Perceptually optimized image rendering

VALERO LAPARRA,[1,2,*] ALEXANDER BERARDINO,[2] JOHANNES BALLÉ,[2] AND EERO P. SIMONCELLI[2,3]

[1]*Image Processing Laboratory, Universitat de València, 46980 Paterna, Spain*
[2]*Center for Neural Science, New York University, New York, New York 10003, USA*
[3]*Courant Institute of Mathematical Sciences, New York University, New York, New York 10003, USA*
*Corresponding author: valero.laparra@uv.es

We develop a framework for rendering photographic images by directly optimizing their perceptual similarity to the original visual scene. Specifically, over the set of all images that can be rendered on a given display, we minimize the normalized Laplacian pyramid distance (NLPD), a measure of perceptual dissimilarity that is derived from a simple model of the early stages of the human visual system. When rendering images acquired with a higher dynamic range than that of the display, we find that the optimization boosts the contrast of low-contrast features without introducing significant artifacts, yielding results of comparable visual quality to current state-of-the-art methods, but without manual intervention or parameter adjustment. We also demonstrate the effectiveness of the framework for a variety of other display constraints, including limitations on minimum luminance (black point), mean luminance (as a proxy for energy consumption), and quantized luminance levels (halftoning). We show that the method may generally be used to enhance details and contrast, and, in particular, can be used on images degraded by optical scattering (e.g., fog). Finally, we demonstrate the necessity of each of the NLPD components—an initial power function, a multiscale transform, and local contrast gain control—in achieving these results and we show that NLPD is competitive with the current state-of-the-art image quality metrics.   © 2017 Optical Society of America

1. INTRODUCTION

A general goal in designing a pipeline for the capture and display of photographic images is to remain as faithful to the original source as possible, minimizing distortions introduced by the sensor, coding, transmission, or display processes. If images are meant for presentation to human observers, distortion should be measured accordingly, penalizing errors that are most visually noticeable and/or disturbing, while permitting those that are perceptually unnoticeable. This strategy is most evident in the handling of color in which both sensors and displays are designed so as to accurately capture and render the three-dimensional subspace of wavelengths relevant for human trichromatic visual representation, while allowing significant distortion outside of this subspace.

Arguably the most significant limitations of current sensors and displays are with regard to dynamic range. Early digital sensors were restricted to capturing a limited luminance range and were unable to adequately capture the majority of realistic natural scenes, which contain luminances spanning up to roughly 20 orders of magnitude. In contrast, the human visual system is capable of sensing fixed scenes with a range of over

5 orders of magnitude in real time, up to 8 orders of magnitude in the photopic regime when the effects of extended temporal adaptation mechanisms are incorporated [1], and up to 14 orders of magnitude when including the scotopic and mesopic regimes (see Fig. 1). The dynamic range of sensors has steadily improved, and current sensors (often augmented with software solutions that fuse images captured at different exposures) are capable of acquiring images with dynamic range approximating the sensitivity of human vision. Despite this, even the best display devices are limited to a significantly lower dynamic range than these sensors can capture.

The simplest solution to the problem of displaying high dynamic range (HDR) images on a low dynamic range (LDR) rendering device is to linearly rescale the luminance values recorded by the sensor into the display's reproducible range of luminances. This, however, produces images that are quite different from the original scene—typically all of the low-luminance information is lost. A variety of "tone-mapping" methods have been proposed to solve this problem by nonlinearly remapping the intensities of the original image into the output range in a way that least interferes with the visual appearance of the original scene. Most of these are based on heuristics, and require manual parameter

adjustments for best results. In addition, many displays introduce constraints other than the global luminance range, such as restriction to discrete luminance levels (i.e., halftoning), maximal average power consumption, and interactions between pixel values over space or time. Separate methods have been developed for solving each of these problems.

Perceptual optimization of tone mapping was introduced in a seminal paper by Tumblin and Rushmeier, who proposed the selection of a tone mapping transformation from HDR images to LDR displays to best match the appearance of the original scene [2]. A variety of tone mapping papers have followed this framework (see, for instance, [3–7]). These methods are dependent on the parametric function used as a tone mapping operator, which restricts the space of possible solutions: a given functional form may not be able to achieve a perceptually optimal solution, or may only work satisfactorily for a particular type of rendering constraint.

Here, we formulate a more general solution for perceptually accurate rendering, directly optimizing the rendered image to minimize perceptual differences with the light intensities of the original scene, subject to all constraints imposed by the display (Fig. 1). This constrained optimization formulation relies on four ingredients: knowledge of the original scene luminances (or calibration information that allows calculation of those luminances), a measure of the perceptual similarity between images, knowledge of the display constraints, and a method for optimizing the image to be rendered. We use a model of perceptual similarity loosely based on the transformations of the early stages of the human visual system [specifically, the retina and lateral geniculate nucleus (LGN)], that has previously been fit to a database of human psychophysical judgments. Because this model is continuous and differentiable, our method can be efficiently solved by first-order constrained optimization techniques. We show that the solution is well defined and general, and therefore represents a framework for solving a wide class of rendering problems.
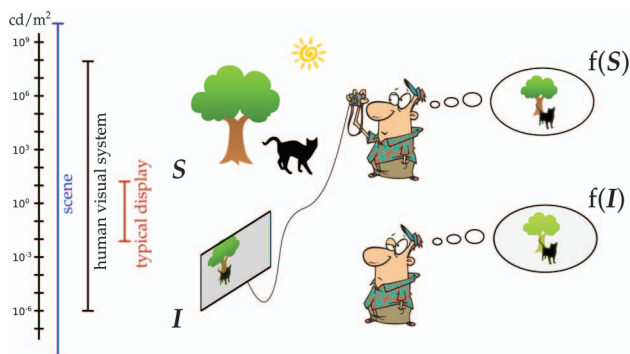

Fig. 1. Perceptually optimized rendering framework. When we view a real-world scene, the luminances, specified by a vector S, give rise to an internal perceptual representation $f(S)$. While luminances in the real world can range from complete darkness (0 cd/m²) to extremely bright (e.g., midday sun, roughly 10^9 cd/m²), a typical display can generate a relatively narrow range of roughly 5 to 300 cd/m². The optimization goal is to adjust the luminances I generated by the display to minimize the difference between the perceptual representations $f(S)$ and $f(I)$ while remaining within the set of images that can be generated by the display.

In Section 3, we optimize images captured under differing acquisition conditions for rendering on the same display. We show one result per experiment; more images can be found at http://www.cns.nyu.edu/~lcv/perceptualRendering/. We start with calibrated images, where the original scene luminances are known. We also deal with the more common scenario in which the exact luminances of the original scene are unknown (the tone mapping problem). In this scenario, we have to make some educated guesses about the luminance range of the original scene, and we demonstrate the effect that different assumptions have on the optimized images. Moreover, we take advantage of these effects to solve other image processing problems, such as detail enhancement and haze removal, by manipulating these source assumptions. For each of these tasks, we compare the results with state-of-the-art algorithms designed to solve each specific case. In Section 4, we optimize images to be displayed under differing display restrictions, including luminance limited displays, power limited displays, and displays restricted to a small set of output values. Finally, we analyze the effect that each component of our perceptual measurement has on the quality of our optimized images.

2. OPTIMAL RENDERING FRAMEWORK

Optimally rendering an image I on a given display means displaying it in such a way that it remains faithful to the human perception of the original scene S. Here, S and I are vectors representing the luminances of all pixels in the respective images. We formalize this as a constrained optimization problem,

$$\hat{I}_{\mathcal{C}}(S) = \arg\min_{I} D(S, I), \qquad \text{s.t. } I \in \mathcal{C}, \tag{1}$$

where $D(\cdot, \cdot)$ is a measure of human perceptual dissimilarity, and $\mathcal{C}$ is the set of all images that can be rendered on the display. This formulation can express many well-known rendering problems, such as tone mapping or dithering, which differ only in the specification of $\mathcal{C}$. In general, the optimization problem expressed in Eq. (1) cannot be solved analytically, and thus we will not obtain an explicit function to compute $\hat{I}_{\mathcal{C}}(S)$, given S and $\mathcal{C}$. Instead, we choose a perceptual measure that is differentiable with respect to I, and use modern high-dimensional optimization tools to numerically solve for $\hat{I}_{\mathcal{C}}(S)$. Specifically, we descend the objective function, alternating between minimizing the perceptual distance and projecting the image back onto the constraint set. Specific formulations for different example problems can be found online at http://www.cns.nyu.edu/~lcv/perceptualRendering/.

We follow a principled, two-step approach to quantify perceptual distance. Rather than defining a perceptual distance directly (as in SSIM [8], for example) we first define a nonlinear perceptual transform $f(\cdot)$ that approximates the computations performed within the early stages of the human visual system. We apply this to both the original scene luminances S and the rendered image I and then measure the distance between $f(S)$ and $f(I)$. We refer to this casually as a "metric" (as is common in the image quality assessment literature), even though it is not guaranteed to satisfy all requirements of the mathematical definition of a metric. Specifically, it is symmetric and yields a value of zero for identical images, but for some parameter values
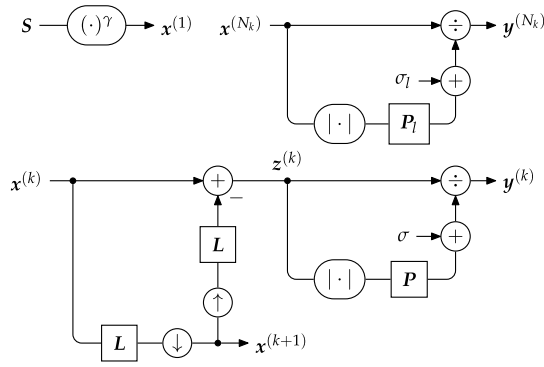
Fig. 2. Perceptual transform $f(S)$ constructed as a NLP [9]. The scene luminances S (in cd/m^2) are first transformed using a power function (top left). The transformed luminance image is then decomposed into frequency channels using the recursive implementation of the Laplacian pyramid [10]. Each channel $z^{(k)}$ is then divided by a weighted sum of local amplitudes (computed with lowpass filter P) plus a constant σ. The final lowpass channel $x^{(N_k)}$ is also normalized, but with distinct parameters (top right). Symbols $\uparrow$ and $\downarrow$ indicate upsampling and downsampling by a factor of 2, respectively.

the transformation can discard information (allowing it to produce a zero distance for non-identical images), and it also may not satisfy the triangle inequality.

Figure 2 illustrates the components of the perceptual transform, which we call the normalized Laplacian pyramid (NLP), a multiscale nonlinear representation that mimics the operations of the early stages of the human visual system. This representation is inspired by a model for responses of the LGN [11], and includes both luminance and contrast gain control mechanisms. The former is primarily attributed to the photoreceptors and initial transformations within the retina whereas the latter is likely initiated in the retina and is enhanced/amplified in the LGN [12]. This transform bears some resemblance to previously published image metrics that utilize local normalization but differs in motivation, structure, and implementation [8,9,13–15]. Here, we adapt this model to operate directly on luminances (in cd/m^2), rather than values that have been gamma-adjusted for a particular display, which provides a standardized set of units for defining constraints on acquisition and display.

Luminances are first transformed elementwise using a power law that approximates the transformation of light to response of retinal photoreceptors:

$$x^{(1)} = S^\gamma. \qquad (2)$$

This initial nonlinear transformation is followed by a recursive partition into frequency channels, as in the Laplacian pyramid [10], mimicking the center-surround receptive fields found in the retina (and LGN),

$$x^{(k+1)} = DLx^{(k)}, \qquad k \in \{1, \ldots, N_k - 1\}, \qquad (3)$$

$$z^{(k)} = x^{(k)} - LUx^{(k+1)}, \qquad (4)$$

$$z^{(N_k)} = x^{(N_k)}, \qquad (5)$$

where D and U indicate down/upsampling by a factor of two, respectively (Fig. 2). For the filtering operation L, we apply a

spatially separable 5-tap filter (0.05, 0.25, 0.4, 0.25, 0.05), as originally specified in [10].

Within each channel, each coefficient is divided by a weighted local sum of the elementwise amplitudes (absolute values) plus a constant:

$$y^{(k)} = z^{(k)} \oslash (\sigma + P|z^{(k)}|), \qquad (6)$$

where P indicates convolution with a filter, and $\oslash$ indicates pointwise division. All bandpass channels and the highpass channel share the same parameters P and σ, whereas the lowpass channel ($k = N_k$) has its own parameter set P_l and σ_l. This function is a simplified variant of divisive normalization used to describe the responses of neurons in different parts of the visual system [16–18]. The NLP coefficients of all channels $y^{(k)}$ combined represent the response of the perceptual transform:

$$f(S) = \{y^{(k)}; k = 1, \ldots, N_k\}. \qquad (7)$$

Figure 3 illustrates the construction of the metric employed in the perceptual space. We compute the L_α-norm of the differences between NLP coefficients within each frequency channel (that is, we raise the absolute value of each coefficient difference to the power α, sum over the entire channel, and take the αth root). These values are then combined across channels using an L_β-norm, to yield the final NLP distance (NLPD):

$$D(S, I) = \left[\frac{1}{N_k} \sum_{k=1}^{N_k} \left(\frac{1}{N_c^{(k)}} \sum_{i=1}^{N_c^{(k)}} \left| y_i^{(k)} - \tilde{y}_i^{(k)} \right|^\alpha \right)^{\frac{\beta}{\alpha}} \right]^{\frac{1}{\beta}}, \qquad (8)$$

where $\tilde{y}_i^{(k)}$ indicates the kth subband arising from the displayed image I (i.e., $f(I) = \{\tilde{y}^{(k)}; k = 1, \ldots, N_k\}$) and $N_c^{(k)}$ is the number of coefficients in that subband. A similar summation model has been employed in previous perceptual quality metrics [20,21].

All parameters of the perceptual transform and metric were optimized to best explain human perceptual ratings of distorted images in a public database of grayscale images [19]. Specifically, we chose parameters to maximize the correlation between the mean opinion scores from the human observers and the distance computed by the metric. The optimized
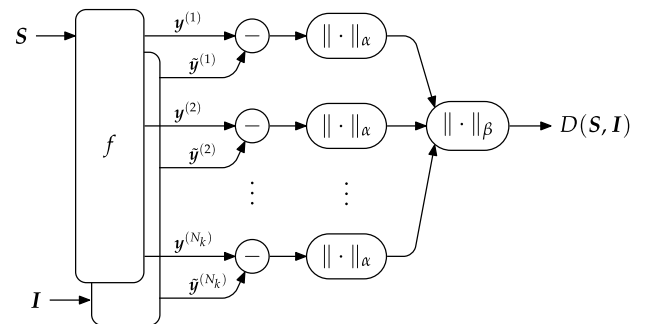

Fig. 3. Construction of the NLPD measure. Two images are transformed by $f(\cdot)$ to a perceptual representation, yielding two NLPs (see Fig. 2). We compute the α-norm over the vector of differences for each frequency channel, and then combine these over channels using a β-norm. For all rendering results, we use $\alpha = 2.0$ and $\beta = 0.6$, which are optimized to fit the human perceptual ratings of the distorted images reported in [19].

exponent for the front-end nonlinearity was $\gamma = \frac{1}{2.6}$. Unlike previous work [9], we set the normalization parameters to be identical for all bandpass channels (assuming scale invariance), but allowed a different set for the lowpass channel. For bandpass channels, the additive constant was $\sigma = 0.17$, and the local weighting functions P were filters with 5×5 support, with values

$$P = \begin{bmatrix} 4 & 4 & 5 & 4 & 4 \\ 4 & 3 & 4 & 3 & 4 \\ 5 & 4 & 5 & 4 & 5 \\ 4 & 3 & 4 & 3 & 4 \\ 4 & 4 & 5 & 4 & 4 \end{bmatrix} \cdot 10^{-2}. \tag{9}$$

The parameters for the lowpass channel were $P_l = 1$ (the identity) and $\sigma_l = 4.86$. Optimized exponents for the metric were $\alpha = 2.0$ and $\beta = 0.6$. Appendix B shows that the performance of this extended and optimized version of the NLP metric surpasses that of state-of-the-art image quality metrics. This metric, with parameters held fixed at their optimized values, was used to optimize all of the rendering results presented below.

3. VARYING IMAGE ACQUISITION CONDITIONS

We performed a set of experiments to test the capabilities of our optimization framework over different image acquisition conditions. We begin with calibrated images, for which we know the exact luminance values (in cd/m^2) of the original scene. We then move on to uncalibrated images, for which we need to make an assumption about the luminance values in the original scene. Finally, we close this section by demonstrating that the method is stable and flexible enough that it can be used to solve other rendering problems, such as haze removal and artificial detail enhancement.

Each example requires us to minimize the perceptual distance with respect to the rendered image I, subject to the display constraints. In general, this is accomplished by alternating between projection onto the constraint set and minimization of the distance using the adaptive moment estimation (Adam) algorithm [22]. The gradient of the perceptual distance with respect to I is described in Appendix A. Implementation of the derivatives, along with additional optimized examples, are provided on the project webpage http://www.cns.nyu.edu/~lcv/perceptualRendering/. All images presented here are intended for viewing on a display with luminance ranging from 5 to 300 cd/m^2, and a gamma value of 2.2. Computation time scales linearly with the size of the image. When optimized on a Tesla K40 GPU card, it takes approximately 1 s per 10000 pixels (i.e., an image of 1000×1000 requires less than 2 min).

A. Rendering Calibrated HDR Luminances

We begin by considering the rendering of images S obtained from a calibrated HDR imaging device, such that we know the true luminance values of all pixels. As an example, Fig. 4 shows an image from the database of Mark Fairchild [24], with luminance range $S_{min} = 0.78$ to $S_{max} = 16,200$ cd/m^2. We wish to display this image on a device with a limited luminance range of $I_{min} = 5$ to $I_{max} = 300$ cd/m^2 (typical values for a computer monitor). We solve for the perceptually optimal rendered image:

$$\hat{I}(S) = \arg\min_I D(S, I), \quad \text{s.t.} \quad \forall i: I_{min} \leq I_i \leq I_{max}. \tag{10}$$

Figure 4 shows the original image intensities, linearly rescaled to fit within the luminance range $[I_{min}, I_{max}]$, an image tone-mapped using a recent state-of-the-art method by Paris *et al.* [23], and our perceptually optimized image $\hat{I}(S)$. The second image was computed using the default
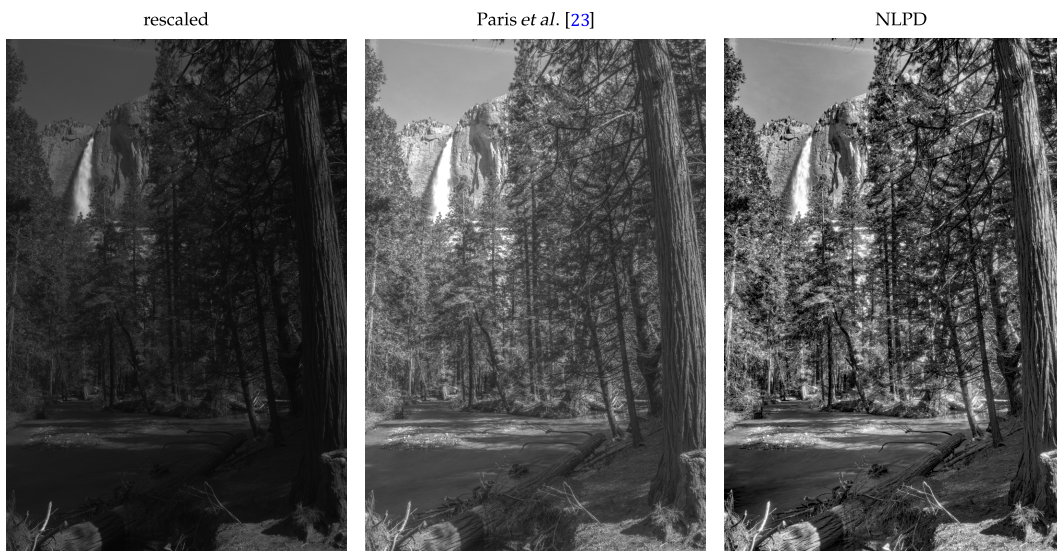

| rescaled | Paris *et al.* [23] | NLPD |

Fig. 4.   Rendering of a calibrated HDR image on a display with a limited luminance range. The scene luminances for this image spanned the range from $S_{min} = 0.78$ cd/m^2 to $S_{max} = 16,200$ cd/m^2, whereas the display luminances are assumed to lie between 5 cd/m^2 and 300 cd/m^2. Left: the image rendered by linear rescaling of luminance values into the display range. Center: the image rendered using a state-of-the-art tone mapping algorithm [23]. Right: the image rendered using the proposed method of minimizing the NLPD metric subject to the display constraints.

parameters recommended by the authors for tone mapping of HDR images: $\alpha = 1$, $\beta = 0$, and $\sigma = \log 2.5$. Linearly rescaling yields a rendered image in which most of the details cannot be seen or differentiated. The algorithm by Paris *et al.* [23] does an excellent job in mitigating this problem, rendering an image that reveals detail in both the dark and bright regions. Nevertheless, the solution appears less detailed and lower in contrast than the image computed using our method. This is mostly because the Paris algorithm does not take into account the display luminance range. Although it (and most other tone-mapping algorithms) has additional parameters that can be adjusted, it is not obvious to a naive user how to select these parameters based on the display properties. In contrast, our solution is fully automatic (assuming the luminance values of the source image and the range of the display are known), albeit at the expense of significantly more computation.

B. Rendering LDR Images with an Image Acquisition Model

Our method can also be used to improve the appearance of images acquired with a conventional LDR digital camera that has been calibrated to allow the recovery of luminance values from recorded pixel values $\boldsymbol{R}$. For most modern digital cameras, the acquisition luminance range is still generally much larger

than the display range, and in any case, is unlikely to be exactly matched. Thus, we need to solve the following optimization problem analogous to the previous section:

$$\hat{I}(\boldsymbol{R}) = \arg\min_{I} D(g(\boldsymbol{R}), I), \quad \text{s.t.} \ \ \forall i : I_{\min} \leq I_i \leq I_{\max}, \quad \textbf{(11)}$$

where g is the mapping from the recorded pixel values to the estimated scene luminances (in cd/m^2).

The results for two example grayscale images from the McGill database [25] are shown in Fig. 5. For each image, we again compare the original image intensities, linearly rescaled to fit within the luminance range $[I_{\min}, I_{\max}]$, to our perceptually optimized image $\hat{I}(\boldsymbol{R})$, and a tone-mapped image computed using the Paris *et al.* method [23]. For the latter, we have again used the parameters recommended by the authors for tone mapping of HDR images: $\alpha = 1$, $\beta = 0$, and $\sigma = \log 2.5$. Our method again offers a visual advantage, producing higher contrast and more visible details. The improvement here is perhaps even more noticeable than in the HDR case, for which the Paris *et al.* algorithm was developed.

C. Rendering Uncalibrated HDR Images

Unlike the situation in Section 3.A, the typical scenario for images acquired from HDR cameras is that they are uncalibrated. That means that we have access to measurements $\boldsymbol{L}$ that are
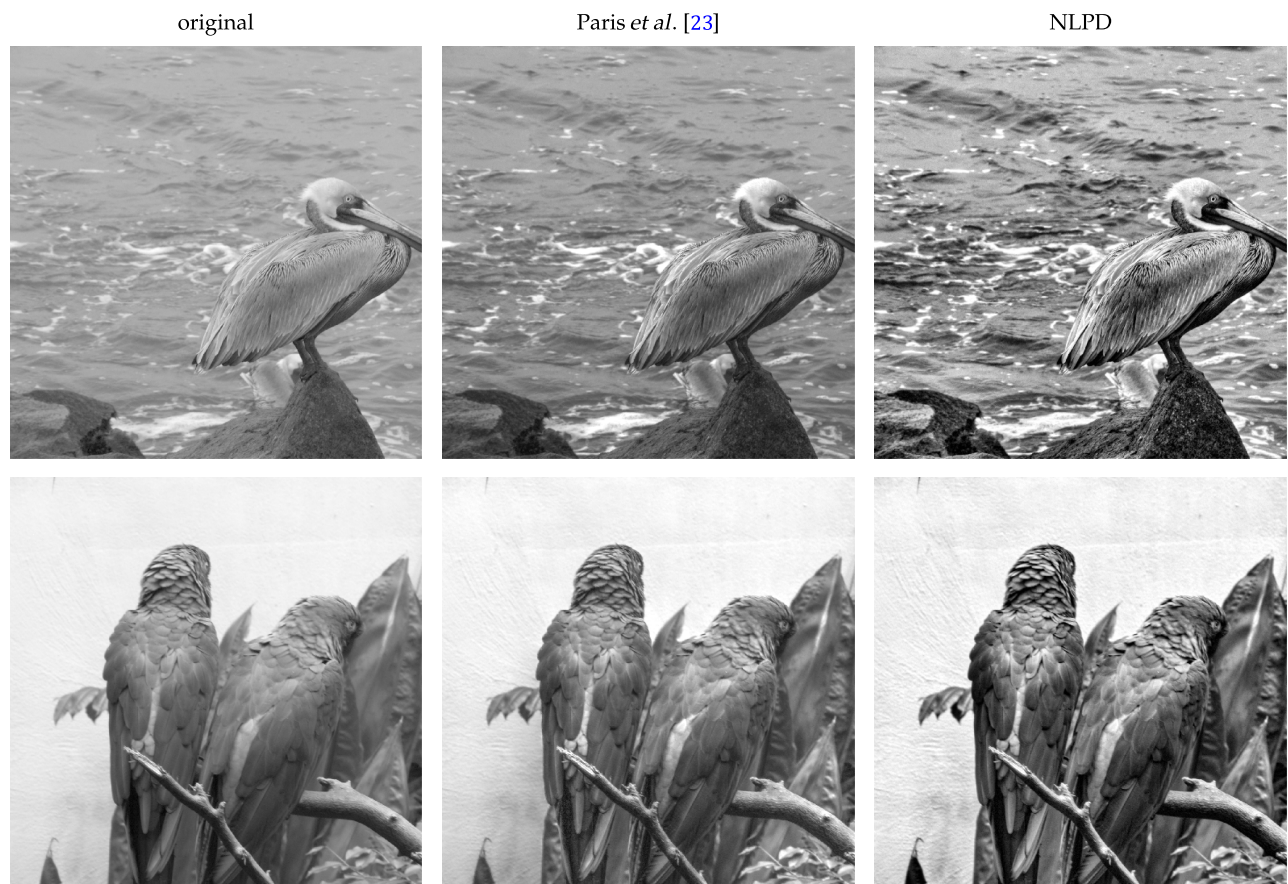

original	Paris *et al.* [23]	NLPD

Fig. 5. Rendering of two calibrated LDR images to a display with a limited luminance range of $[5, 300]$ cd/m^2 (see caption of Fig. 4).

linearly related to actual luminances, but we do not have access to the scaling parameters (for instance, they might be normalized values, lying between 0 and 1). To apply our method, the measurements need to be linearly rescaled to luminance values, which amounts to estimating the minimum and the maximum luminance in the original scene (S_{min} and S_{max}, respectively). One can often use an educated guess for those values given the content of the image; for instance, the luminance of a filament of a clear incandescent lamp is roughly 10^6 cd/m^2. As in the previous experiments, we solve the resulting optimization problem as follows:

$$\hat{I}(S) = \arg\min_{I} D(S, I), \qquad \text{s.t. } \forall\, i : I_{min} \leq I_i \leq I_{max},$$
$$\text{where } S = (S_{max} - S_{min}) \cdot L + S_{min}. \qquad \textbf{(12)}$$

Figure 6 shows the results for the widely used HDR image "Memorial" for different values of S_{max} (and a fixed value of $S_{min} = 5$). The proposed method converges to an image that exhibits enhanced contrast in all the regions, preserving the details, but also preserving the relative contrast and luminance between regions. This is particularly evident in high-luminance regions (for instance, the bright window behind the altar, or the round window in the top of the dome), where both the perceived details and luminance intensity are effectively portrayed.

As we increase the assumed maximum luminance of the original scene (while fixing the display restrictions), our algorithm further amplifies the contrast of details in the image. This makes sense from a perceptual perspective. If the original scene was brighter, an observer would be able to perceive more details within the scene. Therefore, the method has to
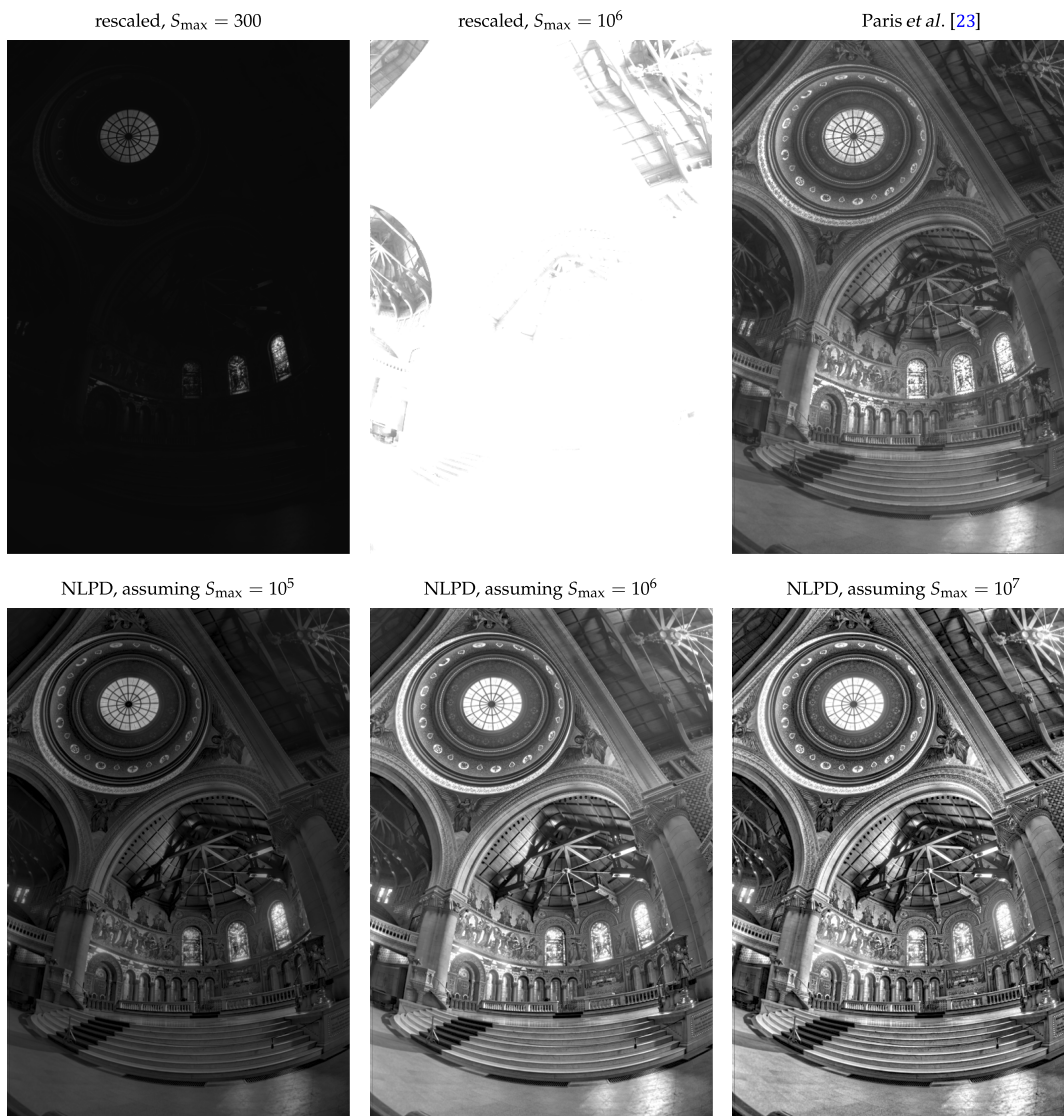

Fig. 6. Rendering of an uncalibrated HDR image on a display with a limited luminance range. Linear mapping of luminances leads to loss of detail (top left: rescaling of luminances to the display range, assuming $S_{max} = 300$ cd/m^2; top center: rescaling of luminances, assuming a more realistic value of $S_{max} = 10^6$ cd/m^2). Top right: the image rendered using [23]. Bottom: the image optimized for NLPD, with different assumed maximum luminance values.

artificially enhance these details to mimic the appearance of the original scene. In the next two sections we take advantage of this behavior.

D. Artificial Detail Enhancement

We showed in the preceding sections that using knowledge about the image acquisition process helps greatly in automatically rendering images, given the display constraints. In some cases, however, detail visibility in the scene might be unsatisfactory. Intuitively, photographers know that the amount of detail visible in a scene depends on the amount of available light. If the image has already been acquired, it is of course not possible to alter the light sources. However, since the scene luminances scale linearly with the intensity of the light sources, our method allows us to simulate increased intensity post hoc, by linearly re-scaling the luminances of the scene S.

Figure 7 shows the results of modifying our choice of S_{max} (as in the previous experiment, we fixed $S_{min} = 5$). Note that with increasing values of S_{max}, details become more visible. We also show the results of applying the Paris *et al.* algorithm for which we have employed the following parameters proposed in their paper for the detail enhancement problem: $\alpha = 0.25$, $\beta = 1$, and $\sigma = 0.3$.

E. Haze Removal

Surprisingly, this same method of detail enhancement can also be used for the problem of haze removal. In a hazy scene, the local contrast has effectively been reduced (roughly speaking, by adding a constant level of scattered light) which makes detail more difficult to discern. In this experiment, we also choose $S_{min} = 5$ (we find that results are fairly robust with the selection of this parameter) and $S_{max} = 10^4$.

Figure 8 compares the performance of our method with two other methods [26,27]. Our algorithm converges on an image that greatly enhances the details of the original hazy image, boosting the contrast and reducing the perception of haze within the image. Although the other two methods are specifically designed for this particular problem, our method obtains a similar result without modification.

4. VARYING DISPLAY CONSTRAINTS

While examining the effects of various image acquisition scenarios in the previous section, we assumed only that the display luminance is bounded. The upper bound is a natural constraint for any real display. The lower bound is also relevant for a wide range of practical display devices, and arises from reflected ambient light and scatter within the
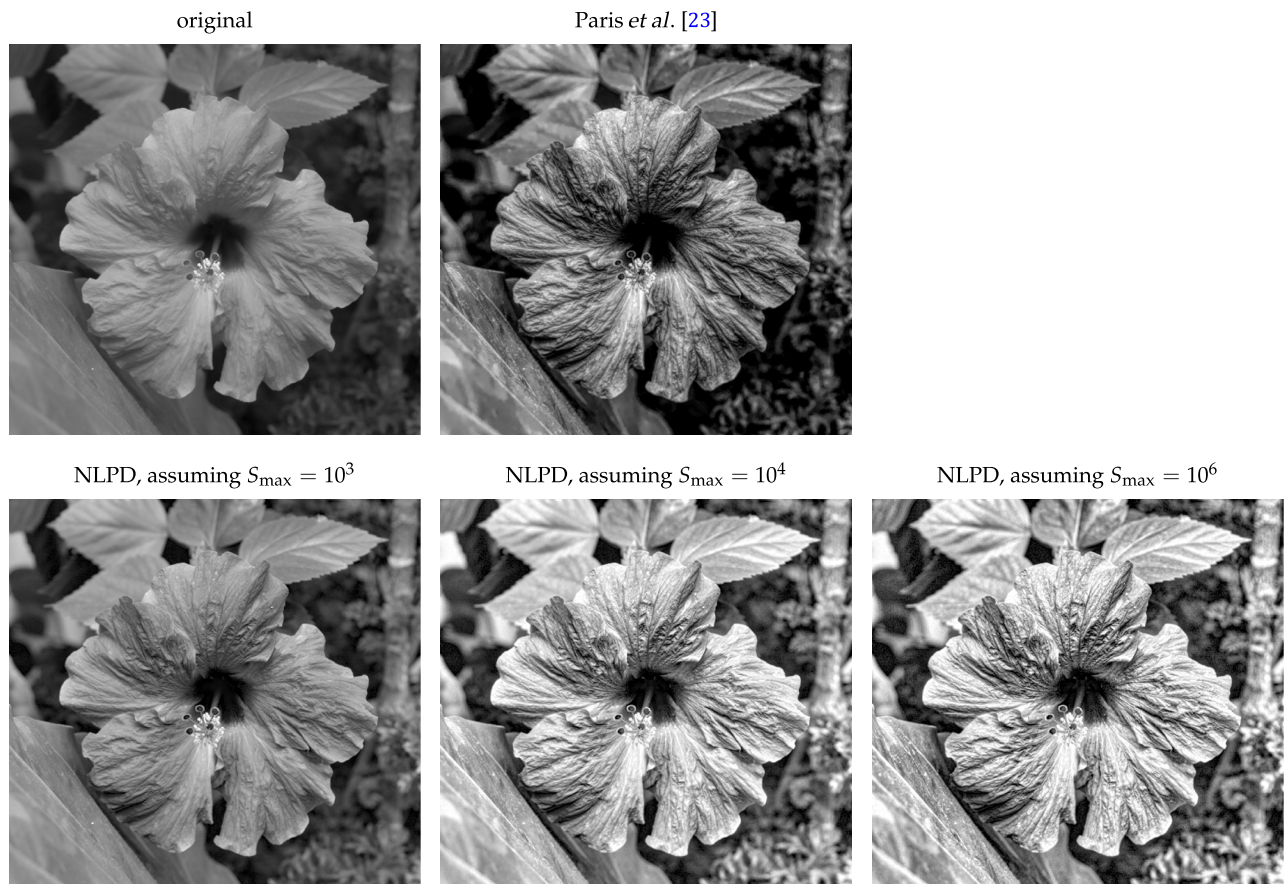

Fig. 7. Example of artificial detail enhancement by simulating more light in the original scene. Top left: the original image. Top center: the image enhanced using [23]. Bottom: the image optimized for NLPD, with different assumed values of maximum luminance.
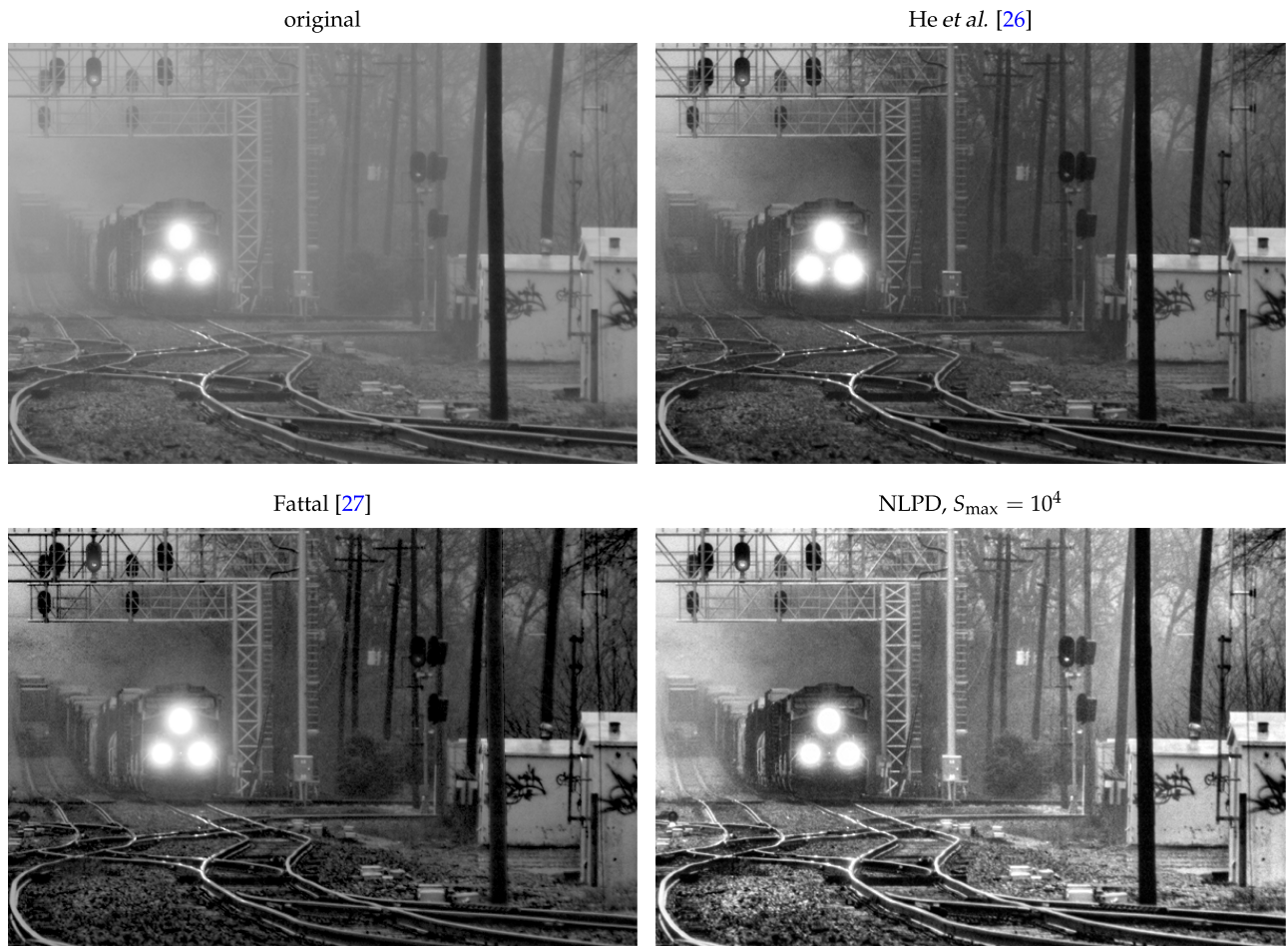
original

He *et al.* [26]

Fattal [27]

NLPD, $S_{\max} = 10^4$

Fig. 8. Example of haze removal. Top left: the original image. Top right: the image processed using He *et al.* algorithm [26]. Bottom left: the image processed using the Fattal algorithm [27]. Bottom right: the image processed by optimizing NLPD, with $S_{\min} = 5$ and $S_{\max} = 10^4$.

display. In this section, we examine the effect of each of these constraints independently, along with a few more complex constraints.

Figure 9 shows the results for different minimum and maximum luminance bounds, $(I_{\min}, I_{\max})$. Our method enhances local contrast, whereas linear rescaling can only manipulate contrast globally. For a wide range of display characteristics, optimizing the image to minimize the NLP distance reduces distortion in the rendered images and increases the visibility of perceptually relevant features.

A. Rendering with Limited Power Consumption

The proposed framework allows us to seamlessly introduce arbitrary display constraints. For example, we can optimize the trade-off between image quality and power consumption. To illustrate this, we assume that power consumption is proportional to mean display luminance (as, for instance, in organic light-emitting diode displays used in cell phones; if the relationship were nonlinear, that could also be incorporated into the problem), and optimize the NLPD while constraining both the mean luminance and the range as follows:

$$\hat{I}(S) = \arg\min_{I} D(S, I), \qquad \text{s.t.} \ \forall \ i : I_{\min} \leq I_i \leq I_{\max}$$

$$\text{and} \quad \frac{1}{N_i} \sum_i I_i = I_{\text{mean}}. \quad (13)$$

Figure 10 shows images optimized for different mean luminance values compared to images linearly rescaled to achieve the same target mean luminance. For each mean luminance value, the NLPD-optimized images retain more detail from the original scene than the rescaled images. In Fig. 11, we plot the mean luminance as a function of the perceptual distortion (NLPD) for both methods. Optimizing the images yields a clear benefit in terms of the trade-off between the mean luminance and perceptual distortion. Over a wide range of distortion levels we see that the NLPD-optimized images reduce the power consumption by roughly 80% compared to linear rescaling.

B. Rendering with a Discrete Set of Gray Levels (Dithering)

Most displays have a limited number of available gray levels. In the extreme case, this can be as few as two
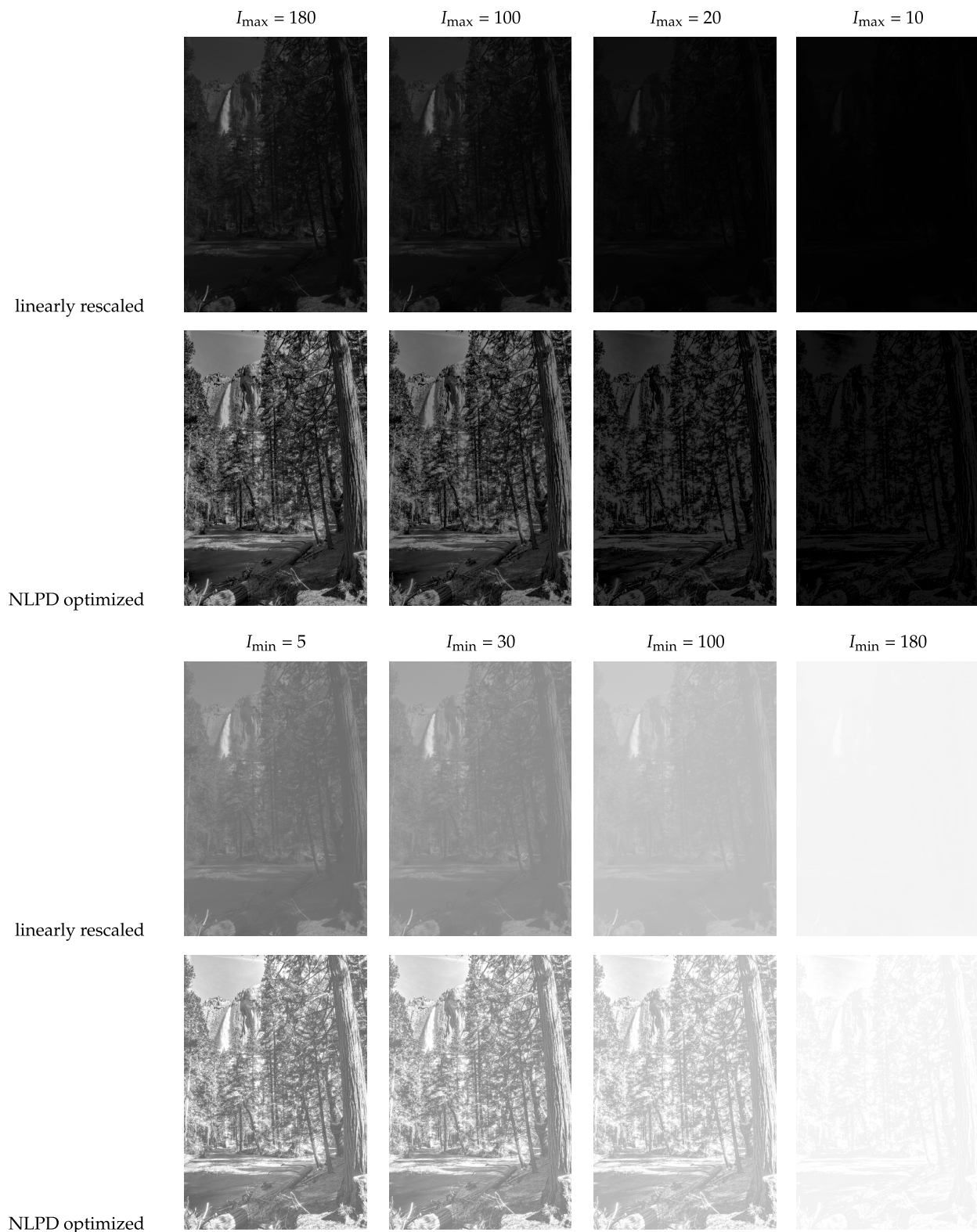
Fig. 9. Effect of different maximum and minimum display luminance constraints. Top two rows: the image rendered for different levels of maximum luminance (assuming $I_{min} = 5$) by linearly rescaling (1st row) versus the proposed NLPD-optimization method (2nd row). Bottom two rows: analogous, but for different levels of minimum luminance (assuming $I_{max} = 300$).

Fig. 10. Rendering with a power consumption constraint. Top left: the image at full luminance (smartphone screenshot). Top row: the image linearly rescaled to achieve the target mean luminance. Bottom row: the image optimized for NLPD with the target mean luminance constraint. Assuming the power consumption is proportional to the mean luminance, the NLPD-optimized renderings convey more detail than their linearly-rescaled counterparts, while consuming the same power.

(e.g., black-and-white printers, e-ink devices, etc.). Here, we illustrate that the proposed method is flexible enough to produce good results even under such extreme constraints. The optimization problem is the same as before, but here we restrict the pixel values to be taken from a discrete set:

$$\hat{I}(S) = \arg\min_{I} D(S, I), \quad \text{s.t.} \quad \forall \, i : I_i \in \{I_{\min}, \ldots, I_{\max}\}. \tag{14}$$

The discrete nature of the optimization problem prevents us from using a gradient-based method. Instead, we use a greedy error-diffusion algorithm, analogous to the classic Floyd–Steinberg method. We first initialize the image to the solution

obtained for a continuous range of luminances, as in previous experiments. Then, we iteratively select the discrete value for each pixel of the image in raster-scan order, each time picking the discrete value that minimizes the NLP distance of the intermediate result to the original scene.

Figure 12 shows the results for images rendered using two and four gray levels. In low-contrast regions, our method is seen to preserve significantly more detail than the Floyd–Steinberg method. In addition, the Floyd–Steinberg algorithm tends to generate artificial patterns in extensive regions of slowly varying luminance, which can be seen in the dark regions of the bird's wings. Our method, however, does not generate these artificial patterns.
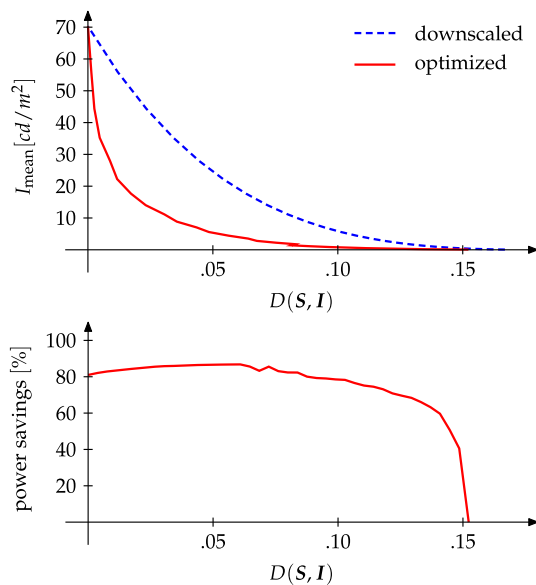
Fig. 11. Trade-off between the power consumption and the image quality, comparing the linear luminance rescaling to the optimization of the perceptual distortion with a mean luminance constraint. Top: the relationship between perceptual distortion $D(S, I)$ and the mean display luminance I_{mean}. For any given acceptable distortion level, the optimization method requires only a fraction of the display luminance, hence significantly decreasing the power consumption. Bottom: power savings, quantified as one minus the ratio of the required mean display luminances for the two methods.

5. CONTRIBUTION OF PERCEPTUAL METRIC COMPONENTS

To provide intuition regarding the effect of each of the primary components of the NLP, we optimized images for rendering while removing one of three components of the transform: the initial pointwise nonlinearity (set $\gamma = 1$), the multiscale decomposition (set $N_k = 1$), and divisive normalization (set $P = 0$ and $\sigma = 1$). Figure 13 shows the results for each manipulation. Note that we did not refit each of the partial transforms to predict human perceptual judgments; therefore, these results should be seen as a way to understand the importance of each computation and not as a quantitative comparison of image quality assessment performance (see details in Appendix B).

Each of the three images differs noticeably from the one optimized with the full transform. Without the initial pointwise nonlinearity, the algorithm produces images in which low to medium luminance patches of an image are misrepresented. The high-luminance areas are detailed but some parts with medium or low luminance are reduced in contrast. Without the multiscale decomposition, the algorithm produces images in which extremely high and extremely low frequencies are well preserved, but intermediate frequencies are underrepresented, and in some cases nearly disappear. Without the contrast normalization, the algorithm converges to images that saturate at the luminance boundary constraints of the display. Normalization preserves the relative luminance changes between coefficients while allowing the absolute luminance

to be modified. This allows the rendered image pixel intensities to be proportional to the relative power in each local region. Moreover, this ensures that regions with similar content scale in a similar way.

6. DISCUSSION

We have described a framework for directly optimizing rendered images, account for display limitations, to minimize perceptual differences between the rendered image and the original scene. The method is parameter free and only requires knowledge of the display restrictions and the original scene intensities. Since these restrictions are expressed in standard physical units (cd/m^2), if either is missing, suitable values can be estimated easily. We have shown that our method matches or exceeds the state-of-the-art for rendering across a variety of acquisition conditions and display restrictions.

We have employed a perceptual metric based on an abstraction of the transformations implemented in the early stages of the human visual system. The metric is an extension of the NLP distance presented in [9], adapted to deal directly with luminances and images of any size. We fit the parameters of this metric to optimize its ability to predict human distortion ratings. We have shown that this metric is consistent with human perception, exhibiting correlation with human quality ratings that is similar to or better than full-reference models specifically designed to assess perceptual quality (see Appendix B). It is continuous and has well-behaved gradients, making it easy to incorporate into a rendering optimization framework. In addition, it has also been previously employed to optimize an image compression algorithm [29].

Most contemporary tone mapping methods do not make explicit use of perceptual metrics (see [30] for a nice review), but rather provide the user with a small set of free parameters to hand-adjust the mapping from the scene to the displayed image. These methods are conceptually simpler than ours, and some of them can produce high quality results in controlled situations (see for instance [23]). Nevertheless, their parameters are often difficult to interpret (and thus, to set), and the restriction to particular functional forms may limit their applicability to specific rendering problems.

In contrast, by directly optimizing the rendered image itself, our method is able to take into account different display constraints without requiring manual selection of an appropriate parametric mapping for each situation and without requiring a human operator to adjust any parameters. The downside of this approach is computational cost: optimization over the high-dimensional space of feasible rendered images is expensive, and although both hardware and software continue to improve, this optimization will always be significantly more expensive than optimizing a small set of parameters for a fixed transformation. Even if the computational costs prevent the use of this method in a real-world application, the results can still serve as a benchmark for what is possible, thus facilitating the development of alternative methods.

Although our framework may be applied to any display problem, the solution can depend heavily on both the perceptual metric employed and the method used to solve the constrained optimization (for example, if the constraints force the
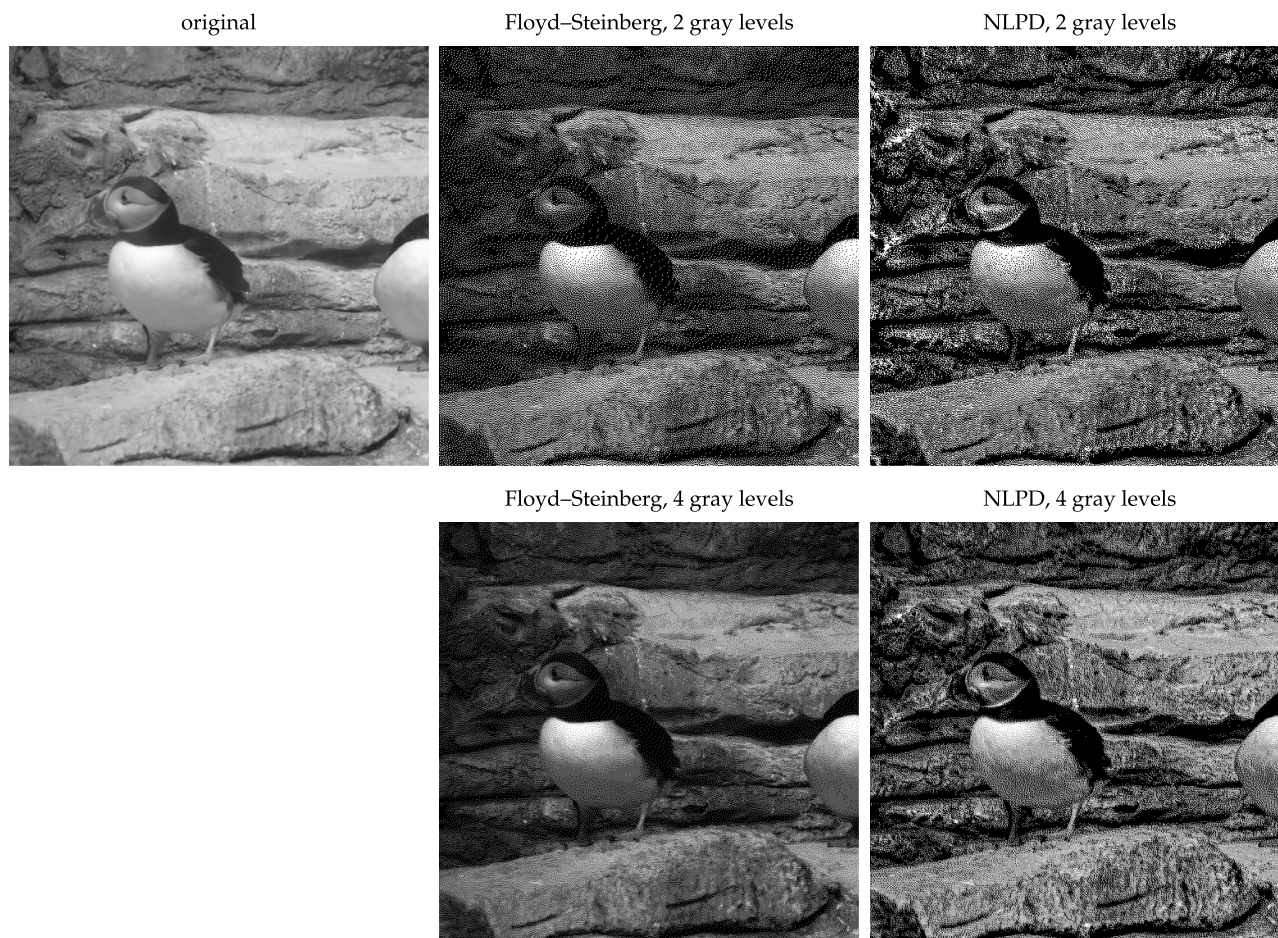
Fig. 12. Rendering with a discrete set of gray levels. Top left: the original image. Center column: the image rendered with 2 or 4 gray levels using a standard error diffusion (Floyd–Steinberg) method [28]. Right column: the image rendered with NLPD error diffusion.

problem into nonconvex or discrete regimes). Optimization has undergone dramatic changes in the past decade, and methods for handling nonconvex and discrete problems have become more reliable and efficient. As an example, we believe it will

be possible to improve on our halftoning solution (for which we used a simple greedy method with error diffusion).

Our use of a simple physiologically-inspired model for assessing perceptual distortion also offers opportunities for
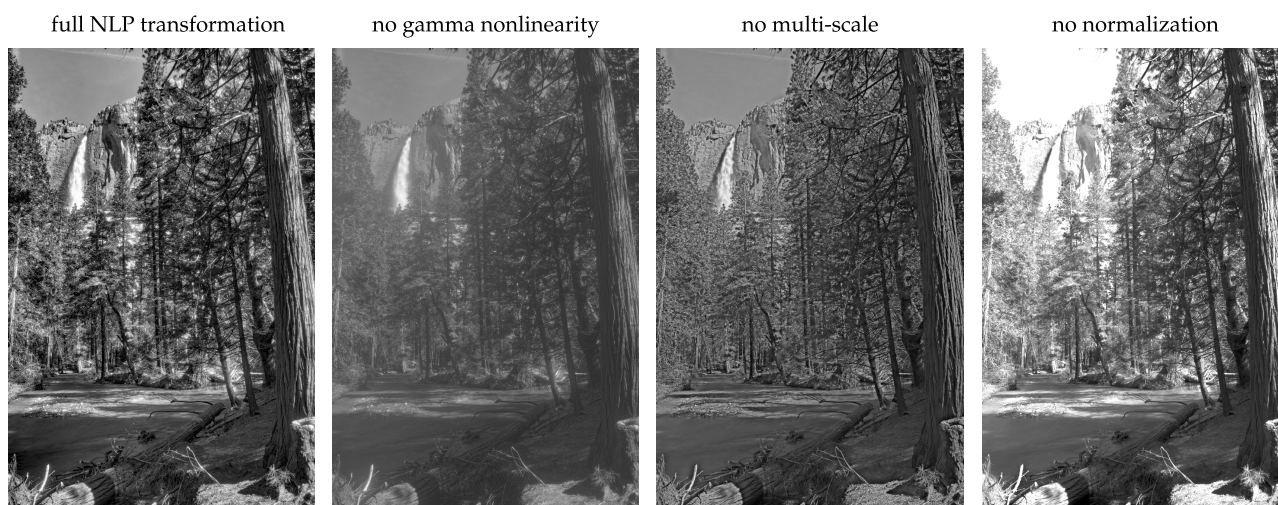

Fig. 13. Rendering of an HDR image with different parts of the NLP transformation removed (see text).

improvement (note that most image quality models are less physiologically motivated [8,31,32]). For example, the NLPD can likely be improved by including the relationships between frequency channels, which could help to control artifacts such as halos that sometimes appear around high-contrast edges. In addition, the NLP model should be extended to operate on color images, and to include another stage of processing corresponding to primary visual cortex (for example, using oriented, multiscale, derivative filters with cross-scale and cross-orientation normalization). All of these improvements can be made following the same framework that we have presented for the current model: defining a functional form based on the transformations of sensory neurobiology, fitting the parameters using human perceptual data, and using this model with fixed parameters to optimize the rendering of images.

APPENDIX A. DERIVATIVE OF THE DISTANCE WITH RESPECT TO THE RENDERED IMAGE

Here we provide, for the interested reader, the gradient of the perceptual distance $D(S, I)$ with respect to the rendered image I. The distance is given by

$$D(S, I) = \left[\frac{1}{N_k} \sum_{k=1}^{N_k} \left(\frac{1}{N_c^{(k)}} \sum_{i=1}^{N_c} (d_i^{(k)})^\alpha \right)^{\frac{\beta}{\alpha}} \right]^{\frac{1}{\beta}},$$

where $d_i^{(k)}$ is the distance between the transformed images, $y = f(S)$ and $\tilde{y} = f(I)$:

$$d_i^{(k)} = |y_i^{(k)} - \tilde{y}_i^{(k)}|.$$

Differentiating with respect to the image pixels gives

$$\frac{\partial D(S, I)}{\partial I_j} = \frac{1}{\beta} D(S, I)^{1-\beta} \frac{\partial}{\partial I_j} \left[\frac{1}{N_k} \sum_{k=1}^{N_k} \left(\frac{1}{N_c^{(k)}} \sum_{i=1}^{N_c} d_i^{(k)\alpha} \right)^{\frac{\beta}{\alpha}} \right].$$

From here we apply the chain rule to expand the full equation. The derivative of the bracketed expression is

$$\frac{\partial}{\partial I_j} \left[\frac{1}{N_k} \sum_{k=1}^{N_k} \left(\frac{1}{N_c^{(k)}} \sum_{i=1}^{N_c} d_i^{(k)\alpha} \right)^{\frac{\beta}{\alpha}} \right]$$
$$= \frac{\beta}{\alpha N_k} \sum_{k=1}^{N_k} \left(\frac{1}{N_c^{(k)}} \sum_{i=1}^{N_c} d_i^{(k)\alpha} \right)^{\frac{\beta}{\alpha}-1} \frac{\partial}{\partial I_j} \left[\frac{1}{N_c^{(k)}} \sum_{i=1}^{N_c} d_i^{(k)\alpha} \right],$$

and the derivative of this bracketed expression can be written in terms of the derivative of the distance:

$$\frac{\partial}{\partial I_j} \left[\frac{1}{N_c^{(k)}} \sum_{i=1}^{N_c} d_i^{(k)\alpha} \right] = \frac{\alpha}{N_c^{(k)}} \sum_{i=1}^{N_c} d_i^{(k)(\alpha-1)} \frac{\partial d_i^{(k)}}{\partial I_j}.$$

Combining these equations, we obtain the desired derivative:

$$D(S, I)^{1-\beta} \frac{1}{N_k} \sum_{k=1}^{N_k} \frac{1}{N_c^{(k)\frac{\beta}{\alpha}}} \left(\sum_{i=1}^{N_c} d_i^{(k)\alpha} \right)^{\frac{\beta}{\alpha}-1} \sum_{i=1}^{N_c} d_i^{(k)(\alpha-1)} \frac{\partial d_i^{(k)}}{\partial I_j}.$$

Now, expanding the derivative of the difference

$$\frac{\partial d_i^{(k)}}{\partial I_j} = \operatorname{sgn}(y_i^{(k)} - \tilde{y}_i^{(k)}) \frac{\partial y_i^{(k)}}{\partial I_j},$$

$$\frac{\partial y_i^{(k)}}{\partial I_j} = \left(\frac{\partial y_i^{(k)}}{\partial z^{(k)}} \right) \left(\frac{\partial z^{(k)}}{\partial x_j} \right) \left(\frac{\partial x_j}{\partial I_j} \right),$$

where $\operatorname{sgn}(z_i)$ is the sign of z_i. The first parenthesized derivative is

$$\frac{\partial y_i^{(k)}}{\partial z_i^{(k)}} = \frac{\sigma + P_i|z| - P_{ii} \operatorname{sgn}(z_i) z_i}{(\sigma + P_i|z|)^2},$$

$$\frac{\partial y_i^{(k)}}{\partial z_l^{(k)}} = \frac{-P_{il} \operatorname{sgn}(z_l) z_i}{(\sigma + P_i|z|)^2}, \qquad l \neq i.$$

The second parenthesized derivative is

$$\frac{\partial z^{(k)}}{\partial x_j} = Q_{(:,j)}^{(k)},$$

where Q is the matrix of the linear transformation performed by the Laplacian pyramid $z = Qx$. The third parenthesized derivative is

$$\frac{\partial x_j^{(k)}}{\partial I_j} = \frac{1}{\gamma} I_j^{(\frac{1}{\gamma}-1)}.$$

APPENDIX B. IMAGE QUALITY ASSESSMENT PERFORMANCE OF NORMALIZED LAPLACIAN PYRAMID

Perceptual image quality assessment (IQA), as a means of comparing results obtained by different methods, has become an important topic in image processing. Although the only guaranteed IQA method is through the explicit measurement of human responses, this is a difficult and costly undertaking. An objective measure of perceptual quality alleviates this difficulty. If the measure is differentiable and well behaved, an additional advantage arises from using it to optimize the perceptual performance of algorithms.

The most widely used method of validating IQA models is by measuring their correlation with human quality ratings on a diverse set of distorted images [8,31,32]. Table 1 presents correlation results against five databases of human mean opinion scores: four were measured using LDR displays, and one was targeted at HDR displays. All results are obtained using the achromatic version of the images in the databases (we have not yet extended the NLP metric to handle color), and leaving out the images with chromatic distortions. We also include the results for several widely employed IQA methods. Note that other IQA methods have been proposed that use local gain control mechanisms [13–15,21].

The table shows results for two types of correlation: the Pearson correlation, which measures linear predictability of the human responses, and the Spearman correlation, which is concerned only with the ranking of the responses, and thus more robust to (monotonic) nonlinear distortions. Note that the latter measure is perhaps too flexible, since the nonlinear relationship between the MOS and the predicted value can

Table 1. Evaluation of IQA Methods in Different Databases[a]

	TID 2008 [19]		TID 2013 [33]		LIVE [34]		CSIQ [35]		EPFL [36]	
PSNR	0.52	(0.55)	0.64	(0.67)	0.86	(0.94)	0.76	(0.81)	0.78	(0.79)
SSIM	0.74	(0.78)	0.77	(0.80)	0.83	**(0.97)**	0.79	(0.87)	0.75	(0.95)
MS-SSIM	0.79	(0.85)	0.79	(0.86)	0.77	**(0.97)**	0.77	(0.91)	0.79	(0.94)
VDP 2.2	0.80	(0.85)	0.78	(0.84)	**0.93**	**(0.97)**	**0.90**	(0.92)	0.90	(0.95)
NLPD	**0.89**	**(0.89)**	**0.88**	**(0.88)**	0.89	**(0.97)**	**0.90**	**(0.93)**	**0.93**	**(0.96)**

[a]Pearson correlation and Spearman correlation (in parentheses) of distance metrics versus human perceptual judgments. The numbers were obtained using the gray-scale version of the images in the databases (see the text for details).

be different for each database, it is often reported when evaluating IQA methods, and we include it here for completeness.

Our results indicate that the proposed metric behaves well for both LDR and HDR images. Note that the parameters of our metric were adjusted using the TID 2008 [19] database, the VDP 2.2 metric was trained using HDR images, the TID 2008 [19] and the CSIQ [35] database, and the SSIM and MS-SSIM metrics were trained using the LIVE database [34]. The Pearson (linear) correlation of our proposed metric is clearly better for four of five datasets (including the HDR dataset), and the Spearman (nonlinear) correlation is equal to or better than all the other metrics for all the datasets. We conclude that our proposed NLP metric is competitive with the current state of the art in IQA. In addition, the NLP metric is the only one that has proven to be easily differentiable for incorporation into optimization procedures. For example, the application of SSIM to optimization procedures is not straightforward and involves some modifications of the metric [37].

REFERENCES

1. B. Hoefflinger, ed., *High-Dynamic-Range (HDR) Vision* (Springer, 2007).
2. J. Tumblin and H. Rushmeier, "Tone reproduction for realistic images," IEEE Comput. Graph. Appl. **13**, 42–48 (1993).
3. J. A. Ferwerda, S. N. Pattanaik, P. Shirley, and D. P. Greenberg, "A model of visual adaptation for realistic image synthesis," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)* (1996).
4. S. N. Pattanaik, J. A. Ferwerda, M. D. Fairchild, and D. P. Greenberg, "A multiscale model of adaptation and spatial vision for realistic image display," in *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)* (1998).
5. J. Tumblin, J. K. Hodgins, and B. K. Guenter, "Two methods for display of high contrast images," ACM Trans. Graph. **18**, 56–94 (1999).
6. S. N. Pattanaik, J. Tumblin, H. Yee, and D. P. Greenberg, "Time-dependent visual adaptation for fast realistic image display," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)* (2000).
7. R. Mantiuk, S. Daly, and L. Kerofsky, "Display adaptive tone mapping," ACM Trans. Graph. **27**, 68 (2008).
8. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process. **13**, 600–612 (2004).
9. V. Laparra, J. Ballé, A. Berardino, and E. P. Simoncelli, "Perceptual image quality assessment using a normalized Laplacian pyramid," in *Proceedings of the Human Vision and Electronic Imaging (HVEI 2016)*, San Francisco, California, February 14–18 (Society for Imaging Science & Technology, 2016), paper HVEI-103.
10. P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," IEEE Trans. Commun. **31**, 532–540 (1983).
11. V. Mante, V. Bonin, and M. Carandini, "Functional mechanisms shaping lateral geniculate responses to artificial and natural stimuli," Neuron **58**, 625–638 (2008).
12. G. Sclar, J. H. R. Maunsell, and P. Lennie, "Coding of image contrast in central visual pathways of the macaque monkey," Vis. Res. **30**, 1–10 (1990).
13. P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *Proceedings of 1st International Conference on Image Processing* (1994), Vol. **2**, pp. 982–986.
14. A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," IEEE Trans. Image Process. **21**, 4695–4708 (2012).
15. A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," IEEE Signal Process. Lett. **20**, 209–212 (2013).
16. D. J. Heeger, "Normalization of cell responses in cat striate cortex," Visual Neurosci. **9**, 181–197 (1992).
17. O. Schwartz and E. P. Simoncelli, "Natural signal statistics and sensory gain control," Nat. Neurosci. **4**, 819–825 (2001).
18. M. Carandini and D. J. Heeger, "Normalization as a canonical neural computation," Nat. Rev. Neurosci. **13**, 51–62 (2013).
19. N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008—a database for evaluation of full-reference visual quality assessment metrics," Adv. Mod. Radioelectron. **10**, 30–45 (2009).
20. A. B. Watson, "DCTune: a technique for visual optimization of DCT quantization matrices for individual images," Society for Information Display Digest of Technical Papers 24 (1993), pp. 946–949.
21. V. Laparra, J. Muñoz Marí, and J. Malo, "Divisive normalization image quality metric revisited," J. Opt. Soc. Am. A **27**, 852–864 (2010).
22. D. P. Kingma and J. L. Ba, "Adam: a method for stochastic optimization," arXiv:1412.6980 (2014). Presented at the 3rd International Conference on Learning Representations, 2015.
23. S. Paris, S. W. Hasinoff, and J. Kautz, "Local Laplacian filters: edge-aware image processing with a Laplacian pyramid," ACM Trans. Graph. **30**, 68 (2011).

24. M. D. Fairchild, "The HDR photographic survey," http://www.cis.rit.edu/fairchild/HDR.html.

25. A. Olmos and F. A. A. Kingdom, "A biologically inspired algorithm for the recovery of shading and reflectance images," Perception **33**, 1463–1473 (2004).

26. K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," IEEE Trans. Pattern Anal. Mach. Intell. **33**, 2341–2353 (2011).

27. R. Fattal, "Dehazing using color-lines," ACM Trans. Graph. **34**, 13 (2014).

28. R. W. Floyd and L. Steinberg, "An adaptive algorithm for spatial grey-scale," Proc. J. Soc. Inf. Disp. **17**, 75–77 (1976).

29. J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end optimization of nonlinear transform codes for perceptual quality," in *Proceedings of 32nd Picture Coding Symposium* (2016).

30. X. Cerdá-Company, C. A. Párraga, and X. Otazu, "Which tone-mapping operator is the best? A comparative study of perceptual quality," arXiv:1601.04450 (2016).

31. Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *37th Asilomar Conference on Signals, Systems and Computers* (2003).

32. M. Narwaria, R. K. Mantiuk, M. P. Da Silva, and P. Le Callet, "HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images," J. Electron. Imaging **24**, 010501 (2015).

33. N. N. Ponomarenko, L. Jin, O. Ieremeiev, V. V. Lukin, K. O. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. J. Kuo, "Image database TID2013: peculiarities, results and perspectives," Signal Proc. Image Commun. **30**, 57–77 (2015).

34. H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," IEEE Trans. Image Process. **15**, 3440–3451 (2006).

35. E. Larson and D. Chandler, "Categorical image quality (CSIQ) database," 2010, http://vision.eng.shizuoka.ac.jp/mod/page/view.php?id=23.

36. P. Korshunov, P. Hanhart, T. Richter, A. Artusi, R. Mantiuk, and T. Ebrahimi, "Subjective quality assessment database of HDR images compressed with JPEG XT," in *7th International Workshop on Quality of Multimedia Experience (QoMEX)* (2015).

37. D. Brunet, E. R. Vrscay, and Z. Wang, "On the mathematical properties of the structural similarity index," IEEE Trans. Image Process. **21**, 1488–1499 (2012).