Proceedings of the 2003 IEEE
International Conference on Robotics & Automation
Taipei, Taiwan, September 14-19, 2003

Design of an Artificial Mark to Determine 3D Pose By Monocular Vision

Rie KATSUKI, Jun OTA, Takahisa MIZUTA, Tomomi KITO, Tamio ARAI

Department of Precision Engineering,
School of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, JAPAN
rie@prince.pe.u-tokyo.ac.jp

Tsuyoshi UEYAMA, Tsuyoshi NISHIYAMA

DENSO WAVE Inc.
1-1, Showa-cho, Kariya, Aichi 448-8661, Japan

Abstract

A design for an artificial mark attached to small objects is presented in this paper. Three conditions are required for the design of the marks:(1) short calculation time, (2) easy attachment, and (3) individual calculation for each item. The marks are obtained by using three different extraction methods. The experiments determine the best mark. The mark using the method that extract colors and calculated them gravity has good results. We analyze the errors of the mark that has best experiment. The appropriateness of the experimental results is then confirmed. We have the experiment that the manipulator handles the objects with the best marks. The experiment is successful. Therefore, the usability of the mark is verified.

Key words: Artificial mark, Image processing, Measurement, Recognition, Design, Manipulator

1 Introduction

It is very important to calculate three-dimensional positions and orientation (3D pose) of targets using information from a CCD camera image. There are many applications that require their use, for example, self-positioning of robots and interface for virtual reality. In one approach, a system detects a mark attached by a man on a target. Accordingly, the system calculates the target's 3D pose more easily than it calculates the target itself. This approach has the following advantages: the system is simple; the image processing algorithm is also simple; and the system for changing the light is robust.

There are many studies dealing with marks supporting the self-positioning of robots[1]; however, for the following reasons, such marks are not used here to calculate target positions: (1) the robot only moves horizontally[2], (2) the robot cannot calculate the mark pose from the front view[3], (3) most of the marks can be calculated in only a 2D pose.

There is a mark using a moiré pattern to calculate a 3D pose[4]. The mark is too big to be attached to an object because it has an LED and a power supply. Aoyagi et al. proposed a mark to calculate the 3D pose of dishes[5]. A problem with their mark is that the system is unable to differentiate one mark from another when a dish has more than one color.

Therefore, a design for an artificial mark attached to small objects is presented in this paper. We aim to achieve a mark whose measuring accuracy has been excellent to various observation conditions. The accuracy depends on how completely the features of a mark can be extracted. Therefore, some marks are proposed that require different methods for extracting features, and the best ones are determined by experiments.

A summary of the contents of this study is outlined in the following. In section 2, the specifications of the marks are detailed, and three kinds of marks that satisfy the specifications are proposed. In section 3, experiments on measuring accuracy are presented, and the results and selection of the best mark are considered. In section 4, the errors of the best mark are analyzed using simulation. In section 5, experiments are reported of a manipulator handling small objects with the best marks, and the usability of the marks is confirmed.

To develop a system that is simple, the marks are calculated with a color image captured by a CCD camera.

2 Proposal of marks

2.1 Specifications and guidelines for marks

A mark has many applications. In this study, we assume that a robot works in home with marks. we deal with robot handling to give an actual example because it is one of the most popular work in home. The robot can handle small objects for dairy use as a juice can. These size are decided for the robot handling: the width of handle position are 10-50[mm], the weight is 500[g] maximum. The mark should attach various surface such as curved surface.

The measuring accuracy is determined as shown in Table 1. Accuracy is determined by signals received from a camera positioned on an end effecter of a manipulator on the robot. Actual robot handling of some objects with pseudo pose error brought accuracy value.

Moreover, it is difficult to extract only shape features of a mark from a monocular image because there are many objects with various shapes in home. In this study, colors that are quite rare in home and easy to extract are adopted.

Other specifications include the following: (1) short calculation time, (2) easy attachment of the mark, and (3) individual calculation of the marks when there are some marks in an image. Generally, the system calculates the features of the marks to get the pose. If the features of a mark consist of n points, the system can determine the pose by solving "perspective n-point (PnP) problem"[6]. The system can also use the features that consist of an n-line[7] or an n-angle[8]. As shown in (1), a short calculation time is important. Features that consist of n points were chosen because the ease with which the algorithm can be calculated. When the system uses a CCD image, it requires at least four points to determine a 3D pose. Therefore, the design of the mark includes features consisting of four points. From (2) and (3), marks that are planes, squares, and closed loops are proposed.

2.2 Three different marks are proposed

Measuring the accuracy of the mark depends on how completely the mark features can be extracted. Three methods are compared below.

Table 1 Specifications

item	specifications
maximum distance [mm]	300
maximum angle [deg]	60
error allowance of distance [mm]	±15
error allowance of rotation around x axis [deg]	±15
error allowance of rotation around y axis [deg]	±15
error allowance of rotation around z axis [deg]	±15
calculation time[s]	0.5

(a) Side Extraction: Four sides of a square are extracted, and then, the points' coordinates are calculated from the intersections of sides[9] (extraction from the sides).

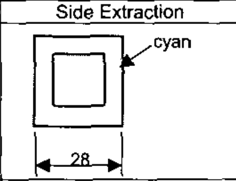
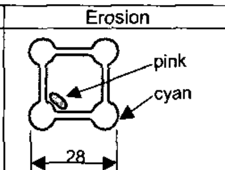
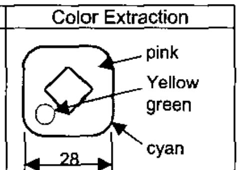
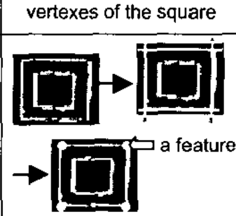
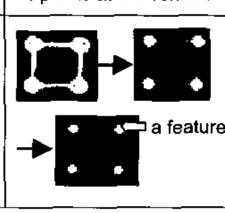
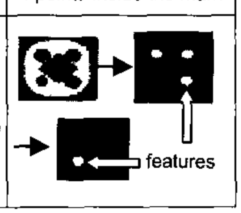
(b) Erosion: Only four points are extracted by eliminating the sides using erosion (extraction using erosion).

(c) Color Extraction: Four points are extracted with the use of color extraction. The colors of the points are different from those of the rest (extraction using color).

The marks that can be obtained using the above three methods are shown in Table 2. (b) and (c) have circular points that remain constant when the angle of the view changes. The mark colors have quite different values in YUV space; furthermore, the colors are not easily influenced by light. The mark (b) has a pink point, and (c) has a yellow-green point. The system is able to calculate the rotation around the normal vector using the points. (a) cannot determine the vector.

The marks sizes are decided by the region where the system can find out the marks and by the size of a place where the mark can be attached. A size of 28mm was determined for small objects. Regular-paper and printed colored marks were used because they may be attached anywhere cheaply and easily. Takahashi's algorithm[10] is used to calculate the 3D pose using four points. The algorithm applies a character so that the adjacent sides and the diagonals intersect perpendicularly. The algorithm has two advantages. One is that the algorithm is hard to have the calculating errors from the errors of 4 points' coordinates. The other is that the algorithm enables a quick calculation despite a convergent operation.

Table 2 Proposed marks

name	Side Extraction	Erosion	Color Extraction
image			
hpoint features	vertexes of the square	4 points at the corners	4 points inside the mark
how to extract features			

3 Experiments

In this section, measuring accuracy is used in experiments to choose the best mark proposed in section 2. First, a 3D pose of parameter estimation is used. Second, for an actual trial, the calculation time when an image has some marks is used, and an experiment on a 3D pose parameter for curvature is examined.

The experimental setup consists of a 1/4-inch color CCD camera SONY FCB-1X10, an image processing board Hitachi IP5005, and a computer with a CPU, Intel Celeron Processor 533MHz. The focal length is fixed at 4.20[mm].

The frame grabber is 512×440[pixel]. The pixel size for each side is 1.30×10^{-3}[mm]. The center of the image is (280, 216)[pixel]. These sizes are determined by the camera calibrations. Image distortion is not considered. Illuminations are fluorescent ceiling lamps. Fig. 1 shows a coordinate system for the experiments.

3.1 Experiments on 3D pose estimations

For the basic error characteristic, 3D pose estimates are used in experiments. A mark is attached to a white plane. The mark then moves along d and rotates around the y axis. d is the distance between the camera and the mark. r_y is the rotation around the mark y axis. The calculation regions are d=150-400[mm] and r_y=0-80[deg]. The pitches are d=50[mm] and r_y=20[deg]. In calculating, r_x and r_z are fixed to 0[deg]. The calculating results of each point are averaged ten times.

The experimental results of r_y are shown in Fig. 2. The squares in the graphs are variances of errors, and the hatchings are the regions in which the results satisfy the specifications of r_y, shown in Table 1. The regions with no squares mean that the system cannot calculate the marks. There are two requirements: (1) the system is unable to extract the point features and (2) the calculating results are varied widely at all times.

The results of the comparisons of the proposed marks are as follows: (a) Side Extraction, r_y variances tend to be bigger, as the value of d or r_y is big. (b) Erosion, the region in which the system can calculate the mark is small. The result at (d, r_y)=(200, 40) has a big variance. (c) Color Extraction, the variances are less than the value of the other marks.

The results of the marks' calculable region show that Side Extraction is the best, Color Extraction, second, and Erosion, third. The hatched regions are nearly identical in the Side and Color Extractions. Meanwhile, the Erosion region is smaller than the others.

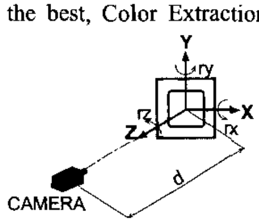


Fig. 1 Coordinate system

The variances in the Color Extraction are minimum 2.39[deg^2], maximum 6.12×10^2[deg^2], and average 7.80 [deg^2]. The absolute errors are minimum 0.760[deg], maximum 5.47×10[deg], and average 0.578[deg]. The results of d are similar to those of r_y in Fig. 2. In these experiments, errors are calculated concerning the change of d and r_y. Moreover, in the case that all pose parameters have errors, the experimental results are similar to those of the experimental results in Fig. 2.

3.2 Consideration of the experiments

In Side Extraction, the error variances are larger when the observation angles and distances are big. This is because the extracted mark outlines have a large error. In Fig. 3 (b), when the observation angle and distance are big, the equations that stand for the square sides have errors. This is because the system cannot extract outlines correctly. The equations make the errors of the intersection coordinates become large. The intersections mean point features. Therefore, the coordinates lead to large calculation errors.

In the Erosion method, the error variances are larger when the observation angles are big. This is because the shapes of the point features are warped by color extraction and erosion. Fig. 4 shows an image of the Erosion method in (d,r_x,r_y,r_z)=(150,30,30,90) after the system extracts the point features. The parts linking the point features with the lines are extracted vaguely at steep observation angles. The system erodes the extracted mark with the parts; the point features are warped from the shapes, such as (b) in Fig. 4. The feature gravities then have errors, and the error variances increase.

In addition, the calculable region in the Erosion method is small. This is because the system cannot extract the lengthwise lines as Fig. 5. This can be explained by the following three reasons: (1) the width of the observed lines is very thin(5[pixel] in Fig. 5) when the observation angle is steep, (2) the edges of the mark are sensitive to the error of color extraction, and

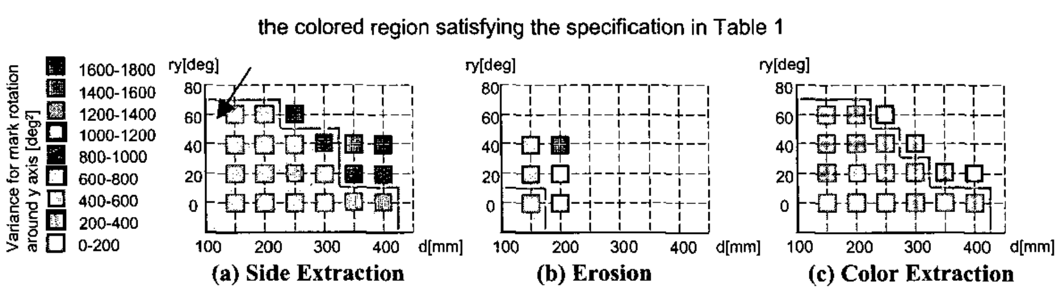


Fig. 2 Experimental results for mark rotation around the y axis

(3) the image processing board recognizes that 2[pixel] is one unit when the image has colors.

On the other hand, the variances of the Color Extraction method are low throughout the experiment. The reasons are that (1) the mark does not use a thin line, (2) the system only extracts colors to get the point features, and (3) in the case of calculating gravity by a circle, the extraction errors of the edges are small. However, the calculable region is smaller than the region in the Side Extraction method. This is because the system cannot extract 4 points from a distance.

3.3 Calculation of some marks

The calculating times for some marks that are 200[mm] parallel away from the camera lens are measured. There are from 1 to 6 marks. The results are shown in Fig. 6.

In the Side Extraction method, calculation times are longer than in the other methods. Despite the fact that the time depends on the number of the algorithm step, this method requires more steps than the others. Therefore, the method does not satisfy 0.5[s] in Table 1. In the Erosion method, when an additional mark is added, a increase in calculation time is the longest. This is because the system erodes each mark individually to keep the point features separate from the other marks. The Color Extraction method takes less than 0.5[s] to calculate six marks all together because the mark has a simple algorithm for image processing.

3.4 Results of attaching to curved surface

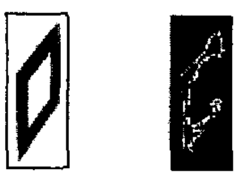
A high degree of accuracy in calculation is required when the marks are attached to a curved surface for actual use. A mark attached on a curved surface is calculated with a radius of 26[mm], from $(d,r_x,r_y,r_z)=(200,0,30,0)$.

Table 3 is a comparison of error variances of a mark on a plane and on a curved surface. In the Side Extraction method for a curved plane, the error variance is larger. The Erosion method on a curved surface cannot be calculated due to the mark extraction. On the other hand, the Color Extraction method has excellent accuracy, which is almost the same as that on the plane. Moreover, the Color Extraction method only satisfies the specifications on d and r_y in Table 1. These results are caused by the phenomenon resulting from the steep observation angle discussed in 3.2.

From these experiments, the accuracy of the Color Extraction method was generally high. Therefore, the Color Extraction method was applied in the following sections.

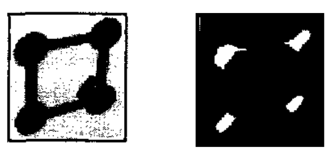
4 Error analysis

The error of the Color Extraction method was analyzed. The algorithm[10] was programmed to calculate a 3D pose using four point features. In the mark calculation, the errors were conjectured by color extraction, change of the discrete quantity, and image deformation to bring the coordinate errors of the point features. The mark calculation was simulated to analyze the influence of the errors. A normal distribution was used to correct the coordinates (x_c, y_c) of each feature. The value is $\sigma=1/6$[pixel], $N_x=(x_c,$



(a) camera image (b) mark outlines

Fig. 3 Mark images using the Side Extraction method $(d, r_x, r_y, r_z)=(200, 35,70,0)$



(a) camera image (b) points after erosion

Fig. 4 Mark images using the Erosion method $(d, r_x, r_y, r_z)=(150, 30,30,90)$

(a) camera image (b) mark after extraction

Fig. 5 Mark images using the Erosion method (d, rx, r_y, r_z)=(250, -40,0,0)

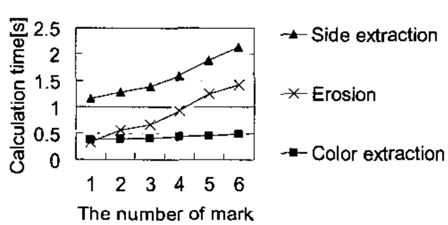


Fig. 6 Relationship between the mark number and the calculation time

Table 3 Comparison of a mark on a plane and a mark on a curvature $(d,r_x,r_y,r_z)=(200,0,30,0)$

		variances		
		Side Extraction	Erosion	Color Extraction
d	for plane	0.356	12.552	37.581
	for a curved surface	2688.749	failure	21.692
r_y	for plane	0.058	1529.098	15.878
	for a curved surface	653.171	failure	39.152

$(1/6)^2$), N_y=(y_c, $(1/6)^2$) for the minimum error. We then simulated 10^3 times. The changed parameters were d, calculation distance and r_y, mark rotation around the y axis, which are defined in section 3.1. The camera parameters are the same as in section 3.1.

Fig. 7 shows the simulation results. The results from the experiments are compared with the models for the error variances of r_y. The models are the simulation results. The parameters are (a) r_y=0[deg], d =150-400[mm] and (b) r_y=40[deg], d =150-250[mm]. The variances are compared when the experiment results satisfy the specifications, as shown in Table 1.

In both r_y=0[deg] and r_y=40[deg], the tendency of the models and the experimental results are almost the same up to 200[mm]. (The values of the experimental results) / (Those of models) are 0.61-1.99. Therefore, the experimental results and the model are appropriate.

However, at over 250[mm], the models are different from the experimental results. This is because the normal distribution models are different from the real error models by reason of the lighting conditions. The larger the calculated distance, the larger the difference. A more complicated and realistic error model would be required if the simulation were made at over 250[mm].

5 Experiment for verification

In one experiment, a manipulator handles objects to verify the usefulness of a proposed mark. The marks are used to recognize objects for the manipulator. The success ratio of handling depends on the calculating accuracy for the mark and the position error attached to the mark. However, in actual use, some situations lead to inaccuracies. Some examples are when a background has many colors, when the marks are not attached to a plane, and when an observation point is bad for calculation results. A mark calculation method was proposed in which a manipulator enabled the accomplishment of the task in such situations.

5.1 Experimental setup

The experiment has a 6-degree-of-freedom manipulator DENSO AM-60A0D, a handmade parallel two-fingered hand (opening width, 5-65[mm]) as a end effecter, and a mark calculation device, as described in section 3, and a barcode reader KEYENCE TL-600 for reading the QR code mentioned below. The manipulator moves 10[mm/s] on the safe side. The CCD camera is set on the hand so that the viewpoint can be changed.

A QR code was placed on the center of the mark so that each mark could be distinguished. A QR code is a kind of two-dimensional barcode. The data of a mark position and the handling position are contained within the QR code. The manipulator can then use the data to manage the handling. The marks are placed on the objects. The attaching errors are within ±3[mm].

Fig. 8 shows the experimental surroundings. The manipulator handles the objects from the table to objects' home position input the mark already. The background is a general room for actual use.

There are three objects for handling: (a) a pen case (size $204\times50\times80$[mm]), (b) a basket (size $\phi164\times210$, width on the grasp point 12), (c) a flashlight (size $\phi64\times168$, width on the grasp point 40). The objects are used for the purposes described below: the pen case is a soft object; the basket has an uneven surface; and the flashlight has a curved surface. They are placed on the table at random.

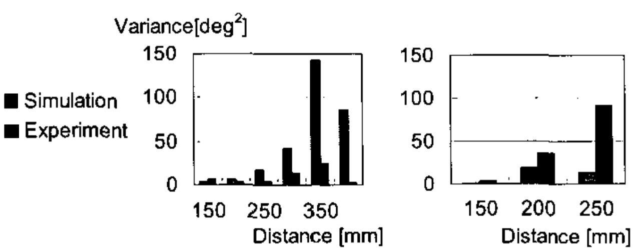
5.2 Mark calculation method

The manipulator calculates the marks to gain high pose accuracy in the following procedures.

(1) SEARCH: to find the marks on the table.

(2) CALCULATION: to move in front of the mark and calculate the mark using a zoom lens. Moreover, an average of five calculations decreases the pose errors.

In SEARCH, the camera is 150[mm] deep, 180[mm] high, and 20[deg] below the table. The camera makes a parallel move along the side of the table. In CALCULATION, the calculating distance is 250[mm], and the zoom value is four.



(a) r_y=0[deg] (b) r_y=40[deg]

Fig. 7 Simulation results

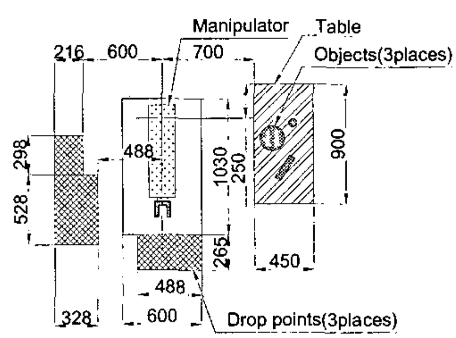


Fig. 8 Experimental environment [mm]

5.3 Experimental results

Fig. 9 shows the experimental progress. The manipulator is (a) mark searching, (b) handling the pen case after handling the basket.

Success is defined when the manipulator puts all the objects away. If the manipulator fails in handling, the handling ceases, even when the task is being processed. Out of six experiments, success was achieved four times. There were two failures (one with the basket and the other with the flashlight). The failure results from the error of mark calculation.

The average time for each task was: (i)16[min] and 3[s] to accomplish the entire task and (ii) 2[min] and 29[s] to do the CALCULATION and put one object away. Table 4 shows a breakdown for (ii). The task time can be shortened when the manipulator moves quickly.

The manipulator was able to put the objects with the marks away. The task time and accuracy were appropriate. Therefore, the usability of the marks is verified.

6 Conclusion

In this study, an artificial mark was designed and attached to small objects to calculate a 3D pose. The extraction method was assumed to be important in order to obtain an accurate pose. Therefore, three proposals for marks were established, and the best one was determined through experimentation.

A mark using Color Extraction had high calculation accuracy in almost all calculable regions. This method extracts the colors of four points and calculates the gravities to obtain point features. On almost all calculation points, the mark had errors of rotation around the y axis at less than ±15[deg]. The error variances were a minimum of.2.385[deg^2] and averaged 77.956[deg^2]. The error of the mark was analyzed, and the appropriateness of the experimental results was confirmed. Finally, a manipulator could do the handling when marks were attached to the objects. Therefore, the usability of the mark was verified.

Measuring occluded marks would be a topic for further research.

Table 4 Analysis of task time

time elements	time[min:s]
manipulator moving	2:14
focus and serial connection	0:13
reading the QRcode	0:001
mark calculation (the sum of 5times)	0:02

(a) mark searching (c) pen case handling

Fig. 9 Experimental results

Acknowledgements

We wish to thank Ms. Yukiyou Iguchi for reading this manuscript and making a number of helpful suggestions.

References

[1] DeSouza GN, Kak AC: "Vision for mobile robot navigation: A survey", IEEE Trans. Pattern Anal. & Machine Intel., Vol. 24, No.2, pp.237-267,2002.

[2] Liqiang Feng, Yeshaiahu Fainman, and Yoram Koren: "Estimation of the Absolute Position of Mobile Systems by an Optoelectronic Processor," IEEE Trans. Syst. Man. Cybern., Vol. 22, No.5, pp.953-963, 1992.

[3] Mansur Kabuka, Alvaro E. Arenas, "Position Verification of a mobile robot using standard pattern," IEEE Trans. Robotics & Autom., Vol.3, No.6, pp.505-516, 1987.

[4] Braian Armstrong, Thomas Verron, Lee Reynolds, Karl Schmidt: "RGR-3D: Simple, Cheap detection of 6-DOF Pose for Tele-Operation, and Robot Programming and Calibration," Proc. IEEE Int. Conf. Robotics & Autom., pp.2938-2943, 2002.

[5] Seiji Aoyagi, Hiroshi Kinomoto, Shinya Ieuji, Masaharu Takano: "Recognition and Handling of Tableware by Using a Robot Based on RECS Concept," T.IEE Japan, Vol.120-C, No.5, pp.615-624, 2000(in Japanese).

[6] R.M. Haralick, C. Lee, K. Ottenberg, and M.Nolle, "Review and analysis of solutions of the three point perspective pose estimation problem," Proc. Computer Vision and Pattern Recognition '91, 1991.

[7] Homer H. Chen: "Pose Determination from Line-to-Plane Correspondence: Existence Condition and Closed-From Solutions," IEEE Trans. Pattern Anal. Machine Intell., Vol. 13, No.6, pp.530-541, 1991.

[8] Takeshi Shakunaga and Hiroshi Kaneko: "Perspec-tive Angle Transform: Principle of Shape from Angles," Int. J. of Comput. Vision, No.3, pp.239-254, 1989.

[9] Jun Ota: "Service Robot as an Intermediate agent among functional objects," AI*IA 98, pp.259-264, 1998.

[10] Akira Takahashi, Ikuo Ishii, Hideo Makino and Makoto Nakashizuka: "A High Accuracy Realtime 3D Measuring Method of Marker for VR Interface by Monocular Vision," 3D Image Conference'96, pp.167-172, 1996(in Japanese).