Electronics and Communications in Japan, Part 2, Vol. 90, No. 12, 2007
Translated from Denshi Joho Tsushin Gakkai Ronbunshi, Vol. J89-D, No. 2, February 2006, pp. 381–392

Color Photometric Stereo and Virtual Image Rendering Using Neural Networks

Haruki Kawanaka,[1] Yuji Iwahori,[1] Robert J. Woodham,[2] and Kenji Funahashi[3]

[1]Department of Computer Science, Chubu University, Kasugai, 487-8501 Japan

[2]Department of Computer Science, University of British Columbia, Vancouver, B.C. Canada V6T 1Z4

[3]Graduate School of Engineering, Nagoya Institute of Technology, Nagoya, 466-8555 Japan

SUMMARY

In this paper we extend the application of neural network-based photometric stereo founded on the principle of empirical photometric stereo to color images proposing a method for computing both the normal vectors of a target object and its color reflectance coefficients. This method is able to render objects that have non-Lambert reflectance properties without using any parametric reflectance function as a reflectance model. In addition, we propose a novel neural network-based rendering method that allows the generation of realistic virtual images of an object with arbitrary light source direction and from arbitrary viewpoints based on the physical reflectance properties of the actual object and perform a comparative evaluation with approximations by existing models, the Phong model, and the Torrance–Sparrow model. © 2007 Wiley Periodicals, Inc. Electron Comm Jpn Pt 2, 90(12): 47–60, 2007; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/ecjb.20423

Key words: neural networks; photometric stereo; mixed reality; model-based rendering.

1. Introduction

In the fields of computer vision and mixed reality, research was originally focused primarily on monochrome grayscale images; in recent years, however, color images have gradually become the focus of such work [1–4].

With a fixed single viewpoint, Woodham [5] proposed a method known as *photometric stereo* for estimating local surface normal vectors for the surface of an object from multiple grayscale images taken under different light source conditions; in addition, he proposed [6] empirical photometric stereo, a method that does not assume a function for the reflectance properties of the surface of an object and does not require information of the light source directions. These methods also fix the viewpoint while the images are taken under different directions of the light source.

Iwahori and colleagues have previously developed a neural network-based photometric stereo method [7, 8]. This method is based on empirical photometric stereo [6] by training a neural network on a sphere with the same surface reflectance properties as the test object; it is possible to obtain the normal vectors for the surface of the object and monochrome reflectance coefficients without assum-

Contrat grant sponsors: Scientific Research Grant [Basic Research (C) (2) 16500108], the Kayamori Foundation of Informational Science Advancement (Yuji Iwahori), the Hori Information Science Promotion Foundation (Kenji Funahashi), and the Institute for Robotics and Intelligent Systems and the Natural Sciences and Engineering Research Council (Robert J. Woodham).

ing any parametric function for the reflectance properties of the surface of the object nor explicitly using a value for the lighting directions. However, since the method given in Ref. 8 makes use of an assumption that the reflectance factor is a constant parameter proportional to the reflectance property function, there is a problem remaining in that areas of the sharp specular reflections give some errors in the surface normal obtained due to the effects of this specular reflectance.

Against this background, in this paper we first extend the neural network-based photometric stereo to color images and then such that it can be applied to objects that give rise to sharp specular reflections. Both Refs. 9 and 10 have proposed photometric stereo for color images; however, there are problems and limitations such that the former uses a particular color light source while the latter assumes completely diffuse reflectance (Lambert reflectance). Our proposed method uses a standard white light source and is applicable to objects with non-Lambert reflectance properties. In addition to extending photometric stereo to color images, the method also extends to objects on which sharp specular reflections arise.

Model rendering, on the other hand, consists of the creation of virtual images of an object given a mathematical formulation of the reflectance properties on the object's surface as a parametric reflectance function (such as the Phong model) and the geometric shape of the object. In Ref. 11 a range finder was used to reconstruct the three-dimensional shape of an object and by assuming a Torrance–Sparrow model as the surface reflectance function of the object, a method was proposed to estimate its parameters from a color image of the object. In this paper we propose a neural network-based rendering method that generates the appearance of an object using the viewing direction vector, the light source direction vector, the object's surface normal vector, and color reflectance coefficients as an inverse system of neural network-based photometric stereo. This method makes no use of any parametric reflectance function and can generate realistic virtual images of an object from arbitrary viewpoints and based on arbitrary lighting directions.

2. Principles of Dichromatic Reflectance Model and Neural Network-Based Photometric Stereo

2.1. The dichromatic reflectance model

With substances such as plastic, incident rays (the illuminating rays) can be reflected in one of two ways. First they can result in a diffuse reflectance, the illuminating rays enter the inside of the object from the object's surface and undergo a modulation according to the inherent reflectivity rate of the substance and then the reflected rays are projected uniformly in all directions. The color of these reflected rays is treated as the standard color of the object. Another type of reflection is specular reflectance; due to relations of the inflection rate between the media, the rays do not enter the inside of the object and are reflected on the surface. Since these incident rays do not enter the object, they do not undergo any absorption and the reflecting rays are projected as a direct specular reflection along the surface normal vector. Since specular reflections do not undergo the influence of the reflecting substance, they are the same color as the light source. The observed light is a linear sum of these two types of ray and can be expressed as follows [12]:

$$E = d\rho E_w + mE_w \tag{1}$$

Here E is the grayscale value observed, E_w is the intensity of the light source, and ρ is the reflectivity rate (the reflection coefficient). The first term represents the diffuse reflection component while the second term represents the specular reflection component. d and m are the mixing proportions; these are used to express the differences in color by location on the image. d and m are the coefficients of the surface normal vector $\mathbf{n} = (n_x, n_y, n_z)$, the light source direction vector $\mathbf{l} = (l_x, l_y, l_z)$ and the viewing direction vector $\mathbf{v} = (v_x, v_y, v_z)$. In other words we may write

$$E = R_d(\mathbf{n}, \mathbf{l})\rho + R_m(\mathbf{n}, \mathbf{l}, \mathbf{v}) = R(\mathbf{n}, \mathbf{l}, \mathbf{v}, \rho) \tag{2}$$

Here R_d and R_m represent the diffuse reflection component and the specular reflection component respectively of the reflectance properties.

2.2. Neural network-based photometric stereo

This method proposed in Ref. 8 involves fixing the location of the camera and the object and estimating the gradients and the reflectance coefficients of the object's surface from four monochrome grayscale images taken under four different light source directions. However, here the normal vectors and reflectance coefficients are estimated without computing the precise directions of the light sources.

From Eq. (2), when the light source direction and viewing direction are fixed, variation in brightness values will be due to differences in the direction of the surface normal vector. In other words, when the light source direction and viewing direction are fixed, then Eq. (2) becomes

$$E = R(\mathbf{n}, \rho) \tag{3}$$

and when we have four observations each taken with a different light source direction, then the following set of photometric image equations can be established:

$$\begin{bmatrix} E_1 \\ E_2 \\ E_3 \\ E_4 \end{bmatrix} = \begin{bmatrix} R_1(\mathbf{n}, \rho) \\ R_2(\mathbf{n}, \rho) \\ R_3(\mathbf{n}, \rho) \\ R_4(\mathbf{n}, \rho) \end{bmatrix} \tag{4}$$

The neural network-based photometric stereo method constructs a radial basis function neural network (RBF-NN) that takes $\{E_1, E_2, E_3, E_4\}$ as input and produces $\{\mathbf{n}, \rho\}$ as output [13]. Here in using monochrome grayscale images, ρ will be a scalar value from 0 to 1. RBF-NN is able to perform nonlinear nonparametric approximation. Therefore, we are able to estimate the surface normal vector $\mathbf{n}$ and the reflectance coefficient ρ from the set of four brightness values without the need to assume a particular parametric form of the reflectance function. The observation environment is shown in Fig. 1 and the structure of this RBF-NN is shown in Fig. 2.

The neural network is trained by using images of a sphere with the same reflectance properties as the object that we intend to perform reconstruction for. The neural network is able to sufficiently learn the relationship between the surface normal vectors and the brightness levels by using a sphere for training since in addition to the fact that the surface normal vectors of the sphere can be computed easily from its radius and the image coordinates, the sphere also possesses all different kinds of surface normal vectors. For training data we select a single value for the reflectance coefficient ρ from the set of values 1.0, 0.8, 0.6, and 0.4 and then combine this value of ρ with four images consisting of a set of brightness levels for identical pixels $\{\rho E_1, \rho E_2, \rho E_3, \rho E_4\}$ and take this as input and then set the surface normal vector and reflectance coefficient $\{n_x, n_y, n_z, \rho\}$ as output and construct training data from various different points in the image of the sphere. NN then learns the relationship between the set of brightness levels and the corresponding surface normal vector for various different settings of the reflectance coefficient. We construct the training set for the neural network by sampling uniformly from the space on the hemisphere with various surface normal vectors as the outputs. These training data are used to train the neural network.

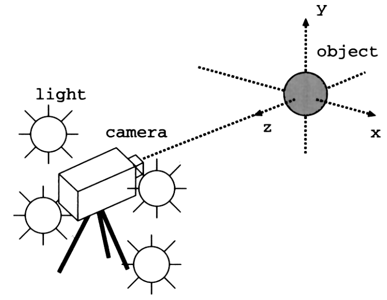

Fig. 1. Observation setup system 1.

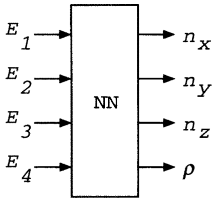

Fig. 2. RBF-NN for monochrome photometric stereo.

When estimating the surface normal vectors of the object and the reflectance coefficients, we give the set of four grayscale values for the same pixel $\{E_1, E_2, E_3, E_4\}$ from four images of the object that were taken with the same light source directions as the sphere used in the trained neural network. NN outputs the local surface normal vector and the reflectance coefficient. By performing this procedure for all pixels of the object, we are able to obtain the surface normal vectors and reflectance coefficients for the object in the image.

3. Neural Network-Based Color Photometric Stereo

3.1. Applying neural network-based photometric stereo to color images

When photometric stereo is applied to the surface with completely diffuse reflection (Lambert reflectance), the reflectance coefficients and surface normal vectors can be determined analytically from three grayscale images [6]. When the reflectance coefficients of the target object are non-Lambert with nonuniform reflectance in general, these can be estimated using four or more images from the viewpoint of unknown parameters and nonlinear simultaneous equations. In Ref. 8, four images are used, based on this reasoning. When applying photometric stereo to color images in this work we also use four different images taken with four different directions for the light source in order to estimate the reflectance coefficients and the surface normal vectors as given in Ref. 8.

Equation (1) showing the reflection model for the bicolor case is rewritten as the following equation when applied to color images:

$$\mathbf{E} = d\boldsymbol{\rho}\mathbf{E_w} + m\ \mathbf{E_w} \tag{5}$$

Here $\mathbf{E} = \{E_R, E_G, E_B\}$ are the RGB values for an image, $\mathbf{E_w} = \{E_{wR}, E_{wG}, E_{wB}\}$ are the intensities of the light source, and $\boldsymbol{\rho} = \{\rho_R, \rho_G, \rho_B\}$ are the color diffuser reflectance coefficients. As in the monochrome case, when the light source directions and the viewpoint directions are fixed, we can write

$$\mathbf{E} = R(\mathbf{n}, \boldsymbol{\rho}) = \{R_d(\mathbf{n})\boldsymbol{\rho} + R_m(\mathbf{n})\}\,\mathbf{C_L} \qquad (6)$$

Here $\mathbf{C}_\mathrm{L}$ represents the RGB value for the light source color. We can extend photometric stereo to color images on the basis of Eq. (6).

Constructing a neural network that can compute the surface normal vectors and color diffuse reflectance coefficients simultaneously from the set of brightness values as in the monochrome neural network-based photometric stereo, would require a large amount of training data and is unrealistic. If the training data consisted of images represented in 24 bits (full color), then there are already around 167 million possible color diffuse reflectance coefficient combinations and the scope of training is 65,536 times greater than that of the monochrome case. When we consider how these would combine with the surface normal vectors, the training scope becomes extremely large and it is clearly impractical to extend the monochrome method directly to the case of color images in terms of both the amount of training data required and the amount of time needed for training. Therefore, in order to perform training efficiently we introduce the two-step neural network and propose a method whereby the color diffuse reflectance coefficients and the surface normal vectors are computed using separate neural networks.

3.2. A neural network to estimate color reflectance coefficients

As shown in Fig. 3 we use an RBF-NN that is constructed such that the inputs are given as $\{E_{1R}, E_{1G}, E_{1B}, E_{2R}, E_{2G}, E_{2B}, E_{3R}, E_{3G}, E_{3B}, E_{4R}, E_{4G}, E_{4B}\}$ and the outputs obtained are $\{\rho_R, \rho_G, \rho_B\}$ in order to compute the color diffuse reflectance coefficients. Both the input and output values are color RGB values.

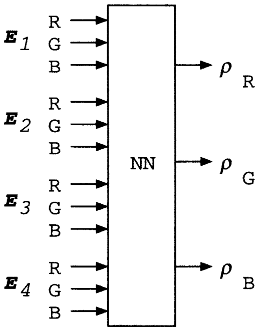

Fig. 3. RBF-NN for reflectance coefficient estimation.

The training data are constructed using images of a sphere as in the original method. We need to create these training data to be a uniform sampling from the whole space of values that can be obtained, but it is physically impossible to actually give diffuse reflectance coefficients with different RGB values to the sphere. In addition, it is not desired to require the taking of images of spheres of various different colors. Therefore, the sphere used for the training is white ($\boldsymbol{\rho}$ = {1, 1, 1}) but we set the color reflectance coefficients $\boldsymbol{\rho}$ as follows to be $\boldsymbol{\rho}'$ whereby each pixel is randomly assigned a coloration:

$$\boldsymbol{\rho}' = \left\{\rho'_R, \rho'_G, \rho'_B\right\} = \{rand_1, rand_2, rand_3\} \quad (7)$$

Here *rand* is a uniform random variable (a real number) between 0 and 1. Using Eqs. (5), (6), and (7), we also transform the set of brightness values into the following $\mathbf{E}'$ that reflects the effects of randomly generated color reflectance coefficients:

$$\mathbf{E}' = \left\{R_d(\mathbf{n})\boldsymbol{\rho}' + R_m(\mathbf{n})\right\}\mathbf{C_L} \qquad (8)$$

When applying this coloration, it is necessary first to partition each of the brightness level values of the images of the sphere used for training into a diffuse reflectance component R_d and a specular reflectance component R_m. In the method presented here we partition the reflectance components of the images of spheres in the following manner. We make use of four color images taken under four different light source directions and also four more color images of the same sphere from a different viewpoint (a total of eight images). Even if the light source directions are not given explicitly as long as they are fixed, then if the viewpoint is changed significantly the locations in which specular reflections occur will also change significantly; in addition, the diffuse reflectance component that does not depend on viewing direction will remain constant. Based on these conditions and prior knowledge of shape consisting of the sphere, we are able to determine the diffuse reflectance component by observing corresponding points in the two images between the viewpoint change, and take the darker pixel value for each light source. Depending on the surface material, specular reflectance components may be observed over a fairly wide range but in order to perform the decomposition of components into diffuse reflectance and specular reflectance as described in this paper, the specular component must be somewhat sharp. This is a condition on the objects to which the method can be applied. Figure 17 shows an example of the appearance of the decomposition of reflectance components for the images of spheres used for training with a randomly selected RGB reflectance coefficient added based on $\boldsymbol{\rho}'$.

To train the neural network we make use of images of the sphere that have random reflectance coefficients. The training data consist of inputs $\{E'_{1R}, E'_{1G}, E'_{1B}, E'_{2R}, E'_{2G}, E'_{2B}, E'_{3R}, E'_{3G}, E'_{3B}, E'_{4R}, E'_{4G}, E'_{4B}\}$ and outputs corresponding to the color reflectance coefficient for each of these pixels $\{\rho'_R, \rho'_G, \rho'_B\}$; the neural network is then trained on the sphere with various different points.

When estimating the color reflectance coefficients for an object, we input the set of RGB values for each pixel in four images of the target object taken with the same four light source directions as used for the sphere $\{E_{1R}, E_{1G}, E_{1B}, E_{2R}, E_{2G}, E_{2B}, E_{3R}, E_{3G}, E_{3B}, E_{4R}, E_{4G}, E_{4B}\}$ to the trained neural network. Color reflectance coefficient will be obtained for each pixel as output. By applying this procedure to all pixels on the object, we can obtain all the color reflectance coefficients.

3.3. A neural network to compute surface normal vectors

We now describe the neural network for computing the surface normal vectors. When estimating the surface normal vectors, the color image is converted into a monochrome image using the following equation. This equation is generally used when converting color images into monochrome (grayscale) images:

$$E{=}0.299{\times}E_R{+}0.587{\times}E_G{+}0.114{\times}E_B \qquad (9)$$

However, the grayscale transformation resulting from this equation makes the dynamic range of the blue and the red small. With neural network-based photometric stereo, the larger the dynamic range is, the easier it is to learn the grayscale variations. Therefore, for each light source direction, we assume that the grayscale value is proportional to the maximum of the RGB values and transform the color image to a grayscale image using the following equation:

$$E = max(E_R, E_G, E_B) \qquad (10)$$

We then train the neural network taking the maximum value among the RGB components as the monochrome value for the corresponding pixel.

We could consider a method that uses the original RGB values for input to the neural network; however, by using the color RGB values instead of the grayscale values, a larger amount of training data will be needed and consequently it may result in a lower precision.

Figure 4 shows the structure of a neural network that uses the grayscale values to estimate surface normal vectors. The neural network has a structure that allows the set of grayscale values $\{E_1, E_2, E_3, E_4\}$ to be given as input and the surface normal vector $\{n_x, n_y, n_z\}$ to be obtained as output. In this way, since we convert to monochrome images as were used in the original method, we are able to have the neural network estimate surface normal vectors for color images without any hindrance.

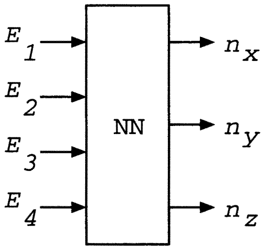

Fig. 4. RBF-NN for estimating vector surface normal vector.

This neural network learns the mapping between the set of grayscale values and the surface normal vectors; there we add the random monochrome reflectance coefficients with 256 levels (8 bits) to a white sphere used for training. Training is performed by creating training data that consist of sets of grayscale values of the image of the sphere $\{E'_1, E'_2, E'_3, E'_4\}$ to which are randomly added monochrome reflectance coefficients corresponding to the diffuse reflectance components as inputs and surface normal vectors $\{n_x, n_y, n_z\}$ that correspond to this pixel as outputs.

When estimating the surface normal vectors for an object we transform four color images of the object in monochrome images and input the set of brightness levels for each pixel $\{E_1, E_2, E_3, E_4\}$ into the trained neural network. This results in the estimate of the surface normal vector corresponding to this pixel being output. By performing the procedure for all pixels on the object, we can obtain the distribution of surface normal vectors for the object.

3.4. Positioning of this method

Our method implements empirical photometric stereo that does not make explicit use of information regarding the direction of the light source using a neural network. In order to create the training data for the neural network, we partition the image of a sphere into the diffuse reflectance component and the specular reflectance component and then randomly add a reflectance coefficient to the diffuse reflectance component fusing these together; here, we decompose the reflectance components using prior knowledge of shape of the sphere and features of the diffuse reflectance. This method makes use of a general assumption, namely, the dichromatic reflection assumption, but specifically does not assume a parametric characteristic reflectance function to describe the specular reflectance. In general decomposing the reflectance components of an object of unknown shape is difficult using only the assumption of dichromatic reflection when the light source direction is not given explicitly, and the diffuse reflectance coefficients and the surface normal vectors cannot be calculated analytically.

An empirical approach to solve this problem that is not analytically solvable under the condition that the light source direction is not given explicitly and no parametric function is assumed for the reflectance properties has been given in Ref. 6 whereby the relationship between the actual observed intensities and surface normal vectors is mapped using a look-up table. Neural network-based photometric stereo (Refs. 7, 8, and this paper) extends the method of Ref. 6 by training a neural network to allow generalization.

4. Neural Network-Based Rendering

4.1. A neural network used in rendering

Here as an application to mixed reality, we describe a method for generating virtual images of an object from an arbitrary viewpoint and with an arbitrary light source direction using obtained surface normal vectors and the color reflectance coefficients.

The neural network-based photometric stereo described above is a method for estimating the diffuse reflectance coefficients and the surface normal vectors of an object; however, the neural network can also be used for the appearance of the object with realistic images generated. We refer to this as *neural network-based rendering*.

Neural network-based rendering is a method for generating realistic virtual images of an object from an arbitrary viewpoint and with an arbitrary light source direction without assuming any parametric function characterizing the reflections as in the Phong model or the Torrance–Sparrow model.

From Eq. (2) we see that the brightness of a given pixel is determined by the angles formed by the light source direction vector **l**, the surface normal vector **n**, and the viewpoint direction vector **v**. The incident angle i, the emittance angle e, and the phase angle g are each functions of **n**, **l**, and **v** given respectively by the following equations:

$$i = \cos^{-1}(\mathbf{n} \cdot \mathbf{l}) \tag{11}$$

$$e = \cos^{-1}(\mathbf{n} \cdot \mathbf{v}) \tag{12}$$

$$g = \cos^{-1}(\mathbf{v} \cdot \mathbf{l}) \tag{13}$$

The reflectance property R can be expressed as a function of i, e, and g shown in Fig. 5 and the diffuse reflectance coefficients $\boldsymbol{\rho}$:

$$\mathbf{E} = R(\mathbf{l}, \mathbf{n}, \mathbf{v}, \boldsymbol{\rho}) = R(i, e, g, \boldsymbol{\rho}) \tag{14}$$

Here since the objects have a dichromatic reflection, we estimate **E** using the following equation decomposing the reflection into the specular component R_m and the diffuse component R_d and calculating each separately:

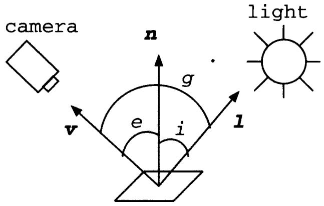

Fig. 5. Incident angle i, emittance angle e, and phase angle g.

$$\mathbf{E} = R_d(i)\boldsymbol{\rho} + R_m(i, e, g) \tag{15}$$

Since R_d can be computed analytically as the inner product $\mathbf{n} \cdot \mathbf{l} = \cos i$ of the surface normal vector **n** obtained via photometric stereo and the explicitly specified light source direction vector **l**, the neural network is used simply to estimate the term R_m. We construct a neural network that takes i, e, and g that depend on **l**, **n**, and **v** as input and give as the output R_m consisting of the RGB value for each point of the object. This neural network is shown in Fig. 6.

4.2. Creating training data and neural network-based generalization

Since only the specular reflectance component is estimated using the neural network, there is no need to train the network for the diffuse reflectance component. Consequently, the neural network training procedure differs from that used for photometric stereo and is performed using only white sphere images. The training data are created by sampling various different combinations of i, e, and g. Here the ranges that these angles can take are as follows:

$$0 \le i \le 90 \quad 0 \le e \le 90 \quad 0 \le g \le 180 \tag{16}$$

In order to create sufficient training data, we use images of the sphere from seven different directions as shown in Fig. 7. The light source is on the xz plane and the angle between the straight lines from the origin to the light source for adjacent locations is every 15°. In addition, as in the case

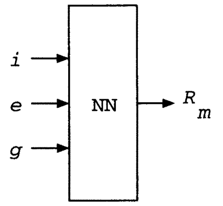

Fig. 6. RBF-NN used for neural network-based rendering.

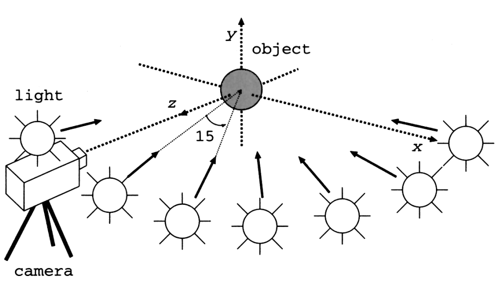

Fig. 7. Observation setup system 2.

of photometric stereo, we decompose the image of the sphere into a specular reflectance component and a diffuse reflectance component.

The training data consist of sets of inputs $\{i, e, g\}$ and the corresponding specular reflectance component of the pixel of the sphere $\{R_m\}$ as output sampled for various different points of the image of the sphere. It is clearly straightforward to have i and e take a variety of different values and varying the location of the light source as shown in Fig. 7 enables us to create a variety of different combinations of i and e with seven different types of g thereby allowing sufficient combinations for training data. In general, g can take values in the range shown in Eq. (16); however, in this paper we created training data with g in the range $0 \leq g \leq 90$. Although it is not a substantial problem, if it is necessary to generate images with a value of g greater than 90°, then it would suffice to increase the number of distinct light source locations to 12. We use these training data to train the neural network.

When generating a virtual image of an object using the trained neural network, we use Eq. (13) to compute $\{i, e, g\}$ from the light source direction vector $\mathbf{l}$, the surface normal vector $\mathbf{n}$, and the viewing direction vector $\mathbf{v}$ for each pixel in the image and then give these as input to the neural network. This results in the specular reflectance component for the pixel R_m as output. By using this R_m, the corresponding R_d and the color reflectance coefficients $\{\rho_R, \rho_G, \rho_B\}$ obtained from the neural network in Fig. 4, we can obtain the RGB value for the pixel via Eq. (15). By repeating this procedure for each pixel in the object, we are able to create a virtual image of the object.

4.3. Generating virtual images from different viewpoints

To generate images in which the viewpoint has been changed (these are equivalent to images in which the object and the light source have undergone a rotation), it is not sufficient simply to have the neural network generalize the reflectance function; rather, in order to generate images in which the object is rotated, it is necessary to obtain surface normal vectors and color reflectance coefficients for each pixel of the object in the pose to be generated.

Therefore, we first compute the object's height distribution by integrating the surface normal vectors obtained from the photometric stereo. We then create another height distribution in a different pose by rotating the height distribution over the x, y, and z axes. In addition, we compute the surface normal vectors and color diffuse reflectance coefficients by determining correspondences between object point during rotation. By giving the intended light source direction and the values of i, e, and g computed from the surface normal vectors after the rotation as input to the neural network, we obtain the specular reflection component for each corresponding pixel and then the RGB values for each pixel from Eq. (15).

5. Experiments and Discussion

In this work we also performed an evaluation of the precision and merits of neural network-based color photometric stereo based on the approach of Ref. 15. The following experiments were performed on an AthlonXp2500+ machine with main memory of 512 MB; we used MATLAB to train the neural network.

5.1. Simulation experiments with neural network-based photometric stereo

5.1.1. Color diffuse reflectance coefficients

Figures 18(a) to 18(d) show four input images (the target object for reconstruction is a sphere). In order to perform a quantitative evaluation of the precision of the proposed method, these images of spheres were created using the Torrance–Sparrow model. Without assuming the reflectance function or the light source direction information explicitly, neural network-based photometric stereo is used for evaluation. We set the number of training examples for the neural network to approximately 2000, the number of training iterations was 400, and the spread constant was 0.4. Training took approximately 15 minutes while diffuse reflectance coefficients took approximately 10 seconds to be estimated.

Figure 18(e) shows the color of the object displaying the color diffuse reflectance coefficient estimated by neural network-based photometric stereo as RGB values. We compute the error between the RGB values of the estimated reflectance coefficients (a real-valued vector normalized over the range 0 to 1) and the theoretical value as

$$error = \frac{\Sigma\|\mathbf{E}_{RGB} - \mathbf{T}_{RGB}\|}{N} \tag{17}$$

Here $\mathbf{E}_{RGB}$ is the vector of the RGB values of the estimated reflectance coefficients, $\mathbf{T}_{RGB}$ is the vector of theoretical

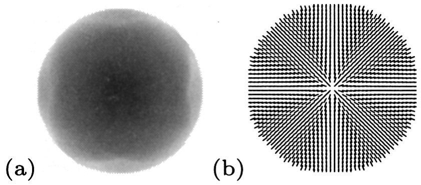

Fig. 8. Results of the proposed method. (a) Slope; (b) aspect.

values, and N is the number of pixels in the target object. The average RGB error with 256 levels is between 10 and 15; this corresponds to when the error values were aiming for between 0.0391 and 0.0586, and the actual error was 0.0373. This result confirms that reconstruction for color diffuse reflectance coefficients was performed with a high level of precision.

5.1.2. Surface normal vectors

Unlike the previous method given in Ref. 8, the proposed method adopts a dichromatic reflection model that allows it to deal with sharp specular reflections. Therefore, we compare the estimation accuracy of the proposed method and previous method over a region of sharp specular reflections.

Both methods make use of a neural network with the structure shown in Fig. 4; in these experiments as with the previous method we did not make use of a dichromatic reflection model and also used Eq. (9) to transform images into monochrome. The previous method does not assume a dichromatic color model. In addition, we performed training for both methods using a set of 2000 examples, the number of training iterations was 400, and the spread constant was 0.4. The training time for both methods was approximately 10 minutes; the estimation of surface normal vectors took approximately 10 seconds for both methods.

The estimation results using the proposed method are shown in Fig. 8. Panel (a) shows larger angles between the surface normal vector and the viewpoint direction vector with a brighter coloring. Panel (b) depicts the directional gradient of the object's aspect. The estimation results using the previous method are shown in Fig. 9.

We aimed for a higher precision value than that of the previous method [8]. While the average angle of error for the previous method was 2.310°, the proposed method could estimate the angle with an error of 0.371°. The use of the dichromatic reflection model resulted in an improvement in the precision for the object where sharp specular reflections occurred.

Errors were large for the previous method on the region where sharp specular reflections occurred. The reflectance property model for the previous method is

$$R(\mathbf{n}, \boldsymbol{\rho}) = \boldsymbol{\rho}\,\{R_d + R_m\} \tag{18}$$

and since this does not comply with the dichromatic reflection model, in regions where a sharp specular reflection occurs, errors appear due to the influence of specular reflections on the corresponding surface normal vectors. In contrast, the proposed method is able to perform better estimation in these regions as well without a large error arising. With the proposed method the error angle for estimated vectors in regions with specular reflections was also around 0.4° giving a higher accuracy than the previous method.

We next conducted estimation experiments using color images themselves. The estimation results are shown in Fig. 10. The overall precision gives an average error angle of 19.630°; this is significantly worse than the proposed method. Smooth estimation errors were obtained even in the vicinity of boundaries between diffuse and specular reflections but overall an error that drew the surface normal vectors toward the viewing direction occurred. When color RGB values are input to the neural network, a significantly larger amount of training data is required compared to when monochrome grayscale values are used; as a result, the estimation precision becomes low. This confirms that by converting these values to monochrome grayscale values and then estimating the surface normal vectors, estimation is performed efficiently in terms of both training time and precision.

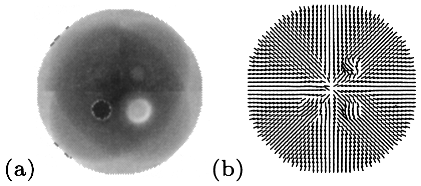

Fig. 9. Results of the previous method [8]. (a) Slope; (b) aspect.

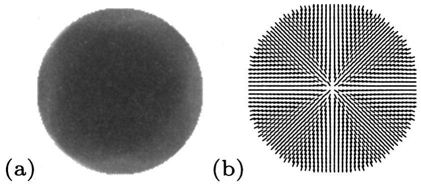

Fig. 10. Results of the proposed method using color RGB values as input. (a) Slope; (b) aspect.

5.1.3. Relationship between reflection component decomposition accuracy and estimation precision

In order to create training data for the proposed method, we decompose the reflection components in training images of a sphere. We performed experiments to evaluate what effect the accuracy with which this decomposition was performed had on the precision with which the surface normal vectors are estimated.

The proposed method treats objects on which sharp specular reflections occur; however, we now evaluated the error when we had erroneously decomposed the reflection component setting it up to 50 percent larger than its actual value and as low as 50 percent less than this. The results are shown in Fig. 11. The horizontal axis shows the proportion of error in the decomposition of the specular reflection component while the vertical axis shows the average and the maximum error angle given this error in the decomposition. From these results we can see that while an error of plus or minus 10 percent away from the true specular reflection component does not produce a significant change in the estimation results, a larger error in this decomposition of the components will result in a large error in the estimation errors around boundary regions of specular reflection.

5.2. Experiments using neural network-based color photometric stereo with real objects

Figure 17(a) shows one of the images of a sphere used for training; this image was decomposed in its diffuse reflection components [Fig. 17(b)] and its specular reflection components [Fig. 17(c)] and color reflectance coefficients were added to it at random [Fig. 17(d)] and then the neural network training was performed. In the training of both the neural network for computing color reflectance coefficients and the neural network for computing surface normal vectors we used approximately 1000 training data sets, 400 training iterations, and a spread constant of 0.4. The training took approximately 15 minutes for each of these networks and the estimation of both the surface normal vectors and the color reflectance coefficients took approximately 10 seconds in each case.

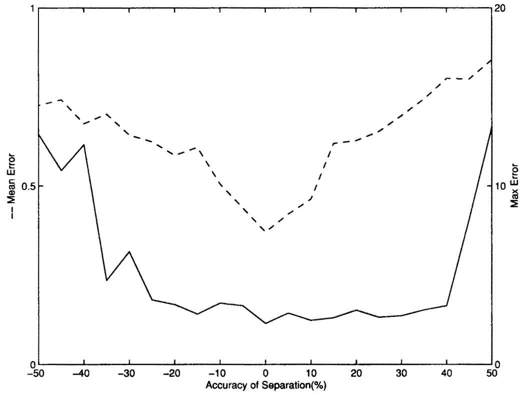

Fig. 11. Relation between the precision of decomposing of components and the estimated results.

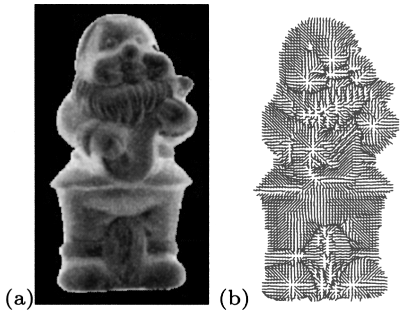

Fig. 12. (a) Slope and (b) aspect estimated by the proposed method.

Figures 19(a) to 19(d) show four of the images of the object used as input. The object used in these experiments was a glazed ceramic container. The surface of the sphere used in training was coated with the same type of paint in order to create the same reflectance properties. These experiments were performed with the assumption that interreflections and cast shadows did not occur.

Figures 12 and 19(e) show the results of the neural network-based photometric stereo. Color diffuse reflectance coefficients and surface normal vectors are estimated also for a real object.

However, in the clothing or wrinkles seen in human forms there are regions that are locally concave and therefore the effects of interreflection may arise; as a result there may be some slight errors in the neural network-based photometric stereo. In addition, parts that have shadows cast on them by the nose or the cheekbones are also theoretically in a similar manner. This is because no secondary reflections or cast shadows will arise on the sphere used in training. The problem of cast shadows is another issue and is not dealt with in this paper; however, it would be possible to obtain some improvements by also applying a method such as the illumination planning method proposed by Fukui and colleagues [14].

5.3. Evaluation of the accuracy of neural network-based rendering using spheres

Next we evaluated the accuracy of our proposed method for neural network-based rendering. Since it is not possible to perform a quantitative evaluation with a real

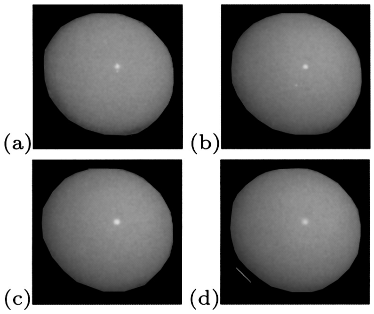

Fig. 13. (a) Real image of sphere. (b) Image estimated by proposed method, (c) estimated by Torrance–Sparrow model, and (d) estimated by Phong model.

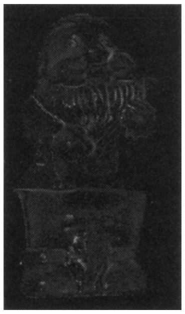

Fig. 15. Residual between real and virtual image with an arbitrary light source direction.

object of unknown shape, we used a known shape, a sphere, to perform these experiments and compared the rendering results with the Torrance–Sparrow model and the Phong model. Figure 13(a) shows real images taken of the sphere used here. The light source direction **l** here is **l** = (0.36, 0.31, 0.88); this is not one of the directions used during the training of the neural network. During the training of the neural network we use images of the sphere taken with seven different light source directions. The training data consist of approximately 1000 examples, the number of training iterations was 200, and the spread constant was set to 0.4. The training took approximately 10 minutes. Given the results obtained from the neural network-based color photometric stereo, we estimated the parameters for the Torrance–Sparrow model and the Phong model by using the least squared method on the seven images of the sphere used for training the neural network. It was assumed that these parameters were constant over the whole sphere. The geometric decay coefficient was estimated as $G = 1$, the Fresnel reflection coefficient as $F = 1.06$, and σ representing the roughness of the object's surface as 0.0272. The parameter representing sharpness of the specular reflections in the Phong model was $n = 289$.

The results of rendering with the proposed model, the Torrance–Sparrow model, and the Phong model are shown in Figs. 13(b) to 13(d). Figure 14 shows the results of displaying the brightness distribution for the specular reflection regions of each of the images in terms of height. The mean error in the specular reflection region for the proposed model was 0.005 while its variance was 0.075. The mean error for the Torrance–Sparrow model was 0.093 with a variance of 0.102; the mean error of the Phong model was 0.102 with a variance of 0.058.

The results of the evaluation were that the proposed method generated quantitatively better results than the Phong model and the Torrance–Sparrow model.

5.4. Experiments generating virtual images with an arbitrary light source

We performed experiments generating images of the target object with a light source direction that was entirely different from the directions used in the images of spheres

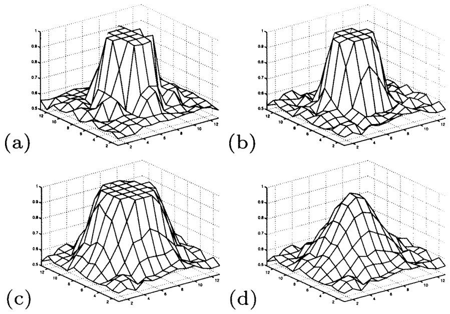

Fig. 14. Specular reflection region. (a) Real image; (b) proposed method; (c) Torrance–Sparrow model; (d) Phong model.

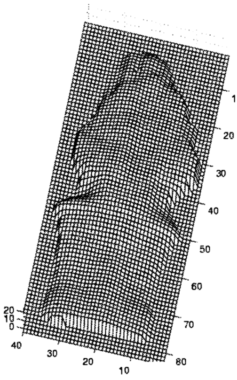

Fig. 16. Height distribution.

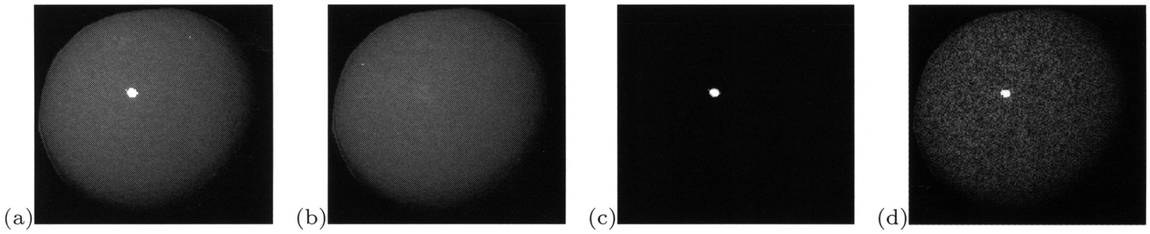

Fig. 17. (a) Image of a sphere used for training. (b) Diffuse reflection component. (c) Specular reflection component. (d) Example with randomly added reflection coefficient.

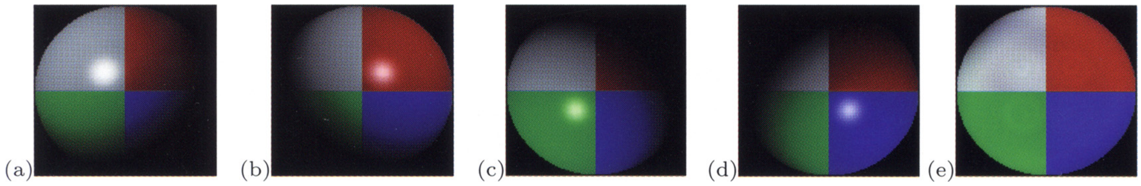

Fig. 18. (a–d) Four input images; (e) the resulting color reflectance coefficient. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

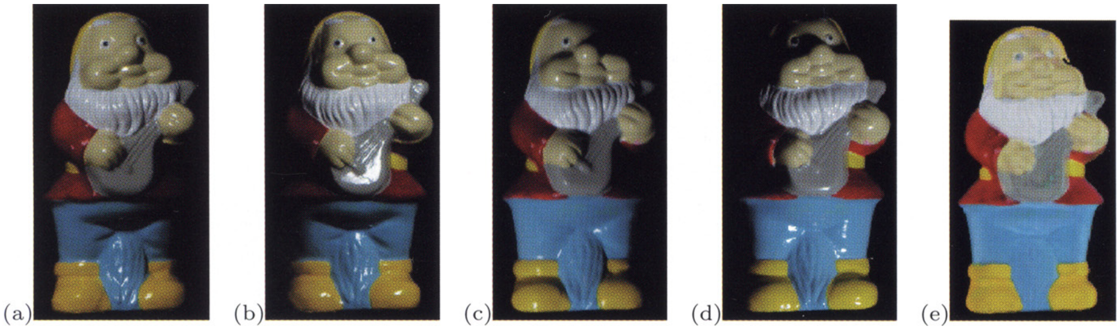

Fig. 19. (a–d) Four input images; (e) the resulting color reflectance coefficient. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

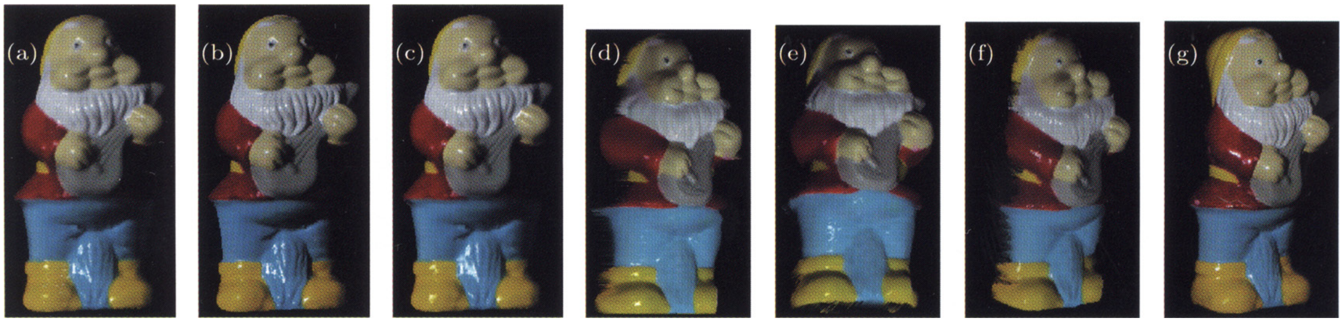

Fig. 20. (a) Virtual image generated by the proposed method; (c) Torrance–Sparrow model virtual image; (d)–(f) virtual images from arbitrary viewpoints; (b, g) real images. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

for training. The light source direction of the image generated was **l** = (–0.55, 0.35, 0.75) (this was not a light source direction that occurred in the training data).

Figure 20(a) shows the results of rendering with the proposed method. The time for rendering an image of 340 × 160 pixels was approximately 2 to 3 seconds. Figure 20(b) shows the real image with this light source direction. The difference between Figs. 20(a) and 20(b) is shown in Fig. 15. The average error between Figs. 20(a) and 20(b) is 0.073 with a variance of 0.067. For comparative purposes we show the results of rendering with the Torrance–Sparrow model in Fig. 20(c). The average error between Figs. 20(c) and 20(b) is 0.140 with a variance of 0.185.

Slight differences between the image generated by the proposed method, as with the other models, can be seen when there are local interreflections and cast shadows. Except for these points, it is confirmed that the trained neural network is able to generate virtual images that have a different light source direction than that of the training data.

5.5. Experiments generating virtual images with arbitrary viewpoints

Figure 16 shows the height distribution of the object obtained by integrating the surface normal vectors obtained from the neural network-based photometric stereo. From the results of neural network-based color photometric stereo for the target object in various different poses, we computed the integrals and performed a reciprocal rotation of each height distribution and examined the concordance rate of these; from the fact that for areas with continuous height the error was within 3 percent, we judged that these height distributions that we had integrated and created were a suitable shape. We then rotated these height distributions to generate images in different poses.

We set the light source direction as **l** = (0.40, –0.05, 0.92) and generated virtual images with the viewing direction being varied; these results are shown in Figs. 20(d) to 20(f). Figure 20(g) shows the actual object after a rotation (the viewpoint was changed); Fig. 20(f) is a virtual image for the corresponding viewpoint. The results of evaluating the error using Eq. (16) for Figs. 20(d) to 20(f) were an average error of 0.059 with a variance of 0.010. On the other hand, the Torrance–Sparrow model results in an average error of 0.084 with a variance of 0.017. Based on a comparison of the precision of our proposed method and that in Ref. 8, we have obtained better results for both the surface normal vectors and the surface reflectance coefficients for our method. In addition, as shown by Fig. 14 and the results of the precision comparison with the Torrance–Sparrow model, realistic virtual images can be generated due to the generalization capacity of the neural network-based reflection characteristic function working effectively.

6. Considerations from a Practical Perspective

The proposed method has extended a previous method to allow it to be applied to color images, but in its requirement that a sphere with the same reflectance properties as the target object be prepared for the training of the neural network it remains unchanged. Either of the following methods could be used to prepare a sphere for training purposes.

- Creating a sphere from the same material as the target object
- Painting the sphere and the target object with the same paint

If the composition and the material used in the target object are known, then it is possible to create a sphere with the same material. In addition, if the surface of the target object is coated in paint, then we can create the same surface reflectance properties by coating the sphere in the same paint. However, there will be cases where we do not know the material from which the object is made and the surface of the object cannot be coated in paint. For objects for which we cannot create a sphere to be used as reconstruction targets, we require a method for creating training data that does not require the preparation of a sphere.

In previous work, self-calibration neural network-based photometric stereo has been proposed as an attempt at finding a method that does not require a sphere [16]. Self-calibration is a method for generating training data for the neural network based on the application of geometric constraints and optical constraints on multiple images taken while rotating the object around one of its axes. However, the training data generated by this method is designed to be used with monochrome neural network-based photometric stereo [8] and it would be difficult to apply it directly to the method presented here. From a practical perspective, based on the approach presented in Ref. 16, the development of a method that does not require a sphere for training is an important issue and one we leave for future work.

7. Conclusion

In this paper we have extended neural network-based photometric stereo to allow its application to color images and presented a method for estimating the surface normal vectors of a target object as well as its color reflectance coefficients. This method does not make use of a parametric function to characterize the reflection of the object simply using the general assumption of a dichromatic reflection model. In addition, by splitting the neural network into two stages, we have been able to avoid an explosion in the amount of combinations that must be used as training data, and been able to perform training with the same scale of training data and the same amount of time as the previous method while achieving higher estimation accuracy.

In addition, based on the results of neural network-based color photometric stereo, we have proposed a neural network-based rendering method to enable the generation of virtual images of an object with an arbitrary light source direction and from an arbitrary viewpoint. The results of comparisons of these virtual images that we have generated with the actual images, confirmed that we were able to render specular reflections with a higher level of accuracy than models using a parametric reflectance function.

However, since the proposed method assumed that cast shadows and interreflections do not occur on the target object, these problems are left to future work. Besides this, future work will include reconstruction and virtual image generation for the entire circumference of the object and the development of a method that does not make use of a sphere in training.

Acknowledgments. We thank Professor S. Sato of Nagoya Institute of Technology for help and advice during this work. This research was supported by a Scientific Research Grant [Basic Research (C) (2) 16500108], the Kayamori Foundation of Informational Science Advancement (Yuji Iwahori), the Hori Information Science Promotion Foundation (Kenji Funahashi), and the Institute for Robotics and Intelligent Systems and the Natural Sciences and Engineering Research Council (Robert J. Woodham).

REFERENCES

1. Sato Y, Ikeuchi K. Temporal-color space analysis of reflection. J Opt Soc Am A 1994;11:2990–3002.
2. Sato Y, Wheeler MD, Ikeuchi K. Object shape and reflectance modeling from observation. Proc ACM SIGGRAPH 97, p 379–387.
3. Nishiyama Y, Saito H, Ozawa S. Estimation of 3-dimensional shape and object color from multiple viewpoint color images. Trans IEICE 1999;J82-D-II:990–1000. (in Japanese)
4. Omata K, Saito H, Ozawa S. Estimation of object shape and surface reflectivity characteristics using relative rotations of light source. Trans IEICE 2000;J83-D-II:927–937. (in Japanese)
5. Woodham RJ. Photometric method for determining surface orientation from multiple images. Opt Eng, p 139–144, 1980.
6. Woodham RJ. Gradient and curvature from the photometric-stereo method, including local confidence estimation. J Opt Soc Am A, p 3050–3068, 1994.
7. Iwahori Y, Woodham RJ, Ozaki M, Tanaka H, Ishii N. Neural network based photometric stereo with a nearby rotational moving light source. IEICE Trans Inf Syst 1997;E80-D:948–957.
8. Iwahori Y, Woodham RJ, Bhuiyan MS, Ishii N. Neural network based photometric stereo for object with non-uniform reflectance factor. Proc 6th International Conference on Neural Information Processing (ICONIP'99), Vol. III, p 1213–1218.
9. Tomita Y, Kaneko S, Honda T. 3-Dimensional shape reconstruction using color photometric stereo. Technical Report of the SIG Computer Vision of the Information Processing Society of Japan, 080-016, p 115–120, 1992. (in Japanese)
10. Barsky S, Petrou M. Color photometric stereo: Simultaneous reconstruction of local gradient and color of rough textured surfaces. Eighth IEEE Int Conf on Computer Vision, Vol. II, p 600–605, 2001.
11. Ikeuchi K, Sato Y, Nishino W, Sato I. Creating optical consistency for mixed reality images. Journal of the Virtual Reality Society of Japan, Special Issue on Mixed Reality 1999;4:623–630. (in Japanese)
12. Tominaga M, Ohashi S. A model of object color reflectivity. J Inf Process Soc Japan 1992;33:37–45. (in Japanese)
13. Chen S, Cowan CFN, Grant PM. Orthogonal least squares learning algorithm for radial basis function networks. IEEE Trans Neural Networks 1991;2:302–309.
14. Fukui S, Iwahori H, Woodham RJ, Iwada A. Classifying curves from monochrome image using neural networks and illumination planning. Trans IEICE 2000;J83-D-II:610–622. (in Japanese)
15. Iwahori Y, Kawanaka H, Woodham RJ, Funahashi K. Neural network based modeling and color rendering for mixed reality. Proc 17th International Conference on Pattern Recognition (ICPR2004).
16. Iwahori Y, Watanabe Y, Woodham RJ, Iwata A. Self-calibration and neural network implementation of photometric stereo. Proc 16th International Conference on Pattern Recognition, 2002.

AUTHORS (from left to right)

Haruki Kawanaka (full member) graduated from the Department of Engineering, Electrical Engineering and Information Science, Nagoya Institute of Technology, in 1999 and completed the master's and doctoral programs in electrical engineering and information science in 2001 and 2005. Currently he is a postdoctoral researcher in the Center for Advanced Measurement Research at Chubu University. His research interests are in electron microscopy, neural networks, computer vision, and mixed virtual reality. He holds a D.Eng. deree, and is a member of the Institute of Image Electronics Engineers of Japan.

Yuji Iwahori (full member) graduated from the Department of Information Sciences at Nagoya Institute of Technology in 1983, completed the master's and doctoral programs in electronics and electrical engineering at Tokyo Institute of Technology in 1985 and 1988, and became a research associate in the Center for Information Processing Education at Nagoya Institute of Technology. He was appointed a professor at the Center for Information and Media Studies in 2002. He has been a professor in the Department of Computer Science at Chubu University since 2004, and also a visiting researcher at the University of British Columbia. His research interests are in computer vision, neural networks, and mixed reality. He holds a D.Eng. degree, and is a member of the Information Processing Society of Japan and IEEE.

Robert J. Woodham graduated from the Department of Mathematics at the University of Western Ontario in 1971 and completed the master's program in mathematics and electrical engineering at Massachusetts Institute of Technology (MIT) in 1974 and the doctoral program at the same university (AI Lab) in 1977. After serving as a doctoral researcher at MIT, he joined the University of British Columbia in 1978, and has been a professor of computer science there since 1990. His research interests are in computer vision and pattern recognition. He holds a Ph.D. degree, and is a member of IEEE.

Kenji Funahashi (full member) graduated from the Department of Electrical Engineering and Information Science at Gifu University in 1993, completed the master's and doctoral programs in information engineering at Nagoya University in 1995 and 1998, and became a research associate at the Center for Information Processing Education at Nagoya Institute of Technology. He moved to the Center for Information Media Education in 2000, and has been an associate professor there since 2005. His research interests are in virtual reality and computer graphics. He holds a D.Eng. degree, and is a member of the Society for Virtual Reality of Japan.