

Issues with Common Assumptions about the Camera Pipeline and Their Impact in HDR Imaging from Multiple Exposures*

R. Gil Rodríguez[†], J. Vazquez-Corral[†], and M. Bertalmío[†]

Abstract. Multiple-exposure approaches for high dynamic range (HDR) image generation share a set of building assumptions: that color channels are independent and that the camera response function (CRF) remains constant while changing the exposure. The first contribution of this paper is to highlight how these assumptions, which were correct for film photography, do not hold in general for digital cameras. As a consequence, results of multiexposure HDR methods are less accurate, and when tone-mapped they often present problems like hue shifts and color artifacts. The second contribution is to propose a method to stabilize the CRF while coupling all color channels, which can be applied to both static and dynamic scenes, and yield artifact-free results that are more accurate than those obtained with state-of-the-art methods according to several image metrics.

Key words. high dynamic range, multiexposure sequence, projective transformation, color stabilization

AMS subject classifications. 68U10, 62H35, 65D18, 94A08

DOI. 10.1137/19M1250248

1. Introduction. The dynamic range of light intensities in a natural scene, defined as the ratio between the highest and the lowest luminance values, may easily span five orders of magnitude or more. While in most common situations the light coming from a scene is of high dynamic range (HDR), the vast majority of sensors (and displays) are of low dynamic range (LDR). The net result is that standard cameras are only able to capture different intervals of the luminance range at different exposure times; in particular, bright areas are properly captured using short exposure times, while dark areas are better captured using longer exposure times.

To overcome this limitation, Mann and Picard, in their seminal work [24], introduced the idea of creating an HDR picture of a static scene by combining a set of LDR images taken with different exposure times, proposing a parametric method to estimate the camera response function (CRF) that transforms the linear data into nonlinear form. This was soon followed by other very influential approaches for the problem that differ in the way the constant CRF is estimated, e.g., Debevec and Malik [5]; Mitsunaga and Nayar [28]; and Tsing, Ramesh, and Kanade [35]. Later works tackled more general cases, like dynamic scenes with camera

*Received by the editors March 19, 2019; accepted for publication (in revised form) July 15, 2019; published electronically October 1, 2019.

<https://doi.org/10.1137/19M1250248>

Funding: The work of the authors was supported by the European Unions Horizon 2020 research and innovation programme grant 761544 (project HDR4EU) and grant 780470 (project SAUCE) and by the Spanish government and FEDER fund grant TIN2015-71537-P (MINECO/FEDER,UE). The work of the second author was supported by the Spanish government under grant IJCI-2014-19516.

[†]Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona 08018, Spain (raquel.gil@upf.edu, javier.vazquez@upf.edu, marcelo.bertalmio@upf.edu).

and/or object motion [23, 14, 15, 10, 13, 29, 18] or video [34, 16, 11], and it can be said that multiexposure HDR creation is an ongoing research topic [27, 4, 21, 31].

Multiple-exposure approaches that use nonlinear input pictures assume the following image formation model:

$$(1.1) \quad J(p) = f(E(p)\Delta t),$$

where Δt is the exposure time, p is a pixel location, $E(p)$ is the scene radiance value at p , f is a nonlinear transform usually denoted as the CRF, and $J(p)$ is the resulting LDR image value, corresponding to one color channel. Analogous expressions hold for each of the three color channels, for which the function f might be different. In a static scene the values $E(p)$ remain constant, so taking a stack of N pictures by varying the exposure times gives us for each image

$$(1.2) \quad J_i(p) = f(E(p)\Delta t_i), \quad i = 1, \dots, N,$$

where the subindex i denotes the different exposures *and it is also assumed that the function f remains constant as Δt_i changes*. Multiple-exposure methods estimate the inverse g of the CRF f , $g \equiv f^{-1}$, apply it to the image values $J_i(p)$, and then divide by the exposure time Δt_i so as to obtain one estimate of $E(p)$ for each image i in the stack:

$$(1.3) \quad \frac{g(J_i(p))}{\Delta t_i} = E(p), \quad g \equiv f^{-1}.$$

These N estimates of $E(p)$ are then averaged in order to provide the final output, the HDR value for pixel p .

We can see then how most multiple-exposure approaches share a set of building assumptions for the camera capture:

1. Different color channels are independent.
2. The camera response remains constant while changing the exposure.

The contributions of this paper are, first, to highlight how these assumptions do not hold in general for digital cameras, so multiexposure HDR methods based on them often produce HDR pictures that when tone-mapped show hue shifts, color artifacts, or contrast problems. As a second contribution we propose a method to improve multiple-exposure combination, compensating for the violations of assumptions (1) and (2) above and allowing us to obtain more precise HDR images from nonlinear LDR inputs, both for static and for dynamic scenes, that when tone-mapped show no signs of spatiotemporal artifacts of any kind.

2. The response function of digital cameras. The assumptions that different color channels are independent and that the camera response remains constant while changing the exposure are correct in the case of film photography, but they are not an accurate model of how digital cameras work. Digital cameras follow a typical camera color processing pipeline [3] that can be expressed as

$$(2.1) \quad \begin{bmatrix} R \\ G \\ B \end{bmatrix}_{out} = \left(A \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}_{in} \right)^\gamma,$$

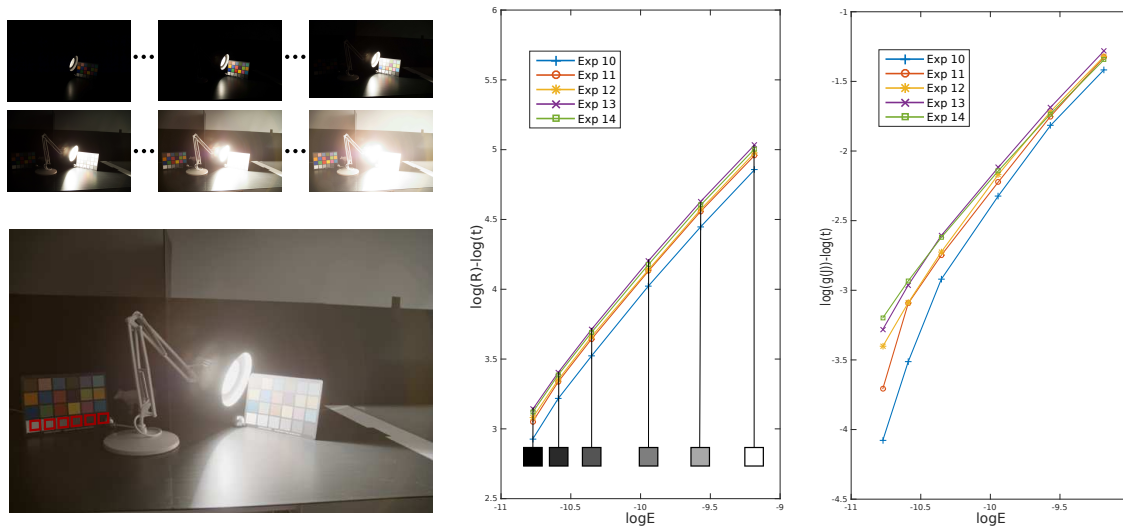


Figure 1. The *LuxoDoubleChecker* scene from [7] (a) on top the LDR image sequence and on the bottom the final HDR image (tone-mapped). Plots computed for points in the gray squares in the dark color checker (red squares) using (b) the RAW values and (c) the linearized JPEG values. While in both cases the plots should theoretically be a single line of slope one, it can be seen that in the linearized JPEG case the points are more dispersed than in the RAW case.

where $[R, G, B]_{in}^t$ is the sensor triplet (usually in 12 or 14 bits); $[R, G, B]_{out}^t$ is the pixel value at the end of the pipeline (in 8 bits per channel); A is a 3×3 matrix that combines the different color channels taking into account white balance, color encoding, color characterization, and a gain value; and γ is a value, typically between $1/1.8$ and $1/3$, performing gamma correction (notice that we omit demosaicking, denoising, compression, etc.; for a complete explanation of these pipeline processes, see [2]).

We claim that, contrary to what has been assumed in the multiexposure literature,

1. the three channels R , G , B are not independent because the matrix A is not diagonal, as it incorporates color processing steps, like color characterization, that involve all channels;
2. the CRF changes from one picture to the next in the multiexposure scenario. The camera automatically modifies the γ value and the A matrix, implying that the non-linear transform f in (1.2) is not constant and therefore that *there is no CRF to speak of because the camera response has changed from image to image.*

The next example illustrates the latter point. Figure 1(a) reproduces a few individual exposures (top) used by M. Fairchild in [7] to create the *LuxoDoubleChecker* HDR image (bottom). The images were captured in RAW format alongside the nonlinearly corrected counterparts. For a stack of N RAW pictures R_i , the image formation model is

$$(2.2) \quad R_i(p) = E(p)\Delta t_i, \quad i = 1, \dots, N,$$

and this equation is valid for the range of luminances for which the sensor operates in the linear range, above the black pedestal and below saturation. This is why, when creating an

HDR image through multiple-exposure combination, professional users prefer to take RAW pictures; in this way, there is no need to estimate and invert the CRF that is applied to the nonlinearly modified pictures stored in 8 bits per channel form; this also involves camera manufacturers. Applying the logarithm to both sides of (2.2) and leaving only the exposure term on the right, we get

$$(2.3) \quad \log\left(\frac{R_i(p)}{\Delta t_i}\right) = \log(E(p)), \quad i = 1, \dots, N;$$

therefore, if we plot $\log(\frac{R_i}{\Delta t_i})$ versus $\log(E)$, we should get a single line of slope one. This is indeed approximately the case, as we can see in Figure 1(b). In principle the same could be said in the nonlinear case when applying the logarithm to (1.3):

$$(2.4) \quad \log\left(\frac{g(J_i(p))}{\Delta t_i}\right) = \log(E(p)), \quad i = 1, \dots, N.$$

This is so because if we plot $\log(\frac{g(J_i)}{\Delta t_i})$ versus $\log(E)$, we should also get a single line of slope one. In practice, though, this does not always happen, as Figure 1(c) shows. The fact that the values for $\log(\frac{g(J_i)}{\Delta t_i})$ are rather spread implies that it was wrong to assume that f (as well as its inverse g) was constant, and therefore the conclusion is that the camera must have modified the values for some of its parameters, A , γ , when the exposure time Δt_i is changed.

To support this claim and highlight how generalized this camera behavior is, we have performed tests on multiple-exposure sequences coming from four different camera models, where during capture, only the exposure time changed, and with results recorded both in linear (RAW) and in nonlinear (JPEG) form. Having the same picture in these two versions allows us to estimate the values for γ and the matrix A with (2.1), using the RAW data for the $[R, G, B]_{in}^t$ values and the JPEG data for the $[R, G, B]_{out}^t$ values.

In Figure 2, columns 1–4 correspond to different sequences (shown in rows 1–4 of Figure 3) taken with different cameras, while the fifth column corresponds to the average over the 105-image HDR Survey database [7]. The first row of Figure 2 plots, for different sequences, a value that measures how far the matrix A of each image in the sequence departs from being diagonal: We have chosen for this to compute the average of the absolute value of the nondiagonal elements of A normalized by its maximum value. The fact that these values are consistently above 0.1 shows that the three channels R , G , B are not independent. The second row of Figure 2 plots the value of $1/\gamma$ for each image in the sequence, which is ordered from shortest to longest exposure time. We can see that, for all sequences, as the exposure time increases, the value of $1/\gamma$ also increases, and the change is quite substantial. The third row of Figure 2 plots the difference between the matrix A of each image in the sequence with respect to the matrix A of the middle image in the stack.¹ Again we see that the cameras are changing A from one exposure to the next.

In Figure 3, each row corresponds to a camera model from a different camera maker. The columns show tone-mapped results of the HDR pictures obtained with different multiexposure combination methods: from the RAW pictures (first column), from the JPEG pictures

¹We compute this difference as the Frobenius norm $\|\cdot\|_F$ of the difference between the matrices.

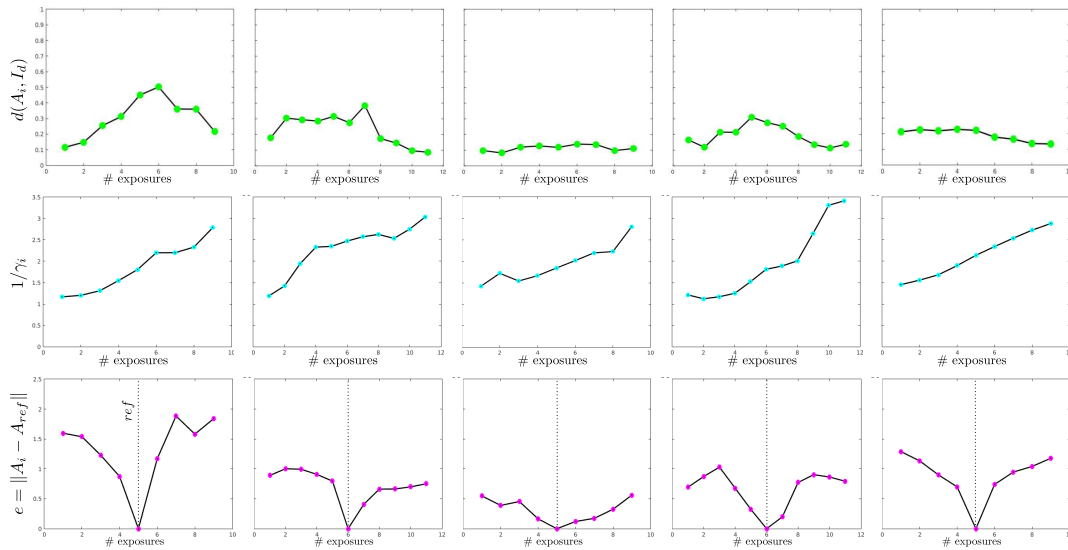


Figure 2. Columns 1–4 correspond to different sequences taken with different cameras (Nikon, Pentax, Panasonic, Canon), shown in rows 1–4 of Figure 3. The last column corresponds to the average over the 105-image HDR Survey database [7]. The x-axes correspond to the number of exposure images (denoted by i), and the y-axes correspond to the corresponding error. The first row shows the distance error between the color correction matrix A_i and a diagonal one I_d , the second row plots the estimated $1/\gamma_i$ for each image, and the last row presents the difference error between the A_{ref} matrix of the middle exposure and the rest of the matrices A_i . Note that the A and γ values correspond to the model in (2.1).

using the multiple-exposure combination methods of Debevec and Malik [5] (column 2), and Mitsunaga and Nayar [28] (column 3), from Lee et al. [18] (column 4), and from the method proposed in this paper (last column). We can see that the previous multiple-exposure methods that take the nonlinear JPEG inputs produce results which have visible problems, like hue shifts and color artifacts; to underline that these artifacts are not due to the particular tone-mapping method used, each row employs a different, state-of-the-art tone-mapping algorithm.

In the next section we will introduce a method that, considering all three channels simultaneously, removes the fluctuations in γ and A , effectively making the CRF constant for the whole sequence. Its results for the above sequences are shown in the last column of Figure 3.

3. Proposed method to make the CRF constant. A schematic of our method is presented in Figure 4. The input is a set of N nonlinear LDR images acquired under different exposure times, and we select P of them as reference images. Our algorithm consists of two steps:

- **Step 1.** For each nonlinear reference image I_{ref_k} , all images in the stack are color-matched and linearized with respect to I_{ref_k} and then averaged to produce an intermediate HDR picture HDR_{ref_k} .
- **Step 2.** The P intermediate HDR images HDR_{ref_k} for $k = 1, \dots, P$ are combined to produce the final HDR output.

We have made the code for our implementation available at http://ip4ec.upf.edu/HDR_code. Let us now see in detail the steps of the proposed method.

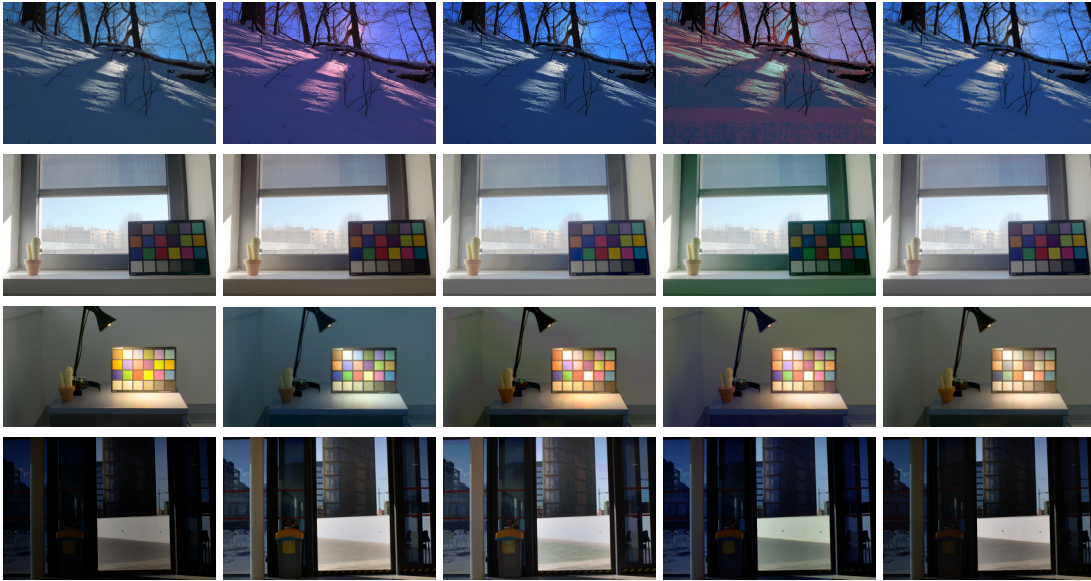


Figure 3. From left to right: tone-mapped HDR result obtained from RAW pictures (i.e., ground truth) (first column) and tone-mapped HDR results obtained with the multiple-exposure combination methods of Debevec and Malik [5] (column 2), Mitsunaga and Nayar [28] (column 3), Lee et al. [18] (column 4), and our proposed approach (last column). Cameras used: Nikon (first row, from the HDR Survey [7]), Pentax (second row), Panasonic (third row), Canon (fourth row). Tone-mapping methods used: Drago et al. [6] (row 1); Mai et al. [22] (row 2); Mantiuk, Daly, and Kerofsky [25] (row 3); and Ferradans et al. [8] (row 4).

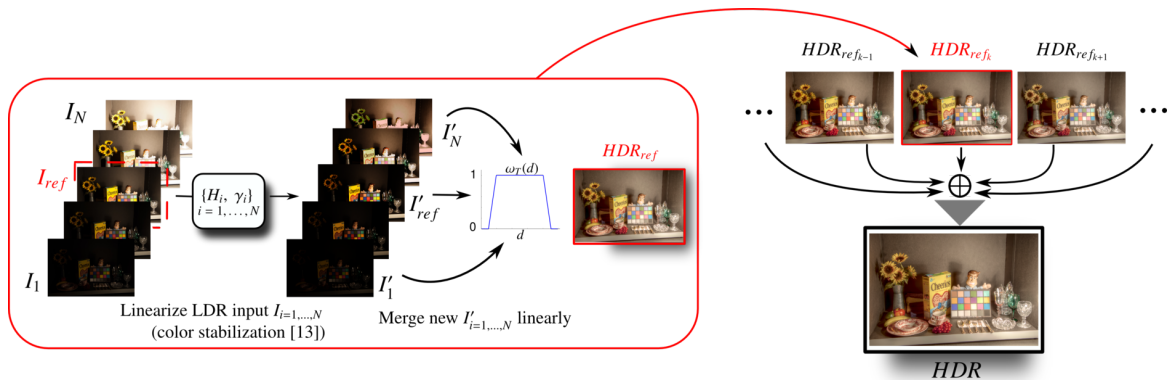


Figure 4. Diagram of our method. The input is a set of N nonlinear images acquired under different exposure times. We select P reference images; for each nonlinear reference image $I_{ref,k}$, all images in the stack are color-matched and linearized with respect to $I_{ref,k}$ and then averaged to produce an intermediate HDR picture $HDR_{ref,k}$. Then these P intermediate HDR pictures are averaged to produce the final output.

3.1. Step 1. Inspired by the digital camera color processing pipeline in (2.1) and the color stabilization model proposed in [36], we consider the color-matching process as key to our method because it is the one that removes fluctuations in γ and A and in practice turns the CRF constant for all images in the multiexposure sequence. The idea from the color stabilization model [36] is to obtain a 3×3 matrix H_{src} and two $\gamma_{ref}, \gamma_{src}$ gamma correction

values such that a source image I_{src} can be color-corrected to match the colors of a given reference I_{ref}

$$(3.1) \quad \left(H_{src} \cdot I_{src}^{1/\gamma_{src}} \right)^{\gamma_{ref}} \sim I_{ref}.$$

Given the reference image I_{ref} , for each other image I_i in the sequence we do the following. We compute a set of correspondences pts_{ref} and pts_i ; we use SIFT [20], although it can be exchanged by any other method.

Then, from the set of correspondences, we build a system of equations,

$$(3.2) \quad \begin{aligned} \left(H_i \cdot I_i(pts_i)^{1/\gamma_{src}} \right)^{\gamma_{ref}} - I_{ref}(pts_{ref}) &= \vec{0}, \\ \left(H_{ref} \cdot I_{ref}(pts_{ref})^{1/\gamma_{ref}} \right)^{\gamma_i} - I_i(pts_i) &= \vec{0}, \end{aligned}$$

where $\{\gamma_i, \gamma_{ref}, H_i, H_{ref}\}$ are the unknowns and $I_i(pts_i)$, $I_{ref}(pts_{ref})$ are the pixel values of corresponding points. Note that we perform a single optimization process, where the only constraint is $H_{ref} \cdot H_i \sim Id$ (the identity). This constraint ensures that the transformation H_i has an inverse, and that is represented by H_{ref} (which corresponds to the matrix that would transfer the colors of the source into the reference). Thus, we define the objective function as

$$E(\mathcal{V}) = \sum_{\substack{pts_i \\ pts_{ref}}} \left\| I_{ref}(pts_{ref}) - g_v^1(I_i(pts_i)) \right\|_2 + \left\| I_i(pts_i) - g_v^2(I_{ref}(pts_{ref})) \right\|_2,$$

where $\mathcal{V} = \{\gamma_{ref}, \gamma_{src}, H_{ref}, H_{src}\}$ is the set of unknowns, $I_i(pts_i)$, $I_{ref}(pts_{ref})$ are the pixel values of corresponding points, and finally the functions $g_v^1(\cdot)$ and $g_v^2(\cdot)$ are defined as

$$(3.3) \quad g_v^1(I_i(pts_i)) = \left(H_i \cdot I_i(pts_i)^{1/\gamma_i} \right)^{\gamma_{ref}} \quad \text{and} \quad g_v^2(I_{ref}(pts_{ref})) = \left(H_{ref} \cdot I_{ref}(pts_{ref})^{1/\gamma_{ref}} \right)^{\gamma_i}.$$

At this point, we want to mention that we have considered different sizes for the matrices H_i and H_{ref} , in particular 3×3 and 4×4 . In section 4.1, we will explain the specific implementation details for each case, and in section 4.4, we will present the results and discussion. Finally, the matrix and nonlinearity $\{\gamma_i, H_i\}$ are applied to the entire image I_i as in (3.1), and we obtain *the linear color-matched image*:

$$(3.4) \quad I'_i = H_i \cdot I_i^{1/\gamma_i}.$$

In Figure 5 we show an example of this procedure.

After we have linearized and color-corrected all images in the sequence, obtaining $I'_{i=1, \dots, N}$, we produce an intermediate HDR result HDR_{ref} by performing a weighted average with a trapezoidal weighting function ω_T , in the range $[0, 1]$, that discards extreme pixel values:

$$(3.5) \quad HDR_{ref} = \frac{\sum_{i=1}^N \omega_T^i I'_i}{\sum_{i=1}^N \omega_T^i}.$$

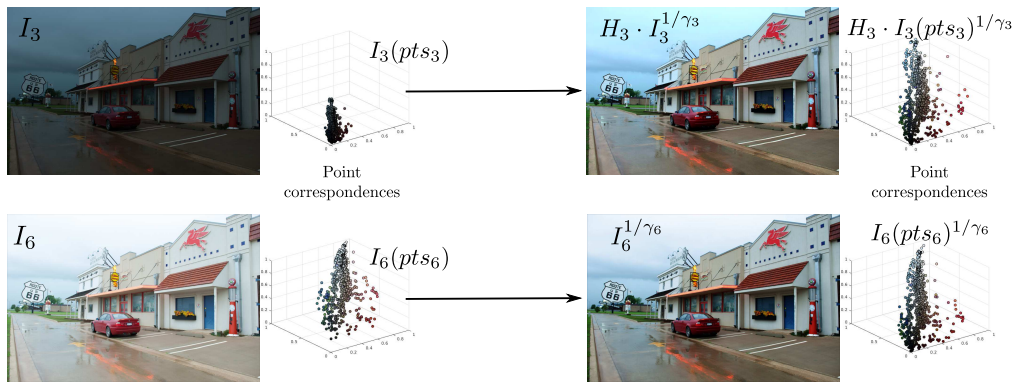


Figure 5. Top row: the image I_3 on the left and the corresponding pixel values $I_3(pts_3)$ with respect to the reference image, and, on the right, its linearized version $H_3 I_3^{1/\gamma_3}$. Bottom row: the reference image I_6 , together with the corresponding pixel values $I_6(pts_6)$ on the left and its linearized version I_6^{1/γ_6} on the right.

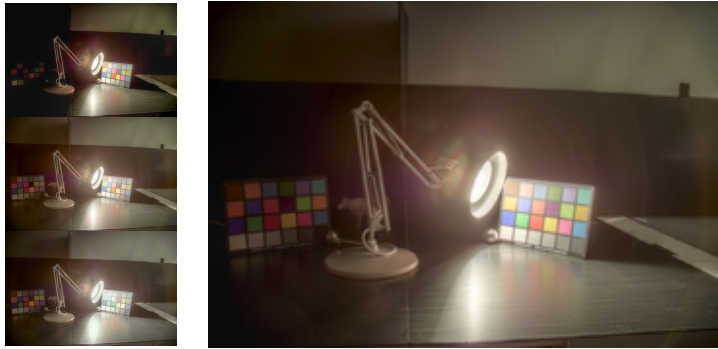


Figure 6. Top row: intermediate HDR images HRD_{ref_6} , HRD_{ref_9} , and $HRD_{ref_{12}}$ computed taking as reference exposures 6, 9, and 12, respectively. Bottom row: final HDR output of proposed method after linear combination of intermediate results. All images have been tone-mapped using [25].

3.2. Step 2. The intermediate HDR results present differences, as Figure 6 shows. In the top row, the leftmost image correctly captures the bright color checker while missing out the details on the dark color checker, and the reverse situation occurs with the middle and rightmost intermediate HDR results. Thus, we propose combining the different $\{HDR_{ref_i}\}_{i=1,\dots,P}$ images to produce the final HDR output. We scale each of them so that they are all in the same range since they have been computed from different reference images captured with different exposures. In order to do this, for each HDR_{ref_i} we compute the trimean, defined as $\frac{1}{4}(Q_1 + 2Q_2 + Q_3)$, where Q_1 , Q_2 , Q_3 are the quartiles; we choose the trimean in order to avoid outliers and take into account the distribution of the image data. Once all trimeans are obtained, we scale the values of each HDR_{ref_i} so that the resulting image has the same trimean as the one of a selected $HDR_{ref_{sel}}$, which in our case has been the middle-exposure reference (image 5 in a 9-image sequence). Finally, we sum the scaled set $\{HDR_{ref_i}\}_{i=1,\dots,P}$ to obtain the final HDR image, as shown in Figure 4. By fusing them, we achieve a final HDR image with more details in both bright and dark areas.

4. Results and comparisons.

4.1. Experiments. As explained above in section 3.1, we linearize the stack of images with respect to a reference one, as expressed in (3.4), by estimating a nonlinearity (power-law function γ) and a linear transformation (matrix H). In this section, we conduct the following experiments regarding the dimension of matrix H : (1) Estimate H as a 3×3 matrix and (2) estimate H as a projective 4×4 matrix.

In the case of considering H a 3×3 matrix, we perform three intermediate HDR images to obtain the final one. We select the middle exposure together with the two around it: $HDR_{ref_{4,5,6}}$. This experiment is our proposed method *Proposed*, and the quantitative and qualitative results are introduced below in section 4.4. In the optimization process in Step 1, we estimate 11 unknowns for each image pair (2 γ values and 9 elements of the matrix).

The choice of a 4×4 matrix will allow the model to be more flexible and to deal with all those pixels that lay close to the border of the color gamut. In this case, we perform as well three intermediate HDR images, but this time we select the middle exposure and the first and last exposures: $HDR_{ref_{1,5,9}}$. We call this experiment *Projective*. In the optimization process in Step 1, we estimate 17 unknowns for each image pair (2 γ values and 15 elements of the matrix). In addition, we consider homogeneous coordinates, by adding 1 to the $[R, G, B]^T$ color vector, for the matrix multiplication. After that, we divide the resulting vector by its last element, and we only keep the first three coordinates of vector.

The way we estimate the linearized images is slightly different than the one presented for the 3×3 case. In order to compute the new linear sequence, we first estimate the set of transformations for each consecutive pair of exposed images: for I_1 and I_2 the set $\{H_1, \gamma_1, \gamma_2\}$, for I_2 and I_3 the set $\{H_2, \gamma_2, \gamma_3\}$, for I_3 and I_4 the set $\{H_3, \gamma_3, \gamma_4\}$, and so on. Finally, the linearization from one I_i image to the given reference would be the composition of all intermediate transformations.

4.2. Database. We performed our experiments using the HDR Survey data set by Mark Fairchild [7]. The online public-domain database contains 105 different scenes acquired using a *Nikon D2x* DSLR camera. The data consist of corresponding JPEG and RAW images for different exposures. In each scene, images in the sequence are numbered going from shortest to longest exposure time. All the scenes except two are composed of 9 images; the other two have, respectively, and 18 images. For the experiments and evaluation we reduce all images by a factor of 1/4, so the image size equals 1072×712 .

4.3. Ground-truth generation. Let us consider N RAW images acquired with different exposure times Δt_i . From the header of the RAW file we read the following parameters: (1) dark and saturation values, which are the minimum and maximum values that the camera produces; (2) a 3×1 array containing the white balance values for each channel; and (3) the CFA Bayer pattern, e.g., “*rggb*.” The ground-truth (GT) construction is defined in two stages. The first one is the merging step, where the N RAW images are combined to obtain a RAW HDR image HDR_L ,

$$(4.1) \quad HDR_L = \frac{\sum_{i=1}^P \omega_i (RAW_i / \Delta t_i)}{\sum_{i=1}^P \omega_i},$$

where ω_i is a weighting function.

We have chosen to use as weighting function t^2 , where t is the exposure time, for its simplicity and good performance, as shown in [30].

The second stage of GT creation converts the obtained HDR_L into a color image by applying a number of steps: First HDR_L is linearly scaled to range $[0, 1]$. Next, white balance is applied. Then demosaicking using the method proposed by Zhang and Xiaolin [38] is used. Finally, the color transformation is done in which each pixel triplet $[R, G, B]^t$ is multiplied by the 3×3 matrix $M_{color} = E \cdot C$, where E is the matrix converting XYZ values into sRGB values and C is the sensor characterization matrix that transforms RGB sensor values into standard XYZ values and is the one described in [7] for the camera used to acquire the data.

4.4. Evaluation. We compare our approach (Proposed), together with the 4×4 experiment (Projective) in section 4.1, against seven multiple-exposure HDR methods. Four of them operate only on static scenes: the classical methods of Debevec and Malik (DM) [5] and Mitsunaga and Nayar (MN) [28] and the recent methods by Lee et al. (Lee13) [18] and Gil, Vazquez-Corral, and Bertalmío [9]. The other three work also on dynamic scenes: Lee, Li, and Monga (Lee14) [17]; Sen et al. (Sen) [32]; and Hu, Pulli, and Sun (Hu) [13] (linearization performed with Debevec and Malik [5]). We use public available codes either from the authors (Sen et al. [32]; Lee, Li, and Monga [17]; and Lee et al. [18]; and Hu, Pulli, and Sun [13]) or by the *HDR Toolbox* from Banterle et al. [1] for Debevec and Malik [5] and Mitsunaga and Nayar [28].

Starting with quantitative, objective evaluation, we compare the HDR outputs of each algorithm (with JPEG sequences as input) versus the computed GT using six standard metrics suggested in Hanhart et al. [12] for this purpose: for luminance, peak signal-to-noise ratio (PSNR), structural similarity metric (SSIM) [37], and HDR quality assessment HDR-VDP-2 [26]; for color, the color version of PSNR (CPSNR), the color extension of SSIM called CID [19], and the color difference measure $CIEDE2000$ (ΔE_{00}^*) [33]. Finally, we also compute the l^2 -norm on RGB space of the difference between a given method and the GT. The results, averaged over the data set,² are presented in Table 1. We can see that our method outperforms the others according to all metrics, except for color metrics CID and ΔE_{00}^* , where Projective is superior, and HDR-VDP-2, where DM [5] performs better. Our algorithm comes second in both cases. Let us note that for HDR-VDP-2 the average is done over the 42 images for which photometer readings exist for the minimum and maximum absolute luminances of the scene, as these values are required by the metric.

To show that the errors due to fluctuations in camera parameters can result in very visible artifacts, in Figure 7, top to bottom, we compare the outputs of DM [5], Lee13 [18], Sen [32], and our approach for the scenes *RITTiger*, *HancockKitchenInside*, *TheNarrows2*, and *MasonLake1*, from left to right. All results have been tone-mapped with the method in [25]. For the scene *RITTiger* we can see that DM [5] (first row) presents a red cast in the image; therefore, Sen [32] (third row) has the same color cast since it uses DM as input. In *TheNarrows2*, Lee13 [18] (second row) shows very noticeable color issues. Finally, for the

²For all metrics apart from HDR-VDP-2 we use all the images except *Zentrum*, since MN [28] is not able to produce a reliable result for this image. Also, the algorithm of Hu, Pulli, and Sun [13] only produces an output for 64 out of the total 105 scenes, so we have chosen to present just a qualitative comparison, in Figure 8.

Table 1

Mean (\bar{x}) and median (\tilde{x}) results from all the presented methods versus GT among the 104 scenes. The last metric, HDR-VDP-2, is computed only among the 42 images containing the values for the minimum and maximum absolute luminances.

		PSNR L	CPSNR	SSIM L	CID	L2	ΔE_{00}^*	HDR-VDP-2
DM	\bar{x}	29.48	28.77	0.890	0.179	0.052	4.41	60.37
	\tilde{x}	27.44	27.29	0.912	0.150	0.032	3.87	61.32
MN	\bar{x}	26.50	25.91	0.863	0.194	0.071	5.32	57.75
	\tilde{x}	25.01	24.79	0.872	0.175	0.071	5.02	57.79
Lee13	\bar{x}	30.91	30.24	0.905	0.179	0.043	4.56	59.48
	\tilde{x}	29.29	28.94	0.929	0.125	0.030	3.37	59.88
Sen	\bar{x}	28.59	27.90	0.860	0.208	0.058	4.89	56.96
	\tilde{x}	27.11	26.79	0.899	0.189	0.034	4.25	57.17
Lee14	\bar{x}	28.66	28.01	0.868	0.169	0.048	4.45	59.00
	\tilde{x}	27.89	27.22	0.927	0.123	0.037	3.72	58.23
Gil	\bar{x}	30.37	29.43	0.899	0.164	0.044	4.74	58.44
	\tilde{x}	29.97	29.21	0.924	0.141	0.032	4.19	60.16
Projective	\bar{x}	32.18	31.62	0.882	0.093	0.041	2.44	59.22
	\tilde{x}	31.40	30.40	0.931	0.064	0.028	1.54	59.50
Proposed	\bar{x}	33.45	32.68	0.930	0.110	0.033	3.39	60.00
	\tilde{x}	34.02	32.28	0.942	0.098	0.025	3.36	60.98

MasonLake1 scene, the method of Lee13 presents a blue cast, while DM presents a reddish cast, and the method of Sen shows a banding artifact effect on the sky region.

To highlight that the visual problems described before are not due to a particular choice of tone-mapping algorithm, Figure 8 shows the same HDR results but tone-mapped with two different methods: [25] for the first two columns and [6] for the last ones. The scenes are *AirBellowsGap* (columns 1 and 3) and *LabWindow* (columns 2 and 4), while the multiple-exposure HDR methods to compare are, from top to bottom; MN [28], Lee14 [17], Hu [13], and our approach. We can see how the previously existing methods produce color artifacts in the sky and sun of the *AirBellowsGap* scene and in the curtains, sky, and background of the *LabWindow* scene, which are apparent for both of the tone-mapping methods used.

4.4.1. Experimental discussion: 3×3 versus 4×4 matrix H_i . Next, we point out interesting outcomes from the comparison of both experiments. Although the 4×4 results were expected to outperform the 3×3 ones, the quantitative results suggest otherwise (see Table 1). The explanation for that relies on the matrix definition. Let us introduce a projective transformation,

$$(4.2) \quad \left(\begin{array}{c|c} A_{3 \times 3} & w \\ \hline \mathbf{v}^T & v \end{array} \right),$$

where $(A_{3 \times 3}|w)$ is an affine transformation and \mathbf{v}^T represents the transformation of the “line at infinity.” Notice that the information about exposure time is carried by matrix $A_{3 \times 3}$. In the context of HDR acquisition, in images with a large difference in exposure time, the offset w might affect dark/bright areas containing relevant data by making them darker/brighter

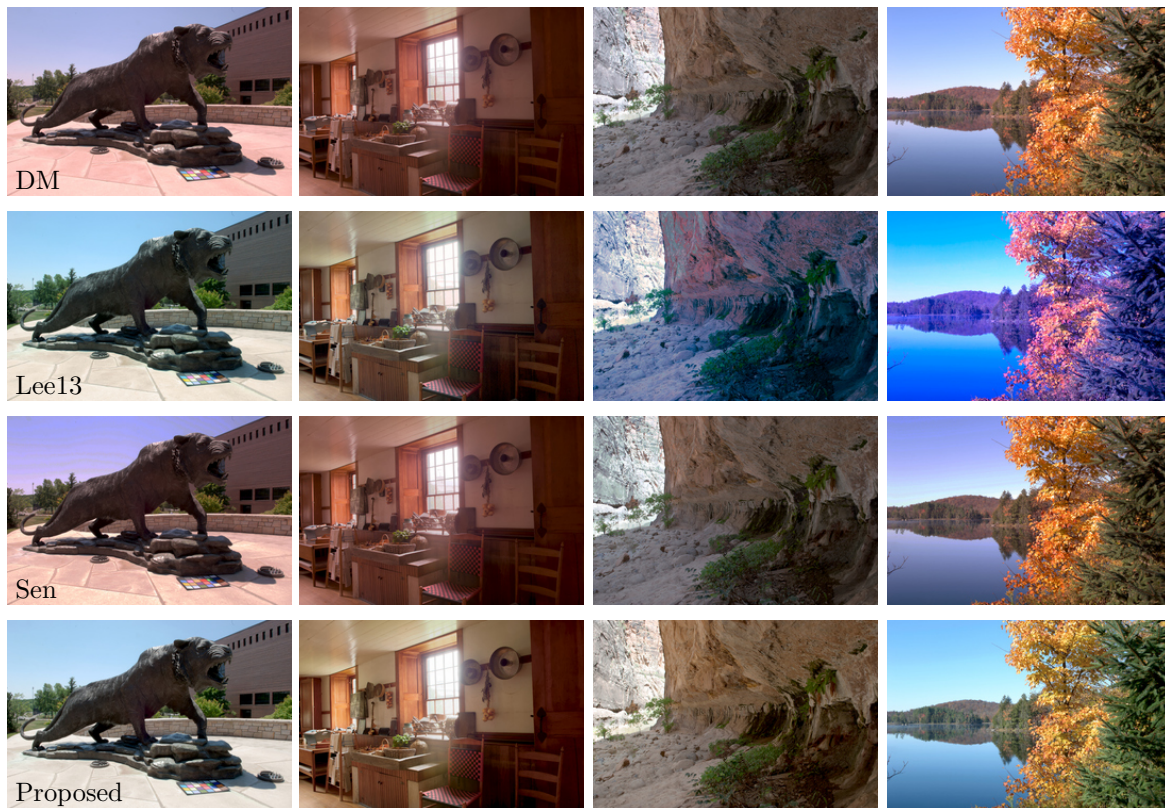


Figure 7. From left to right: scenes *RITTiger*, *HancockKitchenInside*, *TheNarrows2*, and *MasonLake1* from the HDR Survey. From top to bottom: results from DM [5], Lee13 [18], Sen [32], and our approach. All images tone-mapped with [25].

during the optimization process. This fact has a huge impact on the recovery of the final HDR. As an example, let us consider a well-exposed image (as the reference) and an underexposed image (which captures information on bright scene areas). In this case, the estimated 4×4 matrix might brighten all the underexposed regions by selecting a very high value of the offset parameter and not by selecting high values in the first three elements of the diagonal. When this happens, the high values of the w offset clearly compromise the dynamic range of the final HDR by reducing it.

Nevertheless, the Projective choice shows better performance in terms of color. As an example, Figure 9 shows the linearization results of a set of images, considering 3×3 and 4×4 matrices. From the 507 scene, we select two images, I_5 , and I_8 ; see first row. We set image 5 as the reference one. This example helps us to show that the projective transformation (third row) outperforms qualitatively the 3×3 case (second row) in terms of color. Let us focus on the frontal part of the car: On the one hand, the color-matched image I_8 using 3×3 matrix (second row, third column) presents a brighter red. On the other hand, the projective transformation allows recovering the darker red color of the car (third row, third column). The same occurs in the blue digits 507. In the 3×3 case (second row, last column), the blue

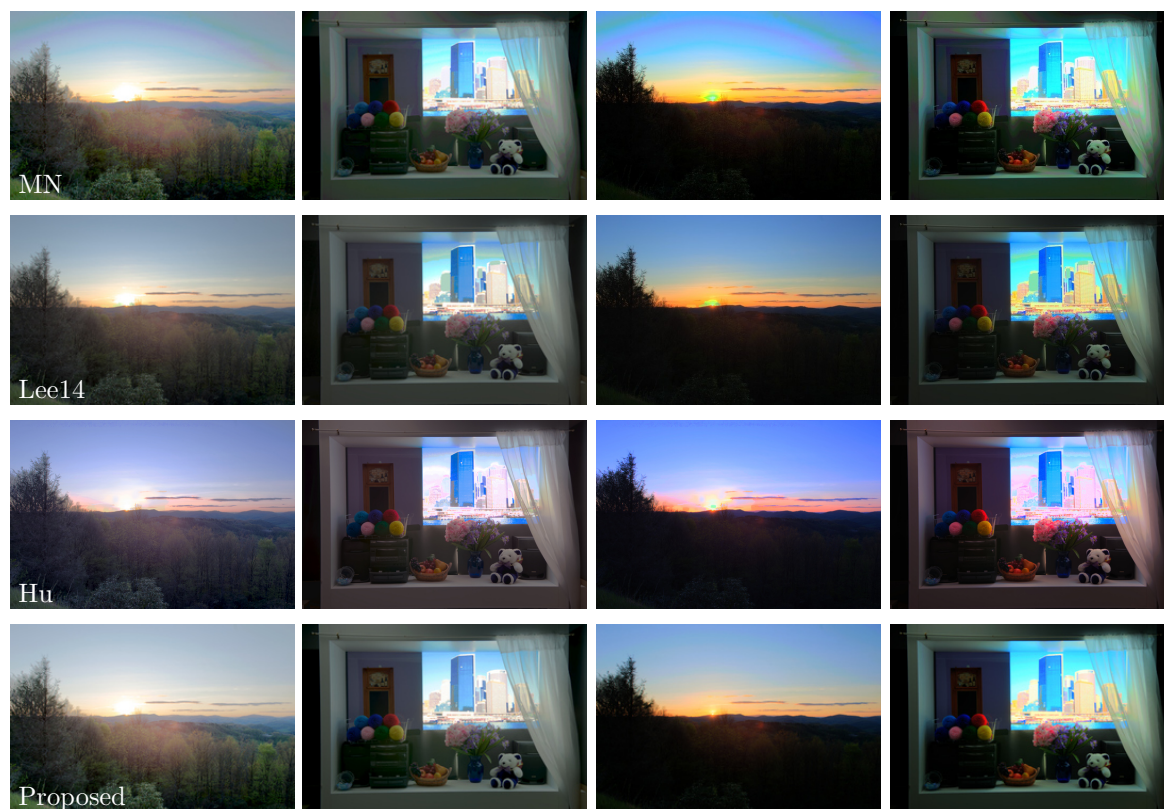


Figure 8. First and third columns: *AirBellowsGap* scene. Second and fourth columns: *LabWindow* scene. Columns 1 and 2: results tone-mapped with [25]. Columns 3 and 4: results tone-mapped with [6]. From top to bottom: results from MN [28], Lee14 [17], Hu [13], and our approach.

color appears brighter than the reference image, whereas in the projective case, the blue is recovered. This explains why the Projective outperforms the Proposed one in terms of the color metrics (CID and ΔE_{00}^*).

4.4.2. Dynamic scenes. It is worth emphasizing that the proposed algorithm does not require image registration, only a set of pixel correspondences. Therefore, it can be used on dynamic scenes as well: In particular, Step 1 of our method can be employed as a pre-processing step to color-stabilize the inputs of HDR methods operating on dynamic scenes, enhancing their performance. To illustrate this, we consider the algorithm of Sen et al. [32], which receives linearized images as input. Consequently, given a stack of nonlinear images, we compare three linearizing approaches: (1) CRF computed by Debevec and Malik [5], (2) radiometric calibration by Lee et al. [18], and (3) applying Step 1 of our method, using as a reference the image in the midpoint of the sequence and finding pixel correspondences with SIFT [20]. We conducted this experiment on a stack of five images from the data set presented in [32], the *Skater* sequence (which comes in JPEG format). In Figure 10 we present the HDR outputs obtained using the three different linearization approaches. The zoomed-in details allow us to see how linearization by [5] (left) produces artifacts in overexposed areas,

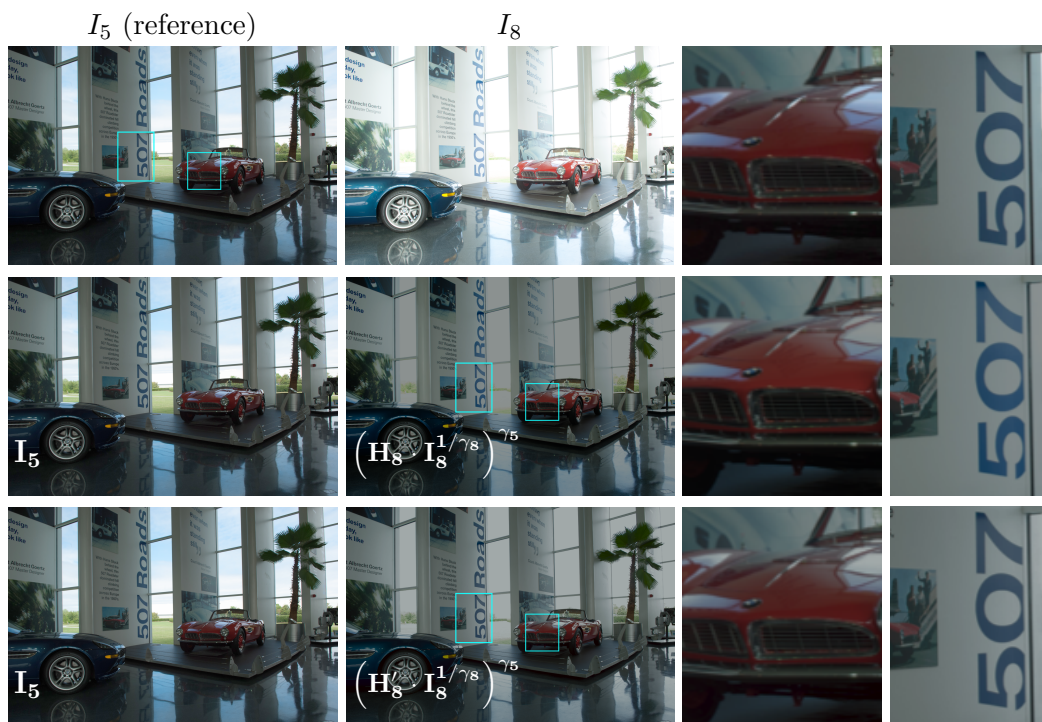


Figure 9. Top row: the LDR images 5 (reference) and 8 from the 507 scene [7]. Middle row: the color-matched results considering a 3×3 matrix (H). Bottom row: the color-matched results considering a 4×4 matrix (H'). The last two columns present the ROI shown as cyan rectangles in each corresponding row. Notice the red color in the front of the car and the blue numbers 507. The 4×4 results are closer to the reference than the 3×3 ones.



Figure 10. HDR results on a dynamic scene applying the HDR creation method of Sen et al. [32] with three different linearization techniques: Debevec and Malik [5] (left), Lee et al. [18] (middle), and step 1 of our proposed method (right), taking as reference the image in the midpoint of the sequence. HDR results tone-mapped with Mantiuk, Daly, and Kerofsky [25].

whereas linearization with [18] produces results that, although free from artifacts, have lower contrast and less saturated colors than what can be obtained with our method.

5. Conclusions. Our experiments show that the camera response function changes with the exposure and depends on the three color channels simultaneously. For this reason, multi-exposure HDR approaches based on estimating and inverting a CRF that is supposed to be

constant may have substantially more error than if computed directly from the linear data, and when tone-mapped, they commonly show hue shifts, color artifacts, or contrast problems. In this work we have proposed a method for removing the fluctuations in the internal settings that the camera has automatically modified, so that our approach effectively makes the CRF constant for the whole sequence. It can be applied to both static and dynamic scenes. Our results are more accurate than those obtained with state-of-the-art methods and show no visual problems.

Acknowledgments. The authors would like to thank Mark Fairchild for his kind help and John McCann for interesting discussions.

REFERENCES

- [1] F. BANTERLE, A. ARTUSI, K. DEBATTISTA, AND A. CHALMERS, *Advanced High Dynamic Range Imaging: Theory and Practice*, A. K. Peters, Ltd., Natick, MA, 2011.
- [2] M. BERTALMIÓ, *Image Processing for Cinema*, Chapman & Hall/CRC Mathematical and Computational Imaging Sciences Series, CRC Press, Taylor & Francis, Boca Raon, FL, 2014.
- [3] S. BIANCO, A. BRUNA, F. NACCARI, AND R. SCETTINI, *Color space transformations for digital photography exploiting information about the illuminant estimation process*, *J. Opt. Soc. Amer.*, 29 (2012), pp. 374–384.
- [4] M. CHANG, H. FENG, Z. XU, AND Q. LI, *Robust ghost-free multiexposure fusion for dynamic scenes*, *J. Electron. Imaging*, 27 (2018), 3.
- [5] P. E. DEBEVEC AND J. MALIK, *Recovering high dynamic range radiance maps from photographs*, in *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH*, New York, 1997, ACM Press/Addison-Wesley, New York, pp. 369–378.
- [6] F. DRAGO, K. MYSZKOWSKI, T. ANNEN, AND N. CHIBA, *Adaptive logarithmic mapping for displaying high contrast scenes*, in *Computer Graphics Forum*, Wiley Online Library, New York, 2003, pp. 419–426.
- [7] M. D. FAIRCHILD, *The HDR photographic survey*, *Color and Imaging Conference*, 2007 (2007), pp. 233–238.
- [8] S. FERRADANS, M. BERTALMIÓ, E. PROVENZI, AND V. CASELLES, *An analysis of visual adaptation and contrast perception for tone mapping*, *IEEE Trans. Pattern Anal. Mach. Intell.*, 33 (2011), pp. 2002–2012.
- [9] R. GIL RODRÍGUEZ, J. VAZQUEZ-CORRAL, AND M. BERTALMIÓ, *The intrinsic error of exposure fusion for HDR imaging, and a way to reduce it*, in *Proceedings of the British Machine Vision Conference (BMVC)*, BMVA Press, Swansea, UK, 2015, 126.
- [10] M. D. GROSSBERG AND S. K. NAYAR, *High dynamic range from multiple images: Which exposures to combine*, in *Proceedings International Conference on Computer Vision Workshop (ICCVW) on Color and Photometric Methods in Computer Vision*, Nice, France, 2003, IEEE, Piscataway, NJ.
- [11] B. GUTHIER, S. KOPF, M. WICHTLHUBER, AND W. EFFELSBERG, *Parallel implementation of a real-time high dynamic range video system*, *Integr. Computer Aided Eng.*, 21 (2014), pp. 189–202.
- [12] P. HANHART, M. BERNARDO, M. PEREIRA, A. PINHEIRO, AND T. EBRAHIMI, *Benchmarking of objective quality metrics for HDR image quality assessment*, *EURASIP J. Image Video Process.*, 2015 (2015).
- [13] J. HU, O. GALLO, K. PULLI, AND X. SUN, *HDR deghosting: How to deal with saturation?*, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Washington, DC, 2013, IEEE, Piscataway, NJ, pp. 1163–1170.
- [14] S. B. KANG, M. UYTENDAELE, S. WINDER, AND R. SZELISKI, *High dynamic range video*, *ACM Trans. Graphics*, 22 (2003), pp. 319–325.
- [15] S. J. KIM AND M. POLLEFEYS, *Robust radiometric calibration and vignetting correction*, *IEEE Trans. Pattern Anal. Mach. Intell.*, 30 (2008), pp. 562–576.

- [16] J. KRONANDER, S. GUSTAVSON, AND J. UNGER, *Real-time HDR video reconstruction for multi-sensor systems*, in ACM SIGGRAPH Posters, ACM, New York, 2012, 65.
- [17] C. LEE, Y. LI, AND V. MONGA, *Ghost-free high dynamic range imaging via rank minimization*, IEEE Signal Process. Lett., 21 (2014), pp. 1045–1049.
- [18] J.-Y. LEE, Y. MATSUSHITA, B. SHI, I. S. KWEON, AND K. IKEUCHI, *Radiometric calibration by rank minimization*, IEEE Trans. Pattern Anal. Mach. Intell., 35 (2013), pp. 144–156.
- [19] I. LISSNER, J. PREISS, P. URBAN, M. S. LICHTENAUER, AND P. ZOLLIKER, *Image-difference prediction: From grayscale to color*, IEEE Trans. Image Process., 22 (2013), pp. 435–446.
- [20] D. G. LOWE, *Object recognition from local scale-invariant features*, in IEEE International Conference on Computer Vision (ICCV), Washington, DC, 1999, IEEE, Piscataway, NJ, Vol. 2, pp. 1150–1157.
- [21] K. MA, H. LI, H. YONG, Z. WANG, D. MENG, AND L. ZHANG, *Robust multi-exposure image fusion: A structural patch decomposition approach*, IEEE Trans. Image Process., 26 (2017), pp. 2519–2532.
- [22] Z. MAI, H. MANSOUR, R. MANTIUK, P. NASIOPOULOS, R. K. WARD, AND W. HEIDRICH, *Optimizing a tone curve for backward-compatible high dynamic range image and video compression*, IEEE Trans. Image Process., 20 (2011), pp. 1558–1571.
- [23] S. MANN, *Comparametric equations with practical applications in quantigraphic image processing*, IEEE Trans. Image Process., 9 (2000), pp. 1389–1406.
- [24] S. MANN AND R. W. PICARD, *On being “undigital” with digital cameras: Extending dynamic range by combining differently exposed pictures*, in Proceedings of IS&T, Springfield, VA, 1995, pp. 442–448.
- [25] R. MANTIUK, S. DALY, AND L. KEROFISKY, *Display adaptive tone mapping*, in ACM Trans. Graphics, Vol. 27, ACM, New York, 2008, 68.
- [26] R. MANTIUK, K. J. KIM, A. G. REMPEL, AND W. HEIDRICH, *HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions*, ACM Trans. Graphics, 30 (2011), 40.
- [27] I. MERIANOS AND N. MITIANOUDIS, *Multiple-exposure image fusion for HDR image synthesis using learned analysis transformations*, J. Imaging, 5 (2019), 32.
- [28] T. MITSUNAGA AND S. NAYAR, *Radiometric self calibration*, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 1, 1999, Ft. Collins, CO, IEEE, Piscataway, NJ, pp. 374–380.
- [29] T.-H. OH, J.-Y. LEE, Y.-W. TAI, AND I. S. KWEON, *Robust high dynamic range imaging by rank minimization*, IEEE Trans. Pattern Anal. Mach. Intell., 37 (2015), pp. 1219–1232.
- [30] M. A. ROBERTSON, S. BORMAN, AND R. L. STEVENSON, *Estimation-theoretic approach to dynamic range enhancement using multiple exposures*, J. Electron. Imaging, 12 (2003), pp. 219–228.
- [31] P. SEN AND C. AGUERREBERE, *Practical high dynamic range imaging of everyday scenes: Photographing the world as we see it with our own eyes*, IEEE Signal Process. Mag., 33 (2016), pp. 36–44.
- [32] P. SEN, N. K. KALANTARI, M. YAESOUBI, S. DARABI, D. B. GOLDMAN, AND E. SHECHTMAN, *Robust patch-based HDR reconstruction of dynamic scenes*, ACM Trans. Graphics, 31 (2012), 203.
- [33] G. SHARMA, W. WU, AND E. N. DALAL, *The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations*, Color Res. Appl., 30 (2005), pp. 21–30.
- [34] M. D. TOCCI, C. KISER, N. TOCCI, AND P. SEN, *A versatile HDR video production system*, ACM Trans. Graphics, 30 (2011), 41.
- [35] Y. TSIN, V. RAMESH, AND T. KANADE, *Statistical calibration of CCD imaging process*, in IEEE International Conference on Computer Vision (ICCV), Vol. 1, Vancouver, Canada, 2001, IEEE, Piscataway, NJ, pp. 480–487.
- [36] J. VAZQUEZ-CORRAL AND M. BERTALMÍO, *Color stabilization along time and across shots of the same scene, for one or several cameras of unknown specifications*, IEEE Trans. Image Process., 23 (2014), pp. 4564–4575.
- [37] 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 (2004), pp. 600–612.
- [38] D. ZHANG AND X. WU, *Color demosaicking via directional linear minimum mean square-error estimation*, IEEE Trans. Image Process., 14 (2005), pp. 2167–2178.