Pattern Recognition Pergamon Press 1973. Vol. 5, pp. 99–108. Printed in Great Britain

Guiding a Robot by Visual Feedback in Assembling Tasks

Y. SHIRAI and H. INOUE

Electrotechnical Laboratory, Chiyoda-ku, Tokyo, Japan

(*Received* 2 *March* 1971)

Abstract—A system whereby a robot puts a block into a box using visual feedback is described. In this experiment, the difference between the desired position and the actual position of the object is recognized visually and corrected by moving the manipulator. The experimental results show that such visual feedback is effective in performing the precise positioning.

1. INTRODUCTION

INTELLIGENT robots which manipulate objects using visual information have been studied and experimental models have already been developed in various laboratories.[1–3] Generally the intelligent robot consists of the visual system, the handling system and the supervisory system that links these systems.

The visual system and handling system are usually combined together as an open loop system. Once the visual information is processed and the parameters are obtained, the visual system is completely separated from the handling system, and while the objects are handled, the visual system is not employed. In this open loop system, the accuracy of the operation depends directly on the errors of both the visual input device and the digitization in the visual system, the positioning errors in the handling system and the errors in co-ordinate transformation.

In order to compensate for those errors, a visual feedback loop can detect the difference between the actual position and the desired position of the manipulator. If the manipulator's position is adjusted to the desired position by the visual feedback loop, the accuracy of the task is raised to the point of the resolution of both the visual and handling systems.

In this paper, the assembly task of placing a square prism block into a square box is described as an example of an operation with a visual feedback loop. In this operation, the recognition problem is easy; however, it requires the accurate control of position and orientation.

2. OUTLINE OF THE EXPERIMENT

The processing of this experiment is composed of the following five successive stages. A flow diagram is shown in Fig. 1.

(A) *Visual recognition of the position of the box.* In this stage, the robot finds where the box is through its eye. The result of this recognition is used to determine the reference position in the next stage.

(B) *Manipulation to carry a block over its mating box.* Robot grasps a block and tries to carry it toward the desired position which is determined in the preceding stage.

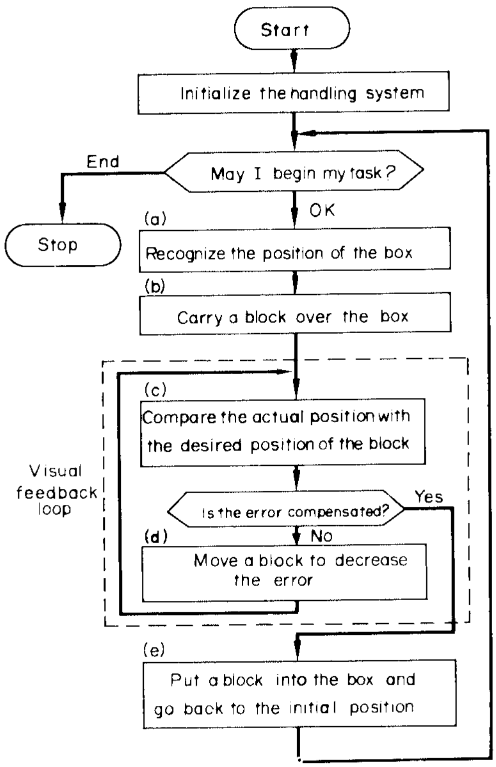


FIG. 1. Flow diagram of this experiment.

(C) *Recognition for visual feedback.* This stage gives the difference between the desired and the actual position of the block which has been moved by the manipulator.

(D) *Movement to decrease the errors.* In this stage, the robot moves its hand to compensate the errors recognized in previous stage. The procedures (C) and (D) are combined to constitute visual feedback loop.

(E) *Manipulation to insert a block into a box.* After the precise positioning has been performed by means of visual feedback, the robot completes the task and goes back to the initial position.

3. VISUAL INFORMATION PROCESSING

In Fig. 1, visual information processing corresponds to stage A and C. Both stages include the recognition of the object and the measurement of its position. The method of processing is common to both stages and it is described in the following Sections 3.1 to 3.3.

3.1 *Visual input*

A commercial vidicon TV camera is employed as a visual input device. The video signal is converted into a 6 bit digital signal by the preprocessor, and transmitted to the computer. The preprocessor divides the whole picture area into 256×256 picture elements and 64 × 64 elements are sampled in one scanning interval. The sampling area (so called "window") can be varied by the software. In this experiment, either $\frac{1}{4}$ or $\frac{1}{16}$ of the area of the whole picture ($\frac{1}{4}$ mode or $\frac{1}{16}$ mode) is selected as the sampling area. The $\frac{1}{4}$ mode is selected to detect the approximate position of the object. In this mode, every fourth element is sampled, one from each 2 × 2 element section of the picture over one quarter of the area of the picture. The digital picture thus obtained has a low resolution. On the other hand, the $\frac{1}{16}$ mode is selected to obtain the precise position of an object. This mode samples every picture element over $\frac{1}{16}$ of the area of the picture.

3.2 *Extraction of a line drawing*

This process extracts a line drawing of an object from the visual input data obtained from the eye. Because the background and the inside of the box are black and the outside of the box and the block are bright colored, the light–dark boundary lines constitute a line drawing.

First, the boundary points between dark and bright parts are obtained as shown in Fig. 2. While roughly scanning the picture elements (each of which is denoted as C_{ij}), if a

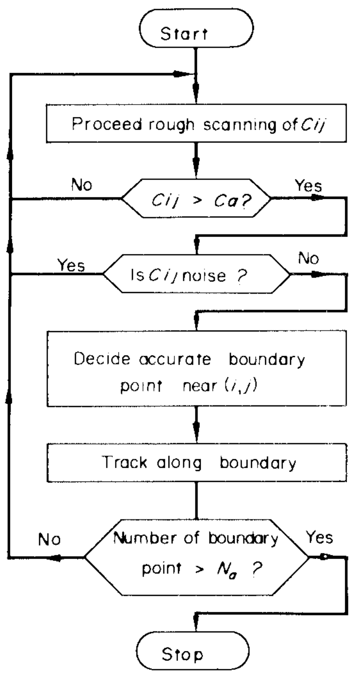


FIG. 2. Procedure to obtain boundary points.

possible boundary point is found, its neighboring elements are examined in detail to accurately locate the boundary point. Then starting from this point, the sequence of the boundary points are decided by the tracking algorithm.

Secondly, the equations of lines are calculated from the sequence of the boundary points. This procedure is shown in Fig. 3. The judgment whether a point is on the line or not is based upon the distance between the point and line represented by the equation. If the

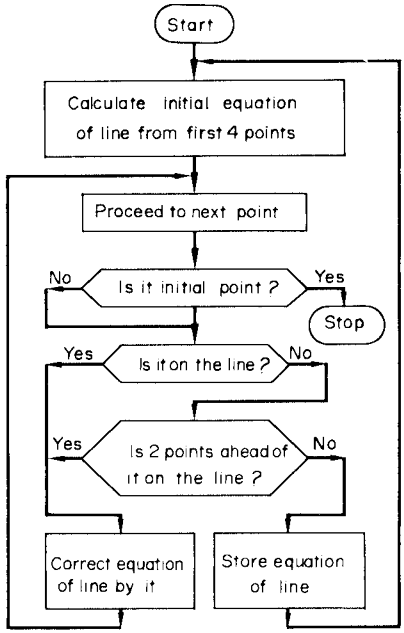


FIG. 3. Procedure to extract line drawing.

distance is less than a threshold, the point is regarded as on the line. The point might not be on the line because of noise. To help compensate for this, consideration is given to not only the point examined but also to the point which is 2 units ahead in this sequence of the boundary points. Specifically, if the second point ahead of the point is on the line, the examined point is also regarded as on the line. If the point is on the line, the equation of the line and the threshold of judgment are appropriately adjusted.

When all the equations of the lines are obtained, the cross points of the adjacent lines are calculated and the line drawing is completed.

3.3 *Transformation into the 3-dimensional space*

It is required for our robot to determine the 3-dimensional coordinates from the 2-dimensional picture data. When only one camera is employed as an input device, an additional constraint is necessary to obtain 3-dimensional coordinates. In this system, the constraint is the vertical position of the objects. That is, the height of the plane where the box is placed and the vertical position of the block is fixed. With this constraint, a point in the picture frame can be easily mapped into 3-dimensional space if the position and the attitude of the TV camera, parameters of lens system and the distortion of the image by the input device are measured beforehand.

3.4 *Determination of the position and the attitude of the box*

The procedure to obtain the position and the attitude is shown in Fig. 4. First, visual data is input by the preprocessor with the $\frac{1}{4}$ mode, and the approximate position is obtained. Then, the position of the window is adjusted to the center of the box and again visual data is input with the $\frac{1}{16}$ mode. From this data the line drawing of the box is obtained by the

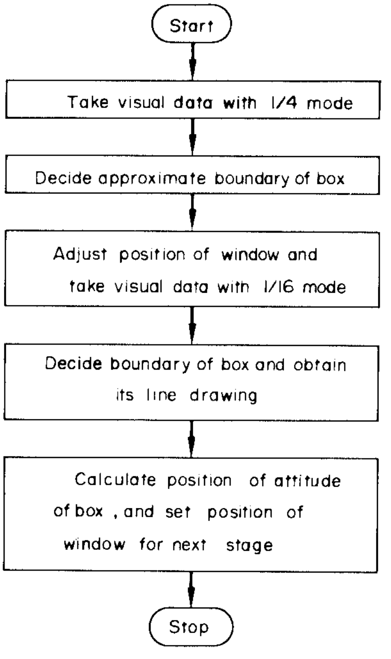


FIG. 4. Procedure to obtain position and attitude of box.

method described in 3.3. If the visual system is noiseless, this line drawing consists generally of 6 lines as AB, BC, CD, DE, EF and FA in Fig. 5. Taken from a front view, only one side of the box is seen and the line drawing consists of 4 lines forming a parallelogram. However, because of the noise of the vidicon or the error of the digitization, the complete line drawing is not always obtainable. For example, a small line may disappear, or one line may be divided into two adjacent lines; furthermore, two lines may become one line if the angle between them is very small. On the other hand, the visual data contains redundant information. Considering the fact that the outline of the box is a square prism, the position and the orientation of the box are determined by two vertices. Since the bottom edge which makes a greater angle to the line of vision is extracted most precisely, the end points of the line corresponding to this edge (BC in Fig. 5) are selected as the basis of calculation. Thus, the position and the attitude of the box are calculated based upon the 3-dimensional position of these two points.

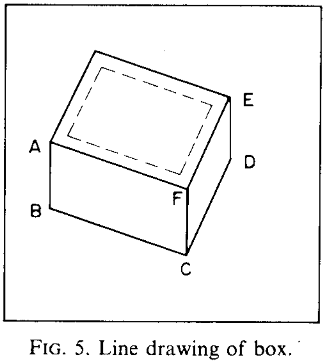


FIG. 5. Line drawing of box.

3.5 *The correction of the position and the orientation of the block by visual feedback*

The procedure to obtain the deviations of the position and the attitude of the block from the first approximations is shown in Fig. 6. The initial position of the window to get the visual data is set in accordance with the position of the box as shown in Fig. 4. If there

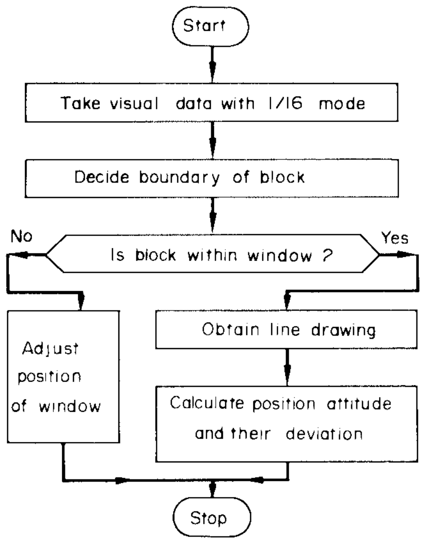


FIG. 6. Procedure to obtain deviations of block.

exists no errors in the whole system, the block might be at the center of the window. However, part of the block may actually be out of the window. The position of the block is then adjusted. The direction of this correction is determined by the position of the boundary of the block. The amount of the correction is constant. The block is brought inside the window after several corrections. If the block is within the window, its line drawing is extracted by the previously described procedure. The block is long enough that the manipulator is not in the window as shown in Fig. 7. Generally, the line drawing consists of 5 lines as shown in Fig. 7. The bottom edge of the block which makes a larger angle to the line of vision is selected as the base to determine the position and attitude as in the case of the box.

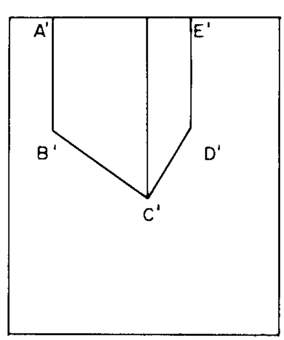


FIG. 7. Line drawing of block.

Since the height of the blocks is kept constant, the position and the attitude of the block are obtained from the end points of the line corresponding to this edge (B′C′ in Fig. 7), and their deviations from the desired values are calculated. The amount of the rotation about the vertical axis is limited to less than 45°.

4. MANIPULATOR AND ITS CONTROL

4.1 *Manipulator*

The manipulator of this robot, shown in Fig. 8, has seven degrees of freedom which are denoted as body rotation, shoulder pivot, elbow pivot, elbow rotation, wrist pivot, wrist rotation and squeezing motion. The first six axes are driven by the hydraulical position servomechanisms, and the last one is positioned by means of an electrical servomechanism.

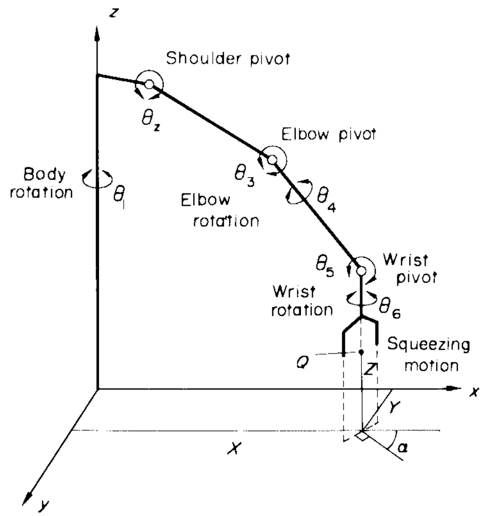


FIG. 8. The freedom of the manipulator and its coordinate system.

4.2 *Basic subroutines of motion*

As described in Section 3.3, the visual recognition is performed under the following conditions.

(1) The axis of the block is vertical to the working table.

(2) The vertical position of the block is kept constant.

Therefore, the attitude and the height of the hand must be always controlled to satisfy these conditions. To facilitate such control of the manipulator, the following subroutines are prepared.

NMOVE (J, X, Y, Z, α, 0, 0). Under the explained conditions, the coordinates X, Y, Z of point Q and the attitude in Fig. 8 are transformed into the corresponding six angles of the servomechanisms ($\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6$). Then the axes of the manipulator are actuated till they coincide with these computed reference angles. The first argument "J" is used to designate whether the successive six arguments are taken as relative displacement or as absolute displacement.

AMOVE (J, $\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6$). This subroutine is prepared to control the axes of the manipulator to the corresponding reference angles which are given as six arguments.

OPEN. This subroutine makes jaw open.

CLOSE. This is used to grasp an object.

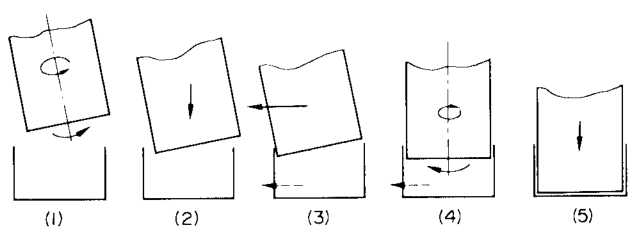
4.3 *Motions of the manipulator*

The task of assembly in this demonstration is composed of the following three manipulations. In the first and the third manipulation, the desired motion of the manipulator is prepared on the sequence table as the appropriate sequence of the above mentioned basic subroutines.

(I) *Manipulation to carry a block over its mating box.* After the visual recognition of the box is completed, the hand moves to reach the block placed at the predetermined location, grasps it and carries it over the box. The block is not always carried to the exact position decided by the visual system, because it is grasped without help of an eye. However, the variation in this positioning can be well corrected using visual feedback compensation in the next stage.

(II) *Visual feedback loop.* In this visual feedback loop, when the block is moved by the manipulator, its position and attitude are recognized by robot's eye to determine the difference between the actual and the desired position of the block. Then, the position and the orientation of the manipulator are changed to compensate for this difference. After several repetitions of such recognition-handling compensation cycles, a precise positioning is attained.

(III) *Manipulation to put a block into the box.* A simple method to assemble a block with a box is to put a block down vertically. But the validity of this method is limited by the resolution of visual processor and of position control system. In order to resolve the difficulty due to this resolution, another method shown in Fig. 9 is attempted. In this method, the block is tilted so that the box follows the movement of the block and they are automatically coupled.



(1) Tilt a block. (2) Put it down.
(3) Slide it to align at the edge of the box.
(4) Reset its attitude to vertical.
(5) Put it into the box.

FIG. 9. Advanced motion for assembling.

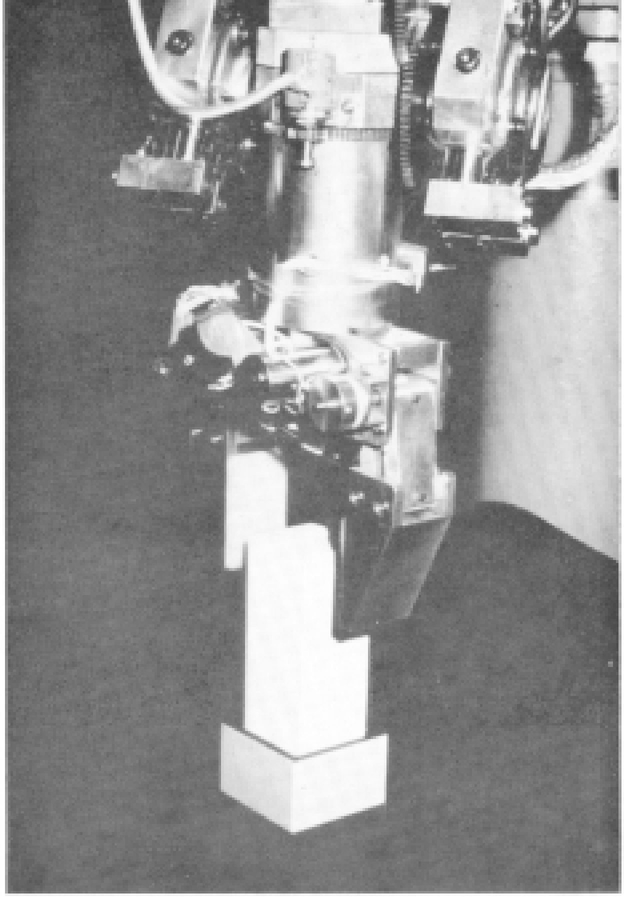


FIG. 10. Assembly manipulation by robot.

[*facing* p. 106]

The first method is enough to insert a 40 mm square block into a box whose interior size is 45 mm square. If the second method is used, the same block can be coupled with a box of 43 mm square interior.

5. ON THE PROGRAM OF THIS EXPERIMENT

The core memory unit of our computer contains 32K words of 18 bits. It is not large enough for all the programs of this experiment to be accommodated in it. Therefore, the programs are divided into four phases. One is a core-resident phase of the supervisory program. The others are three overlayed phases corresponding to the visual recognition program, the handling program and the visual feedback compensation program. The flow diagram of this experiment is shown in Fig. 1.

Before beginning the task, system initialization is made and the manipulator is moved to outside of the television picture.

The task is carried out as follows. First, the visual recognition program is loaded to the core memory unit from the magnetic disk unit. It determines the position and the attitude of the box. According to this result, supervisory program calculates a sequence of data which are necessary to carry a block over the box. This information is given to the handling program and the manipulation described in 4.3-I is carried out. Next, visual feedback compensation program is loaded and the accurate positioning is performed. At last, handling program is loaded again and the block is coupled with the box.

It requires about 80 sec to complete the task by use of NEAC-3100 (cycle time: 2 μsec). The time to recognize the box is about 10 sec and to carry a block above the box is also about 10 sec. One cycle of the visual feedback loop requires about 10 sec. Three or four cycles are enough to decrease the error within a limit. About 20 sec is used to swap the overlayed phase.

The photograph of this experiment is shown in Fig. 10.

6. CONCLUSION

It is demonstrated by the experiment that the visual feedback system enables an intelligent robot to perform assembly tasks with high accuracy. The given task is to put a square prism block into a square box with 5 mm clearance. First, the box is recognized and the desired position of the block is obtained, and the block is carried off to the indicated position. Then, the block is recognized and the deviation from the desired position is calculated. The block is moved to compensate this deviation. This visual feedback loop is repeated until the deviation is within a predetermined limit, and the block is inserted into the box. If the block is tilted during insertion, the task with 3 mm clearance is achieved.

In this experiment, only one TV camera is used to obtain the position of the objects in the 3-dimensional space. If 3-dimensional position is directly detected by the input device, the range of the possible tasks might be increased. Not only the visual information but also the tactile information is effective in giving more ability to the robot. Furthermore, by employing bilateral manipulator,[5] the performance of more complicated and precise tasks might be achieved.

Acknowledgements—The authors would like to express appreciation to Professor S. Tsuji of Osaka University who was Chief of Bionics Section in 1970, and Dr. K. Sato, Chief of System and Control Section, for kind guidance and indication, and the researchers in the Electrotechnical Laboratory who provided an excellent experimental system.

REFERENCES

1. J. McCARTHY *et al.*, A computer with hands, eyes and ears, AFIPS Conf. Proc., Vol. 33 (1968).
2. L. L. SUTRO *et al.*, Assembly of computers to command and control a robot, Report R-582, Instrumentation Lab. M.I.T. (1968).
3. M. EJIRI *et al.*, Development of an adaptive robot with visual sensor, Tokyo Conf. of I.E.E.J. (1970).
4. M. SUWA *et al.*, A hardware system of an industrial eye, *Bull. Electrotech. Lab.* **35** (3) (1971).
5. H. INOUE, Computer controlled bilateral manipulator, *Bull. JSME* **14** (69) (1971).