

Shape Reconstruction based on Diffusion of Captured Images through Deep Learning using Position Data of The Light Sources

Hiroto SUTO*, Ikumi HIROSE* and Norimichi TSUMURA

Abstract: Three-dimensional reconstruction techniques are used in various situations such as visual inspections in a factory and digital archiving in a museum. One such method is photometric stereo, in which a change in the direction of the light source is used to illuminate an object. This method can reconstruct objects with high accuracy. By contrast, this approaches have certain limitations, such as the requirement for a darkroom environment with no more light sources than necessary, and the need for a Lambert diffuse reflection material that reflects the incident light equally in all reflection directions. Therefore, a method for transforming the specular reflection components of an object into a diffuse form using deep learning was previously developed. However, this method applies a network in which the object surface is not changed despite the change in the position of the light source. It is therefore not applicable to a shape reconstruction using photometric stereo, which applies a change of the position of the light source. In this study, we propose an improved learning network that transforms captured specular images into a diffuse form while reflecting the changes in the surface luminance by applying images that differ in the position of the light source and its corresponding data. In addition, we verified the accuracy and effectiveness of a conventional method for both the transformed 2D images and the 3D shape model reconstructed from these images. As a result, we confirmed that the accuracy of our method is improved in both cases.

Key words: 3D reconstruction, Shape reconstruction, Photometric stereo, Deep learning, GAN

1. Introduction

There has been a need for 3D shape measurements of objects and a 3D reconstruction on a computer, which applies measurement information such as a visual inspection of industrial products and the digital archiving of cultural properties. Photometric stereo^{1,2)} is one of the most popular methods for a 3D reconstruction. This method applies multiple images in which the direction of the illuminating light source is changed, and utilizes the relational expression estimating the luminance by the inner product of the light source vector and the angle of the normal of the object for reconstructing the shape. It therefore implements a highly accurate shape reconstruction because the normal is calculated for each pixel of the image. By contrast, other approaches have certain limitations, such as the requirement of a darkroom environment with no more light sources than necessary, and the need for a Lambert diffuse reflection material that reflects the incident light equally in all reflection directions. Therefore, a method for transforming the specular reflection components of an object into a diffuse form using a generative adversarial network (GAN), which is a leaning network developed by Wu *et al.* However, this method does not change the object surface despite the change in the position of the light source. Thus, it is inapplicable to a shape reconstruction using photometric stereo, which uses a change in the position of the light source.

In this study, we propose a new learning network that transforms

captured specular images into a diffuse form while reflecting changes in the surface luminance by using images that differ in terms of the position of the light source and position data of the light source. This learning network utilizes the position data of the light source described as category information of a one-hot vector in addition to images that differ in the position of the light source. Thus, it enables a shape reconstruction of the specular surface from environments in which more than a single light source is illuminated, such as in photometric stereo. This network allows us to convert the specular part of the captured image into a diffuse form, which is necessary for photometric stereo, and enables a shape reconstruction by applying the photometric stereo for objects with specular surfaces using these images transformed through our learning network. In addition, it can avoid the limitations of not only surface reflections but also the capturing condition because the diffuse and specular images used in this network are respectively darkroom and non-darkroom images.

2. Related work

This section describes the related research on GANs, which is the basis of the proposed learning network. The basic structure is based on the Cycle-GAN proposed by Zhu *et al.*³⁾, which is a network that trains a generator and a discriminator by using two sets of images with features before and after the transformation, transforming one image into the other, and then reconverting the transformed image

Received 6th October, 2021; Accepted 12th March, 2022; Published 24th March, 2022

Department of Imaging Sciences, Graduate school of Engineering, Chiba University
1-33, Yayoi-cho, Inage-ku, Chiba-shi, Chiba 263-8522, Japan

* Hiroto SUTO and Ikumi HIROSE contributed equally to this work.

to be closer to the image prior to the transformation. This network features a mechanism that returns the image transformed by the generator back into the image prior to transformation. It is possible to train the learning network such that this loss decreases by calculating the cycle consistency loss, which is represented by the difference between the dataset image before the transformation and the dataset image that has been re-transformed to be closer to the image prior to the transformation using the generator.

Our method was also inspired by the network developed by Wu *et al.*, which uses this Cycle-GAN mechanism and transforms the image from a specular form into a diffuse form for a 3D shape reconstruction using images⁴⁾. This method applies a learning network that specializes in converting the surface reflection of a specular reflecting object into a diffuse reflection and introduces cycle consistency loss similar to that of a Cycle-GAN. However, because this method is intended to be applied along with an approach combining two types of shape reconstruction^{5, 6)}, i.e., Structure from Motion⁷⁾ and Shape from Shading⁸⁾, it is unsuitable for a shape restoration in photometric stereo simply because of a problem in which the image transformation cannot reflect the direction of the light source position as shown in Fig. 1.

In this study, we use a StarGAN⁹⁾ learning network to achieve an image transformation, which includes the information of the light source position to solve this problem. In this research, the feature of adding category information describing the information of the light source position at the time of training makes it easier to determine the change in the light source position from the image alone. In StarGAN, the category information is used to update the weights using the classification information of the image transformation. In addition, to transform images by manipulating the category information, StarGAN has the structure of a single generator. The categorical information in StarGAN prevents learning from proceeding by focusing on features that are not the target of the transformation, which results in a higher accuracy than learning from images alone. In addition, the discriminator in StarGAN not only identifies whether an image has been transformed, by applying the ACGAN¹⁰⁾ method as an auxiliary approach allowing it to discriminate fine details, it also identifies the category of the image classification. Using this method to classify images, it is able to discriminate small changes in comparison to a general GAN. In this study, we incorporate the network structure of StarGAN into the method developed by Wu *et al.* and propose a new image transformation network that can transform captured images into images suitable for photometric stereo.

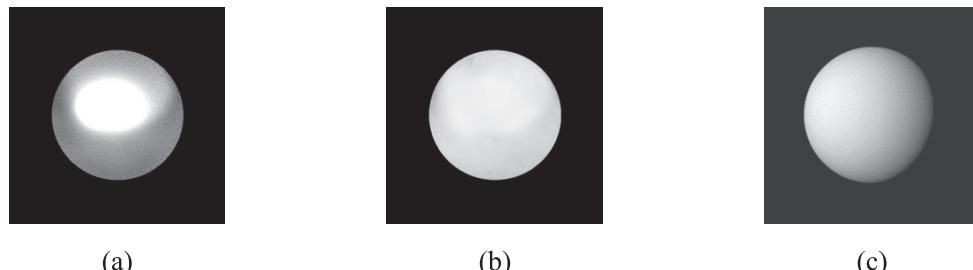


Fig. 1 Results of image transformation with learning network of Wu *et al.*⁴⁾ (a) Original image.
(b) Transformed image. (c) Ground truth.

3. Training dataset creation

In this study, we prepared a custom dataset to implement an image transformation that can be used for a shape reconstruction in photometric stereo. This dataset consists of two types of assets: images of objects with specular and diffuse surface reflections, and category information that describes the surface reflection and light source position within the images. In this section, we describe how these datasets were constructed.

3.1 Image dataset creation

We created the dataset images by rendering a 3D CG model in a computer. In this study, 10 different simple primitive shapes were applied, which can be divided into two main types of surface features, planar and curved, at a 1:1 ratio. The shape models include typical primitive shapes such as polyhedra, cones, or similar shapes such as spheres, pyramids, cones, cylinders, regular polyhedra, and torus. Some of the rendering images used are shown in Fig. 2.

As shown in Fig. 3, the captured object was set on the floor. In addition, the camera was placed at a fixed position in a direction vertical to the object.

The material of the object surface was set through the rendering on the computer, as shown in Fig. 3. In this study, we use physically based rendering¹¹⁾, which can quantitatively set the position, angle, and gloss of an object. We applied silver as the material for the specular reflection, and a matte material with a white color for the Lambert diffuse reflection. We set the parallel light to change the surface luminance, which is necessary for photometric stereo, and set the environmental light from outside for the diffuse material.

We applied the parallel light by setting up a directional light at infinity. We set up the light sources at six locations where the angle between the light sources and the floor surface is 60°, using the center of the object as the base axis, as shown in Fig. 4. We capture the images by turning on one light per shot. A 3D reconstruction can be achieved using images with light sources located in at least three positions using photometric stereo. However, we have frequently observed a collapsed estimation of the normal where there is no light exposure, and thus we used six light positions to ensure that the entire object is illuminated.

The environmental light is simulated on a computer using five different HDR environment maps. These environment maps are limited to 32-bit HDR images captured indoors. This is because, in this study, we suppose the use of photometric stereo in an indoor

environment. In this case, environment maps are rotated at a random angle based on the axis vertical from the floor to accommodate images captured in a variety of situations. The rendered image is a 24-bit color image. Although an image containing specular reflection should normally be a 32-bit image, because the purpose of this study is to convert a specular reflection image captured using an ordinary camera, we decided to apply a 24-bit color image.

3.2 Category information creation

In this study, we describe information regarding the surface reflection and light source positions as the category information. The learning network applied in this study uses such information to transform and classify the images. The category information is described through a one-hot vector using "1"s and "0"s. A vector with a 1 indicates a corresponding category, whereas a vector with a "0" represents a noncorresponding category. In addition, we classified images into two types of categories: surface reflection and light source position. As an example, the category information under a certain situation is as shown in Fig. 5. The value on the left side describes the surface reflection, and the other values indicate the light

source position. In addition, the generator can switch the target category to a different category by switching the "0" or "1" one-hot vectors using such information. In this study, we use this structure to transform images with a generator.

4. Building of a learning network

The learning model used for this approach applies a GAN consisting of two networks: a generator for image transformation and a discriminator for image discrimination. The architecture of our approach is shown in Fig. 6.

4.1 Generator and discriminator

The generator serves to transform image x in the dataset into image y , corresponding to the transformation target. The training network is based on the StarGAN structure and uses two types of information input into the generator, i.e., the images contained in the dataset and the category information as the target of the transformation. The architecture of the generator is shown in Fig. 7.

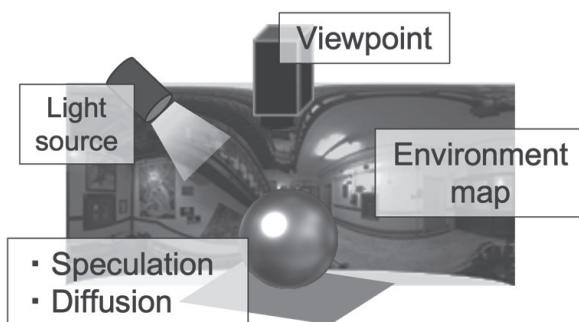


Fig. 3 Rendering environment.

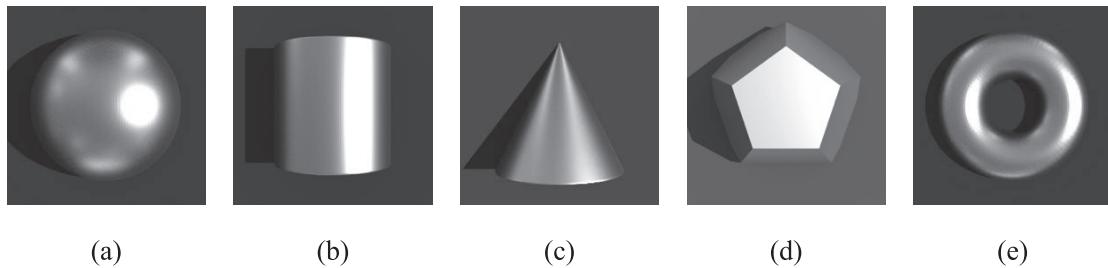
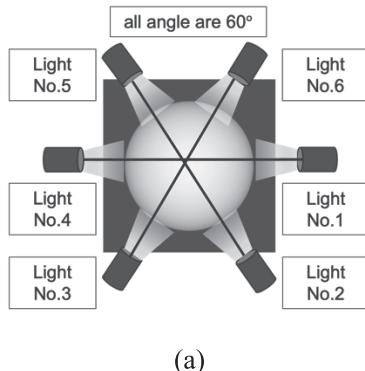
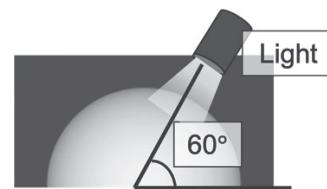


Fig. 2 Example of rendering images with specular surface used. (a) Sphere. (b) Cylinder. (c) Cone. (d) Regular dodecahedron. (e) Torus.



(a)



(b)

Fig. 4 Position of light source at rendering. (a) View from the shooting direction. (b) View from the direction horizontal to the floor.

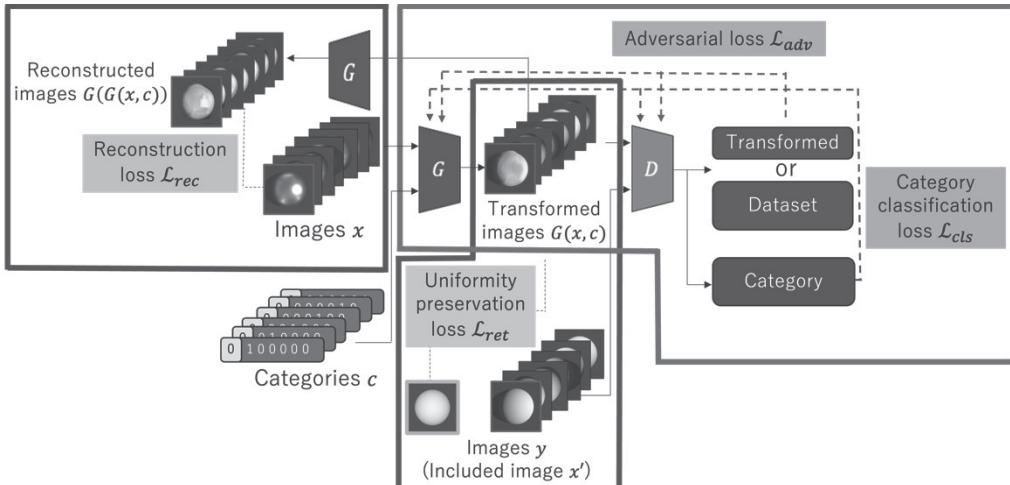


Fig. 6 Architecture and loss function of the proposed method.

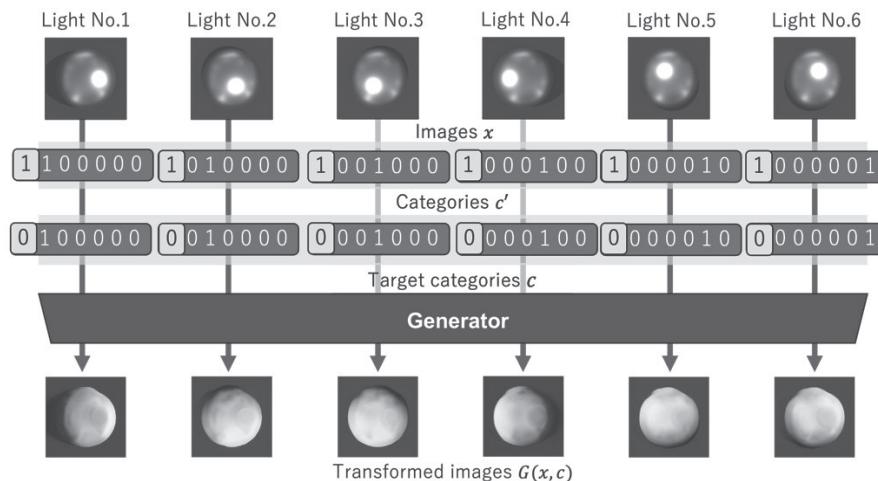


Fig. 7 Architecture of the generator using category information.

The target category information for the transformation is created by switching the one-hot vectors of the category information associated with the images in the dataset, as shown in Fig. 7. By inputting the image into the generator, the image can be converted into the target image. To achieve the purpose of this study in terms of the image transformation, we set some constraints related to the condition when choosing the images and category information, including always switching the one-hot vector of the surface reflection before and after the input of the generator, and fixing the light source position without such switching. In addition, as with a Cycle-GAN, this method introduces a consistency loss mechanism in which the image transformed by the generator is transformed again into the original image prior to the transformation, and the differences between them are calculated. Therefore, even when the surface reflection is diffuse, the transformation from a diffuse to specular surface is set up in the same way as when an image with a specular surface is input. This makes it possible to transform the image by changing the surface reflection while maintaining the information of the light source position.

The discriminator inputs the image $G(x, c)$ transformed by the generator and image y contained in the dataset, and then differentiates the input images. In this study, we discriminate the images in two ways: using the image itself, and by classifying the image based on category information. The architecture of the discriminator is shown in Fig. 8. The discriminator alternately inputs the image $G(x, c)$ transformed by the generator and dataset image y .

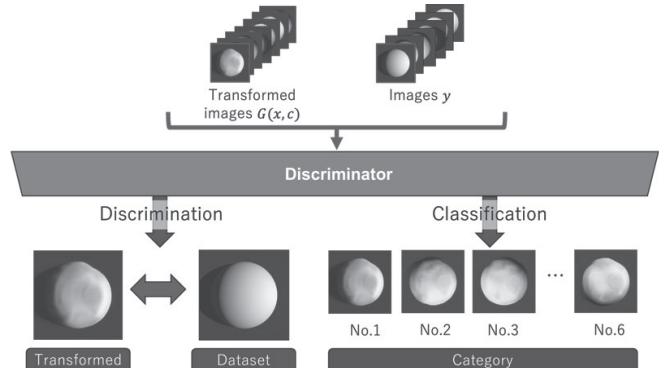


Fig. 8 Architecture of the discriminator with discrimination and classification functions.

First, it identifies whether the image input into the discriminator is an image transformed by the generator or a dataset image, as with a conventional GAN. However, using this method alone makes it

difficult to discriminate fine details such as changes in luminance. Thus, as an auxiliary approach, we use the method applying an AC-GAN¹⁰ to categorize the images. Such an approach is applied to classify the image input into the discriminator corresponding to the light source position. This makes it possible to discriminate small changes such as the position of the light source in comparison to a general GAN.

4.2 Loss function

We calculate the loss function to find the optimal weights for achieving a higher accuracy of the generator and the discriminator. The learning applied in this study is based on four loss functions, i.e., adversarial loss, category classification loss, and reconstruction loss based on a conventional StarGAN method, and uniformity preservation loss, which was newly developed for this study. When the adversarial loss is \mathcal{L}_{adv} ; the category classification loss is $\mathcal{L}_{\text{cls}}^G$ and $\mathcal{L}_{\text{cls}}^D$ for the generator and discriminator, respectively; the reconstruction loss is \mathcal{L}_{rec} ; and the uniformity preservation loss is \mathcal{L}_{ret} . In addition, the hyperparameters of each loss function are λ_{adv} , λ_{cls} , λ_{rec} , and λ_{ret} , respectively. The final loss functions of the generator and discriminator, \mathcal{L}_G and \mathcal{L}_D , respectively, are as shown in the following equations.

$$\mathcal{L}_G = \lambda_{\text{adv}} \mathcal{L}_{\text{adv}} + \lambda_{\text{cls}} \mathcal{L}_{\text{cls}}^G + \lambda_{\text{rec}} \mathcal{L}_{\text{rec}} + \lambda_{\text{ret}} \mathcal{L}_{\text{ret}} \quad (1)$$

$$\mathcal{L}_D = -\lambda_{\text{adv}} \mathcal{L}_{\text{adv}} + \lambda_{\text{cls}} \mathcal{L}_{\text{cls}}^D \quad (2)$$

Adversarial and category classification losses are functions that use the results of an image discrimination and classification by a discriminator. Adversarial loss is the loss of computing the probability distribution of whether image x input into the discriminator is contained in the original dataset or is transformed by the generator. The category classification loss is a function that calculates the probability distribution of whether image x input into the discriminator will be classified as category information c or c' . Thus, if expected value is \mathbb{E} , the adversarial and category classification losses can be expressed through the following equations. For example, if $D_{\text{adv}}(x)$ is able to output data close to reality, $\mathbb{E}_x[\log D_{\text{adv}}(x)]$ becomes larger. In addition, if $G(x, c)$ generates an output close to reality, $\mathbb{E}_{x,c} [\log(1 - D_{\text{adv}}(G(x, c)))]$ becomes larger. Thus, \mathcal{L}_{adv} will be close to zero.

$$\mathcal{L}_{\text{adv}} = \mathbb{E}_x[\log D_{\text{adv}}(x)] + \mathbb{E}_{x,c} [\log(1 - D_{\text{adv}}(G(x, c)))] \quad (3)$$

$$\mathcal{L}_{\text{cls}}^G = \mathbb{E}_{x,c}[-\log D_{\text{cls}}(c|G(x, c))] \quad (4)$$

$$\mathcal{L}_{\text{cls}}^D = \mathbb{E}_{x,c'}[-\log D_{\text{cls}}(c'|x)] \quad (5)$$

As shown in Eqs. (4) and (5), the loss function of the generator calculates the probability for the transformed image, and the discriminator calculates the probability for the original image to learn separately in the generator and discriminator.

The reconstruction loss is a function for the generator. It computes the difference between images transformed by the generator and those transformed again using the generator such that the original image x prior to input is restored. This takes on the role of creating correspondence between the images before and after the transformation. When the target category information of the transformation is c , and the original category information contained in the dataset is c' , the reconstruction loss \mathcal{L}_{rec} is expressed through the following equation.

$$\mathcal{L}_{\text{rec}} = \mathbb{E}_{x,c,c'} [\|x - G(G(x, c), c')\|_1] \quad (6)$$

We calculate the difference between the image transformed by the generator and the image prior to the transformation with the L1 norm. We train the generator such that the difference in the above equation decreases, and then build a learning network that enables an image transformation that maintains the consistency of the image.

There are problems with the loss function described above. First, images are transformed with high or low luminance regardless of changes in the position of the light source. Second, the transformation fails considerably in areas with low luminance, such as areas corresponding to shadows. Therefore, in this study, we introduce a uniformity preservation loss as a new loss function for applying a diffuse image illuminated toward an object from the same direction as the captured position, and then correcting the image from each light source position to be closer to this image. An image with a uniformity preservation loss is shown in Fig. 9.

This is a function that corrects for changes in the luminance of the surface to be transformed, with the goal of improving the accuracy of the transformation. The formula is computed using the difference between images based on the L1 norm, as with Eq. (6) for the reconstruction loss. When the image in the dataset is x , the categorical information targeted for transformation is c , and the image illuminated from the same direction as the captured direction is x' , the

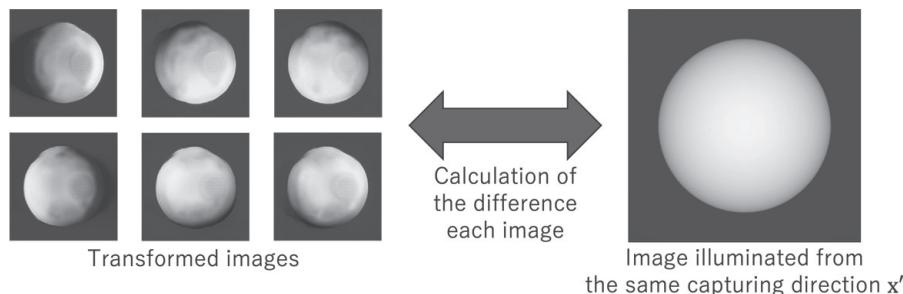


Fig. 9 Mechanism of uniformity preservation loss.

uniformity preservation loss \mathcal{L}_{ret} is expressed through the following equation.

$$\mathcal{L}_{ret} = \mathbb{E}_{x, x', c} [\|x' - G(x)\|_1] \quad (7)$$

This makes it possible to transform an image reflecting the luminance transformation of the surface owing to changes in the light source position while correcting the luminance of the transformed image.

4.3 Network structure

The network structure of the learning network is based on that used in StarGAN. The network structure of the generator and discriminator are shown in Fig. 10. The structure of the generator is roughly divided into three layers: a downsampling layer, a bottleneck layer, and an upsampling layer. The downsampling layer convolves images and extracts image features, the bottleneck layer transforms features based on the weights, and the upsampling layer deconvolves the features and restores the images from the feature map.

The ReLU function is used as the activation function for each layer. In addition, in the bottleneck layer, the structure of the residual network (ResNet)¹²⁾ is adopted to efficiently learn more detailed features. In addition, as with a general neural network, the structure of the discriminator consists of three layers: an input layer, a hidden layer, and an output layer. The input layer obtains a feature map from the image, the five hidden layers convolute to extract the features, and the output layer classifies and identifies such features.

4.4 Learning parameters

Two types of images, specular and diffuse, are used along with 200 patterns. In the case of specular images, we prepared six light sourc-

es positions, whereas in the case of diffuse images, to calculate the difference between the image illuminated from the same capturing direction from the same direction, seven images, including one additional image, were used for the calculation of the uniformity preservation loss. The resolution of each image was set to 256×256 owing to the learning time. In addition, although 32-bit images should be used for images containing specular reflections, 8-bit images are used for training in this study because the purpose is to convert images taken under general conditions. We set the learning rate to 0.0001 for both the generator and discriminator; the hyperparameters of each loss function to 1 for the adversarial and uniformity preservation losses, and to 10 for the category classification and reconstruction losses; and the number of iterations to 100,000, which is the number of times the best accuracy was achieved when we repeated the training of the model and checked the accuracy of the image transformation and shape reconstruction.

5. Verification of learning results using the proposed method

We transformed the images using the learning network developed in this study, reconstructed the shape, and compared the results with the ground truth. In addition, we verified the effectiveness of the newly introduced uniformity preservation loss. This time, among the polyhedra and cones included in the dataset for primitive shapes, we used pentagonal pyramids, faceted icosahedra, and faceted cones that were not applied for training, as shown in Fig. 11.

The evaluation of the transformed images is carried out in two ways: a qualitative evaluation based on the transformation results of the images, and a quantitative evaluation based on the root mean square error (RMSE) of the ground truth images. The ground truth image is an image of the target shape of the object created in a ren-

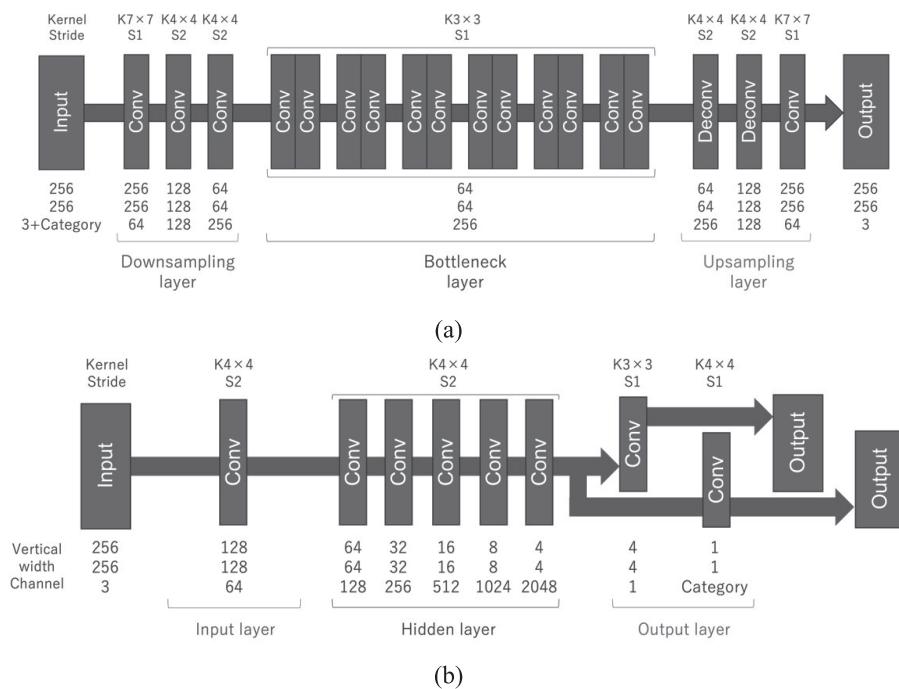


Fig. 10 Network structure. (a) Generator. (b) Discriminator

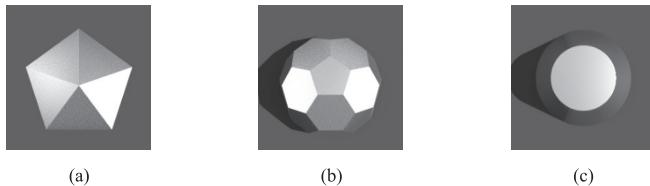


Fig. 11 Shapes for verification. (a) Pentagonal pyramid. (b) Faceted icosahedra. (c) Faceted cone.

dering environment in which the surface reflection is Lambert diffuse reflection, as in the training dataset, and with no environment map used in PBRT. Because the normal is estimated using only the part of the object that is not included in the target object in illuminance stereo, the RMSE is calculated by masking the part of the object that is not the target object in the quantitative evaluation as well.

As with the evaluation of the image transformation, the evaluation of the restored shape is carried out in two ways: a qualitative evaluation using the estimated normal map showing the direction of the normal xyz as an RGB value, as well as a 3D model of the restored shape, and a quantitative evaluation using the RMSE between the point cloud and the restored shape from the ground-truth image. The RMSE between point clouds is synonymous with the calculation of the Euclidean distance between point clouds. To derive this Euclidean distance, we use the iterative closest point (ICP) algorithm¹³⁾, which is a method for aligning point clouds. This method is used to align the point cloud of the target shape by repeatedly detecting the nearest neighbor points between the point cloud of the ground truth shape and the point cloud estimated from the transformed image, and by estimating the rigid body transformation sequence such that the error in the distance between the nearest neighbor points is minimized. The process is repeated until the error in the distance between the point clouds is less than the threshold for convergence, and the average error in the distance between the point clouds at the time of convergence is derived for each corresponding point cloud. In practice, if the points to be mapped are p , the number of point groups is N , the points to be mapped are q , and the rotation and translation of the rigid body transformation sequence are R and t , respectively, the Euclidean distance between the point groups is d_{211} , which is calculated as follows:

$$d_{err} = \frac{1}{N_p} \sum_{i=1}^{N_p} \|Rp_i + t - q_i\|^2. \quad (8)$$

5.1 Qualitative evaluation

In Fig. 12, (1) is a captured image to be input into the learning model, (2) is an image used as the ground truth, (3) is an image transformed using S2Dnet as a conventional approach, (4) is an image transformed using the learning model without uniformity preservation loss, and (5) is the images transformed using our approach. Nos. 1-6 indicate the horizontal direction.

The result of a shape reconstruction when applying photometric stereo using the transformed images is shown in Fig. 13. This result of the shape reconstruction is a view from the horizontal direction.

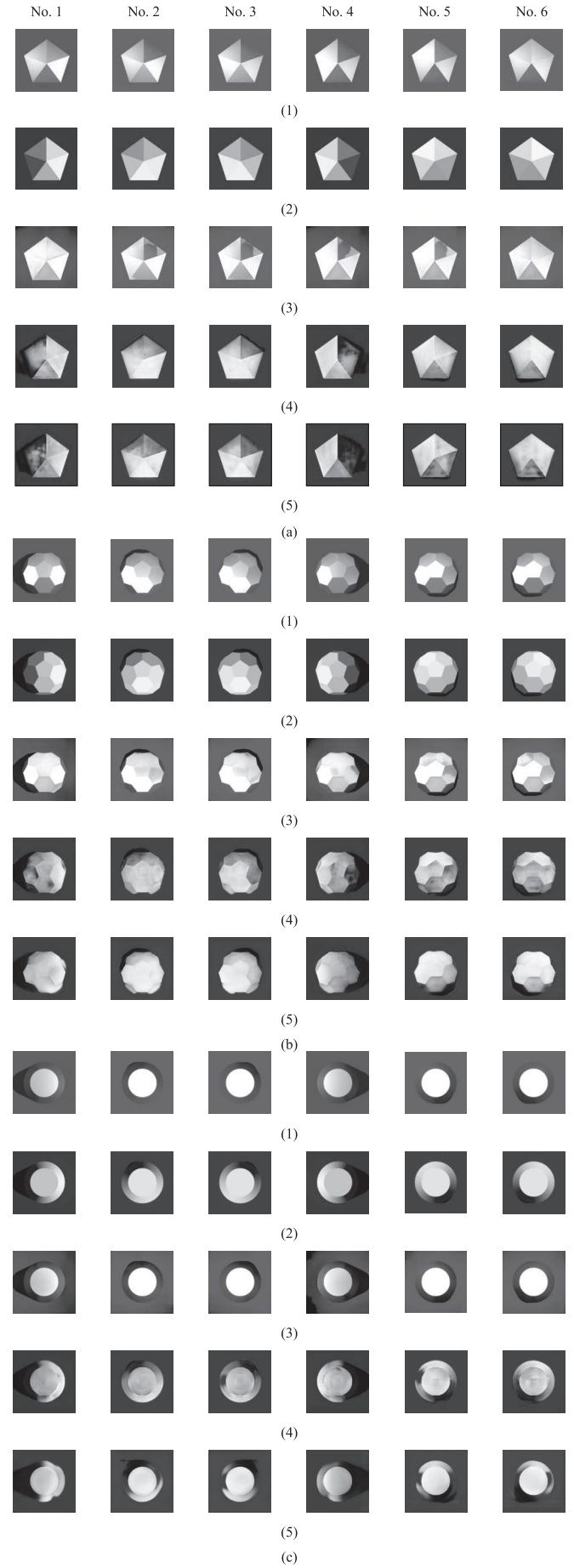


Fig. 12 Images transformed from each light position. (a) Pentagonal pyramid. (b) Faceted icosahedra. (c) Faceted cone. ((1) Input images. (2) Ground truth. (3) Wu et al.'s approach.⁴⁾ (4) Our method (with no uniformity preservation loss). (5) Our method.)

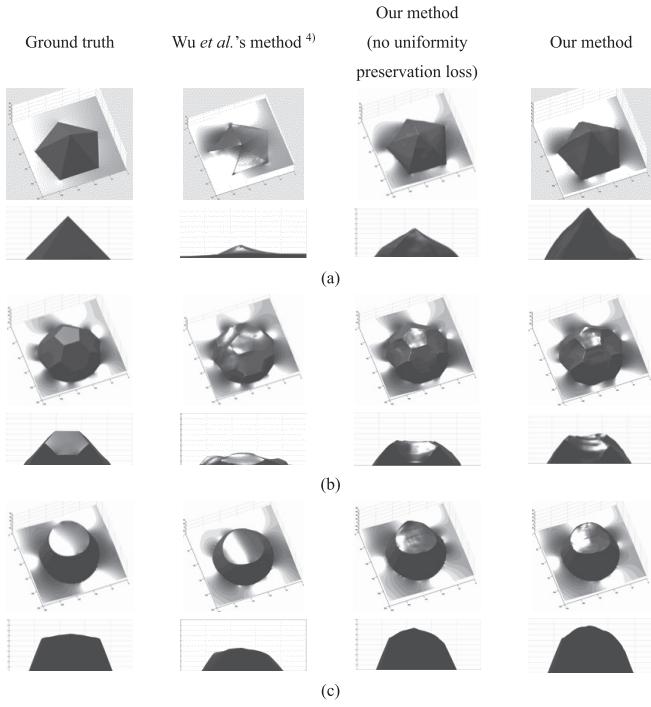


Fig. 13 Ground truth shape and reconstructed shape with Wu *et al.*'s approach⁴⁾ and our method without and with uniformity preservation loss. (The upper level is a normal view. The lower level is a horizontal view.) (a) Pentagonal pyramid. (b) Faceted icosahedra. (c) Faceted cone.

Based on the results, we confirmed that our method can transform images reflecting changes in the luminance of the surface owing to changes in the light source position better than the conventional method. Although we could not confirm significant changes in the image transformation with or without a loss of uniformity preservation, we confirmed that the reconstructed shape was closer to ground truth in terms of the height from the floor.

5.2 Quantitative evaluation

A graphical representation of the RMSE of the transformed images using the S2Dnet when applying the conventional method, the method with no uniformity preservation loss, and our proposed method, along with the ground truth image of each shape, are shown in Fig. 14. Each graph is the result of averaging six images formed into each shape, and the average for each method is shown on the right. The standard errors based on the shape applied are also shown in the same graph along with error bars.

In addition, Fig. 15 shows a graphical representation of the RMSE results when applying the conventional method, the method without uniformity preservation loss, and our proposed approach, along with the ground truth images, using photometric stereo. These results indicate the individual errors for the three validation geometries, and the mean and standard error based on the geometry are also shown on the right side.

The results show that the proposed method has a smaller RMSE error than the conventional method for both the transformed image and the reconstructed shape, and the accuracy of the proposed method has been improved. In addition, there were no significant changes in the transformed image with or without a loss of unifor-

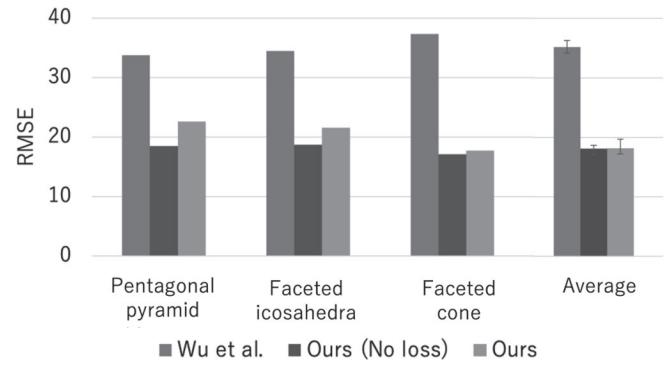


Fig. 14 RMSE between ground truth and transformed images with Wu *et al.*'s approach⁸⁾ and our method without and with uniformity preservation loss.

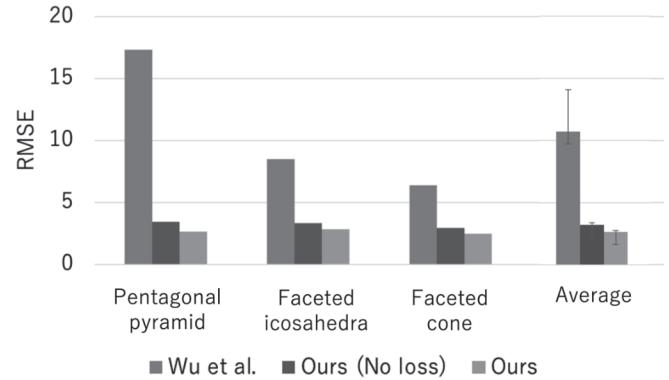


Fig. 15 RMSE between ground truth and reconstructed shapes with Wu *et al.*'s approach⁸⁾ and our method without and with uniformity preservation loss.

mity preservation, where the error in the reconstructed shape was slightly smaller with the method that introduced our proposed loss.

6. Discussion

From the verification results described in section 5, it can be stated that our method can convert an image to reflect the changes in luminance caused by the change in the direction of the light source and can restore the shape close to the ground truth in comparison to the conventional method. In particular, as one of the major differences between our method and the conventional approach, the thickness of the shape from the floor surface can be restored using our method. As the reason for this, the image used in this study, which is applied as an index for calculating the loss of uniformity, is illuminated from the same direction as the shooting direction. Although the distance between the floor surface and the light source was not changed in the images applied for this training, the training was conducted such that the images were closer to those irradiated from the same direction as the shooting direction. Because of this, the luminance was converted to be higher at positions closer to the shooting direction and lower at positions farther away. As a result, it is therefore highly possible that the luminance representation reflected the shape of the vertical direction from the floor surface. From this perspective, it can be stated that the introduction of the uniformity retention loss is effective.

By contrast, it was confirmed that the surface was distorted in comparison to the results of the ground truth, and that the incorrect conversion has not yet been sufficiently improved. This problem is thought to be largely due to the fact that the calculation of the uniformity preservation loss is from a difference in subtraction. Therefore, it will be necessary to improve this problem by reducing the number of convolutional layers, introducing the structure of a Self-Attention GAN¹⁴⁾ that can generate images based on global features and allowing such features to be learned, improving the hyperparameter of the uniformity preservation loss, and increasing the variation of the height of the light source from the floor.

7. Conclusions and future work

In this study, we proposed a new learning network that can reflect the changes in luminance of a surface owing to changes in the light source position by adding the position data of the light source as category information. The results of photometric stereo using the images transformed by our learning network showed that it is possible to reconstruct a shape more accurately than through a conventional image transformation method.

In the future, we will solve the problem of distortions appearing on the surface of the shape and improve the method to address the incorrectly transformed parts of the image. In addition, it is expected that the approach will handle 3D reconstructions with images under a wide range of conditions by increasing the number of light sources, materials, and shapes included in the dataset images. Moreover, this research was aimed at converting specular images captured by popular cameras. Because of this, we used 24-bit images in the training set. However, specular images require a very wide dynamic range, and 24-bit images may not be able to accurately represent specular

reflection. Therefore, in order to construct a conversion network that can handle specular reflection images more accurately, it will be necessary to conduct learning using HDR format images of approximately 32 bits in the future.

References

- 1) W. M. Silver, "Determining shape and reflectance using multiple images," PhD thesis (Massachusetts Institute of Technology, 1980), 1980.
- 2) R. J. Woodham, Optical Engineering, vol. 19, pp. 139-144 (1980)
- 3) J. Y. Zhu, T. Park, P. Isola, A. A. Efros, Proceedings of IEEE ICCV, pp. 2223-2232 (2017)
- 4) S. Wu, H. Huang, T. Portenier, M. Sela, D. Cohen-Or, R. Kimmel, M. Zwicker, Proceedings of ECCV, pp. 183-200 (2018)
- 5) F. Langguth, K. Sunkavalli, S. Hadap, M. Goesele, ECCV, pp. 469-485 (2016)
- 6) R. Maier, K. Kim, D. Cremers, J. Kautz, M. Nießner, Proceedings of IEEE ICCV, pp. 3114-3122 (2017)
- 7) O. Özyeşil, V. Voroninski, R. Basri, A. Singer, Acta Numerica, vol. 26, pp. 305-364 (2017)
- 8) R. Zhang, P. S. Tsai, J. E. Cryer, M. Shah, IEEE TPAMI, vol. 21, pp. 690-706, 1999
- 9) Y. Choi, M. Choi, M. Kim, J. W. Ha, S. Kim, J. Choo, Proceedings of IEEE CVPR, pp. 8789- 8797 (2018)
- 10) A. Odena, C. Olah, J. Shlens, PMLR, vol. 70, pp. 2642-2651 (2017)
- 11) M. Pharr, W. Jakob, G. Humphreys, "Physically based rendering: From theory to implementation," 3rd ed., by J. Pierce, Morgan Kaufmann, Massachusetts, 2016
- 12) K. He, X. Zhang, S. Ren, J. Sun, Proceedings of IEEE CVPR, pp. 770-778 (2016) 13) P. J. Besl, N. D. McKay, SPIE, vol. 1611, pp. 586-606 (1992)
- 14) H. Zhang, I. Goodfellow, D. Metaxas, A. Odena, PMLR, vol. 97, pp. 7354-7363 (2019)