Temporal Video Filtering and Exposure Control for Perceptual Motion Blur

Michael Stengel, *Member, IEEE,* Pablo Bauszat, Martin Eisemann, Elmar Eisemann, and Marcus Magnor

Abstract—We propose the computation of a perceptual motion blur in videos. Our technique takes the predicted eye motion into account when watching the video. Compared to traditional motion blur recorded by a video camera our approach results in a perceptual blur that is closer to reality. This postprocess can also be used to simulate different shutter effects or for other artistic purposes. It handles real and artificial video input, is easy to compute and has a low additional cost for rendered content. We illustrate its advantages in a user study using eye tracking.

Index Terms—high frame rate, temporal filtering, perception, sharpening and blur

1 Introduction

When observing a scene, we tend to track features to resolve details. In contrast, this tracking leads to a perceived blur of the non-tracked background. A similar observation holds for cameras which exhibit motion blur based on exposure time. When watching a video, the recorded motion blur does not necessarily reflect the expected perceived blur. Especially on a large display, additional eye movement influences perception and can result in visible artifacts such as ghosting, judder, edge banding and a significant loss of detail.

Composition, motion, aperture, focus, gain, and exposure time are well-known parameters to artistically influence video recordings [1]. Especially exposure time is an important element as it is inherently related to frame rate. Short exposures lead to discontinuous motion (strobing effects), while a longer exposure creates motion blur, resulting in detail loss [2]. Blur, also in the context of depth of field, can be of significant importance for artistic purposes, for example, to attract attention [1, p. 299], [3, p. 51], to increase motion perception and liveliness [2, p. 129], or to serve in story telling [3, p. 62].

It must be realized that perceived blur in the real world will always differ from camera-recorded blur. One of the major reasons is that we as humans tend to track the interesting elements in the scene whereas a video camera does not necessarily follow the same object. Consequently, eye motion and camera motion differ, and so does the corresponding motion blur. For example, vertical eye movement induces vertical motion blur, but if the camera was panning left, horizontal motion blur was recorded. Especially for larger screens and low frame rates, this mismatch can result in visible artifacts.

Additionally, hold-type blur might occur due to a mismatch between the continuous eye movement when tracking an object on the screen and discontinuous movement of the target due to limited frame rate. The latter can be very confusing as the human visual system (HVS) expects tracked objects to appear sharper than non-tracked objects. For sharp images recorded at the traditionally used low frame rate (LFR) for movies being 24 Hz, this condition is not fulfilled due to the hold-type blur.

Current high frame rate (HFR) videos, with a typical frame rate of 48 Hz to 60 Hz, reduce recorded motion blur and hold-type blur, leading to sharper perceived images. For this reason, they have become popular in the consumer market; specialized upsampling techniques are integrated into standard TV equipment, and high frame rate movies are being explored by movie directors (e.g., *The Hobbit*). Nevertheless, the consequences are not always beneficial. An HFR video must be recorded at lower exposure times and, because there has to be a minimal time to store a frame (or to open the shutter), usually only 60% of the overall time is captured on film [3]. Temporal replicates caused by the high sampling rate and perceivable as shifted ghost images may appear. Some viewers even reported perceiving a distracting speedup of the video [4]. For this reason, recent HFR television shows such as *Video Game High School* added hand-tuned blur to some scenes, thereby removing many of the details. Such solutions are rather ad-hoc and not always successful.

Whether considering motion or hand-created blur, the blur does not lead to the expected perceptual blur induced by eye movement. Even a frame rate

- *M. Stengel, P. Bauszat, M.Eisemann and M. Magnor are with the Computer Graphics Lab, TU Braunschweig, Germany.*
- *E. Eisemann and M. Eisemann are with the Computer Graphics and Visualization Lab, TU Delft, Netherlands.*

of 120 Hz—far beyond the 48 Hz used in current HFR movies—is insufficient to allow for natural perception of blur, not to mention the high bandwidth requirements and lack of support by current displays [5]. To remove the camera induced motion blur, we would require an infinitely high frame rate. Hence, we believe the only practical solution is to include the respective eye motion into the blur model and create a displayable lower frame rate video from an ultra-high frame rate video (UHFR) including the expected eye motion. We use the term UHFR to depict videos with a frame rate higher than 1000 Hz.

In this paper, to solve these problems, we show how to adapt exposure and motion blur in a postprocess taking eye motion into account. We propose a model to explain the perception of a scene with a standard video camera and the HVS (Section 3) and show that the camera itself is a insufficient approximation of human perception when content is tracked in the image plane by the observer (Section 4). We then derive a filtering technique that takes predicted eye motion into account, which leads to a more faithful image reconstruction on the retina (Section 5). Our solution can be used to artistically manipulate frame rate and exposure time in a postprocess, which goes beyond the possibilities of a standard camera.

Specifically, we present the following contributions:

- a model for perceived motion blur
- an estimation of eye movement and corresponding filtering process for a more faithful retinal image

Our approach results in a variety of benefits and applications, including downsampling for real world and CG generated UHFR videos, virtual shutter simulation, motion stills generation, subtle gaze direction. Our technique is applicable to high-speed footage as well as traditional LFR camera output (24–30 Hz) or synthetic content and leads to improvements in perceived video quality (Section 6). For rendered scenes, we also show that our solution does not necessarily require higher computation times. The benefit of our approach for subtle gaze direction are illustrated in a user study.

2 Related Work

Shutter and exposure are usually controlled during the capture process. In contrast, our work modifies these parameters in a postprocess. Our contributions can be seen as a generalization and extension of synthetic shutter speed [6], which imitates a long exposure shot by taking a series of short exposure photographs and aligning them, reducing noise and camera shake while preserving motion blur. Our filtering also differs from traditional temporal downsampling of HFR videos [7] by taking the viewer into account.

We also draw inspiration from rendering techniques to simulate shutter effects [8]. Previously, the desired shutter type had to be chosen beforehand, so that any change implied a costly reshooting of the scene. We offer a postprocessing solution and make it possible to test different shutter types to define the wanted final appearance. A rolling shutter allows the definition of exposure time based on an opening of a rotating disc. By default, an open arc of $180°$ has been established for 24 Hz shots [2]. Shorter shutters create stuttered motion, which can be used for artistic purposes (e.g. $45°$, Saving Private Ryan, Gladiator, Three Kings). A longer exposure ($210°$), especially, in combination with low frame rates (6–12 Hz), creates dramatic blur effects.

Blur can also help guiding the observer's gaze. Our foveal vision contains a high density of cones and leads to high acuity compared to peripheral vision [9]. The latter is still sensitive to subtle temporal changes, which can attract the viewer's attention [10], [11], which is another motivation to avoid temporal artifacts. The HVS is attracted mostly to salient regions. Special blur and sharpening filters [12], depth-of-field effects [13], or observer-driven simplification [14] are good means to accentuate or subdue saliency and to guide gaze. Our solution also allows us to add such indications.

Temporal processing influences our blur perception; strongly blurred (>10 arcmin) moving patterns appear sharper than their static counterpart, yet for a small blur (<10 arcmin) stationary edges seem sharper than moving ones [15]. This observation relates to motion sharpening. It is not a mechanism that removes blur, but results from the HVS's inability to discriminate whether or not the moving object is indeed sharp [16]. This theory has been strengthened by the fact that observers tend to match blurred peripheral stimuli with sharper foveal stimuli [17]. Temporal and spatial coherence as well as motion contrast are important factors for the HVS as well as in video saliency [18], [19]. We, thus, propose to blur the video according to the predicted/intended eye motion instead of the camera motion.

Although eye-movements are not known a priori, [20] report that in natural movies up to 80% of the subjects look at the same image region. Especially in Hollywood movies, the coherence was very high due to camera work and scene cuts. Most likely, target regions are of high saliency [21], [22]. As our approach reduces saliency in the areas outside the object of interest, it can be seen as an extension to those classic movie techniques to subtly draw attention to specific regions.

In movie productions, the widest rolling shutter is usually still limited to $210°$ due to a minimum sensor read-out time. Consequently, each recording shows gaps that can result in motion artifacts such as judder (unsmooth motion) or edge banding being perceived

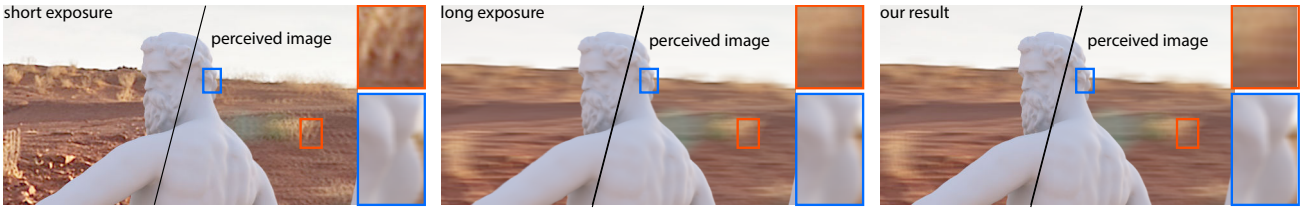

Fig. 1: **Exposure comparison.** Ghosting artifacts can usually be perceived when exposure time or frame rate are insufficient (left). Longer exposure times (middle) avoid ghosting, but details in the scene suffer due to motion blur. Further, this blur does not match the expected motion blur of an observer watching the scene. Our solution (right), is a temporal downsampling method taking eye motion into account and leads to sharp tracked objects and a consistent motion blur in the rest of the scene. Our method also makes it possible to subtly guide gaze or for artistic purposes.

as overlapping edge replicas at the border of a moving object.

It is possible to analyze the required sampling rate and maximal motion between two images to prevent these replicas [23]. Alternatively, if supported by the hardware using a multi-flash protocol that results in showing a video frame multiple times reduces artifact visibility at a given capture rate. [24]. Another option is to employ an appropriate bandlimiting filter [25]. We build upon this observation when constructing an initial ultra-high frame rate video, free from temporal artifacts.

3 Image Formation Model

Here, we describe a model to predict how a scene and a video are perceived by the human visual system (HVS). Even though many directors consider the video camera as the observing eye, watching the video afterwards does not create a perfect illusion of viewing the recorded scene as in the real world.

For ease of explanation, let us define the irradiance that is recorded on the sensor plane of a video camera is a function $\mathbf{S}(x,t)$, where x is sensor position and t is time. Ideally, the recorded video $\mathbf{I}$ would be equal to $\mathbf{S}$, but it is only a discretized version. We focus on temporal discretization and assume resolution to be sufficient, which in natural viewing conditions is often the case for full-HD content. A frame $\mathbf{I}_i$ is described as

$$\mathbf{I}_i(x) := \int_{t_i}^{t_i+T_V} \mathbf{S}(x,t)dt \quad , \tag{1}$$

where the camera shutter opens at time t_i and closes again at time t_i+T_V. A shorter open shutter would reduce the intensity, but we can assume that gain is used to counterbalance the frame duration. Based on the common usage of a $180°$ shutter in traditional film-making the exposure time T_V is chosen in accordance with the simple equation $T_V = 1/(2 * \text{frame rate})$.

Analogously, we want to define the retinal image $\mathbf{R}$ via the intensities perceived by the retina. We define the $\mathbf{R}$ for a retinal location x as

$$\mathbf{R}_i(x) := \int_{t_i}^{t_i+T_R} \mathbf{S}(x+p(t),t)dt \quad , \tag{2}$$

where $p(t)$ describes eye's path due to tracking and T_R is a small period of time over which the information is integrated by the HVS. T_R is refered to as the *critical duration* or *critical period*. The critical duration is an empirically estimated value which describes for how long a receptor of the retina counts incoming photons before an electrical stimulus is triggered for higher level processing in the retinal ganglion cells. The critical duration depends on the incoming light intensity. This dependency is described by Bloch's Law of Vision [26] . Bloch's Law can be expressed by the simple equation $R = L \cdot T_R$ and states that the product of luminance L and stimulus duration T_R is a constant R as long as the maximum critical duration of 10–15 ms for cones and 100 ms for rods is not reached. [27]. Since for photopic vision rods are fully saturated the perceived information only depends on the output of cones. Considering these aspects in Eq. (2) we assume the critical duration T_R to be 15 ms, which is the longest temporal summation time for cones. p is closely linked to feature tracking—saccades and tremors can be neglected because *smooth pursuit eye motion* allows us to almost perfectly track targets up to object speeds of about 10 deg/s. Higher speeds may lead to significant interindividual differences in perception [28]. The combination of smooth pursuit eye motion and the integration time of the HVS also explains hold-type blur [29], which results from a mismatch between (discontinuous) object motion on the screen and (continuous) eye tracking. It is particularly pronounced for low frame rates. For more details on capture and display of movies in signal processing terms we refer the interested reader to [30]

Photoreceptors do not move independently but are fixed on the rigid retina [31]. Hence, we can safely assume that p is the same for all locations on the retina. While this is not exactly true (the induced error depends on eye shape, viewing conditions, and camera settings), these deviations are negligible for our purposes.

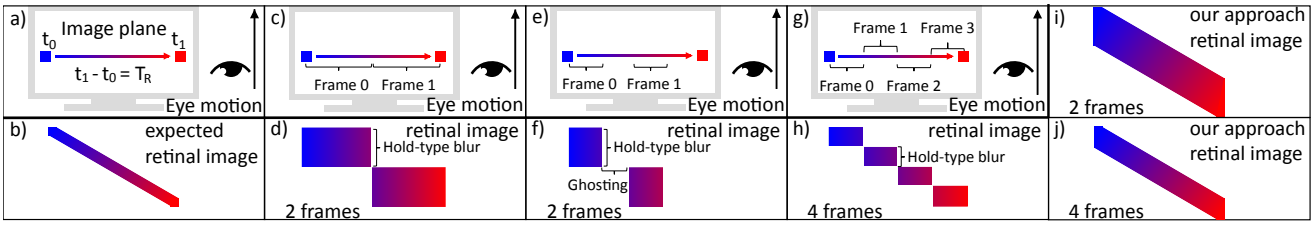

Fig. 2: **Temporal artifacts.** (a) A small object moves horizontally, the eye upwards. (b) Expected perceived image. (c) Long-exposure recordings with a static camera transform the point in a horizontal streak. (d) However, due to eye integration, an observer perceives two rectangles, which are only loosely connected, although they come from the same object. (e) Shorter exposure times lead to temporal artifacts; separate components are perceived and temporal information is lost. (f) Here, two disconnected rectangles appear on the retina. (g) High frame rate videos exhibit more frames and less motion blur, which can reduce the problem. (h) By increasing the frame rate, one can approximate the expected retinal image, but the needed frame rates are not possible to capture without time gaps, and displaying them is challenging. Our approach makes use of the eye motion and results in a closer approximation to the expected retinal image (b) for (i) low as well as (j) high frame rates.

4 Blur Mismatch of Camera and Eye

A typical low frame rate video camera is an imperfect substitute for the human eye when an object of interest (OOI) moves in the image plane. The reason is that the additional eye movement while watching the video should have been taken into account during the recording. Let the duration of the video be equal to the integration time T_R of the HVS which is reasonable for 30–60 Hz videos [26]. Then, only in the absence of any eye movement, Eq. (2) is equivalent to Eq. (1), i.e., $p(t) = 0$.

A simple example is shown in Fig. 1. The camera motion is an off-axis rotation around the OOI resulting in both a rotation and translation of the Neptune statue in image-space. For a short exposure, the OOI is detailed, but the background exhibits temporal artifacts (left). These artifacts appear only on the retina of the observer; they reveal themselves as an unnaturally sharp background or even ghosting whenever the integration time of the eye crosses frame boundaries. For a long exposure, the OOI suffers from motion blur (middle).

To explain the temporal artifacts, we consider a simple scene (Fig. 2); a small object moves horizontally from left to right while the eye tracks an OOI that moves vertically in the image plane. The correct integration on the retina should result in a diagonal line (Fig. 2b). If the camera is static and captures at a frame rate equal to $1/2\ T_R$ the object is smeared along a horizontal line due to the motion blur of the object (Fig. 2c). When watching the video, however, the eye tracks the upward moving object. This eye movement results in hold-type blur. At each location of the retina pixels that are crossed are integrated, resulting in two square-shaped features on the retina, one for each frame (Fig. 2d).

To reduce motion blur for important objects, a shorter exposure time can be used, but then some temporal information is lost (Fig. 2e). This leads to perceivable ghosting and features seem to *jump* (Fig. 2f). Only when increasing the frame rate (Fig. 2h), hold-type blur is reduced linearly and, in the limit, converges to zero. Nonetheless, it is difficult to record videos at such high frame rates without gaps. Also, displaying the content is challenging due to the necessary high bandwidth. Our temporal downsampling, explained in the next section, simulates the perceptual blur of the HVS and delivers a more faithful image on the retina (Fig. 2i,Fig. 2j).

5 Gaze-guided Downsampling

Temporal edge banding or ghosting artifacts in videos can only appear if the motion between two frames exceeds one pixel, a result that can be derived from similar findings for light field rendering [23]. Hence, although our goal is consistent filtering and temporal downsampling, if LFR or HFR video footage is given as input we first transform a given video sequence into an ultra-high frame rate (UHFR) video $\mathbf{I}_{\text{UHFR}}$ indicating a frame rate of 1000+ Hz. This footage is computed by using an interpolation algorithm based on image similarity [32]. This upsampling process is usually robust, but can fail for blurry edges. Fortunately, blurry edges are not very problematic for the eye integration process [33] and are unlikely to exhibit temporal edge banding. Nonetheless, to achieve sharp images, the original exposure time should be low.

If the eye motion p is known or user-defined, the retinal image $\mathbf{R}$ can be computed for any point in time and any desired integration time T_R by integrating $\mathbf{I}_{\text{UHFR}}$ along p. More intuitively, this is equal to translating each frame in the opposite direction of p. Basically this compensates the eye motion and sets the spatial components of p to 0. Next, integrating along the temporal axis and translating the frames back to their original position results in the filtered video frames.

For a perceptionally plausible reconstruction when temporally downsampling $\mathbf{I}_{\text{UHFR}}$, we could proceed as for apparent resolution enhancement [33], [34], i.e., compute the retinal image and optimize the output video frames such that the integration in the eye best matches the target image. Nevertheless, because our output is not necessarily an UHFR sequence, exploiting eye integration is difficult and we cannot create frequencies that would cancel out any hold-type blur.

Instead, we opt for a frame $\mathbf{R}_i$ of the output video sequence via

$$\mathbf{R}_i(x) := \int_{t_i}^{t_i+T} \mathbf{I}_{\text{UHFR}}(x + p(t), t)dt \qquad (3)$$

where T is the desired output frame duration (inversely related to the frame rate). Hereby, we produce an image similar to the expected retinal image. As p is defined via the OOI, this part of the output stays sharp. Further, when assuming hold-type blur, our result will stay consistent; non-tracked objects will consistently be blurrier than the OOI. Our definition also leads to robustness with respect to eye tracking because an eye deviation from the intended path will not introduce additional high frequencies, as could be the case in [33]. In Section 6, we show that these properties help us avoid temporal artifacts and improve perceived quality. Further, the effect is successful in guiding the gaze of observers.

Saliency-based Temporal Integration Eq. (3) requires the eye's motion path p to be known, which is the case for rendered content, or if it was intentionally created and imposed for artistic purposes. In all other cases, we need a robust solution to estimate p.

We first compute the saliency $\mathbf{A}$ for each frame of $\mathbf{I}_{\text{UHFR}}$ [35], with $\mathbf{A}$ being normalized to the range $[0, 1]$, which gives us a direct measure of how probable it is for a viewer to track a feature in $\mathbf{I}_{\text{UHFR}}$. We then assume $p(x, t)$ is defined by the optical flow $\mathbf{F}_{i\to i+1}(x)$ [32], [36] for any two frames i and $i+1$ in $\mathbf{I}_{\text{UHFR}}$, which proved successful in related work as well [34], [37]. Nonetheless, it is not a robust solution to simply select the path $p(x)$ of the most salient x to define the global p. Instead, we first mark a set of pixels describing the OOI. The mask is derived from the likelihood of each pixel to belong to the OOI. Precisely, we look for high saliency and similar pixel motion [35]. For frame i, we define the mask $\mathbf{O}_i$ as:

$$\mathbf{O}_i(x) := \begin{cases} 1, & \text{if } ||\mathbf{F}_{i\to i+1}(x) - \frac{\sum_x \mathbf{A}_i(x)\mathbf{F}_{i\to i+1}(x)}{\sum_x \mathbf{A}_i(x)}|| < \tau \\ 0, & \text{else} \end{cases}$$

In practice, a value of $\tau = 7$ worked well and was used for all examples. We then set $p(t_i)$ to the movement of the center of masses from $\mathbf{O}_i$ to $\mathbf{O}_{i+1}$. Additionally, we also allow for manual restriction of the OOI to regions defined by an optional binary mask $\mathbf{M}$ that is transferred from one frame to the next via rotoscoping [38]. Hereby, we can disambiguate multiple OOIs when needed, which is also useful for artistic purposes and allows us to choose and guide gaze direction (Section 6.6).

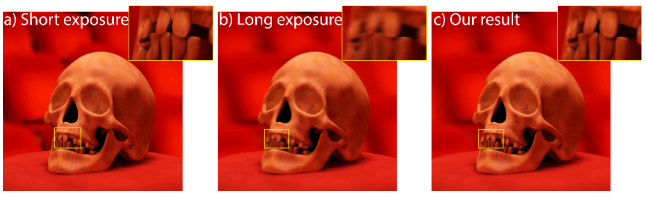

Fig. 3: **Synthetic Ultra-high Frame Rate Video:** (a) short exposure, (b) long exposure, (c) our result using a rendered video (60 Hz).

Fig. 4: **Real-World Ultra-high Frame Rate Videos:** short exposure (left column), long exposure (middle column), our result (60 Hz, right column) for a ultra-high frame rate input (3000 Hz).

Although, we only consider translational motion for the eye integration, this choice is not very restricting. The translations are applied to an $\mathbf{I}_{\text{UHFR}}$ sequence. Hence, each frame exhibits minimal motion. Furthermore, the sequence itself was constructed via an upsampling technique that assumes general motion. Although it is true that different parts of the OOI can undergo different motion, ultimately, our eyes can only follow one path, which for a single OOI is usually well detected by our method [20].

6 Applications

Here, we present results and applications of our method. Precisely, we compare to video sequences captured with *short exposure* times (pinpoint-sharp images with the typical 180° shutter, resulting in an exposure time of half the frame duration) and to *long exposure* shots where the shutter was kept open for as long as possible. We show synthetic as well as real-world sequences captured with traditional low frame rate or high-speed cameras. We refer to the supplemental material for the videos.

6.1 Ultra-high Frame Rate Videos

First, we illustrate our downsampling for two synthetic "hero" shots (Fig. 1&3): the camera moves around the object of interest in an ellipse creating

opposing foreground and background motion. Short exposure leads to temporal artifacts in the background, which results in ghosting artifacts on the retina when tracking the foreground. A long exposure shot removes these artifacts but blurs the foreground. Our result leads to a sharp foreground, while avoiding background ghosting. We explicitly used a static camera to show the difference between a short/long exposure shot and our approach. The short exposure shot keeps both fore- and background in focus which results in unnaturally sharp images. The long exposure, on the other hand, removes most details from the foreground object. Our solution filters the background slightly to counteract the hold-type and put emphasis on the main elements.

6.2 Stochastic Ultra-high Frame Rate Videos

For CG-generated movies, we can modify the upsampling process and derive an UHFR sequence more easily by relying on existing temporal coherence methods [39]. In particular, given a physically-based renderer, which have become common in production rendering [40], [41], we can create a low quality UHFR video as input to our algorithm (Fig. 5b). We render each frame of the low quality video with just a fraction of the samples required for a high quality solution, thus not changing the overall number of samples required for rendering the LFR. Our proposed filter kernel gathers samples over multiple frames of the UHFR video resulting in a high quality LFR video with the desired motion blur (Fig. 5c). The validation of this step relates to distributed ray tracing [42]. It also implies that for physically impossible exposure times exceeding the duration of a frame, compute time decreases using our filter (Fig. 5d). This behavior is different from most Monte Carlo-based motion-blur rendering techniques, where stronger motion blur tends to increase render times [43].

6.3 Low Frame Rate Real-World Videos

Low frame rate videos are first upsampled to UHFR relying on a standard temporal upsampling technique (Sec. 5), then the downsampling is applied (Fig. 6). If the target frame rate is equal to the original frame rate, our algorithm uses the original frames inside the OOI and the interpolated frames in the background. This strategy avoids artifacts in the OOI induced by potentially imperfect upsampling, e.g., upsampling may fail to produce faithful results in case of occlusions.

6.4 Virtual Shutter

Our approach is compatible with virtual shutter simulations. Rolling, focal plane shutters, or even artistic shutters can be obtained. On a per-pixel basis, we define the exposure interval of Eq. (3), resulting in a direct integration into our solution. In Fig. 7, a focal plane shutter was used to imply speed by producing a tilting effect. In this case, the per-pixel definition was given by shifting the time interval in each row (top to bottom).

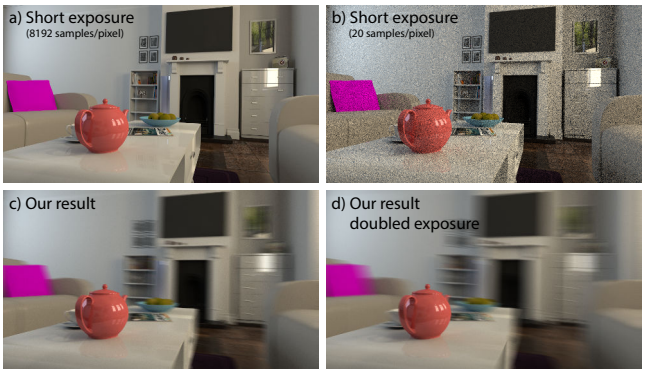

Fig. 5: **Stochastic Ultra-high Frame Rate Videos:** (a) High-quality short exposure (8192 samples per pixel). (b) Image from low-quality, high frame rate video (20 samples per pixel). (c) Applying our temporal downsampling to (b) leads to approximately similar quality as in (a). (d) Physically impossible exposure (twice the frame time) using only 10 samples per pixel as input (both 30Hz).

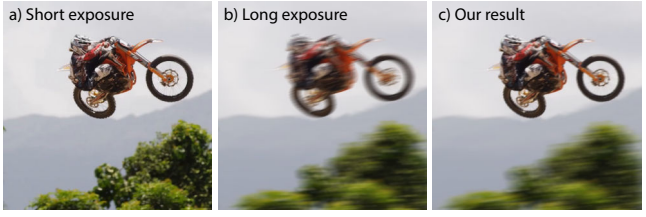

Fig. 6: **Low Frame Rate Video:** The 60 Hz video was upsampled to 3000 Hz, then downsampled to 60 Hz to simulate different exposure times. (a) Original. (b) Long exposure; entire image is blurred. (c) Our result; OOI is kept sharp while background is blurred.

6.5 Motion Stills

Images represent a snapshot in time, however, a single time slice or pinpoint sharp image does not convey any information about the motion in the scene. In contrast to short or long exposed traditional imagery our approach keeps the OOI in focus but still preserves this important motion information (Fig. 1–7). This is especially interesting for advertisements or movie descriptions in magazines where one wants to convey the dynamics in the scene.

6.6 Subtle Gaze Direction

To investigate the influence of our approach on the gaze behavior of an observer, we performed a user study. As stimuli, we chose identical balls moving at equal speed in different directions, Fig. 8a&b. We

intentionally chose a simple artificial scene to reduce the influence of as many higher-level perception mechanisms as possible. Each video was twelve seconds long, created via our downsampling method from a 3000 Hz input video to 60 Hz, but focusing on different balls as OOI (Fig. 8b). We created two sets of these videos one with a $1/60$ s and one with a $1/30$ s exposure time. For both, an additional long exposure shot was created as a reference (Fig. 8a) to validate whether our filter has a measurable influence on the eye-motion behavior. Additionally, we showed the stochastic rendering *Room* scene Fig. 5, once with a $1/30$ s traditional long exposure and once with our downsampled version using the same exposure time but focusing on the teapot. These eight videos were shown three times to each participant in randomized order. Each sequence was about one second long (76 frames, 60 frames per second).

14 participants, unaware of the goal of the experiment and with normal or corrected-to-normal vision took part in the experiment. We used a Samsung RZ2233 Full-HD screen in a darkened room to present video footage and an EyeLink 1000 eye tracker to record fixation times on the screen. The eye tracker records at 1000 Hz. The participants were seated in 60 cm distance to the screen. The participants did not receive any specific task except for watching the videos to prevent any task-specific influence of the results. The test took around ten minutes for each participant.

Fig. 8 shows snapshots and results of the experiment in the form of gaze heat maps. The top row shows the long exposure shot (left) and one version using our downsampling (right), where the focus was on the diagonally moving ball marked on the left. Below are heat maps describing the average gaze distribution for all participants, again for the long exposure shot and our approach with the $1/30$ s exposure settings. Our downsampling to $1/30$ s increases the fixation times for the intended objects of interest (Fig. 8d). A statistical evaluation (Fig. 10) revealed that the fixation time was roughly even among all balls in the long exposure video, with a bias towards horizontal motion or motion through the center of the screen. Further, subjects reported that following motion at the screen border is more demanding due to head fixation. For a simulated standard camera with $1/30$ s exposure, participants followed the object of interest (OOI) for 14% (standard deviation SD=7.2%) of the time, on average, in the test sequence with focus on the diagonally-moving ball (Ball A, increase for 11 out of 14 participants) near the left border and 26% (SD=9.0%) with focus on the horizontally-moving ball (Ball B, increase for 11 out of 14 participants). Using our approach the average percentage increased to 32% for Ball A (two-tailed t-test p=0.0015) and to 44% for Ball B (p=0.0013) which is a significant relative increase by 124% and 68%. These results

Fig. 7: **Shutter postprocessing** Our method allows to redefine shutter types after recording. Here, a focal plane shutter with different speeds is applied to a synthetic scene.

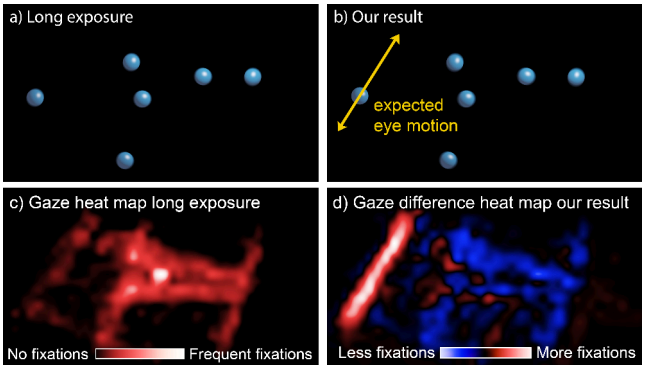

Fig. 8: **Subtle Gaze Direction:** (a) Image from the long-exposure sequence. (b) Image from our result. The expected eye motion induced by the OOI is shown as an arrow. (c) Gaze heat map of tracked gaze direction for the long-exposure sequence. (d) Gaze difference heat map for our result related to long-exposure sequence. The OOI (Ball A) exhibits an increased fixation time (hot areas). Other balls are fixated less (cool areas).

strongly indicate the ability of our approach to direct gaze. For an exposure time of $1/60$ s, the effect is more subtle, changing from 12% (SD=5.9%) to 16% (SD=6.2%) (Ball A, increase for 10 out of 14 participants, two-tailed t-test p=0.12) and 21% (SD=7.6%) to 29% (SD=10.8%) (Ball B, increase for 12 out of 14 participants, p=0.04) which is still a relative increase by 31% and 36%, respectively, despite the subtlety of the effect which is hardly noticeable even with a priori knowledge.

In the more realistic Room scene (Fig. 5) the camera rotates around a view-centerd object (fruit bowl) while we apply our filter to focus on an off-centered object (teapot). Since the camera rotates quickly around the center of the scene in this video, off-center objects appear strongly blurred in the long exposure video. However, the object in the center only suffers from blur caused by rotation and therefore remains sharper. In our study we showed two versions of the scene, one long exposure video and one video filtered by our method. In the filtered version the teapot was selected being the object of interest. We hypothesized that the central object (being a fruit bowl as visible in Fig. 5a) would mainly attract the attention of the viewer in

the long exposure version.

A statistical evaluation of the fixation time on the OOI is shown in Fig. 9 (significant differences marked by $*$). In the very beginning (frames 1 to 5) of both versions of the video the participants fixate on the center of the screen due to the earlier calibration step. In the following (frames 6-20) the teapot moves into the central part of the screen caused by rotation of the camera. Thus, it starts attracting attention. As the teapot moves off the center again from frame 25 on, fixation time on it decreases. In the long exposure video most of the participants focus on the sharper center object. The amount of time the subjects fixate the teapot in frames 21-26 and 27-66 decreases to 28% and 7%, respectively.

In the filtered version the fixation time in frames 21-26 and 27-66 increases significantly to 67% and 25% in total, which is a relative gain of 139% (M_{long}=27.9%, SD_{long}=12.1%, M_{our}=66.7%, SD_{our}=15.6%, two-tailed t-test p=0.0016) and 223% (M_{long}=7.8%, SD_{long}=5.8%, M_{our}=25.2%, SD_{our}=6.1%, p=0.001). Towards the end of the video (frames 67-76) the rotation of the camera slows down and finally stops. Since there is no motion blur without motion, all of the scene objects appear sharp. Our gaze analysis reveals that the participants change their focus to diverse objects in the environment of the scene. Accordingly, to fixation times of the long exposure video and the filtered version converge to the same level as there is no visual difference without motion. In total the overall fixation time on the OOI increased for 12 out of 14 participants.

It is likely that the temporal sensitivity of the human peripheral vision influences the participants' focus because high frequency video content tends to attract gaze [44]. The results of both user studies suggest that the effectiveness of gaze guidance using our filter increases in congruence with the exposure time, even beyond physically possible exposure times emphasizing the importance of being able to adjust exposure in a post-process.

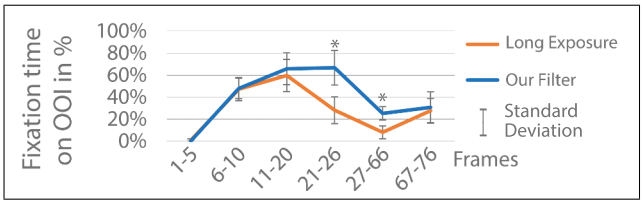

Fig. 9: **Quantitative evaluation of *Room* sequence:** In the respective frame ranges of the video the lines represent the average amount of time the user spent on the specified object of interest in the two version of the video (long exposure, lower line; perceptually filtered, upper line).

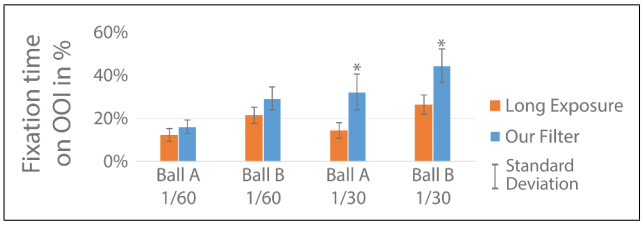

Fig. 10: **Quantitative evaluation of *Balls* sequence:** The time the user spent on the specified ball for the given exposure time is given in percents of the total video length. In the perceptually filtered video (right bars) the fixation time is increased compared to the long exposure video (left bars).

7 Discussion

As indicated by the results and user study, our work makes artistic postprocessing possible and can successfully influence observers' gaze. Humans rather focus on sharp and moving objects when watching videos. Hence, knowledge of a reasonable scanpath is not necessarily required, but can also be created with our technique. This is an important tool for movie production.

There is a trend towards large-screen home theater systems. Since the artifacts induced by traditional cameras are more obvious on larger screens, the difference of a long-exposure and a perceptually-filtered video becomes more pronounced, which renders our approach increasingly interesting.

The required UHFR videos have a non-negligible memory cost. In most cases creating a full UHFR video can be avoided by using a cache maintaining the necessary UHFR frames whose size depends on the desired integration time.

Our work does not yet address hold-type blur, which reduces higher frequencies in the direction of eye motion. For a lower frame rate, one would need to introduce higher frequencies into the images that would become visible as artifacts in the still images [33]. Instead, our solution opts for a consistent image, reducing any temporal edge banding artifacts, while keeping the OOI sharpest.

Our method assumes frame duration times below the integration time of the HVS, a video played at very slow frame rates may result in a discontinuous motion on the retina due to an insufficient eye integration in this case. A solution for this case is a problem on its own.

The accuracy of the automatic saliency metric works well for our scenarios but is not perfect. If it fails, in the worst case, attention would be drawn to different parts of the video. In addition, standard tools for matting and rotoscoping can always be used to correct or manually define saliency masks. Often these have already been created for other post-processing steps such as color grading or 2D to stereo conversion and should therefore be available in most production settings.

Assuming a single OOI should not be considered a strong limitation because this restriction holds simi-

larly for a standard camera. Further, in our approach, selection of multiple OOI computes an average motion for all OOIs creating a video that keeps them in focus as good as possible, at least as good as for the case of a traditional camera.

Another assumption inherent to the OOI is that it will not be occluded by another fast moving object. Potentially, these situations can lead to conflicts. In practice, these are typically the situations when tracking becomes more difficult for an observer and they tend to be more forgiving with respect to temporal artifacts, as their tracked signal is expected to be discontinuous.

Not following the intended scanpath has an effect, but it does not necessarily deteriorate the viewing experience. Our perceptual blur does not produce a blurrier overall image; objects moving in approximately the same direction as the OOI will appear sharper than with a standard long exposure. However, if the viewer deviates from the intended scanpath a temporal flickering could theoretically occur for the OOI - although no participant reported such observations in our user study. This subtle effect might actually help guiding the gaze towards the OOI, similar to [10]. It is convenient that as soon as the observer follows the OOI, the gaze respects the intended path and potential artifacts will disappear.

Having access to UHFR footage is a benefit because the right tradeoff between exposure time and motion blur is often difficult to decide upon when capturing a scene. Especially for stunt shots, there are many fast movements and repeating the action can be very costly. HFR equipment is currently expensive, but hardware prices already dropped and movie makers start recognizing the new possibilities and advantages. It is difficult to answer, whether a higher frame rate movie or a perceptually-motivated motion blur "looks better". We are conditioned to Hollywood movies recorded at 24 Hz and the audience reacted reluctantly at first to the 48 Hz version of "The Hobbit" as they were not used to the viewing experience. However, there is a clear tendency towards higher frame rates (e.g. "Avatar 2" by James Cameron will be shot at 60 fps) and it is crucial to investigate this area in depth. We think that our solution is a first significant step in this research field.

8 Conclusion

The temporal integration in traditional camera recordings does not correspond to the integration of the human visual system when watching the movie. Our work proposes a gaze-guided as well as gaze-guiding, temporal downsampling to achieve consistent results without edge banding or judder artifacts for real and synthetic video input of arbitrary frame rate. We introduced a model for video perception based on the human visual system. We then described our approach for gaze-guided downsampling using video saliency. We presented different applications for our approach, including downsampling of real world and CG generated ultra-high frame rate videos, virtual shutter simulation, motion stills generation, and subtle gaze direction. Our conducted user study confirms the effectiveness of our approach to influence observers' gaze.

In the future, we want to support multiple objects of interest also via an interpolation of the eye motion vectors over the image plane. One option is a Poisson reconstruction using the OOIs as boundary conditions. Nonetheless, in practice, assuming a single OOI currently leads to better results and our method is robust with respect to deviating eye motion.

We showed that our approach enables novel and interesting post-processing possibilities. We believe there are many more possible applications related to our perceived blur, for example in the field of high dynamic range video reconstruction.

Acknowledgments

The research leading to these results has received funding from the European Union's Seventh Framework Programme FP7/2007-2013 under grant agreement no. 256941, Reality CG. The authors thank Red Digital Cinema Camera Company for offering usage of the bike video.

References

[1] N. Proferes, *Film Directing Fundamentals: See Your Film Before Shooting*. Focal Press, 2008.

[2] M. O'Brien and N. Sibley, *The Photographic Eye SE: Learning to See with a Camera*, ser. Studio Textbooks Series. Davis Publications, Incorporated, 1995.

[3] B. Brown, *Cinematography: Theory and Practice : Imagemaking for Cinematographers, Directors & Videographers*. Focal Press, 2002.

[4] W. Fenlon, "48 FPS and Beyond: How High Frame Rate Films Affect Perception," http://tiny.cc/beyondHFR, 2014, [Online; accessed 23-April-2014].

[5] D. Trumbull, "Douglas Trumbull On High Frame Rate Filmmaking," http://tiny.cc/HFRFilmmaking, 2013, [Online; accessed 23-April-2014].

[6] J. Telleen, A. Sullivan, J. Yee, O. Wang, P. Gunawardane, I. Collins, and J. Davis, "Synthetic shutter speed imaging," *Comput. Graph. Forum*, vol. 26, no. 3, pp. 591–598, 2007.

[7] M. Fuchs, T. Chen, O. Wang, R. Raskar, H.-P. Seidel, and H. P. Lensch, "Real-time temporal shaping of high-speed video streams," *Computers & Graphics*, vol. 34, no. 5, pp. 575 – 584, 2010.

[8] A. Glassner, "An open and shut case," *IEEE Comput. Graph. Appl.*, vol. 19, no. 3, pp. 82–92, May 1999. [Online]. Available: http://dx.doi.org/10.1109/38.761554

[9] G. Osterberg, "Topography of the layer of rods and cones in the human retina," *Acta Ophthal. Suppl.*, no. 6, pp. 11–97, 1935.

[10] R. Bailey, A. McNamara, N. Sudarsanam, and C. Grimm, "Subtle gaze direction," *ACM Trans. Graph.*, vol. 28, no. 4, pp. 100:1–100:14, 2009.

[11] A. McNamara, R. Bailey, and C. Grimm, "Improving search task performance using subtle gaze direction," in *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization*, ser. APGV '08, 2008, pp. 51–56.

[12] G. E. Mitchell, "Taking control over depth of field: Using the lens blur filter in adobe photoshop cs," 2004, http://www.outbackphoto.com/workflow/wf 51/essay.html.

[13] R. Kosara, S. Miksch, and H. Hauser, "Semantic depth of field," in *Proc. of the IEEE Symposium on Information Visualization*, 2001, pp. 97–104.
[14] D. DeCarlo and A. Santella, "Stylization and abstraction of photographs," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 769–776, 2002.
[15] S. T. Hammett, M. A. Georgeson, and A. Gorea, "Motion blur and motion sharpening: temporal smear and local contrast non-linearity," *Vision research*, vol. 38, no. 14, pp. 2099–2108, 1998.
[16] D. C. Burr and M. J. Morgan, "Motion deblurring in human vision," *Proc. Biol. Sci.*, vol. 264, no. 1380, pp. 431–436, 1997.
[17] S. J. Galvin, R. P. O'Shea, A. M. Squire, and D. G. Govan, "Sharpness overconstancy in peripheral vision," *Vision Res*, vol. 37, no. 15, pp. 2035–2044, 1997.
[18] T. Liu, N. Zheng, W. Ding, and Z. Yuan, "Video attention: Learning to detect a salient object sequence," in *International Conference on Pattern Recognition*, 2008, pp. 1–4.
[19] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proceedings of the 14th Annual ACM International Conference on Multimedia*, ser. MULTIMEDIA '06, 2006, pp. 815–824.
[20] M. Dorr, T. Martinetz, K. R. Gegenfurtner, and E. Barth, "Variability of eye movements when viewing dynamic natural scenes." *Journal of vision*, vol. 10, 2010.
[21] M. Böhme, M. Dorr, C. Krause, T. Martinetz, and E. Barth, "Eye movement predictions on natural videos." *Neurocomputing*, vol. 69, no. 16-18, pp. 1996–2004, 2006.
[22] M. Böhme, C. Krause, E. Barth, and T. Martinetz, "Eye movement predictions enhanced by saccade detection," in *In: Brain Inspired Cognitive Systems*, 2004, pp. 1–7.
[23] J.-X. Chai, X. Tong, S.-C. Chan, and H.-Y. Shum, "Plenoptic sampling," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '00, 2000, pp. 307–318.
[24] D. M. Hoffman, V. I. Karasev, and M. S. Banks, "Temporal presentation protocols in stereoscopic displays: Flicker visibility, perceived motion, and perceived depth," *Journal of the Society for Information Display*, vol. 19, no. 3, pp. 271–297, 2011.
[25] J. Stewart, J. Yu, S. J. Gortler, and L. McMillan, "A new reconstruction filter for undersampled light fields," in *Proceedings of the 14th Eurographics Workshop on Rendering*, 2003, pp. 150–156.
[26] A. Bloch, "Experiences sur la vision," *CR Seances Soc. Biol. Paris*, vol. 37, pp. 493–495, 1885.
[27] F. H. Adler, P. L. Kaufman, L. A. Levin, and A. Alm, *Adler's Physiology of the Eye*. Elsevier Health Sciences, 2011.
[28] C. Rasche and K. R. Gegenfurtner, "Precision of speed discrimination and smooth pursuit eye movements," *Vision research*, vol. 49, no. 5, pp. 514–523, 2009.
[29] H. Pan, X.-F. Feng, and S. J. Daly, "Lcd motion blur modeling and analysis." in *Proc. of International Conference on Image Processing*, 2005, pp. 21–24.
[30] A. B. Watson, "High frame rates and human vision: A view through the window of visibility," *SMPTE Motion Imaging Journal*, vol. 122, no. 2, pp. 18–32, 2013.
[31] C. A. Curcio, K. R. Sloan, R. E. Kalina, and A. E. Hendrickson, "Human photoreceptor topography," *Journal of Comparative Neurology*, vol. 292, no. 4, pp. 497–523, 1990.
[32] C. Lipski, C. Linz, T. Neumann, M. Wacker, and M. Magnor, "High resolution image correspondences for video post-production," in *Proc. European Conference on Visual Media Production (CVMP) 2010*, vol. 7. IEEE Computer Society, 2010, pp. 33–39.
[33] P. Didyk, E. Eisemann, T. Ritschel, K. Myszkowski, and H.-P. Seidel, "Apparent display resolution enhancement for moving images," *ACM Trans. Graph.*, vol. 29, no. 4, pp. 113:1–113:8, 2010.
[34] K. Templin, P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, and H.-P. Seidel, "Apparent resolution enhancement for animations," in *Proc. of the 27th Spring Conference on Computer Graphics*, 2011, pp. 85–92.
[35] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in Neural Information Processing Systems 19*, 2007, pp. 545–552.
[36] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime tv-l1 optical flow," in *Proceedings of the 29th DAGM Conference on Pattern Recognition*, 2007, pp. 214–223.
[37] M. Stengel, M. Eisemann, S. Wenger, B. Hell, and M. Magnor, "Optimizing apparent display resolution enhancement for arbitrary videos," *IEEE Transactions on Image Processing (TIP)*, vol. 22, no. 9, pp. 3604–3613, 2013.
[38] A. Agarwala, A. Hertzmann, D. H. Salesin, and S. M. Seitz, "Keyframe-based tracking for rotoscoping and animation," in *ACM SIGGRAPH 2004 Papers*, ser. SIGGRAPH '04. New York, NY, USA: ACM, 2004, pp. 584–591. [Online]. Available: http://doi.acm.org/10.1145/1186562.1015764
[39] D. Scherzer, L. Yang, O. Mattausch, D. Nehab, P. V. Sander, M. Wimmer, and E. Eisemann, "A survey on temporal coherence methods in real-time rendering," in *EUROGRAPHICS 2011 State of the Art Reports*. Eurographics Association, 2011, pp. 101–126. [Online]. Available: http://www.cg.tuwien.ac.at/research/publications/2011/scherzer2011c/
[40] Solid Angle, "Arnold renderer," http://solidangle.com.
[41] C. Eisenacher, G. Nichols, A. Selle, and B. Burley, "Sorted deferred shading for production path tracing." *Comput. Graph. Forum*, vol. 32, no. 4, pp. 125–132, 2013.
[42] R. L. Cook, T. Porter, and L. Carpenter, "Distributed ray tracing," *SIGGRAPH Comput. Graph.*, vol. 18, no. 3, pp. 137–145, 1984.
[43] F. Navarro, F. J. Serón, and D. Gutierrez, "Motion blur rendering: State of the art." *Comput. Graph. Forum*, vol. 30, no. 1, pp. 3–26, 2011.
[44] D. Burr, "Temporal summation of moving images by the human visual system," in *Proc. of the Royal Society of London*, vol. B 211, 1981, pp. 321–339.

Michael Stengel received his Diploma degree in Computational Visualistics from the University of Magdeburg, Germany, in 2011. From 2010 to 2011 he worked at the Virtual Reality Lab at Volkswagen AG. He is currently pursuing his PhD in Computer Graphics at Technische Universität (TU) Braunschweig, Germany. His research interests include visual perception, human-computer-interaction and visualization.

Pablo Bauszat received his Diploma degree in Computer Science from the Technische Universität (TU) Braunschweig, Germany, in 2011. He is currently pursuing his PhD in Computer Graphics at Technische Universität (TU) Braunschweig, Germany. His research interests include real-time rendering, ray tracing, light transport simulation and perceptual rendering.

Martin Eisemann received a Diploma degree in Computational Visualistics from the University of Koblenz-Landau, Germany, in 2006 and his PhD degree in Computer Graphics from the TU Braunschweig, Germany, and received the best student paper award at the annual conference of the European Association for Computer Graphics (Eurographics) in 2008. Since 2011 he is Akademischer Rat (Postdoctoral Researcher) at the Computer Graphics Lab at the TU Braunschweig. His main research interests include image- and video-based rendering and editing, visual analytics, and realistic and interactive rendering.

Elmar Eisemann is a professor at TU Delft, heading the Computer Graphics and Visualization Group. Before he was an associated professor at Telecom ParisTech (until 2012) and a senior scientist heading a research group in the Cluster of Excellence (Saarland University / MPI Informatik) (until 2009). He studied at the École Normale Supérieure in Paris (2001-2005) and received his PhD from the University of Grenoble at INRIA Rhône-Alpes (2005-2008). He spent several research visits abroad; at the Massachusetts Institute of Technology (2003), University of Illinois Urbana-Champaign (2006), Adobe Systems Inc. (2007,2008). His interests include real-time and perceptual rendering, alternative representations, shadow algorithms, global illumination, and GPU acceleration techniques. He coauthored the book ”Real-time shadows” and participated in various committees and editorial boards. He was local organizer of EGSR 2010, 2012 and HPG 2012. His work received several distinction awards and he was honored with the Eurographics Young Researcher Award 2011.

Marcus Magnor is full professor and head of the Computer Graphics Lab at Braunschweig University of Technology. He holds a Diploma degree in Physics (1997) and a PhD in Electrical Engineering (2000). After his postgraduate time as Research Associate in the Graphics Lab at Stanford University, he established his own research group at the Max-Planck-Institut Informatik in Saarbrücken. He completed his habilitation and received the venia legendi in Computer Science from Saarland University in 2005. In 2009, he spent one semester as Fulbright scholar and Visiting Associate Professor at the University of New Mexico. His research interests meander along the visual information processing pipeline, from image formation, acquisition, and analysis to image synthesis, display, perception, and cognition. Ongoing research topics include image-based measuring and modeling, photorealistic and real-time rendering, and perception in graphics.