PAPER

Auto-Radiometric Calibration in Photometric Stereo

Wiennat MONGKULMANN^{†a)}, Nonmember, Takahiro OKABE^{††b)}, and Yoichi SATO^{†c)}, Members

SUMMARY We propose a framework to perform auto-radiometric calibration in photometric stereo methods to estimate surface orientations of an object from a sequence of images taken using a radiometrically uncalibrated camera under varying illumination conditions. Our proposed framework allows the simultaneous estimation of surface normals and radiometric responses, and as a result can avoid cumbersome and time-consuming radiometric calibration. The key idea of our framework is to use the consistency between the irradiance values converted from pixel values by using the inverse response function and those computed from the surface normals. Consequently, a linear optimization problem is formulated to estimate the surface normals and the response function simultaneously. Finally, experiments on both synthetic and real images demonstrate that our framework enables photometric stereo methods to accurately estimate surface normals even when the images are captured using cameras with unknown and nonlinear response functions.

key words: photometric stereo, surface orientations, radiometric calibration, radiometric response function, specularity

1. Introduction

Photometric stereo is a method for estimating the surface orientations of a static object from a set of images taken from a fixed viewpoint but under different illumination conditions. The classic Woodham's photometric stereo method [1] assumes that surfaces follow the Lambertian reflection model, and the directions and intensities of directional light sources are known a priori. These assumptions are difficult to achieve outside the laboratory and prevent photometric stereo from being used outside the computer vision community. Thus, current research trends in photometric stereo emphasize relaxing such assumptions; *i.e.*, generalizing materials and light sources.

Another assumption that is usually taken for granted is that images are captured using a camera with a linear radiometric response function; *i.e.*, pixel values in the images are linearly proportional to irradiance values. Unfortunately, this assumption is not always true because response functions in consumer cameras are often nonlinear to enhance image quality visually, *e.g.*, compensating for a nonlinear

mapping of a display, and simulating traditional films response. Moreover, these nonlinear relationships vary among camera models and manufacturers. Therefore, cumbersome and time-consuming radiometric calibration is required to cancel the effect of the nonlinear relationships so that irradiance values can be subsumed from pixel values.

We propose an auto-radiometric calibration framework for photometric stereo techniques to estimate surface orientations from a sequence of radiometrically uncalibrated images. The main idea is to use the consistency of irradiance; the irradiance values converted from pixel values by using the inverse response function (IRF) must be consistent with the irradiance values computed by the corresponding surface normals and reflection property, specifically the Lambertian reflection model. In other words, we exploit the cues inherent in the image formation process to estimate the IRF as well as surface normals.

The preliminary version of our framework [2] represents an IRF as a linear combination of basis functions and uses diffuse pixels to estimate the surface normals and the coefficients of the basis functions. We show that the estimation of surface normals and the coefficients results in a linear least-square problem with linear constraints. In this work, we extend our method to estimate surface normals of non-Lambertian surfaces. Specular highlights observed in many real-world surfaces do not follow the Lambertian reflection model, and those pixels distort the estimated surface normals as well as the estimated IRF. We also introduce prior knowledge derived from a real-world response function database into our IRF estimation in this work to avoid unrealistic IRFs. Our experiments show that our framework allows surface normals to be estimated from a set of radiometrically uncalibrated images regardless of nonlinearity of a response function.

The contribution of this study is an auto-radiometric calibration framework for photometric stereo methods. This framework requires no additional images for radiometric calibration. Therefore, it allows one to estimate surface orientations from images taken using consumer cameras without additional effort to calibrate nonlinearity of radiometric response functions.

This paper is structured as follows: We discuss previous studies on radiometric calibration, photometric stereo methods, and relationships to our method in Sect. 2. Our auto-radiometric calibration method for Lambertian surfaces is explained in Sect. 3. We illustrate an integration of our auto-radiometric calibration framework and a photomet-

Manuscript received April 7, 2016.

Manuscript revised August 19, 2016.

Manuscript publicized September 1, 2016.

[†]The authors are with Institute of Industrial Science, The University of Tokyo, Tokyo, 153–8505 Japan.

^{††}The author is with Kyushu Institute of Technology, Iizuka-shi, 820–8502 Japan.

a) E-mail: wiennat@iis.u-tokyo.ac.jp

b) E-mail: okabe@ai.kyutech.ac.jp

c) E-mail: ysato@iis.u-tokyo.ac.jp

DOI: 10.1587/transinf.2016EDP7150

ric stereo for non-Lambertian surfaces technique in Sect. 4. We discuss the experiment results on both Lambertian and non-Lambertian objects in Sect. 5 and 6. Finally, conclusions and future work are presented in Sect. 7.

2. Related Work

In this section, we briefly review previous works on radiometric camera calibration and how they are related to our present work. Originally, one can use an image of a calibration target with known reflectances. The response function can be estimated from the relationship between the reflectances of the calibration target and corresponding pixel values in the image.

An alternative approach uses multiple images of a static scene taken with varying exposure times [3]–[5]. Manning and Picard [3] used a regression method to estimate parameters of a response function represented by a gamma correction curve. Debevec and Malik [4] used a non-parametric model to approximate a response function with smoothness property. Mitsunaga and Nayar [5] proposed a method that represents a response function with a linear combination of polynomials. This method can accurately estimate a response function even when only rough estimates of exposure times are available.

Recent works in radiometric calibration have fewer restrictions in input images, *e.g.*, requiring only a single input image, or allowing camera movement. Lin *et al.* [6] proposed a method for estimating a radiometric response function from a single color image. Color mixtures in edge regions lie on a straight line in the RGB space. A nonlinear response function bends this line into a nonlinear curve. The IRF can be estimated as a function that bends the nonlinear curve back into a straight line. For grayscale images, one can use distribution of intensity mixtures along edge boundaries instead of using color mixtures [7]. Wilburn *et al.* [8] conducted radiometric calibration on the basis of motion blur in a single image. They made use of temporal irradiance mixtures instead of spatial ones. Kim and Pollefeys [9] estimated a radiometric response function from an image sequence captured using a moving camera. Because irradiance values observed at a certain point should not change when the camera moves, the changes in pixel values can be used to compute the changes in exposure times and the response function. Grossberg and Nayar [10] argued that radiometric calibration can be done using the pixel value histograms of two image frames without exact registration. Matsushita and Lin [11] suggested that noise distributions in uncalibrated images are not symmetry due to nonlinear response functions. They developed a calibration method to estimate an IRF that makes the noise distribution symmetry. Recently, Lee *et al.* [12] have formulated the radiometric calibration from multiple images with different exposure times as a rank minimization problem. This method is more robust to noise and can estimate response functions more accurately than the current methods.

Although we can conduct radiometric calibration by

using these existing methods separately, they are often cumbersome and require additional images used for radiometric calibration. Unlike these methods, our proposed method does not require additional images. More importantly, our method is based on a different cue; we take advantage of the image formation process for radiometric calibration.

Previously, radiometric calibration and photometric stereo have been studied separately. As far as we know, Shi *et al.* [13] was the first to use photometric stereo images for radiometric calibration. They calibrate input images for photometric stereo by using color profile linearity. A color profile is the RGB values of a surface point under varying illumination conditions. They observed that a color profile of a Lambertian point forms a straight line under the RGB space if the response function is linear. In contrast, color profiles become nonlinear curves if the response function is nonlinear. Therefore, the IRF can be estimated as a function that makes the nonlinear curve into a straight line. Although their method does not require additional images, the radiometric calibration step can be considered as a preprocessing, and requires nonlinear optimization. More importantly, their method cannot handle a certain class of radiometric response functions such as gamma correction curves as well as gray objects whose color profiles remain in straight lines even when the response function is nonlinear.

Recently, Abram *et al.* [14] have incorporated the idea of auto-radiometric calibration [2] into outdoor photometric stereo method. They formulated a nonlinear optimization problem to simultaneously recover surface normals, exposures, and a response function from images captured with a webcam over many months. Unfortunately, both of those methods [13], [14] assume that input images consist of only diffuse pixels, which is not always true.

There are a lot of techniques proposed to relax the assumption of surface materials which can be classified into two approaches: modeling approach and statistical approach. The modeling approach assumes target objects follows a specific reflection model, and estimates parameters of the reflectance model as well as surfaces. *E.g.*, [15] assume that surfaces are composed of multiple materials, each described with a parametric *Bidirectional reflectance distribution function (BRDF)* model. [16] uses Torrance-Sparrow model to estimate surface normals as well as the model parameters without knowing light source directions. In contrast, the statistical approach determine specular highlights as outliers and excludes them from the surface estimation. There are a lot of works that use linear models derived from Lambertian reflectance model [17]–[19] to determine and remove outliers from the estimation. Instead of removing pixels with specular components, [20] casts photometric stereo problem into a low-rank matrix completion and recovery technique problem to determine diffuse components in each pixel value and estimates surface normals from the diffuse components. [21] approximates reflectance of non-Lambertian surfaces with piecewise linear functions and then uses a sparse Bayesian learning-based technique to estimate surface from the diffuse components more accu-

ately.

However, these methods have one common assumption, the input images must be radiometrically calibrated. On the other hand, our framework can estimate surface normals of a non-Lambertian object and a radiometric response function simultaneously. In addition, it does not require color images and can handle gamma correction curves and gray objects.

3. Recovering Lambertian Surfaces

3.1 Photometric Stereo

We briefly summarize the classic photometric stereo method [1] in which radiometrically calibrated input images of Lambert surfaces illuminated by known light sources are assumed. Let us consider a surface observed under orthographic projection and light sources far from the surface, then the viewing direction and light source directions are constant across the surface.

Let us denote the irradiance value at the p -th pixel ($p = 1, 2, \dots, P$) under the d -th light source ($d = 1, 2, 3, \dots, D$) by $E_{p,d}$. Assuming that the surface is Lambertian, the irradiance value $E_{p,d}$ can be described as

$$E_{p,d} = \mathbf{n}_p^T \mathbf{l}_d, \quad (1)$$

where $\mathbf{n}_p$ is the surface normal at the p -th pixel scaled by its albedo and $\mathbf{l}_d$ is the direction of the d -th light source scaled by its intensity.

The classic photometric stereo method estimates the scaled surface normal $\mathbf{n}_p$ from the irradiance value $E_{p,d}$ with known light sources $\mathbf{l}_d$. Because a scaled surface normal $\mathbf{n}_p$ has three degrees of freedom, *i.e.*, two for a normal with unit length and one for an albedo, the surface normal can be estimated from at least three images.

Conventionally, the irradiance value $E_{p,d}$ in (1) is described in a matrix form:

$$\begin{pmatrix} E_{p,1} \\ \vdots \\ E_{p,D} \end{pmatrix} = \begin{pmatrix} l_{1,x} & l_{1,y} & l_{1,z} \\ \vdots & \vdots & \vdots \\ l_{D,x} & l_{D,y} & l_{D,z} \end{pmatrix} \begin{pmatrix} n_{p,x} \\ n_{p,y} \\ n_{p,z} \end{pmatrix}, \quad (2)$$

$$\mathbf{E}_p = \mathbf{L} \mathbf{n}_p,$$

where $\mathbf{l}_d = (l_{d,x}, l_{d,y}, l_{d,z})^T$ and $\mathbf{n}_p = (n_{p,x}, n_{p,y}, n_{p,z})^T$. The estimate of the scaled surface normal $\hat{\mathbf{n}}_p$ is given by the least-square method:

$$\hat{\mathbf{n}}_p = \arg \min_{\mathbf{n}_p} \sum_{d=1}^D (E_{p,d} - \mathbf{n}_p^T \mathbf{l}_d)^2. \quad (3)$$

The surface normal and albedo are computed from the estimated scaled surface normal $\hat{\mathbf{n}}_p$ as $\hat{\mathbf{n}}_p / |\hat{\mathbf{n}}_p|$ and $|\hat{\mathbf{n}}_p|$ respectively.

3.2 Radiometric Response Function

Suppose a radiometric response function f maps an irradiance value E into a pixel value I , *i.e.*, $I = f(E)$. Since the

radiometric response function is a monotonically increasing function, there is a unique inverse function $g = f^{-1}$ that maps an I to an E , *i.e.*, $g(I) = f^{-1}(I) = E$. Hereafter, we normalize the ranges of pixel values and irradiance values to $[0, 1]$ without loss of generality.

Assume an IRF g can be approximated as a linear combination of basis functions such as polynomials [5], or eigenvectors (or eigenfunctions in this context) of response curve data [22]. In this work, we use the K -parameters EMoR approximation [22] in which the basis functions are derived from Principle Component Analysis (PCA) on the real world response function database. The approximated IRF is in the form:

$$g(I) = g_0(I) + \sum_{k=1}^K c_k g_k(I), \quad (4)$$

subjects to the boundary conditions, $g(0) = 0$ and $g(1) = 1$. Here g_0 is the mean curve, and g_k is the k -th basis functions with their coefficients c_k .

3.3 Simultaneous Estimation

We propose a technique to estimate both the surface normals $\mathbf{n}_p$ ($p = 1, 2, \dots, P$) and coefficients of the IRF c_k ($k = 1, 2, \dots, K$) at the same time.

In a similar manner to (3), the simultaneous estimation of surface normals $\hat{\mathbf{n}}_p$ and IRF $\hat{g}$ can be estimated at the same time by minimizing the difference in irradiance values estimated from surface normals and IRF. We combine the PCA approximation (4), and (3) to obtain the estimates of the surface normals and coefficients of the IRF $\{\hat{c}_k\}$:

$$\{\hat{\mathbf{n}}_p, \hat{c}_k\} = \arg \min_{\{\mathbf{n}_p, c_k\}} \sum_{p=1}^P \sum_{d=1}^D \left[g_0(I_{p,d}) + \sum_{k=1}^K c_k g_k(I_{p,d}) - \mathbf{n}_p^T \mathbf{l}_d \right]^2, \quad (5)$$

subjects to the monotonicity constraints, which can be given as discrete derivatives of g as $g(i_{s+1}) - g(i_s) \geq 0$. The monotonicity constraints can be expressed as follows:

$$\begin{aligned} g(i_{s+1}) - g(i_s) &\geq 0, \\ \left(g_0(i_{s+1}) + \sum_{k=1}^K c_k g_k(i_{s+1}) \right) - \left(g_0(i_s) + \sum_{k=1}^K c_k g_k(i_s) \right) &\geq 0, \\ \sum_{k=1}^K (c_k g_k(i_{s+1}) - c_k g_k(i_s)) &\geq (g_0(i_s) - g_0(i_{s+1})), \end{aligned} \quad (6)$$

$\forall i_s \in I_S$, where $I_S = \{i_0, \dots, i_S\}$ is the dense sampling of pixel intensities in the range $[0, 1]$ in monotonically increasing order, *e.g.*, $I_S = \{0, 1/255, 2/255, \dots, 255/255\}$ for 8-bit images.

Thus, the simultaneous estimation of the surface normals and the IRF results in the linear least-square problem in (5) with the linear constraints in (6) given that the input pixel values distributed uniformly to cover the whole range

of pixel intensity [0, 1].

When input images are under-exposed, it is possible that foreground pixel values in the images do not cover the pixel intensity levels. We can use additional constraints such as smoothness, and integrability in order to avoid unrealistic IRFs as in [6], [11]. Let $P(\mathbf{c})$ a prior model of IRF coefficients constructed by fitting the probabilistic distribution of coefficients of IRF in DoRF to a multivariate Gaussian mixture model:

$$P(\mathbf{c}) = \sum_{m=1}^M \alpha_m \mathcal{N}(\mathbf{c}; \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m), \quad (7)$$

where $\mathcal{N}$ is a Gaussian distribution with mean $\boldsymbol{\mu}_m$, and covariance matrix $\boldsymbol{\Sigma}_m$. α_m is the weighting factor for the m -th distribution, and $\mathbf{c} = (c_1, \dots, c_k)^T$ are the coefficients of IRF. The model parameters $\boldsymbol{\mu}_m$, $\boldsymbol{\Sigma}_m$ and α_m can be obtained using EM algorithm or cross-entropy method [23]. Then, the prior term can be added to (5) as following:

$$\{\hat{\mathbf{n}}_p, \hat{\mathbf{c}}\} = \arg \min_{\{\mathbf{n}_p, \mathbf{c}\}} \left\{ \frac{1}{2PD} \sum_{p=1}^P \sum_{d=1}^D w(I_{p,d} | \mathbf{n}_p, \mathbf{c}) - \lambda \log(P(\mathbf{c})) \right\}, \quad (8)$$

where $w(I_{p,d} | \mathbf{n}_p, \mathbf{c}) = (g_0(I_{p,d}) + \sum_{k=1}^K c_k g_k(I_{p,d}) - \mathbf{n}_p^T \mathbf{l}_d)^2$, and λ is a regularization factor for the prior model term. Although arbitrary number of normal distributions can be used, we assume the number of normal distributions to $M = 5$ similar to the previous studies [6], [11]. We detect shadow pixels by thresholding and remove them from the summations in (3) and (8).

When the radiometric response function is linear, we can estimate a surface normal at each surface point independently, as shown in (2). On the other hand, when it is nonlinear, we cannot estimate each surface point independently as shown in (8). Therefore, the naive optimization of (8) subject to the constraints of (6) is computationally expensive when the number of pixels increases. To reduce the computational cost, we can estimate the IRF (and the surface normals) by using a small number of randomly selected pixels, then all the pixel values are converted to irradiance values with the estimated IRF, and finally the surface normals are estimated using (3).

Note that our simultaneous estimation also has degenerate cases; it fails to estimate an IRF if pixel values are not well distributed, *e.g.*, estimating an IRF from images of a plane which is illuminated by light sources rotated around its perpendicular axis so the pixel values remain constant across all images. In such scenes, the nonlinear relationship between irradiance and pixel values cannot be observed from the images; thus the IRF cannot be estimated.

4. Recovering Non-Lambertian Surfaces

We present our RANSAC-based framework to integrate photometric stereo for non-Lambertian surfaces methods into the simultaneous estimation of surfaces and IRF.

Since specular highlights observed in many real-world surfaces do not obey the Lambertian reflection model, including such pixels in the simultaneous estimation leads to distorted surfaces and response functions. Fortunately, many non-Lambertian surfaces behave similarly to Lambertian surfaces where specular highlights do not exist. Assuming that specular highlights can be observed only within limited angles, we can treat highlights as outliers that deviate from the Lambertian reflection model. Consequently, it would be possible to integrate a robust estimation technique based on RANSAC [24] into our framework to estimate the surfaces of a non-Lambertian object and the response function of a camera.

4.1 Estimation Framework for Non-Lambertian Surfaces

Our framework consists of two layers as shown in Fig. 1: an outer layer for IRF estimation, and an inner layer for surface estimation. The outer layer is a RANSAC loop to estimate a lot of candidate IRFs. With the estimated IRFs, the inner layer estimates surface normals using an existing photometric stereo method, and then we determine inliers which are diffuse pixel values that follow the Lambertian reflection model with respect to each estimated IRF. The inlier set with the maximum number of supporting inliers is considered as the consensus diffuse pixel values and we can use these pixel values to estimate the IRF and surfaces without being affected by specularly.

The rest of this section explains the detailed algorithm of our framework.

1) Randomly Selecting Pixels and Light Sources: Our method begins by randomly selecting pixel values for the IRF estimation using RANSAC technique. So we first randomly select one foreground pixel. For the pixel, there are 3 unknowns because a scaled surface normal has 3 degrees of freedom. In additional, there are K unknowns if we approximate an IRF with K basis functions. Therefore, we select $(3 + K)$ light sources and their corresponding pixel values for the simultaneous estimation.

2) Estimating IRF from Pixel Values of Selected Pixels Under Selected Light Sources: After the pixel values have been selected, we estimate an IRF that satisfies the pixel values with (8) and (6). Although one might argue that one of the selected pixel values is a specular highlight, assume that all selected pixel values are diffuse so the IRF can be estimated with no distortion from specularly. The goodness of the estimated IRF will be determined in a later step.

3) Converting All Pixel Values Into Irradiance Values: In this step, we convert the pixel values in the input images into irradiance values by using the estimated IRF.

4) Detecting Outliers such as Specular Pixels: In this step, we determine outlier irradiance values that violate Lambertian reflection property. We first estimate surface normals from the calibrated input images by using an photometric stereo method such as [19], or [21]. Assuming the images are correctly calibrated, the estimated surface normals have no distortion from both specularly and nonlinear response

Fig. 1 Overview of our framework for photometric stereo with auto-radiometric calibration. The outer layer is used for IRF estimation and the inner layer is used for surface estimation.

function. Then, we estimate irradiance values with (1) and compare them to the observed irradiance values. An observed irradiance value is a supporting inlier of the estimated surface normal if it is equal to the corresponding estimated irradiance value.

In practice, the estimates might contain small errors due to noise in the input images. We relax the equality constraint by introducing a threshold for the error. An observed irradiance value is considered as an inlier if the difference between the observed and estimated irradiance values is less than this threshold. Let $\tilde{\mathbf{n}}_p$ denote the estimated surface normal scaled by its albedo. Assuming that input images are contaminated with photon shot noise[†], the variance of noise is proportional to the irradiance value [25]. An observed irradiance value $E_{p,d}$ supports the estimated normal $\tilde{\mathbf{n}}_p$ if

$$(E_{p,d} - \tilde{\mathbf{n}}_p^T \mathbf{l}_d)^2 \leq \tau_s E_{p,d}, \quad (9)$$

where τ_s is a specified parameter.

5) Counting Number of Inliers: We then count the number of inliers to evaluate the goodness of the estimated IRFs. If the IRF is estimated without outlier, it must be consistent with all diffuse irradiance values in the images. In contrast, if the IRF is distorted, it will be consistent to only a few irradiance values. Therefore, the best IRF should maximize the number of supporting diffuse irradiance values.

6) Estimating Surface Normals and IRF from Largest Consensus Set: Up to now, we use only a few randomly selected pixel values to estimate an IRF. However, it is possible that one of the selected pixel values is a specular highlight. Therefore, we repeat the whole process multiple times to produce a set of candidate IRFs. Given the number of iterations is sufficiently large, there probably exists an iteration in which all selected pixel values are diffuse. Without distortion from specularity, the IRF should have largest supporting inlier set. The inliers in the largest inlier set are then determined as the consensus-diffuse pixel values.

Finally, we estimate the final IRF from the maximum

consensus-diffuse pixel values by using (8) and (6). Because the pixel values are all diffuse, the final IRF have no distortion from specularity and surface normals can be estimated using a photometric stereo method.

5. Experiments on Diffuse Objects

5.1 Experiments on Synthetic Images

We validated the surface and IRF estimation of our method with the experiments on synthetic images of a Lambertian sphere. The images were illuminated under 20 directional light sources whose directions were uniformly selected from a hemisphere. The uncalibrated images were obtained by applying nonlinear response functions from DoRF. We detected shadows in all images with thresholding, *i.e.*, a pixel value is considered in shadow if its intensity is less than $5/255$. Then, we discarded all pixels that consisted of only one or two non-shadow pixel values from the estimation.

We implemented the optimization of (8) and its constraint (6) by using a MATLAB's built-in function *fmincon*. The number of basis functions was fixed to $K = 4$. To make the estimation tractable, we first estimated an IRF from 50 randomly selected pixels to calibrate the images then estimated surface normals with (3). We used $\lambda = 0$ in the experiments with diffuse objects since pixel values in our synthetic input images cover over the whole range of the pixel intensities $[0, 1]$. Therefore, it is unnecessary to use the prior model in this experiment. In contrast, when there are no pixels whose pixel values are in a certain range, the estimated radiometric response function is not constrained sufficiently in that range, and as a result the estimation tends to be unstable without the prior term.

To evaluate surface estimation accuracy, we compared the surfaces estimated using our method (ours) to the surfaces estimated from radiometrically uncalibrated images with the classic photometric stereo (PS-LS) [1], and photometric stereo with sparse Bayesian learning (PS-SBL) [21] to illustrate the advantage over the state-of-the-art method.

[†]Poisson noise due to random arrivals of photons

Fig. 2 Samples of surface estimated from synthetic data sets and the comparison to the ground truth. The response function used to generate the images are specified over the results. (a) ground truth (b) examples of input images (c)-(e) Estimated surface normals and difference to the ground truth in degrees.

Table 1 Comparison of average of mean angular error and their SD of the surface normals estimated from the synthetic data sets using our proposed method (ours), photometric stereo with sparse Bayesian regression (PS-SBL), and the classic photometric stereo (PS-LS).

Method	Mean angular error (degrees)		Elapsed Time (secs.)
	Uncalib. images	Calib. images	
ours	0.68 ± 0.71	0.53 ± 0.23	2.18 ± 1.35
PS-SBL	4.42 ± 4.203	0.002 ± 0.003	27.83 ± 0.15
PS-LS	12.09 ± 3.11	0.0002 ± 0.004	0.39 ± 0.01

The parameters for PS-SBL were configured as suggested in the original paper, *i.e.*, $p = 3$, $\lambda_{sbl} = 10^{-2}$, $\sigma_a^2 = 10^{-2}$, and $\sigma_n^2 = 10^6$.

The mean angular error of the estimated surfaces are summarized in Table 1 and few examples of the estimated surfaces and their difference in degrees to the ground truth are shown in Fig. 2 as RGB color-coded surface normals along with their difference to the ground truth. It is clear that our auto-calibration framework compensates nonlinear response functions so the estimated surfaces are more similar to ground truth than one of those estimated using PS-LS and PS-SBL. Meanwhile, the nonlinear reflectance model used in PS-SBL compensates the nonlinearity of response functions, therefore the mean of angular errors of the estimated surfaces were significantly lower than those estimated by PS-LS.

We also conducted experiments when the images are radiometrically calibrated. Here, we fixed the number of piecewise reflectance function $p = 1$ for PS-SBL. The difference of the surfaces estimated by the proposed method from calibrated images is comparable to (or slightly smaller than) that for uncalibrated images. However, the surfaces estimated by PS-SBL and PS-LS have less error than those estimated from our proposed method. This is because the basis functions can not represent a linear function well. Therefore, the estimated IRFs had subtle deviation from the linear IRF. Those small errors eventually propagated to the estimated surfaces. In contrast, PS-LS and PS-SBL assume calibrated images and the input images were images of Lambertian surfaces without noise so that the only precision error become the source of error in the estimated surfaces.

Also, we evaluated the IRF estimation accuracy of our method to [5] with EMoR representation [22]. Note that we used images of a static scene with varying exposure times

Fig. 3 The IRFs estimated by *ours* from synthetic data sets. The dotted lines show ground truth and the solid lines show the estimated functions. The RMSE for each IRF are depicted over the graph.

to estimate IRF with [5], hence, indirect comparison. Here, we show some examples of the estimated IRFs fitted to their corresponding ground truth in Fig. 3. The root mean square error (RMSE) of the fitted IRFs are also shown in the figure. The average of RMSE for all synthetic images sets is 0.0134 while the RMSE of the IRFs estimated with [5] is 0.0098.

However, it is worth to mention that our method can estimate the IRFs accurately up to scale. When selected irradiance values do not well cover the entire range of irradiance values, there is few information to constrain the estimated IRF. Therefore it is possible to obtain multiple IRFs with different scales that satisfy the selected irradiance values. In our experiments on synthetic images, this scaling ambiguity did not affect the estimated IRF but affects the overall scale of the estimated albedos.

5.2 Experiments on Real Images

We evaluated the accuracy of our method through the experiments on images of two real-world objects: *sphere*, and *statue*. We used 20 light sources placed randomly over the objects and calculate the light source direction from images of a chrome sphere. Then we used images of a Lambertian sphere taken with a calibrated camera to calculate the light source intensities with (1). Instead of using calibrated images, ones can use a luminometer to directly measure the light source intensities.

We captured images of the objects with a Point Grey Flea 2 camera. Although the camera provides linear measurement of light intensity, we can configure its intensity lookup table so it acts like a nonlinear camera. We configured the lookup table with the measured agfapan-apx-100CD, agfa-scala-200xCDStandard, and gamma+2.2 nonlinear response functions from DoRF database [22]. These functions represent three common shapes of nonlinear functions in the database: concave, convex, and sigmoid. Since the ground truth of *sphere*, and *statue* were not available,

Fig. 4 The estimated surface normals *statue*. The name of response functions are specified in the top row and the method used to estimate the surfaces are specified in the bottom. (a) ground truth (b) sample input image (c)-(f) estimated surfaces and difference to the ground truth in degrees.

Fig. 5 The IRFs estimated from *statue* with [5] and ours compare to the ground truth. The RMSEs of the estimated IRFs are included in the legend.

we used the surfaces estimated from the radiometrically calibrated input images with PS-SBL [21] as ground truth.

The mean of angular error of the estimated surfaces are presented in Fig. 6. Similar to the experiments with the synthetic images, our auto-radiometric calibration compensated the nonlinearity of response functions so the estimated surfaces were more similar to the ground truth than those estimated by PS-LS and PS-SBL from the uncalibrated images. However, it is possible that shadow pixel values were raised by nonlinear response functions and became non-shadow in pixel value space. Eventually, those false pixel values were included in the surface estimation and deviated the surfaces. This can be observed in the side parts of the estimated *statue* (*agfapan-apx-100CD*) in Fig. 4. Moreover, noises in the input images also caused the estimated IRF to deviate from the ground truth IRF. The images calibrated with such IRFs were marginally different to the ground truth images so that the surfaces estimated from those images had small angular

Fig. 6 Mean angular error of the surfaces estimated from real images with ours, PS-SBL, and PS-LS methods. Note that the error for *sphere (calib.)* is near zero and surfaces estimated from PS-SBL are ground truth.

error over the surfaces.

We also compared the accuracy of the estimated IRFs to the ones estimated from the images of a static scene using radiometric calibration technique [5]. Instead of images with varying light sources, we captured the images of a static scene with different exposure times under the same response functions used for capturing the input images for our method. The IRFs estimated from both [5] and ours are shown in Fig. 5. The RMSE of the estimated IRFs show that our method can estimate IRFs accurately without any additional image for radiometric calibration.

6. Experiments on Specular Objects

6.1 Experiments on Synthetic Images

We conducted experiments on synthetic and real images of non-Lambertian surfaces to validate the estimation accuracy of our framework. We synthesized 20 images of a sphere under the assumptions of Torrance-Sparrow model [26] and directional light sources with the same intensities. The light directions were uniformly distributed over a hemisphere. The images were applied with response functions from DoRF to obtain uncalibrated images.

We implemented two variants of the inner layer of our framework with two different photometric stereo methods: a straightforward RANSAC-based method (ours-RS), and a sparse Bayesian learning-based method (ours-SBL) based on [21]. Unlike Mukaigawa's [19], we directly applied RANSAC to Lambertian photometric stereo to estimate surface normals. More specifically, we select three light directions and their corresponding pixel values to estimate surface normals with (3). Then the supporting inliers are determined as the observed irradiance values that matches ones predicted with the estimated surface normals. This process is repeated for many iterations and the support inliers of the iteration with maximum number of inliers are selected as consensus inliers. Then the final surface normals are estimated using only consensus inliers.

To calculate the RMSE of the IRFs, we discarded top ten percent of the brightest selected pixel values and evaluated the RMSE with the rest. This is because the number of bright diffuse pixel values was small due to specularly.

Fig. 7 The average of the estimated surfaces angular errors and RMSE of the IRFs with respect to the value of λ .

Therefore, the estimated IRF where the pixel value was near 1 could not be constrained well; thus, not accurate.

First, we observed the effect of λ to the estimated surfaces and IRF. Here we picked a small positive value to factor the prior term to match the magnitude of the surfaces estimation term in (8). More specifically, we used various λ to perform surfaces and IRFs estimation from the synthetic images generated with a selected set of response functions, then we selected the most appropriate value based on the estimation results.

Figure 7 shows that the value of $\lambda = 10^{-5}$ gave the best balance of the IRF and surface estimation accuracy. We found that the prior term started to dominate when $\lambda = 10^{-2}$ and the IRFs is overfitted to the prior model for larger λ . As the pixel values no longer follow the Lambertian model when the IRFs is overfitted, ours-SBL misclassified all foreground pixel values, *i.e.*, no result when $\lambda \geq 10^{-2}$. In contrast, the sampling process in ours-RS guarantees that there are enough at least 4 pixel values determined as inliers. Therefore, the surfaces could be estimated even when the IRFs were overfitted. The effect of the threshold τ_s on the surface and IRF estimation performance is shown in Fig. 8. A large τ_s leads to more false positive, *i.e.*, more outliers are misclassified as inliers. In contrast, a smaller τ_s causes less pixel values to be classified as inliers. Here we varied the value of τ_s and performed experiments on the synthetic data sets. We found that $\tau_s = 0.05$ gave the best balance between the classification accuracy, angular error, and RMSE of the estimated IRFs. Therefore, we fixed the threshold to $\tau_s = 0.05$ for the rest of the experiments. With the given λ and τ_s , the average RMSE of the IRFs estimated from ours, ours-RS, and ours-SBL for all synthetic data sets were 0.0128, 0.0156, and 0.0208 respectively. It took about 53 seconds for Ours-RS and 414 seconds for Ours-SBL to estimate an IRF and surface normals from a set of 20 images with 3,228 foreground pixels.

We compared the surface normals estimated using our proposed framework to those estimated from PS-SBL, and our method without outlier detection. Figure 10 shows an example of input image we used for the surface estimation and the estimation results. The bright areas shown in the error maps correspond to the specular highlight areas in the input images, *e.g.*, the specular highlights in the right-top area in Fig. 10(b) caused error on the same region in the

Fig. 8 Experimental results with respect to the value of τ_s .

Fig. 9 Comparison of mean angular error in degrees of the surfaces estimated from synthetic images using the specified methods. The error bar shows SD of the errors.

Fig. 10 Qualitative results from the synthetic specular sphere. (a) ground truth, (b) example of an input image from a synthetic images set, (c)-(j) estimated surfaces estimated using the specified methods. The bottom row shows angular differences to the ground truth in degrees.

estimated normal map. One can see that Ours-RS and Ours-SBL can remove the specular pixel values from the surface estimation so it reduces distortion caused by specular highlights on the estimated surfaces. We summarized the mean angular errors of the surfaces estimated from all data sets and their standard deviation in Fig. 9. The mean angular error of the surfaces estimated from ours, ours-RS, and ours-SBL are 1.32°, 0.99°, and 3.19° respectively.

We observed that ours-SBL produced surfaces and

IRFs with larger errors than the ones estimated by ours without outlier detection. This was due to scaling ambiguity that distorted the overall shape of the estimated IRF. When the outlier detection removes bright specular highlights, the range of pixel values is narrower so there is less information to constrain the IRF that are outside of the range of inliers values. Eventually, our method further minimized (8) by minimizing the scale of IRF while still satisfying both input pixel values and the boundary conditions. Without enough pixel intensities to constraint the whole range of IRF, the IRFs near the end of the range of inliers values are therefore incorrect. Since PS-SBL uses diffuse components in the pixel with specular highlight to achieve better results, Ours-SBL took the disrupted bright pixel values into account and it led to more angular errors. In contrast, Ours-RS was more capable to handle these kind of distortions since it aggressively discarded outliers such as the incorrect calibrated pixel values from the surface estimation, therefore specular highlights were removed more properly.

6.2 Experiments on Real Images

We validated our auto-radiometric framework using the experiments on real-world objects to observe the robustness of our framework when the objects do not follow Lambertian reflection model. We used four objects made from different materials; matte ceramic with painted smooth areas *seal*, polished ceramic *ghost*, opaque plastic *tomato*, and glossy-painted ceramic *fish*. The images were captured in the same lighting environment and nonlinear response functions used in the experiments with diffuse objects. With no ground truth available, we used surface normals estimated using PS-SBL from radiometrically calibrated images as ground truth.

We show the qualitative results of the experiments on *ghost* in Fig. 11. It can be seen that our framework compensated nonlinearity of the response functions and estimated surfaces with significantly less angular differences than the method without auto-radiometric calibration. Moreover, ours-RS and ours-SBL also removed specular highlights to estimate more reasonable surfaces as the estimated surfaces have significant less distortion from specularity. However, we still observed large errors along the concave regions due to inter-reflection and cast shadows.

Also, we noticed small errors scattered where specular highlights existed on the surfaces estimated by ours-SBL. This is due to the scaling ambiguity which distorts the IRFs near the end of intensities range. This error also can be observed on the experimental results from *seal* and *tomato* data set as well (Fig. 13). Although selecting optimal parameter values (ρ , σ_a^2 , and σ_n^2) for ours-SBL might substantially improve the estimation accuracy in this situation, it is highly scene-dependent hence a difficult problem.

The results for the experiments with calibrated images are illustrated in Fig. 11. As expected, PS-LS outperformed our auto-radiometric calibration when input images were calibrated despite of strong distortion from specularity. As our methods approximated the linear IRF with nonlinear

Fig. 11 Qualitative results from the data set *ghost*. The response functions used for capturing the images are specified over the results. (a) ground truth (b) sample of input images (c)-(j) the top row shows the normal maps estimated from the specified method and the bottom row shows the angular differences to the ground truth.

Fig. 12 The IRFs estimated from *ghost* with [5] and ours compare to the ground truth. The RMSEs of the estimated IRFs are included in the legend.

Fig. 13 Samples of qualitative results from the data set *seal* and *tomato*. The response functions used for capturing the images are specified over the results. (a) ground truth (b) sample of input images (c)-(j) the top row shows the normal maps estimated from the specified method and the bottom row shows the error maps.

Fig. 14 The quantitative results from the experiments with real images. (top) Mean angular error the surfaces estimated with specified methods. (bottom) RMSE of the IRFs agfapan-apx-100CD (agfapan), agfa-scala-200xCDStandard (agfascala), gamma+2.2, and linear function. The RMSE of the IRF estimated by [5] are 0.0034, 0.0032, 0.0117, and 0.0066 respectively.

functions, the estimated IRF slightly differed to a linear function so the auto-radiometrically calibrated images had marginal errors. Hence, the estimated surfaces were considerably different to the ground truth.

Figure 12 shows the estimated IRFs compare to their corresponding ground truth and IRF estimated with [5].

Here, the estimated IRFs are plotted up to the 90-percentile brightest pixel values used for the IRF estimation. The experimental results show that Ours, Our-RS, and Ours-SBL can estimate the IRFs reasonably accurate in the presence of specular highlights. Similar to the experiments with synthetic images, the RMSE of the IRFs estimated by Ours-RS and Ours-SBL were slightly higher than the ones estimated by Ours.

We show the quantitative results of the experiments in Fig. 14. While our auto-radiometric calibration framework reduced the distortion from nonlinear response functions, Ours-RS also successfully removed the outliers such as specular highlights, noise, and brightened shadows to achieve lower mean angular errors. However, we still observed that there are small errors in the estimated surfaces which were caused by overly removing pixel intensities that exceed the threshold in (9). These small errors could be observed over the estimated surfaces from Ours-RS: therefore, mean angular error remained relatively high, even though specular highlights were removed.

7. Conclusion

We proposed a framework for photometric stereo to recover surface normals from images captured using a camera with an unknown nonlinear response function. This uses the consistency of the irradiance value calculated from a reflection model and the irradiance value calculated from an IRF. With this framework, one can perform surface estimation from uncalibrated images without any additional images for radiometric calibration. Our framework can be integrated with an existing photometric stereo method to handle outliers such as specular pixels. Experiments show that our method can estimate the surface normals of the non-Lambertian surfaces more accurately than the existing methods when images are radiometrically uncalibrated.

There are two limitations in our method. First, it can estimate response functions and surface reflectance up to scale. Since our method has no constraint regarding neither scale of albedos nor scale of response function, the scale of albedo can propagate to the scale of response function and vice versa. Therefore, it cannot determine the correct scale of estimated IRF and albedos without additional cues. Second, our shadow thresholding works incorrectly if noisy pixels are greatly modified by the response function. From the experimental results, noisy pixels in shadow areas were amplified by the agfapan-apx-100CD response function, so that they exceeded the shadow threshold and became non-shadow in pixel-value space. Those non-shadow pixels were eventually included in the simultaneous estimation and caused distortion in such areas. Similarly, non-shadow pixels were darkened by gamma+2.2 function so they became shadows.

As for future work, extension for highly reflective materials, how to sample pixels for reducing computational cost while maintaining the accuracy of the estimation still remain to be addressed.

References

- [1] R.J. Woodham, "Photometric method for determining surface orientation from multiple images," *Optical Engineering*, vol.19, no.1, pp.139–144, 1980.
- [2] W. Mongkulmann, T. Okabe, and Y. Sato, "Photometric stereo with auto-radiometric calibration," *Proc. IEEE Workshop on Color and Photometry in Computer Vision*, pp.753–758, Nov. 2011.
- [3] S. Mann and R.W. Picard, "On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures," *Proc. IS&T's 48th Annu. Conf.*, pp.442–448, 1995.
- [4] P.E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," *Proc. 24th annual conference on Computer graphics and interactive techniques - SIGGRAPH '97*, pp.369–378, 1997.
- [5] T. Mitsunaga and S.K. Nayar, "Radiometric self calibration," *Proc. IEEE Conf. on CVPR*, pp.374–380, 1999.
- [6] S. Lin, J. Gu, S. Yamazaki, and H.-Y. Shum, "Radiometric calibration from a single image," *Proc. IEEE Conf. on CVPR*, pp.938–945, 2004.
- [7] S. Lin and L. Zhang, "Determining the radiometric response function from a single grayscale image," *Proc. IEEE Conf. on CVPR*, pp.66–73, 2005.
- [8] B. Wilburn, H. Xu, and Y. Matsushita, "Radiometric calibration using temporal irradiance mixtures," *Proc. IEEE Conf. on CVPR*, pp.1–7, 2008.
- [9] S.J. Kim and M. Pollefeys, "Radiometric alignment of image sequences," *Proc. IEEE Conf. on CVPR*, pp.645–651, 2004.
- [10] M.D. Grossberg and S.K. Nayar, "Determining the camera response from images: What is knowable?," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.25, no.11, pp.1455–1467, Nov. 2003.
- [11] Y. Matsushita and S. Lin, "Radiometric calibration from noise distributions," *Proc. IEEE Conf. on CVPR*, pp.1–8, 2007.
- [12] J.-Y. Lee, B. Shi, Y. Matsushita, I.-S. Kweon, and K. Ikeuchi, "Radiometric calibration by transform invariant low-rank structure," *Proc. IEEE Conf. on CVPR*, pp.2337–2344, 2011.
- [13] B. Shi, Y. Matsushita, Y. Wei, C. Xu, and P. Tan, "Self-calibrating photometric stereo," *Proc. IEEE Conf. on CVPR*, pp.1118–1125, 2010.
- [14] A. Abrams, C. Hawley, and R. Pless, "Heliometric stereo: Shape from sun position," *Proc. 12th European Conference on Computer Vision, ECCV'12, Berlin, Heidelberg*, vol.7573, pp.357–370, Springer-Verlag, 2012.
- [15] D.B. Goldman, B. Curless, A. Hertzmann, and S.M. Seitz, "Shape and spatially-varying brdfs from photometric stereo," *Proc. Int'l Conf. on Computer Vision*, vol.1, pp.341–348, 2005.
- [16] A. Georgiades, "Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo," *Proc. Int'l Conf. on Computer Vision*, vol.2, pp.816–823, 2003.
- [17] E.N. Coleman Jr. and R. Jain, "Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry," *Comput. Vision. Graph.*, vol.18, no.4, pp.309–328, 1982.
- [18] S. Barsky and M. Petrou, "The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.25, no.10, pp.1239–1252, Oct. 2003.
- [19] Y. Mukaigawa, Y. Ishii, and T. Shakunaga, "Analysis of photometric factors based on photometric linearization," *J. Opt. Soc. Am. A*, vol.24, no.10, pp.3326–3334, Oct. 2007.
- [20] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma, "Robust photometric stereo via low-rank matrix completion and recovery," *Proc. Asian Conference on Computer vision*, vol.6494, pp.703–717, Springer-Verlag, 2011.
- [21] S. Ikehata, D. Wipf, Y. Matsushita, and K. Aizawa, "Photometric stereo using sparse bayesian regression for general diffuse surfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.36, no.9, pp.1816–1831, Sept 2014.
- [22] M. Grossberg and S. Nayar, "What is the Space of Camera Response Functions?," *Proc. IEEE Conf. on CVPR*, pp.602–609, 2003.
- [23] Z. Botev and D.P. Kroese, "Global likelihood optimization via the cross-entropy method with an application to mixture models," *Proc., 2004 Winter, Simulation Conference*, pp.517–523, 2004.
- [24] M.A. Fischler and R.C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol.24, no.6, pp.381–395, June 1981.
- [25] N. Campbell, "Discontinuities in light emission," *Proc. Camb. Phil. Soc. Math. Phys. Sci.*, vol.15, pp.310–328, 1909.
- [26] K.E. Torrance and E.M. Sparrow, "Theory for off-specular reflection from roughened surfaces," *J. Opt. Soc. Am.*, vol.57, no.9, pp.1105–1112, Sep. 1967.

Wiennat Mongkulmann received his B.Eng. in computer engineering in 2008 from Chulalongkorn University, Thailand, and the M.S. degrees in information science and technology in 2013 from the University of Tokyo. He is working for Ph.D. at the University of Tokyo under supervision of Prof. Yoichi Sato. His research interests include photometry-based scene reconstruction, computational photography, and image processing.

Takahiro Okabe received the B.S. and M.S. degrees in physics, and the Ph.D. degree in information science and technology from the University of Tokyo, in 1997, 1999, and 2011 respectively. After working at the Institute of Industrial Science, the University of Tokyo, he joined Kyushu Institute of Technology, Japan, as an associate professor in 2013. His research interests include computer vision, computational photography, image processing, pattern recognition, and computer graphics, in particular their physical and mathematical aspects.

Yoichi Sato is a professor at Institute of Industrial Science, the University of Tokyo. He received his B.S. degree from the University of Tokyo in 1990, and his M.S. and Ph.D. degrees in robotics from School of Computer Science, Carnegie Mellon University in 1993 and 1997 respectively. His research interests include physics-based vision, reflectance analysis, image-based modeling and rendering, and gaze and gesture analysis. He served/is serving in several conference organization and journal editorial roles including IEEE Transactions on Pattern Analysis and Machine Intelligence, International Journal of Computer Vision, Computer Vision and Image Understanding, IET Computer Vision, IPSJ Journal of Computer Vision and Applications, ECCV2012 Program Co-Chair, ACCV2016 Program Co-Chair and MVA2013 General Chair.