

# Ambiguity-Free Radiometric Calibration for Internet Photo Collections

Zhipeng Mo, Boxin Shi, *Member, IEEE*,  
Sai-Kit Yeung, *Member, IEEE*, and Yasuyuki Matsushita, *Senior Member, IEEE*

**Abstract**—Radiometrically calibrating nonlinear images from Internet photo collections makes photometric analysis applicable not only to lab data but also to big image data in the wild. However, conventional calibration methods cannot be directly applied to such photo collections. This paper presents a method to jointly perform radiometric calibration for a set of nonlinear images in Internet photo collections. By incorporating the consistency of scene reflectance of corresponding pixels across nonlinear images, the proposed method first estimates radiometric response functions of all the nonlinear images up to a unique exponential ambiguity using a rank minimization framework. The ambiguity is then resolved using the linear edge color blending constraint. Quantitative evaluation using both synthetic and real-world data shows the effectiveness of the proposed method.

**Index Terms**—Radiometric Calibration, Photo Collections, Internet Images, Exponential Ambiguity, Edge Color Blending

## 1 INTRODUCTION

FOR a popular landmark, millions of pictures are recorded and shared through the Internet. Such Internet images and community photo collections serve as a comprehensive image resource for computer vision research, because they contain images captured from different viewpoints, at different times, under different illumination conditions, and using diverse types of cameras and settings. By exploring the interrelationship of these images, geometric analyses such as geometric camera calibration and 3D reconstruction, which are generally infeasible using a single image, become tractable by establishing correspondences across multiple images. Recent progress on structure from motion (SfM) [8] and multi-view stereo (MVS) [9, 10] shows successful applications using Internet photos.

Photometric analysis is another important problem for analyzing images from Internet photo collections organized by a set of cameras and images sharing similar contents, and radiometric calibration is a key prerequisite for photometric analysis. Many computer vision problems, such as intrinsic image decomposition [11] and photometric stereo [12], when they are to be applied to Internet photos, require input images to be radiometrically linearized (*linear images*). However, a commercial camera usually maps the scene radiance to its pixel values in a nonlinear manner (*nonlinear images*) for compressing the dynamic range and an aesthetic purpose. Such a mapping is unknown in most cases and treated as business secrets of 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.

Traditional radiometric calibration approaches capture a static scene under various exposure times [1]. For an outdoor scene with illumination changes, given a nonlinear image sequence from a fixed viewpoint, the problem could also be solved by modeling the image formation and using the assumption of consistent scene albedo across linearized images [7]. Radiometrically calibrating Internet photos shares a similar spirit to [7], but it exhibits further challenge due to that (1) the scenes are captured from multiple diverse viewpoints, (2) each nonlinear image is captured by an unknown camera with unknown settings, *e.g.*, white balance and exposure times, and (3) a set of distinct response functions needs to be estimated simultaneously.

Theoretically, single image based methods [13, 14] can be applied to calibrate the response functions of all nonlinear images in an Internet photo collection one by one. But the constraints they rely on such as linear edge color blending [13] and symmetric distribution of noise [14] can only be observed in high-quality images without compression artifacts. Therefore, in practice, it is not straightforward to apply these methods for Internet photos that are mostly degraded.

This paper proposes a method to jointly perform radiometric calibration of all cameras using a collection of Internet photos of the same scene. Our key assumption is that the scene reflectance (albedo) is the same for corresponding pixels across linearized images, and we have an access to the geometric information of the scene that are computed by SfM and MVS. By selecting pixels that correspond to the same surface normal in a pairwise manner in *each* nonlinear image, we compute a vector of *ratios* of the observed pixel values. The ratio operation cancels the influence of different white balance settings, exposure times, and environment illumination conditions. By stacking the ratio vectors of all the nonlinear images, we form a matrix. If correct inverse radiometric response functions are applied to all the nonlinear images, the matrix should exhibit a rank-1 structure. Capitalizing on this observation, we develop a method for estimating the inverse response functions, up to a *unified exponential ambiguity*, based on rank minimization.

- Z. Mo is with Pillar of Information Systems Technology and Design, Singapore University of Technology and Design. E-mail: zhipeng\_mo@mymail.sutd.edu.sg.
- B. Shi (corresponding author) is with National Engineering Laboratory for Video Technology, School of Electronics Engineering and Computer Science, Peking University, China. E-mail: shiboxin@pku.edu.cn.
- S.-K. Yeung is with Division of Integrative Systems and Design and Department of Computer Science and Engineering, Hong Kong University of Science and Technology. E-mail: saikit@ust.hk.
- Y. Matsushita is with Graduate School of Information Science and Technology, Osaka University, Japan. E-mail: yasumat@ist.osaka-u.ac.jp.

TABLE 1  
A unified summarization of various radiometric calibration methods.

Assumption	Equation	Solution	Ambiguity
Unified representation	$g(B_m) = K_m R$ , $g(B_n) = K_n R$ $\Rightarrow g(B_m)/g(B_n) = K_m/K_n$	(Non)linear equations (EQ) Rank minimization (RM)	Exponential Additional constraint
Different exposure	$R$ : radiance values as a row vector $K_{m,n}$ : exposure time for $m$ - $n$ -th image	EQ: [1–3] RM: [4, 5]	Exponential Known exposure ratio
Local smoothness	$R$ : matched patch radiance $K_{m,n}$ : approx. linear scale for patch $m, n$	EQ: [6] RM: N. A.	Exponential One linear image
Photometric image	$R$ : $c_j t_j (\mathbf{n}^\top \mathbf{1})$ $K_{m,n}$ : different $\rho_{m,n}$ with the same $\mathbf{n}$	EQ: [7] RM: Our method	Exponential Known albedo ratio, linear edge color blending, etc.

Such a unified ambiguity consists of only one unknown parameter. We develop a method for resolving the ambiguity by adapting the method of [13] to Internet photos. The method of [13] assumes a linear blending of edge color triplets in the  $RGB$  space when a response function is linear for noise-free images. To work well with heavily degraded Internet photos, we develop an outlier rejection method by taking the advantage of many photos. We show that the disambiguation problem for a set of inverse response functions in our context reduces to a 1D search problem, and present a stable estimation technique for obtaining the *ambiguity-free* solution. Figure 1 illustrates key ideas and the pipeline of our method.

This paper extends its preliminary version [15] in the following three aspects:

- We propose the ambiguity-free solution for radiometric calibration of Internet photo collections given the unified exponential ambiguity [15] by developing a robust 1D search method.
- We analyze the applicability of single image radiometric calibration method to images with compression noise and adapt the linear edge color blending constraint [13] to Internet photos by selecting reliable edge color triplets.
- We summarize a unified framework for existing radiometric calibration methods with similar formulation and constraint for deeper understanding the nature of relevant problems.

The rest of the paper is organized as follows. Section 2 reviews related work. In Sections 3 and 4, we introduce the image formation model and the proposed method with ambiguity-free solution. Section 5 shows quantitative results of our approach on various synthetic and real-world datasets. We conclude this paper in Section 6.

## 2 RELATED WORK

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

### 2.1 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 [16]. Single image calibration without calibration chart is possible by assuming different constraints. Lin *et al.* [13] assume color blending at edge pixels so the radiance distribution for pixels along the edge in  $RGB$  space should be linear. This assumption is also useful for gray images by calculating the histogram of radiance for edge pixels [17]. Matsushita and Lin [14] propose to use the symmetric distribution of noise

on imaging process. Other single image constraints like the linear intensity profiles of locally planar irradiance points [18] and skin structure and pigment components [19] are also introduced in recent works.

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 [20], and Mitsunaga and Nayar’s method adopting a polynomial model [1]. The problem in [1] can be formulated as a rank minimization one to achieve superior robustness [4]. The response functions can also be represented using a more realistic model by using the database of measured response functions (DoRF) [3]. Such a representation can also be used in log-space to deal with moving cameras [2, 21], varying illumination conditions [7], and dynamic scenes in the wild [5]. In addition to multiple exposure constraint, other constraints such as the statistical model of CCD imaging [22], temporal mixture of motion blur [23], image vignetting effect [24], known albedo ratio of two-colored surface under near-lighting condition [25], multiple directional lighting [26] and polarization effect [27] 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. Instead of only considering the radiometric response function, the in-camera processing can be modeled in more comprehensive pipelines [28–30], but these models require more complicated calibration procedures as well.

There are existing works that perform radiometric calibration for Internet photos. Kuthirummal *et al.* [31] 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 nonlinear 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 [8] and MVS [9, 10]. Diaz and Sturm [32, 33] 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 [6] 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 [2], which could potentially be applied to Internet photos.

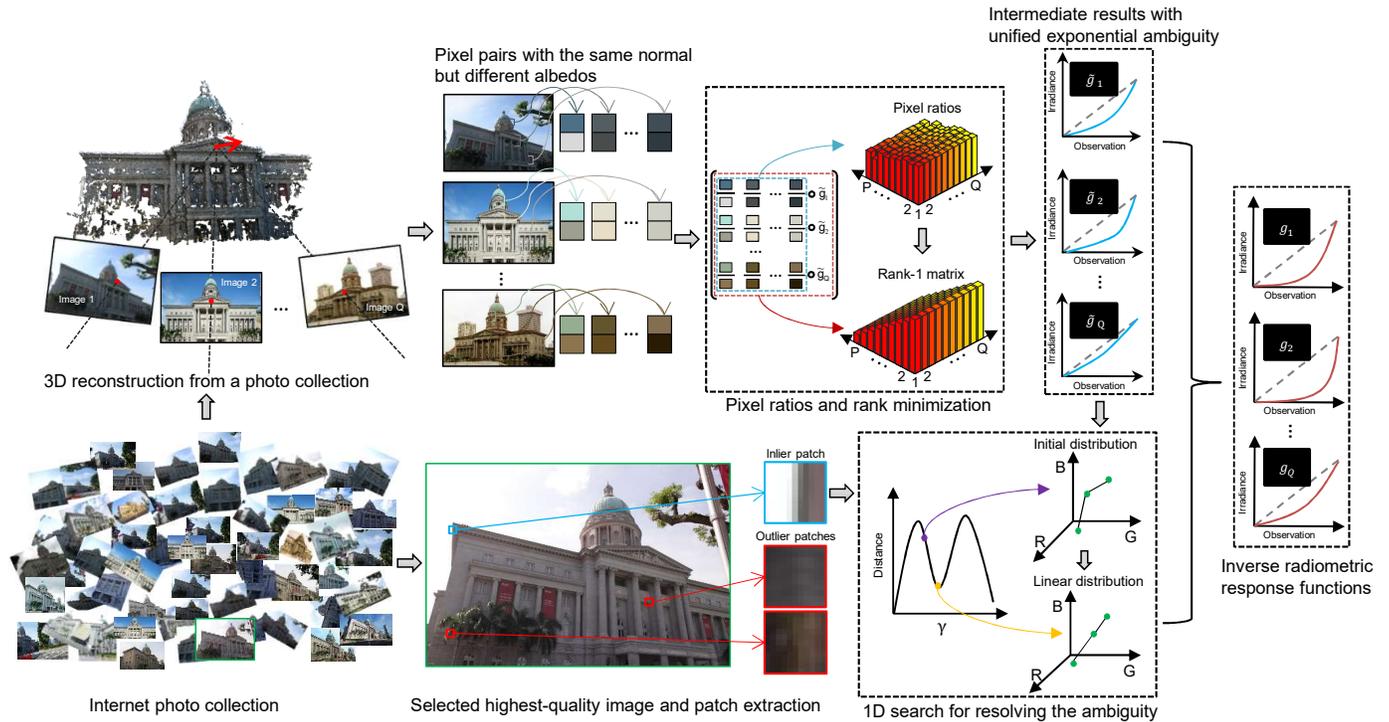


Fig. 1. Pipeline of our method. We estimate the (inverse) radiometric response functions for each nonlinear image  $\{\tilde{g}_1, \tilde{g}_2, \dots, \tilde{g}_Q\}$  by rank minimization over the stacks of pixel pairs up to a unified exponential ambiguity. The  $\circ$  operator applies each intermediate inverse response function  $\tilde{g}$  to both the numerator and denominator of ratio terms in the same row. The correct  $\tilde{g}$  transforms each row of the matrix to the same vector (up to a scale) to make the matrix rank-1. Then, the nonlinear image with highest quality from the photo collection is selected to extract patches. Only inlier patches are kept to find edge color triplets, among which the triplet with linear distribution tells the correct value of  $\gamma$  for the unified exponential ambiguity above. Finally, we obtain inverse radiometric response functions for all photos free from ambiguity as  $\{g_1, g_2, \dots, g_Q\}$ .

## 2.2 A unified framework.

We found many methods mentioned in Sec. 2.1 fall into a unified framework and summarize some representative methods in a coherent manner in Table 1 to compare their similarities and differences.

**Unified representation:** The unified representation summarizes the common form used in different radiometric calibration methods. The equation  $g(B_m) = K_m R$  and  $g(B_n) = K_n R$  indicate that the correct inverse response function transforms the observation values to a vector  $R$  multiplied by a scalar  $K$  where  $R$  encodes some intrinsic information of the scene and  $K$  is a camera-dependent factor. By taking the ratio between two linearized images,  $R$  is canceled out so that the ratio  $K_m/K_n$  connects the left and right side of the equation as  $g(B_m)/g(B_n) = K_m/K_n$ , where  $g$  is encoded on the left side. Different methods use different terms for  $K$  and  $R$ . Such ratio constraints could be used to build nonlinear or linear optimization problems, or solved by rank minimization. Since both sides of the equation are ratios, one can take both ratios to an arbitrary power and keep the equality, i.e.,  $g(B_m)/g(B_n) = K_m/K_n$  implies that  $g^\gamma(B_m)/g^\gamma(B_n) = (K_m/K_n)^\gamma$  for any  $\gamma$ . So all methods adopt such a representation suffer from the exponential ambiguity, and additional constraints have to be introduced to remove such an ambiguity.

**Different exposure times:** Given a sequence of nonlinear images with different exposure times,  $R$  is vectorized scene radiance for all pixels, and  $K$  is the exposure time. Since all nonlinear images are naturally aligned, millions of pixels in  $R$  provide more than enough equations than unknowns. The problem can be solved using nonlinear optimization [1, 3] or linear least squares [2]

or rank minimization [4, 5]. The exact exposure time ratios are required to remove the exponential ambiguity.

**Local smoothness:** A typical example is the method in [6] where  $K$  is a constant for a small patch of pixels. The problem can be solved by solving linear equations. The exponential ambiguity should be removed by using a radiometrically calibrated image (“example” image as mentioned in [6] with  $K = 1$ ).

**Photometric image formation:** The 3D points sharing the same surface normal and receiving the same amount of lighting are used to compose  $R$ , which are further scaled by different albedos  $K$ . For single-view method [7], the nonlinear images are naturally aligned and surface normals are estimated by appearance clustering. It solves linear equations in the log-space for only one response function per image set, but an alternative rank minimization formulation can be similarly applied [4]. [7] requires additional constraint, such as true albedo values or linear edge color blending, to resolve the ambiguity.

## 2.3 Computer vision meets Internet photos

Various computer vision problems could be extended to deal with Internet photos. Successful applications include scene completion [34], virtual tourism [35], weather estimation [36], composing pictures from sketches [37], image restoration [38], generating face animations [39], image colorization [40], intrinsic image decomposition [11], color consistency [41, 42], 3D reconstruction using multi-view stereo [43], and photometric stereo [12], 3D face reconstruction [44], synthesizing time-lapse video [45], 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 the 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 nonlinear 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  $\rho_i$  as

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

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

$$R_{ij} = \rho_i (\mathbf{n}_i^T \bar{\mathbf{I}}_j), \quad (2)$$

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

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., *RGB*) can be represented as

$$I_{ij}^k = c_j^k t_j \rho_i^k (\mathbf{n}_i^T \bar{\mathbf{I}}_j), \quad (3)$$

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

$$B_{ij}^k = f_j^k(I_{ij}^k) = f_j^k(c_j^k t_j \rho_i^k (\mathbf{n}_i^T \bar{\mathbf{I}}_j)). \quad (4)$$

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

$$g_j^k(B_{ij}^k) = c_j^k t_j \rho_i^k (\mathbf{n}_i^T \bar{\mathbf{I}}_j). \quad (5)$$

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 the image-dependent scaling factor, then Eq. (5) is simplified as

$$g_j(B_{ij}) = q_j \rho_i (\mathbf{n}_i^T \bar{\mathbf{I}}_j). \quad (6)$$

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

## 4 RADIOMETRIC CALIBRATION METHOD

Our method first identifies pixel pairs with same normal and different albedos across nonlinear images using 3D information obtained from 3D reconstruction (SfM [8] and MVS [10]). By stacking these pixel pairs into a matrix, we are able to solve all the inverse response functions up to the same ambiguity via rank minimization. The ambiguity is further removed by assuming linear color edge blending [13]. The above operation and process is illustrated in Fig. 1.

### 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 nonlinear 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  $\bar{\mathbf{I}}$ , and their albedo values be  $\rho_m$  and  $\rho_n$  ( $\rho_m \neq \rho_n$ ) respectively. Substituting these two points into Eq. (6) and taking the ratio between them, we obtain

$$\frac{g_j(B_{mj})}{g_j(B_{nj})} = \frac{q_j \rho_m (\mathbf{n}^T \bar{\mathbf{I}})}{q_j \rho_n (\mathbf{n}^T \bar{\mathbf{I}})} = \frac{\rho_m}{\rho_n}. \quad (7)$$

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 [1]. Recent progress in radiometric calibration shows that the rank minimization could solve such a problem in a more robust manner and effectively avoid overfitting [4]. 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 nonlinear images, or equivalently pairs of lighting conditions and cameras. We arrange these pixels according to the ratio format of Eq. (7) and stack them as the following matrix:

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

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

$$\{\tilde{g}_1, \tilde{g}_2, \dots, \tilde{g}_Q\} = \underset{\{g_1, g_2, \dots, g_Q\}}{\operatorname{argmin}} \operatorname{rank}(\mathbf{A}). \quad (9)$$

Similar to existing radiometric calibration methods [1–5, 7], the results generated by directly optimizing Eq. (9) also suffer from the exponential ambiguity. This is because for a set of optimized solution  $\tilde{g}$ ,  $\tilde{g}^\gamma$  for any unknown  $\gamma$  also keeps the ratio consistent in Eq. (7) and makes  $\mathbf{A}$  rank-1. Note there only exists one  $\gamma$  for all nonlinear images. In the following to subsections, we

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

will first introduce how to solve the problem up to the  $\gamma$  ambiguity, and then resolve the ambiguity by incorporating the linear edge color blending constraint [13].

## 4.2 Solution up to unified exponential ambiguity

We solve the above rank minimization using a similar approach as in [4], which is represented by the condition number as

$$\{\tilde{g}_1, \tilde{g}_2, \dots, \tilde{g}_Q\} = \operatorname{argmin}_{\{g_1, g_2, \dots, g_Q\}} \frac{\sigma_2(\mathbf{A})}{\sigma_1(\mathbf{A})}, \quad (10)$$

where  $\{\tilde{g}_1, \tilde{g}_2, \dots, \tilde{g}_Q\}$  denote the set of intermediate results with unified exponential ambiguity and  $\sigma_i(\mathbf{A})$  is the  $i$ -th singular value of  $\mathbf{A}$ .

We choose to use the polynomial representation for  $g$  as suggested by [4]. 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

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

where  $\{p_1, p_2, \dots, p_{S-1}\}$  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 [4] to solve Eq. (10). In fact, our problem is much more challenging than [4] 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 that directly solving such a problem like [4] 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 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). This section is summarized in Algorithm 1.

## 4.3 Ambiguity-free solution

Algorithm 1 successfully unifies all intermediate results of inverse response functions up to the same exponential ambiguity. Now we only need to solve the unified exponential ambiguity, and we adopt the linear edge color blending constraint in [13].

**Linear edge color blending.** To be self-contained, we briefly summarize the definition of linear edge color blending here<sup>2</sup>.

2. Please refer to [13] for further details.

### Algorithm 1 Solution up to unified exponential ambiguity

- 1: **INPUT:** Input nonlinear images, with pixels selected and stacked as the matrix of Eq. (8).
- 2: // *Pairwise optimization:*
- 3: **for** all pairwise combinations using two rows of  $\mathbf{A}$  **do**
- 4:   Solve for two  $g$  using Eq. (10);
- 5:   Apply these two  $g$  to the corresponding rows in  $\mathbf{A}$ ;
- 6: **end for**
- 7: Select  $g_m^0$  and  $g_n^0$  that make corresponding rows of  $\mathbf{A}$  have the minimum difference;
- 8: **for**  $k = \{1, 2, \dots, Q\} \wedge k \neq \{m, n\}$  **do**
- 9:   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;
- 10: **end for**
- 11: // *Global refinement:*
- 12: Solve Eq. (10) for all  $g$  simultaneously using  $\{g_1^0, g_2^0, \dots, g_Q^0\}$  as initial values;
- 13: **OUTPUT:** Intermediate results with unified ambiguity  $\{\tilde{g}_1, \tilde{g}_2, \dots, \tilde{g}_Q\}$ .

Suppose that we have a rectangular local patch  $S(x)$  centered at  $x$  and  $x$  is an edge pixel that separates  $S(x)$  into two regions  $S_1(x)$  and  $S_2(x)$  whose radiance values are  $I_1(\lambda)$  and  $I_2(\lambda)$ . The radiance  $I$  for edge pixel  $x$  can be represented as:

$$\begin{aligned} \int_{a \in S(x)} I(a, \lambda) da &= \int_{a \in S_1(x)} I_1(\lambda) da + \int_{a \in S_2(x)} I_2(\lambda) da \\ &= \alpha I_1(\lambda) + (1 - \alpha) I_2(\lambda), \end{aligned} \quad (12)$$

where  $\alpha = \int_{a \in S(x)} da$  and  $\lambda$  is the light wavelength.

Eq. (12) indicates that the radiances for edge pixels are the linear combination of radiances of two distinct regions  $S_1$  and  $S_2$  in the patch. Any nonlinear response function will damage this property and make observed brightness  $B$ ,  $B_1$ ,  $B_2$  not a linearly weighted summation like Eq. (12). This is called linear edge color blending constraint as proposed in [13] and it will be adapted to our problem.

**Image quality vs. linear edge color blending.** We define an edge color triplet extracted from a local patch  $S(x)$  as  $T = \{M_1, M_2, M_x\}$  where  $M_1 = \{I_1^R, I_1^G, I_1^B\}$  and  $M_2 = \{I_2^R, I_2^G, I_2^B\}$  are the *RGB* radiance in each region and  $M_x$  is the *RGB* radiance of an edge pixel. In a noise-free environment,  $T$  forms a straight line in the *RGB* space, as illustrated at bottom right of Fig. 1. So the method in [13] works well if high-quality nonlinear images are captured using a high-end DLSR under good lighting conditions. Such a condition is hardly satisfied for nonlinear images in an Internet photo collection, where compressed and degraded nonlinear images are commonly uploaded.

We illustrate how compression distorts the linear distribution of an edge color triplet in Fig. 2. We use the Matlab built-in function “imwrite” to compress a raw image<sup>3</sup> to different levels by setting the “quality” parameter to  $\{100, 75, 50, 25, 10\}$  sequentially with 100 meaning lossless compression. As demonstrated in Fig. 2, the linear blending property of edge color triplet becomes harder to observe with an increased compression rate, as stronger compression blurs the edge color by fusing pixels along the edge. This is one of the main reasons we cannot directly

3. Collected from a professional photo sharing website: <https://www.preview.com/>.

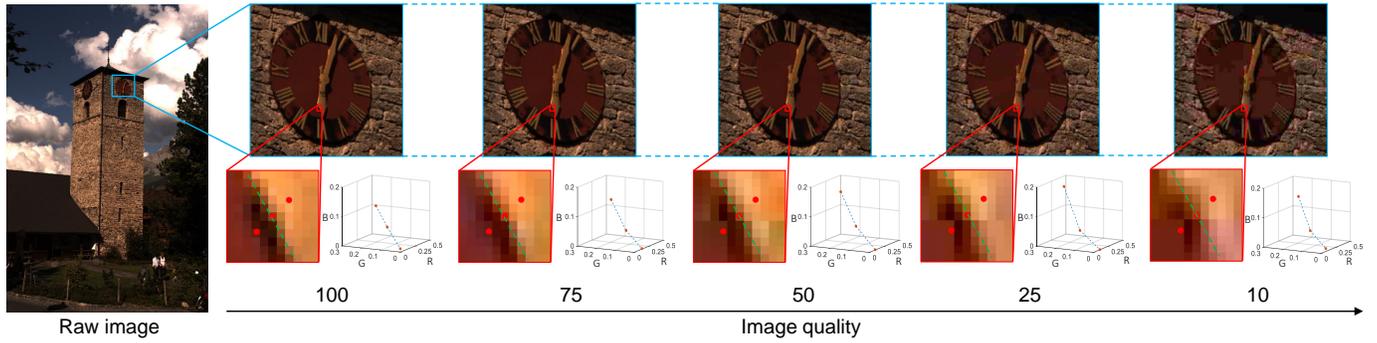


Fig. 2. Compression (a raw image compressed at different levels) distorts the linear edge color blending property. Close-up views of part of the compressed image (in blue box; multiplied by 6 for better visualization) are demonstrated in the top row. The image patches below (in red box; multiplied by 6 for better visualization) show the distributions of edge color triplets in the  $RGB$  space (the three red dots are pixels of an edge color triplet with the green line as the corresponding edge).

### Algorithm 2 Ambiguity-free solution

- 1: **INPUT:** Input nonlinear images and intermediate results with unified exponential ambiguity  $\{\tilde{g}_1, \tilde{g}_2, \dots, \tilde{g}_Q\}$  from Algorithm 1.
- 2: // Image selection and edge color triplets extraction:
- 3: **for**  $k = \{1, 2, \dots, Q\}$  **do**
- 4:     Calculate per-pixel storage space;
- 5: **end for**
- 6: Return a nonlinear image  $I_{HQ}$  with the biggest per-pixel storage space;
- 7: Extract edge color triplets  $\{T_1, T_2, \dots, T_T\}$  from  $I_{HQ}$ ;
- 8: // Coarse search (treat  $RGB$  as the same):
- 9: Use  $R$  channel as base and set  $\{\tilde{g}_R, \tilde{g}_G, \tilde{g}_B\} = \{\tilde{g}_R, \tilde{g}_R, \tilde{g}_R\}$ ;
- 10: **for**  $t = \{1, 2, \dots, T\}$  **do**
- 11:     1-D search of  $\gamma$  that minimizes Eq. (14);
- 12:     **if** A non-degenerated local min.  $\gamma^*$  (Figure 3 left) exists **then**
- 13:          $t$  is a reliable edge color triplet;
- 14:     **end if**
- 15: **end for**
- 16: Return  $\gamma^*$  as the average across all reliable triplets;
- 17: // Fine search (update each channel independently):
- 18: Set  $\gamma_R^* = \gamma^*$ ,  $\tilde{g}_R^* = \tilde{g}_R^{\gamma_R^*}$ ;
- 19:     Set  $\gamma_B^* = \operatorname{argmin}_{\tilde{\gamma}} \|\tilde{g}_B^{\tilde{\gamma}} - \tilde{g}_R^*\|$  and  $\gamma_G^* = \operatorname{argmin}_{\tilde{\gamma}} \|\tilde{g}_G^{\tilde{\gamma}} - \tilde{g}_R^*\|$ ;
- 20: **while** not converged **do**
- 21:     Fix  $\gamma_R^*$  and  $\gamma_B^*$  and perform 1-D search of  $\gamma_R^* \in [\gamma_R^* - \Delta, \gamma_R^* + \Delta]$  that minimizes Eq. (14);
- 22:     Fix  $\gamma_R^*$  and  $\gamma_B^*$  and perform 1-D search of  $\gamma_G^* \in [\gamma_G^* - \Delta, \gamma_G^* + \Delta]$  that minimizes Eq. (14);
- 23:     Fix  $\gamma_R^*$  and  $\gamma_G^*$  and perform 1-D search of  $\gamma_B^* \in [\gamma_B^* - \Delta, \gamma_B^* + \Delta]$  that minimizes Eq. (14);
- 24: **end while**
- 25: Return  $\{\gamma_R^*, \gamma_G^*, \gamma_B^*\}$ ;
- 26: **OUTPUT:** Inverse response functions  $\{g_1, g_2, \dots, g_Q\}$ .

apply [13] to Internet photo collections for each nonlinear image one by one.

Thanks to the problem definition of working with multiple nonlinear images and the unified ambiguity, we can avoid the distortion due to compression by using higher-quality nonlinear images in the photo collection whose linear edge color blending properties are better preserved. We simply assume the quality of

edge color triplet is proportional to the image quality represented by per-pixel storage space (the size of the compressed image divided by the number of pixels in the image). We then order nonlinear images in the photo collection according to their per-pixel storage space and choose the largest one for the following computation.

**Edge color triplet extraction.** Given the highest-quality nonlinear image, we extract edge color triplets using the same method as in [13]. Given a triplet  $T = \{M_1, M_2, M_x\}$ , we define the distance function similarly to [13] as:

$$D(T) = \frac{|[g(M_1) - g(M_2)] \times [g(M_x) - g(M_2)]|}{|g(M_1) - g(M_2)|}. \quad (13)$$

Substituting the intermediate results from Algorithm 1 into the distance function, we can rewrite Eq. (13) as:

$$D(T) = \frac{|[\tilde{g}^\gamma(M_1) - \tilde{g}^\gamma(M_2)] \times [\tilde{g}^\gamma(M_x) - \tilde{g}^\gamma(M_2)]|}{|\tilde{g}^\gamma(M_1) - \tilde{g}^\gamma(M_2)|}, \quad (14)$$

where  $\tilde{g} = \{\tilde{g}_R, \tilde{g}_G, \tilde{g}_B\}$  are the intermediate results in  $RGB$  channels and  $\gamma = \{\gamma_R, \gamma_G, \gamma_B\}$ <sup>4</sup> are the solution of exponential ambiguities in  $RGB$  channels, *i.e.*,  $\tilde{g}_R^{\gamma_R} = g_R$ ,  $\tilde{g}_G^{\gamma_G} = g_G$ , and  $\tilde{g}_B^{\gamma_B} = g_B$ , respectively.

Our problem is different from [13] that needs to solve all parameters in response functions from scratch, and we are dealing with a better constrained problem with only three unknowns ( $\gamma_R, \gamma_G$ , and  $\gamma_B$ ). So we only need several highly reliable edge color triplets to ensure the robustness against noise in Internet images, while the original problem in [13] needs much more triplets that are expected to cover the whole intensity range. Therefore, instead of minimizing the sum of distances of all triplets as in [13], we evaluate each edge color triplet independently using the distance calculated from Eq. (14) to select the most reliable estimates, as we will introduce in the next step.

**Search for  $\gamma$  using reliable triplets.** We assume that inverse response functions are the same for  $RGB$  channels for now and use one channel as base to start the search. In this case the search space are reduced from three-dimensional to one-dimensional and we can easily identify reliable and discard unreliable triplets by analyzing the shape of distance functions as shown in Fig. 3. For a reliable edge color triplet, a clear local minimum point

4. The ambiguity for each channel across nonlinear images in the photo collections are unified, but ambiguities in different channels can be different.

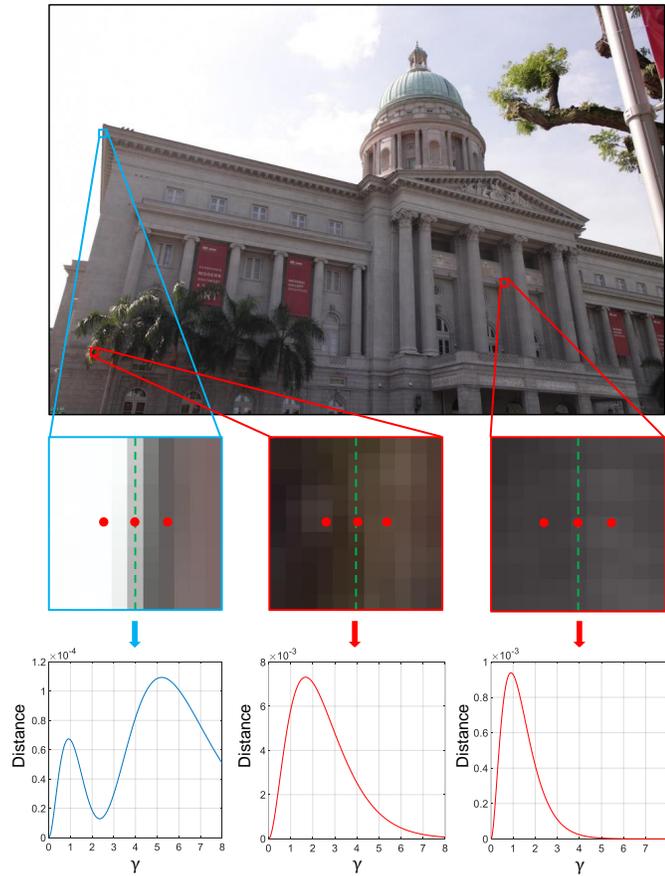


Fig. 3. Reliable edge color triplet is marked with the blue box and unreliable ones are marked with red boxes, and their corresponding distance values (according to Equation (14)) varying with the searching of  $\gamma$  are plotted in the bottom row. The reliable triplet observes a local minimum not appearing at the extreme values.

can be found in the middle of its distance function curve while unreliable triplets do not show distance curves with such a shape. In extreme condition, when the ambiguity term  $\gamma$  is set to 0 or  $+\infty$ , the disambiguated inverse response functions will map observed colors to extreme radiance colors as  $(0, 0, 0)$  or  $(1, 1, 1)$ . Such extreme values also make the cost of Eq. (14) approach zero, which should be discarded. Without losing generality, we use the average of estimated  $\gamma$  obtained from all reliable triplets as initial guess of  $\gamma_R$  and set initial guess of  $\gamma_G$  and  $\gamma_B$  by applying  $\gamma_R$  to disambiguate the intermediate results in  $G$  and  $B$  channels.

To refine the initial guess in  $RGB$  channels, we perform an iterative update. In each iteration there will be three rounds of one-dimensional search in  $RGB$  channels independently, and in each round of search one channel will be refined while the other two channels remain unchanged. Given the initial guess, the fine search is performed centered at current  $\gamma$  value within  $[\gamma - \Delta, \gamma + \Delta]$  ( $\Delta = 0.5$  in our experiments) to find an updated  $\gamma$  that minimizes Eq. (14). We set current  $\gamma$  to its initial value if the extreme values are encountered. The iteration is stopped when  $\gamma$  is no longer changed (usually after two or three iterations). The complete solution to resolve the exponential ambiguity is summarized in Algorithm 2.

#### 4.4 Implementation details

**3D reconstruction.** To build the matrix in Eq. (8), we need to extract corresponding pixels in all nonlinear images that have the same surface normal, under the same lighting condition, but with

different albedos. We first perform 3D reconstruction (SfM [8] 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, due to the noise in real data brought by 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 nonlinear 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 nonlinear 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 a divide-and-conquer strategy gives a more stable solution, and this property potentially allows the parallel processing of large amount data.

We use a 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 nonlinear images in the pairwise optimization step. A monotonicity constraint is added to penalize non-monotonic estimates similarly as adopted in [4]. 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 [3], but they are rarely observed in common digital cameras.

For coarse search, we set the range to  $[0, 5]$  and the step size to 0.1, and the step size is set to 0.01 for fine search. The current implementation of our method takes about 5 minutes for image size of  $2000 \times 1334$  with an unoptimized Matlab implementation running on a single core of i7 8700 processor (4.3 GHz) and we only need to search the unified ambiguity once for one set of Internet photo collection.

**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 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  $\mathbf{n}^T \mathbf{l}$  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.

Another degenerate case of our method is the same as in [13]. When the distribution of extracted edge color triplets lies along the  $R = G = B$  line and sensor response functions are the same in three color channels, the distribution of edge color triplets will always be linear, thus they will not provide any useful cue for solving the ambiguity.

## 5 EXPERIMENTAL RESULTS

We conduct 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 Verification for Algorithm 1

We first evaluate Algorithm 1 using synthetic data. Since Algorithm 1 can only solve all the inverse response functions up to the same exponential ambiguity, we remove such ambiguity by directly aligning our results with ground truth.

**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,  $1 \times 2$  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 [3] to generate our observation values and the RMSE and disparity are summarized in Fig. 4.

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. 4, 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 trade-off between accuracy and 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 [4, 46], 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 Eqs. (9)-(11) in [4] for the noise model representation. We perform 20 trials and each test contains 10 randomly selected response functions from DoRF with  $4 \times 6$  pixel pairs.

		Polynomial order					
		3	4	5	6	7	8
# pixel pairs	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

		Polynomial order					
		3	4	5	6	7	8
# pixel pairs	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

Fig. 4. The average RMSE/disparity w.r.t. the number of pixel pairs (row-wise) and order of polynomials (column-wise). “Red” means larger and “blue” means smaller errors.

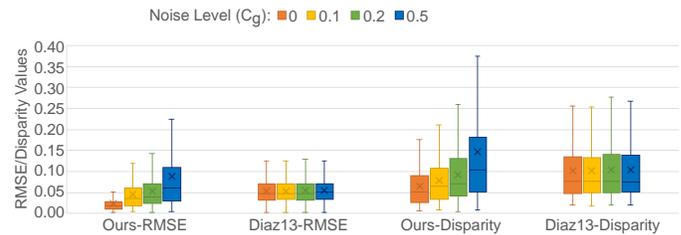


Fig. 5. Evaluation under various noise levels and comparison between our method and Diaz13 [33]. The box-and-whisker plot shows the mean (indicated as “x”), median, the first and third quartile, and the minimum and maximum values for RMSE and disparity for 200 ( $20 \times 10$ ) estimated response functions.

We evaluate another photometric image formation based method [33] (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 [33] and use this as the stable performance of Diaz13 [33] 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. 5. Our method outperforms Diaz13 [33] 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 [33] requires the dense reconstruction for lighting estimation, while we can only work on a few selected pixel pairs.

**Results on single-view images.** A single-view synthetic test is performed to provide an intuitive example with quantitative analysis, which is free of errors from 3D reconstruction. We use the data from [11], which is generated using a physics-based renderer. Given the ground truth reflectance and shading images, as show in the upper part of Fig. 6, 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

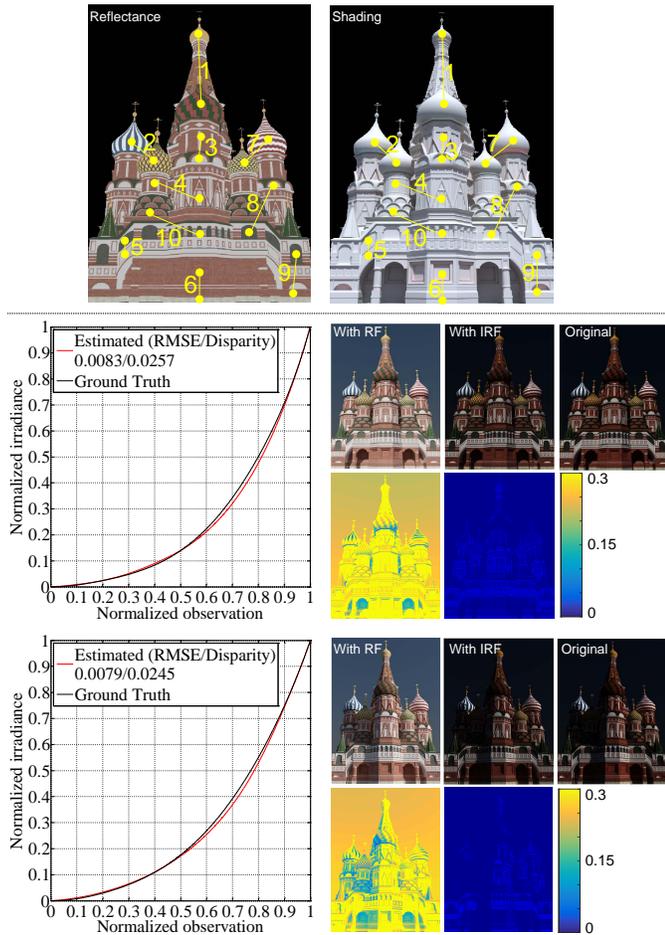


Fig. 6. Radiometric calibration results using a synthetic dataset. The upper row shows the ground truth reflectance, shading images, and the ten selected pixel pairs (yellow numbers) with the same normal but different albedo values. Two example results of the estimated inverse response functions and 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.

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. 6. 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 Verification for Algorithm 2

We use similar data as being tested in Figure 2, which compress raw images to different levels, for this verification. Such professionally captured images provide clean observation of linear edge color blending. Randomly selected response functions from DoRF [3] are added to compressed images to introduce the nonlinearity. We then obtain intermediate results from Algorithm 1 by manually selecting pixel pairs on the color checker and resolve the exponential ambiguity using Algorithm 2. We compare our method with a single image calibration method [13]<sup>5</sup> (denoted as “Lin04”). The quantitative results are shown in Fig. 7.

5. Implemented by Jean-François Lalonde and available at: <https://github.com/jflalonde/radiometricCalibration>

As we can observe from the results, “Lin04” [13] works well on non-compressed image (quality set to 100), but it is very sensitive to image quality degradation as shown by the increasing errors in calibration results. Our method outperforms “Lin04” [13] in all situations mainly because we are able to discard unreliable edge color triplets distorted by compression and only use the reliable ones for optimization. While our method shows tolerance to certain levels of compression, for Internet photo collection, we only choose one highest-quality image for resolving the unified ambiguity to ensure the most stable solution.

## 5.3 Performance on Internet photo collections

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. 8, 44 for the dataset used in Fig. 9, and 31 for the dataset used in Fig. 10. 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 [4]. 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. After applying Algorithm 1 and obtain intermediate results which share the same exponential ambiguity, we first use the calibrated response functions to remove the exponential ambiguity for all intermediate results, to provide upper bound references for our method, before evaluating the ambiguity-free solution. The exponential ambiguity is then solved using Algorithm 2. In addition to comparing with “Lin04” [13], we also run our own implementation of “Diaz13” [33] on the same dataset. This time the lighting coefficients are also part of the optimization which are initialized randomly, since they cannot be fixed to ground truth as using the synthetic data.

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 all three scenes are plotted in Fig. 8, Fig. 9, and Fig. 10, respectively. All our estimates (both the intermediate results aligning to the ground truth and the ambiguity-free solution) show closer shapes of estimated inverse response functions (with smaller RMSE/Disparity) to the calibrated ground truth than other methods, which indicates that our method is able to jointly estimate the inverse response functions for a set of Internet images reliably.

By a closer check between the results generated using our ambiguity-free pipeline (Algorithm 1 and Algorithm 2) and the results generated by aligning our intermediate results (Algorithm 1) to ground truth, the average RMSE between the two results for three datasets are 0.0053, 0.0078, and 0.0029, respectively, which proves the effectiveness of Algorithm 2 in resolving exponential ambiguities.

“Lin04” [13] tends to generate unstable results because their method uses all extracted edge color triplets which contain many triplets violating their assumptions due to compressed and degraded images from the Internet. For “Diaz13” [33], the unstable estimation mainly comes from joint estimation of too many unknown parameters.

To further investigate the reason for poor performance of “Lin04” [13] on Internet photo collections, we conducted an experiment by providing only reliable color triplets to “Lin04” [13]. Images with inverse response functions that are obtained

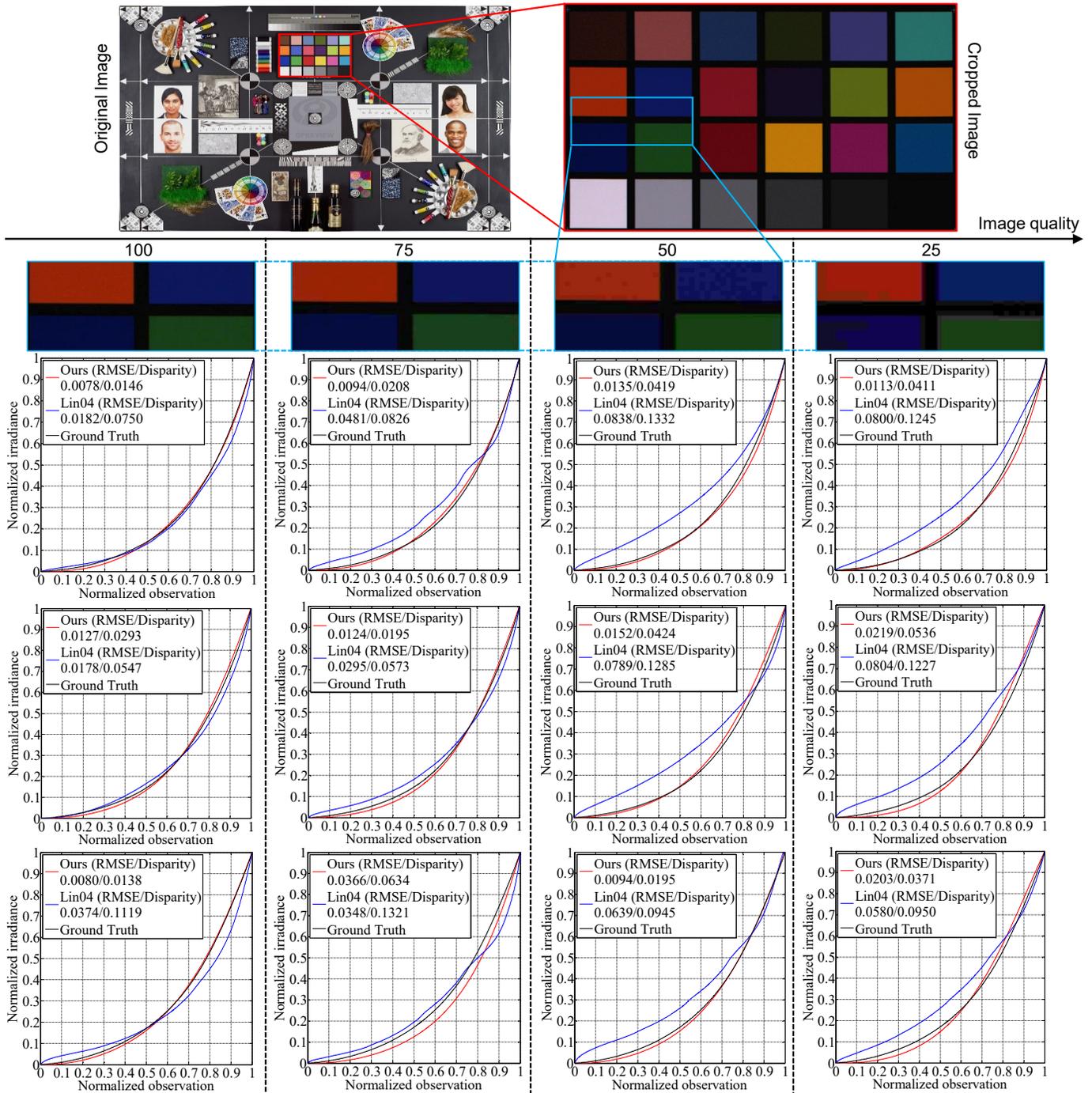


Fig. 7. Comparison of our method and “Lin04” [13] with different image qualities. Three examples (applied with different response functions in  $R$  channel) are shown for each image quality level in each column. The RMSE and Disparity are in the legend of each plot.

by the multiple exposure approach [4] are used for quantitative evaluation. We first compress high-quality images to different levels using Matlab built-in function “imwrite” so that both our method and “Lin04” [13] can be evaluated under different image qualities. Note that this operation only changes image qualities and the response functions are not affected. Then Algorithm 1 is performed to obtain intermediate results. Incorporating these intermediate results, Algorithm 2 is partially performed until Line 15 to obtain reliable color triplets. The selected reliable color triplets are then used to resolve the inverse response function using “Lin04” [13]. We show the results in Fig. 11. With unreliable color triplets discarded, “Lin04” [13] produces comparable results to ours on

high-quality images (Fig. 11 left), but its performance greatly degrades with the image quality getting worse (Fig. 11 middle and right), since lower-quality images have fewer reliable color triplets along edges. This is the main reason that “Lin04” [13] cannot be reliably applied to Internet images where compressed images widely exist. Instead, our method utilizes relationship of images in a set, thus performs much more robust than calibrating each single image one by one.

#### 5.4 Radiometric alignment

Finally, we show how our calibration results benefit radiometric alignment [47] using Internet images. We first apply the estimated

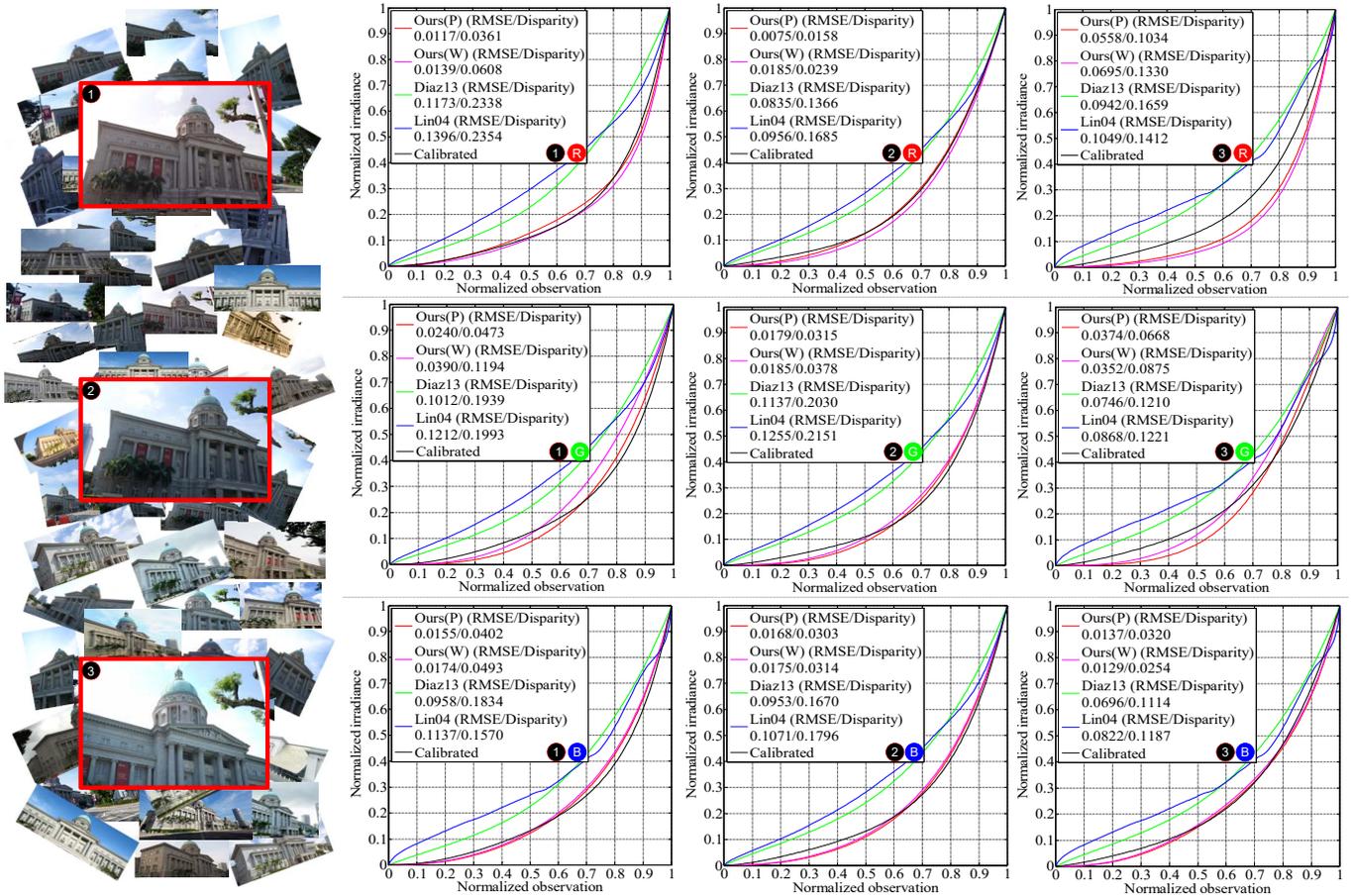


Fig. 8. Estimated inverse radiometric response functions (denoted with “W”, Algorithm 1 and Algorithm 2) using an image collection mixed with Internet photos and captured images (in red box). The results compared with the intermediate results generated by Algorithm 1 (aligned with ground truth and denoted with “P”), “Diaz13” [33] and “Lin04” [13] on *RGB* channels of three images and calibrated ground truth are shown. The RMSE and Disparity are in the legend of each plot.

inverse response function to each nonlinear image to linearize the image intensities. The linearized images may still show different color appearance due to different and unknown exposure time and white balance settings for Internet images. To visualize the effect of radiometric alignment, we warp all images to the same reference viewpoint using the method in [12], and a linear radiometric alignment is further conducted on *RGB* channels independently to get rid of the influence from different exposure time and white balance settings. The results are shown in Fig. 12. The image color appearances look rather consistent after removing the nonlinear response function using our method, and the remaining differences are mainly caused by different illumination (shading and shadow).

We use RMS intensity difference as error metric [47] to evaluate the radiometric alignment results. The average error dropped from 0.1165 (without applying inverse response function) to 0.0558 (with our estimated inverse response function applied) on the first dataset and from 0.1919 to 0.0897 on the second dataset.

## 6 DISCUSSION

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. We further combine our intermediate results with linearity of edge color blending to resolve the exponential ambiguity, which is a common problem for radiometric calibration methods using ratio constraints. The effectiveness of the proposed method is verified quantitatively using both synthetic and real-world data.

Currently, we need to assume the 3D reconstruction is sufficiently reliable for extracting points with the same surface normal under the same visibility condition. With radiometric calibration problem solved, it will be interesting to combine with photometric 3D reconstruction using Internet photos [12] to further improve the quality of 3D reconstruction. Current method requires at least one high-quality image in the dataset, so it cannot handle a photo collection with all images containing severe noise. An even more robust solution that compensates the noise magnification issue caused by ratio operation may help further improve the accuracy of the proposed method. Explicitly inferring the white balance and exposure time settings of Internet photos is another interesting future topic, which could provide additional information to unorganized Internet photo collections. Our method could also be applied to calibrate multi-spectral cameras and images by extending the linearity constraint from *RGB* space to a multi-dimensional space, except for a degenerate case where the spectral channels contain all-zero observations.

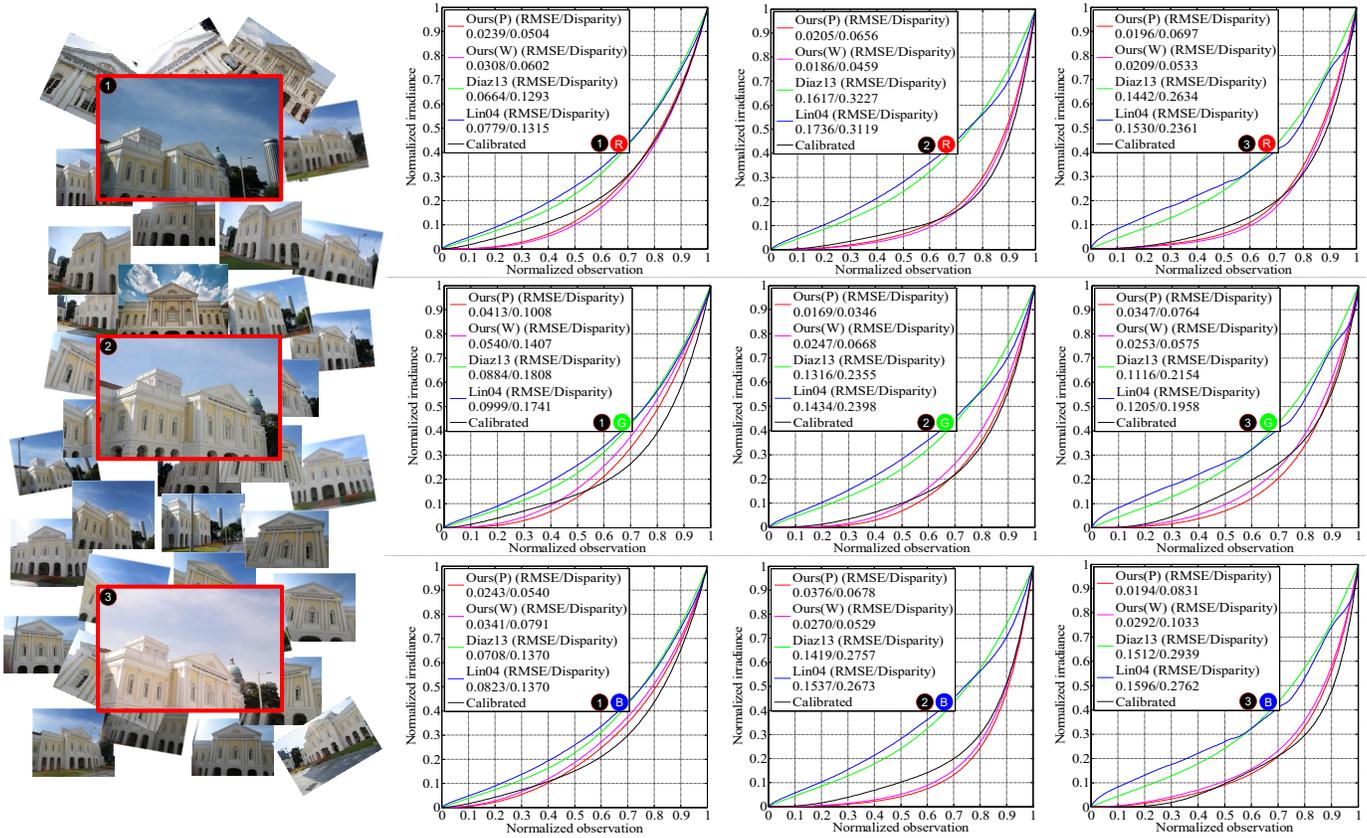


Fig. 9. Estimated inverse radiometric response functions (denoted with "W", Algorithm 1 and Algorithm 2) using an image collection mixed with Internet photos and captured images (in red box). The results compared with the intermediate results generated by Algorithm 1 (aligned with ground truth and denoted with "P"), "Diaz13" [33] and "Lin04" [13] on RGB channels of three images and calibrated ground truth are shown. The RMSE and Disparity are in the legend of each plot.

### ACKNOWLEDGMENT

We thank Joon-Young Lee for providing the code of [4]. This research is supported in part by National Science Foundation of China under Grant No. 61872012 and Singapore MOE Academic Research Fund MOE2016-T2-2-154. Sai-Kit Yeung is supported by an internal grant from HKUST (R9429). Yasuyuki Matsushita is supported by JSPS KAKENHI Grant Number JP16H01732.

### REFERENCES

- [1] T. Mitsunaga and S. K. Nayar. Radiometric self-calibration. In *Proc. of Computer Vision and Pattern Recognition*, 1999. 1, 2, 3, 4
- [2] S. J. Kim and M. Pollefeys. Radiometric self-alignment of image sequences. In *Proc. of Computer Vision and Pattern Recognition*, 2004. 2, 3
- [3] 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, 3, 7, 8, 9
- [4] 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, 7, 8, 9, 10, 12
- [5] A. Badki, N. K. Kalantari, and P. Sen. Robust radiometric calibration for dynamic scenes in the wild. In *Proc. of International Conference on Computational Photography*, 2015. 2, 3, 4
- [6] 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, 3
- [7] 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, 3, 4

- [8] 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, 4, 7
- [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, 4, 7
- [11] 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, 3, 8
- [12] 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, 3, 11
- [13] 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. 1, 2, 4, 5, 6, 8, 9, 10, 11, 12, 13
- [14] Y. Matsushita and S. Lin. Radiometric calibration from noise distributions. In *Proc. of Computer Vision and Pattern Recognition*, 2007. 1, 2
- [15] Z. Mo, B. Shi, S.-K. Yeung, and Y. Matsushita. Radiometric calibration for internet photo collections. In *Proc. of Computer Vision and Pattern Recognition*, 2017. 2
- [16] 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
- [17] 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
- [18] 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

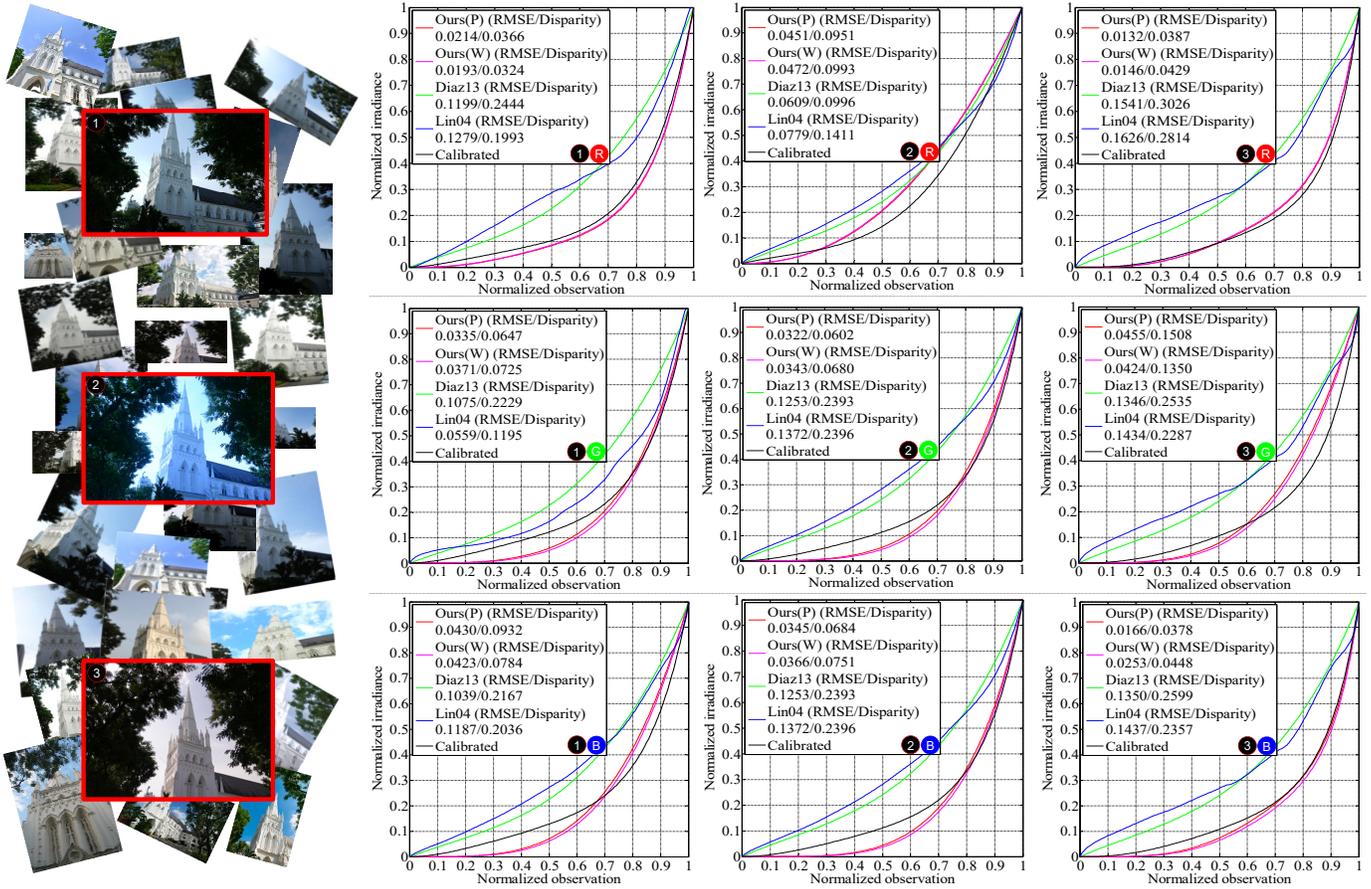


Fig. 10. Estimated inverse radiometric response functions (denoted with “W”, Algorithm 1 and Algorithm 2) using an image collection mixed with Internet photos and captured images (in red box). The results compared with the intermediate results generated by Algorithm 1 (aligned with ground truth and denoted with “P”), “Diaz13” [33] and “Lin04” [13] on *RGB* channels of three images and calibrated ground truth are shown. The RMSE and Disparity are in the legend of each plot.

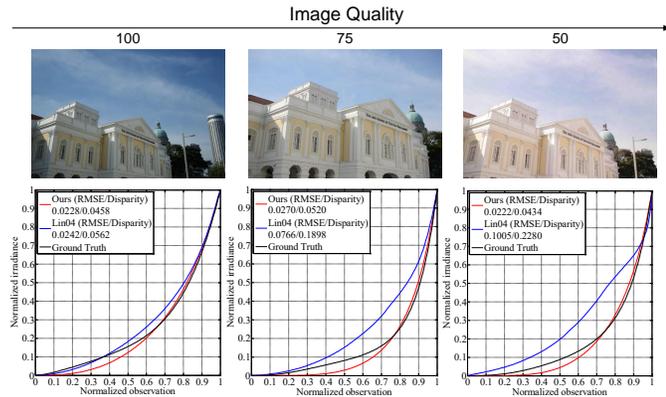


Fig. 11. Comparison between our method and “Lin04” [13] using reliable color triplets. The RMSE and Disparity are in the legend of each plot.

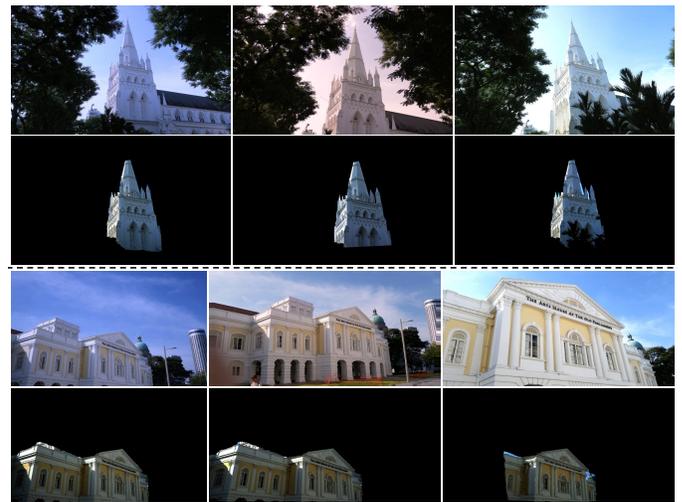


Fig. 12. Radiometric alignment using Internet images. The first and third row are the input images. The second and fourth row are the radiometrically aligned images.

[19] C. Li, S. Lin, K. Zhou, and K. Ikeuchi. Radiometric calibration from faces in images. In *Proc. of Computer Vision and Pattern Recognition*, 2017. 2

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

[21] M. D. Grossberg and S. K. Nayar. Determining the camera response from images: What is knowable? *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(11):1455–1467, 2003. 2

[22] Y. Tsin, V. Ramesh, and T. Kanade. Statistical calibration of CCD imaging process. In *Proc. of International Conference on*

*Computer Vision*, 2001. 2

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

[24] S. J. Kim and M. Pollefeys. Robust radiometric calibration and vignetting correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(4):562–576, 2008. 2

[25] P. Rodrigues and J. P. Barreto. Single-image estimation of the

camera response function in near-lighting. In *Proc. of Computer Vision and Pattern Recognition*, pp. 1028–1036, 2015. 2

[26] 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

[27] D. Teo, B. Shi, Y. Zheng, and S.-K. Yeung. Self-calibrating polarising radiometric calibration. In *Proc. of Computer Vision and Pattern Recognition*, 2018. 2

[28] A. Chakrabarti, D. Scharstein, and T. E. Zickler. An empirical camera model for internet color vision. In *Proc. of British Machine Vision Conference*, 2009. 2

[29] H. Lin, S. J. Kim, S. Süsstrunk, and M. S. Brown. Revisiting radiometric calibration for color computer vision. In *Proc. of International Conference on Computer Vision*, 2011.

[30] S. J. Kim, H. T. Lin, Z. Lu, S. Süsstrunk, S. Lin, and M. S. Brown. A new in-camera imaging model for color computer vision and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(12):2289–2302, 2012. 2

[31] 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

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

[33] M. Diaz and P. Sturm. Estimating photometric properties from image collections. *Journal of Mathematical Imaging and Vision* 43:93–107, 2013. 2, 8, 9, 11, 12, 13

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

[35] 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. 3

[36] L. Shen and P. Tan. Photometric stereo and weather estimation using internet images. In *Proc. of Computer Vision and Pattern Recognition*, 2009. 3

[37] 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. 3

[38] K. Dale, M. K. Johnson, K. Sunkavalli, W. Matusik, and H. Pfister. Image restoration using online photo collections. In *Proc. of International Conference on Computer Vision*, 2009. 3

[39] 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. 3

[40] 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. 3

[41] 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. 3

[42] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski. Optimizing color consistency in photo collections. *ACM Transactions on Graphics (Proc. of ACM SIGGRAPH)* 32(4):38, 2013. 3

[43] Q. Shan, R. Adams, B. Curless, Y. Furukawa, and S. M. Seitz. The visual turing test for scene reconstruction. In *Proc. of International Conference on 3D Vision*, 2013. 3

[44] J. Roth, Y. Tong, and X. Liu. Adaptive 3d face reconstruction from unconstrained photo collections. In *Proc. of Computer Vision and Pattern Recognition*, 2016. 3

[45] 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. 3

[46] 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. 8

[47] S. Kim and M. Pollefeys. Radiometric self-alignment of image

sequences. In *Proc. of Computer Vision and Pattern Recognition*, 2004. 10, 11



**Zhipeng Mo** received his B.E. degree from Tianjin University in 2014. He is currently pursuing his Ph.D. degree at Singapore University of Technology and Design, under the supervision of Prof. Sai-Kit Yeung and Prof. Boxin Shi. His research interest is photometric methods in computer vision.



**Boxin Shi** is currently an Assistant Professor (1000 Youth Talents Professorship) at Peking University, where he leads the Camera Intelligence Group. Before joining PKU, he did post-doctoral research at MIT Media Lab, Singapore University of Technology and Design, Nanyang Technological University from 2013 to 2016, and worked as a Researcher at the National Institute of Advanced Industrial Science and Technology from 2016 to 2017. He received the B.E. degree from Beijing University of Posts and Telecommunications in 2007, M.E. degree from Peking University in 2010, and Ph.D. degree from the University of Tokyo in 2013. He won the Best Paper Runner-up award at International Conference on Computational Photography 2015. He served/is serving as Area Chair for ACCV 2018, MVA 2019, 2017.



**Sai-Kit Yeung** is currently an Associate Professor at the Division of Integrative Systems and Design (ISD) at the Hong Kong University of Science and Technology (HKUST). Before joining HKUST, he was an Assistant Professor at the Singapore University of Technology and Design (SUTD) and founded the Vision, Graphics and Computational Design (VGD) Group. During his time at SUTD, he was also a Visiting Assistant Professor at Stanford University and MIT. Prior to that, he had been a Postdoctoral Scholar in the Department of Mathematics, University of California, Los Angeles (UCLA). He was a visiting student at the Image Processing Research Group at UCLA in 2008 and the Image Sciences Institute, University Medical Center Utrecht, the Netherlands in 2007. He received his PhD degree in Electronic and Computer Engineering, MPhil degree in Bioengineering, and BEng degree (First Class Honors) from HKUST, respectively in 2009, 2005 and 2003.



**Yasuyuki Matsushita** received his B.S., M.S. and Ph.D. degrees in EECS from the University of Tokyo in 1998, 2000, and 2003, respectively. From April 2003 to March 2015, he was with Visual Computing group at Microsoft Research Asia. In April 2015, he joined Osaka University as a professor. His research area includes computer vision, machine learning and optimization. He is on the editorial board of IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), International Journal of Computer Vision (IJCV), IPSJ Journal of Computer Vision and Applications (CVA), The Visual Computer Journal, and Encyclopedia of Computer Vision. He served/is serving as a Program Co-Chair of PSIVT 2010, 3DIMPVT 2011, ACCV 2012, ICCV 2017, and a General Co-Chair for ACCV 2014. He is a senior member of IEEE.