Photometric stereo under a light source with arbitrary motion

Hideki Hayakawa*

ATR Human Information Processing Research Laboratories, 2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02, Japan

Received September 27, 1993; revised manuscript received March 28, 1994; accepted April 6, 1994

A new photometric-stereo method for estimating the surface normal and the surface reflectance of objects without *a priori* knowledge of the light-source direction or the light-source intensity is proposed. First, I construct a $p \times f$ image data matrix I from p pixel image intensity data through f frames by moving a light source arbitrarily. Under the Lambertian assumption the image data matrix I can be written as the product of two matrices S and L, with S representing the surface normal and the surface reflectance and L representing the light-source direction and the light-source intensity. Using this formulation, I show that the image data matrix I is of rank 3. On the basis of this observation, I use a singular-value decomposition technique and useful constraints to factorize the image data matrix. This method can also be used to treat cast shadows and self-shadows without assumptions. The effectiveness of this method is demonstrated through performance analysis, laboratory experiment, and out-of-laboratory experiment.

1. INTRODUCTION

Photometric stereo is a method for shape estimation that uses several intensity images obtained under different lighting conditions.[1] The method can be used for specular surfaces if extended light sources with spatially varying brightness are used.[2,3] Surface curvature can be obtained from photometric-stereo data if one uses brightness gradients in addition to brightness itself.[4] Surface depth can also be estimated from photometric-stereo data when a point light source is relatively near the surface and near the camera.[5,6] The photometric-stereo technique has been practically applied to scanning electron microscopic images,[7] and data provided by photometric stereo have been used for some industrial inspection and part-alignment tasks.[8,9] These applications show that photometric stereo is a good candidate for propelling the commercialization of machine vision techniques. However, with previous methods it is impossible to estimate the surface normal and the surface reflectance without *a priori* knowledge of both the light-source direction and the light-source intensity.

In this paper I propose a new photometric-stereo method for estimating the surface normal and the surface reflectance of objects without *a priori* knowledge of the light-source direction or light-source intensity. In this method, assuming only that the object's surface is Lambertian, the surface normal, the surface reflectance, the light-source direction, and the light-source intensity can be determined simultaneously. For example, we can estimate these four parameters from intensity images that are obtained under a light source arbitrarily moved by a human. This method does not rely on any smoothness assumptions for these parameters. Furthermore, as the number of intensity images increases, the surface normal, the surface reflectance, the light-source direction, and the light-source intensity errors become smaller even if the intensity images are taken in a noisy environment.

The foundation of this method is similar to that of a factorization method for shape and motion estimation from image streams.[10] In this method, I construct a $p \times f$ image data matrix I from p pixel image intensity data through f frames by moving a light source arbitrarily. Under the Lambertian assumption, the image data matrix I can be written as the product of two matrices S and L, with S representing the surface characteristics (surface normal and reflectance) and L representing the light characteristics (light-source direction and intensity). Using this formulation, I show that the image data matrix I is of rank 3. On the basis of this observation, I use a singular-value decomposition (SVD) technique and one of the two following constraints to factorize the image data matrix. One constraint is the constraint of surface reflectance. This constraint can be used when there are at least 6 pixels in which the relative value of the surface reflectance is constant or is known. The other is the constraint of light-source intensity. This constraint can be used when there are at least 6 frames in which the relative value of the light-source intensity is constant or is known. The idea of applying the SVD technique to observed image intensities to separate them into fundamental components was introduced previously.[11,12]

The present method also describes how to deal with shadow regions. In an intensity image, there are two types of shadows: self-shadows and cast shadows. Previous methods could not deal with either shadow type without relying on certain assumptions. This method can easily treat both types of shadow without the use of assumptions. In the image data matrix, we select an initial submatrix having no shadowed data. We can then estimate the surface normal and reflectance in shadow regions by growing a partial solution obtained from the initial submatrix.

The effectiveness of this method is demonstrated through performance analysis and through a laboratory experiment on Lambertian reflectance objects. Further-

0740-3232/94/113079-11$06.00

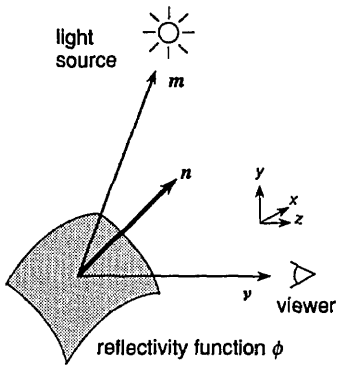

Fig. 1. Geometric reflectance model for image generation in the viewer-oriented coordinate system. Here **n**, **m**, and **v** denote the three-dimensional (3D) unit vectors of surface normal, light-source direction, and viewer direction, respectively.

more, this method is applied to scenes of an architectural model of the ATR building that are photographed with a hand-held camera.

In Section 2 the photometric stereo, which is the computational basis of this method, is described. A method for factorization of the image data matrix and a method for dealing with shadow regions are proposed in Sections 3 and 4, respectively. In Section 5 the performance of this method by computer simulation is analyzed, and the experimental results in real intensity images are discussed.

2. PHOTOMETRIC STEREO REVISITED

Photometric stereo can estimate the surface normal and the surface reflectance of objects from three or more intensity images under different lighting conditions.[1] The images are obtained with a fixed camera and fixed objects.

Figure 1 shows a geometric reflectance model for image generation. The fraction of light reflected by an object's surface in a given direction depends on the optical properties of the surface material, the surface microstructure, and the spatial and the spectral distribution of the incident illumination. For many surfaces the fraction of incident illumination reflected in a particular direction depends only on the surface normal. The reflectance characteristics of such a surface can be represented as a reflectance function ϕ of the three unit vectors: surface normal $\mathbf{n} = (n_x, n_y, n_z)^T$, light-source direction $\mathbf{m} = (m_x, m_y, m_z)^T$, and viewer direction $\mathbf{v} = (\nu_x, \nu_y, \nu_z)^T$, defined in Fig. 1. Using the reflectance function ϕ, we can write the basic equation describing the image-generation process as

$$i = t \cdot \phi(\mathbf{n}, \mathbf{m}, \mathbf{v}), \tag{1}$$

where i denotes the image intensity and t denotes the light-source intensity. Assuming that the image projection is orthographic and that the incident illumination is provided by a single distant point light source, the viewer direction and the light-source direction can be considered to be constant over the image plane. When both the light-source directions and the light-source intensities are known and three light-source directions do not lie in a plane, the surface normal can be estimated. This is the principle of the photometric-stereo method.

One simple idealized model of surface reflectance is given by

$$i(x, y) = r(x, y)t[\mathbf{n}(x, y) \cdot \mathbf{m}] \tag{2}$$

for $\mathbf{n}(x, y) \cdot \mathbf{m} \geq 0$. This reflectance function corresponds to the phenomenological model of a Lambertian surface. Here r is the surface reflectance and (x, y) is the xy coordinate of the viewer-oriented coordinate system, which is defined in Fig. 1. Assuming that each light-source intensity in three intensity images is known, three normalized image intensities $i'(x, y)$ can be defined with each light-source intensity:

$$i_k'(x, y) = r(x, y)[\mathbf{n}(x, y) \cdot \mathbf{m}_k], \qquad k = 1, 2, 3. \tag{3}$$

If we construct a matrix M by using these light-source direction vectors as columns, Eq. (3) is rewritten as

$$I'(x, y) = r(x, y)[\mathbf{n}(x, y)]^T M, \tag{4}$$

where

$$I'(x, y) = [i_1'(x, y) \quad i_2'(x, y) \quad i_3'(x, y)],$$
$$M = [\mathbf{m}_1 \quad \mathbf{m}_2 \quad \mathbf{m}_3].$$

Because the three known light-source directions $\mathbf{m}_1$, $\mathbf{m}_2$, and $\mathbf{m}_3$ do not lie in a plane, the inverse of the matrix M exists. In this case the surface reflectance r and the surface normal $\mathbf{n}$ at (x, y) are given by

$$r(x, y) = \|I'(x, y)M^{-1}\|, \tag{5}$$

$$\mathbf{n}(x, y) = \frac{1}{r(x, y)} I'(x, y)M^{-1}. \tag{6}$$

The photometric-stereo method was improved and was extended as described in Section 1. At this time, this method requires *a priori* knowledge of both the light-source direction and the light-source intensity. This method also requires that the surface point not be in the shadow of any light source.

3. FACTORIZATION OF IMAGE DATA MATRIX

A. Image Data Matrix Formulation

Assume that we measure the image intensity data i at p pixels through f frames by moving only a light source. We write the image intensity data i into a $p \times f$ matrix I, with 1 row/pixel and 1 column/frame:

$$I = \begin{bmatrix} i_{11} & \cdots & i_{1f} \\ \cdot & \cdots & \cdot \\ i_{p1} & \cdots & i_{pf} \end{bmatrix}. \tag{7}$$

The matrix I is called the image data matrix. Assuming a Lambertian shading model with a single distant point

light source in the measurement, we can express the image data matrix I in matrix form, as in Eq. (2):

$$I = RNMT, \tag{8}$$

where

$$R = \begin{bmatrix} r_1 & & 0 \\ & \ddots & \\ 0 & & r_p \end{bmatrix} \tag{9}$$

is the surface reflectance matrix,

$$N = [\mathbf{n}_1 \quad \cdots \quad \mathbf{n}_p]^T = \begin{bmatrix} n_{1x} & n_{1y} & n_{1z} \\ \vdots & \vdots & \vdots \\ n_{px} & n_{py} & n_{pz} \end{bmatrix} \tag{10}$$

is the surface normal matrix,

$$M = [\mathbf{m}_1 \quad \cdots \quad \mathbf{m}_f] = \begin{bmatrix} m_{x1} & \cdots & m_{xf} \\ m_{y1} & \cdots & m_{yf} \\ m_{z1} & \cdots & m_{zf} \end{bmatrix} \tag{11}$$

is the light-source direction matrix, and

$$T = \begin{bmatrix} t_1 & & 0 \\ & \ddots & \\ 0 & & t_f \end{bmatrix} \tag{12}$$

is the light-source intensity matrix. r and $\mathbf{n}$ represent the surface reflectance and the surface normal at each pixel, respectively. t and $\mathbf{m}$ represent the light-source intensity and the light-source direction at each frame, respectively. Because subscripts x, y, z in Eqs. (10) and (11) represent an arbitrary three-dimensional (3D) coordinate system, they do not, in general, correspond to the viewer-oriented coordinate system shown in Fig. 1.

On the basis of these equations, the surface matrix S and the light-source matrix L are defined as follows:

$$S = [\mathbf{s}_1 \quad \cdots \quad \mathbf{s}_p]^T = \begin{bmatrix} s_{1x} & s_{1y} & s_{1z} \\ \vdots & \vdots & \vdots \\ s_{px} & s_{py} & s_{pz} \end{bmatrix} = RN,$$

$$L = [\mathbf{l}_1 \quad \cdots \quad \mathbf{l}_f] = \begin{bmatrix} l_{x1} & \cdots & l_{xf} \\ l_{y1} & \cdots & l_{yf} \\ l_{z1} & \cdots & l_{zf} \end{bmatrix} = MT. \tag{13}$$

The magnitude of surface vector $\mathbf{s}$ in surface matrix S represents the surface reflectance. The magnitude of light-source vector $\mathbf{l}$ in light-source matrix L represents the light-source intensity. Using Eqs. (13), one obtains the other description of Eq. (8) as follows:

$$I = SL. \tag{14}$$

Assuming that three surface normals do not lie in a plane, the surface matrix S is of rank 3. In the same way, assuming that three light-source directions do not lie in a plane, the light-source matrix L is of rank 3. Therefore, without noise, the image data matrix I is also of rank 3.

B. Singular-Value Decomposition and Approximate Rank

Assuming that $p \geq f$, the matrix can be decomposed into a $p \times f$ matrix U, a diagonal $f \times f$ matrix Σ, and an $f \times f$ matrix V:

$$I = U\Sigma V, \tag{15}$$

where $U^TU = V^TV = VV^T = E$, where E is the $f \times f$ identity matrix. Here the assumption $p \geq f$ is not crucial: if $p < f$, everything can be repeated for the transposed matrix of I. Σ is a nonnegative diagonal matrix whose diagonal entries are the singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_f \geq 0$ sorted in nonincreasing order. This is the SVD of the matrix I. Focusing only on the first three columns of U, the first 3×3 submatrix of Σ, and the first three rows of V, we can partition these matrices as follows:

$$U = [\underbrace{U'}_{3} \quad \underbrace{U''}_{f-3}]\}p,$$

$$\Sigma = \begin{bmatrix} \Sigma' & 0 \\ \underbrace{0}_{3} & \underbrace{\Sigma''}_{f-3} \end{bmatrix} \begin{matrix} \}3 \\ \}f-3 \end{matrix},$$

$$V = \begin{bmatrix} V' \\ \underbrace{V''}_{f} \end{bmatrix} \begin{matrix} \}3 \\ \}f-3 \end{matrix}. \tag{16}$$

Taking into consideration the fact that the image data matrix I has image intensity noise, we let I^* be the ideal image data matrix, i.e., the matrix that we would obtain in the absence of noise. In the case in which Σ' contains all the singular values of I^* that exceed the noise level, the best possible rank-3 approximation to the ideal image data matrix I^* is the product

$$\hat{I} = U'\Sigma'V'. \tag{17}$$

Thus $\hat{I}$ is the best estimate of I^*. Whether the noise level is low enough to be ignored for this approximation also depends on the surface normal at p pixels and on the light-source direction through f frames. The SVD method yields sufficient information for judging the approximation. The requirement is that the ratio between the third and the fourth largest singular values of I be sufficiently large.

Using Eq. (17), we define the pseudo surface matrix $\hat{S}$ and the pseudo light-source matrix $\hat{L}$ as

$$\hat{S} = U'(\pm[\Sigma']^{1/2}),$$

$$\hat{L} = (\pm[\Sigma']^{1/2})V', \tag{18}$$

and we can write

$$\hat{I} = \hat{S}\hat{L}. \tag{19}$$

Here Eqs. (18) have a sign ambiguity. There are two different signs for the pseudo solutions in Eqs. (18). They correspond to the solutions in the right-handed and the

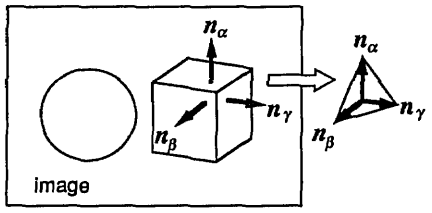

Fig. 2. Relationship of a triplet of surface normals. $\mathbf{n}_\alpha$, $\mathbf{n}_\beta$, $\mathbf{n}_\gamma$ are the assumed three surface normals in an image, which do not lie in a plane. A triangle, whose vertices correspond to the terminal points of the three assumed surface normals, is drawn.

left-handed coordinate systems. However, without any knowledge, we cannot judge which sign corresponds to the solution in the right-handed coordinate system.

Using the relationship of an arbitrary triplet of the surface normals in an image, which do not lie in a plane, we can choose the solution in the right-handed coordinate system. As shown in Fig. 2, we assume three surface normals in the image, which do not lie in a plane, and draw a triangle, whose vertices correspond to the terminal points of these assumed surface normals. We extract three pseudo surface vectors that correspond to the assumed surface normals from two pairs of pseudo surface matrices $\hat{S}$ in Eqs. (18). Then we calculate the determinant of the 3×3 matrices, which are constructed from the three pseudo surface vectors in the counterclockwise direction of the triangle. Consequently the accuracy of the assumption requires that we have the ability to determine this order correctly. We choose the pseudo surface matrix $\hat{S}$ that corresponds to the 3×3 matrix whose determinant's sign is positive, and we also select the paired pseudo light-source matrix $\hat{L}$ with the chosen pseudo surface matrix $\hat{S}$. The method of selection depends on the fact that all the surface normals in the image have a positive z element in the viewer-oriented coordinate system. Note that we can choose the solution in the right-handed coordinate system from the two solutions in Eqs. (18) by using the relationship of an arbitrary triplet of light-source directions, which do not lie in a plane.

C. Useful Constraints

The two pseudo matrices $\hat{S}$ and $\hat{L}$ are of the same size as the desired surface matrix S and light-source matrix L. However, the decomposition in Eq. (19) is not unique. This is because, if A is an arbitrary invertible 3×3 matrix, then the matrices $\hat{S}A$ and $A^{-1}\hat{L}$ are also a valid decomposition of $\hat{I}$ as follows:

$$\hat{S}AA^{-1}\hat{L} = \hat{S}(AA^{-1})\hat{L} = \hat{S}\hat{L} = \hat{I}.$$

Thus, because $\hat{S}$ and $\hat{L}$ are different from S and L in general, we must find the matrix A such that

$$\begin{aligned} S &= \hat{S}A, \\ L &= A^{-1}\hat{L}. \end{aligned} \tag{20}$$

However, it is impossible to find the matrix A without any knowledge of the surface and the light characteristics. Accordingly, we use one of the two following useful constraints:

(C1) We can find at least 6 pixels in which the relative value of the surface reflectance is constant or known.

(C2) We can find at least 6 frames in which the relative value of the light-source intensity is constant or known.

A way of finding the matrix A by use of constraint (C1) is as follows. From the pseudo surface matrix $\hat{S}$, we extract p' (≥ 6) pseudo surface vectors $\hat{\mathbf{s}}$, of which the relative value of the surface reflectance is constant or known. Assuming that three vectors among them do not lie in a plane, the following system of p' equations can be solved when the relative value of the surface reflectance is constant:

$$\hat{\mathbf{s}}_k{}^T AA^T \hat{\mathbf{s}}_k = 1, \qquad k = 1, \cdots, p'. \tag{21}$$

Here, when the relative value of the surface reflectance for 6 or more pixels is known, each numeral of the ratio is set to the right-hand side of Eqs. (21) in place of 1. If we introduce the symmetric matrix $B = AA^T$, we can rewrite Eqs. (21) as

$$\hat{\mathbf{s}}_k{}^T B \hat{\mathbf{s}}_k = 1, \qquad k = 1, \cdots, p'. \tag{22}$$

This is a linear problem, so the solution is straightforward. Once B is determined, we can find a matrix A such that $B = AA^T$ by taking the SVD of B. The SVD of a symmetric matrix is itself symmetric: $B = W\Pi W^T$, where Π is diagonal and W is orthonormal. We let $A = W[\Pi]^{1/2}$. In the decomposition, we do not use the other sign $W(-[\Pi]^{1/2})$ to preserve the relationship of the right-handed coordinate system. The calculation for constraint (C2) is identical.

D. Outline of the Algorithm

From the above development, we have an algorithm to factorize image data matrix I into surface matrix S and light-source matrix L.

Step 1. Compute the SVD of the image data matrix I as

$$I = U\Sigma V.$$

Step 2. Define

$$\begin{aligned} \hat{S} &= U'(\pm[\Sigma']^{1/2}), \\ \hat{L} &= (\pm[\Sigma']^{1/2})V', \end{aligned}$$

where the primes are defined in Eqs. (16), and choose a solution in the right-handed coordinate system.

Step 3. Compute the matrix A in Eqs. (20) by using constraint (C1) or (C2).

Step 4. Compute the surface matrix S and the light-source matrix L as

$$\begin{aligned} S &= \hat{S}A, \\ L &= A^{-1}\hat{L}. \end{aligned}$$

The derived surface matrix S and light-source matrix L are represented in an arbitrary 3D coordinate system. If we know three surface normals or three light-source directions that are represented in the viewer-oriented coordinate system, then these matrices S and L can be automatically aligned to the viewer-oriented coordinate system by solution of an absolute orientation problem.[13]

4. SHADOWS

A prerequisite for using methods based on the photometric-stereo method is that the surface point not be in the shadow of any light source. In reality, the light-source position in relation to the object and the camera is changed, causing frequent self-shadows and cast shadows in the image data. Because this phenomenon is frequent enough to render a shape-computation method impossible, I propose a method of dealing with shadow regions in the factorization of the image data matrix I.

For preparation, we classify the image intensity data of the $p \times f$ matrix I into shadowed data and illuminated data. For noise-free images without ambient illumination, the intensity level of the shadowed data is 0, and the intensity level of the illuminated data is larger than 0. In the presence of noise and ambient illumination, we can easily determine an appropriate threshold value for the classification by confirming that the region in which the image intensity is less than the threshold value coincides with the actual shadow region.

First, we extract a $q \times g$ submatrix $\tilde{I}$ of the $p \times f$ image data matrix I, in which all the elements are illuminated data. In practice, to extract a possibly large submatrix $\tilde{I}$, the following two processes are repeated iteratively: while $q > g$, rows with the most shadowed data are removed; while $q \leq g$, columns with the most shadowed data are removed. Here we consider q and g to be monotonically decreasing variables whose initial values are p and f, respectively. The extracted submatrix $\tilde{I}$ is decomposed by the method in Subsection 3.B, and the q pseudo surface vectors and the g pseudo light-source vectors can be derived.

Next, using the derived q pseudo surface vectors and g pseudo light-source vectors, we estimate the unknown $p-q$ pseudo surface vectors and the $f-g$ pseudo light-source vectors. A way to solve an unknown pseudo surface vector $\hat{\mathbf{s}}_i$ in the least-squares-error sense is as follows: Assuming the existence of a known pseudo light-source vector $\hat{\mathbf{l}}_j$, which generates illuminated data with the unknown pseudo surface vector $\hat{\mathbf{s}}_i$, we have the next equation from Eq. (14)

$$i = \hat{\mathbf{s}}_i^T \cdot \hat{\mathbf{l}}_j . \tag{23}$$

If we have three or more of such known pseudo light-source vectors, where at least three vectors do not lie in a plane,[14] we can obtain the unknown pseudo surface vector $\hat{\mathbf{s}}_i$ by solving a linear problem. In contrast, the way to obtain an unknown pseudo light-source vector with Eq. (23) is identical to the above-mentioned method.

During estimation of the unknown $p-q$ pseudo surface vectors and $f-g$ pseudo light-source vectors, the order with which we solve these unknown pseudo vectors is the reverse of that previously used to remove the rows (corresponding to the unknown pseudo surface vectors) and the columns (corresponding to the unknown pseudo light-source vectors). In the estimation, we first calculate the values of more reliable pseudo vectors, i.e., those having more illuminated data and less shadowed data. And we can use not only the known q pseudo surface vectors and g pseudo light-source vectors but also the calculated values of the more reliable pseudo vectors to estimate the values of less reliable pseudo vectors. As a result, the total value of the estimation error can become small in comparison with that for random-order estimation.

Finally, with the derived p pseudo surface vectors and f pseudo light-source vectors, the p surface vectors and the f light-source vectors can be calculated as described in Subsection 3.C.[15]

5. EXPERIMENTS

A. Performance Analysis

To analyze experimentally the performance of our method, synthetic intensity images are employed. Sixteen examples of the synthetic intensity images are shown in Fig. 3. We generated random surface normals at 64 × 64 pixels, with a uniform distribution on the Gaussian

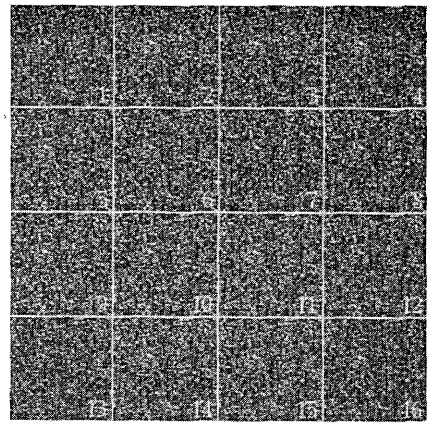

Fig. 3. Examples of synthetic intensity images used for performance analysis.

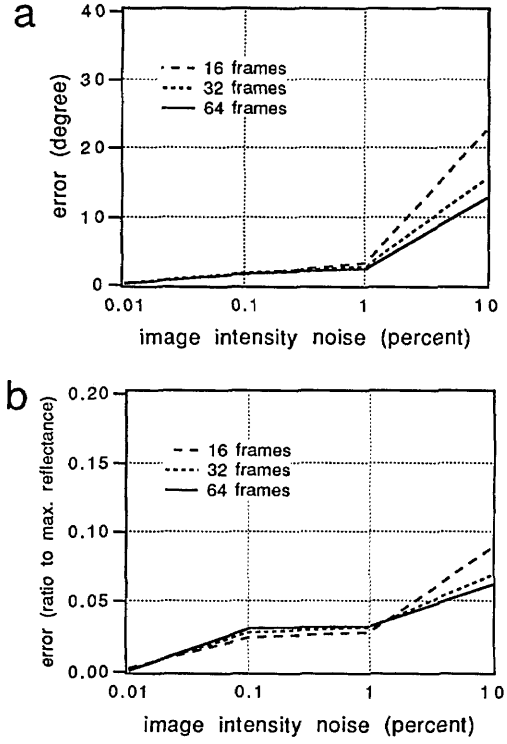

Fig. 4. Estimation error versus image intensity noise for different frame counts: (a) surface normal error, (b) surface reflectance error.

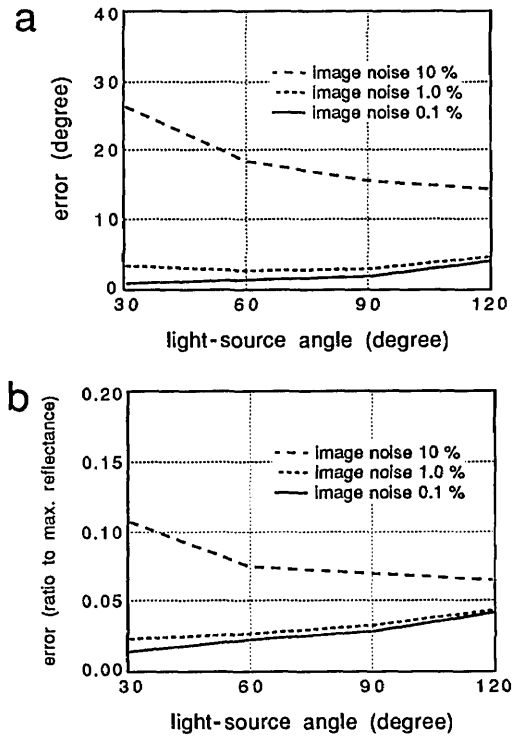

Fig. 5. Estimation error versus light-source angle for different standard deviations of image intensity noise: (a) surface normal error, (b) surface reflectance error.

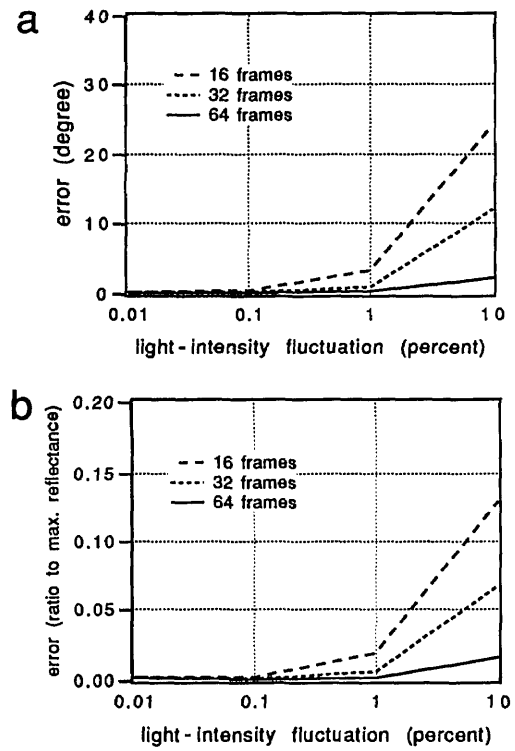

Fig. 6. Estimation error versus light-source intensity fluctuation among images for different frame counts when the constraint of light-source intensity is used: (a) surface normal error, (b) surface reflectance error.

sphere; the surfaces are limited to visible surfaces from the camera. Random surface reflectances are also generated at 64 × 64 pixels, with a uniform distribution from the value 0.1 to the value 1.0. We define light-source angle as the range in which the light source can be moved. The directions of the moving light source have a uniform distribution on the Gaussian sphere, and they are limited to within the light-source angle. The subsequent intensity images have an 8-bit resolution. Because the light-source intensities in the intensity image generation are assumed to be a constant whose value is 1.0, the constraint of light-source intensity (C2) is used for subsequent factorizations. Furthermore, by using the actual three light-source directions, which are represented in the viewer-oriented coordinate system, we can select the solution of the right-handed coordinate system from the two solutions in Eqs. (18), and the estimated surface normals and light-source directions can be transformed to the representation in the viewer-oriented coordinate system.

First, we evaluate the effect of image intensity noise for different frame counts. We generate 64 intensity images when the light-source angle is fixed at 90° and add Gaussian image intensity noise to the images with various standard deviations. Each standard deviation is normalized by the maximum image intensity of the images. Figure 4 shows surface normal error and reflectance error versus image intensity noise. The surface normal error

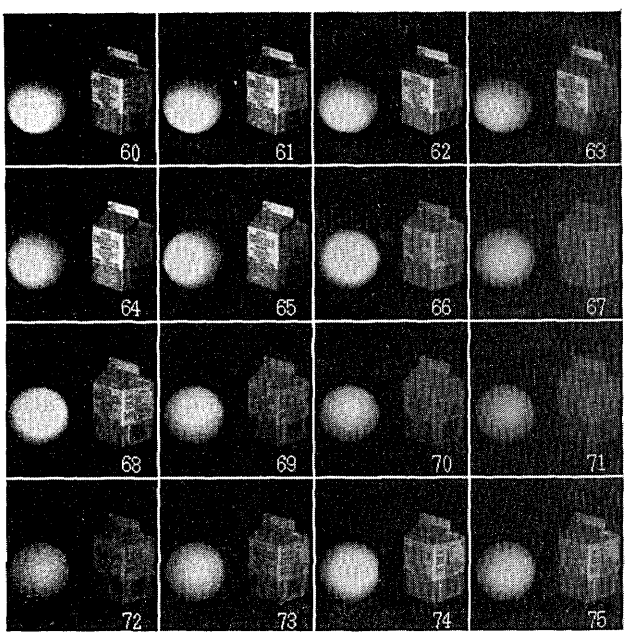

Fig. 7. Intensity images of a sphere and a milk carton that are derived when a light source is arbitrarily moved by a human.

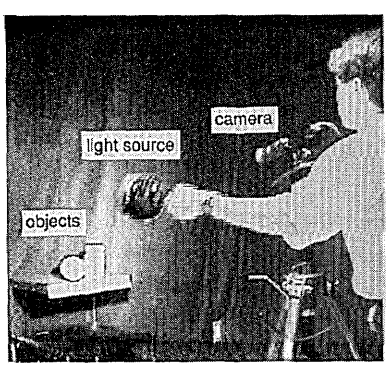

Fig. 8. Measurement scene for obtaining intensity images with a CCD camera when a light source is arbitrarily moved by a human.

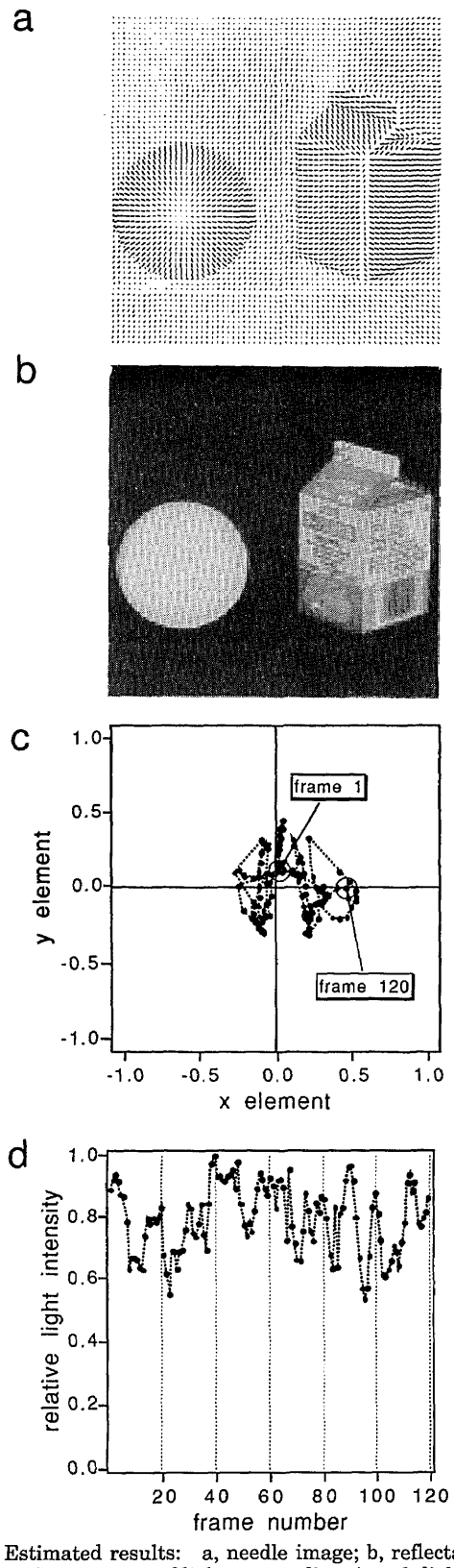

Fig. 9. Estimated results: a, needle image; b, reflectance image; c, x and y elements of light-source direction; d, light-source intensity.

is defined as the 3D angle error on the estimated surface normal, averaged over all the surface data. The surface reflectance error is defined as the absolute value of the error on the estimated surface reflectance, averaged over all the surface data and divided by the maximum value of the true surface reflectances. The diagrams show that the errors increase with more image intensity noise and fewer frames of intensity images.

Next, we evaluate the effect of light-source angle for different standard deviations of image intensity noise. We generate 4 sets each with 32 intensity images while changing the light-source angle and add Gaussian image intensity noise to the images with various standard deviations. Figure 5 shows surface normal error and reflectance error versus light-source angle. The errors decrease with large light-source angle when the image intensity noise is relatively high. The errors increase slightly when the image intensity noise is relatively low because a larger light-source angle makes shadow regions larger.

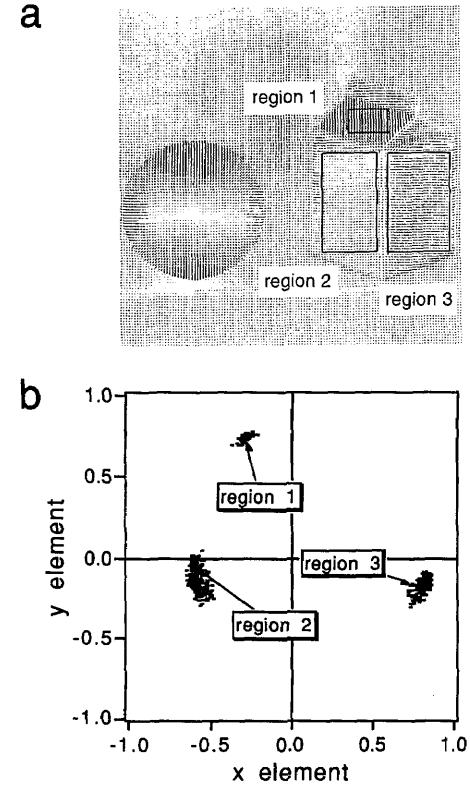

Fig. 10. Quantitative evaluation of the estimated surface normal: a, region of extracted surface normal; b, distribution of x and y elements of the extracted surface normal.

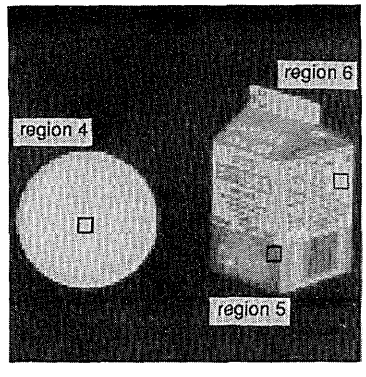

Fig. 11. Region of extracted surface reflectance.

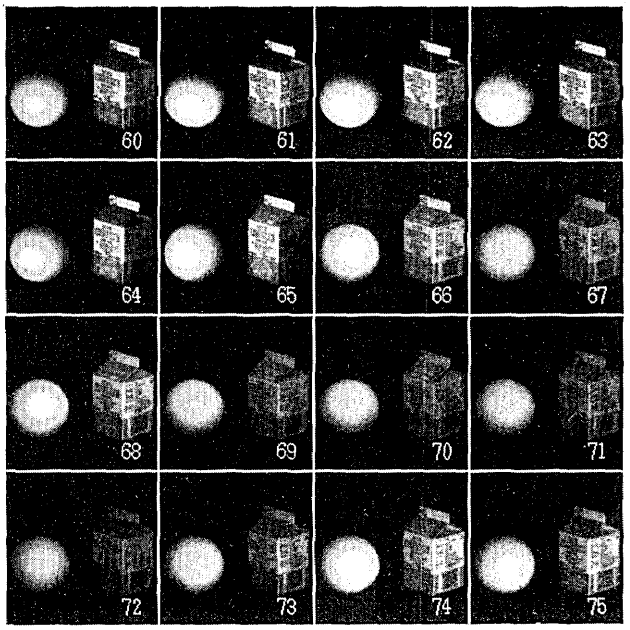

Fig. 12. Reconstructed intensity images. The frame numbers correspond to those shown in Fig. 7.

Finally, we evaluate the effect of light-source intensity fluctuation among images when the constraint of light-source intensity (C2) is used. We generate 64 intensity images when the light-source angle is fixed at 90° and add light-source intensity fluctuations among images with Gaussian distributions. Each standard deviation of the fluctuations is normalized by the actual light-source intensity value, 1.0. Figure 6 shows surface normal error and reflectance error versus light-source intensity fluctuation among images for different frame counts. When the constraint of light-source intensity (C2) is used with the light-source intensity fluctuation whose standard deviation is 10% of the light-source intensity, 64 or more intensity images are required.

B. Laboratory Experiment

In this subsection a laboratory experiment on 120 intensity images of a sphere and a milk carton is described. The intensity images are 128 × 128 pixel images with an 8-bit resolution. Sixteen images (frame 60 to frame 75) are shown in Fig. 7. The sphere is painted with dull white paint to give it a Lambertian reflectance, and the reflectance of the milk carton can be roughly regarded as Lambertian. These intensity images are derived when a light source is arbitrarily moved by a human for approximately 1 min. The intensity image acquisition can be performed in real time because an image-measurement system with a real-time disk device is used. Figure 8 shows a measurement scene for this experiment.

We set the threshold value for the classification into shadowed data and illuminated data at a tenth of the maximum intensity among all the image data. Because the light-source intensity is not constant in the experiment, the constraint of surface reflectance (C1) is applied to approximately 1200 pixels on a part of the sphere. The surface normal of the sphere can be roughly calculated from the coordinate of the center and from the radius of the sphere in the images. By using the result of the rough calculation, we can transform the estimated surface normals and light-source directions, which are derived from the factorization of the image data matrix, to the representation in the viewer-oriented coordinate system. Here, because no choice has yet been made for the solution in the right-handed coordinate system from the two solutions in Eqs. (18), both solutions are transformed. After the transformation, the average angular errors between the surface normal derived from the rough calculation and the estimated surface normal of each of the two solutions derived from the factorization are 3.7°

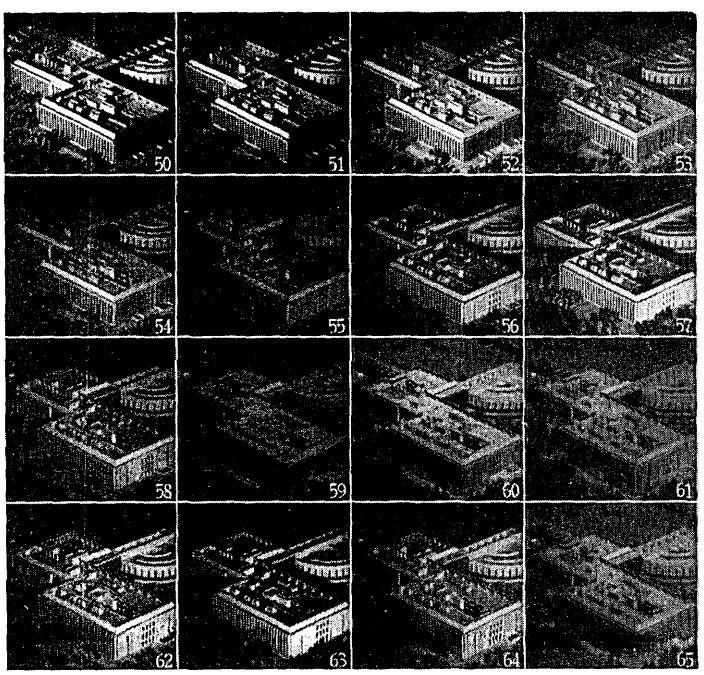

Fig. 13. Intensity images of an architectural model of the ATR building that are derived when a light source is arbitrarily moved by a human.

and 43.1°, respectively. We choose the solution with the smaller error because the solution in the left-handed coordinate system continues to have a large error despite the transformation to the viewer-oriented coordinate system. Figure 9 shows the estimated results. Figure 9a shows the needle image produced by the estimated surface normal. The needle image is subsampled with a 1:2 ratio for clarity. Figure 9c shows only the x and the y elements of the estimated light-source directions, inasmuch as the z element can be uniquely determined from these elements. The two points enclosed by circles in Fig. 9c represent the estimated value of the first and the last frames, and all the points in the figure are connected by a dotted line in the order of the frame number.

Next, we evaluate the estimated surface normals and reflectances. The estimated surface normal is extracted in three regions, as shown in Fig. 10a. Figure 10b shows the distribution of x and y elements for all the extracted surface normals. The averages and the standard deviations of the x and the y elements in each region are region 1 (−0.273, 0.737), (0.03, 0.02); region 2 (−0.565, −0.158), (0.04, 0.07); and region 3 (0.805, −0.182), (0.03, 0.04). The angle between the average surface normal in region 1 and that in region 2 is 57.4° (actual angle: 60°). In the same way, the angle between the average surface normal in region 2 and that in region 3 is 88.2° (actual: 90), and the angle between that in region 3 and that in region 1 is 90.3° (actual: 90°). As a result, these estimated surface normals are fairly correct.

The estimated surface reflectance is extracted in three regions, as shown in Fig. 11, and is averaged in each region. These averaged relative values are region 4, 0.930; region 5, 0.481; and region 6, 0.910. To evaluate these estimated values, we measure the luminance in each region under constant illumination. Because the measured values are region 4, 2.2×10^2 cd/m^2; region 5, 9.5×10 cd/m^2; and region 6, 2.3×10^2 cd/m^2, the estimated surface reflectances are roughly correct.

Finally, after substituting the estimated values of surface normal, surface reflectance, light-source direction, and light-source intensity into Eq. (8), we obtain the reconstructed intensity images as shown in Fig. 12. The frame numbers in Fig. 12 correspond to the same frame numbers in Fig. 7, and the image intensity noise of the original intensity images is removed.

C. Out-of-Laboratory Experiment

In this subsection an out-of-laboratory experiment on 100 intensity images of the architectural model of the ATR building (outer dimensions of the model: 30 cm × 30 cm × 6 cm, support table: 150 cm × 90 cm × 100 cm) is described. The scenes of the ATR model are photographed with a fixed hand-held camera in a lobby, when a light source is arbitrarily moved by a human for approximately 1 min in the evening. Of course, if the ambient illumination is subtracted from the derived scenes, daytime measurement is also possible. The measured scene data recorded on magnetic tape are transformed to digital intensity image data through a videotape recorder, a time-based corrector, and a real-time disk device. We subsample the derived intensity image data with a 1:2 ratio and slightly blur all the subsampled intensity images with a 3 × 3 Gaussian low-pass filter whose standard deviation is 0.85 pixel. The smoothed intensity images are 256 × 240 pixel images with an 8-bit resolution. Sixteen images (frame 50 to frame 65) are shown in Fig. 13.

We set the threshold value for classification into shadowed data and illuminated data at a tenth of the maximum intensity of all the intensity image data. The constraint of surface reflectance (C1) is applied to roughly 700 pixels at the foreside corner of the ATR model, and the relationship of the three surface normals at the corner is used for the choice of the solution in the right-handed coordinate system. The estimated surface normals are manually transformed to the viewer-oriented coordinate system. Figure 14 shows the estimated surface normal and surface reflectance. The needle image produced by the estimated surface normal is subsampled with a 1:2 ratio for clarity. Furthermore, using the estimated surface normal and reflectance, we synthesize four night scenes in which the ATR model is illuminated by a few virtual point light sources that are relatively near the ATR model (see Fig. 15).

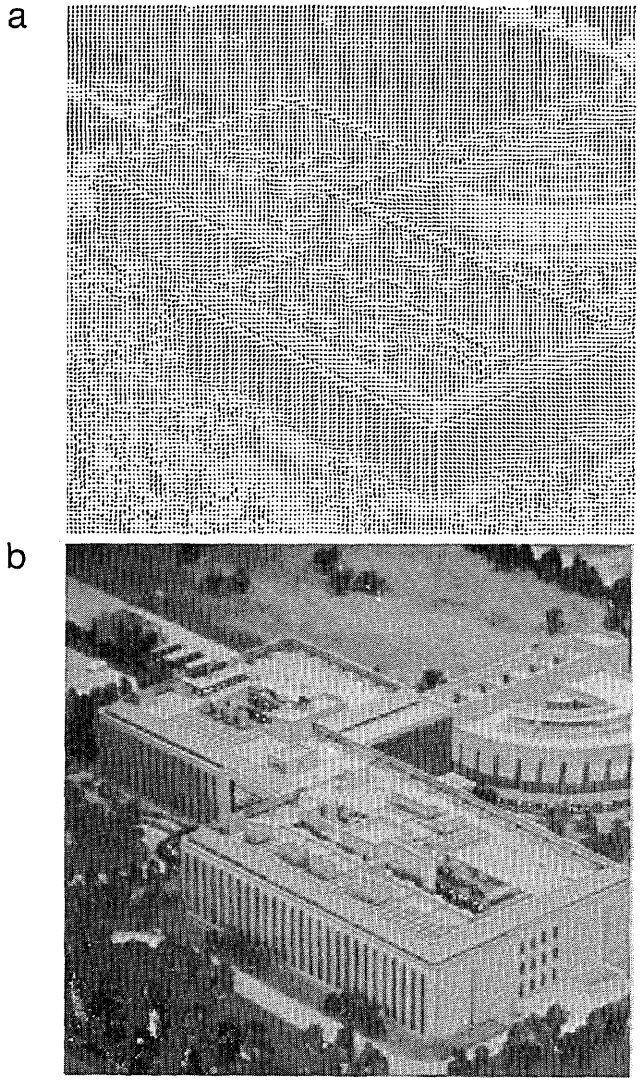

Fig. 14. Estimated results: a, needle image; b, reflectance image.

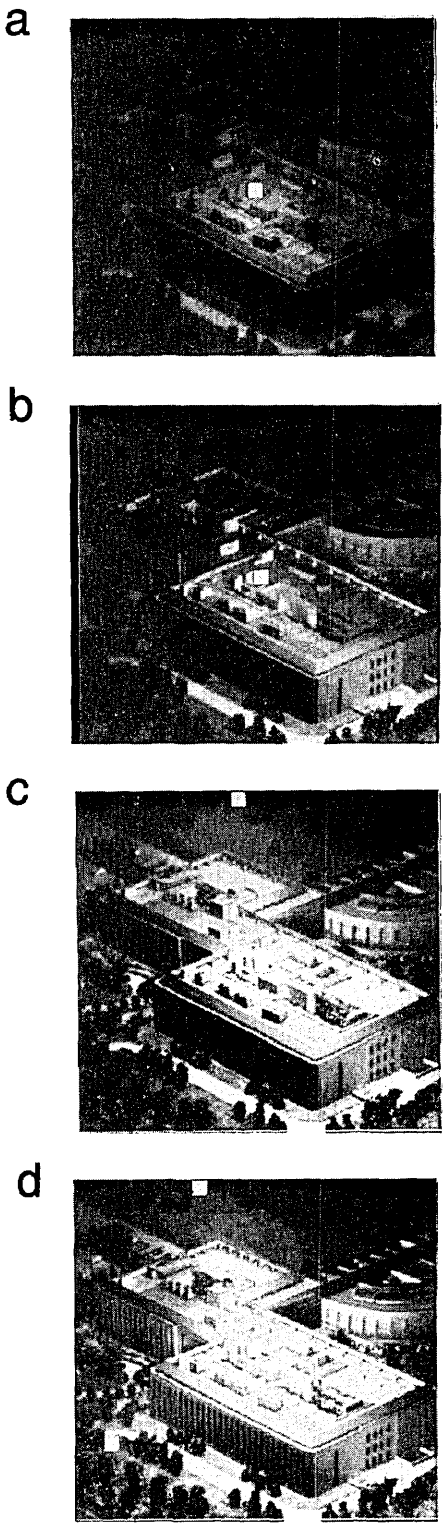

Fig. 15. Synthesized night scenes of the ATR model illuminated by a few virtual point light sources that are relatively near the model: a, with one point light source; b, with two point light sources; c, with three point light sources; d, with four point light sources. The small white square denotes the location of a point light source.

6. CONCLUSION

This paper describes a new photometric-stereo method for estimation of the surface normal and the surface reflectance of objects without *a priori* knowledge of the light-source direction or the light-source intensity. Assuming only that the object's surface is Lambertian, it is shown that the image data matrix I, which can be written as the product of surface matrix S and light-source matrix L, is of rank 3. Based on this observation, singular-value decomposition and useful constraints are used for the factorization of the image data matrix. This method does not rely on any smoothness assumption for the surface normal, the surface reflectance, the light-source direction, or the light-source intensity. In addition, as the number of intensity images increases, the estimated errors become smaller even if the intensity images are obtained in a noisy environment.

Shadow regions in intensity images are frequent enough to render a shape computation method impossible. I describe how to deal with shadow regions in the factorization of the image data matrix I. In the image data matrix, one selects an initial submatrix that has no shadowed data. One can then estimate the surface normal and reflectance in shadow regions by growing a partial solution obtained from the initial submatrix.

The useful constraints employed in this method require qualitative *a priori* knowledge of surface reflectance or light-source intensity. It is not difficult, however, to find regions of constant reflectance in images or to find image frames of constant light-source intensity among intensity images when measuring objects commonly seen in daily life. Because the proposed method can estimate the surface normal and reflectance from intensity images that are obtained with a fixed camera under a light source of arbitrary motion in a noisy environment, this method makes it possible for us to measure object shapes outside the laboratory.

ACKNOWLEDGMENTS

Portions of this research were performed at ATR Auditory and Visual Perception Research Laboratories. Sinjiro Kawato and Masahiko Shizawa provided not only many valuable comments but also experimental support. Yoichi Sato, Fredric Solomon, and Katsushi Ikeuchi of Carnegie Mellon University provided many useful comments, which have improved the readability of this paper. Toshinori Yoshioka of CSK Corporation helped us develop the software.

*Present address, Computer Science Laboratory, Fundamental Research Laboratories, Osaka Gas Company, Ltd., 6-19-9, Torishima Konohana-ku, Osaka 554, Japan.

REFERENCES AND NOTES

1. R. J. Woodham, "Photometric method for determining surface orientation from multiple images," Opt. Eng. **19,** 139–144 (1980).
2. K. Ikeuchi, "Determining surface orientation of specular surfaces by using the photometric stereo method," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-3,** 661–669 (1981).
3. S. K. Nayer, K. Ikeuchi, and T. Kanade, "Determining shape and reflectance of hybrid surfaces by photometric sampling," IEEE Trans. Robotics Autom. **6,** 418–431 (1990).
4. L. B. Wolff, "Surface curvature and contour from photometric stereo," in *Proceedings of the Defense Advanced Research Projects Agency Image Understanding Workshop* (Morgan Kaufmann, Los Altos, Calif., 1987), pp. 821–824.

5. B. Kim and P. Burger, "Depth and shape from shading using the photometric stereo method," Comput. Vision Graph. Image Proc. **54,** 416–427 (1991).
6. J. J. Clark, "Active photometric stereo," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Institute of Electrical and Electronics Engineers, New York, 1992), pp. 29–34.
7. L. Reimer, R. Bongeler, and V. Desai, "Shape from shading using multiple detector signals in scanning electron microscopy," Scanning Microsc. **1,** 963–975 (1987).
8. B. K. P. Horn and K. Ikeuchi, "The mechanical manipulation of randomly oriented parts," Sci. Am. **251,** 100–111 (1984).
9. K. Ikeuchi, "Determining a depth map using a dual photometric stereo system," Int. J. Robotics Res. **6,** 15–31 (1987).
10. C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method," Int. J. Comput. Vision **9,** 137–154 (1992).
11. Y. Sato and K. Ikeuchi, "Temporal-color space analysis of reflection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Institute of Electrical and Electronics Engineers, New York, 1993), pp. 570–576.
12. B. V. Funt and M. S. Drew, "Color space analysis of mutual illumination," IEEE Trans. Pattern Anal. Mach. Intell. **15,** 1319–1326 (1993).
13. B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," J. Opt. Soc. Am. A **4,** 629–642 (1987).
14. In experiments, I confirmed that the ratio between the largest and the third largest singular values of the matrix, which is constructed from known light-source vectors, is greater than 0.05.
15. However, initially estimated q pseudo surface vectors and g pseudo light source vectors are not affected by errors caused by shadowed regions. Consequently, once the p surface vectors and the f light-source vectors are derived, it may be necessary to refine the results with a method of steepest descent.