Radiometric Calibration for Internet Photo Collections

Zhipeng Mo[1]   Boxin Shi[2*]   Sai-Kit Yeung[1]   Yasuyuki Matsushita[2,3]

[1]Singapore University of Technology and Design

[2]Artificial Intelligence Research Center, National Institute of AIST

[3]Osaka University

Abstract

Radiometrically calibrating the images from Internet photo collections brings photometric analysis from lab data to big image data in the wild, but conventional calibration methods cannot be directly applied to such image data. This paper presents a method to jointly perform radiometric calibration for a set of images in an Internet photo collection. By incorporating the consistency of scene reflectance for corresponding pixels in multiple images, the proposed method estimates radiometric response functions of all the images using a rank minimization framework. Our calibration aligns all response functions in an image set up to the same exponential ambiguity in a robust manner. Quantitative results using both synthetic and real data show the effectiveness of the proposed method.

1. Introduction

For a popular landmark, millions of pictures are captured and shared through the Internet. Such Internet images and community photo collections provide comprehensive image resources to computer vision research, because they contain images captured from different viewpoints, at different time, under different illumination, and using different types of cameras and settings. The cameras capturing these images naturally organize a "network". By exploring such a network, geometric analyses such as geometric camera calibration and 3D reconstruction, which are generally infeasible using a single image, become tractable by establishing correspondences for recovering 3D points using multi-view observations. Recent progress on structure from motion (SfM) [32] and multi-view stereo (MVS) [9, 10] show successful applications using Internet photos.

Photometric analysis is an important complement problem for analyzing images from Internet photo collections organized by such a camera network, and radiometric calibration is a key prerequisite for photometric analysis. Many

computer vision problems, when they are to be applied to Internet photos, *e.g.*, intrinsic image decomposition [18] and photometric stereo [29], require input images to be radiometrically linearized. However, a commercial camera usually maps the scene radiance to its pixel values in a nonlinear manner for compressing the dynamic range and aesthetic purpose. Such a mapping is unknown in most cases and treated as business secrets by camera manufactures. The goal of radiometric calibration is to estimate the camera's (inverse) radiometric response function so that the observed pixel values linearly relate the scene radiance.

Classic radiometric calibration approaches capture a static scene under various exposure times [25]. For outdoor scene with illumination changes, given an image sequence from a fixed viewpoint, the problem could also be solved by physically modeling the image formation and using the consistent scene albedo across images as the key constraint [15]. Radiometrically calibrating Internet photos shares a similar spirit to [15], but it is much more challenging due to that 1) the scenes are captured from multiple diverse viewpoints, 2) each image is captured by an unknown camera with unknown settings (*e.g.*, white balance and exposure time), and 3) a batch of different response functions need to be estimated simultaneously.

In this paper, we propose a new method to jointly perform radiometric calibration to all cameras using a collection of Internet photos for the same scene. We utilize the geometric information calculated from SfM and MVS, and assume the scene reflectance (albedo) is the same for corresponding pixels in all images. Our key observation is that the ratio of albedo values for pixels in the same image with the same surface normal should be equal across the image set, only when each image is applied with a correctly calibrated inverse radiometric response function. Such a ratio operation cancels the influence of different white balance settings, exposure times, and environment illumination conditions which are usually arbitrary for each image (and camera). The albedo ratios are stacked as a matrix whose rank should be 1. We minimize the rank of such a matrix to achieve robust estimation that aligns all response func-

---

1

tions up to a global exponential ambiguity. Figure 1 illustrates key ideas and the pipeline of our method. The performance of our approach is evaluated using various synthetic and real-world datasets in a quantitative manner.

2. Related Work

Our work is related to conventional radiometric calibration methods and computer vision applications to Internet photo collections.

Radiometric calibration. A Gretag Macbeth twenty-four patch color checker, whose reflectance value for each patch is known, is a commonly used tool for radiometric calibration [2]. Single image calibration without calibration chart is possible by assuming linear radiance distribution at edges [21, 22]. Using multiple images with different exposure times is a popular and practical approach for radiometric calibration. Classic methods include Debevec and Malik's approach fitting a non-parametric, smooth response function [6], and Mitsunaga and Nayar's method adopting a polynomial model [25]. The problem in [25] can be formulated as a rank minimization one to achieve superior robustness [19]. The response functions can also be represented using a more realistic model by using the database of measured response functions (DoRF) [11]. Such a representation can also be used in log-space to deal with moving cameras [16], varying illumination conditions [15], and dynamic scenes in the wild [1]. In addition to multiple exposure constraint, other constraints such as the statistical model of CCD imaging [33], geometry-invariant property [26], symmetric distribution of noise [24], temporal mixture of motion blur [34], and multiple directional lighting [30], are proposed for different imaging setups and applications. Note in all these approaches only one camera is calibrated using one image or multiple images, and the camera is controlled to adjust its settings (*e.g.*, manual mode with fixed white balance, ISO, but varying exposure times) for calibration purpose.

There are existing works that perform radiometric calibration for Internet photos. Kuthirummal *et al.* [17] explore priors on large image collections. By assuming the same camera model has a consistent response function and some radiometrically calibrated images of that camera model are available, a camera-specific response function could be estimated for the image collection according to the deviation from statistical priors. Due to the improvement of 3D reconstruction techniques on Internet-scale image sets, radiometric calibration becomes feasible by using the scene geometry estimated from SfM [32] and MVS [9, 10]. Diaz and Sturm [7, 8] jointly solve for the albedos, response functions, and illuminations by using nonlinear optimization and priors from DoRF. A more recent work by Li and

Peers [20] assume a local smoothness of image patch appearances and a 1D linear relationship over corresponding image patches, so that the radiometric calibration for multi-view images under varying illumination could be recast as the classic multiple exposure one [16], which could potentially be applied to Internet photos.

Computer vision meets Internet photos. Various computer vision problems could be extended to deal with Internet photos. Successful applications include scene completion [12], virtual tourism [31], weather estimation [28], composing pictures from sketches [3], image restoration [5], generating face animations [14], image colorization [4], intrinsic image decomposition [18], color consistency [27], photometric stereo [29], synthesizing time-lapse video [23], and so on and so forth. In many of these applications, the arbitrary nonlinear response functions for all cameras are simply approximated as a global Gamma correction, or even completely ignored as a linear one. So we believe a radiometric calibration solution for Internet photos is a very important technique that could potentially benefit miscellaneous application scenarios relying on photometric analysis.

3. Image Formation Model

We assume the scene reflectance follows the Lambertian model, and we know the scene geometry (surface normal) from SfM and MVS. The correspondence between 3D scene points and 2D pixels in all images are also obtained from 3D reconstruction. We take the j-th image in the image collection as an example, in which the scene is illuminated under the j-th natural lighting $L_j(\omega)$. Then the scene radiance of the i-th 3D point is determined by the interaction of lighting with its surface normal $\mathbf{n}_i \in \mathbb{R}^{3 \times 1}$ scaled by Lambertian albedo ρ_i as

$$R_{ij} = \int_\Omega v_{ij}(\omega)\rho_i L_j(\omega) \max((\mathbf{n}_i^\top \omega), 0)\mathrm{d}\omega, \qquad (1)$$

where $\omega \in \mathbb{R}^{3 \times 1}$ is a unit vector of spherical directions Ω, and $L_j(\omega)$ is the environment map for the j-th image which encodes the light intensity from the direction ω. $v_{ij}(\omega)$ is the visibility function which is set to 0 if the illumination from direction ω is not visible for the i-th 3D point projected to the j-th image or 1 otherwise. For any $\mathbf{n}_i$ with visibility function being equal to 1, it receives the light from its visible hemisphere Ω_i, and the integration over the visible hemisphere is simplified as

$$R_{ij} = \rho_i(\mathbf{n}_i^\top \bar{\mathbf{l}}_j), \qquad (2)$$

where $\bar{\mathbf{l}}_j = \int_{\Omega_i} L_j(\omega)\mathrm{d}\omega$.

When a scene is captured by the j-th camera, the image irradiance for the i-th pixel in the k-th color channel (*e.g.*,
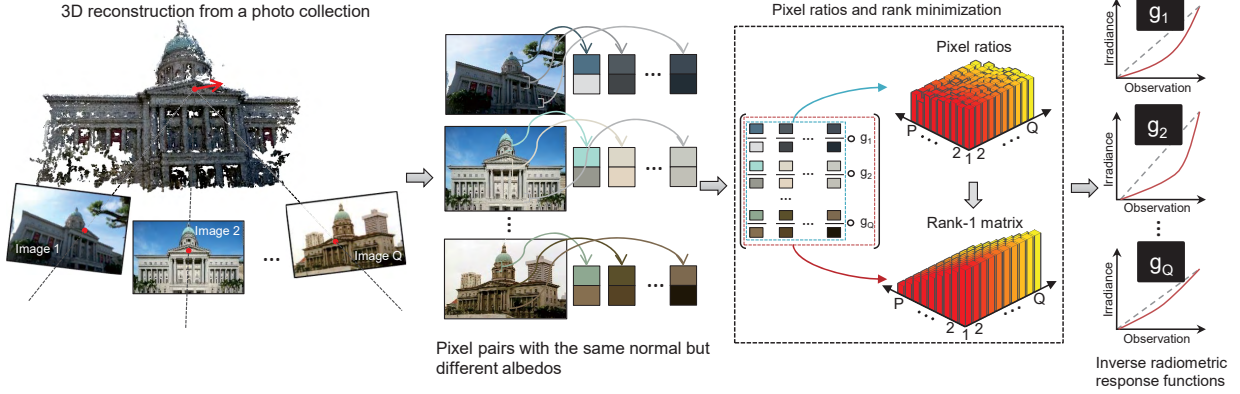
Figure 1. Pipeline of our method. We estimate the (inverse) radiometric response functions for each image $\{g_1, g_2, \cdots, g_Q\}$ by rank minimization over the stacks of pixel pairs. The "∘" operator applies each inverse response function g to both the numerator and denominator of ratio terms in the same row. The correct g transforms each row of the matrix to the same vector (up to a scale) to make the matrix rank-1.

RGB) can be represented as

$$I_{ij}^k = c_j^k t_j \rho_i^k (\mathbf{n}_i^\top \bar{\mathbf{l}}_j), \tag{3}$$

where c_j^k is a white balance scale and t_j is an exposure time for the j-th camera. Due to a nonlinear mapping of the camera radiometric response function $f(\cdot)$, the observations are distorted by f as

$$B_{ij}^k = f_j^k(I_{ij}^k) = f_j^k(c_j^k t_j \rho_i^k (\mathbf{n}_i^\top \bar{\mathbf{l}}_j)). \tag{4}$$

The response function is a monotonic function, so there exists a unique inverse function $g = f^{-1}$ to map the observation values to image irradiance values. By applying the inverse response function g to both sides of Eq. (4), we obtain

$$g_j^k(B_{ij}^k) = c_j^k t_j \rho_i^k (\mathbf{n}_i^\top \bar{\mathbf{l}}_j). \tag{5}$$

Radiometric calibration could be performed for three different color channels independently, so we drop the k-related terms thereafter. Denote $q_j = c_j t_j$ as an image-dependent scaling factor, then Eq. (5) is simplified as

$$g_j(B_{ij}) = q_j \rho_i (\mathbf{n}_i^\top \bar{\mathbf{l}}_j). \tag{6}$$

In the context of radiometric calibration for Internet photos, each image in the photo collection has its own g. Our goal is to simultaneously estimate g for all the images in a photo collection.

4. Radiometric Calibration Method

4.1. Formulation

With the scene 3D information available, it is possible to find points with the same surface normal, receiving the

same amount of light, but with different albedos in each image. We assume such pixels are identified for now, and the pixel selection method will be introduced in Sec. 4.4. Let a pair of such 3D points have normal $\mathbf{n}$ and lighting $\bar{\mathbf{l}}$, and their albedo values be ρ_m and ρ_n ($\rho_m \neq \rho_n$) respectively. Substituting these two points into Eq. (6) and taking the ratio between them, we obtain

$$\frac{g_j(B_{mj})}{g_j(B_{nj})} = \frac{q_j \rho_m (\mathbf{n}^\top \bar{\mathbf{l}})}{q_j \rho_n (\mathbf{n}^\top \bar{\mathbf{l}})} = \frac{\rho_m}{\rho_n}. \tag{7}$$

The albedo ratio consistency above is the key constraint we employ for radiometric calibration. Given a sufficient number of observation values B that cover a broad intensity range, we can build a system of equations solved by nonlinear optimization [25]. Recent progress in radiometric calibration shows that the rank minimization could solve such a problem in a more robust manner and effectively avoid overfitting [19]. Therefore, we formulate our problem in a matrix form, whose minimum rank corresponds to the correct estimates of inverse response functions.

Denote P as the total number of pixel pairs with the same normal (and lighting) but different albedos, and Q as the total number of images, or equivalently pairs of lighting conditions and cameras. We could arrange these pixels according to the ratio format of Eq. (7) and stack them as the following matrix:

$$\mathbf{A}_{Q \times P} = \begin{pmatrix} \frac{g_1(B_{11})}{g_1(B_{01})} & \frac{g_1(B_{21})}{g_1(B_{11})} & \cdots & \frac{g_1(B_{P1})}{g_1(B_{(P-1)1})} \\ \frac{g_2(B_{12})}{g_2(B_{02})} & \frac{g_2(B_{22})}{g_2(B_{12})} & \cdots & \frac{g_2(B_{P2})}{g_2(B_{(P-1)2})} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{g_Q(B_{1Q})}{g_Q(B_{0Q})} & \frac{g_Q(B_{2Q})}{g_Q(B_{1Q})} & \cdots & \frac{g_Q(B_{PQ})}{g_Q(B_{(P-1)Q})} \end{pmatrix}. \tag{8}$$

The optimal inverse response function g_j for each row transforms each pixel ratio in the matrix to its corresponding albedo ratio, so that each row of $\mathbf{A}$ becomes

$\left(\frac{\rho_1}{\rho_0}, \frac{\rho_2}{\rho_1}, \cdots, \frac{\rho_P}{\rho_{(P-1)}}\right)$[1], which obviously makes $\mathbf{A}$ a rank-1 matrix. Thus, our radiometric calibration problem becomes the following rank minimization one:

$$\{g_1^*, g_2^*, \cdots, g_Q^*\} = \underset{\{g_1, g_2, \cdots, g_Q\}}{\operatorname{argmin}} \quad \operatorname{rank}(\mathbf{A}). \qquad (9)$$

The above operation and process is illustrated in Fig. 1.

4.2. Optimization

We solve the above rank minimization using a similar approach as in [19], which is represented by the condition number as

$$\{g_1^*, g_2^*, \cdots, g_Q^*\} = \underset{\{g_1, g_2, \cdots, g_Q\}}{\operatorname{argmin}} \frac{\sigma_2(\mathbf{A})}{\sigma_1(\mathbf{A})}, \qquad (10)$$

where $\sigma_i(\mathbf{A})$ is the i-th singular value of $\mathbf{A}$.

We choose to use the polynomial representation for g as suggested by [19]. The main consideration is that polynomial representation is more appropriate for gradient-based convex optimization because of its smoothness. Both irradiance and observation values are normalized in the range of 0 to 1. Then the polynomial representation of g becomes

$$g(B) = B + B(B-1)\sum_{i=1}^{S-1} p_i B^{S-i-1}, \qquad (11)$$

where $\{p_1, p_2, \cdots, p_{S-1}\}$ are the polynomial coefficients to be estimated. Such an expression uses only $S-1$ unknowns to represent an S-order polynomial. The end point constraints for inverse response functions are explicitly enforced, since Eq. (11) satisfies $g(0) = 0$ and $g(1) = 1$.

Note that we only borrow the optimization strategy from [19] to solve Eq. (10). In fact, our problem is much more challenging than [19] due to the joint estimation of many different response functions, and the structure of the matrix whose rank needs to be minimized is completely different due to pixel ratios. We find such a problem cannot be directly solved like [19] as directly solving for all g simultaneously is quite unstable, because each g_j transforms one row of $\mathbf{A}$ independently and this significantly increases the search space for minimum rank. So we solve this issue by using a *pairwise optimization* followed by a *global refinement*.

The pairwise optimization means we select two rows as "base" image pair and align all the other rows to the base in an incremental manner. The base image pair is selected as the two rows of $\mathbf{A}$ with the minimum difference after applying the estimated inverse response functions, through

---

[1] For easy representation, we only show one option for arranging the pixel pairs (there could be ratio terms like $\frac{\rho_2}{\rho_0}$, $\frac{\rho_P}{\rho_1}$, *etc.*). Given $p+1$ different ρ values, there are C_{P+1}^2 possible combinations of taking the ratio.

Algorithm 1 Radiometric calibration algorithm

INPUT: Input images, with pixels selected and stacked as the matrix of Eq. (8).
// Pairwise optimization:
for all pairwise combinations using two rows of $\mathbf{A}$ **do**
    Solve two g using Eq. (10);
    Apply these two g to the corresponding rows in $\mathbf{A}$;
end for
Select g_m^0 and g_n^0 that make corresponding rows of $\mathbf{A}$ have the minimum difference;
for $k = \{1, 2, \cdots, Q\} \wedge k \neq \{m, n\}$ **do**
    Build a matrix with the $\{m, n, k\}$-th rows of $\mathbf{A}$ and solve g_k^0 using Eq. (10), with g_m^0 and g_n^0 fixed;
end for
// Global refinement:
Solve Eq. (10) for all g simultaneously using $\{g_1^0, g_2^0, \cdots, g_Q^0\}$ as initial values;
OUTPUT: Inverse response functions $\{g_1^*, g_2^*, \cdots, g_Q^*\}$.

---

solving Eq. (10) for all C_Q^2 submatrices composed by two rows of $\mathbf{A}$. Then we add one row at a time to solve for the remaining $Q-2$ rows for submatrices with three rows. The estimated inverse response functions here are denoted as g^0. The global refinement takes g^0 as initial values to solve for all g simultaneously using Eq. (10). Our complete algorithm is summarized as Algorithm 1.

4.3. Exponential ambiguity

Similar to existing radiometric calibration methods [1, 11, 15, 16, 19, 25], our method also suffers from the exponential ambiguity. This is because for a set of optimized solution g, g^γ for any unknown γ also keeps the ratio consistent in Eq. (7) and makes $\mathbf{A}$ rank-1. Note there only exists one γ for all images. The uncertainty of γ brings challenges to joint estimation of all g and this is part of the reason that we have to start from a pairwise optimization approach in Algorithm 1. Although infinite many solutions exist for one set of images without additional constraint, our method naturally unifies all response functions to the same ambiguity and such results could be useful for tasks such as high dynamic range imaging and panorama image stitching.

We could remove such a one-parameter ambiguity by inserting only one radiometrically calibrated image into the photo collection. Without losing generality, we put pixel pairs from this image in the first row and fix g_1^* as a linear function during the optimization. Then the optimized results for $\{g_2^*, g_3^*, \cdots, g_Q^*\}$ automatically align all the other rows of $\mathbf{A}$ with the linearized values to form a set of solution free of exponential ambiguity.

4.4. Implementation details

3D reconstruction. To build the matrix in Eq. (8), we need to extract corresponding pixels in all images that have the same surface normal, under the same lighting condition, but with different albedos. We first perform 3D reconstruction (SfM [32] and MVS [10]) using the input photo collection. The 3D points with the same surface normal are selected and projected onto 2D images. We then calculate pairwise ratio for these pixels as initial guess of albedo ratios. The selected pixels in each pair should receive the same amount of environment illumination, if their visibility function v_{ij} defined in Eq. (1) were the same. However, the visibility information cannot be accurately estimated through sparse 3D reconstruction and unknown environment illumination, which results in noise in real data such as cast shadow and local illumination variations. Therefore, we propose a simple outlier rejection approach to deal with this issue. We find that the majority of such initially selected pixel pairs show similar ratio values, and any noisy pixels appearing in either numerator or denominator cause the ratio significantly different from others. Such outliers could be easily identified and discarded by a line fitting using RANSAC. Finally, remaining pixel pairs observed in all images are stacked as the matrix in Eq. (8) for optimization.

Details of optimization. In practice, we only require dozens of images as input for the 3D reconstruction by SfM and MVS. According to Algorithm 1, it is not necessary to optimize Q response functions simultaneously, since we use an incremental approach to estimate all response functions except for the two that are selected as base. Given a complete set of images, we divide it into several subgroups (*e.g.*, 10 images in each subgroup), and solve for each subgroup using Algorithm 1. We empirically find such divide-and-conquer strategy gives a more stable solution, and this property allows the parallel processing of large amount data.

We use the Matlab build-in function "lsqnonlin" to solve our nonlinear optimization. The initial guess for inverse response functions are chosen as a linear function for all images in the pairwise optimization step. A monotonicity constraint is added to penalize non-monotonic estimates similarly as adopted in [19]. We further add a second-order derivative constraint by assuming most response functions have either concave or convex shapes. There are response functions with more irregular shapes according to [11], but they are rarely observed in common digital cameras.

Degenerate case. One obvious degenerate case for our problem is a scene with uniform albedo. Because the uniform albedo causes $\rho_m = \rho_n$ in Eq. (7), and every element in $\mathbf{A}$ becomes one and thus $\mathbf{A}$ becomes rank-1. Therefore, we need at least two different albedo values in the scene.

However, if pixels with two different albedos are all on the same plane (assuming there is no shadow in the scene, so that the same surface normal receives the same amount of environment lighting), it falls into a similar degenerate case. Therefore, the minimum requirement to avoid the degenerate case is a pair of different albedos on two different planes. This is because even if the albedo ratio is the same after optimization, if two pixel pairs are scaled by different shading terms $\mathbf{n}^\top \mathbf{l}$ and then nonlinearly mapped by the same response function, their ratios become different before optimization. Fortunately, a wild scene may contain much more variations in either albedo or normal than our minimum requirement. Since the problem formulation is highly nonlinear, it is non-trivial to provide analytical proof for the number of different albedo or surface normal required, but we will experimentally analyze such an issue in the next section.

5. Experimental Results

We perform quantitative evaluation of our method using both synthetic and real data. The error metric used for evaluation is the rooted mean square errors (RMSE) of the estimated inverse response function w.r.t. the ground truth and the disparity, *i.e.*, the maximum absolute difference between the estimated and the ground truth curves.

5.1. Experiment with synthetic data

Number of pixel pairs vs. order of polynomial. Our method is expected to be more stable and accurate given more diverse values of pixel pairs (albedo and normal variations) and fitted with higher order polynomials. We use synthetic data to verify the accuracy under different number of albedo values and polynomial orders by testing 6 types of pixel pairs and 6 different polynomial orders. The 6 groups of input data are generated by changing the number of different albedo values multiplied by different normals as $\{1 \times 2, 2 \times 3, 3 \times 4, 4 \times 6, 5 \times 10, 6 \times 15\}$. Here, 1×2 means one pair of different albedo values on two different planes. We then apply 10 different lighting conditions and 10 different response functions from the DoRF database [11] to generate our observation values. The inverse response functions are estimated using our methods, and the RMSE and disparity are summarized in Fig. 2.

As expected, the average errors show a row-wise decreasing tendency due to more diverse input data variations. It is interesting to note that only 24 pairs of points from each image (four pairs of albedo values on six different planes) produce reasonably small error (RMSE around 0.01) for a joint estimation of 10 different response functions. From Fig. 2, we can also see our method is not sensitive to the choice of polynomial order and in general a polynomial order larger than 5 works well. We fix the polynomial order to 7 in all experiments as a tradeoff between accuracy and

Figure 2. The average RMSE/disparity w.r.t. number of pixel pairs (row-wise) and order of polynomials (column-wise). "Red" means larger and "blue" means smaller errors.

RMSE	Polynomial order					
# pixel pairs	3	4	5	6	7	8
1×2	0.0761	0.0433	0.0539	0.0566	0.0554	0.0522
2×3	0.0662	0.0224	0.0185	0.0171	0.0149	0.0168
3×4	0.1458	0.0167	0.0162	0.0194	0.0170	0.0149
4×6	0.0742	0.0144	0.0144	0.0130	0.0094	0.0075
5×10	0.0628	0.0159	0.0076	0.0066	0.0095	0.0075
6×15	0.0665	0.0194	0.0094	0.0079	0.0067	0.0073

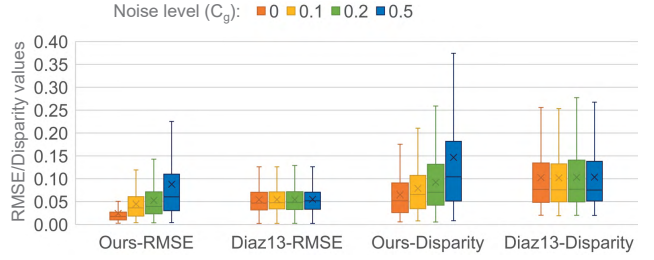| Disparity	Polynomial order					
# pixel pairs	3	4	5	6	7	8
1×2	0.1334	0.0713	0.0940	0.1014	0.0990	0.0925
2×3	0.1162	0.0412	0.0351	0.0329	0.0287	0.0319
3×4	0.2681	0.0320	0.0312	0.0377	0.0339	0.0288
4×6	0.1334	0.0288	0.0288	0.0260	0.0195	0.0169
5×10	0.1117	0.0325	0.0175	0.0152	0.0206	0.0169
6×15	0.1174	0.0398	0.0206	0.0175	0.0166	0.0180

Figure 3. Evaluation under various noise levels and comparison between our method and Diaz13 [8]. The box-and-whisker plot shows the mean (indicated as "×"), median, the first and third quartile, and the minimum and maximum values for RMSE and disparity for 200 (20 × 10) estimated response functions.

complexity.

Performance with various noise. We first add quantization to mimic the 8-bit image formation, and then we add Poisson noise as suggested in [13, 19], which describes the imaging noise in a more realistic manner by considering signal-dependent shot and dark current noise. The noise level is controlled by the camera gain parameter C_g, and larger C_g means more severe noise. Please refer to Equations (9)-(11) in [19] for the noise model representation. We perform 20 trials and each test contains 10 randomly selected response functions from DoRF with 4×6 pixel pairs.

We evaluate another photometric image formation based method [8] (denoted as "Diaz13") implemented by ourselves using the same data. We find a joint estimation to all variables (albedos, lighting and response function coefficients) produces unreliable results, due to the nonlinear optimization over too many variables. Therefore, we provide the ground truth lighting coefficients in our implementation of Diaz13 and use this as the stable performance of Diaz13 for comparison. The results under various $C_g = \{0, 0.1, 0.2, 0.5\}$ (where $C_g = 0$ means only quantization noise) for both methods are plotted in Fig. 3. Our method outperforms Diaz13 for most noise levels, but Diaz13 shows more stable but less accurate performance under these noise levels. When the noise is large, our method shows degraded performance partially due to that ratio operation magnifies the noise. Note that in real case, Diaz13 requires the dense reconstruction for lighting estimation, while we can only work on a few selected pixel pairs.

Results on single-view images. Single-view synthetic test is performed to provide an intuitive example with quan-
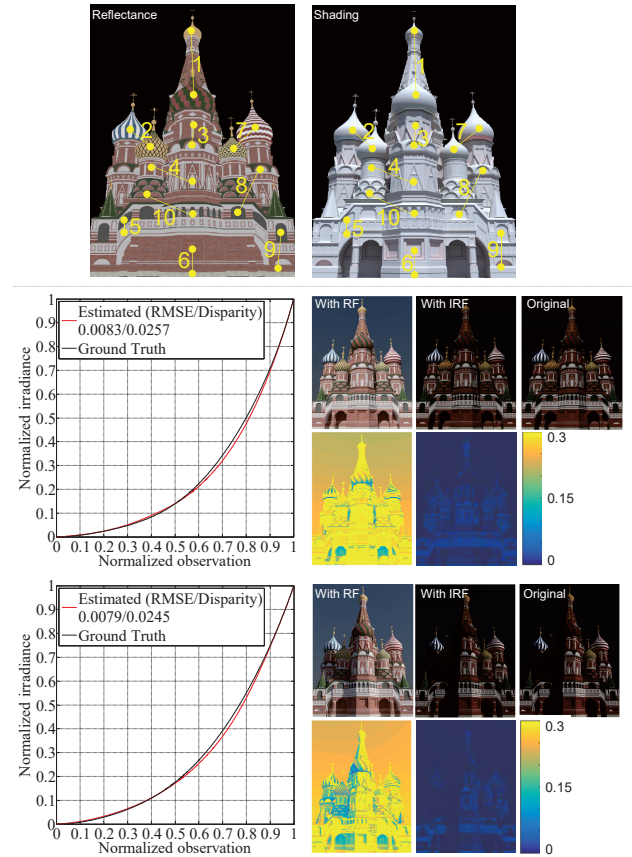

Figure 4. Radiometric calibration results using a synthetic dataset. The upper row shows the ground truth reflectance, shading images, and the ten selected pixel pairs (yellow numbers) with the same normal but different albedo values. Two example results of the estimated inverse response functions and the ground truth curves are plotted, with the RMSE and disparity values shown in the legend. The nonlinear observations ("With RF"), linearized images ("With IRF") and their absolute difference maps w.r.t. the "Original" images are shown next to the inverse response curve plots.

titative analysis, which is free of errors from 3D reconstruction. We use the data from [18], which is generated using a
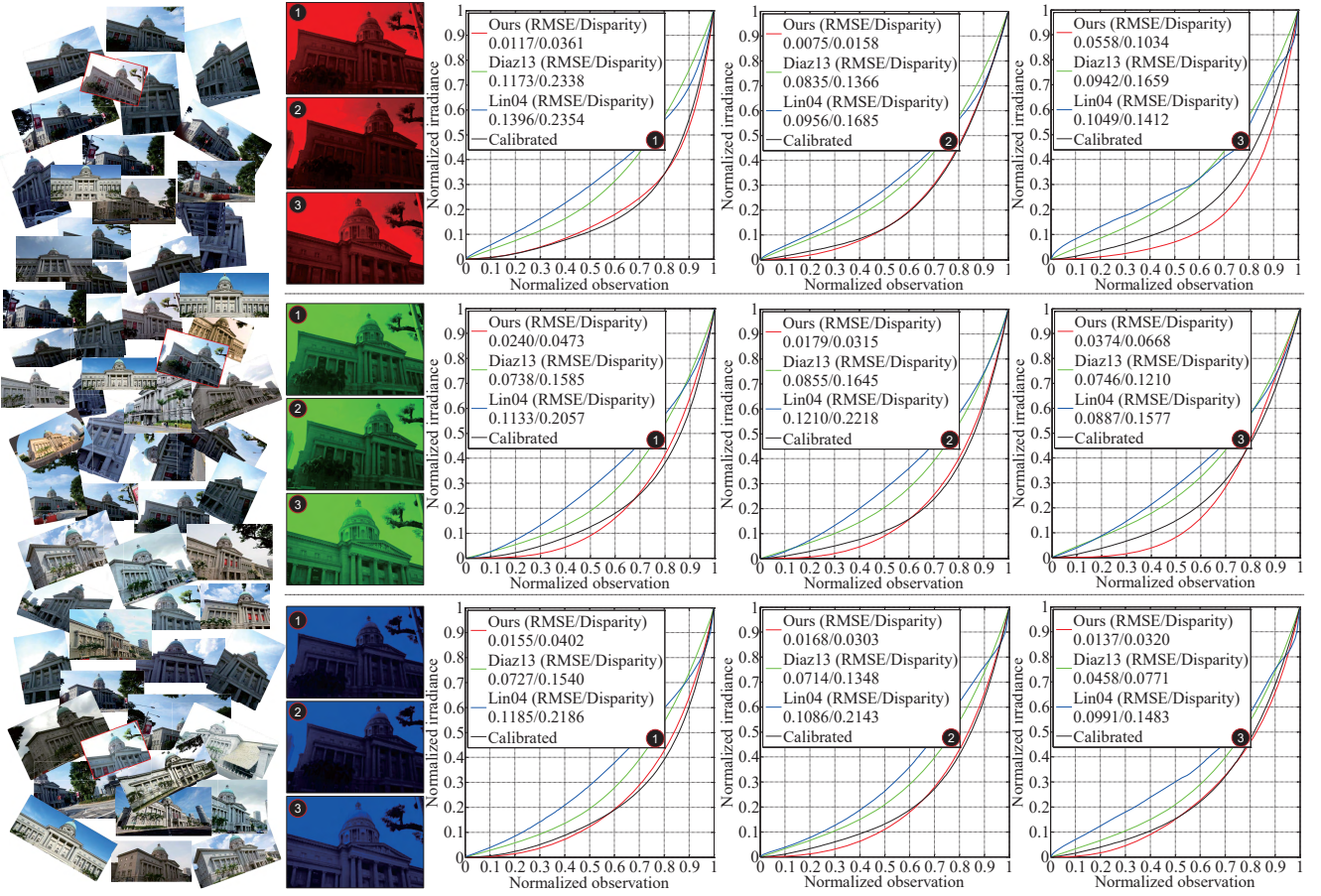
Figure 5. Estimated inverse radiometric response functions using an image collection mixed with Internet photos and captured images (in red box). The results compared with "Diaz13" [8] and "Lin04" [21] on RGB channels of three images and calibrated ground truth are shown. The RMSE and Disparity are in the legend of each plot.

physics-based renderer. Given the ground truth reflectance and shading images, as show in the upper part of Fig. 4, we manually select 10 pairs of pixels with the same surface normal but different albedo values (labeled using yellow numbers on the reflectance and shading images). We randomly apply 10 different response functions to produce 10 observation images, and then use the selected pixels as input to perform radiometric calibration. Two typical results of the estimated inverse radiometric response functions w.r.t. the ground truth curves are shown in Fig. 4. We further apply the inverse response functions to the observation images, and the close appearances between the linearized images and the original images show the correctness of our radiometric calibration method.

5.2. Real data experiment

To perform quantitative evaluation using real data, we create datasets containing mixture of Internet photos and images captured using controlled cameras for three different scenes. The total numbers of images are 55 for the dataset used in Fig. 5, and 44, 31 respectively for the other two datasets used in Fig. 6. We use three controlled cameras (① Sony Alpha7, ② Nikon D800, and ③ Canon EOS M2) for all datasets, and calibrate their response functions using multiple exposure approach [19]. We perform 3D reconstruction and radiometric calibration for all images in each dataset. Two (out of three) captured images are used as the base image pair for pairwise optimization, and the calibrated response functions are used to remove the exponential ambiguity for all estimates from our method. We also run our own implementation of "Diaz13" [8][2] and [21] (denoted as "Lin04")[3] on the same dataset for comparison.

The quantitative results (estimated inverse radiometric response functions and their RMSE/Disparity w.r.t. the ground truth) on RGB channels of three images (captured by three different camera models) from the first scene are plotted in Fig. 5, and we show similar plots only using the R channel of the other two scenes in Fig. 6. Except for the

---

[2]This time the lighting coefficients are also part of the optimization which are initialized randomly.

[3]Implemented by Jean-Franois Lalonde and available from: https://github.com/jflalonde/radiometricCalibration
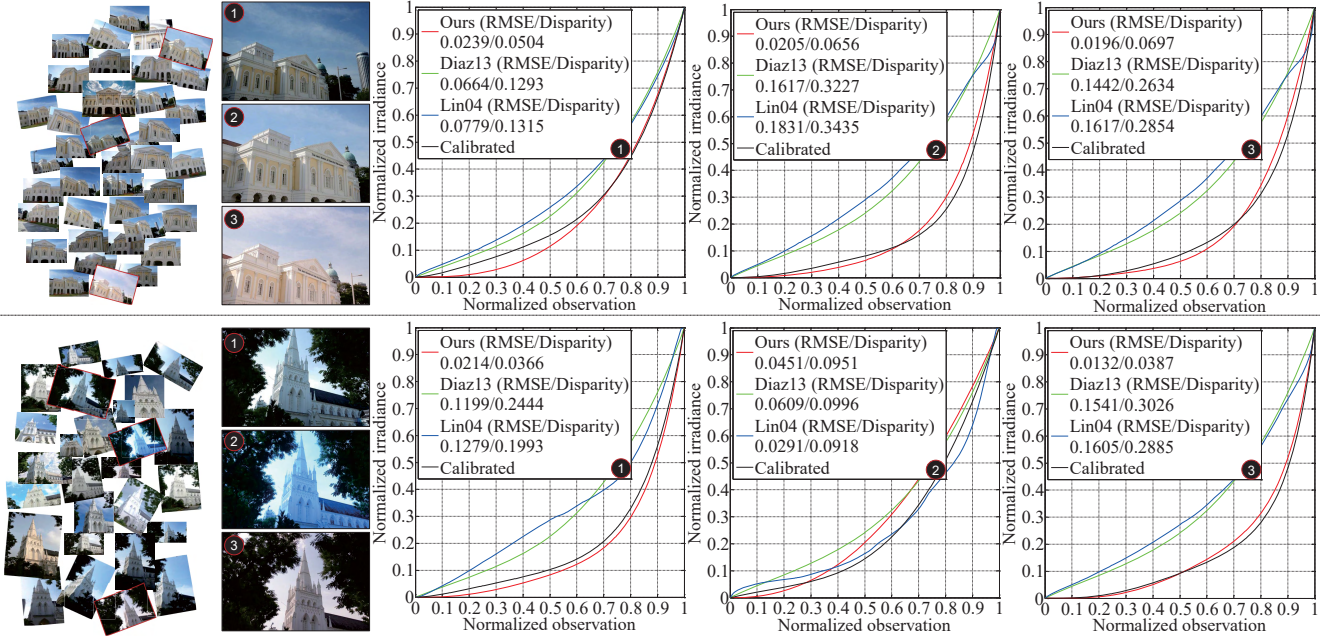
Figure 6. Estimated inverse radiometric response functions using another two image collections mixed with Internet photos and captured images (in red box). The results compared with "Diaz13" [8] and "Lin04" [21] on R channel of three images and calibrated ground truth are shown. The RMSE and Disparity are in the legend of each plot.

second example in the bottom row of Fig. 6 where "Lin04" [21] shows better result than ours, all our estimates show the closest shape (with minimum RMSE/Disparity) to the calibrated ground truth. The high accuracy of our method is partly due to that we use calibrated ground truth to resolve the exponential ambiguity, but the experiment here proves that our method is able to jointly estimate the inverse response functions for a set of images up to the same exponential ambiguity reliably.

6. Discussions

We present a method to perform radiometric calibration for images from Internet photo collections. Compared to the conventional radiometric calibration problem, the Internet photos have a wide range of unknown camera settings and response functions, which are neither accessible nor adjustable. We solve this challenging problem by using the scene albedo ratio, which is assumed to be consistent in all images. We develop an optimization method based on rank minimization for jointly estimating multiple response functions. The effectiveness of the proposed method is demonstrated using both synthetic and real-world data.

Currently, we need to assume the 3D reconstruction is sufficiently reliable for extracting points that share the same surface normal. With radiometric calibration problem solved, it will be interesting to combine with photometric 3D reconstruction using Internet photos [29] to further improve the quality of 3D reconstruction. Although our solution is naturally invariant to white balance and exposure-

time settings of individual images, we can only recover the scene radiance up to a linearly scaled value (for each color channel) through the radiometric calibration. Explicitly inferring the white balance and exposure-time settings of Internet photos is an interesting future topic. Current method cannot handle images with severe noise, and an even more robust solution that compensates the noise magnification issue caused by ratio operation is desired.

References

[1] A. Badki, N. K. Kalantari, and P. Sen. Robust radiometric calibration for dynamic scenes in the wild. In *Proc. of Internatoinal Conference on Computational Photography*, 2015. 2, 4

[2] Y.-C. Chang and J. F. Reid. RGB calibration for color image analysis in machine vision. *IEEE Transactions on Image Processing*, 5(10):1414–1422, 1996. 2

[3] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu. Sketch2photo: Internet image montage. *ACM Transactions on Graphics (Proc. of ACM SIGGRAPH Asia)*, 28(5):124:1–124:10, 2009. 2

[4] A. Y.-S. Chia, S. Zhuo, R. K. Gupta, Y.-W. Tai, S.-Y. Cho, P. Tan, and S. Lin. Semantic colorization with internet images. *ACM Transactions on Graphics (Proc. of ACM SIGGRAPH Asia)*, 30(6):156:1–156:8, 2011. 2

[5] K. Dale, M. K. Johnson, K. Sunkavalli, W. Matusik, and H. Pfister. Image restoration using online photo collections. In *Proc. of Internatoinal Conference on Computer Vision*, 2009. 2

[6] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proc. of ACM SIGGRAPH*, 1997. 2

[7] M. Diaz and P. Sturm. Radiometric calibration using photo collections. In *Proc. of Internatoinal Conference on Computational Photography*, 2011. 2

[8] M. Diaz and P. Sturm. Estimating photometric properties from image collections. *Journal of Mathematical Imaging and Vision*, 43:93–107, 2013. 2, 6, 7, 8

[9] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski. Towards internet-scale multi-view stereo. In *Proc. of Computer Vision and Pattern Recognition*, 2010. 1, 2

[10] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376, 2010. 1, 2, 5

[11] M. D. Grossberg and S. K. Nayar. Modeling the space of camera response functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1272–1282, 2004. 2, 4, 5

[12] J. Hays and A. A. Efros. Scene completion using millions of photographs. *ACM Transactions on Graphics (Proc. of ACM SIGGRAPH)*, 26(3), 2007. 2

[13] Y. Hwang, J.-S. Kim, and I. S. Kweon. Difference-based image noise modeling using Skellam distribution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7):1329–1341, 2012. 6

[14] I. Kemelmacher-Shlizerman, E. Shechtman, R. Garg, and S. M. Seitz. Exploring photobios. *ACM Transactions on Graphics (Proc. of ACM SIGGRAPH)*, 30(4):61:1–61:10, 2011. 2

[15] S. J. Kim, J. Frahm, and M. Pollefeys. Radiometric calibration with illumination change for outdoor scene analysis. In *Proc. of Computer Vision and Pattern Recognition*, 2008. 1, 2, 4

[16] S. J. Kim and M. Pollefeys. Radiometric self-alignment of image sequences. In *Proc. of Computer Vision and Pattern Recognition*, 2004. 2, 4

[17] S. Kuthirummal, A. Agarwala, D. B. Goldman, and S. K. Nayar. Priors for large photo collections and what they reveal about cameras. In *Proc. of European Conference on Computer Vision*, 2008. 2

[18] P.-Y. Laffont, A. Bousseau, S. Paris, F. Durand, and G. Drettakis. Coherent intrinsic images from photo collections. *ACM Transactions on Graphics (Proc. of ACM SIGGRAPH)*, 31(6):202:1–202:11, 2012. 1, 2, 6

[19] J.-Y. Lee, Y. Matsushita, B. Shi, I. S. Kweon, and K. Ikeuchi. Radiometric calibration by rank minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):144–156, 2013. 2, 3, 4, 5, 6, 7, 8

[20] H. Li and P. Peers. Radiometric transfer: Example-based radiometric linearization of photographs. *Computer Graphics Forum (Proc. of Eurographics Symposium on Rendering)*, 34(4):109–118, 2015. 2

[21] S. Lin, J. Gu, S. Yamazaki, and H.-Y. Shum. Radiometric calibration from a single image. In *Proc. of Computer Vision and Pattern Recognition*, 2004. 2, 7, 8

[22] S. Lin and L. Zhang. Determining the radiometric response function from a single grayscale image. In *Proc. of Computer Vision and Pattern Recognition*, 2005. 2

[23] R. Martin-Brualla, D. Gallup, and S. M. Seitz. Time-lapse mining from internet photos. *ACM Transactions on Graphics (Proc. of ACM SIGGRAPH)*, 34(4):62:1–62:8, 2015. 2

[24] Y. Matsushita and S. Lin. Radiometric calibration from noise distributions. In *Proc. of Computer Vision and Pattern Recognition*, 2007. 2

[25] T. Mitsunaga and S. K. Nayar. Radiometric self-calibration. In *Proc. of Computer Vision and Pattern Recognition*, 1999. 1, 2, 3, 4

[26] T.-T. Ng, S.-F. Chang, and M.-P. Tsui. Using geometry invariants for camera response function estimation. In *Proc. of Computer Vision and Pattern Recognition*, 2007. 2

[27] J. Park, Y.-W. Tai, S. N. Sinha, and I. S. Kweon. Efficient and robust color consistency for community photo collections. In *Proc. of Computer Vision and Pattern Recognition*, 2016. 2

[28] L. Shen and P. Tan. Photometric stereo and weather estimation using internet images. In *Proc. of Computer Vision and Pattern Recognition*, 2009. 2

[29] B. Shi, K. Inose, Y. Matsushita, P. Tan, S.-K. Yeung, and K. Ikeuchi. Photometric stereo using internet images. In *Proc. of International Conference on 3D Vision*, 2014. 1, 2, 8

[30] B. Shi, Y. Matsushita, Y. Wei, C. Xu, and P. Tan. Self-calibrating photometric stereo. In *Proc. of Computer Vision and Pattern Recognition*, 2010. 2

[31] N. Snavely, R. Garg, S. M. Seitz, and R. Szeliski. Finding paths through the world's photos. *ACM Transactions on Graphics (Proc. of ACM SIGGRAPH)*, 27(3):15:1–15:11, 2008. 2

[32] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3D. *ACM Transactions on Graphics (Proc. of ACM SIGGRAPH)*, 25(3):835–846, 2006. 1, 2, 5

[33] Y. Tsin, V. Ramesh, and T. Kanade. Statistical calibration of CCD imaging process. In *Proc. of Internatoinal Conference on Computer Vision*, 2001. 2

[34] B. Wilburn, H. Xu, and Y. Matsushita. Radiometric calibration using temporal irradiance mixtures. In *Proc. of Computer Vision and Pattern Recognition*, 2008. 2