Self-calibrating Photometric Stereo

Boxin Shi1∗ Yasuyuki Matsushita2 Yichen Wei2 Chao Xu1 Ping Tan3

1Key Lab of Machine Perception (MOE), Peking University 2Microsoft Research Asia 3Department of Electrical & Computer Engineering, National University of Singapore

Abstract

We present a self-calibrating photometric stereo method. From a set of images taken from a fixed viewpoint under different and unknown lighting conditions, our method automatically determines a radiometric response function and resolves the generalized bas-relief ambiguity for estimating accurate surface normals and albedos. We show that color and intensity profiles, which are obtained from registered pixels across images, serve as effective cues for addressing these two calibration problems. As a result, we develop a complete auto-calibration method for photometric stereo. The proposed method is useful in many practical scenarios where calibrations are difficult. Experimental results validate the accuracy of the proposed method using various real-world scenes.

1. Introduction

Photometric stereo [23] recovers surface orientations and albedos from a set of images taken from a fixed viewpoint under different lighting directions. There are two necessary calibrations for a standard photometric stereo algorithm to work correctly. First, the camera needs to be radiometrically calibrated to obtain the scene irradiance from measured pixel values. Second, lighting directions and intensities need to be known to uniquely determine the surface. With these two calibrations, surface orientations and albedos can be estimated uniquely from three images for Lambertian scenes.

Typically, these calibrations are performed separately before applying photometric stereo algorithms. The radiometric camera calibration is often performed as a completely independent preprocessing step by, for example, using multiple images [15] captured from a fixed viewpoint under a static lighting configuration with different exposure times. In practice, however, this involves an additional data measurement that is not always possible. The lighting direc-

∗This work was done while the first author was visiting Microsoft Research Asia.
1.1. Related works

Our work is related to both radiometric calibration and GBR disambiguation. Several methods have been proposed to estimate camera response functions. One of the most widely used methods is Mitsunaga and Nayar’s method [15], which uses multiple images of a static scene with multiple different known exposure times. More recent methods work under more general imaging conditions. Lin et al. [13] proposed a method that only requires a single color image. Later they extended it to work with a single gray-scale image [14]. These methods are based on color/intensity mixtures on edge boundaries. There are a few works that allow scene/camera movements during data capturing. Kim and Pollefeys [11] computed pixel correspondence across image frames for radiometric calibration. Grossberg and Nayar [8] used intensity histograms of two images of different exposure rather than exact pixel correspondence. Wilburn et al. [22] showed the response function can be recovered from a motion blur in a single image. Unlike all of these methods, our radiometric calibration method uses photometric stereo images captured by a static camera under varying lighting conditions.

The GBR ambiguity is an intrinsic shape/lighting ambiguity for uncalibrated photometric stereo with unknown lighting parameters. This ambiguity arises because different combinations of shapes and lightings can produce the same appearance. To resolve the GBR ambiguity, additional cues are necessary. Prior knowledge about surface albedo can be used to resolve the GBR [10, 1]. Chandraker et al. [4] used interreflections in a Lambertian scene to resolve the ambiguity. It has been shown that the GBR can also be resolved if the surface reflectance is known to be Lambertian combined with a Torrance-Sparrow specular reflection [7], or with specular reflections containing specular spikes [6, 5]. More recently, Tan et al. [20, 21] showed that the GBR can be reduced and resolved in a stratified manner for isotropic surfaces. Besides resolving the GBR ambiguity, Basti et al. [2] proposed a general solution under environment lighting. Furthermore, to handle various BRDFs, Sato et al. [19] suggested using the similarity in radiance changes and Okabe et al. [17] utilized the attached shadow. Our method takes a different approach from these. As we will see later in this paper, our method can achieve a high-accuracy estimation using the intensity profile as a cue.

The radiometric calibration and GBR disambiguation have been discussed separately in prior works. To achieve a complete auto-calibration in photometric stereo, we solve the mixture of these problems. We take a fundamentally different approach to achieve this goal using the color and intensity profiles.

2. Problem statement

We perform photometric stereo with images taken from a fixed viewpoint under varying and unknown directional lighting (both intensities and directions are unknown) using a camera where the response function is also unknown. We address the two calibration problems in this setting, i.e., radiometric camera calibration and GBR disambiguation. In this paper, we use the following imaging model:

\[ M = f(I) = f((\rho n) \cdot (EI)), \]  

where \( M \) is the recorded pixel value, and \( \rho \) and \( n \) are the surface reflectance (or albedo) and the surface normal vector at a pixel. \( E \) and \( I \) are the intensity and direction of the directional lighting. \( I \) is the irradiance, and \( f \) is the camera’s radiometric response function. Since the albedo \( \rho \) and surface normal vector \( n \) are location dependent and lighting intensity \( E \) and direction \( I \) are image dependent, we use subscripts \( i \) and \( j \) to index pixels and images respectively. For multiple images under varying directional lighting, the imaging model can be written as

\[ M_{ij} = f(I_{ij}) = f((\rho n_i) \cdot (E_j I_j)). \]  

Our goal is to estimate \( \rho_i, n_i, E_j, I_j \) and \( f \) from the observations \( M_{ij} \).
Radiometric response function  The goal of radiometric calibration is to estimate the radiometric response function $f$ that maps the irradiance $I$ to a pixel intensity $M$. As the response function $f$ is a monotonic function, there is always a unique inverse mapping $g(= f^{-1})$. Radiometric calibration amounts to estimating the inverse response function $g$.

The generalized bas-relief ambiguity As in previous work [10], the photometric stereo problem is formulated in a matrix form:

$$ I = SL, $$

where $I$ is a $P \times F$ matrix containing irradiance values from all images, and $P$ and $F$ are the number of pixels and images respectively. Each row of $I$ corresponds to a pixel position in an image, and each column corresponds to a different image. The surface matrix $S \in \mathbb{R}^{P \times 3}$ represents surface properties, i.e., the $i$-th row of $S$ encodes the surface normal at the $i$-th pixel scaled by its albedo as $S_{i*} = \rho_i n_i$. The lighting matrix $L \in \mathbb{R}^{3 \times F}$ represents the lighting directions and intensities, i.e., the $j$-th column of the matrix $L$ corresponds to the lighting direction in the $j$-th image scaled by its intensity as $L_{*j} = E_j I_j$. When $L$ is unknown, $S$ and $L$ can be estimated only up to a GBR transformation $G$ by enforcing integrability over the surface as $I = SL = (S'G)(G^{-1}L')$. The GBR transformation $G$ is a three-parameter matrix:

$$ G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & \nu & \lambda \end{pmatrix}, \quad G^{-1} = \frac{1}{\lambda} \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ -\mu & -\nu & 1 \end{pmatrix}. $$

The overview of the proposed method is illustrated in Fig. 1. Our method uses color profiles for radiometric calibration (Section 3) and intensity profiles for resolving GBR ambiguity (Section 4).

3. Radiometric calibration using color profiles

This section describes our method for determining the radiometric response function using color profiles. A color profile is a set of RGB color values at the same coordinate in a set of photometric stereo images. Our method capitalizes on the fact that color profiles form straight lines in the RGB color space when the radiometric response function $f$ is linear. On the other hand, if $f$ is nonlinear, color profiles form nonlinear curves in the RGB space.

To understand this relationship between color profiles and the radiometric response function $f$, let us consider the ratio between red, green and blue colors at a pixel. In a color image, Eq. (2) is evaluated at each color channel, i.e., $M^r = f(I^r)$, $M^g = f(I^g)$ and $M^b = f(I^b)$, where superscripts $r$, $g$, and $b$ correspond to the RGB color channels. If the radiometric response function $f$ is linear, i.e., $f(x) = kx$, the color ratio at a Lambertian point is

$$ (M^r : M^g : M^b) = (kI^r : kI^g : kI^b) = (\rho^r : \rho^g : \rho^b), $$

which is constant over the photometric stereo image sequences. Hence, in this case, a color profile is a straight line with direction $(\rho^r, \rho^g, \rho^b)$ passing through the origin in the RGB space. On the other hand, if $f$ is nonlinear, this ratio will change over the image sequence. As a result, the color profile bends to a nonlinear curve in the RGB space. This result is consistent with the empirical observation in [18]. Examples of color profiles obtained from a real-world scene are shown in Fig. 2.

The correct inverse response function $g$ ($g(M) = I$) should linearize the curved color profiles. Using this property as a constraint, we estimate the inverse response function $g$ that maps curved color profiles to straight lines. We first define a nonlinearity measurement for color profiles to assess how the profiles are bended. We then formulate the estimation problem in an energy minimization framework to derive the inverse response function $g$.

Degree of nonlinearity  We assess the degree of nonlinearity of a transformed color profile by its line fitting error. Let us denote the RGB triplet of the measured color value as $M$ and use $g(M)$ to represent that the inverse response function $g$ is applied to each element of $M$ independently. For a color profile at the $i$-th pixel, it is computed as

$$ D_i(g) = \sum_j \frac{||(X^j_0 - X^j_1) \times (g(M_{ij}) - X^j_1)||}{||X^j_0 - X^j_1||}, $$

where $X^j_0$ and $X^j_1$ are two points in the RGB space that define the line best fitting to the transformed color profile $g(M_{ij})$. Intuitively, $D$ measures the summed Euclidean distance between $g(M_{ij})$ to the fitted line. It approaches to zero when $g$ is the correct inverse function.
Energy function We compute the inverse response function \( g \) from a set of color profiles \( \Omega \) by minimizing the non-linearity. We use a polynomial representation for the inverse response function to reduce the computational complexity as in [15]. Thus, the inverse response function \( g \) is described by a small number of unknown coefficients \( c_k \) in the form of \( K \)-order polynomial as

\[
I = g(M) = \sum_{k=0}^{K} c_k M^k. \quad (7)
\]

We estimate the coefficients \( c = \{c_0, ..., c_K\} \) and fitted lines \( \chi = \{X_i^0, X_i^1\} | i \in \Omega \) by minimizing the summed nonlinearity for each color profile \( i \in \Omega \):

\[
\{\hat{c}, \hat{\chi}\} = \arg\min_{c, \chi} \frac{1}{|\Omega|} \sum_{i \in \Omega} D_i(e), \quad (8)
\]

where \( |\Omega| \) is the number of color profiles used for the computation. The minimization is performed with normalized color values in the range \([0, 1]\).

For minimization, we put two constraints on the inverse radiometric response function \( g \), namely monotonicity and boundary conditions, which are general for any inverse response functions. The monotonicity constraint is represented as \( \frac{\partial g}{\partial M} > 0 \), and the boundary conditions are described as \( g(0) = 0 \) and \( g(1) = 1 \). We use these boundary conditions as a hard constraint by setting \( c_0 = 0 \) and \( c_K = 1 - \sum_{k=1}^{K-1} c_k \) because \( g(1) = \sum_{k=1}^{K} c_k = 1 \). The final formulation of the energy minimization becomes

\[
\{\hat{c}, \hat{\chi}\} = \arg\min_{c, \chi} \frac{1}{|\Omega|} \sum_{i \in \Omega} D_i(e) + \lambda \sum_i H\left(-\frac{\partial g(t)}{\partial M}\right). \quad (9)
\]

Here, \( H(\cdot) \) is the Heaviside step function \( (H(x) = 1 \text{ when } x \geq 0, \text{ and } H(x) = 0, \text{ otherwise}) \). The derivative \( \frac{\partial g}{\partial M} \) are assessed at various \( t \) in the range of \([0, 1]\).

Optimization of Eq. (9) is performed with the Nelder-Mead simplex method [16] due to its simplicity, but other similar methods could be used as well. We initialize \( g \) as a linear function and \( \chi \) as least square fitted lines from color profiles.

4. GBR disambiguation using intensity profiles

This section describes the proposed method for resolving the GBR ambiguity using intensity profiles. By Hayakawa’s singular value decomposition (SVD) method [10], the irradiance \( I \) in Eq. (3) can be factorized into a surface component \( S \) and a lighting component \( L \), up to a linear ambiguity. If albedos are known at 6 different normals or lighting intensity is known at 6 images, this linear ambiguity can be reduced to a rotational ambiguity. On the other hand,

Suppose we are given pixels that have the same albedo \( a \) but different normals. The GBR transformation should satisfy \( ||s_iG|| = a \), i.e.,

\[
s_i CG^T = a^2, \quad (10)
\]

where

\[
C = GG^T = \begin{pmatrix} 1 & 0 & \mu \\ 0 & 1 & \nu \\ \mu & \nu & \mu^2 + \nu^2 + \lambda^2 \end{pmatrix}, \quad (11)
\]

and \( s_i \) is the \( i \)-th row vector of \( S \). Let \( \tilde{\lambda} = \mu^2 + \nu^2 + \lambda^2 \) be an auxiliary variable. The constraint in Eq. (10) is linear in all 4 unknowns \( \{\mu, \nu, \tilde{\lambda}, a^2\} \). Therefore, given at least 4 pixels with different normals but with the common albedo, the GBR ambiguity can be resolved. Our goal is to automatically identify these pixels to resolve the GBR ambiguity.

Surface normal grouping Following Koppal and Narasimhan’s work [12], we use intensity profiles to cluster surface normals. As shown in Fig. 3, pixels with the same surface normal show strong correlation in their intensity profiles, while pixels with different surface normals become less correlated. Since we work in the irradiance domain, instead of using the extrema of the profiles as in [12], we simply use the correlation of the entire profiles. This allows us to use fewer images for surface normal clustering. The Pearson’s correlation \( r_{iij'} \) between two intensity profiles \( I_i \),
and $I_i'$ is assessed by
\[
\tau_{i\nu'} = \frac{\sum_{j=1}^{F} (I_{ij} - \bar{I}_i)(I_{ij'} - \bar{I}_{\nu'})}{(F-1)\sigma_i\sigma_{\nu'}}, \tag{12}
\]
where $\bar{I}_i$ is the mean value of the $i$-th row vector of matrix $I$, and $\sigma_i$ is its standard deviation. Using the correlation as a distance, we perform $K$-means clustering to partition images into a set of groups that have similar surface normals. Hence, by selecting from different groups, we can obtain points with different normal directions. For this purpose, we use a large $K$. An example of the surface normal grouping result is shown in Fig. 4 (middle).

**Albedo grouping** Our albedo grouping aims at identifying pixels with the same albedo. As we have discussed in Section 3, the RGB ratio of a Lambertian scene point should be constant ($\rho^r : \rho^g : \rho^b$) across images in the irradiance domain. Hence, we can approximately measure the similarity of pixel albedos by their difference of chromaticity. To reduce the influence of shadows and noise, we compute the chromaticity of each pixel from its averaged irradiance value $I$. The chromaticity is defined as a vector of two elements \([I^r/(I^r + I^g + I^b), I^g/(I^r + I^g + I^b)]\). We then cluster pixels in the chromaticity domain to obtain pixels with similar albedo. Similar to the surface normal grouping, we apply $K$-means clustering using the Euclidean distance of chromaticity vectors. As a result, we obtain albedo groups as shown in Fig. 4 (right).

**Pixel selection** We combine the two grouping results to automatically select pixels with the same albedo but different surface normals.

First, our method selects reliable albedo groups. All the albedo groups are sorted by their variances, and half of the groups with larger variances are eliminated. After that, in each albedo group, we further remove pixels that have a large distance from the group center in the chromaticity domain. Specifically, pixels that have a distance larger than the average are discarded. The selected groups are split into subgroups based on the surface normal grouping result. The resulting subgroups become the groups of pixels with the same albedo and the same surface normal. Since we aim at finding pixels with the same albedo but different surface normals, only one pixel from each subgroup is selected.

To implement this pixel selection algorithm, we have tested two different approaches. One is random selection from subgroups, and the other is an optimization approach. The second approach selects pixels that minimize the summation of the surface normal similarity among selected pixels. This optimization problem is discrete and can be efficiently solved using a belief propagation algorithm [24]. Although the optimization-based method theoretically guarantees to select pixels with diverse surface normals, we find that the random sampling produces as good result as the optimization-based method when the number of surface normal groups is high. For this reason, we use the random sampling method in this work.

Finally, a set of pixels are selected from $R$ albedo groups of unknown albedo values \(\{a_1, a_2, \ldots, a_R\}\). From these pixels, we obtain multiple linear equations Eq. (10) with unknowns \(\{\mu, \nu, \bar{\lambda}, a_1^2, \ldots, a_K^2\}\) to resolve the GBR ambiguity by a least square solution.

### 5. Experiments

We use four real-world scenes, CAT, OWL (courtesy of Dan Goldman and Steven Seitz\(^1\)), PILLOW and SHEEP scenes, to verify our method. The scenes are shown in Fig. 5. CAT and OWL are radiometrically calibrated, while PILLOW and SHEEP are recorded by a Nikon D70 and Canon 20D, respectively. Both of these cameras have nonlinear radiometric response functions. In this experiment, we first quantitatively evaluate the radiometric calibration part. Second, we use the CAT and OWL scenes to assess the proposed GBR disambiguation method. Finally, we evaluate the entire pipeline with both radiometric calibration and GBR disambiguation. We implement our method using Matlab. The optimization of Eq. (9) can be implemented simply by a Matlab built-in function “fminsearch”. Throughout our experiments, we set $\lambda = 5000$, $|\Omega| = 100$ (100 color profiles), $K = 200$ for the surface normal grouping, and $K = 20$ for the albedo grouping in two $K$-means stages. With an Intel E6550 (2.33GHz) CPU, the entire process took 10 minutes without optimization.

\(^1\)http://www.cs.washington.edu/education/courses/csep576/05wi/projects/project3/project3.htm
Radiometric calibration result Using a Nikon D70 and Canon 20D, we take 50 datasets for testing the radiometric calibration method. From each dataset, we randomly select pixels to produce color profiles. The result is shown in Fig. 6 in comparison with the ground truth obtained by Mitsunaga and Nayar’s method [15]. The averages of root mean squared error (RMSE) and disparity (the maximum difference) are computed from all the 50 datasets. The statistics show that our method consistently works well with various scenes and different cameras.

Surface normal and albedo grouping result We show an example of the grouping result of the SHEEP scene in Fig. 7. It depicts the accuracy of the grouping result for the following GBR disambiguation.

GBR disambiguation result We evaluate the proposed GBR disambiguation method using the CAT and OWL scenes. Fig. 8 and Fig. 9 show the results. On the top row, the color coded normal maps (x, y, z components are linearly mapped in the RGB channel) are shown. On the bottom row is a depth map integrated from the normal map by Poisson solver. For a validation, we show the ground truth obtained by the calibrated photometric stereo on the left. We compare our method with the minimum entropy method [1]. Because our method uses surface normal groups in addition to the albedo information, it performs well in GBR disambiguation. The average angular errors are summarized in Table 1. Comparing the results between the CAT and OWL scenes, because of the smoothly varying albedo of OWL scene, the performances of both our method and the minimum entropy method degrade.

Self-calibrating photometric stereo result To test the entire pipeline, we use the PILLOW and SHEEP scenes taken by a Nikon D70 and Canon 20D, respectively. The results are shown in Fig. 10 and Fig. 11. In the figures, from left to right, we show the calibrated result (known lighting and response function), the result of our method with and without the radiometric calibration step. These results show that our method produces results that are close to the calibrated data, although we only use the input images taken under unknown lighting parameters with an unknown response function. The importance of radiometric calibration can also be seen clearly in the side-by-side comparison. Table 2 shows the quantitative comparison. Our self-calibration method significantly improves the accuracy. Fig. 12 shows the rendering of the reconstructed surface.

We recommend viewing the electronic version of this paper for better visualization.
6. Conclusion and discussion

Auto-calibration for photometric stereo is important for photometric stereo to be applicable in various settings. In this paper, we propose a self-calibrating photometric stereo method that handles both radiometric calibration and GBR disambiguation. The mixture of these problems is solved by analyzing color/intensity profiles in the RGB and irradiance-time domains. We show that radiometric calibration can be performed from images from a fixed viewpoint under varying lightings. We also develop a method to automatically resolve the GBR ambiguity from at least four pixels with different normals but the same albedo.

Limitations Our radiometric calibration method has a couple of limitations. First, it cannot work with gray-scale images and scenes. Second, it cannot handle a certain class of response functions $f$ where $f(x)/f(ax) = \text{const.}$ holds.

<table>
<thead>
<tr>
<th></th>
<th>Angular error</th>
<th>PILLOW</th>
<th>SHEEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our method</td>
<td>Mean</td>
<td>5.63</td>
<td>7.30</td>
</tr>
<tr>
<td></td>
<td>Std. dev.</td>
<td>3.71</td>
<td>3.02</td>
</tr>
<tr>
<td>Without radiometric calib.</td>
<td>Mean</td>
<td>16.08</td>
<td>20.18</td>
</tr>
<tr>
<td></td>
<td>Std. dev.</td>
<td>7.35</td>
<td>6.86</td>
</tr>
</tbody>
</table>

Table 2. Mean angular error and standard deviation of self-calibrating photometric stereo result, and the result without radiometric calibration.
This is a special case when the response function is non-linear while the color profiles remain straight. This class of function includes $f(x) = x^n, n \in \mathbb{R}$, when boundary conditions $f(0) = 0$ and $f(1) = 1$ is imposed.

To assess the practical limitation of our method, we tested our method using the database of response functions (DoRF) [9] with a simulation dataset. We have confirmed that all the response functions in the database, including gamma correcting functions, bend the color profiles in the RGB color space. Among the degenerate case where the response function is in the form of $f(x) = x^n$, only the linear response function $f(x) = x$ is likely dominant in practice.\(^3\)

Therefore, by setting the initial guess of the response function as a linear function, our method works well for a wide variety of real-world camera response functions.

Acknowledgement

The authors would like to thank Neil Alldrin for providing the code to compare the results. Boxin Shi and Chao Xu was partially supported by the Chinese and Beijing Natural Science Foundation (Grant No. 60975014 and 4102024). Ping Tan was supported by Singapore MOE Grant R-263-000-555-112.

References


\(^3\)The linear response as well as other response functions in the form of $f(x) = x^n$ does not exist in the DoRF.