Systems and Computers in Japan, Vol. 33, No. 11, 2002
Translated from Denshi Joho Tsushin Gakkai Ronbunshi, Vol. J83-D-II, No. 12, December 2000, pp. 2652–2662

# Shape from Contour Using Adaptive Image Selection

Takashi Akutsu[1], Ken-ichi Arakawa,[1,*] and Hiroshi Murase[2]

[1]NTT Cyber Space Laboratories, Yokosuka, 239-0842 Japan

[2]NTT Communication Science Laboratories, Atsugi, 243-0198 Japan

## SUMMARY

The "shape-from-contour method" reconstructs the 3D shape of the surface of an object by extracting its contour from each of a series of successive images of the object. This can be realized by using a CCD camera, and is a relatively accurate method of obtaining environmental information. However, to obtain an accurate result, many images must be processed. Therefore, it is important to select the images depending on their effect on the final result, particularly for high-speed processing.

This paper proposes an adaptive image selection (AISE) method which depends on the required accuracy. The method has been applied to objects having various cross-sectional shapes. The errors in shape reconstruction and the number of images required in the conventional method and the proposed method are compared. The experimental results show that the proposed method requires fewer images than the conventional method, particularly when the surface curvature of the object has large variation. The theoretical relationship between the accuracy of the reconstructed shape and the number of images required is also derived. © 2002 Wiley Periodicals, Inc. Syst Comp Jpn, 33(11): 50–60, 2002; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/scj.1163

**Key words:** contour; measurement of shape; evaluation of error; adaptive processing; variation of curvature.

## 1. Introduction

Obtaining information on the real environment from images has been a useful method for use with intelligent robots designed for high-order tasks. The extraction of 3D data from images that is required for this technique is an important subject in the field of computer vision. The "shape-from-contour" method has been found to be a relatively simple and yet effective [1–3] method that reconstructs the 3D shape of the surface of an object by extracting the contours of the object from sequential images. This method has recently been further improved [4–10].

The measured 3D shape data are useful in these applications. However, the quality and amount of such data vary depending on the application. Generally, rapid extraction of the shape of an object, with a specified accuracy, from the original images is essential. The accuracy of this kind of measurement has been studied [4, 7, 8], but the relationship between the accuracy and the number of images required has not been considered. In addition, Refs. 11 and 12 have described a way to obtain high accuracy by using sampling and interpolation, but for image coding or image generation, not measurement of shape. Reference 6 has described the interpolation of obtained images, but not the relationship between processing efficiency and errors. There have been no published papers describing the rela-

*Now affiliated with NTT, Third R&D Promotion Department.

tionship between the errors of measurement, the number of images required, and the shapes of objects.

This paper describes a method which measures the 3D shape of an object by extracting its contours from fewer images with a specific accuracy, so that less time is required to obtain the results. The measurement of the shape is performed by using the contour of the object and the position of the camera calculating each point on the surface of the object. The contour is a function of factors such as the relative position of the object seen from the camera, and the shape of the object. Generally, a reduction in the number of images reduces the accuracy of measurement, since the intervals of change of the image become coarse. However, the accuracy depends on the choice of images selected. It is important to select images which contribute to high-accuracy results.

This paper proposes the "adaptive image selection with error estimation" method (AISE), which selects images by using the maximum error, indicating the size of the uncertainty region in the result. We theoretically determine the maximum number of images required when the AISE is applied. The theory has been tested experimentally by using several objects which were prepared with high precision. The results show that the proposed method is especially effective for objects whose surface curvatures vary significantly, that is, an accurate shape is reconstructed with fewer images.

## 2. Principle of Shape from Contours

Measuring the surface shape of an object by using its contour extracted from sequential images is called the "shape-from-contour" method. The procedure is as follows:

a) The camera used for the measurements is calibrated.

b) Sequential images of an object are obtained with adequately fine intervals.

c) The contour of the object is extracted from each image.

d) Points on the surface of the object are calculated from the changes of contours.

e) The shape of the object is obtained by forming a surface which contains the points.

Figure 1 shows the principle of the shape-from-contour method, where B is an object, S is its surface, $\mathbf{v}(t)$ is the position of the camera center, and $t$ is time. The position of point $\mathbf{r}(t_0)$ on B at $t_0$ is calculated as follows: When $\mathbf{v}(t)$ moves, the contour also moves. Curve $\mathbf{r}(t)$ passing through $\mathbf{r}(t_0)$ is determined by using point $\mathbf{r}(t)$ at each time. Let the unit vector in the viewing direction from $\mathbf{v}(t)$ to $\mathbf{r}(t)$ on each

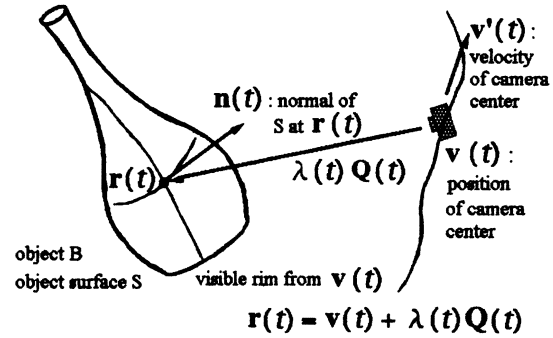


Fig. 1. Principle of shape from contour.

occasion be $\mathbf{Q}(t)$, and let the distance between $\mathbf{v}(t)$ and $\mathbf{r}(t)$ be $\lambda(t)$. At each time increment,

$$\mathbf{r}(t) = \mathbf{v}(t) + \lambda(t)\,\mathbf{Q}(t) \tag{1}$$

Letting the derivative of $\mathbf{r}(t)$ be $\mathbf{r}'(t)$, it is normal to $\mathbf{n}(t)$, since $\mathbf{r}'(t)$ is contained in the tangential plane of S. $\mathbf{Q}(t)$ is also normal to $\mathbf{n}(t)$. Therefore, by differentiating each term of Eq. (1) with respect to $t$, and taking the inner product with $\mathbf{n}(t)$, the following equation is obtained:

$$\lambda(t) = \frac{-\mathbf{v}'(t)\cdot\mathbf{n}(t)}{\mathbf{Q}'(t)\cdot\mathbf{n}(t)} \tag{2}$$

where $\mathbf{v}'(t)$ and $\mathbf{Q}'(t)$ represent the derivatives of $\mathbf{v}(t)$ and $\mathbf{Q}(t)$, respectively.

$\lambda(t)$ can be obtained from Eq. (2), and $\mathbf{r}(t_0)$ can be obtained from Eq. (1), by calculating $\mathbf{v}(t)$, $\mathbf{Q}(t)$, and $\mathbf{n}(t)$ with sufficiently small intervals of $t$. The shape of S is obtained by applying this procedure to each point on S. This procedure can be listed in the previously described order as:

a) Find a correspondence between the viewing direction $\mathbf{Q}(t)$ and a pixel in the camera.

b) An image is obtained while calculating $\mathbf{v}(t)$ each time.

c) A contour is calculated from the image, and $\mathbf{Q}(t)$ and $\mathbf{n}(t)$ are calculated.

d) $\lambda(t)$ and $\mathbf{r}(t)$ are calculated from $\mathbf{v}(t)$, $\mathbf{Q}(t)$, and $\mathbf{n}(t)$.

e) The shape of surface S is calculated from the $\mathbf{r}(t)$.

## 3. Intervals of Sequential Images, and Uncertainty Region

In procedure d) in Section 2, the position of each point is calculated to obtain shape S. Equation (2) in this

process becomes a limit-value operation with respect to $t$. Since it is impossible in practice to repeat the procedure infinite times, Eq. (2) is applied to the sampling of images with sufficiently small intervals of the camera positions. It is more practical for reducing the amount of computation to select the intervals of $\mathbf{v}(t)$ and $\mathbf{Q}(t)$ in Eq. (2).

Figure 2 shows the relationship between the intervals of the images taken and the accuracy of $\lambda$ and explains how to calculate the position of point $\mathbf{r}(t_2)$ on a visible rim when $t = t_2$. Let a straight line passing through $\mathbf{v}(t_2)$ and $\mathbf{r}(t_2)$ in the viewing direction be $\ell(t_2)$. The cross point between the visible rim from $\mathbf{v}(t)$ and a plane containing $\mathbf{v}(t)$ and $\ell(t_2)$ is $\mathbf{r}(t)$ when $t \neq t_2$. Let the straight line passing through $\mathbf{v}(t)$ and $\mathbf{r}(t)$ be $\ell(t)$. From this definition of $\mathbf{r}(t)$, each straight line $\ell(t)$ has a cross point with $\ell(t_2)$. Referring to $t_1 < t_2 < t_3$, $\ell(t_1)$ and $\ell(t_3)$ are chosen. Let the cross point between $\ell(t_1)$ and $\ell(t_3)$ be A, and let the cross point between $\ell(t_3)$ and $\ell(t_2)$ be B. Let the distance between $\mathbf{v}(t_2)$ and A be $\lambda_A$, and let that between $\mathbf{v}(t_2)$ and B be $\lambda_B$. Then from Eq. (2), $\lambda_A \geq \lambda(t_2) \geq \lambda_B$. Generally, only the values of $\lambda_A$ and $\lambda_B$ are obtained by processing three images at $t_1$, $t_2$, and $t_3$, and not $\lambda(t_2)$. Therefore, the calculated result contains an uncertainty, depending on the value of $\lambda_A - \lambda_B$.

The question is what effect on the calculated shape is produced by the uncertainty of $\lambda$ which is affected by the intervals of the images. Plane $\alpha$ containing $\mathbf{v}(t_1)$, $\mathbf{v}(t_2)$, and $\mathbf{r}(t_2)$ (see Fig. 3) is called an epipolar plane [13] of point $\mathbf{r}(t_2)$ in the space. Now, let us consider the result of calculation of the shape of the common part of this $\alpha$ and the object. When $t_0 < t_1 < t_2 < t_3$, let the cross point between the visible rim from $\mathbf{v}(t_1)$ and $\alpha$ be $\mathbf{r}(t_1)$. Let the straight lines in the viewing direction be $\ell(t_1)$ and $\ell(t_2)$ (as in Fig. 2). It is assumed that $\mathbf{r}(t_0)$ is a cross point between the visible rim from $\mathbf{v}(t_0)$ and a plane containing $\mathbf{v}(t_0)$ and $\ell(t_1)$. It is also assumed that $\mathbf{r}(t_3)$ is a cross point between the visible rim

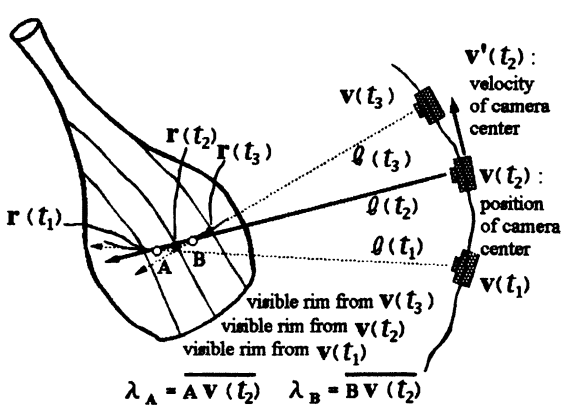


Fig. 2. Relationship between the distance of positions where processed images are taken and calculated lambda.

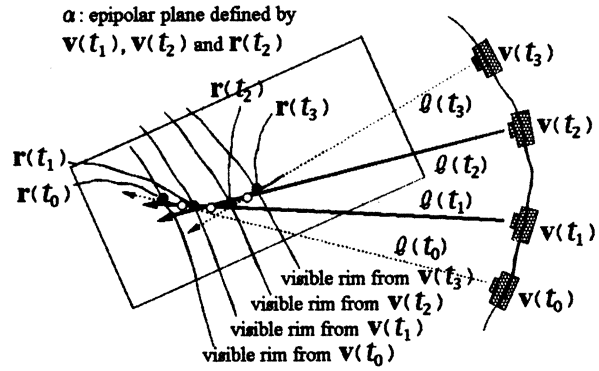


Fig. 3. Relation between the distance of positions where processed images are taken and calculated shape.

from $\mathbf{v}(t_3)$ and a plane containing $\mathbf{v}(t_3)$ and $\ell(t_2)$. Let the cross point between $\ell(t_0)$ and $\ell(t_1)$ be $P_{01}$. Let the cross point between $\ell(t_1)$ and $\ell(t_2)$ be $P_{12}$. Let the cross point between $\ell(t_2)$ and $\ell(t_3)$ be $P_{23}$. Then, $\mathbf{r}(t_1)$ is somewhere on the segment from $P_{01}$ to $P_{12}$, and $\mathbf{r}(t_2)$ is somewhere on the segment from $P_{12}$ to $P_{23}$. The actual shape of an object cannot be exactly determined due to the uncertainty caused by the intervals of image processing. Let us call such a region an "uncertainty region." Figures 4(a) and 4(b) show examples of different shapes giving the same measured result. Note that these four images were processed from

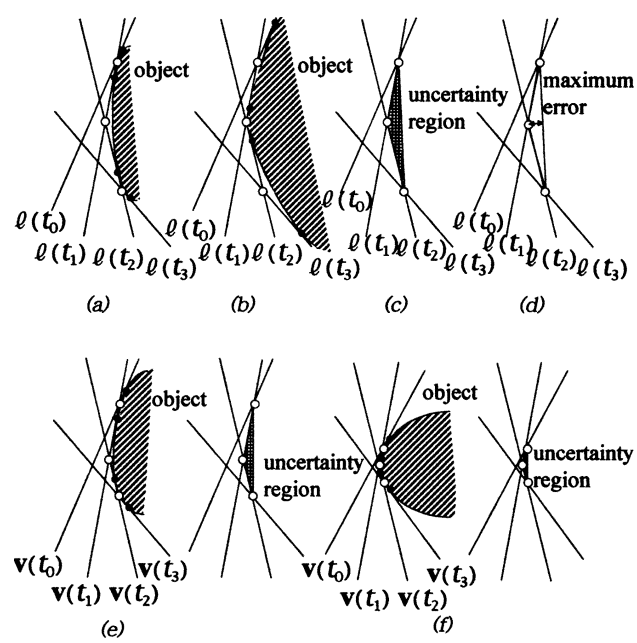


Fig. 4. Uncertainty region and maximum error.

$\mathbf{v}(t_0)$, $\mathbf{v}(t_1)$, $\mathbf{v}(t_2)$, and $\mathbf{v}(t_3)$. The shaded area in Fig. 4(c) shows the uncertainty region. To measure this region more precisely, it is necessary to use finer intervals for the processing. Let us call the height of the uncertainty region the "maximum error" [see Fig. 4(d)], since the measured result possibly contains this amount of error.

The size of the uncertainty region differs with the shape of the object, even if the image intervals and the position of the camera are the same [Figs. 4(e) and 4(f)]. Generally, if the number of images is reduced, the size of the uncertainty region increases.

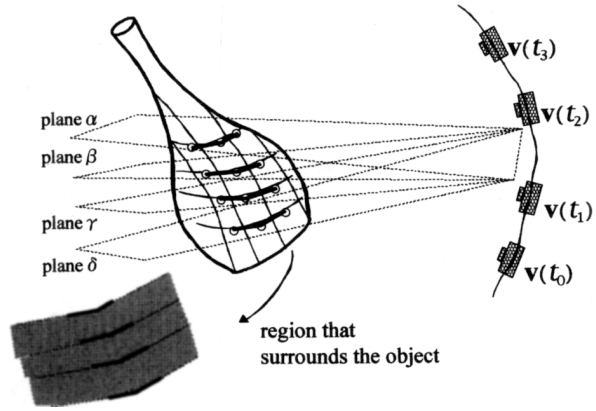


Fig. 6. Object shape reconstruction in each plane.

## 4. Selection of Images Using Maximum Error Evaluation

### 4.1. Proposal of AISE method

We propose the AISE method, which selects images to process using the maximum error, so that the measurement is carried out with fewer images and smaller error. Figure 5 shows the notations of the proposed method. A sphere (radius $d$) containing an object to be measured is considered. The admissible error $E_{ad}$ is determined with reference to the required accuracy of measurement of the shape. The intervals of sampling of the images should be sufficiently small (as described in Section 5). Assuming that the position of the center of the camera $\mathbf{v}(t)$ is given, the shape of the common part of the epipolar plane $\alpha$ defined by $\mathbf{v}(t_1)$, $\mathbf{v}(t_2)$, and $\mathbf{r}(t_2)$ and the object is considered, by using the four processed images at $\mathbf{v}(t_0)$, $\mathbf{v}(t_1)$, $\mathbf{v}(t_2)$, and $\mathbf{v}(t_3)$. Then the uncertainty region on $\alpha$ and the maximum error $E_{max,\alpha}$ are obtained. The viewing angles of $\mathbf{r}(t_2)$ at $\mathbf{v}(t_1)$ and $\mathbf{v}(t_2)$ are changed, and we denote the epipolar planes after the changes as $\beta$, $\gamma$, $\delta$, . . . , respectively (Fig. 6). The uncertainty regions on these planes, and the maximum errors $E_{max,\beta}$, $E_{max,\gamma}$, and $E_{max,\delta}$ are calculated. Let the maximum values of $E_{max,\alpha}$, $E_{max,\beta}$, $E_{max,\gamma}$, . . . , be $E_{max}(\mathbf{v}(t_1), \mathbf{v}(t_2))$. If $E_{max} > E_{ad}$, an image obtained between $\mathbf{v}(t_1)$ and $\mathbf{v}(t_2)$ is processed additionally.

In the proposed method, the above process is repeated. Figure 7 shows the order of processing of images. The images to be processed are selected by using a predetermined "standard interval." The value $E_{max}$ of each section is calculated from each set of four images, and this is compared with the admissible error $E_{ad}$ (① to ⑤ in Fig. 7). If $E_{max} > E_{ad}$ (② and ④ in Fig. 7), another image which is obtained from the center of the four images is added, and the above steps are applied to each of the four images (⑥ and ⑦; and ⑧ and ⑨ in Fig. 7). If still $E_{max} > E_{ad}$ (⑧ in Fig. 7), the above steps are repeated (⑩ and ⑪ in Fig. 7).

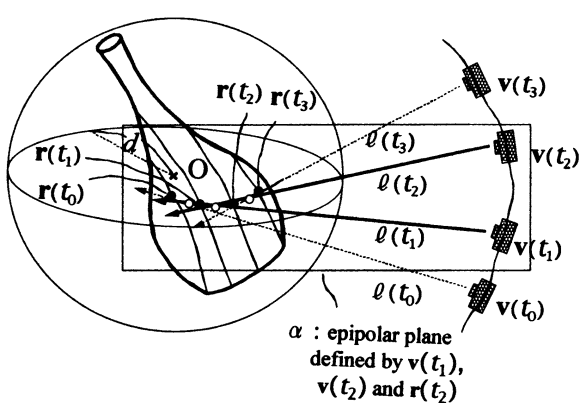


Fig. 5. Notations for AISE method.

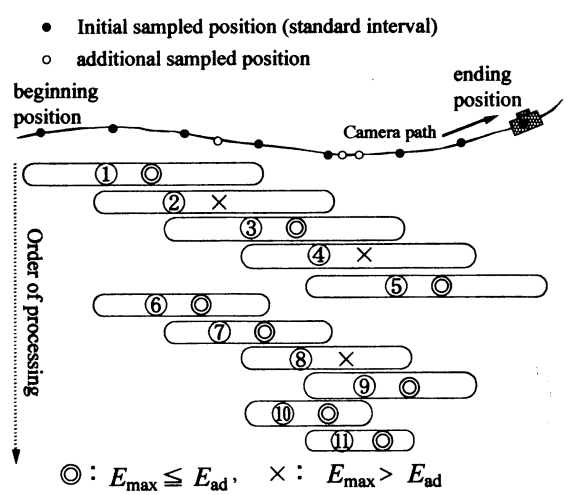


Fig. 7. Order of processing images in AISE method.

Generally, this repetition is continued until $E_{max} \leqq E_{ad}$ is satisfied at each section. Then, the final shape of the object is obtained by using the range containing the object as shown in Fig. 6.

Figure 8 shows how the initial standard interval is set. The point on $\mathbf{v}(t)$ is set so that the angle spanned by a pair of adjacent points becomes a certain standard angle $\theta_{std}$, at the center of the sphere containing the object. As shown in Fig. 8, $\theta_{std}$ is chosen so that

$$E_{ad} = 2d \tan\frac{\theta_{std}}{2} \sin\frac{\theta_{std}}{2} \tag{3}$$

where the distances between the center of the sphere O and each $\mathbf{v}(t)$ are the same. The added point for processing is selected so that it equally divides the angle spanned from the center of the sphere to the points.

### 4.2. Number of images with AISE method

The possible reduction of images by the AISE method is now theoretically examined. For simplicity, it is assumed that a single cross section is enough for the purpose, and that images are obtained from a circle on the plane containing the cross section. The number of images $N_{prp}$ processed by using the proposed method, and the number of images $N_{sam}$ required to process a shape with uniform intervals and an error less than $E_{ad}$, are evaluated.

Let the radius of a circle on which the center of the camera shifts be $D$, assuming that the object is contained in a sphere of radius $d$ from the center of the circle. It is assumed that $E_{ad} << d << D$, where $E_{ad}$ is the admissible error. Also, $pE_{ad} < d < D/q$ if $N_{prp}$ and $N_{sam}$ are quantitatively

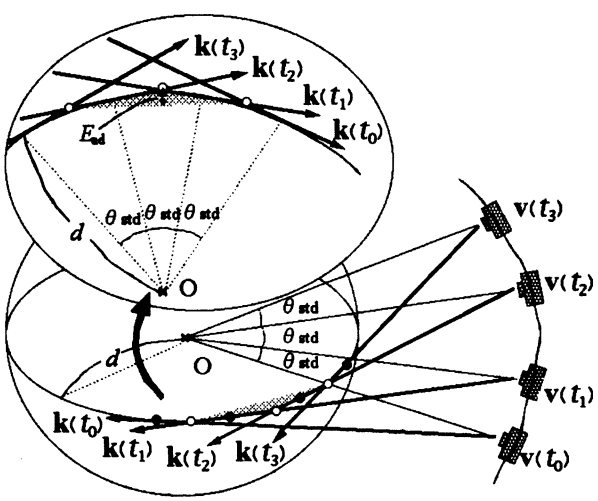


Fig. 8. Admissible error and standard interval.

expressed where $p$ and $q$ are positive. When the method described in section 4.1 is applied to a section, $\mathbf{r}(t_2)$ for each of four points, $\mathbf{v}(t_0)$, $\mathbf{v}(t_1)$, $\mathbf{v}(t_2)$, and $\mathbf{v}(t_3)$, can be chosen in two ways: from the left or right contour (one of them is used in this section for evaluating $N_{prp}$, for simplicity). Also, for images with uniform intervals, $N_{sam}$ using one of the contours alone is evaluated, and compared with the results obtained by using $N_{prp}$.

The procedure for applying the proposed method is as follows. First, $\theta_{std}$ is calculated from Eq. (3) as

$$\cos\frac{\theta_{std}}{2} = \sqrt{1+\left(\frac{a}{2}\right)^2} - \frac{a}{2} \tag{4}$$

where $E_{ad}/2d = a$. The number of images obtained by processing the images with the interval of $\theta_{std}$ on the circle on which the camera center shifts is $N_{std} = [2\pi/\theta_{std}] + 1$, where $[\cdot]$ denotes the floor (the largest integer smaller than the argument). From $N_{std} - 1 < 2\pi/\theta_{std}$, $\cos(\theta_{std}/2) > \cos(\pi/(N_{std} - 1))$ holds. After substitution of Eq. (4) both sides squared and the result is rearranged by using $\cos^2 = 1 - \sin^2$ so that

$$\sin^2\frac{\pi}{N_{std}-1} > \frac{a}{\sqrt{1+\left(\frac{a}{2}\right)^2} + \frac{a}{2}}$$

is obtained. Taking the square roots of both sides, $\pi/(N_{std} - 1) > \sin(\pi/(N_{std} - 1))$ and $0 < a < 1/2p$,

$$N_{std} < 1 + \frac{\pi}{\sqrt{a}}\sqrt{\sqrt{1+\frac{1}{16p^2}} + \frac{1}{4p}} \tag{5}$$

is obtained. By using Eq. (5), the assertion below, and constant $C(p, q)$,

$$N_{prp} < C(p,q)\left(1 + \frac{\pi}{\sqrt{a}}\sqrt{\sqrt{1+\frac{1}{16p^2}} + \frac{1}{4p}}\right) \tag{6}$$

is obtained. If $p$ and $q$ are sufficiently large, $C(p, q)$ becomes close to $1 + 2/e \log_e 2$.

Assertion: For the proposed method, with the arrangement described in this section, the following equation holds (see Appendix):

$$N_{prp} < C(p,q)\, N_{std} \quad \left(p, q \to \infty, C(p,q) \to 1 + \frac{2}{e \log_e 2}\right)$$

Figure 9 shows an example of an object which is likely to require many images to process at equal intervals.

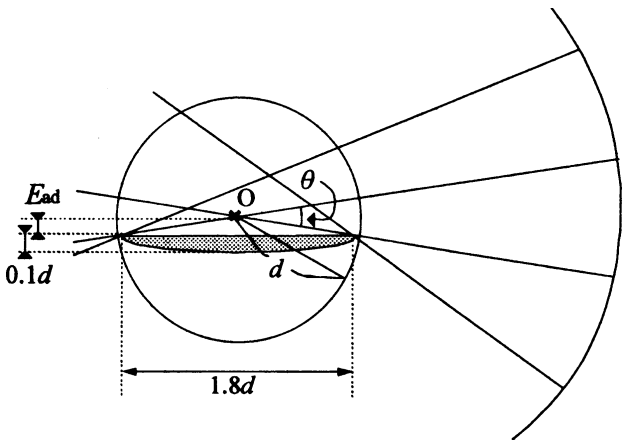


Fig. 9. An example of the sectional shape of the object for which many images to process can be necessary in nonadaptive method.

In this figure, an object is at a distance $E_{ad}$ from the center of camera movement. To measure a shape in this arrangement with the accuracy of $E_{ad}$, the angle θ must satisfy $E_{ad}/0.9d > \tan(\theta/2)$, where θ is an angle formed by two image sampling points (uniform intervals) and the center of the camera movement. Using $E_{ad}/2d = a$ and $\tan x > x$ ($x > 0$), we obtain

$$2\,a > 0.9 \tan\frac{\theta}{2} > 0.45\,\theta$$

Using this condition, the number of processed images is obtained as

$$N_{sam} \geqq \frac{2\pi}{\theta} > \frac{0.45\pi}{a} \tag{7}$$

When the variation of curvature is large, as in the example shown in Fig. 9, generally $N_{sam}$ becomes large. If an object to be measured coincides with a sphere having a constant radius $d$ (from the center of camera movement), $N_{sam} = N_{prp} = N_{std}$ holds. Then the value of $N_{sam}$ is small so that the proposed method does not shows its effect.

Let us compare Eqs. (6) and (7): When $p = q = 9$ and $E_{ad}/d = 1/25$, then $N_{prp} < 70.6$ and $N_{sam} > 70.7$ are obtained, showing that their difference is small. When $E_{ad}/d = 1/50$, then $N_{prp} < 98.6$ and $N_{sam} > 141.4$ are obtained, showing that their difference is large. From Eqs. (6) and (7), it is found theoretically that a smaller value of $E_{ad}/d$ makes the proposed method more effective as

$$\left(\frac{N_{prp}}{N_{sam}} \propto \sqrt{\frac{E_{ad}}{d}}\right)$$

## 5. Experimental Evaluation of Method Using Actual Objects

Experiments were carried out with actual objects under the conditions described in Section 4.2. The effects of the proposed method were evaluated by comparing the number of images required and the error of shape measurements in the proposed method and a conventional method. In Sections 2 to 4, errors other than image-sampling intervals were not included for simplicity. They are included in experimental evaluations in this section. In Section 4.2, an image was sampled with a camera whose center shifts on a circle (radius $D$). An equivalent effect can be obtained by using an object on a rotating disk and a CCD camera fixed on a plane which passes through the object. Figure 10 shows the nine objects used for the experiment. They were manufactured with an error less than 0.1 mm.

The CCD camera system was calibrated for the experiment by using a board perpendicular to the optical axis of the camera.

Nine images with different depths were obtained. Taking 12 feature points on each image (108 points in all), the correspondences between these points and the actual coordinates were found. Choosing 8 images from them (96

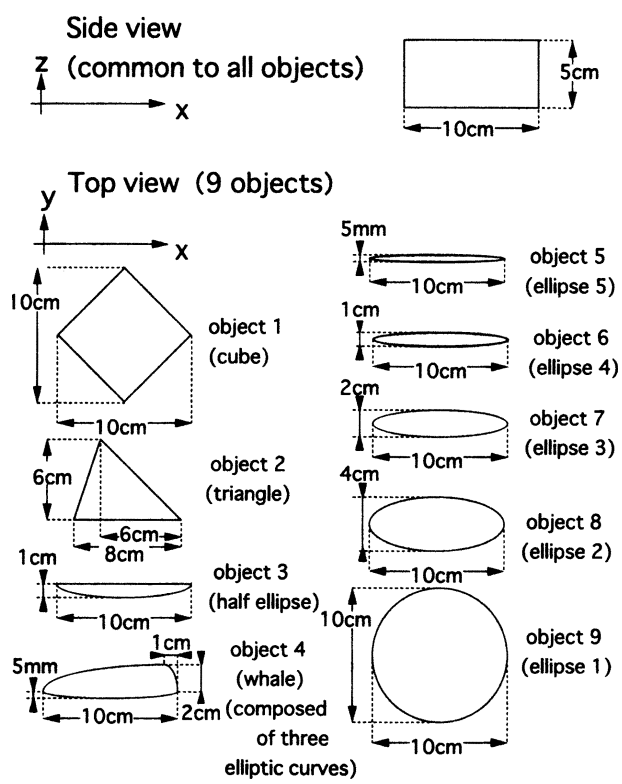


Fig. 10. Objects' shape used in the experiment.

points in all), camera parameters (e.g., the position of the camera center, and the focal length of the lens) were calculated by using Tsai's calibration method [11].

The calibration showed that the distance $D$ (see Section 4.2) between the rotation shaft and the camera center was 1106.6 mm. Each object (in Fig. 10) was set on the rotation table so that the radius $d$ of a sphere containing the object was 80 mm. The object was turned through 230° at intervals of 0.2°. 1151 images were taken for each object. The intervals were chosen for the objects, including those shown in Fig. 10, so that the maximum error was less than $E_{ad}$ satisfying $E_{ad}/d > \tan(\theta/2)$. For example, if $d = 80$ and $E_{ad} \geq 0.5$, then the interval is 0.7° from $\theta° < 0.71° \leqq 360\ E_{ad}/\pi d$. In practice, however, the interval was chosen as 0.2° for safety, considering the error in the calibration. Each object was uniformly illuminated by halogen lamps. Each object was white on a dark background. Each image had a size of 640 × 480 pixels in RGB form. The system was adjusted so that the pixel line in the horizontal direction was parallel to the object rotation table.

Figure 11 shows the luminance variation on *object*-9 (see Fig. 10), obtained by converting the RGB value of each pixel along the section. The contour of each object was obtained by taking the points at which the luminance changed sharply (the left and right edges of an object). Due to the structure of the rotation table, it was difficult to rotate the table accurately through a full 360°, and only one side of each contour was measured, while the other side of the contour was used as the available contour for the unrotatable side.

Figure 12 shows the results of the experiment with nine objects (Fig. 10) processed using the conventional method and the proposed method. For the proposed method, the admissible errors $E_{ad}$ were chosen to be from 0.5 mm to 3.0 mm with an interval of 0.5 mm (six values of $E_{ad}$). The experiment was performed with 10 to 70 images in the conventional method. In the experiment shown in Fig. 12, 15 different positions of the initial processing images were chosen, and the average value of each result was plotted.

For the reference value in the comparison of measuring errors in Fig. 12, a shape which minimized the parallax in the direction of a contour (derived from all of the data) was used. This is because the purpose of this experiment is to demonstrate a certain accuracy with a small number of images, and the objects could not all be set with an accuracy of 0.1 mm. The measuring error in Fig. 12 indicates the Hausdorff distance between the comparison shape of each object and the measured shape.

This paper proposes the AISE method, which can obtain results with an appropriate accuracy by using a relatively small number of images, and mainly discusses the errors due to the spacing of the sampling positions of the images to be processed. However, errors due to other causes cannot be ignored in practice. For example, the maximum error in camera calibration is 0.3 mm, and that in contour extraction is 0.6 mm. The error in calibration was estimated by using 12 feature points (which were not used for parameters) out of 108 feature points. The error in contour extraction was estimated by using a shift of 2 pixels. When the sampling points of two adjacent images are close in the experiment, the calculation of the cross point of the viewing line for the corresponding point can contain unacceptable error by errors due to other causes [12]. To avoid this, the order of crossing of the view lines was checked before the maximum error was calculated. If the order was not correct, the maximum error was regarded as zero.

The experiment reveals that objects having large variations of the surface curvature (e.g., objects 1, 2, 3, and 5) show a higher effect of reduction of the number of images, but otherwise there is no significant effect in the shape with little variation of the curvature.

The theoretical evaluation of the proposed method shows that the proposed method is more effective with decreasing measurement error. However, this is not recognized in the experiment, due to the fact that other kinds of errors increase in such a case.

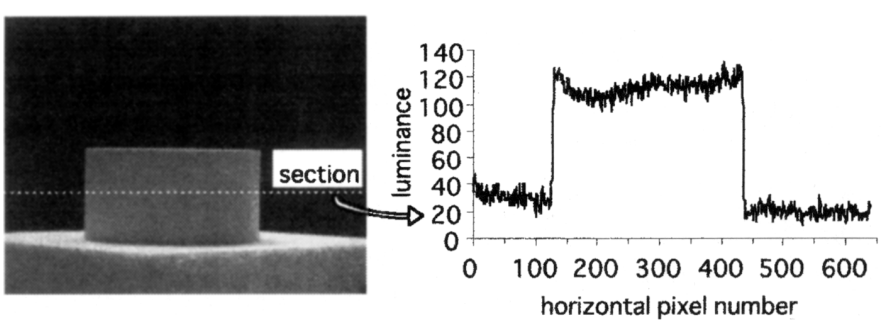


Fig. 11. Luminance on the section of the object's surface to reconstruct.

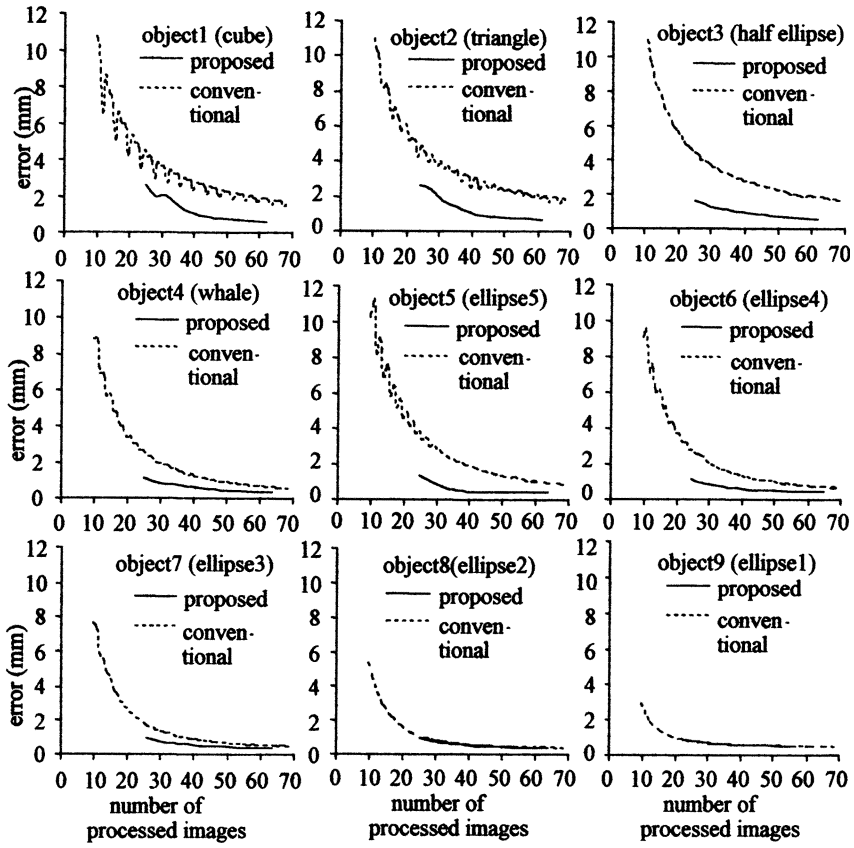


Fig. 12. Results of the experiment.

## 6. Conclusion

This paper proposes the method of "adaptive image selection with error estimation" (AISE), which accurately measures the shape of an object by using its contours and processing the maximum error.

The number of images to be processed using the proposed method has been shown to be less than three times the number of images processed using the standard intervals, and the number of images required for the same accuracy as the conventional method has been theoretically evaluated. The results show that the proposed method can reduce the number of images to be processed even for an object having a large variation in surface curvature. The results also show theoretically that the proposed method is more effective in cases where a higher accuracy is required.

The measured results were experimentally evaluated by using several objects specially made with high accuracy for the purpose. The proposed method has been shown to be more effective for objects having large variations in surface curvature.

**Acknowledgments.** The authors thank Dr. Kenji Kogure (Director, NTT Cyber Space Laboratories) and Dr. Yo-ichi Tohkura (Director, NTT Communication Science Laboratories) for their approval of this project, and Dr. Masahiko Hase (Project Manager, NTT), Dr. Naoki Kobayashi (Group Leader, NTT), and many other people, for their advice on the project.

## REFERENCES

1. Ballard DH, Brown CM. Computer vision. Prentice–Hall; 1982.
2. Koenderink JJ. What does the occluding contour tell us about solid shape? Perception, 1984, p 321–330.
3. Giblin P, Weiss R. Reconstruction of surfaces from profiles. Proc 1st Int Conference on Computer Vision, p 136–144, 1987.
4. Cipolla R, Blake A. Surface shape from the deformation of apparent contours. Int J Computer Vision 1992;9:83–112.

5. Kato K, Nakanishi T, Shio A, Ishii K. Structure from image sequences captured through a monocular extra-wide angle lens. Proc CVPR, p 919–924, 1994.
6. Zhao C, Mohr R. Relative 3D regularized B-spline surface reconstruction through image sequences. Proc 3rd ECCV, p 417–426, 1994.
7. Zheng JY. Acquiring 3-D models from sequences of contours. IEEE Trans Pattern Anal Mach Intell 1994;16:163–178.
8. Boyer E, Berger MO. 3D surface reconstruction using occluding contours. Int J Computer Vision 1997;22:219–233.
9. Sato J, Kinoshita K, Cipolla R. Epipolar geometry on curved surfaces and visual navigation from uncalibrated cameras. Trans IEICE Japan 1999;J82-D-II:1401–1410.
10. Matsuyama T, Hisano Y, Inomiya J. Computer vision: Status of technology and future prospects. New Technology Communications; 1998.
11. Belfor RAF, Hesp MPA, Lagendijk RL, Biemond J. Spatially adaptive subsampling of image sequences. IEEE Trans Image Processing 1994;3:492–500.
12. Kobayashi T, Quan L, Ohta T. Construction of the virtual image of an object viewing from an arbitrary point using an adaptive selection. Tech Rep IEICE 2000;PRMU99-195.
13. Gang X, Tsuji S. 3-D vision. Kyoritsu Press; 1998.
14. Tsai RY. An efficient and accurate camera calibration technique for 3D machine vision. Proc. CVPR, p 364–374, 1986.
15. Kanatani K. Geometric information criterion for model selection. Int J Computer Vision 1998;26: 171–189.

## Appendix

### Proof of Assertion in Section 4.2

Since the convex closure of the sectional shape of an object is contained in a circle of radius $d$, its circumference $L_{\mathrm{std,srd}}$ (processed with the standard interval) is shorter than the thick line in Fig. A.1(a). Since $L_{\mathrm{srd}}$ (circumference with additional processing) is shorter than $L_{\mathrm{std,srd}}$, the following equation holds:

$$L_{\mathrm{srd}} \leqq L_{\mathrm{std,\,srd}} < 2d\, N_{\mathrm{std}} \tan\frac{\theta_{\mathrm{std}}}{2} \tag{A.1}$$

The assertion holds for $N_{\mathrm{prp}} = N_{\mathrm{std}}$, when the images are processed with the standard intervals, and when $E_{\max} \leq E_{\mathrm{ad}}$ at each $N_{\mathrm{std}}$ section. If $N_{\mathrm{prp}} > N_{\mathrm{std}}$, the number of images with additional processing is calculated using Eq. (A.1), since images to be additionally processed exist.

Figure A.1(b) shows the process of the proposed method. □ represents $N_{\mathrm{std}}$ images which are initially processed. ○ and ● represent images which have been processed when $E_{\max}$ exceeds $E_{\mathrm{ad}}$. ● represents the terminal point of an additional procedure (an image whose additional sections show $E_{\max} \leq E_{\mathrm{ad}}$).

Now, let us evaluate the circumference ($\leq L_{\mathrm{std,srd}}$) of a shape formed by connecting the center of each side of the sections in which □ and ○ have been processed but ● has not been processed.

Figure A.1(c) shows a section in which □ and ○ have been processed but ● has not been processed. In such a section, $E_{\max} > E_{\mathrm{ad}}$. The region surrounded by the thick line is the uncertainty region.

Let the length of a segment connecting the center of each side [dotted line in Fig. A.1(c)] be $M_{\phi,\rho}$. Then

$$\begin{aligned} M_{\phi,\rho} &= \frac{E_{\max}}{2}\left(\frac{1}{\tan\rho} + \frac{1}{\tan(\phi-\rho)}\right) \\ &\geqq \frac{E_{\max}}{\tan\frac{\phi}{2}} > \frac{E_{\mathrm{ad}}}{\tan\frac{\phi}{2}} \end{aligned} \tag{A.2}$$

Therefore, $M_{\phi,\rho}$ can be evaluated by using $\phi$ alone. Taking the section of ● in Fig. A.1(b) which has not processed "additional processing (1st step)", let the length of a segment connecting the center of adjacent edges be $M_1$. From $\phi_1$ in Fig. A.1(d),

$$\begin{aligned} M_1 > \frac{E_{\mathrm{ad}}}{\tan\frac{\phi_1}{2}} &= E_{\mathrm{ad}} \frac{D\cos\frac{\theta_{\mathrm{std}}}{2} - d}{D\cos\frac{\theta_{\mathrm{std}}}{2}\tan\frac{\theta_{\mathrm{std}}}{2}} \\ &= \frac{E_{\mathrm{ad}}}{\tan\frac{\theta_{\mathrm{std}}}{2}}\left(1 - \frac{d}{D\cos\frac{\theta_{\mathrm{std}}}{2}}\right) \end{aligned}$$

Using $0 < a < 1/2p$, $d < D/q$, and Eq. (4),

$$\begin{aligned} M_1 &> \frac{E_{\mathrm{ad}}}{\tan\frac{\phi_1}{2}} > \frac{C_1(p,q)\,E_{\mathrm{ad}}}{\tan\frac{\theta_{\mathrm{std}}}{2}}, \\ C_1(p,q) &= 1 - \frac{1}{q}\left(\sqrt{1+\frac{1}{16p^2}} + \frac{1}{4p}\right) \end{aligned} \tag{A.3}$$

Similarly, taking the section of ● in Fig. A.1(b) which has not processed "additional processing ($n$-th step)", let the length of a segment connecting the center of adjacent edges be $M_n$ ($n = 2, 3, \ldots$). Using $\varphi \leqq \theta_{\mathrm{std}}/2$, Eq. (4), $0 < a < 1/2p$, and $d < D/q$, from $\phi_{n-1}$ and $\phi_n$ in Fig. A.1(e),

$$\frac{\tan\frac{\phi_n}{2}}{\tan\frac{\phi_{n-1}}{2}} \leqq \frac{\dfrac{D\sin\frac{\varphi}{2}}{\cos\frac{\varphi}{2} - d}}{\dfrac{D\sin\varphi}{D+d}} < \frac{\dfrac{q+1}{q\cos\frac{\varphi}{2} - 1}}{2\cos\frac{\varphi}{2}} < C_2(p,q)$$

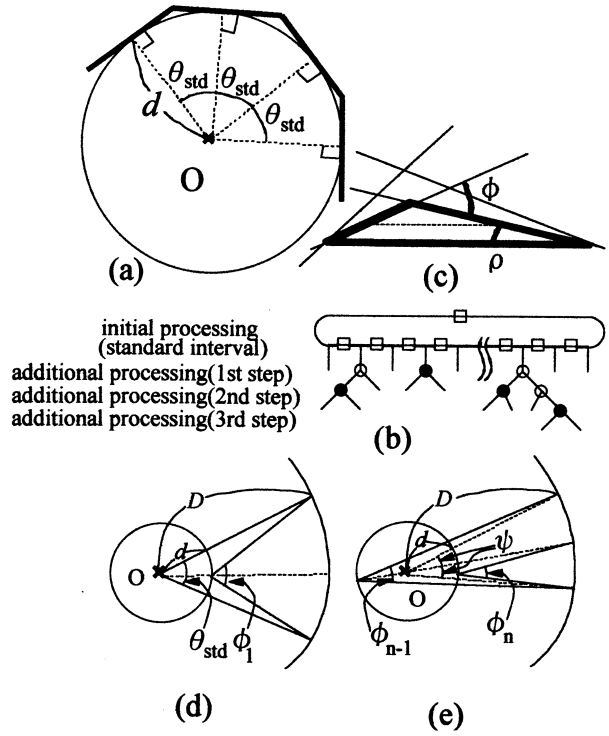


Fig. A.1 AISE method and surrounding length of the calculated shape.

where

$$C_2(p,q) = \frac{q+1}{\sqrt{1-\frac{1}{4p}+\sqrt{1+\frac{1}{16p^2}}}\left(q\sqrt{1-\frac{1}{4p}+\sqrt{1+\frac{1}{16p^2}}}-\sqrt{2}\right)}$$

When $n \geq 1$, from this equation, Eqs. (A.2) and (A.3),

$$M_n > \frac{E_{ad}}{\tan\frac{\phi_n}{2}} > \frac{E_{ad}}{C_2(p,q)\tan\frac{\phi_{n-1}}{2}} > \cdots$$
$$> \frac{E_{ad}}{C_2(p,q)^{n-1}\tan\frac{\phi_1}{2}} > \frac{C_1(p,q)\,E_{ad}}{C_2(p,q)^{n-1}\tan\frac{\theta_{std}}{2}}$$

This also holds from Eq. (A.3), when $n = 1$. The length of a segment per processed image is

$$\frac{M_n}{n} > \frac{C_1(p,q)\,E_{ad}}{C_2(p,q)^{n-1}\,n\tan\frac{\theta_{std}}{2}} \tag{A.4}$$

Using $f(x) = 1/C_2(p, q)^{x-1}x$, we obtain

$$f'(x) = -\frac{\frac{1}{x}+\log_e C_2(p,q)}{C_2(p,q)^{x-1}\,x}$$

Therefore, using $x_{min} = -1/\log_e C_2(p, q)$,

$$f(x) \geqq f(x_{min}) = -e\,C_2(p,q)\log_e C_2(p,q) \quad (x \geqq 1)$$

From this equation and Eq. (A.4),

$$\frac{M_n}{n} > \frac{-E_{ad}\,e\,C_1(p,q)\,C_2(p,q)\log_e C_2(p,q)}{\tan\frac{\theta_{std}}{2}} \tag{A.5}$$

Letting the number of additionally processed images be $k\,N_{std}$ ($k > 0$), from Eq. (A.1), we have

$$\frac{k\,N_{std}\,M_n}{n} < 2d\,N_{std}\tan\frac{\theta_{std}}{2} \tag{A.6}$$

Comparing Eqs. (A.5) and (A.6), from Eq. (4) and $0 < E_{ad}/2d = a < 1/2p$,

$$k < -\frac{\sqrt{1+\frac{1}{16p^2}}+\frac{1}{4p}}{e\,C_1(p,q)\,C_2(p,q)\log_e C_2(p,q)} = C(p,q)-1 \tag{A.7}$$

The right-hand terms of this equation depend only on $p$ and $q$. When $p, q \to \infty$, using $C(p, q) - 1$ in Eq. (A.7) from Eq. (A.3) and the definition of $C_2(p, q)$, $C(p, q) - 1 \to 2/e \log_e 2$. The total number of images is given by $N_{prp} = (k + 1)N_{std} < C(p, q)N_{std}$.

## **AUTHORS** (from left to right)



**Takashi Akutsu** graduated from the Mathematics Department of the University of Tokyo in 1992 and completed the M.E. program (mathematical science) in 1994. He then joined NTT, where he has been engaged in research on robot vision and image coding. He was awarded an IEICE Young Engineer Award in 1998. He is now a research manager at NTT Cyber-Space Laboratories. He is a member of the Institute of Information Processing, Japan, the Robotics Society of Japan, and IEICE Japan.

**Ken-ichi Arakawa** graduated from the Information Engineering Department of Kyoto University in 1984 and completed the M.E. program in 1986. He then joined NTT, where he has been engaged in research on computer vision and robot vision at NTT Basic Laboratories and Human Interface Laboratories. He was a guest researcher at Carnegie–Mellon University in 1990–1991. He is now acting head of the Third R&D Promotion Department at NTT. He is a member of the Information Processing Society, the Japan Robotics Society, AVIRG, IEEE, and IEICE Japan.

**Hiroshi Murase** graduated from the Electronics Department of Nagoya University in 1978 and completed the M.E. program in 1980. He then joined NTT, where he has been engaged in research on character/image recognition and multimedia recognition. He was a visiting researcher at Columbia University in 1992. He is now a research group leader (media recognition) at NTT Communication Science Laboratories. He holds a D.Eng. degree. He has been awarded an IEICE Young Engineer Award (1985), a Telecommunications Prize of the Electrical Communications Promotion Society (1992), a Paper Award at the IEEE–CVPR International Meeting (1994), a Yamashita Memorial Prize of the Information Processing Society (1995), and a Video Prize at the IEEE–CVPR International Meeting (1996). He is a member of the Information Processing Society of Japan, IEEE, and IEICE Japan.