Does Inverse Lighting Work Well under Unknown Response Function?

Shuya Ohta and Takahiro Okabe
Department of Artificial Intelligence, Kyushu Institute of Technology, Iizuka, Japan

Keywords: Single-image Understanding, Inverse Lighting, Response Function, Lambertian Model.

Abstract: Inverse lighting is a technique for recovering the lighting environment of a scene from a single image of an object. Conventionally, inverse lighting assumes that a pixel value is proportional to radiance value, *i.e.* the response function of a camera is linear. Unfortunately, however, consumer cameras usually have unknown and nonlinear response functions, and therefore conventional inverse lighting does not work well for images taken by those cameras. In this study, we propose a method for simultaneously recovering the lighting environment of a scene and the response function of a camera from a single image. Through a number of experiments using synthetic images, we demonstrate that the performance of our proposed method depends on the lighting distribution, response function, and surface albedo, and address under what conditions the simultaneous recovery of the lighting environment and response function works well.

1 INTRODUCTION

Humans seem to extract rich information about an object of interest even from a single image of the object. Since the dawn of computer vision, understanding a single image is one of the most important and challenging research tasks.

The appearance of an object depends on the shape and reflectance of the object as well as the lighting environment of a scene. From the viewpoint of physical image understanding, this means that the problem of single-image understanding results in recovering those three descriptions of a scene from a single image. Unfortunately, however, such a problem is terribly underconstrained; we have multiple unknowns but only a single constraint per pixel. Therefore, conventional techniques such as shape from shading (Horn, 1986) assume that two out of the three descriptions are known and then recover the remaining one.

One direction of generalization of single-image understanding is to recover two or three descriptions from a single image. For example, Romeiro and Zickler (Romeiro and Zickler, 2010) propose a method for simultaneously recovering the reflectance of an object and the lighting environment of a scene from a single image of the object with known shape (sphere). Barron and Malik (Barron and Malik, 2012) propose a method for simultaneously recovering the shape, albedo, and lighting environment from a single image. Those methods exploit the priors, *i.e.* the statistical models of those descriptions, and search the most likely explanation of a single image.

In this study, we focus on single-image understanding under unknown camera properties, that is another direction of generalization. Existing methods for single-image understanding assume that a single image is taken by an ideal camera and ignore the effects of in-camera processing. They represent the pixel values of the image as a function with respect to the descriptions of a scene, *i.e.* shape, reflectance, and lighting environment, and then recover some of those descriptions from a single image. On the other hand, we assume that a single image is taken by a consumer camera and take in-camera processing in particular tone mapping into consideration. We represent the pixel values of the image as a function with respect to both the scene descriptions and the properties of a camera and camera setting, and then address the recovery of scene descriptions from a single image under unknown camera properties.

Inverse lighting (Marschner and Greenberg, 1997) is a technique for recovering the lighting environment of a scene from a single image under the assumptions that the shape and reflectance (or basis images) of the scene are known. Comparing with active techniques using devices such as a spherical mirror (Debevec, 1998) and a camera with a fish-eye lens (Sato et al., 1999) placed in a scene of interest, inverse lighting is a passive technique and therefore could recover the lighting environment from a given image taken even in the past, although it is applicable to scenes with relatively simple shape and reflectance.

Conventionally, inverse lighting assumes that a pixel value is proportional to a radiance value at the

corresponding point in a scene. The relationship between radiance values and pixel values is described by a *radiometric response function*, and the above assumption means that the response function of a camera is linear. Unfortunately, however, consumer cameras usually have nonlinear response functions in order to improve perceived image quality via tone mapping (Grossberg and Nayar, 2003). Therefore, conventional inverse lighting requires a machine-vision camera with a linear response function or radiometric calibration of a response function in advance.

Accordingly, we propose a method for recovering the lighting environment of a scene from a single image taken by a camera with an unknown and nonlinear response function. Our proposed method also assumes that the shape and reflectance of an object are known, and then simultaneously estimates both the lighting environment of a scene and the response function of a camera from a single image. Specifically, our method represents a lighting distribution as a linear combination of the basis functions for lighting and also represents a response function as a linear combination of the basis functions for response, and then estimates those coefficients from a single image. We conduct a number of experiments using synthetic images, and investigate the stability of our method. We demonstrate that the performance of our method depends on the lighting environment, response function, and surface albedo, and show experimentally under what conditions the simultaneous recovery of the lighting environment and response function from a single image works well.

The main contribution of this study is twofold; (i) the novel method for simultaneously recovering the lighting environment of a scene and the response function of a camera from a single image, and (ii) empirical insights as to under what conditions inverse lighting from a single image with an unknown response function works well.

2 INVERSE LIGHTING

In this section, we explain the framework of inverse lighting on the basis of the original work by Marschner and Greenberg (Marschner and Greenberg, 1997). Inverse lighting assumes that the shape and reflectance of an object are known, and then recovers the lighting environment of a scene from a single image of the object. We assume that the object is illuminated by a set of directional light sources, and describe the intensity of the incident light from the direction (θ,ϕ) to the object as $L(\theta,\phi)$. Here, θ and ϕ are the zenith and azimuth angles in the spherical coordinate system centered at the object. Hereafter, we call $L(\theta,\phi)$ a lighting distribution.

Specifically, we represent a lighting distribution $L(\theta,\phi)$ by a linear combination of basis functions as

$$L(\theta,\phi)=\sum_{n=1}^{N}\alpha_n L_n(\theta,\phi), \tag{1}$$

where α_n and $L_n(\theta,\phi)$ $(n=1,2,3,...,N)$ are the coefficients and basis functions for lighting. Then, based on the assumption of known shape and reflectance, we synthesize the basis images, *i.e.* the images of the object when the lighting distributions are equal to the basis functions for lighting $L_n(\theta,\phi)$. We denote the p-th $(p=1,2,3,...,P)$ pixel value of the n-th basis image by $R_p(L_n)$. According to the superposition principle, the p-th pixel value of an input single image I_p $(p=1,2,3,...,P)$ is described as

$$I_p=\sum_{n=1}^{N}\alpha_n R_p(L_n). \tag{2}$$

This means that we obtain a single constraint on the coefficients of lighting per pixel.

Rewriting the above constraints in a matrix form, we obtain

$$\begin{pmatrix}\vdots\\ I_p\\ \vdots\end{pmatrix} = \begin{pmatrix}\vdots\\ \cdots\ R_p(L_n)\ \cdots\\ \vdots\end{pmatrix}\begin{pmatrix}\vdots\\ \alpha_n\\ \vdots\end{pmatrix}, \tag{3}$$

$$\boldsymbol{I} = R\boldsymbol{\alpha}. \tag{4}$$

Since the $P\times N$ matrix R is known and the number of pixels (constraints) P is larger than the number of basis functions (unknowns) N in general, we can estimate the coefficients of lighting $\boldsymbol{\alpha}$ by solving the above set of linear equations. Specifically, the coefficients are computed by using the pseudo inverse matrix R^+ as

$$\boldsymbol{\alpha}=R^+\boldsymbol{I}=(R^\top R)^{-1}R^\top\boldsymbol{I}, \tag{5}$$

if R is full rank, *i.e.* the rank of R is equal to N. This solution is equivalent to that of the least-square method;

$$\boldsymbol{\alpha}=\arg\min_{\hat{\boldsymbol{\alpha}}}\sum_{p=1}^{P}\left[I_p-\sum_{n=1}^{N}\hat{\alpha}_n R_p(L_n)\right]^2. \tag{6}$$

Once the coefficients of lighting $\boldsymbol{\alpha}$ are computed, we can obtain the lighting distribution by substituting them into eq.(1).

3 PROPOSED METHOD

In this section, we propose a method for simultaneously recovering both the lighting environment of a

scene and the response function of a camera from a single image taken under an unknown and nonlinear response function.

3.1 Representation of Lighting Distribution

Our proposed method assumes that the shape of an object is convex and the reflectance obeys the Lambertian model, and then recovers the lighting distribution of a scene on the basis of the diffuse reflection components observed on the object surface. It is known that the image of a convex Lambertian object under an arbitrary lighting distribution is approximately represented by a linear combination of 9 basis images when low-order spherical harmonics $Y_{lm}(\theta,\phi)$ $(l = 0,1,2;\ m = -l,-l+1,...,l-1,l)$ are used as the basis functions for lighting (Ramamoorthi and Hanrahan, 2001). For the sake of simplicity, we denote the spherical harmonics $Y_{lm}(\theta,\phi)$ by $Y_n(\theta,\phi)$, where $n = (l+1)^2 - l + m$.

Substituting the spherical harmonics into eq.(1), a lighting distribution $L(\theta,\phi)$ is represented as

$$L(\theta,\phi) = \sum_{n=1}^{9} \alpha_n Y_n(\theta,\phi). \tag{7}$$

In a similar manner to eq.(2), the p-th pixel value of the n-th basis image is represented as

$$I_p = \sum_{n=1}^{9} \alpha_n R_p(Y_n). \tag{8}$$

Note that the pixel value I_p is equivalent to the radiance value when the response function of a camera is linear.

The reason why the image of a convex Lambertian object under an arbitrary lighting distribution is approximated by using low-order spherical harmonics is that high-frequency components of a lighting distribution have no/little contribution to pixel values. In an opposite manner, we cannot recover the high-frequency components of a lighting distribution from pixel values (Ramamoorthi and Hanrahan, 2001). This is the limitation of inverse lighting from diffuse reflection components.

3.2 Representation of Response Function

The radiance values of a scene are converted to pixel values via in-camera processing such as demosaicing, white balancing, and tone mapping. In this study, we focus on tone mapping because it is widely used in consumer cameras in order to improve perceived image quality and the pixel values converted by tone mapping are significantly different from the radiance values as shown in Figure 1 (a) and (b). As mentioned in Section 1, the relationship between radiance values and pixel values is described by a radiometric response function. In general, the response function depends on cameras and camera settings.

Let us denote a response function by f, and assume that a radiance value I is converted to a pixel value I' by using the response function as $I' = f(I)$. Since the response function is monotonically increasing, there exists the inverse of f, *i.e.* an inverse response function g. The inverse response function converts a pixel value I' to a radiance value I as $I = g(I')$, and therefore has 255 degrees of freedom for 8-bit images. Such a high degree of freedom makes the simultaneous recovery of a lighting distribution and a response function from a single image intractable.

Accordingly, our proposed method uses an efficient representation of response functions by constraining the space of response functions on the basis of the statistical characteristics. Specifically, we make use of the EMoR (Empirical Model of Response) proposed by Greenberg and Nayar (Grossberg and Nayar, 2003). They apply PCA to the dataset of response functions, and show that any inverse response function is approximately represented by a linear combination of basis functions as

$$I = g(I') = g_0(I') + \sum_{m=1}^{M} \beta_m g_m(I'). \tag{9}$$

Here, β_m and $g_m(I')$ are the coefficients and basis functions for response. Since the inverse response function is also monotonically increasing, the coefficients has to satisfy

$$g_0(I') + \sum_{m=1}^{M} \beta_m g_m(I') < g_0(I'+1) + \sum_{m=1}^{M} \beta_m g_m(I'+1), \tag{10}$$

where $I' = 0,1,2,...,254$ for 8-bit images.

3.3 Simultaneous Recovery

Substituting eq.(9) into the left-hand side of eq.(8), we can derive

$$g_0(I'_p) = \sum_{n=1}^{9} \alpha_n R_p(Y_n) - \sum_{m=1}^{M} \beta_m g_m(I'_p), \tag{11}$$

i.e. a single constraint on the coefficients of lighting and response per pixel. We can rewrite the above constraints in a matrix form as

$$\boldsymbol{g}_0 = (R|G) \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix}, \tag{12}$$

where $\boldsymbol{g}_0 = (g_0(I'_1), g_0(I'_2), \cdots, g_0(I'_P))^\top$ and G is a $P \times M$ matrix;

$$G = \begin{pmatrix} & \vdots & \\ \cdots & -g_m(I'_p) & \cdots \\ & \vdots & \end{pmatrix}. \quad (13)$$

If the $P \times (9+M)$ matrix $(R|G)$ is full rank, *i.e.* the rank of $(R|G)$ is $(9+M)$, we can compute both the coefficients of the lighting distribution and those of the response function by using its pseudo inverse matrix in a similar manner to Section 2;

$$\begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} = (R|G)^+ \boldsymbol{g}_0. \quad (14)$$

This solution is equivalent to that of the least-square method;

$$\{\boldsymbol{\alpha}, \boldsymbol{\beta}\} = \arg\min_{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}} \sum_{p=1}^{P} \left[g_0(I'_p) - \sum_{n=1}^{9} \hat{\alpha}_n R_p(L_n) + \sum_{m=1}^{M} \hat{\beta}_m g_m(I'_p) \right]^2 \quad (15)$$

Since the response function is monotonically increasing, we solve eq.(15) subject to eq.(10). We used the MATLAB implementation of the trust-region reflective algorithm for optimization. In the experiments, we set $M = 5$. Once the coefficients of lighting $\boldsymbol{\alpha}$ and those of response $\boldsymbol{\beta}$ are computed, we can obtain the lighting distribution and response function by substituting them into eq.(7) and eq.(9).

In order to make the simultaneous recovery of lighting and response more stable, we can incorporate the priors of lighting distributions and response functions into the optimization. For example, we can add the smoothness term with respect to the response function

$$w \sum_{l=1}^{255} \left[\frac{\partial^2 g(I')}{\partial I'^2} \bigg|_{I' = \frac{l}{255}} \right]^2 \quad (16)$$

to eq.(15), where w is a parameter that balances the likelihood term and the smoothness term. In our preliminary experiments, we tested some simple priors and found that we often need to fine-tune the parameters of the priors according to input images. Therefore, we do not use any priors in this study and investigate the stability of the linear least-square problem in eq.(15) with the linear constraints in eq.(10) instead.

4 EXPERIMENTS AND DISCUSSION

In this section, we conduct a number of experiments using synthetic images, and investigate the stability of our proposed method. We can see from eq.(12) and eq.(14) that the stability of our method is described by the full rankness of the matrix $(R|G)$ or the condition number of the pseudo inverse matrix $(R|G)^+$.

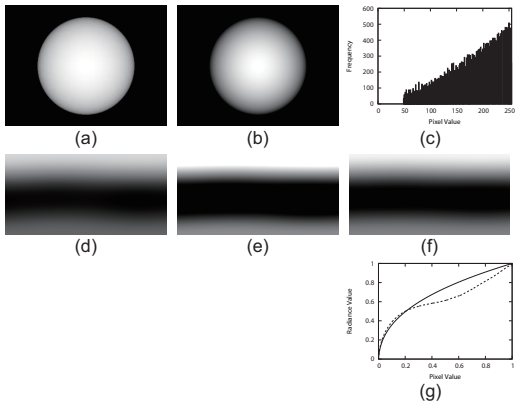

Figure 1: Recovered lighting distributions and response function from a single image of an object under a single directional light source.

Table 1: The RMS errors of the estimated response functions.

Case	A	B	C	D	E
RMSE	0.09	0.10	0.54	0.04	0.13

Each column of the sub-matrix R corresponds to a basis image of an object illuminated by spherical-harmonics lighting. It is known that the basis images are orthogonal to each other if the surface normals of the object distribute uniformly on a unit sphere because spherical harmonics are orthonormal basis functions on a unit sphere (Ramamoorthi and Hanrahan, 2001). Intuitively, inverse lighting works well for spherical objects but does not work for planar objects. In this study, we assume spherical objects and therefore the sub-matrix R is full rank.

Each column of the sub-matrix G corresponds to an eigenvector of response functions, and therefore the columns are orthogonal to each other if the pixel values in a single image distribute uniformly from 0 to 255. Intuitively, the simultaneous recovery of lighting distribution and response function works well when the histogram of the pixel values is uniform. Because the pixel values in the image of a spherical object depend on the lighting distribution of a scene, the response function of a camera, and the surface albedo of the object, we demonstrate how the performance of our method changes depending on the lighting distribution, response function, and surface albedo.

Case A:
In Figure 1, we show images of a sphere under a single directional light source with a linear response function (a) and with a nonlinear response function

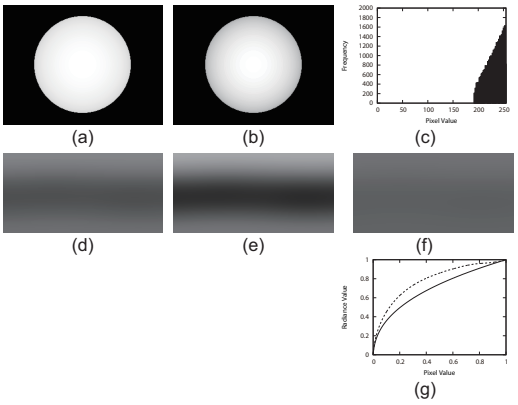

Figure 2: Results when only the lighting distribution is different from Figure 1.

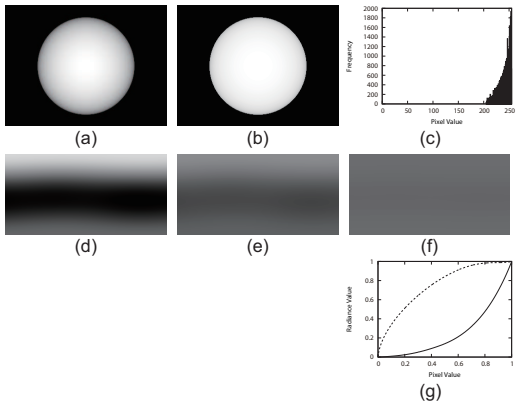

Figure 3: Results when only the response function is different from Figure 1.

(b). The histogram of the pixel values in the tone-mapped image (b) is shown in (c). Figure 1 (d) shows the lighting distribution estimated from the linear image (a) by using the conventional method assuming a linear response function (see Section 2). Here, the lighting distribution $L(\theta, \phi)$ is represented by a 2D map whose vertical and horizontal axes correspond to the zenith angle θ and azimuth angle ϕ respectively. Figure 1 (e) and (f) show the lighting distributions estimated from the tone-mapped image (b) by using the conventional method and our proposed method respectively. In Figure 1 (g), the solid and dotted lines stand for the ground truth and estimated response function by using our method.

We can see that the lighting distribution estimated by using our proposed method (f) looks more similar to (d) than that estimated by using the conventional method (e). Since inverse lighting based on diffuse reflection components cannot estimate the high-frequency components of a lighting distribution as mentioned in Subsection 3.1, (d) is considered to be the best possible result. Therefore, those results demonstrate that our method works better than the conventional method for the tone-mapped image. In addition, (g) demonstrates that the response function estimated by using our method is similar to the ground truth in some degree. The root-mean-square (RMS) errors of the estimated response functions are shown in Table 1.

Case B:

In Figure 2, we show the results when only the lighting distribution is different from Figure 1. Specifically, a single directional light source and a uniform ambient light are assumed. Comparing the lighting distributions estimated by using the conventional method (e) and our proposed method (f) with the best possible result (d), we can see that our method outperforms the conventional method. In addition, (g) shows that the response function estimated by using our method is similar to the ground truth in some degree.

Those results demonstrate that the performance of our method depends on lighting distributions. Although the performance of our method is not perfect, our method works well for both the case A and the case B, and outperforms the conventional method.

Case C:

In Figure 3, we show the results when only the response function is different from Figure 1. Comparing (e) and (f) with (d), it is clear that our method does not work well. In addition, (g) shows that the response function estimated by using our method is completely different from the ground truth.

Those results demonstrate that the performance of our method depends also on response functions and that our method does not work well for the case C. The histogram of the pixel values in the tone-mapped image (c) shows that the range of pixel values is significantly reduced due to the tone mapping. Comparing Figure 1 (g) with Figure 3 (g), we can see that the simultaneous recovery works well when the inverse response function is convex upward, *i.e.* expands the range of pixel values, but does not work well when it is convex downward, *i.e.* shrinks the range of pixel values.

Case D:

In Figure 4, we show the results when the image of a textured sphere under four directional light sources is used. Comparing (e) and (f) with (d), we can see that our method works better than the conventional method. In addition, (g) shows that our method can estimate the response function accurately.

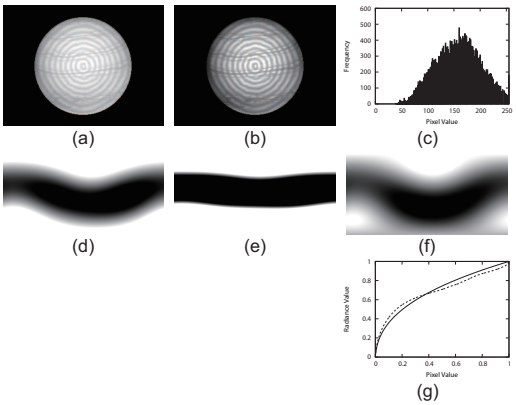

Figure 4: Recovered lighting distributions and response function from a single image of a textured object under four directional light sources.

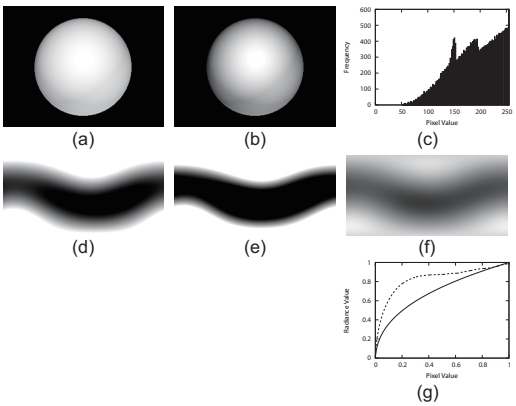

Figure 5: Results when only the surface albedo is different from Figure 4.

Case E:
In Figure 5, we show the results when only the surface albedo is different from Figure 4. Comparing (e) and (f) with (d), we can see that our method does not necessarily work well. In addition, (g) shows that the response function estimated by using our method deviates from the ground truth to some extent.

Those results demonstrate that the performance of our method depends also on the surface albedo of an object and that our method works better for textured objects. The effects of texture can be explained as follows. First, non-uniform albedo makes the distribution of pixel values diverse. Second, more importantly, two pixels with similar surface normals but different reflectance values yield a strong constraint on the response function. Since the irradiance values at the pixels with similar surface normals are also similar to each other, the radiance values converted from the pixel values at those pixels by using the inverse response function should be proportional to their reflectance values.

5 CONCLUSION AND FUTURE WORK

In this study, we extended inverse lighting by taking an unknown and nonlinear response function of a camera into consideration, and proposed a method for simultaneously recovering the lighting environment of a scene and the response function of a camera from a single image of an object. Through a number of experiments, we demonstrated that the performance of our proposed method depends on the lighting distribution, response function, and surface albedo, and addressed under what conditions the simultaneous recovery works well.

One of the future directions of this study is to incorporate sophisticated priors in order to make the simultaneous recovery more stable. Another direction is to make use of other cues such as specular reflection components and cast shadows in order to recover high-frequency components of a lighting distribution.

ACKNOWLEDGEMENTS

A part of this work was supported by JSPS KAKENHI Grant No. 26540088.

REFERENCES

Barron, J. and Malik, J. (2012). Shape, albedo, and illumination from a single image of an unknown object. In *Proc. IEEE CVPR2012*, pages 334–341.

Debevec, P. (1998). Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proc. ACM SIGGRAPH'98*, pages 189–198.

Grossberg, M. and Nayar, S. (2003). What is the space of camera response functions? In *Proc. IEEE CVPR2003*, pages 602–609.

Horn, B. (1986). *Robot vision*. MIT Press.

Marschner, S. and Greenberg, D. (1997). Inverse lighting for photography. In *Proc. IS&T/SID Fifth Color Imaging Conference*, pages 262–265.

Ramamoorthi, R. and Hanrahan, P. (2001). A signal-processing framework for inverse rendering. In *Proc. ACM SIGGRAPH'01*, pages 117–128.

Romeiro, F. and Zickler, T. (2010). Blind reflectometry. In *Proc. ECCV2010*, pages 45–58.

Sato, I., Sato, Y., and Ikeuchi, K. (1999). Acquiring a radiance distribution to superimpose virtual objects onto a real scene. *IEEE TVCG*, 5(1):1–12.