EUROGRAPHICS 2012 / P. Cignoni, T. Ertl
(Guest Editors)

Volume 31 (2012), Number 2

NoRM: No-Reference Image Quality Metric for Realistic Image Synthesis *(Supplementary Material)*

Robert Herzog[1] and Martin Čadík[1] and Tunç O. Aydın[1,2] and Kwang In Kim[1] and Karol Myszkowski[1] and Hans-P. Seidel[1]

[1] MPI Informatik Saarbrücken, Germany, [2] Disney Research Zurich, Switzerland

Abstract

In this document we provide specific details and further results of our method, in particular for the user study. Please notice that Table 3 is reproduced here from the main manuscript for the sake of the reading convenience, and Fig. 1 and Figs. 7, 8 represent the full versions of Fig. 1 and Fig. 8, respectively, in the main paper. To accelerate the progress in visual metrics research, we make the subjective experimental data gathered in scope of this study publicly available (http://www.mpi-inf.mpg.de/resources/hdr/norm/).

1. Synthetic Image Acquisition

We generated a set of 24 synthetic artifact images with references, which are shown in Fig. 1. The scene and the rendering parameters of each image are given in Table 1. However, only the parameters relevant to the artifact type are provided in the table. For the Lightcuts rendering algorithm [WFA*05] we used the proposed default parameters (see [WFA*05]) with automatic VPL clamping (1% of the total result and at most 1500 VPLs per pixel).

2. Classification Details

We use the C-library `libsvm 3.0` [CL11] for support vector classification, and `ANN 1.1.2` [MA10] (Approximate nearest neighbor) for (approximate) k-nearest neighbor searching. The test and training images were generated using several global illumination rendering softwares and exported to `OpenEXR` [Ope] HDR-image files. Rendered images and material buffers were stored in 16 bit RGB format whereas depth buffers use 32 bit per pixel.

During the training phase we do not compute features for every pixel in the images but for a uniformly sampled set of pixels in the image. However, since usually fewer pixels are labeled as artifacts we sample the artifact labels denser than the remaining pixels. To counter-balance this unequal sampling we have to re-weight the posterior probability of the k-nn prediction results by the ratio of the label sampling-densities. In the k-nearest neighbor search we use the L_1 distance norm, which produced better results than the L_2 norm

Img.	Scene	Renderer	Settings (Artifact/Ref.
#1	Disney	IGI (PT)	1M VPLs (43K spp)
#2	Kitchen	IGI (PT)	1M VPLs (380K spp)
#3	Tab	IGI (PT)	1M VPLs (97K spp)
#4	Bar	IGI (PT)	1M VPLs (200K spp)
#5	Sponza	GL (GL)	Shadowmap: 1$K \times 1K$ ($4K \times 4K$)
#6	Sponza Trees	GL (GL)	Shadowmap (PCF): 1$K \times 1K$ ($4K \times 4K$)
#7	Sponza Trees	GL (GL)	Shadowmap (PCF): 1$K \times 1K$ ($4K \times 4K$)
#8	Sponza Trees	GL (GL)	Shadowmap: 1$K \times 1K$ ($4K \times 4K$)
#9	Apartment	LC (LC)	100K VPLs (2M VPLs)
#10	Sponza Trees	IGI (LC)	60K VPLs (2M VPLs)
#11	Apartment	IGI (LC)	40K VPLs (2M VPLs)
#12	Sibenik	GL (GL)	Shadowmap: 1$K \times 1K$ ($4K \times 4K$)
#13	Icido	LC (LC)	100K VPLs (1M VPLs)
#14	Fairy	GL (GL)	Shadowmap: 1$K \times 1K$ ($4K \times 4K$)
#15	Sibenik	GL (GL)	Shadowmap: 1$K \times 1K$ ($4K \times 4K$)
#16	Fairy	GL (GL)	Shadowmap (PCF): 1$K \times 1K$ ($4K \times 4K$)
#17	Box	IGI (PT)	1M VPLs (60K spp)
#18	Apartment	LC (LC)	200K VPLs (2M VPLs)
#19	Sibenik	GL (GL)	Shadowmap: 1$K \times 1K$ ($4K \times 4K$)
#20	Conference	LC (LC)	200K VPLs (2M VPLs)
#21	Fairy	GL (GL)	Shadowmap: 512 $\times$ 512 ($4K \times 4K$)
#22	Sibenik	GL (GL)	Shadowmap: 1$K \times 1K$ ($4K \times 4K$)
#23	Apartment	LC (LC)	200K VPLs (2M VPLs)
#24	Fairy	GL (GL)	Shadowmap: 512 $\times$ 512 ($4K \times 4K$)

Table 1: *The scene identifier and rendering parameters of our image data set shown in Fig. 1 for the artifacts rendering and reference rendering algorithm (shown in parenthesis). GL stands for the OpenGL based deferred renderer based on shadow maps with percentage closer filtering (PCF) and screen-space ambient occlusion. IGI is an instant global illumination renderer, which supports glossy VPLs. The reference solutions are computed either by pathtracing (PT) with a constant number of samples per pixel (spp) or by the lightcuts algorithm (LC) [WFA*05].*

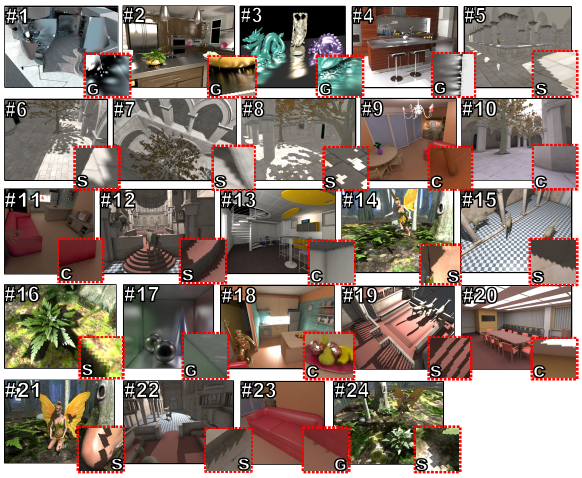

Figure 1: *The set of images used for our tests consisting of images exhibiting VPL clamping bias artifacts (C), glossy VPL noise (G), and shadow map aliasing (S). The corresponding scene and rendering settings are given in Table 1.*

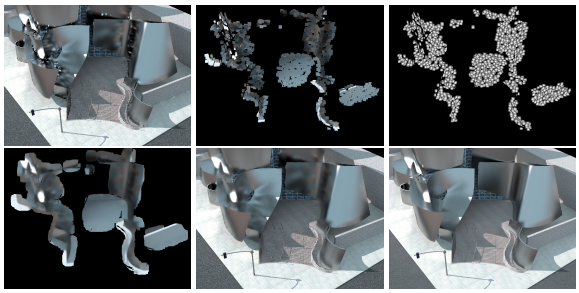

Figure 2: *Visualization of the inpainting steps (from top-left to bottom-right): original (exaggerated) artifact image for the scene #1, rectified splatting footprints (1 nearest neighbor), the corresponding Gaussian splat weights, the refined and normalized result (10× more splats), the blended result, and the reference image. Note the strong chromatic noise in this example.*

for both tasks, the classification and the inpainting, probably because a few outliers in the feature vectors are not dominating the final result.

3. Inpainting Details

The main difficulty in the inpainting step is to find artifact-free image blocks that fit the local image configuration. Here we give a more detailed description of our automatic procedure.

First, we remove the material-dependent signal from the color image and obtain HDR lighting as already discussed in the main paper. Next, the resulting HDR lighting image is gamma corrected (with $\gamma = 2.2$), scaled to LDR (we use the inverse of the 95th-percentile of all HDR luminance pixels as scale factor), and converted to *YCbCr* color space. We also construct a Gaussian pyramid for the lighting image.

We assume that chroma is less affected by artifacts and focus only on the more important luminance during inpainting. However, to remove chroma noise, the two chroma channels are filtered with a joint-bilateral filter, which is described in the main paper. Then, the initially detected artifact-pixels are sampled uniformly ($\frac{1}{10}$ samples per pixel) and small square blocks of constant size ($w \times w$, $w := 16$ pixels) are extracted from the depth, normal and color buffer around each sampled pixel. After rectification based on the depth and normal block (transformation to texture space), a local feature descriptor is constructed, which is used as an index to query a database for the k-nearest neighbors (k-nn). This database is initially constructed from our training image pairs using the same patch extraction procedure and contains tens of thousands of rectified reference lighting patches together with the artifact descriptor index. However, in order to achieve a larger variation in the lighting patches, we randomly rotate each rectified patch during database construction (for that the patches are generated with $w^* = \sqrt{2} * w$ larger size and cropped after rotation).

As descriptor we use the downsampled luminance (8×8) of the rectified artifact patch multiplied with a Gaussian envelope to penalize off-center pixels. In order to deal with large-scale artifacts we employ a multi-scale approach during the k-nn search. We extract rectified image patches with constant size w from multiple levels of the Gaussian pyramid (e.g., first 2 levels) of the LDR lighting image. The k-nearest retrieved reference patches are first upsampled (bicubic) to the corresponding scale of the search descriptor, then cropped to our patch resolution w, and blended according to their k-nn distance norm (L_1). This linear combination of patches is warped to image space using the computed texture parametrization and splatted into the luminance image with weights computed from a radial Gaussian window that is warped via the same texture parametrization (see Fig. 2, top-right).

The whole sampling and splatting process is repeated until each pixel under the artifact label mask has a weight larger than a minimum threshold (we use 0.5). Next, the luminance image is normalized by dividing by the accumulated splat weights and chrominance is added from the prefiltered original image. Remaining cracks are filled using a push-pull approach and the result is modulated with the material buffer. Finally, the image is blended with the original image as described in the main paper. In Figs. 2 and 3 we show a failure case and a good example, respectively.

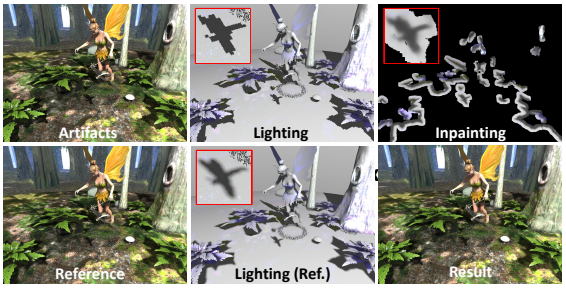

Figure 3: *Inpainting in a vivid textured scene (from top-left to bottom-right): original artifact image, the extracted lighting image, the inpainted splats from 3 nearest neighbors, the reference, with extracted lighting, and the inpainting result with textures blended with the original image.*

4. Human Visual System Details

In this section we describe implementation details of the human visual system (HVS) model used in the perceptual normalization step (Section 8 of the main manuscript). Principally, we take advantage of the observation that the rendering artifacts we consider are of medium to high frequency, and we can thus approximate the reference image I_{ref} using the distorted image I_{dst} after the inpainting correction (Section 7 of the main manuscript).

Our HVS model operates as follows: given an image (I_{ref}, I_{dst}), we first compute a 6-level Laplacian Pyramid of image luminance L. Then, the Wilson's transducer [Wil80] function T is applied at each pyramid level L_k as follows:

$$T_k(L_k,S) = \frac{3.291\,[(1+(SL_k)^3)^{1/3}-1]}{0.2599\,(3.433+SL_k)^{0.8}}. \tag{1}$$

The transducer function additionally takes a HVS sensitivity parameter S as input that is computed as:

$$S = CSF(\rho_k, L_{adapt}), \tag{2}$$

where CSF denotes the Contrast Sensitivity Function discussed in [Dal93] and described below. The spatial frequency ρ_k depends on the angular resolution of the input image in pixels per visual degree units (n_{ppd}), and is given by $n_{ppd}/2^k$, where $k=1$ for the highest frequency band [MKRH11]. The local luminance adaptation map L_{adapt} is approximated by the low-pass residue of the Laplacian Pyramid.

The Contrast Sensitivity Function (CSF) accounts for sensitivity changes due to the luminance adaptation and spatial frequency of the image:

$$CSF^S(\rho, L_a, \theta, i^2, d, c) = P \cdot \min\left[S_1\left(\frac{\rho}{r_a \cdot r_c \cdot r_\theta}\right), S_1(\rho)\right], \tag{3}$$

where

$$\begin{aligned} r_a &= 0.856 \cdot d^{0.14} \\ r_c &= \tfrac{1}{1+0.24c} \\ r_\theta &= 0.11\cos(4\theta)+0.11 \\ S_1(\rho) &= \frac{\left[\left(3.23(\rho^2 i^2)^{-0.3}\right)^5+1\right]^{-\frac{1}{5}}}{A_l \varepsilon\rho e^{-(B_l\varepsilon\rho)}\sqrt{1+0.06e^{B_l\varepsilon\rho}}} \cdot \\ A_l &= 0.801\left(1+0.7\,L_a^{-1}\right)^{-0.2} \\ B_l &= 0.3\left(1+100\,L_a^{-1}\right)^{-0.15}. \end{aligned} \tag{4}$$

The parameters are:

- ρ – spatial frequency in cycles per visual degree,
- L_a – light adaptation level in cd/m^2,
- θ – orientation ($\theta = 1$),
- i^2 – stimulus size in deg^2 ($i^2 = 1$),
- d – distance in meters,
- c – eccentricity ($c = 0$),
- ε – constant ($\varepsilon = 0.9$),
- P – absolute peak sensitivity ($P = 250$).

Given the transducer responses T_k^{ref} (for (I_{ref})) and T_k^{dst} (for (I_{dst})) at each pyramid level k, the differences of HVS responses are combined using a Minkowski summation with exponent 2:

$$R = \left[\sum_k^K \left|T_k^{ref} - T_k^{dst}\right|^e\right]^{(1/e)}. \tag{5}$$

Finally, the perceptually weighted response R is masked by the original binary distortion map $NoRM$ as follows:

$$NoRM_{perc.} = R \cdot NoRM. \tag{6}$$

5. User Study Details

In pure image classification providing the labels is relatively easy. The user sets a discrete label for each image. In our approach labels are continuous over the entire image and each pixel must be labeled, which is a tedious task prone to errors. However, providing good labels significantly affects the performance of the classification. Therefore in the following section we describe the design, procedure and results of a perceptual user study that we conducted to gather subjective labellings of artifacts in rendered images.

There were several motivations for the study: on one hand we needed the subjective data (labels) for training the classifier and to provide the guidance for inpainting, on the other hand we used part of the acquired data to validate the proposed no-reference metric. Furthermore, the performed study was an interesting probe to the perception of rendering artifacts on its own – to our best knowledge this was the first attempt to subjectively label locations of visual artifacts caused by various rendering techniques both in with- and without- the reference setups. Finally, results of the study are

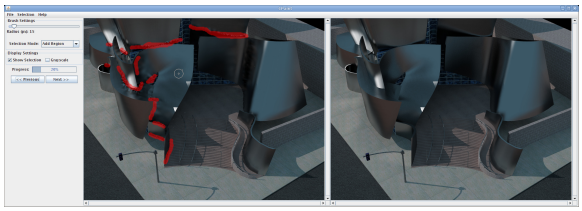

Figure 4: *User interface of the custom scribbling application used in the user study (with-the-reference version).*

valuable for the evaluation of existing full-reference metrics, which were usually not validated for detection of rendering artifacts, as we show below.

5.1. Experimental Design

The evaluated images were displayed on a characterized and calibrated LCD display. The calibration was performed using X-Rite i1 Display Pro colorimeter (to D65, 120 cd/m^2, colorimetric characterization by means of measured ICC profiles). The experimentation room was neutrally painted, darkened (measured light level: 2 lux), and observers sat approximately 60 cm from the display. The total of 20 observers took part in our experiments. They were computer graphics students and researchers, proficient in computer graphics, but naïve to the purpose of the experiment. We involved both male and female observers between the ages of 21 to 38, and all of them reported to have normal, or corrected-to-normal vision.

5.2. Experimental Procedure

In the experiment we were presenting selected images to users (see Fig. 5), who were asked to mark the regions they perceived as artifacts. For the purpose of the study we implemented a custom scribbling application, see Fig. 4. Using this application, we performed two experiments: in the first experiment (*with-the-reference*), an image exhibiting rendering artifacts (distorted image) was presented along with the artifact-free (reference) image; in the second experiment (*without-the-reference*), subjects saw only the distorted image. Each subject was introduced to the problem before the experiment as follows: in with-the-reference experiment, observers were asked to mark those regions in the distorted image, where they saw the difference to the reference image shown beside. For without-the-reference experiment, observers were instructed to label all the areas in the image that they found disturbing, while they were encouraged to concentrate on the lighting artifacts only. The sequence of images and the type of the experiment (with- or without-the-reference) were randomized, but for a given observer the type of experiment remained constant. Specifically, the half of the subjects performed the with-reference experiment, the other half did the no-reference part. The whole experiment took on average 25 minutes per subject. According to the post-test discussions with our subjects, the no-reference experiment was slightly more demanding than the with-reference one.

Image #	subj. no-ref.	HDR-VDP2	SSIM	NoRM	NoRM perc.
1	0.903	**0.725**	0.674	0.628	0.662
2	0.908	0.579	0.538	0.558	**0.590**
3	0.828	**0.778**	0.643	0.682	0.727
4	0.913	**0.495**	0.469	0.298	0.436
5	0.769	0.542	0.602	0.677	**0.748**
6	0.772	0.669	0.742	0.638	**0.767**
7	0.857	0.390	0.374	0.383	**0.479**
8	0.805	0.618	**0.692**	0.607	0.657
9	0.510	**0.418**	0.231	0.416	0.320
10	0.186	0.134	**0.637**	0.450	0.470
Average	0.745	0.535	0.560	0.534	**0.586**

Table 3: *Correlations of subjective responses in with-the-reference experiment with subjective responses in no-reference experiment and with the predictions of HDR-VDP2, SSIM, NoRM and NoRM after the perceptual normalization. The last row shows the average correlations over the test set. The best correlations (excluding the no-reference subjective experiment) for each stimulus are printed in bold.*

5.3. Study Results and Discussion

Results of the study are summarized in Figs. 7 and 8. In the first and second rows we show the input reference and distorted stimuli (#1–10), respectively. Images 1–4 exhibit glossy VPL artifacts [Kel97], images 5–8 suffer from shadow map discretization [RSC87], and images 9–10 show clamping bias artifacts [Kel97]. All the stimuli were tone mapped for the presentation purposes (images 1–4 using logarithm-based global tone reproduction curve [DMAC03], and the rest by means of photographic tone mapping operator [RSSF02]). Third and fourth rows show average subjective with-the-reference and no-reference experimental distortion maps, respectively. The maps were obtained by averaging over all observers in each study. This way we can compare the outcome of the artifact perception experiment in the presence and absence of the reference.

Interestingly, the visual inspection of distortion maps reveals apparent agreement between the subjective experiments. This is corroborated by the numerical analysis, where the correlation between the average distortion maps of the with-the-reference and no-reference experiments is markedly high (second column in Table 3). The exceptions are images 9 and 10, where the correlation is clearly weak. This is caused by rather low perceptual strength of clamping bias artifacts. Our subjects had apparently troubles to mark artifact regions here and some of them even left those images intact as artifact-free cases (happened in both experiments, see standard deviation images Table 6). This is also reflected in low maximal values in average distortion maps (see Table 2). In any case, sensitive subjects in the with-the-reference experiment were guided by the reference image

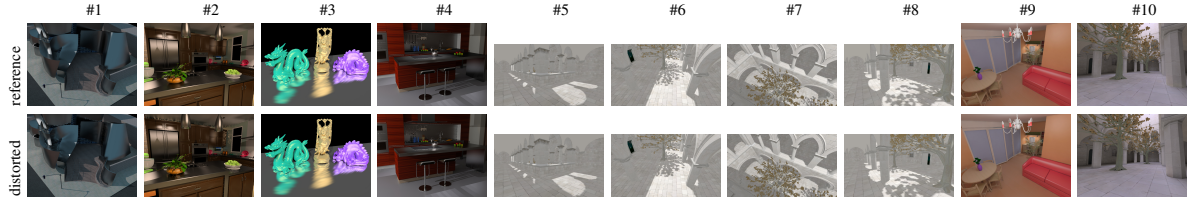

Figure 5: *The test set consists of 2 × 10 rendered images. Top row: reference images, bottom row: images containing rendering artifacts.*

Image #	subjects with-ref.	subjects no-ref.	HDR-VDP2	SSIM	NoRM	NoRM perc.
	[min, max]; avg; std	[min, max]; avg; std	[min, max]; avg; std	[min, max]; avg; std	[min, max]; avg; std	[min, max]; avg; std
1	[0.000, 1.000]; 0.065; 0.164	[0.000, 1.000]; 0.074; 0.167	[0.000, 1.000]; 0.048; 0.170	[0.000, 1.000]; 0.056; 0.125	[0.000, 1.000]; 0.130; 0.296	[0.000, 1.000]; 0.072; 0.175
2	[0.000, 1.000]; 0.160; 0.241	[0.000, 1.000]; 0.169; 0.260	[0.000, 1.000]; 0.189; 0.320	[0.000, 1.000]; 0.143; 0.210	[0.000, 1.000]; 0.208; 0.345	[0.000, 1.000]; 0.134; 0.246
3	[0.000, 1.000]; 0.041; 0.148	[0.000, 0.900]; 0.074; 0.166	[0.000, 1.000]; 0.052; 0.185	[0.000, 1.000]; 0.044; 0.117	[0.000, 1.000]; 0.091; 0.265	[0.000, 1.000]; 0.090; 0.263
4	[0.000, 1.000]; 0.067; 0.160	[0.000, 1.000]; 0.084; 0.196	[0.000, 1.000]; 0.134; 0.254	[0.000, 1.000]; 0.094; 0.175	[0.000, 1.000]; 0.227; 0.354	[0.000, 1.000]; 0.127; 0.214
5	[0.000, 1.000]; 0.041; 0.145	[0.000, 0.700]; 0.065; 0.119	[0.000, 1.000]; 0.113; 0.258	[0.000, 1.000]; 0.036; 0.139	[0.000, 1.000]; 0.058; 0.207	[0.000, 1.000]; 0.053; 0.203
6	[0.000, 0.800]; 0.046; 0.134	[0.000, 0.500]; 0.047; 0.084	[0.000, 1.000]; 0.115; 0.254	[0.000, 1.000]; 0.079; 0.209	[0.000, 1.000]; 0.050; 0.196	[0.000, 1.000]; 0.060; 0.224
7	[0.000, 1.000]; 0.043; 0.136	[0.000, 0.800]; 0.039; 0.108	[0.000, 1.000]; 0.173; 0.304	[0.000, 0.999]; 0.089; 0.190	[0.000, 1.000]; 0.046; 0.168	[0.000, 1.000]; 0.055; 0.196
8	[0.000, 0.900]; 0.083; 0.173	[0.000, 0.700]; 0.060; 0.112	[0.000, 1.000]; 0.300; 0.388	[0.000, 1.000]; 0.119; 0.254	[0.000, 1.000]; 0.074; 0.239	[0.000, 1.000]; 0.094; 0.281
9	[0.000, 0.300]; 0.008; 0.034	[0.000, 0.600]; 0.039; 0.081	[0.000, 1.000]; 0.045; 0.131	[0.000, 1.000]; 0.033; 0.088	[0.000, 1.000]; 0.072; 0.219	[0.000, 1.000]; 0.053; 0.177
10	[0.000, 0.500]; 0.010; 0.045	[0.000, 0.300]; 0.021; 0.042	[0.000, 1.000]; 0.465; 0.420	[0.000, 0.998]; 0.043; 0.104	[0.000, 1.000]; 0.053; 0.184	[0.000, 1.000]; 0.055; 0.200

Table 2: *Descriptive statistics of the distortion maps (depicted in Fig. 7 and Fig. 8) for each input image. Used abbreaviations: min=minimal value, max=maximal value, avg=average value, std=standard deviation, of a distortion map for particular image (1-10). (Maps in subjective experiments were obtained by averaging over all observers).*

which resulted in reasonable distortion maps. On the other hand, average subjective no-reference distortions for images 9 and 10 seem to be more random. For this reason, we use average with-the-reference subjective experiment results as a reference in the following evaluation of objective metrics.

Average subjective distortion maps (i.e. the 'ground truth') enable us to qualitatively evaluate predictions of the proposed no-reference metric NoRM, as well as the predictions of other state-of-the art metrics. We are not aware of any no-reference metric capable of providing distortion maps, i.e. the locations of artifacts. Therefore, we compare our predictions to the results of widely accepted full-reference metrics HDR-VDP2 [MKRH11] and SSIM [WBS*04], although they are often not appropriate for capturing the quality of rendered 3D scenes [RRP00, RR01]. Please note that full-reference metrics have the obvious advantage of knowing both the distorted and reference images. Our technique inputs the distorted image only, making the prediction of artifact locations and visibility much harder task. On the other hand, we make use of the knowledge of the scene depth and diffuse material albedo.

The predictions of HDR-VDP2, SSIM and NoRM are shown in rows 5, 6 and 7 of Figs. 7 and 8. Similarly to the analysis above, we show correlations of prediction maps with the subjective experimental results in Table 3 (columns 3, 4, 5). We run all the metrics with the default parameter settings to make the comparison fair, however better results may be achieved in some cases after manual parameter tweaking. Specifically, for *HDR-VDP2* we make use of the probability map output (P_{map}), which was calibrated for stimuli close to the visibility threshold. As some of our images show quite supra-threshold distortions, the response of HDR-VDP2 saturates. This is especially visible for large low-frequency supra-threshold differences, that are actually perceptually much less influential than as predicted by HDR-VDP2 (see e.g. image #10). On the other hand, HDR-VDP2's predictions are deliberately conservative so if there is a perceivable visible difference, though tiny, it should be reported. Apart from several ringing artifacts (probably due to the steerable pyramid decomposition), the metric performs well in this conservative sense.

In case of *SSIM,* we utilize the version of the SSIM index with automatic downsampling. The resulting SSIM distortion map is weighted according to recommendations [WBS*04] as follows: $SSIM_{out} = \max(0, SSIM_{map})^4$, and finally upsampled to the original image stimulus size (which is the reason of 'smooth' appearance of SSIM distortion maps). Some of the SSIM predicted artifacts are exaggerated (e.g. the shadow mapping artifacts masked by leafs of the tree in the image #7), probably due to absence of a high-level visual masking model, but in general the predictions of SSIM are considerably close to the subjective ground truth.

Distortion maps produced by the proposed no-reference metric *NoRM* are binary, meaning the presence or absence of an artifact. However, we assume the artifacts are spatially coherent and hence, for classification we sampled the test image uniformly (approx. every 10-th pixel) and interpolated

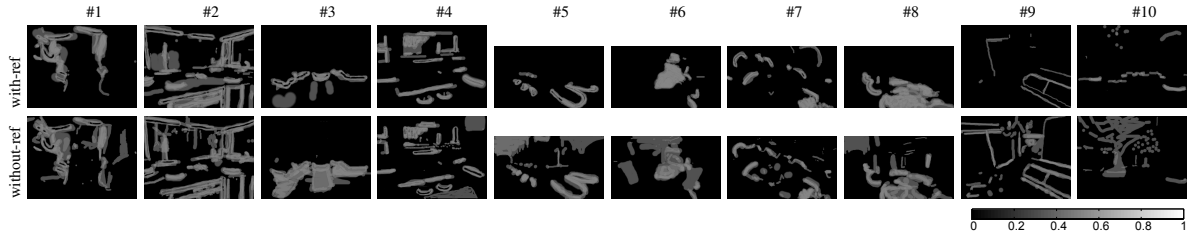

Figure 6: *Standard deviations over subjects for each input image. First row: standard deviation maps for with-the-reference experiment, second row: standard deviation maps for no-reference subjective experiment.*

the result, which produced continuous values at the edges of artifact predictions in the distortion maps. These distortion maps sometimes tend to show also too many locations (see e.g. the image #2), which may be correct, but the artifact severity is in reality obviously not uniform. However, thanks to the inpainting procedure, we are able to perform the perceptual normalization step (*NoRM perc.*), which makes the strength of detected artifacts substantially closer to average subjective distortion maps. Strictly speaking, the prediction we obtain after the perceptual normalization step is a suprathreshold distortion map calibrated in JND (just noticeable differences) units. To convert those values to probabilities, we employ a mapping function similar to the one proposed by Lubin [Lub95], where the value of 1 JND is mapped to the probability of $P = 0.5$ (see also [MKRH11]). According to Table 3, the prediction of NoRM does not correlate well with the ground truth in particular for the image #9 (as well as the predictions of other metrics). Peculiarly, this is the case where also the observers (in without-the-reference experiment) had troubles to find an agreement and to concisely mark the distorted regions.

In conclusion, neither HDR-VDP2 nor SSIM were designed or calibrated to predict the strength of rendering artifacts, but the distortion maps they produce are quite plausible. According to average correlations with measured ground-truth distortion maps, SSIM only slightly outperforms HDR-VDP2 (0.56 vs 0.535). The result of our metric (0.534) is qualitatively quite similar, making it competitive with current state-of-the-art full-reference metrics. The perceptual normalization step makes predictions of NoRM even closer to the experimental ground truth, resulting in the highest average correlation (0.586) of all the tested metrics. However, as one may observe in Table 3, each of the tested metrics fail for some, as well as perform the best for another input stimulus. Accordingly, there is still the need to improve the robustness of the metrics and a space for future research.

References

[CL11] CHANG C.-C., LIN C.-J.: LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology 2* (2011), 27:1–27:27.

[Dal93] DALY S.: The Visible Differences Predictor: An algorithm for the assessment of image fidelity. In *Digital Images and Human Vision* (1993), MIT Press, pp. 179–206.

[DMAC03] DRAGO F., MYSZKOWSKI K., ANNEN T., CHIBA N.: Adaptive logarithmic mapping for displaying high contrast scenes. In *Proc. of EUROGRAPHICS 2003* (Granada, Spain, 2003), Brunet P., Fellner D. W., (Eds.), vol. 22 of *Computer Graphics Forum*, Blackwell, pp. 419–426.

[Kel97] KELLER A.: Instant radiosity. In *Proceedings of SIGGRAPH* (1997), pp. 49–56.

[Lub95] LUBIN J.: *Vision Models for Target Detection and Recognition*. World Scientific, 1995, ch. A Visual Discrimination Model for Imaging System Design and Evaluation, pp. 245–283.

[MA10] MOUNT D. M., ARYA S.: ANN: A library for approximate nearest neighbor searching, January 2010. Version 1.1.2 available at http://www.cs.umd.edu/~mount/ANN/.

[MKRH11] MANTIUK R., KIM K. J., REMPEL A. G., HEIDRICH W.: HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on Graphics (Proc. of SIGGRAPH)* (2011), 40:1–40:14.

[Ope] OPENEXR: A hdr image file format. C++ library available at http://www.openexr.com.

[RR01] ROGOWITZ B. E., RUSHMEIER H.: Are image quality metrics adequate to evaluate the quality of geometric objects? In *Proc. of Human Vision and Electronic Imaging VI* (2001), SPIE, vol. 4299.

[RRP00] RUSHMEIER H., ROGOWITZ B. E., PIATKO C.: Perceptual issues in substituting texture for geometry. In *Proc. of Human Vision and Electronic Imaging V* (2000), SPIE, vol. 3959.

[RSC87] REEVES W. T., SALESIN D. H., COOK R. L.: Rendering antialiased shadows with depth maps. In *Proc. of SIGGRAPH* (1987).

[RSSF02] REINHARD E., STARK M. M., SHIRLEY P., FERWERDA J. A.: Photographic tone reproduction for digital images. In *SIGGRAPH* (2002), pp. 267–276.

[WBS*04] WANG Z., BOVIK A. C., SHEIKH H. R., MEMBER S., SIMONCELLI E. P., MEMBER S.: Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing 13* (2004), 600–612.

[WFA*05] WALTER B., FERNANDEZ S., ARBREE A., BALA K., DONIKIAN M., GREENBERG D.: Lightcuts: A scalable approach to illumination. *ACM Transactions on Graphics (Proc. of SIGGRAPH)* (2005), 1098–1107.

[Wil80] WILSON H.: A transducer function for threshold and suprathreshold human vision. *Biological Cybernetics 38* (1980), 171–178.

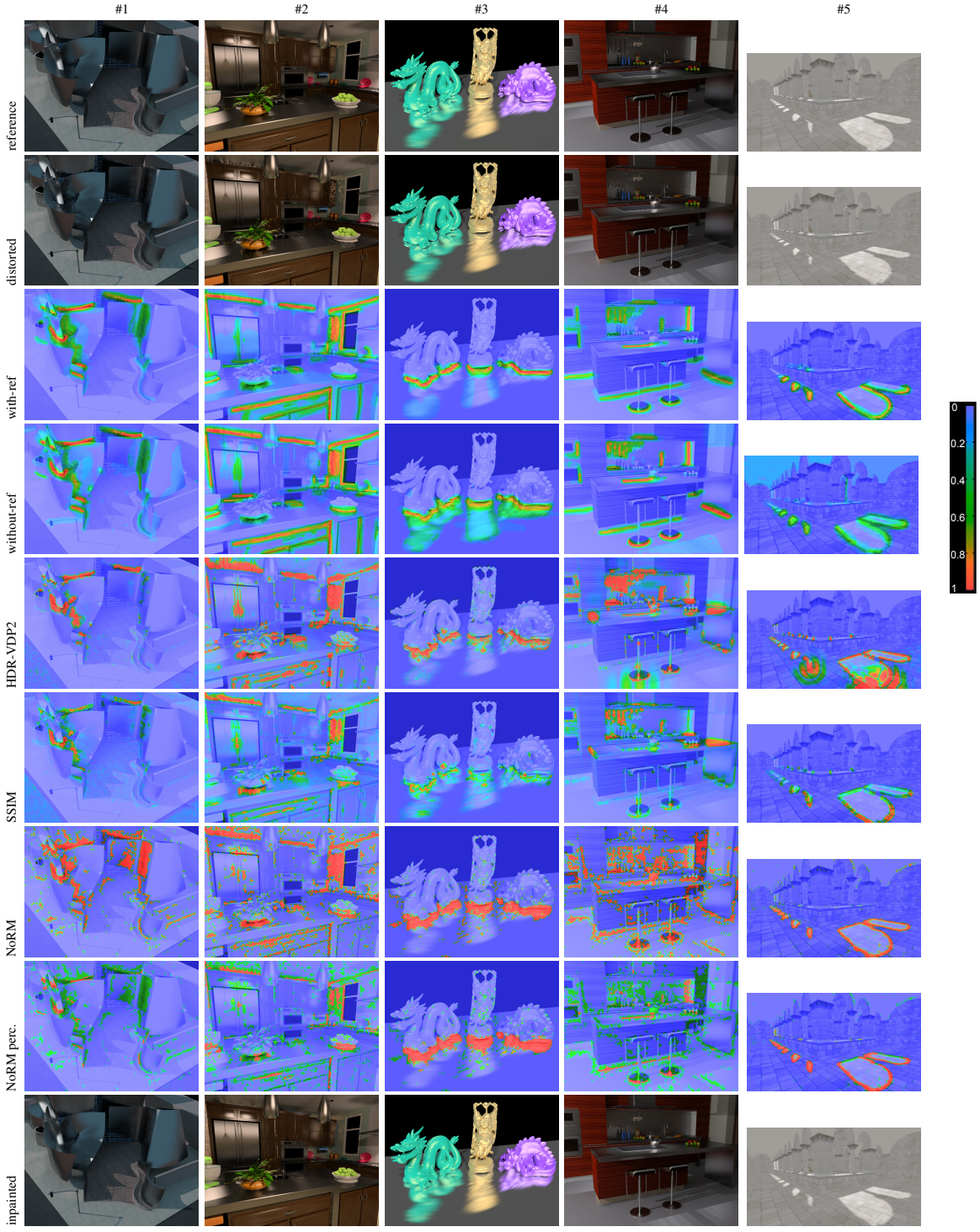

Figure 7: *Results of the user study: average subjective artifact strenghts, and the comparison to predictions of current state-of-the art full reference metrics as well as the proposed no-reference technique. First row: reference images, second row: images containing rendering artifacts, third row: average distortion maps for with-the-reference experiment, fourth row: average distortion maps for no-reference experiment, fifth row: predictions of the full-reference metric HDR-VDP2 [MKRH11], sixth row: predictions of the full-reference metric SSIM [WBS*04], seventh row: predictions of the proposed no-reference metric NoRM, eighth row: predictions of NoRM after the perceptual normalization, ninth row: artifact correction using inpainting.*

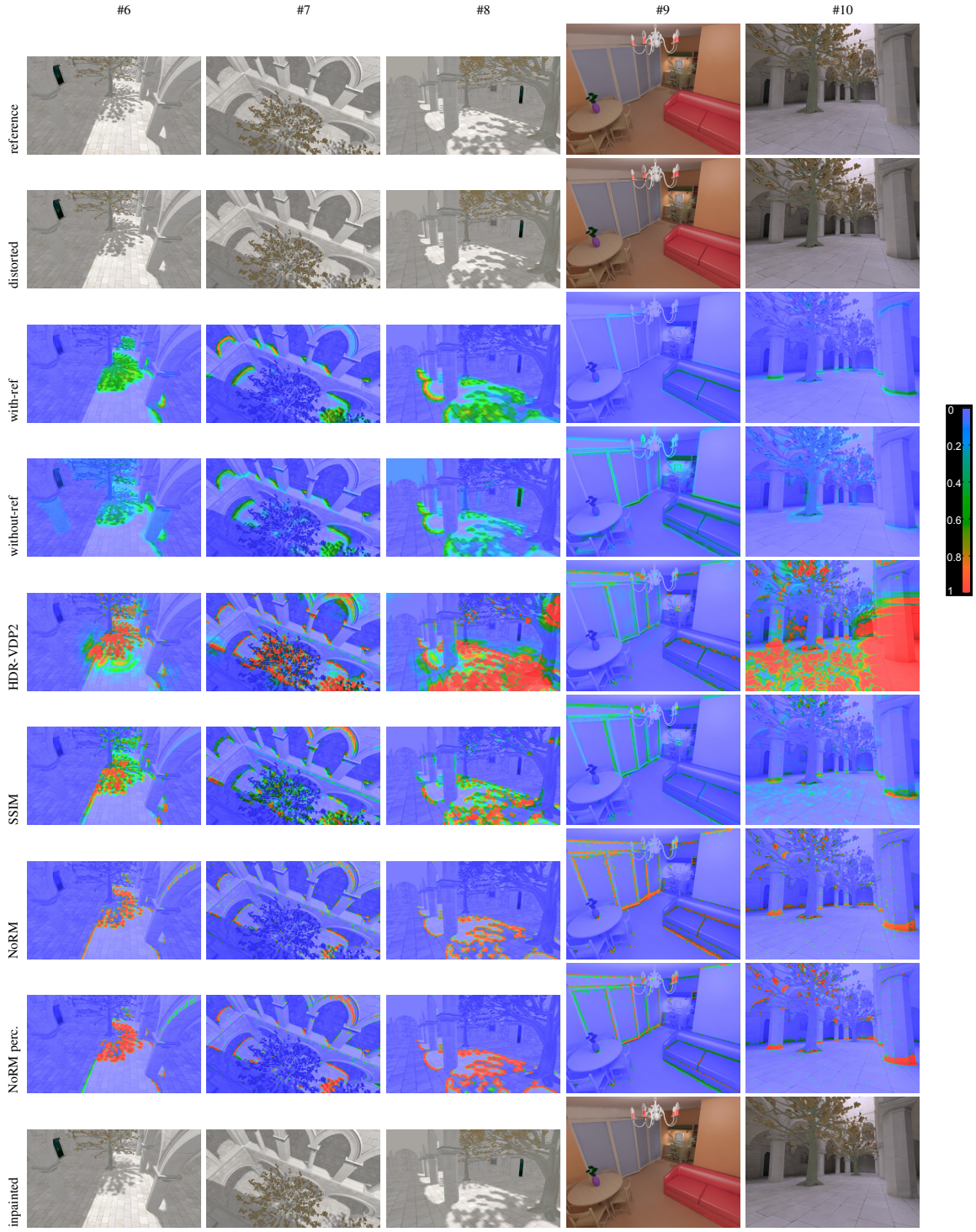

Figure 8: *Results of the user study: average subjective artifact strenghts, and the comparison to predictions of current state-of-the art full reference metrics as well as the proposed no-reference technique. First row: reference images, second row: images containing rendering artifacts, third row: average distortion maps for with-the-reference experiment, fourth row: average distortion maps for no-reference experiment, fifth row: predictions of the full-reference metric HDR-VDP2 [MKRH11], sixth row: predictions of the full-reference metric SSIM [WBS*04], seventh row: predictions of the proposed no-reference metric NoRM, eighth row: predictions of NoRM after the perceptual normalization, ninth row: artifact correction using inpainting.*