Dataset and metric for predicting visible differences Suplemental material

KRZYSZTOF WOLSKI, MPI Informatik
DANIELE GIUNCHI, University College London
NANYANG YE, University of Cambridge
PIOTR DIDYK, Saarland University, MMCI, MPI Informatik, Università della Svizzera italiana, Switzerland
KAROL MYSZKOWSKI, MPI Informatik
RADOSŁAW MANTIUK, West Pomeranian University of Technology, Szczecin
HANS-PETER SEIDEL, MPI Informatik
ANTHONY STEED, University College London
RAFAŁ K. MANTIUK, University of Cambridge

Additional Key Words and Phrases: visual perception, deep learning, artifact detection, dataset collection, visual metric

ACM Reference Format:
Krzysztof Wolski, Daniele Giunchi, Nanyang Ye, Piotr Didyk, Karol Myszkowski, Radosław Mantiuk, Hans-Peter Seidel, Anthony Steed, and Rafał K. Mantiuk. 2018. Dataset and metric for predicting visible differences Suplemental material. *ACM Trans. Graph.* n, n, Article n (February 2018), 14 pages. https://doi.org/nnnnnnnn.nnnnnnnn

Section 1 provides additional information about dataset, including detailed description of distortions and visual examples of problematic areas. Section 2 consists of images used for compression application and more results of watermark application. Finally, in Section 3 we provide all metrics' coefficients which are results of training with k-fold cross validation procedure.

1 DATASET

The dataset consists of 557 low dynamic range (LDR) images with 170 unique scenes. Many of them are generated for up to 3 distortion levels. The scenes were selected to cover many common and specialized computer graphic artifacts such as noise, image compression, shadow acne etc. This variety makes our data challenging for the state-of-the-art image quality metrics. During the selection process that aimed to generate a large database collecting images from different sources, we gathered scenes with multiple artifact types in a specific category named mixed (Section 1.1). In this way we avoided ambiguities in all the other set or possible hiding effects by one artifact over another one. To make data more clear, we organized the images into the subsets that are summarized in Table 1. At the end of this section, we present all the images from the dataset except TID2013 subset due to big volume and lack of locality in images. Full datatset will is available at: https://doi.org/10.17863/CAM.21484/.

Authors' addresses: Krzysztof Wolski, MPI Informatik, P.O. Box 1212, Saarbrücken, 66123, kwolski@mpi-inf.mpg.de; Daniele Giunchi, University College London, Gower Street, London, WC1E 6BT, d.giunchi@cs.ucl.ac.uk; Nanyang Ye, University of Cambridge, 15 JJ Thomson Avenue, Cambridge, CB3 0FD, yn272@cam.ac.uk; Piotr Didyk, Saarland University, MMCI, MPI Informatik, Università della Svizzera italiana, Via G. Buffi 13, Lugano, CH-6900, Switzerland, pdidyk@mmci.uni-saarland.de; Karol Myszkowski, MPI Informatik, P.O. Box 1212, Saarbrücken, 66123, karol@mpi-sb.mpg.de; Radosław Mantiuk, West Pomeranian University of Technology, Szczecin, al. Piastów 17, Szczecin, 70-310, rmantiuk@wi.zut.edu.pl; Hans-Peter Seidel, MPI Informatik, P.O. Box 1212, Saarbrücken, 66123, hpseidel@mpi-inf.mpg.de; Anthony Steed, University College London, Gower Street, London, WC1E 6BT, a.steed@ucl.ac.uk; Rafał K. Mantiuk, University of Cambridge, 15 JJ Thomson Avenue, Cambridge, CB3 0FD, rafal.mantiuk@cl.cam.ac.uk.

0730-0301/2018/2-ARTn $15.00
https://doi.org/nnnnnnnn.nnnnnnnn

1.1 Mixed

This set consists of 59 images that come from LOCCG data set [Čadík et al. 2012] and contain more than one distortion type. The distortions include high-frequency and low-frequency noise, structured noise, virtual point lights (VPL) artifacts, clamping, downsampling, blurring and light leaking artifacts. One example of those artifacts is presented in Figure 1.

High frequency noise is the common artifact in many global illumination methods e.g. ray tracing, path tracing, radiosity etc. It is caused by low number of samples and appears usually in shadowed areas. Some types of materials cause high-frequency noise more often than the others. We can see it when scene consists of high-glossy surfaces, curved transparent objects that scatter the light, or translucent surfaces that need a lot of light samples to be rendered properly.

Structured noise is a distortion that results from correlated pixel errors. Both noise and bias are showed. Instant radiosity [Keller 1997], photon mapping [Jensen 2001] and radiance caching algorithm [Ward et al. 1988] [Krivánek et al. 2005] can exhibit interpolation and caching artifacts.

VPL is one of many global illumination methods. In the first pass rays are cast from the light source. In the intersection point between ray and geometry a new virtual point lights are generated. Therefore the scene is rendered multiple times for every light generated in the first pass. Because the virtual lights are on the surfaces, the intensity of the light in their 3D surroundings is higher. This causes local brightness changes and low-frequency noise, that spoil the overall look of the image. Due to its high computational complexity, this

Subset name	Scenes	Images	Distortion levels	levels generation method	Resolution [px]	Source
Mixed	20	59	2-3	blending	800×600	custom software, real photos
Perception patterns	12	34	1,3	blending	800×800	matlab
Aliasing	14	22	1-3	varying sample number	800×600	Unity, CryEngine
Peter panning	10	10	1	n/a	800×600	Unity, CryEngine
Shadow map downsampling	9	27	3	varying shadow map resolution	800×600	Custom OpenGL app
Shadow acne	9	9	1	n/a	800×600	Unity, CryEngine
Z fighting	10	10	1	n/a	800×600	Unity, CryEngine
Compression	25	71	2-3	varying bit-rates	512×512	real photos
Deghosting	12	12	1	n/a	900×900	real photos
IBR	18	36	1,3	varying distance between key frames	960×720	custom software
CGIBR	6	6	1	n/a	960×720	custom software
TID2013	25	261	n/a	n/a	512×384	real photos

Table 1. Dataset details.

method is not used in real time computer graphics but in production renderings (movies, animations, architectural visualizations).

Light leaking is one of the photon mapping artifact and appears like an area clearly brighter than normal. It depends on illumination of particular geometries in the scene, like corners in a room or related to specific attributes of a material like smoothness, showing reflected light even if the object is closed off by other geometries.

These last two techniques (VPL and photon mapping) are classified as approximate global illumination algorithms. Locally they inject errors, sometimes deliberately to camouflage more evident artifacts. Artifacts like VPL clamping in instant radiosity, light leaking in photon mapping and irradiance caching belong to this set.

Fig. 1. Example of high frequency noise, common artifact in VPL method.

1.2 Perception patterns

Contrast-Luminance-Frequency-Masking (CLFM) dataset consists of 34 images from [Čadík et al. 2013] that are artificial patterns designed to expose well known perceptual phenomena, such as luminance masking, contrast masking and contrast sensitivity. The images are generated in the luminance domain (linear) and converted to gray scale images (luma) using the sRGB color space. Differently from other sets in our collection this one includes abstract patterns like blobs or stripes with different contrast values. For those scenes we prepared three distortion levels by blending linearly reference and distorted image. One of the scenes is shown in Figure 2.

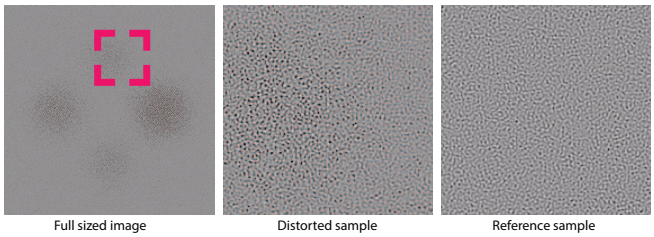

Fig. 2. Example of perception patterns dataset.

1.3 Aliasing

Generally aliasing is a phenomenon which happens when sampling frequency of a signal is too low to reproduce high frequency details accordingly to the Nyquist criterion. In the image domain, aliasing appears as an effect that include jagged profiles, improperly rendered details, and stair step artifacts on the edges. All these images are rendered starting from 3D scenes of interior rooms or outdoor environments. In this category we created from one up to three distortion levels and for this purpose we used different sample numbers for multi-sampling anti-aliasing method. Images source: [Piórkowski et al. 2017]. Figure 3 shows an example of aliasing artifact.

Fig. 3. Example of aliasing artifact.

1.4 Shadow acne

Shadow acne is an effect caused by the discrete nature and limited resolution of the shadow map. During depth map generation the angle between surface and ray of light has to be taken into account. Tilted depth texel can cross the surface having a part above and a part below it. The resulting effect is a striped Moire pattern. This

supra-threshold type of artifact is commonly seen in computer games and can be the reason of unnatural looking image. Images come from [Piórkowski et al. 2017]. An example of this kind of artifact is presented in Figure 4.

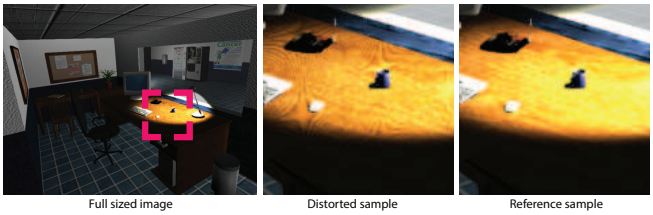

Fig. 4. Example of Moire pattern known as shadow acne artifact.

1.5 Peter panning

This artifact appears clearly as supra-threshold distortion and is related to objects with missing shadows or part of it, which look like detached from the surface, conveying the illusion of floating above the surface. Peter Panning arises from a correction of another problem. Since adding a depth offset is a technique for removing shadow acne (Section 1.4), this increment is related to pixel position in light space. Peter Panning results results from too large depth offset which causes errors in the depth test. Like shadow acne, peter panning is aggravated when there is insufficient precision in the depth buffer. Calculating tight near planes and far planes also helps avoid peter panning. Figure 5 shows an example of peter panning artifact.

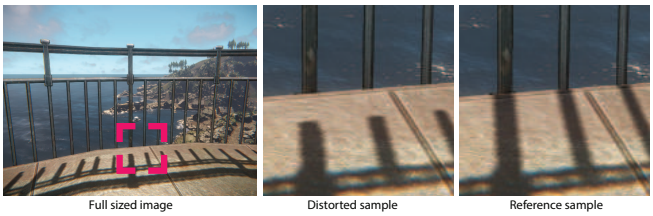

Fig. 5. Example of peter panning artifact - typical 'detached' shadows.

1.6 Shadow map downsampling

Downsampling of an image is the process of information reduction. Using lower resolution of shadow maps results in a loss of shadows accuracy but improves the computational performance. Game rendering benefits a lot from this technique. In order to maintain a fast refresh rate during game rendering, it is common practice to use the possibly smallest shadow maps. If the map used for generating shadows is too small, it appears on the screen as the jagginess of shadows' edges. This is kind of a supra-threshold artifact, but since it is localized only on the shadows' edges it is quite difficult to notice, especially without any reference image. An example of artifact caused by too low shadow map resolution is shown in Figure 6.

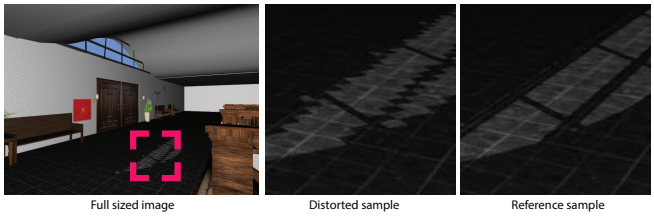

Fig. 6. Example of jagged shadows' edges as typical artifact that appears when the shadow map resolution is too low.

1.7 Z-fighting

Z-fighting (or stitching) is a 3D rendering effect that happens where two or more primitives have close values in the z-buffer. This causes annoying flickering issue since one primitive can be displayed in front or behind the other inconsistently. Several techniques can mitigate the problem as increasing depth buffer resolution or changing slightly the position of the objects. Since it is usual in game engines to deal with very complex scene with many objects, it is quite common to experience this kind of artifact. All 10 images were render in Unity or CryEngine and come from [Piórkowski et al. 2017]. Figure 7 shows an example of z-fighting artifact.

Fig. 7. Example of Z-fighting artifact caused by small precision of the depth buffer.

1.8 Compression

Compression dataset consists of 71 images and contains distortions due to experimental low-complexity image compression, operating at several bit-rates. Compression artifacts are the most common ones in computer graphics. Too low quality settings of compression result in very well known blockiness or mosaic artifact (Figure 8) which has a great impact on overall image quality. The distortion appear globally on the whole image and its visibility depends on the local image content. Usually it is well seen on gradients and can be easily masked by some specific frequencies. This set is an important source of near-threshold distortions.

Fig. 8. Example of blockiness effect - compression artifact.

1.9 Deghosting

High dynamic range (HDR) images become very popular in the recent years. Merging multiple exposures is a common method for generating HDR images. During acquisition, in the presence of a dynamic scene, non static objects can cause ghosting artifacts. Usually deghosting algorithms replace detected motion pixels either with pixels from only one exposure, or from multiple exposures. The main drawback of these methods is the reduction of high dynamic range of the moving object and local color or brightness deviation (Figure 9). Some other common artifacts that could be introduced by deghosting process are motion artifacts, and noise. This set of artifacts consists of either supra-threshold and near-threshold distortions. Images come from [Karađuzović-Hadžiabdić et al. 2017].

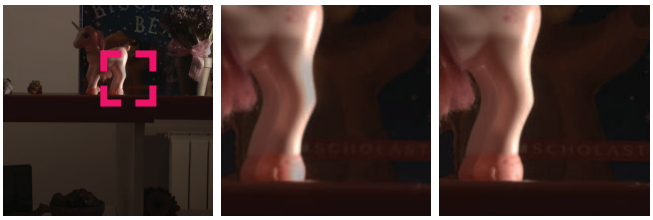

Fig. 9. Example shows local color and brightness deviations.

1.10 IBR

This set contains typical optical flow warping artifacts (Figure 10) and small shifts caused by nearest neighbor warping methods. Optical flow warping results usually in deformation of the objects, slight ghosting and discontinuities. The images created with NN method do not contain any artifacts, but they are slightly shifted according to the reference image. This effect is almost unnoticeable for a human, even when he compares testing image with the reference one. This kind of distortions were prepared to make our metric invulnerable to slightly misaligned images. The images come from dense light-field camera acquisition, followed by a process of images reduction and subsequently a reconstruction process with an optical flow or a nearest neighbor (NN) policy. NN images have only one distortion level and optical flow ones have three of them. All the images originate from [Adhikarla et al. 2017].

Fig. 10. Double edges and discontinuities caused by optical flow warping method.

1.11 CGIBR

This set consists of 5 rendered images and one photography that contain optical flow and linear warping artifacts (ghosting and discontinuity presented in Figure 11). Ghosting artifact results in the image as the objects with double edges and semi-transparent areas between those edges. Ghosting is a common artifact of HDR bracketing, but appears also in warping methods, especially in linear interpolation. In this case objects closer to the camera have stronger ghosting effect than objects in background. Each image has only one distortion level. All images of this set come from [Adhikarla et al. 2017].

Fig. 11. Example of ghosting as typical artifact of linear warping.

1.12 TID2013

In addition to 296 newly marked images, we added 261 images from the TID2013 image quality dataset [Ponomarenko et al. 2015], for which we could automatically generate marking. We selected from that dataset a subset of images that did not contain noticeable differences and assigned them marking maps set to 0s (no user markings). Then we selected another subset with well-noticeable distortions and set corresponding marking maps to 1s (distortions visible in the entire image). To ensure that both subsets were correctly selected, we compared the four least severe distortion levels with the reference images in an additional pairwise comparison experiment (comparisons missing in the original dataset) and scaled the original (per-observer) pairwise data together with additional measurements using methods described in [Perez-Ortiz and Mantiuk 2017] and assuming Thurstone Case V observer model. Then, we selected for the first subset the images with the score of less than 0.2 just-objectionable-difference (JOD) to the reference, and for the second subset the images with the difference larger than 3 JODs. We also excluded the distortion types that affected only small image regions, such as JPEG transmission errors, and left the distortions that affected all pixels.

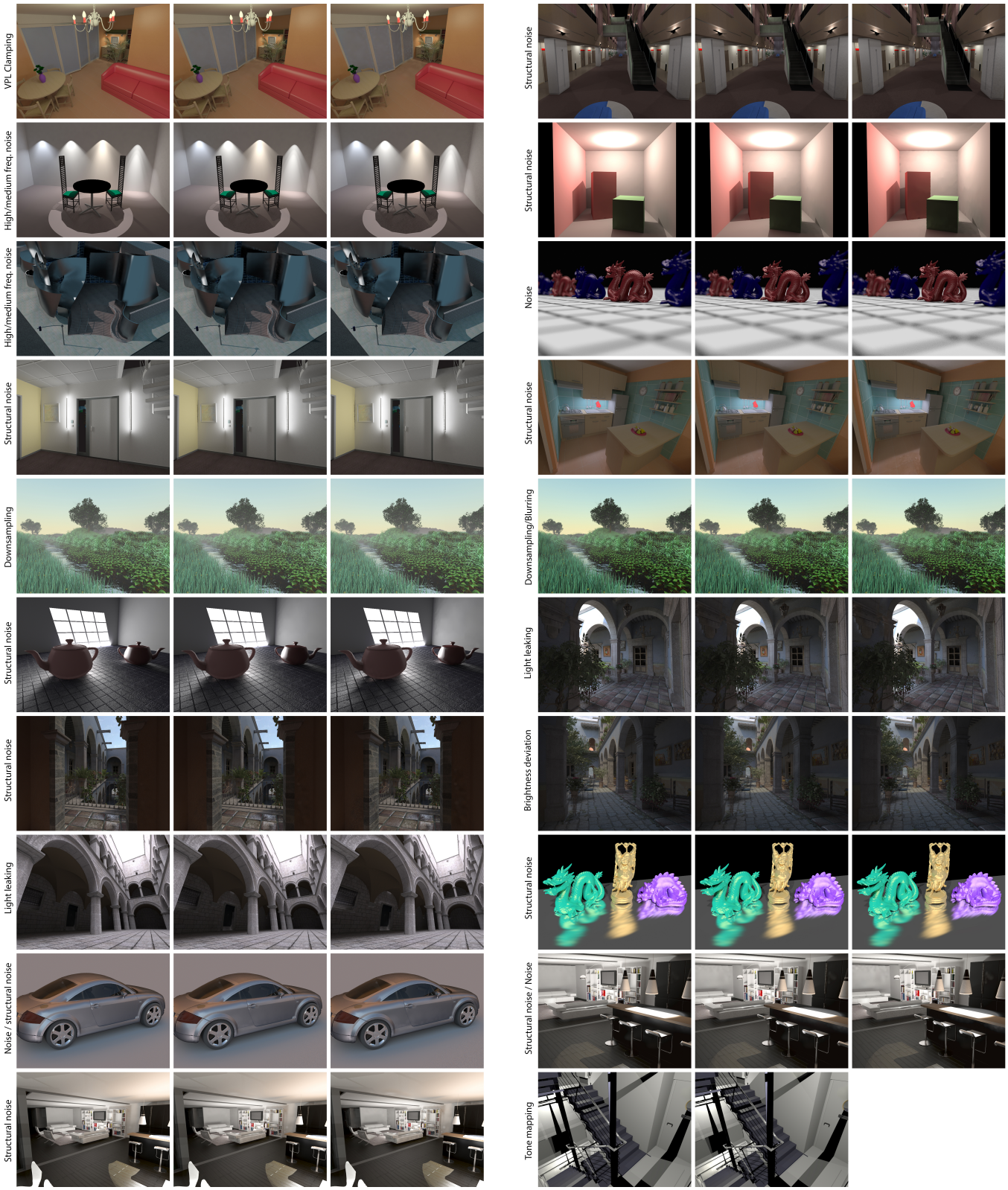

Fig. 12. Mixed subset.

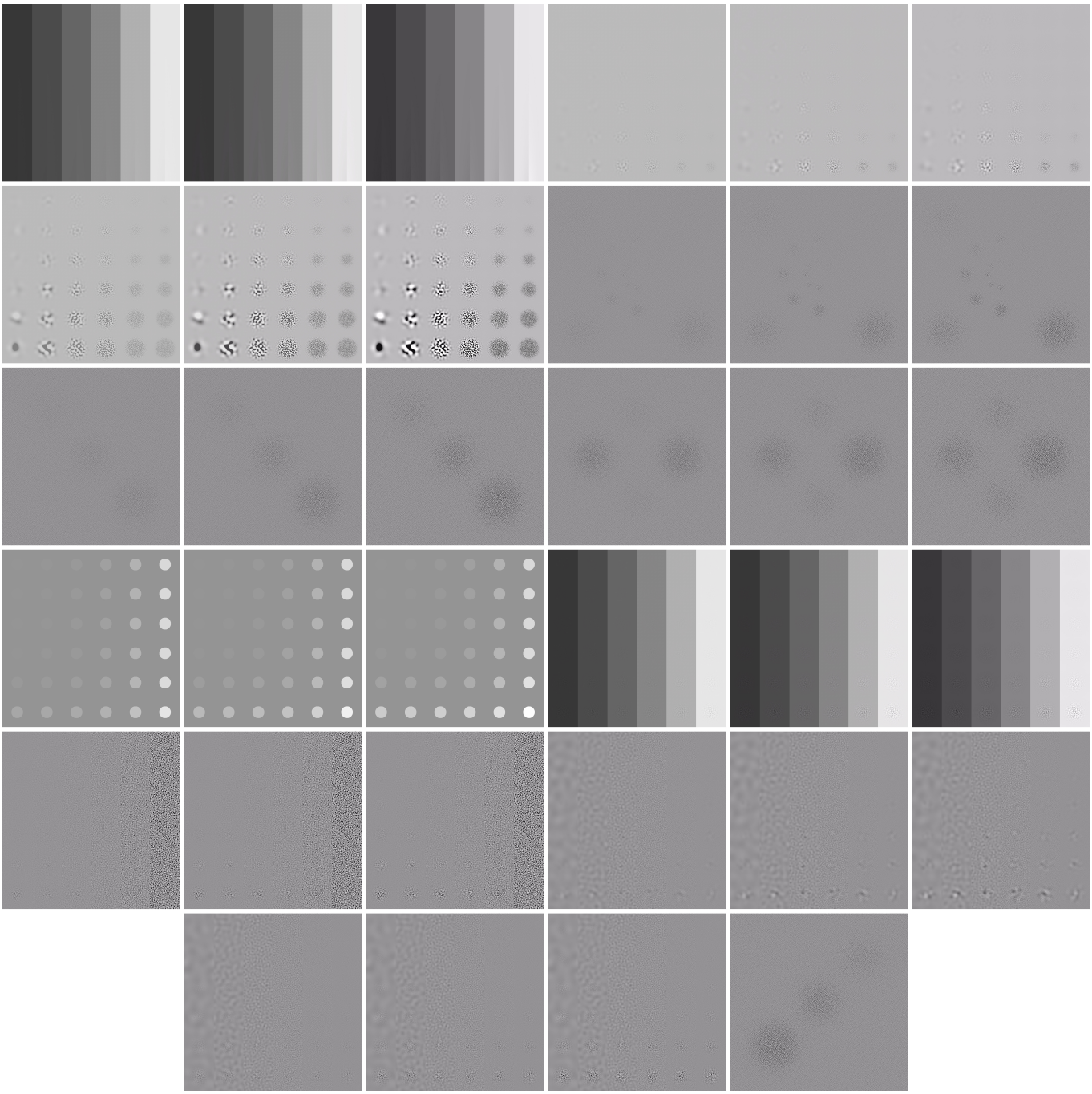

Fig. 13. Perception patterns subset.

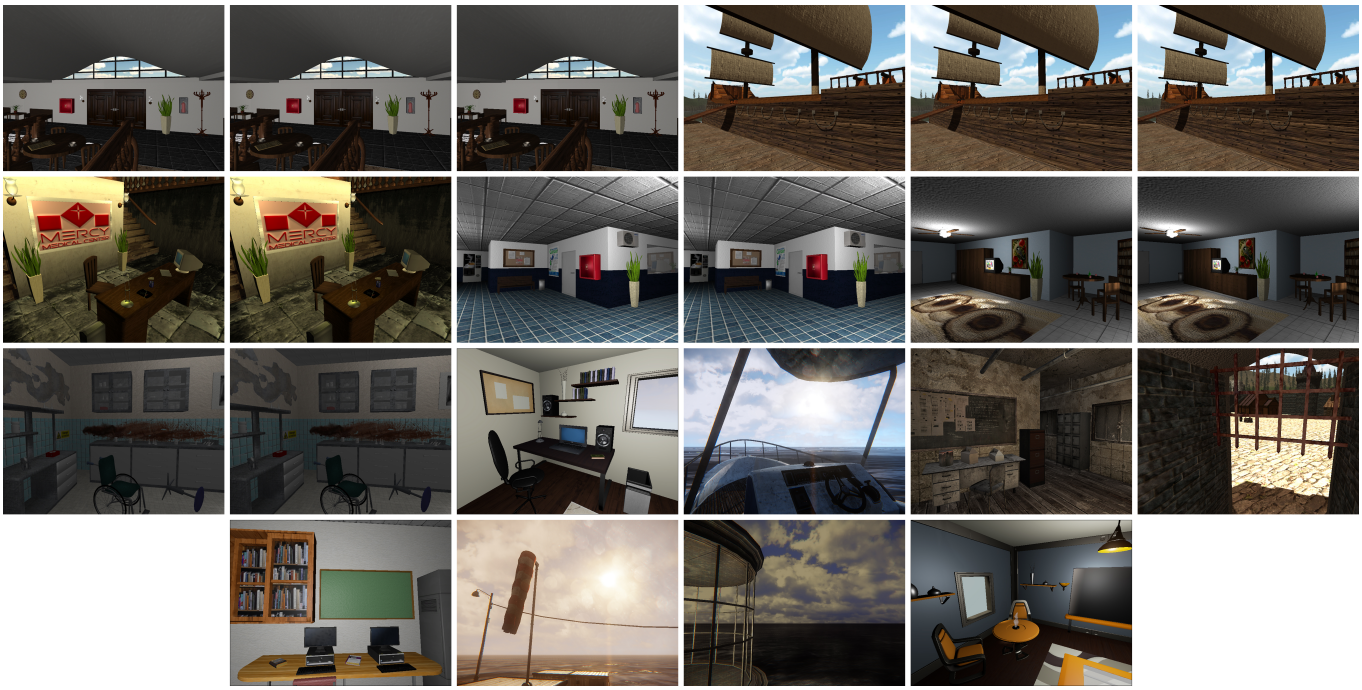

Fig. 14. Aliasing subset.

Fig. 15. Shadow acne subset.

Fig. 16. Peter panning subset.

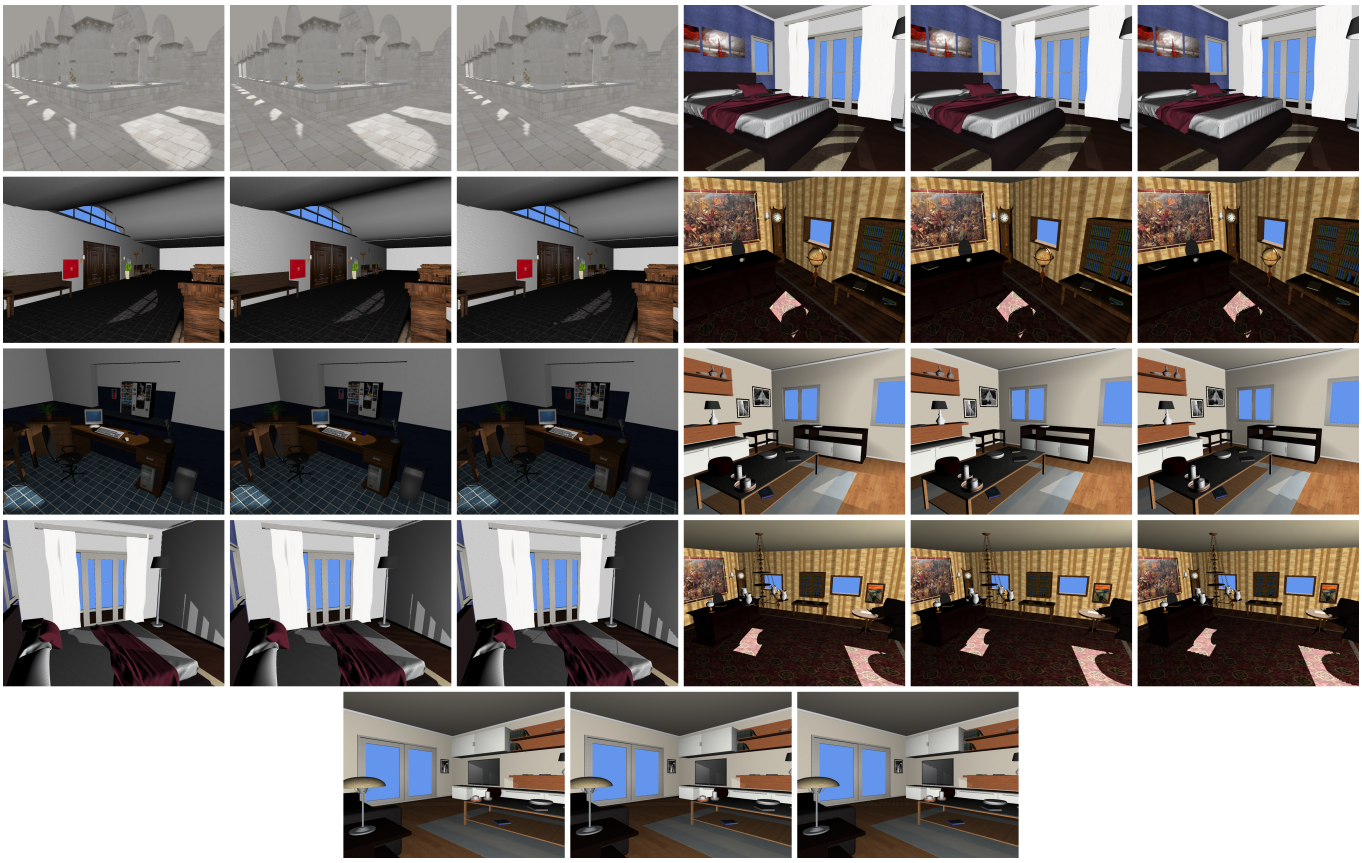

Fig. 17. Shadow map downsampling subset.

Fig. 18. Z-fighting subset.

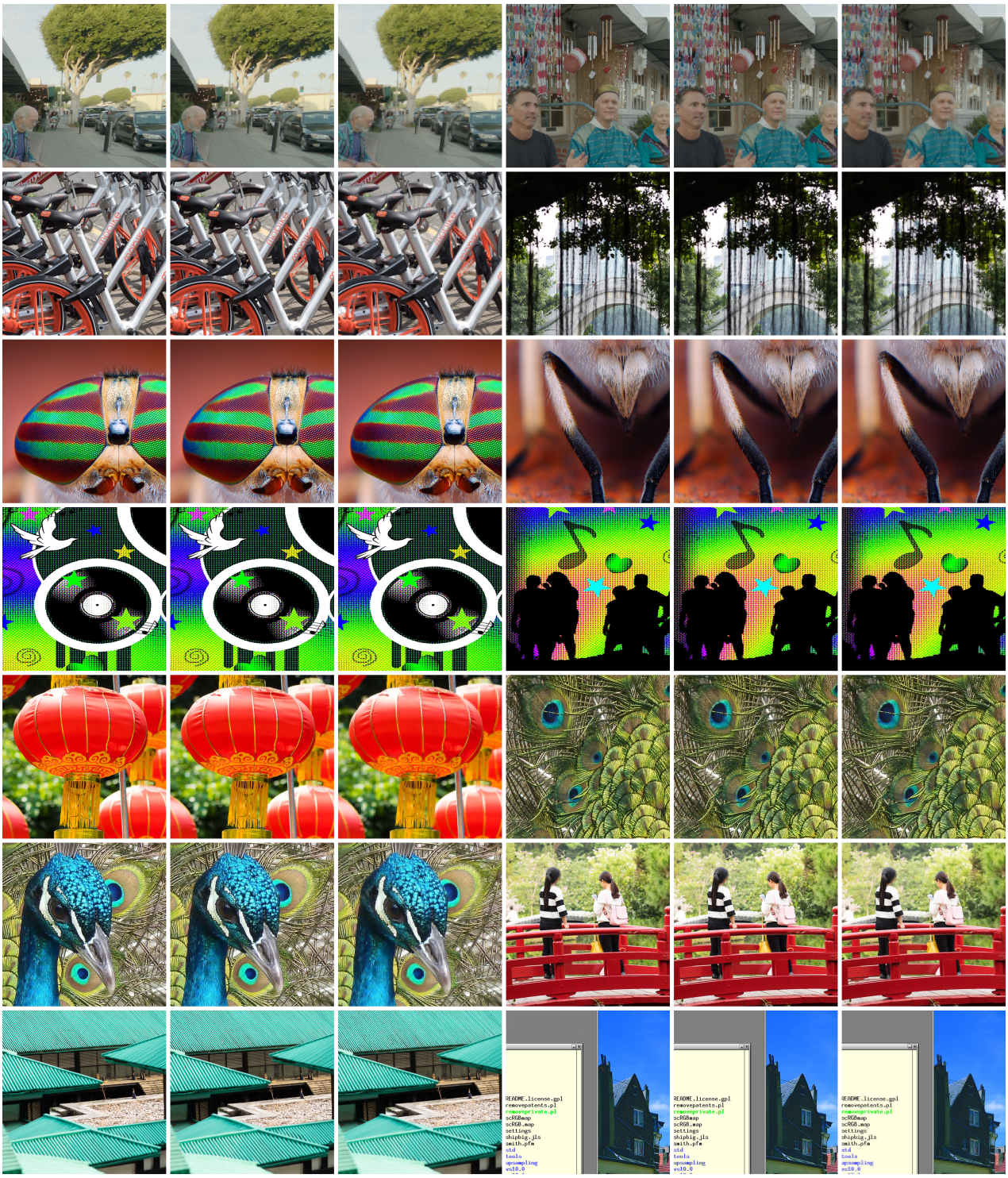

Fig. 19. Compression subset (continued on the next page).

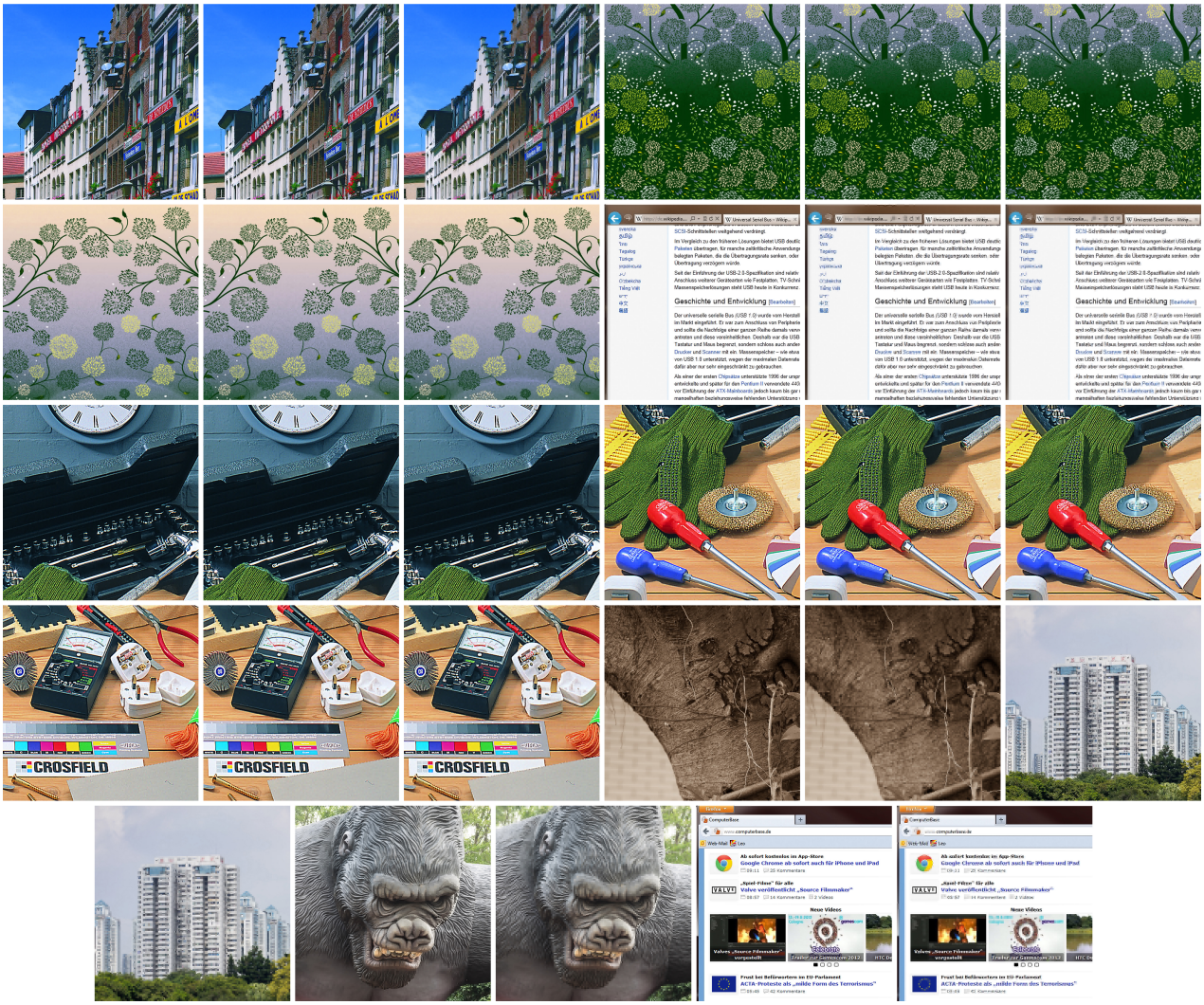

Fig. 19. Compression subset. (cont.)

Fig. 20. Deghosting subset.

Fig. 21. IBR subset.

Fig. 22. CGIBR subset.

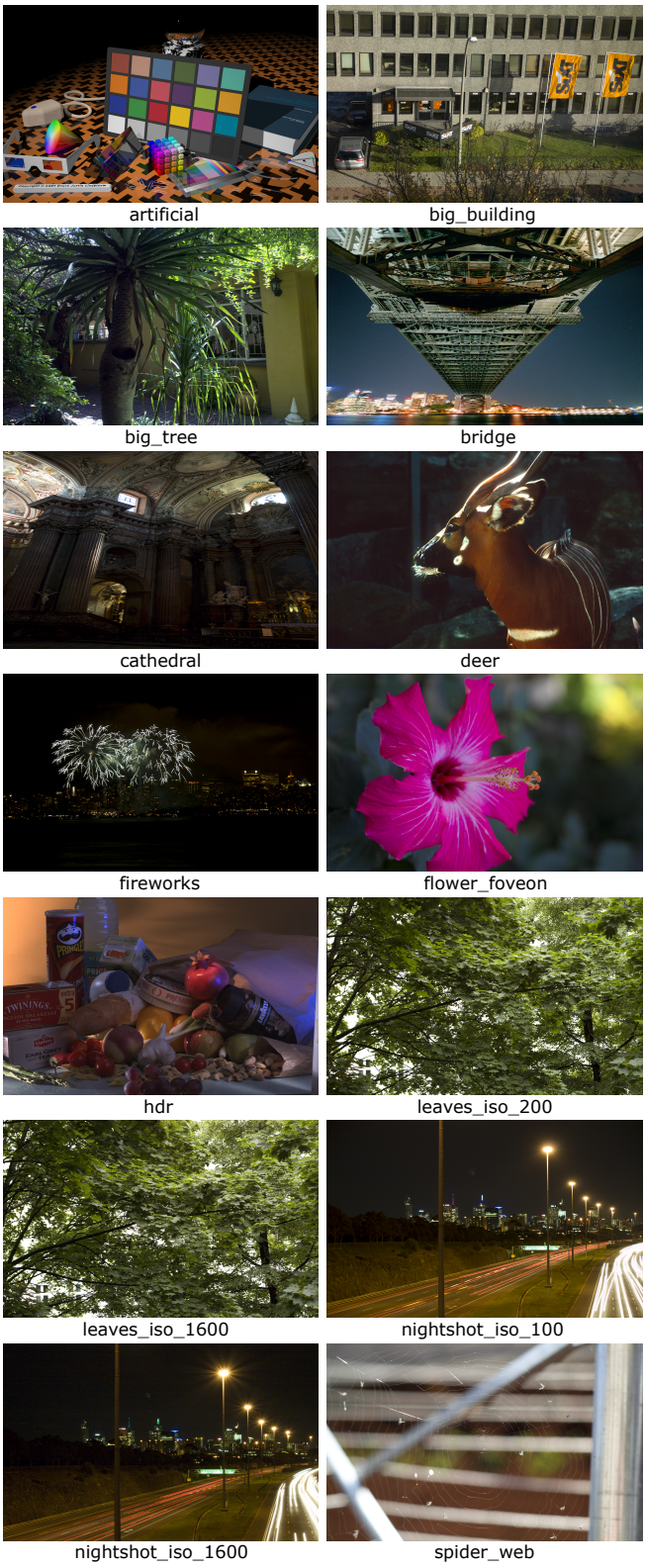

Fig. 23. Images used for compression application

2 APPLICATION SUPPLEMENT

In this section we present the materials used for our application and additional results for our applications that were not included to the paper.

2.1 Visually lossless compression

For application purpose, we used Rawzor's free dataset [1] which contains a rich kind of image contexts. The images are cropped to 960×600 to fit into our screen. Then we run JPEG standard compression experiment on this dataset with different compression qualities. We use the standard JPEG library for running JPEG compression on images. The cropped images from Rawzor's dataset are shown in Figure 23.

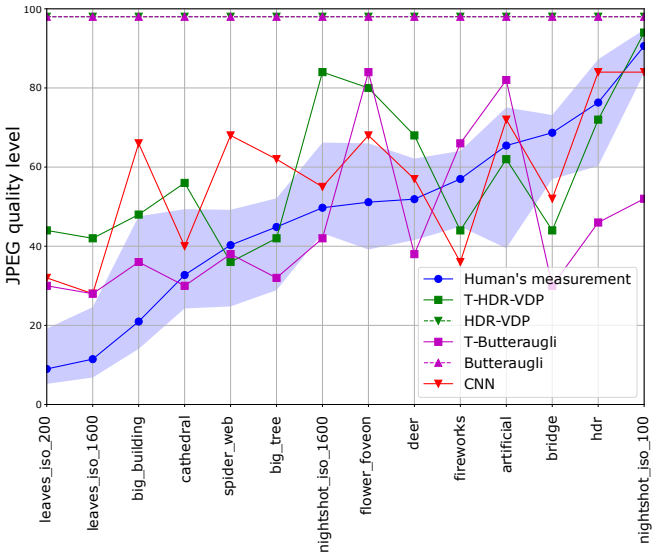

Fig. 24. Visually lossless compression results

In Figure 24 we present results of our visually lossless compression with images' names in x axis labels.

2.2 Content-adaptive watermark

In Figure 25 we present more results of our watermarking application.

3 TRAINED METRICS COEFFICIENTS

The appendix contains tables with the trained parameters for each metric. Values were rounded to the fifth decimal place.

3.1 T-ABS

Fold	1	2	3	4	5
thr	0.0321	0.02942	0.02417	0.03639	0.04059
beta	1.2217	1.08502	1.12467	1.06505	1.02051

[1] (*http://imagecompression.info/testimages/*)

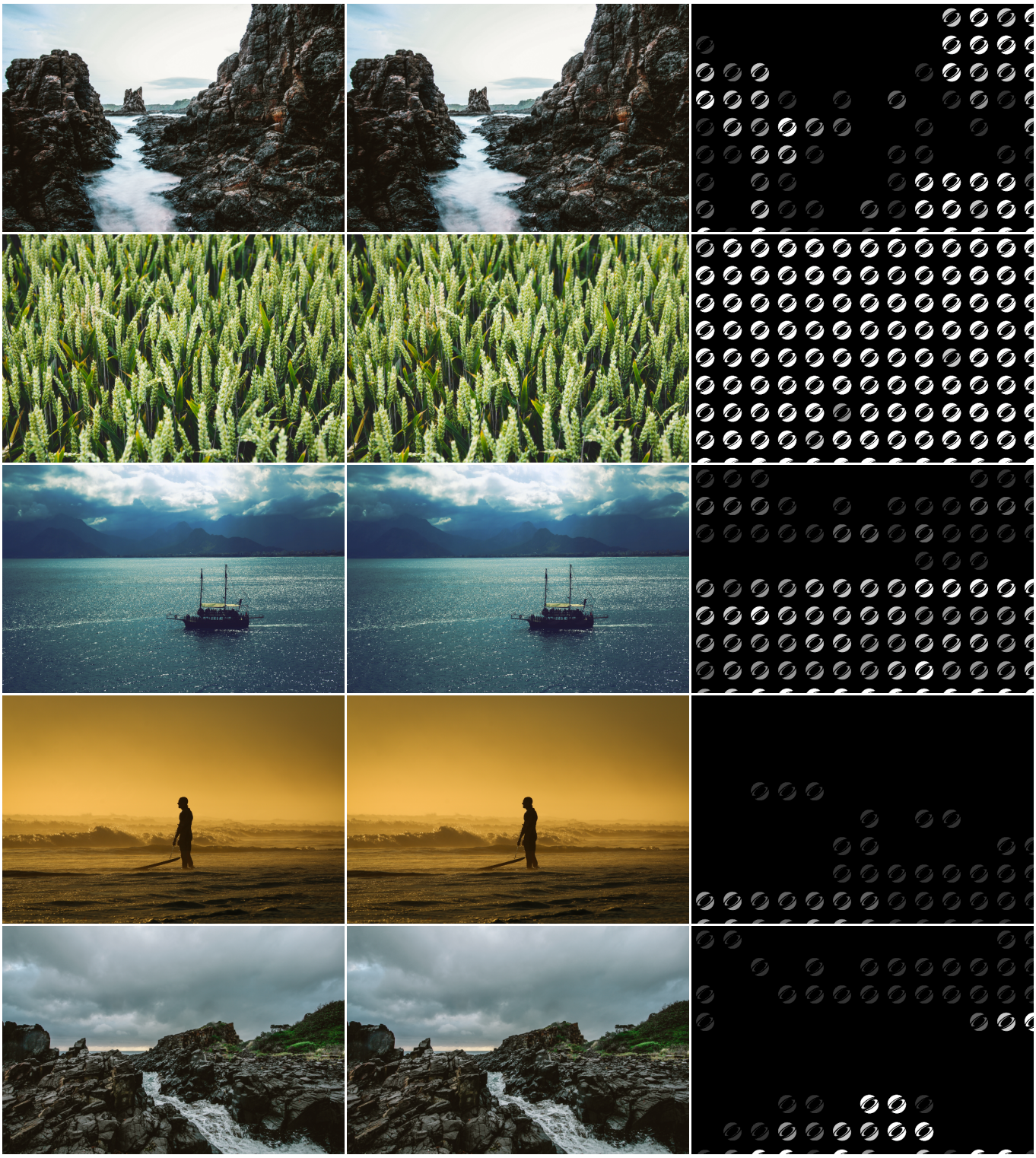

Fig. 25. Results of watermarking application. Left column presents reference image, center column - watermarked image, right column - watermark amplified for the presentation purpose

3.2 T-CIEDE2000

Fold	1	2	3	4	5
thr	6.55466	6.25582	5.13933	7.22037	8.29298
beta	1.27386	1.23024	1.38956	1.15686	1.13099

3.3 T-sCIELab

Fold	1	2	3	4	5
thr	2.98988	2.87221	2.55975	3.01634	3.57053
beta	1.44352	1.61917	1.60862	1.49786	1.27204

3.4 T-SSIM

Fold	1	2	3	4	5
thr	0.41701	0.36027	0.33016	0.44991	0.36851
beta	3.79632	3.87352	2.47467	3.22844	3.60306
k1	0.13532	0.10600	0.18738	0.13919	0.17895
k2	0.76799	0.99830	0.99365	0.61453	0.99557

3.5 T-FSIM

Fold	1	2	3	4	5
thr	0.01233	0.01401	0.02391	0.02896	0.02893
beta	1.11681	1.19778	0.96535	0.95018	0.87426
k1	1.54232	0.11217	1.79319	0.53842	1.45343
k2	465.272	507.059	139.883	155.229	139.053

3.6 T-VSI

Fold	1	2	3	4	5
thr	0.00640	0.01523	0.00811	0.00998	0.01313
beta	1.13423	1.10442	0.80044	0.99466	0.98652
k1	11.4743	5.76465	15.5094	262.608	2.75587
k2	212.161	38.7918	412.850	411.786	296.098
k2	249.399	49.2923	204.810	126.973	85.4913

3.7 T-Butteraugli

Fold	1	2	3	4	5
thr	4.57724	4.39537	5.13794	5.27943	4.97349
beta	2.17965	2.15649	1.88865	1.95179	1.99546

3.8 T-HDR-VDP

Fold	1	2	3	4	5
peak_sensitivity	3.14234	3.12003	3.03552	3.08941	3.07797
mask_self	1.34086	1.10952	1.34726	1.38707	1.31143
mask_xn	-3.59236	-1.67806	-1.93849	-1.27285	-1.56384
mask_p	0.47948	0.43410	0.57700	0.57052	0.53681
mask_q	0.10744	0.11764	0.26642	0.29759	0.18823
psych_func_slope	0.37482	0.35478	0.30360	0.36325	0.32973
si_sigma	-0.49228	-0.45199	-0.49665	-0.49300	-0.48688

REFERENCES

Vamsi Kiran Adhikarla, Marek Vinkler, Denis Sumin, Rafał Mantiuk, Karol Myszkowski, Hans-Peter Seidel, and Piotr Didyk. 2017. Towards a Quality Metric for Dense Light Fields. In *30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*. IEEE, Honolulu, HI, USA.

Martin Čadík, Robert Herzog, Rafał K. Mantiuk, Radosław Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. 2013. Learning to predict localized distortions in rendered images. In *Computer Graphics Forum*, Vol. 32. 401–410.

Martin Čadík, Robert Herzog, Rafał K. Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. 2012. New Measurements Reveal Weaknesses of Image Quality Metrics in Evaluating Graphics Artifacts. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)* 31, 6 (2012), 147.

Henrik Wann Jensen. 2001. *Realistic Image Synthesis Using Photon Mapping*. A K Peters, Ltd.

Kanita Karađuzović-Hadžiabdić, Jasminka Hasić Telalović, and Rafał K Mantiuk. 2017. Assessment of multi-exposure HDR image deghosting methods. *Computers & Graphics* 63 (2017), 1–17.

Alexander Keller. 1997. Instant Radiosity. *SIGGRAPH '97 (Proceedings of the 24th annual conference on Computer graphics and interactive techniques)* (1997).

Jaroslav Křivánek, Pascal Gautron, Sumanta N. Pattanaik, and Kadi Bouatouch. 2005. Radiance Caching for Efficient Global Illumination Computation. *IEEE Trans. Vis. Comput. Graph.* 11, 5 (2005), 550–561. https://doi.org/10.1109/TVCG.2005.83

Maria Perez-Ortiz and Rafal K. Mantiuk. 2017. A practical guide and software for analysing pairwise comparison experiments. *arXiv preprint* (dec 2017). arXiv:1712.03686 http://arxiv.org/abs/1712.03686

Rafał Piórkowski, Radosław Mantiuk, and Adam Siekawa. 2017. Automatic Detection of Game Engine Artifacts Using Full Reference Image Quality Metrics. *ACM Transactions on Applied Perception (TAP)* 14, 3 (2017), 14.

Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, and C.-C. Jay Kuo. 2015. Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication* 30 (jan 2015), 57–77. https://doi.org/10.1016/j.image.2014.10.009

Gregory J. Ward, Francis M. Rubinstein, and Robert D. Clear. 1988. A Ray Tracing Solution for Diffuse Interreflection. *Computer Graphics* 22, 4 (1988).

Received January 2018; final version January 2018