Dynamic Range Independent Image Quality Assessment

Tunç Ozan Aydın * Rafał Mantiuk * Karol Myszkowski * Hans-Peter Seidel *

MPI Informatik

Figure 1: *Quality assessment of an LDR image (left), generated by tone-mapping the reference HDR (center) using Pattanaik's tone-mapping operator. Our metric detects loss of visible contrast (green) and contrast reversal (red), visualized as an in-context distortion map (right).*

Abstract

The diversity of display technologies and introduction of high dynamic range imagery introduces the necessity of comparing images of radically different dynamic ranges. Current quality assessment metrics are not suitable for this task, as they assume that both reference and test images have the same dynamic range. Image fidelity measures employed by a majority of current metrics, based on the difference of pixel intensity or contrast values between test and reference images, result in meaningless predictions if this assumption does not hold. We present a novel image quality metric capable of operating on an image pair where both images have arbitrary dynamic ranges. Our metric utilizes a model of the human visual system, and its central idea is a new definition of visible distortion based on the detection and classification of visible changes in the image structure. Our metric is carefully calibrated and its performance is validated through perceptual experiments. We demonstrate possible applications of our metric to the evaluation of direct and inverse tone mapping operators as well as the analysis of the image appearance on displays with various characteristics.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image generation—display algorithms,viewing algorithms

Keywords: image quality metrics, high dynamic range images, visual perception, tone reproduction

1 Introduction

In recent years we have witnessed a significant increase in the variation of display technology, ranging from sophisticated high dynamic range (HDR) displays [Seetzen et al. 2004] and digital cinema projectors to small displays on mobile devices. In parallel to the developments in display technologies, the quality of electronic content quickly improves. For example luminance and contrast values, which are encoded in the so-called HDR images [Reinhard et al. 2005] correspond well with real world scenes. HDR images are already being utilized in numerous applications because of their extra precision, but reproduction of these images is only possible by adjusting their dynamic range to the capabilities of the display device using tone mapping operators (TMO) [Reinhard et al. 2002; Durand and Dorsey 2002; Fattal et al. 2002; Pattanaik et al. 2000]. The proliferation of new generation display devices featuring higher dynamic range introduces the problem of enhancing legacy 8-bit images, which requires the use of so-called inverse tone mapping operators (iTMO) [Rempel et al. 2007; Meylan et al. 2007]. An essential, but yet unaddressed problem is *how to measure the effect of a dynamic range modification on the perceived image quality*.

Typical image quality metrics commonly assume that the dynamic range of compared images is similar [Daly 1993; Lubin 1995; Wang and Bovik 2002]. They predict visible distortion using measures based on the magnitude of pixel intensity or normalized contrast differences between the two input images, which become meaningless when input images have significantly different contrast or luminance ranges. However, when we look at images on a computer screen or even on traditional photographs we often have an impression of plausible real world depiction, although luminance and contrast ranges are far lower than in reality. So, *the key issue in image reproduction is not obtaining an optical match, but rather plausible reproduction of all important image features and preserving overall image structure*. Such features improve the discrimination and identification of objects depicted in the image, which are important factors in image quality judgment [Janssen 2001]. The processed image structure can be affected by introducing visible artifacts such as blur, ringing, ghosting, halo, noise, contouring and blocking, which distort structure of the original image and may degrade the overall impression of image quality.

In this paper we present a novel image quality metric that can compare a pair of images with significantly different dynamic ranges. Our metric includes a model of the human visual system (HVS), and its main contribution is a new visible distortion concept based on the visibility of image features and the integrity of image structure (Section 3). The metric generates a distortion map that shows

*e-mail: {tunc, mantiuk, karol, hpseidel}@mpi-inf.mpg.de

the loss of visible features, the amplification of invisible features, and reversal of contrast polarity (Section 4). All these distortions are considered at various scales and orientations that correspond to the visual channels in the HVS. Novel features of our metric are tested (Section 5), and the overall metric performance confirmed in a psychophysical study (Section 6). We demonstrate application examples of our metric to predict distortions in feature visibility introduced by the state-of-the-art TMOs (Section 7.1) and inverse-TMOs (Section 7.2). Also, we analyze the influence of display dynamic range on the visibility of such distortions for three different displays (Section 7.3).

2 Previous Work

Image quality evaluation is important in many applications such as image acquisition, synthesis, compression, restoration, enhancement and reproduction. The topic is relatively well covered in a number of textbooks [Winkler 2005; Wang and Bovik 2006; Wu and Rao 2005]. Three important metric categories can be distinguished: metrics measuring contrast distortions, detecting changes in the image structure, and judging visual equivalence between images. In this section we discuss all these metric categories from the standpoint of their ability to handle image pairs of significantly different dynamic ranges.

The most prominent contrast distortion metrics such as the *visible difference predictor* (VDP) [Daly 1993] and the *Sarnoff visual discrimination model* (VDM) [Lubin 1995] are based on advanced models of the HVS and are capable of capturing just visible (near threshold) differences or even measuring the magnitude of such differences and scale them in JND (just noticeable difference) units. While these metrics are designed for LDR images, Mantiuk et al. [2005] proposed an HDR extension of VDP, that can handle the full luminance range visible to the human eye. iCAM06 [Kuang et al. 2007] has similar capabilities, but it also models important aspects of color appearance. While, the iCAM06 framework has been mostly applied in tone mapping applications, it has a clear potential to compute HDR image difference statistics and to derive from them image quality metrics. Recently, Smith et al. [2006] proposed an objective tone mapping evaluation tool, which focuses on measuring suprathreshold contrast distortions between the source HDR image and its tone mapped LDR version. The main limitation of this metric is that it is based on the contrast measure for neighboring pixels only, which effectively means that its sensitivity is limited to high frequency details. Also, the metric calibration procedure has not been reported, while it may be expected that the metric may be excessively sensitive for small near-threshold distortions because the peak sensitivity is assumed for each luminance adaptation level instead of using contrast sensitivity function.

An important trend in quality metrics has been established with the development of *structural similarity index metric* (SSIM) by Wang and Bovik [2002]. Since the HVS is strongly specialized in learning about the scenes through extracting structural information, it can be expected that the perceived image quality can be well approximated by measuring structural similarity between images. SSIM proved to be extremely successful in many image processing applications, it is easy to implement, and very fast to compute. As the authors admit [Wang et al. 2003], a challenging problem is to calibrate its parameters, which are quite "abstract" and thus difficult to derive from simple-stimulus subjective experiments as it is typically performed for contrast-based metrics. For this reason it is difficult to tell apart visible and non-visible (just below threshold) structure changes, even for multi-scale SSIM incarnations [Wang et al. 2003]. SSIM is sensitive for local average luminance and contrast values, which makes it inadequate for comparing LDR and HDR images. Recently, Wang and Simoncelli [2005] proposed the CW-SSIM metric, which in its formulation uses complex wavelet coefficients instead of pixel intensities employed in the SSIM. Since in CW-SSIM bandpass wavelet filters are applied, the mean of the wavelet coefficients is equal to zero in each band, which significantly simplifies the metric formulation with respect to the SSIM and makes it less sensitive to uniform contrast and luminance changes. However, this reduced sensitivity concerns rather small changes of the order 10–20%, which are not adequate for comparing HDR and LDR images.

An interesting concept of *the visual equivalence predictor* (VEP) has been recently presented by Ramanarayanan et al. [2007]. VEP is intended to judge whether two images convey the same impression of scene appearance, which is possible even if clearly visible differences in contrast and structure are apparent in a side-by-side comparison of the images. The authors stress the role of higher order aspects in visual coding, but developing general computational model for the VEP is a very difficult task. The authors show successful cases of the VEP models for different illumination map distortions, which also requires some knowledge about the scene geometry and materials. While the goals of VEP and our metric are different, both approaches tend to ignore certain types of visual differences, which seem to be unimportant both for the scene appearance and image structure similarity judgements.

Our metric can be considered as a hybrid of contrast detection and structural similarity metrics. Careful HVS modeling enables precise detection of only visible contrast changes, but instead of reporting such changes immediately as VDP, HDR-VDP, and VDM metrics, we use the visibility information to analyze only visible structure changes. We distinguish three classes of structure changes, which provides with additional insight into the nature of structural changes compared to SSIM. Finally, what makes our approach clearly different from existing solutions is the ability to compare images of drastically different dynamic ranges, which broadens the range of possible applications.

3 Image Distortion Assessment

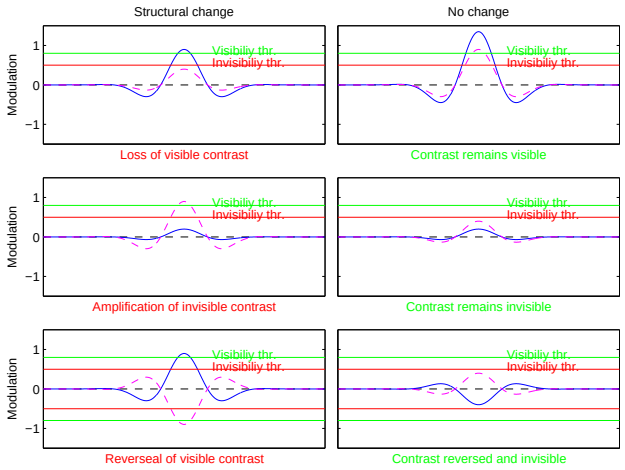

Figure 2: *Several cases of contrast modification, that our metric classifies as a structural change (left) or a lack of structural change (right). Blue continuous line – reference signal; magenta dashed line – test signal. For the explanation of visibility and invisibility threshold (50% probability) refer to the text and Figure 5.*

Instead of detecting contrast changes, our metric is sensitive to three types of structural changes:

Loss of visible contrast happens when a contrast that was visible in the reference image becomes invisible in the test image. This typically happens when a TMO compresses details to the level that they become invisible.

Amplification of invisible contrast happens when a contrast that was invisible in the reference image becomes visible in the test image. For instance, this can happen when contouring artifacts starts to appear due to contrast stretching in the inverse TMO application.

Reversal of visible contrast happens when a contrast is visible in both reference and test images, but has different polarity. This can be observed at image locations with strong distortions, such as clipping or salient compression artifacts.

An intuitive illustration of the three types of distortions is shown in Figure 2 [1]. Note that this formulation makes our metric invariant to differences in dynamic range or to small changes in the tone-curve.

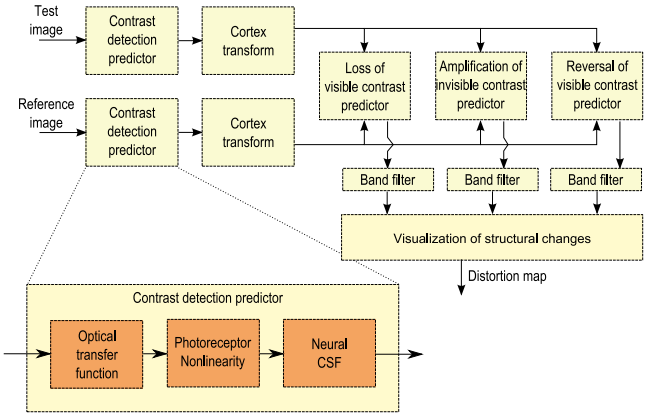

Figure 3: *The data flow diagram of our metric.*

Before we can detect any of the three types of distortions, we need to predict whether a contrast is visible or not. This is achieved with the metric outlined in Figure 3. The input to our metric are two luminance maps, one for a reference image (usually an HDR image), and one for a test image (usually an image shown on the display). 8-bit images must be transformed using the display luminance response function to give actual luminance values shown on a screen. In the first step we predict detection thresholds and produce a perceptually normalized response map, in which the amplitudes equal to 1 correspond to the detection threshold at $P_{det} = 75\%$ (1 JND). Although several such predictors have been proposed in the literature, we found the HDR-VDP detection model [Mantiuk et al. 2005], designed especially for HDR images, the most appropriate. The predictor takes into account light scattering in the eye's optics, non-linear response of the photoreceptors and spatial-sensitivity changes due to local adaptation. For completeness, we summarize the HDR-VDP contrast detection predictor in the Appendix.

To ensure accurate predictions, we calibrated the HDR-VDP detection model with the ModelFest [Watson 2000] measurements. The ModelFest data set was collected in a number of different laboratories to enhance both the generality and accuracy, and was especially designed to calibrate and validate vision models. Figure 4 shows a few examples of the detection probability maps for stimuli below, at and above the detection threshold. All results were generated by setting the pixels per visual degree to 120, and observer distance

[1] Refer to supplemental material for metric responses to similar distortions

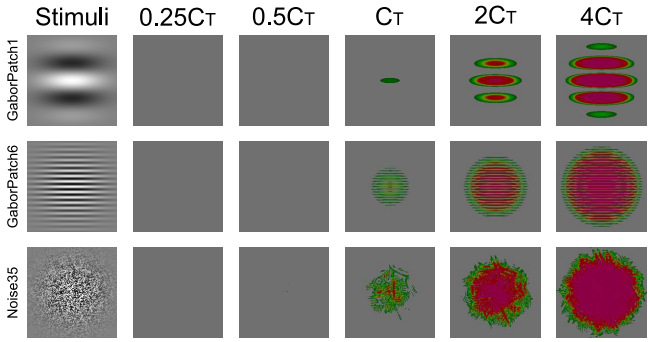

Figure 4: *The output of the detection predictor for the selected ModelFest stimuli at 0.25, 0.5, 1, 2 and 4 times the detection threshold, C_T. The first column shows the original stimuli at high contrast. The predictor is well calibrated if the visible contrast starts to be signalled in the C_T column.*

to $2m$. The model fitting error for 0.25% peak sensitivity was below 2dB contrast units. The errors were the largest for the stimuli "GaborPatch14" and "Dipole32", for which our predictor was too sensitive.

In the second step, we split the perceptually normalized response into several bands of different orientation and spatial bandwidth. We employ the cortex transform [Watson 1987] with the modifications from [Daly 1993], given in the Appendix. Then, to predict three types of distortions separately for each band, we compute conditional probabilities of

$$\begin{aligned} &\text{loss of visible contrast:} && P^{k,l}_{loss} = P^{k,l}_{r/v} \cdot P^{k,l}_{t/i}, \\ &\text{amplification of invisible contrast:} && P^{k,l}_{ampl} = P^{k,l}_{r/i} \cdot P^{k,l}_{t/v}, \\ &\text{and reversal of visible contrast:} && P^{k,l}_{rev} = P^{k,l}_{r/v} \cdot P^{k,l}_{t/v} \cdot R^{k,l} \end{aligned} \tag{1}$$

where k and l are the spatial band and orientation indices, the subscript $r/\cdot$ denotes reference and $t/\cdot$ test image, the subscript $\cdot/v$ visible and $\cdot/i$ invisible contrast. R equals 1 if the polarity of contrast in the reference and test images differ:

$$R^{k,l} = \left[C^{k,l}_r \cdot C^{k,l}_t < 0 \right] \tag{2}$$

For simplicity we omit the pixel indices (x, y). The above formulation assumes that that contrast detection process is performed in the visual system separately for each visual channel.

The probabilities $P_{\cdot/v}$ and $P_{\cdot/i}$ are found from the detection probabilities, as shown in Figure 5. The visual models commonly assume that a contrast is visible when it is detectable ($P_{det} \geq 75\%$), as in the two alternative forced choice (2AFC) experiments. We found this assumption to be too conservative, since complex images are never as scrutinously observed as stimuli in such experiments. Therefore, we require a contrast to be detected with a higher probability, to be regarded as visible. From our empirical study on a series of simplified stimuli, we found that a reliable predictor of visible contrast is given by shifting the psychophysical function, so that a contrast magnitude is *visible* with 50% probability, if it can be *detected* by our predictor with 95% probability (about 2 JND), as shown in Figure 5. The probability of invisible contrast is given by the negation of the probability of detection.

The rules from Equation 1 contain the non-linear operators, therefore the resulting probability map $P^{k,l}_{\cdot}$ can contain features of spatial frequency that do not belong to a particular subband. This leads

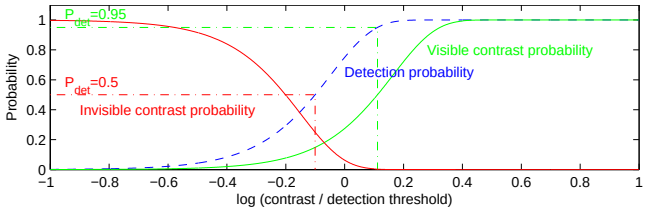

Figure 5: *Probability functions for a normalized contrast magnitude being visible (green) and invisible (red).*

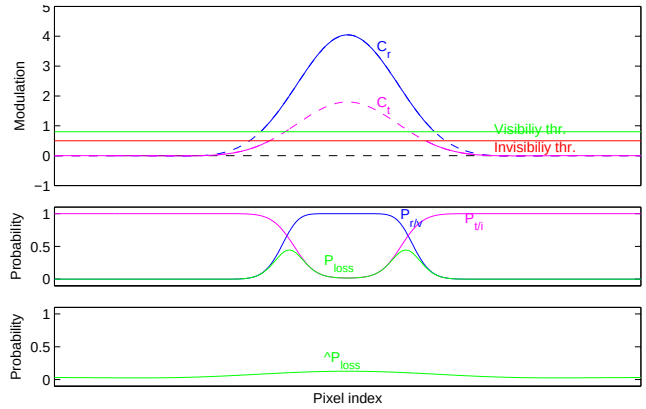

Figure 6: *The probability rules may produce response that do not belong to a particular frequency band. Top pane: although a contrast magnitudes are well above visibility threshold, there is a small part in which contrast is visible in the reference image (C_r) but not visible in a test image (C_t). Center pane: this triggers higher values of the P_{loss} in these regions. Bottom pane: the spurious responses can be eliminated with a band-pass filter.*

to spurious distortions, as shown in Figure 6. To avoid this problem, each probability map is filtered once more using the corresponding cortex filter $B^{k,l}$:

$$\hat{P}^{k,l}_{loss} = \mathscr{F}^{-1}\left\{\mathscr{F}\{P^{k,l}_{loss}\} \cdot B^{k,l}\right\} \tag{3}$$

where $\mathscr{F}$ and $\mathscr{F}^{-1}$ are the 2D Fourier transforms. Formulas for $B^{k,l}$ can be found in the Appendix.

Assuming that detection of each distortion in each band is an independent process, the probability that a distortion will be detected in any band is given by:

$$P_{loss} = 1 - \prod_{k=1}^{N}\prod_{l=1}^{M}\left(1 - \hat{P}^{k,l}_{loss}\right) \tag{4}$$

The probability maps P_{ampl} and P_{rev} are computed in a similar way.

Unlike typical HVS-based contrast difference predictors, our metric does not model contrast masking (decrease in sensitivity with increase of contrast amplitude). Since our metric is invariant to suprathreshold contrast modifications, contrast masking does not affect its result. For example, if we compare two visible contrast stimuli, like the ones shown in top-right pane of Figure 2, the contrast masking can predict by how many JNDs their amplitudes differ. But the contrast difference is not relevant for our metric, therefore there is no need to estimate the magnitude of suprathreshold contrast in JND units.

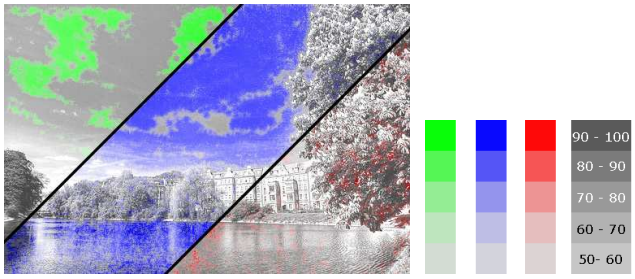

Figure 7: *Three distortion maps shown partially (left). We arbitrarily chose green for loss of visible contrast, blue for amplification of invisible contrast, and red for reversal of visible contrast. The saturation of each color indicates the magnitude of detection probability, as shown in the respective scales.*

4 Visualization of Distortions

The multitude of distortion types detected by our metric makes visualization of the outcome on a single image a challenging task. We employ an in-context distortion map [Daly 1993] approach to provide an overview of distortions, but also introduce a custom viewer application for more detailed inspections.

To generate the in-context map, luminance of the distorted image is copied to all three RGB channels, and each channel is scaled by the detection probabilities of corresponding distortion type. We observed that using multiple colors for each type of distortion makes it hard to memorize the association of each color to the correct distortion type. We also found that in regions where multiple distortions overlap, the simple approach of blending the colors makes the final map less intuitive by increasing the number of colors. We therefore show only the distortion with the highest detection probability at each pixel location. We arbitrarily chose **green** for loss of visible contrast, **blue** for amplification of invisible contrast, and **red** for reversal of visible contrast (Figure 7).

In cases where the test image is heavily distorted the in-context map representation may become too cluttered, and there may be significant overlaps between different distortion types. On the other hand, one may simply be interested in a closer examination of each distortion type present in the image. Using the viewer application one can dynamically set the opacity values of distortion types and the background image to a legible configuration, that allows to investigate distortions separately (Figure 8). In the rest of this paper, all metric responses are presented as in-context maps. The viewer application can be used for any further investigation of the results[2].

5 Evaluation and Results

In the following sections, we present results and demonstrate advantages of our metric to previous work [3].

5.1 Dynamic Range Independence

We claim that our metric generates meaningful results even if the input images have different dynamic ranges, in addition to the case where both have the same dynamic range. In Figure 9, we show the distortion maps resulting from the comparison of all variations of an HDR and LDR image. The LDR image is generated by applying

[2]Refer to the supplemental material for the viewer application

[3]Refer to the supplemental material for a simple stimuli experiment

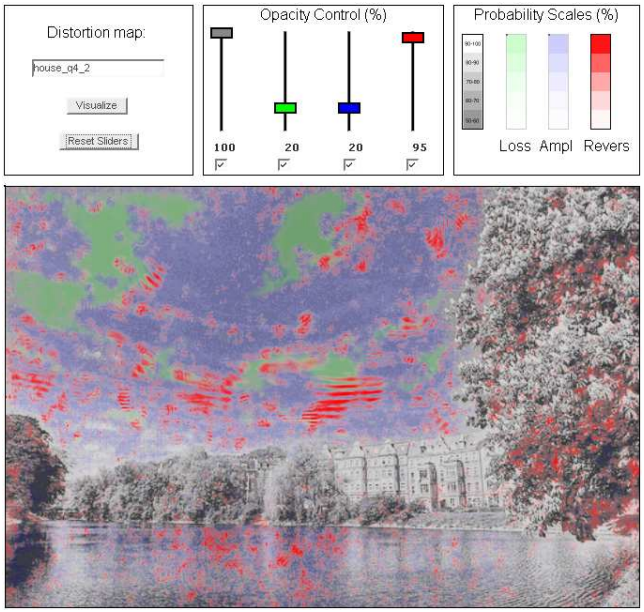

Figure 8: *Our distortion viewer. Users can adjust opacities of distortion maps and background image. The respective scales (top right) are adjusted accordingly by the tool. In this example setting, the user emphasizes contrast reversal, while keeping the other distortions barely visible.*

a compressive power function to the HDR reference (more sophisticated tone-mapping operators are discussed in Section 7.1). We always distort the test image by locally adding random pixel noise, whose magnitude is modulated with a Gaussian that has its peak at the center of the distorted region.

Randomly distributed pixels in the distorted region both introduce previously non-existent contrast and invert the polarity of the contrast proportional to the magnitude of the distortion. Consequently, for both HDR-HDR and LDR-LDR cases (first two rows) our metric reports visible contrast reversal and amplification of invisible contrast confined in the distorted region. Similar responses are also observed in LDR-HDR and HDR-LDR cases. Additionally, a comparison of the distorted LDR image with an HDR reference yields to an overall loss of visible contrast spread across the entire image, indicating the effect of contrast compression applied to the test image (third row). When we compare the HDR test image with the LDR reference, visible contrast of the reference lost during compression manifests itself this time as amplification of invisible contrast in the distortion map (last row).

5.2 Comparison with Other Metrics

Our metric has two major advantages to the previous work: classification of distortion types, and dynamic range independence. In this section, we compare responses of our metric with a pair of state-of-the-art metrics, namely SSIM [Wang and Bovik 2002] that predicts changes in the image structure, and HDR-VDP [Mantiuk et al. 2005] that is explicitly designed for HDR images. Figure 10 shows a side-by-side comparison of the three metrics where a blurred and a sharpened version of the reference was used as test image. The reference is an 8-bit image, which is linearized and converted to luminance for HDR-VDP and our metric. The outcome of SSIM is a simple matrix of probability values with the same size as the input images, to which we applied HDR-VDP's visualization algorithm to make it legible. The spatial distribution of the responses from all

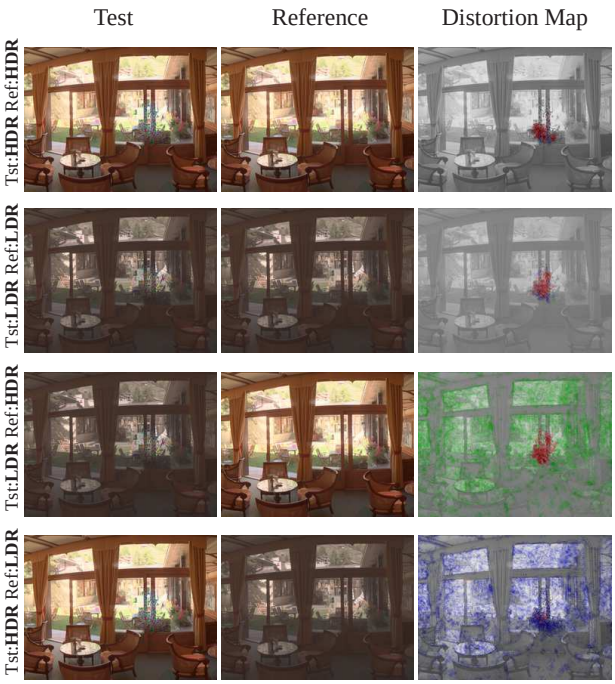

Figure 9: *Comparing images with different dynamic ranges. While distortions caused by the local distortion are visible in all results, in the LDR-HDR and HDR-LDR cases, additional visible contrast loss and invisible contrast amplification can be observed due to the contrast lost through dynamic range compression. HDR images are tone-mapped using Reinhard's photographic tone reproduction for printing purposes.*

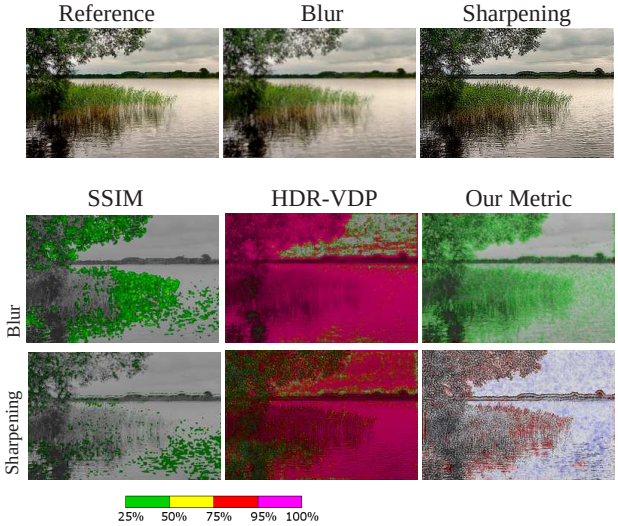

Figure 10: *The reference, blurred and sharpened test images (top row), and metric responses to blurring (middle row) and sharpening (bottom row). Color coding for SSIM and HDR-VDP are given in the scale. Our metric is visualized as discussed in Section 4*

.

three metrics to blurring and sharpening is similar, with the overall tendency of HDR-VDP's response being stronger (due to reporting all visible differences) and SSIM's response being weaker (due to the difficulty of calibration) than that of our metric.

The important difference between the proposed metric and others is the classification of distortion types. That is, in case of blurring our metric classifies all distortions as loss of visible contrast, confirming the fact that high frequency details are lost. On the other hand, in the sharpening case we observe contrast reversal and amplification of invisible contrast, both of which are expected effects of unsharp masking. Such a classification gives insight about the nature of the image processing algorithm and enables distortion-type-specific further processing.

The second major advantage of our metric is that it enables a meaningful comparison of images with different dynamic ranges (Section 5.1). We ran all three metrics on a test set, that is generated using a similar procedure as used for Figure 9, with the only difference being the use of Gaussian blur as the distortion type. HDR images in the test set were calibrated to absolute luminance values of the scene, and were directly passed to both our metric and HDR-VDP. For SSIM, we took the 10-base logarithm of the HDR images to compensate for the Weber law, and mapped them to pixel values within 0-255 to prevent an ambiguity in the dynamic range parameter of the metric.

Figure 11 shows a comparison of images with same dynamic range results in all three metrics reporting distortions in the blurred region with slightly different magnitudes (first two rows). One important difference between our metric's and HDR-VDP's responses is that the distorted area reported by HDR-VDP is larger than that of our metric's. HDR-VDP simply reports all visible differences of the blurred test images with respect to their references, while our metric ignores the differences in the periphery of the Gaussian, where the magnitude of the blur is weaker and details in the distorted image are still visible. This example shows a case where our metric provides complementary information to well established metrics. In the different dynamic range case, the distortion maps of SSIM and HDR-VDP are entirely dominated by contrast change due to the dynamic range compression (last two rows). Similar to the results for different dynamic range case in Figure 9, our metric reports an overall loss of visible contrast in the LDR-HDR case, and an overall amplification of invisible contrast in the HDR-LDR case, both due to the dynamic range compression. These responses, however, do not mask the response at the blurred region, as they do with the other metrics.

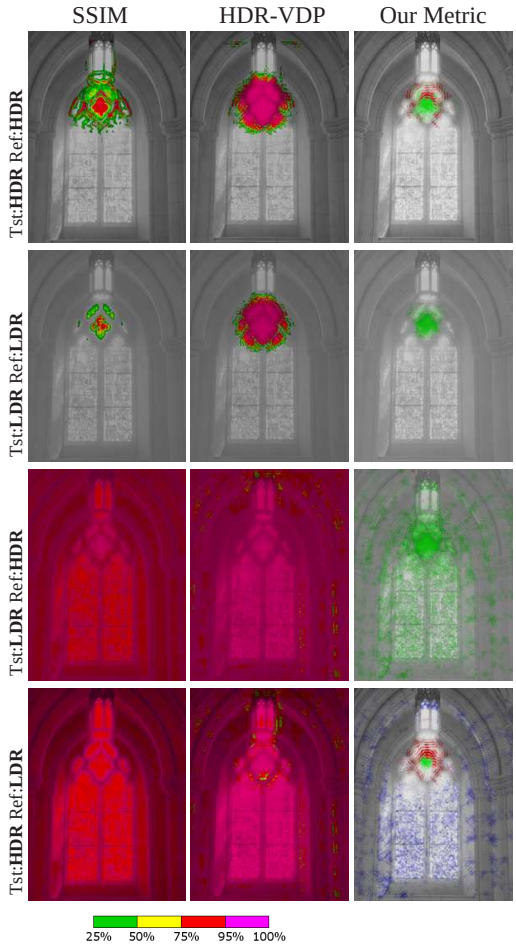

Figure 11: *A comparison of SSIM, HDR-VDP and our metric on all dynamic range combinations. Results for the same dynamic range case are comparable (first two rows), whereas in the different dynamic range case SSIM and HDR-VDP responses are dominated by the dynamic range difference (last two rows). The scale shows the color coding for SSIM and HDR-VDP. Our metric is visualized as discussed in Section 4*

6 Validation

Validation of the metric is performed by comparing the metric responses to subjective distortion assessments. We generated a test set containing permutations of 3 images of natural scenes, 3 types of distortions and 3 levels of distortions. Each subject evaluated the entire test set twice to ensure reliability, leading to 54 images per subject. Gaussian blur that produces visible contrast loss, and unsharp masking that mostly produces invisible contrast amplification were chosen as distortions. Another type of distortion was considered to specifically produce contrast reversal, where we calculate a bandpass image pyramid, invert the signs of a number of layers proportional to desired distortion level, and recombine the pyramid to get the distorted image. All distorted images were generated to dominantly produce a metric response of the desired type.

We asked 14 subjects within the ages $23-48$, with all nearly perfect or corrected vision, to *identify the type of distortion* they see on a number of test images. Possible answers were *blur*, *sharpening*, *contrast reversal* or *no distortion*. We assumed no prior knowledge

Figure 12: *A sample image from the validation set, showing three levels of sharpening (top row), and the corresponding metric responses (bottom row) increasing from left to right.*

of the subjects about the distortion types. Therefore, a short training section preceded the actual experiment, where subjects were shown a series of images that contain strong distortions of each of the three types, together with the correct distortion labels.

In order to account for the variation of subject responses to different distortion magnitudes, we applied all distortions at three different levels, from which the first is selected to generate no metric response at all. The second level was chosen to generate a weak metric response of the desired type, where the detection probability at most of the distorted pixels is less than one. Similarly, the third level was chosen to generate a strong metric response in a noticeably large region. In our statistical analysis, we considered the first level as invisible, and the other two as visible. Since our metric is not intended to produce a single number, we restrained ourselves from using an average of the detection probabilities within the distorted region.

First, we examined subject reliability by testing the stochastic independence of the consecutive iterations for each subject. Using the χ^2 test we obtained a $\chi^2(9)$ value of 739.105, where the value in parenthesis denotes the number of degrees of freedom. The corresponding $p - value$ was found to be $\ll 0.05$, indicating that the null-hypothesis can safely be rejected. The Cramer's V [Cramér 1999], that measures the association between two categorical variables, is found to be 0.807 which is considered a large effect size. Next, we investigated the main effect of factors using the ANalysis Of VAriance (ANOVA) method (See [D'Agostino 1972] for the use of ANOVA on nominal data). We found that distortion type and level to have a significant effect on the subject response ($F(2) = 179.96$ and $F(2) = 456.20$ respectively, and $p \ll 0.01$ for both). We also found that the test image factor ($F(2) = 4.97$ and $p = 0.02$) to have an effect on the final outcome, which is hard to avoid when experimenting with complex stimuli. Finally, we analyzed the statistical dependency between the subject and metric responses. For the null-hypothesis that these responses are independent, we found $\chi^2(9) = 1511.306$ and $p \ll 0.05$, showing that it is unlikely that the initial assumption holds. The corresponding Cramer's V of 0.816 signals a strong dependency between the metric and subject responses.

7 Applications

In this section, we present several application areas of our metric, where a comparison of images with different dynamic ranges is required.

7.1 Tone Mapping Operator Comparison

Tone mapping operators (TMO) are commonly used for contrast compression of HDR images to reproduce them properly on conventional media. This is a lossy process by definition. From a functional point of view, information reproduction capability of a TMO is a suitable measure of its performance. Figure 13 shows the comparison result of an HDR image with the corresponding tone mapped images. The luminance ranges of 0.24–89,300 and 0.1–80 cd/m^2 have been assumed for the original scene and displayed tone mapped image, respectively. Five TMOs (2 global and 3 local operators) have been considered: Drago's adaptive logarithmic mapping [2003], Pattanaik's visual adaptation model [2000], Fattal's gradient domain compression [2002], Durand's bilateral filtering [2002], and Reinhard's photographic tone reproduction [2002].

For all studied TMOs certain detail loss can be observed in the bright lamp region due to strong contrast compression. Pixel intensity clipping also causes visible contrast reversal in the lamp region, which is reported for some pixels as the strongest distortion. Drago's operator reproduces contrast relatively well in dark image regions and tends to wash out image details in bright regions due to logarithmic shape of the tone mapping curve. Pattanaik's operator, which is based on the sigmoid photoreceptor response (mostly adapted to the luminance levels at the illuminated table regions), tends to strongly suppress image details in dark regions, but also in very bright highlights. The detail amplification typical for Fattal's operator can be seen in non-illuminated scene regions, which in real-world observation conditions are not visible due to insufficient HVS sensitivity. Our metric takes into account this sensitivity by modeling the dependence of contrast sensitivity function on luminance values in the HDR image. Durand's operator uniformly compresses lower spatial frequencies across the entire image, which means that resulting contrast loss will be more likely visible in dark display regions in which the HVS sensitivity is lower. The compression of low frequency features leads also to the reversal of visible contrast. The default parameters used for Reinhard's operator tend to excessively saturate bright image regions for this particular scene. Also, in the full size image it can be seen that contrast of certain pixels representing the table and paper page textures has been magnified due to local dodging and burning mechanism. Our results are consistent with the expected outcomes of the TMO's, indicating the potential use of our metric as a diagnostic tool for such algorithms.

7.2 Inverse Tone Mapping Evaluation

Recently, [Meylan et al. 2007] and [Rempel et al. 2007] attacked the problem of recovering the contrast in LDR images that has been clipped and/or compressed due to the limited dynamic range. These algorithms should be validated by costly subjective user studies to assess the plausibility of the results and the amount of visible artifacts [Akyüz et al. 2007]. The latter task can be fulfilled much more efficiently by our metric.

The response of our metric to simple contrast stretching with clipping is shown in Figure 14. To exaggerate the contouring artifacts, we use a 4-bit quantized version of the 8-bit reference as our test image. We observe that the more we increase image contrast, the more visible contrast in the bright sky region is lost, and invisible contrast in the darker horizon line is amplified, both due to clipping on both sides of the expanded image histogram. Our metric also reports contrast reversal on the boundaries within the visible and clipped contrast regions. In Figure 15, we show the comparison of an HDR image reconstructed by Ldr2Hdr [Rempel et al. 2007] algorithm, with the reference LDR image image. The increase in contrast due to stretching reveals some previously invisible details around the trees in the foreground, which is correctly reported by our metric. Contrast content amplified in bright regions, however, was already visible, and therefore is not interpreted as a structural change.

7.3 Simulation of Displays

The highly diverse characteristics of today's display devices make an objective analysis of their reproduction capability an interesting problem. Our metric can be used as a measure of how well the structural information of the image is preserved when it is viewed on different displays, to ensure that important features of the image are preserved regardless of the display type.

In Figure 16 we show the distortion maps for an HDR reference image that is viewed on an BrightSide DR37-P HDR display ($2,005 cd/m^2$), Barco Coronis 3MP LCD display ($400 cd/m^2$), and a Samsung SGH-D500 cell phone display ($30 cd/m^2$). To simulate the HDR and LCD displays, we apply the respective display

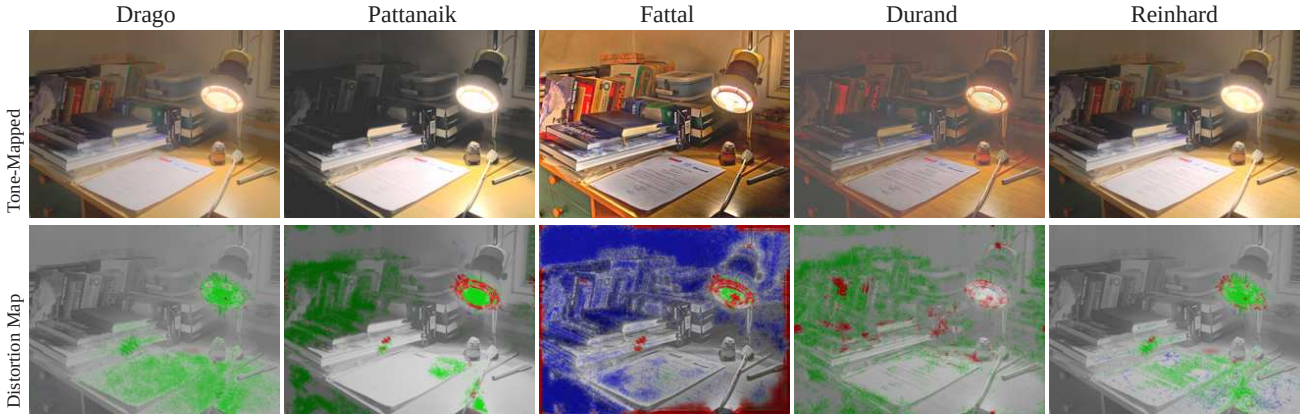

Figure 13: *Comparison of Tone-Mapping Operators*

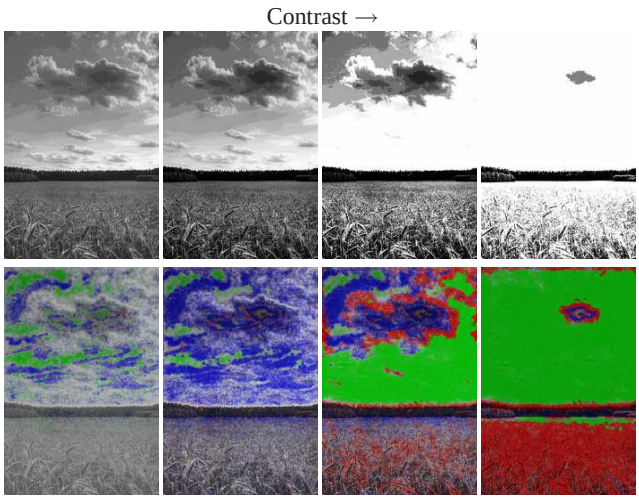

Figure 14: *Response of the metric to simple contrast stretching with clipping. Contrast is increased from left to right, which results in more clipping and generates stronger visible contrast loss and reversal responses.*

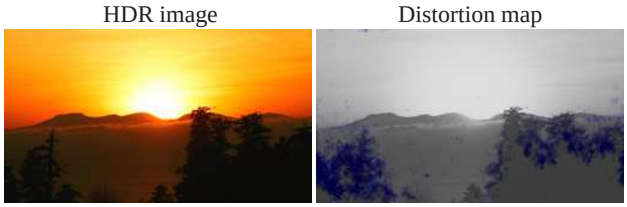

Figure 15: *HDR image generated by Ldr2Hdr algorithm (left), and the distortion map obtained by comparing the HDR image with the LDR reference (right). Both images are taken from the original author's website.*

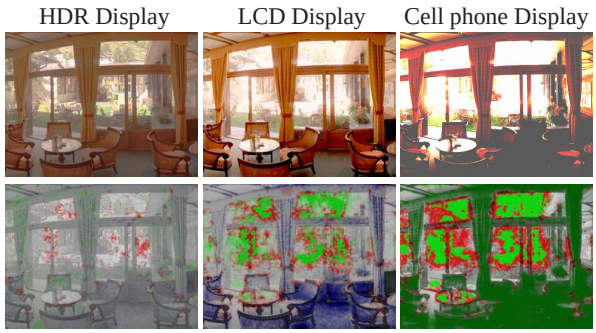

Figure 16: *Display Comparison. The brightness of the LCD (first row center) and Cell phone (first row right) display images are artificially enhanced for maximum detail visibility.*

response functions to image luminance values using a Minolta LS-100 luminance meter.

The results show that the HDR display faithfully reproduces most of the visible and invisible contrast. The small amount of distortion is expected, as even the dynamic range of the HDR display does not span the entire visible luminance range. The distortion map for the LCD display shows visible contrast loss in the outside region directly illuminated by sunlight. This luminance level exceeds the capabilities of the display device and therefore details are clipped. On the other hand, we observe invisible contrast amplification in parts of the darker interior region. This is because these regions in the reference image are so dark that the fine details at the chairs and floor are not visible. But since the LCD display is not capable of displaying such low luminance, those details are amplified above the visibility threshold. Finally, the cell phone display fails to reproduce most of the visible contrast, and hence we observe strong visible contrast loss in both the interior and exterior regions, as well as contrast reversal around the borders of the clipped regions.

8 Conclusion

We presented a quality assessment metric capable of handling image pairs with arbitrarily different dynamic ranges. Our metric classifies structural image distortions into three intuitive categories, revealing additional information about the nature of the test image compared to previous approaches. To visualize all distortion types

legibly, we provide a visualization tool in addition to the commonly used in-context map. We carefully calibrated the human visual system model employed in the metric, and performed a series of psychophysical experiments for statistical validation. We presented successful applications of our metric to TMO and iTMO operator evaluation, and comparison of various types of displays.

As future work, we intend to test our metric in medical applications which require faithful reproduction of details captured by HDR sensors in the displayed images. It would be also interesting to try our metric in watermarking applications, which require reproducing images on various media.

Acknowledgements

The authors would like to thank Matthias Ihrke for his valuable contribution to the statistical evaluation of the validation data.

Appendix

We summarize a complete contrast detection predictor employed in the HDR-VDP [Mantiuk et al. 2005], including the models from the VDP [Daly 1993]. The structure follows the three processing blocks shown in the bottom-left of Figure 3. For simplicity, we omit pixel and frequency component indices.

To account for light scattering in the eye's optics, the initial luminance map L is filtered with the Optical Transfer Function (OTF):

$$L_{OTF} = \mathscr{F}^{-1}\left\{\mathscr{F}\{L\} \cdot OTF\right\} \tag{5}$$

using the Deeley et al. model:

$$OTF = \exp\left[-\left(\frac{\rho}{20.9 - 2.1d}\right)^{1.3-0.07d}\right] \tag{6}$$

where d is a pupil diameter in mm and ρ is spatial frequency in cycles per degree. The pupil diameter is calculated for a global adaptation luminance L_{ga} using the formula of Moon and Spencer [1944]:

$$d = 4.9 - 3\ tanh[0.4(\log_{10}(L_{ga}) + 1)] \tag{7}$$

The global adaptation luminance L_{ga} is a geometric mean of the luminance map L.

Then, to account for lower sensitivity of the photoreceptors at low luminance, the map L_{OTF} is transformed using a transducer function constructed from the peak detection thresholds. The easiest way to find such a transducer function is to use the recursive formula:

$$T_{inv}[i] = T_{inv}[i-1] + cvi(T_{inv}[i-1])\ T_{inv}[i-1] \quad \text{for } i = 2..N \tag{8}$$

where $T_{inv}[1]$ is the minimum luminance we want to consider ($10^{-5}\ cd/m^2$ in our case). The actual photoreceptor response R is found by linear interpolation between the pair of i values corresponding to particular luminance L_{OTF}.

The contrast versus intensity function cvi used in the recursive formula above estimates the lowest detection threshold at a particular adaptation level:

$$cvi(L_{la}) = \left(\max_{\mathbf{x}}\left[CSF(L_{la}, \mathbf{x})\right]\right)^{-1} \tag{9}$$

where CSF is the contrast sensitivity function and $\mathbf{x}$ are all its parameters except adapting luminance. If perfect local adaptation is assumed, then $L_{la} = L_{OTF}$.

The CSF [Daly 1993] is given by:

$$CSF(\rho, \theta, L_a, i^2, d, c) = P \cdot \min\left[S_1\left(\frac{\rho}{r_a \cdot r_c \cdot r_\theta}\right), S_1(\rho)\right], \tag{10}$$

where

$$\begin{aligned}
r_a &= 0.856 \cdot d^{0.14} \\
r_c &= \tfrac{1}{1+0.24c} \\
r_\theta &= 0.11\cos(4\theta) + 0.89 \\
S_1(\rho) &= \left[\left(3.23(\rho^2 i^2)^{-0.3}\right)^5 + 1\right]^{-\frac{1}{5}} \cdot \\
&\quad \cdot A_l \epsilon \rho e^{-(B_l \epsilon \rho)} \sqrt{1 + 0.06 e^{B_l \epsilon \rho}} \\
A_l &= 0.801\left(1 + 0.7\ L_a^{-1}\right)^{-0.2} \\
B_l &= 0.3\left(1 + 100\ L_a^{-1}\right)^{0.15}
\end{aligned} \tag{11}$$

The parameters are: ρ – spatial frequency in cycles per visual degree, θ – orientation, L_a – the light adaptation level in cd/m^2, i^2 – the stimulus size in deg^2 ($i^2 = 1$), d – distance in meters, c – eccentricity ($c = 0$), ϵ – constant ($\epsilon = 0.9$), and P is the absolute peak sensitivity ($P = 250$). Note that the formulas for A_l and B_l contain the corrections found after the correspondence with the author of the original publication.

In the last step, the photoreceptor response is modulated by the normalized neural contrast sensitivity functions, which excludes the effect of the eye's optics and luminance masking:

$$nCSF(\rho, \theta, L_{la}, i^2, d, c) = \frac{CSF(\rho, \theta, L_a, i^2, d, c) \cdot cvi(L_a)}{OTF(\rho)} \tag{12}$$

Since the filter function depends on the local luminance of adaptation, the same kernel cannot be used for the entire image. To speed up computations, the response map R is filtered six times assuming $L_a = \{\ 0.001,\ 0.01,\ 0.1,\ 1,\ 10,\ 100\ \}\ cd/m^2$ and the final value for each pixels is found by the linear interpolation between two filtered maps closest to the L_{la} for a given pixel. The resulting filtered map has the property that the unit amplitude estimates the detection threshold at $P_{det} = 75\%$.

Another element of the VDP that we use in our metric is the modified cortex transform, which is the collection of the band-pass and orientation selective filters. The band-pass filters are computed as:

$$dom_k = \begin{cases} mesa_{k-1} - mesa_k & \text{for } k = 1..K-2 \\ mesa_{k-1} - base & \text{for } k = K-1 \end{cases} \tag{13}$$

where K is the total number of spatial bands and the low-pass filters $mesa_k$ and $baseband$ have the form:

$$\begin{aligned}
mesa_k &= \begin{cases} 1 & \text{for } \rho \le r - \frac{tw}{2} \\ 0 & \text{for } \rho > r + \frac{tw}{2} \\ \frac{1}{2}\left(1 + \cos\left(\frac{\pi(\rho - r + \frac{tw}{2})}{tw}\right)\right) & \text{otherwise} \end{cases} \\
base &= \begin{cases} e^{-\frac{\rho^2}{2\sigma^2}} & \text{for } \rho < r_{K-1} + \frac{tw}{2} \\ 0 & \text{otherwise} \end{cases}
\end{aligned} \tag{14}$$

where

$$r = 2^{-k}, \quad \sigma = \frac{1}{3}\left(r_{K-1} + \frac{tw}{2}\right) \text{ and } \quad tw = \frac{2}{3}r \tag{15}$$

The orientation-selective filters are defined as:

$$fan_l = \begin{cases} \frac{1}{2}\left(1 + cos\left(\frac{\pi|\theta - \theta_c(l)|}{\theta_{tw}}\right)\right) & \text{for } |\theta - \theta_c(l)| \leq \theta_{tw} \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

where $\theta_c(l)$ is the orientation of the center, $\theta_c(l) = (l-1)\cdot\theta_{tw} - 90$, and θ_{tw} is the transitional width, $\theta_{tw} = 180/L$. The cortex filter is formed by the product of the dom and fan filters:

$$B^{k,l} = \begin{cases} dom_k \cdot fan_l & \text{for } k = 1..K-1 \text{ and } l = 1..L \\ base & \text{for } k = K \end{cases} \tag{17}$$

To compute the detection probability, we use a psychometric function in the form:

$$P(C) = 1.0 - \exp(-|\alpha\ C|)^s) \tag{18}$$

where s is the slope of the function ($s = 3$), and $\alpha = (-\log(1 - 0.75))^{1/s}$ ensures that $P(1) = 0.75$.

References

AKYÜZ, A. O., REINHARD, E., FLEMING, R., RIECKE, B. E., AND BÜLTHOFF, H. H. 2007. Do HDR displays support LDR content? a psychophysical evaluation. *ACM Transactions on Graphics (Proc. SIGGRAPH) 26*, 3. Article 38.

CRAMÉR, H. 1999. *Mathematical Methods of Statistics*. Princeton University Press.

D'AGOSTINO, R. 1972. Relation between the chi-squared and ANOVA test for testing equality of k independent dichotomous populations. *The American Statistician 26*, 30–32.

DALY, S. 1993. The Visible Differences Predictor: An algorithm for the assessment of image fidelity. In *Digital Images and Human Vision*, MIT Press, A. B. Watson, Ed., 179–206.

DRAGO, F., MYSZKOWSKI, K., ANNEN, T., AND CHIBA, N. 2003. Adaptive logarithmic mapping for displaying high contrast scenes. *Computer Graphics Forum (Proc. of EUROGRAPHICS) 24*, 3, 419–426.

DURAND, F., AND DORSEY, J. 2002. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics (Proc. SIGGRAPH) 21*, 3, 257–266.

FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. *ACM Transactions on Graphics (Proc. SIGGRAPH) 21*, 3, 249–256.

JANSSEN, R. 2001. *Computational Image Quality*. SPIE Press.

KUANG, J., JOHNSON, G. M., AND FAIRCHILD, M. D. 2007. iCAM06: A refined image appearance model for hdr image rendering. *Journal of Visual Communication and Image Representation 18*, 5, 406–414.

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

MANTIUK, R., DALY, S., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2005. Predicting visible differences in high dynamic range images - model and its calibration. In *Human Vision and Electronic Imaging X*, vol. 5666 of *SPIE Proceedings Series*, 204–214.

MEYLAN, L., DALY, S., AND SUSSTRUNK, S. 2007. Tone mapping for high dynamic range displays. In *Human Vision and Electronic Imaging XII*, SPIE, volume 6492.

MOON, P., AND SPENCER, D. 1944. On the stiles-crawford effect. *J. Opt. Soc. Am. 34*, 319-329.

PATTANAIK, S. N., TUMBLIN, J. E., YEE, H., AND GREENBERG, D. P. 2000. Time-dependent visual adaptation for fast realistic image display. In *Proc. of ACM SIGGRAPH 2000*, 47–54.

RAMANARAYANAN, G., FERWERDA, J., WALTER, B., AND BALA, K. 2007. Visual Equivalence: Towards a new standard for Image Fidelity. *ACM Transactions on Graphics (Proc. SIGGRAPH) 26*, 3. Article 76.

REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic tone reproduction for digital images. *ACM Transactions on Graphics (Proc. SIGGRAPH) 21*, 3, 267–276.

REINHARD, E., WARD, G., PATTANAIK, S., AND DEBEVEC, P. 2005. *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. Morgan Kauffman.

REMPEL, A. G., TRENTACOSTE, M., SEETZEN, H., YOUNG, H. D., HEIDRICH, W., WHITEHEAD, L., AND WARD, G. 2007. Ldr2Hdr: On-the-fly reverse tone mapping of legacy video and photographs. *ACM Transactions on Graphics (Proc. SIGGRAPH) 26*, 3. Article 39.

SEETZEN, H., HEIDRICH, W., STUERZLINGER, W., WARD, G., WHITEHEAD, L., TRENTACOSTE, M., GHOSH, A., AND VOROZCOVS, A. 2004. High dynamic range display systems. In *Proc. of ACM SIGGRAPH 2004*.

SMITH, K., KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2006. Beyond tone mapping: Enhanced depiction of tone mapped HDR images. *Computer Graphics Forum (Proc. of EUROGRAPHICS) 25*, 3, 427–438.

WANG, Z., AND BOVIK, A. C. 2002. A universal image quality index. *IEEE Signal Processing Letters 9*, 3 (March), 81–84.

WANG, Z., AND BOVIK, A. C. 2006. *Modern Image Quality Assessment*. Morgan & Claypool Publishers.

WANG, Z., AND SIMONCELLI, E. P. 2005. Translation insensitive image similarity in complex wavelet domain. In *IEEE International Conference on Acoustics, Speech, & Signal Processing*, vol. II, 573–576.

WANG, Z., SIMONCELLI, E., AND BOVIK, A., 2003. Multi-scale structural similarity for image quality assessment.

WATSON, A. 1987. The Cortex transform: rapid computation of simulated neural images. *Comp. Vision Graphics and Image Processing 39*, 311–327.

WATSON, A. 2000. Visual detection of spatial contrast patterns: Evaluation of five simple models. *Optics Express 6*, 1, 12–33.

WINKLER, S. 2005. *Digital Video Quality: Vision Models and Metrics*. John Wiley & Sons, Ltd.

WU, H., AND RAO, K. 2005. *Digital Video Image Quality and Perceptual Coding*. CRC Press.