

Robust Radiometric Calibration for Dynamic Scenes in the Wild

Abhishek Badki, Nima Khademi Kalantari and Pradeep Sen
University of California, Santa Barbara

{abhishek_badki, nima}@umail.ucsb.edu, psen@ece.ucsb.edu

Abstract

The camera response function (CRF) that maps linear irradiance to pixel intensities must be known for computational imaging applications that match features in images with different exposures. This function is scene dependent and is difficult to estimate in scenes with significant motion. In this paper, we present a novel algorithm for radiometric calibration from multiple exposure images of a dynamic scene. Our approach is based on two key ideas from the literature: (1) intensity mapping functions which map pixel values in one image to the other without the need for pixel correspondences, and (2) a rank minimization algorithm for radiometric calibration. Although each method has its problems, we show how to combine them in a formulation that leverages their benefits. Our algorithm recovers the CRFs for dynamic scenes better than previous methods, and we show how it can be applied to existing algorithms such as those for high-dynamic range imaging to improve their results.

1. Introduction

Many algorithms in computational photography and computer vision need to combine information from several images of the same scene (e.g., high-dynamic range (HDR) imaging, panorama stitching, 3D structure-from-motion). In some cases the images must be taken at different exposures, so to match features across images these algorithms often assume that pixel intensity values are linearly related to scene radiance. However, this assumption is violated by the non-linear camera response function (CRF) added by camera manufacturers. This CRF is applied to sensor irradiance measurements to compress the dynamic range of the sensor irradiance and make the final images more visually pleasing.

If this non-linearity is not accounted for, errors could occur when matching across images with different exposures. However, the CRFs are generally unknown, considered trade secrets by camera manufacturers. Furthermore, they can be scene dependent [2, 9, 13] making their esti-

mation more difficult. Hence *radiometric calibration*, the process of inverting the CRF to bring values to the linear irradiance domain, is a crucial first step for any algorithm working with images at different exposures.

There have been many different radiometric calibration approaches proposed for the past 20 years, starting with the seminal works of Mann and Picard [19] and Debevec and Malik [4]. However, most of these approaches assume static scenes and tripod-mounted cameras, which is not the typical scenario for casual photographers. Some researchers have proposed methods that attempt to recover the CRF for dynamic scenes, but they fail in the presence of significant scene motion, significant noise, motion blur, and different defocus blur across images. Since highly dynamic scenes are ubiquitous in our world, the problem of automatic radiometric calibration is an open and important problem in computational imaging.

In this paper, we present a new radiometric calibration algorithm specifically designed to handle scenes with significant motion. To do this, we build upon two key ideas from the literature. First, is the concept of intensity mapping functions (IMFs) proposed by Grossberg and Nayar [7], which directly map pixel values in one exposure to another by transforming their histograms. They observed that IMFs can handle small motion because they do not require finding correspondences between images, which can be brittle and error-prone. However, their method for computing IMFs could not handle large scene motions because the histogram of the scene radiance changes. Therefore, we present a new RANSAC-based algorithm for computing IMFs that is more robust to large motions.

Furthermore, Grossberg and Nayar's method used a least-squares optimization which easily overfits to noisy IMF estimates, resulting in incorrect CRFs. To address this problem, we leverage a second key idea: the rank-minimization algorithm for radiometric calibration of Lee et al. [15]. This algorithm avoids overfitting and presents a principled approach for radiometric calibration. However, Lee et al. used pixel-wise correspondences in their optimization which results in artifacts for scenes with significant motion. We address this problem by reformulating the

rank minimization approach to handle IMFs.

By combining the two methods, our approach capitalizes on their benefits. This results in a method that is very stable, avoids overfitting, and is also robust to inaccuracies in IMF estimation. Furthermore, it allows the recovery of response functions up to an exponential ambiguity even when no information is available about the exposure values. We demonstrate that our algorithm can estimate the CRFs from a stack of images of a dynamic scene more accurately than existing methods. Furthermore, we show how it can be used in practice by using it to improve the quality of state-of-the-art HDR reconstruction algorithms.

2. Previous Work

Radiometric calibration is usually performed using multiple aligned images of a static scene taken at known exposures with a fixed camera. The early methods mostly differed in how they represented and solved for the CRF. Mann and Picard [19], for example, used a gamma-correcting function to represent the CRF, while Debevec and Malik [4] used a nonparametric smooth mapping of image irradiance to intensity. Mitsunaga and Nayar [21] solved for the CRF using a high-order polynomial representation, while Grossberg and Nayar [8] leveraged a database of response functions to represent the CRF. All of these approaches assume that the scene is static and the camera is tripod-mounted.

There are a few prior approaches that allow some camera movement or small scene motion. Mann and Mann [20] proposed a method that allows camera rotations and simultaneously estimates the CRF and exposure ratios, but it does not work for general motion. As described earlier, Grossberg and Nayar [7] proposed to estimate the CRF by solving a least-squares optimization after recovering the intensity mapping functions between image pairs using histogram specification. Their IMFs allow for small scene motion, image noise, and blur as long as the histogram of scene radiance does not change much. We extend their approach by proposing a new way to compute IMFs that is more robust to large scene motions.

Kim and Pollefeys’s method [14] allows for free camera movement and some scene motion by computing IMFs using dynamic programming to find maximal values on the joint histogram built from correspondences. However, this approach cannot handle noisy observations that are common in real applications. Methods have also been proposed that estimate the CRF from a single image [17, 18], but they are usually not very robust because they rely strongly on edge information, which can be noisy and is often affected by post-processing in the imaging pipeline.

Finally, Lee et al. [15, 16] introduced a clever radiometric calibration approach that leverages the low-rank structure of sensor irradiances and uses rank minimization to recover the response function. Their basic observation is that

the ideal inverse CRF would map the input images to the image irradiance domain where the values between two images would differ by a single scale factor (the exposure ratio). If these irradiance images are placed into the columns of an observation matrix, this matrix would be rank-one since every successive column is linearly dependent on the first. By solving for the inverse CRF which minimizes the rank of the observation matrix, they avoid the problems with overfitting which affect least-square algorithms. Furthermore, in absence of exposure ratios, they can recover the camera response function up to an exponential ambiguity, which is useful in radiometric alignment applications.

However, Lee et al. use pixel-wise correspondences to find matching intensity values in the different images, which can be prone to error when the scene motion is large or complex. To address this problem, we reformulate the rank-minimization algorithm to leverage IMFs in the optimization directly, allowing it to handle highly dynamic scenes. This results in an algorithm that is more robust to scene motion, camera noise, and image blur than existing methods.

3. Proposed Algorithm

We begin by giving an overview of radiometric calibration algorithms such as ours. Suppose we take two images of a static scene with different exposures by setting the shutter times to t_1 and t_2 . Let z_1 be the image intensity measurement (pixel value) of a point with image irradiance¹ e_1 in the first image, while z_2 and e_2 are the corresponding values of the same pixel in the second image. Assume that all intensity and irradiance values are normalized $[0, 1]$. The exposure ratio between the two images can be expressed as $k_{2,1} = t_1/t_2 = e_1/e_2$. If f is the camera response function that maps linear irradiance to pixel values (e.g., $f(e_1) = z_1$ and $f(e_2) = z_2$), the goal of this work is to find the inverse camera response function $g = f^{-1}$, where in this case $g(z_1) = e_1$ and $g(z_2) = e_2$. Using the fact that $e_1 = k_{2,1} e_2$, we substitute in g to get:

$$g(z_1) = k_{2,1} g(z_2). \quad (1)$$

Let $\tau_{1,2}$ be a function which relates the intensity values of the first image to that of the second, such that $z_2 = \tau_{1,2}(z_1)$. In a general sense, this can be done in many ways such as through pixel correspondences (e.g., optical flow), but in this work we propose to use intensity mapping functions

¹In our paper we call the RAW image value at the photo sensor as the linear irradiance value. Technically, the RAW image value is linearly proportional to the amount of light energy collected at the photo sensor. If we assume the irradiance to be constant over the pixel footprint and shutter integration time then the light energy is linearly proportional to the irradiance at a pixel. Hence, the RAW image value is often simply called as image irradiance value [7, 13, 16, 15]. Although this does not hold when there is a motion blur we stick with this notation in our paper.

(IMFs). Assume for now that the IMF is provided to us, although we show later in Sec. 3.2 how to estimate it from the input images. Substituting this into Eq. 1 gives us:

$$g(z_1) = k_{2,1} g(\tau_{1,2}(z_1)). \quad (2)$$

Let us create a vector $B_1 = [b_1 \ b_2 \ \dots \ b_j \ \dots \ b_K]^T$ that contains the K unique intensity values in the first image such that b_j and $\tau_{1,2}(b_j)$ follow the constraints $0 < b_j < 1$ and $0 < \tau_{1,2}(b_j) < 1$. We could then find g by solving the following least squares optimization, as done in previous approaches [7]:

$$\hat{g} = \arg \min_g \sum_{j=1}^K [g(b_j) - k_{2,1} g(\tau_{1,2}(b_j))]^2. \quad (3)$$

The problem with this least squares optimization is that it can overfit the data and can lead to bad estimates of g , as we will see in Sec. 4. Therefore, we propose instead to use the IMF in a new optimization framework based on the rank minimization work of Lee et al. [15], presented next.

3.1. Our Proposed Rank Minimization Framework

We begin by explaining the basic method of Lee et al. [16, 15]. First, we define a two-column *observation* matrix $D_1 \in \mathbb{R}_{K \times 2}$ as

$$D_1 = [B_1 \ | \ \tau_{1,2} \circ B_1], \quad (4)$$

where \circ is an element wise operator. Note that in Lee et al.'s work the function $\tau_{1,2}$ comes from pixel-wise correspondences, while in ours it is the IMF. We then construct matrix $P = g \circ D_1 = [g \circ B_1 \ | \ g \circ \tau_{1,2} \circ B_1]$, which if g is correct will be a rank-one matrix because by Eq. 2 the second column is a multiple $k_{2,1}$ of the first. Using this property, we can solve for the camera response function g by minimizing the following energy function:

$$\hat{g} = \arg \min_g \left[\text{rank}(g \circ D_1) + \lambda \sum_t H \left(-\frac{\partial g(t)}{\partial B} \right) \right], \quad (5)$$

where $H(\cdot)$ is the Heaviside step function, i.e., $H(x) = 1$ when $x \geq 0$ and $H(x) = 0$ when $x < 0$ and B is a vector that contains all the valid intensity values. Here, the first term minimizes the rank of P while the second term forces the CRF to be monotonically increasing by penalizing solutions with negative gradients.

Eq. 5 estimates the inverse camera response function using a single IMF between a pair of images in a rank minimization framework, but in practice we usually have more than two images with different exposures. Therefore, we need to extend it to include all pairwise IMFs between the adjacent images. This will improve our estimate in two critical ways. First, each IMF contains mapping information

of the intensity values found in only one pair of images, which might not represent the entire range of possible pixel values. By using all the IMFs together we can recover g for a larger range of intensity values. Second, images with different exposures usually have significant overlap in their intensity values, so including them in the observation matrix D_1 improves the robustness of our estimate to noise.

To explain how we do this, we show the process of extending Eq. 5 using three images and generalize it to N images later. This extension can be simply done by including the corresponding two IMFs between the adjacent images in the observation matrix D_1 as follows,

$$D_1 = [\tilde{B}_1 \ | \ \tau_{1,2} \circ \tilde{B}_1 \ | \ \tau_{2,3} \circ \tau_{1,2} \circ \tilde{B}_1],$$

where \tilde{B}_1 is a vector containing a subset of values from vector B_1 , s.t., none of the values in D_1 defined above are saturated, since saturated values break our low-rank assumption because they are no longer linearly related. For the case with N images, we could generate an N -column matrix D_1 where its $j + 1$ th column is computed by applying $\tau_{j,j+1}$ to the j th column.

Using all pairwise intensity mapping functions for constructing observation matrix D_1 could potentially improve the robustness by utilizing the overlap between different images. However, the number of intensity values in vector \tilde{B}_1 that remain non-saturated after successive application of pairwise IMFs will be greatly reduced and could negatively affect the accuracy of the estimation. In the limit, if the range of exposures in the images is great enough, no pixels would exist that would be unsaturated through out the entire range and so the size of \tilde{B}_1 would be effectively zero.

To address this problem we make two key observations. The first is that each specific image has considerable overlap with only a few adjacent images in the exposure stack. Therefore, by using only these neighboring images in the observation matrix we can robustly estimate the camera response function g for the range of intensities covered by that image. Therefore, we propose to divide all the images into overlapping groups of size m (i.e., $m - 1$ pairwise IMFs) and construct an observation matrix D_i for each group. This allows us to use all available IMFs to improve the robustness of our estimation while avoiding the problem of significant reduction of valid intensity values in \tilde{B}_1 . Formally, we create a set of $N - m + 1$ observation matrices as follows,

$$D_i = [d_0 \ | \ d_1 \ | \ \dots \ | \ d_{m-1}],$$

$$d_0 = \tilde{B}_i \quad d_j = \tau_{i+j-1,i+j} \circ d_{j-1}.$$

In our implementation, we select m to be as large as possible such that \tilde{B}_i has at least 20 quantization levels, which is usually around 3 images. Note that this is in contrast to Lee et al.'s method [15], where only a single subset of all the images is used to construct the observation matrix.

Our second key observation is that since the intensity mapping functions are non-linear in nature, uniform quantization of one domain causes highly non-uniform sampling in some parts of the function. To address this discretization problem, we propose to construct an *inverse* observation matrix D'_i corresponding to each D_i as follows,

$$D'_i = [d'_0 \mid d'_1 \mid \cdots \mid d'_{m-1}],$$

$$d'_{m-1} = \tilde{B}_{i+m-1} \quad d'_{j-1} = \tau_{i+j-1, i+j}^{-1} \circ d'_j,$$

where $\tau_{i,i+1}^{-1}$ is the inverse intensity mapping function that maps the intensity values in the $i+1$ th image to the i th image. We combine these two observations into a new energy function which robustly estimates the inverse CRF:

$$\hat{g} = \arg \min_g \sum_{i=1}^{N-m+1} \left(\text{rank}(g \circ D_i) + \text{rank}(g \circ D'_i) \right) + \lambda \sum_t H \left(-\frac{\partial g(t)}{\partial B} \right) \quad (6)$$

In summary, the optimization problem we proposed differs from Lee et al.'s method [15] in two important ways:

- Instead of using pixel correspondences, we propose the use of IMFs to construct the observation matrix D_i . Not only are IMFs more robust to motion than pixel correspondences, but they are also a very compact representation (Lee et al. need thousands of pixel correspondences to construct their observation matrix, whereas the intensity values in our matrices are often less than 256). This makes our optimization step much faster as well.
- Instead of selecting a single subset of images to construct one observation matrix and solving for g to linearize it, we propose to compute multiple D_i and D'_i observation matrices, and solve for g to linearize them all together. As we shall see in Sec. 4, this makes our estimation more robust and improves its performance on average.

3.1.1 Solving the optimization problem

The objective function in Eq. 6 can be solved using standard optimization techniques. We use the Levenberg-Marquardt method in our experiments. While the intensity mapping function estimated by our approach gives a reasonably good approximation, to further improve the robustness and accuracy of our approach we adopted the outliers rejection method proposed by Lee et al. [15]. Please refer to our supplementary material for further details.

Finally, instead of minimizing the rank of a matrix in our optimization problem, we minimize the second condition

number of a matrix, as proposed by Lee et al. [15, 16]. We use a parametric form for g given by n^{th} degree polynomial function as shown below.

$$g(b) = b + b(b-1) \sum_{i=0}^{n-2} c_i b^i$$

The polynomial function defined this way has $n-1$ degrees of freedom and follows the constraints $g(0) = 0$ and $g(1) = 1$. Experimentally, we found $n = 6$ to be a good choice as shown in Sec. 4.3.

3.1.2 Exponential ambiguity

The solution to Eq. 6 recovers the response function up to an exponential ambiguity. This means that if \hat{g} is a solution to Eq. 6 such that $\text{rank}(\hat{g} \circ D_i) = 1$, then \hat{g}^γ is also a solution. For example, if $\hat{g} \circ B_1 = x$ and $\hat{g} \circ \tau_{12} \circ B_1 = k_{2,1}x$, then $\hat{g}^\gamma \circ B_1 = x^\gamma$ and $\hat{g}^\gamma \circ \tau_{12} \circ B_1 = k_{2,1}^\gamma x^\gamma$ are still linearly related, albeit with a different ratio. Therefore, the observation matrix still has rank of one.

This corroborates the observation of Grossberg and Nayar [7], who found it was impossible to recover the exposure ratio and inverse response simultaneously from the IMF without making a prior assumption on either one first. In our case, if the exposure ratios are unknown we cannot use the inverse CRF to recover the actual radiance values, but it is still very useful for radiometric alignment applications like high-dynamic range imaging where we can relax the need to know exposure values (see Sec. 5). Furthermore, if the exposure ratios are available we can resolve the exponential ambiguity and recover the inverse CRF through the following least squares problem:

$$\hat{\gamma} = \arg \min_\gamma \sum_i \sum_j [\hat{g}^\gamma(b_j) - k_{i+1,i} \hat{g}^\gamma(\tau_{i,i+1}(b_j))]^2 \quad (7)$$

In our work, we used Levenberg-Marquardt method to solve this optimization problem. Once we have estimated $\hat{\gamma}$, the inverse camera response function is given by $\hat{g}(\cdot)^{\hat{\gamma}}$.

3.2. Intensity Mapping Functions

We now return to the problem of computing the intensity mapping functions between input images. To do this, there are two general approaches. The first computes the IMF using the joint histogram of two images. These methods first aggregate the intensity values of the corresponding pixels in the two images into a joint image histogram. Then, they estimate the intensity mapping function using regression methods or some kind of fitting (e.g., dynamic programming [14]) over the intensity pairs in the joint histogram. Although these approaches can sometimes handle object motion using an appropriate outliers rejection technique, they are not robust and fail to estimate the intensity mapping function in presence of highly dynamic scene,

noise, motion blur, and different defocus blur across images.

On the other hand, Grossberg and Nayar [7] proposed an alternative approach for estimating IMF using 1D histograms of the two images without the need for pixel correspondences. They showed that if the histograms of scene radiance in two images are similar, the IMF can be computed using the histogram specification between the two 1D histograms. This approach can faithfully recover the IMF in presence of small scene motions, image noise, and blur since they do not change the histogram of scene radiances significantly. However, in case of large motion, the histograms of scene radiance change significantly and hence this method fails to accurately estimate the IMF. Since this approach only uses 1D image histograms detecting the large motion using outliers rejection techniques is difficult.

In order to reliably estimate the IMF in all cases (including large and small motion and in presence of significant image blur and noise) we propose a novel hybrid approach that utilizes these two methods to avoid their problems. In our system, we first detect the large motions in a joint histogram generated from pixel correspondences between two images. Once these pixels are identified, we exclude them and accurately estimate the IMF using histogram specification which is robust to remaining small motions, noise, and blur.

Our proposed approach has the following three steps to estimate an intensity mapping function, $\tau_{i,j}$, between a pair of images:

1) Remove camera motion: We perform the rough registration of the images using a global homography computed with RANSAC [5] from sparse SURF keypoint matches [1].

2) Remove outliers due to large motions: First, we construct a joint histogram $J_{ij}(b_x, b_y)$ for the two globally aligned images using the intensity pairs $(I_i(p), I_j(p))$. We then find the maximum mode of data corresponding to each bin b_x in the joint histogram $J_{ij}(b_x, b_y)$. This is done by performing mean shift clustering [3] over the intensity values of all the pixels in image I_j that contribute to the bin b_x in the joint histogram and selecting the maximum mode. We empirically found that using a fixed kernel bandwidth of 0.15 for mean shift clustering provides the best performance and used it in all of our experiments.

Since large motions are infrequent, most of these modes are expected to lie close to the ground truth IMF and only a few corresponding to the large motions are outliers. Inspired by Hu et al. [10], we detect the inlier modes by fitting cubic Hermite splines in a RANSAC framework. In each RANSAC iteration, we select the control points for the Hermite splines and find the inlier percentage. We select one to four control points depending on the range of bins spanned by the valid mode points in the joint histogram. In addition to this we select two control points at $(0, 0)$ and

$(1, 1)$. In each RANSAC iteration, we select control points such that they are monotonic and well separated from each other. If these conditions are satisfied we fit cubic Hermite splines and calculate the inlier modes percentage. Once we have fit a model using cubic Hermite splines, we find all the pixel correspondences in image I_i and I_j whose intensity pair $(I_i(p), I_j(p))$ lie inside a fixed threshold δ (0.05 in our implementation) from the model. Please refer to [10] for more details on this. The cubic Hermite splines fit is used just to remove the pixel correspondences, whose intensity pair could be large outliers to the ground truth IMF. We remove these outlier pixel correspondences from the images and proceed to next step.

3) Compute the final IMF: The joint histogram computed using these inlier pixel correspondences can still be very noisy due image noise, blur, small mismatches, etc. Hence, we use histogram specification, as proposed by Grossberg and Nayar [7], to compute the final IMF.

We have tested the above approach to compute the IMF on many image pairs and found it to be very robust even in complex scenarios. We compute the intensity mapping functions between adjacent image pairs and use them in our optimization framework (Sec. 3.1) to recover the inverse camera response function.

4. Experiments

We implemented our algorithm in MATLAB and show the performance of our approach through extensive experiments on both synthetic and real world images. We compare against the radiometric calibration algorithms by Lee et al. [15], Grossberg and Nayar [7], and Mitsunaga and Nayar [21]. We use the Lee et al.'s [15] implementation for both their method and Mitsunaga and Nayar's approach, and use our own implementation of the method of Grossberg and Nayar. Since Lee et al.'s and Mitsunaga and Nayar's methods work with pixel correspondences, we compute the inverse camera response function using 1,000 randomly selected samples. We repeat this process five times and use the median of the five computed CRFs in our comparisons.

Although the focus of our approach is handling large motions, we first evaluate our performance in handling image noise on the synthetic dataset by Lee et al. [16]. We demonstrate similar performance in comparison with the state-of-the-art approach by Lee et al. [15]. We then show that our method produces significantly better results than the other approaches on real world images with large motions. Finally, we evaluate the importance of different components in our method through extensive analysis.

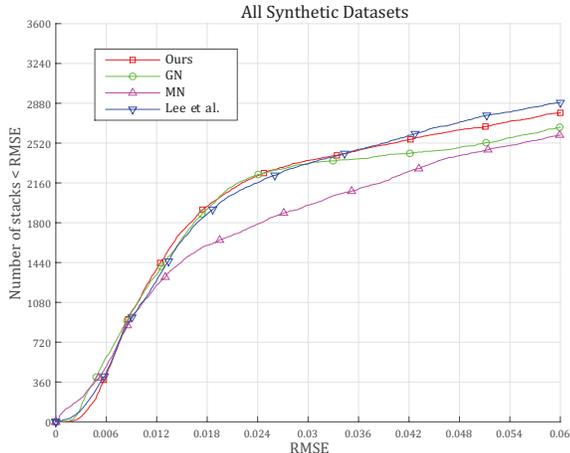


Figure 1. Cumulative histogram of number of successful cases w.r.t. RMSE for the entire synthetic dataset. Larger values shows better performance. Our method has similar performance to the Lee et al.’s method [15] in handling image noise, despite the fact that the focus of the paper is handling large motions.

4.1. Simulations on synthetic dataset

We synthetically generated scene radiances in the range of $[0, 1]$ with four different distributions (see supplementary material), as proposed by Lee et al. [16]. For each distribution, we produced five one dimensional irradiance images of size 100,000 with exposure times of step 0.5 (0.0625, 0.125, 0.25, 0.5, 1). We then added Gaussian noise with five different standard deviations (0, 0.0025, 0.0050, 0.0075, 0.0100) to each images resulting in $4 \times 5 = 20$ irradiance image stacks. Finally, we applied 201 camera curves from the DoRF database [8] and quantized the images to 256 levels to produce $20 \times 201 = 4020$ synthetically generated multiple exposure image stacks.

We demonstrate the overall performance of our approach in comparison against other methods in Fig. 1. We calculate the root mean squared error (RMSE) between the estimated and ground-truth CRFs and for each method we plot the number of image stacks with RMSE less than a specific value. Although the main advantage of our approach is in handling highly dynamic scenes, we demonstrate similar performance in comparison to the state-of-the-art algorithm of Lee et al. [15] on this synthetic dataset with only image noise. Next, we show the performance of our approach on casually captured images of real world scenes.

4.2. Simulations on real dataset

We use 20 RAW multiple exposure image sets from Sen et al. [23]. These images, shown in the supplementary material, cover a variety of different cases (indoor, outdoor, dynamic, and static) and have been taken by hand-held cameras. Moreover, in most cases both the aperture size and the exposure time are varied resulting in having different

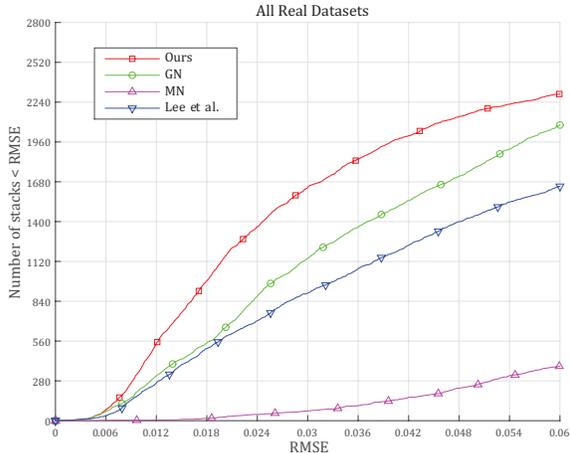


Figure 2. Cumulative histogram of number of successful cases w.r.t. RMSE for the entire real dataset. Our method shows significant improvement over the previous approaches.

defocus blur in the images of an stack. We demosaic the RAW images and apply 201 CRFs from the DoRF database to generate a set of $20 \times 201 = 4020$ multiple exposure image stacks. Note that the radiometric calibration needs to be performed separately for each color channel. To show our results we arbitrarily chose to perform the radiometric calibration for green channel only. To have a fair comparison, we remove the global camera motion in each image stack by homography and use the aligned images as input to all the other approaches.

Fig. 2 shows that our method has superior performance in comparison with the other radiometric calibration approaches. To evaluate the effect of scene properties on the performances, we show comparison on four individual image stacks in Fig. 3. Mitsunaga and Nayar’s approach is designed to work on static images, and thus, performs poorly in all the cases. Note that although the image stack in Fig. 3 (a) is almost static, their method performs poorly because of the images having different defocus blur. Lee et al. use pixel correspondences, and thus, fail to robustly estimate the CRF for the image stack shown in Fig. 3(b) with drastic change in defocus blur. However, since the histogram of scene radiance in different images is fairly similar, the Grossberg and Nayar’s approach performs well. On the other hand, the image stack in Fig. 3 (c) contains large motion and their method fails to robustly estimate the CRF because of the violation of the histogram similarity assumption. Nevertheless, Lee et al.’s approach is able to handle large motion through outliers rejection and performs well in this case. Finally, all the other methods perform poorly in Fig. 3 (d) because of drastic change in defocus blur and large motion. Our approach consistently produces better results in all the cases.

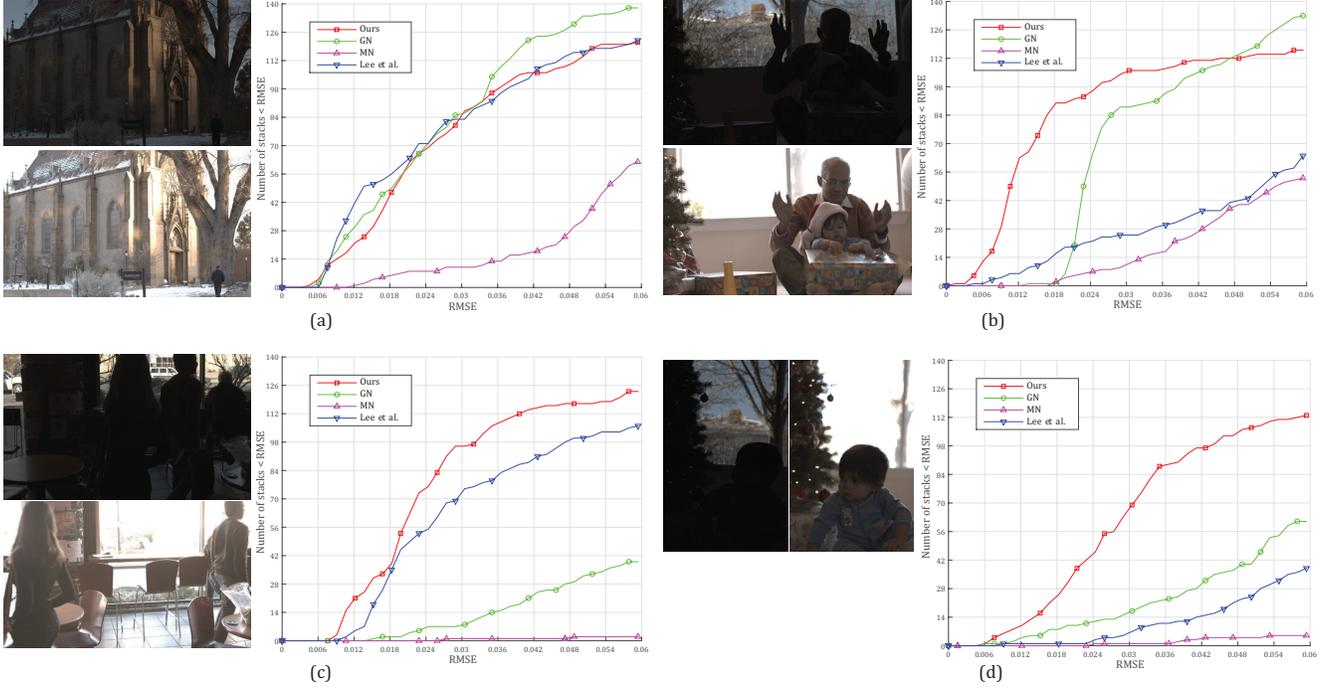


Figure 3. Cumulative histogram of the number of successful cases w.r.t. RMSE for four real exposure stacks from [23]. Only two images for each set are shown for compactness (see supplemental for full sets), and a gamma curve is applied to them for display. Our method performed consistently well in all cases we tested, often much better than existing approaches. Note that scene (a) is almost entirely static so several methods perform well.

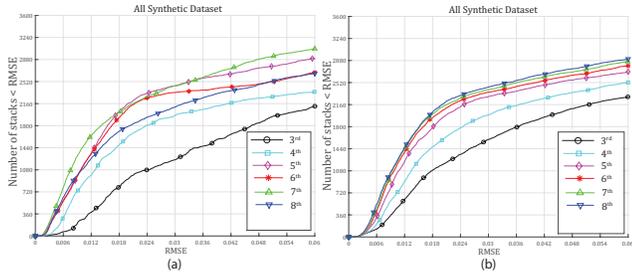


Figure 4. Effect of polynomial order on optimization method. (a) Using our IMFs in least squares framework. (b) Proposed method. Please see the text for description.

4.3. Analysis

Effect of polynomial degree: As explained in Sec. 3.1, we use an n^{th} degree polynomial function to model the CRF. Increasing the polynomial degree n would increase the flexibility of the method to model more complicated CRFs, but increases the chance of overfitting to noise. We evaluate the effect of polynomial degree in our rank minimization framework and compare it to the least squares framework, as proposed by Grossberg and Nayar [7], in Fig. 4. As can be seen, there is a significant drop in performance of the least squares framework from 7 to 8 which shows this approach is prone to overfitting. On the other hand, our rank minimization framework consistently produces better results as

the polynomial degree increases.

Effect of using forward and inverse IMFs: As explained in Eq. 6, we handle the discretization problem using two observation matrices D_i and D'_i computed from forward and inverse intensity mapping functions. We evaluate the effect of each term by comparing the result of our approach using forward, inverse, and both of them. As can be seen in Fig. 5 (a), using both forward and inverse intensity mapping function improves the RMSE performance on an average for the synthetic dataset.

Effect of combining observation matrices: As explained in Sec. 3.1, we divide the N images into overlapping groups of m images. We then compute an observation matrix, D_i , for each group and use all of them in a single energy function. We now compare this approach with an approach where we use just one observation matrix D_i and the corresponding inverse observation matrix D'_i . We perform this analysis on real image dataset. We came up with several strategies to select one observation matrix out of all the valid ones. The RMSE performance of each of this strategy is compared with our approach and is shown in Fig. 5. The different strategies which we used involved: 1) always using the first observation matrix, i.e., using images with lower exposure values, 2) always using last observation matrix, i.e., using images with higher exposure values, 3) always using center observation matrix. On an

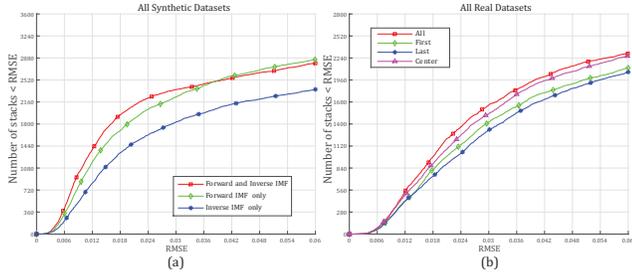


Figure 5. We analyze the effect of different components of our optimization problem on the overall RMSE performance. **(a)** Effect of using forward and inverse IMFs. **(b)** Effect of combining observation matrices. Please see the text for description.

average, our approach gave the best RMSE performance as compared to any of the above discussed strategies. Our approach tries to utilize all the available N images in the multiple exposure image stack by estimating g which linearizes each of the valid observation matrices and hence is more robust as compared to selecting an observation matrix constructed using pairwise IMFs computed on a subset of N images.

5. Application in HDR imaging

We demonstrate the application of our method in high-dynamic range (HDR) imaging by using it as a preprocessing step for HDR reconstruction algorithms of Sen et al. [23] and Oh et al. [22]. These approaches take multiple low-dynamic range (LDR) images at different exposures as input and generate an HDR image, but assume the LDR images are linearized. Therefore, if the input LDR images are in jpeg or other non-linear formats, the CRF need to be estimated by a radiometric calibration approach and be used to linearize the images. Using our method for radiometric calibration enables such HDR reconstruction methods to be applied to any non-linear image set. Working directly with such images eliminates the need for RAW images which usually take a lot of memory (around 100MB for the image stack) and may not be available for all the commercial cameras. This is even more necessary for HDR video algorithms like Kalantari et al. [12] which need RAW video frames, requiring extremely huge memory even for a short video. However, using our method as a preprocessing step, such HDR image and video reconstruction methods can be more generally applied in practice.

As described in Eq. 7, to resolve the exponential ambiguity, exposure ratios should be known. However, in some cases this information is not available and it is important to estimate the exposure ratios directly from the input LDR images. This can be done by modifying Eq. 7 to have a con-

straint on the inverse CRF for each color channel as follows:

$$\hat{\phi} = \arg \min_{\phi} \sum_i \sum_j \sum_l [\hat{g}_l^{\gamma_l}(b_j) - k_{i+1,i} \hat{g}_l^{\gamma_l}(\tau_{l,i,i+1}(b_j))]^2$$

$$s.t. \hat{g}_1^{\gamma_1}(\alpha) = \beta, \hat{g}_2^{\gamma_2}(\alpha) = \beta, \hat{g}_3^{\gamma_3}(\alpha) = \beta \quad (8)$$

where $\phi = [k_{2,1} \ k_{3,2} \ \dots \ k_{N,N-1} \ \gamma_1 \ \gamma_2 \ \gamma_3]$ is an unknown vector and index l represents different color channels. The estimated inverse camera response function for each color channel, $\hat{g}_l^{\gamma_l}(\cdot)$, satisfies the above constraint and differs from the true inverse camera response function by the exponential ambiguity. $[\hat{k}_{2,1} \ \hat{k}_{3,2} \ \dots \ \hat{k}_{N,N-1}]$ are the estimated pseudo exposure ratios. The constraints shown in Eq. 8 are arbitrary and can be set according to the application. In our experiments we set $\alpha = 0.5$ and $\beta = 0.2$. Given non-linear input LDR images, we apply the inverse camera response functions solved using Eq. 6 and 8 to recover linear LDR images and pseudo exposure ratios. We can then use the linearized LDR images and pseudo exposure ratios in an existing HDR reconstruction algorithm.

We show the result of HDR reconstruction methods of Sen et al. [23] and Oh et al. [22] using different radiometric calibration methods as a preprocessing step in Fig. 6 and 7, respectively. Since the HDR reconstruction method of Hu et al. [11] does not require radiometric calibration, we also show their results. Note that since Lee et al. and our approaches are based on rank minimization, we use the above method to estimate the pseudo exposures, but provide the ground-truth exposures for Grossberg and Nayar [7]. As seen, using our method to perform radiometric calibration in the preprocessing step results in artifact-free HDR images in all cases.

We have tested the above approach for many non-linear LDR image sets and were able to achieve a good estimation of inverse camera response function and the pseudo exposure ratios, and therefore, achieve artifact-free reconstruction of HDR images. Hence, using our method to estimate the inverse camera response function, the HDR reconstruction methods, that assume the input LDR images to be linear in nature, can in general be applied to any non-linear input LDR image set, thus improving their applicability.

6. Conclusion

We have presented a new radiometric calibration approach that could robustly recover inverse camera response function from multiple exposure images of a dynamic scene. Our proposed intensity mapping function estimation method is robust to large scene motions and other noisy observations. We have proposed a new optimization method, that uses intensity mapping functions in a rank minimization framework and solves for the inverse camera response function. We showed the superior performance of

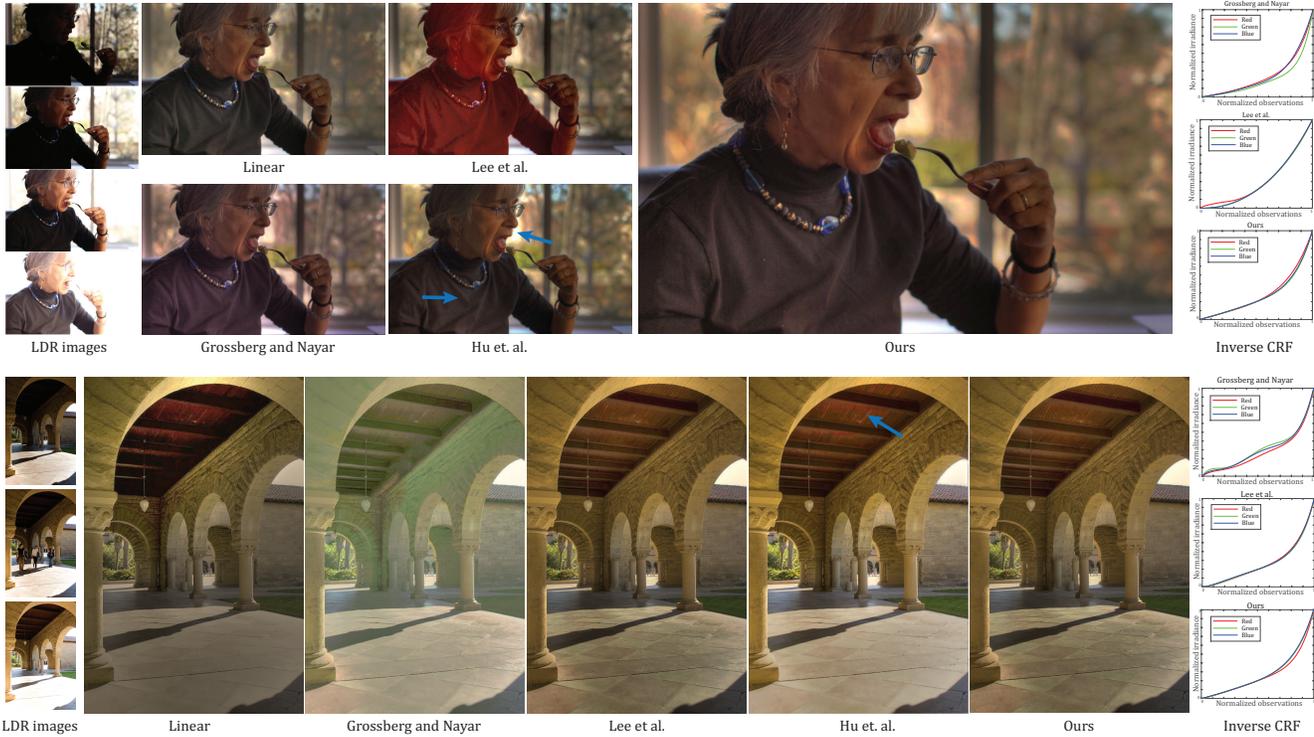


Figure 6. We used different radiometric calibration methods to linearize the jpeg LDR images, shown on left. We then use the linearized images as input to the HDR reconstruction method of Sen et al. [23] to produce the final HDR images. We also include the result of Hu et al.’s HDR reconstruction method [11] which does not require radiometric calibration. Note that all the results are tone mapped with the same setting. “Linear” implies that a linear inverse camera curve was assumed for radiometric calibration, i.e., using the jpeg LDR images directly in Sen et al.’s approach. Since the CRFs are non-linear in both examples, assuming linear CRF introduces artifacts in the results. Grossberg and Nayar’s method can handle small scene motion and performs well on the top example, however, it fails on bottom example due to large motions which cause significant change in the scene across multiple exposure images. On the other hand, Lee et al.’s algorithm is able to reject the outliers due to large motions and successfully handles the bottom example, but fails to provide a good calibration in the top example due to drastic different defocus blur and scene motion. Hu et al.’s HDR reconstruction method introduces artifacts in both the examples (See absence of wrinkles on clothes in the first example and color artifacts in the second example). Our method can handle noisy observations as well as large motions, and thus, using our radiometric calibration method as a preprocessing step results in artifact-free HDR images. The LDR images are obtained from [23] (top) and [6] (bottom).

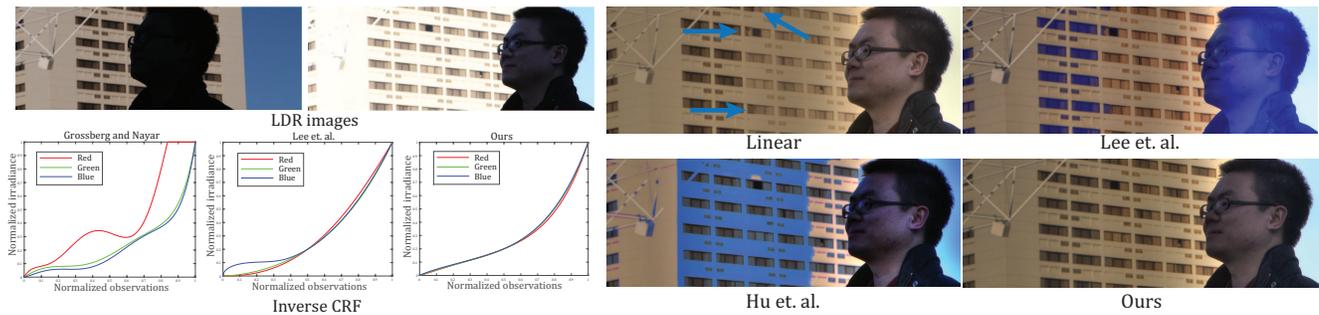


Figure 7. We used different radiometric calibration methods to linearize the jpeg LDR images and then use the linearized images as input to the HDR reconstruction method of Oh et al. [22]. The LDR image stack is obtained from [23]. All the competing radiometric calibration approaches perform poorly on this challenging scene resulting in poor HDR reconstruction results by the method of Oh et al. Moreover, the HDR reconstruction method of Hu et al. cannot handle the scene because of large motions. Note that we were not able to produce HDR result using LDR images linearized by Grossberg and Nayar’s approach since their estimated CRF is not monotonically increasing. Our method is robust to large scene motions and when used as a preprocessing step results in artifact-free HDR image by method of Oh et al.

our method as compared to other radiometric calibration approaches by conducting extensive experiments on synthetic as well as real datasets. Finally, we showed that how using our approach as a preprocessing step improved the quality of some of the state of art algorithms for high dynamic range imaging.

7. Acknowledgments

The authors would like to thank the reviewers for their insightful suggestions. This work was funded by National Science Foundation grants IIS-1321168 and IIS-1342931.

References

- [1] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 110(3):346–359, June 2008.
- [2] A. Chakrabarti, D. Scharstein, and T. Zickler. An empirical camera model for internet color vision. In *Proceedings of the British Machine Vision Conference*, pages 51.1–51.11. BMVA Press, 2009. doi:10.5244/C.23.51.
- [3] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):603–619, May 2002.
- [4] 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 '97*, pages 369–378, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.
- [5] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981.
- [6] O. Gallo, N. Gelfand, W.-C. Chen, M. Tico, and K. Pulli. Artifact-free high dynamic range imaging. In *Computational Photography (ICCP), 2009 IEEE International Conference on*, pages 1–7, April 2009.
- [7] M. Grossberg and S. Nayar. Determining the camera response from images: what is knowable? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(11):1455–1467, Nov 2003.
- [8] M. Grossberg and S. Nayar. Modeling the space of camera response functions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(10):1272–1282, Oct 2004.
- [9] J. Holm. Pictorial digital image processing incorporating adjustments to compensate for dynamic range differences, Sept. 30 2003. US Patent 6,628,823.
- [10] J. Hu, O. Gallo, and K. Pulli. Exposure stacks of live scenes with hand-held cameras. In *Proceedings of the 12th European Conference on Computer Vision - Volume Part I, ECCV'12*, pages 499–512, Berlin, Heidelberg, 2012. Springer-Verlag.
- [11] J. Hu, O. Gallo, K. Pulli, and X. Sun. Hdr deghosting: How to deal with saturation? In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1163–1170, June 2013.
- [12] N. K. Kalantari, E. Shechtman, C. Barnes, S. Darabi, D. B. Goldman, and P. Sen. Patch-based high dynamic range video. *ACM Trans. Graph.*, 32(6):202:1–202:8, Nov. 2013.
- [13] S. J. Kim, H. T. Lin, Z. Lu, S. Süsstrunk, S. Lin, and M. Brown. A new in-camera imaging model for color computer vision and its application. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(12):2289–2302, Dec 2012.
- [14] S. J. Kim and M. Pollefeys. Robust radiometric calibration and vignetting correction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(4):562–576, April 2008.
- [15] J.-Y. Lee, Y. Matsushita, B. Shi, I. S. Kweon, and K. Ikeuchi. Radiometric calibration by rank minimization. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(1):144–156, Jan 2013.
- [16] J.-Y. Lee, B. Shi, Y. Matsushita, I.-S. Kweon, and K. Ikeuchi. Radiometric calibration by transform invariant low-rank structure. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2337–2344, June 2011.
- [17] S. Lin, J. Gu, S. Yamazaki, and H.-Y. Shum. Radiometric calibration from a single image. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–938–II–945 Vol.2, June 2004.
- [18] S. Lin and L. Zhang. Determining the radiometric response function from a single grayscale image. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 66–73 vol. 2, June 2005.
- [19] Mann, Picard, S. Mann, and R. W. Picard. On being ‘undigital’ with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proceedings of IST*, pages 442–448, 1995.
- [20] S. Mann and R. Mann. Quantigraphic imaging: Estimating the camera response and exposures from differently exposed images. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–842–I–849 vol.1, 2001.
- [21] T. Mitsunaga and S. Nayar. Radiometric self calibration. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*, volume 1, pages –380 Vol. 1, 1999.
- [22] T. Oh, J. Lee, Y. Tai, and I. Kweon. Robust high dynamic range imaging by rank minimization. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PP(99):1–1, 2014.
- [23] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman. Robust Patch-Based HDR Reconstruction of Dynamic Scenes. *ACM Transactions on Graphics (TOG) (Proceedings of SIGGRAPH Asia 2012)*, 31(6):203:1–203:11, 2012.