



# Gaze-Dependent Screen Space Ambient Occlusion

Radosław Mantiuk<sup>(✉)</sup>

West Pomeranian University of Technology,  
Szczecin al. Piastów 17, 70–310 Szczecin, Poland  
rmantiuk@zut.edu.pl

**Abstract.** The screen space ambient occlusion (SSAO) is a fast global illumination technique, which approximates interreflections between rendered objects. Due to its simplicity, it is often implemented in commercial computer games. However, despite the fact that SSAO calculations take a few milliseconds per frame, a significant computation load is added to the total rendering time. In this work we propose a technique, which accelerates the SSAO calculations using information about observer's gaze direction captured by the eye tracker. The screen region surrounding the observer's gaze position is rendered with maximum quality, which is reduced gradually for higher eccentricities. The SSAO quality is varying by changing the number of samples that are used to approximate the SSAO occlusion shadows. The reduced sampling results in almost two-fold acceleration of SSAO with negligible deterioration of the image quality.

## 1 Introduction

In the Phong reflection model, diffuse and specular reflections are varying due to observer and lights positions, but ambient light is constant. Having this assumptions, we miss the interreflections between rendered objects. Adding *ambient occlusion* (AO) for varying ambient light creates very convincing soft shadows, that combined with direct lighting give realistic images [1, Sect. 9.2]. The AO technique is faster in comparison to the full global illumination solutions, however, it still needs demanding resources to achieve high quality renderings. An approximation of this technique, called *screen space ambient occlusion* (SSAO) [15] simulates local occlusions in real time. However, the accuracy of approximation strongly depends on the sampling density, which in turn limits its applications. It is especially a drawback in the game engine, in which the rendering time spent for AO computation should use only a fraction of the frame time because other calculations determine the quality of the gameplay. Thus, further acceleration of the SSAO computations is a desirable task which has significant impact on the overall quality of the real-time computer graphics.

In this work we present a gaze-dependent screen space ambient occlusion technique in which information about human viewing direction is employed to vary accuracy of the occlusion factors. The screen region surrounding the

observer’s gaze position is rendered with maximum precision, decreasing gradually towards parafoveal and peripheral regions. The idea of this solution is based on the directional characteristic of the human visual system (HVS). People see the high frequency details only in a small viewing angle subtended 2–3° of the field of view. In this range, people see with a resolution of up to 60 cycles per angular degree, but for a 20-degree viewing angle, this sensitivity is reduced even ten times [8].

The accuracy of the occlusion factors is determined by the number of samples used to calculate these factors. In practice, less samples results in ragged edges of the SSAO shadows. However, this aliasing artifacts are barely visible by humans in peripheral regions of vision. In other words, image deterioration caused by low sampling in the SSAO technique can be clearly visible only in high frequency regions. Number of samples in SSAO can be gradually reduced with distance to a gaze point what significantly speeds-up rendering. We use the *eye tracker* to capture the *human gaze point*. Then, sampling is reduced with eccentricity (deviations from the axis of vision) along the curve determined by the gaze-dependent contrast sensitivity function (GD-CSF) [5]. This perceptual function models loss of contrast sensitivity with *eccentricity*. It can be used to determine the maximum special frequency visible for humans for an arbitrary viewing angle.

Section 2 gives background information on the screen-space ambient occlusion technique and outlines the previous work. Section 3 is focused on our gaze-dependent extension of SSAO and shows how sampling can be reduced without noticeable image deterioration. Section 4 presents the results of the perceptual experiment in which we evaluate the perceptual visibility of the image deteriorations.

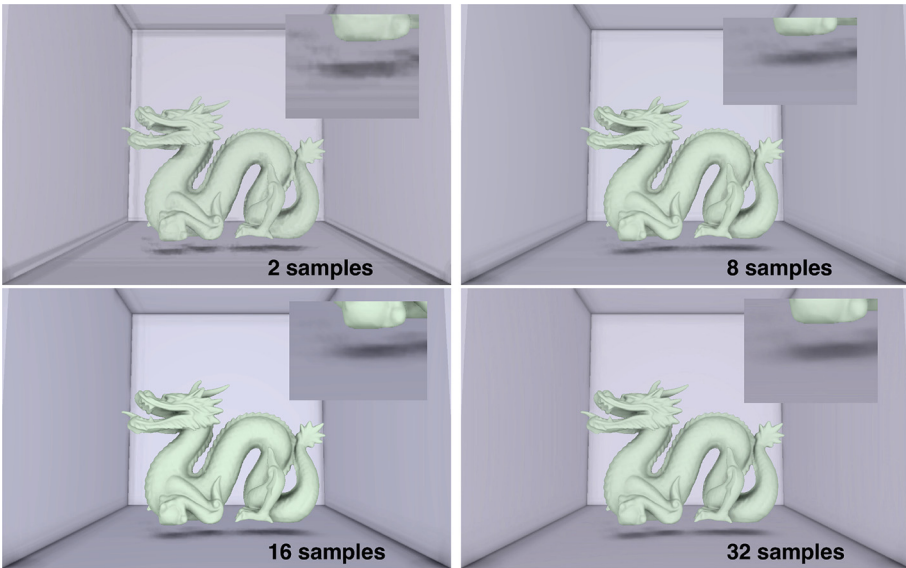
## 2 Background and Previous Work

*Screen space ambient occlusion* (SSAO) is a rendering technique for approximating the global illumination in real time [15]. For every pixel on the screen, the depths of surrounding pixels are analyzed to compute the amount of occlusion, which is proportional to the depth difference between a current pixel and a sampled pixel.

We implemented the *normal-oriented hemisphere* SSAO technique [20]. The hemisphere is oriented along the surface normal at the pixel. The samples from the hemisphere are projected into screen space to get the coordinates into the depth buffer. If the sample position is behind this sample depth (i.e. inside geometry), it contributes to the occlusion factor. The procedure is repeated for every pixel in image to generate the map of occlusion factors (also called the *occlusion shadows*, see example in Fig. 1). This map forms a characteristic shadowing of ambient light, which is visible as a high frequency information in characteristic regions of the scene (e.g. at corners, close to complex objects, etc.). The shadows are blended with the pixel colors computed based on the Phong lighting equation. Wherein, frequency of the shadows is often higher than variability of the Phong ambient shading.



**Fig. 1.** Map of occlusion shadows generated by the normal-oriented hemisphere SSAO technique.



**Fig. 2.** Quality of the occlusion shadows in relation to the number of samples. The image deterioration is especially visible in the insets.

In SSAO the screen space computations are performed rather than tracing new rays in 3-dimensional space as it is done in the original Ambient Occlusion (AO) technique [23]. The AO algorithm generates better results than SSAO, but its complexity prevents the use of AO in the game engines and generally in the real time computer graphics. Even gaze-dependent extension of AO proposed in [9] offers rendering of up to 1–2 frames per second at full GPU load.

The SSAO method introduces a number of visible artifacts like z-fighting caused by the limited resolution of the Z-buffer or unrealistic darkening of the

objects resulting from applying an arbitrary sampling radius around the pixel (see details in [3, 7, 15]). However, the most problematic is a noise and banding in occlusion shadows caused by too few number of samples (see examples in Fig. 2). To avoid noise visibility, hundreds of samples per pixel should be generated. This is too much for the game engines, for which the trade-off between accuracy and computational complexity is required. The number of samples is reduced to 32, while using the bilateral filtering, which is still time consuming. We noticed that fewer samples from these 32, even with the low-pass filtering, results in perceivable quality deterioration of the ambient occlusion shadows and should not be used in the practical applications.

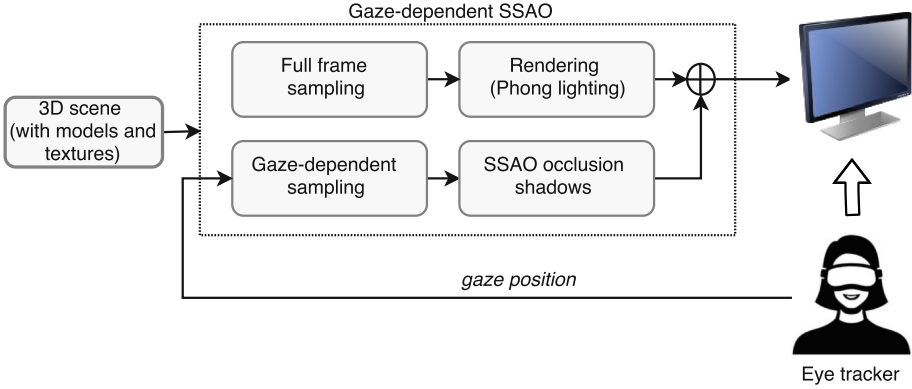
In the following section we propose a technique, which reduces number of samples in the peripheral region of vision without visible degradation of the image quality. This type of image synthesis is called *foveated rendering*. The foveated rendering was proposed to accelerate the ray casting by Murphy et al. [16]. Günter et al. [6] presented a rendering engine, which generates three low-resolution images corresponding to the different fields of view. Then, the wide-angle images are magnified and combined with non-scaled image of the area surrounding the gaze point. Thus, the number of processed pixels can be reduced by 10–15 times, while ensuring the deterioration of image quality invisible for observer. Another foveated rendering technique proposed by Stengel et al. [19] aimed to reduce shading complexity in the deferred shading technique [1]. The spatial sampling is constant for the whole image but the material shaders are simplified for peripheral pixels. According to the authors, this technique reduces the shading time up to 80%. The foveated rendering was also proposed for real time tone mapping [10, 13].

### 3 Gaze-Dependent Rendering of SSAO

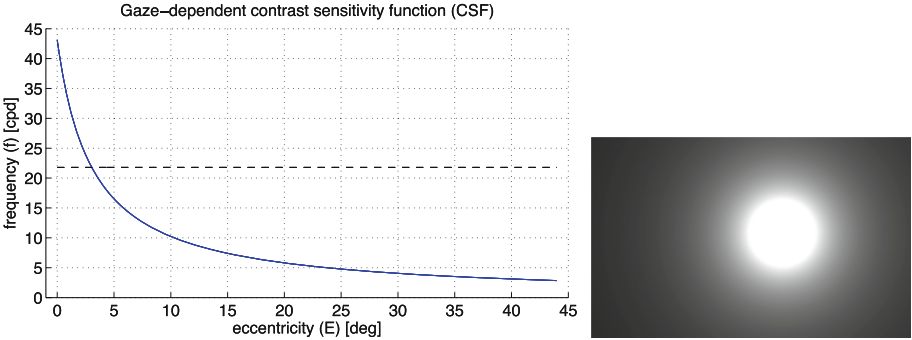
In the gaze-dependent SSAO technique the high frequency spatial sampling is performed in the region of interest. The further from the gaze point, the less detailed ambient factor is rendered saving computation time, while the use of eye tracker leaves observer with a feeling that the sampling is fully detailed. The outline of our gaze-dependent SSAO system is presented in Fig. 3. Observer’s gaze position on the screen of the display is captured by the eye tracker (see Sect. 3.3). At the same time, the 3D scene is rendered using the Phong lighting model. Then, the ambient occlusions are calculated using varying number of samples (see Sect. 3.2). Frequency of sampling depends on the angular distance between a pixel and position of the gaze point (see Sect. 3.1). Finally, the occlusions are blended with the color image and displayed in real time on the screen.

#### 3.1 Gaze-Dependent Contrast Sensitivity Function

The fundamental relationship describing the behavior of the human visual system is the contrast sensitivity function (CSF) [2]. It shows the dependence



**Fig. 3.** Gaze-dependent screen space ambient occlusion rendering system.



**Fig. 4.** Left: the most recognizable stimulus frequency as a function of eccentricity expressed (the dashed line shows the maximum frequency of our display). Right: Region-of-interest mask for an image of  $1920 \times 1080$  pixel resolution (gaze position at  $(1000, 500)$ ), brighter area depicts higher frequency of HVS. The white spot surrounding the gaze position shows an area, in which the maximum resolution of the display is reached.

between the threshold contrast visibility and the frequency of the stimulus. For a frequencies of about 4 cpd (cycles-per-degree), people are the most sensitive to contrast, i.e. they will see the pattern despite the slight differences in the brightness of its individual motifs. The CSF can be used to e.g. better compress the image by removing the high frequency details that would not be seen by humans.

An extension of the CSF, called the gaze-dependent CSF (GD-CSF), is measured for stimuli observed in various viewing angles. Following Peli et al. [5], we model the contrast sensitivity  $C_t$  for spatial frequency  $f$  at an eccentricity  $E$  with the equation:

$$C_t(E, f) = C_t(0, f) * exp(kfE), \tag{1}$$

where  $k$  determines how fast sensitivity drops off with eccentricity (the  $k$  value is ranged from 0.030 to 0.057).  $C_t(0, f)$  is the contrast sensitivity for the foveal vision (equivalent to CSF). The plot of this function is presented in Fig. 4 (left).

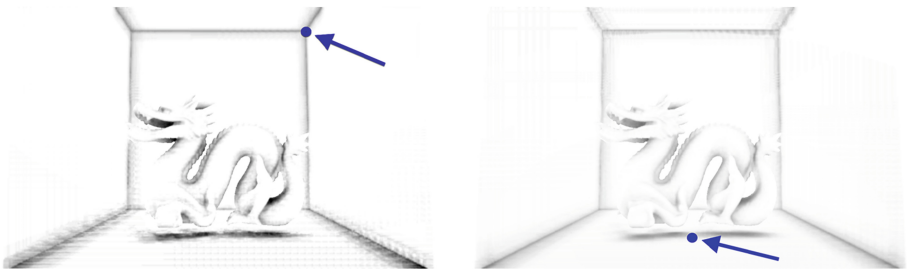
Based on GD-CSF, for a range of eccentricities, the most recognizable stimulus frequency can be modeled by the equation [21]:

$$f_c(E) = E_1 * E_2 / (E_2 + E), \quad (2)$$

where  $f_c$  denotes cut-off spatial frequency (above this frequency observer cannot identify the pattern),  $E_2$  is retinal eccentricity at which the spatial frequency cut-off drops to half its foveal maximum ( $E_1 = 43.1$ ), and  $E_2 = 3.118$  (see details in [22]). An example region-of-interest mask computed for our display based on the above formula is presented in the right image in Fig. 4. Applying this mask, one can sample an image with varying frequency generating less samples for the peripheral regions of vision.

### 3.2 Region-of-Interest Sampling

In the SSAO technique, a number of samples located in the hemisphere oriented along the surface normal at the pixel is analyzed (see Sect. 2 for details). For pixels distant from the gaze point, we reduce a number of samples in the hemisphere according to Eq. 2. For each pixel in the image the eccentricity  $E$  expressed in degrees of the viewing angle is calculated. This transformation must take into account the position of the gaze point as well as physical dimensions of the display, its resolution and viewing distance. Resulting frequency  $f_c(E)$  is normalized to  $\langle 0, 1 \rangle$  and mapped to a number of samples ranging from 2 to 32. The example ambient occlusion maps generated for varying number of samples are presented in Fig. 5.



**Fig. 5.** The occlusion shadows with the gaze-dependent reduction of the sampling frequency. The blue arrows point the location of the gaze points. The deterioration of shadows is clearly visible in the areas farther from the gaze points. (Color figure online)

### 3.3 Eye Tracking

Accuracy of the eye tracking plays a crucial role in our GD-SSAO setup, because even small deviations from the actual gaze position can make the peripheral image deteriorations visible for observer. Eye tracker captures the gaze position indicated by temporary location of the pupil centre [11]. This data must be filtered because saccadic movements of the eye make the gaze position unstable [4]. A typical filtration is based on the *fixation algorithms*, that analyze velocity and/or dispersion of the gaze points and estimates the average gaze position for a time window [17]. However, the fixation techniques are also prone to accuracy errors and cannot be directly used in our system because of flickering they generate [12]. We found that temporal pooling of the fixation points generates satisfactory results. In our setup a 250 Hz eye tracker is used which frequency allowed to average 4 gaze point locations per frame. In cases of persons “incompatible” with the eye tracker (i.e. receiving significant calibration error) we increase the size of the high frequency sampling area by scaling the  $f_c(E)$  value (multiplying by a number greater than one). This solution eliminates visible flickering of the occlusion shadows, however, also reduces the achieved rendering speed-up.

## 4 Experimental Evaluation

The main goal of the experiment was to evaluate how reduction of the SSAO sampling affects the quality of the rendered animation. We wanted to test if the peripheral image deteriorations are visible for observers. In this Section we also present a performance boost achieved due to reduced SSAO sampling.

### 4.1 Stimuli

In the experiment we used the Stanford dragon model<sup>1</sup> enclosed in the 5-walls box and Sibenik cathedral scene<sup>2</sup>. Three camera poses were selected for each scene resulting in 6 different images (see selected shots in Fig. 8). The images were static because we notice that an animation focuses observers’s attention on object movements rather than evaluation of quality of the ambient occlusion effect. Please note, that this assumption leads to more conservative results. The image deteriorations should be less visible in the case of dynamic images because of the visual masking effect.

### 4.2 Procedure

We asked observer to carefully watch two images presented one by one on the screen in random order. One of these images was rendered with the full frame

<sup>1</sup> The Stanford dragon is a test model created with a Cyberware 3030 Model Shop (MS) Color 3D Scanner at Stanford University (100040 triangles).

<sup>2</sup> The Sibenik cathedral is a project by Marko Dabrovic ([www.RNA.HR](http://www.RNA.HR), 80841 triangles).

SSAO with 32 samples per pixel (we called it the reference image). The second image was rendered using our GD-SSAO technique with the gaze point captured by eye tracker. Each image was presented for 10 s and after this time observer were asked to assess the image quality on the 10-points Likert scale ranging from significant deteriorated image (score 0) to excellent quality (score 10). This procedure was repeated for 6 pairs of images twice, resulting in 24 images being evaluated (2 scenes x 3 camera poses x 2 repetitions x 2 images in a pair).

### 4.3 Participants

The experiment was repeated for 9 observers (aged between 21 and 24 years old, 7 males and 2 females). All of them had normal or corrected to normal vision. No session took longer than 10 min. The participants were naïve about the purpose of the experiment. The eye tracker were calibrated at the beginning of the session. Observer did not know if it was used while watching a given image.

### 4.4 Apparatus and Performance Tests

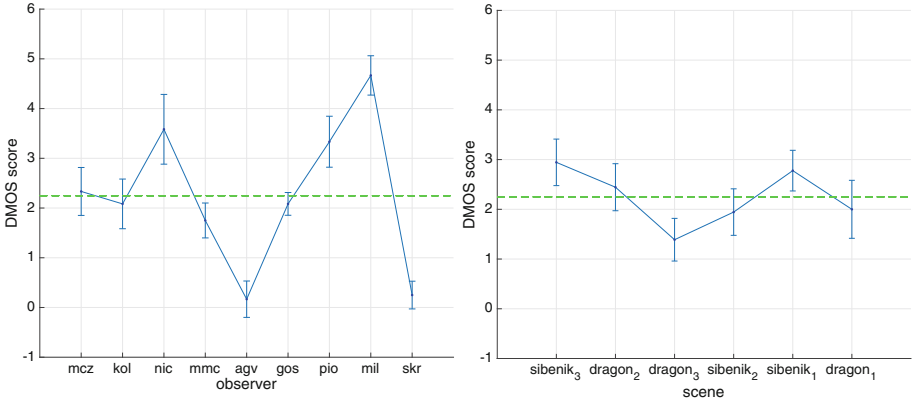
The experiment was conducted in a darkened room. Observers sit in the front of the 22-in. LCD display with the screen dimensions of  $51 \times 28.5$  cm, and the native resolution of  $1920 \times 1080$  pixels. To achieve the rendering framerate of 30 Hz for the full frame SSAO (32 samples per each pixel), we reduced the image resolution to  $1280 \times 720$  pixels. A distance from observer’s eyes to the display screen was restricted to 65 cm with a chin rest. We use SMI RED250 [18] eye tracker working with accuracy close to  $1^\circ$ . Our GD-SSAO renderer was run on PC equipped with 2.66 GHz Intel Xenon W3520 CPU with 8 GB of RAM, Windows 7 64bit OS, and a GPU NVIDIA QUADRO 4000 graphics card.

For the full frame SSAO, our system was able to render 30 fps (frames-per-second) for Sibenik scene, and 35.5 fps for the Stanford Dragon. For GD-SSAO the performance increased to the average frame-rate of 47.7 fps for Sibienik and 54.7 fps for Stanford Dragon (1.59-times and 1.56-times acceleration, respectively).

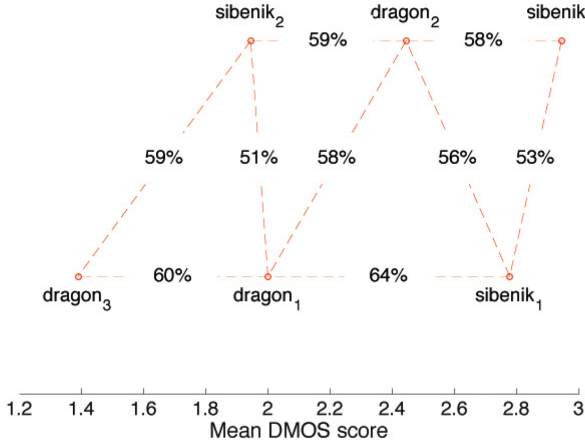
Please note, that the frame-rate depends on location of the gaze point because sampling in image regions corresponding with the complex geometry of the scene is more challenging than for flat regions. For the Stanford Dragon scene the higher frame rates were achieved when observer looks at the corners of the box because the dragon was sampled with the lower frequency. For this scene the frame-rate varied from 48.8 fps to 63.2 fps. For Sibenik scene, it varied from 40.4 fps for observer looking at the centre of the screen, to 52.3 fps for the top-right corner.

It is also worth noting that acceleration in the GD-SSAO technique is related to resolution of the rendered image. Due to hardware limitation, we has to reduce this resolution to  $1280 \times 720$  pixels. For the full HD or 4k resolution displays the performance boost will be correspondingly greater.





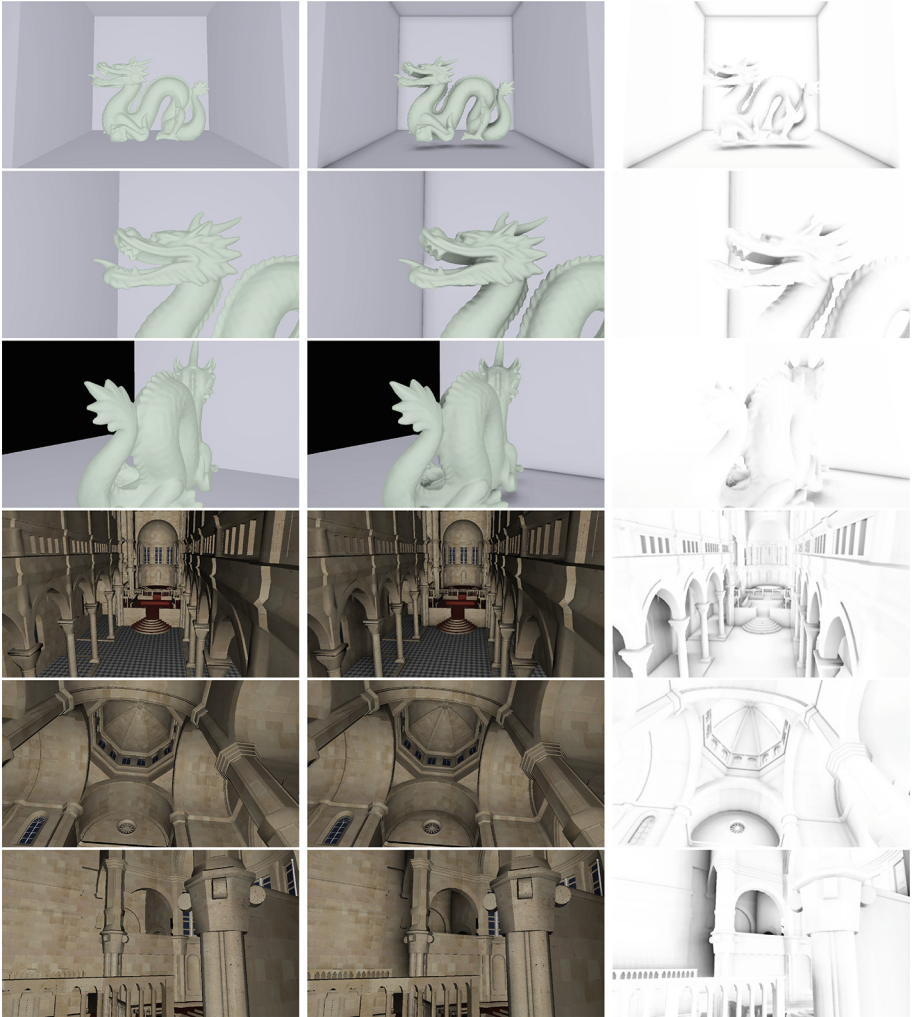
**Fig. 6.** Results of the quality evaluation for individual observers (left) and scene shots (right). The dashed horizontal line depicts average DMOS value. The error bars show standard error of mean. (Color figure online)



**Fig. 7.** Ranking graph illustrating lack of the statistically significant difference between tested scenes.

### 4.5 Quality Evaluation

To evaluate if the peripheral image deterioration was visible for observers we calculated the difference mean opinion score (DMOS) as a difference between the scores given for the reference full-frame SSAO rendering and for GD-SSAO with eye tracker. The score of zero would suggest that observers did not see any difference between techniques, while  $DMOS = 10$  would mean the full disagreement for the GD-SSAO. The DMOS score computed based on the results of our experiment, averaged over all observers and all pairs of images, is equal to 2.25



**Fig. 8.** Images used in the experiment. Left column presents images rendered based on the Phong shading. The ambient occlusions were added in the middle column. Column on the right shows corresponding ambient occlusion maps.

(std = 2.04), which suggests that observers noticed the quality deterioration when using eye tracker but this deterioration was negligible.

Figure 6 shows the DMOS scores for individual observers (left) and individual scene shots (right). The variation of the scores could suggest that there are different opinions between observers and for different scenes. Therefore, we perform the multiple-comparison test, which identifies statistical difference in ranking tests. After [14], the results of this analysis are presented as the ranking of the mean DMOS score for tested scene (see Fig. 7). The scenes are ordered

according to increasing DMOS value, with the smallest DMOS on the left. The percentages indicate the probability that an average observer will choose the scene on the right as better than the scene on the left. If the line connecting two samplings is red and dashed, it indicates that there is no statistical difference between this pair of scenes. The probabilities close to 50% usually result in the lack of statistical significance. For higher probabilities the dashed-lines will start to be replaced with the blue lines but, as can be seen in Fig. 7, the multiple-comparison test confirms lack of the significant statistical difference for all tested scenes.

## 5 Conclusion

We proposed a novel concept of the SSAO technique, in which a rendering speed-up was achieved based on varying sampling of the ambient occlusion shadows. The results of the conducted experiments show that people can experience only a slight deterioration of image quality in comparison to the full frame SSAO. We argue that this deterioration is caused rather by the eye tracking temporal lag than the reduced sampling. In future work we plan to repeat the experiment using better rendering hardware and higher image resolution.

**Acknowledgement.** In this work we used partial results of Andrzej Czajkowski master thesis. We would like to thank you our former student for his excellent work. The project was partially funded by the Polish National Science Centre (grant number DEC-2013/09/B/ST6/02270).

## References

1. Akenine-Möller, T., Haines, E., Hoffman, N.: Real-Time Rendering, 3rd edn. A. K. Peters Ltd., Natick (2008)
2. Barten, P.G.J.: Contrast Sensitivity of the Human Eye and Its Effects on Image Quality. SPIE Press, Bellingham (1999)
3. Bunnell, M.: Dynamic ambient occlusion and indirect lighting. In: GPU Gems 2. Addison Wesley (2005)
4. Duchowski, A.T.: Eye Tracking Methodology: Theory and Practice, 2nd edn. Springer, London (2007). <https://doi.org/10.1007/978-1-84628-609-4>
5. Eli Peli, J.Y., Goldstein, R.B.: Image invariance with changes in size: the role of peripheral contrast thresholds. *JOSA A* **8**(11), 1762 (1991)
6. Guenter, B., Finch, M., Drucker, S., Tan, D., Snyder, J.: Foveated 3D graphics. *ACM Trans. Graph.* **31**(6), 164:1–164:10 (2012)
7. Hegeman, K., Premoze, S., Ashikhmin, M., Drettakis, G.: Approximate ambient occlusion for trees. In: Proceedings of ACM Symposium in Interactive 3D Graphics and Games (I3D 2006), pp. 41–48 (2006)
8. Loschky, L.C., McConkie, G.W., Yang, J., Miller, M.E.: The limits of visual resolution in natural scene viewing. *Vis. Cogn.* **12**, 1057–1092 (2005)
9. Mantiuk, R., Janus, S.: Gaze-dependent ambient occlusion. In: Bebis, G., et al. (eds.) ISVC 2012. LNCS, vol. 7431, pp. 523–532. Springer, Heidelberg (2012). [https://doi.org/10.1007/978-3-642-33179-4\\_50](https://doi.org/10.1007/978-3-642-33179-4_50)

10. Mantiuk, R.: Gaze-dependent tone mapping for HDR video. In: Chalmers, A., Campisi, P., Shirley, P., Olaizola, I. (eds.) *High Dynamic Range Video Concepts, Technologies and Applications*, vol. 10, pp. 189–199. Academic Press (2016)
11. Mantiuk, R.: Accuracy of high-end and self-build eye-tracking systems. In: Kobayashi, S., Piegat, A., Pejaś, J., El Fray, I., Kacprzyk, J. (eds.) *ACS 2016. AISC*, vol. 534, pp. 216–227. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-48429-7\\_20](https://doi.org/10.1007/978-3-319-48429-7_20)
12. Mantiuk, R., Bazyluk, B., Mantiuk, R.K.: Gaze-driven object tracking for real time rendering. *Comput. Graph. Forum* **32**(2), 163–173 (2013). <https://doi.org/10.1111/cgf.12036>, <http://diglib.org/EG/CGF/volume32/issue2/v32i2pp163-173.pdf>
13. Mantiuk, R., Markowski, M.: Gaze-dependent tone mapping. In: Kamel, M., Campilho, A. (eds.) *ICIAR 2013. LNCS*, vol. 7950, pp. 426–433. Springer, Heidelberg (2013). [https://doi.org/10.1007/978-3-642-39094-4\\_48](https://doi.org/10.1007/978-3-642-39094-4_48)
14. Mantiuk, R.K., Tomaszewska, A., Mantiuk, R.: Comparison of four subjective methods for image quality assessment. *Comput. Graph. Forum* **31**(8), 2478–2491 (2012)
15. Mittring, M.: Finding next gen-cryengine 2. In: *SIGGRAPH 2007 Advanced Real-Time Rendering in 3D Graphics and Games Course Notes* (2007)
16. Murphy, H.A., Duchowski, A.T., Tyrrell, R.A.: Hybrid image/model-based gaze-contingent rendering. *ACM Trans. Appl. Percept. (TAP)* **5**(4), 22 (2009)
17. Salvucci, D.D., Goldberg, J.H.: Identifying fixations and saccades in eye-tracking protocols. In: *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (ETRA)*, New York, pp. 71–78 (2000)
18. SMI: RED250 Technical Specification, sensoMotoric Instruments GmbH (2009)
19. Stengel, M., Magnor, M.: Gaze-contingent computational displays: boosting perceptual fidelity. *IEEE Signal Process. Mag.* **33**(5), 139–148 (2016)
20. Vardis, K., Papaioannou, G., Gaitatzes, A.: Multi-view ambient occlusion with importance sampling. In: *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, I3D 2013*, pp. 111–118 (2013)
21. Yang, J., Coia, T., Miller, M.: Subjective evaluation of retinal-dependent image degradations. In: *Proceedings of PICS 2001: Image Processing, Image Quality, Image Capture Systems*, pp. 142–147. Society for Imaging Science and Technology (2001)
22. Yang, J., Qi, X., Makous, W.: Zero frequency masking and a model of contrast sensitivity. *Vis. Res.* **35**, 1965 (1995)
23. Zhukov, S., Iones, A., Kronin, G.: An ambient light illumination model. In: Drettakis, G., Max, N. (eds.) *Rendering Techniques '98. Eurographics*, pp. 45–56. Springer, Vienna (1998). [https://doi.org/10.1007/978-3-7091-6453-2\\_5](https://doi.org/10.1007/978-3-7091-6453-2_5)