The Humanoid Saika that Catches a Thrown Ball

doi: 10.1109/ROMAN.1997.646959

The Humanoid *Saika* that Catches a Thrown Ball

Koichi NISHIWAKI, Atsushi KONNO, Koichi NAGASHIMA,
Masayuki INABA, and Hirochika INOUE
Department of Mechano-Informatics,
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113, Japan

Abstract

This paper addresses ball-catching behavior by the humanoid robot Saika. The aim of this study is to implement human skills on the humanoid in the same manner as human performs them. The behavior of catching a falling and a thrown ball is chosen as an example of dynamic skillful manipulation. Considering the human behavior, we realize ball-catching behavior by three steps: (a) localization of the ball by the vision system, (b) prediction of the ball's path to determine the catching point, and (c) reaching out the hand to the catching point by a neural network inverse kinematics model. Experimental results demonstrate the validity of the catching strategies.

1 Introduction

Future robots are expected to be used in an environment where humans live and work. Humanoid robots of similar size, figure and DOFs arrangement as human beings' are considered to have the advantage of behaving themselves like a human in such tasks as handling tools and communicating by gesture.

The first humanoid robot in the world, WABOT-1 (WAseda roBOT 1), was developed in 1973 by Kato [1]. Many other humanoid robots have subsequently been developed [2, 3, 4].

On the other hand, some dynamic skillful manipulations have been studied such as Ping-Pong [5, 6] and devil sticking [7]. These manipulations were carried out in an environment where the cameras were placed outside of the robot.

In this work we study dynamic skillful manipulation on a humanoid robot where the positional relationship between the eyes and arms is designed to be similar to that of a human. Ball-catching behavior is chosen as an example of dynamic skillful manipulation on the humanoid robot *Saika* which is developed for general study of humanoid robots.

2 The aim of the work

The aim of the work is to implement human dexterity on the humanoid. Especially, our interest is to achieve humanoid's dexterous manipulations in the same manner as human performs them. Therefore, it is desired that the humanoid has similar size, figure, DOFs arrangement and binocular arrangement as that of a human being.

IEEE International Workshop on
Robot and Human Communication
0-7803-4076-0/97/ $10.00 © 1997 IEEE.

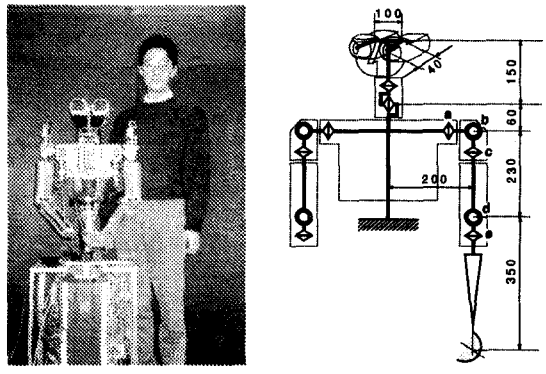

Figure 1: Humanoid Robot *Saika*.

Catching a falling and a thrown ball is chosen as an example of human dexterity. We assume the model of human ball-catching behavior as: (a) the ball's localization in the human's Cartesian coordinate system is done by using visual information, (b) the mapping between the coordinate systems of the arms and the human's Cartesian coordinate system is acquired, and (c) catching behavior is composed of the ball's path prediction and visual feedback.

In this work, we do not deal with visual feedback. Catching a ball by the humanoid is done only by prediction of the ball's path. Based on the assumed model of human catching behavior, the following three steps are considered necessary for the humanoid to achieve the ball-catching behavior;

- localization of the ball using the humanoid's vision system,
- prediction of the ball's path and determination of the catching point,
- positioning the humanoid's hand using the acquired inverse kinematics neural network model.

3 Humanoid *Saika*

3.1 Mechanical structure

The humanoid robot *Saika* is shown in Figure 1. The humanoid consists of a head, a torso, and arms. It has approximately the size of an adult female, and has 12-DOFs in total, 2-DOFs at the neck and 5-DOFs

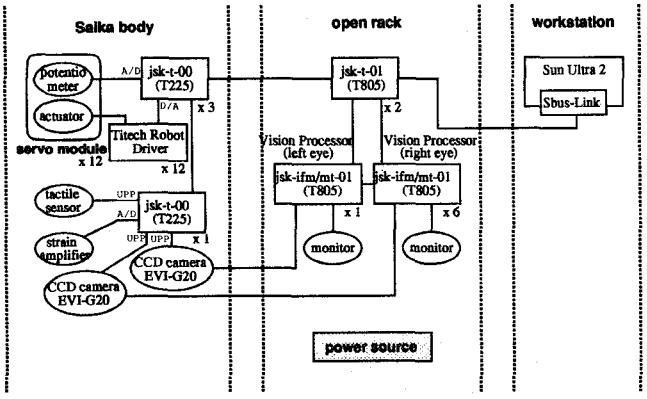

Figure 2: The control system of *Saika*.

at each arm. Two CCD cameras with internal pan and tilt DOFs are placed in parallel on the head. For this work, a cooking basket of 120 [mm] in diameter is attached to the left arm as a hand for ball-catching.

3.2 Control system

The control system is developed on the basis of the distributed computing environment consisting of a workstaion and transputers. The motor controller JSK-T-00s have A/D converters, D/A converters, pulse processors, and transputers (INMOS-T225). The JSK-T-00 also controls pan and tilt DOFs in the cameras by commanding via RS-232C.

Visual processing is carried out by the JSK-IFM/MT-01 [8]. Six JSK-IFM/MT-01 boards are used for the right eye and one is used for the left eye. The JSK-IFM/MT-01 has a transputer and a Motion Estimation Processor (MEP, SGS-Thomson STI3220E) on each board. The MEP can search for the region that is most similar to a previously memorized image. The JSK-IFM/MT-01 is used for the real-time tracking of moving objects.

4 Localization of a ball using visual information

We discuss the localization strategies for the following objects; (i) a vertically falling ball and (ii) a thrown ball. Different methods are used for each object depending on the characteristics of the measurement conditions.

4.1 Localization of a vertically falling ball

The characteristics for the localization of a vertically falling ball are:

- the falling movement is very quick,
- therefore there is not enough time for measurement (it takes about 250 [ms] for the ball to move through the camera view),
- the quick calculation of 3-dimensional position from the measurement is needed (it takes about 400 [ms] for the ball to reach the catching point from the release point),

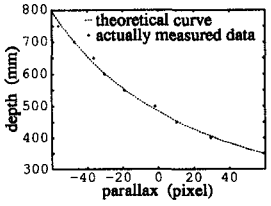

Figure 3: The relationship between the depth and the parallex.

- small change in the distance along the direction of the visual axis of the camera.

Taking these characteristics into consideration, the localization is done using a stereo view obtained by the binocular camera arrangement. Since the cameras do not have a synchronizing mechanism, the vision system cannot obtain the left view image and the right view image at the same time in a strict sense (maximum time lag is 33 [ms]). However, in the case of the falling ball, the distance between the humanoid and the falling ball in depth (distance in Z direction in Figure 9) does not change much. Therefore, the time lag is not a big problem for the stereo viewing in this case. By fixing the relationship of the two cameras, the calculation of 3-dimensional position from the measurement is simplified.

The relationship between the parallax and the depth is shown in Figure 3. The average error is 10 [mm] in the range of 300∼800 [mm] from the camera.

4.2 Localization of a thrown ball

In the localization of a thrown ball, a large change of the distance between the ball and the humanoid in depth could be a big problem. Because the cameras used in this work do not have a synchronizing mechanism as mentioned above, stereo viewing would not be effective for localization. Therefore we use the ball's size to localize the ball in depth. The relationship between the ball size n [pixel] on the image and the distance between the ball and the camera l [mm] in depth is expressed by the following equation,

$$l = \frac{df}{n}, \tag{1}$$

where, f is the focal length (820 [pixel]), and d is the diameter of the ball (70 [mm]). The size of the ball measured at each depth and the curve obtained from of the Eq.(1) are shown in Figure 4. The average error in depth is 13 [mm] in the range of 300∼1200 [mm]. The resolution of the depth (ΔZ) at each point of distance calculated by Eq.(2) is shown in Figure 5.

$$\Delta Z = \frac{df}{n} - \frac{df}{n+1} = Z\left(1 - \frac{1}{1+\frac{1}{df}Z}\right). \tag{2}$$

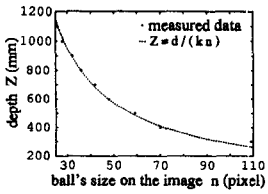

Figure 4: The relationship between the distance and the ball size on the image.

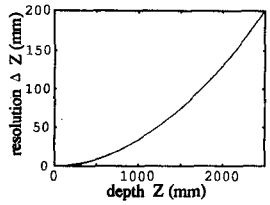

Figure 5: The resolution in depth.

5 Acquisition of inverse kinematics model using the neural network

The humanoid *Saika* has five DOFs for each arm. Four DOFs (a, b, c, and d in Figure 1) are used to determine the hand position and one DOF (e in Figure 1) is used for hand's posture control.

The transformation from the 3-dimensional hand position to the four joint angles (inverse kinematics) is acquired using a neural network model. After the four position controlling joint angles are determined, the posture-controlling joint is controlled so as to keep vertical the plane that includes the normal of the face covering the rim of the cooking basket and the rotation axis of the posture controlling joint. Each joint was then controlled by position feedback.

5.1 Neural network model

A three layered neural network with sigmoid function is used for acquisition of the inverse kinematics. The numbers of neurons are; three for the input layer, twenty-five for the hidden layer, and four for the output layer. The back propagation algorithm is employed to update the weight.

Direct-inverse modeling is used for the acquisition (Figure 6). In the first place, 162 data sets of four joint angles that give a hand position within the humanoid's working area are prepared. In the second place, a set is taken from the 162 data sets and the hand position corresponding to the set is calculated from the previously prepared forward kinematics model. In the third place, the calculated hand position is used as the input signal, while the taken data set is used as the training signal. Furthermore, this procedure is iterated for the next data set.

Since the arm has redundancy, the acquisition of inverse kinematics is an ill-posed problem. In order to avoid this ill-posed problem we prepare the learning data of joint angles so as to have a unique relationship between angles and hand position. To achieve this, we give the the following conditions depending on the hand position. The norm of joint angle vector is minimized when the hand is further out to the side than the shoulder, and the arm dose not interfere with the torso when the hand is closer to the center of the body than the shoulder.

5.2 Learning

The reduction of total error of the normalized output in the learning process is shown in Figures 7.

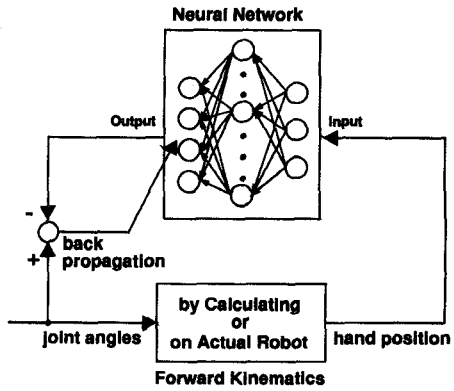

Figure 6: Acquisition of an inverse kinematics neural network model.

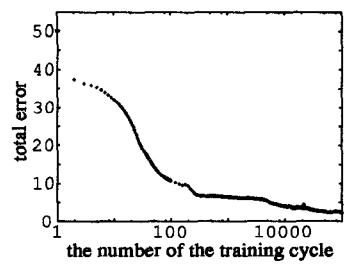

Figure 7: The total error of the normalized output in the learning process.

After learning the 162 data sets for 100,000 cycles the average error becomes 5 [mm] when the output of the neural network is transformed to position by forward kinematics and compared with the input position data.

The actual humanoid is also used, instead of forward kinematics, aiming for the learning without the influence of the backlash of the gears and the offsets of the origins of joint angles (see Figure 8). The hand position is measured by the cameras. After learning the 162 data sets for 100,000 cycles the average error becomes 52 [mm] when the output of the neural network is transformed to position with the actual robot and compared with the input position data. Since the average error is larger than that of the model acquired by calculating forward kinematics, the model acquired by calculating forward kinematics is adopted for the ball-catching experiment.

6 Catching a falling ball

By using the position measurement discussed in section 4.1 and the inverse kinematics neural network model discussed in section 5 the experiment of catching a vertically falling ball is performed by moving the hand to a position vertically under the measured position of the ball (Figure 9).

In the experiment a ball of 7 [cm] in diameter falls

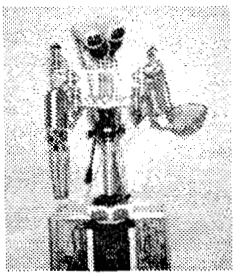

Figure 8: Acquisition of inverse kinematics by seeing its own hand.

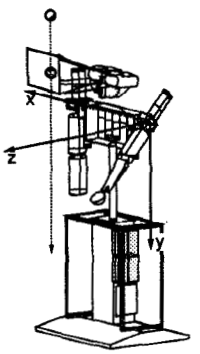

Figure 9: The experiment of catching a vertically falling ball.

Figure 10: A scene of the experiment of catching a vertically falling ball.

from above the upper edge of the camera vision with 0 initial velocity. The distance between the initial position of the ball and the catching point is 800∼ 900 [mm]. Figure 10 shows the Saika's catching a falling ball behavior.

7 Catching a thrown ball

In the catching a thrown ball behavior, prediction of the ball's path is a key issue to determine the hand position to catch a ball. For the ball's path prediction, the ball's path in the sagittal plane (y-z plane in Figure 11) is approximated by a parabola, while the ball's path in the horizontal plane (x-z plane in Figure 11) is approximated by a line. The prediction was carried out by determining the coefficient of the approximated parabola and line using the least squares method. The first prediction is carried out after getting five data points. The prediction is iterated when new data is obtained in order to modify the catching point.

The least squares method is used where the weights of the data changes depending upon the reliability of each data. The lower reliabilities are given to the older data.

7.1 Prediction of the ball's path on the horizontal plane

Let (x_i, y_i, z_i), and p_i represent the i-th observed ball position and the reliability of the i-th data respectively. The fitting line is represented by

$$x = a_0 + a_1 z. \tag{3}$$

After the observation of the n-th point the summation of the square errors is expressed by

$$E_1 = \sum_{i=1}^{n} p_i (a_0 + a_1 z_i - x_i)^2. \tag{4}$$

The prediction is done by solving a_0 and a_1 which minimize the E_1. Consequently we can obtain the fitting line by solving the following equations:

$$\frac{\partial E_1}{\partial a_0} = 0, \quad \frac{\partial E_1}{\partial a_1} = 0. \tag{5}$$

Let

$$\begin{cases} S_0^n = \sum_{i=1}^{n} p_i, \quad S_x^n = \sum_{i=1}^{n} p_i x_i, \quad S_z^n = \sum_{i=1}^{n} p_i z_i, \\ S_{z^2}^n = \sum_{i=1}^{n} p_i z_i^2, \quad S_{xz}^n = \sum_{i=1}^{n} p_i x_i z_i, \end{cases} \tag{6}$$

and then, Eq.(5) becomes

$$\begin{cases} S_0^n a_0 + S_z^n a_1 = S_x^n \\ S_z^n a_0 + S_{z^2}^n a_1 = S_{xz}^n, \end{cases} \tag{7}$$

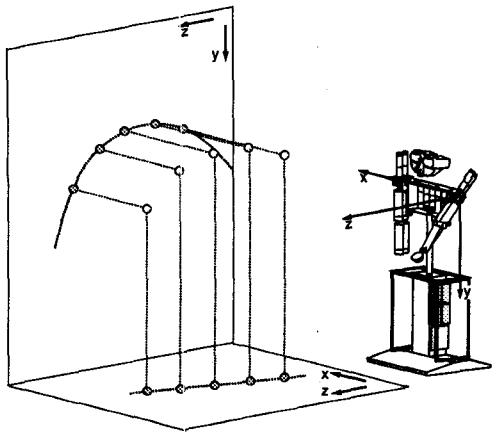

Figure 11: The estimation of the trajectory.

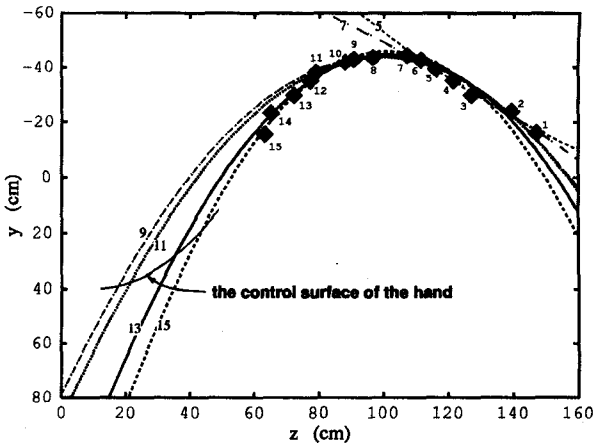

Figure 12: Fitting a parabola to the observed point.

therefore, a_0 and a_1 are given by:

$$\begin{cases} a_1 = \dfrac{S_x^n S_z^n - S_0^n S_{xz}^n}{S_{z^2}^n - S_0^n S_{z^2}^n} \\ a_0 = \dfrac{S_{xz}^n - a_0 S_{z^2}^n}{S_z^n}. \end{cases} \tag{8}$$

In the experiment the reliability of the data is given by:

$$p_i = r^{n-i} \quad (0 < r \leq 1). \tag{9}$$

The relationships between n-th summations S_A^n and $(n-1)$-th summations S_A^{n-1} are expressed by the recurrent formulas:

$$\begin{cases} S_0^n = 1 + rS_0^{n-1} \\ S_x^n = x_n + rS_x^{n-1} \\ \quad \vdots \end{cases} \tag{10}$$

Therefore we can reduce the amount of calculation using the above recurrent formulas.

7.2 Prediction of the ball's path on the sagittal plane

Let (x_i, y_i, z_i), and p_i represent the i-th observed ball position and the reliability of the i-th data respectively, and let the fitting parabola be represented by

$$y = b_0 + b_1 z + b_2 z^2, \tag{11}$$

in the same way as described in the previous section. After getting the n-th data point, the summation of the square errors is expressed by

$$E_2 = \sum_{i=1}^{n} p_i (b_0 + b_1 z_i + b_2 z_i^2 - x_i)^2. \tag{12}$$

The fitting parabola is obtained by solving b_0, b_1, and b_2 which minimize the E_2. Consequently the fitting parabola can be obtained by solving the following equations in the same manner as described in the previous section.

$$\frac{\partial E_2}{\partial b_0} = 0, \quad \frac{\partial E_2}{\partial b_1} = 0, \quad \frac{\partial E_2}{\partial b_2} = 0 \tag{13}$$

The reliability of the data is decided to be the same as used in the previous section.

8 Experiment

By using the position measurement discussed in section 4.2, the prediction of the ball's path discussed in this section, and the inverse kinematics neural network model discussed in section 5, the experiment of catching a thrown ball is performed.

A ball, 7cm in diameter, was thrown from about 2 meters away in front of the humanoid. The panning joint of the neck is turned to the front and the tilting joint of the neck is turned to 15 degrees above the horizontal. Right eye and left eye are turned to the front and 15 degrees above the horizontal respectively. A scene of the experiment is shown in Figure 13.

In this method, the small error in localization makes a significant difference in the predicted ball's path near the catching point. Therefore visual feedback is expected to increase the rate of success.

9 Conclusion

This work aimed to implement human manners in dexterous manipulations on the humanoid. The catching of a falling and a thrown ball behavior was chosen as an example of dexterous manipulation. First we assumed the model of human ball-catching behavior as: (a) the ball's localization in the human's Cartesian coordinate system is done by using visual information, (b) the mapping between the coordinate systems of the arms and the human's Cartesian coordinate system is acquired, and (c) catching behavior is composed of the ball's path prediction and visual feedback. We realized the ball-catching behavior by the humanoid *Saika* based on these assumptions. The experiment

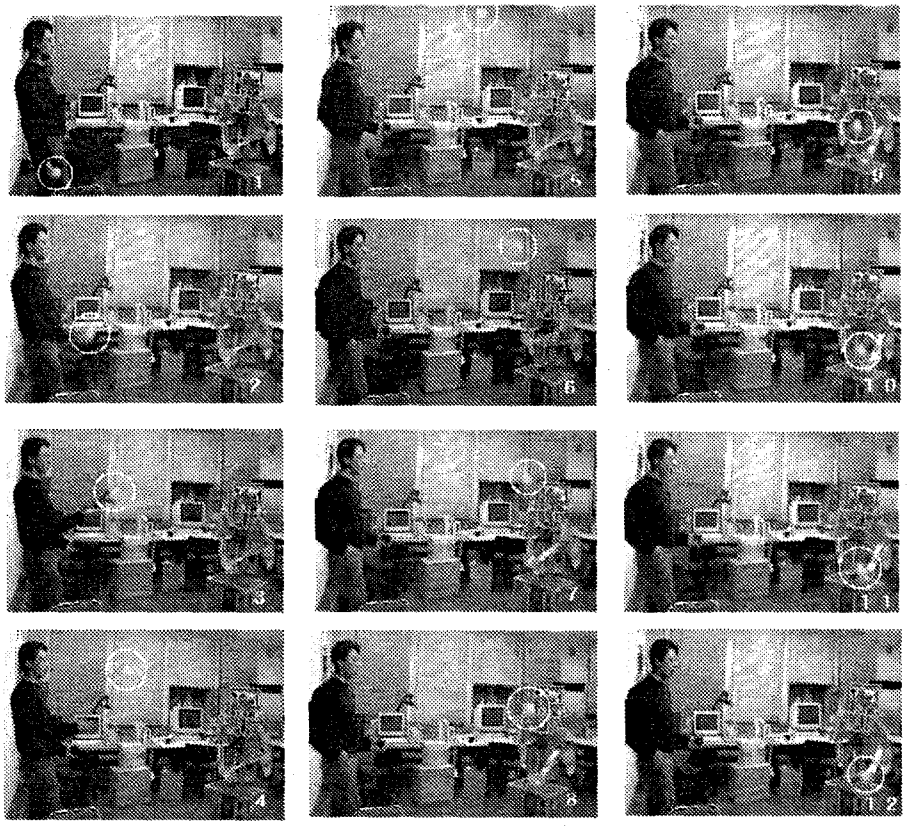

Figure 13: A scene of the experiment of catching a thrown ball.

of catching a vertically dropping ball was successful though the area was restricted. Catching a thrown ball is performed by predicting the ball's path. The path prediction is done by introducing the reliability of data and fitting to the approximated path using the least squares method. Visual feedback in catching will be a topic for future research.

References

[1] Ichiro Kato. Development of wabot-1. *Biomechanism 2, The University of Tokyo Press*, pages 173–214, 1973 (In Japanese).

[2] Rodney A. Brooks. Prospects for human level intelligence for humanoid robots. In *Proc. of the First Int. Symp. on HUmanoid RObots (HURO '96)*, pages 17–24, Tokyo, Japan, October 1996.

[3] K. Kawamura, D. M. Wiles, T. Pack, M. Bishay, and J. Barile. Humanoids: Future robots for home and factory. In *Proc. of the First Int. Symp. on HUmanoid RObots (HURO '96)*, pages 53–62, Tokyo, Japan, October 1996.

[4] John M. Hollerbach and Stephen C. Jacobsen. Anthropomorphic robots and human interactions. In *Proc. of the First Int. Symp. on HUmanoid RObots (HURO '96)*, pages 83–91, Tokyo, Japan, October 1996.

[5] R. L. Andersson. *A robot ping-pong player.* The MIT Press, 1988.

[6] Fumio MIYAZAKI, Eiichi HIROSE, Daisuke NAKAMURA, and Takuya ISHII. Motion estimation of a spinning ball for ping-pong robots. In *Proc. of JSME Conf. on Robotics·Mechatronics (ROBOMECH '97)*, pages 543–544, 1997.

[7] Stefan Schaal and Christopher G. Atkeson. Memory-based robot learning. In *Proc. of IEEE Int. Conf. on Robotics and Automation*, pages 2928–2933, 1994.

[8] Hirochika Inoue, Masayuki Inaba, Taketoshi Mori, and Tetsuya Tachikawa. Real-time robot vision system based on correlation technology. In *Proc. of Int. Symp. on Industrial Robots (ISIR '93)*, pages 675–680, 1993.