# ILLUMINATION LEARNING FROM A SINGLE IMAGE WITH UNKNOWN SHAPE AND TEXTURE

*Tingbo Hou*[1], *Sen Wang*[2], *Hong Qin*[1]

[1]Department of Computer Science, Stony Brook University, USA
[2]Kodak Research Laboratories, Eastman Kodak Company, USA
{thou, qin}@cs.sunysb.edu sen.wang@kodak.com

## ABSTRACT

In this paper, we introduce a method for learning illumination from single images, which can be further applied to illumination-invariant algorithms in computer vision and image-based rendering in graphics. Illumination learning has been widely studied, yet still has some shortcomings such as the restriction of Lambertian surfaces and the prerequisite of known shape or texture. Our method can adaptively learn illumination from images of vehicles with unknown shape and texture. We formulate the illumination model with both diffusion and specularity components using a frequency-space representation, and adopt an iterative strategy to estimate lighting, shape, and texture under a joint energy function. Using our method, we can perform de-lighting and re-lighting on input images, and render other 3D models with learned illumination. Experimental results show that our method can work in a wide range of environments with both indoor and outdoor illumination conditions.

***Index Terms***— Illumination learning, 3D model, de-lighting, re-lighting

## 1. INTRODUCTION

When many computer vision problems such as detection, recognition and tracking, are compounded with varying illumination, they become extremely challenging. The ideal case is to extract illumination from input images, so we can run algorithms on the left de-lighted images, and if applicable, render other objects with the extracted illumination. It is, however, hard to achieve with complicated lightings and unknown shape and texture of the object.

To date, techniques on varying illumination have been widely studied from various aspects. For a number of algorithms in computer vision, only the degraded reflectance images are concerned for their invariance under varying illumination, such as the intrinsic images [1], the illumination ratio map (IRM) [2], etc. The IRM, for example, works on homogeneous Lambertian surfaces with linear responses to illumination changes. Some other work [3, 4] focus on the separation of diffuse and specular reflections using their different characteristics in the chromaticity space, while the illumination has not been fully extracted and therefore can not be transfered. The previous work on inverse rendering [5] showed that the appearance of an object can be described as a spherical convolution of the illumination and Bidirectional Reflectance Distribution Function (BRDF). This framework has been applied in computer vision problems [6, 7, 8], but restricted to Lambertian surfaces where only the first 9 spherical harmonic coefficients are needed. These methods also have prerequisites of known geometry of the object, or a large training dataset of object in the same category.

In this paper, we present a method that adaptively learns illumination from single images of vehicles. Here the term "learning" refers to estimating lighting coefficients from input images. Our method alleviates the Lambertian restriction and the prerequisites on shape and texture. The illumination is formulated with both diffusion and specularity reflections using a frequency-space representation. A joint linear and nonlinear optimization is adopted to estimate the lighting, shape, and texture iteratively. Initial values for shape and texture are obtained using a 3D generic model and a color grouping method. By our method, we can perform de-lighting and re-lighting on input images, and render 3D models with learned illumination from images. Experimental results show that our method can work in a wide range or environments with both indoor and outdoor illumination conditions.

## 2. ANALYTIC FORMULATION OF ILLUMINATION MODEL

We consider the reflected light field $B(x, \vec{w}_o)$ for a surface point $x$ and an outgoing direction $\vec{w}_o$, given by

$$B(x, \vec{w}_o) = \int_{\Omega} T(x)\rho(\vec{w}_i, \vec{w}_o)L(x, \vec{w}_i)(\vec{w}_i \cdot \vec{n})dw_i, \quad (1)$$

where $\vec{w}_i$ is an incoming direction, and $\vec{n}$ is the surface normal. This integrand is the product of three terms: the texture $T(x)$, the BRDF $\rho(\vec{w}_i, \vec{w}_o)$, and the lighting $L(x, \vec{w}_i)$.

Hong Qin and Tingbo Hou (Stony Brook University)'s research reported in this paper is partially supported by NSF grants: IIS-0949467, IIS-0710819, and IIS-0830183.

The reflected light field can be expanded on spherical harmonics $Y_{l,m}$ with spherical coordinates $(\theta, \phi)$, given by

$$B(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} B_{lm} Y_{lm}(\theta, \phi). \tag{2}$$

Spherical harmonics are represented in the spherical coordinates, forming an orthogonal system. In our method, we exploit an effective model of reflected light field which is derived from the microfacet BRDF including two components:

$$B = B_d + B_{s,slow},$$

where $B_d$ is from the diffuse component of the BRDF, and $B_{s,slow}$ represents specularity from the slowly-varying lighting. Assuming the camera has a linear sensitive function in a given spectrum band, we can approximate the image intensity $I$ as

$$I = \sum_{l=0}^{l^*} \sum_{m=-l}^{l} \Lambda_l L_{lm}(K_d \rho_{dl} + K_s \rho_{sl}) Y_{lm}(n), \tag{3}$$

where $\Lambda_l$ is the normalization constant, $\rho_d$ and $\rho_s$ are diffuse and specular BRDF, $K_d$ and $K_s$ are diffuse and specular albedos (texture), $l^*$ is the cutoff for levels of spherical harmonics, and $Y_{lm}(n)$ is the spherical harmonic function of surface normal $n$ in the local coordinate system. The illumination we want to learn is approximated by

$$L = \sum_{l=0}^{l^*} \sum_{m=-l}^{l} L_{lm} Y_{lm}. \tag{4}$$

For the Larmbertian BRDF, an analytic formula of $A_l = \Lambda_l \rho_{dl}$ can be computed numerically as

$$A_0 = \pi,\ A_1 = \frac{2\pi}{3},\ A_2 = \frac{\pi}{4},\ A_3 = 0,\ A_4 = -\frac{\pi}{24},\ ...,$$

and more than 99% of the energy is captured by $l \leq 2$. For the specularity component of the BRDF, we use the Phone illumination model with an approximation given by

$$\Lambda_l \rho_{sl} \approx e^{-\frac{l^2}{2s}}, \tag{5}$$

where $s$ is the Phone exponent which can be approximated according to certain materials or estimated independently by specifying a neighborhood region of specularity under directional source of light. This is a good approximation when $l^* \approx \sqrt{2s}$. Furthermore, we assume $K_d + K_s = 1$. Then we define the basis function as

$$b_{lm}(x) = (K_d A_l + (1 - K_d) e^{-\frac{l^2}{2s}}) Y_{lm}(n), \tag{6}$$

where $x = [n_x, n_y, n_z, K_d]^T$. We use $b(x)$ to represent the transpose of the vector of basis functions $b_{lm}(x)$, and $L$ to represent the vector of lighting coefficients $L_{lm}$. Therefore, assuming $s$ is known or under independent estimation, Eq. (3) can be simplified as

$$I = b(x)L, \tag{7}$$

which is as simple as the product of basis functions and lighting coefficients in a $(l^* + 1)^2$-dimensional space. This function is linear to $L$ but nonlinear to $x$.

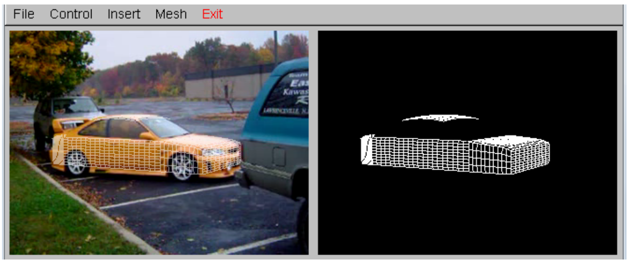



**Fig. 1**. The GUI for shape initialization.

## 3. INITIAL SETTINGS OF SHAPE AND TEXTURE

Without known texture and BRDF, the problem of illumination learning is severely ill-conditioned. In fact for real applications, shapes and reflectance characteristics of objects are usually unknown. Therefore, we specify the object category as vehicle, since it has simple shape that is easy to approximate, and uniform material with the same reflectance characteristic. We also perform optimization on the shape and texture while estimating the lighting coefficients to relieve restrictions on the initial settings.

Observing that only the surface normals are required in Eq. (3), we adopt a generic model that approximates the shape of a given vehicle to initialize its normals. The generic model has 5 surfaces with 6 length parameters. Users can easily adjust those parameters to generate an approximate model through a GUI, as shown in Fig. 1.

For the initial texture, we adopt a color classifier inspired by the work in [1] to group colors. We observe that the changes in color between pixels are due to either illumination or texture. When surfaces are Lambertian, any color changes due to illumination should affect all three color channels proportionally. When surfaces have low specularities, the chromatic changes due to illumination between two adjacent pixels should bring three color channels approximately proportional effect. Assuming two adjacent pixels with the same albedo have color $c_1$ and $c_2$, their directions in RGB color space should be close. Otherwise, the chromaticity of colors is changed and color changes are caused by the changes of texture not illumination. To group colors associated with the same texture class, we treat each color as a vector in RGB space and normalize them. We then compute the angle between normalized color $\vec{c_1}$ and $\vec{c_2}$ as

$$(\vec{c_1}, \vec{c_2}) = \arccos(\vec{c_1} \cdot \vec{c_2}). \tag{8}$$

If the angle $(\vec{c_1}, \vec{c_2})$ is below a threshold, we assign them into the same group. Finally, we compute the average of each group of colors as the texture assigned to this group.

## 4. ILLUMINATION LEARNING

We design a joint energy function of $x$ and $L$ inspired based on the errors-in-variables model [9], given by

$$E(x, L) = \lambda_1 E_x(x) + \lambda_2 E_f(x, L), \tag{9}$$

where $\lambda_1$ and $\lambda_2$ are coefficients. We take $x$ as the variable with corresponding error $\delta x$ in the initial setting, and $L$ as parameter going to be estimated. The first term is sum of square of error $\delta x_i$, expressed as

$$E_x(x) = \sum_i (\delta x_i)^2. \tag{10}$$

The second term is the weighed quadratic sum of residual error $f_i(x_i, L) = b(x_i)L - I_i$, given by

$$E_f(x, L) = \sum_i w_i f_i(x_i, L)^2, \tag{11}$$

where $w_i$ is a weight function defined as

$$w_i = \left[ \left( \frac{\partial f_i}{\partial x_i} \right)^T \left( \frac{\partial f_i}{\partial x_i} \right) \right]^{-1}. \tag{12}$$

We solve this joint minimization problem iteratively, with estimations of $L$ and $x$ in the each iteration. Taking the derivative of energy function $E$ with respect to $L$, it yields

$$\frac{\partial E}{\partial L} = 2[S_L - C_L]L - 2\sum_i w_i b_i I_i, \tag{13}$$

with the weighted scatter matrix and the covariance matrix:

$$S_L = \sum_i w_i b_i^T b_i, \quad C_L = \sum_i (w_i f_i)^2 \left( \frac{\partial b_i}{\partial x_i} \right)^T \left( \frac{\partial b_i}{\partial x_i} \right).$$

Let the derivative in Eq. (13) be zero, and lighting coefficients $L$ can be estimated by solving linear equations,

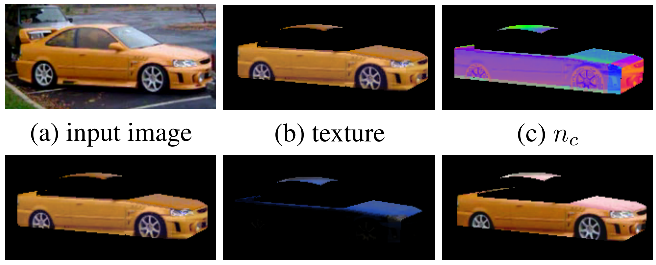
$$[S_L - C_L]\, L = \sum_i w_i b_i I_i. \tag{14}$$

In the same iteration, we also take the derivative with respect to $x_i$, which yields

$$\frac{\partial E_i}{\partial x_i} = 2\lambda_1 \delta x_i + \lambda_2 (2 w_i f_i \frac{\partial f_i}{\partial x_i} + \frac{\partial w_i}{\partial x_i} f_i^2). \tag{15}$$

By ignoring quadratic term and letting the derivative to be zero, we can estimate $\delta x_i$ by

$$\delta x_i = -\frac{\lambda_2}{\lambda_1} w_i f_i \frac{\partial f_i}{\partial x_i}. \tag{16}$$



(a) input image (b) texture (c) $n_c$

(d) diffuse (e) specular (f) rendered

**Fig. 2**. Example of illumination learning: The input image (a) is decomposed as texture (b), normals (shown are the normalized complements) (c), (d) diffuse lighting and (e) low-specular lighting, which generate the rendered result (f).

**Fig. 3**. De-lighting results. The input images (first row) are processed by three estimation methods with the same settings: the method in [7] (second row), the method in [8] (third row), and our method (bottom row).

Then we can update variables $x_i$ by estimated $\delta x_i$. In implementation, we also impose normalization constraints on normals and the white light source.

An example of illumination learning with $s = 8$ is shown in Fig. 2, where the image is spanned in a 25-dimensional space. The estimated diffuse lighting illuminates all directions reflected by the environment, and the specular lighting highlights the top of the car. The complements of recovered normals are normalized $n_c = \frac{1-n}{|1-n|}$ to visualize the details.

## 5. APPLICATIONS AND EXPERIMENTS

We apply our method to image de-lighting and re-lighting, referring to removing and transferring lighting effects of input images. The de-lighting technique can be further applied to illumination-invariant algorithms in computer vision, while the re-lighting technique is an intuitive way to examine the performance of illumination learning, with a byproduct of lightening other 3D models.

**De-lighting** is performed by factorizing texture from input images. To preserve more details, we smooth the input

(a) input image (b) target image (d) rendered image

**Fig. 4**. Re-lighting on a de-lit image: The input image (a) is re-rendered (c) by the illumination learned from (b).

**Fig. 5**. Re-lighting on a 3D model: Illumination of input images (first row) is transferred to a 3D model (second row).

image using a Gaussian filter, and record the pixel-wise ratios between the input image and its smoothed version. Then we apply the ratios to the estimated texture to get the de-lit image. Fig. 3 shows some indoor results of de-lighting. There the input images (first row) are processed by three estimation methods with the same settings: the method in [7] (second row), the method in [8] (third row), and our method (bottom row). The two previous methods only embody the diffuse component in their illumination model. Hence the specularity can be found in their delit images.

**Re-lighting** is performed by permutation of lighting coefficients $L$ and basis functions $b(x)$. A new image $I'$ can be rendered by the combination of its basis function $b(x)$ and learned illumination $L'$:

$$I' = b(x)L', \tag{17}$$

if the BRDF and texture are known. Fig. 4 shows a re-lighting on a de-lit image, where the outdoor image (a) is re-rendered (c) by the illumination learned from (b). It is more clear to examine the illumination learning by transferring illumination to 3D models with known shape and texture, as shown in Fig. 5. There illumination of input images (first row) is transferred to a 3D model (second row). Please note the specular effects are transfered as well.

## 6. CONCLUSION

We have detailed a practical method to adaptively learn illumination from single images of vehicles, which alleviates the demanding requirements on Lambertian restriction and known shape and texture. The goal of this method is to span images on a low-dimensional space and estimate the lighting coefficients. So far cast shadows and occlusions are not considered. It can not guarantee to acquire the accurate shape and texture of the given object. In fact, details of shape and texture are not necessary to be distinguished, since the majority of lighting is approximated by a low-dimensional vector. We apply this method to image de-lighting and re-lighting. Experimental results demonstrate that our method can work in both indoor and outdoor environments. Moreover, our method can couple the vehicle model fitting method [10] to further omit the generic model.

## 7. REFERENCES


[1] M. F. Tappen, W. T. Freeman, and E. H. Adelson, "Recovering intrinsic images from a single image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 9, pp. 1459–1472, 2005.

[2] J. Zhang, L. McMillan, and J. Yu, "Robust tracking and stereo matching under variable illumination," in *CVPR*, 2006, pp. 871–878.

[3] R. T. Tan, K. Nishino, and K. Ikeuchi, "Separating reflection components based on chromaticity and noise analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 10, pp. 1373–1379, 2004.

[4] S. P. Mallick, T. Zickler, D. J. Kriegman, and P. N. Belhumeur, "Beyond lambert: Reconstructing specular surfaces using color," in *CVPR*, 2005.

[5] R. Ramamoorthi and P. Hanrahan, "A signal-processing framework for inverse rendering," in *SIGGRAPH*, 2001, pp. 117–128.

[6] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 218–233, 2003.

[7] L. Zhang, S. Wang, and D. Samaras, "Face synthesis and recognition from a single image under arbitrary unknown lighting using a spherical harmonic basis morphable model," in *CVPR*, 2005, pp. 209–216.

[8] Y. Wang, Z. Liu, G. Hua, Z. Wen, Z. Zhang, and D. Samaras, "Face re-lighting from a single image under harsh lighting conditions," in *CVPR*, 2007, pp. 1343–1350.

[9] B. Matei and P. Meer, "A general method for errors-in-variables problems in computer vision," in *CVPR*, 2000, pp. 18–25.

[10] Z. Kim and J. Malik, "Fast vehicle detection with probabilistic feature grouping and its application to vehicle tracking," in *ICCV*, 2003.