Proposal of Artificial Mark to Measure 3D Pose by Monocular Vision*

Rie KATSUKI**, Jun OTA***, Tamio ARAI*** and Tsuyoshi UYEYAMA****

**Toshiba Corporation,

1 Komukai-Toshiba-cho, Saiwai-ku, Kawasaki, Kanagawa 212-8582 JAPAN

E-mail: rie.katsuki@toshiba.co.jp

***The University of Tokyo,

7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656 JAPAN

****DENSO WAVE INCORPORATED,

1-1 Showa-cho, Kariya, Aichi 448-8661 JAPAN

Abstract

To measure a 3D pose of the objects, we propose an artificial mark attached to objects. We propose three differently shaped marks, all of which have four point features. To measure the marks, the image processing system extracts the point features using edge extraction, erosion, and color extraction. Through measurement experiments, we compared error characteristic of these marks. The results showed that color extraction, which uses the simplest image processing method, was most robust when a mark was measured from various poses (mean of error: 10.4deg, mean of std. dev.: 7.06deg, worst std. dev.: 24.7deg) or was attached to the surface. Moreover, the mark enabled the shortest measurement time (0.478s/6marks). The usefulness of the mark with color extraction was verified through the success of an experiment in which a manipulator handled three marked objects.

Key words: Artificial Mark, Image Processing, Measurement, Manipulator

1 Introduction

Because the measurement of the 3D position and orientation (pose) of various objects is important for handling by a manipulator, many pose-measurement methods have been proposed.⁽¹⁾ Although the methods have shown good results, they require model registration of the objects, which can be troublesome, and a highly-functional image processing system, which is expensive. Therefore, methods that measure marks attached on objects have been proposed.⁽²⁾

However, a mark whose 3D pose can be measured by a simple image processing system and that can be attached to various objects has not been proposed. Many marks for navigation by mobile robots⁽³⁾⁽⁴⁾ can show only a 2D pose. A mark with a moiré-pattern⁽⁵⁾ can show a 3D pose; however, the mark is large because it includes an electric power supply for an LED. Aoyagi et al. proposed a mark⁽⁶⁾ with four barcodes on the edge of a dish. They measured the 3D pose of the dish by determining the center of gravity of the barcodes. Such a mark can be attached to symmetrical objects only, however. Hada et al. proposed a mark consisting of three circles and one barcode,⁽⁷⁾ but three cameras are needed to measure the mark. If a manipulator equips a hand-eye system with the three cameras, its size increases.

In this study, we propose a mark that can be attached to objects and whose 3D pose can be measured by monocular vision. The accuracy of pose measurement depends on

*Received 28 Dec., 2006 (No. T-04-0852)
Japanese Original : Trans. Jpn. Soc. Mech.
Eng., Vol.71, No.706, C (2005),
pp.1927-1936 (Received 22 July, 2004)
[DOI: 10.1299/jamdsm.1.155]

accurately extracting the feature on the mark. Therefore, in Section 2, three marks with different features are proposed. As described in Section 3, the poses of the three marks were measured, and the best mark was then selected according to the accuracy of its measurement. In Section 4, the results of the measurements of the best mark are analyzed through error analysis. In Section 5, an experiment is described which verified the practicality of the mark. In this experiment, a manipulator with a hand-eye system manipulated three marked objects. Conclusions are given in Section 6.

2 Three Proposed Marks

2.1 Usage of the Mark

Our objective was to design a mark that can be measured from various places and put on various objects. To meet this objective, we set up the following situation:

Environment: Home or office

Manipulator: One fixed manipulator with a hand-eye

Task: Transportation of objects

In a home or office environment, it is difficult to adjust the lighting condition for measurement of a mark. The manipulator must handle many kinds of objects, but the objects must be small enough to be grasped and transported. The hand-eye system uses a CCD camera with a zoom lens to search for the mark.

2.2 Specifications for the Mark

The specifications for the mark are listed in Table 1. The coordinate systems used in the table are defined in Fig. 3. d is the distance between the camera's principal point and the center of the mark. r_x , r_y , and r_z are the rotations from Σ_c to Σ_m .

The value of "maximum d " is the distance at which the camera does not zoom in. When the camera measures a mark from a distance over 300[mm], it can zoom in. The value of maximum r_x or r_y is the angle at which the image processing system can detect one or more marks when the marks are attached to each face of a rectangular solid. The values of maximum d , r_x , r_y , and r_z are the tolerable errors when a hand corrects its grasp width using touch sensors.

Table 1 Specifications

Item	Specification
Maximum d [mm]	300
Maximum r_x or r_y [deg]	± 60
Error allowance of d [mm]	± 15
Error allowance of r_x [deg]	± 30
Error allowance of r_y [deg]	± 30
Error allowance of r_z [deg]	± 30

2.3 Requirements for the Mark Design

Attaching marks to objects in a home or office environment has two challenges: customizing the environment for easy detection of the marks and dealing with the many objects necessary for a transportation task. Therefore the following two points are noted: i) The background of the mark is not uniform, and ii) the camera on the manipulator observes some marks that are attached to objects.

Based on i and ii, four requirements are set:

- 1) Speed of measurement: Around 0.5[sec]. After the mark is measured, the manipulator begins its work. The 0.5[sec] pause of the manipulator should not irritate the user of the manipulator. In addition, to hasten measurement, only one image is used.
- 2) Ease of finding: The mark can be extracted from a non-uniform environment.
- 3) Ease of attachment: The mark can be attached to many kinds of objects.
- 4) Ease of identification: The marks can be identified when an image includes plural marks.

2.4 Guidelines for the Design of the Mark

In response to the above four requirements, the following guidelines are defined.

To meet 1), both a simple method for the extraction of the features of the marks and a simple algorithm for the calculation of the pose of the marks were applied. The typical features are point,⁽⁸⁾ line,⁽⁹⁾ and angle.⁽¹⁰⁾ We selected **point features** because of their simple shape. To measure a 3D pose, monocular vision requires at least four point features,⁽¹¹⁾ so the number of point features for the mark is four. In addition, the point features are set on four vertexes on a square. To find a movable object, the camera must measure the poses of the marks regardless of their location. Therefore, a uniform change in the shape of the mark is required when the mark is measured from various viewpoints. A square, which changes its shape uniformly, is thus a suitable shape for the mark.

A modified version of Takahashi's algorithm⁽¹²⁾ was used to calculate the 3D pose of the mark. The algorithm at first calculates the orientation (normal vector) of the mark, and then the orientation is used to calculate the position (a vector from a camera's principal point and the center of the mark). To calculate the orientation, the algorithm gets the 3D positions of the four vertexes on a square, which is projected on an image plane. Next, it calculates the edges and diagonals of the square using the vertexes and then calculates five inner products using four edges and one diagonal. The normal vector of the mark is calculated. The vector satisfies a minimum square sum of each inner product. It computes the minimum value using a downhill simplex method. This method is robust against error and has a high convergence speed.

To determine the position, the algorithm calculates two kinds of normalized vectors. One is from the camera's principal point to the vertexes on the image plane, and the other is from the camera principal point to the real vertexes. To calculate the latter vectors, the algorithm uses a normal vector of the mark. The algorithm then subtracts the former vectors from the latter vectors. The subtractions are done for each vertex. Next, the algorithm calculates a square sum of the x , y , and z components of the subtracted vectors. It then calculates the square sum with a minimum value using the downhill simplex method. Finally, the algorithm identifies the vector that satisfies the minimum square sum as position vector.

Two extraction methods are proposed in order to meet 2): An extraction of colors, an extraction of monochrome patterns. When an image processing system extracts monochrome patterns from a non-uniform background, the processor must process complex images, creating the risk of an overrun of the measurement time dictated by 1); therefore, for the extraction, we apply **rare colors** in a home or office environment.

To meet 3), we use a **flat and flexible material** such as paper.

For 4), we apply a **mark with one region that includes the four point features**.

2.5 Proposal of Three Marks

Table 2 shows the three marks that satisfy the guidelines. The marks have different methods for the extraction of their point features.

- (a) Side Extraction: The features are extracted from four intersections of edges on the outer side of the square.⁽¹³⁾
- (b) Erosion: The features are extracted by erosion operation of image processing system.
- (c) Color Extraction: The features are extracted by color extraction.

Table 2 Proposed marks

Name	Side Extraction	Erosion	Color Extraction
Image
Point features	4 vertexes of the square	4 circles on the corners	4 circles inside the mark
How to extract features

(a) Side Extraction

(b) Erosion

Fig. 1 Image processing algorithm for each mark(1/2)

Figure 1 and Fig. 2 contain flowcharts for each extraction method. At step 8 in Side Extraction, M-estimation is adopted for the estimation using image data of the edges, including errors. Because there are few edges with large errors, Hough transformation is not applied. The calculation time for measuring the mark using M-estimation is short. Six samples were used for the one M-estimation.⁽¹⁴⁾

Because the erosion operation sometimes deforms the four circles, the algorithm of

Erosion is used to reduce the errors of the positions of the circles. In addition, the pose of Erosion has to be measured within 0.5[sec], as defined in 1). To reduce the errors and to satisfy 1), we apply calculations of the gravities of the point features after the erosion operation. To strongly reduce the error resulting from the erosion operation, we discussed the re-estimation of the features using both Hough transformation and the least-squares method after obtaining temporal features by the above calculation of gravities. However, because of the risk involved in using many steps that require more than 0.5[sec], we do not apply the re-estimation.

We chose circles for the point features in Erosion and Color Extraction because a circular image changes its appearance uniformly (i.e., ellipse) even when a camera captures it from various viewpoints. To calculate the orientation around z_m in Fig. 3, we added a pink circle for Side Extraction, a pink region for Erosion, and a yellow-green circle for Color Extraction. To calculate their orientations, an image-processing system requires a base point feature. The system recognizes the nearest point feature from the pink region as the base point. Regarding the pink region of Erosion, to accurately identify the nearest point feature, the region fits the nearest point feature. In the case of Color Extraction, the yellow-green circle is the base point feature.

To determine the colors of the marks, we pick up colors on the edges of the YUV coordinate system; then we choose colors do not exist in our laboratory. To separate the marks and their backgrounds with cyan, we frame the marks in white. Paper is used for the marks because it is inexpensive and flexible. We define 30[mm] as the size of the white frame to attach the mark to a small object. In addition, we add white squares at the center of Side Extraction and Color Extraction. To identify the mark, storage media is placed on the square.

3 Measurement Experiments

We checked the measurement accuracies of the three marks.

3.1 Experimental Setup

The experimental system consists of a color CCD camera (SONY FCB-1X10), an image processing system (Hitachi IP5005), and a computer (CPU: Intel Pentium III 1GHz). The size of the image plane is 512×440 [pixels]. Only fluorescent ceiling lights are used, and the illuminance around the experimental system is 600[lx].

In Fig. 3, the camera coordinate systems Σ_c , image coordinate system Σ_i , and mark

(c) Color Extraction

**Fig. 2 Image processing algorithm for
each mark(2/2)**

coordinate systems Σ_m are defined. The point feature P is projected on P_i in the image plane, and then the coordinate value of $P_i(u, v)$ is transferred to camera coordinate systems: (x, y, z) . The transformation uses formulas (1) and (2).

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{u}{s_x} - c_x \\ -\left(\frac{u}{s_y} - c_y\right) \end{pmatrix} \quad (1)$$

and

$$z = f \quad (2),$$

where f is the focal length, (c_x, c_y) is the center of the image plane, and (s_x, s_y) is the scale factor. These parameters are computed using Halcon image processing software.⁽¹⁶⁾ Image distortion is not corrected because the camera captures a mark at the center of the camera and because the ratio of an area within the mark to the lens plane (768×494 [pixels]) is small: 5.3%. (x, y, z) of all point features are input in the pose calculation algorithm described in Section 2.4, and the mark's pose in the camera coordinate system is then calculated. The dimensions of output are [mm]:position, [deg]:orientation with a roll-pitch-yaw angle.

Fig. 3
Coordinate systems

3.2 Measurement Results for Changing d and r_{ym}

We attached a mark to a white plane and changed d and r_{ym} . The experimental system then measured the pose (d, r_x, r_y, r_z) of the mark. The other parameters, x, y , and z , are not shown because they were calculated with the vectorized d . The changing regions are d : 150–300 [mm] and r_{ym} : 0–60 [deg]; their pitches are d : 50 [mm] and r_{ym} : 20 [deg]. We kept r_{xm} and r_{zm} 0 [deg] throughout the measurements. The ratios of the area of the mark to the plane of the image area are 10.5 [%] at $(d, r_{xm}, r_{ym}, r_{zm}) = (150, 0, 0, 0)$ and 1.31 [%] at $(d, r_{xm}, r_{ym}, r_{zm}) = (300, 0, 60, 0)$. We calculated the absolute values of the error averages and standard deviations of measurement results using 10 measurement poses.

The measurement results are shown in Fig. 4. The hatching indicates acceptable results, namely, the value of absolute error plus double standard deviation satisfies the specifications of r_y in Table 1. The absence of a bar indicates 10 consecutive failures to measure the mark, which was attributed to a failure to extract color.

The area in which the marks can be measured (i.e., number of bars) is small for Erosion but large for Side Extraction and Color Extraction. While almost all standard deviations (i.e., the lengths of the bars) of Side Extraction are smaller than those of other marks, the area that meets the requirement shown in Table 1 (i.e., the number of hatched bars) for Side Extraction is larger than that of Color Extraction. Regarding the trends of the standard

Fig. 4 Measurement results for mark rotation around y_m axis

deviations, there are some high values (i.e., tall bars), such as $(d, r_{ym}) = (250, 60)$ in Side Extraction, $(d, r_{ym}) = (200, 40)$ in Erosion, and $(d, r_{ym}) = (250, 60)$ in Color Extraction. In addition, the standard deviations of $r_{ym} = 20$ in Side Extraction and Color Extraction are large.¹

Regarding d of each mark, the trends of the standard deviations and numbers of hatched bars are similar to the trends of r_y of each mark except for the high standard deviations of r_y at $(d, r_{ym}) = (300, 20)$ in Side Extraction and $(d, r_{ym}) = (200, 20)$ in Color Extraction.

In this experiment, we changed only d and r_{ym} . When we changed only d and r_{xm} , the trends of their standard deviations and the numbers of hatched bars are similar to the case of d and r_{ym} . When we changed only d and r_{zm} with regard to each mark, the standard deviations to the changes of r_{zm} were almost constant because the appearance of the marks did not change although the marks were rotated around r_{zm} .

3.3 Measurement Results for Changing d , r_{xm} , r_{ym} , and r_{zm}

We changed d , r_{xm} , r_{ym} , and r_{zm} of the three marks and then measured the marks. The results at $(d, r_{xm}, r_{ym}, r_{zm}) = (200, -20, 20, 0)$ are shown in Fig. 5 (a). With regard to the standard deviations of r_y for Side Extraction and Erosion, the value shown in Fig. 5 and Fig. 4 at $(d, r_{xm}, r_{ym}, r_{zm}) = (200, 0, 20, 0)$ are quite similar. On the other hand, the value in the case of Color Extraction is 65% lower.

The hatching means that
abs. error + std. dev. $\times 2$ is within 15mm (d), 30deg (r_x, r_y, r_z)

**Fig. 5 Measurement results at $(d, r_{xm}, r_{ym}, r_{zm}) =$
(a) (200, -20, 20, 0), (b) (130, 30, -45, 180)**

The results at $(d, r_{xm}, r_{ym}, r_{zm}) = (130, 30, -45, 180)$ are also shown in Fig. 5 (b). The mark with the least standard deviation is Color Extraction. The standard deviation shown in Fig. 5 for Side Extraction is large, whereas that shown in Fig. 4 is small. In addition, we changed d from 130 to 150 and 200. The image processing system can measure Side Extraction up to 150, Erosion up to 130, and Color Extraction up to 200.

3.4 Discussion of the Measurement Results

The large standard deviations of the three marks have the following three causes: i. Dispersions of coordinate values exist for point features. ii. Large errors in the coordinate values of the point features cause the results of the convergence calculation to vibrate. iii. The coordinate values of the four point features satisfy some poses of a mark.

To validate i, we measured the coordinate values of point features of fixed marks 10 times and then calculated the standard deviations of the coordinate values. The data are shown in Table 3. As shown in the table, the average standard deviations of Erosion and Color Extraction are smaller than Side Extraction. Nakagaki et al. showed that the deviation of the position of the center of gravity of a circle is smaller than the deviation of the position of the intersection of two lines.⁽¹⁷⁾ The standard deviations of point features for Erosion and Color Extraction are, therefore, smaller than that for Side Extraction. However, the accuracy of a pose of a mark also depends on the size of the mark. As shown in Table 3, the standard deviation/square length is smallest for Side Extraction. Because of the large

¹ The standard deviations of $(d, r_{ym}) = (350, 20)$, $(400, 20)$ in Side Extraction are also large.

distance between the two point features, Side Extraction is less influenced by errors the coordinate values of point features. As a result, the standard deviations of the pose for Side Extraction are small as shown in Fig. 4 (a). On the other hand, at $(d, r_{ym}) = (250, 60)$ in Side Extraction, at $(d, r_{ym}) = (200, 40)$ in Erosion, and at $(d, r_{ym}) = (250, 20)$ in Color Extraction, the positions of the point features include large dispersions because of a partial color extraction of the marks when observed from far and steep viewpoints(see Fig. 6 (a)).

To validate ii, we calculated the position errors of point features of the marks. The calculation results are shown in the row labeled “Norm/square length” in Table 3. Norm refers to an average of 10 samples of norms of $(e_x,$

$e_y)$: a position error of a point feature in an image coordinate system. Erosion shows the largest norm/square length. The convergence calculation cannot converge when positions of point features with large errors are input. The large norm/square length for Erosion was caused by the deformation of the four circles by erosion operations, which led to many failures in mark measurement (see Fig. 4 (b)). In the case of a large r_{ym} , the width between the front two circles and the width between the rear two circles are quite different. Therefore, the erosion operation frequently causes the rear circles to vanish before the front circles are extracted. When d is large, the image processing system cannot detect vertical lines and so cannot find the mark. In order to separate the circles and the lines easily, we did not apply thick lines.

As shown in Fig. 4, iii is satisfied by the results for $(d, r_{ym}) = (300, 20)$ in the Side Extraction type and for $(d, r_{ym}) = (200, 20)$ in the Color Extraction type. Regarding $(d, r_{ym}) = (300, 20)$, the size of the mark in an image plane is small; therefore, the moving distances between the point features at $r_{ym} = 0$ and those at $r_{ym} = 20$ are small. On the other hand, regarding $(d, r_{ym}) = (200, 20)$, d is not bigger than $d=300$. Even so, it was difficult for the convergence algorithm to choose the rotative direction(whether r_y is positive or negative) of the mark at $r_{ym} = 20$ because the v directional (see Fig. 3) offsets of the point features were smaller than the u directional offsets. Due to the small offsets, the convergence algorithm converged its answer on $r_y = +20$ or -20 . This alternate convergence did not happen at $(d, r_{xm}, r_{ym}, r_{zm}) = (200, 0, 30, 0)$.

We expected that $(d, r_{ym}) = (300, 0)$ in Side Extraction and $(d, r_{ym}) = (200, 0)$ in Color Extraction would satisfy iii; however, their r_y standard deviations were small. For this reason, we think that the convergence calculation has a local minimum near $r_y = 0$. In addition, regarding $(d, r_{xm}, r_{ym}, r_{zm}) = (200, -20, 20, 0)$ of Color Extraction in Fig. 5, the standard deviation is small even though $r_{ym} = 20$ because of the relatively large offsets of point features in an image plane when the mark is rotated around both r_{xm} and r_{ym} . Moreover, the standard deviations of d , $(d, r_{ym}) = (300, 20)$ in Side Extraction and $(d, r_{ym}) =$

**Table 3 Std. dev. of features at
 $(d, r_x, r_y, r_z) = (150, 0, 0, 0)$**

	Side Ext.	Erosion	Color Ext.
Ave. of std. dev.	0.194	0.165	0.116
Ave. of square length	143	98.2	71.6
Std.dev /square length	0.00135	0.00168	0.00163
Norm/ square length	0.0579	0.0777	0.0576

Unit: pixel

Fig. 6 Images of Side Extraction

(200, 20) in Color Extraction are small because the distances between the camera's principal point and the center of the mark were almost the same when the mark was observed from $r_{ym} = 20$ or $r_{ym} = -20$.

When Side Extraction was measured from an orientation in which r_{xm} , r_{ym} and r_{zm} were not zero, the measurement accuracy was low. On the other hand, at $(d, r_{xm}, r_{ym}, r_{zm}) = (130, 30, -45, 180)$ of Side Extraction in Fig. 5, the standard deviation was high. Moreover, Side Extraction could not be measured from $(d, r_{xm}, r_{ym}, r_{zm}) = (180, 30, -45, 180)$ (see Section 3.3) even though the mark could be measured from $(d, r_{xm}, r_{ym}, r_{zm}) = (300, 0, 40, 0)$ (see Fig. 4 (a)). The reasons for the low accuracies are shown in Fig. 6(b). In this figure, the angles of the corners on the outer square are not 90[deg], and the upper right angle is dented. The image processing system divides adjacent lines from the outer square using the edges, but it is difficult to use the corners to find edges. If the intersection of the lines is calculated using inaccurate lines, high accuracy positions of the point features cannot be attained.

The following is a summary of the points discussed in this subsection. The measurement accuracies for Side Extraction (i.e. the standard deviation and the area within which the mark can be measured) are high when the mark is rotated around only one axis (e.g. only y_m), but the accuracies are low when the mark is rotated around all axes. Erosion operations reduce the accuracy of measurements. Regarding Color Extraction, although the standard deviations when the mark is rotated around only one axis are somewhat lower than those of Side Extraction, the standard deviations when the mark is rotated around all axes are high.

3.5 Measurement Times for Changing the Number of Marks

We attached one kind of mark or marks to a white flat surface, and the experimental system measured a pose of the mark or marks. The surface was set parallel to the camera. The distance between the plane and the camera was 200mm. The number of marks varied from 1 to 6. The measurement times are shown in Fig. 7.

The measurement time of Side Extraction and Erosion increased more steeply than that of Color Extraction because of the increase in the image processing time. In the case of Side Extraction, the time for separation to four lines and M-estimation was increased. In the case of Erosion, the time for erosion operations for some marks was increased. On the other hand, Color Extraction requires only the color extraction operation to detect its point features, and it can extract the colors of multiple marks at one time. Therefore, the measurement times of Color Extraction increased only slightly. The measurement time for 6 marks was below 0.5[sec], the requirement shown in Table 1.

Fig. 7 Relationship between the mark number and measurement time

3.6 Measurement Results When a Mark is Attached to a Curved Surface

To calculate the measurement accuracy when a mark is attached to a curved surface, we attached a mark to a curved surface with a curvature radius of 26[mm](190[ml] coffee can). The experimental system was used to measure the pose of each mark 10 times $(d, r_{xm}, r_{ym}, r_{zm}) = (200, 0, 0, 0)$ and $(200, 0, -30, 0)$. Although the positions of point features in the image plane were shifted due to the curved surface, we did not correct the positions using a 3D model of the curved surface because the objective of applying of a mark is to omit the

registration of the object's 3D model.

The absolute value of the averages of r_y error and standard deviations of r_y are shown in Fig. 8. The errors and standard deviation for Side Extraction are higher on a surface than on a plane; hence, this mark is weak against a curved surface because of the long distance between two point features. The distance is maximized when $r_{ym} = 0$; thus, the r_y error in $(d, r_{xm}, r_{ym}, r_{zm}) = (200, 0, 0, 0)$ is bigger than the r_y in $(d, r_{xm}, r_{ym}, r_{zm}) = (200, 0, -30, 0)$.

Regarding Erosion, the r_y errors in $(d, r_{xm}, r_{ym}, r_{zm}) = (200, 0, 0, 0)$ on the surface and on a plane are similar. However, at $(d, r_{xm}, r_{ym}, r_{zm}) = (200, 0, -30, 0)$ on a surface, the measurement system failed to measure the mark because the system could not detect a vertical line on the rear side.

The errors and standard deviations for Color Extraction on both a plane and a curved surface are low because the distance between point features is short and because the image processing system gets the point features using only extraction of colors of the four circles.

Fig. 8 Relationship between curvatures and measurement results

3.7 Selection of a Mark for a Manipulation Task

The measurement accuracy of a mark that is applied for a manipulation task has to satisfy the requirement shown in Table 1, and the accuracy has to be stable for various attachment conditions, such as on a curved surface or with multiple marks. From the above standpoint, considering Color Extraction to be the best mark, we used Color Extraction for error analysis and a manipulation task.

Side Extraction seems to be suitable when the mark is rotated around only one axis or attached on a plane and when one or a few marks are measured (e.g., landmark for robot self-localization). Erosion is inappropriate for a manipulation task because its measurement accuracy is low.

4 Error Analysis

To verify the above measurement results of Color Extraction, a simulator was used to re-create position errors of point features and then to calculate the pose of mark. Finally, we compared simulated mark measurement errors and the experimental measurement errors. The simulator changed d and r_{ym} , and then calculated its pose 10^3 times. To create the position error, the simulator added the errors that follow a 2-dimensional normal distribution ($N_{ui} = (u_i, \sigma^2)$, $N_{vi} = (v_i, \sigma^2)$) on the position $((u_i, v_i)[\text{pixel}, \text{pixel}])$ of four point features ($P_i, i=1$ to 4). To calculate the pose of the mark, we applied Takahashi's algorithm, which is the same as the measurement experiments discussed in Section 3.

To determine (u_i, v_i) , consideration of the steady-state error is important. To estimate the steady-state error, we measured the position errors of point features when the mark was observed from certain viewpoints. However because we could estimate only unsteady-state errors, we did not include a steady-state error in (u_i, v_i) .

To determine the value of σ , we calculated the standard deviations of positions of point features. However, the values of the deviations were random. We defined $\sigma = 0.16[\text{pixel}]$. To find 0.16, we referred to the information in Section 3.4. First, we supposed that v_i of a point feature on the front upper side is always smaller than that on the rear upper side when a mark is measured from $(d, r_{xm}, r_{ym}, r_{zm}) = (200, 0, 30, 0)$; thus, the convergence calculation does not converge on both $r_y = +r_y$ and $-r_y$. Then we supposed that the error of (u_i, v_i) was within $0.47[\text{pixel}]$, which is half the distance between v_i on the front upper side and v_i on the

rear upper side.

The standard deviations of r_y through the simulation are shown in Fig. 9. To completely model (u_i, v_i) and σ is quite hard because (u_i, v_i) and σ include various kinds of errors, for example, a color extraction error and a position error of a mark's location. Despite this, the simulation data and the experimental data (Fig. 4 (c)) have some similarities. For example, the orders of both data sets are the same, the standard deviations of both data sets are high at $r_{ym} = 20$, and standard deviations of both data sets at $r_{ym} = 40$ increase as d increases. Therefore, the simulated data validate the experimental results. In addition, the simulator was able to reproduce high standard deviations such as $(d, r_{ym}) = (200, 20)$ when values of (u_i, v_i) were the actual measurement coordinate values and $\sigma = 0.16$.

Fig. 9 Simulation results of Color Extraction

5 Manipulation Experiment

To verify the feasibility of a manipulation task with Color Extraction, an experiment in which a manipulator transferred three objects on a table was performed.

5.1 Experimental Setup

We used a 6-d.o.f manipulator (AM-60A0D, Denso Wave Inc.) with a handmade parallel gripper (open width: 5-65 [mm]). For safety, we set the speed of the manipulator at 10 [mm/s]. We added the mark measurement system described in Section 3 to the manipulator. The CCD camera, a component of the mark measurement system, was settled on the gripper as shown in Fig. 10.

Figure 11 shows the experimental environment. Through the marks, the manipulator measures the poses of the objects on a table. It then transfers the objects to specified locations. The manipulator does not re-grasp with a different grasp pose during the transfers. We put the objects on the table in random position.

Fig. 10 Hand-eye system

To validate the marks' applications for soft objects, for objects without flat surface, and for objects with a curved surface, we applied a pen case (size: 204×50×80 [mm], width of grasped region: 25 [mm]), a basket (radius: ϕ164 [mm], height: 210 [mm], width of grasped region: 12 [mm]), and a flashlight (radius: ϕ64 [mm], height: 168 [mm], width of grasped region: 40 [mm]).

Fig. 11 Experimental environment

We attached the mark to each object. To identify the objects through the mark, we added a QR code, a kind of 2-dimensional barcode, to the center of the mark, and then we stored the ID in the QR code. We applied the QR code from various memory storages, such as RFID, for four reasons: i. The QR code decoder can decode only the QR code that a camera is capturing. ii. The decoder can decode a QR code even though the QR code is rotated around a normal line from the plane of the QR code. iii. When the decoder uses a camera with a zoom function, it can decode QR codes from afar.

iv. We can input ID and print a QR code easily. In addition, we input information(e.g., the position of a mark, the grasp pose, and the storage location of an object) in a database. We then matched the information and the ID. The manipulator could get the information through the ID and could manipulate an object using the information. We used a TL-600(KEYENCE Corp.) for the QR code decoder.

For successful mark detection and measurement, the location of a mark on an object is important. We determined the pose of a mark on a surface with a sufficient area to attach a 30[mm] mark and on a surface that is close to a plane. The exact location was determined by sight. The differences between the locations in the database and the real locations were within ± 3 [mm]: translational component and ± 5 [deg]: rotational component. The differences were manually measured and were not corrected.

To correct the difference, the use of an auto-measurement system is effective. The system captures and shows an object with a mark. We indicate some points on edges of the mark and the object. Then the system is used to measure the pose of the mark on an object coordinate system. In addition, the method that calculates a 3D position using 2D points in a camera image including an object was proposed by Saito et al.⁽¹⁸⁾

In order to attach a mark with Scotch tape, we widened the outer white region of the mark to 5[mm] at one edge. The wide region did not affect the results of the measurement of the mark.

5.2 Mark Measurement Procedure

For a manipulation task to be successful, a mark must be measured with high accuracy. On the other hand, real measurement accuracy is dispersed because the marks are measured from various viewpoints. For stable accuracy, we used the mark measurement procedure described below.

(1) Search: The manipulator moves the camera parallel to the width of the table; it then measures the number and poses of marks on the table(Fig. 12). The observation points of the camera are determined manually. At these points, the camera can capture the surface on the table. In addition, the manipulator knows the width of the table and the pose of the table.

Fig. 12 Mark Search

(2) Measurement: The manipulator moves the camera to a specific viewpoint, and the mark measurement system measures a mark (Fig. 13). The measurement experiment described in Section 3 revealed that highly accurate viewpoints were in front of or near the mark. Therefore, for the specific viewpoint, $(x, y, z, r_x, r_y, r_z) = (250, 0, 0, 0, 0, 0)$, which is a pose between the mark and the camera. From this viewpoint, the hand does not collide with a target object. After the manipulator moves to the specific viewpoint, the mark measurement system zooms in to 4 $\times$, and the size of the mark is quadrupled. The 4 $\times$ zoom is determined manually. After the 4 $\times$ zoom, the camera focuses automatically on the mark. In addition, the mark measurement system measures the mark five times, and then it calculates the average of the five data. The manipulator stops after it measures all marks.

(a) Move to mark center

(b) Zoom

Fig. 13 Mark Measurement

5.3 Results of the Manipulation Experiment

The manipulator tried to transfer the three objects, as shown in Fig. 14, and we recorded the number of successes and the execution time. When the manipulator transferred three objects in one trial, we considered the trial successful. When the manipulator failed, we quit the experiment even if there were objects remaining. Six manipulation experiments were conducted, and four were successful. In the two failed experiments, the manipulator failed to grasp a basket in one case and a flashlight in the other case. The failures resulted from measurement errors involving the mark. The measurement errors were caused by search failures. During Search, the orientation between the camera and a mark was very steep; thus, the mark measurement system output an inaccurate pose. During Measurement, the manipulator changed its viewpoint derived from the inaccurate pose. The viewpoint was not in front of the mark, and it caused an inaccurate measurement of the mark. To avoid such an inaccuracy, attaching some marks to one object⁽¹⁹⁾ is effective because the measurement system can select an accurate pose.

(a) Pen case grasping (b) Flashlight grasping

Fig. 14 Experimental results

The average task execution time for the three objects was 16[min] and 3[sec]. The breakdown of the time is as follows: the search time on the table was 1[min] and 32[sec] for each Search, and the time from Measurement to the placement of an object at a specific location was 2[min] and 29[sec] for each object. The remainder of the time was spent moving the manipulator from the specific location to the home position. The breakdown of the above 2[min] and 29[sec] was follows: moving time of the manipulator and the gripper, 2[min] and 14[sec]; waiting for focusing after zooming in and for the synchronization of serial communications, 13[sec]; and decoding time for QR code, 0.05[sec]. The mark measurement time was 2.0[sec], the sum of five mark measurements. The measurement time per mark was 0.4[sec]. A considerable amount of time was required to move the manipulator; therefore, the transfer time could be shortened by increasing the manipulator speed.

From the above experiment, we validated the usability of manipulation tasks with Color Extraction. During this experiment, objects were placed only on a table. However, in a real manipulation in which locations of objects can change, dynamic active search⁽²⁰⁾ is useful.

6 Conclusion

In this study of the measurement of object poses using monocular vision, we designed and attached to objects these three marks: Side Extraction, Erosion, and Color Extraction. All marks have four point features. The differences in the marks are the image processing used to extract the features: edge extraction(Side Extraction), erosion(Erosion), and color extraction(Color Extraction). The error characteristics of these marks were compared in measurement experiments. The results of the comparison are described below.

- Regarding Side Extraction, high accuracy was achieved when the mark was attached to a plane surface because the distances between point features were the longest of the all marks. On the other hand, the mark had low accuracy when the mark was attached to a curved surface. Furthermore, the mark resulted in inaccuracy when it was rotated around all axes of a mark coordinate system because it was difficult to detect corners. The measurement time increased in proportion to the number of marks because the algorithm for detecting point features involves many steps.

- Regarding Erosion, the disadvantages were the inaccurate determination of the positions of point features and the lack of success in extracting the features.
- Color Extraction produced fewer inaccuracies than Side Extraction because the distance between two point features was shorter than that of Side Extraction. The r_y accuracies of Color Extraction were as follows: Mean of error: 10.4[deg]; mean of standard deviation: 7.06[deg]; worst standard deviation: 24.7[deg]. On the other hand, the mark had high accuracy when it was attached to a curved surface and when it was rotated around all axes of the coordinate system. Six marks required less time (0.478[sec]) to measure because the extraction method for point features, which is limited to the extraction of colors, is simple.
- We defined the usage of a mark as object recognition for a manipulation task in a home or office. For this task, the manipulator handles many objects at various locations. Color Extraction worked best because multiple marks are measured the fastest and because the mark has high accuracy when the mark is rotated around all axes. When we applied Color Extraction to the manipulation task, the manipulator succeeded in four of six attempts. Therefore, we validated the use of Color Extraction for the manipulation tasks.

For the practical application of Color Extraction, robustness against a change of lighting conditions is important. Two methods are useful for achieving robustness. One is a change in the YUV value for color detection corresponding to lighting conditions; the other is irradiation of the light to a mark to maintain a specific lighting condition. The methods allow a mark measurement system to measure a mark at different times and places.⁽²¹⁾

Small and invisible Color Extraction is also desired. The size of a mark depends largely on the size of the QR code. The minimum size of the QR code is 2.55[mm] only when an ID with five characters is recorded. In the case using 2.55[mm], the minimum size of Color Extraction will be 6.95[mm] on a side. When the mark is smaller, it is difficult to measure. To solve this problem, the use of a dynamic active search⁽²⁰⁾ and the zoom function is effective. On the other hand, to make an invisible mark, the application of paint that glows in black light is effective. The use of small and invisible Color Extraction is earmarked for future research.

References

- (1) Sumi, Y., Kawai, Y., Yoshimi, T. and Tomita, F., Recognition of 3D Free-Form Objects Using Segment-Based Stereo Vision, *Proc. Sixth Int'l Conf. on Computer Vision, ICCV'98*(1998), pp.668-674.
- (2) Ota, J., Service Robot as an Intermediate Agent among Functional Objects, *AI*IA 98*(1998), pp.259-264.
- (3) Feng, L., Fainman, Y. and Koren, Y., Estimation of the Absolute Position of Mobile Systems by an Optoelectronic Processor, *IEEE Trans. Syst. Man. Cybern.*, Vol. 22, No.5 (1992), pp.953-963.
- (4) Briggs, A. J. and Scharstein, D., Mobile Robot Navigation Using Self-Similar Landmarks, *Proc. IEEE Int. Conf. Robotics & Autom.*(2000), pp.1428-1434.
- (5) Armstrong, B., Verron, T., Reynolds, L. and Schmidt, K., RGR-3D: Simple, Cheap Detection of 6-DOF Pose for Tele-Operation, and Robot Programming and Calibration, *Proc. IEEE Int. Conf. Robotics & Autom.* (2002), pp.2938-2943.
- (6) Aoyagi, S., Kinomoto, H., Ieuji, S. and Takano, M., Recognition and Handling of Tableware by Using a Robot Based on RECS Concept, *T.IEE Japan Vol.120-C*, No.5(2000), pp.615-624.(in Japanese)
- (7) Hada, Y., Takase, K., and Ayusawa, T., Development of a Mark-based 3D Vision System, *T.IEE Japan Vol.120-C*, No.5(2000), pp.625-633.(in Japanese)

- (8) Abidi, M.A., and Chandra, T., Pose Estimation for Camera Calibration and Landmark Tracking, *Proc. IEEE Int. Conf. Robotics & Autom.*(1990), pp.420–426.
- (9) Chen, H. H., Pose Determination from Line-to-Plane Correspondence: Existence Condition and Closed-Form Solutions, *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 13, No.6(1991), pp.530–541.
- (10) Shakunaga, T. and Kaneko, H., Perspective Angle Transform: Principle of Shape from Angles, *Int. J. of Comput. Vision*, 3(1989), pp.239–254.
- (11) Lin, Y.C., Measurement and Design of Artificial Landmark, *doctoral dissertation*, The Univ. of Tokyo(1996).
- (12) Takahashi, A., Ishii, I., Makino, H., and Nakashizuka, M., A High Accuracy Realtime 3D Measuring Method of Marker for VR Interface by Monocular Vision, *Proc. 3D Image Conference'96*(1996), pp.167-172.(in Japanese)
- (13) Ota, J., Yamamoto, M., Ikeda, K., Aiyama, Y., and Arai, T., Environmental Support Method for Mobile Robots Using Visual Marks with Memory Storage, *Proc. 1999 IEEE Int. Conf. Robotics and Automat.*, 4(1999), pp.2976-2981.
- (14) Nakagawa, T., and Koyanagi, Y., Analysis of Experimental Data Using the Least-Squares Method, University of Tokyo Press(1982), pp.164–168.(in Japanese)
- (15) Watanabe, T., and Shibata, T., Detection of Broken Ellipses by the Hough Transforms and Multiresolutional Images, *IEICE Trans.* Vol. J73-D-II, No.2(1990), pp.159-166.(in Japanese)
- (16) FEST Project, Practices for Image Processing, Springer-Verlag Tokyo Inc(2000), pp. 216-222.(in Japanese)
- (17) Nakagaki, H., Takahashi H., Suehiro, T. and Ogasawara, T., Hand-Eye System Calibration by using Conic Pattern, *JRSJ* Vol.15, No.5(1997), pp.781–788.(in Japanese)
- (18) Saito, F., and Suehiro, T., Toward Telemanipulation via 2-D Interface — Concept and First Result of “Titi” —, *Proc. of IECON–2002*, Vol.3(2002), pp.2243–2248.
- (19) Katsuki, R., Ota, J., Tamura, Y., Mizuta, T., Kito, T., Arai, T., Ueyama, T., and Nishiyama, T., Handling of Objects with Marks by a Robot, *Proc. 2003 IEEE/RSJ Int. Conf. Intell. Robots and Systems* (2003), pp. 130-135.
- (20) Kawanishi T., Murase, H., Takagi, S. and Werner, M., Dynamic Active Search for Quick Object Detection with Pan-tilt-zoom Camera, *Proc. IEEE International Conference on Image Processing (ICIP)*, Vol.3(2001), pp.716-719.
- (21) Katsuki, R., Siegwart, R., Ota, J. and Arai, T., Reasoning of Abstract Motion of Target Object through Task Order with Natural Language - As Preknowledge of Object-Handling-Task Programming for Service Robot, *Advanced Robotics*, 20,4, 391/412 (2006).