Automatic Detection of Shadow Acne and Peter Panning Artefacts in Computer Games

Rafał Piórkowski*
Supervised by: Radosław Mantiuk †

West Pomeranian University of Technology, Szczecin
Poland

Abstract

Contemporary game engines offer an outstanding graphics quality but they are not free from typical graphics artefacts. Essential deteriorations are the shadow acne and peter panning artefacts related to deficiency of the shadow mapping technique. In this work we assess whether the objective image quality metrics (IQMs) are suitable for automatic detection and evaluation of these artefacts. We conduct subjective experiments in which people manually mark the visible local artefacts. Then, the detection maps averaged over a number of observers are compared with results generated by IQMs. We evaluate effectiveness of the mathematically-based objective metric - MSE, and advanced IQMs: S-CIELAB, SSIM, MSSIM, and HDR-VDP-2. The achieved results reveal that MSSIM and SSIM metrics outperforms other techniques and are the most suitable for automatic detection of the shadow acne and peter panning.

Keywords: image quality metrics, game engine artefacts, shadow acne, peter panning, perceptual experiments, image quality

1 Introduction

Graphics artefacts are anomalies found in rendered images. They can significantly degrade an image reception and reduce the overall quality of graphics. Interestingly, contemporary advanced game engines are not free from presence of the visually confusing artefacts. In this work we evaluate two types of such deteriorations: *shadow acne* and *peter panning*. Shadow acne (see Fig. 1) is caused by limited depth resolution of the depth maps used in the shadow maps technique [1, Sect.Shadow map]. This artefact can be reduced applying the bias shift to the depth computation. However, too excessive displacement can cause the discontinuity of shadows, i.e. the peter panning deterioration (see Fig. 3). The latter one does not degrade the graphics quality directly but can be perceived by humans as something unnatural.

*rpiorkowski@wi.zut.edu.pl
†rmantiuk@wi.zut.edu.pl

Our goal in this paper is to find out whether the objective image quality metrics (IQMs) [7] are suitable for detection of the shadow acne and peter panning artefacts. The primary application of this concept is an automatic detection of the artefacts during the game production process. Another important issue is evaluation of the perceptual importance of an artefact. If it is barely visible for human observers its correction can be neglected to save the GPU resources.

The image quality assessment revealed its usefulness in the computer graphics applications. The extensive studies were performed in the area of 3D mesh quality assessment [7]. The mesh simplification causes such artefacts as geometric quantisation noise or texture deteriorations. The first attempts to evaluate the visual fidelity of these types of artefacts were simple geometric distance metrics [12]. The advanced IQMs were also tested with the conclusion that better detection of the mesh simplification deterioration can be achieved using the model-based metrics [10]. A comprehensive review of other assessment techniques in this field has been published in [6].

Rushmeier et al. [11] studied the effectiveness of replacing geometric detail with texture maps as a method of simplification. They used a psychophysical scaling procedure to measure the perceived fidelity of simplified geometry and textures relative to the reference representation. They focused on a user study and analysis of its results rather than using the objective metrics.

In [2] a quality metric for stereoscopic images was proposed. It combines the typical 2D image quality metric (SSIM or C4) with the depth information. Another idea was presented in [9], in which the depth map is compressed based on the results of the visual masking experiment. Differences in depth, which are invisible to the human and not caused the visible artefact in the stereo images, are masked out to reduce the size of the depth map.

We focus on the static artefacts that are visible in a separate frame of the game animation. Even more prominent are the artefacts occurring in the temporal domain, that cause the flickering. Analysis of this type of deteriorations requires different quality metrics and a separate experimental methodology (see examples in [13, 6]). We address this issue to future work.

In this paper we describe conducted subjective experiment in which observer manually marks the visible local artefacts in snapshots from the games. It is done in the presence of the reference image without artefacts (ground truth). Hence, we perform the full-reference experiment compatible with the full-reference IQMs. This methodology follows the technique introduced in [5] and [4]. However in the mentioned papers, Čadík et al. [5] evaluate artefacts caused by limitation of the global illumination rendering techniques. The shadow acne and peter panning have a different characteristic and are common for the real-time rendering systems. We evaluate if they are identified by simple arithmetic difference (MSE), colour difference metric S-CIELAB [18], texture statistics SSIM [17], MSSIM [15], and metric based on perceptual models HDR-VDP-2 [8].

The achieved results reveal that MSSIM and SSIM metrics outperform other techniques and are the most suitable for automatic detection of the shadow acne and peter panning artefacts.

The paper is organised in the following way. In Sect. 2 the shadow acne and peter panning are outlined. Sect. 3 presents details on the conducted perceptual experiments. We compare the detection maps marked by human observers with the maps generated by IQMs in Sect. 4. In this section we also briefly describe the advantages of individual objective metrics (Sect. 4.2). The paper ends with conclusions and providing directions for further work in Sect. 5.

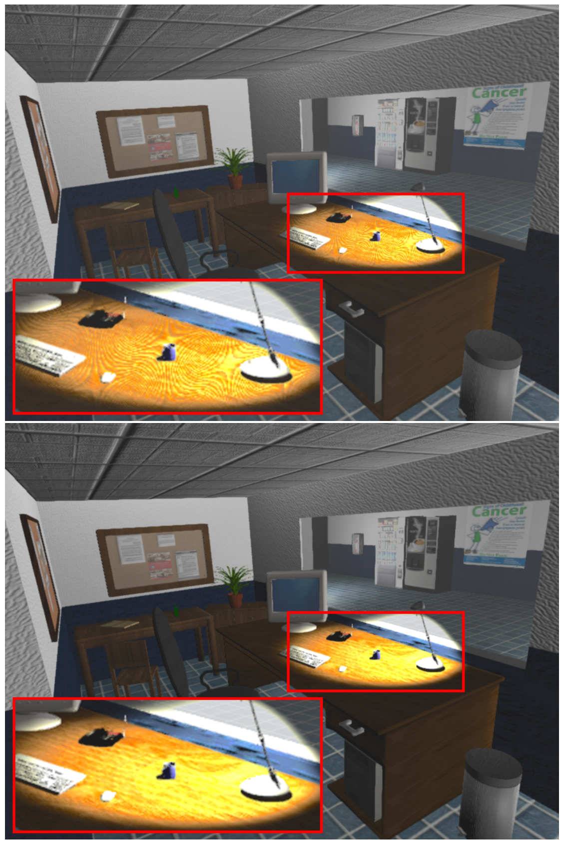

Figure 1: Shadow acne artefacts (top) and the corrected frame (bottom). The inset depicts the surface which is self-shadowed but it should not be shadowed at all.

2 Shadow acne and peter panning

In this section two prominent graphics artefacts are presented: shadow acne and peter panning. We discuss the reasons for their occurrence in the graphics engines and how to prevent them from occurring.

Shadow acne

Shadow acne also called *erroneous self-shadowing*, may occur when shadow depth map algorithm [1, Sect.Shadow map] is used in order to add shadows into the scene in the real-time graphics engine. This artefact manifests itself as moire patterns on surfaces (see Fig. 1). The shadow maps technique consists of two passes. In first, scene is rendered from the light source point of view. Information about the distances between light source and objects is stored as texture called *shadow map*. Those distances are called depth. The more distant is the object from the light source the brighter is texel in the shadow map. During the second pass, when the scene is rendered from the camera point of view the location of each pixel is compared to the corresponding texel in the shadow map. If a rendered point is farther away from the light source than the corresponding value in the shadow map, that point is in the shadow, otherwise it is not.

The shadow acne artefact can be caused by two factors. The first reason is a limited computation precision of the depth maps. When both depth values compared in the second phase of the shadow map algorithm are close to each other, the depth test may fail for selected pixels. The second issue is geometrical - shadow map quantises the depth over an entire texel (see Fig. 2), while a surface is smooth. Due to this fact, a depth test can erroneously give over- and under-the-surface result for the same texel, resulting in self-shadowing.

The most common method to mitigate shadow acne artefact is adding small value - *bias* into light space when depth test is computing (see Fig. 2).

Peter panning

Peter panning is another artefact connected with shadow depth maps algorithm (see Fig. 3). This term derives from the book character - Peter Pan, who could fly and his shadow was detached from body. When this artefact occurs, shadow is detached from the object which seems to hover above surface. Peter panning appears when too large bias is used to prevent the shadow acne occurrence.

Finding proper bias value in order to simultaneously avoid shadow acne and peter panning artefacts for whole

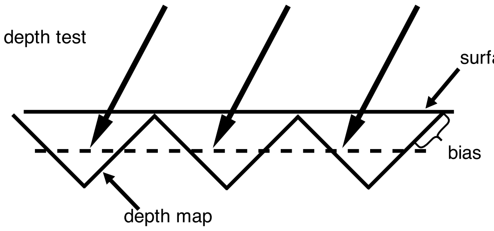

Figure 2: Depth values are stored in the depth map with limited spatial precision. Adding *bias* shift prevents erroneous depth tests.

scene and each frame could be computationally expensive and affect on performance.

3 Experimental study

The goal of the experiment was to create the reference maps that identify the artefacts seen by the people in the game screenshots.

3.1 Stimuli

Even the most prominent and popular graphics engines are not free from the rendering artefacts. We selected three contemporary graphics engines that deliver the development environment for independent developers: Unity 3d [1], CryEngine 3 [2], and Unreal Engine 4 [3]. In these engines it was possible to model a scene based on external graphics objects or some examples delivered with the engine. Then we ran the game changing the rendering parameters. In particular, the shadow mapping was activated with different bias levels to test the shadow acne and peter panning deteriorations.

We modelled 20 different scenes. We used the static camera to avoid motion in the scene. Scene objects and game engine parameters were combined in a way resulted in the appearance of shadow acne or peter panning artefacts. It was done using the *bias* coefficient that was set to too low value for the shadow acne and too high for the peter panning. In our stimuli for 10 scenes we forced the shadow acne and in remaining 10 the peter panning. For each scene the reference image was generated with the correct *bias*. In real-world games it is often challenging to automatically find a correct *bias*, which can differ for various scenes and even various camera shots.

The screenshots of the scenes were captured using the FRAPS program [4], which saved images in 800x600 pixel resolution.

[1] `http:\\www.unity3d.com`
[2] `http:\\www.cryengine.com`
[3] `http:\\www.unrealengine.com`
[4] `http:\\www.fraps.com`

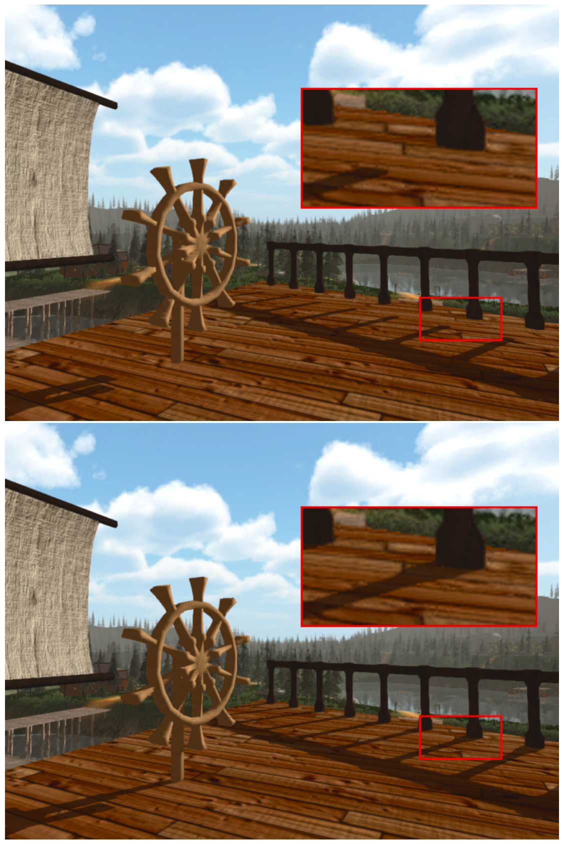

Figure 3: Peter panning artefact (top) and the reference frame (bottom). The shadow discontinuity can be seen in shadows cast by the railing posts.

3.2 Experimental procedure

We asked people to manually mark visible differences between the reference image and an image with a particular artefact. Observers used a custom brush-paint interface controlled by the computer mouse. The brush size could be reduced up to per-pixel resolution. This procedure was repeated for every scene, resulting in 20 comparisons and finally 20 binary difference maps generated per observer.

The experiment was performed in a darkened room. Images were displayed on 24” Eizo ColorEdge CG245W monitor with native resolution of 1920 x 1200 pixels. This display is equipped with the hardware colour calibration module and was calibrated before each experimental session to sRGB colour profile with the maximum luminance level increased to 110 cd/m2. During the experiment, an observer was sitting in front of the display at a distance of 70 cm. This distance was not stabilized by a chin rest but we asked observers to keep it approximately constant.

3.3 Participants

We repeated the experiment for 25 volunteer observers (age between 20 and 23 years, 23 males and 2 females). They declared normal or corrected to normal vision and

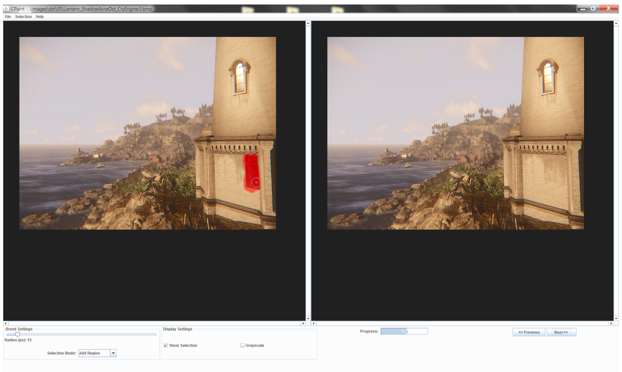

Figure 4: Screenshot from the experiment. Notice the red vertical mark in the left image made by an observer.

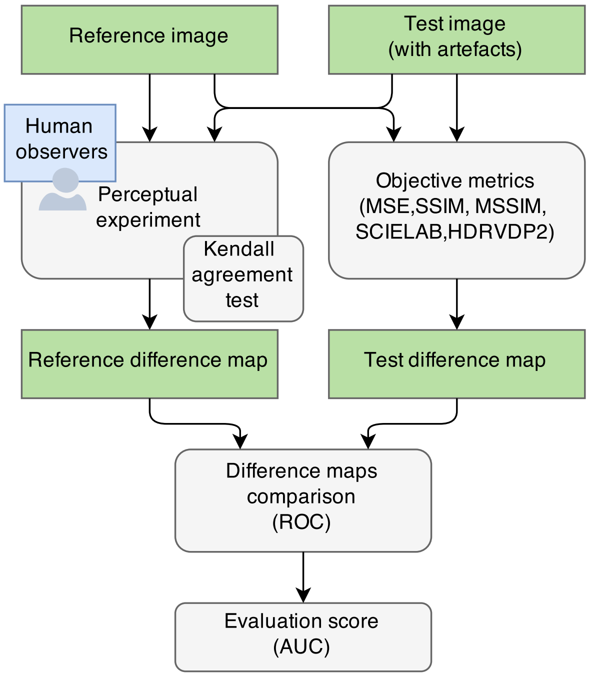

Figure 5: Evaluation procedure.

correct colour vision. The participants were aware that the image quality is evaluated, but they were naïve about the purpose of the experiment.

The experiment is time consuming, therefore we clustered the stimuli images into packages consisting of 10 pairs of images (tested and reference). While there were no time limitations to our study, the average subject finished marking a package in approximately 15 minutes.

4 Results

A goal of the experiment was to test which of the full-reference IQMs is the most suitable for testing the game engine artefacts. We achieved the *reference difference maps* from results of the perceptual experiment described in Sect. 3 and analysed in Sect. 4.1. These results were compared with the *test reference maps* generated by IQMs in Sect. 4.2.

Both reference and test maps are compared using the receiver-operator-characteristic (ROC) technique and their coherence was expressed as the Area Under Curve (AUC) value (see Sect. 4.3). The whole procedure is outlined in Fig. 5.

4.1 Reference difference maps

Example difference maps created by a single observer during the experiment are presented in Fig. 6. The white background represents untouched pixels while the pixels marked by observer are drawn in grey. Latter pixels depict the areas in the test image recognised as artefacts by the human observer.

Kendall analysis

The Kendall rank correlation coefficient (or Kendall's tau (τ)) is a statistic used to measure the association between two measured quantities. In our case, it assesses the inter-observer agreement, i.e. the similarity of the difference maps created by individual observers. As shown by Cadic et al. [5], we used the τ value to assess whether people marked similar areas for a given pair of test and reference images. The coefficient τ ranges from $\tau = -1/(o-1)$, which indicates no agreement between o observers, to $\tau = 1$ indicating that all observers responded the same. Examples of the coefficient maps are shown in Fig. 7.

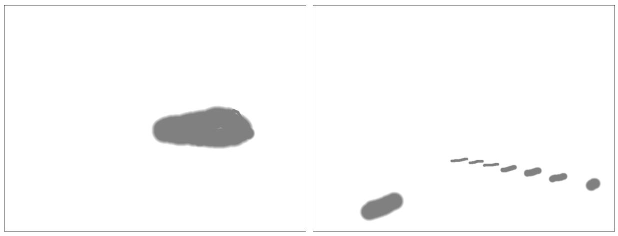

Figure 6: Example difference maps created by a single observer for images with the shadow acne (left) and peter panning (right) artefacts.

We computed average coefficients $\overline{\tau}$ for each scene containing artefacts. However, these values tend to skewed toward very high values because most pixels did not contain any distortion and were consistently left unmarked by all observers. Therefore, we also compute a $\overline{\tau}_{masked}$, which considers only those pixels that were marked as distorted by at least two observers. We achieved $\overline{\tau}$ equal to 0.97 and $\overline{\tau}_{masked}$ to 0.44, averaged over all scenes. These values indicate a high inter-observer agreement. For comparison, in the similar experiment described in [5] and assumed as a high consistent, the $\overline{\tau}$ and $\overline{\tau}_{masked}$ equaled to 0.78 and 0.41, respectively.

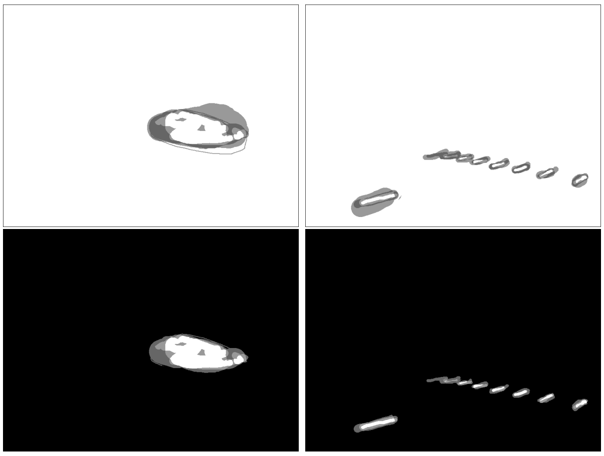

Figure 7: Examples of Kendall coefficient maps (top row) and Kendall maps after masking (bottom row). The white pixels depict good agreement between observers. The maps correspond to shadow acne example from Fig. 1 (left column), and peter panning from Fig. 3 (right column).

Averaged difference maps

We averaged the difference maps related to individual test image over all observers to achieve the reference difference maps. Then, these maps were binarised with the 0.5 threshold. In other words, the pixels marked by 50% of observers were set to 1 and remaining pixels to 0. This thresholding gives reliable result during further statistical analysis, because it eliminates strong deviations in markings. Example reference difference maps are shown in Fig. 8.

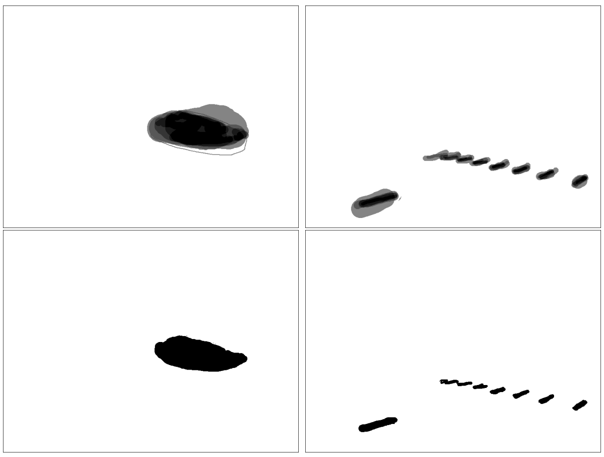

Figure 8: Example difference maps averaged over all observers before binarization (top row) and after (bottom row). Presented artefacts are shadow acne (left column) and peter panning (right column).

4.2 Objective metrics

Objective Image Quality Metrics (IQM) deliver quantitative assessment of the perceptual quality of images [14][16]. In our studies we chose four representative IQMs: S-CIELAB (Spatial-CIELAB), SSIM (Structural SIMilarity Index), MSSIM, and HDR-VDP-2 (High Dynamic Range Visual Difference Predicator) that prove their efficacy in perceptual comparison of images. Additionally, we evaluated the results of the MSE metric to give a background for comparison. The S-CIELAB metric [18] is a spatial extension of standard CIELAB colour difference. SSIM [17] and MSSIM [15] detect structural changes in the image. They are sensitive to difference in the mean intensity and contrast but the main factors are local correlations of pixel values. These dependencies carry information about the structure of the objects and reveal structural image difference between tested and reference images. HDR-VDP-2 [8] predicts the quality degradation expressed as a mean option score of the human observers and visibility (detection/discrimination) of the differences between tested and reference images. It takes into account the contrast sensitivity function measured for variable background luminance and spatial frequencies. The sensitivity to light is modelled separately for cones and rods resulting in correct prediction for mesopic and scotopic light conditions.

For each IQM, we generated 20 test difference maps that were compared to corresponding reference difference maps. Example maps computed using each mentioned metric are presented in Fig. 10 and Fig. 9.

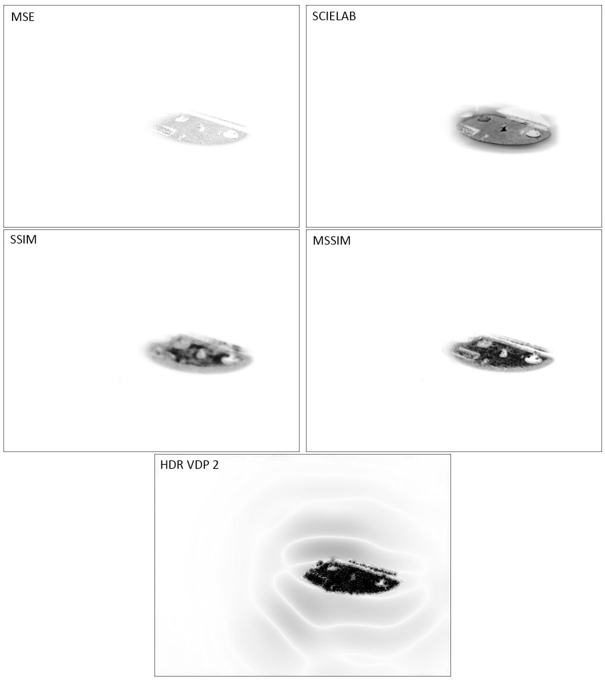

Figure 9: Difference maps automatically generated by IQMs for the shadow acne artefact.

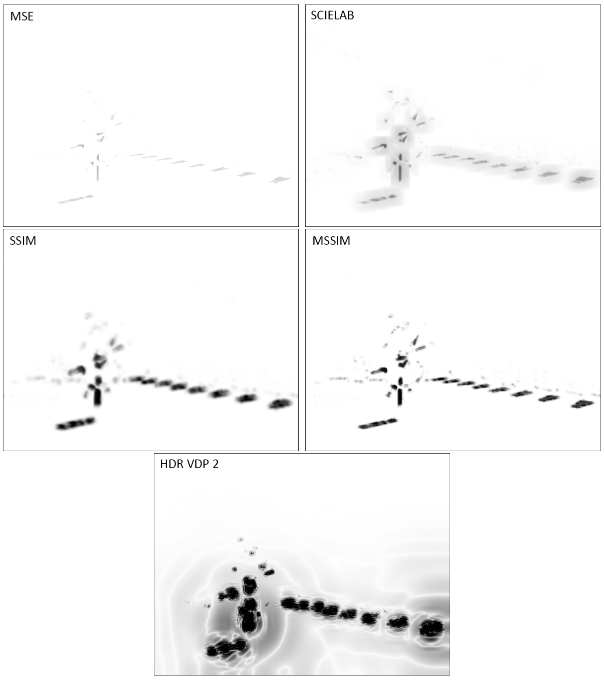

Figure 10: Difference maps automatically generated by IQMs for the peter panning artefact.

4.3 ROC analysis

The key question is whether any of the IQM performs significantly better than the others in terms of detecting the shadow acne and peter panning artefacts. In our experiment, observers binary classified pixels that contained artefacts. The performance of such classification can be analysed using the receiver-operator-characteristic (ROC) [3]. ROC captures the relation between the size of artefacts that were correctly marked by a IQM (true positives), and the regions that do not contain artefacts but were still marked (false positives). The metric that produces a larger *area under the ROC curve* (AUC) is assumed to perform better.

The ROC plots for individual metrics are presented in Fig. 11. We achieved the best results for the SSIM and MSSIM metrics (AUC=99.38 and 98.69, respectively). Interestingly, simpler metric SCIELAB works comparatively well (AUC=99.33). As it was expected MSE gave the worst results with AUC=89.34. We achieved also poor result for HDR-VDP-2 (AUC=90.26). This metric incorrectly detected artefacts covering the large areas, which is common e.g. for the peter panning. Analysis of this issue we address for further work.

5 Conclusions

In this work we asked group of observers to find local artefacts in images rendered by the real time game engines. We focused on two types od artefacts: shadow acne and peter panning, accompanying the shadow mapping technique. The reference difference maps derived from this perceptual experiment were compared with the difference maps generated by the most recognised objective image quality metrics. The ROC analysis revealed the best accuracy of the SSIM metric with the effectiveness of detection close to 100%.

In future work we plan to analyse other artefacts, especially aliasing and z-fighting. We are also interested in the flickering resulting from the motion on the scene. However, localised analysis of such artefacts seems to be challenging.

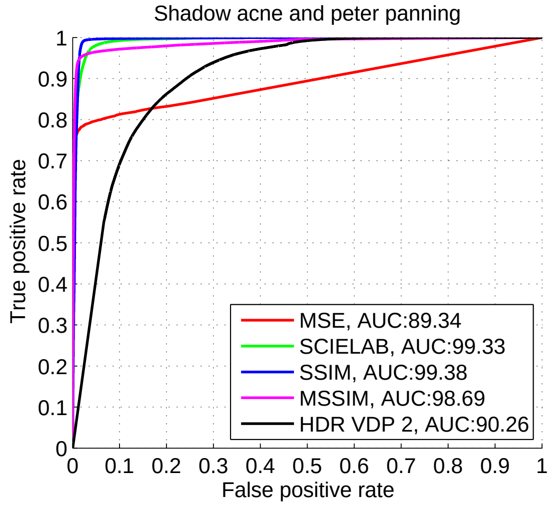

Figure 11: Results of the ROC analysis for individual IQMs.

References

[1] Tomas Akenine-Möller, Eric Haines, and Naty Hoffman. *Real-Time Rendering*. A K Peters, Ltd., Wallesley, Massachusetts, third edition, 2008.

[2] Benoit Alexandre, Le Callet Patrick, Campisi Patrizio, and Cousseau Romain. Quality assessment of stereoscopic images. *EURASIP journal on image and video processing*, 2008, 2009.

[3] P. Baldi, S. Brunak, Y. Chauvin, C. A. F. Anderson, and H. Nielsen. Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics*, 16(5):640–648, 2000.

[4] Martin Čadík, Robert Herzog, Rafał Mantiuk, Radosław Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. Learning to predict localized distortions in rendered images. In *Computer Graphics Forum*, volume 32, pages 401–410. Wiley Online Library, 2013.

[5] Martin Čadík, Robert Herzog, Rafał Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. New measurements reveal weaknesses of image quality metrics in

evaluating graphics artifacts. *ACM Transactions on Graphics (TOG)*, 31(6):147, 2012.

[6] Massimiliano Corsini, Mohamed-Chaker Larabi, Guillaume Lavoué, Oldřich Petřík, Libor Váša, and Kai Wang. Perceptual metrics for static and dynamic triangle meshes. In *Computer Graphics Forum*, volume 32, pages 101–125. Blackwell Publishing Ltd, 2013.

[7] Guillaume Lavoué and Rafał Mantiuk. Quality assessment in computer graphics. *Visual Signal Quality Assessment*, pages 243–286, 2015.

[8] Rafal Mantiuk, Kil Joong Kim, Allan G. Rempel, and Wolfgang Heidrich. Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph.*, 30(4):40:1–40:14, July 2011.

[9] Dawid Pajak, Robert Herzog, Radosław Mantiuk, Piotr Didyk, Elmar Eisemann, Karol Myszkowski, and Kari Pulli. Perceptual depth compression for stereo applications. *Computer Graphics Forum*, 33(2):195–204, 2014.

[10] Bernice E Rogowitz and Holly E Rushmeier. Are image quality metrics adequate to evaluate the quality of geometric objects? In *Photonics West 2001-Electronic Imaging*, pages 340–348. International Society for Optics and Photonics, 2001.

[11] Holly E Rushmeier, Bernice E Rogowitz, and Christine Piatko. Perceptual issues in substituting texture for geometry. In *Electronic Imaging*, pages 372–383. International Society for Optics and Photonics, 2000.

[12] William J Schroeder, Jonathan A Zarge, and William E Lorensen. Decimation of triangle meshes. In *ACM Siggraph Computer Graphics*, volume 26, pages 65–70. ACM, 1992.

[13] L. Vasa and V. Skala. A perception correlated comparison method for dynamic meshes. *IEEE Trans. on Visualization and Computer Graphics*, 17(2):220–230, 2011.

[14] Zhou Wang and Alan Bovik. *Modern Image Quality Assessment*. Morgan & Claypool Publishers, 2006.

[15] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, volume 2, pages 1398–1402. Ieee, 2003.

[16] H. Wu and K. Rao. *Digital Video Image Quality and Perceptual Coding*. CRC Press, 2005.

[17] Wang Z., Bovik A.C., Sheikh H.R., and Simoncelli E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions On Image Processing*, 13(4):600–612, 2004.

[18] X. M. Zhang and B. A. Wandell. A spatial extension to cielab for digital color image reproduction. *Proceedings of the SID Symposiums*, pages 731–734, 1996.