Inverse Rendering Based on Compressed Spatiotemporal Infomation by Neural Networks

Eito Itonaga, Fumihiko Sakaue and Jun Sato
Nagoya Institute of Technology, Nagoya, Japan

Keywords: Inverse Rendering, Photometric Stereo, Light Distribution Estimation.

Abstract: This paper proposes a method for simultaneous estimation of time variation of the light source distribution, and object shape of a target object from time-series images. This method focuses on the representational capability of neural networks, which can represent arbitrarily complex functions, and efficiently represent light source distribution, object shape, and reflection characteristics using neural networks. Using this method, we show how to stably estimate the time variation of light source distribution, and object shape simultaneously.

1 INTRODUCTION

Inverse rendering, which is the estimation of object shape, reflection characteristics, and light distribution from a single image or multiple images, has been actively studied in the fields of computer graphics as well as computer vision in recent years. Inverse rendering from a single image is generally regarded as an ill-posed problem because it requires more information than the input information as an image. Even when multiple images are used as input, it is known that estimating many unknown parameters such as object shape, reflection characteristics, and light source distribution simultaneously is difficult due to the degrees of freedom and representation methods of these parameters. Therefore, the way these parameters are represented is very important from the standpoint of estimation stability.

Therefore, in this study, we focus on the ability of neural networks to represent arbitrarily complex functions and represent object shape and light source distribution using neural networks. This enables all of the object shapes and light distribution to be represented as model parameters of the neural network. By simultaneously optimizing these parameters, we can simultaneously estimate object shape, reflection characteristics, and light source distribution from time series images observed in an environment where the light source distribution changes over time.

2 RELATED WORKS

Many methods for estimating object shape, reflection characteristics, and light distribution from images have been studied. One method for estimating light distribution from images is based on the diffuse reflection component on the object's surface. In this method, the distribution of light sources is estimated by expressing the light distribution using a spherical harmonic function and obtaining the expansion coefficients of the spherical harmonic function. However, the light distribution estimation method based on diffuse reflections is known to be prone to an ill-posed problem because high-frequency components are missing from the image(Marschner and Greenberg, 1997). In recent years, there have also been methods that estimate light distribution directly from images using convolutional neural networks based on a large amount of training data(Georgoulis et al., 2018). In addition, the method (LeGendre et al., 2019) has also been proposed to estimate the light source distribution by calculating the difference between the image generated from the estimated light source distribution and the image actually taken as a loss and training the network. These methods have been confirmed to have higher performance than conventional methods that use physics-based vision. However, they require the collection of a very large amount of data in order to perform proper training.

Several methods for estimating object shape and reflection characteristics have been proposed using the photometric stereo method, which estimates the normal vector of the object surface representing the

Itonaga, E., Sakaue, F. and Sato, J.
Inverse Rendering Based on Compressed Spatiotemporal Infomation by Neural Networks.
DOI: 10.5220/0011792200003417
In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) - Volume 4: VISAPP*, pages 467-474
ISBN: 978-989-758-634-7; ISSN: 2184-4321

object shape from multiple images taken at different light source positions(HIGO, 2010). These methods can simultaneously estimate object shape and reflection characteristics under conditions where the light source is known. Similar to light source information estimation, a neural network-based method has also been proposed to simultaneously estimate object shape, reflection characteristics, and light distribution from a single image taken of an indoor scene(Sengupta et al., 2019). However, this method also requires a large amount of training data to achieve appropriate learning.

Therefore, in this paper, we consider the use of neural networks to represent object shape and light distribution without using training data and to simultaneously estimate object shape, reflection characteristics, and light distribution from time-series images. This method aims to realize inverse rendering with properties that are intermediate between physics-based vision and learning-based vision.

3 OBSERVATION MODEL

3.1 General Rendering Model

First, we discuss the general rendering model. Rendering is the computation of light reflections in a scene based on specular reflections on object surfaces, shadows, inter-reflections between objects, etc. The rendering equation proposed by Kajiya(Kajiya, 1986) is well-known as a basic mathematical model. Therefore, rendering in computer graphics corresponds to solving this rendering equation.

Assuming that no light penetrates into the interior of the object, the reflectance at the observation point x as $f(x,\vec{\omega},\vec{\omega'})$. This reflectance indicates the ratio of rays incident from the $\vec{\omega'}$ direction reflected in the $\vec{\omega}$ direction. The angle between the incident direction $\vec{\omega'}$) and the plane normal direction $\vec{n}$ is shown by θ. The set of ray directions incident on point x is denoted by Ω. In this case, the observed intensity L_o is expressed by the rendering equation as follows:

$$L_o(x,\vec{\omega}) = L_e(x,\vec{\omega}) + \int_\Omega f_r(x,\vec{\omega'},\vec{\omega}) L_i(x,\vec{\omega'}) \cos\theta d\vec{\omega} \quad (1)$$

where L_e is the amount of light emitted from the point $\vec{x}$. This equation indicates that the observed intensity is determined by the integral of the reflected light incident from all directions and the light $L_e(x,\vec{\omega})$ emitted from the object in the $\vec{\omega}$ direction. Therefore, the observed intensity information includes not only the reflectance characteristics at the point, but also information about the surrounding lighting environment. Therefore, if we can analyze this information appropriately, we can reconstruct various types of information from the intensity information.

3.2 Intensity Model

Next, we describe the image observation model used in this study. In equation (1), L_o is the luminance emitted from the observation point x on the object surface in the scene in the direction ω, and L_e is the radiance of light emitted from the object interior. *Omega* is the direction of incidence of the light at the observation point x, which coincides with the hemisphere. Here, L_e can be assumed to be zero regardless of the direction because the light is rarely emitted from the interior of a typical object. The $\cos\theta$, is the just inner product of the normal vector $\vec{n}$ and the direction of incidence $\vec{\omega'}$. Thus, the following equation, which focuses only on the reflection component, is used as the observation model of the image in this study.

$$L_r(x,\vec{\omega}) = \int_\Omega f_r(x,\vec{\omega'},\vec{\omega}) L_i(x,\vec{\omega'}) (\vec{n}\cdot\vec{\omega'}) d\vec{\omega} \quad (2)$$

When using such an observation model, light distribution and object shape information for the entire scene, as well as reflectance property information, are required to render the image. Conversely, it is possible to estimate these information from the observed intensity information.

3.3 Light Distribution Representation

Next, we describe the representation model of light distribution used in this study. As described in the equation (2), the light distribution around the object is necessary to determine the intensity observed on the object's surface. However, it is difficult to directly express this as a continuous quantity. Therefore, the In this study, we use a geodesic dome as shown in Fig 1. A geodesic dome is an approximate model of a sphere composed of triangular patches, which can discretely represent the spread of points on a sphere. The vertices are distributed with equal density on the sphere, so it is possible to represent light distribution equally at isosceles angles. In this study, we assume that a point light source exists at the center of gravity of these triangular patches, and represents light distribution by changing the intensity of each light source. In this study, the temporal variation of light distribution is represented by changing the intensity of each light source on the sphere according to the time.

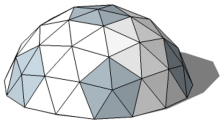


Figure 1: Geodesic Dome.

3.4 Reflectance Model

Next, we describe the representation of reflectance characteristics used in this study. The BRDF (Bidirectional Reflectance Distribution Function) is a typical model of the reflectance characteristics of an object. The BRDF shows how incident light reflects from one direction to another on the surface of an object in each direction. Therefore, the BRDF can be used to describe how an object is observed from various directions under various light sources. Considering the reflection at the observation point x on the object surface, the BRDF at the observation point x depends on the incident direction of light (θ_i,ϕ_i) and the viewpoint direction (θ_r,ϕ_r), as shown in Figure2. Although the BRDF is strictly wavelength-specific, it is often redundant to describe the reflectance of each wavelength for rendering purposes, so it is common practice to define a BRDF for each of the three RGB channels. With the BRDF defined in this way, the ratio of the intensity in the incident direction $\vec{\omega}'$ to the intensity in the view direction $\vec{\omega}$ at the observation point x can be described as follows:

$$f_r(x, \vec{\omega}', \vec{\omega}) \tag{3}$$

By this function f, the reflectance property of the surface can be described.

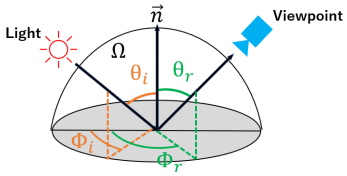


Figure 2: Directions for BRDF.

4 SIMULTANEOUS ESTIMATION OF LIGHT DISTRIBUTION AND SHAPE INFORMATION

Inverse rendering is to obtain the object shape $\vec{n}$, reflection property f_r, and light distribution L_i in Eq.(2) from the observed intensity, and the difficulty of the problem varies greatly depending on the representation method and degree of freedom. Many of the methods proposed so far have focused on the frequency characteristics of the light distribution and used spherical harmonic functions to represent the light distribution. Such methods have been effective in light distribution estimation focusing on diffuse reflections, in which the reflected components consist only of low-frequency components. However, in analyses focusing on specular reflections, which contain higher-frequency components, the number of degrees of freedom increases significantly when attempting to represent the high-frequency components of the light distribution, and this makes the estimation unstable. Furthermore, when time series information is used, the number of degrees of freedom increases further, making stable estimation difficult. Therefore, we propose a method to simultaneously estimate light distribution, object shape, and reflection characteristics from time-series images by using a neural network that can represent arbitrarily complex functions.

4.1 Definition

Based on the above, we define and formulate the problem addressed in this paper. In this paper, we take as input multiple images taken continuously in a situation where the light source environment changes with time, and estimate the temporal change of the light distribution and the shape of the photographed object from these images. The object shape and light distribution are represented as a regression using a neural network. That is, given a certain coordinate on the light distribution, there is a function that returns the intensity of the light source. Similarly, for object shape, given a certain coordinate on the image, there is a function that returns the normal direction. Therefore, the estimation of object shape and light distribution corresponds to training a neural network that can appropriately represent the input image.

To solve such a problem, Eq.(2) is redefined in terms of geodesic domes and neural networks. Assuming that the object-camera relationship is fixed and the shape of the object does not change with time, the viewpoint direction in the observed image can be considered fixed. Assume that T time images are available as input. Assume that the light distribution at time $t(t = 1,2,\cdots,T)$ is sampled using a geodesic dome with G sampling points as $(\theta_j,\phi_j)(j = 1,2,\cdots,G)$ and the direction from the object to each light source is $vecs_j$ and the light source intensity $\hat{E}_{j_t}$ at time t, the observed intensity can be expressed as follows:

$$\hat{I}_t = \sum_{j=1}^{G} f_r(\vec{n}(x), s_j)\hat{E}_{j_t}(\vec{n}\cdot\vec{s_j}) \tag{4}$$

Here $\hat{I}_t$ is the re-rendered image at time t rendered from the estimated object shape $\vec{n}$, reflection property f_r and light distribution $\hat{E}_{j_t}$. In this case, $\hat{E}_{j_t}$ and $\vec{n}(x)$ are functions represented by a neural network, and estimating them is the objective of this study.

4.2 Light Distribution Representation by Neural Networks

This section describes the details of the representation of light distribution using neural networks. As mentioned above, in this study, light distribution is represented as a regression function using a neural network. In this case, estimating the light distribution corresponds to estimating the parameters of the neural network that composes this regression function. Such a regression function representation using a neural network has been used for various applications, such as 3D shape representation and estimation called NeRF(Mildenhall et al., 2020), and is known for its stable and appropriate representation of functions with a high degree of freedom. However, it is known that high-frequency components within a function cannot be represented properly when coordinates representing space are used as direct input to the function. To avoid this problem, a method called Positional Encoding is used. In this method, each coordinate is mapped to a higher-order, higher-dimensional space using trigonometric functions, etc., and they are used as input to the regression function. In this study, Positional Encoding is applied to the light source direction $\vec{s}_j$ and time t, and the computed values are used as input to a function that expresses the time variation of the light distribution. The details of Positional Encoding are described in the next section, as there are several possible variations depending on the function used for the mapping.

4.3 Positional Encoding

Let us describe details of the Positional Encoding. This method is used to represent the 3D shape as a model parameter of a neural network in NeRF(Mildenhall et al., 2020) that has recently attracted attention for restoring the 3D shape of a scene from multiple image data and generating images from a new viewpoint. This method is used to represent 3D shapes as parameters. Neural networks are known to have a bias to learn only low-frequency components, as it is difficult to represent high-frequency functions whose outputs vary in a variety of ways to minute changes in the input when learning. Therefore, by embedding the input to the neural network in a higher space with a function such as the one shown in Eq.(5), the neural network itself only needs to represent low-frequency functions, and the neural network as a whole can represent higher frequencies.

$$\gamma(\vec{p}) = ((sin(2^0\pi p), cos(2^0\pi p), sin(2^0\pi p), (cos(2^0\pi p), \cdots, sin(2^{L-1}\pi p), cos(2^{L-1}\pi p), sin(2^{L-1}\pi p), cos(2^{L-1}\pi p)) \quad (5)$$

where $\vec{p}$ is the input vector of the neural network, and Equation (5) is applied to each element of the input vector separately. Therefore, as patterns of Positional Encoding in this study, we consider two patterns of functions that apply Positional Encoding to the light source direction $\vec{s}_j$ and time t as shown in Eqs(6) (7).

$$\gamma(\vec{s}_j, t) = ((sin(2^0\vec{s}_j), cos(2^0\vec{s}_j), sin(2^0\pi\frac{t}{T}), cos(2^0\pi\frac{t}{T})\cdots, sin(2^{L-1}\vec{s}_j), cos(2^{L-1}\vec{s}_j), sin(2^{L-1}\pi\frac{t}{T}), cos(2^{L-1}\pi\frac{t}{T})) \quad (6)$$

$$\gamma(\vec{s}_j, t) = ((sin(2^0\vec{s}_j)sin(2^0\pi\frac{t}{T}), cos(2^0\vec{s}_j)cos(2^0\pi\frac{t}{T})\cdots, sin(2^{L-1}\vec{s}_j)sin(2^{L-1}\pi\frac{t}{T}), cos(2^{L-1}\vec{s}_j)cos(2^{L-1}\pi\frac{t}{T})) \quad (7)$$

4.4 Shape Representation by Neural Networks

4.4.1 Representation of Object Shape

In this study, object shapes are represented using normal maps, which directly represent the normals of object shapes. The normal map is a map showing the normal direction $\vec{n} = (n_x, n_y, n_z)$ for each point in the image, and in this research, the normal direction is expressed using polar coordinate representation as follows.

$$\begin{cases} n_x = \sin\theta + \cos\phi \\ n_y = \sin\theta + \sin\phi \\ n_z = \cos\theta \end{cases} \quad (8)$$

where, θ and ϕ represent latitude and longitude, respectively, and the normal direction can be determined by determining these two parameters. In this research, this normal map is represented by a neural network in the same way as light distribution. That is, we define a neural network that takes the coordinates (x, y) in the image as input and these two parameters as output, and perform the estimation.

$$\gamma(x,y) = ((sin(2^0\frac{x}{H}), cos(2^0\frac{x}{H}), sin(2^0\pi\frac{y}{W}), cos(2^0\pi\frac{y}{W})\cdots, sin(2^{L-1}\frac{x}{H}), cos(2^{L-1}\frac{x}{H}), sin(2^{L-1}\pi\frac{y}{W}), cos(2^{L-1}\pi\frac{y}{W})) \quad (9)$$

The encoded $\gamma(x, y)$ is input to the neural network and normal map will be obtained from the network.

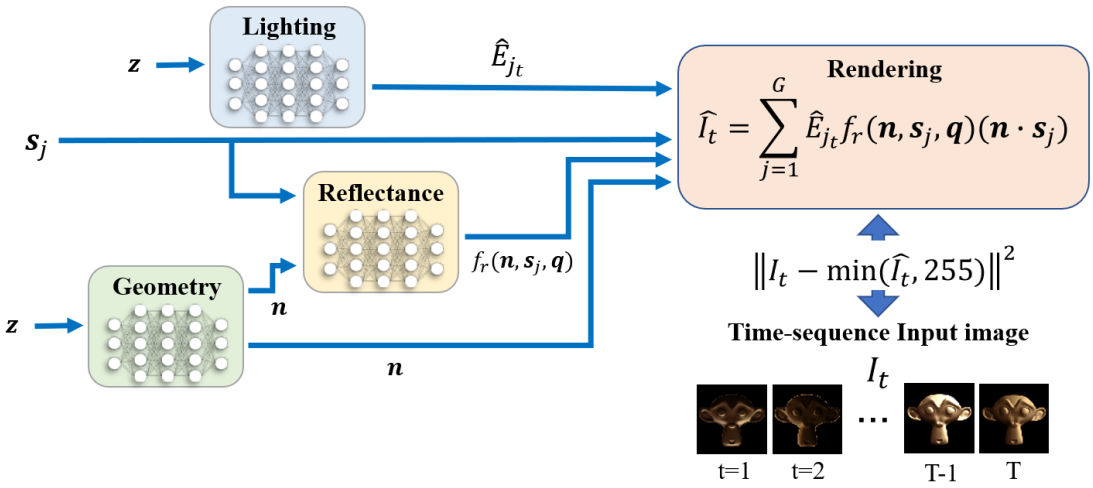


Figure 3: Intensity representation by several neural networks for representing phisics parameters.

4.4.2 Representation of BRDF

Reflection characteristics can also be represented using neural networks in the same way as the light distribution and object shape described above. However, the BRDF representing reflection characteristics is considered to have fewer degrees of freedom than the light distribution and normal distribution described above, so it is not efficient to represent them in the same way. Therefore, a neural network that takes as input the embedding vector, the incident direction, and the emission direction that represents the reflection characteristics is trained in advance, and the reflection characteristics are estimated by estimating the parameters of the embedding vector. Considering that the coordinate system in the BRDF is determined by the normal direction of the object surface and that the viewpoint direction is fixed in the scenes in this study, the reflectance determined by the BRDF can be redefined as a function of the normal direction and light source direction as inputs. Therefore, the function f representing the BRDF is redefined as follows:

$$f(\vec{n}, \vec{s}, g(\vec{q})) \tag{10}$$

where, $\vec{q}$ is a representation of the object's material label as a one-hot vector, and g is an embedding function that transforms it into a feature vector. By learning these f and g using a database of reflection characteristics measured in advance, a neural network that can efficiently represent BRDF can be constructed. This learning can be achieved by minimizing the following loss function.

$$\varepsilon = \sum_{\vec{q}} \sum_{\vec{n}} \sum_{\vec{s}} (\hat{f}(\vec{n}, \vec{s}, g(\vec{q})) - (f(\vec{n}, \vec{s}, g(\vec{q})))^2 \tag{11}$$

where $\hat{f}$ is the BRDF given by the training data. Using a neural network trained in this way, an efficient representation of the BRDF can be achieved. In addition, by estimating the embedding vector according to the input image, the estimation of reflection characteristics can be performed.

Instead of estimating one-hot vectors in the estimation of reflection characteristics, feature vectors representing reflection characteristics embedded in lower dimensions are estimated. This is in anticipation of the possibility of representing various BRDFs by combining the characteristics of existing BRDFs, even when the reflective properties of the input object are not included in the BRDF database used for training.

We have confirmed that the method described in this section can efficiently represent reflection properties and that it can appropriately estimate reflection properties when the normal direction and light distribution are known. However, the simultaneous estimation of object shape, light distribution, and reflection characteristics has not been sufficiently verified. Therefore, in the experiments described below, the simultaneous estimation of reflection characteristics is not performed, but light distribution and object shape are estimated assuming that the BRDF of the target is known. Simultaneous estimation including reflection characteristics is a subject for future work.

4.5 Estimation by Training

The light distribution, normal distribution, and reflection characteristics can be expressed by the above. Using these, the observed intensity in this study can be shown as Figure3. This figure shows that the observed intensity is represented as the output of a neural network that follows a physical model. Therefore, the estimation of each parameter is equivalent to train-

Figure 5: Input images.

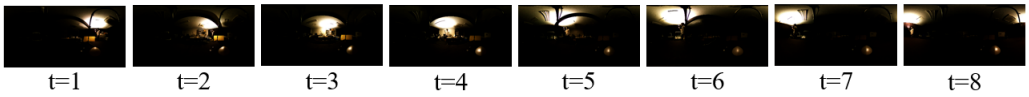


Figure 6: Obtained light distribution by omnidirectional camera.

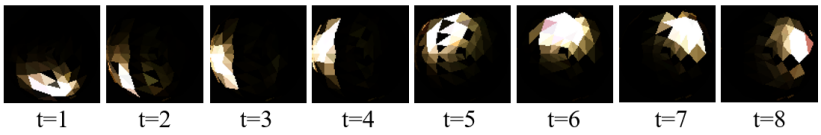


Figure 7: Light distribution map using geodesic dome.

ing this entire neural network with the input images. In other words, all parameters can be estimated by computing $\hat{I}_t$ based on Eq.(4) as the output of each network and training the neural network so that the error between the result and the input image is minimized.

However, when specular reflection is observed using a camera that does not have sufficient dynamic range, the observed values may exceed the range of image representation. In this case, the intensity values in the input image do not follow the physical model, and a loss function that takes this into account is required. To consider the case, we define the loss function as follows:

$$\varepsilon_e = \sum_t \sum_x \sum_y \| I_t(x,y) - min(\hat{I}(x,y)_t, I_{max}) \|^2 \quad (12)$$

In this loss function, the re-rendering result is replaced by I_{max}, the maximum value that can be represented by the image, so that the re-rendering result itself can output values that exceed the maximum value. By simultaneously optimizing the model parameters of the neural networks and the vector $\vec{q}$ to minimize this loss function, object shape, reflection characteristics, and light distribution can be simultaneously estimated from time series images only.


(a) Normal map


(b) Reflection characteristics

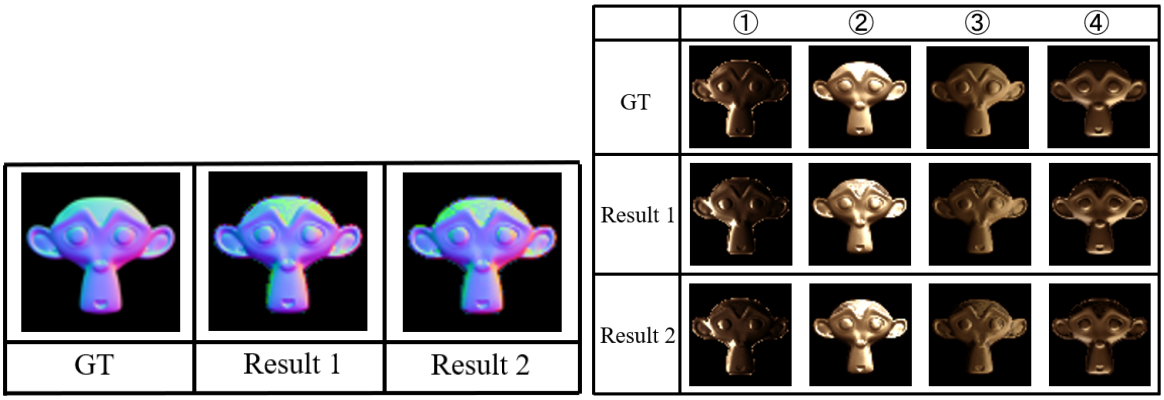
Figure 4: Information of the target object.

5 EXPERIMENT

5.1 Environment

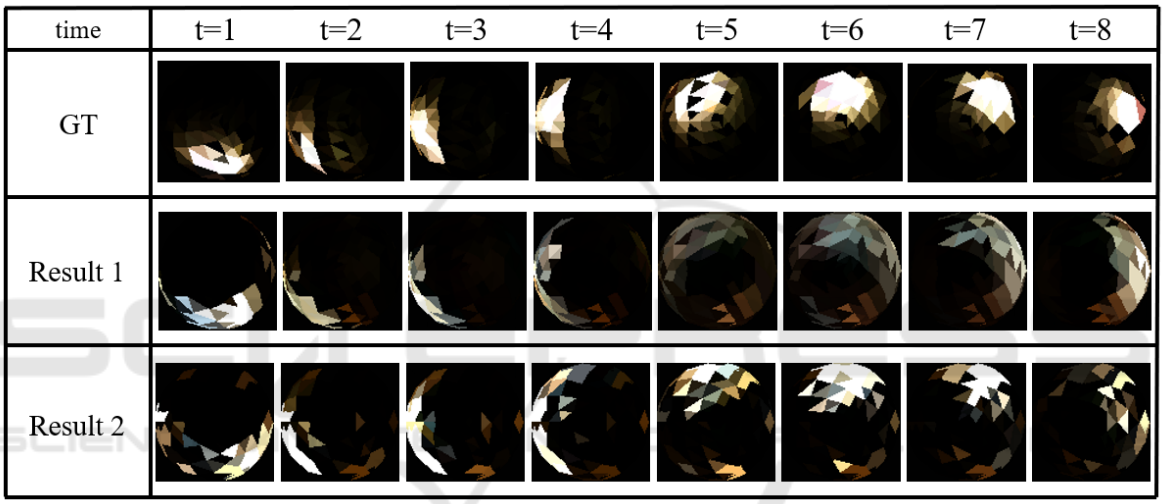
The results of inverse rendering using the previously described methods are presented. First, we describe the experimental environment. In this experiment, we used the input images for 8 time periods as shown in Fig.reffig:input.This time-series input image is a composite image rendered using the normal map shown in Fig.4(a), the BRDF representing the texture of leather in the UTIA BRDF Databaset(Filip and Vávra, 2014) shown in Fig.4(b), and the light distribution in Fig.7.The light distribution was created from an environment map obtained by using an omnidirectional camera to capture a light source environment that changes with time in a darkroom environment, as shown in Figure6. The number of geodesic dome samplings used to represent the light distribution was set to 192. In the environment described above, the proposed method was used to simultaneously estimate light distribution and normal distribution.

The neural network representing object shape was tested by inputting the values obtained by applying the Positional Encoding of Eq.(9) to the coordinate (x,y) on the image, and the neural network representing light source distribution was tested by inputting the values obtained by applying the Positional Encoding of Eq.(6) and (7) to the light source direction $\vec{s_j}$ and time t.

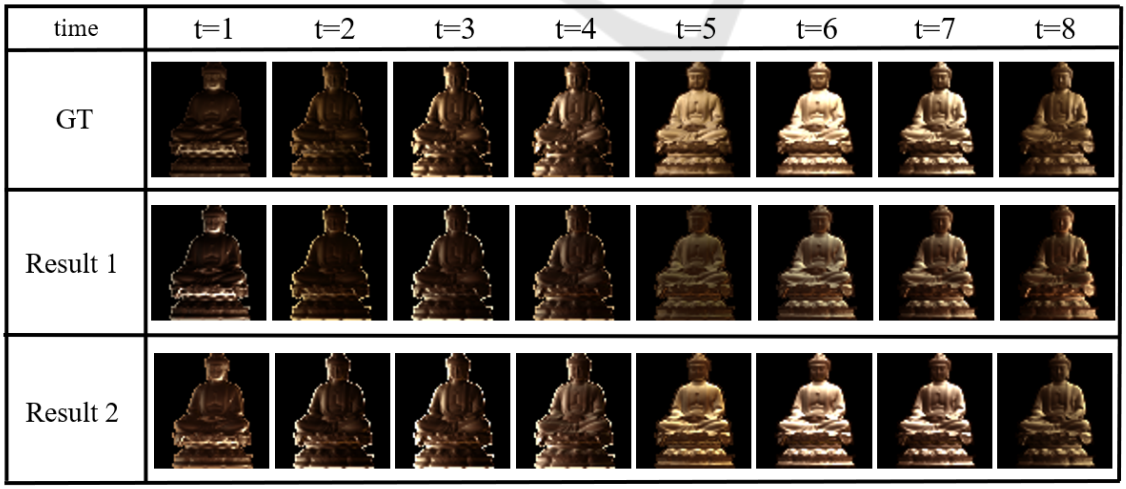


(a) Estimated object shape. (b) Re-rendered image.

Figure 8: Estimated results (object shape).



(a) Estimated light distribution.



(b) Re-rendered image.

Figure 9: Estimated results (light distribution).

5.2 Results

The results of the object shape estimation are shown in Figure 8(a), and the results of the re-rendering under different 4 pattern light distributions using the estimated object shape are shown in Figure 8(b). The results of the light distribution estimation are shown in Figure 9(a), and the results of the re-rendering for different object shapes using the estimated light distribution are shown in Figure 8(b). Result 1 is the estimation result when Positional Encoding of the Eq.(6) is applied to the light direction $\vec{s_j}$ and time t of the light distribution, and Result 2 is the estimation result when Positional Encoding of the Eq.(7) is applied to the neural network representing the light distribution.

From the results of object shape estimation, it can be confirmed that both Result 1 and Result 2 are close to the ground truth. The results of re-rendering under different light distributions using the estimated object shape also show a representation that is close to the ground truth, confirming that the object shape is correctly estimated. The estimated light distribution is also confirmed to be close to the ground truth, especially in Result 2. It can also be confirmed that the estimated light distribution is correctly re-rendered for different object shapes. From the above, it can be confirmed that the object shape and light distribution are generally correctly represented even when neural networks are used, and that the object shape and light distribution are correctly estimated in the simultaneous estimation of object shape and light distribution. Thus, the effectiveness of the proposed method was confirmed.

6 CONCLUSION

In this study, we proposed a method to simultaneously estimate object shape and light distribution in inverse rendering using only time-series images by representing object shape and light distribution using a neural network, without using training data. Experiments were conducted to confirm the effectiveness of the proposed method. In the future, we will consider a method for simultaneous estimation including reflection characteristics.

REFERENCES

Filip, J. and Vávra, R. (2014). Template-based sampling of anisotropic BRDFs. *Computer Graphics Forum*, 33(7):91–99.

Georgoulis, S., Rematas, K., Ritschel, T., Gavves, E., Fritz, M., Van Gool, L., and Tuytelaars, T. (2018). Reflectance and natural illumination from single-material specular objects using deep learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(8):1932–1947.

HIGO, T. (2010). *Generalization of Photometric Stereo: Fusing Photometric and Geometric Approaches for 3-D Modeling*. PhD thesis, University of Tokyo.

Kajiya, J. T. (1986). The rendering equation. In *Proceedings of the 13th annual conference on Computer graphics and interactive techniques*, pages 143–150.

LeGendre, C., Ma, W., Fyffe, G., Flynn, J., Charbonnel, L., Busch, J., and Debevec, P. E. (2019). Deeplight: Learning illumination for unconstrained mobile mixed reality. *CoRR*, abs/1904.01175.

Marschner, S. R. and Greenberg, D. P. (1997). Inverse lighting for photography. In *Color and Imaging Conference*, volume 1997, pages 262–265. Society for Imaging Science and Technology.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. (2020). Nerf: Representing scenes as neural radiance fields for view synthesis. *CoRR*, abs/2003.08934.

Sengupta, S., Gu, J., Kim, K., Liu, G., Jacobs, D. W., and Kautz, J. (2019). Neural inverse rendering of an indoor scene from a single image. In *International Conference on Computer Vision (ICCV)*.