TR－H－024

0045

Shape, surface reflectance, light intensity and light direction from shading images

Hideki HAYAKAWA

1993. 9. 1

ATR人間情報通信研究所

〒619-02 京都府相楽郡精華町光台2-2 ☎07749-5-1011

ATR Human Information Processing Research Laboratories

2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

Telephone: +81-7749-5-1011
Facsimile: +81-7749-5-1008

Abstract

We propose a method for estimating the surface normal and surface reflectance of objects without a priori knowledge of the light intensities and light directions. First, we construct a $p \times f$ image data matrix I from p pixel image intensity data through f frames by moving a light source arbitrarily. Using the Lambertian assumption, the image data matrix I can be written as the product of two matrices S and L, S representing the surface normal and surface reflectance, and L representing the light intensities and light directions. Based on this formulation, we show that the image data matrix I is of rank 3. Based on the observation, we use a singular value decomposition technique and useful constraints to factorize the image data matrix. Our method can also treat both self-shadow and cast-shadow without assumptions. We demonstrate the effectiveness of our method through performance analysis, laboratory experiments on Lambertian reflectance objects and hybrid reflectance objects, and an out-of-laboratory experiment on a miniature model in a lobby.

1. INTRODUCTION

Photometric stereo is a method that uses several shading images obtained under different lighting conditions.[1] The method can be used for specular surfaces provided that 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 close to the surface and the camera.[5,6] The photometric stereo technique has been applied practically to scanning electron microscopic (SEM) images,[7] and the data provided by photometric stereo has been used for some industrial inspection and part-alignment tasks.[8,9] This shows that photometric stereo is a good candidate for propelling the commercialization of machine vision techniques. However, it is impossible for the previous methods to estimate the surface normal and surface reflectance without a priori knowledge of the light directions or arrangements.

In this paper, we propose a new photometric stereo method for estimating the surface normal and surface reflectance of objects without a priori knowledge of the light intensities and light directions. In our method, assuming only that the object's surface is Lambertian, the surface normal, surface reflectance, light intensities and light directions can be solved simultaneously. For example, we can estimate these four parameters from shading images that are obtained under a light source arbitrarily moved by a human arm. Our method does not rely on any smoothness assumption for these parameters; thus, they are very easy to treat even if they vary significantly. Furthermore, as the number of shading images increases, the surface normal, surface reflectance, light intensity and light direction errors become smaller even if the shading images are taken in a noisy environment.

The foundation of our method is similar to that of a factorization method for shape and motion estimation from image streams.[10] In our method, we construct a $p \times f$ image data matrix I from p pixel image intensity data through f frames by moving a light source

arbitrarily. Using the Lambertian assumption, the image data matrix I can be written as the product of two matrices S and L, S representing the surface characteristics (surface normal and reflectance) and L representing the light characteristics (light intensity and direction). Based on this formulation, we show that the image data matrix I is of rank 3. Based on the observation, we use a singular value decomposition technique and one of the two following constraints to factorize the image data matrix. One is the constraint of surface reflectance which can be used when there are at least 6 pixels in the images, in which the surface reflectance is constant or the ratio of surface reflectance is known. The other is the constraint of light intensity which can be used when there are at least 6 frames among the images, in which the light intensity is constant or the ratio of light intensity is known.

Our method also describes how to deal with shadow regions. In an intensity image, there are two types of shadows: self-shadow and cast-shadow. Previous methods could not deal with either shadow type without certain assumptions. Our method, however, can easily treat both types without assumptions, because many image frames are taken under the light source at different directions. First, we select an initial submatrix in the image data matrix having no shadowed data. The surface normal and reflectance in shadow regions can then be estimated by growing a partial solution obtained from the estimation result of the initial submatrix with an iterative procedure.

To analyze the performance of our method, we evaluate the surface normal and reflectance errors versus image noise, light rotation angle and light intensity fluctuation. Then, we demonstrate the effectiveness of our method with a series of experiments on Lambertian reflectance objects and hybrid reflectance objects. Furthermore, our method is applied to scenes of the "ATR miniature model" in a lobby that are taken with a hand-held camera.

In Section 2, the photometric stereo method which is the computational basis of our method is described. A factorization of image data matrix and a method to deal with shadow regions are proposed in Section 3 and Section 4, respectively. In Section 5, the

performance analysis of our method by computer simulation is described, and the experimental results are demonstrated.

2. PHOTOMETRIC STEREO REVISITED

Photometric stereo can estimate the surface normal and surface reflectance of objects from three or more shading images; the images are derived with a moving light source and a fixed camera and object.[1]

Figure 1 shows a geometric reflectance model for image generation. The fraction of light reflected by an object surface in a given direction depends upon the optical properties of the surface material, the surface microstructure and the spatial and 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 vectors: surface normal $\boldsymbol{n} = (n_x, n_y, n_z)^T$, light direction $\boldsymbol{m} = (m_x, m_y, m_z)^T$, and viewer direction $\boldsymbol{v} = (v_x, v_y, v_z)^T$, defined in Fig. 1. Using the reflectance function ϕ, the basic equation describing the image generation process can be written as:

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

where i denotes the image intensity and t denotes the light 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 direction can be considered to be constant over the image plane. In the case that these directions and the light intensities are known and that three light 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:

$$i(x,y) = r(x,y)t\big(\boldsymbol{n}(x,y) \cdot \boldsymbol{m}\big), \tag{2}$$

for $\boldsymbol{n}(x,y) \cdot \boldsymbol{m} \geq 0$. 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 the

three light intensities are known, three normalized image intensities $i'(x, y)$ can be defined by using each light intensity.

$$i'_k(x,y) = r(x,y)\big(\boldsymbol{n}(x,y)\cdot \boldsymbol{m}_k\big) \qquad k = 1,2,3. \tag{3}$$

If we construct the matrix M using these light direction vectors as columns, Eq. (3) is rewritten as follows:

$$I'(x,y) = r(x,y)\big(\boldsymbol{n}(x,y)\big)^T M, \tag{4}$$

where

$$I'(x,y) = \begin{bmatrix} i'_1(x,y) & i'_2(x,y) & i'_3(x,y) \end{bmatrix},$$

$$M = \begin{bmatrix} \boldsymbol{m}_1 & \boldsymbol{m}_2 & \boldsymbol{m}_3 \end{bmatrix}.$$

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

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

and

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

The photometric stereo method had been improved and extended as described in Section 1; at the time, however, applications using this method had required a priori knowledge of the light directions or arrangements. This method had also required that the surface point not be in shadows for any light source.

3. FACTORIZATION OF IMAGE DATA MATRIX

A. Image Data Matrix Formulation

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

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

The matrix I is called the *image data matrix*. Assuming a Lambertian shading model with a single distant point light source in the measurement, the image data matrix I can be expressed in a matrix form, like 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;

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

is the surface normal matrix;

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

is the light direction matrix; and

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

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

Based on these equations, the *surface matrix S* and *light matrix L* are defined as follows:

$$S = \begin{bmatrix} s_1 & \cdots & s_p \end{bmatrix}^T = \begin{bmatrix} s_{1x} & s_{1y} & s_{1z} \\ \vdots & \vdots & \vdots \\ s_{px} & s_{py} & s_{pz} \end{bmatrix} = RN \quad ,$$

$$L = \begin{bmatrix} l_1 & \cdots & l_f \end{bmatrix} = \begin{bmatrix} l_{x1} & \cdots & l_{xf} \\ l_{y1} & \cdots & l_{yf} \\ l_{z1} & \cdots & l_{zf} \end{bmatrix} = MT \quad . \tag{13}$$

The magnitude of *surface vector s* in surface matrix S represents the surface reflectance. The magnitude of *light vector l* in light matrix L represents the light intensity. By using Eq. (13), the other description of Eq. (8) is obtained as follows:

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

Assuming that three surface normals do not lie in a plane at least, the surface matrix S is of rank 3. In the same way, assuming that three light directions do not lie in a plane at least, the light 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

In this Subsection, the notion of approximate rank is introduced, using the concept of singular value decomposition.[11] 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 a $f \times f$ matrix V,

$$I = U \Sigma V \quad . \tag{15}$$

where $U^T U = V^T V = V V^T = E$, 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 transpose 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 *singular value decomposition* (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, these matrices can be partitioned as follows:

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

$$\Sigma = \underbrace{\begin{bmatrix} \Sigma' & 0 \\ 0 & \Sigma'' \end{bmatrix}}_{3 \quad f-3} \begin{matrix} \} 3 \\ \} f-3 \end{matrix} \,, \tag{16}$$

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

Considering that the image data matrix I has image noise, let I^* be the ideal image data matrix, that is, the matrix we would obtain in the absence of noise. In the case that 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 depends also on the surface normal at p pixel and on the light direction through f frames. The singular value decomposition yields sufficient information to judge the approximation: the requirement is that the ratio between the third and fourth largest singular values of I be sufficiently large.

By using Eq. (17), if we define the *pseudo surface matrix* $\hat{S}$ and *pseudo light matrix* $\hat{L}$ as follows:

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

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

we can write

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

Here, two pairs of pseudo solutions in Eq. (18) correspond to the solutions in the coordinates of the right handed system and left handed system. However, we can not judge which of the two pairs corresponds to the solution in the coordinate of the right handed system without any knowledge.

By using the relationship of an arbitrary triplet of surface normals in an image, which do not lie in a plane, we can choose the solution in the coordinate of the right handed system from the two solutions in Eq. (18). 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 Eq. (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 the grade that can determine this order uniquely. The pseudo surface matrix $\hat{S}$ that corresponds to the 3×3 matrix whose determinant's sign is positive is chosen. And the pair pseudo light matrix $\hat{L}$ to the chosen pseudo surface matrix $\hat{S}$ is also selected. The method of selection depends on the fact that all surface normal in the image have a positive z element in the viewer-oriented coordinate system. Note that the solution in the coordinate of the right handed system can be chosen from the two solutions in Eq. (18) by using the relationship of an arbitrary triplet of light 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 matrix L, respectively. 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, since $\hat{S}$ and $\hat{L}$ are different from S and L in general, we must find the matrix A such that

$$S = \hat{S}A, \qquad L = A^{-1}\hat{L}. \tag{20}$$

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

[C1] We can find 6 pixels at least, in which the surface reflectance is constant or the ratio of surface reflectance is known.

[C2] We can find 6 frames at least, in which the light intensity is constant or the ratio of light intensity is known.

Let us explain a way of finding the matrix A by using constraint [C1]. From the pseudo surface matrix $\hat{S}$, we extract $p'(\geq 6)$ pseudo surface vectors $\hat{s}$, of which the surface reflectance is constant or the ratio of surface reflectance is known. Assuming that the three vectors among them do not lie in a plane, the following system of p' equations can be solved in the case that the surface reflectance is constant.

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

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

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

This is a linear problem, so the solution is straightforward. Once B is determined, a matrix A such that $B = AA^T$ can be found by taking the singular value decomposition of B. The singular value decomposition of a symmetric matrix is itself symmetric: $B = W\Pi W^T$, where Π is diagonal and W is orthonormal, and we let $A = W[\Pi]^{1/2}$. In the

decomposition, we do not use the other sign $W\left(-[\Pi]^{1/2}\right)$ to keep the relationship of the coordinate in the right handed system. The calculation for constraint [C2] is identical.

D. Outline of the Algorithm

Based on the above development, we have an algorithm for the factorization of the image data matrix I into the surface matrix S and the light matrix L.

step1 Compute the singular value decomposition of the image data matrix I as

$$I = U\Sigma V.$$

step2 Define

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

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

step3 Compute the matrix A in Eq. (20) by using constraint [C1] or [C2].

step4 Compute the surface matrix S and the light matrix L as

$$S = \hat{S}A,$$
$$L = A^{-1}\hat{L}.$$

The derived surface matrix S and light matrix L are represented in an arbitrary 3-dimensional coordinate system. If we know three surface normals or three light 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 solving an absolute orientation problem.[12]

4. SHADOWS

A prerequisite for using methods based on the photometric stereo method is that the surface point not be in shadows for any light source. In reality, however, the light source position to the object and camera is changed, causing frequent self-shadow and cast-shadow in the image data. Since this phenomenon is frequent enough to make a

shape computation method unrealistic, we propose how to deal 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 the 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 where 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 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 are removed in the order of large to small amount of shadowed data. While $q \leq g$, columns are removed in the order of large to small amount of shadowed data. Here, we consider that q and g are monotonically decreasing variables and that their initial values are p and f, respectively. The extracted submatrix $\tilde{I}$ is decomposed by the method in Subsection 3.2, and the q pseudo surface vectors and g pseudo light vectors can be derived.

Next, by using the derived q pseudo surface vectors and g pseudo light vectors, we estimate the unknown $p-q$ pseudo surface vectors and $f-g$ pseudo light vectors. Let us explain the way to solve an unknown pseudo surface vector $\hat{s}_i$ in the least-square-error sense [14]. Assuming that a known pseudo light vector $\hat{l}_j$ which generates illuminated data with the unknown pseudo surface vector $\hat{s}_i$ exists, the next equation is obtained by using Eq. (14).

$$i = \hat{s}_i \cdot \hat{l}_j \tag{23}$$

If we have three or more of such known pseudo light vectors, where three vectors do not lie in a plane at least[13], the unknown pseudo surface vector $\hat{s}_i$ can be obtained by solving

a linear problem. On the other hand, the way to obtain an unknown pseudo light vector by using Eq. (23) is identical to the above mentioned method.

When estimating the unknown p–q pseudo surface vectors and f–g pseudo light vectors, the order with which to solve these unknown pseudo vectors is the reverse order of that previously used to remove the rows (corresponding to unknown pseudo surface vectors) and columns (corresponding to unknown pseudo light vectors). Such order estimation calculates the values of more reliable pseudo vectors, those having more illuminated data and less shadowed data, early. Since we can use not only the known q pseudo surface vectors and g pseudo light vectors but also the calculated values of the more reliable pseudo vectors to estimate the values of less reliable pseudo vectors, the total value of the estimation error becomes small in comparison with that for random order estimation.

Finally, by using the derived p pseudo surface vectors and f pseudo light vectors, the p surface vectors and f light vectors can be calculated in the way described in Subsection 3.3.

5. EXPERIMENTS

A. Performance Analysis

To experimentally analyze the performance of our method, synthetic shading images are employed. The surface normals have a uniform distribution on Gaussian sphere; the surface are limited to visible surfaces from the camera. The surface reflectances have a uniform distribution from the value 0.1 to 1.0. We define *light source angle* as the range in which the light source can be moved, and the directions of the light source have a uniform distribution on Gaussian sphere; they are limited to within the light source angle. The subsequent shading images have 8-bit resolution, and since the light intensities are assumed to be a constant value of 1.0 in the shading image generation, the constraint of light intensity [C2] is used for subsequent factorizations. Furthermore, by using three light directions, which are represented in the viewer-oriented coordinate system, the

solution of the right handed system can be selected from the two solutions in Eq. (18) and the estimated surface normals and light directions can be transformed to the representation in the viewer-oriented coordinate system.

First, we evaluate the effect of image noise for different frame numbers. We generate 64 shading images when the light source angle is fixed to 90 degrees, and add Gaussian image noise to the images with various standard deviations. Each standard deviation is normalized by the maximum image intensity of the shading images. Figure 3 shows surface normal and reflectance errors versus image noise. The surface normal error is defined as the 3-dimensional angle error on the estimated surface normal, averaged over all surface data. The surface reflectance error is defined as the absolute error on the estimated surface reflectance, averaged over all surface data. The diagrams show that the errors increase with more image noise and fewer frames of shading images.

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

Finally, we evaluate the effect of light intensity fluctuation when the constraint of light intensity [C2] is used. We generate 64 shading images when the light source angle is fixed to 90 degrees, and add light intensity fluctuations with Gaussian distributions. Each standard deviation of the fluctuations is normalized by the true light intensity value 1.0. Figure 5 shows surface normal and reflectance errors versus light intensity fluctuation for different frame numbers. When the constraint of light intensity [C2] is used with the light intensity fluctuation whose standard deviation is 10 percent of the light intensity, 64 or more shading images are needed.

B. Lambertian Reflectance Object

In this Subsection, we describe an experiment on 120 shading images of a sphere and milk carton. The shading images are 128×128 pixel images having 8-bit resolution, and 16 images (*frame 60* to *frame 75*) from among them are shown in Fig. 6. 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 shading images are derived when a light source is arbitrarily moved by a human arm for about one minute. The shading image acquisition can be carried out in real time, because we use an image measurement system with a real-time-disk device. Figure 7 shows a measurement scene in this experiment.

We set the threshold value for the classification into the shadowed data and the illuminated data at a tenth of the maximum intensity among all image data. Since the light intensity is not constant in the experiment, the constraint of surface reflectance [C1] is applied to about 1200 pixels at a part of the sphere. The surface normal of the sphere can be roughly calculated from the coordinate of the center and 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 directions, which are derived from the factorization of the image data matrix, to the representation in the viewer-oriented coordinate system. Here, since no choice has yet been made for the solution in the coordinate of the right handed system from the two solutions in Eq. (18), both solutions are transformed. After the transformation, the average angle errors in the sphere between the surface normal derived from the rough calculation and the estimated surface normal of the two solutions derived from the factorizations are 3.7 degrees and 43.1 degrees, respectively. We choose the solution with the smaller error, because the solution in the coordinate of the left handed system continues to have a large error in spite of the transformation to the representation in the viewer-oriented coordinate system. Figure 8 shows the estimated surface normal, surface reflectance, light directions and light intensities. The two points respectively enclosed by circles in Fig. 8(c) represent the estimated value of the first and last frames.

Next, we evaluate the estimated surface normals and reflectances. The estimated surface normal is extracted in three regions as shown in Fig. 9(a). Figure 9(b) shows the distribution of x and y elements for all extracted surface normals. The averages and standard deviations of the x and 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 degrees (actually 60 degrees). In the same way, the angle between the average surface normal in *region 2* and that in *region 3* is 88.2 degrees (actually 90 degrees), and the angle between that in *region 3* and that in *region 1* is 90.3 degrees (actually 90 degrees). As a result, these estimated surface normals are fairly correct.

The estimated surface reflectance is extracted in three regions as shown in Fig. 10 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. Since the measured values are *region 4* 2.2×10^{2} cd/m^2, *region 5* 9.5×10^{1} 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 directions and light intensities into Eq. (8), the reconstructed shading images are obtained as shown in Fig. 11. The frame numbers in Fig. 11 correspond to the same frame numbers in Fig. 6, and the image noise of the original shading images is excepted in all of the reconstructed images.

C. Hybrid Reflectance Object

In this Subsection, we describe an experiment on 90 shading images of two spheres and a rectangular solid. The shading images are 128×128 pixel images having 8-bit resolution, and 16 images (*frame 50* to *frame 65*) from among them are shown in Fig. 12. The left-side sphere is the Lambertian sphere used in Subsection 5.1. On the other

hand, the rectangular solid and the right-side sphere each have a hybrid reflectance surface. These shading images are derived by the same way as in Subsection 5.1.

We set the threshold value for the classification into the shadowed data and the illuminated data at a twentieth of the maximum intensity among all image data. The constraint of surface reflectance [C1] is applied to about 1200 pixels at a part of the Lambertian sphere, and the estimated surface normals and light directions are transformed to the representation in the viewer-oriented coordinate system by the same way as in Subsection 5.1. After the transformation, the average angle errors of the estimated surface normals of the two solutions in the region of the Lambertian sphere are 4.4 degrees and 46.4 degrees, respectively. We choose the solution with the smaller error. Figure 13 shows the estimated surface normal, surface reflectance, light directions and light intensities. The two points respectively enclosed by circles in Fig. 13(c) represent the estimated values of the first and last frames.

Next, we evaluate the estimated surface normals and reflectances. The estimated surface normal is extracted in three regions as shown in Fig. 14(a). Figure 14(b) shows the distribution of x and y elements for all extracted surface normals. The averages and standard deviations of the x and y elements in each region are *region 1* (0.056, 0.866), (0.02, 0.01); *region 2* (–0.737, –0.163), (0.03, 0.05); and *region 3* (0.516, –0.365), (0.04, 0.05). The angle between the average surface normal in *region 1* and that in *region 2* is 81.7 degrees (actually 90 degrees). In the same way, the angle between the average surface normal in *region 2* and that in *region 3* is 79.2 degrees (actually 90 degrees), and the angle between that in *region 3* and that in *region 1* is 84.4 degrees (actually 90 degrees). Furthermore, the surface normal of the hybrid surface sphere can be roughly calculated from the coordinate of the center and the radius of the sphere in the images. In the region of the hybrid surface sphere, the average angle error between the surface normal derived from the rough calculation and the estimated surface normal is 12.7 degrees. As a result, these estimated surface normals are almost correct even though the objects have a hybrid reflectance surface.

The estimated surface reflectance (Lambertian reflectance) is extracted in three regions as shown in Fig. 15 and is averaged in each region. These averaged relative values are *region 4* 0.530, *region 5* 0.492, *region 6* 0.239, and *region 7* 0.465. To evaluate these estimated values, we measure the luminance in each region under constant illumination, which does not generate specular reflection. Since the measured values are *region 4* 2.2×10^2 cd/m^2, *region 5* 2.2×10^2 cd/m^2, *region 6* 8.4×10^1 cd/m^2, and *region 7* 1.8×10^2 cd/m^2, the estimated Lambertian reflectances are roughly correct.

Finally, after substituting the estimated values of surface normal, surface reflectance, light directions and light intensities into Eq. (8), we calculate the reconstructed shading images shown in Fig. 16(a). The frame numbers in Fig. 16(a) correspond to the same frame numbers in Fig. 12. Furthermore, we calculate the absolute subtracted images between these reconstructed shading images and the original shading images as shown in Fig. 16(b). The absolute subtracted images are normalized by the maximum of the absolute subtracted value. The regions where the absolute subtracted values is high correspond to specular pixels in the shading images.

D. "ATR Miniature Model" in a Lobby

In this Subsection, we describe an experiment on 100 shading images of the "ATR miniature model" (outer dimensions of the miniature model 30cm×30cm×6cm, support table 150cm×90cm×100cm) in a lobby. The scenes of the "ATR miniature model" are taken with a fixed hand-held camera, when a light source is arbitrarily moved by a human arm for about one minute in an evening. Of course, if ambient illuminations are subtracted from the derived scenes, daytime measurement is also possible. The measured scene data on a magnetic tape is transformed to digital shading image data through a videotape recorder, a time-base-corrector and a real-time-disk device. We subsample the derived shading image data with a 1:2 ratio, and slightly blur all of the subsampled shading images with a 3×3 Gaussian low-pass filter whose standard deviation is 0.85

pixels. The smoothed shading images are 256×240 pixel images having 8-bit resolution, and 16 images (*frame 50* to *frame 65*) from among them are shown in Fig. 17.

We set the threshold value for the classification into the shadowed data and the illuminated data at a tenth of the maximum intensity among all shading image data. The constraint of surface reflectance [C1] is applied to about 700 pixels at the foreside corner of the miniature model, and the relationship of three surface normals at the corner is used for the choice of the solution in the coordinate of the right handed system. The estimated surface normals are manually transformed to the representation in the viewer-oriented coordinate system. Figure 18 shows the estimated surface normal and surface reflectance. Furthermore, using the estimated surface normal and reflectance, we synthesize four night scenes, in which the "ATR miniature model" is illuminated by a few virtual point light sources that are relatively close to the miniature model (Fig. 19).

6. CONCLUSION

This paper describes a new photometric stereo method for estimating the surface normal and surface reflectance of objects without a priori knowledge of the light intensities and light directions. 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 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. Our method does not rely on any smoothness assumption for surface normal, surface reflectance, light direction and light intensity. In addition, as the number of shading images increases, the estimated errors become smaller even if the shading images are taken in a noisy environment.

Shadow regions in shading images are frequent enough to make a shape computation method unrealistic. Therefore, how to deal with shadow regions in the factorization of the image data matrix I is also described. We select an initial submatrix in the image data matrix that has no shadowed data. The surface normal and reflectance in shadow regions

can then be estimated by growing a partial solution obtained from the estimation result of the initial submatrix with an iterative procedure.

A Lambertian assumption makes it possible to simplify the reflectance model and to develop our factorization method, but the assumption has a shortcoming in that specular reflection is not considered at all. However, our method can roughly estimate the surface normal and Lambertian reflectance of hybrid reflectance objects as shown in Subsection 5.3, though the examination of our method's limitation to hybrid reflectance objects remains for future study. Furthermore, the extraction of the specular pixel in shading images shown in Fig. 16(b) will contribute to the improvement of the surface normal and Lambertian reflectance estimation accuracies in our method.[14,15,16] Of course, if our proposed method is combined with the interreflections algorithm,[17] the accuracies will be improved in addition.

The useful constraints employed in our method require qualitative a priori knowledge of surface reflectance or light intensity. It is not difficult, however, to find regions of constant reflectance in images or to find the image frames of constant light intensity among shading images, when measuring common objects that can be generally seen in daily life. Since our proposed method can estimate the surface normal and reflectance from shading images that are obtained with a fixed camera under a light source of arbitrary motion in a noisy environment, the measurement of object shape out of the laboratory is possible.

ACKNOWLEDGEMENTS

Portions of this work 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. Toshinori Yoshioka of CSK helped us to develop the software.

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. Pattern Anal. Mach. Intell. **PAMI-6**, 418-431 (1990).
4. L. B. Wolff, "Surface curvature and contour from photometric stereo," in *Proceedings of the DARPA Image Understanding Workshop* (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 Graphics Image process.: Image Understanding **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 Microscopy **1**, 963-975 (1987).
8. B. K. P. Horn and K. Ikeuchi, "The mechanical manipulation of randomly oriented parts," Scientific American **251**, 100-111 (1984).
9. K. Ikeuchi, "Determining a depth map using a dual photometric stereo system," Internat. J. Robot. Res. **6**, 15-31 (1987).
10. C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method," Internat. J. Comput. Vision **9**, 137-154 (1992).

11. G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," in *Handbook for Automatic Computation* **2** (Springer Verlag, New York, 1971), pp.134-151.

12. B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *J. Opt. Soc. Am. A* **4**, 629-642 (1987).

13 In experiments, we confirmed that the ratio between the first and third largest singular values of the matrix, which is constructed from the known light vectors, is greater than 0.05.

14. G. Brelstaff and A. Blake, "Detecting specular reflections using Lambertian constraints," *Proceedings of the IEEE International Conference on Computer Vision* (Institute of Electrical and Electronics Engineers, New York, 1988), pp.297-302.

15. K. Ikeuchi and K. Sato, "Determining reflectance parameters using range and brightness images," *Proceedings of the IEEE International Conference on Computer Vision* (Institute of Electrical and Electronics Engineers, New York, 1990), pp.12-20.

16. F. Solomon and K. Ikeuchi, "Extracting the shape and roughness of specular lobe objects using four light photometric stereo," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Institute of Electrical and Electronics Engineers, New York, 1992), pp.466-471.

17. S. K. Nayer, K. Ikeuchi, and T. Kanade, "Shape from interreflections," *Proceedings of the IEEE International Conference on Computer Vision* (Institute of Electrical and Electronics Engineers, New York, 1990), pp.2-11.

FIGURE CAPTION

FIG. 1.

Geometric reflectance model for image generation in the viewer-oriented coordinate system. Here, $\boldsymbol{n}$, $\boldsymbol{m}$ and $\boldsymbol{v}$ denote the unit vectors of surface normal, light direction and viewer direction, respectively.

FIG. 2.

Relationship of a triplet of surface normals. $\boldsymbol{n}_\alpha$, $\boldsymbol{n}_\beta$, $\boldsymbol{n}_\gamma$ 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.

FIG. 3.

Estimation error versus image intensity noise for different frame numbers: (a) surface normal error and (b) surface reflectance error.

FIG. 4.

Estimation error versus light source angle for different standard deviations of image noise: (a) surface normal error and (b) surface reflectance error.

FIG. 5.

Estimation error versus light intensity fluctuation for different frame numbers when the constraint of light intensity is used: (a) surface normal error and (b) surface reflectance error.

FIG. 6.

Shading images of a sphere and milk carton that are derived when a light source is arbitrarily moved by a human arm.

FIG. 7.

Measurement scene of shading images that are taken with a CCD camera when a light source is arbitrarily moved by a human arm.

FIG. 8.

Estimated results: (a) surface normal; (b) surface reflectance; (c) light direction; and (d) light intensity.

FIG. 9.

Quantitative evaluation of estimated surface normal: (a) region of extracted surface normal and (b) distribution of extracted surface normal.

FIG. 10.

Region of extracted surface reflectance.

FIG. 11.

Reconstructed shading images. The frame numbers correspond to the same frame numbers in Fig. 6.

FIG. 12.

Shading images of two spheres and a rectangular solid that are derived when a light source is arbitrarily moved by a human arm.

FIG. 13.

Estimated results: (a) surface normal; (b) surface reflectance; (c) light direction; and (d) light intensity.

FIG. 14.

Quantitative evaluation of estimated surface normal: (a) region of extracted surface normal and (b) distribution of extracted surface normal.

FIG. 15.

Region of extracted surface reflectance.

FIG. 16.

(a) Reconstructed shading images. The frame numbers correspond to the same frame numbers in Fig. 12. (b) Absolute subtracted images between each reconstructed shading image and each input shading image.

FIG. 17.

Shading images of the "ATR miniature model" in a lobby that are derived when a light source is arbitrarily moved by a human arm.

FIG. 18.

Estimated results: (a) surface normal and (b) surface reflectance.

FIG. 19.

Synthesized night scenes of the "ATR miniature model" illuminated by a few virtual point light sources that are relatively close to the miniature model. The small white square represents the location of a point light source. (a) scene under one point light source; (b) scene under two point light sources; (c) scene under three point light sources; and (d) scene under four point light sources.

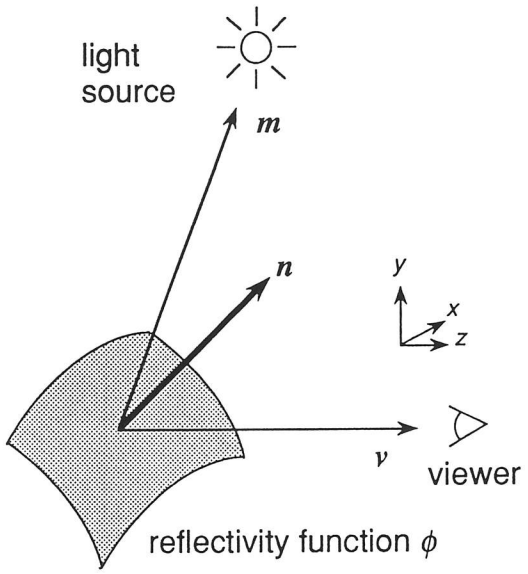


Fig. 1

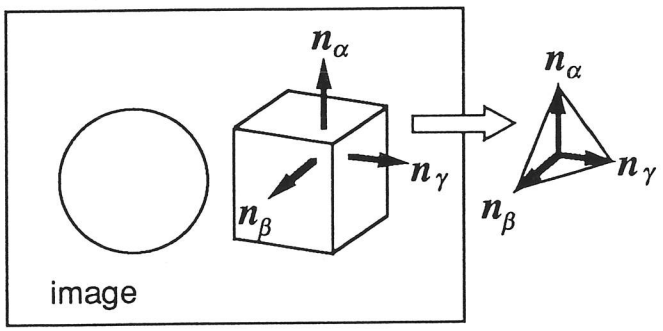


Fig. 2

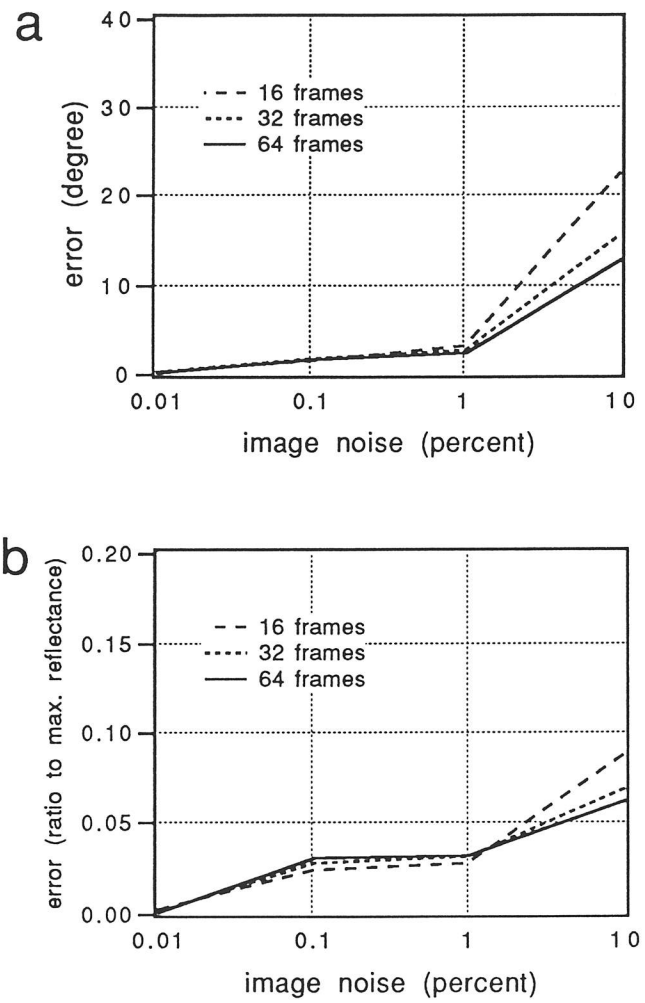

a
40
30
20
10
0
error (degree)
16 frames
32 frames
64 frames
0.01
0.1
1
10
image noise (percent)
b
0.20
0.15
0.10
0.05
0.00
error (ratio to max. reflectance)
16 frames
32 frames
64 frames
0.01
0.1
1
10
image noise (percent)

Fig. 3

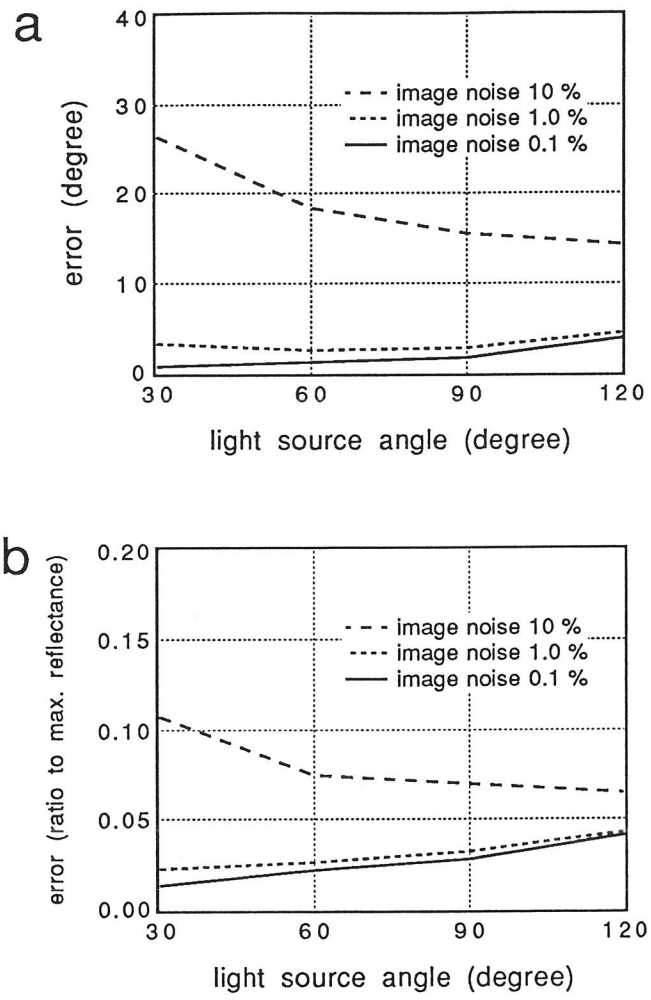

a
40
30
20
10
0
error (degree)
image noise 10 %
image noise 1.0 %
image noise 0.1 %
30
60
90
120
light source angle (degree)
b
0.20
0.15
0.10
0.05
0.00
error (ratio to max. reflectance)
image noise 10 %
image noise 1.0 %
image noise 0.1 %
30
60
90
120
light source angle (degree)

Fig. 4

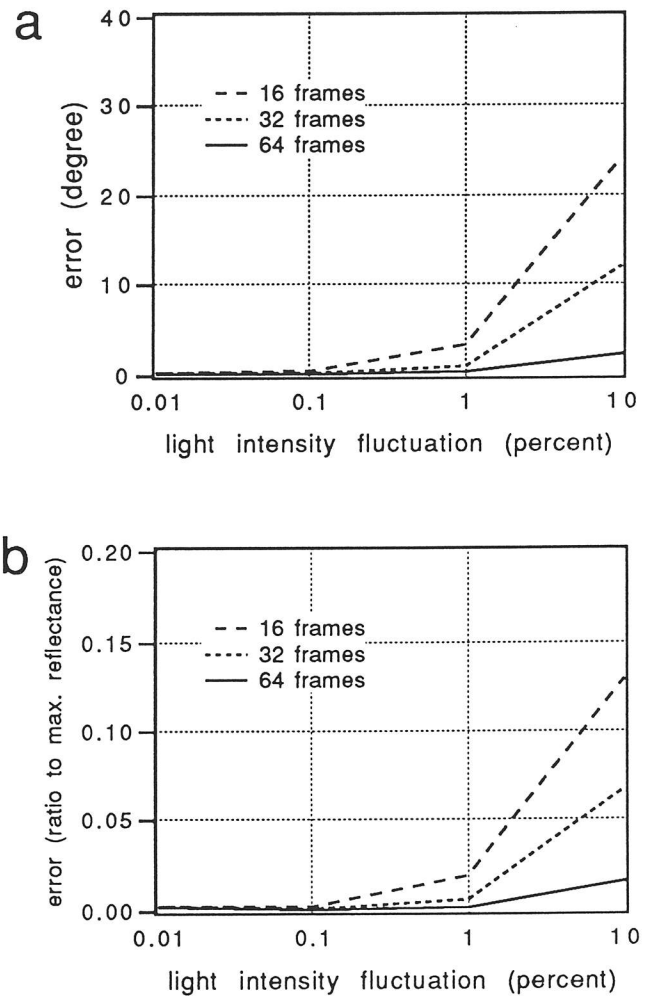

a
40
30
20
10
0
error (degree)
16 frames
32 frames
64 frames
0.01
0.1
1
10
light intensity fluctuation (percent)
b
0.20
0.15
0.10
0.05
0.00
error (ratio to max. reflectance)
16 frames
32 frames
64 frames
0.01
0.1
1
10
light intensity fluctuation (percent)

Fig. 5

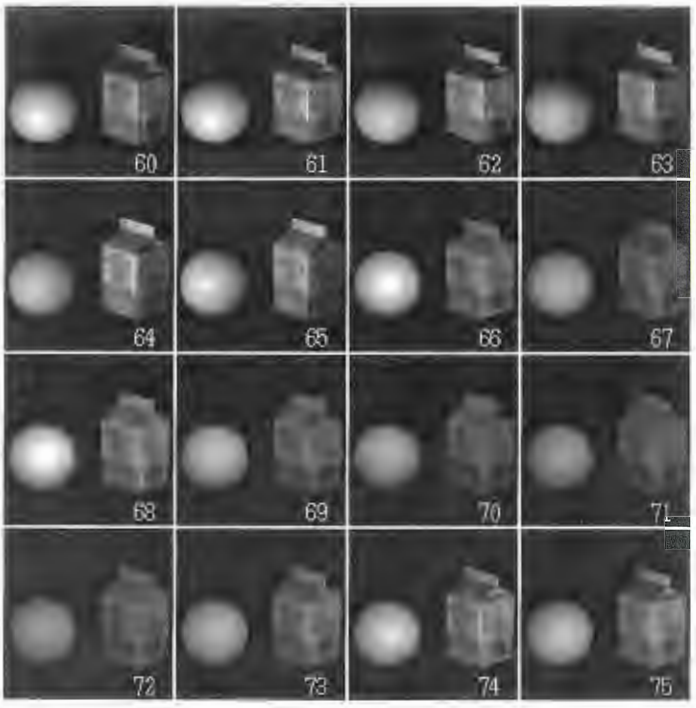


Fig. 6

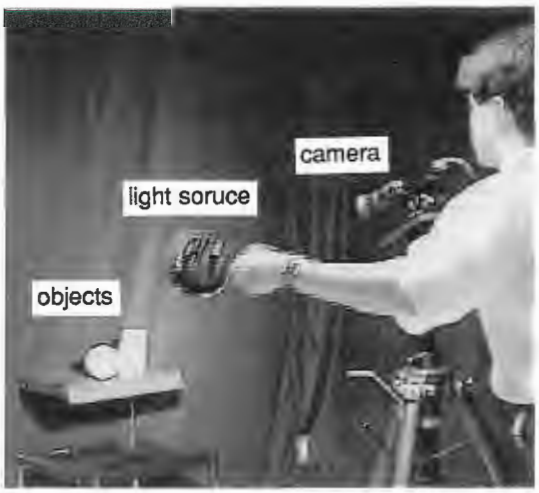


Fig. 7

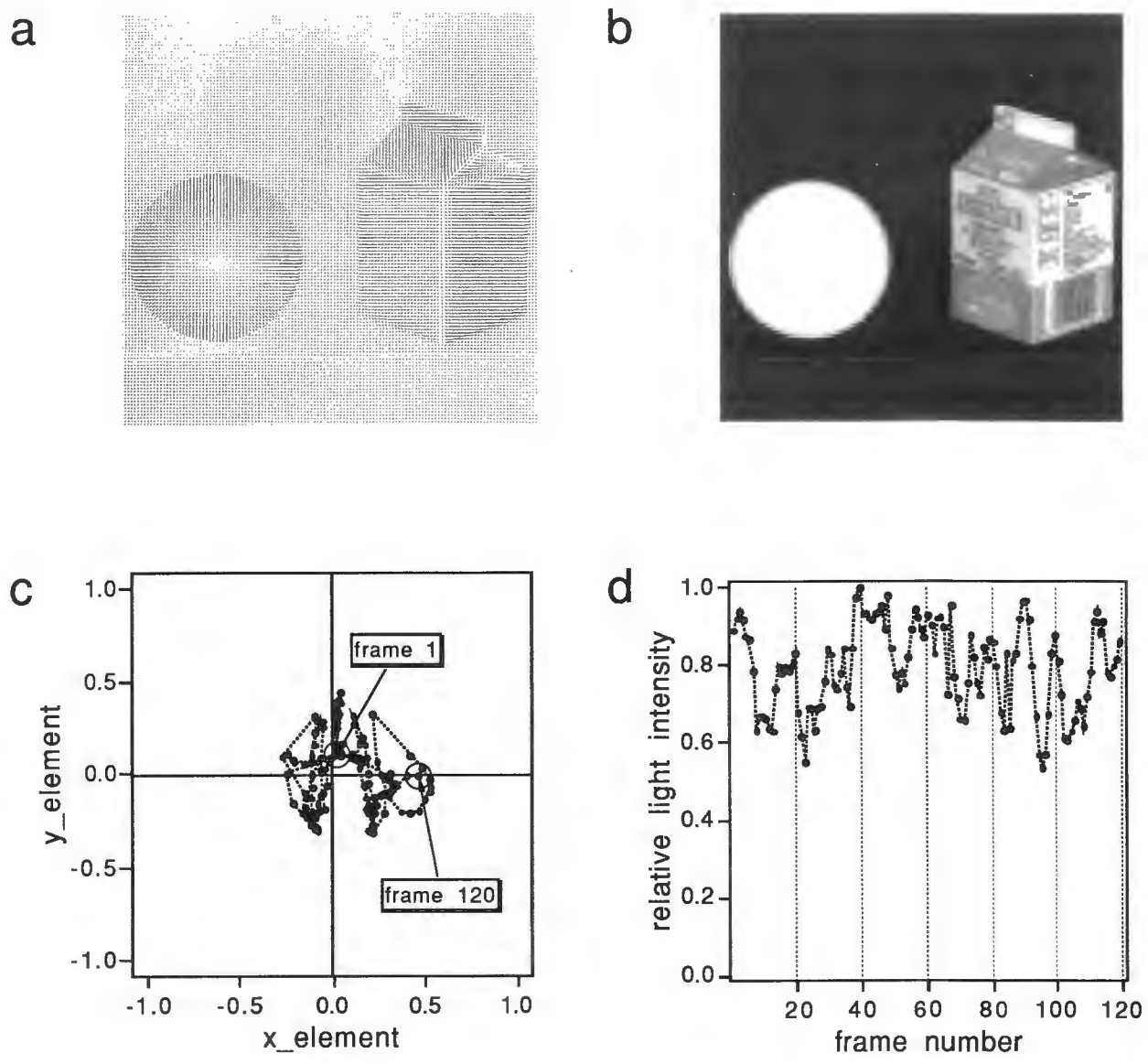

a
b
c
1.0
0.5
0.0
-0.5
-1.0
y_element
frame 1
frame 120
-1.0
-0.5
0.0
0.5
1.0
x_element
d
1.0
0.8
0.6
0.4
0.2
0.0
relative light intensity
20
40
60
80
100
120
frame number

Fig. 8

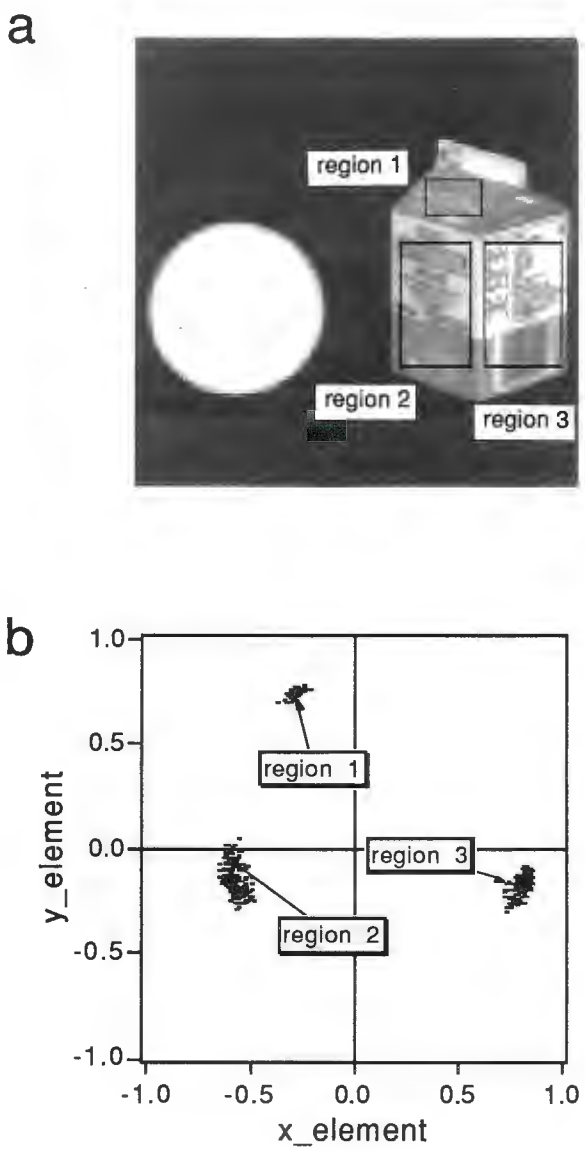

a
region 1
region 2
region 3
b
1.0
0.5
0.0
-0.5
-1.0
y_element
region 1
region 3
region 2
-1.0
-0.5
0.0
0.5
1.0
x_element

Fig. 9

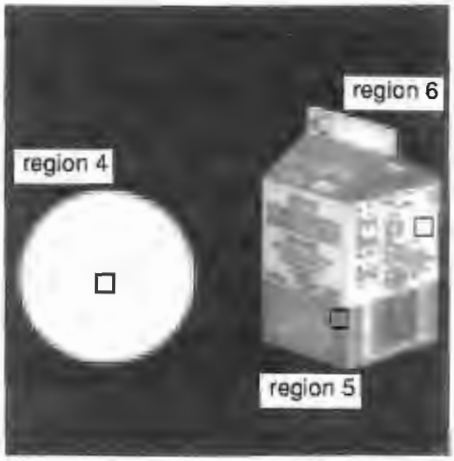


Fig. 10

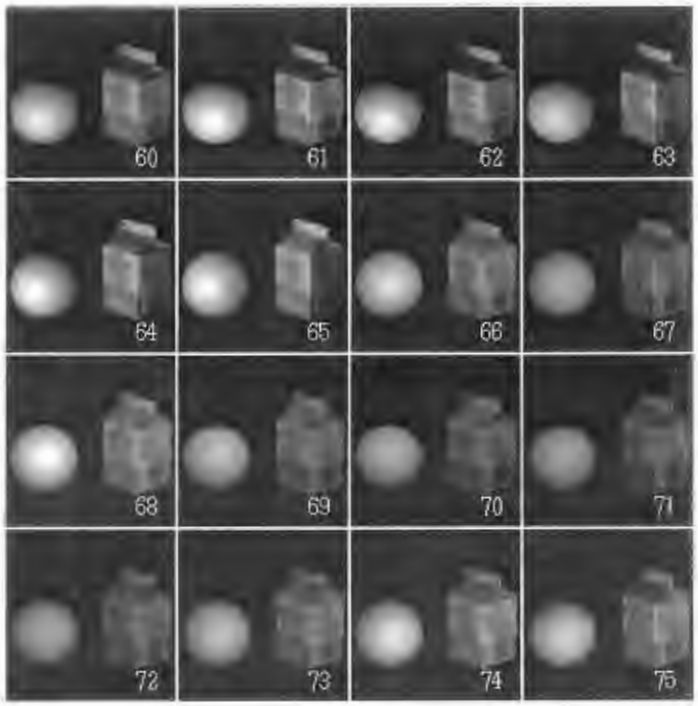


Fig. 11

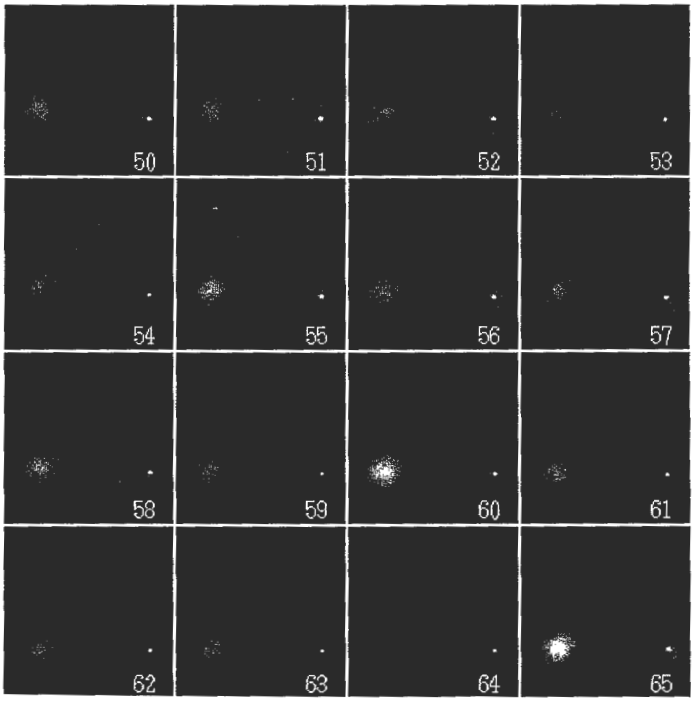


Fig. 12

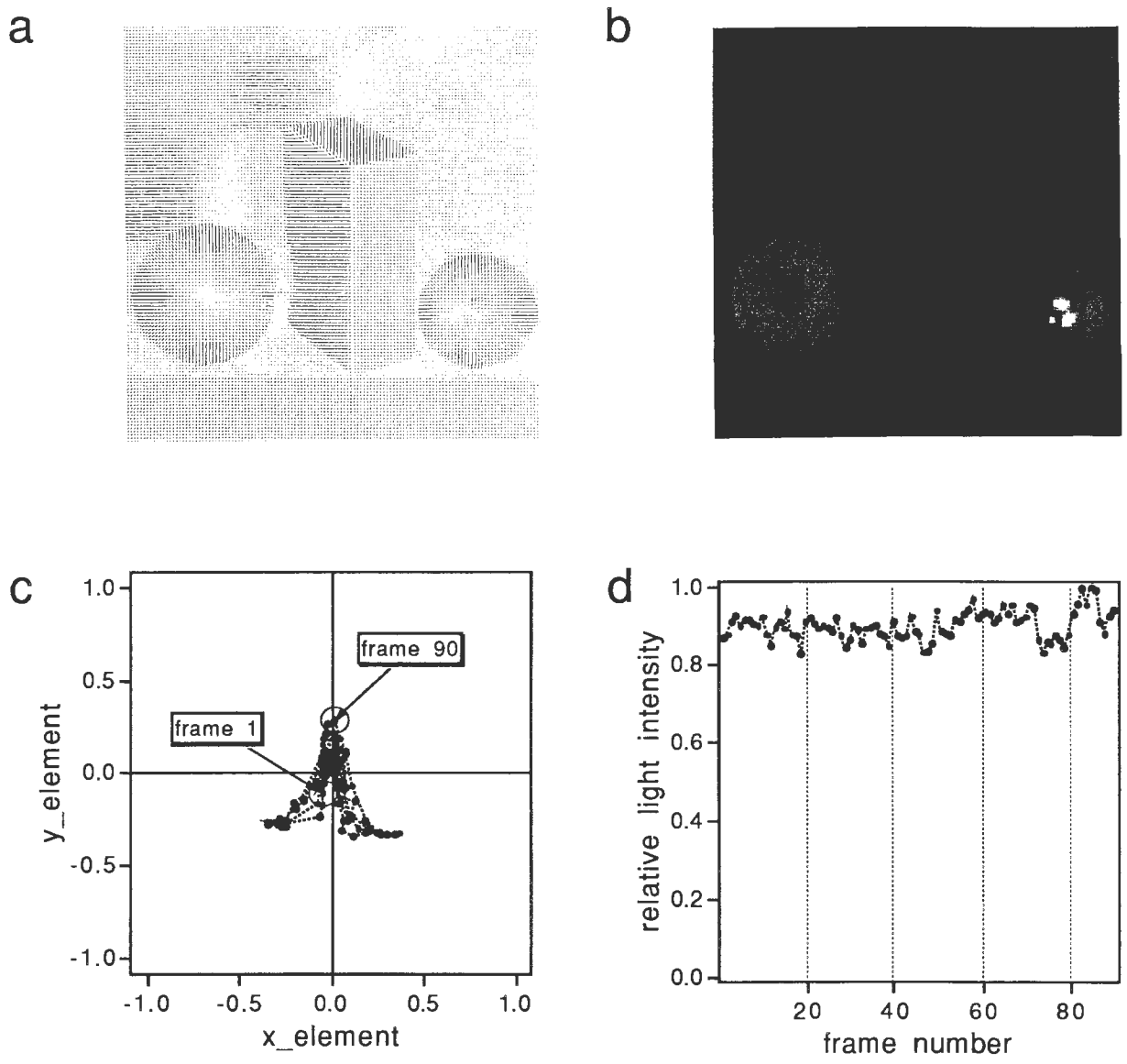

a
b
c
d
frame 90
frame 1
y_element
x_element
1.0
0.5
0.0
-0.5
-1.0
relative light intensity
0.8
0.6
0.4
0.2
20
40
60
80
frame number

Fig. 13

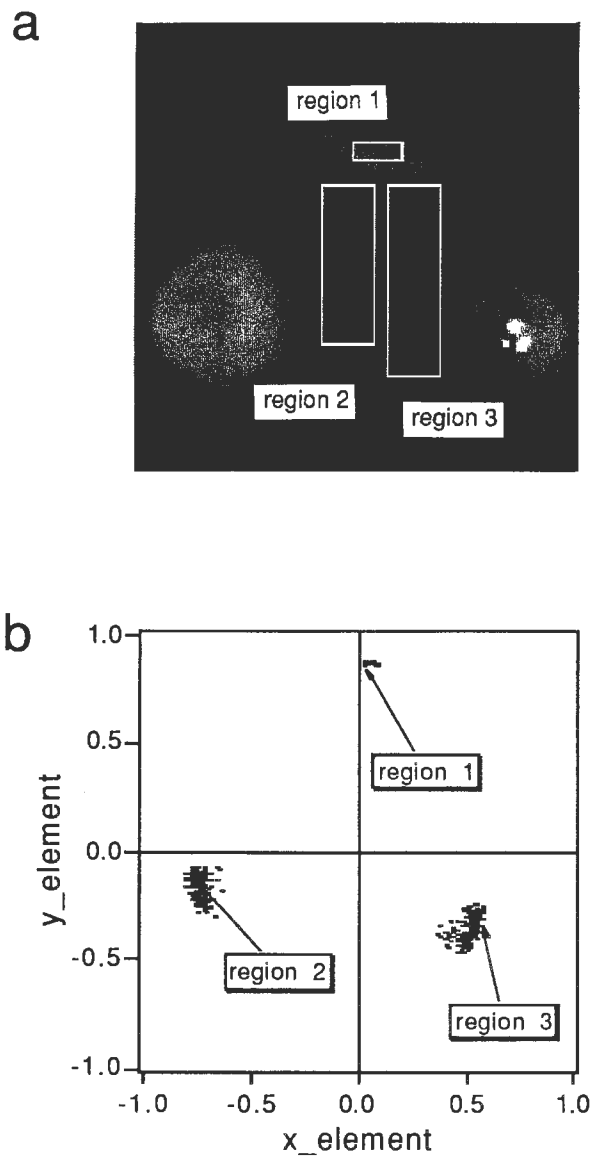

a
region 1
region 2
region 3
b
1.0
0.5
0.0
-0.5
-1.0
y_element
region 1
region 2
region 3
-1.0
-0.5
0.0
0.5
1.0
x_element

Fig. 14

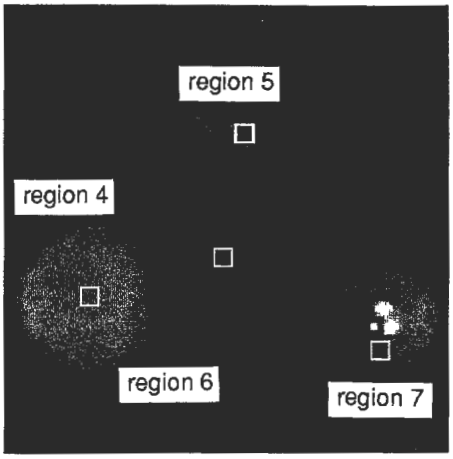
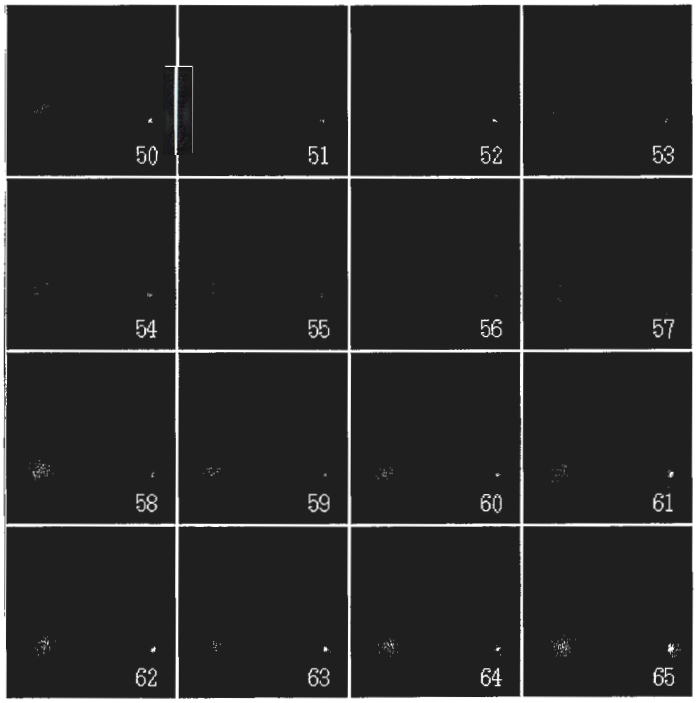


Fig. 15

a


50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

b

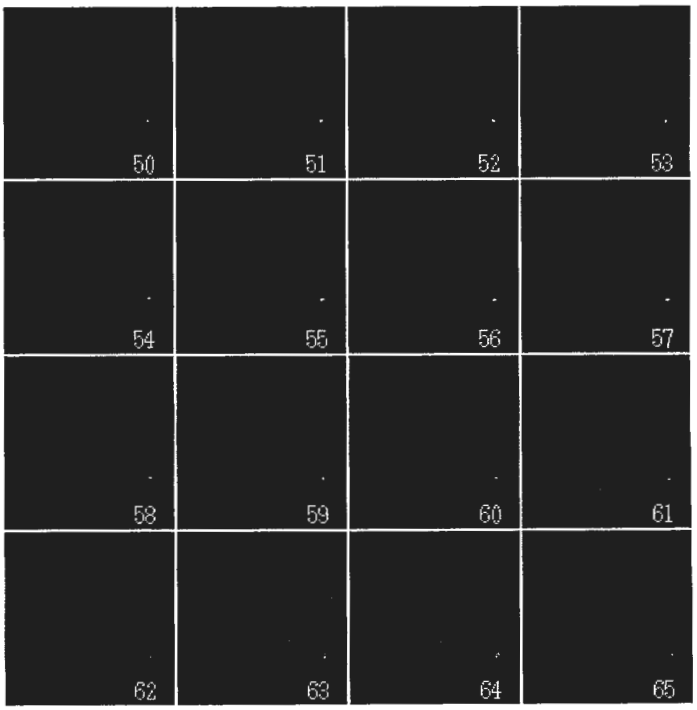

50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Fig. 16

Fig. 17

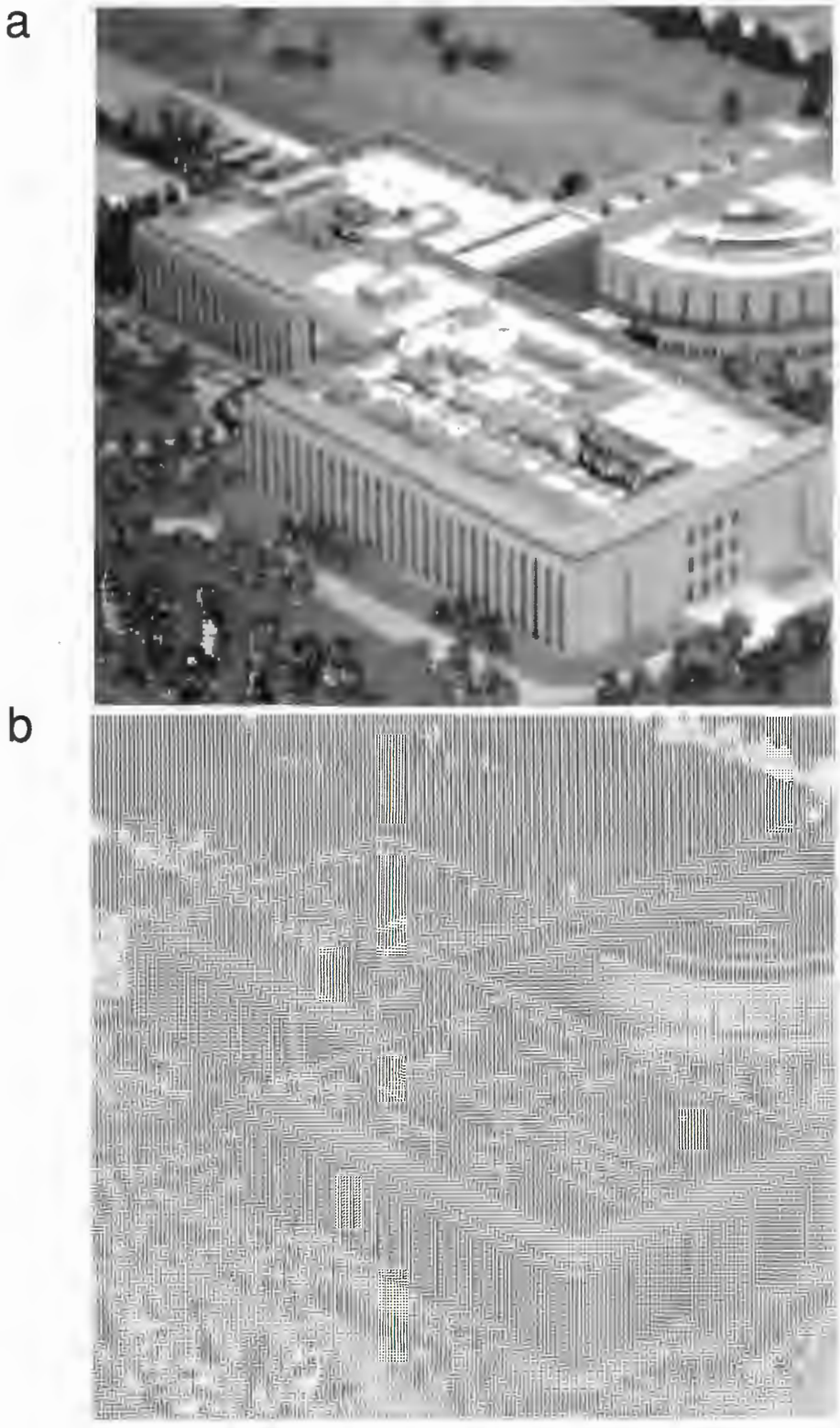


Fig. 18

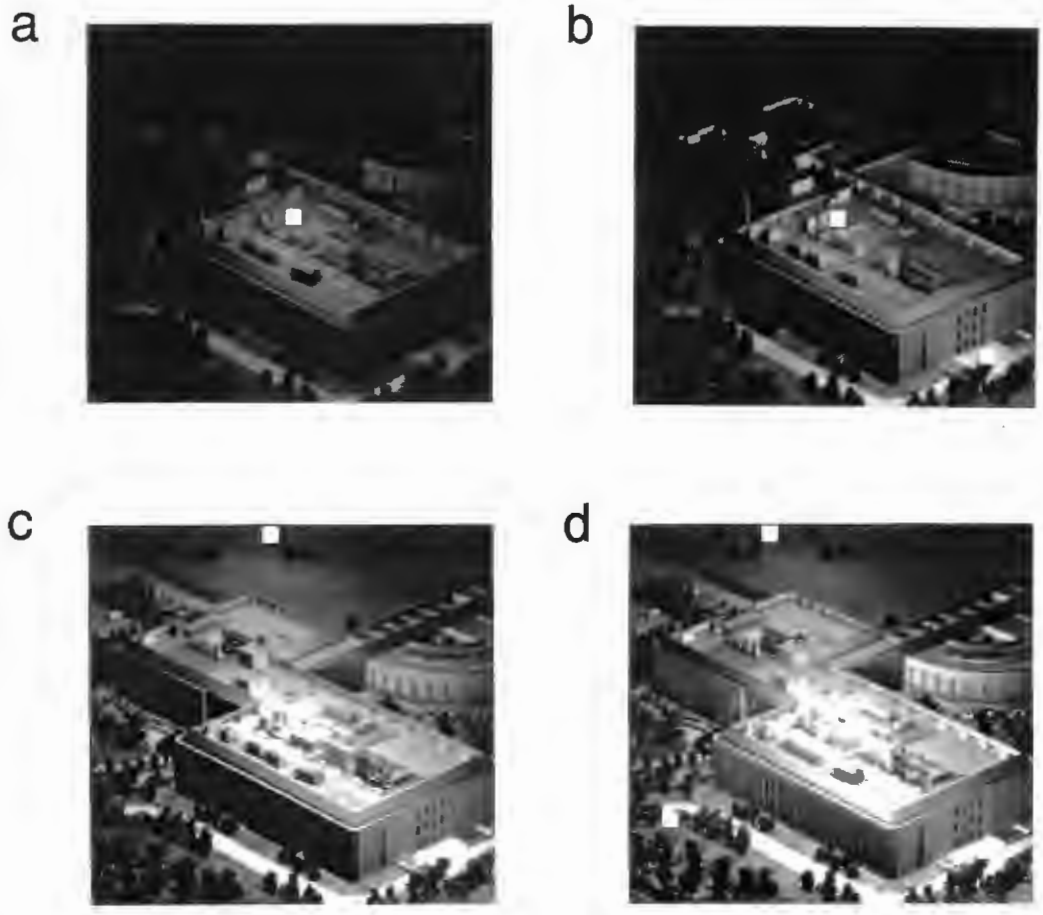


Fig. 19