted from Adler et al. [AKLA11, p. 705]

as stable rather than seeing a sequence of flashes. The CFF is dependent on different features. For photopic lighting conditions the CFF increases linearly with the logarithm of luminance of the flickering light over a dark background (Ferry-Porter law) [Por02]. The Granit-Harper law states that the CFF increases linearly with size of the stimulus area [GvA30]. Rovamo and Raninen [RR88] have shown that for constant stimulus size and luminance, the CFF increases with eccentricities up to 55° . Towards the far periphery, the CFF decreases again. Hence, mid-peripheral vision has better temporal resolution than foveal vision and far-peripheral vision. If the CFF is plotted against the number of stimulated retinal ganglion cells, the resulting function is linear across all eccentricities [RR88]. This directly relates to the number of frames per second we need to render in order to perceive an animation rather than a sequence of individual images. A high number of frames combats the flickering and decreases the motion-induced blur. Therefore, temporal upsampling is often used to artificially increase the frame rate [DER*10].

One can ask if there is a dependency between the HVS' ability to detect motion and certain eccentricities. McKee [MN84] specifically looked at motion detection in the visual periphery. They conducted several experiments showing that the peripheral visual field has no special ability to detect motion; the threshold to detect motion and accelerations are not better in the periphery than in the fovea. However, they are not worse either. Interestingly the threshold to detect motion is much smaller than the MAR. This means the peripheral visual field is as good as the fovea when it comes to motion detection.

4.2. High-level Models

High-level models of the HVS try to model the entire human visual system to find a perceptual measure for the image quality and

to detect perceptual differences. The most common image quality metrics and the easiest to derive are *full reference metrics*, which require the availability of a reference. Often the outcome of such metrics is a measure of the overall perceptual difference and a map that provides localized information about the strength of the differences. While such metrics are sufficient for comparing different rendering algorithms, it is challenging to use these metrics to guide rendering. One option is to compare frames from the subsequent rendering stages to gain an insight regarding the rendering convergence. *No-reference quality metrics* can directly judge the quality of single images, which makes them better suited for rendering applications. However, they usually provide less reliable predictions. In the following sections, we describe representative full- and no-reference image quality metrics in greater detail.

4.2.1. Full-Reference Metrics

Daly's Visible Differences Predictor (VDP) [Dal93] introduces a model of the HVS to compare two input images and derive a *difference map*. Each input image undergoes identical processing. First, the retinal response and luminance adaptation are simulated. Then, the images are converted into the frequency domain, where CSF (Sec. 4.1.5) filtering is performed. This step scales pixel values into perceptually meaningful detection threshold units. Such perceptually-scaled input images are decomposed into spatial and orientation channels to account for per-channel visual masking (Sec. 4.1.5). Per-channel differences are transformed into the probability of perceiving the differences by means of a psychometric function and then accumulated in an aggregated difference map. VDP is particularly sensitive in detecting image differences near the visibility thresholds (Sec. 6.2). Since rendering artifacts, such as Monte-Carlo noise, typically cannot be tolerated, VDP is a useful tool for guiding such artifact suppression below the visibility level. If the goal is to measure the magnitude of clearly visible (supra-threshold) artifacts, the precision of VDP is limited.

Daly's original VDP has been extended by various researchers. Jin et al. [JFN98] and Tolhurst et al. [TRP*05] extend the model to consider the eye's color sensitivities (Sec. 4.1.3). To this end, they follow the color-opponent theory by Hering [HJ57] and use the chromatic CSFs [Mul85, MS99] to process color information separately. In practice, they first transform the input images into luminance, red-green, and blue-yellow channels, and then, apply VDP separately to each of them using corresponding CSF. To compute the final difference, the results from all channels are combined.

Mantiuk et al. [MDMS05] improve the prediction of perceivable differences in HDR images (HDR-VDP). They integrate several aspects of high contrast vision, e.g., light scattering by the eye optics, nonlinear light response for full-range luminance, and local adaptation. In particular, light scattering is important for HDR signals, as strong light sources or highlights can lead to significant glare, even for remote image regions. In a follow-up paper Mantiuk et al. [MKRH11] introduce HDR-VDP2. It improves on the original metric, among others, by using another visual model for varying luminance conditions, derived from contrast sensitivity measurements as performed by Barten [Bar99] (Sec. 4.1.2). Moreover, they have calibrated and validated the model using several image quality and contrast databases. Swaffort et al. further improve HDR-

VDP2 by providing an image metric for gaze-contingent rendering quality [SCM15, SIGK*16], which adds measurements for the peripheral vision degradation at increasing eccentricities. The contrast sensitivity is thus decreased with visual eccentricity based on a tuned model of the CMF. Narwaria et al. [NPDSLCP14] improve the accuracy of HDR-VDP2 prediction by employing a comprehensive database of HDR images along with their corresponding subjective ratings. The same group provides a quality measure for HDR videos [NDSL15]. This is based on a spatio-temporal analysis focused on fixation behavior when viewing videos. The performance of the method is verified using a HDR video database and their subjective ratings.

Despite VDP being widely used for various rendering systems (Sec. 6.2) it is computationally expensive, due to the individual processing of each band within the frequency domain [LMK01]. The VDM [Lub95] is simpler, requires less computational effort and GPU-based implementations are available [WM04]. In contrast to VDP, where image filtering is performed in the frequency domain, VDM uses convolutions and down-sampling in the spatial domain only [Čad04, p. 18]. A transducer function (Sec. 4.1.5) is applied to account for visual masking [BM98]. VDM derives two measures from two input images. The first is a single measure of the JND. The other is a map containing the locations of regions with a high predicted visual difference. In contrast to VDP, VDM is an example of a metric specifically designed to account for the magnitude of supra-threshold image differences. However, this comes at the expense of precision loss, when near threshold differences need to be judged.

A metric specifically designed for realistic image synthesis and inspired by VDP and VDM, was introduced by Ramasubramanian et al. [RPG99]. Their metric tries to predict thresholds for detecting artifacts and spends most computational effort in regions where the visibility of artifacts is the highest. The metric models the adaptation processes, contrast sensitivity and visual masking. The key idea is to precompute the most expensive metric components for direct lighting as a per pixel contrast threshold elevation map. Such a map is directly used to guide the costly computation of indirect lighting. Similarly Walter et al. [WPG02] analyze texture information to find tolerance for visual error. The tolerance can be stored as a standard mip-map, along with the texture, and efficiently used as a lookup table during rendering.

Ramanarayanan et al. [RFWB07] pointed out that, even though some visible image differences can be predicted by VDP, they simply do not matter to human observers. They try to focus on visual equivalency and determine whether two images convey the same impression regarding scene appearance. A couple of psychophysical experiments alongside with a validating study led to the *visual equivalence predictor* (VEP) metric. Later, Křivánek et al. [KFC*10] investigated visual equivalence for instant radiosity (virtual point light) algorithms and proposed a number of useful rendering heuristic, which were difficult to formalize into a ready-to-use computational metric.

Vangorp et al. [VCL*11] propose a perceptual metric for measuring the perceptual impact of image artifacts generated by approximative image-based rendering methods. Considered artifacts include blurring, ghosting, parallax distortions and popping. For

the evaluation the authors generated viewpoint-interpolated image datasets containing different levels of distortions and respective artifact combinations.

The metrics discussed so far are deeply rooted in advanced models of early vision in the HVS and are capable of capturing just visible (near threshold) differences or even measuring the magnitude of such (supra-threshold) differences and scaling them in JND units. The Structural SIMilarity (SSIM) index [WBS04] is one of the most popular and influential quality metrics in recent years. It emphasizes less on the precise perceptual scaling but is still sensitive to the differences in the image brightness, contrast and structure. In particular, the structure modeling component plays an important role in achieving a high accuracy [CHM*12]. Since the HVS is strongly specialized in learning about the scenes through extracting structural information, it can be expected that the perceived image quality can be well approximated by measuring structural similarity between images.

Several other perception-based image and video quality metrics have been developed, e.g., the Moving Picture Quality Metric (MPQM) [VdBLV96] and the Multi Scale Structural Similarity Index Metric (MS-SSIM) [WSB03] (based on SSIM) using structural similarities of the image at various scales. Other modern video quality metrics, such as the Visual Information Fidelity (VIF) index [SB05], rely on natural-scene statistics and employ an information-theoretic approach to measure the amount of information that is shared between pairs of frames. A survey on video quality metrics can be found in the work by Wang [Wan06].

The above metrics work assume that both reference and test images are perfectly aligned. However, human perception compensates for geometric transformations. For example, we can easily tell that an image is identical to its rotated copy. Kellnhofer et al. [KRMS15] present a metric that quantifies the effect of transformations not only on the perception of image differences but also on saliency and motion parallax.

Cadik et al. [CHM*12] compare a large number of state-of-the-art image quality metrics, including the discussed SSIM, MS-SSIM, HDR-VDP2, and evaluate their suitability for detecting rendering artifacts. The conducted user experiments show that the most problematic features for existing metrics are excessive sensitivity to brightness and contrast changes, calibration for near visibility-threshold distortions, lack of discrimination between plausible/implausible illumination and poor spatial localization of distortions for multi-scale metrics. Based on these observations, the authors have developed a test data set to allow for the development of further improved metrics.

The current trend is to employ deep machine learning methods to derive full reference metrics [ZK15, ZWF16, APY16]. So far, the existing metrics are generally successful in predicting the mean opinion score (MOS) value, i.e., a single number that characterizes the overall image quality, without producing detailed error maps. While in terms of the computation performance such metrics can be a viable option for rendering applications, it remains to be seen how they will perform in such contexts.

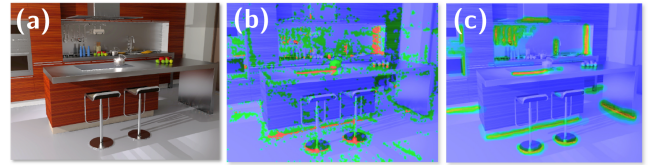


Figure 7: Example for a no-reference metric. No-reference metrics derive a measure (b) of perceived image quality based on a single image (a). Results can be close to ground truth (c) often determined in psychophysical experiments. Image from Herzog et al. [HČA*12]

4.2.2. No-Reference Metrics

All of the above models are comparative approaches that assume the reference image is given as an input. However, in a vast majority of computer graphics applications, the goal is to synthesize a new image. In such a situation, the reference image is missing, and it is desirable to have a method that can blindly estimate the quality of the image or an animation sequence (Fig. 7).

Chandler and Hemami [CH07] quantify the visual fidelity of natural images based on near-threshold and supra-threshold properties of human vision. Their *Visual Signal-to Noise Ratio* (VNSR) first uses contrast thresholds for detection of distortion and wavelet-based models of visual masking and visual summation to determine whether the distortions are visible. If the distortion is above the threshold, a second stage uses low-level vision models and accommodates different viewing conditions and contrasts to compute the final VSNR value.

Stokes et al. [SFWG04] try to predict the perceptual importance of the indirect illumination components with respect to image quality by conducting a series of psychophysical experiments. Their idea is based on the observation that the different direct and indirect illumination components are likely to be not equally important with respect to their contributions to the visual quality. Their metric is solely based on simple measures of scene reflectances that are gathered during computation of the direct illumination component. Hence, a lightweight progressive update during the integration of the indirect illumination component is possible. The proposed metric cannot detect local artifacts, which would sometimes be desirable for local image enhancement.

Such local error maps, as shown in Fig. 7b, are generated by the no-reference metric NoRM as proposed by Herzog et al. [HČA*12]. They use a machine learning system, trained with various types of rendering artifacts that are locally marked by the subjects in a perceptual experiment. At both the training and error prediction stage, they actively use feature descriptors based on 3D scene information (Fig. 8-top row) to compensate for the lack of a reference image. They also employ low-level models of the HVS to predict the perceived strength of rendering artifacts in the error map (Fig. 7b).

Since it is difficult to build a general-purpose no-reference quality metric, a number of attempts have been made to focus on specific artifacts such as ghosting [BLL*10] or camera-shake blur [LWC*13] that cause specific and relatively easy to isolate

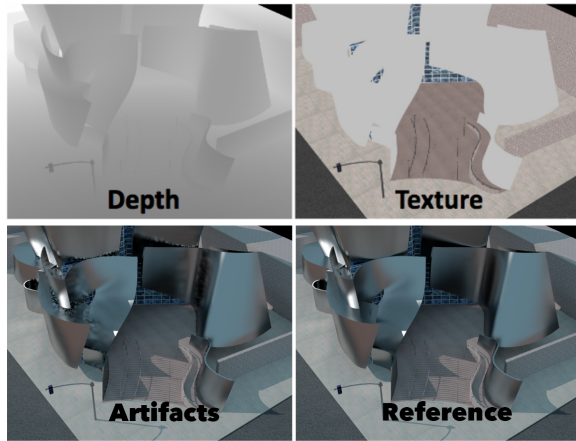


Figure 8: NoRM – training [HČA*12]. An example scene from the data set used to train a Support Vector Machine (SVM) to derive the no-reference metric NoRM. Each case consists of a reference color image and a test image with different rendering artifacts. Moreover, explicit 3D scene information that is readily available in rendering such as the depth and diffuse material/texture buffers are employed. Image from Herzog et al. [HČA*12]

changes in the image signal. Similar to NoRM, Support Vector Machine (SVM) and other classic machine learning techniques have been employed to derive a number of no-reference metrics that typically rely on natural image statistics and are focused on predicting various incarnations of noise and compression artifacts such as ringing, blur or blocking [MB10] (refer also to the work by Liu et al. [LWC*13] for a more complete survey). Departs from the natural image statistics can be analyzed to judge the image “naturalness” [MMB12]. Similar to the full-reference metrics (Sec. 4.2.1), deep machine learning might provide a viable tool for robust, no-reference artifact detection in the years to come [BMW16, KYLD14, BCNS16].

5. Attentional Models

Stimulus salience, or saliency, refers to the visual “attractiveness” or importance of parts of the environment. Attentional models are used to determine this saliency, i.e., what draws attention and what fixates a viewer’s gaze in a scene. We previously defined attention in Sec. 3.5 as a visual selection process. In this section, we present an overview of the most common and successful models to describe attention. Based on psychophysical literature, saliency models can be subdivided in *bottom-up* and *top-down* models: saliency models are either driven by basic visual stimuli of the HVS such as contrast, edges or boundaries (*bottom-up*) or by the task and intention of the subject understanding the scene (*top-down*). Attention can be guided by information stored in so-called pre-attentive object files - in particular new objects may attract attention [WB97]. Stimulus saliency modulates pre-attentive processing speed in the human visual cortex [TZGM11]. Furthermore, pre-attentive object files that are created before actual attention is placed upon an object may direct eye scan movements in conscious attentive processing of information [PRC84, PRH90] and have often be used to predict fixation locations [JDT12, VDC14, BJD*15]. Methods for visual

attention modeling and gaze prediction model common gaze properties and try to estimate from an image or video where fixations will be placed in a scene.

A strong interrelationship exists between pre-attentive object files and saliency: pre-attentive segmentation (the process of creating “figural units”) is based either on perceptual grouping (object shapes are integrated with surface details) or saliency [Edw09, pp. 57–59]. Saliency is generally recorded in saliency maps, in which the probability that a certain image region is actually observed is encoded. This is often visualized as a greyscale image or heat map. Inspired by feature integration theory [Tre88], the saliency map can be thought of as a summary of the conspicuities of all visual stimuli. Visual saliency may directly be learned from large amounts of eye tracking data [ZK13]. Visual perception research has discovered gaze patterns that are common for healthy adult humans, although differences exist between cultural environments [CBN05] and gender [VCD09, SI10]. Humans are similarly attracted by faces and objects that are located in the line of sight of such faces [Gol10a, p.823]. Analyzing scan paths using active eye tracking has revealed many more similarities [CW11, p. 160]. Recent survey papers on visual attention modeling have been provided by Scholl [Sch01] and Borji and Itti. [BI13].

5.1. Bottom-up Saliency Models

Bottom-up models are motivated by the feature integration theory for early vision [Tre88]. In this theory, salience of a stimulus is only affected by low-level features such as color, orientation, brightness and contrast of the stimulus. Individual stimulus features are added linearly resulting in a normalized saliency map (Fig. 9b). However, the bottom-up perspective is not sufficient for prediction of the actual sequence of fixations known as the *scan path* and fixation duration since selective attention is not just based on low-level features [OTCH03, HBCM07]. According to bottom-up theory, the detection of objects across the visual field is assumed to be sub-conscious and does not depend on attention (pre-attentive processing) [WDW99].

The predominant, biologically inspired bottom-up model was provided by Itti and Koch [IKN98]. The model measures local center-surround contrast on different scales, simulating the receptive fields of ganglion cells in the retina and neurons in the visual cortex. This model can be quickly evaluated, allowing for real-time computation of saliency. Based on this model, various models that target different aspects of human visual system have been developed. They investigate the following aspects: depth, motion, proximity and habituation components [LDC06], high edge density regions [MRW96], low frequency edges [MVOP11], luminance contrast [PN03], low-level saliency features and central vision bias [HKP07], binocular disparity [JOK09], local contrast, orientation and eccentricity [OPPU09] and blurriness [MDWE02].

5.2. Top-down Saliency Models

Top-down methods model scene understanding based on the observation that humans are biased towards object features while performing particular tasks [NI07]. Hence, top-down attention models commonly introduce a visual feature bias with respect to

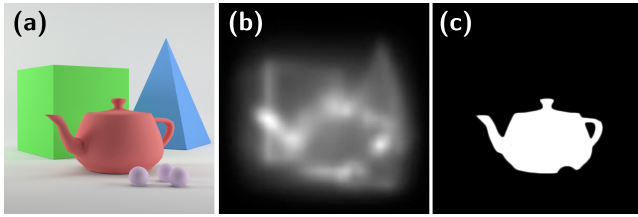


Figure 9: Bottom-up and top-down saliency. Given input image (a) the method by Harel et al. [HKP07] based on bottom-up saliency can predict fixations in a free-viewing task (b). Top-down prediction in a visual search task for the teapot can result in (c).

known objects in the scene [MTT04]. The saliency methods briefly summarized in this section derive scene knowledge from figure-background segmentation [FWMG15], face detection [VJ04], person detection [FMR08], object detection [CHEK08] or manually defined task-specific location bias [CCW03] (Fig. 9c).

Top-down gaze prediction is usually combined with bottom-up approaches in order to derive the overall saliency of a pixel, resulting in higher prediction accuracy for task-based scenarios. Various models have been developed, addressing the following aspects: task related feature values [IK01], color opponent images and task information [GVC03], importance map for task-relevant objects combined with a bottom-up saliency computation step [SDL*05], edge intensities of detected objects [WK06], visual distractors [NI07], gaze behavior for natural scenes including face detection [CHEK08], task and scene-dependent cues [JEDT09] and the usage of probabilistic methods [HHQ*13].

Recently, deep convolutional networks trained on large image data sets have shown impressive improvements in fixation prediction [VDC14, KTB14, KAB15]. In comparison to dedicated feature detectors, trained networks can model the influence of high-level features (faces, text) and abstract features such as popout better. Kümmerer et al. [KTB14] reuse existing neural networks to decrease the computational effort in creating a network for saliency prediction. Kruthiventi et al. [KAB15] introduced location-biased convolutional filters, which enables the deep network to learn location-dependent patterns of fixations such as the center bias observed by Judd et al. [KAB15].

Apart from those learning-based approaches, findings in cognitive science remain important to improve modern high-level saliency predictors. Koulieris et al. [KDCM14a] make use of the *scene schema hypothesis* [HH99] and the *singleton hypothesis* [TG02] to improve saliency prediction. The scene schema hypothesis states that a scene is comprised of objects we expect to find in a specific context. Salient objects that are not expected in a scene, e.g., a lawn mower in the kitchen, have a high saliency. The singleton hypothesis is based on the observation that the HVS is more sensitive to features that are singular across the field of view and suppresses prevalent features [Wol94]. Hence the singleton hypothesis states that the viewer's attention is drawn by stimuli that are locally unique and globally rare. In this context, Frintrop et al. [FWMG15] showed that, in combination with a top-down component, the bottom-up model by Itti and Koch [IKN98] is still competitive with other computationally more complex methods.

5.3. Attention Model Quality

Attention models for passive gaze prediction do not provide exact solutions. In terms of accuracy, fixations from saliency maps are not comparable to active gaze tracking. However, knowing the approximate gaze location may be sufficient for some applications. The prediction accuracy for bottom-up saliency is typically evaluated by a free-viewing task in which participants look at photographs and watch videos for the very first time [JDT12]. However, there is some controversy about the role of bottom-up versus top-down mechanisms in the context of gaze prediction [JDT12, VDC14, BJD*15]. Free-viewing experiments assume controlled conditions to be comparable, which is difficult to achieve since participants may be biased by cognitive load when performing the tests.

The influence of the task on bottom-up saliency prediction has been the focus of a variety of studies, looking into the following aspects: the irrelevance of low-level features in relation to a certain task that may be affect in inattentive blindness [CCW03, MR98] (Sec. 3.5), fixation and saccade differences between task-based tests and free-viewing experiments, comparing bottom-up saliency [IKN98] vs. top-down saliency showing that saliency is a much better predictor if the user's task is known [NI02, SC06], the overriding of low-level features in top-down mechanisms in task-based scenarios, increasing the relevance of bottom-up saliency maps [ERK08] and the existence of stimuli that capture the attention irrelevance of tasks [KSR*03]. In several other studies, the overall results have also been confirmed for natural scenes [GVC03, SU07, STNE15]. Judd et al. [JDT12] indicated that there is no single method equally suitable for all types of scenes and situations, with accuracy significantly increasing with an additional face detection step. They also find that the ground truth gaze data from two observers already gives more accurate results than the best tested bottom-up gaze predictors. Later benchmarks using different metrics show that, for free-viewing tasks, saliency prediction based on convolutional networks learned from gaze-labeled natural images often outperforms traditional "hand-crafted" saliency predictors [VDC14, KWB14, BJD*15].

Saliency prediction seldom results in a single, distinct salient region. To estimate the sequence of fixation locations of an observer is therefore a much more difficult problem and has largely been ignored in saliency research [NSEH10]. Scene-viewing models have primarily been designed to predict *potential* fixation locations. Henderson et al. [HBCM07] confirm that scan paths generated by bottom-up saliency maps do not correlate well with ground truth. Apparently, when performing a saccade, humans are biased towards making horizontal or vertical saccades. Le Meur et al. [LMC16, OLME16] exploit this bias in combination with bottom-up feature detection, resulting in a saccadic model for free-viewing scenarios. The authors make use of their model in a method for spatial fixation prediction and scanpath generation. While the previous method is optimized for simulating spatial eye motion, the saccadic model by Trukenbrod and Engbert predicts fixation durations by combining a task timer, randomly timed saccades and inhibition effects, varying the fixation time with respect to foveal processing effort [TE14]. Nuthmann et al. [NSEH10, NH12] significantly improved computational modeling of fixation durations

for scene viewing, by using an algorithm to automatically compute scan paths, including gaze locations, fixation durations, and inhibition of return. Furthermore, Further research present approaches for extracting the scan path from a given video using machine learning on gaze data in combination with a perceptually inspired color space [BDK*06, DMGB10, DVB12].

In summary, most successful saliency models balance the complex interaction of low-level and high-level processes in visual perception. Deep learning using convolutional neural networks have initiated a new direction for saliency estimation to obtain high model complexity. Yet, robust, accurate, fast and temporally stable scan path prediction remains a topic of ongoing research [VDMB12, VDC14, HLSR13, NE15].

6. Model-based Perceptual Approaches

In this section, we discuss work that directly makes use of the perception (Sec. 4) and attention models (Sec. 5) without considering any active measurements such as eye tracking. We present how these models can be used to accelerate rendering (RQ3), either by simplifying the models to be rendered or by adapting the sampling based on the characteristics of the HVS.

6.1. Scene Simplification

Reducing the complexity of a scene accelerates rendering. This is done by culling invisible objects, adapting object details or by directly employing multiple representations, at a different level-of-detail (LoD), e.g. by reducing the number of polygons. However, perceptual appearance needs to be preserved to avoid visual artifacts such as popping or flickering geometry. Even though beyond the scope of this report, semi-automatic LoD techniques allow for manual guidance of mesh simplification by identifying and weighting mesh regions considered critical to perception [PS03, KG03, HLC*06, GGC*09]. Historically, most automatic methods either rely on perceptual (Sec. 4) or on attentional and saliency-based models (Sec. 5). However, in recent years, methods for geometric simplification have begun to blur the lines by combining knowledge from perceptual and attentional models in single systems.

Early approaches use rendered images and compare them to simplification candidates using perceptual models to guide LoD selection and generation [Red97, LT00, LH01]. These approaches are coupled to visual acuity (Sec. 4.1.1) and CSF models, e.g., the one described by Mannos and Sakrison (Sec. 4.1.2). An overview of such LoD methods can be found in the book by Luebke et al. [Lue03, pp. 264–278]. Similar to the above methods, Scoggins et al. [SMM00] present an approach that transforms images to the frequency domain to develop a relationship between sampling rate, viewing distance, object projection, and the expected image error caused by LoD approximations. LoD selection can be matched with perceptual limits using the CSF model by Mannos and Sakrison (Sec. 4.1.2). All of these systems try to measure the perceived quality of the output based on image contrast and the spatial frequency of the resulting LoD changes. However, they do not look specifically at textures and effects caused by dynamic lighting.

Williams et al. [WLC*03] extend these approaches to estimate the degradation of textures and lighting changes. Their technique creates view-dependent LoD representations, sensitive to silhouettes, underlying texture content and illumination, and simplifies regions of imperceptibly low contrast first. Drettakis et al. [DBD*07] and Qu and Meyer [QM08] (Fig. 10) further improve on Williams et al. by incorporating visual masking (Sec. 4.1.5). Qu and Meyer [QM08] speed up this computationally demanding process by using a pre-processing step that computes an importance map which indicates the visual masking potential of a surface. Here, they use a model derived from JPEG 2000 (Sec. 4.1.5) and the Sarnoff VDM (Sec. 4.2). However, image space metrics are computationally involved and resolution-dependent. Therefore, Menzel and Guthe [MG10] contribute by presenting a method to move the error computation from image-space to vertex space. The authors perform LoD evaluation by measuring changes in contrast, curvature, and lighting, as well as the effect of visual masking.

Several high-level perceptual metrics exist to compare the visual quality of an image to ground truth data (Sec. 4.2). A couple of those metrics have been used to generate LoDs by rendering and comparing degenerated models. An extensive overview of investigated metrics applied to mesh compression and mesh watermarking is given in the reports from Corsini et al. [CGEB07, CLL*13]. Choosing the strategy and metric in terms of perception is still under research. Cleju and Saupé [CS06] evaluate metrics predicting quality differences of meshes at different LoDs. They find evidence that common image-based metrics perform better than geometry-based metrics. Schwarz and Stamminger [SS09] propose rendering the image twice at successive LoDs and then determining whether the LoD switch is visible due to undesired popping artifacts. Guo et al. [GVBL15] study the visibility of LoD distortions by asking subjects to mark perceivable distortions on a mesh. The derived ground truth is used to evaluate different error metrics. In the study, perception-based metrics outperform purely geometry-based approaches. Recently, Lavoué et al. [LLV16] concluded that purely image-based metrics including HDR-VDP2 (Sec. 4.2) perform sub-optimally. Based on these observations, Nader et al. [NWHWD16a, NWHWD16b] perform an experimental study of the HVS' low-level properties and derive a contrast sensitivity and contrast masking function. These can be used to identify a super threshold when vertex changes on a 3D mesh become visible, so that a JND profile can be computed on the mesh to guide a simplification algorithm.

Besides such measures based on low- and high-level vision models, geometry can be simplified while trying to preserve salient features of the mesh using attention mechanisms. Those parts that are likely to draw attention should be degraded more slowly (Sec. 5). An early approach for automatic LoD generation and selection based on attentional models is proposed by Horvitz and Lengyel [Eri97]. The authors evaluate the trade-off between the high-resolution mesh's visual quality and computational savings using a cost-function-based on mesh degradation and a probability distribution over the attentional focus of the viewer. Lee et al. [LVJ05] make use of a top-down attention model (Sec. 5.2) to preserve salient mesh features defining *mesh saliency*. A strong change in curvature is considered to result in high local saliency. For mesh simplification mesh reduction is steered by evaluating

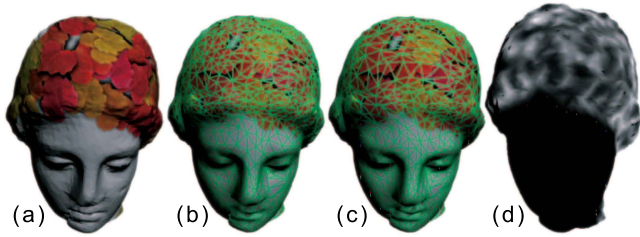


Figure 10: Perception-based mesh simplification by Qu and Meyer [QM08]. For a textured model (a) visual masking is evaluated (d). Compared to traditional simplification (b) including visual masking enables stronger simplification without affecting perceived mesh quality (c). Image from Qu and Meyer [QM08]

such a geometric saliency. For partial shape matching of meshes, Gal et al. [GCO06] define the mesh saliency as a function of geometric features and determined by clustering regions of high curvature relative to their surroundings, and a high variance of curvature values. Lavoué [Lav07] presents an extended curvature-based measure for model roughness and shows how mesh saliency can be applied to compute the visual masking potential of the geometry. Later, a study by Kim et al. [KVJG10] confirmed that *mesh saliency* better describes human fixations than random models of eye fixation, validating the importance of the local curvature measure. The approach by Wu et al. [WSZL13] extends the ideas by looking at two more aspects: Local contrast and *global rarity* based on the singleton hypothesis (Sec. 5.2). The authors introduce a multi-scale shape descriptor to estimate saliency locally, and in a rotationally invariant way. Yang et al. [YLW*16] combine *mesh saliency* with texture contrast resulting in saliency texture, which is used to simplify textured models (Fig. 11).

Ramanarayanan et al. [RFWB07] do not reduce the complexity of the model by geometric means; they reduce the complexity of the materials. Based on their VEP metric (Sec. 4.2), the authors show that the complexity of certain maps and materials can be greatly reduced when rendering Lightcuts [WFA*05] and Precomputed Radiance Transfer [NRH04], without sacrificing the visual appearance. The system by Koulieris et al. [KDCM14b] provides an LoD approach for materials, building upon the ideas of their high-level saliency predictor [KDCM14a] and extended by using object-intrinsic and contextual information (Sec. 5.2). The first type of information accounts for the fact that an object pops out if it is rotated in a way that violates its expected posture. The latter provides a measure of its contextual isolation, i.e., is a specific object showing in parts of the scene where you would not typically expect it in accordance to the *scene schema hypothesis* (Sec. 5.2). This allows for continuous adaptation of material quality.

Only few systems make use of cross-modal effects. Such a system was presented by Grelaud et al. [GBW*09]. They use both audio and graphics to guide LoD selection, and jointly adapt auditory and visual quality. Besides the limited knowledge in systems that rely on similar effects, the boundaries of perceptual and attentional models are further blurred by coupling low-level knowledge of the HVS to attentional models. Moreover, improvements on LoD systems for high-level scene properties, e.g., materials and

the incorporated lighting, will further improve on reduction rates and quality.

6.2. Adaptive Sampling

In this section, we explore the field of adaptive sampling in rendering. Modifications of the originally uniform sampling process may happen at (sub-)pixel level as well as at image level. While the former methods are mainly concerned with reducing aliasing, the latter may also take higher-level perceptual properties into account, making it possible to adaptively sample an image plane based on knowledge about low- and high-level models of the HVS and saliency estimates.

6.2.1. (Sub-)Pixel Level

As rendering is a sampling process, aliasing is an inherent issue. It is beneficial to specifically design anti-aliasing (AA) methods by taking perceptual aspects into account. A general overview of AA can be found in the work from Maule et al. [MCTB12]. Supersampling the image plane is computationally demanding. Moreover, regular sampling patterns can lead to artifacts such as Moiré patterns or temporal flickering; artifacts to which the eye is highly sensitive. The spectral properties of the spatial distribution of the photosensitive cells on the retina (Sec. 3.2) suggest turning regular patterns in less perceivable high-frequency noise. It is often better to sample using a random, pseudo-random or non-uniform pattern [DW85]. However, an efficient random (re-)sampling of individual pixels is not possible with rasterization, which is always tied to a fixed resolution. Therefore, several strategies using pseudo-random patterns [AM03, HAML05, JGY*11] and temporally jittered pixel locations [HA90] have been developed.

Methods that can adaptively resample an image in areas that matter most, such as ray tracing, are well-suited for perception-driven rendering. An early approach involving perceptual aspects was introduced by Mitchell et al. [Mit87]. After initially sampling the image plane with one sample per pixel, a simple error metric using contrast thresholds for the RGB color values is used to guide a resampling process (Sec. 4.1.3). More samples are generated for regions that expose a large error. Painter and Sloan [PS89] presented hierarchical adaptive stochastic sampling for ray tracing that worked in a progressive manner. They stop if drawing further samples is below a supra-threshold for the user. Other work in the field focuses on the change in the acuity of color perception with increasing eccentricities [ML92]. Bolin and Meyer [BM95] use a ray tracer that transforms images into the frequency domain. It is coupled to an adaptive quadtree in image space and a simple vision model, that controls where new rays are cast by accounting for the perception of colors and frequency of the content within each image block (Sec. 4.1.3). Bolin and Meyer [BM98] extend their work by developing a more elaborate adaptive sampling algorithm, based on a simplified model of Sarnoff's VDM (Sec. 4.2). Jin et al. [JIC*09] and Shevtsov et al. [SLR10] propose adaptive supersampling schemes for efficient ray tracing on many-core architectures such as GPUs and an SIMD friendly version for CPUs. After an initial sampling step, a discontinuity detection performs a pair-wise computation of gradients based on luminance or per color channel, similar to Mitchell et al. [Mit87].

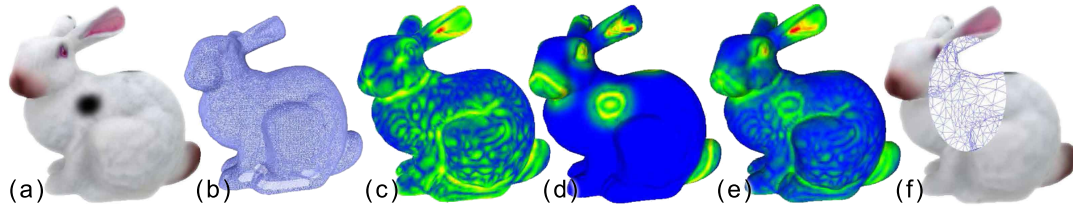


Figure 11: Mesh saliency method by Yang et al. [YLW*16]. The approach takes a textured mesh (a),(b) and measures local geometric entropy (c), color and intensity (d). The features are combined into a final saliency map (e) used to produce a simplified textured model (f). Image from Yang et al. [YLW*16]

Efficient (re-)sampling of individual pixels is hardly possible using rasterization and less interesting due to the raw processing power of modern rasterization pipelines. Hence, there are no methods that allow for selectively (re-)sampling individual pixels. However, it could be interesting to couple perception-based insights to post-processing approaches, where the final aliased image is filtered by finding and cleverly blurring jagged edges in image-space [JGY*11, JESG12].

6.2.2. Image Level

While AA usually works at the pixel/subpixel level, adaptive sampling may happen on a higher level by considering more than just the individual pixels when guiding the sampling process (selective rendering). Methods from the field of selective rendering take perceptual implications of the generated image into account, allowing more computational effort to be put into important regions of an image. Perception-critical regions are determined by looking for salient features. Selective rendering is mostly used as a flexible rendering method in ray tracers to steer the number of samples per pixel or recursion depth. A common goal is to obtain an image that is perceptually indistinguishable from a fully converged, but expensive, rendering solution.

Myszkowski et al. [Mys98, HMYS01] use the VDP (Sec. 4.2.1) as an image metric to selectively stop rendering in a Monte-Carlo path tracer for global illumination rendering. In another approach, Farrugia et al. [FP04] make use of a perceptually inspired metric based on eye adaptation for a progressive rendering method that allows global illumination computation to be stopped early. Yu et al. [YCK*09] analyze the influence of visibility approximations on the perception of global illumination renderings. They conduct a study on the perceived realism of scenes rendered with imperfect visibility, (directional) ambient occlusion and another study where renderings using visibility approximations are compared to reference renderings. The authors conclude that using appropriate visibility approximations leads to results that are perceived as realistic despite perceptible differences between approximate and reference renderings. Jarabo et al. [JES*12] take a closer look at the importance of accurate lighting and its effect on perceived realism when rendering crowds. They employ an approximation based on spherical harmonics, which is used to compute a temporal interpolation of the full radiance transfer matrix. The essential factors influencing scene fidelity found by the authors are geometric complexity, the presence or absence of color, the movement of individual crowd entities and the movement of the crowd as whole, in accordance with the effects of *crowding* (Sec. 3.4). Dachsbacher [Dac11] shows how

the analysis of visibility configurations can be used for adapting the sampling process in ray tracing, improving perceptually motivated level-of-detail approaches in real-time rendering and extending visibility classifications in radiosity methods. To determine the kernel size of the density estimation kernel (bandwidth) in a photon tracing framework, Walter [Wal98, p. 87ff] makes use of a perceptual error taking into account luminance and chrominance to compute a JND value. Guo [Guo98] developed a progressive refinement algorithm for Monte-Carlo rendering that stops refining image blocks based on a simple CSF model (Sec. 4.1.2). In the work from Ferwerda et al. [FPSG96] the authors take a closer look at the eye's adaptation process and the influence of realistic image synthesis. By performing a psycho-physical experiment, they developed a model to display and combine the results of global-illumination simulations at different illumination levels. The resulting images better capture the visual characteristics of scenes viewed over a wide range of illumination intensities (Sec. 4.1.4).

Models of visual attention (Sec. 5) can also be used to improve the quality of global-illumination renderings for animations and dynamic scenes [Mys02]. As rendering quality can be decreased for moving objects or patterns, they use temporal reprojection alongside a temporal extension of the VDP (Sec. 4.2.1) called Animation Quality Metric Algorithm (AQM), which accounts for motion when computing new samples.

The perceptual importance of the final image is often approximated by saliency extracted from previews rendered at lower quality, where the initial image estimate requires at least one sample per pixel. For decreased computation times, Longhurst et al. [LDC06] present a method that computes such a preview frame by rasterization. This frame is then used to extract saliency including different low-level features such as edges, contrasts, motion, depth, color contrast and scene habituation. The generated saliency map is used to steer the number of samples distributed on each pixel of the image. However in this approach, a high (re-)sampling weight needed for advanced global-illumination effects, such as caustics, cannot be accounted for. In contrast, Cater et al. [CCW03] and Sundstedt et al. [SDL*05] do not focus on low-level features such as edges and contrast but selectively render task-relevant salient objects and features in high-quality and reduce rendering quality for the remaining parts by adapting the resolution or the number of rays per pixel in a global-illumination renderer. In the studies performed by Cater et al. and Sundsted et al. the subjects were not able to distinguish high-fidelity rendering from selective rendering result. The experiments demonstrate the suitability of perceptual rendering if selective attention can be predicted.

One aspect of such a saliency computation using attentional models (Sec. 5) is that movement in the background of a scene may substantially influence how humans perceive foreground objects, e.g., if objects start moving in the midst of a sequence. Yee et al. [YPG01] use a model of visual attention for moving objects to speed up rendering of animations. In doing this, they introduce a method to compute a spatiotemporal error tolerance map, based on a velocity-dependent CSF. This CSF is augmented by a top-down model of visual attention (Sec. 5.2) to account for the tracking behavior of the eye, guiding the sampling of a global illumination renderer. Another system that makes use of attentional models has been developed by Chalmers et al. [CDdS06]. The authors investigate several ideas such as importance-based sampling for on-screen-distractors, e.g., sound-emitting objects, or sorting of effects to compute the most visually-important paths first and postponing less important reflections or global illumination. Hasic et al. [HCS10] show the importance of visual tasks and motion for selective rendering, as both attract the viewer's attention. They present various types of movements with varied acceleration in a psychophysical experiment to a group of subjects.

Still, ray-tracing systems are often solely used to get the highest image quality for production rendering, even though interactive rates depending on the scene complexity [ALK12, PKC15] are achievable. However, ray-based approaches do not yet reach the performance and convenience of rasterization when it comes to real-time rendering. A hybrid approach for HMDs is introduced by Pohl et al. [PBNG15]. They adapt rendering in HMDs by exploiting their lens astigmatism, which leads to a decrease in image quality towards outer regions. The proposed system reduces the sampling rate for areas outside the lens center by deploying ray tracing on the CPU combined with rasterization.

However, all of these approaches treat the HVS as a single-eyed system, though healthy humans are capable of stereopsis due to binocular vision. Rendering images for both eyes independently doubles the computational effort. The ray-tracing approach by Lo et al. [LCDC10] exploits perceptual limits that arise from the brain being able to fuse information from both eyes separately. They show that the resolution of one of the images of a stereo pair could be reduced by a factor of 6 without being noticed by the viewer. The authors also observed that shadow cues and disparity cues perform equally well when judging depth.

Besides its wide use, a disadvantage of rasterization over ray-based approaches is that, for efficiency, rasterization and the corresponding shading pipelines are traditionally tied to a fix resolution. In the last years several approaches have been introduced that allow for *multi-rate* and *multi-resolution shading*: an enabling technology for perception-driven selective rendering systems using rasterization. Clarberg et al. [CTH*14] propose a modification to current rendering pipelines, which enables varying shading rates on a per-patch basis to reuse shading within tessellated primitives. He et al. [HGF14] present a system that uses a coarse grid to reuse shading samples within grid cells. Vaidyanathan et al. [VST*14] introduce coarse pixels and tiles, which allow shading samples to be reused in a multi-level grid like fashion. NVIDIA recently proposed a multi-resolution shading approach drawing different resolutions within a single pass on their latest GPU hardware [Ree15].

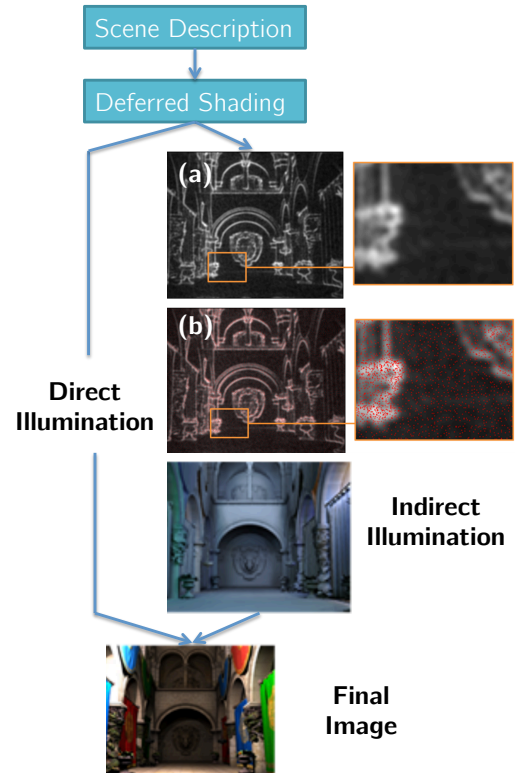


Figure 12: Sampling adaptation method by Galea et al. [GDS14]. A deferred rendering system decouples direct and indirect illumination components. To speed up the computation of the indirect illumination, a saliency map (a) is computed and sparsely evaluated (b). Inpainting computes a dense representation of indirect lighting that is combined with the direct lighting for the final image. Image from Galea et al. [GDS14]

Even though these techniques are widely used for gaze-contingent rendering (Sec. 7.2), they are rarely used for solely model-based approaches. An example of such a system has been proposed by Galea et al. [GDS14]. They describe a GPU-based selective rendering algorithm for high-quality rasterization. Their sparse sampling approach makes use of a saliency model to evaluate only a set of sparse sample locations that can be used to compute an indirect lighting solution that is perceptually equivalent to full sampling. An inpainting algorithm is used to reconstruct a dense representation of the indirect lighting component, which is then combined with direct lighting to produce the final image (Fig. 12).

Instead of only accounting for visual perception, several selective rendering systems have been introduced that also consider multi-modal aspects. A survey by Hulusić et al. [HHD*12] gives information on the perceptual and cross-modal influences that have to be considered in the course of generating spatialized sound. Harvey et al. [HDBRC16] investigate the effect of spatialized directional sound on the visual attention of a user towards certain objects contained in the rendered imagery. Hulusić et al. [HCD*09] show that the beat-rate of an audio cue has a substantial impact on viewer perception of a video and video frame rate, allowing for the manipulation of the temporal visual perception. Bonneel et

al. [BSVDD10] analyze how the auditory and visual level of detail influence the perceived quality of audio-visual rendering methods. They show strong interactions between auditory and visual levels of detail in the process of material similarity perception.

The current state-of-the-art shows that perception-based approaches that rely on models of the HVS have traditionally been used in global illumination computation, e.g., by adapting sampling for (bi-directional) path tracing. Due to increasing processing powers, these methods are on the brink of appearing in real-time ray tracing systems as well. Moreover, and in contrast to methods targeting the sub-pixel level, deferred rendering systems and developments on multi-resolution shading allow efficient selective rendering using rasterization. Those methods will be further improved to enhance the visual quality and performance in consumer level VR and AR devices like HMDs. Along similar lines, further exploiting other perceptual channels and their cross-modal interaction will continue to improve presence in virtual environments and help to increase the overall performance of modern rendering systems. One example of a rather exotic approach is the acceleration method by Ramic et al. [RCHR07], where visual attention is drawn to certain objects within the scene using olfaction.

7. Measurement-based Perceptual Approaches

In this section, we present work that exploits limitations of visual perception (RQ3) by using an active measurement process. Such a process utilizes data from devices such as head- and eye trackers or inertial measurement units often integrated in modern head-mounted devices [LYKA14,SGE*15]. Using eye tracking, the gaze location is derived from the location of the pupil center mapped into screen space.

To use active measurements for rendering while avoiding visible artifacts, both latency and accuracy of the measuring process have to be considered. Fei-Fei et al. [FFIKP07] study the detection of animated objects in shortly presented scenes. The authors find that an animated object in a static environment can be detected in less than 27 ms. Dedicated measurements of acceptable latencies for gaze-contingent displays have been conducted in several studies [LW07, SRIR07, SW14, RJG*14]. The measured *end-to-end latency* comprises the full gaze capture and rendering pipeline, starting with capturing the tracking camera frame and ending with the reception of display-emitted photons of the retina. The gaze-contingent display system presented by Santini et al. [SRIR07] renders at a frame rate of 200 Hz and achieves an end-to-end latency of only 10 ms with dedicated hardware. Loschky and Wolverton [LW07] test for perceptually acceptable latencies with respect to peripheral image blur of different sizes. Their study reveals, that for an image change to go undetected, the update must be started at 5 ms to 60 ms after an eye movement, depending on the blurred area's angular distance from the fovea. In addition, the acceptable delay depends on the task of the application and the stimulus size in the visual field. Beyond that delay, the likelihood of detection increases quickly [LM00,LW07]. A VR application that uses inertial measures was presented by Xiao and Benko [XB16]. Sparse measurements of scene radiance are used to control colors of an LED matrix in the periphery in order to widen the FOV of VR headsets, thus exploiting the low acuity limit at high eccentricities.

7.1. Scene Simplification

Perceptual models can be used to reduce the number of polygons in areas of lower acuity, making view-dependent geometric LoD a typical example of gaze-contingent rendering, e.g., [OYT96, Hop98, HSH10].

Ohshima et al. [OYT96] employ gaze-aware LoD rendering in order to interact with multiple objects in a virtual environment. Besides a simple model of the visual acuity and eccentricity (Sec. 4.1.1), the authors take additional perceptual clues from kinetic and binocular vision into account when selecting a model from a set of precomputed models at different LoDs. In contrast, Luebke et al. [DLW00] simplify geometry directly in accordance with gaze. To remain visually imperceptible, the degree of mesh simplification is controlled by a perceptual model that exploits the CMF and Kelly's [Kel79] temporal CSF (Sec. 4.1.6). This is done by considering the eye's motion-induced sensitivity loss in spatial acuity. Reddy [Red97, pp. 105–129] proposes a two-stage approach for generating and selecting LoDs. In the offline stage, each object is analyzed in the spatial as well as the frequency domain. A perception-inspired color metric (Sec. 4.1.3) is used to generate LoDs with defined maximum spatial frequencies. In the online stage, a perceptual model (including visual acuity) and a custom CSF model are used to select the appropriate object LoD based on the projected object rotation, relative size, the user's gaze direction and pre-computed object data. Along similar lines, Howlett et al. [HHO04, HHO05] use eye tracking to detect salient features that can be improved by better geometric approximations during a mesh simplification process. Murphy and Duchowski [MD01] propose a non-isotropic LoD rendering approach using eye tracking for meshes based on a 3D spatial degradation function derived via a user study. Reddy [Red01] describes a system that recursively subdivides terrain meshes until the projected polygon size reaches an imperceptibility threshold that is coupled to a spatio-temporal CSF-model based on Mannos and Sakrison and a CMF model (Sec. 4.1.2). Papadopoulos and Kaufmann [PK13] use tracking in front of a large high-resolution display wall to adapt the visualization of gigapixel images to the user's physiological capabilities and FOV. Along similar lines, Weier et al. [WHS15] employ head tracking and a hybrid voxel-/polygonal representation for rendering to such a VR system. The LoD is adaptively degraded based on the user's position and FOV, modeling the acuity loss at increasing eccentricities (Sec. 4.1.1).

Another study on the influence of the LoD reduction on visual search tasks was performed by Parkhurst and Niebur [PN04]. The experiments are based on two measures, rotational velocity of the viewport and eccentricity. Even though the authors found that search times increase with decreasing LoDs beyond a critical threshold, the increase in the responsiveness of the system is usually more important than the decrease in search times. In a later study by Watson et al. [WWH04], the goal was to identify the conditions in which a stimulus first becomes perceivable (supra-threshold). Overall, they found contrast to be a better predictor of the overall search performance and perceptibility than feature size. Thus, detail should be added to low contrast regions first. Additionally, adding detail to peripheral before foveal regions when design-

ing LoD controls has a higher impact on search times and capturing a scene's gist.

Geometric techniques that reduce the scene's complexity drastically reduce the workload for geometry processing. However, systems that adapt a scene's complexity according to the user's gaze have not seen much attention lately. Due to the limited rasterization and ray-casting performance of graphics hardware of previous generations, scene simplification techniques usually led to huge speed-up factors. However, in current pipelines for real-time rendering, shading often dominates rendering costs [VST*14, HGF14]. As consumer-level eye tracking devices become available, we expect to see novel approaches targeting methods for gaze-contingent geometric simplification, tessellation and a gaze-aware adaptation of other features like materials.

7.2. Adaptive Sampling

Adaptation of the shading quality is another important aspect of rendering systems that exploit active measurements. We present rasterization and ray-based methods to adapt sampling based on the user's gaze (gaze-contingent rendering). Their goal is to exploit the spatial falloff of the visual acuity (Sec. 3.2). Early works in gaze-contingent rendering primarily observed the general detectability and influence of the quality degradation on visual performance [PP99, PLN01, NNB*04, DBMB06]. However, those approaches have not boosted rendering performance. Watson et al. [WWHW97] evaluated the effect of acuity degradation in the periphery of head-mounted displays on search performance. The results indicated that image resolution, i.e., spatial acuity and color, could be reduced by half in the periphery without a significant loss of performance. These results have been confirmed by a study presented by Duchowski et al. [DBS*09] that estimates visual search times for an anisotropic color-degraded periphery. Additionally, the results imply that chromatic detail levels cannot be reduced as readily as geometric or pixel detail at increasing eccentricities. "Foveated 3D graphics" simulates the acuity fall-off by rendering three nested layers of increasing angular diameter and decreasing resolution around the gaze direction [GFD*12]. These are blended together (Fig. 13). For the decrease in resolution, a model based on the CMF (Sec. 4.1.1) is employed. The technique achieves impressive shading reductions but also introduces overheads by repeating the rasterization for each nested layer.

Adapting sampling rates and shading complexity to the scene content for rasterization requires efficient *multi-rate* and *multi-resolution shading*. Vaidyanathan et al. [VST*14] (6.2.2) tested their approach for gaze-contingent rendering by using a simplified acuity model. Assuming static gaze and a constant radial acuity function, shading is computed at full-resolution in the foveal region and at a lower rate towards the periphery. The Valve Corporation integrated a practical implementation of multi-resolution rendering into the Source Engine™ [Vla16]. Most recently, another prototype by NVIDIA from Patney et al. [PSK*16] was introduced. They carefully investigate the impact of several effects induced by a quality degeneration in the periphery by distorting images. The result of these preliminary studies led them to develop of a foveated renderer with MSAA and a saccade-aware temporal antialiasing (TAA) [Kar14] strategy to improve temporal stability

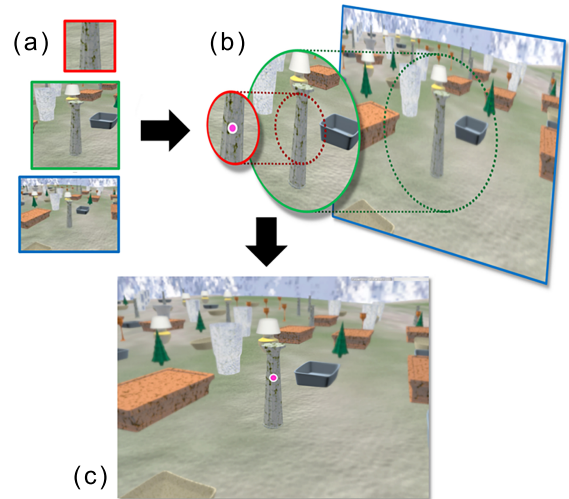


Figure 13: Foveated 3D Graphics by Guenter et al. [GFD*12]. Rasterization is performed at three different resolutions (d) according to acuity fall-off across the visual field. This reduces the number of shaded pixels. The results are then blended together (b). The combined image (c) approximates perceivable detail and is faster to compute than traditional full-resolution rendering. Image from Guenter et al. [GFD*12]

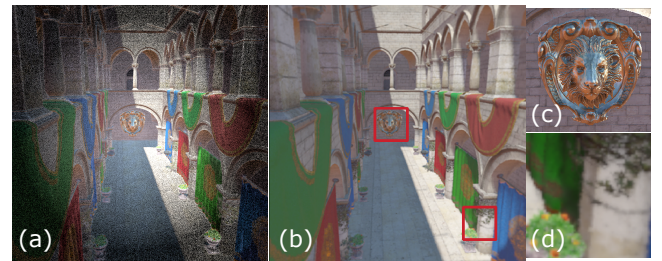


Figure 14: Gaze-contingent Adaptive Sampling by Stengel et al. [SGEM16]. Incorporating visual cues such as acuity, eye motion, adaptation and contrast, a perceptually-adaptive sampling pattern is computed and used for sparse shading (a). Fast image interpolation (b) achieves the same perceived quality at a fraction of the costs of shading each fragment. The resulting image contains high detail in the foveal region (c) and reduced detail in the periphery (d). Image from Stengel et al. [SGEM16]

and suppress aliasing artifacts critical to peripheral vision. Stengel et al. [SGEM16] present a paradigm for gaze-contingent rendering, which combines the benefits of sampling flexibility and fast rendering based on a deferred shading rasterization pipeline (Fig. 14). The sampling is steered by a perceptual model including acuity, active gaze motion and brightness adaptation. The selected samples are shaded and completed by fast image interpolation, resulting in images that are *perceptually equal* to images rendered with full per-pixel shading but at significantly reduced shading costs.

In contrast to rasterization, ray-based methods are inherently well-suited for the arbitrary sampling patterns needed for foveated rendering. An early system for ray-based volume rendering that

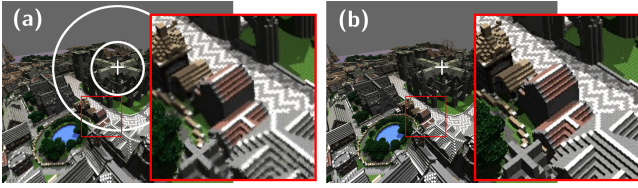


Figure 15: Ray Tracing-based foveated rendering in HMDs by Weier et al. [WRK*16] A gaze-adaptive and contrast-aware sampling is used to stochastically update pixels. The new samples are combined with reprojected samples from the previous frames to generate the final image (a). The presented user study shows, that even for small configurations of the foveal acuity, user's barely notice any difference to full rendering (b) Image from Weier et al. [WRK*16]

adapts sampling using eye tracking was developed by Levoy and Whitaker [LW90]. Here, the number of rays cast through the image plane and the number of samples drawn along each ray are adapted based on eccentricity. As sparsely sampling the image plane adaptive to the user's gaze leads to unsampled pixels and thus missing information, the authors additionally reconstruct a dense image, adapting masks and filtering kernels to the eccentricity. Fujita and Harada [FH14] present an approach to perform gaze-contingent rendering inside HMDs using ray tracing. They use a precomputed sampling pattern together with a kNN scheme to reconstruct a dense image. To improve image quality of foveated ray tracing in HMDs, Weier et al. [WRK*16, RWH*16] couple their system to a reprojection scheme (Fig. 15). The reprojection can be combined into a smoothly refined image allowing for TAA, where parts of the image that expose high contrasts are resampled, as these are critical to peripheral vision. Pohl et al. [PZB16] have extended their system to cope with lens astigmatism in HMDs to account for eye tracking as well (Sec. 6.2.2). They have implemented their gaze-contingent renderer as hybrid approach that uses rasterization and ray tracing in a single system by utilizing a precomputed sampling map.

Due to vast improvements in eye tracking solutions integrated in modern HMDs, research on gaze-contingent rendering is gaining increasing popularity. Although rasterization makes it inherently difficult to sample individual patterns in accordance to acuity models such as the CMF, advances such as deferred rendering and multi-resolution shading are already showing their potential to increase rendering performance. The question for the future is: How can locally changing rendering and shading quality make the most effective use of perceptual limits to produce photo-realistic scenes with the required flexibility? Capability for efficient low-latency rendering has been shown for both main rendering strategies, rasterization and ray-tracing. Although rasterization methods are faster on GPUs today, ray tracing does provide a higher degree of flexibility. Accordingly, ray-based methods could become the first choice for performance critical real-time VR rendering in head mounted devices [Hun15, FSTG16].

8. Summary

Knowledge of human perception can greatly improve the performance and quality of image synthesis. We have presented a general

overview of the human visual system and its limitations associated models to describe key visual processes or mechanisms, and we have described how they are used in various rendering techniques. We have seen that perceptual and cognitive models can be used directly to drive the design of optimized rendering techniques and how active measurements - in particular eye tracking - can aid (further) optimizing rendering.

In the following, we provide a résumé with respect to the initial research questions introduced in Section 2.

RQ1: What are the limitations of human visual perception?

After two centuries of research, the capabilities of human vision have been precisely measured. In particular, low-level knowledge on the eyes optical abilities (Sec. 3.1) and knowledge such as sensitivity, distribution, and interconnection of retinal photoreceptors are well-known (Sec. 3.2). As a result, models for contrast sensitivity, brightness adaptation and visual acuity work impressively well. Nonetheless, higher-level perception such as attention (Sec. 3.5) is difficult to measure and still not well-understood due the complexity of the involved parts of the brain (Sec. 3.4). In addition, individual differences between subjects may vary widely. The interplay of different senses is even more complex and largely unexplored. However, games and VR applications are breaking many boundaries and will increasingly provide multisensory experiences including vision, audio, and haptics, for which more research will be required. Even more so, we expect that further rendering optimization can be achieved by "tuning" the non-visual senses.

RQ2: How are these limitations modeled?

Models for low-level vision features (Sec. 4.1) such as spatial and temporal contrast sensitivity are fairly well studied. However, integrating them into high-level models (Sec. 4.2) is more complex. Often, these models describe high dimensional functions involving parameters such as environment lighting and display properties. Moreover, the process of calibrating spatio-temporal models is cumbersome and error-prone. The entire process is even more challenging when no reference images or information of the output devices are available (Sec. 4.2.2).

Saliency models (Sec. 5) combining bottom-up and top-down perception greatly increased accuracy compared to previous (purely bottom-up) saliency approaches. Still, predicting perception and attention from saliency computation is not accurate. Most models neither provide temporal stability nor are they able to provide a distinct gaze direction. This is a problem, as models for attention strongly depend on the gaze direction due to differences in foveal and peripheral vision. Furthermore, models for scan path prediction have received less attention in research and current models are far from viable. However, attentional blindness and inhibition concepts require knowledge about previously attended fixations.

In recent years, deep learning has shown great potential in modeling the HVS with higher precision and for general image data. Hence, it might provide a viable tool for more robust reference and non-reference metrics leading to novel methods in next years. Saliency evaluations could become more reliable. In addition, networks – once trained – are usually fast, compared to approaches

evaluating every parameter from scratch. Last but not least, models for multisensory perception hardly exist. A naïve combination of models for single modalities have been rendered non-optimal.

RQ3: How are limitations used in current state-of-the-art methods to accelerate rendering?

Perceptual models and insights of human perception already enable effective rendering algorithms. Due to the increase in geometric processing power of modern GPUs and the fact that shading cost often dominated in modern rendering pipelines, a shift of focus on methods that adapt sampling can be observed. State-of-the-art algorithms spend most of the computational effort where it matters most – namely in those regions that are critical to perception. Perceptual models allow guidance for ray-tracing processes into regions that need more samples, resulting in faster convergence to perceptually high-quality images. Recently, visual acuity models have also proven to be successful in rasterization (Sec. 6.2). Moreover, these approaches allow for the development of novel gaze-contingent rendering methods (Sec. 7.2). Adaptive local tone-mapping and brightness adaptation is another example of convincingly adopting perceptual properties and has led to a wide-spread use in games and movies (Sec. 4.1.4). Although perception models based on CSF and acuity are able to greatly reduce rendering cost, current applications perform sub-optimally due to hardware constraints (lacking active measures such as eye tracking) and implementation constraints. Hardware improvements with respect to perceptual algorithms, such as efficient pixel-precise shading, multi-resolution rendering and low-latency eye tracking, may significantly push the efficiency of current and upcoming algorithms. In direct comparison, methods that allow for gaze-contingent adaptive sampling or tone mapping by using such active measurements work dramatically better than their solely model-based counterparts. Other examples that can highly benefit from gaze tracking are subject calibration or methods optimized for interaction with wide field-of-view displays.

Properties of the HVS especially need to be considered for systems that require low-latency rendering to reduce nausea. Here rendering quality is ideally adapted to both the computational power of the rendering system and the capabilities of the human perception, guaranteeing lower bounds for the refresh rates. Current attentional methods have shown great potential to accelerate rendering. However, success of these methods is currently a trade-off between implementation effort for task description and response quality for rendering. In addition, complexity of fully automatic methods limits applicability to offline rendering. Massive use of eye tracking may simplify the creation of large-scale gaze databases and could lead to significant improvements in learning-based saliency methods. Unfortunately, current eye tracking devices are either costly or lack precision, have a high latency and are lacking a simple calibration procedure. Next-generation HMDs might better handle these requirements [SGE*15].

Perception-driven rendering has become a very important topic and new ways to simulate and exploit human vision are certain to be discovered in the coming years. With affordable tracking de-

vices, non-obtrusive ways to capture human attention and perception, novel VR devices available on a consumer level and ever increasing displays technologies, the evolvement of richer, more immersive, perception-driven computer graphics is only a question of time.

9. Acknowledgments

André Hinkenjann, Ernst Kruijff, Martin Weier and Thorsten Roth gratefully acknowledge funding by the Federal Ministry of Education and Research (13N13161 “OLIVE”). Marcus Magnor and Steve Grogoric gratefully acknowledge funding by the German Science Foundation (DFG MA2555/15-1 “Immersive Digital Reality”).

Glossary

Accommodation Mechanical ability of the eye to compress and relax the lens, enabling the eye to maintain focus on an object, so that a sharp image appears on the *retina*.

Adaptation Automatically triggered and time-dependent process of tuning sensitivity of the photosensitive *retina* to the amount of incoming light, also includes the *pupillary light reflex*.

Central Vision Part of the visual field that is projected onto the *fovea*, *parafovea* and *perifovea*, i.e., up to an eccentricity of up to 17°.

Cones Cone-shaped *photoreceptors* on the *retina* responsible for (*Photopic Vision*). They are tightly packed in the *fovea centralis* with their density decreasing quickly towards the periphery. Cones can be subdivided into **Long**, **Medium** and **Short**-Cones according to the band of the visual spectrum they are sensitive to.

Contrast Sensitivity Sensitivity to the difference in the light intensities of two adjacent areas [Gol10b, p. 411].

Contrast Sensitivity Function (CSF) A function defined over spatial frequency of a *sinusoidal grating pattern* yielding a subject's *contrast sensitivity*.

Cortical Magnification Factor (CMF) The linear extent of visual striate cortex to which each degree of the *retina* projects. It is directly proportional to visual acuity [CR74].

Critical Flicker Frequency (CFF) The frame rate at which a sequentially presented series of images appears as continuous, or is perceptually fused. Measured in Hertz (Hz).

Cross-modal Interaction Effects between various perceptual channels, e.g. visual stimuli might be missed when an auditory distractor is active.

Cycles per Degree (cpd) A unit to describe spatial frequency, defined as one period of a *sinusoidal grating pattern* at the projected size of 1 degree of the visual field.

Depth Cues Strategies such as eye convergence (binocular depth cue), motion parallax (monocular depth cue) and perspective for estimating the distance of an object.

Eccentricity Angular deviation from the center of the *fovea*.

Field of View (FoV) A measure describing the extent of the world

- observable by an optical system at one specific point in time, given in degrees. Using both eyes and looking straight ahead humans have an almost 180° horizontal field of view. If the eyeball rotation is included (and with the temporal restriction being relaxed) the horizontal field of view extends to 270°.
- Fixation** Gazing at a point of the scene or display for a certain time (fixation duration).
- Fovea (Centralis)** The area of the retina that is able to perceive and resolve visual information at the highest possible detail from approx. 5.2° around the central optical axis.
- High-level Perception** A field concerned with how known objects are recognized. The "top-down" processing of the human visual system.
- Human Visual System (HVS)** A model that describes the entire system that enables humans to perceive and process visual input including the eyes, visual pathways, visual cortex and the deeper neural processing.
- Hyperacuity** Perception of features that exceed the *visual acuity* .
- Inattentive Blindness Effect** A psychological lack of attention in which an individual fails to recognize an unexpected stimulus that is in plain sight.
- Interpupillary Distance (IPD)** The distance between the optical centers of the pupils.
- Just Noticeable Difference (JND)** A psycho-physical measure of how much a stimulus has to be changed in order for a difference to be perceivable in at least 50% of the cases.
- LMS Color Space** Represents colors, separated by their distribution into Long, Medium and Short wavelengths, corresponding to the *cone* types in the human eye.
- Low-level Perception** The "bottom-level" processing in the early stages of the human visual system. Models allow *saliency* estimation.
- Luminance** A photometric measure of the intensity per unit area of light emitted in a specific direction.
- M-Scaling Hypothesis** States that visual performance degradation with increasing eccentricity can be canceled out by spatial scaling of stimuli, by the inverse of the *CMF* .
- Mesopic Vision** A combination of photopic and scotopic vision occurring at dim light levels where both rods and cones are active.
- Minimum Angular Resolution (MAR)** Property to describe the resolution of an optical system. Resolution is expressed as the minimum angle allowing for the distinction of two points. For the eye and with normal vision this corresponds to about 1° when mapped to the *fovea* and decreases with increasing *eccentricity*.
- Motion Sickness** Over time conflicting visual and motion cues can result in motion sickness.
- Object of Interest (OOI)** An object or part of a scene the user is looking at. It can be either measured by using active *eye tracking* or approximated by *saliency* analysis.
- Parafovea** The area of the retina from approx. 5.2° to 9° around the central optical axis.
- Perifovea** The area of the retina from approx. 9° to 17° around the central optical axis.
- Peripheral Vision** Visual stimuli that are not within *central vision*.
- Photopic Vision** Color vision using the cone receptors under normal lighting conditions (daylight). *Rods* are permanently saturated and therefore deactivated under these conditions.
- Photoreceptor** Retinal cells (rods and cones) that convert light received at the *retina* into nerve signals. *Rods* are achromatic and sensitive to motion, while *cones* provide color sensitivity.
- Pupillary Light Reflex** The process of adjusting the pupil's diameter to the amount of incoming light as a part of *adaptation* .
- Receptive Field** A particular part of the sensory space in which a stimulus triggers a neuron. The receptive field of a *photoreceptor* can be described as a cone-shaped volume representing the directions in which light can trigger a response. For the *retina* it is the entire *visual field*.
- Retina** Photosensitive layer of the eye containing *photoreceptors*.
- Retinal Ganglion Cells** The output neurons containing circular *receptive fields* in order to encode and transmit information from the eye to the brain.
- Rods** Rod-shaped achromatic *photoreceptors* in the *retina* that are especially important in dim lighting conditions (*scotopic vision*).
- Saccade** A small rapid movement of the eye that occurs during the scanning of a scene and fixation changes.
- Saccadic Suppression** The effect that the visual system seems to shut down to some degree during *saccades*. That is, even though the point of fixation moves at very high velocities during a *saccade*, blurred vision is not experienced.
- Saliency** The perceptual importance of parts in a scene and their likelihood to capture attention .
- Scan Path** A description for captured gaze behavior usually including spatial fixation locations and fixation durations .
- Scene Schema Hypothesis** States that objects that are unexpected/unusual in a specific context have a high *saliency* [HH99].
- Scotopic Vision** Monochromatic vision under low light-level conditions making use of the *rod receptors* exclusively.
- Simultaneous Contrasts** The effect that two colors when viewed side-by-side interact with each other and can lead to a different visual sensation.
- Simultaneous Masking** see *Visual Masking* .
- Singleton Hypothesis** States that the viewer's attention is drawn by stimuli that are locally unique and globally rare [TG02].
- Sinusoidal Grating Pattern** An alternating pattern of bright and dark areas at a specific or increasing frequency of a sine function. Used to measure a subject's contrast sensitivity.
- Smooth Pursuit Eye Motion (SPEM)** Smooth movement of the eyes when following a moving object, stands contrary to saccadic movements. Smooth pursuit and saccadic movements may occur in conjunction when an object is moving fast, so catch-up saccades may be required .
- Stereo Vision** Describes the human ability to combine two visual streams (*Stereopsis*) to improve visual performance, e.g., depth perception.

Stereopsis Process that fuses the visual input from both eyes to allow for *stereo vision*.

Supra-threshold A term to describe a stimulus large enough to produce a response. This can be an action potential in a sensory cell or even just a perceivable difference of a stimulus (see *Just Noticeable Difference*).

Tone Mapping Operator (TMO) A computational method to compute *Tone Mapping*. This includes methods for compressing the dynamic range of a high-dynamic-range image in order to display it on a low-dynamic-range device such as a typical computer screen.

Vergence Describes the process that is required to simultaneously rotate both eyes into opposite directions to fixate an object.

Vergence-accommodation conflict Describes the discomfiting situation when stereo images are generated that convey depth information, which needs a conflicting *vergence* and *accommodation* to the one given by the actual screen's focal distance [SKHB11].

Vestibular System The mechanism in the ear to monitor the body's acceleration, equilibrium and relationship with the earth's gravitational field.

vestibular-ocular reflex Keeps the orientation of the eyes aligned with the current *OOI*, based on acceleration information from the vestibular system, amount of head rotation and retinal velocity.

Visual Acuity The ability to resolve small detail under ideal illumination conditions, i.e., the ability to detect and distinguish two points close to each other.

Visual Cortex The main part of the brain concerned with the sense of sight and the processing of visual information.

Visual Cues Signals or prompts derived from visual input. Such cues are preattentive by providing information from the environment subconsciously. Moreover, they might bring knowledge from previous experiences to mind.

Visual Difference Predictor (VDP) Daly's Visible Differences Predictor [Dal93] introduces a psycho-physical computational model of the HVS to compare two input images and derive a measure of perceivable differences. VDP processes images in the frequency domain. In contrast to *VDM*, it is particularly sensitive to differences near the visibility threshold.

Visual Discrimination Metric (VDM) The Sarnoff Visual Discrimination Metric [Lub95] introduces a psycho-physical computational model of the HVS to compare two input images. VDM derives a single *JND* value and a difference map. VDM processes images by convolution and down-sampling. In contrast to *VDP*, it is designed to generate a response above the *supra-threshold* at the expense of precision loss, when near threshold differences need to be judged.

Visual Equivalence Predictor (VEP) The VEP metric by Ramnarayanan et al. [RFWB07] introduces a psycho-physical computational model with the goal of measuring the visual equivalency of input images. Visually equivalency means the same impression of scene appearance is conveyed even though there can be measurable perceptual differences.

Visual Field see *Field of View*.

Visual Masking The reduction or elimination of a stimulus (target) by the presentation of a second stimulus (mask). The de-

tection threshold of the target can be affected by the interfering masking stimulus when closely coupled in space and time.

References

- [Ade82] ADELSON E. H.: Saturation and adaptation in the rod system. *Vision Research* 22, 10 (1982), pp. 1299–1312. 4
- [AET96] ANDERSON R. S., EVANS D. W., THIBOS L. N.: Effect of window size on detection acuity and resolution acuity for sinusoidal gratings in central and peripheral vision. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision* 13, 4 (1996), pp. 697–706. 7
- [AKF13] ADZIC V., KALVA H., FURHT B.: Exploring Visual Temporal Masking for Video Compression. In *2013 IEEE International Conference on Consumer Electronics* (2013), ICCE 2013, pp. 590–591. 9
- [AKLA11] ADLER F. H., KAUFMAN P. L., LEVIN L. A., ALM A.: *Adler's Physiology of the Eye*. Elsevier Health Sciences, 2011. 3, 4, 5, 9
- [AL73] ARDEN G. B., LEWIS D. R.: The pattern visual evoked response in the assessment of visual acuity. *Transactions of the Ophthalmological Societies of the United Kingdom* 93, 0 (1973), pp. 39–48. 7
- [ALK12] AILA T., LAINE S., KARRAS T.: *Understanding the Efficiency of Ray Traversal on GPUs - Kepler and Fermi Addendum*. Technical Report NVR-2012-02, NVIDIA, 2012. HPG2012 poster. 17
- [AM03] AKENINE-MÖLLER T.: *An Extremely Inexpensive Multisampling Scheme*. Technical Report No. 03-14, Ericsson Mobile Platforms AB and Chalmers University of Technology, Aug 2003. 15
- [APY16] AMIRSHAHI S. A., PEDERSEN M., YU S. X.: Image Quality Assessment by Comparing CNN Features between Images. *Journal of Imaging Science and Technology* 60, 6 (2016), 60410. 11
- [ARH78] ANSTIS S., ROGERS B., HENRY J.: Interactions between simultaneous contrast and coloured afterimages. *Vision Research* 18, 8 (1978), pp. 899–911. 8
- [ASG15] AYDIN T. O., SMOLIC A., GROSS M.: Automated aesthetic analysis of photographic images. *IEEE Transactions on Visualization and Computer Graphics* 21, 1 (2015), pp. 31–42. 7
- [BAHLC09] BENOIT A., ALLEYSSON D., HERAULT J., LE CALLET P.: *Spatio-temporal Tone Mapping Operator Based on a Retina Model*. Springer, Berlin, Heidelberg, Mar 2009, pp. 12–22. Computational Color Imaging: Second International Workshop, CCIW. 8
- [Bak49] BAKER H. D.: The course of foveal light adaptation measured by the threshold intensity increment. *Journal of the Optical Society of America* 39, 2 (1949), pp. 172–179. 4
- [Bar99] BARTEN P. G.: *Contrast sensitivity of the human eye and its effects on image quality*, vol. 72. HV Press, Knegsel, 1999. 7, 10
- [BBS09] BRATKOVA M., BOULOS S., SHIRLEY P.: oRGB: A Practical Opponent Color Space for Computer Graphics. *IEEE Computer Graphics and Applications* 29 (2009), pp. 42–55. 8
- [BCNS16] BIANCO S., CELONA L., NAPOLETANO P., SCHETTINI R.: On the use of deep learning for blind image quality assessment. *CoRR, preprint arXiv:1602.05531* (2016). 12
- [BDK*06] BOEHME M., DORR M., KRAUSE C., MARTINETZ T., BARTH E.: Eye movement predictions on natural videos. *Neurocomputing* 69 (Oct 2006), pp. 1996–2004. 14
- [BI13] BORJI A., ITTI L.: State-of-the-Art in Visual Attention Modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 35, 1 (2013), pp. 185–207. 12
- [BJD*15] BYLINSKII Z., JUDD T., DURAND F., OLIVA A., TORRALBA A.: MIT Saliency Benchmark, 2015. URL: http://saliency.mit.edu/results_mit300.html. 12, 13
- [BKM05] BATTISTA J., KALLONIATIS M., METHA A.: Visual function: the problem with eccentricity. *Clinical & Experimental Optometry* 88, 5 (2005), pp. 313–321. 7

- [BLBI13] BEGHDAI A., LARABI M.-C., BOUZERDOUM A., IFTEKHARUDDIN K.: A survey of perceptual image processing methods. *Signal Processing: Image Communication* 28, 8 (Sep 2013), pp. 811–831. 9
- [BLL*10] BERGER K., LIPSKI C., LINZ C., SELLENT A., MAGNOR M.: A ghosting artifact detector for interpolated image quality assessment. In *Proc. IEEE International Symposium on Consumer Electronics (ISCE)* (2010). 11
- [BM95] BOLIN M. R., MEYER G. W.: A Frequency Based Ray Tracer. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques* (1995), SIGGRAPH '95, ACM, pp. 409–418. 8, 15
- [BM98] BOLIN M. R., MEYER G. W.: A Perceptually Based Adaptive Sampling Algorithm. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques* (1998), SIGGRAPH '98, ACM, pp. 299–309. 8, 10, 15
- [BMW516] BOSSE S., MANIRY D., WIEGAND T., SAMEK W.: A deep neural network for image quality assessment. In *Image Processing (ICIP), 2016 IEEE International Conference on* (2016), IEEE, pp. 3773–3777. 12
- [BNR09] BALAS B., NAKANO L., ROSENHOLTZ R.: A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision* 9, 12 (2009), pp. 11–18. 6
- [BO00] BREITMEYER B. G., OGMEN H.: Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & Psychophysics* 62, 8 (2000), pp. 1572–1595. 8
- [Bou70] BOUMA H.: Interaction Effects in Parafoveal Letter Recognition. *Nature* 226, 5241 (1970), pp. 177–178. 6
- [BSA91] BANKS M. S., SEKULER A. B., ANDERSON S. J.: Peripheral spatial vision: Limits imposed by optics, photoreceptors, and receptor pooling. *Journal of the Optical Society of America* 8, 11 (1991), pp. 1775–1787. 4
- [BSVDD10] BONNEEL N., SUIED C., VIAUD-DELMON I., DRETTAKIS G.: Bimodal Perception of Audio-visual Material Properties for Virtual Environments. *ACM Transactions on Applied Perception (TAP)* 7, 1 (2010), pp. 1–16. 18
- [Čad04] ČADÍK M.: Human Perception and Computer Graphics, 2004. Postgraduate Study Report DC-PSR-2004-06, Czech Technical University. 10
- [CBN05] CHUA H. F., BOLAND J. E., NISBETT R. E.: Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences of the United States of America* 102, 35 (2005), pp. 12629–12633. 12
- [CCW03] CATER K., CHALMERS A., WARD G.: Detail to Attention: Exploiting Visual Tasks for Selective Rendering. In *Proceedings of the 14th ACM Eurographics Symposium on Rendering* (2003), vol. 44 of EGSR '03, pp. 270–280. 13, 16
- [CDdS06] CHALMERS A., DEBATTISTA K., DOS SANTOS L. P.: Selective rendering: Computing only what you see. In *Proceedings of the 4th International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia* (2006), GRAPHITE '06, pp. 9–18. 17
- [CGEB07] CORSINI M., GELASCA E. D., EBRAHIMI T., BARNI M.: Watermarked 3D Mesh Quality Assessment. *IEEE Transactions on Multimedia (TOMM)* 9, 2 (2007), pp. 247–256. 14
- [CH07] CHANDLER D. M., HEMAMI S. S.: VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images. *IEEE Transactions on Image Processing (TIP)* 16, 9 (2007), pp. 2284–2298. 11
- [CHEK08] CERF M., HAREL J., EINHÄUSER W., KOCH C.: Predicting human gaze using low-level saliency combined with face detection. *Advances in Neural Information Processing Systems* 20 (2008), pp. 1–7. 13
- [ČHM*12] ČADÍK M., HERZOG R., MANTIUK R., MYSZKOWSKI K., SEIDEL H.-P.: New Measurements Reveal Weaknesses of Image Quality Metrics in Evaluating Graphics Artifacts. *ACM SIGGRAPH Asia 2012, Transactions on Graphics (TOG)* 31, 6 (2012), pp. 147–157. 11
- [CLL*13] CORSINI M., LARABI M.-C., LAVOUÉ G., PETŘÍK O., VÁŠA L., WANG K.: Perceptual metrics for static and dynamic triangle meshes. *ACM Eurographics '12 - STAR, Computer Graphics Forum* 32, 1 (2013), pp. 101–125. 2, 14
- [CR74] COWEY A., ROLLS E. T.: Human Cortical Magnification Factor and its Relation to Visual Acuity. *Experimental Brain Research* 21, 5 (1974), pp. 447–454. 7, 21
- [CS06] CLEJU I., SAUPE D.: Evaluation of supra-threshold perceptual metrics for 3D models. In *Proceedings of the 3rd symposium on Applied perception in graphics and visualization* (2006), ACM APGV '06, pp. 41–44. 14
- [CSKH90] CURCIO C. A., SLOAN K. R., KALINA R. E., HENDRICKSON A. E.: Human photoreceptor topography. *Journal of Comparative Neurology* 292, 4 (1990), pp. 497–523. 3, 4
- [CTH*14] CLARBERG P., TOTH R., HASSELGREN J., NILSSON J., AKENINE-MÖLLER T.: AMFS: Adaptive Multi-Frequency Shading for Future Graphics Processors. *SIGGRAPH '14, Transactions on Graphics (TOG)* 33, 4 (2014), pp. 141–152. 17
- [CV95] CUTTING J. E., VISHTON P. M.: Perceiving layout and knowing distances: the integration, relative potency and contextual use of different information about depth. In *Handbook of Perception and Cognition*., Epstein W., Rogers S., (Eds.), vol. 5: Perception of Space and Motion. 1995, pp. 69–117. 3
- [CW11] CUNNINGHAM D. W., WALLRAVEN C.: *Experimental Design: From User Studies to Psychophysics*. CRC Press, Taylor & Francis Group, Nov 2011. 12
- [Dac11] DACHSBACHER C.: Analyzing visibility configurations. *IEEE Transactions on Visualization and Computer Graphics, TVCG '17* 17, 4 (Apr. 2011), pp. 475–486. 16
- [Dal93] DALY S.: *Digital Images and Human Vision*. MIT Press, Cambridge, MA, USA, 1993, ch. The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity, pp. 179–206. 9, 10, 23
- [Dal98] DALY S. J.: Engineering observations from spatiovelocity and spatiotemporal visual models. In *Proceedings of SPIE - The International Society for Optical Engineering* 3299 (1998), pp. 180–191. 7
- [DBD*07] DRETTAKIS G., BONNEEL N., DACHSBACHER C., LEFEBVRE S., SCHWARZ M., VIAUD-DELMON I.: An Interactive Perceptual Rendering Pipeline using Contrast and Spatial Masking. In *Proceedings of the 18th Eurographics conference on Rendering Techniques* (2007), ACM EGSR '07, pp. 297–308. 14
- [DBMB06] DORR M., BÖHME M., MARTINETZ T., BARTH E.: Gaze-Contingent Spatio-Temporal Filtering in a Head-Mounted Display. André E., Dybkjær L., Minker W., Neumann H., Weber M., (Eds.), In *Proceedings of Perception and Interactive Technologies: International Tutorial and Research Workshop, PIT '06*, Springer Berlin Heidelberg, pp. 205–207. 19
- [DBS*09] DUCHOWSKI A. T., BATE D., STRINGFELLOW P., THAKUR K., MELLOY B. J., GRAMOPADHYE A. K.: On Spatiochromatic Visual Sensitivity and Peripheral Color LOD Management. *ACM Transactions on Applied Perception (TAP)* 6, 2 (2009), 9. 19
- [DD00] DURAND F., DORSEY J.: Interactive Tone Mapping. In *Proceedings of the Eurographics Workshop on Rendering, Rendering Techniques 2000*, EGWR '00. Springer, 2000, pp. 219–230. 8
- [DER*10] DIDYK P., EISEMANN E., RITSCHEL T., MYSZKOWSKI K., SEIDEL H.-P.: Perceptually-motivated Real-time Temporal Upsampling of 3D Content for High-refresh-rate Displays. In *Computer Graphics Forum* (2010), vol. 29, pp. 713–722. Eurographics '10. 9
- [DLW00] DAVID LUEBKE BENJAMIN HALLEN D. N., WATSON B.: *Perceptually Driven Simplification Using Gaze-Directed Rendering*. Tech. rep., University of Virginia, 2000. CS-2000-04. 18

- [DMGB10] DORR M., MARTINETZ T., GEGENFURTNER K. R., BARTH E.: Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision* 10, 10 (2010), pp. 28–44. 14
- [DP08] D.G. PELLI K. T.: The uncrowded window of object recognition. *Nature Neuroscience* 11, 10 (2008), pp. 1129–1135. 6
- [DRE*11] DIDYK P., RITSCHER T., EISEMANN E., MYSZKOWSKI K., SEIDEL H.-P.: A Perceptual Model for Disparity. In *SIGGRAPH '11, ACM Transactions on Graphics (TOG)* (2011), vol. 30, ACM, pp. 96–104. 6, 9
- [DVB12] DORR M., VIG E., BARTH E.: Eye movement prediction and variability on natural video data sets. *Visual Cognition* 20, 4-5 (2012), pp. 495–514. 14
- [DW61] DANIEL P. M., WHITTERIDGE D.: The representation of the visual field on the cerebral cortex in monkeys. *The Journal of Physiology* 159, 2 (1961), pp. 203–221. 7
- [DW85] DIPPÉ M. A. Z., WOLD E. H.: Antialiasing Through Stochastic Sampling. *Proceedings of the 12th annual conference on Computer graphics and interactive techniques, SIGGRAPH '85* 19, 3 (1985), pp. 69–78. 15
- [EDL00] ENNS J. T., DI LOLLO V.: What's new in visual masking? *Trends in Cognitive Sciences* 4, 9 (2000), pp. 345–352. 8
- [Edw09] EDWARDS K. H.: *Optometry: Science, Techniques and Clinical Management*. Elsevier Health Sciences, 2009. 12
- [EJGAC*15] E. JACOBS D., GALLO O., A. COOPER E., PULLI K., LEVOY M.: Simulating the visual experience of very bright and very dark scenes. *ACM Transactions on Graphics (TOG)* 34, 3 (2015), pp. 25:1–25:15. 8
- [EMU15] EILERTSEN G., MANTIUK R. K., UNGER J.: Real-time noise-aware tone mapping. *ACM Transactions on Graphics (TOG), SIGGRAPH Asia '15* 34, 6 (2015), pp. 198–222. 7
- [Eri97] ERIC HORVITZ AND JED LENGYEL: Perception, Attention, and Resources: A Decision-Theoretic Approach to Graphics Rendering. In *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence, UAI '97* (1997), Morgan Kaufmann, pp. 238–249. 14
- [ERK08] EINHÄUSER W., RUTISHAUSER U., KOCH C.: Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision* 8, 2 (2008), pp. 2:1–2:19. 13
- [EUWM13] EILERTSEN G., UNGER J., WANAT R., MANTIUK R.: Survey and Evaluation of Tone Mapping Operators for HDR Video. In *ACM SIGGRAPH 2013 Talks* (2013), ACM, pp. 11:1–11:1. 8
- [Fai05] FAIRCHILD M. D.: *Color Appearance Models*. John Wiley & Sons, 2005. 8
- [Fai15] FAIRCHILD M. D.: Seeing, adapting to, and reproducing the appearance of nature. *Applied Optics* 54, 4 (2015), pp. 107–116. 8
- [FFIKP07] FEI-FEI L., IYER A., KOCH C., PERONA P.: What do we perceive in a glance of a real-world scene? *Journal of Vision* 7, 1 (2007), pp. 1–29. 18
- [FH14] FUJITA M., HARADA T.: Foveated Real-Time Ray Tracing for Virtual Reality Headset, 2014. *SIGGRAPH Asia '14* - Poster. 20
- [FMR08] FELZENSZWALB P., MCALLESTER D., RAMANAN D.: A Discriminatively Trained, Multiscale, Deformable Part Model. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR '08* (2008), IEEE, pp. 1–8. 13
- [FP04] FARRUGIA J.-P., PÉROCHE B.: A Progressive Rendering Algorithm Using an Adaptive Perceptually Based Image Metric. In *Computer Graphics Forum* (2004), vol. 23 of *Eurographics '04*, pp. 605–614. 16
- [FSPG96] FERWERDA J. A., PATTANAİK S. N., SHIRLEY P., GREENBERG D. P.: A Model of Visual Adaptation for Realistic Image Synthesis. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques* (1996), *SIGGRAPH '96*, ACM, pp. 249–258. 4, 8, 16
- [FR84] FISCHER B., RAMSPERGER E.: Human express saccades: extremely short reaction times of goal directed eye movements. *Experimental Brain Research* 57, 1 (1984), pp. 191–195. 5
- [FSPG97] FERWERDA J. A., SHIRLEY P., PATTANAİK S. N., GREENBERG D. P.: A model of visual masking for computer graphics. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques* (1997), *SIGGRAPH '97*, ACM, pp. 143–152. 9
- [FSTG16] FRISTON S., STEED A., TILBURY S., GAYDADJIEV G.: Construction and Evaluation of an Ultra Low Latency Frameless Renderer for VR. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 22, 4 (2016), pp. 1377–1386. 20
- [FWK63] FLOM M. C., WEYMOUTH F. W., KAHNEMAN D.: Visual Resolution And Contour Interaction. *Journal of the Optical Society of America* 53 (1963), pp. 1026–1032. 7
- [FWMG15] FRINTROP S., WERNER T., MARTIN GARCIA G.: Traditional Saliency Reloaded: A Good Old Model in New Shape. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR '15* (2015), pp. 82–90. 13
- [GAMS05] GUTIERREZ D., ANSON O., MUNOZ A., SERON F.: Perception-based Rendering: Eyes Wide Bleached. *Eurographics '05, Short Presentations* 5 (2005), pp. 49–52. 8
- [GBW*09] GRELAUD D., BONNEEL N., WIMMER M., ASSELOT M., DRETTAKIS G.: Efficient and Practical Audio-visual Rendering for Games Using Crossmodal Perception. In *Proceedings of the 2009 Symposium on Interactive 3D Graphics and Games* (2009), *I3D '09*, ACM, pp. 177–182. 15
- [GCO06] GAL R., COHEN-OR D.: Salient Geometric Features for Partial Shape Matching and Similarity. *ACM Transactions on Graphics (TOG)* 25, 1 (Jan. 2006), pp. 130–150. 15
- [GDS14] GALEA S., DEBATTISTA K., SPINA S.: GPU-Based Selective Sparse Sampling for Interactive High-Fidelity Rendering. In *6th International Conference on Games and Virtual Worlds for Serious Applications* (2014), *VS-GAMES '14*, IEEE, pp. 1–8. 17
- [GFD*12] GUENTER B., FINCH M., DRUCKER S., TAN D., SNYDER J.: Foveated 3D Graphics. *SIGGRAPH Asia '12, ACM Transactions on Graphics (TOG)* 31, 6 (Nov. 2012), pp. 164:1–164:10. 19
- [GGC*09] GONZÁLEZ C., GUMBAU J., CHOVER M., RAMOS F., QUIRÓS R.: User-assisted Simplification Method for Triangle Meshes Preserving Boundaries. *Computer-Aided Design* 41, 12 (Dec. 2009), pp. 1095–1106. 14
- [GHR84] GERVAIS M. J., HARVEY L. O., ROBERTS J. O.: Identification confusions among letters of the alphabet. *Journal of Experimental Psychology. Human Perception and Performance* 10, 5 (Oct. 1984), pp. 655–666. 7
- [Gol10a] GOLDSTEIN E. B.: *Encyclopedia of perception*. SAGE Publications, Inc, 2010. 3, 5, 6, 7, 12
- [Gol10b] GOLDSTEIN E. B.: *Sensation and Perception*, 8th ed. Wadsworth-Thomson Learning, Pacific Grove, 2010. 3, 6, 21
- [Gre70] GREEN D. G.: Regional variations in the visual acuity for interference fringes on the retina. *The Journal of Physiology* 207, 2 (1970), pp. 351–356. 7
- [Guo98] GUO B.: Progressive Radiance Evaluation Using Directional Coherence Maps. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques* (1998), *SIGGRAPH '98*, ACM, pp. 255–266. 16
- [GvA30] GRANIT R., VON AMMON W.: Comparative Studies On The Peripheral And Central Retina. *American Journal of Physiology – Legacy Content* 95, 1 (1930), pp. 229–241. 9
- [GVBL15] GUO J., VIDAL V., BASKURT A., LAVOUÉ G.: Evaluating the Local Visibility of Geometric Artifacts. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception* (2015), *SAP '15*, ACM, pp. 91–98. 14

- [GVC03] GABORSKI R., VAINGANKAR V. S., CANOSA R.: Goal directed visual search based on color cues: Cooperative effects of top-down & bottom-up visual attention. *Proceedings of the Artificial Neural Networks in Engineering, Rolla, Missouri 13* (2003), pp. 613–618. 13
- [GW07] GONZALEZ R. C., WOODS R. E.: *Digital Image Processing*, 3 ed. Prentice Hall, Upper Saddle River, N.J, 2007. 8
- [HA90] HAEBERLI P., AKELEY K.: The Accumulation Buffer: Hardware Support for High-quality Rendering. In *Proceedings of the 17th Annual Conference on Computer Graphics and Interactive Techniques* (1990), vol. 24 of *SIGGRAPH '90*, pp. 309–318. 15
- [HAML05] HASSELGREN J., AKENINE-MÖLLER T., LAINE S.: A Family of Inexpensive Sampling Schemes. In *Computer Graphics Forum* (2005), vol. 24, pp. 843–848. 15
- [HBCM07] HENDERSON J. M., BROCKMOLE J. R., CASTELHANO M. S., MACK M.: Visual saliency does not account for eye movements during visual search in real-world scenes. *Eye movements: A window on mind and brain* (2007), pp. 537–562. 12, 13
- [HČA*12] HERZOG R., ČADÍK M., AYDČIN T. O., KIM K. I., MYSZKOWSKI K., SEIDEL H.-P.: NoRM: No-Reference Image Quality Metric for Realistic Image Synthesis. In *Computer Graphics Forum* (2012), vol. 31, pp. 545–554. 7, 11, 12
- [HCD*09] HULUSIĆ V., CZANNER G., DEBATTISTA K., SIKUDOVA E., DUBLA P., CHALMERS A.: Investigation of the beat rate effect on frame rate for animated content. In *Proceedings of the 25th Spring Conference on Computer Graphics* (2009), SCCG '09, ACM, pp. 151–159. 17
- [HCOB10] HELD R. T., COOPER E. A., O'BRIEN J. F., BANKS M. S.: Using blur to affect perceived distance and size. *ACM Transactions on Graphics (TOG)* 29, 2 (2010). 3
- [HCS10] HASIC J., CHALMERS A., SIKUDOVA E.: Perceptually Guided High-fidelity Rendering Exploiting Movement Bias in Visual Attention. *ACM Transactions on Applied Perception (TAP)* 8, 1 (2010), pp. 6:1–6:19. 17
- [HDBRC16] HARVEY C., DEBATTISTA K., BASHFORD-ROGERS T., CHALMERS A.: Multi-Modal Perception for Selective Rendering. In *Computer Graphics Forum* (2016). 17
- [Her20] HERING E.: *Grundzüge der Lehre vom Lichtsinn*. Springer, 1920. 8
- [HGF14] HE Y., GU Y., FATAHALIAN K.: Extending the Graphics Pipeline with Adaptive, Multi-rate Shading. In *ACM Transactions on Graphics (TOG)* (2014), vol. 33 of *SIGGRAPH '14*, ACM, pp. 142:1–142:12. 17, 19
- [HH99] HENDERSON J. M., HOLLINGWORTH A.: High-Level Scene Perception. *Annual Review of Psychology* 50, 1 (Feb 1999), pp. 243–271. 13, 22
- [HHD*12] HULUSIC V., HARVEY C., DEBATTISTA K., TSINGOS N., WALKER S., HOWARD D., CHALMERS A.: Acoustic Rendering and Auditory-Visual Cross-Modal Perception and Interaction. *Computer Graphics Forum* 31, 1 (2012), pp. 102–131. 17
- [HHO04] HOWLETT S., HAMILL J., O'SULLIVAN C.: An experimental approach to predicting saliency for simplified polygonal models. In *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization* (New York, NY, USA, 2004), APGV '04, ACM, pp. 57–64. 18
- [HHO05] HOWLETT S., HAMILL J., O'SULLIVAN C.: Predicting and evaluating saliency for simplified polygonal models. *ACM Transactions on Applied Perception (TAP)* 2, 3 (July 2005), pp. 286–308. 18
- [HHQ*13] HAN J., HE S., QIAN X., WANG D., GUO L., LIU T.: An Object-Oriented Visual Saliency Detection Framework Based on Sparse Coding Representations. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)* 23, 12 (2013), pp. 2009–2021. 13
- [HJ57] HURVICH L. M., JAMESON D.: An opponent-process theory of color vision. *Psychological review* 64, 1(6) (1957), pp. 384–404. 8, 10
- [HK97] HONTSCH I., KARAM L. J.: APIC: adaptive perceptual image coding based on subband decomposition with locally adaptive perceptual weighting. In *Proceedings of the International Conference on Image Processing, 1997* (Oct. 1997), vol. 1, pp. 37–40. 9
- [HKP07] HAREL J., KOCH C., PERONA P.: Graph-based visual saliency. In *Advances in Neural Information Processing Systems 19* (2007), pp. 545–552. 12, 13
- [HLC*06] HO T.-C., LIN Y.-C., CHUANG J.-H., PENG C.-H., CHENG Y.-J.: User-assisted Mesh Simplification. In *Proceedings of the 2006 ACM International Conference on Virtual Reality Continuum and Its Applications* (2006), VRCIA '06, ACM, pp. 59–66. 14
- [HLSR13] HENDERSON J. M., LUKE S. G., SCHMIDT J., RICHARDS J. E.: Co-registration of eye movements and event-related potentials in connected-text paragraph reading. *Frontiers in Systems Neuroscience* 7 (2013), pp. 28:1 – 28:13. 14
- [HMYS01] HABER J., MYSZKOWSKI K., YAMAUCHI H., SEIDEL H.-P.: Perceptually guided corrective splatting. In *The European Association for Computer Graphics 22th Annual Conference, Computer Graphics Forum* (2001), vol. 20 of *Eurographics '01*, pp. 142–153. 16
- [Hop98] HOPPE H.: Smooth View-dependent Level-of-detail Control and Its Application to Terrain Rendering. In *Proceedings of the Conference on Visualization '98* (1998), VIS '98, IEEE, pp. 35–42. 18
- [How12] HOWARD I. P.: *Perceiving in Depth, Volume 1: Basic Mechanisms*. Oxford University Press, 2012. 5
- [HPG09] HANSEN T., PRACEJUS L., GEGENFURTNER K. R.: Color perception in the intermediate periphery of the visual field. *Journal of Vision* 9, 4 (2009), pp. 26:1–26:12. 4
- [HR95] HOWARD I. P., ROGERS B. J.: *Binocular vision and stereopsis*. Oxford University Press, USA, 1995. 3
- [HSH10] HU L., SANDER P. V., HOPPE H.: Parallel View-Dependent Level-of-Detail Control. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 16, 5 (2010), pp. 718–728. 18
- [Hun15] HUNT W.: Virtual Reality: The Next Great Graphics Revolution. Keynote Talk HPG, 2015. URL: <http://www.highperformancegraphics.org/wp-content/uploads/2015/Keynote1/WarrenHuntHPGKeynote.pptx>. 20
- [IK01] ITTI L., KOCH C.: Computational modelling of visual attention. *Nature Reviews Neuroscience* 2, 3 (2001), pp. 194–203. 13
- [IKN98] ITTI L., KOCH C., NIEBUR E.: A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 20, 11 (1998), pp. 1254–1259. 12, 13
- [JDT12] JUDD T., DURAND F., TORRALBA A.: A Benchmark of Computational Models of Saliency to Predict Human Fixations. *MIT Computer Science and Artificial Intelligence Laboratory Technical Report* (2012). MIT-CSAIL-TR-2012-001. 12, 13
- [JEDT09] JUDD T., EHINGER K., DURAND F., TORRALBA A.: Learning to Predict Where Humans Look. In *Proceedings of the IEEE Conference on Computer Vision* (2009), ICCV '09, IEEE, pp. 2106–2113. 13
- [JES*12] JARABO A., EYCK T. V., SUNDSTEDT V., BALA K., GUTIERREZ D., O'SULLIVAN C.: Crowd light: Evaluating the perceived fidelity of illuminated dynamic scenes. *Computer Graphics Forum, Eurographics '12* 31, 2, 4 (May 2012), pp. 565–574. 16
- [JESG12] JIMENEZ J., ECHEVARRIA J. I., SOUSA T., GUTIERREZ D.: SMAA: Enhanced Subpixel Morphological Antialiasing. In *Computer Graphics Forum* (2012), vol. 31 of *Eurographics '12*, pp. 355–364. 16
- [JF02] JOHNSON G. M., FAIRCHILD M. D.: On Contrast Sensitivity in an Image Difference Model. In *International Technical Conference on Digital Image Capture and Associated System, Reproduction and Image Quality Technologies* (2002), IS&T's Pics Conference '02, Society for Imaging Science & Technology, pp. 18–23. 7

- [JFN98] JIN E. W., FENG X.-F., NEWELL J.: The development of a color visual difference model (CVDm). In *International Technical Conference on Digital Image Capture and Associated System, Reproduction and Image Quality Technologies* (1998), IS&T's Pics Conference '98, Society for Imaging Science & Technology, pp. 154–158. 10
- [JGY*11] JIMENEZ J., GUTIERREZ D., YANG J., RESHETOV A., DEMOREUILLE P., BERGHOF T., PERTHUIS C., YU H., MCGUIRE M., LOTTES T., MALAN H., PERSSON E., ANDREEV D., TIAGO S.: Filtering Approaches for Real-Time Anti-Aliasing. *SIGGRAPH Courses 2*, 3 (2011), 4. 15, 16
- [JH91] J.C. HORTON W. H.: The representation of the visual field in human striate cortex. A revision of the classic Holmes map. *Archives of Ophthalmology 109*, 6 (1991), pp. 816–824. 5
- [JIC*09] JIN B., IHM I., CHANG B., PARK C., LEE W., JUNG S.: Selective and Adaptive Supersampling for Real-Time Ray Tracing. In *Proceedings of the Conference on High Performance Graphics* (2009), HPG '09, ACM, pp. 117–125. 15
- [JOK09] JANSEN L., ONAT S., KÖNIG P.: Influence of disparity on fixation and saccades in free viewing of natural scenes. *Journal of Vision 9*, 1 (2009), pp. 1–19. 12
- [KAB15] KRUTHIVENTI S. S., AYUSH K., BABU R. V.: DeepFix: A Fully Convolutional Neural Network for predicting Human Eye Fixations. *arXiv* (2015). 1510.02927. 13
- [Kar14] KARIS B.: High Quality Temporal Supersampling. *Advances in Real-Time Rendering in Games, ACM SIGGRAPH Courses*. (2014). 19
- [KDCM14a] KOULIERIS G. A., DRETTAKIS G., CUNNINGHAM D., MANIA K.: An Automated High-Level Saliency Predictor for Smart Game Balancing. *ACM Transactions on Applied Perception (TAP) 11*, 4 (Dec. 2014), pp. 17:1–17:21. 13, 15
- [KDCM14b] KOULIERIS G. A., DRETTAKIS G., CUNNINGHAM D., MANIA K.: C-LOD: Context-aware Material Level-of-Detail applied to Mobile Graphics. *Computer Graphics Forum, Eurographics Symposium on Rendering, EGSR '14 33*, 4 (2014), pp. 41–49. 15
- [Kel61] KELLY D. H.: *Visual Responses to Time-Dependent Stimuli*, vol. 51. *Journal of the Optical Society of America*, jul 1961, ch. II Single-Channel Model of the Photopic Visual System, p. 747. 9
- [Kel79] KELLY D. H.: *Motion and vision*, vol. 69. *Journal of the Optical Society of America*, Oct 1979, ch. II. Stabilized spatio-temporal threshold surface, pp. 1340–1349. 9, 18
- [KFC*10] KRÍVÁNEK J., FAJARDO M., CHRISTENSEN P. H., TABELLION E., BUNNELL M., LARSSON D., KAPLAYAN A.: Global illumination across industries. In *ACM SIGGRAPH 2010 Courses* (New York, NY, USA, 2010), SIGGRAPH '10, ACM. 10
- [KFSW09] KIENZLE W., FRANZ M. O., SCHÖLKOPF B., WICHMANN F. A.: Center-surround patterns emerge as optimal predictors for human saccade targets. *Journal of Vision 9*, 5 (2009), pp. 7:1–7:15. 5
- [KG03] KHO Y., GARLAND M.: User-guided Simplification. In *Proceedings of the Symposium on Interactive 3D Graphics* (2003), I3D '03, ACM, pp. 123–126. 14
- [KKK*14] KISHISHITA N., KIYOKAWA K., KRUIJFF E., ORLOSKY J., MASHITA T., TAKEMURA H.: Analysing the Effects of a Wide Field of View Augmented Reality Display on Search Performance in Divided Attention Tasks. In *IEEE International Symposium on Mixed and Augmented Reality* (2014), ISMAR '14, IEEE, pp. 177–186. 5
- [KMS05] KRAWCZYK G., MYSZKOWSKI K., SEIDEL H.-P.: Lightness Perception in Tone Reproduction for High Dynamic Range Images: Lightness Perception in Tone Reproduction. *Computer Graphics Forum 24*, 3 (2005), pp. 635–645. 8
- [Kow11] KOWLER E.: Eye movements: The past 25years. *Vision Research 51*, 13 (2011), 1457–1483. 5
- [KRMS15] KELLNHOFFER P., RITSCHEL T., MYSZKOWSKI K., SEIDEL H.-P.: A Transformation-Aware Perceptual Image Metric. In *SPIE/IS&T Electronic Imaging* (2015), International Society for Optics and Photonics, pp. 939408, pp. 1–14. 11
- [KSR*03] KINGSTONE A., SMILEK D., RISTIC J., FRIESEN C. K., EASTWOOD J. D.: Attention, researchers! It is time to take a look at the real world. *Current Directions in Psychological Science 12*, 5 (2003), pp. 176–180. 13
- [KTB14] KÜMMERER M., THEIS L., BETHGE M.: Deep Gaze I: Boosting Saliency Prediction with Feature Maps Trained on ImageNet. *arXiv:1411.1045* (2014). 13
- [KVJG10] KIM Y., VARSHNEY A., JACOBS D. W., GUIMBRETIÈRE F.: Mesh Saliency and Human Eye Fixations. *ACM Transactions on Applied Perception (TAP) 7*, 2 (2010), pp. 12:1–12:13. 15
- [KWB14] KÜMMERER M., WALLIS T., BETHGE M.: How close are we to understanding image-based saliency? *arXiv:1409.7686* (2014). 13
- [KYLD14] KANG L., YE P., LI Y., DOERMANN D.: Convolutional neural networks for no-reference image quality assessment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2014), pp. 1733–1740. 12
- [Lav07] LAVOUÉ G.: A Roughness Measure for 3D Mesh Visual Masking. In *Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization* (2007), APGV '07, ACM, pp. 57–60. 15
- [LCDC10] LO C.-H., CHU C.-H., DEBATTISTA K., CHALMERS A.: Selective rendering for efficient ray traced stereoscopic images. *The Visual Computer 26*, 2 (2010), pp. 97–107. 17
- [LDC06] LONGHURST P., DEBATTISTA K., CHALMERS A.: A GPU Based Saliency Map for High-fidelity Selective Rendering. In *Proceedings of the 4th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa* (2006), AFRIGRAPH '06, ACM, pp. 21–29. 12, 16
- [LE13] LENKIC P., ENNS J.: Apparent Motion Can Impair and Enhance Target Visibility: The Role of Shape in Predicting and Postdicting Object Continuity. *Frontiers in Psychology 4* (2013), 35. 6
- [LF80] LEGGE G. E., FOLEY J. M.: Contrast masking in human vision. *Journal of the Optical Society of America 70*, 12 (1980), pp. 1458–1471. 8
- [LH01] LUEBKE D. P., HALLEN B.: Perceptually-Driven Simplification for Interactive Rendering. In *Proceedings of the 12th Eurographics Workshop on Rendering Techniques* (2001), EGWR '01, Springer-Verlag, pp. 223–234. 14
- [LLV16] LAVOUÉ G., LARABI M. C., VÁŠA L.: On the Efficiency of Image Metrics for Evaluating the Visual Quality of 3D Models. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 22*, 8 (Aug 2016), pp. 1987–1999. 14
- [LM00] LOSCHKY L. C., MCCONKIE G. W.: User Performance With Gaze Contingent Multiresolutional Displays. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (2000), ETRA '00, ACM, pp. 97–103. 18
- [LMC16] LE MEUR O., COUTROT A.: Introducing context-dependent and spatially-variant viewing biases in saccadic models. *Vision research 121* (2016), 72–84. 13
- [LMK01] LI B., MEYER G. W., KLASSEN R. V.: A Comparison of Two Image Quality Models. In *Proceedings of SPIE* (2001), Rogowitz B. E., Pappas T. N., (Eds.), The International Society for Optical Engineering, pp. 98–109. 10
- [LMS10] LOPEZ F., MOLLA R., SUNDSTEDT V.: Exploring Peripheral LOD Change Detections during Interactive Gaming Tasks. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization* (2010), APGV '10, ACM, pp. 73–80. 6
- [LN07] LINDEMAN R. W., NOMA H.: A Classification Scheme for Multi-sensory Augmented Reality. In *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology* (2007), VRST '07, ACM, pp. 175–178. 6
- [LSC04] LEDDA P., SANTOS L. P., CHALMERS A.: A local model of eye adaptation for high dynamic range images. In *Proceedings of the 3rd*

- International Conference on Computer Graphics, Virtual Reality, Visualization and Interaction in Africa* (2004), Afrigraph '04, ACM, pp. 151–160. 5, 8
- [LT00] LINDSTROM P., TURK G.: Image-driven Simplification. *ACM Transactions on Graphics (TOG)* 19, 3 (July 2000), pp. 204–241. 14
- [Lub95] LUBIN J.: A visual discrimination model for imaging system design and evaluation. *Vision models for target detection and recognition 2* (1995), pp. 245–357. 9, 10, 23
- [Lue03] LUEBKE D. P.: *Level of Detail for 3D Graphics*. Morgan Kaufmann Publishers, 2003. 14
- [Luk12] LUKAC R.: *Perceptual Digital Imaging: Methods and Applications*. CRC Press, 2012. 7, 8
- [LVJ05] LEE C. H., VARSHNEY A., JACOBS D. W.: Mesh Saliency. In *ACM Transactions on Graphics (TOG)* (2005), vol. 24, ACM, pp. 659–666. 14
- [LW90] LEVOY M., WHITAKER R.: Gaze-directed Volume Rendering. In *Proceedings of the 1990 Symposium on Interactive 3D Graphics* (1990), I3D '90, ACM, pp. 217–223. 20
- [LW07] LOSCHKY L. C., WOLVERTON G. S.: How late can you update gaze-contingent multiresolutional displays without detection? *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 3, 4 (2007), 7. 18
- [LWC*13] LIU Y., WANG J., CHO S., FINKELSTEIN A., RUSINKIEWICZ S.: A no-reference metric for evaluating the quality of motion deblurring. *ACM Transactions on Graphics (TOG)* 32, 6 (2013), 175:1–175:12. 11, 12
- [LYKA14] LAVALLE S. M., YERSHOVA A., KATSEV M., ANTONOV M.: Head tracking for the Oculus Rift. In *IEEE International Conference on Robotics and Automation* (May 2014), ICRA '14, pp. 187–194. 18
- [MB10] MOORTHY A. K., BOVIK A. C.: A two-step framework for constructing blind image quality indices. *IEEE Signal Processing Letters* 17, 5 (2010), 513–516. 12
- [MCNV14] MAUDERER M., CONTE S., NACENTA M., VISHWANATH D.: Using Gaze-Contingent Depth of Field to Facilitate Depth Perception. *i-Perception* 5, 5 (2014), pp. 473–473. 8
- [MCTB12] MAULE M., COMBA J. L., TORCHELSEN R., BASTOS R.: Transparency and anti-aliasing techniques for real-time rendering. In *25th Conference on Graphics, Patterns and Images, Tutorials* (2012), SIBGRAPI-T, IEEE, pp. 50–59. 15
- [MD01] MURPHY H., DUCHOWSKI A. T.: Gaze-contingent level of detail rendering. *EuroGraphics 2001* (2001). 18
- [MDMS05] MANTIUK R., DALY S. J., MYSZKOWSKI K., SEIDEL H.-P.: Predicting Visible Differences in High Dynamic Range Images: Model and its Calibration. In *Electronic Imaging 2005* (2005), International Society for Optics and Photonics, pp. 204–214. 10
- [MDWE02] MARZILIANO P., DUFAUX F., WINKLER S., EBRAHIMI T.: A no-reference perceptual blur metric. In *Image Processing. 2002. Proceedings. 2002 International Conference on* (2002), vol. 3 of *ICIP '02*, IEEE, pp. 57–60, III. 12
- [MG10] MENZEL N., GUTHE M.: Towards Perceptual Simplification of Models with Arbitrary Materials. In *Computer Graphics Forum* (2010), vol. 29 of *Pacific Graphics '10*, pp. 2261–2270. 14
- [Mit87] MITCHELL D. P.: Generating Antialiased Images at Low Sampling Densities. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques* (1987), vol. 21 of *SIGGRAPH '87*, ACM, pp. 65–72. 8, 15
- [MKRH11] MANTIUK R., KIM K. J., REMPEL A. G., HEIDRICH W.: HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions. In *ACM Transactions on Graphics (TOG)* (2011), vol. 30, ACM, p. 40. 7, 10
- [ML92] MEYER G. W., LIU A.: Color spatial acuity control of a screen subdivision image synthesis algorithm. In *Human Vision, Visual Processing, and Digital Display III* (1992), Rogowitz B. E., (Ed.), SPIE '92, The International Society for Optical Engineering, pp. 387–399. 15
- [MM05] M. MARTELLI N.J. MAJAJ D. P.: Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision* 5, 1 (2005), pp. 58–70. 6
- [MM13] MANTIUK R., MARKOWSKI M.: Gaze-Dependent Tone Mapping. In *Image Analysis and Recognition*. Springer, 2013, pp. 426–433. 8
- [MMB12] MITTAL A., MOORTHY A. K., BOVIK A. C.: No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Transactions on Image Processing* 21, 12 (2012), pp. 4695–4708. 12
- [MMBH10] MCNAMARA A., MANIA K., BANKS M., HEALEY C.: Perceptually-motivated Graphics, Visualization and 3D Displays. In *SIGGRAPH '10, Courses* (2010), ACM, pp. 7:1–7:159. 2
- [MMS15] MANTIUK R., MYSZKOWSKI K., SEIDEL H.-P.: *High Dynamic Range Imaging*. Wiley Encyclopedia of Electrical and Electronics Engineering, 2015. 8
- [MN84] MCKEE S., NAKAYAMA K.: The detection of motion in the peripheral visual field. *Vision Research* 24, 1 (1984), pp. 25–32. 9
- [MR98] MACK A., ROCK I.: Inattention blindness: Perception without attention. *Visual Attention* 8 (1998), pp. 55–76. 13
- [MRW96] MANNAN S. K., RUDDOCK K. H., WOODING D. S.: The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision* 10, 3 (1996), pp. 165–188. 12
- [MS74] MANNOS J., SAKRISON D.: The Effects of a Visual Fidelity Criterion on the Encoding of Images. *IEEE Transactions on Information Theory (TIT)* 20, 4 (Sept. 1974), pp. 525–536. 7
- [MS99] MULLEN K. T., SANKERALLI M. J.: Evidence for the stochastic independence of the blue-yellow, red-green and luminance detection mechanisms revealed by subthreshold summation. *Vision Research* 39, 4 (1999), pp. 733–745. 10
- [MSR*13] MIKAMO M., SLOMP M., RAYTCHEV B., TAMAKI T., KANEDA K.: Technical Section: Perceptually Inspired Afterimage Synthesis. *Computer & Graphics* 37, 4 (2013), pp. 247–255. 8
- [MTT04] MARTINEZ-TRUJILLO J. C., TREUE S.: Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current Biology* 14, 9 (2004), pp. 744–751. 13
- [Mul85] MULLEN K. T.: The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology* 359 (Feb. 1985), pp. 381–400. 4, 10
- [MVOP11] MURRAY N., VANRELL M., OTAZU X., PARRAGA C. A.: Saliency estimation using a non-parametric low-level vision model. In *IEEE Conference on Computer Vision and Pattern Recognition* (2011), CVPR '11, IEEE, pp. 433–440. 12
- [MWDG13] MASIA B., WETZSTEIN G., DIDYK P., GUTIERREZ D.: A survey on computational displays: Pushing the boundaries of optics, computation, and perception. *Computers & Graphics* 37, 8 (2013), pp. 1012–1038. 2
- [Mys98] MYSZKOWSKI K.: The visible differences predictor: applications to global illumination problems. In *Rendering Techniques '98*. Springer, 1998, pp. 223–236. 16
- [Mys02] MYSZKOWSKI K.: Perception-based Global Illumination, Rendering, and Animation Techniques. In *Proceedings of the 18th Spring Conference on Computer Graphics* (2002), SCCG '02, ACM, pp. 13–24. 16
- [NDSL15] NARWARIA M., DA SILVA M. P., LE CALLET P.: HDR-VQM: An objective quality measure for high dynamic range video. *Signal Processing: Image Communication* 35 (2015), pp. 46–60. 10
- [NE15] NUTHMANN A., EINHÄUSER W.: A new approach to modeling the influence of image features on fixation selection in scenes. *Annals of the New York Academy of Sciences* 1339, 1 (2015), pp. 82–96. 14
- [NH12] NUTHMANN A., HENDERSON J. M.: Using CRISP to model global characteristics of fixation durations in scene viewing and reading

- with a common mechanism. *Visual Cognition* 20, 4-5 (2012), pp. 457–494. 13
- [NI02] NAVALPAKKAM V., ITTI L.: A goal oriented attention guidance model. In *International Workshop on Biologically Motivated Computer Vision* (2002), BMCH '02, Springer, pp. 453–461. 13
- [NI07] NAVALPAKKAM V., ITTI L.: Search goal tunes visual features optimally. *Neuron* 53, 4 (2007), pp. 605–617. 12, 13
- [NKOE83] NOORLANDER C., KOENDERINK J. J., OLDEN R. J. D., EDENS B. W.: Sensitivity to spatiotemporal colour contrast in the peripheral visual field. *Vision Research* 23, 1 (1983), pp. 1–11. 4
- [NNB*04] NIKOLOV S. G., NEWMAN T. D., BULL D. R., CANAGARAJAH N. C., JONES M. G., GILCHRIST I. D.: Gaze-contingent display using texture mapping and OpenGL: System and Applications. In *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications* (2004), ETRA '04, ACM, pp. 11–18. 19
- [NPDSLCP14] NARWARIA M., PERREIRA DA SILVA M., LE CALLET P., PÉPION R.: On Improving the Pooling in HDR-VDP-2 towards Better HDR Perceptual Quality Assessment. In *Human Vision and Electronic Imaging 2014* (San Francisco, United States, Feb. 2014), pp. 1–6. 10
- [NRH04] NG R., RAMAMOORTHY R., HANRAHAN P.: Triple product wavelet integrals for all-frequency relighting. In *ACM Transaction on Graphics (TOG)* (2004), SIGGRAPH '04, ACM, pp. 477–487. 15
- [NSEH10] NUTHMANN A., SMITH T. J., ENGBERT R., HENDERSON J. M.: CRISP: a computational model of fixation durations in scene viewing. *Psychological Review* 117, 2 (2010), pp. 382–405. 13
- [NWHWD16a] NADER G., WANG K., HÉTROUY-WHEELER F., DUPONT F.: Just noticeable distortion profile for flat-shaded 3d mesh surfaces. *IEEE Transactions on Visualization & Computer Graphics (TVCG)* 22, 11 (2016), pp. 2423–2436. 14
- [NWHWD16b] NADER G., WANG K., HETROY-WHEELER F., DUPONT F.: Visual Contrast Sensitivity and Discrimination for 3D Meshes and their Applications. *Computer Graphics Forum (CGF), Pacific Graphics '16* (2016). 14
- [OLME16] OLIVIER LE MEUR S. N. P., EAKTA J.: Visual attention from a graphics point of view. In *Eurographics '16 Tutorial* (2016). URL: <http://jainlab.cise.ufl.edu/visual-attention-graphics-pov.html>. 13
- [OPPU09] OPREA C., PIRNOG I., PALEOLOGU C., UDREA M.: Perceptual video quality assessment based on salient region detection. In *Fifth Advanced International Conference on Telecommunications* (2009), AICT'09, IEEE, pp. 232–236. 12
- [OTCH03] OLIVA A., TORRALBA A., CASTELHANO M. S., HENDERSON J. M.: Top-down control of visual attention in object detection. In *Proceedings of the International Conference on Image Processing* (2003), vol. 1 of *ICIP '03*, IEEE, pp. 253–256. 12
- [OYT96] OHSHIMA T., YAMAMOTO H., TAMURA H.: Gaze-directed adaptive rendering for interacting with virtual space. In *Proceedings of the IEEE Virtual Reality Annual International Symposium* (1996), IEEE VR '96, IEEE, pp. 103–110. 18
- [Pai05] PAI D. K.: Multisensory Interaction: Real and virtual. In *The Eleventh International Symposium on Robotics Research* (2005), Springer, pp. 489–498. 6
- [Pal99] PALMER S. E.: *Vision Science: Photons to Phenomenology*. MIT Press, 1999. 3, 6
- [PB71] POSNER M. I., BOIES S. J.: Components of attention. *Psychological Review* 78, 5 (1971), pp. 391–408. 6
- [PBNG15] POHL D., BOLKART T., NICKELS S., GRAU O.: Using astigmatism in wide angle HMDs to improve rendering. In *Annual International Symposium on Virtual Reality* (2015), IEEE VR '15, pp. 263–264. 17
- [PK13] PAPADOPOULOS C., KAUFMAN A. E.: Acuity-Driven Gigapixel Visualization. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 19, 12 (2013), pp. 2886–2895. 18
- [PKC15] PIETRZAK J., KACPERSKI K., CIEŚLAR M.: NVIDIA OptiX ray-tracing engine as a new tool for modelling medical imaging systems. In *SPIE Medical Imaging* (2015), International Society for Optics and Photonics, p. 94122P. 17
- [PLN01] PARKHURST D., LAW I., NIEBUR E.: Evaluating Gaze-Contingent Level of Detail Rendering of Virtual Environments using Visual Search, 2001. http://cnslab.mb.jhu.edu/publications/Parkhurst_etal01c.pdf. 19
- [PM13] PETIT J., MANTIUK R. K.: Assessment of video tone-mapping: Are cameras' S-shaped tone-curves good enough? *Journal of Visual Communication and Image Representation* 24, 7 (2013), pp. 1020–1030. 7
- [PN03] PARKHURST D. J., NIEBUR E.: Scene content selected by active vision. *Spatial Vision* 16, 2 (2003), pp. 125–154. 12
- [PN04] PARKHURST D., NIEBUR E.: A feasibility test for perceptually adaptive level of detail rendering on desktop systems. In *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization* (2004), APGV '04, ACM, pp. 49–56. 18
- [Por02] PORTER T. C.: Contributions to the Study of Flicker. Paper II. *Proceedings of the Royal Society of London* 70 (1902), pp. 313–329. 9
- [PP99] PRIKRYL J., PURGATHOFER W.: Perceptually-driven termination for stochastic radiosity. In *Seventh International Conference in Central Europe on Computer Graphics and Visualization (Winter School on Computer Graphics)* (1999), WSCG '99. 19
- [PR98] PRINCE S. J., ROGERS B. J.: Sensitivity to disparity corrugations in peripheral vision. *Vision Research* 38, 17 (1998), pp. 2533–2537. 5
- [PRC84] POLLATSEK A., RAYNER K., COLLINS W. E.: Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General* 113, 3 (1984), pp. 426–442. 12
- [PRH90] POLLATSEK A., RAYNER K., HENDERSON J. M.: Role of spatial location in integration of pictorial information across saccades. *Journal of Experimental Psychology: Human Perception and Performance* 16, 1 (1990), pp. 199–210. 12
- [PS89] PAINTER J., SLOAN K.: Antialiased Ray Tracing by Adaptive Progressive Refinement. In *Proceedings of the 16th Annual Conference on Computer Graphics and Interactive Techniques* (1989), vol. 23 of *SIGGRAPH '89*, ACM, pp. 281–288. 15
- [PS03] POJAR E., SCHMALSTIEG D.: User-controlled Creation of Multiresolution Meshes. In *Proceedings of the 2003 Symposium on Interactive 3D Graphics* (2003), I3D '03, ACM, pp. 127–130. 14
- [PSK*16] PATNEY A., SALVI M., KIM J., KAPLANYAN A., WYMAN C., BENTY N., LUEBKE D., LEFOHN A.: Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics (TOG), SIGGRAPH Asia '16* 35, 6 (Nov. 2016), pp. 179:1–179:12. 19
- [PTYG00] PATTANAIK S. N., TUMBLIN J., YEE H., GREENBERG D. P.: Time-dependent Visual Adaptation for Fast Realistic Image Display. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques* (2000), SIGGRAPH '00, ACM, pp. 47–54. 8
- [PZB16] POHL D., ZHANG X., BULLING A.: Combining eye tracking with optimizations for lens astigmatism in modern wide-angle hmds. In *2016 IEEE Virtual Reality (VR)* (March 2016), pp. 269–270. doi:10.1109/VR.2016.7504757. 20
- [QM08] QU L., MEYER G. W.: Perceptually Guided Polygon Reduction. In *IEEE Transactions on Visualization and Computer Graphics (TVCG)* (2008), vol. 14, pp. 1015–1029. 14, 15
- [RCHR07] RAMIC B., CHALMERS A., HASIC J., RIZVIC S.: Selective Rendering in a Multi-modal Environment: Scent and Graphics. In *Proceedings of the 23rd Spring Conference on Computer Graphics* (2007), SCCG '07, ACM, pp. 147–151. 18
- [RE12] RITSCHEL T., EISEMANN E.: A Computational Model of Afterimages. *Computer Graphics Forum* 31, 2 (2012), pp. 529–534. Eurographics '12. 8

- [Red97] REDDY M.: *Perceptually Modulated Level of Detail for Virtual Environments*. PhD thesis, 1997. 14, 18
- [Red01] REDDY M.: Perceptually Optimized 3D Graphics. *IEEE Computer Graphics and Applications* 21, 5 (Sep 2001), pp. 68–75. 18
- [Ree15] REED N.: *Gameworks VR*. Technical slides, NVIDIA, 2015. https://developer.nvidia.com/sites/default/files/akamai/gameworks/vr/GameWorks_VR_2015_Final_handouts.pdf. 17
- [RFWB07] RAMANARAYANAN G., FERWERDA J., WALTER B., BALAK.: Visual Equivalence: Towards a New Standard for Image Fidelity. *ACM Transaction on Graphics (TOG), SIGGRAPH '07* 26, 3 (2007), pp. 76:1–76:12. 10, 15, 23
- [RJG*14] RINGER R. V., JOHNSON A. P., GASPAR J. G., NEIDER M. B., CROWELL J., KRAMER A. F., LOSCHKY L. C.: Creating a New Dynamic Measure of the Useful Field of View Using Gaze-Contingent Displays. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (2014), ETRA '14, ACM, pp. 59–66. 18
- [RMGB01] ROSS J., MORRONE M., GOLDBERG M. E., BURR D. C.: Changes in visual perception at the time of saccades. *Trends in Neurosciences* 24, 2 (2001), pp. 113–121. 5
- [RPG99] RAMASUBRAMANIAN M., PATTANAIK S. N., GREENBERG D. P.: A perceptually based physical error metric for realistic image synthesis. In *Proceedings of the 26th annual conference on Computer graphics and Interactive Techniques* (1999), SIGGRAPH '99, ACM, pp. 73–82. 10
- [RR88] ROVAMO J., RANINEN A.: Critical flicker frequency as a function of stimulus area and luminance at various eccentricities in human cone vision: a revision of Granit-Harper and Ferry-Porter laws. *Vision Research* 28, 7 (1988), pp. 785–790. 9
- [RR07] R. ROSENHOLTZ Y. LI L. N.: Measuring visual clutter. *Journal of Vision* 7, 2 (2007), pp. 11–22. 6
- [RVN78] ROVAMO J., VIRSU V., NÄSÄNEN R.: Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision. *Nature* 271, 5640 (Jan. 1978), pp. 54–56. 4, 7
- [RWH*16] ROTH T., WEIER M., HINKENJANN A., LI Y., SLUSALLEK P.: An Analysis of Eye-Tracking Data in Foveated Ray Tracing. In *ETVIS 2016: Second Workshop on Eye-Tracking and Visualization, Baltimore, USA* (2016). 20
- [RWP*10] REINHARD E., WARD G., PATTANAIK S., DEBEVEC P., HEIDRICH W., MYSZKOWSKI K.: *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. The Morgan Kaufmann Series in Computer Graphics. Morgan Kaufmann, San Francisco, 2010. 8
- [SB05] SHEIKH H. R., BOVIK A. C.: A visual information fidelity approach to video quality assessment. In *The First International Workshop on Video Processing and Quality Metrics for Consumer Electronics* (2005), pp. 23–25. 11
- [SC06] SUNDSTEDT V., CHALMERS A.: Evaluation of Perceptually-based Selective Rendering Techniques Using Eye-movements Analysis. In *Proceedings of the 22nd Spring Conference on Computer Graphics* (2006), SCCG '06, ACM, pp. 153–160. 13
- [Sch56] SCHADE O. H.: Optical and Photoelectric Analog of the Eye. *Journal of the Optical Society of America* 46, 9 (1956), pp. 721–739. 7
- [Sch01] SCHOLL B. J.: Objects and attention: the state of the art. *Cognition* 80, 1–2 (2001), pp. 1–46. Objects and Attention. 12
- [SCM15] SWAFFORD N. T., COSKER D., MITCHELL K.: Latency Aware Foveated Rendering in Unreal Engine 4. In *Proceedings of the 12th European Conference on Visual Media Production* (2015), CVMP '15, ACM, pp. 17:1–17:1. 10
- [SDL*05] SUNDSTEDT V., DEBATTISTA K., LONGHURST P., CHALMERS A., TROSCIANKO T.: Visual attention for efficient high-fidelity graphics. In *Proceedings of the 21st Spring Conference on Computer graphics* (2005), SCCG '05, ACM, pp. 169–175. 13, 16
- [SFWG04] STOKES W. A., FERWERDA J. A., WALTER B., GREENBERG D. P.: Perceptual illumination components: A new approach to efficient, high quality global illumination rendering. In *ACM Transactions on Graphics (TOG)* (2004), vol. 23, ACM, pp. 742–749. 11
- [SGE*15] STENGEL M., GROGORICK S., EISEMANN M., EISEMANN E., MAGNOR M.: An Affordable Solution for Binocular Eye Tracking and Calibration in Head-mounted Displays. In *Proceedings of the 23rd ACM international conference on Multimedia 2015* (2015), MM'15, pp. 15–24. 18, 21
- [SGEM16] STENGEL M., GROGORICK S., EISEMANN M., MAGNOR M.: Adaptive Image-Space Sampling for Gaze-Contingent Real-time Rendering. *Proceedings of the Eurographics Symposium on Rendering* 35, 4 (2016). EGSR '16. 19
- [Sha49] SHANNON C. E.: Communication in the presence of noise. *Proceedings of the Institute of Radio Engineers* 37, 1 (1949), pp. 10–21. 3
- [SH10] SHEN J., ITTI L.: Gender differences in visual attention during listening as measured by neuromorphic saliency: What women (and men) watch. *Journal of Vision* 10, 7 (2010), pp. 159–159. 12
- [SIGK*16] SWAFFORD N. T., IGLESIAS-GUITIAN J. A., KONIARIAS C., MOON B., COSKER D., MITCHELL K.: User, Metric, and Computational Evaluation of Foveated Rendering Methods. In *Proceedings of the ACM Symposium on Applied Perception* (2016), SAP '16, ACM, pp. 7–14. 10
- [SK07] SMITH E., KOSSLYN S.: *Cognitive Psychology: Mind and Brain*, 1 ed. Pearson/Prentice Hall, 2007. 6
- [SK13] SMITH E. E., KOSSLYN S. M.: *Cognitive Psychology: Mind and Brain*, Pearson new international ed. 2013. 6
- [SKHB11] SHIBATA T., KIM J., HOFFMAN D. M., BANKS M. S.: Visual discomfort with stereo displays: effects of viewing distance and direction of vergence-accommodation conflict. In *Proceeding of SPIE* (Feb. 2011), Woods A. J., Holliman N. S., Dodgson N. A., (Eds.), p. 78630P. 23
- [SLR10] SHEVTSOV M., LETAVIN M., RUKHLINSKIY A.: Low Cost Adaptive Anti-Aliasing for Real-Time Ray-Tracing. In *Proceedings of the 20th International Conference on Computer Graphics and Vision* (St. Petersburg, Russia, 2010), GraphiCon '10, pp. 45–49. 15
- [SMM00] SCOGGINS R. K., MOORHEAD R. J., MACHIRAJU R.: Enabling level-of-detail matching for exterior scene synthesis. In *Proceedings of the IEEE Conference on Visualization* (2000), IEEE VIS '00, pp. 171–178. 14
- [SRIR07] SANTINI F., REDNER G., IOVIN R., RUCCI M.: EyeRIS: a general-purpose system for eye-movement-contingent display control. *Behavior Research Methods* 39, 3 (2007), pp. 350–364. 18
- [SRJ11] STRASBURGER H., RENTSCHLER I., JÜTTNER M.: Peripheral vision and pattern recognition: a review. *Journal of Vision* 11, 5 (2011). 6, 7
- [SS01] SHIMOJO S., SHAMS L.: Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology* 11, 4 (2001), pp. 505–509. 6
- [SS03] SPENCE C., SQUIRE S.: Multisensory Integration: Maintaining the Perception of Synchrony. *Current Biology* 13, 13 (July 2003), pp. 519–521. 6
- [SS09] SCHWARZ M., STAMMINGER M.: On Predicting Visual Popping in Dynamic Scenes. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization* (2009), APGV '09, ACM, pp. 93–100. 14
- [STNE15] STOLL J., THRUN M., NUTHMANN A., EINHÄUSER W.: Overt attention in natural scenes: objects dominate features. *Vision Research* 107 (2015), pp. 36–48. 13
- [STT12] SNOWDEN R., THOMPSON P., TROSCIANKO T.: *Basic Vision: An Introduction To Visual Perception*, 2nd edition. ed. Oxford University Press, Oxford, 2012. 4

- [SU07] STIRK J. A., UNDERWOOD G.: Low-level visual saliency does not predict change detection in natural scenes. *Journal of Vision* 7, 10 (2007), 3, 13
- [Sut02] SUTCLIFFE A.: *Multimedia and Virtual Reality: Designing Usable Multisensory User Interfaces*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 2002. 6
- [SW14] SAUNDERS D. R., WOODS R. L.: Direct measurement of the system latency of gaze-contingent displays. *Behavior Research Methods* 46, 2 (2014), pp. 439–447. 18
- [SYM*12] SCHERZER D., YANG L., MATTAUSCH O., NEHAB D., SANDER P. V., WIMMER M., EISEMANN E.: Temporal Coherence Methods in Real-Time Rendering. *Computer Graphics Forum* 31, 8 (2012), pp. 2378–2408. Survey. 2
- [TE14] TRUKENBROD H. A., ENGBERT R.: ICAT: A computational model for the adaptive control of fixation durations. *Psychonomic bulletin & review* 21, 4 (2014), pp. 907–934. 13
- [TG80] TREISMAN A. M., GELADE G.: A feature-integration theory of attention. *Cognitive Psychology* 12, 1 (1980), pp. 97–136. 6
- [TG02] THEEUWES J., GODJIN R.: Irrelevant singletons capture attention: Evidence from inhibition of return. *Perception & Psychophysics* 64, 5 (2002), pp. 764–770. 13, 22
- [TGFTB01] THORPE S., GEGENFURTNER K., FABRE-THORPE M., BÜLTHOFF H.: Detection of animals in natural images using far peripheral vision. *European Journal of Neuroscience* 14, 5 (2001), pp. 869–876. 5
- [TGTT11] TO M., GILCHRIST I., TROSCIANKO T., TOLHURST D.: Discrimination of natural scenes in central and peripheral vision. *Vision Research* 51, 14 (2011), pp. 1686–1698. 5, 6
- [Tre88] TREISMAN A.: Features and objects: The fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology* 40, 2 (1988), pp. 201–237. 12
- [TRP*05] TOLHURST D. J., RIPAMONTI C., PÁARRAGA C. A., LOVELL P. G., TROSCIANKO T.: A Multiresolution Color Model for Visual Difference Prediction. In *Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization* (2005), APGV '05, ACM, pp. 135–138. 10
- [TW01] THOMAS L. C., WICKENS C. D.: Visual Displays and Cognitive Tunneling: Frames of Reference Effects on Spatial Judgments and Change Detection. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 45, 4 (oct 2001), pp. 336–340. 6
- [TW06] THOMAS L. C., WICKENS C. D.: Effects of battlefield display frames of reference on navigation tasks, spatial judgements, and change detection. *Ergonomics* 49, 12-13 (2006), pp. 1154–1173. 6
- [TZGM11] TÖLLNER T., ZEHETLEITNER M., GRAMANN K., MÜLLER H. J.: Stimulus saliency modulates pre-attentive processing speed in human visual cortex. *PLoS ONE* 6, 1 (Jan 2011), e16276. 12
- [VCD09] VASSALLO S., COOPER S. L., DOUGLAS J. M.: Visual scanning in the recognition of facial affect: Is there an observer sex difference? *Journal of Vision* 9, 3 (2009), pp. 1–11. 12
- [VCL*11] VANGORP P., CHAURASIA G., LAFFONT P.-Y., FLEMING R. W., DRETTAKIS G.: Perception of visual artifacts in image-based rendering of façades. In *Proceedings of the Twenty-second Eurographics Conference on Rendering* (Aire-la-Ville, Switzerland, Switzerland, 2011), EGSR '11, Eurographics Association, pp. 1241–1250. 10
- [VdBLV96] VAN DEN BRANDEN LAMBRECHT C. J., VERSCHURE O.: Perceptual quality measure using a spatio-temporal model of the human visual system. In *Electronic Imaging: Science & Technology* (1996), SPIE '96, International Society for Optics and Photonics, pp. 450–461. 11
- [VDC14] VIG E., DORR M., COX D.: Large-Scale Optimization of Hierarchical Features for Saliency Prediction in Natural Images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2014), CVPR '14, pp. 2798–2805. 12, 13, 14
- [VDMB12] VIG E., DORR M., MARTINETZ T., BARTH E.: Intrinsic dimensionality predicts the saliency of natural dynamic scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 34, 6 (2012), pp. 1080–1091. 14
- [VDSL14] VALENZISE G., DE SIMONE F., LAUGA P., DUFAUX F.: Performance evaluation of objective quality metrics for hdr image compression. In *SPIE Optical Engineering+ Applications, 92170C* (2014), International Society for Optics and Photonics. 7
- [VJ04] VIOLA P., JONES M.: Robust Real-Time Face Detection. *International Journal of Computer Vision* (2004), pp. 137–154. 13
- [Vla16] VLACHOS A.: Advanced VR Rendering Performance. In *Game Developer Conference 2016 - Slides* (2016), GDC '16. 19
- [VMGM15] VANGORP P., MYSZKOWSKI K., GRAF E. W., MANTIUK R. K.: A Model of Local Adaptation. *ACM Transactions on Graphics (TOG), SIGGRAPH Asia 2015* 34, 6 (2015), pp. 166:1–166:13. 8
- [VRWM78] VOLKMAN F. C., RIGGS L. A., WHITE K. D., MOORE R. K.: Contrast sensitivity during saccadic eye movements. *Vision Research* 18, 9 (1978), pp. 1193–1199. 5
- [VST*14] VAIDYANATHAN K., SALVI M., TOTH R., FOLEY T., AKENINE-MÖLLER T., NILSSON J., MUNKBERG J., HASSELGREN J., SUGIHARA M., CLARBERG P., ET AL.: Coarse Pixel Shading. In *Eurographics/ACM SIGGRAPH Symposium on High Performance Graphics* (2014), HPG '14, pp. 9–18. 17, 19
- [WA09] WICKENS C. D., ALEXANDER A. L.: Attentional tunneling and task management in synthetic vision displays. *The International Journal of Aviation Psychology* 19, 2 (2009), pp. 182–199. 6
- [Wal98] WALTER B.: *Density Estimation Techniques for Global Illumination*. PhD thesis, aug 1998. 16
- [Wan95] WANDELL B. A.: *Foundations of Vision*. Stanford University, 1995. 3, 4
- [Wan06] WANG Y.: *Survey of Objective Video Quality Measurements*. Tech. rep., Worcester Polytechnic Institute, EMC Corporation Hopkinton, MA 01748, USA, Jan 2006. 11
- [Wat93] WATSON A. B.: DCT quantization matrices visually optimized for individual images. In *Human Vision, Visual Processing, and Digital Display IV* (1993), vol. 1913 of *SPIE '93*, pp. 202–216. 9
- [WB97] WOLFE J. M., BENNETT S. C.: Preattentive Object Files: Shapeless Bundles of Basic Features. *Vision Research* 37, 1 (1997), pp. 25–43. 5, 12
- [WBSS04] WANG Z., BOVIK A. C., SHEIKH H. R., SIMONCELLI E. P.: Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 4 (2004), pp. 600–612. 11
- [WC97] WICKENS C. D., CARSWELL C. M.: *Information Processing*. John Wiley & Sons, Inc., 1997, pp. 130–149. 6
- [WDW99] WILLIAMS A. M., DAVIDS K., WILLIAMS J. G. P.: *Visual Perception & Action in Sport*. Taylor & Francis, 1999. 5, 12
- [Wey58] WEYMOUTH F. W.: Visual sensory units and the minimal angle of resolution. *American Journal of Ophthalmology* 46, 1 (1958), pp. 102–113. 7
- [WFA*05] WALTER B., FERNANDEZ S., ARBREE A., BALA K., DONIKIAN M., GREENBERG D. P.: Lightcuts: A Scalable Approach to Illumination. In *ACM Transaction on Graphics (TOG)* (2005), SIGGRAPH '05, ACM, pp. 1098–1107. 15
- [WHS15] WEIER M., HINKENJANN A., SLUSALLEK P.: A Unified Triangle/Voxel Structure for GPUs and its Applications. *Journal of WSCG (Cumulative issue)* 24, No. 1-2 (2015), pp. 83–90. 18
- [WK06] WALTHER D., KOCH C.: Modeling attention to salient proto-objects. *Neural Networks* 19, 9 (2006), pp. 1395–1407. 13
- [WLC*03] WILLIAMS N., LUEBKE D., COHEN J. D., KELLEY M., SCHUBERT B.: Perceptually Guided Simplification of Lit, Textured Meshes. In *Proceedings of the 2003 Symposium on Interactive 3D Graphics* (2003), I3D '03, ACM, pp. 113–121. 14

- [WM78] WESTHEIMER G., MCKEE S. P.: Steroscopic acuity for moving retinal images. *Journal of the Optical Society of America* 68, 4 (Apr. 1978), pp. 450–455. 7
- [WM04] WINDSHEIMER J. E., MEYER G. W.: Implementation of a visual difference metric using commodity graphics hardware. In *Electronic Imaging 2004* (2004), International Society for Optics and Photonics, pp. 150–161. 10
- [WM14] WANAT R., MANTIUK R. K.: Simulating and Compensating Changes in Appearance Between Day and Night Vision. In *ACM Transactions on Graphics (TOG)* (July 2014), vol. 33 of SIGGRAPH '14, ACM, pp. 147:1–147:12. 8
- [Wol94] WOLFE J. M.: Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review* 1, 2 (1994), pp. 202–238. 13
- [WP04] WEISENBERGER J. M., POLING G. L.: Multisensory roughness perception of virtual surfaces: effects of correlated cues. In *Proceedings 12th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems* (2004), HAPTICS '04, IEEE, pp. 161–168. 6
- [WPG02] WALTER B., PATTANAIK S. N., GREENBERG D. P.: Using perceptual texture masking for efficient image synthesis. *Computer Graphics Forum* 21, 3 (2002). 10
- [WRK*16] WEIER M., ROTH T., KRUIJFF E., HINKENJANN A., PÉRARD-GAYOT A., SLUSALLEK P., LI Y.: Foveated Real-Time Ray Tracing for Head-Mounted Displays. In *Computer Graphics Forum* (2016), Pacific Graphics '16. 20
- [WS13] WADE N. J., SWANSTON M.: *Visual Perception: An Introduction*. Psychology Press, 2013. 8
- [WSB03] WANG Z., SIMONCELLI E. P., BOVIK A. C.: Multiscale structural similarity for image quality assessment. In *Conference Record of the 37th Asilomar Conference on Signals, Systems and Computers* (2003), vol. 2, IEEE, pp. 1398–1402. 11
- [WSZL13] WU J., SHEN X., ZHU W., LIU L.: Mesh saliency with global rarity. *Graphical Models* 75, 5 (2013), pp. 255–264. 15
- [WWH04] WATSON B., WALKER N., HODGES L. F.: Supra-threshold Control of Peripheral LOD. *ACM Transactions on Graphics (TOG)* 23, 3 (2004), 750–759. 18
- [WWHW97] WATSON B., WALKER N., HODGES L. F., WORDEN A.: Managing Level of Detail Through Peripheral Degradation: Effects on Search Performance with a Head-mounted Display. *ACM Transactions on Computer-Human Interaction* 4, 4 (1997), 323–346. 19
- [XB16] XIAO R., BENKO H.: Augmenting the Field-of-View of Head-Mounted Displays with Sparse Peripheral Displays. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), ACM, pp. 1221–1232. 18
- [Yan95] YANTIS S.: Perceived Continuity of Occluded Visual Objects. *Psychological Science* 6, 3 (1995), pp. 182–186. 6
- [YCK*09] YU I., COX A., KIM M. H., RITSCHER T., GROSCH T., DACHSBACHER C., KAUTZ J.: Perceptual influence of approximate visibility in indirect illumination. *ACM Transactions on Applied Perception (TAP)* 6, 4 (Oct. 2009), pp. 24:1–24:14. 16
- [Yel83] YELLOTT J.: Spectral consequences of photoreceptor sampling in the rhesus retina. *Science* 221, 4608 (1983), pp. 382–385. 4
- [YLW*16] YANG B., LI F. W., WANG X., XU M., LIANG X., JIANG Z., JIANG Y.: Visual saliency guided textured model simplification. *The Visual Computer* 32, 11 (2016), pp. 1415–1432. 15, 16
- [YPG01] YEE H., PATTANAIK S., GREENBERG D. P.: Spatiotemporal Sensitivity and Visual Attention for Efficient Rendering of Dynamic Environments. *ACM Transactions on Graphics (TOG)* 20, 1 (2001), pp. 39–65. 17
- [YZWH12] YANG X. S., ZHANG L., WONG T.-T., HENG P.-A.: Binocular Tone Mapping. *ACM Transactions on Graphics (TOG), SIGGRAPH '12* 31, 4 (2012), pp. 93:1–93:10. 7
- [ZC01] ZAVAGNO D., CAPUTO G.: The glare effect and the perception of luminosity. *Perception* 30, 2 (2001), pp. 209–222. 8
- [ZDL02] ZENG W., DALY S., LEI S.: An overview of the visual optimization tools in JPEG 2000. *Signal Processing: Image Communication* 17, 1 (Jan. 2002), pp. 85–104. 9
- [ZK13] ZHAO Q., KOCH C.: Learning saliency-based visual attention: A review. *Signal Processing: Special issue on Machine Learning in Intelligent Image Processing* 93, 6 (2013), pp. 1401–1407. 12
- [ZK15] ZAGORUYKO S., KOMODAKIS N.: Learning to compare image patches via convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 4353–4361. 11
- [ZWF16] ZUO L., WANG H., FU J.: Screen content image quality assessment via convolutional neural network. In *Image Processing (ICIP), 2016 IEEE International Conference on* (2016), IEEE, pp. 2082–2086. 11



Martin Weier got his MSc in Computer Science from the Bonn-Rhein-Sieg University (BRSU), Sankt Augustin, Germany in 2011. Currently, he is a research associate at the BRSU's Institute of Visual Computing in Sankt Augustin and a PhD student at Saarland University, Germany. His research interests include perception and gaze-contingent rendering, large model visualization, compiler technologies and virtual and augmented reality systems.



Thorsten Roth got his MSc in Computer Science from the Bonn-Rhein-Sieg University (BRSU), Sankt Augustin, Germany in 2011. He is a research associate at the BRSU's Institute of Visual Computing and a PhD student at Brunel University London. His research interests include global illumination, screen-space filtering techniques and perception-based rendering.



Elmar Eisemann, professor at Delft University of Technology, is heading the Computer Graphics and Visualization group. He was associate professor at Telecom ParisTech and senior researcher in the Cluster of Excellence at MPII/Saarland University. He co-authored the book "Real-time Shadows", was paper chair for HPG 2015, EGSR 2016, GI2017, and will be general chair of Eurographics 2018 in Delft. In 2011, he was honored with the Eurographics Young Researcher Award.



Steve Grogorick is a research associate and PhD student at the Computer Graphics Lab at Technische Universität Braunschweig, Germany. He holds a Master of Science in Computer Science (2015) from the TU Braunschweig. His research interests include virtual reality, gaze tracking and gaze-contingent rendering, real-time rendering and visual perception.



Ernst Kruijff is interim professor for computer graphics and interactive systems at the Bonn-Rhein-Sieg University (BRSU), and heads the 3DMI group at the Institute of Visual Computing. He holds a MA from Utrecht University and a PhD from TU Graz. His research interests include 3D user interfaces, human factors, and multisensory feedback systems. He co-authored the book "3D User Interfaces - Theory and Practice".



Karol Myszkowski Karol Myszkowski is a senior researcher at the MPI Informatik, Germany. He received his PhD (1991) and habilitation (2001) degrees in computer science from Warsaw University of Technology. In 2011 he was awarded with a lifetime professor title by the President of Poland. He co-authored the book "High Dynamic Range Imaging", and co-chaired EGSR 2001, ACM SAP 2008, SCCG 2008, and Graphicon 2012.



Michael Stengel is a postdoctoral researcher at TU Delft. He holds a Diploma degree (2011) in Computational Visualistics from the University of Magdeburg, Germany, and a PhD degree (2017) from the TU Braunschweig, Germany. From 2010 to 2011 he worked at the Virtual Reality Lab at Volkswagen AG. His research interests include virtual reality, real-time rendering, gaze tracking, visual perception, visualization, and natural human-computer interaction.



Piotr Didyk is an independent research group leader at the Excellence Cluster for "Multimodal Computing and Interaction" where he is a head of "Perception, Display, and Fabrication Group". Prior to this, he spent two years as a postdoctoral associate at Massachusetts Institute of Technology. In 2012, he obtained his PhD from the Max Planck Institute for Informatics and the Saarland University for his work on perceptual displays.



Martin Eisemann is a professor at the TH Köln, Germany. He holds a Diploma (2006) from the University of Koblenz-Landau, and a PhD (2011) from the TU Braunschweig. Between 2007 and 2014 he was a visiting researcher at several institutions including TU Delft, Saarland University, EDM Hasselt and the Max-Planck Institut für Informatik.



André Hinkenjann is the founding director of the Institute of Visual Computing and research professor for Computer Graphics and Interactive Systems at the Bonn-Rhein-Sieg University (BRSU). He holds a PhD and Diploma in Computer Science from TU Dortmund. He is also an Adjunct Professor at the Department of Computer Science at the University of New Brunswick in Fredericton, Canada.



Marcus Magnor heads the Computer Graphics Lab at Technische Universität Braunschweig, Germany. He holds degrees in physics (MSc), electrical engineering (PhD), and computer science (habil.). He was Fulbright scholar at the University of New Mexico, USA, where he is adjunct professor at the Department of Physics and Astronomy. He is laureate of the Wissenschaftspreis Niedersachsen 2012 and elected member of the Braunschweigische Wissenschaftliche Gesellschaft.



Philipp Slusallek is Scientific Director at the German Research Center for Artificial Intelligence (DFKI), Research Director at the Intel Visual Computing Institute and professor for Computer Graphics at Saarland University, and PI at the German Excellence Cluster for "Multimodal Computing and Interaction". He was a Visiting Assistant Professor at Stanford University and is a Fellow of the Eurographics Association. He holds a PhD in Computer Science and a Master in Physics.