Saccade Landing Position Prediction for Gaze-Contingent Rendering

ELENA ARABADZHIYSKA, Saarland University, MMCI and Intel Visual Computing Institute
OKAN TARHAN TURSUN, MPI Informatik
KAROL MYSZKOWSKI, MPI Informatik
HANS-PETER SEIDEL, MPI Informatik
PIOTR DIDYK, Saarland University, MMCI and MPI Informatik

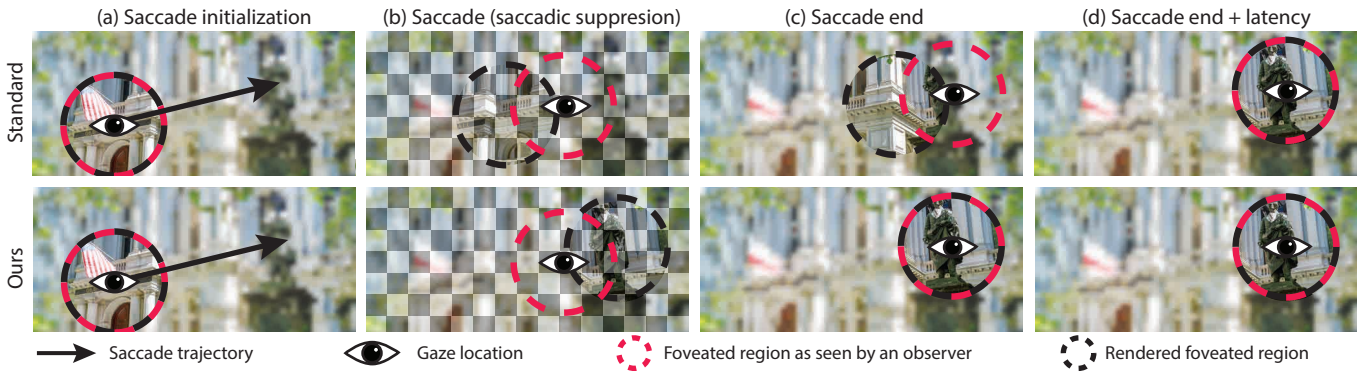


Fig. 1. Standard gaze-contingent rendering (top row) updates the image according to the current gaze prediction. Due to the system latency, during a saccade, there is a significant mismatch between the rendering and the actual gaze position (b, c). The method moves the foveated region to the actual gaze position only after a delay equal to the system latency (d). Our technique (bottom row) predicts the ending position of the saccade at its early stage and updates the image according to the new prediction as soon as it is available (b). Due to the saccadic suppression the user cannot observe the image manipulations during the saccade (b). When the saccade ends and the suppression is deactivated (c), the observer sees the correct image at the new gaze position with our method.

Gaze-contingent rendering shows promise in improving perceived quality by providing a better match between image quality and the human visual system requirements. For example, information about fixation allows rendering quality to be reduced in peripheral vision, and the additional resources can be used to improve the quality in the foveal region. Gaze-contingent rendering can also be used to compensate for certain limitations of display devices, such as reduced dynamic range or lack of accommodation cues. Despite this potential and the recent drop in the prices of eye trackers, the adoption of such solutions is hampered by system latency which leads to a mismatch between image quality and the actual gaze location. This is especially apparent during fast saccadic movements when the information about gaze location is significantly delayed, and the quality mismatch can be noticed. To address this problem, we suggest a new way of updating images in gaze-contingent rendering during saccades. Instead of rendering according to the current gaze position, our technique predicts where the saccade is likely to end and provides an image for the new fixation location as soon as the prediction is available. While the quality mismatch during the saccade remains unnoticed due to saccadic suppression, a correct image for the new fixation is provided before the fixation is established. This paper describes the derivation of a model for predicting saccade landing positions and demonstrates how it can be used in the context of gaze-contingent rendering to reduce the influence of system latency on the perceived quality. The technique is validated in a series of experiments for various combinations of display frame rate and eye-tracker sampling rate.

CCS Concepts: • **Computing methodologies** → **Perception**; **Rendering**; **Image manipulation**;

Additional Key Words and Phrases: gaze-contingent rendering, perception, saccadic suppression, saccade prediction, new display technology, virtual reality

ACM Reference format:
Elena Arabadzhiyska, Okan Tarhan Tursun, Karol Myszkowski, Hans-Peter Seidel, and Piotr Didyk. 2017. Saccade Landing Position Prediction for Gaze-Contingent Rendering. *ACM Trans. Graph.* 36, 4, Article 50 (July 2017), 12 pages.
DOI: http://dx.doi.org/10.1145/3072959.3073642

0730-0301/2017/7-ART50 $15.00
DOI: http://dx.doi.org/10.1145/3072959.3073642

1 INTRODUCTION

Despite the constant improvement of hardware, meeting the quality demands regarding spatial and temporal resolutions, stereoscopic presentation, and scene complexity required in current applications is still a challenging problem. This is manifested in the recent developments of new mobile platforms as well as virtual and augmented reality (VR/AR) systems, where both the quality and energy efficiency are limiting factors that have to be tackled to enable full adoption of these technologies. With the recent advances in affordable eye-utracking technology, the above problems can be addressed by exploiting properties of the human visual system (HVS). The most

prominent example is foveated rendering techniques [Guenter et al. 2012; Murphy and Duchowski 2001; Patney et al. 2016; Vaidyanathan et al. 2014] which take advantage of the decay in visual acuity towards the periphery [Banks et al. 1991; Curcio et al. 1990; Grüsser and Grüsser-Cornehls 1986], and provide high image quality only for the fovea. This leads to improvements both in the rendering time and quality. A similar principle can be employed to improve realism and viewer experience by simulating depth-of-field effects [Mantiuk et al. 2011] and local luminance adaptation [Jacobs et al. 2015], reducing the vergence-accommodation conflict during a stereoscopic presentation [Duchowski et al. 2014], or enhancing depth impression by gaze-driven disparity manipulations [Kellnhofer et al. 2016].

Although gaze-contingent rendering can lead to significant improvements, it is very sensitive to system latency [Saunders and Woods 2014]. When gaze location changes rapidly during saccades, even short delays may result in visible artifacts which make the gaze-contingent rendering unfavorable. This problem is alleviated to some degree by the saccadic suppression, which lowers the sensitivity of the HVS during saccades [Ross et al. 1996, 2001; Volkmann et al. 1978]. However, the HVS sensitivity is fully restored just 40–60 ms after the saccade ends, and this process is gradual, so even shorter delays can be noticed depending on the magnitude of the changes [Loschky and Wolverton 2007]. Therefore, displaying an updated image after saccade completion is critical, and any delays may limit the benefits of gaze-contingent rendering. In practice, to prevent problems with system latency, high-end equipment has to be used. Guenter et al. [2012] performed their experiments using a 300 Hz eye tracker with a 120 Hz display, which resulted in an overall latency of 20–40 ms. While such displays are becoming a commodity, high-quality rendering rarely achieves such frame rates. An alternative to keep both the rendering quality and the frame rate high is to use efficient upsampling techniques, such as "Spacewarp" and "Timewarp" [Oculus VR 2016a,b]. This, however, does not overcome the problem of latency. Also, 300 Hz eye trackers are prohibitively expensive for regular users. As a result, the gaze-contingent systems introduce significant latency which leads to visible artifacts such as the perception of low-quality image from the peripheral rendering.

To address this problem, we propose a new technique for controlling gaze-contingent rendering during saccades. The key idea is to maximally exploit the saccadic suppression when it is the strongest, i.e., during saccades. Instead of placing the foveated region at each gaze position as soon as it is provided by the eye tracker, our method fixes it to a predicted saccade landing position (Figure 1). The prediction is continuously adjusted during the saccade for new gaze direction samples so that the delay with which the correct image appears is minimized when the saccade ends. During the saccade, the mismatch between the actual gaze direction and the rendering is hidden by the saccadic suppression. In this work, we propose a method for predicting the saccade landing position. It accounts for both within- and between-participant saccade variability as well as inaccuracies of modern eye trackers. We demonstrate how the prediction can be used in the context of gaze-contingent rendering to alleviate the problem of system latency. Our user experiments validate the accuracy of the predictions and the quality improvements when our strategy is applied. To provide further insights, we use different combinations of display frame rate and eye-tracking sampling rate in our tests. We present the following contributions:

- measurements and analysis of saccade trajectories,
- a new model for predicting the landing position of saccades,
- a comparison of several prediction techniques based on our measurements, and
- two experiments which validate both a subjective and an objective quality increase when our method is applied in a gaze-contingent rendering system.

2 BACKGROUND AND PREVIOUS WORK

In this section, we motivate our work by discussing recent advances in gaze-contingent rendering. We also provide basic facts on saccadic suppression, which enables our technique. Finally, we describe the mechanics of saccadic eye motion trajectories and discuss previous work on their prediction.

2.1 Gaze-Contingent Rendering

The sensitivity of the HVS to luminance contrast, color, and depth is not equal across the visual field. It is highest in the fovea and declines significantly towards the periphery [Noorlander et al. 1983; Prince and Rogers 1998; Strasburger et al. 2011]. This is in agreement with the variation of photoreceptor density across the retina [Curcio et al. 1990], and the volume of neural circuitry that further processes the foveal and peripheral signal, i.e., cortical magnification [Grüsser and Grüsser-Cornehls 1986]. This leads to distinct characteristics of *foveal* (high-quality) and *peripheral* (low-quality) vision [Banks et al. 1991], where the foveal vision is usually assumed to span 1–5°.

These findings can be beneficial in rendering systems that use eye tracking devices. In foveated rendering, the key idea is to increase rendering performance by providing lower quality to the peripheral vision. This has been successfully demonstrated for spatial resolution [Guenter et al. 2012; Stengel et al. 2016; Swafford et al. 2016; Vaidyanathan et al. 2014], level-of-detail control [Duchowski et al. 2009], and color [Duchowski et al. 2009]. Gaze-contingent rendering can also help overcome certain display limitations. Simulating the depth-of-field effect leads to improved realism [Mantiuk et al. 2011; Mauderer et al. 2014], as well as better depth perception [Vinnikov and Allison 2014] and reduction of visual discomfort [Duchowski et al. 2014] for stereoscopic displays. Gaze-driven stereo disparity manipulations can not only improve visual comfort by reducing the vergence-accommodation conflict, frame violation, and crosstalk [Hanhart and Ebrahimi 2014], but also enhance depth perception [Kellnhofer et al. 2016]. In tone mapping, the dynamic range can be allocated effectively by reducing the image contrast with eccentricity [Jacobs et al. 2015].

All these techniques respond to the gaze direction updates as provided by an eye tracker. In practice, this information is delayed by an accumulated latency of the eye tracker, rendering, and display system [Saunders and Woods 2014]. Many authors report problems with excessive latency which may lead to user dissatisfaction [Duchowski et al. 2014; Mantiuk et al. 2011], reduced task performance, or even user sickness [Draper et al. 2001; Frank et al. 1988]. In this work, we address this problem for situations where the delays in gaze direction are the most significant, i.e., during fast eye

movements called saccades. To this end, we propose to predict the saccade landing location (Section 2.3) at its early stage and provide an image for the new fixation location before the fixation. Additionally, the image update is performed during the saccade when the HVS sensitivity is reduced by the saccadic suppression (Section 2.2). This makes the update imperceptible. As shown in our experiments, this strategy significantly reduces the impact of accumulated system latency and improves the quality of gaze-contingent rendering.

2.2 Saccadic and Post-Saccadic Suppression

Saccades are rapid eye motions which align the fovea with the region of interest for the best resolution of spatial details. During the saccades, a stable depiction of the surrounding world is attributed to the so-called saccadic suppression, which strongly compresses the sensory information [Ross et al. 2001; Volkmann et al. 1978]. There are three main factors contributing to this suppression. The first is the retinal smear due to fast eyeball rotation and relatively long time integration, which results in luminance contrast reduction for higher spatial frequencies [Volkmann et al. 1978]. The second is the visual masking that occurs in the presence of retinal-image motion [Diamond et al. 2000]. The third is the neural inhibition of luminance contrast, where the strongest compression is observed for motion sensing mechanisms, in particular for lower spatial frequencies that are usually enhanced due to retinal pattern motion [Ross et al. 2001]. The saccadic suppression is substantial but not complete. It is initialized around 50 ms before the saccade, and, at its peak, amounts to 0.5 log units of threshold elevation [Volkmann et al. 1978]. It is known that the saccadic suppression diminishes towards the end of the saccade, but there is a disagreement regarding its peak. While Volkmann et al. [1978] report the peak of suppression in mid-saccade, Ross et al. [2001] observe that the suppression is strongest at the beginning.

A weak suppression also extends beyond the saccade – a so-called post-saccadic suppression. Loschky and Wolverton investigated the effect in the context of gaze-contingent rendering [Loschky and Wolverton 2007, Figure 3] and showed that a local change in the spatial resolution could be perceived when performed 5 ms after the saccade completion. Similarly, Bodelón et al. [2007] demonstrated that grating orientation can be detected already after 8.4 ms. The HVS sensitivity is fully restored in 40–60 ms after the saccade ends [Diamond et al. 2000; Loschky and Wolverton 2007; Volkmann et al. 1978]. Shortly after that, the sensitivity can even be enhanced [Volkmann et al. 1978, Figure 5].

From the above discussion, it is clear that the HVS sensitivity to image changes continuously increases starting from mid-saccade, and it quickly recovers or even exhibits enhancements at the early post-saccadic stage. Current gaze-contingent rendering techniques rely on both saccadic and post-saccadic suppression (Section 2.1) and continuously update images according to new eye-tracking samples even during the post-saccadic stage. However, the higher the latency of the system, the more likely such updates will be detected as distortions. In practice, the updates are quite limited, which greatly reduces potential benefits from gaze-contingent rendering. The goal of our work is to avoid significant image updates towards the end of the saccade and later on. This reduces the risk of performing major changes in the post-saccadic stage.

2.3 Saccade Landing Position Prediction

Saccade velocity and duration cannot be voluntarily controlled, and normally, the oculomotor system follows a preprogrammed ballistic motion trajectory [Kowler 2011]. Although it has been demonstrated that, in some cases, the central nervous system is capable of changing the trajectory of the saccades in flight, it takes approximately 70 ms for visual information to travel from the retina to oculomotor mechanisms of the brain [Leigh and Zee 2015, Chapter 3]. Since the duration of saccades usually falls between 20 and 80 ms, there is not enough time to respond to stimuli during the saccade. A large body of work has been dedicated to analyzing saccade ballistics. For example, it has been demonstrated that there is a linear relationship between the duration and the amplitude of saccades [Bahill et al. 1975]. On the other hand, the same work showed that the relation between the duration and the peak velocity (*the main sequence*) is nonlinear. Moreover, velocity profiles of short saccades are symmetric. This, however, does not hold for medium and long saccades whose profiles are skewed towards their beginning [Van Opstal and Van Gisbergen 1987].

The ballistic characteristics have been exploited for modeling saccades. Anliker [1976] and Paeye et al. [2016] assume symmetrical velocity profiles, and predict the landing position essentially by doubling the distance traveled until peak velocity. Yeo et al. [2012] define saccade velocity profiles using a bell-shaped curve for simulating HVS dynamics during object tracking. As all of these solutions assume symmetry of the velocity profiles, they can be used only for short saccades [Van Opstal and Van Gisbergen 1987]. Han et al. [2013] propose a method based on fitting a compressed exponential function to the eye trajectory. In contrast to our work, they focus on providing short-term predictions (10 ms), while we aim to predict landing positions. Komogortsev and Khan [2009] propose the Oculomotor Plant Kalman Filter (OPKF) that can handle both tasks, and accounts for many anatomical eye properties. A number of anatomy-inspired complex plant saccade models exist, such as [Zhou et al. 2009], but they are less suitable for real-time, gaze-contingent applications as they are not designed to perform prediction [Han et al. 2013]. Based on the OPKF model, Komogortsev and Khan proposed a computationally efficient chi-square test whose peak value is correlated with the saccade amplitude [Komogortsev et al. 2009b]. This model can predict the saccade landing position at an early stage and can be used for fast target selection in gaze-guided computer interaction [Komogortsev et al. 2009a]. However, the model was tested only for a single, large (5°) target, and only horizontal saccades were considered. In Komogortsev et al. [2009b], an average prediction error of 5.41° was reported for a similar model, where again only horizontal saccades were considered, while the model proposed by Anliker [1976] performed best in such conditions with an average error of 3.46°.

In contrast to the previous work, we propose a simple yet efficient and accurate model which provides landing position prediction at early stages of the saccade with a continuous refinement. The model can be integrated into any gaze-contingent rendering system at a negligible computational cost. Additionally, our model covers a wide range of saccades, and it is independent of their direction. We argue that these are crucial features for gaze-contingent rendering. To

our knowledge, we are also the first to demonstrate and analyze the performance of gaze-contingent rendering with this type of prediction.

3 ONLINE PREDICTION OF THE SACCADE LANDING POSITION

In our modeling, we aim to predict the landing position of saccades to update gaze-contingent rendering early enough to reduce the delays coming from the latency of the system. The greatest challenge in building such a system is robustness. Standard gaze-contingent rendering suffers from latency; however, the introduced delay is always consistent. When a prediction is used, even small errors may lead to catastrophic failures which will result in clearly visible artifacts, and therefore, user dissatisfaction. Our goal is to create a system which is robust to fluctuations of eye-tracking data arising from instabilities in eye movements, as well as from eye-tracker errors.

To this end, we follow the assumption that saccades obey ballistic trajectories which are determined mainly by saccade amplitude. Even though this assumption does not hold completely, as other factors may influence saccades (Section 5), we demonstrate that it leads to a simple yet powerful and robust model that provides significant improvements in gaze-contingent applications (Section 4). To gain knowledge about saccade characteristics, we first perform measurements to collect samples of many saccades performed by several participants (Section 3.1). After analyzing the collected data (Section 3.2), we construct a computational model (Section 3.3) that captures the characteristics of different saccades and uses eye-tracker samples to predict their landing positions.

3.1 Measurements

We conducted a user experiment to collect a large amount of data associated with different saccades. To evoke saccades, participants were asked to focus on a target stimulus, which was a white dot on a uniform 50%-gray background shown on a screen operating at 60 FPS. The target changed its position when the user pressed a key after fixation. The target positions were pre-generated and shuffled so that different saccade amplitudes, spanning a range of 5°–45°, were equally represented in the collected data. The data collection was performed with an eye tracker at 300 Hz sampling frequency. The eye tracker allows free head movement during tracking; nevertheless, we used a chin-rest to improve the tracking accuracy. The viewing distance was fixed to 70 cm, which resulted in a coverage of 46×27 visual degrees. 22 participants with normal or corrected-to-normal vision took part in our measurements. The eye tracker was calibrated for each participant. Every participant performed at least 300 saccades which were recorded during a 5-minute session. Figure 2 shows our setup and gaze data recorded from one of our participants.

System Details. In all our experiments, including the validation described in Section 4, we used a Tobii TX300 eye tracker, capable of 300 Hz, 120 Hz and 60 Hz sampling rates. The display was a 27" 2560×1440 ASUS ROG Swift PG278Q which supports 60 Hz, 120 Hz and 144 Hz refresh rates. The CPU and GPU of our platform were an Intel(R) Xeon(R) E5-1620 v3 @ 3.50 GHz and NVIDIA GeForce GTX 660, respectively. The software system used was our own C++ implementation based on OpenGL.

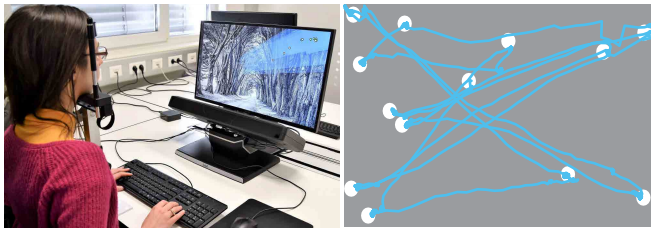


Fig. 2. Our measurement setup and a sample subset of eye-tracking data from one participant. The white circles visualize saccade targets shown during the experiment, whereas the blue traces correspond to eye-tracker samples. For simplicity, only 12 targets are visualized here. Throughout the actual experiment, 300 consecutive targets were shown to each observer.

3.2 Data Processing

The data recorded in the experiments includes both saccades and fixations. To build our model, we need to extract eye-tracking samples corresponding to saccades. Detecting saccades can be easily performed in a post-processing step by analyzing velocity profiles. However, in order to be consistent with how the saccades will be detected in gaze-contingent rendering techniques, we opt for a robust analysis that is suitable for on-line data. Several techniques have been proposed and analyzed in this context [Andersson et al. 2016; Salvucci and Goldberg 2000]. For our purpose, we found that the velocity threshold method (I-VT) provides satisfactory results. I-VT relies on the fact that the saccades are very fast eye movements and detects them as soon as the high velocity of eye movement is observed. We denote the detection threshold as V_d and refer to the first gaze sample whose velocity exceeds V_d as *the detection point*.

For the saccade detection to be robust, V_d has to be relatively high, usually above 100°/s. This means that the true beginning of the saccade is much earlier than the detection point. To obtain all the past samples of the saccade, we scan the gaze samples backward in time to find the beginning. Due to the inherent noise of eye trackers and small movements of the eye, the velocity is always positive even during fixations. Therefore, we employ a two-step procedure similar to Dorr et al. [2010] by introducing another velocity threshold, V_a. The first sample where the measured gaze velocity is equal to V_a is the *anchor point*, i.e., beginning of the saccade. Due to discrete sampling of eye trackers, it is unlikely that a sample with this exact velocity will be found in practice. Therefore, we introduce an additional sample for velocity V_a, by interpolating between the first sample for which the velocity is over V_a and the previous sample. We also use a velocity threshold to detect the end of the saccade and refer to the first sample whose velocity drops below V_f as *the end point*. Additionally, we treat samples occurring up to 15 ms after *the end point* as a part of the saccade to account for potential corrective saccades called *glissades*, which are typically not larger than 0.5°–2° [Holmqvist et al. 2011, Ch. 2]. We ignore the *glissades* detected separately by removing the detections which are shorter than 15 ms. For increased robustness, we only take the saccades whose *anchor points* are found within a 30 ms interval

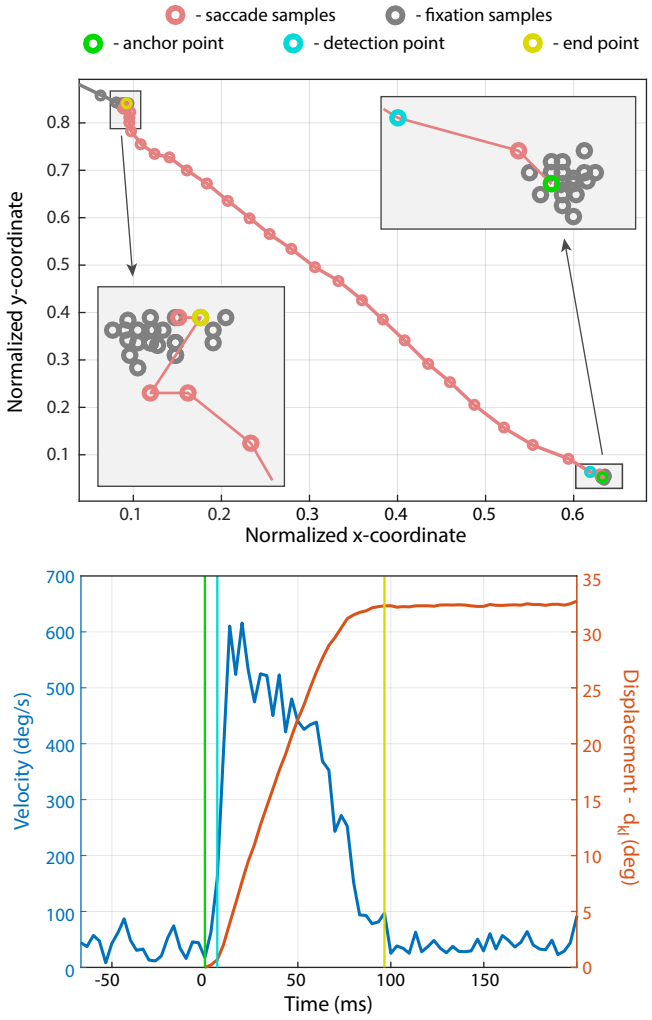


Fig. 3. Visualization of our data processing for one sample saccade. Top: Gaze samples for a saccade from the bottom-right corner to the top-left corner of the screen. Bottom: The gaze velocity and displacements. Samples corresponding to the anchor, detection and the end points are indicated in green, cyan and yellow colors, respectively. Captured with a Tobii TX300 eye tracker at 300 Hz sampling frequency.

before the detection. If tracking is lost during a saccade, it is not used for training our model. Please see Figure 3, which shows the anchor, detection, and end points for a sample saccade.

We define each saccade, S_k, as a set of subsequent gaze samples from the eye tracker:

$$S_k = \{s_{k0}, s_{k1}, s_{k2}, ..., s_{kN}\}, \tag{1}$$

where s_{k0} is the gaze sample corresponding to *the anchor point* of the saccade and s_{kN} is *the end point* of the saccade. The gaze samples are expressed in terms of triplets:

$$s_{kl} = \langle t_{kl}, d_{kl}, \theta_{kl} \rangle, \tag{2}$$

where t_{kl} is the timestamp, d_{kl} is the displacement and θ_{kl} is the direction of the sample. t_{kl}, d_{kl} and θ_{kl} are measured with respect to *the anchor point*; therefore,

$$t_{k0} = 0,\ d_{k0} = 0,\ \theta_{k0} = 0. \tag{3}$$

This representation is similar to the polar coordinate system where the origin is *the anchor point* of the saccade and d_{kl} corresponds to the radial coordinate, while θ_{kl} corresponds to the polar angle. d_{kl} is measured in terms of visual degrees instead of pixels to make it independent of the distance between the observer and the screen.

Given our representation, the amplitude of saccade S_k is defined as

$$|S_k| = d_{kN}, \tag{4}$$

and the direction as

$$\angle S_k = \theta_{kN}. \tag{5}$$

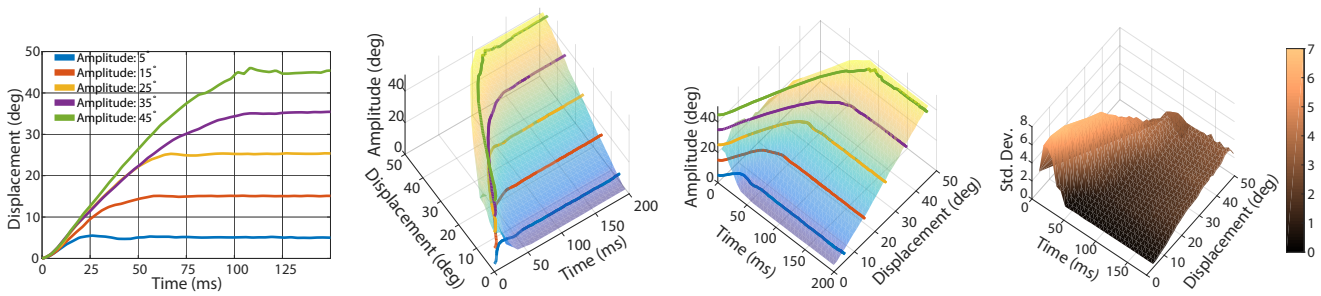
The choice of velocity threshold is crucial for robustness to noise and small involuntary eye movements. Saccades involve very fast eye motion; therefore, they achieve speeds that cannot be observed during other eye movements. Usually, the achieved velocity during a saccade exceeds 100°/s. Consequently, we set $V_d = 130$°/s. Possible misses of very short saccades are not very problematic as these usually do not lead to problems in gaze-contingent rendering. Also, smooth-pursuit eye movements are successfully filtered out by this detection threshold as they rarely reach velocities above 80°/s [Daly 1998; Meyer et al. 1985]. For the anchor point, the velocity threshold V_a needs to be chosen such that we do not include eye-tracking samples corresponding to fixations. We observed that a good choice of the threshold may depend on the participant, as different eye-tracking noise and involuntary movements can be observed for different participants. However, for better generalization, we decided to choose a conservative value that reduces the probability of including fixation samples as saccades in our training data. Consequntly, we set $V_a = 60$°/s, which is used in all our experiments. We set the threshold for the end point accordingly as $V_f = 60$°/s. All our choices remain in agreement with the general characteristics of saccades' velocity profiles described by Boghen et al. [1974].

3.3 Model

The problem of predicting the landing position can be defined as estimating $|S_k|$ and $\angle S_k$ for every timestamp t_{kl} during the saccade, i.e., before d_{kN} and θ_{kN} are actually observed. As the trajectories of most saccades are linear or approximately linear [Leigh and Zee 2015, Ch. 3], our estimate for the direction of the saccade, $\widehat{\angle S_k}$, at a point with timestamp t_{kl} is equal to the direction of the last observed gaze sample:

$$\widehat{\angle S_k}(t_{kl}) = \theta_{kl}. \tag{6}$$

Because of the different acceleration, deceleration, peak velocity and duration characteristics, the displacement observed at a specific time after *the anchor point* depends on the amplitude of a saccade. Figure 4a shows how displacement profiles change with respect to different saccade amplitudes. The characteristics of this change are captured in the data that we collected in our experiment (Section 3.1). The displacement profiles tend to form a consistent surface in the 3D space where the x- and y-axis are the displacement and time axes, while the z-axis corresponds to the amplitude values (see Figure 4b-4c).



(a) (b) (c) (d)

Fig. 4. Displacement profiles of one participant for different saccade amplitudes are given in (a). The corresponding prediction surface is shown from two different viewing angles in (b) and (c) with the saccade amplitude in the z-axis. Standard deviation of the prediction surface across all participants is given in (d). For simplicity, only 5 individual saccades are shown in (a), (b) and (c). We collected more than 300 saccades from each participant.

We treat the amplitude prediction as a regression problem. Based on the collected data, we seek a function f that estimates the saccade amplitude given a displacement d_{kl} at a given time period t_{kl} from the beginning of the saccade to the current samples. Formally, we define our prediction as:

$$\widehat{|S_k|}(t_{kl}) = f(t_{kl}, d_{kl}), \tag{7}$$

and require function f to minimize:

$$\sum_{k,l} (|S_k| - f(t_{kl}, d_{kl}))^2. \tag{8}$$

While the amplitude function $f(t_{kl}, d_{kl})$ gives us a direct prediction of saccade length, it is also possible to model the displacement during saccades as a function of time and amplitude. This may potentially provide a more stable solution, as modeling the steep part of f for small values of displacement and time (Figure 4) might be problematic. To this end, we also consider a function g that minimizes:

$$\sum_{k,l} (d_{kl} - g(|S_k|, t_{kl}))^2. \tag{9}$$

The prediction of saccade length using g requires a linear search that, for a given timestamp t_{kl} and displacement d_{kl}, finds $\widehat{|S_k|}$ such that $\widehat{|S_k|} = g(|S_k|, t_{kl})$. More formally,

$$\widehat{|S_k|}(t_{kl}) = \arg\min_x \left(d_{kl} - g(x, t_{kl})\right)^2. \tag{10}$$

Both Equations 7 and 10 perform prediction based only on the last sample from the eye tracker. This does not account for information provided by all saccade samples. To investigate whether all samples can provide a better prediction, we modify the prediction in Equation 10 to account for all the samples:

$$\widehat{|S_k|}(t_{kl}) = \arg\min_x \sum_{n=1}^{l} \left(d_{kn} - g(x, t_{kn})\right)^2. \tag{11}$$

Effectively, this prediction takes all saccade samples observed until time t_{kl} and tries to find $\widehat{|S_k|}$ such that the saccade profile best fits function g.

Both of the functions f and g can be realized using different parameterization techniques. In order to find the best technique, we compare predictions provided by a polynomial fitting and an interpolation. To this end, for each method, we treat every participant separately and find a personalized polynomial fit or an interpolation grid that minimizes Equation 8 or Equation 11. To determine the degree of the polynomials and to avoid over-fitting, we analyze cross-validation errors for polynomial degrees ranging from 1 to 7. According to this analysis, polynomial degrees 5 for t and 2 for d provide the lowest cross-validation error for the largest number of participants. Although it is possible to optimize the polynomial degrees for each participant separately, this procedure did not provide better fits in our experiments. Consequently, we used a fixed degree for all polynomial-based models. To enable interpolation, we remove multiple measurements corresponding to the same pairs (t_{kl}, d_{kl}) by applying a 2D median filter on saccade amplitudes. This step also reduces the effect of noise and outliers. The size of the filter is 0.5° for the displacement and 2 ms for the time dimension.

We use 80% of our saccade measurement data (Section 3.1) to build prediction models for all methods described above. Later, we use the remaining 20% for testing to compare their performance. The mean absolute error as well as the standard deviation of the error are shown in Figure 5 for all prediction methods. Both the prediction accuracy and standard deviation of the prediction error improve during the course of the saccade. This, as expected, is a result of diverging behavior of displacement profiles for different saccade amplitudes (Figure 4a). The best performing method in terms of error and standard deviation is the interpolation based on Equation 7 whose mean error drops below 4° in the middle of the saccade and is less than 1° at 80% of the normalized saccade time. This magnitude of prediction error is significantly smaller than the typical foveal region size in gaze-contingent rendering [Patney et al. 2016, Fig. 8]. Interestingly, the best prediction is based on only one (the last) sample from the eye-tracking data, which means that our prediction does not benefit from including all of the saccade samples. This might be related to less reliable information provided by the samples at the early stage of the saccades. In addition, we measure the variability of interpolation-based model between participants (see Figure 4d). The high variance at the beginning of the saccades suggests that their characteristics differ significantly between subjects, which supports the choice of using personalized models.

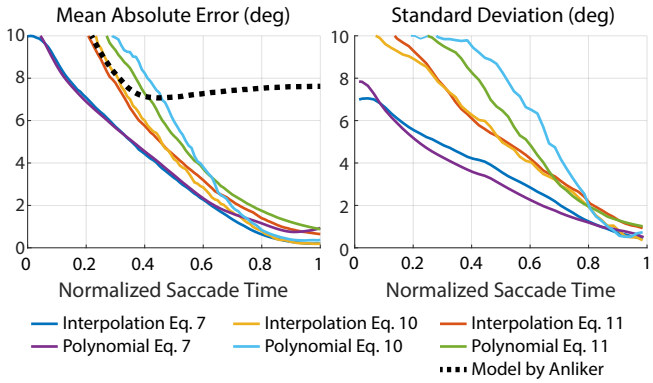


Fig. 5. Comparison of different prediction modeling methods. Left: the mean absolute error as a function of normalized saccade time. Right: the standard deviation of the mean absolute error.

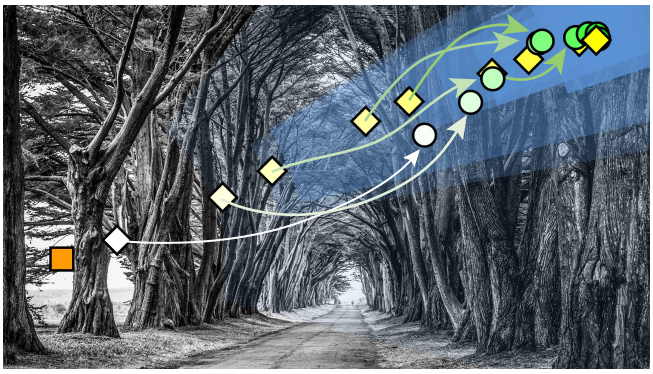


Fig. 6. Gaze samples (yellow), landing position predictions (green) and corresponding prediction intervals (blue) for a sample saccade. The beginning of the saccade (orange), gaze samples and predictions are shown as the square, diamond and circular shapes, respectively. Color saturation level of the points indicates the time when each sample is observed and each prediction is made (more saturated color indicates more recent sample and prediction). Arrows connect gaze samples with our model's corresponding predictions for the landing position.

We compared the prediction performance of our method with that of Anliker [1976]. We observe that this method results in high prediction errors caused by severe undershooting for large saccades due to the violation of symmetry in velocity profiles (Figure 5, left). This suggests that any prediction method based on the symmetry assumption (including symmetric curve fitting approaches [Paeye et al. 2016]) would likely suffer from the same type of inaccuracy.

Average Model. In addition to personalized models, we further investigated the possibility of replacing them with one averaged model which gives the flexibility of predicting saccade amplitudes without the training step. To derive the average models, we used the same procedure as for the personalized models, but with a leave-one-out cross-validation strategy which removes one participant for the training stage. Figure 7 shows how much personalized models improve the mean absolute error compared to the average model both for interpolation and the polynomial fit approach. As expected, for most of the participants personalized models provide better prediction, by up to 30% (1°). This might be explained by the high variance between pesonalized models (Figure 4d). Interestingly, the personalized model of one participant performs about 10% worse than the average model, which might be related to the high measurement noise observed during the data recording session for this participant. The comparison suggests that although in most cases the personalized model leads to improved predictions, the average model is a practical alternative. Here we report the average model that results from the polynomial fit approach:

$$\begin{aligned} f_{poly}(t, \acute{d}) = &-10.19t + 19.11t^2 - 17.15t^3 + 6.251t^4 - 0.7552t^5 \\ &- 23\acute{d}t + 26.89\acute{d}t^2 - 14.74\acute{d}t^3 + 3.882\acute{d}t^4 - 0.4005\acute{d}t^5 \\ &- 11.84\acute{d}^2t + 9.326\acute{d}^2t^2 - 2.583\acute{d}^2t^3 - 0.01058\acute{d}^2t^4 \\ &+ 0.06587\acute{d}^2t^5 + 20.18\acute{d} + 5.071\acute{d}^2 + 18.15, \end{aligned} \tag{12}$$

where $t = (t_{kl} - 47.39)/33.4$ and $\acute{d} = (d_{kl} - 14.67)/11.72$ are the time and displacement measurements normalized with their respective means and standard deviations from the training data. For the personalized models and the gaze data collected from all participants, please refer to the supplementary materials of this paper.

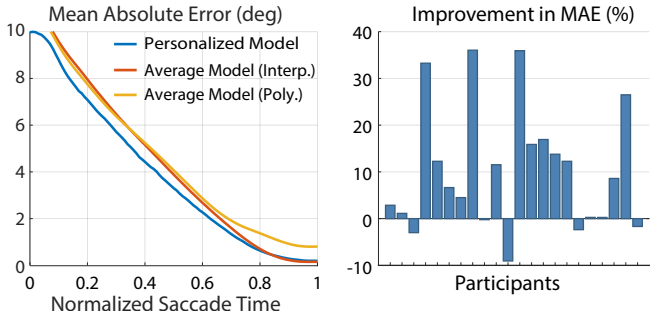


Fig. 7. Left: The mean absolute error of personalized models is compared with that of the average models using interpolation and polynomial fitting. Right: The amount of improvement in the mean absolute errors when personalized models are used instead of the average model for each participant.

Prediction Intervals. Saccades may exhibit variance due to the noise in the firing of motoneurons, target detection inaccuracy of the participant, and measurement inaccuracies [Smeets and Hooge 2003] (Section 5). The information about the within-participant variance is captured in our data. We compute 95^{th}-percentile prediction intervals over the interpolation grid for both the direction and amplitude predictions. Figure 6 demonstrates the prediction intervals for each prediction made during the saccade. Please note how the size of the intervals gets small very quickly. The animation visualizing the process can be seen in the supplementary video. Although we use these intervals only for visualization purposes, we believe that they provide additional information that could be used in the gaze-contingent rendering or gaze-driven interaction techniques.

Application of the Model. To apply our prediction during a gaze-contingent rendering, we replace eye-tracker samples that correspond to saccade samples with our prediction. To detect saccade samples, we apply the same strategy as described in Section 3.2. We start using our technique as soon as the saccade is detected at sample s_{k1}. From that point, all predictions are accepted until we detect the end of the saccade, making s_{kN-1} the last substituted sample. After this sample, we switch to the standard gaze-contingent rendering method which directly uses the samples from the eye tracker. To make the method more robust to noise, we exclude saccades for which the direction changes by more than 5° for the first three consecutive gaze samples at the 300 Hz sample rate and by more than 12.5° at the 120 Hz sample rate. In such cases, the detection is regarded as a false positive, and we switch to the standard gaze-contingent rendering which updates the position of the foveated region according to the most recent gaze sample. For all the samples that are classified as saccade samples, we compute the prediction based on the predicted direction (Equation 6) and the amplitude (Equation 7 and 10). The predicted saccade amplitude $\widehat{|S_k|}(t_{kl})$, and direction $\widehat{\angle S_k}(t_{kl})$, are transformed into a vector representing 2D screen coordinates as follows:

$$p(t_{kl}, d_{kl}) = \left[s_{k0}^{(x)} + h\left(\widehat{|S_k|}(t_{kl})\right) \cos\left(\widehat{\angle S_k}(t_{kl})\right), \; s_{k0}^{(y)} + h\left(\widehat{|S_k|}(t_{kl})\right) \sin\left(\widehat{\angle S_k}(t_{kl})\right) \right]^{\mathsf{T}}, \tag{13}$$

where $s_{k0}^{(x)}$ and $s_{k0}^{(y)}$ are the horizontal and vertical coordinates of the *anchor point* on the screen, respectively, and $h(.)$ is the function which converts the displacement in visual angles to pixel displacement on the screen plane. Next, instead of using the current eye-tracker sample for updating gaze-contingent rendering, we use our prediction. We store the prediction model as an interpolation look-up table or polynomial coefficients for each participant. The whole prediction adds a negligible cost to the gaze-contingent method as it involves a simple lookup and interpolation, or evaluating a polynomial.

4 VALIDATION

To validate our strategy for updating gaze-contingent rendering, we performed two user experiments. In the first one, we used simple, synthetic stimuli to demonstrate that our prediction can significantly reduce the delay in updating gaze-contingent rendering after saccades. The second experiment demonstrates a more natural scenario when foveated rendering is used to render natural scenes. By allowing the user to freely explore the content, we measured the influence of our technique on the user experience. Furthermore, we validated our technique for three different combinations of display frame rates and eye-tracker sampling frequencies to investigate the influence of system latency on the performance of our method. The viewing setup as well as the hardware used in these experiments was the same as in our measurements (Section 3.1).

4.1 Guided-Viewing Experiment

The goal of the first experiment was to demonstrate that our technique can lead to quicker updates of the foveated region. To this end, we designed a simple experiment which was inspired by foveated rendering techniques. The stimulus consisted of four differently oriented Landolt C shapes arranged in a 2×2 grid. Three of them were sharp, and one was blurred (Figure 8, top). The size of each shape was 0.4°, the entire 2×2 grid was 1.2°, and the standard deviation of the Gaussian blur was 0.04°. During the experiment, the stimuli appeared in different locations. Each time the orientations of the Landolt C shapes (up, down, left, right), as well as the position of the blurred one, were chosen randomly. Each time the gaze prediction was in some proximity to the stimulus center (7°) the blur was removed to reveal the masked shape. In half of the cases, appearing in random order, the gaze sampling from the eye tracker was used directly; in the rest of the cases, our prediction was used. The task of the participants was to indicate the orientation of the Landolt C shape that was blurred. After receiving the choice of the participant, a new stimulus was displayed at a different position. The idea was that any delays in gaze-contingent rendering would result in an associated delay in removing the blur. Therefore, the longer the delay in the image update, the easier it was to spot which shape was blurred.

This type of blur-removal strategy simulates the foveated region update scheme used in gaze-contingent rendering. If the update lags behind the actual gaze position as in the standard gaze-contingent applications, the participant will arrive at the stimulus position before the blur is removed and will be able to indicate the shape that was blurred. On the other hand, if our method is used, the foveated region will already be moved to the stimulus position before the arrival of the observer; therefore, the position of the blur will be harder to determine.

Nine participants with normal vision took part in this experiment. For each participant, we used his personalized model to predict saccade landing positions. Each participant saw 300 stimuli for which they indicated the orientation of the blurred Landolt C shape. The experiment took approximately five minutes. We used an eye-tracking sampling frequency and display frame rate of 120 Hz and 60 FPS, respectively. Figure 8 (bottom) shows the ratio of correct responses averaged across all participants. For presentation purposes, we grouped the data according to the amplitude of saccades. For all cases, the number of correct answers was lower when our prediction was used. This indicates that for our update strategy, the blur was removed earlier, effectively reducing the influence of the system latency. The results for medium and large saccades are statistically significant according to Fisher's exact test ($p < 0.001$). For large saccades, the ratio of the correct responses for our method is very close to the expected value of purely random choice (0.25). The difference between the two update strategies is not significant for the shortest saccades. Probably this is related to the proximity of the stimulus to the foveated region, i.e., the stimulus that appears very close to the current fixation location is already in the foveated region and no blur is applied.

The results of this experiment demonstrate that our prediction is fast and accurate enough to reduce the influence of the system latency causing delays in gaze-contingent rendering. As we were not able to demonstrate an advantage for our technique for the shortest saccades, in the rest of our experiments, we modify our prediction model to respond only for medium and long saccades. This is done by raising the detection velocity threshold V_d to 180°/s.

This, in accordance with Boghen et al. [1974], eliminates saccades shorter than 10°.

4.2 Free-Viewing Experiment

The first experiment confirms that our prediction technique allows us to move the region with high rendering quality to the new fixation locations sooner than standard gaze-contingent rendering. However, this does not guarantee that our technique is free of any artifacts and that it provides better perceived quality in general. Therefore, in the second experiment, we validated the performance of our technique in a free-viewing scenario. To this end, we implemented a simple method that imitates a standard foveated rendering technique. Inspired by recent gaze-contingent rendering techniques [Guenter et al. 2012; Patney et al. 2016], the radius of the high-quality rendering region was set to 6.5°, and the render quality of the periphery was reduced by applying a Gaussian filter with the standard deviation of 0.5°. There was a smooth transition between the peripheral and foveal regions which spanned 9°. Using this rendering technique, we conducted a pairwise comparison experiment in two scenarios: one when our prediction was used and the other one without the prediction. In addition to evaluating our prediction using the personalized model, we also measured the prediction performance of the average model. Furthermore, to investigate how the performance changes for systems with different latency characteristics, we considered three different combinations of eye-tracker sampling frequency and display frame rate settings: (120 Hz, 60 FPS), (120 Hz, 30 FPS), and (60 Hz, 30 FPS). As our display could not run at a native 30 FPS, the frame rate was simulated by frame repetition. Figure 9 (top) shows 10 images used in this experiment.

Nine participants with normal vision and naïve to the purpose of the test performed the experiment with their personalized models. In each trial, participants were shown an image rendered with our foveated rendering method described above. They could switch freely between the version with and without prediction and investigate the quality without any time constraints. The participants were asked to indicate which version provided higher-quality and more stable images. Each participant investigated all ten images, which took approximately 5 minutes. Images were shown in the native resolution of the display and spanned a visual field of 46°×27°. The experiment was performed separately for each combination of display frame rate and eye-tracker sampling frequency settings. 6 participants repeated the whole session with the average prediction model, in addition to the personalized model. Participants were allowed to take breaks between sessions or continue the experiment at a different time if they felt fatigue. The eye tracker was re-calibrated at the beginning of each experiment and when the frame and sampling rate settings were changed.

We collected all the responses and analyzed them using a two-sample t-test with a significance level of 95%. Figure 9 (bottom) shows the ratio of cases where the foveated rendering with our prediction was chosen. In all cases, the prediction led to a higher preference. For all scenarios, the improvement provided by the personalized model over the no-prediction scenario is statistically significant. The personalized model also obtained higher scores than the average model; however, we fail to show a significant difference between the performance of the personalized and the average

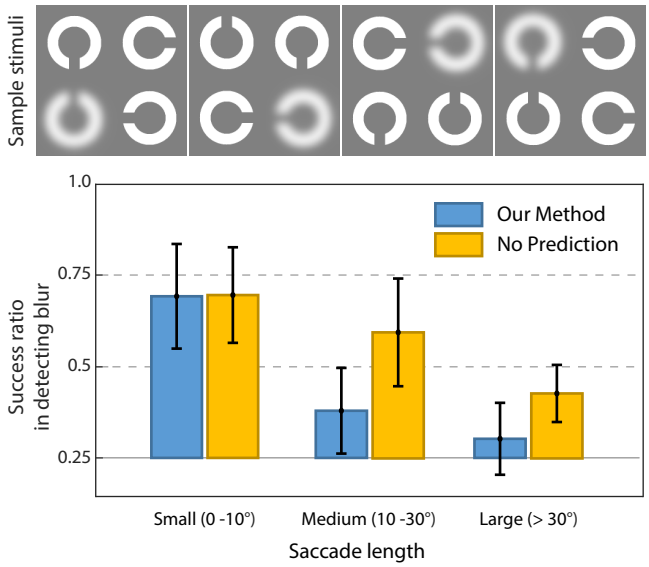


Fig. 8. Top: Four sample synthetic stimuli, each consisting of four Landolt's C shapes. The size of each stimulus was 1.2°. During the experiment, the orientations and the index of the blurry shape was randomly chosen. The blur was removed when the foveated region moved to the position of the stimulus. Bottom: Results of the user experiment. A value closer to 0.25 (expected success ratio with purely random choice) is considered more successful at hiding potential artifacts due to the transition from non-foveated to foveated rendering at the position of the stimulus. The error bars correspond to the standard deviation across participants.

model ($p > 0.45$). This observation, combined with our previous findings (see Figure 7), suggests that the average model is a feasible alternative for the personalized model, even though personalization has the potential to improve the prediction for some of the participants noticeably as shown in Figure 7 (right). We also observed that the advantage of using our techniques decreased for setups with higher latency: (120 Hz, 30 FPS), and (60 Hz, 30 FPS). Although the participants still showed a significant preference for rendering with the prediction ($p < 0.03$), the results are not as remarkable as in the case of (120 Hz, 60 FPS). In fact, the difference for (120 Hz, 30 FPS) and (60 Hz, 30 FPS) setups becomes statistically insignificant for the average model ($p > 0.65$). In an informal interview after the experiment, participants frequently reported "tunnel vision" or sudden "pop-up" effects for both ours and standard foveated renderings in the 60 Hz sampling rate and 30 FPS case. Many people suggested that the instabilities are visible in both techniques, and therefore, the difference between them becomes less obvious. This agrees with the fact that for displays with a lower frame rate, e.g., 30 FPS, an image update requires 33.3 ms in the most optimistic case, which is less than the duration of short saccades. This time might therefore be insufficient to provide image updates before the shorter saccades end. This suggests that for the (120 Hz, 60 FPS) setup our technique successfully overcame the problem of latency. For the systems with higher latency, even though we provided an earlier estimation of the next fixation location, this might not have been early enough to overcome the latency problem completely.

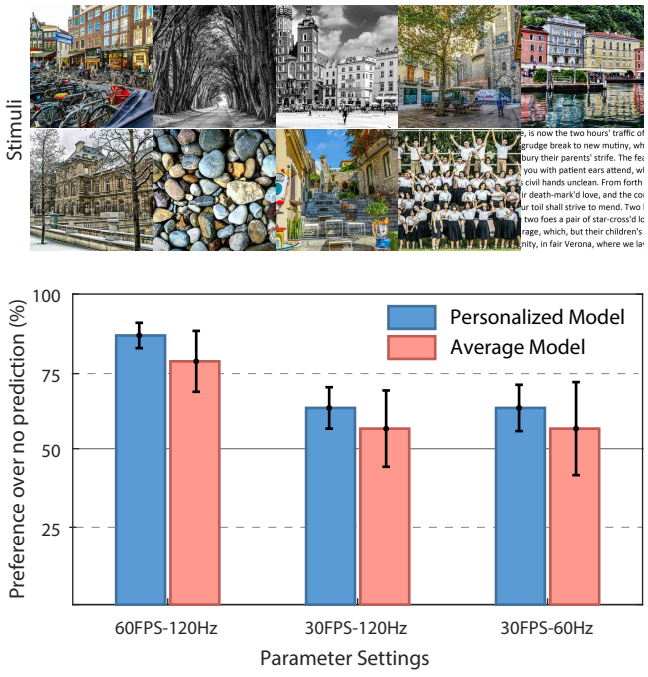


Fig. 9. Top: Insets of the images used in the free-viewing experiment. The size of each image is 2560 × 1440 pixels. Bottom: Results of the free-viewing experiment. The error bars correspond to the standard deviation across participants.

Head-Mounted Display. We also have experimented with a head-mounted system (Oculus DK2 + Pupil Labs' eye-tracker). We ran a preliminary experiment, similar to the one which was used for the stationary system but with stereoscopic images. Five naïve participants took part in this experiment; they compared standard foveated rendering with a method using our personalized prediction derived with the desktop setup (Section 3.3). We repeated the experiment for 75 FPS and 45 FPS. The sampling rate of the eye tracker was 120 Hz. Each participant performed 28 comparisons in total. The rendering with prediction was preferred in 54 % and 69 % of all comparisons for the first and the second setup. Although these results already suggest that our prediction can be directly applied to head-mounted displays, we encountered several hardware problems which, we believe, affect the results. First, the quality and the resolution of the screen, especially the blur towards the boundaries of the visual field introduced by the lenses, make the effect of gaze-contingent rendering more subtle. Second, the quality of the data provided by the eye tracker is very sensitive to tiny movements of the headset with respect to the head of the viewer. In particular, small changes in the relative distance between the observer and the screen introduce errors in the conversion from the on-screen location to visual angles. Moreover, we also observed that our eye tracker often loses the gaze direction and takes a significant amount of time to recover. Some of these problems are already addressed in stationary eye trackers, and we believe that this will also be the case for future HMDs as the technology matures. For these reasons, although the initial results are promising, we leave the application of our prediction strategy to head-mounted setups as future work.

5 DISCUSSION AND FUTURE WORK

As demonstrated in the previous section, our technique can provide significant gains when compared to a standard gaze-contingent rendering. In this section, we discuss potential improvements that could be made based on our assumptions by further extending our model.

Within-subject Variability. Our data-driven prediction model is based on a ballistic saccade approximation, and for a given pair of time stamp and gaze displacement, it returns a single prediction for the saccade amplitude value. However, it is known that there is a statistical within-subject saccade variability [Leigh and Zee 2015, Ch. 3]. For example, the peak velocity of saccades with similar amplitude may depend on the given task, saccade direction, initial and final orbital orientations of the eye, learning, or even on a day-by-day basis [Bollen et al. 1993; Smeets and Hooge 2003]. The overall good performance of our average prediction model shows that, in practical applications, such a saccade variability may be regarded as another source of noise similar to the measurement noise of the eye tracker. We initially experimented to incorporate such factors into our model, but the amount of improvement was observed to be insignificant. In addition, our performance measurements already included the effects of such deviations from the ideal saccade behavior by having participants in our experiments take part in multiple sessions, at different times of day and with different tasks. While inaccurate predictions cannot be removed online, we make an effort to minimize the effect of the conditions which are known to result in unreliable predictions. In such cases, we fall back to the standard gaze-contingent rendering. One example of this is raising the detection threshold to $V_d = 180°/s$ to prevent participant-dependent tremor-like eye motions from triggering a false saccade detection.

Between-Subject Variability. The elimination of between-subject variability by the personalization of the prediction model leads to a further improvement in the prediction accuracy (Figure 7, right). The potential drawback is that the method requires a model-fitting step. At present, we train the personalized models offline. However, our initial experiments indicate that this can be done while running a particular application of gaze-contingent rendering. Such a method could start with our average prediction model and then fine-tune it on the fly using new saccade samples as they are detected. This could also potentially account for some factors that affect within-subject variability, such as fatigue or task-dependent variability. Another advantage of the personalized data-driven model is that it naturally accounts for certain aspects of saccade variability such as corrective glissades [Holmqvist et al. 2011, Ch. 2].

Users with Corrective Glasses. The refractive power of corrective glasses bends the light rays on the way from the screen to the eyeball, which affects the measurements of saccade amplitudes when expressed in terms of visual angle. For example, a basic spherical lens that is used to correct nearsightedness (myopia) minifies the world, which effectively means that the time needed to complete the saccade between a pair of points on the screen is shorter for the same person with glasses than without glasses. This obviously affects the precision of our average prediction model, which was obtained only from participants with normal vision. In an informal

study we observed that our personalized model could compensate for a medium spherical lens correction, while it failed for more complex progressive glasses with an astigmatic component. We relegate to future work the extension of our personalized model to handle diverse prescriptions of corrective glasses. Use of contact lenses does not affect the performance of our prediction model.

Fixation Prediction and Visual Attention. In this work, we focus on predicting the landing position for a single saccade. A significant body of research investigates the saccade planning problem based on the image content and user's task [Kowler 2011], where an attempt at predicting the saccade sequence is made based on visual attention and "saliency map" modeling [Borji and Itti 2013; Katti et al. 2014]. This is a far more complex problem which also involves cognitive and application-dependent aspects, but in principle, the saliency consideration along the saccade trajectory and in the proximity of the predicted landing position could contribute towards an improvement of the prediction accuracy by effectively "snapping" the fixation to the locally most salient feature. We relegate this promising research avenue to future work.

6 CONCLUSION

Gaze-contingent methods promise an improvement in user experience while exploring and interacting with digital content. To provide superior quality, the image updates have to be performed on time according to the current viewer's gaze direction, as any delays may lead to dissatisfaction. In this work, we presented an end-to-end system that uses a saccade landing prediction to combat system latency. The main idea is to take advantage of saccadic suppression and update the image to the new fixation location as soon as the saccade starts. Effectively, such an approach provides an update to the new fixation location before the saccade ends, which leads to less visible delays. To this end, we propose a measurement-driven saccade model that can predict the landing position before the next fixation is established. Our prediction provides better accuracy than existing techniques. Also, an important feature is the continuous prediction refinement as new eye-tracking samples arrive. We applied the model in a simple foveated rendering system and demonstrated significant improvements compared to a system without prediction. A great advantage of our technique is that it comes at almost no additional cost and can be integrated into any existing gaze-contingent system in a straightforward manner. To our knowledge, this is the first work that presents and evaluates a real-time gaze-contingent system with the prediction of saccade landing position.

REFERENCES

Richard Andersson, Linnéa Larsson, Kenneth Holmqvist, Martin Stridh, and Marcus Nyström. 2016. One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior Research Methods* (2016).

James Anliker. 1976. Eye movement: On-line measurement, analysis, and control. *R. A. Monty & J. W. Senders (Eds.), Eye movements and psychological processes* (1976), 185–202.

A Terry Bahill, Michael R Clark, and Lawrence Stark. 1975. The main sequence, a tool for studying human eye movements. *Mathematical Biosciences* 24, 3-4 (1975), 191–204.

Martin S Banks, Allison B Sekuler, and Stephen J Anderson. 1991. Peripheral spatial vision: Limits imposed by optics, photoreceptors, and receptor pooling. *J. Opt. Soc. Am.* 8, 11 (1991), 1775–1787.

Clara Bodelón, Mazyar Fallah, and John H. Reynolds. 2007. Temporal resolution for the perception of features and conjunctions. *Journal of Neuroscience* 27, 4 (2007), 725–730.

D Boghen, BT Troost, RB Daroff, LF Dell'Osso, and JE Birkett. 1974. Velocity characteristics of normal human saccades. *Invest Ophthalmology & Vis Science* 13, 8 (1974), 619–623.

E. Bollen, J. Bax, J.G. Van Dijk, M. Koning, J.E. Bos, C.G. Kramer, and E.A. Van Der Velde. 1993. Variability of the main sequence. *Invest Ophthalmology & Vis Science* 34, 13 (1993), 3700–3704.

A. Borji and L. Itti. 2013. State-of-the-art in visual attention modeling. *IEEE PAMI* 35, 1 (2013), 185–207.

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

Scott J Daly. 1998. Engineering observations from spatiovelocity and spatiotemporal visual models. In *Photonics West'98 Electronic Imaging*. International Society for Optics and Photonics, 180–191.

Mark R. Diamond, John Ross, and M. C. Morrone. 2000. Extraretinal control of saccadic suppression. *Journal of Neuroscience* 20, 9 (2000), 3449–3455.

Michael Dorr, Thomas Martinetz, Karl R. Gegenfurtner, and Erhardt Barth. 2010. Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision* 10, 10 (2010), 28. https://doi.org/10.1167/10.10.28 arXiv:/data/journals/jov/932797/jov-10-10-28.pdf

Mark H. Draper, Erik S. Viirre, Thomas A. Furness, and Valerie J. Gawron. 2001. Effects of image scale and system time delay on simulator sickness within head-coupled virtual environments. *Human Factors* 43, 1 (2001), 129–146.

Andrew T. Duchowski, David Bate, Paris Stringfellow, Kaveri Thakur, Brian J. Melloy, and Anand K. Gramopadhye. 2009. On spatiochromatic visual sensitivity and peripheral color LOD management. *ACM Trans. Appl. Percept.* 6, 2 (2009), 9:1–9:18.

Andrew T. Duchowski, Donald H. House, Jordan Gestring, Rui I. Wang, Krzysztof Krejtz, Izabela Krejtz, Radosław Mantiuk, and Bartosz Bazyluk. 2014. Reducing visual discomfort of 3D stereoscopic sisplays with gaze-contingent depth-of-field. In *Proc. ACM Symp. on Appl. Perc. (SAP)*. 39–46.

LH Frank, JG Casali, and WW Wierwille. 1988. Effects of visual display and motion system delays on operator performance and uneasiness in a driving simulator. *Human Factors* 30, 2 (1988), 201–217.

O-J Grüsser and U Grüsser-Cornehls. 1986. Physiology of vision. In *Fundamentals of Sensory Physiology*. 144–198.

Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. 2012. Foveated 3D graphics. *ACM Trans Graph (Proc SIGGRAPH Asia)* 31, 6 (2012), 164.

Peng Han, Daniel R Saunders, Russell L Woods, and Gang Luo. 2013. Trajectory prediction of saccadic eye movements using a compressed exponential model. *Journal of Vision* 13, 8 (2013), 27–27.

Philippe Hanhart and Touradj Ebrahimi. 2014. Subjective evaluation of two stereoscopic imaging systems exploiting visual attention to improve 3D quality of experience. In *Proc. SPIE vol. 9011*.

Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. 2011. *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

David Jacobs, Orazio Gallo, Emily Cooper, Kari Pulli, and Marc Levoy. 2015. Simulating the visual experience of very bright and very dark scenes. *ACM Trans Graph (TOG)* 34, 3 (2015), 25.

Harish Katti, Anoop Kolar Rajagopal, Mohan Kankanhalli, and Ramakrishnan Kalpathi. 2014. Online estimation of evolving human visual interest. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 11, 1 (2014), 8.

Petr Kellnhofer, Piotr Didyk, Karol Myszkowski, Mohamed M Hefeeda, Hans-Peter Seidel, and Wojciech Matusik. 2016. GazeStereo3D: Seamless disparity manipulations. *ACM Trans Graph (Proc SIGGRAPH)* 35, 4 (2016), 68.

Oleg V Komogortsev and Javed I Khan. 2009. Eye movement prediction by oculomotor plant Kalman filter with brainstem control. *Journal of Control Theory and Applications* 7, 1 (2009), 14–22.

Oleg V Komogortsev, Young Sam Ryu, Do Hyong Koh, and Sandeep M Gowda. 2009a. Instantaneous saccade driven eye gaze interaction. In *Proceedings of the International Conference on Advances in Computer Enterntainment Technology*. ACM, 140–147.

Oleg V Komogortsev, Young Sam Ryu, San Marcos, and Do Hyong Koh. 2009b. Quick models for saccade amplitude prediction. *Journal of Eye Movement Research* 3, 1 (2009).

Eileen Kowler. 2011. Eye movements: The past 25 years. *Vision Research* 51, 13 (2011), 1457–1483.

R John Leigh and David S Zee. 2015. *The neurology of eye movements*. Vol. 90. Oxford University Press, USA.

Lester C Loschky and Gary S Wolverton. 2007. How late can you update gaze-contingent multiresolutional displays without detection? *ACM Trans Multimedia Comput, Comm, and Appl (TOMM)* 3, 4 (2007), 25:10.

Radosław Mantiuk, Bartosz Bazyluk, and Anna Tomaszewska. 2011. Gaze-dependent depth-of-field effect rendering in virtual environments. In *Int Conf on Serious Games*

Dev & Appl. 1–12.

Michael Mauderer, Simone Conte, Miguel A. Nacenta, and Dhanraj Vishwanath. 2014. Depth perception with gaze-contingent depth of field. In *Proc Human Fact in Comp Sys (CHI)*. 217–226.

Craig H Meyer, Adrian G Lasker, and David A Robinson. 1985. The upper limit of human smooth pursuit velocity. *Vision research* 25, 4 (1985), 561–563.

Hunter Murphy and Andrew T. Duchowski. 2001. Gaze-contingent level of detail rendering. *Eurographics Short Presentations* (2001).

Cornelis Noorlander, Jan J. Koenderink, Ron J. Den Olden, and B. Wigbold Edens. 1983. Sensitivity to spatiotemporal colour contrast in the peripheral visual field. *Vision Research* 23, 1 (1983), 1–11.

Oculus VR. 2016a. Asynchronous spacewarp. https://developer.oculus.com/blog/asynchronous-spacewarp/. (2016). Accessed: 2017-04-18.

Oculus VR. 2016b. Asynchronous timewarp. https://developer3.oculus.com/documentation/mobilesdk/latest/concepts/mobile-timewarp-overview/. (2016).

Céline Paeye, Alexander C Schütz, and Karl R Gegenfurtner. 2016. Visual reinforcement shapes eye movements in visual search. *Journal of Vision* 16, 10 (2016), 15–15.

Anjul Patney, Marco Salvi, Joohwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. 2016. Towards foveated rendering for gaze-tracked virtual reality. *ACM Trans Graph (Proc SIGGRAPH Asia)* 35, 6 (2016), 179.

Simon JD Prince and Brian J Rogers. 1998. Sensitivity to disparity corrugations in peripheral vision. *Vision Res* 38, 17 (1998), 2533–2537.

John Ross, David Burr, and Concetta Morrone. 1996. Suppression of the magnocellular pathway during saccades. *Behavioural Brain Research* 80, 1 (1996), 1–8.

John Ross, M.Concetta Morrone, Michael E Goldberg, and David C Burr. 2001. Changes in visual perception at the time of saccades. *Trends in Neurosciences* 24, 2 (2001), 113–121.

Dario D Salvucci and Joseph H Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proc. Symp. on Eye Tracking Res. and Appl. (ETRA)*. 71–78.

DR Saunders and RL Woods. 2014. Direct measurement of the system latency of gaze-contingent displays. *Behavior Research Methods* 46, 2 (2014), 439–447.

Jeroen BJ Smeets and Ignace TC Hooge. 2003. Nature of variability in saccades. *Journal of Neurophysiology* 90, 1 (2003), 12–20.

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

Hans Strasburger, Ingo Rentschler, and Martin Jüttner. 2011. Peripheral vision and pattern recognition: A review. *Journal of Vision* 11, 5 (2011), 13–13.

Nicholas T. Swafford, José A. Iglesias-Guitian, Charalampos Koniaris, Bochang Moon, Darren Cosker, and Kenny Mitchell. 2016. User, metric, and computational evaluation of foveated rendering methods. In *Proc. ACM Symp. on Appl. Perc. (SAP)*. 7–14.

Karthik Vaidyanathan, Marco Salvi, Robert Toth, Tim Foley, Tomas Akenine-Möller, Jim Nilsson, Jacob Munkberg, Jon Hasselgren, Masamichi Sugihara, Petrik Clarberg, and others. 2014. Coarse pixel shading. In *High Performance Graphics*.

AJ Van Opstal and JAM Van Gisbergen. 1987. Skewness of saccadic velocity profiles: A unifying parameter for normal and slow saccades. *Vision Research* 27, 5 (1987), 731–745.

Margarita Vinnikov and Robert S. Allison. 2014. Gaze-contingent depth of field in realistic scenes: the user experience. In *Proc. Symp. on Eye Tracking Res. and Appl. (ETRA)*. 119–126.

Frances C. Volkmann, Lorrin A. Riggs, Keith D. White, and Robert K. Moore. 1978. Contrast sensitivity during saccadic eye movements. *Vision Research* 18, 9 (1978), 1193–1199.

Sang Hoon Yeo, Martin Lesmana, Debanga R Neog, and Dinesh K Pai. 2012. Eyecatch: Simulating visuomotor coordination for object interception. *ACM Transactions on Graphics (TOG)* 31, 4 (2012), 42.

Wei Zhou, Xinnian Chen, and John Enderle. 2009. An updated time-optimal 3rd-order linear saccadic eye plant model. *International Journal of Neural Systems* 19, 05 (2009), 309–330.