STUDY ON AN AUTONOMOUS INTELLIGENT CONTROL SYSTEM FOR AN ARM ROBOT WITH SPEECH RECOGNITION/SYNTHESIS THAT DOES NOT REQUIRE TEACHING

Yoshishige Sato
Dept. of Mechanical Engineering,
Tsuruoka National College of Technology
Tsuruoka City, Yamagata Prefecture 997-8511
E-mail: y-sato@tsuruoka-nct.ac.jp

Abstract

This paper describes research conducted on an advanced, autonomous intelligent-type arm robot that does not require teaching, and that can execute spoken work commands and work reports through speech synthesis. This intelligent arm robot has been developed with the ability to recognize and determine the shape, color, and position of an object, and subsequently grasp the object.

Keywords: image process, arm robot, intelligent control, speech recognition

1. Introduction

In recent years, the act of training robot behavior, called teaching, has become necessary in the operation of arm robots. The overseers of production sites have been tasked with this chore, but this research has eliminated the need for teaching by equipping an arm robot with intelligence and constructing a system that can execute spoken work commands and work reports through speech synthesis. This was accomplished by giving the robot the ability to recognize and determine the shape, color, and position of an object, then automatically create control software that allows it to grasp the object. Thus, an intelligent arm robot was developed that can autonomously act.

2. Autonomous Intelligent System for an Arm Robot

2.1 Summary of the autonomous intelligent system

The autonomous intelligent control system for an arm robot equipped with speech recognition/synthesis that does not require teaching is a control system that utilizes spoken human commands so that an arm robot recognizes and determines the shape, color, and position of an object, then automatically moves to grasp the object. Fig. 1 is a block diagram of the overall system, which calculates the hue of a block from image data captured by a C-MOS camera, and sends the data to a PC using an external Bluetooth wireless module. Pixel information is obtained from image data gained from the PC's wireless module, then digitalized. Next, colors are distinguished to determine red, yellow, and blue, and establish the position of the object through image segmentation. Furthermore, a correction program is created to judge whether the object lies in the center (median point) of the image, and the arm robot moves.

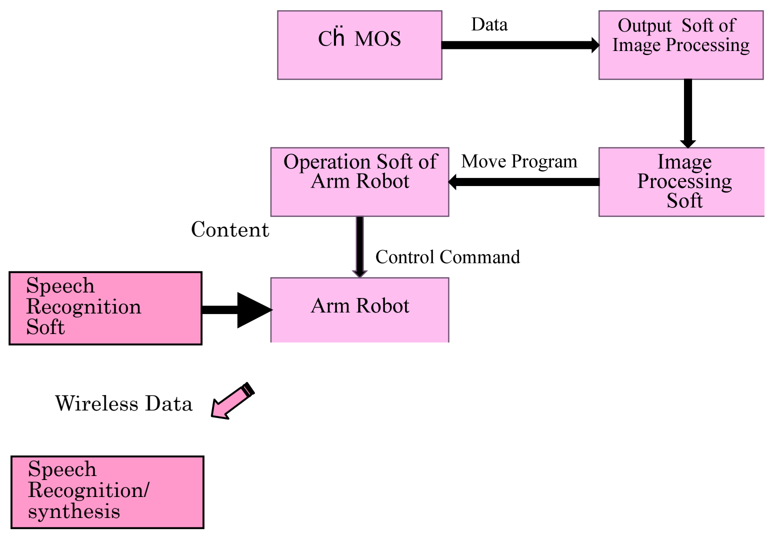

Fig. 1 Block diagram of the overall system

The mounted C-MOS camera is attached at a 25° angle relative to the hand axle so that it can film the tip of the hand. The tip of the arm can be seen in Fig. 2.

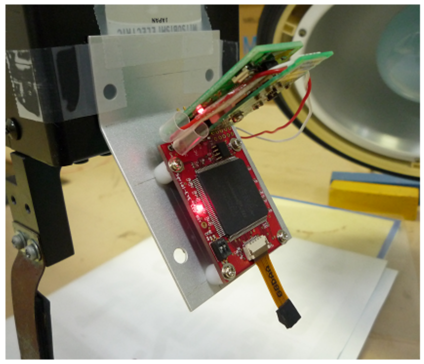

Fig. 2 Tip of the arm robot

2.2 Intelligent Design

2.2.1 Image processing method

The mounted C-MOS camera is attached at a 25° angle relative to the hand axle so that it can film the tip of the hand. An image of the tip of the robot arm is first captured with the C-MOS Eye using wireless Bluetooth, as shown in Fig. 2. The captured image is shown in Fig. 3. The image is dark, so it was reconfigured for a balanced brightness. Fig. 4 is a histogram illustrating the brightness of the initial image. The horizontal axis indicates degree of light and the vertical axis indicates the number of pixels of that brightness.

Fig. 4 reveals that there are a disproportionate number of low-light pixels. The image is altered using a linear density conversion equation (1).

$$Z' = \begin{cases} 0 & (0 \le Z < low) \\ \dfrac{(Z - low)}{(high - low)} & (low \le Z \le high) \\ Z_m & (high < Z \le Z_m) \end{cases} \qquad (1)$$

Z’: brightness post-conversion

Z_m: maximum brightness

Low: minimum brightness in initial image

High: maximum brightness in initial image

Fig. 5 and 6 show the post-conversion histogram and image.

As the converted histogram illustrates, the balance of brightness has clearly improved. Next, areas similar to red in the RGB are extracted from the image using an

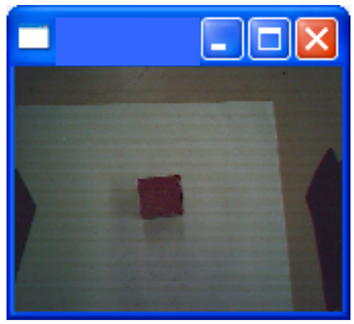

Fig. 3 Initial image

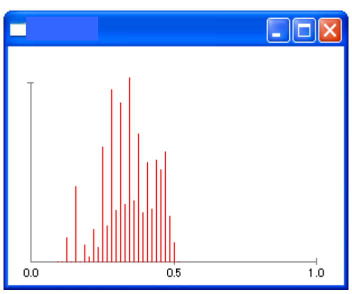

Fig. 4 Histogram of initial image

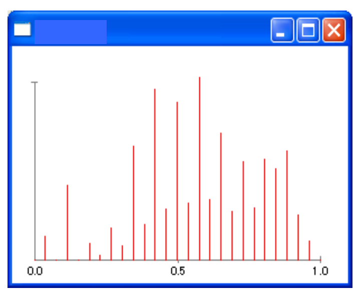

Fig. 5 Post-conversion histogram

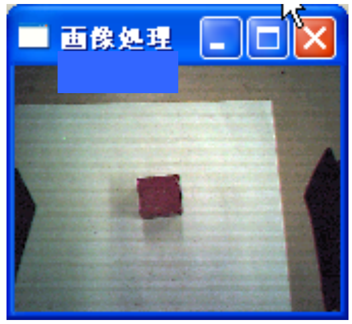

Fig. 6 Post-conversion image

equation for Euclidean distance (2), and the image is then binarized.

$$\sqrt{(R_1-R_2)^2+(G_1-G_2)^2+(B_1-B_2)^2} \le threshold \quad (2)$$

(R_1, G_1, B_1): color information for image

(R_2, G_2, B_2): reference colors

When the value is lower than the threshold it is 1, and when higher it is 0. Fig. 7 shows the image after binarization. Using this image, only the object is detected from the area, which can be found by summing the connecting pixels in the Fig. 7 image. Fig. 8 is the binary image produced after this process takes place.

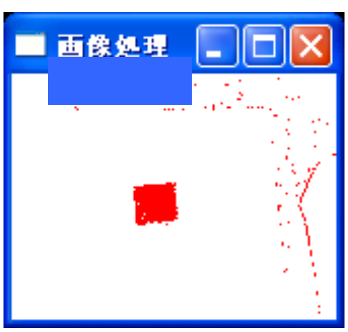

Fig. 7 Euclidean binarization

Fig. 8 Area binarization

2.2.2 Object center & actual position

The object center can be found from the binary image in Fig. 8 using the following equation (3).

$$(cx, cy) = \frac{\sum_{x=0}^{width} \sum_{y=0}^{height} f(x, y)(x, y)}{area} \qquad (3)$$

(cx, cy): object center

$f(x, y)$: value of 0 or 1 in binary image, arbitrary pixels

$(Width, Height)$: image size

area: object area

However, because this is the center of the image, it must be converted to the actual object center, which can be accomplished with the following equation (4).

$$(dx, dy) = \left(\frac{cx - rx}{iWiddth} vWidth, \frac{ry - cy}{iHeight} vHeight\right)$$

(dx, dy): object position relative to the hand position

(cx, cy): object center in image

(rx, ry): object center that ought to be in image

$(iWidth, iHeight)$: image resolution

$(vWidth, vHeight)$: scope captured by camera

With this equation, the object's relative position coordinates (dx, dy) can be found.

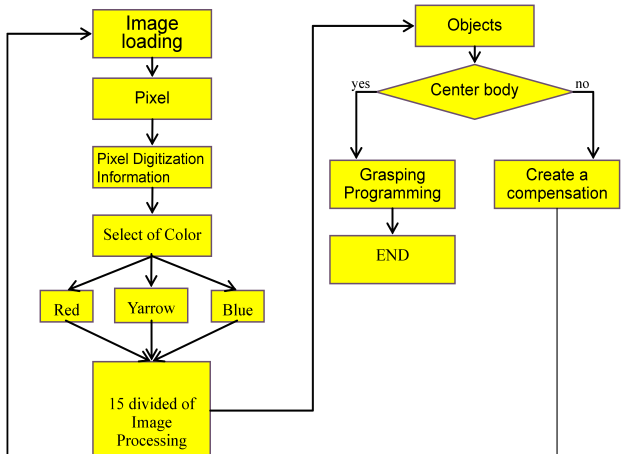

Fig. 9 Robot control flowchart

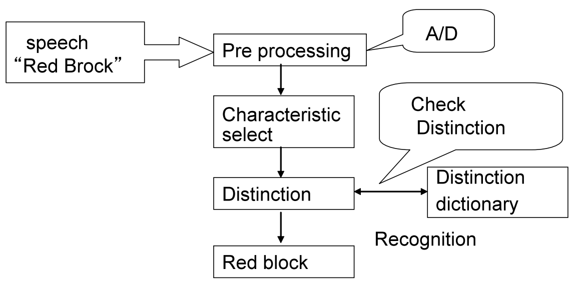

Fig. 10 Speech recognition flowchart

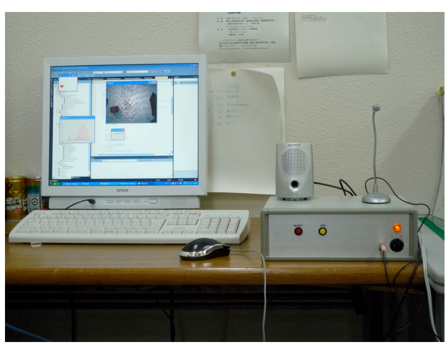

Fig. 11 Computer & speech recognition device

2.2.3 Robot control method

Fig. 9 shows the overall flowchart of robot control. When the operator speaks commands into the computer microphone, based on the speech recognition flowchart illustrated in Fig. 10, the robot recognizes the voice through a unit that utilizes the speaker-independent speech recognition model, Hidden Markov Modeling (HMM) and neural network technology for a faster high recognition rate. Fig. 11 is a picture of the speech recognition device. The robot finds the indicated color block and places that block into a nearby box. The robot recognizes and determines the shape, color, and position of an object, then automatically creates control software that allows it to grasp the object, thereby achieving an intelligent arm robot that can autonomously act.

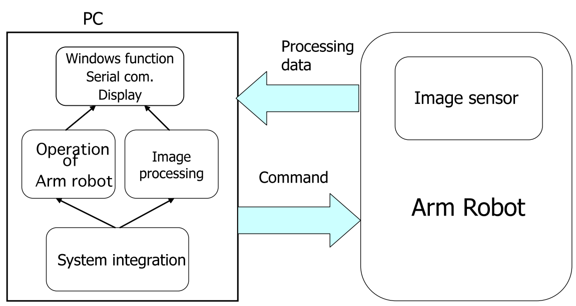

Fig. 12 Block chart for the intelligent control system

The flowchart in Fig. 13 illustrates the algorithm up to the point where the developed robot arm grabs the object.

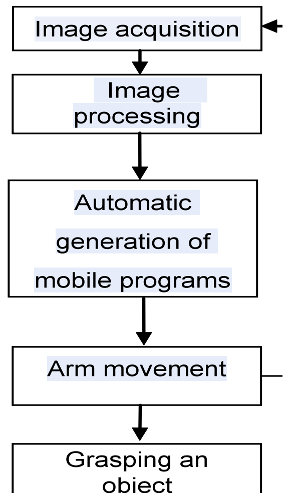

Fig. 13 Basic flowchart for robot control

3. Modeling & implementation of the intelligent control system

3.1Intelligent control system model

The image taken by the image sensor (C-MOS Eye) is sent to the intelligent control software run on the computer. When the image is captured, the Windows serial communications function is used. This function is also utilized for the window displaying communications with the arm robot and image processing results, as well as for graphics processing. The arm robot calculates kinematics and puts together functions that set up necessary commands. It also has an image processing function. The software that carries out the intelligent control system integrates these individual systems. Fig. 12 shows the system block chart.

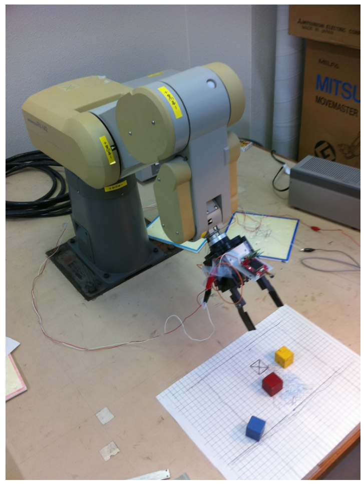

Fig. 14 External view of robot

4. Intelligent control system construction

4.1 Summary of system construction

Equipping the arm robot with a C-MOS image sensor allows real-time image reading so it can recognize object position. The object is then grasped by manipulating the arm robot. There are 3 principal devices and software that carry out this sequence for the intelligent control system—the arm robot, the C-MOS image sensor, and the intelligent control software.

4.2 Arm robot kinematics

The hand position is set as (x, y, z), the length of the link intervals is L_1, L_2, L_3, L_4, L_5, the angle of each link is θ_1, θ_2, θ_3, θ_4, and the attitude angle is the pitch.

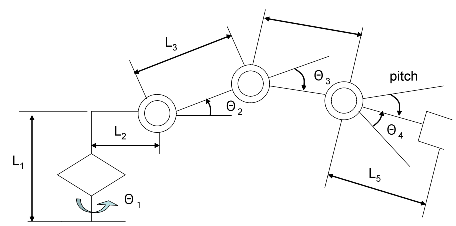

Fig. 15 Robot system

4.3 Direct kinematics equation

In the rectangular coordinate system, when a certain arbitrary coordinate system is rotated to θ on the Z axis, it is written in matrix as equation (5).

$$Rot\ Z(\vartheta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 & 0 \\ \sin(\theta) & \cos(\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (5)$$

In addition, the positional relationship of link intervals can be written in matrix with Denavit Hertenberg notation.

Using DH notation, the matrix expressing the relationship among link intervals becomes equation (6).

$$DH(\vartheta, a, d, \alpha)$$

$$= \begin{pmatrix} \cos(\vartheta) & -\sin(\vartheta)\cos(\alpha) & \sin(\vartheta)\sin(\alpha) & a*\cos(\vartheta) \\ \sin(\vartheta) & \cos(\vartheta)\cos(\alpha) & -\cos(\vartheta)\sin(\alpha) & a*\sin(\vartheta) \\ 0 & \sin(\alpha) & \cos(\alpha) & d \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (6)$$

Using this, solving the arm robot direct kinematics equation gives equation (7), and the left-hand side of x, y, z corresponds to the hand coordinate values.

$$\begin{pmatrix} * & * & * & x \\ * & * & * & y \\ * & * & * & z \\ 0 & 0 & 0 & 1 \end{pmatrix} = Rot\,Z(\pi/2) \qquad (7)$$

$$* DH(-\vartheta_1, L_2, L_1, \pi/2)$$
$$* DH(\vartheta_2, L_3, 0, 0)$$
$$* DH(\vartheta_3, L_4, 0, 0)$$
$$* DH(\theta_4, L_5, 0, 0)$$

5. Experiment

The experiment used the arm robot shown in Fig. 14. Three cubic blocks colored red, blue, and yellow were prepared, and when students indicated a color using the microphone, the robot placed that color block into the nearby box.

6. Conclusion

This paper describes an intelligent, autonomous grasping system for an arm robot using image recognition with speech recognition/synthesis. Experiment results showed that spoken work commands and work reports through speech synthesis were executed without teaching. The above outcomes are anticipated to be applicable to academic fields and industry.

References

Hayakawa, Y. *Robotics*, Corona Publishing Co., Ltd.

Kidono, K. and Ninomiya, Y. (2006). Estimated Night Visibility Using a Multiband Camera, *VIEW*, pp.208–212.

Komiya, K., et al. (2006). Image Input, *Journal for the Institute of Image Electronics Engineers of Japan*. 35, 6, pp.758–760.

Koshimizu H. and Hata, S. (2001.5). Prospecting for New Horizon of Machine Vision Applications, *The Institute of Electrical Engineers of Japan*, (C)Vol.121-C, No.5, pp.835–840.

Koshimizu, H., et al. (2004.1). Prospect on Current Topics of Machine Vision Technologies, *The Institute of Electrical Engineers of Japan*, (C)Vol.123-C, No.1, pp.1–6.

Sato, Y. (2010). Study of Intelligent Control of an Arm Robot Equipped with a C-MOS Camera Using a New Type of Image Processing. *ISSS*, Canada.

Sato, Y. and Sakaue, K. (2006). Development of an Omni-directional Stereo Vision-based Smart Wheelchair, *VIEW*, pp.231–236.