Relighting for an Arbitrary Shape Object Under Unknown Illumination Environment

Yohei Ogura(✉) and Hideo Saito

Keio University, 3-14-1 Hiyoshi, Kohoku, Yokohama, Kanagawa 223-8522, Japan
{y.ogura,saito}@hvrl.ics.keio.ac.jp

Abstract. Relighting techniques can achieve the photometric consistency in synthesizing a composite images. Relighting generally need the object's shape and illumination environment. Recent research demonstrates possibilities of relighting for an unknown shape object or the relighting the object under unknown illumination environment. However, achieving both tasks are still challenging issue. In this paper, we propose a relighting method for an unknown shape object captured under unknown illumination environment by using an RGB-D camera. The relighted object can be rendered from pixel intensity, surface albedo and shape of the object and illumination environment. The pixel intensity and the shape of the object can simultaneously be obtained from an RGB-D camera. Then surface albedo and illumination environment are iteratively estimated from the pixel intensity and shape of the object. We demonstrate that our method can perform relighting for a dynamic shape object captured under unknown illumination using an RGB-D camera.

1 Introduction

Relighting is a technique to change appearance of objects in images captured in arbitrary illumination. We can achieve the photometric consistency in composite images synthesized from different images taken under different illuminations by using the relighting technique. Users can change the lighting condition for the object as if they are in other illumination environment. We consider that this technique can be applied for the entertainment experience or lighting simulation.

Relighting technique generally requires the shape and the surface albedo of the target object, and the illumination environment of the scene. There are some relighting approach with off-line processing or on-line processing with known geometry owing to the difficulty of exacting the object's shape and the surface albedo. Therefore, these methods generally can handle a single image or a static object in a movie. Not only shape and surface albedo of the target object, but also illumination environment are indispensable to relighting. Illumination estimation has been one of the important research issues in computer vision field. Illumination environment can be estimated based on inverse rendering, without any light probes. When we implement the illumination estimation method based on inverse rendering, the pixel intensity, the shape, and the surface albedo of the object are needed. However, the surface albedo of arbitrary object is unknown

G. Bebis et al. (Eds.): ISVC 2015, Part II, LNCS 9475, pp. 433–442, 2015.
DOI: 10.1007/978-3-319-27863-6_40

in many cases. It is not easy to estimate the illumination environment without any known surface albedo or light probes such as a mirror ball or cast shadows. In other words, achieving relighting methods which can be used for arbitrary shape object captured under unknown illumination is still a challenging task.

In this paper, we propose a relighting method for an arbitrary shape object captured under unknown illumination environment by using an RGB-D camera. A relighted object is rendered from pixel intensity, surface albedo and shape of the object and illumination environment. The pixel intensity and shape of the object are obtained from RGB-D camera. The surface albedo and the illumination environment are iteratively estimated from the pixel intensity and shape of the object. All of those properties can be estimated via on-line process for each frame in a movie sequence so that our implementation can be used for a dynamic shape object.

2 Related Work

Debevec et al. proposed a method to acquire the reflectance field of a target object using Light Stage, which has one spotlight which spirals around a target [1]. They take 2048 images under the different light conditions to estimate the reflectance function of the object. Not only they can get relighting result, but also they can change the viewpoint from the reflectance function. Wenger et al. implement newer Light Stage with high speed cameras to take larger number of images [2]. Their method can be applied to the dynamic shape object by using their new Light Stage with high speed cameras. These methods can obtain high quality relighting results thanks to their special equipment.

Zhen et al. proposed a relighting technique for human faces using 3D morphable face model [3]. This method builds on a ratio-image based technique. The advantage of this method is that single face image is required as input. However, the normal map should be estimated from generic human face model in off-line process. Aldrian et al. proposed a face relighting method considering not only diffuse component but also specular component so that more natural relighting results can be obtained [4]. These techniques don't consider acquisition of the illumination environment but simply assume that Illumination environment should be obtained in an arbitrary manner.

Wang et al. proposed a method that relights human faces from a single image under unknown lighting conditions [5]. They integrate spherical harmonics into the 3D morphable model to represent a human face captured under arbitrary unknown lighting and pose by representing them with three dimensional parameters.

However, those methods [3–5] assume that the input is a static human face image. Obtaining shape of the target from 3D morphable models, it is difficult to apply to the whole body of human or other objects.

3 Proposed Method

Our goal is to relight an arbitrary shape object recorded under unknown illumination environment. Our system consists from two parts: illumination estimation part and relighting part. Input data are color images and depth images from an RGB-D camera. Illumination environment is estimated from pixel intensity, surface albedo, and normal map of an arbitrary object existing in the image based on an inverse rendering technique without any light probes that has been proposed by Ramamoorthi and Hanrahan [6]. The normal map of a target object is obtained from depth images. After the normal map estimation, we segment the input color image into some regions. On each region, we set the average pixel intensity as initial albedo. After estimating illumination environment from the normal map and the initial albedo, we re-estimate the surface albedo from the estimated illumination and the normal map. We repeat this process until the improvement of the estimation is lower than a threshold (Fig. 1).

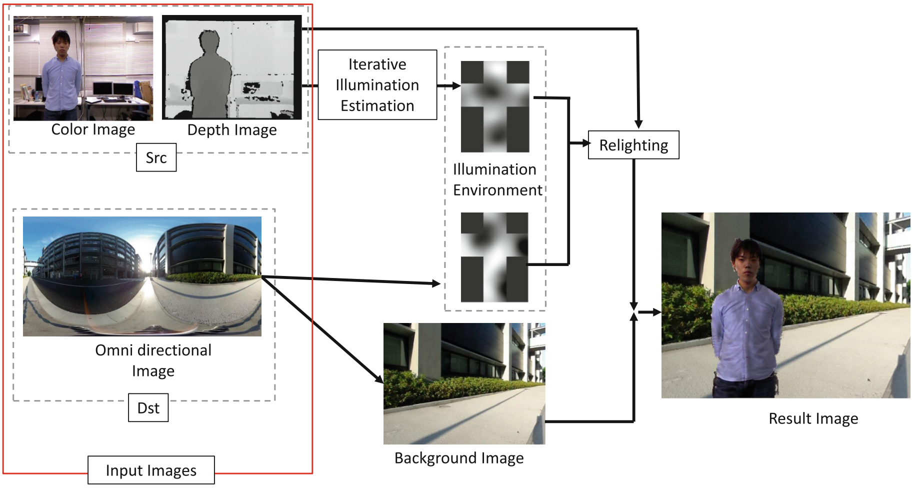


Fig. 1. System overview

After estimating illumination environment, we relight the object to match the illumination of the object with the background scene. Relighting process is done with pixel intensity, normal map and estimated illumination environment. We implement Image based rendering and Spherical Harmonics approximation. We calculate the ratio of the inverse rendering equation of the both scene to get the relighting result. Finally, the target object is superposed to the background image to get the final result. From Sects. 3.1 to 3.4, we introduce the theory of normal map estimation and spherical harmonics relighting.

3.1 Normal Map Estimation

Our relighting method uses normal map in the illumination estimation section and the relighting part. Simply calculating the normal map from two vectors calculated from neighbor pixels, the result may not be good because of noises on depth images. Before the normal map estimation, we apply bilateral filter and temporal filter to depth images. The bilateral filtered depth map D_b is obtained from raw depth map$D$ by using following equation:

$$D_b(\boldsymbol{u}) = \frac{1}{k'(\boldsymbol{u})} \sum_{\boldsymbol{v} \in \Omega_{g'}} g_{s'}(\boldsymbol{u}, \boldsymbol{v}) g_d(D(\boldsymbol{u}), D(\boldsymbol{v})) D(\boldsymbol{v}) \tag{1}$$

Note that $g_{s'}(\boldsymbol{u}, \boldsymbol{v})$ is the spatial Gaussian weight and $g_d(I(\boldsymbol{u}), I(\boldsymbol{v}))D$ is the color similarity Gaussian weight. $k'(\boldsymbol{u})$ is a normalize factor and $\Omega_{g'}$ is a square window whose center is $\boldsymbol{u}$. After applying bilateral filter, we also apply the temporal filter [7]. A current depth image is denoised by using a current frame and a previous frame.

$$D_{tf}(\boldsymbol{u}) = \frac{(D_b(\boldsymbol{u})(w+1) + D_{b-1}(\boldsymbol{u})w)}{2w+1} \tag{2}$$

w is a constant weight term. After denoising the depth image, we can obtain a vertex map corresponding to a camera coordinate since we assume that the camera's intrinsic parameters are known. We estimate the normal map based on the method proposed by Holzer et al. [8]. This method can generate the smooth normal map. However, this method cannot estimate the normal vector around the boundary of the object where the difference of the depth value is too large. We obtain normal vectors by calculating a cross product of two vectors from the neighbor points on these areas. Normal vector $\boldsymbol{N}(\boldsymbol{u})$ at a point $\boldsymbol{u} = (u, v)$

$$\boldsymbol{N}(\boldsymbol{u}) = (V(u+1, v) - V(u, v)) \times (V(u, v+1) - V(u, v)) \tag{3}$$

$V(u, v)$ is a vertex map that corresponds to a camera coordinate. Normal map for illumination estimation and relighting is obtained by merging Holzer's method [8] and Eq. (3).

3.2 Spherical Harmonics Lighting

The purpose of this section is to explain the illumination estimation theory using Spherical Harmonics. Relationship between pixel intensity, albedo and normal vector is presented by Ramamoorthi et al. [6]. This is called "Inverse Rendering". We assume that the light source is distant and objects in the scene have lambertian surfaces.

The irradiance $E(\boldsymbol{x})$ observed at a point $\boldsymbol{x}$ is given by an integral on the distant sphere Ω

$$E(\boldsymbol{x}) = \int_{\Omega} L(\boldsymbol{\omega}) max((\boldsymbol{\omega} \cdot \boldsymbol{n(x)}), 0) d\boldsymbol{\omega} \tag{4}$$

$L(\boldsymbol{\omega})$ is incoming light intensity along the direction vector $\boldsymbol{\omega} = (\theta, \varphi)$ and $\boldsymbol{n}(\boldsymbol{x})$ is normal vector at a point $\boldsymbol{x}$. Then, $\boldsymbol{\omega} \cdot \boldsymbol{n}(\boldsymbol{x})$ is a dot product of a normal vector and incoming light direction. $max((\boldsymbol{\omega} \cdot \boldsymbol{n}(\boldsymbol{x})), 0)$ means that if a normal vector and incoming light vector has the same direction, the light from that direction is fully considered, but if the angle of these two vector is more than 90 degree, we exclude the light from that direction.

It takes too much cost to calculate Eq. (4) regarding the illumination as a aggregate of point sources. Spherical Harmonics(SH) approximation is good way to reduce the calculating cost. The illumination is shown with SH basis function y and the coefficients c The Eq. (4) will be represented in the following equation.

$$E(\boldsymbol{x}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} A_l(\theta) L_{l,m} Y_{l,m}(\boldsymbol{\omega}) \tag{5}$$

l denotes the SH band. There are $2l + 1$ functions in band l, and m denotes the index in a band. $A_l(\theta)$ is the SH projection of $max((\boldsymbol{\omega} \cdot \boldsymbol{n}(\boldsymbol{x})), 0)$. It is obtained by rotating the standard cosine term A_l^{std} which is equal to $A_l(0)$ [9]. $Y_{l,m}$ is the basis function of SH and $L_{l,m}$ is the coefficient of each SH basis function. The pixel intensity $I(\boldsymbol{x})$ from a point $\boldsymbol{x}$ on a image is written as

$$I(\boldsymbol{x}) = R_d \sum_{l=0}^{\infty} \sum_{m=-l}^{l} A_l(\theta) L_{l,m} Y_{l,m}(\boldsymbol{\omega}) \tag{6}$$

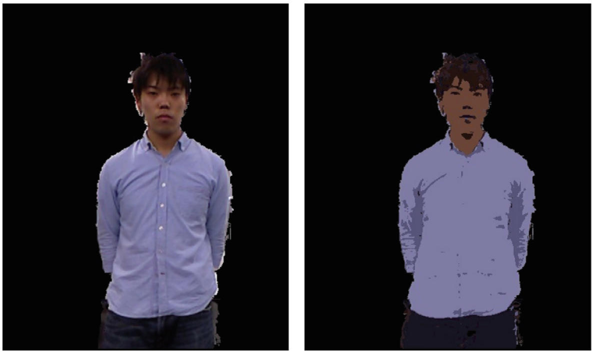
R_d represents albedo. Since we consider lambertian surface uniformly reflects the incoming light to all direction, R_d is set to constant values.

3.3 Albedo and Illumination Estimation

The purpose of this part is to estimate illumination environment by employing inverse rendering method represented as Eq. (6). Illumination environment can be estimated from the pixel intensity, the normal map and the surface albedo. However, albedo is also unknown property in most cases. Estimating these two properties at the completely same time is ill-posed problem. So, we employ an iterative method to obtain both albedo and illumination environment.

Objects in real scene generally have varying albedo on each point. However, precisely adopting this assumption takes too much cost and makes our method hard to be done. We assume that the region with the same color and the same material, such like a monochromatic T-shirt, have uniform albedo value. We subtract the target region by thresholding depth values. After that, we segment a color image by k-means segmentation to obtain divided regions depending on objects as a first step as shown in Fig. 2. For example in case of that the target is a human, a shirt, a jacket, pants, etc., that can be regions with uniform albedo. For each region, we set an average pixel value as an initial albedo of each region.

After that, we iteratively solve Eq. (6) by using the least-square method to obtain illumination environment. Such iterative approach to estimate albedo



(a) Color Image (b) Segmented Image

Fig. 2. Image segmentation

under unknown illumination is based on Zou et al. [10]. We implement this method to estimate illumination with unknown albedo values. The iterative illumination estimation is performed as the following procedure. First, initial illumination is estimated with initial albedo value. Next, albedo is updated by fixing the illumination that was estimated in previous step. Then, illumination is also updated with updated albedo. This procedure is repeated until SH coefficients values are converged. We assume that the illumination color is white, so we handle Eq. (6) in gray scale.

3.4 Relighting Using Spherical Harmonics

In this part, we explain the relighting theory. At first, we define the name of each scene. "Src" is the scene where has the object to be relighted and "Dst" is the other illumination environment that is different from Src. We can calculate the pixel intensity of the relighted object fitting to Dst illumination environment from the ratio of Eq. (6) [3]. Relighting result is obtained from following equation.

$$\boldsymbol{I}_{Dst}(\boldsymbol{x}) = \boldsymbol{I}_{Src}(\boldsymbol{x}) \frac{\sum_{l=0}^{2} \sum_{m=-l}^{l} A_l(\theta) L_{l,m}^{Dst} Y_{l,m}(\boldsymbol{\omega})}{\sum_{l=0}^{2} \sum_{m=-l}^{l} A_l(\theta) L_{l,m}^{Src} Y_{l,m}(\boldsymbol{\omega})} \tag{7}$$

$\boldsymbol{I}_{Dst}(\boldsymbol{x})$ is a pixel intensity of the relighted object and $\boldsymbol{I}_{Src}(\boldsymbol{x})$ is an original pixel intensity of the object to be relighted. We can get the final result image by superposing the relighted target object on Dst background image. Ramamoorthi et al. showed that nine SH coefficients are needed to approximate diffuse reflectance [11]. Based on this method, we use nine SH coefficients and set the supremum of l to 2.

Illumination of Dst scene can be obtained by applying the same illumination estimation method as Src, by using SH projected Omnidirectional images, or synthesized cubemap images. In the case of using the same way as Src to get Dst illumination, we need at least one lambertian object in Dst scene.

4 Experiments

In this section, we present experiments performed to demonstrate the effectiveness of the proposed method. We capture an image of a person under unknown illumination environment by using an RGB-D camera. Albedo of objects such as a blue shirt in the scene are also unknown. We iteratively estimate albedo and illumination environment, and relight the person subsequently. Finally, we get the final result by superposing the person to the background images.

4.1 Experiment Condition

We relight a person who is in a room but the illumination environment is unknown. He is wearing a blue shirt. We estimate the illumination from the region of a target person. Note that 3D reconstruction is hard to be applied because the person moves his body. We also need the illumination environment of Dst scene. In this experience, we obtained the illumination from omnidirectional images captured by omnidirectional camera. The background for resulting composite images is also obtained from these omnidirectional images. Dst scene has sunlight casted from left side. Both images are shown in Fig. 3. We use Microsoft Kinect as an RGB-D Camera and assume that camera intrinsic parameters for converting depth image to vertex map are known.

(a) Src scene (b) Dst scene

Fig. 3. Src and Dst scenes

4.2 Experiment Result

Src scene is indoor scene that has fluorescent lamps but Dst scene is outdoor scene. As shown in Fig. 4, superposing a person without relighting causes unnatural appearance, because his body does not have any shadow area, while the building wall behind the person strongly shaded. Note that our method doesn't have off-line process because we estimate the surface albedo and illumination iteratively, and obtain the normal map from depth images on each frame. Therefore, our method can also be applied for dynamic shape target. Focusing on front

15^{st} frame 182^{th} frame 222^{th} frame 319^{th} frame

Fig. 4. Relighting results. Each caption shows frame number. Upper column shows composite images without relighting. Lower column shows relighting results.

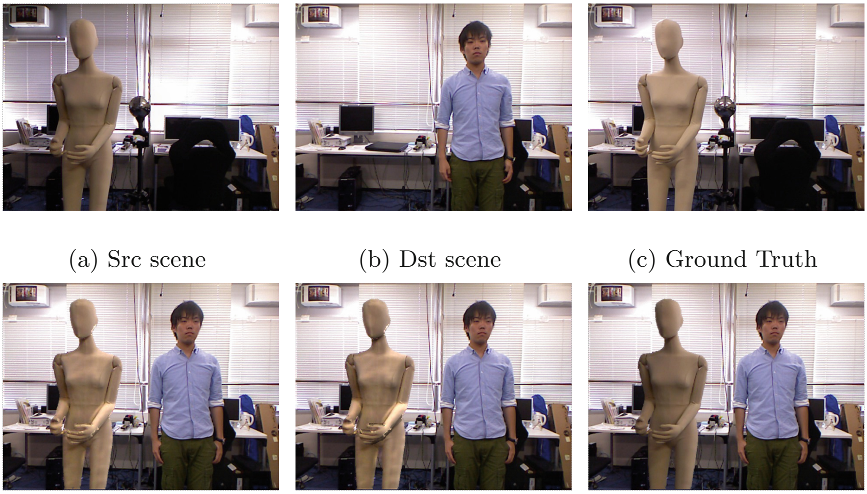
of his body, the brightness are varying according to the angle of the body. Our method can relight dynamic shape targets under unknown illumination to fit other scene which has different illumination and can realize the photometric consistency on composite images.

4.3 Quantitative Evaluation

To evaluate the performance of our method, we compare the results with a ground truth data. We calculate root mean square error values between each result and the ground truth image. The ground truth is captured in the same room with Src scene, but consists of the same illumination with Dst scene. We relight the mannequin and calculate RMS error values on each of pixels of the mannequin region and get average values. Root mean square error values are shown in Table 1 with other two results. The one is relighting result with illumination estimation method proposed by Gruber et al. [12]. Since this method doesn't provide the way of estimating albedo, we estimate the albedo of mannequin in Src scene and a blue shirt in Dst scene in off-line process. The other one is the result of a composite image without relighting. Those two relighted result images are using the same relighting algorithm [3]. The difference is the method of illumination estimation. Input scenes, the ground truth image and result images are shown in Fig. 5.

Table 1. Root mean square error value comparing with Ground Truth

	RMS Error(pixel value)
Proposed method	13.8
Illumination estimation by [12]	13.7
Without relighting	18.6



(a) Src scene (b) Dst scene (c) Ground Truth

(d) Proposed method (e) Illumination estimation by [12] (f) Without Relighting

Fig. 5. Comparison with ground truth

An advantage of our method is that we don't need obtain any albedo beforehand. The error value of the proposed method was similar to the result of illumination estimation by Gruber et al. [12]. We can say that we could estimate the illumination environment appropriately from this point. On the other hand, the error value of proposed method was lower than "Without relighting". Our method could also relight an unknown shape object appropriately.

5 Conclusion

In this paper, we proposed a relighting method for arbitrary shape objects captured under unknown illumination by using an RGB-D camera. We can obtain illumination environment without any light probes by the iterative illumination and albedo estimation. This means that we don't need to estimate albedo in off-line process. Thus our method can be used for arbitrary shape objects under unknown illumination. After the iterative illumination estimation, relighting process is done with estimated illumination data. Since our method obtains the normal map from depth images on each frame, a dynamic shape object can be applied to our method. In experiment, we tested our method to show the utility and quantitative evaluation. We will improve our method more robust, and also apply relighting to other purpose such as object tracking.

References

1. Debevec, P., Hawkins, T., Tchou, C., Duiker, H.P., Sarokin, W., Sagar, M.: Acquiring the reflectance field of a human face. In: ACM SIGGRAPH. ACM (2000)
2. Wenger, A., Gardner, A., Tchou, C., Unger, J., Hawkins, T., Debevec, P.: Performance relighting and reflectance transformation with time-multiplexed illumination. ACM Trans. Graph. **24**(3), 756–764 (2005)
3. Zhen, W., Liu, Z., Huang, T.S.: Face relighting with radiance environment maps. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE (2003)
4. Aldrian, O., Smith, W.: Inverse rendering with a morphable model: A multilinear approach. In: Proceedings of the British Machine Vision Conference. BMVA Press (2011)
5. Wang, Y., Zhang, L., Liu, Z., Hua, G., Wen, Z., Zhang, Z., Samaras, D.: Face relighting from a single image under arbitrary unknown lighting conditions. IEEE Trans. Pattern Anal. Mach. Intell. **31**, 1968–1984 (2009)
6. Ramamoorthi, R., Hanrahan, P.: A signal-processing framework for inverse rendering. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. ACM (2001)
7. Matsumoto, K., Song, C., de Sorbier, F., Saito, H.: Joint upsampling and noise reduction for real-time depth map enhancement. In: Proceedings of IS&T/SPIE Electronic Imaging, SPIE (2014)
8. Holzer, S., Rusu, R.B., Dixon, M., Gedikli, S., Navab, N.: Adaptive neighborhood selection for real-time surface normal estimation from organized point cloud data using integral images. In: IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE (2012)
9. Nowrouzezahrai, D., Simari, P., Fiume, E.: Sparse zonal harmonic factorization for efficient SH rotation. ACM Trans. Graph. **31**, 23 (2012)
10. Zou, X., Kittler, J., Hamouz, M., Tena, J.R.: Robust albedo estimation from face image under unknown illumination. In: SPIE Defense and Security Symposium, SPIE (2008)
11. Ramamoorthi, R., Hanrahan, P.: An efficient representation for irradiance environment maps. In: ACM SIGGRAPH. ACM (2001)
12. Gruber, L., Richter-Trummer, T., Schmalstieg, D.: Real-time photometric registration from arbitrary geometry. In: IEEE International Symposium on Mixed and Augmented Reality. IEEE (2011)