2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

High-Speed Catching by Multi-Vision Robot Hand

Masaki Sato, Akira Takahashi and Akio Namiki

***Abstract*— In this paper, we propose the "multi-vision hand", in which a number of small high-speed cameras are mounted on the robot hand of a common 7 degrees-of-freedom robot. Also, we propose visual-servoing control by using a multi-vision system that combines the multi-vision hand and external fixed high-speed cameras. The target task was ball catching motion, which requires high-speed operation. In the proposed catching control, the catch position of the ball, which is estimated by the external fixed high-speed cameras, is corrected by the multi-vision hand in real-time. In experiments, the catch operation was successfully implemented by correcting the catch position, thus confirming the effectiveness of the multi-vision hand system.**

I. Introduction

In recent years, the size and price of visual sensors have been reduced, making it easier to use multiple cameras simultaneously. In addition, the widespread use of parallel computers such as graphics processing units (GPUs) has enabled the processing of visual information in real-time. The purpose of this study is to develop a so-called "multi-vision hand" in which a number of small high-speed cameras are arranged on the surface of a conventional robot hand, and to achieve high-speed, high-precision manipulation while eliminating blind spots by integrating information from the multi-vision hand and other external high-speed cameras. In particular, our aim is to demonstrate ball catching control. This task has been studied in our research group as part of ball juggling control performed by a multi-fingered hand-arm robot[1][2][3]. This ball juggling control system has succeeded in catching three balls up to four times, but it was difficult to operate the system for a long time due to the poor ball position estimation accuracy, which led to a decrease in the ball juggling accuracy [4].

The control concept of our system is shown in Fig. 1. Visual servoing control is performed by setting the target as the estimated ball position projected onto images on the multi-vision hand. In addition, the hand orientation is also controlled at the same time so as to maintain the catchable posture.

Because multiple cameras are attached to the surface of the hand, blind areas can be eliminated, and robust, precise catching can be achieved.

M. Sato, A. Takahashi, and A. Namiki are with the Graduate School of Engineering, Chiba University, 1-33, Yayoicho, Inage-ku, 263-0022, Chiba, Japan m.sato0409@chiba-u.jp, a.takahashi876@chiba-u.jp, namiki@faculty.chiba-u.jp

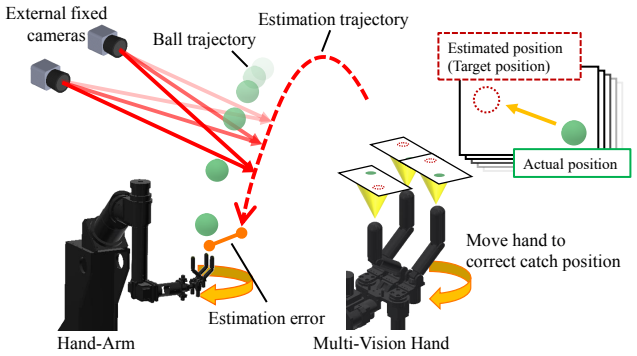


Fig. 1. Concept.

II. Related work

A. Eye-in-hand

Eye-in-hand is a commonly used configuration, and it is used for visual servoing control [5][6]. In conventional eye-in-hand systems, the visual field is so narrow that the visual servoing can only be designed so that the camera approaches the target along a straight line. On the other hand, our multi-vision hand can observe the surrounding area in the form of visual information, helping to eliminate blind areas. As a result, the visual field is wide, so that there is also a lot of flexibility in the design of the visual servoing. This is also useful for achieving rapid capturing or collision avoidance in manipulation.

Our group previously developed a robot system with high-speed cameras mounted on the links of a robot arm for real-time collision avoidance[7]. However, because the cameras were large, the system was not suitable for practical uses.

B. Sensing by robot hand

There are several studies on sensing by placing sensors on robot hands. Yamaguchi et al.[8] proposed Finger Vision, a method for measuring the force applied to a vision-based gripper, and succeeded in gentle grasping and in-hand manipulation. In that system, cameras are primarily used to track markers and provide a tactile sense, but can also see the external scene through transparent skin.

Also, there are some applications of proximity sensors for robotic manipulation. For example, Patel et al.[9] proposed a dynamic tactile sensor that combines proximity, contact and force (PCF) and applied it to robotic grasping. Koyama et al.[10] demonstrated high-speed grasping of flexible objects such as marshmallows and paper balloons by combining high-speed vision with individual sensors in proximity to a robot.

978-1-7281-6211-9/20/$31.00 ©2020 IEEE

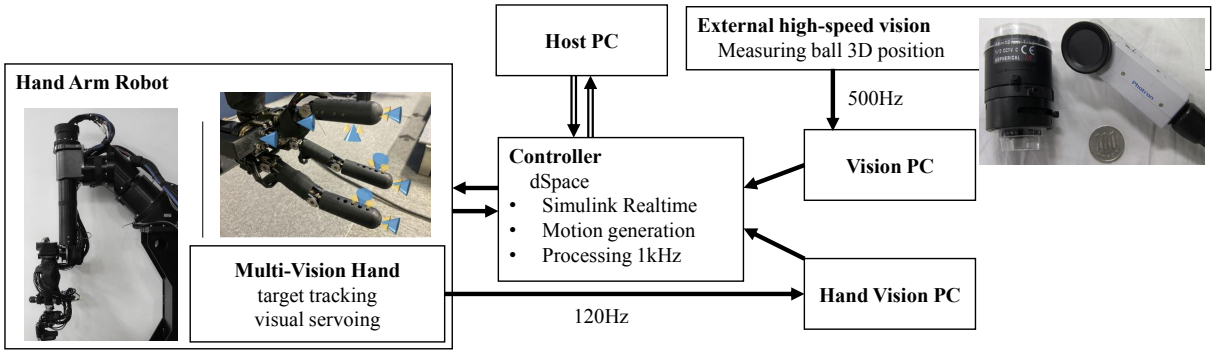
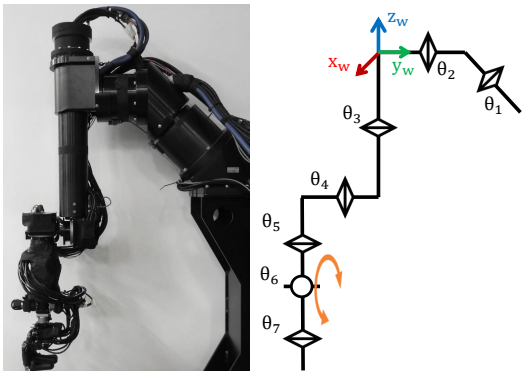


Fig. 2. System Configuration.



(1) Appearance (2) Joint configuration

Fig. 3. High-Speed Hand-Arm.

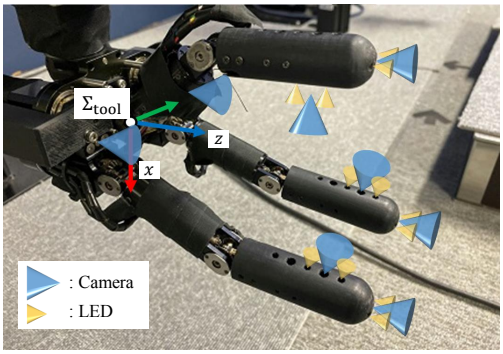


Fig. 4. Multi-Vision Hand.

On the other hand, our system aims to realize a wide field of vision and to realize quick grasping. In general, the depth range of a camera is wider than that of a proximity sensor and the amount of information is also larger. Therefore, this enables our multi-vision hand to achieve complex tasks.

C. Ball catching

Bäuml et al.[11] demonstrated catching of two simultaneously thrown balls by a two-armed mobile humanoid robot. Birbach et al.[12] demonstrated ball catching with an eye-in-hand single-camera configuration on a common industrial robot.

In the present study, we realized a catching motion using a general-purpose seven degrees-of-freedom (DoF) robot. In particular, this paper focuses on information loss due to occlusion.

III. System Configuration

The system configuration is shown in Fig. 2. The multi-vision hand system consists of a multifingered hand-arm on which a number of small cameras are attached and two external fixed high-speed cameras. The controller estimates the trajectory of a ball and the desired catching position and orientation based on the visual information from the external fixed cameras. Next, using the visual information from the multi-vision hand, the hand position and orientation are corrected by visual servoing control.

A. High-Speed Hand-Arm

The robot arm has 7 DoF (Fig. 3). The first axis has a diagonal joint so that the arm can support its own weight and move at high speed. Also, since the reduction ratio of each joint is small, at about 50:1, the arm is suitable for high-speed operation. The multi-vision hand is attached to the tip of the arm. The details of the hand are described in the next section.

B. Real-time Control System

In this study, the multi-vision system is controlled with a dSPACE modular system. The system can be controlled at 1 kHz, and the speed of the control loop is fast enough to operate the hand-arm smoothly. The control software was developed using MATLAB/Simulink.

C. External high-speed vision

We use the high-speed vision platform called IDP Express RF2000F [13]. The platform can acquire 512x512 pixel, 8-bit Bayer pattern information at a rate of up to 2000 Hz. In the experiments, the rate was set to 500 Hz. The captured Bayer pattern image is transferred to the main memory of the Vision PC at high speed through a PCIe board with double channel camera inputs (Fig. 2).

IV. Multi-Vision Hand

The multi-vision hand is shown in Fig. 4. It is a modified version of a multi-fingered high-speed hand developed to

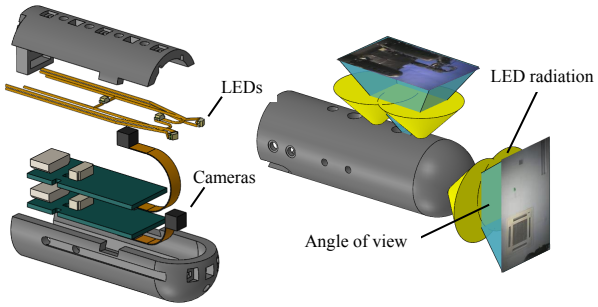
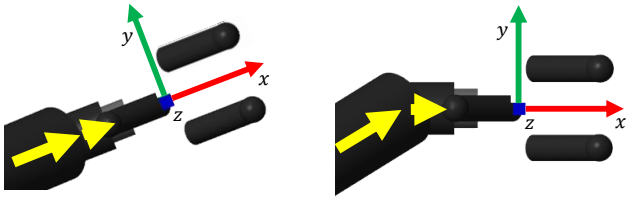


Fig. 5. Finger configuration.



(a) Not considering z-axis rotation (b) Considering z-axis rotation

Fig. 6. Effect of allowing Z-axis rotation.

achieve dynamic high-speed operations [14]. The hand consists of 3-DoF right and left fingers and a 2-DoF center finger, and its weight is about 600 g, excluding the wiring cables.

Smaller cameras are installed at the tip and the surface of the top link of each finger, and at the base of the left and right fingers (Fig. 5). In addition, LEDs are installed around the cameras to assist image capturing.

The cameras used are IU233N2-Z developed by Sony Inc. They are very small, with dimensions 2.6 mm x 3.3 mm x 2.32 mm. They can output images at up to 240 Hz in CSI-2 format. In our system, visual information from each camera is sent to the Vision PC through a USB converter. As a result, the actual rate was 120 Hz for 320x200 pixel 8-bit RGB images.

The small cameras, the LEDs, and the signal conversion boards are installed inside the fingers. Thus, they can continue to acquire images stably even when a shock is applied to the fingers, such as from a ball collision. In addition, each finger is made from ABS resin by using a 3D printer, and the weight of each finger is only 14 g.

V. Visual Servoing Control

As shown in Fig. 1, the concept of the proposed control involves correcting the catch position estimated by the external fixed-position cameras using visual servoing of the multi-vision hand.

A. Estimation by external high-speed vision system

The 3-D position of the ball ${}^{\mathrm{w}}\boldsymbol{x}$ is calculated using a stereo method with the visual information from the external fixed high-speed vision system, where the upper left w indicates the world coordinate system Σ_{w}. The estimated states ${}^{\mathrm{w}}\hat{\boldsymbol{x}}$, ${}^{\mathrm{w}}\dot{\hat{\boldsymbol{x}}}$, ${}^{\mathrm{w}}\ddot{\hat{\boldsymbol{x}}}$ of the ball are obtained from ${}^{\mathrm{w}}\boldsymbol{x}$ by using a Kalman filter. The trajectory of the ball is calculated from the estimated state, and the catch position $\boldsymbol{p}_{\mathrm{catch}}$ and the catch time t_{catch} are estimated. The orientation of the hand $\boldsymbol{R}_{\mathrm{catch}}$ is given so that the hand faces upward.

$$\boldsymbol{R}_{\mathrm{catch}} = \begin{bmatrix} 0 & \sin\phi & \cos\phi \\ 0 & -\cos\phi & \sin\phi \\ 1 & 0 & 0 \end{bmatrix}, \tag{1}$$

$$\sin\phi = \frac{x_{\mathrm{catch}}}{\left\| [x_{\mathrm{catch}}, y_{\mathrm{catch}}]^{\mathrm{T}} \right\|}, \cos\phi = \frac{y_{\mathrm{catch}}}{\left\| [x_{\mathrm{catch}}, y_{\mathrm{catch}}]^{\mathrm{T}} \right\|}.$$

The desired joint angles $\boldsymbol{q}_{\mathrm{catch}}$ are calculated by solving the inverse kinematics using an iterative calculation [15], from the desired hand position/orientation. The desired joint angle trajectory is generated using a fifth-order polynomial.

B. Correction of Catching by Multi-Vision Hand

One approach is to integrate all the visual information from the multi-vision hand with the Kalman filter described in the previous section, but this may result in a delay in estimation. In this paper, we take the approach of correcting the predicted catch position calculated by using the Kalman filter with visual servoing of the multi-vision system. High-speed, high-accuracy catching is realized by using the images from the cameras placed on the fingers directly for visual servoing.

1) Correction Without Considering Orientation: First, consider one of the cameras on the multi-vision hand. Let $\left[{}^{\mathrm{cam}}\dot{\boldsymbol{t}}_{\mathrm{w}}\ {}^{\mathrm{cam}}\boldsymbol{\omega}_{\mathrm{w}}\right]^{\mathrm{T}} \in \mathbb{R}^6$ be the camera velocity/angular velocity and ${}^{\mathrm{w}}\dot{\boldsymbol{x}} \in \mathbb{R}^3$ be the ball velocity. The velocity of the ball projected on the image $\dot{\boldsymbol{m}} \in \mathbb{R}^2$ is expressed as follows:

$$\dot{\boldsymbol{m}} = \boldsymbol{L}_{\boldsymbol{m}} \begin{bmatrix} {}^{\mathrm{cam}}\dot{\boldsymbol{t}}_{\mathrm{w}} \\ {}^{\mathrm{cam}}\boldsymbol{\omega}_{\mathrm{w}} \end{bmatrix} + \boldsymbol{L}_{\boldsymbol{b}}\, {}^{\mathrm{w}}\dot{\boldsymbol{x}}, \tag{2}$$

$$\boldsymbol{L}_{\boldsymbol{m}} \coloneqq \frac{1}{{}^{\mathrm{cam}}z} \overline{\boldsymbol{A}} \left[\boldsymbol{I}_3 \,|\, -[{}^{\mathrm{cam}}\boldsymbol{R}_{\mathrm{w}}\, {}^{\mathrm{w}}\boldsymbol{x}]_{\times} \right] \in \mathbb{R}^{2\times 7} \tag{3}$$

$$\boldsymbol{L}_{\boldsymbol{b}} \coloneqq \frac{1}{{}^{\mathrm{cam}}z} \overline{\boldsymbol{A}}\, {}^{\mathrm{cam}}\boldsymbol{R}_{\mathrm{w}}\, {}^{\mathrm{w}}\dot{\boldsymbol{x}} \ \in \mathbb{R}^{2\times 3} \tag{4}$$

where $\overline{\boldsymbol{A}} \in \mathbb{R}^{2\times 3}$ indicates the top two rows of the internal parameter $\boldsymbol{A} \in \mathbb{R}^{3\times 3}$, ${}^{\mathrm{cam}}\boldsymbol{R}_{\mathrm{w}} \in \mathbb{R}^{3\times 3}$ is the rotation matrix from the camera-specific coordinate system (Σ_{cam}) to Σ_{w}, ${}^{\mathrm{cam}}z$ is the ball depth position in Σ_{cam}, $\boldsymbol{I}_{\mathrm{n}} \in \mathbb{R}^{n\times n}$ is the identity matrix, and $[\,]_{\times}$ is a skew matrix equivalent to the outer product calculation.

If the Jacobian matrix $\boldsymbol{J}$ is defined as $\left[{}^{\mathrm{cam}}\dot{\boldsymbol{t}}_{\mathrm{w}}, {}^{\mathrm{cam}}\boldsymbol{\omega}_{\mathrm{w}}\right]^{\mathrm{T}} = \boldsymbol{J}\dot{\boldsymbol{q}}_{\text{cam-0}}$ with the joint angle from the hand-eye to the robot base $\boldsymbol{q}_{\text{cam-0}}$, the displacement of the joint angle change $\Delta\boldsymbol{q}_{\text{cam-0}}$, the ball displacement in the world coordinate $\Delta{}^{\mathrm{w}}\boldsymbol{x} \in \mathbb{R}^3$, and the ball displacement on the image $\Delta\boldsymbol{m} \in \mathbb{R}^2$ are related as follows:

$$\Delta\boldsymbol{m} = \boldsymbol{L}_{\boldsymbol{m}}\boldsymbol{J}\Delta\boldsymbol{q}_{\text{cam-0}} + \boldsymbol{L}_{\boldsymbol{b}}\Delta{}^{\mathrm{w}}\boldsymbol{x}. \tag{5}$$

However, if this method is extended to the multi-vision system, it is necessary to calculate the Jacobian matrix $\boldsymbol{J}$ from all cameras to the robot base, and thus the computation cost becomes heavy. Therefore, assuming that the joint

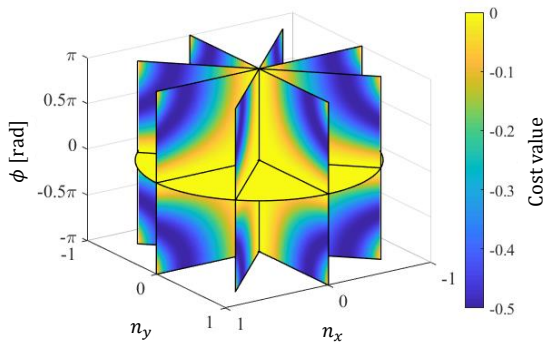


Fig. 7. Cost function E_{Rz}.

angles of the hand are constant during the catching operation will simplify the calculation. This indicates that the relative relationship of the coordinate system fixed to the hand (Σ_{tool}) and Σ_{cam} is constant.

Under this assumption, the relative movement of Σ_{w}, Σ_{tool}, and Σ_{cam} is expressed as follows:

$$\begin{bmatrix} {}^{\mathrm{cam}}\dot{\boldsymbol{t}}_{\mathrm{w}} \\ {}^{\mathrm{cam}}\boldsymbol{\omega}_{\mathrm{w}} \end{bmatrix} = \boldsymbol{V} \begin{bmatrix} {}^{\mathrm{w}}\dot{\boldsymbol{t}}_{\mathrm{tool}} \\ {}^{\mathrm{w}}\boldsymbol{\omega}_{\mathrm{tool}} \end{bmatrix}, \tag{6}$$

$$\boldsymbol{V} = \begin{bmatrix} -{}^{\mathrm{cam}}\boldsymbol{R}_{\mathrm{w}} & -{}^{\mathrm{cam}}\boldsymbol{R}_{\mathrm{w}}[{}^{\mathrm{w}}\boldsymbol{t}_{\mathrm{tool}}]_{\times} \\ \boldsymbol{0} & -{}^{\mathrm{cam}}\boldsymbol{R}_{\mathrm{w}} \end{bmatrix} \begin{bmatrix} {}^{\mathrm{w}}\dot{\boldsymbol{t}}_{\mathrm{tool}} \\ {}^{\mathrm{w}}\boldsymbol{\omega}_{\mathrm{tool}} \end{bmatrix}. \tag{7}$$

Thus, the relationship among $\Delta\boldsymbol{m}$, $\Delta^{\mathrm{w}}\boldsymbol{x}$, and the displacement of the joint angle from the robot hand to the robot base $\Delta\boldsymbol{q}$ is expressed as follows:

$$\Delta\boldsymbol{m} = \boldsymbol{L_m} \boldsymbol{V} \boldsymbol{J_{tool}} \Delta\boldsymbol{q} + \boldsymbol{L_b} \Delta^{\mathrm{w}}\boldsymbol{x}, \tag{8}$$

where $\boldsymbol{J_{tool}}$ is the Jacobian matrix between the joint angle $\Delta\boldsymbol{q}$ and the position/orientation of Σ_{tool}.

By applying (8) to N cameras that can capture the target ball, the following equation is obtained:

$$\Delta\boldsymbol{m}_{\mathrm{all}} = \boldsymbol{J_m} \Delta\boldsymbol{q} + \boldsymbol{J_b} \Delta^{\mathrm{w}}\boldsymbol{x}, \tag{9}$$

where

$$\boldsymbol{m}_{\mathrm{all}} \coloneqq \left[\boldsymbol{m}_1{}^{\mathrm{T}} \ \cdots \ \boldsymbol{m}_{\mathrm{N}}{}^{\mathrm{T}} \right]^{\mathrm{T}}, \in \mathbb{R}^{2\mathrm{N}} \tag{10}$$

$$\boldsymbol{J_m} \coloneqq \left[(\boldsymbol{L}_{\boldsymbol{m}1}\boldsymbol{V}_1)^{\mathrm{T}} \ \cdots \ (\boldsymbol{L}_{\boldsymbol{m}\mathrm{N}}\boldsymbol{V}_{\mathrm{N}})^{\mathrm{T}} \right]^{\mathrm{T}} \boldsymbol{J_{tool}} \tag{11}$$

$$\boldsymbol{V}_{\mathrm{N}} \coloneqq \begin{bmatrix} -{}^{\mathrm{camN}}\boldsymbol{R}_{\mathrm{w}} & -{}^{\mathrm{camN}}\boldsymbol{R}_{\mathrm{w}}[{}^{\mathrm{w}}\boldsymbol{t}_{\mathrm{tool}}]_{\times} \\ \boldsymbol{0} & -{}^{\mathrm{camN}}\boldsymbol{R}_{\mathrm{w}} \end{bmatrix} \tag{12}$$

$$\boldsymbol{J_b} \coloneqq \left[\boldsymbol{L}_{\boldsymbol{b}1}^{\mathrm{T}} \ \cdots \ \boldsymbol{L}_{\boldsymbol{b}\mathrm{N}\cdot}^{\mathrm{T}} \right]^{\mathrm{T}} \tag{13}$$

By solving (9), the following simple control method is obtained.

$$\Delta\boldsymbol{q} = \boldsymbol{J_m^+} \left[\Delta\boldsymbol{m}_{\mathrm{all}} - \boldsymbol{J_b} \Delta^{\mathrm{w}}\boldsymbol{x} \right] \tag{14}$$

$$= \boldsymbol{J_m^+} \left[(\boldsymbol{m}_{\mathrm{d}} - \boldsymbol{m}_{\mathrm{all}}) - \boldsymbol{J_b} \Delta^{\mathrm{w}}\boldsymbol{x}, \right] \tag{15}$$

where $\boldsymbol{m}_{\mathrm{d}} = [\hat{\boldsymbol{m}}_1 \ \cdots \ \hat{\boldsymbol{m}}_{\mathrm{N}}] \in \mathbb{R}^{2\mathrm{N}}$ is the desired position on the image planes, and $\boldsymbol{J_m^+}$ means the pseudo inverse of $\boldsymbol{J_m}$. $\boldsymbol{m}_{\mathrm{d}}$ is given as the virtual perspective projection from the

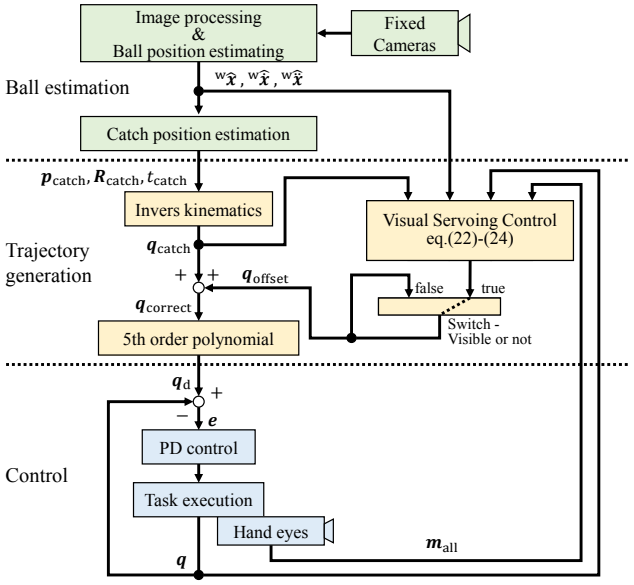


Fig. 8. Ball catching control flow.

estimated position of the ball ${}^{\mathrm{w}}\hat{\boldsymbol{x}}$ to the image planes as follows:

$$\begin{bmatrix} \boldsymbol{m}_{\mathrm{d}} \\ 1 \end{bmatrix} = \frac{1}{{}^{\mathrm{cam}}\hat{z}(\boldsymbol{q}_{\mathrm{catch}})} \boldsymbol{P} \left[{}^{\mathrm{cam}}\boldsymbol{R}_{\mathrm{tool}} \mid {}^{\mathrm{cam}}\boldsymbol{t}_{\mathrm{tool}} \right] {}^{\mathrm{tool}}\boldsymbol{T}_{\mathrm{w}}(\boldsymbol{q}_{\mathrm{catch}}) \begin{bmatrix} {}^{\mathrm{w}}\hat{\boldsymbol{x}} \\ 1, \end{bmatrix} \tag{16}$$

where ${}^{\mathrm{tool}}\boldsymbol{T}_{\mathrm{w}} \in \mathbb{R}^{4\times4}$ is a homogeneous transformation matrix from Σ_{tool} to Σ_{w}, and $\boldsymbol{q}_{\mathrm{catch}}$ is the joint angle in the uncorrected state.

2) Correction considering orientation: In the method described in the previous section, the orientation of the hand is not uniquely determined. Therefore, a cost function of the orientation is added in this section. Let ${}^{\mathrm{w}}\boldsymbol{R}_{\mathrm{tool}} \in \mathbb{R}^{3\times3}$ be the current orientation and ${}^{\mathrm{w}}\boldsymbol{R}_{\mathrm{catch}} \in \mathbb{R}^{3\times3}$ be the orientation in the uncorrected condition:

$${}^{\mathrm{w}}\boldsymbol{R}_{\mathrm{tool}} = \left[{}^{\mathrm{R}}\boldsymbol{x} \, , {}^{\mathrm{R}}\boldsymbol{y} \, , {}^{\mathrm{R}}\boldsymbol{z} \right] \tag{17}$$

$${}^{\mathrm{w}}\boldsymbol{R}_{\mathrm{catch}} = \left[{}^{\mathrm{R}}\boldsymbol{x}_{\mathrm{catch}}, {}^{\mathrm{R}}\boldsymbol{y}_{\mathrm{catch}}, {}^{\mathrm{R}}\boldsymbol{z}_{\mathrm{catch}} \right]. \tag{18}$$

The displacement of the orientation of the z-axis ${}^{\mathrm{Rz}}\boldsymbol{e} \in \mathbb{R}^3$ is given as follows:

$${}^{\mathrm{Rz}}\boldsymbol{e} = {}^{\mathrm{R}}\boldsymbol{z}_{\mathrm{catch}} \times {}^{\mathrm{R}}\boldsymbol{z}. \tag{19}$$

Here, ${}^{\mathrm{Rz}}\boldsymbol{e}$ is not influenced by rotation around the z-axis. The z-axis represents the normal vector in the palm plane. By permitting rotation around the z-axis, as shown in Fig. 6, the orientation of the hand is flexibly adjusted corresponding to the correction of the catch position.

The cost function E_{Rz} is defined as

$$E_{\mathrm{Rz}} \coloneqq -\frac{1}{2} \left({}^{\mathrm{Rz}}\boldsymbol{e} \right)^{\mathrm{T}} \left({}^{\mathrm{Rz}}\boldsymbol{e} \right). \tag{20}$$

The potential of E_{Rz} is shown in Fig. 7 as a color map. The current orientation ${}^{\mathrm{w}}\boldsymbol{R}_{\mathrm{tool}}$ is given as $\boldsymbol{R}_{\phi} \in \mathbb{R}^{3\times3}$. It is expressed with the rotation angle ϕ and the rotation axis $\boldsymbol{n} = \left[n_{\mathrm{x}}, n_{\mathrm{y}}, 1 - (n_{\mathrm{x}}^2 + n_{\mathrm{y}}^2)^{1/2} \right]^{\mathrm{T}}$. The desired posture is set to

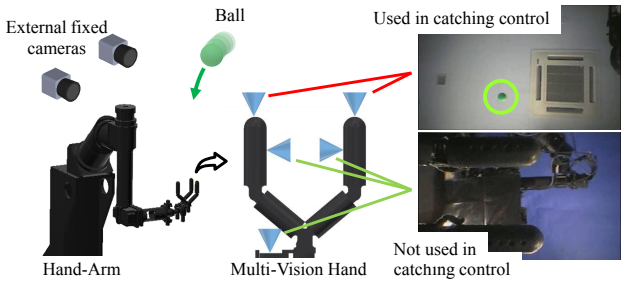


Fig. 9. Experimental setup.

$\boldsymbol{R}_{\phi=0} = \boldsymbol{I}_3$. In Fig. 7, the cost value is high even if not only ϕ is small but also n_x or n_y is small. This is means that when the rotational displacement of x- or y-axis rotation is large, the cost function becomes large.

The Jacobian matrix between the cost function and the joint angle vector $\boldsymbol{q}$ is obtained as:

$$\boldsymbol{J}_e = \frac{\partial^{\mathrm{Rz}} e}{\partial \boldsymbol{q}} = {}^{\mathrm{R}}\boldsymbol{z}_{\mathrm{catch}} \times \frac{\partial^{\mathrm{R}} \boldsymbol{z}}{\partial \boldsymbol{q}}. \tag{21}$$

Thus, the correction value of the joint angle is obtained by the inverse calculation of (9) and (21):

$$\Delta \boldsymbol{q} = \begin{bmatrix} \boldsymbol{J}_m \\ \boldsymbol{J}_e \end{bmatrix}^{+} \begin{bmatrix} \Delta \boldsymbol{m}_{\mathrm{all}} - \boldsymbol{J}_b \Delta^{\mathrm{w}} \boldsymbol{x} \\ \Delta^{\mathrm{Rz}} e. \end{bmatrix} \tag{22}$$

In this method, if visual information is not obtained, $\Delta \boldsymbol{q}$ cannot be calculated, and the desired joint angle cannot be corrected. This happens when the ball is out of sight or the system fails to visually detect the ball. To solve this problem, we define the offset angle as follows:

$$\boldsymbol{q}_{\mathrm{offset}} = \Delta \boldsymbol{q} - \boldsymbol{q}_{\mathrm{catch}} + \boldsymbol{q}, \tag{23}$$

where $\boldsymbol{q}$ is the current joint angle, $\Delta \boldsymbol{q}$ is the value calculated by (22), and $\boldsymbol{q}_{\mathrm{catch}}$ is the joint angle in the uncorrected state. By updating the offset only when visual information is obtained, a sudden change of the desired joint angle is suppressed.

From (23), the catching position was corrected by correcting the joint angle as follows:

$$\boldsymbol{q}_{\mathrm{correct}} = \boldsymbol{q}_{\mathrm{catch}} + \boldsymbol{q}_{\mathrm{offset}}. \tag{24}$$

The flowchart for ball catching is shown in Fig. 8.

VI. Experiment

The experimental conditions were as follows:

- The ball was thrown from a toss machine ($t = 0$ s), and the catch position was determined after convergence of the Kalman filter. Then, when the ball entered the field of view of the multi-vision hand, correction of the catch position started.
- The position of the ball was estimated using the external fixed cameras. In this catch task, only the three cameras at the fingertips of the multi-vision hand were used because the target was not observed in the field of view of the other cameras. (Fig. 9).

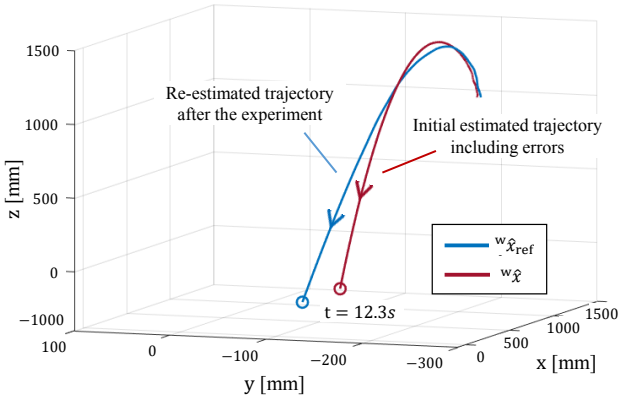


Fig. 10. Estimated ball position.

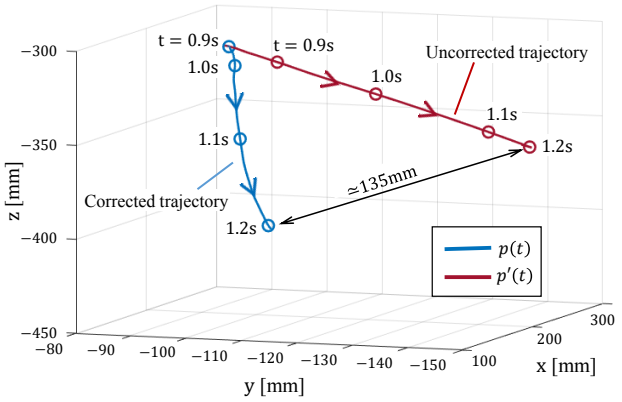


Fig. 11. Trajectory of hand position.

- The experiment was performed while intentionally lowering the estimation accuracy of the ball state calculated by using external fixed cameras, and it was confirmed whether the correction of the catch position was appropriate.
- The frame rate of the external fixed cameras were 500 Hz, and that of the multi-vision hand was 60 Hz, for stability in the experiment.

This section shows the results when the effect of correcting the catch position was remarkable.

The experimental result is shown as a sequence of continuous images in Fig. 12. The Kalman filter converged at $t = 0.69$ s, and the initial catch position was determined. Because the ball was within the field of view of the multi-vision hand at $t = 0.69$ s, the joint correction started at the same time. After that, the ball catching was successful at $t = 1.23$ s.

Fig. 10 shows the initial estimated ball position ${}^{\mathrm{w}}\hat{\boldsymbol{x}}$ and the corrected ball position ${}^{\mathrm{w}}\hat{\boldsymbol{x}}_{\mathrm{ref}}$. ${}^{\mathrm{w}}\hat{\boldsymbol{x}}_{\mathrm{ref}}$ was not used in the visual servoing, and it was re-estimated offline for comparison with ${}^{\mathrm{w}}\boldsymbol{x}$. Also, the corrected hand trajectory $\boldsymbol{p}(t)$ and the uncorrected hand trajectory $\boldsymbol{p}'(t)$ are shown in Fig. 11.

In Fig. 10, the error between the initial estimated position and the re-estimated position gradually increases. This is the error in the initial estimation of the Kalman filter, and is also the amount corrected by the visual servoing. The size

is about 60 mm in the x-direction and about 130 mm in the z-direction at the catching time.

If the error between the actual catch position and the estimated catch position is less than 15 mm, the hand can catch the ball without correction of the catch position. On the other hand, the catch position was corrected up to about 135 mm in the experiment. This means that the proposed correction by the multi-vision hand was effective for achieving robust catching.

VII. Conclusion

In this paper, we have proposed a multi-vision hand system in which a number of small high-speed cameras are arranged on the surface of a robot hand. We have also proposed correction of the catching position by visual servoing control using a multi-vision system combining the multi-vision hand and fixed external high-speed cameras. The experiment was performed under the condition that the ball state estimation included an error. As a result, the catch operation was successful by applying correction by the visual servoing, and the effectiveness of the proposed method was confirmed. Our future work will focus on extension of the proposed multi-vision hand.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP1069176

References

[1] A. Namiki and M. Ishikawa, Robotic catching using a direct mapping from visual information to motor command, 2003 IEEE International Conference on Robotics and Automation, pp. 2400-2405.
[2] T. Kizaki and A. Namiki, Two Ball Juggling with High-speed Hand-Arm and High-speed Vision System, IEEE International Conference on Robotics and Automation, pp. 1372-1377,June, 2012
[3] T. Oka and A. Namiki, Ball Juggling Robot System Controlled by High-Speed Vision, 2017 IEEE International Conference on Cyborg and Bionic Systems, pp. 91-96, January, 2017.
[4] A. Takahashi, N. Koumura and A. Namiki, Optimal Trajectory for High-Speed Dual Arm Juggling Robot, the Robotics Society of Japan Annu. Conf., 1H1-06, 2019(in Japanese).
[5] F. Chaumette, S. Hutchinson. Visual Servo Control, Part I: Basic Approaches, IEEE Robotics and Automation Magazine, pp. 82-90, December 2006.
[6] F. Chaumette, S. Hutchinson. Visual Servo Control, Part II: Advanced Approaches, IEEE Robotics and Automation Magazine, pp. 109-118, March 2007.
[7] S. Morikawa, T. Senoo, A. Namiki, and M. Ishikawa, Realtime collision avoidance using a robot manipulator with light-weight small high-speed vision systems, IEEE International Conference on Robotics and Automation (ICRA'07), pp. 794-799, April, 2007.
[8] A. Yamaguchi and C. G. Atkeson, Combining Finger Vision and Optical Tactile Sensing: Reducing and Handling Errors While Cutting Vegetables, IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids), pp. 1045-1051, November, 2016
[9] R. Patel, R. Cox and N. Correll, Integrated proximity, contact and force sensing using elastomer-embedded commodity proximity sensors, Autonomous Robots, vol. 42, no. 7 pp. 1443-1458, 2018.
[10] K. Koyama, K. Murakami, T. Senoo, M. Shimojo, M. Ishikawa, High-Speed, Small-Deformation Catching of Soft Objects Based on Active Vision and Proximity Sensing, IEEE Robotics and Automation letters, Vol. 4, No. 2, pp. 578-585, April, 2019.
[11] B. Bäuml, O. Birbach, T. Wimböck, U. Frese, A. Dietrich, and G. Hirzinger, Catching flying balls with a mobile humanoid: Systemoverview and design considerations, IEEE-RAS International Conference on Humanoid Robots, pp.513-520, Octorber, 2011.
[12] P. Cigliano, V. Lippiello, F. Ruggiero, and B. Siciliano, Robotic Ball Catching with an Eye-in-Hand Single-Camera System, IEEE Transactions on Control Systems Technology, Vol. 23, No. 5, September, 2015.
[13] I. Ishii, T. Tatebe, Q. Gu, Y. Moriue, T. Takaki, and K. Tajima, 2000fps Real-time Vision System with High-frame-rate Video Recording, IEEE Int. Conf. on Robotics and Automation Anchorage Convention District, pp.1536-1541, July, 2010.
[14] A. Namiki, Y. Imai, M. Ishikawa, and M. Kaneko, Development of a High-speed Multi-Fingered Hand System and Its Application to Catching, EEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 2666-2671, October, 2003.
[15] T. Sugihara, Solvability-Unconcerned Inverse Kinematics by the Levenberg-Marquardt Method, IEEE Transactions on Robotics, Volume. 27, Issue. 5, October, 2011.

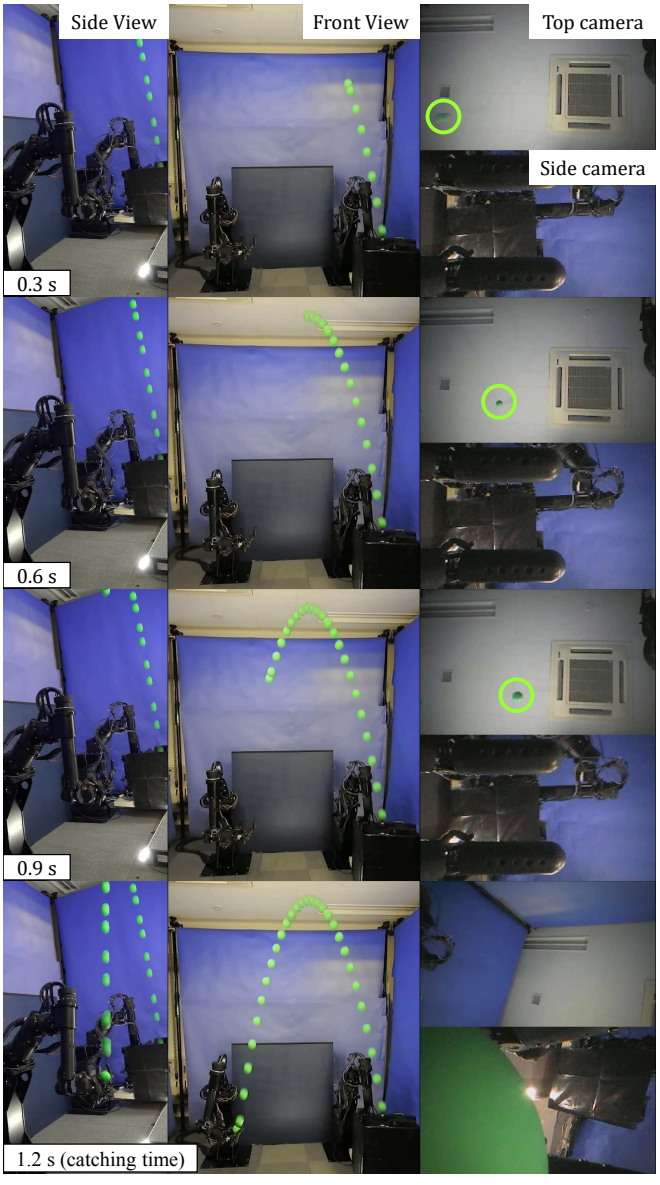


Fig. 12. Catching motion Continuous photos of catching: (Left) Side view, (Center) Front View, (Upper right) tip camera of multi-vision hand, (Lower right) link surface camera of multi-vision hand.