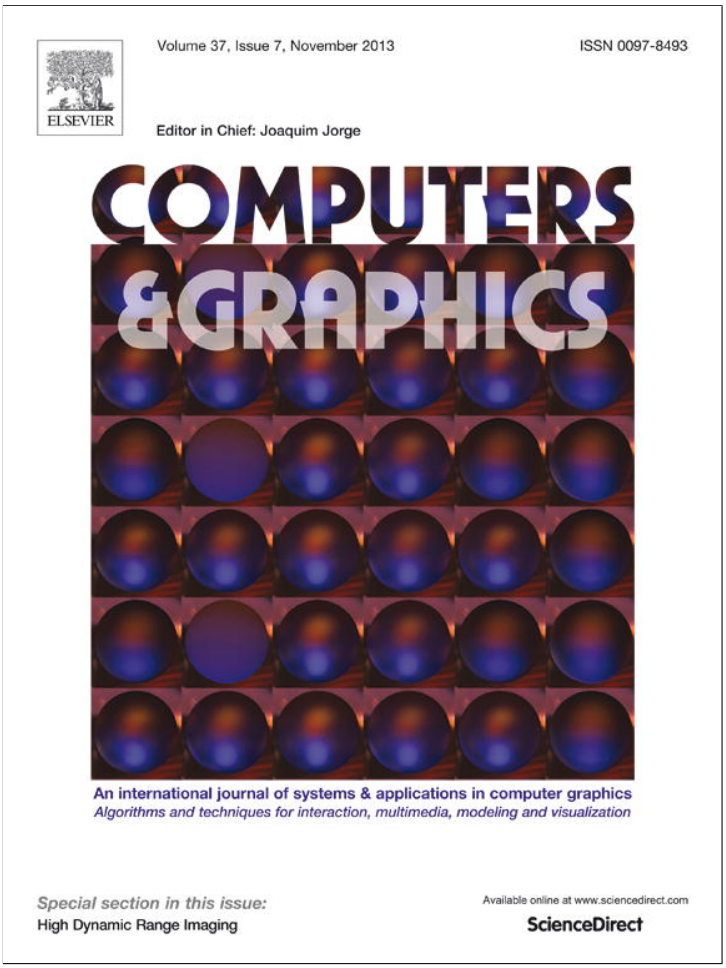

Computers & Graphics 37 (2013) 935–943

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on HDR Imaging

A reality check for radiometric camera response recovery algorithms

Ahmet Oğuz Akyüz*, Aslı Gençtav

Department of Computer Engineering, Middle East Technical University, Universiteler Mahallesi, Dumlupinar Bulvari No: 1, 06800 Cankaya Ankara, Turkey

ARTICLE INFO

Article history:
Received 1 March 2013
Received in revised form
31 May 2013
Accepted 5 June 2013
Available online 12 July 2013

Keywords:
HDR imaging
Camera response functions
Radiometry

ABSTRACT

The radiometric response of a camera governs the relationship between the incident light on the camera sensor and the output pixel values that are produced. This relationship, which is typically unknown and nonlinear, needs to be estimated for applications that require accurate measurement of scene radiance. Until now, various camera response recovery algorithms have been proposed each with different merits and drawbacks. However, an evaluation study that compares these algorithms has not been presented. In this work, we aim to fill this gap by conducting a rigorous experiment that evaluates the selected algorithms with respect to three metrics: consistency, accuracy, and robustness. In particular, we seek the answer of the following four questions: (1) Which camera response recovery algorithm gives the most accurate results? (2) Which algorithm produces the camera response most consistently for different scenes? (3) Which algorithm performs better under varying degrees of noise? (4) Does the sRGB assumption hold in practice? Our findings indicate that Grossberg and Nayar's (GN) algorithm (2004 [1]) is the most accurate; Mitsunaga and Nayar's (MN) algorithm (1999 [2]) is the most consistent; and Debevec and Malik's (DM) algorithm (1997 [3]) is the most resistant to noise together with MN. We also find that the studied algorithms are not statistically better than each other in terms of accuracy although all of them statistically outperform the sRGB assumption. By answering these questions, we aim to help the researchers and practitioners in the high dynamic range (HDR) imaging community to make better choices when choosing an algorithm for camera response recovery.

1. Introduction

Photographic images captured by most cameras are typically stored in a nonlinear color space. In film cameras, this nonlinearity is a result of the nonlinear response of the chemicals used in the film to light. In digital cameras, on the other hand, the nonlinearity is intentionally introduced by the electronics and the firmware during the analog-to-digital conversion and remapping, as optical elements and sensors are inherently linear (see Fig. 1).

Using a nonlinear color space not only serves the purpose of gamma-correction, but also mimics the light response of the human visual system. The human visual system is highly nonlinear, and it is theorized that this nonlinearity allows better utilization of the limited bandwidth of the retinal pathways [4]. Similar to the human eye, digital cameras can encode a large range of incoming light values to a limited number of bits by using a nonlinear color space. Nonlinear encoding also serves to reduce the quantization artifacts and noise as it uses more bits in darker regions for which the human eye is more sensitive to intensity transitions [5]. Finally, nonlinearity is utilized for aesthetic purposes which can be a distinguishing factor between the images produced by different camera manufacturers.

Ideally, digital cameras are expected to adhere to the sRGB standard which has well-defined color primaries and nonlinearity [6]. However, in practice, most digital and film cameras have response curves that are widely different from the sRGB standard [7] (see Fig. 2). Therefore, in applications that require high radiometric precision, such as creating radiance maps from multiple exposures [8,3,2,9], shape from shading algorithms [10,11], and computational photography [12], it is vital to recover the response curve of the camera used rather than relying on the sRGB assumption.

Until now, several methods have been proposed that attempt to recover the unknown response of a digital camera from a set of bracketed exposures [1–3,13–15]. Each method approaches the problem from a different standpoint and makes assumptions about the shape of response curves. However, a formal evaluation of these algorithms in terms of how accurately they estimate an unknown camera response is not available. Our goal in this study is to fill this gap by comparing the performance of the algorithms with respect to three important metrics. Our contributions can be summarized as:

- Developing three metrics that can be used to compare the performance of radiometric response recovery algorithms.

* Corresponding author. Tel.: +90 312 210 5565.
E-mail address: akyuz@ceng.metu.edu.tr (A.O. Akyüz).

0097-8493/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cag.2013.06.003

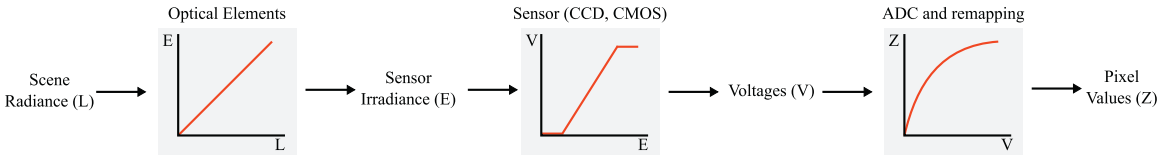

Fig. 1. An abstraction of the photographic pipeline. Typically, the optical and sensor elements are linear, but nonlinearity is introduced during analog to digital conversion and remapping.

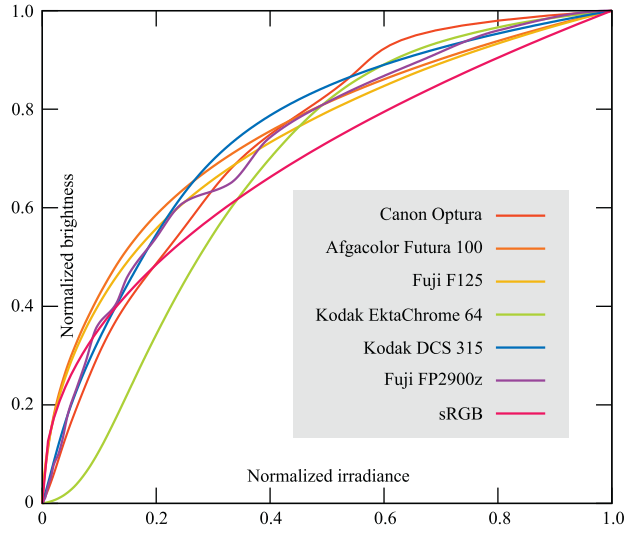

Fig. 2. Response curves of various digital and film cameras. The data source is obtained from the DoRF database (http://www.cs.columbia.edu/CAVE/databases).

- Using these metrics to conduct a rigorous evaluation of four commonly used response recovery algorithms.

2. Background

Many methods exist to recover the unknown radiometric response of a photographic camera. These methods can generally be classified as multi-image or single-image methods. In multi-image methods, multiple exposures of the same scene are used to determine the change in pixel values with respect to a change in exposure enabling one to determine the camera response. These methods rely on the *reciprocity* principle which states that the total exposure X is equal to the product of the image irradiance E and the exposure duration Δt:

$$X = E\Delta t. \tag{1}$$

However, pixel values, Z, are not linearly related to X but rather to a function of it:

$$Z = f(X) = f(E\Delta t) \tag{2}$$

Thus, by varying Δt for a pixel with constant irradiance, one can detect the change in Z, and from there infer the shape of the camera response function, f.

Single-image methods, on the other hand, cannot make use of the reciprocity principle but rely on other cues. Farid observed that nonlinear processing causes specific higher-order distortions in the frequency domain [16]. By detecting and minimizing these distortions one can recover the radiometric response of a camera. Lin et al. [17] argued that edge colors should change linearly between regions of different uniform intensities. Thus, they proposed a function that maps the nonlinear distribution of edge colors to a linear distribution. Later, their method is extended to work on edge histograms for grayscale images [18].

Table 1
Definition of terms used in the equations.

I_p	Irradiance of pixel p
Z_{qp}	Intensity of pixel p in image q
M_{qp}	Normalized intensity in the range $[0, 1]$
t_q	Exposure time of image q
$R_{q,q'}$	Exposure ratio between images q and q'
$w(x)$	Weighting function
$f(x)$	Camera response function
$g(x)$	Inverse camera response function, $f^{-1}(x)$
Q	Number of exposures to combine
P	Number of pixels in each image

In this work, we focus on multi-image response recovery algorithms that are commonly used to create high dynamic range (HDR) images [9]. The specific algorithms that we evaluated are Debevec and Malik's method [3] (abbreviated as DM), Mitsunaga and Nayar's radiometric self calibration [2] (MN), Robertson et al.'s estimation theoretic approach [15] (RBS), and Grossberg and Nayar's [1] principle component analysis based algorithm (GN). Each algorithm is briefly reviewed in the following subsections. The terms used in the following equations are given in Table 1.

2.1. Debevec and Malik's method (DM)

Debevec and Malik present the response recovery problem as the minimization of the following quadratic objective function [3]:

$$\mathcal{O} = \sum_{q=1}^{Q}\sum_{p=1}^{P}\{w(Z_{qp})[\tilde{g}(Z_{qp}) - \ln I_p t_q]\}^2 + \lambda \sum_{z=1}^{254}[w(z)\tilde{g}''(z)]^2, \tag{3}$$

where $\tilde{g} = \ln f^{-1}$ and w is a tent shaped weighting function defined as

$$w(z) = \begin{cases} z/127.5 & \text{for } z \le 127.5, \quad \text{(a)} \\ (255-z)/127.5 & \text{for } z > 127.5. \quad \text{(b)} \end{cases} \tag{4}$$

The first term in Eq. (3) is the data fitting term and the second term is used to force smoothness. Increasing the value of λ brings the recovered response closer to a more idealized logarithmic shape at the cost of deviating it from the actual observations. As this formulation yields an overdetermined system of equations, the unknowns $\tilde{g}$ and I_p can be found in the least squared sense using singular value decomposition.

2.2. Mitsunaga and Nayar's method

Mitsunaga and Nayar, on the other hand, argue that (the inverse of) any response function can be modeled using a higher order polynomial:

$$f^{-1}(x) = \sum_{n=0}^{N} c_n x^n. \tag{5}$$

This reduces the problem of response recovery to determining the coefficients, c_n, and the degree, N, of the polynomial that

minimizes the following objective function [2]:

$$\mathcal{O} = \sum_{q=1}^{Q-1} \sum_{p=1}^{P} \left[\sum_{n=0}^{N} c_n M_{qp}^n - R_{q,q+1} \sum_{n=0}^{N} c_n M_{q+1p}^n \right]^2. \quad (6)$$

The coefficients, c_0 to c_{N-1}, are found by solving the linear system of equations that result from setting $\delta\mathcal{O}/\delta c_i = 0$ for $i = 0, 1 \ldots N-1$. To fix the curve's scale and determine the last coefficient, c_N, the authors set $f(1) = I_{max}$, which gives the additional constraint:

$$c_N = I_{max} - \sum_{n=0}^{N-1} c_n. \quad (7)$$

2.3. Robertson et al.'s method

Robertson et al. assume that the values of $f^{-1}(x)$ are independent Gaussian random variables, which leads to the following objective function to be minimized [15]:

$$\mathcal{O} = \sum_{q=1}^{Q} \sum_{p=1}^{P} w(Z_{qp})(f^{-1}(Z_{qp}) - t_q I_p)^2. \quad (8)$$

Note that, this objective function is similar to that of Debevec and Malik's except that it is formulated in the linear domain and it does not have a smoothing term. Also it uses a different weighting function which is defined as

$$w(z) = e^{-4(z-127.5)^2/127.5^2}. \quad (9)$$

This weighting function together with Debevec and Malik's weighting function are plotted in Fig. 3. Both functions are characteristically similar in that they give higher weights to pixels closer to the middle of the exposure range. The authors solve for the values of the response curve by using a form of Gauss–Seidel relaxation.

2.4. Grossberg and Nayar's method

Grossberg and Nayar approach the problem from an empirical standpoint. They argue that instead of making assumptions about the shape of a response curve (e.g., polynomial as in [2] or logarithmic as in [3]), the real world responses of a multitude of films and digital cameras can be subjected to principal component analysis (PCA) to find the best basis. From this assumption, they model any response curve (and its inverse) using the following formulae:

$$f(x) = f_0(x) + \sum_{n=1}^{N} c_n h_n(x), \quad (10)$$

$$f^{-1}(x) = f_0^{inv}(x) + \sum_{n=1}^{N} c_n h_n^{inv}(x), \quad (11)$$

where $f_0(x)$ and $f_0^{inv}(x)$ are the means of all response and inverse response curves respectively and $\{h_1, h_1^{inv}, h_2, h_2^{inv}, \ldots, h_N, h_N^{inv}\}$ are the basis functions as determined by PCA. The authors performed this analysis on 201 real world response curves (as well as their inverses) to compute the mean curves and the basis functions, and made this data available in a public database.

To recover the response of a camera using this data, one can use an objective function similar to the one used by Mitsunaga and Nayar,

$$\mathcal{O} = \sum_{q=1}^{Q-1} \sum_{p=1}^{P} \left[f_0^{inv}(M_{qp}) + \sum_{n=1}^{N} c_n h_n^{inv}(M_{qp}) - R_{q,q+1} \left\{ f_0^{inv}(M_{q+1p}) + \sum_{n=1}^{N} c_n h_n^{inv}(M_{q+1p}) \right\} \right]^2, \quad (12)$$

and find the coefficients, c_n, using least squares techniques.

2.5. Implementation details

We implemented all four algorithms in C++ using the exact procedures outlined in the original papers. Each of these methods requires a carefully selected set of image samples. Using all pixels from all exposures not only increases the computational cost, but also negatively affects the performance of the algorithms due to possible misalignment and noise in the data. We followed the sample selection procedure suggested in Reinhard et al. [9], and chose 300 distinct sample positions from the uniform regions of the middle three exposures for each sequence using randomized rejection sampling. This gave rise to a total of $300 \times Q$ samples for an image sequence with Q images. We ensured that the samples were not selected from under- and over-exposed regions, and that they were not clumped together. A representative set of samples found using our algorithm is shown in Fig. 4.

In Debevec and Malik's algorithm, we set the value of the smoothing term λ equal to 10. In the original paper, there is no suggested value for this parameter. We decided on this value after experimenting with several other values. We found that smaller values tend to result in jagged curves and larger values make the curves too smooth. In Mitsunaga and Nayar's algorithm, we tested polynomials from degree 2 to 6 to find the best matching polynomial. We set the upper limit to 6 as in our experiments

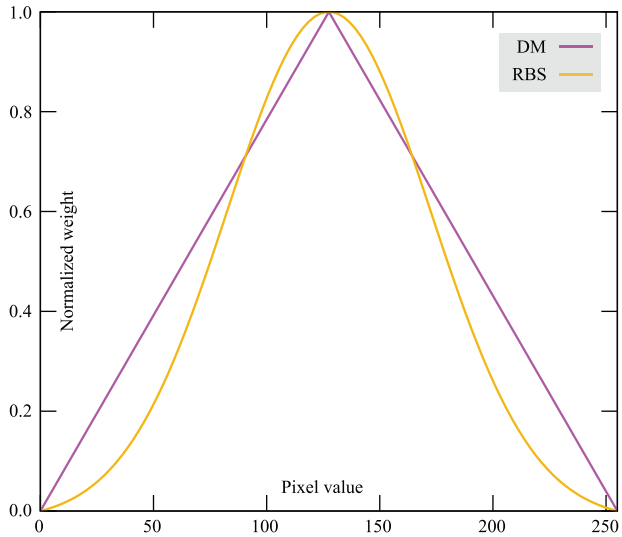

Fig. 3. The weighting functions used by Debevec and Malik, and Robertson et al. during camera response recovery.

Fig. 4. The red squares indicate a representative set of selected samples for one of the test images. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

using higher degree polynomials always resulted in over-fitting resulting in implausible response curves. In Grossberg and Nayar's algorithm, we modeled the inverse response function using 4 basis functions as these capture more than 99.5 of the total energy [1].

We observed that the recovered responses occasionally had problems around very low and high pixel values, and for values that are not represented by enough samples. For instance, Mitsunaga and Nayar's algorithm occasionally produced negative values for pixel values close to zero. When this occurred, we clamped these implausible values to zero. Also, if some parts of the response curve violated the monotonicity assumption such that lower pixel values mapped to higher irradiances than higher pixel values, we fixed these regions by linear interpolation using the surrounding good regions. These types of artifacts were rare and only affected small portions of the response curve when they occurred. Finally, we scaled all response curves to the range of $[0,1]$ to bring them to the same domain.

3. Evaluation

We evaluated the performance of the selected algorithms using three different metrics, namely *consistency*, *accuracy*, and *robustness*. Below, we motivate and explain the importance of each of these metrics and how we derived them.

3.1. Metric 1: consistency

We define consistency as the repeatability of an algorithm's result with changing data. More specifically, we consider an algorithm as consistent if it produces the same camera response using different image sets. Consistency relies on the idea that the camera response should stay the same if one only changes the exposure while keeping other camera settings intact. Thus, a consistent algorithm allows a user to generate a response curve once from an image set, and use it to linearize other images taken with the same camera. Our metric for consistency is defined as

$$\overline{\sigma} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{1}{256}\sum_{z=0}^{255}(\overline{g}(z)-g_i(z))^2} \tag{13}$$

where g_i denotes the inverse camera response recovered from image set i, $\overline{g}$ is the mean inverse camera response obtained from all calculated curves, and N is the number of image sets. This metric computes the mean standard deviation of individual camera responses.

3.2. Metric 2: accuracy

Accuracy measures the closeness of a recovered camera response to the ground truth. The ground truth can be determined by imaging a color chart with known reflectances, such as the Macbeth color chart, to determine the pixel values to which these regions map [20]. However, as there are limited number of color patches with known reflectance, the full camera response can only be found by interpolation. Therefore, whether the interpolated values correspond to the actual response of the camera cannot be guaranteed [1].

In our evaluation, we devised an alternative accuracy metric which is based on the reciprocity principle. The reciprocity principle states that the measured pixel values are proportional to the product of the sensor irradiance by the image exposure. On a real camera, this product is modified by the camera response function to yield the observed pixel values [3]:

$$Z_{qp} = f(E_p t_q), \tag{14}$$

where E_p is the sensor irradiance of pixel p. Thus, if one undoes the effect of the camera response, the resulting value should be linearly proportional to the product of the sensor irradiance by the image exposure. Therefore, if the camera response is recovered perfectly, the following relationship should hold between two corresponding pixels in two images:

$$g(Z_{qp}) - g(Z_{q'p})\frac{t_q}{t_{q'}} = 0 \tag{15}$$

Our metric is simply the generalization of this relationship to include all pixels from all exposures:

$$\mathcal{E} = \frac{1}{W}\sqrt{\sum_{q=1}^{Q-1}\sum_{q'=q+1}^{Q}\sum_{p=1}^{P}K^2} \tag{16}$$

where K is the weighted sum of reciprocity errors:

$$K = w(Z_{qp}, Z_{q'p})[g(Z_{qp}) - R_{q,q'}\,g(Z_{q'p})], \tag{17}$$

and W is the sum of all weights:

$$W = \sum_{q=1}^{Q-1}\sum_{q'=q+1}^{Q}\sum_{p=1}^{P} w(Z_{qp}, Z_{q'p}). \tag{18}$$

The purpose of the weighting function w is to minimize the effect of saturated pixels as the relationship in Eq. (15) will not hold if any of the pixels in question is saturated. We defined w as

$$w(x,y) = \begin{cases} 0 & \text{if } x \text{ or } y \text{ not in } [5,250] \\ (x+y)/255 & \text{if } (x+y)/2 \le 127.5 \\ (255-(x+y)/2)/127.5 & \text{if} (x+y)/2 > 127.5, \end{cases} \tag{19}$$

to give the highest weight to the middle of the exposure range and underplay the influence of the over- and under-exposed pixels.

At this point, one may question why we did not directly compare the recovered camera responses with the actual responses of the camera. Our answer is two-fold. First, recovering the actual response of a camera is an immensely challenging task due to the effects of optical glare [21,22]. Second, we observed that one does not strictly need the ground truth curve to measure accuracy. It is because, if the camera response is the aggregate effect of nonlinearities in an imaging system, determining and reversing its effects should give rise to linear images. The linear images should satisfy the reciprocity principle in that the measured pixel values are proportional to the product of the sensor irradiance by the image exposure. If this proportionality does not hold, one can attribute this to the imperfections in the recovered camera response. In other words, accuracy can also be measured indirectly by comparing the observations rather than directly comparing the camera response curves.

3.3. Metric 3: robustness

Robustness can be defined as how well an algorithm continues to produce acceptable results as the imaging conditions depart from the ideal. To measure robustness, we used a ray-traced HDR image as an artificial scene. We created multiple exposures from this scene as if by taking photographs with a virtual camera (see Fig. 5). The virtual camera was designed to have a perfect sRGB response. During the imaging process, we introduced varying degrees of Gaussian noise to simulate various sources of noise that may infect a real imaging system (Fig. 6). Finally, we used these noise added sequences to recover the response curves of the imaginary camera, and compared them against the ground truth. Here it is important to note that we used the same sample positions selected from the noise-free sequence to recover the responses using all sequences. This was done to avoid changing both source images and sample positions, and to perform a controlled experiment by isolating the effect of noise.

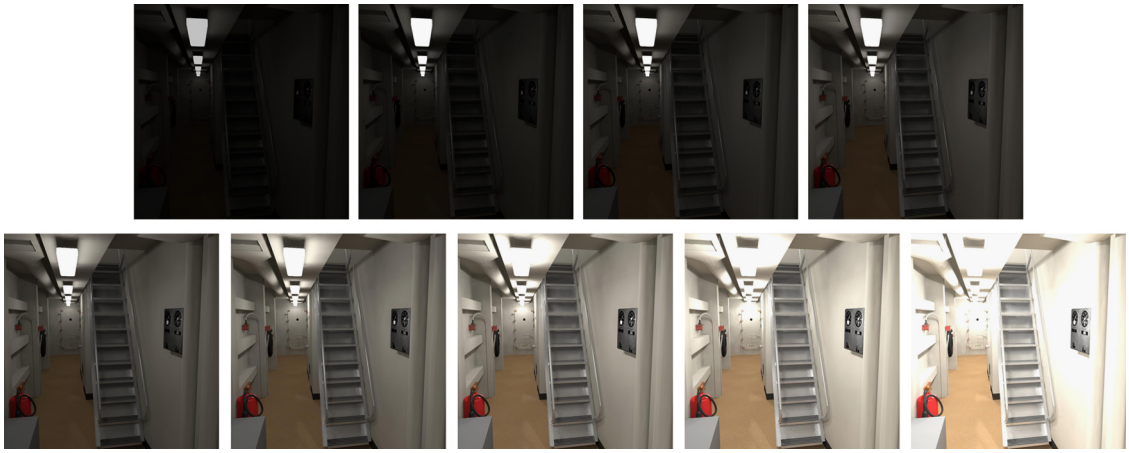

Fig. 5. Virtual exposures created by taking a virtual photograph from a ray-traced HDR image. The HDR image is rendered by the Radiance software [19] (courtesy of Greg Ward). The virtual camera was designed to have a perfect sRGB response.

Fig. 6. From left to right the images show the effect of adding Gaussian noise with $\sigma = 0$ (no noise), $\sigma = 0.25$, $\sigma = 0.50$, $\sigma = 1$, and $\sigma = 1.5$.

To quantify the performance of the algorithms under varying degrees of noise, we used a standard deviation metric:

$$\sigma = \sqrt{\frac{1}{256}\sum_{z=0}^{255}(g_{sRGB}(z) - g_{\sigma}(z))^2} \tag{20}$$

where g_{sRGB} is the response of the sRGB specification and g_{σ} is the response recovered by an algorithm from the image sequence that was created by adding a Gaussian noise with a standard deviation of σ.

4. Results

To measure the performance of the algorithms with respect to the metrics defined above we used a large set of images captured using three different digital cameras. The cameras we used were Canon EOS 550D, Nikon D2H, and Nikon Coolpix E5400. With each camera, we captured bracketed sequences by mounting the camera on a tripod to avoid camera movement during the capture process. We installed the Magic Lantern firmware patch on Canon EOS 550D to allow auto-bracketing (AEB) up to 9 exposures with each exposure 1 f-stop apart (by default the camera allows up to three exposures). The other two cameras, Nikon D2H and Nikon Coolpix E5400, by default allow 9 and 5 exposure AEB respectively and therefore we used the default camera firmware. When 5 exposures were not enough to capture the luminance variation in a given scene for Coolpix E5400, we took another 5 exposures by shifting the value of the center exposure. The resolutions of the images were 5184×3456 for EOS 550D, 1840×1224 for D2H, and 1600×1200 for E5400. Our entire data set is illustrated in Fig. 9.

Our experiments gave rise to a large number of plots and tables, and for lack of space and brevity we will only present representative results. Most notably we only show the results of the green channel as they are representative of all channels. Our full set of results will be shared on the project website.

4.1. Consistency

The consistency plots for the green channel of Canon EOS 550D for each algorithm is shown in Fig. 8(a). Here, we can see that Debevec and Malik's and Robertson et al.'s methods produce more jagged response curves while Mitsunaga and Nayar's and Grossberg and Nayar's methods produce smoother ones. This is expected due to smooth fitting functions used by the latter two (i.e., polynomials and basis functions). Debevec and Malik's method has a smoothing parameter, λ, which we set to 10 in our experiments. We observed that increasing the value of this parameter makes the resulting curves smoother, albeit at the cost of deviating it from the actual observations. Robertson et al.'s method has no built-in smoothing mechanism and therefore the jagged curves were expected. However, despite variations between the response curves, each algorithm produced relatively similar results for different scenes. On average, we found that Mitsunaga and Nayar's method was the most consistent followed by Grossberg and Nayar's algorithm as shown in Table 2. They were followed by Debevec and Malik's, and Robertson et al.'s algorithms.

Here, one may wonder why the recovered camera response changes at all with respect to changing capture conditions. An important reason for this is the *optical glare* which represents the

Table 2
Consistency error of the algorithms as computed according to Eq. (13) using the green channel of all cameras. The smaller figures indicate higher consistency.

Camera	DM	RBS	MN	GN
Canon EOS 550D	0.011	0.011	0.016	0.011
Nikon D2H	0.018	0.020	0.004	0.028
Nikon Coolpix E5400	0.059	0.091	0.016	0.018
Average	0.029	0.041	**0.012**	0.019

scattered light falling on the sensor. As demonstrated by earlier studies [21], glare is scene-, exposure-, lens-, aperture-, and camera-dependent. While methods exist to eliminate glare, they either make assumptions on the shape and extent of the point spread function of the glare [9] or require extra equipment to be placed between the camera and the captured scene [23]. In this study, we did not use any method to account for the glare as our goal was to evaluate the performance of the camera response recovery algorithms under normal imaging conditions.

4.2. Accuracy

To evaluate accuracy, we computed the average response curves generated by each method for all scenes captured by a given camera. The average response curves for the green channel of Canon EOS 550D are shown in Fig. 8(b) together with the sRGB curve for reference. Note that these curves deviate from the sRGB specification at various degrees although each camera was set to output in the sRGB color space. We also note that, as the camera shifts from professional to consumer, the resulting response curves deviate more significantly from the sRGB standard. This is particularly notable for the response curves generated by Debevec and Malik's and Robertson et al.'s methods, where the response of the most professional camera, the D2H, is closest to sRGB, while the response of Coolpix E5400 is the furthest. This may be an expected result as the increased quality generally results in a better matching of the standards.

The accuracy error of each algorithm computed according to Eq. (16) from the green channel of Canon EOS 550D is tabulated in Table 3. This table shows that, on average, the highest accuracy is achieved by Grossberg and Nayar's method followed by Robertson et al.'s algorithm. Debevec and Malik's algorithm is also close to Robertson et al.'s results. We found Mitsunaga and Nayar's method to be the least accurate. However, as can be seen from the rightmost column, the error associated with the sRGB assumption is the highest. This suggests that the sRGB assumption does not hold in practice and that, for applications requiring radiometric precision, it is crucial to recover the actual response of the camera.

It is interesting to note that the accuracy values vary almost uniformly across the algorithms for different images. We attribute this to the scene content. The scenes with static backgrounds appear to have higher accuracy (lower error) such as the *apples*, *miki*, and *workshop*. The scenes with more dynamic backgrounds (those that include trees, leaves, etc.), such as *window2*, *statue*, and *hill*, give rise to higher errors. This is expected because for dynamic scenes the errors due to small object movements contribute to the errors due to imperfect recovery of the camera response.

We further analyzed these results to determine if the differences between the algorithms are statistically significant. The box plots for the reported accuracy error values are shown in Fig. 7. In this figure, one can better see the distribution of accuracy errors yielded by each algorithm. The sRGB curve results in the highest error with other algorithms clumped close together. A one way ANOVA test confirms that the differences are significant, $F(4,50) = 35.341$, $p < 0.001$. Perhaps surprisingly, a multiple comparisons t-test with Bonferroni correction indicates that the differences between the algorithms are not significant with all high p values. However, each algorithm is found to be significantly better than the sRGB approximation.

Table 3
Accuracy error of the algorithms computed according to Eq. (16) from the green channel of Canon EOS 550D. The smaller figures indicate higher accuracy.

Scene	DM	RBS	MN	GN	sRGB
apples	0.500	0.489	0.530	0.453	1.715
corridor	0.323	0.324	0.398	0.328	1.827
kenya	0.439	0.427	0.506	0.411	1.598
window1	0.524	0.513	0.562	0.467	1.759
window2	1.325	1.287	1.349	1.123	2.497
workshop	0.464	0.463	0.487	0.426	1.756
hill	1.022	1.008	?1.036	0.870	2.194
stadium	0.454	0.453	0.476	0.436	1.615
statue	1.105	1.070	1.137	0.937	2.478
miki	0.487	0.493	0.577	0.451	2.080
boat	0.995	0.962	1.028	0.841	2.160
Average	0.694	0.681	0.735	**0.613**	1.971

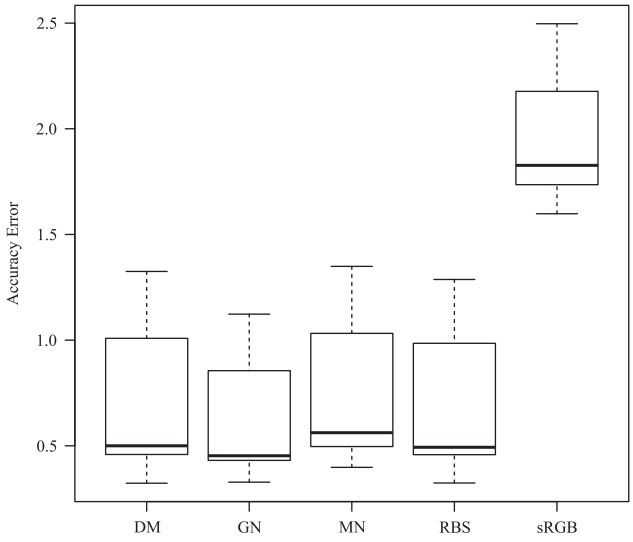

Fig. 7. The box plots showing the accuracy error of all algorithms for all test scenes. As can be seen from the plot, the sRGB curve yields the highest error with other algorithms clumped close together.

4.3. Robustness

The response curves recovered by each algorithm under varying degrees of noise are depicted in Fig. 8(c). Here, we see both expected and surprising results. For instance, the response curves recovered by Mitsunaga and Nayar's algorithm deviate more from sRGB with increasing noise. This pattern can also be seen in Grossberg and Nayar's algorithm. The other two algorithms also produce good results when there is no noise. However, with the addition of noise their behavior changes unexpectedly in that the least amount of noise ($\sigma = 0.25$) causes the largest deviation from the ground truth. While we do not have a full explanation for this phenomenon, we hypothesize that this could be due to the possibility that even the smallest amount of noise that we added was too high for these algorithms. If we added smaller amounts of noise (e.g., $\sigma = 0.05$, $\sigma = 0.10$, etc.), we could see a more expected pattern. However, we did not test this hypothesis.

Our average results for robustness are shown in Table 4. Here, we observe that Debevec and Malik's and Mitsunaga and Nayar's algorithms are the most robust under increasing noise. The other

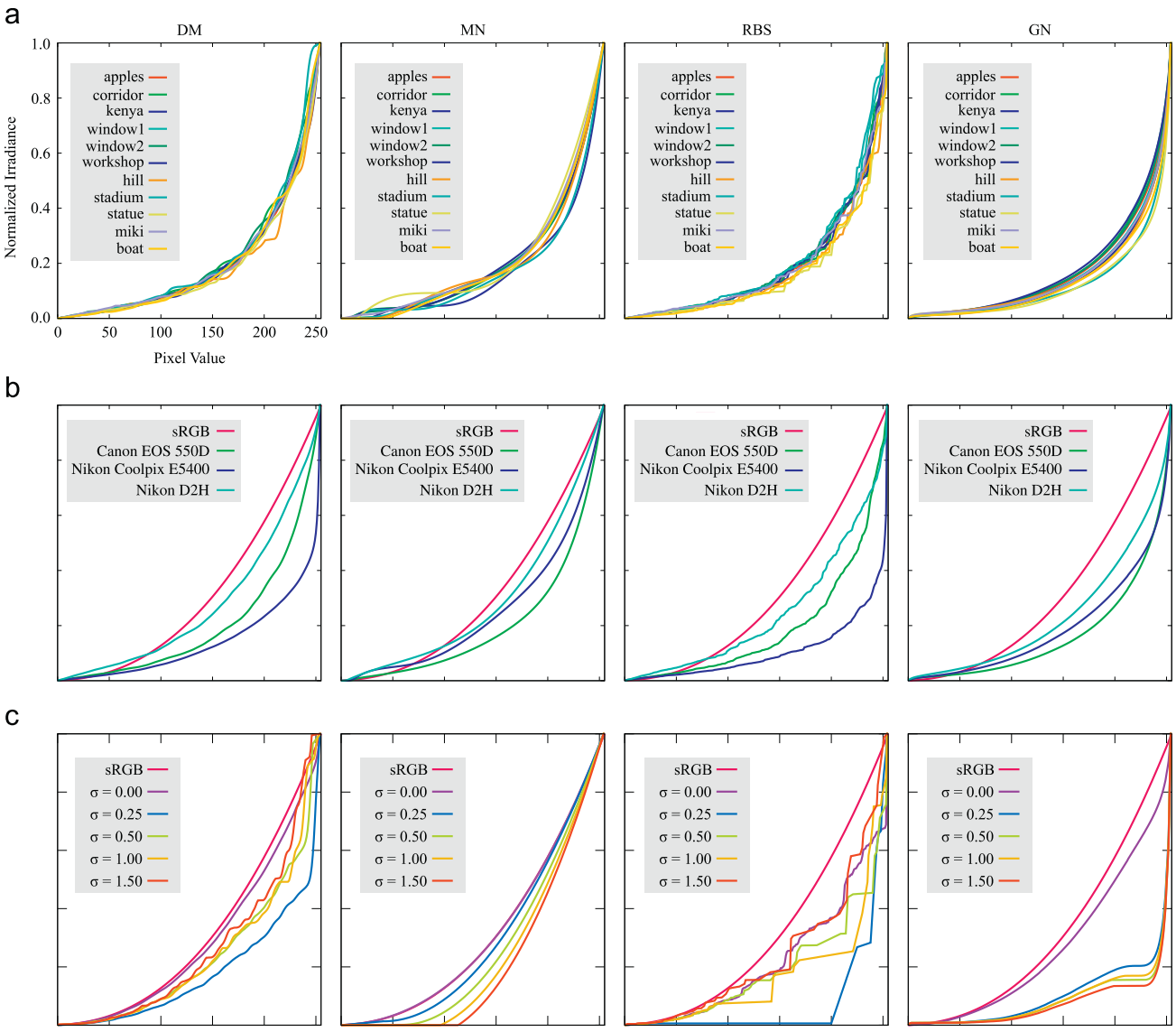

Fig. 8. (a) Response curve of Canon EOS 550D recovered for various image sequences. (b) Mean response curves for each camera. (c) Effect of increasing noise on the recovered response curves. In all plots, only the green channel responses are shown.

Table 4
Robustness error of each algorithm under varying degrees of noise. The smaller figures indicate higher robustness.

Std. dev.	DM	RBS	MN	GN
$\sigma = 0.00$	0.028	0.124	**0.002**	0.058
$\sigma = 0.25$	0.179	0.306	**0.024**	0.280
$\sigma = 0.50$	0.106	0.168	**0.059**	0.304
$\sigma = 1.00$	0.101	0.195	**0.089**	0.297
$\sigma = 1.50$	**0.071**	0.103	0.111	0.315

Table 5
Comparison of the response recovery methods in terms of running time. Performance testing was made on an Intel Core i7 CPU at 3.20 GHz to recover the response curve from a sequence captured by Canon EOS 550D.

Method	Running time (s)
DM	74.68
MN	0.29
RBS	8.39
GN	0.06

two methods are impacted more significantly as the observations depart from the ideal.

4.4. Time performance

Finally, we report the running time of each algorithm in Table 5 to give an insight into the algorithms' efficiencies. As the table shows, Grossberg and Nayar's and Mitsunaga and Nayar's methods perform orders of magnitude faster than the other methods. Debevec and Malik's method performs the slowest as it involves solving a bigger matrix than other methods (approximately 3000×500 for 300 samples and 9 exposures). These results suggest that the former two algorithms may be the preferred choice if they were to be used in an interactive environment.

5. Discussion and conclusion

In this study, we compared the performance of four commonly used radiometric response recovery algorithms using three different metrics. In our experiments we used a large set of images captured using three different digital cameras. Based on these experiments,

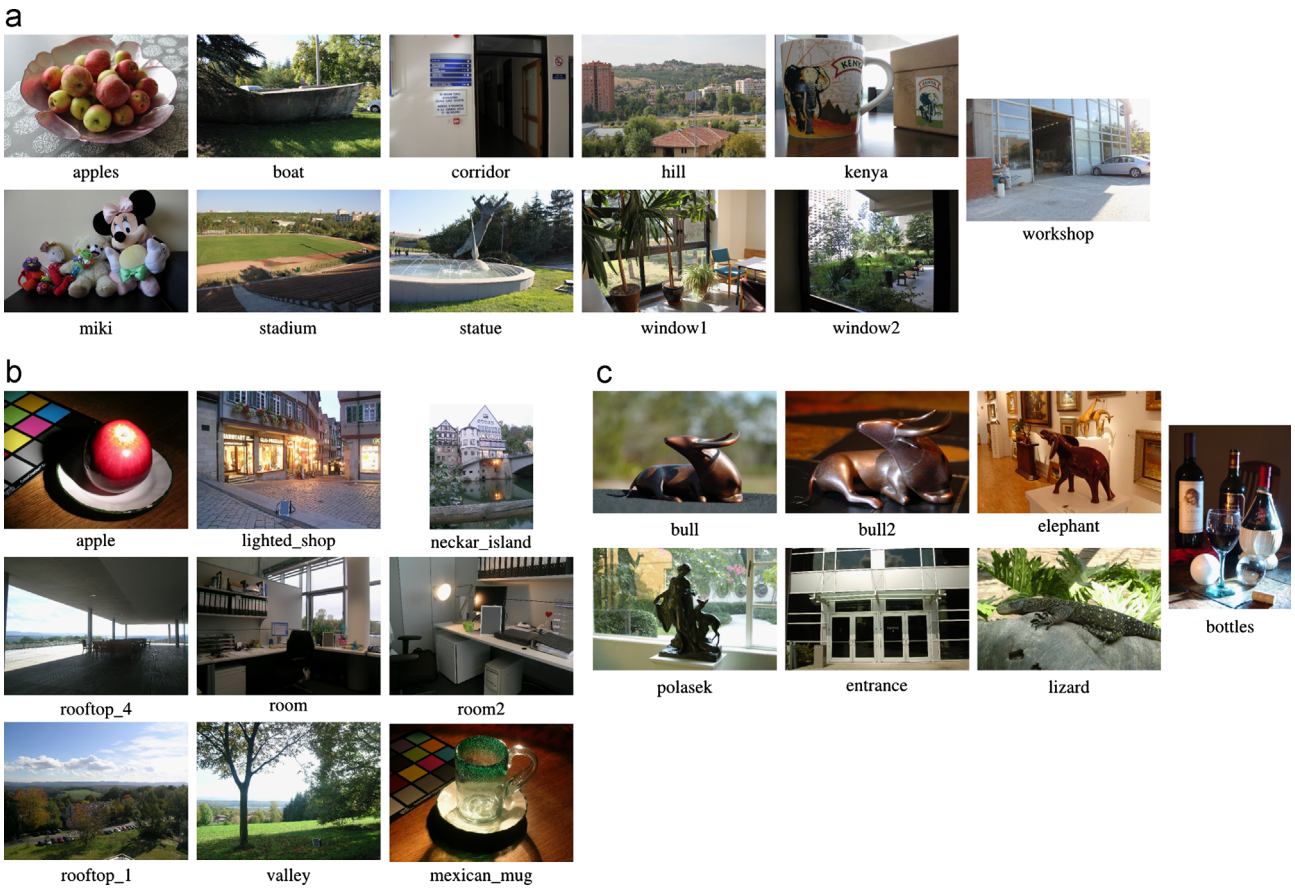

Fig. 9. Data set used in our experiments. Each image belongs to a bracketed sequence of 5 or 9 exposures. Nikon D2H images are courtesy of Erik Reinhard. (a) Canon EOS 550D, (b) Nikon Coolpix E5400 and (c) Nikon D2H.

we found that (1) Grossberg and Nayar's algorithm, on average, is the most accurate in terms of how well the images processed by it preserve the reciprocity principle, (2) Mitsunaga and Nayar's algorithm is the most consistent across different image sequences, (3) Debevec and Malik's and Mitsunaga and Nayar's algorithms are the most robust under increasing noise, and (4) the sRGB assumption fails to hold in practice especially for lower-end cameras. We also found that the accuracy of all four algorithms are not statistically better than each other when their variances are considered for different scenes. However, all four algorithms outperform the sRGB assumption in the statistical sense.

While these results may not give a single answer to which algorithm to choose when recovering the response of a camera, they, nevertheless, shed light on the behavior of radiometric response recovery algorithms. For instance, we argue that the response of a camera should not be recovered only once from a single image sequence, but instead should be allowed to evolve over time with new image sequences. This approach can be put into use in software programs, for instance, to create more accurate HDR images.

We also find that not every scene is equally appropriate to recover a camera response. The scenes that are too dark and too light can particularly be problematic. Also if the dynamic range of a scene is not high enough, it may be difficult to find enough samples to represent all parts of the camera response. In our experience, we observed that relatively lighter scenes with large white regions of different luminance levels were more suitable.

The metrics that we proposed in our study can be used to evaluate the performance of future response recovery algorithms. As future work, we are planning to validate the accuracy of these metrics by using ground truth observations obtained by light measurement devices and surfaces with known reflectances. Finally, we would like to compare single-image methods with multi-image methods.

References

[1] Grossberg M, Nayar S. Modeling the space of camera response functions. IEEE Trans Pattern Anal Mach Intell 2004;26(10):1272–82.
[2] Mitsunaga T, Nayar SK. Radiometric self calibration. In: Proceedings of CVPR, vol. 2, 1999. p. 374–80.
[3] Debevec PE, Malik J. Recovering high dynamic range radiance maps from photographs. In: SIGGRAPH 97 conference proceedings, 1997. p. 369–78.
[4] Palmer SE. Vision science: photons to phenomenology. Cambridge, MA: The MIT Press; 1999.
[5] Poynton C. Digital video and HDTV: algorithms and interfaces. Morgan Kaufmann Publishers; 2003.
[6] ITU-R Recommendation BT.709-5, Parameter values for the HDTV standards for production and for international programme exchange. Geneva: ITU (International Telecommunication Union); 2002.
[7] Grossberg MD, Nayar SK. What is the space of camera response functions? In: IEEE CVPR, vol. 2, 2003. p. 602–9.
[8] Mann S, Picard RW. Being 'undigital' with digital cameras: extending dynamic range by combining differently exposed pictures, vol. 323. Boston, MA: M.I.T. Media Lab Perceptual Computing Section; 1994. p. 422–8.
[9] Reinhard E, Ward G, Pattanaik S, Debevec P. High dynamic range imaging: acquisition, display and image-based lighting. 2nd ed. San Francisco: Morgan Kaufmann; 2010.
[10] Horn BKP, Brooks MJ. Shape from shading. MIT Press; 1989.
[11] Zhang R, Tsai PS, Cryer J, Shah M. Shape-from-shading: a survey. IEEE Trans Pattern Anal Mach Intell 1999;21(8):690–706.
[12] Adams A, Talvala EV, Park SH, Jacobs DE, Ajdin B, Gelfand N, et al. The Frankencamera: an experimental platform for computational photography. ACM Trans Graph 2010;29(4):29:1–12.
[13] Mann S. Comparametric equations with practical applications in quantigraphic image processing. IEEE Trans Image Process 2000;9(8):1389–96.
[14] Tsin Y, Ramesh V, Kanade T. Statistical calibration of ccd imaging process. In: IEEE international conference on computer vision, vol. 1, 2001. p. 480.

[15] Robertson M, Borman S, Stevenson R. Estimation-theoretic approach to dynamic range enhancement using multiple exposures. Journal of Electronic Imaging 2003;12(April (2)):219–28.
[16] Farid H. Blind inverse gamma correction. IEEE Trans Image Process 2001;10(10):1428–33 URL ⟨http://www.cs.dartmouth.edu/farid/publications/ip01.html⟩.
[17] Lin S, Gu J, Yamazaki S, Shum H. Radiometric calibration from a single image. In: IEEE computer society conference on computer vision and pattern recognition, vol. 2, 2004. p. 938–45.
[18] Lin S, Zhang L. Determining the radiometric response function from a single grayscale image. In: IEEE computer society conference on computer vision and pattern recognition, 2005. CVPR 2005, vol. 2. 2005. p. 66–73 http://dx.doi.org/10.1109/CVPR.2005.128.
[19] Ward GJ. The RADIANCE lighting simulation and rendering system. In: Glassner A. editor. Proceedings of SIGGRAPH '94, 1994. p. 459–72.
[20] Chang YC, Reid J. Rgb calibration for color image analysis in machine vision. IEEE Trans Image Process 1996;5(10):1414–22.
[21] McCann JJ, Rizzi A. Camera and visual veiling glare in hdr images. J Soc Inf Disp 2007;15(9):721–30.
[22] McCann JJ, Rizzi A. The art and science of HDR imaging, vol. 26. Wiley; 2011.
[23] Talvala EV, Adams A, Horowitz M, Levoy M. Veiling glare in high dynamic range imaging. ACM Trans Graph (TOG) 2007;26(3).