IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 31 December 2023, accepted 23 February 2024, date of publication 27 March 2024, date of current version 4 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3382204

RESEARCH ARTICLE

Optimal Projection Pattern for Active Visual Servoing (AVS)

SHOGO ARAI[1], (Member, IEEE), YOSHIHIRO MIYAMOTO[2], AKINARI KOBAYASHI[3,4], AND KAZUHIRO KOSUGE[4,5], (Life Fellow, IEEE)

[1]Department of Mechanical and Aerospace Engineering, Faculty of Science and Engineering, Tokyo University of Science, Tokyo 162-8601, Japan
[2]Graduate School of Engineering, Tohoku University, Sendai 980-8577, Japan
[3]Centre for Garment Production Ltd., Hong Kong, SAR
[4]JC STEM of Laboratory of Robotics for Soft Materials, Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, SAR
[5]Transformative AI and Robotics International Research Center, Tohoku University, Sendai 980-8577, Japan

Corresponding author: Shogo Arai (arai.shogo@rs.tus.ac.jp)

ABSTRACT Visual servo control uses images that are obtained by a camera for robotic control. This study focuses on the problem of positioning a target object using a robotic manipulator with image-based visual servo (IBVS) control. To perform the positioning task, IBVS requires visual features that can be extracted from the appearance of the target object. Therefore, a positioning error tends to increase especially for textureless objects, such as industrial parts, since it is difficult to extract differences of the visual features between current and goal images. To solve these problems, this paper presents a novel visual servoing named "Active Visual Servoing." Active Visual Servoing (AVS) projects patterned light onto the target object using a projector. The design of the projection pattern affects the positioning error. AVS uses an optimal pattern which is theoretically derived and maximizes differences between current and goal images. The experimental results show that the proposed AVS reduces the positioning error by more than 59% compared to conventional IBVS.

INDEX TERMS Grasping error, kitting, positioning, visual servoing.

I. INTRODUCTION

Visual servoing is a widely used method for positioning [1], [2], [3], [4], [5], [6]. One important application of visual servoing is positioning task in factory automation since the positioning task is required after random bin-picking [7], [8] to correct grasping error deviated from grasp planning [9] and perform assembling tasks [10]. This study, in particular, considers the problem of positioning an object grasped by a hand attached to a manipulator. The visual servoing is roughly classified into position-based (PBVS) [11] and image-based visual servoing (IBVS) [12], and this paper focuses on IBVS.

IBVS calculates the control input such that the error of the visual feature between the image of the target object at the current time and the goal image converges to zero [13], as shown in Fig. 1. Because IBVS does not need to estimate the pose of the target object from images, the positioning accuracy of IBVS is not influenced by the camera-robot calibration error, the image quantization error, and the modeling error of the camera.[1]

Numerous visual servo techniques, some leveraging deep learning, have been introduced and effectively employed to address relatively significant positioning errors. However, these methods aren't apt for FA because of positioning discrepancies spanning from 5 mm to several centimeters. We term this the "last one-inch problem." Additionally, achieving precise positioning for objects lacking distinct visual features, like those without texture, presents challenges. Even if the visual features can be extracted for the object, a convergence error (positioning error) tends to remain due to small deviation of the visual feature between the target and current pose. Further, it is necessary to extract the image

The associate editor coordinating the review of this manuscript and approving it for publication was Shun-Feng Su.

[1]All of them are problems of PBVS.

features,[2] to convert them to visual features for IBVS, and to track them between the goal and current images at each time. These processes are not feasible especially for objects with poor image features. These factors frequently hinder the integration of visual servo technology into pick-and-place tasks and the assembly of industrial components in factory automation processes.

To solve these problems, we propose a novel visual servoing named "Active Visual Servoing" (AVS). AVS uses a projector to irradiate an object with a light pattern and a camera to capture the reflected light. By directly using the intensity values of each pixel in the captured image as the visual feature, visual servoing is performed, as shown in Fig. 2.

The positioning accuracy of the proposed active visual servoing mainly depends on the projection pattern. Image deviation of the objects with poor features becomes too small near the goal pose. We theoretically derive the optimal light pattern that maximizes image deviation near the goal pose to reduce the convergence of the positioning error. From this point of view, AVS proposed in this paper is a completely new attempt to achieve high accuracy positioning, while existing method uses the structured light which in not optimized for IBVS.

AVS has the following properties:

1) The proposed active visual servoing retains the advantages of IBVS, that is, the positioning accuracy does not depend on the calibration, quantization, and modeling errors of the camera.
2) Since the proposed method uses the intensity values of the image as the visual feature and does not require extraction of the feature from the image, the proposed active visual servoing can easily applied to the objects with poor image features, such as textureless objects, industrial parts, and so on.
3) On the basis of some assumptions, we have derived the optimal light pattern that maximizes image deviation near the goal pose of the target object.
4) The proposed method has high positioning accuracy compared to the conventional IBVS, which was verified through comparative experiments using actual industrial parts.
5) Active Visual Servo method presented in this paper is comparatively straightforward. Notably, for pre-constructed visual servoing systems, the only hardware modification required is the addition of a projector. Time-consuming and costly calibrations are entirely unnecessary.

This paper is organized as follows: Sec. II describes related works. Sec. III proposes active visual servo control and sec. IV derives the optimal light pattern irradiating from the projector for active visual servoing. Sec. V presents the experimental results for the validation of AVS and sec. VI concludes this paper.

This paper uses the following notations. The symbols $\mathbf{0}_n$ and $\mathbf{1}_n$ represent n-dimensional vectors whose all elements are 0 and 1, respectively.

II. RELATED WORKS

In the field of three-dimensional (3D) measurement, a number of methods have been proposed [14], [15], [16]. Besides, the obtained 3D point cloud or depth image are utilized for various robotic applications involving robot vision, such as bin-picking [17], [18], prediction of reaching motion [19], [20], 3D keypoint detection [21], [22], 3D feature description [23], segmentation [24], [25], [26], and SLAM [27]. An example of the representative three-dimensional measurement method is the light-section method, in which 3D measurement is performed by projecting a line of light and measuring its reflected light. The light-section method takes some time for the measurement because it is necessary to scan multiple lines of the light for measuring the entire surface of the target object. This paper focuses on active stereo methods that do not require line by line scanning. Active stereo methods perform three-dimensional measurement by irradiating a target object with a pre-designed light pattern and by measuring the light reflected by the object [28], [29], [30]. As the representative methods, space coding [30] and phase shift methods [31] have been proposed and various other methods using the camera and projector have been developed [29], [32], [33]. In these methods, the design of the patterned light is a major factor determining the accuracy and measurement time. Major advantages of the active stereo methods are that they are robust to changes in ambient light and they can perform 3D measurement of an object irrespective of the presence or absence of a surface pattern on the object. Inspired by a concept of the active stereo methods, this paper proposes the active visual servoing that accomplishes the object positioning with higher accuracy than the conventional IBVS.

Image-based visual servoing using a laser have been proposed in [34], [35], and [36]. Pagès et al. have proposed visual servoing with lasers for positioning the camera attached to the tip of the end effector [35]. This method involves the use of four lasers to irradiate a planar target, and visual feature is calculated by using parts of images surrounding the irradiated points. The authors have shown that this method can achieve robust positioning of the camera and decouple control of rotation around the optical axis of the camera from the other two axes. However, because of the complexity of the interaction matrix, this method does not guarantee global stability. They have subsequently proposed a solution of this problem by using a simple visual feature composed of only the image coordinates of irradiated points [36]. This approach has shown global stability under ideal circumstances, i.e., global stability is guaranteed when

[2]In this paper, the image feature represents the feature directly extracted from captured image, such as point, edge, SIFT, SURF. On the other hand, the visual feature is computed from the image feature and used for IBVS, such as the image coordinates of the corners, edges, and keypoints for SIFT or SURF.

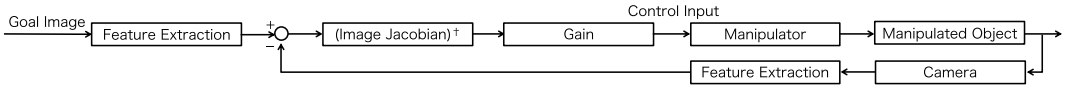


FIGURE 1. Block diagram of image-based visual servo system.

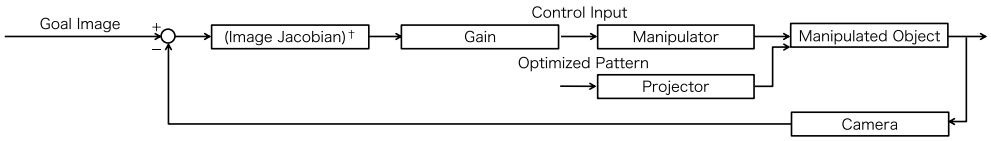


FIGURE 2. Block diagram of active visual servo system proposed in this paper.

a placement error between the camera and laser is zero. Experimental results have shown that these visual servoing methods performed positioning of the camera with respect to a planar target object.

In addition to the above-mentioned work, a visual servo control method that projects a pattern onto an object by using a projector has been proposed [37], [38]. This method entails the generation of dot patterns with three colors, red, green, blue, based on the M-array and the projection of these patterns onto the target object. The method functions by matching the projected dots between the current and goal images; then, by using the matching results, the method performs positioning based on IBVS. The projected patterns based on the M-array facilitate identification of the correspondence dots among images captured by cameras with different viewpoints. The method requires processes of segmentation to extract dots, computing the position of the centroid of the dots, finding of the dots nearest to each of the dots, and checking the graph consistency. The effectiveness of the method for objects with complicated shapes have not been verified, although verification experiments with planar and elliptic cylinders have been conducted.

The advantages of the method proposed in this paper are as follows:

1) Image processing for the proposed active visual servo method is high speed, since the intensity value of each pixel is directly used as the visual feature;
2) Existing structured light is not used. Instead, we analytically derive the optimal patterned light to be irradiated by the projector such that the cost function is minimized;
3) The experimental results confirm that the proposed method can perform positioning of complex-shaped objects with higher accuracy than the conventional visual servoing.

III. PRINCIPLE OF IMAGE-BASED VISUAL SERVO CONTROL

This section describes the basic principles of IBVS. The purpose here is to position a target object grasped by a robotic hand equipped with a manipulator. A camera fixed to the world system captures the target object. The goal image is obtained by capturing the target object located at the goal pose.

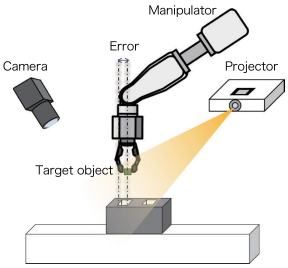


FIGURE 3. System configuration for proposed active visual servoing.

IBVS provides the control input to the manipulator by using the visual features computed from the captured image, as shown in Fig. 1. The visual feature is a multidimensional vector and can be computed based on the image feature, such as the edges, counters, corners, SIFT [39], SURF [40], AKAZE [41], etc., without using the relative pose of the camera to the manipulator and the intrinsic parameter of the camera. The purpose of IBVS is to minimize the objective function defined by

$$e(t) = f(I(t)) - f(I^*), \tag{1}$$

where I and I^* represent vectors whose elements are intensities of each pixel in the current and goal images, respectively. The symbol $f(I)$ denotes the mapping from image I to the visual feature. The simplest control law to achieve this purpose is represented by

$$\dot{\theta}_d = -\lambda J^{\dagger} \left(f(I(t)) - f(I^*) \right), \tag{2}$$

where $\dot{\theta}_d \in \mathbb{R}^n$ is the control input of the joint angular velocity for the manipulator, n is the degree of freedom (DOF) of the manipulator, $\lambda \in \mathbb{R}$ is the gain, and $J^{\dagger}$ is the pseudo-inverse matrix of the image Jacobian.

IV. ACTIVE VISUAL SERVO CONTROL SCHEME AND OPTIMAL PROJECTION PATTERN

This section proposes the active visual servo control scheme. The proposed method uses a projector to project optimized patterned light onto the target object and a camera to capture an image of the reflected light, as shown in Fig. 3. Then the control input is computed by multiplying a pseudo-inverse

matrix of image Jacobian by an image difference between current and goal images, as shown in Fig. 2. This section presents the control law for active visual servoing and derives the optimal projection pattern that affects the positioning accuracy.

A. CONTROL LAW OF ACTIVE VISUAL SERVOING

This subsection explains the control law of the active visual servoing. Active visual servo control directly utilizes the intensity of each pixel in the captured image as the visual feature, which is inspired by direct visual servoing [42]. The control law of AVS is shown below:

$$\dot{\theta}_d = -\lambda J^{\dagger}(I(t) - I^*). \tag{3}$$

The control law (3) can be obtained by replacing the visual feature f with the image I in (2). The control law (3) does not require computation for the extraction of image features and thus can be computed faster than one by (2).

The pseudo-inverse matrix of the image Jacobian $J^{\dagger}$ in (3) represents a mapping from the error between the current and goal images to the joint angular velocity and depends on the joint angles of the manipulator. In this paper, the image Jacobian is assumed to be the constant and approximated by the one at the goal pose.

B. OPTIMAL PROJECTION PATTERN FOR ACTIVE VISUAL SERVOING

This subsection derives the optimal projection pattern for AVS. Let us consider a three-dimensional space as shown in Fig. 4. The manipulator, projector, and camera are fixed in the space and the coordinate system Σ_c and Σ_p are set such that the origins coincide with the their optical centers and the z-axes are parallel to the optical axis of the camera and projector, respectively. In the following discussions, poses are represented based on the camera coordinate system unless otherwise noted.

Now we assume the simplest pinhole camera model, that is, a 3D point located at (x, y, z) is projected at

$$X_c = \frac{f_{cx}}{z}x, \tag{4}$$

$$Y_c = \frac{f_{cy}}{z}y \tag{5}$$

in the image plane of the camera, where f_{cx} and f_{cy} are the focal lengths.

The projector is set at $\xi_p := [x_p, y_p, z_p, \theta_p, \phi_p, \psi_p]^{\top}$, where (x_p, y_p, z_p) is the position of the projector and $(\theta_p, \phi_p, \psi_p)$ is the attitude of the projector. In other words, the projector coordinate system is set at ξ_p in the camera coordinate system. It is assumed that the optical model of the projector also follows the pinhole camera model.

We denote the poses of the target object at the current time and the goal pose by $\xi_o := [x_o, y_o, z_o, \theta_o, \phi_o, \psi_o]^{\top}$ and $\xi_o^* := [x_o^*, y_o^*, z_o^*, \theta_o^*, \phi_o^*, \psi_o^*]^{\top}$, respectively. The surface shape of the target object at the pose $[0, 0, 0, 0, 0, 0]^{\top}$ is represented by $[\alpha(\tau), \beta(\tau), \gamma(\tau)]^{\top}$, where $\tau \in \mathbb{R}^2$, $\mathbf{0}_2 \leq \tau \leq \mathbf{1}_2$ is a

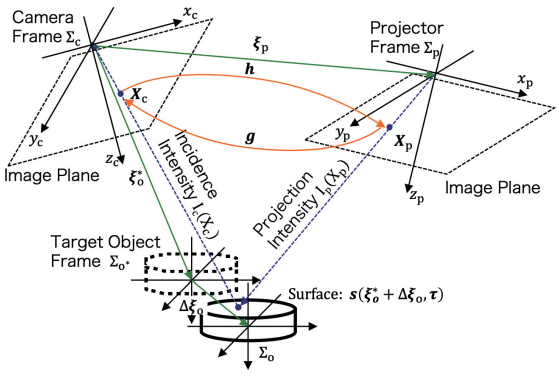


FIGURE 4. Arrangement of a camera, projector, and target object in a problem to the derive optimal projection pattern.

parameter used for representing a curved surface. By using this notation, the surface of the target object at pose ξ_o can be represented by

$$s(\xi_o, \tau) := R(\theta_o, \phi_o, \psi_o)\begin{bmatrix}\alpha(\tau)\\ \beta(\tau)\\ \gamma(\tau)\end{bmatrix} + \begin{bmatrix}x_o\\ y_o\\ z_o\end{bmatrix}, \tag{6}$$

where $R(\theta_o, \phi_o, \psi_o)$ represents a rotation matrix.

The light projected from $\boldsymbol{X}_p := [X_p, Y_p]^{\top} \in \mathbb{R}^2$ in the image plane of the projector is reflected at $s(\xi_o^*, \tau)$ on the surface of the object with pose ξo^* and then reaches $\boldsymbol{X}_c := [X_c, Y_c]^{\top} \in \mathbb{R}^2$ in the image plane of the camera. By using a mapping g from the image plane of the projector to that of the camera, this relationship can be denoted by

$$\boldsymbol{X}_c = g(\boldsymbol{X}_p, s(\xi_o^*, \tau), \xi_p). \tag{7}$$

Next, we consider the pattern of the light projected by the projector. The function for the projected light pattern is defined by $I_p(\boldsymbol{X}_p)$, where $I_p(\boldsymbol{X}_p)$ is the intensity of the projection at $\boldsymbol{X}_p$ in the image of the projector.

To simplify the discussion, it is assumed that the intensity of the light incident on a certain pixel in the image plane of the camera is equal to the one from the corresponding pixel in the image plane of the projector. By considering that the incident ray on $\boldsymbol{X}_c$ in the image plane of the camera is projected from $\boldsymbol{X}_p$ in the image plane of the projector, the relationship for the intensity

$$I_c(\boldsymbol{X}_c) = I_p(\boldsymbol{X}_p) \tag{8}$$

holds. By using the inverse function of g, which is denoted by h, eq. (8) can be rewritten by

$$I_c(\boldsymbol{X}_c) = I_p(h(\boldsymbol{X}_c, s(\xi_o^*, \tau), \xi_p)). \tag{9}$$

As shown in the above equation, I_c actually relies on the pose and the shape of the object. For later discussion, $I_c(\boldsymbol{X}_c)$ is denoted by $I_c(\boldsymbol{X}_c, s, \xi_o^*)$.

When the target object moves by $\Delta\xi_o$ from the target pose ξ_o^*, the incident ray on the same pixel $\boldsymbol{X}_c$ is considered to be

projected from

$$\boldsymbol{X}_p + \Delta\boldsymbol{X}_p = h(\boldsymbol{X}_c, s(\xi_o^* + \Delta\xi_o, \tau), \xi_p) \tag{10}$$

in the image plane of the projector. Therefore, the intensity observed at $\boldsymbol{X}_c$ in the image plane of the camera is

$$I_c(\boldsymbol{X}_c, s, \xi_o^* + \Delta\xi_o) = I_p(h(\boldsymbol{X}_c, s(\xi_o^* + \Delta\xi_o, \tau), \xi_p)). \tag{11}$$

As shown in eq. (3), active visual servo control uses the intensities of the image as the visual features. To decrease the positioning error and the time required for positioning, we derive the optimal projection pattern that maximizes the image error in (3), that is, the optimal projection pattern is denoted by

$$I_p^* = \arg\max_{I_p} c(I_p), \tag{12}$$

where $c(I_p) = |J(I_p)|^2$. The symbol $J(I_p)$ represents the error of luminance captured by the camera between the goal and current poses of the object, which is defined by

$$J(I_p) = I_c(\boldsymbol{X}_c, s, \xi_o) - I_c(\boldsymbol{X}_c, s, \xi_o^*). \tag{13}$$

By using the relationship given by (9), the Taylor expansion of $J(I_p)$ for in the vicinity of the goal pose is derived as follows:

$$\begin{aligned} J(I_p) &= I_p(h(\boldsymbol{X}_c, s(\xi_o^* + \Delta\xi_o), \xi_p)) - I_p(h(\boldsymbol{X}_c, s(\xi_o^*), \xi_p)) \\ &= I_p\left(h\left(\boldsymbol{X}_c, s(\xi_o^*), \xi_p\right) + \left.\frac{\partial h}{\partial s^\top}\right|_{s=s(\xi_o^*)} \left.\frac{\partial s}{\partial \xi_o^\top}\right|_{\xi_o=\xi_o^*} \Delta\xi_o\right) \\ &\quad - I_p(h(\boldsymbol{X}_c, s(\xi_o^*), \xi_p)) + O(\Delta\xi_o^2) \qquad (14) \\ &= \left.\frac{\partial I_p}{\partial h^\top}\right|_{h=h(\boldsymbol{X}_c, s(\xi_o^*), \xi_p)} \underbrace{\left.\frac{\partial h}{\partial s^\top}\right|_{s=s(\xi_o^*)}}_{A} \underbrace{\left.\frac{\partial s}{\partial \xi_o^\top}\right|_{\xi_o=\xi_o^*}}_{B} \Delta\xi_o \\ &\quad + O(\Delta\xi_o^2), \qquad (15) \end{aligned}$$

where $O(\Delta\xi_o^2)$ is the secondary or higher remainder term of the Taylor expansion and $s(\xi, \tau)$ is denoted by $s(\xi)$ to simplify notations. Neglecting $O(\Delta\xi_o^2)$ and substituting (15) to $c(I)$ derives

$$\begin{aligned} &c(I_p) \\ &= \Delta\xi_o^\top B^\top A^\top \underbrace{\left(\left.\frac{\partial I_p}{\partial \boldsymbol{X}_p^\top}\right|_{\boldsymbol{X}_p=h(\xi_o^*)}\right)^\top \left.\frac{\partial I_p}{\partial \boldsymbol{X}_p^\top}\right|_{\boldsymbol{X}_p=h(\xi_o^*)}}_{X} AB\Delta\xi_o, \end{aligned} \tag{16}$$

where we use the relation $\boldsymbol{X}_p = h(\xi_o^*)$ and $h(\boldsymbol{X}_c, s(\xi_o^*, \tau), \xi_p)$ is denoted by $h(\xi_o^*)$ for simplifying notation. In (15) and (16), the matrices A, B, and X depend on the shape of the target object, the poses of the object, and the projected pattern, respectively. Because we can control only matrix X in $c(I_p)$ by changing the projection pattern. For maximizing $c(I_p)$ for arbitrary vector $\varepsilon := AB\Delta\xi_o$, we needs to maximize the eigenvalues of X that are given by

$$\lambda(X) = \left\{0, \quad \left(\frac{\partial I_p}{\partial X_p}\right)^2 + \left(\frac{\partial I_p}{\partial Y_p}\right)^2\right\}. \tag{17}$$

As shown in eq. (17), first eigenvalue is equal to zero.[3] Hence, we consider the projected pattern which maximize a second eigenvalue:

$$I_p^* = \arg\max_{I_p} \left\{\left(\frac{\partial I_p}{\partial X_p}\right)^2 + \left(\frac{\partial I_p}{\partial Y_p}\right)^2\right\} \tag{18}$$

Eq. (18) means that the optimized projection pattern maximizes the square of first derivative of intensity with the image coordinates X_p and Y_p in the image plane of the projector.

In the above discussion, we assumed the resolutions of both the camera and projector to be infinite. In practice, the resolutions are finite and the image planes of both the camera and projector consist of the finite number of pixels. Considering this point, we denote the projection pattern by $I_{pd}(X_{pd}, Y_{pd})$, $X_{pd} \in \{1, 2, \cdots, w_p\}$, $Y_{pd} \in \{1, 2, \cdots, h_p\}$ where w_p and h_p are the resolutions of the projector. Two terms in right hand side of eq. (18) can be approximated by forward difference as follows:

$$\left(\frac{\partial I_p}{\partial X_p}\right)^2 \approx \left(I_{pd}(X_{pd}^* + 1, Y_{pd}^*) - I_{pd}(X_{pd}^*, Y_{pd}^*)\right)^2 =: \bar{I}_X, \tag{19}$$

$$\left(\frac{\partial I_p}{\partial Y_p}\right)^2 \approx \left(I_{pd}(X_{pd}^*, Y_{pd}^* + 1) - I_{pd}(X_{pd}^*, Y_{pd}^*)\right)^2 =: \bar{I}_Y, \tag{20}$$

where

$$X_{pd}^* = f_{sX}(h_X(\boldsymbol{X}_c, s(\xi_o^*, \tau), \xi_p)), \tag{21}$$

$$Y_{pd}^* = f_{sY}(h_Y(\boldsymbol{X}_c, s(\xi_o^*, \tau), \xi_p)). \tag{22}$$

Here h_X and h_Y are first and second element of vector function h, and f_{sX}, f_{sY} are step functions defined by

$$f_{sX}(\zeta) = \sum_{i=1}^{w_p} i\chi_i(\zeta), \tag{23}$$

$$f_{sY}(\zeta) = \sum_{i=1}^{h_p} i\chi_i(\zeta). \tag{24}$$

Here χ_i is an indicator function represented by

$$\chi_i(\zeta) = \begin{cases} 1, & (\zeta \in [i, i+1)), \\ 0, & (\zeta \notin [i, i+1)). \end{cases} \tag{25}$$

Now we can obtain the maxima of $\bar{I}_X$ in eq. (19) and $\bar{I}_Y$ in eq. (20) by

$$\bar{I}_X = (I_{p,\max} - I_{p,\min})^2, \tag{26}$$

$$\bar{I}_Y = (I_{p,\max} - I_{p,\min})^2, \tag{27}$$

[3]We can interpret this fact as follows: regardless of projected patterns, no image difference occurs near the goal pose in a special case where a target object has planar surface and the surface of the object is perpendicular to the optical axis of the camera in the goal pose.

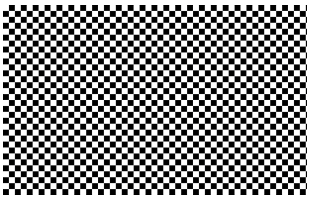


FIGURE 5. Checkered pattern.

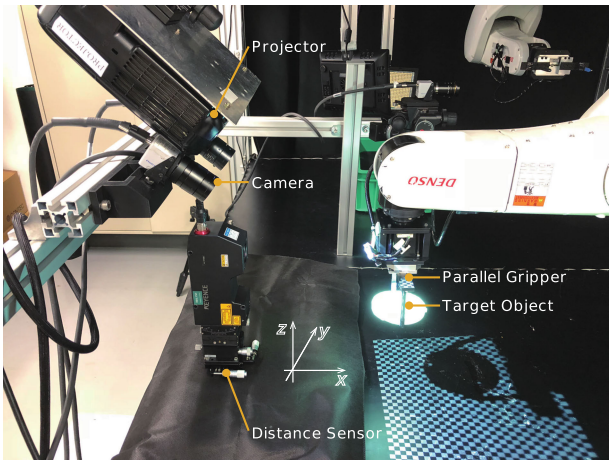


FIGURE 6. Experimental environment.

where $I_{p,\max}$ and $I_{p,\min}$ are the maximum and minimum intensities of the projection, respectively. It is obvious that a checkered pattern as shown in Fig. 5 can satisfies conditions represented by eqs. (26) and (27) simultaneously. Hence AVS adopts the striped patterns and we validates the positioning performance of AVS with the striped patterns in the following section.

V. EXPERIMENT

This section validates the performance of the proposed AVS and compares its performance to the conventional IBVS.

Fig. 6 shows an experimental system used for validation. The validation experiment involves positioning tasks for target objects. The target objects are made of plastic material and has complex shapes, as shown in Fig. 7. The objects utilized in this paper are modeled after real automobile parts. This experiments are designed to automate the kitting process found in actual factories. In the factories, workers manually select a part from the accumulated pile and carefully place it into trays outfitted with positioning fixtures. These trays are then positioned at a designated location in front of the assembly-line robots. After this setup, the robots take over, automating the subsequent stages of the assembly process. The manipulator was VS068 from DENSO WAVE INC. with six DOFs. A parallel gripper was attached to the manipulator. To measure the positioning error, a laser displacement meter LJ-V7300 from KEYENCE CORP. was set up at the position shown in Fig. 6. The direction of the laser irradiation was parallel to the x-axis. The projector, an EB-W420 (Seiko Epson Corp.) with resolution of 1280 × 800 (pixel), was set up at a distance of (−0.5[m], 0.0[m], 0.5[m]) from the

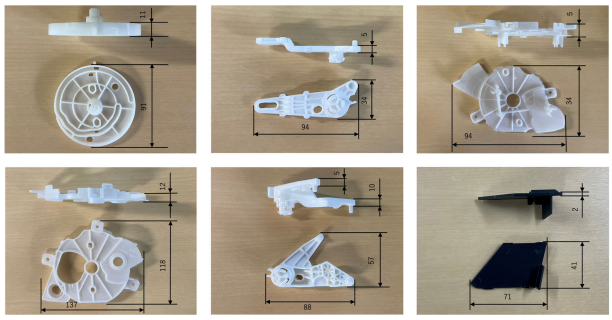


FIGURE 7. Target objects. Top-left: target object A. Top-middle: target object B. Top-right: target object C. Bottom-left: target object D. Bottom-middle: target object E. Bottom-right: target object F.

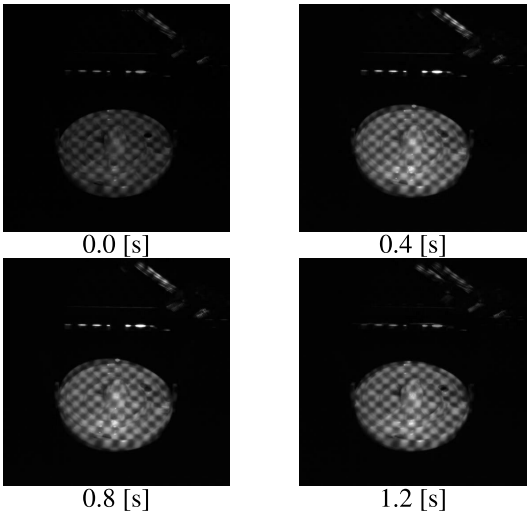


FIGURE 8. Time series images captured from the camera used for active visual servoing for target object A. Goal image of the target object is overlaid on each image. The optimal pattern is irradiated from the projector. The images shows how the goal image and the current image match.

target position. The camera, IDP-Express R2000 (PHOTRON LTD.), which was located at a distance of 0.1[m] below the projector, was used to capture images sized 512 × 512 (pixel) at a frequency of 50 (Hz). The captured images were sent to a PC with an Intel Core i7-4720HQ CPU and 16 GB of DDR3 SDRAM, and processed for control. OpenCV library [43] was used to process the captured images. We computed the image Jacobian by using numerical differentiation.

The purpose of this experiment was to perform positioning of the target objects shown in Fig. 7. The goal pose of target object A was set as indicated in Fig. 6. The initial position and posture were set to be (5[cm], 5[cm], 5[cm]) and (10deg., 10deg., 10deg.) away from the target position and posture, respectively. Here the posture is represented by XYZ Euler angles. In addition, the grasping error was set to be (10deg., 10deg., 10deg.) away from the ideal posture.[4]

The control for the the conventional and active visual servo is given by (3). The gain λ is set to be 2.0.

[4]Assuming the object is grasped by the hand at the ideal posture, we computed the image Jacobian by numerical differentiation.

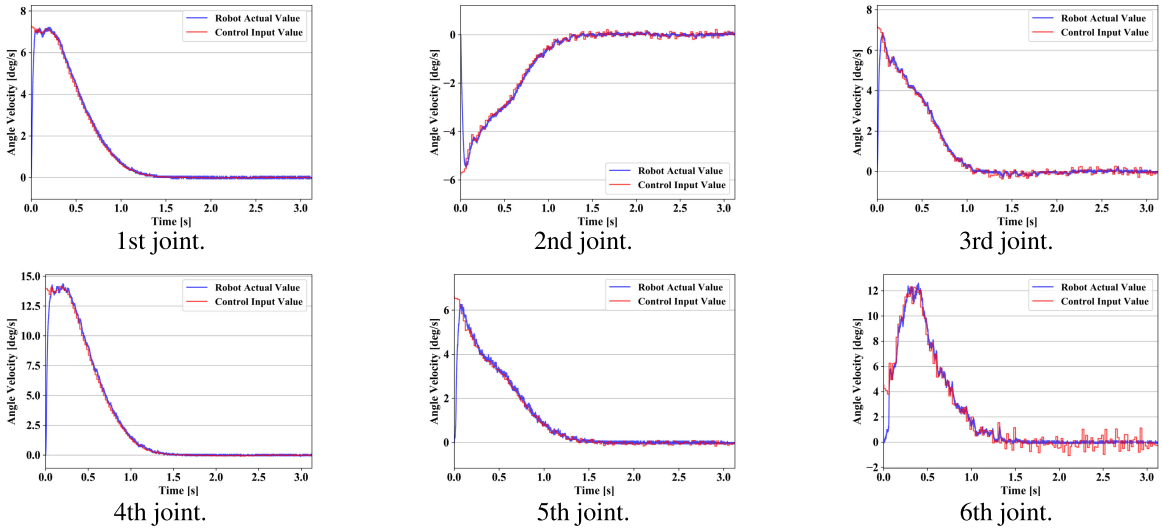


FIGURE 9. Time series data of joint angular velocity of the manipulator for the proposed active visual servoing.

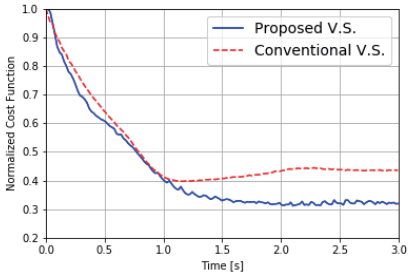


FIGURE 10. Time series data of the normalized sum of squared differences between the goal and the captured images at each time. Blue: proposed AVS. Red: conventional IBVS.

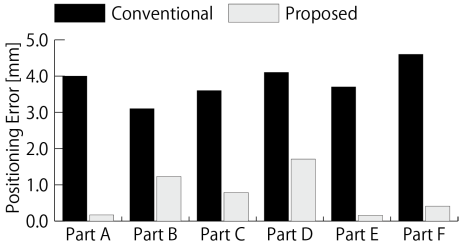


FIGURE 11. Positioning errors measured by a distance sensor.

The projector projected the checkered pattern, which is the optimal projection pattern as shown in sec. IV.

In the experimental system, the robot controller requires the control input in a cycle of 1 [ms]. Because the frame rate of the camera is 50 (fps), we need to interpolate the control input. Since we here use first-order hold, the control input at time t is computed by

$$\bar{u}(t) = u(nT) + \frac{u\,(nT) - u\,((n-1)T)}{T}\bar{t}, \qquad (28)$$

where $t = nT + \bar{t}, \bar{t} \in \{0, 1, \cdots, T-1\}, n \in \{0, 1, \cdots\}$, and T is the sampling period of the camera (=20[ms]).

The system terminates when sum of squared differences between the current and goal images is less than the specified threshold.

Fig. 8 depicts the time series images of one representative AVS sequence for target object A. The images are captured captured by of the camera used for the active visual servoing. The goal image of the target object are overlaid on the images. It can be seen that the images at each time are changing to coincide with the goal image.

Fig. 9 presents the time-series data of the joint angular velocity for the manipulator under the proposed active visual servoing technique. The graph demonstrates that the velocity of each joint approaches zero within approximately 1.2 to 1.5 seconds, which corresponds to the results depicted in Figs. 8 and 10.

The sum of squared differences between the goal and the captured image at each time point is presented in Fig. 10. In the figure, the image error is normalized to the error present at the initial time. The image errors for both the conventional and the proposed AVS converge at approximately 1.5[s]. The steady-state normalized difference for AVS is approximately 12% smaller than that for the conventional IBVS. In addition, while the image error for the conventional IBVS increases from $t = 1.0$[s] to $t = 2.3$[s], that for active visual servo control decreases almost monotonically. Here, note that Fig. 10 shows only the image error and does not show the positioning error in 3D space.

To evaluate the positioning error in 3D space, the laser displacement sensor was used. The positioning errors are measured 10 times for for each target object in both the AVS and the conventional IBVS; the results are plotted in Fig. 11. From Fig. 11, it is evident that for objects A, E, and F, the positioning errors were successfully reduced by 96%, 96%, and 91%, respectively, by using AVS. However, for objects B, C, and D, the reduction rate was 60%, 78%, and 57%, respectively. The rates are comparatively lower than that of A, E, and F. This observation can be rephrased to say that objects with a smaller proportion of holes in relation to their surface area exhibited a substantial

improvement in positioning accuracy when contrasted with conventional IBVS. Areas with holes lack three-dimensional shape characteristics. Hence, these results suggest that AVS enhances positioning accuracy notably for objects with rich three-dimensional features compared to conventional methods. This insight aligns with the theoretical perspective of AVS presented in sec. IV, which derives the projecting patterns that maximize image deviations while taking into account the three-dimensional shape.

VI. CONCLUSION

This paper proposes the active visual servo control (AVS) to perform the positioning of an object, especially a textureless object, with high accuracy. The proposed method uses a projector to project a light pattern onto the target object to increase the image error between the current and goal images. In this method, the light pattern used for irradiation have a great influence on the positioning accuracy. Thus, this study derives the optimal pattern that maximizes the image error. The experimental results demonstrate that the proposed method achieves higher positioning accuracy and yields a positioning error that is more than 57% smaller compared to that of conventional IBVS. Furthermore, the experimental results suggest that the proposed Active Visual Servoing technique provides more accurate positioning than Image-Based Visual Servoing for objects with complex and uneven surfaces. The superior performance of AVS can be credited to its distinctive method of projecting patterns onto objects. These patterns are theoretically derived to emphasize the image errors by intensifying the object's three-dimensional features. The theoretical foundations of AVS are congruent with the experimental results.

In this paper, stability and convergence were experimentally verified. In the future, we plan to theoretically examine the relationship between object shape and stability.

REFERENCES

[1] W. J. Wilson, C. C. Williams Hulls, and G. S. Bell, "Relative end-effector control using Cartesian position based visual servoing," *IEEE Trans. Robot. Autom.*, vol. 12, no. 5, pp. 684–696, Oct. 1996.

[2] R. Vidal, O. Shakernia, and S. Sastry, "Formation control of nonholonomic mobile robots with omnidirectional visual servoing and motion segmentation," in *Proc. IEEE Int. Conf. Robot. Autom.*, Sep. 2003, pp. 584–589.

[3] F. Tokuda, S. Arai, and K. Kosuge, "Object positioning by visual servoing based on deep learning," *Trans. Soc. Instrum. Control Eng.*, vol. 55, no. 11, pp. 717–725, 2019.

[4] R. Ginhoux, J. A. Gangloff, M. F. de Mathelin, L. Soler, M. M. A. Sanchez, and J. Marescaux, "Beating heart tracking in robotic surgery using 500 Hz visual servoing, model predictive control and an adaptive observer," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Apr. 2004, pp. 274–279.

[5] N. Guenard, T. Hamel, and R. Mahony, "A practical visual servo control for an unmanned aerial vehicle," *IEEE Trans. Robot.*, vol. 24, no. 2, pp. 331–340, Apr. 2008.

[6] C. Kingkan, S. Ito, S. Arai, T. Nammoto, and K. Hashimoto, "Model-based virtual visual servoing with point cloud data," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2016, pp. 5549–5555.

[7] M. Fujita, Y. Domae, A. Noda, G. A. Garcia Ricardez, T. Nagatani, A. Zeng, S. Song, A. Rodriguez, A. Causo, I. M. Chen, and T. Ogasawara, "What are the important technologies for bin picking? Technology analysis of robots in competitions based on a set of performance metrics," *Adv. Robot.*, pp. 1–15, Dec. 2019, doi: 10.1080/01691864.2019.1698463.

[8] D. Liu, S. Arai, J. Miao, J. Kinugawa, Z. Wang, and K. Kosuge, "Point pair feature-based pose estimation with multiple edge appearance models (PPF-MEAM) for robotic bin picking," *Sensors*, vol. 18, no. 8, p. 2719, Aug. 2018, doi: 10.3390/s18082719.

[9] T. Yamawaki and M. Yashima, "Grasp planning based on dynamics shaping," in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatronics (AIM)*, Jul. 2011, pp. 617–622.

[10] Y. Yokokohji, Y. Kawai, M. Shibata, Y. Aiyama, S. Kotosaka, W. Uemura, A. Noda, H. Dobashi, T. Sakaguchi, and K. Yokoi, "Assembly challenge: A robot competition of the industrial robotics category, world robot summit—Summary of the pre-competition in 2018," *Adv. Robot.*, vol. 33, no. 17, pp. 876–899, Sep. 2019.

[11] V. Lippiello, B. Siciliano, and L. Villani, "Position-based visual servoing in industrial multirobot cells using a hybrid camera configuration," *IEEE Trans. Robot.*, vol. 23, no. 1, pp. 73–86, Feb. 2007.

[12] F. Chaumette and S. Hutchinson, "Visual servo control. I. Basic approaches," *IEEE Robot. Autom. Mag.*, vol. 13, no. 4, pp. 82–90, Dec. 2006.

[13] K. Hashimoto, T. Kimoto, T. Ebine, and H. Kimura, "Manipulator control with image-based visual servo," in *Proc. IEEE Int. Conf. Robot. Autom.*, Jan. 1991, pp. 2267–2271.

[14] M. Mimou, T. Kanade, and T. Sakai, "A method of time-coded parallel planes of light for depth measurement," *IEICE Trans.*, vol. 64, no. 8, pp. 521–528, Aug. 1981.

[15] T. Chen, H.-P. Seidel, and H. P. A. Lensch, "Modulated phase-shifting for 3D scanning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[16] S. Arai, Y. Iwatani, and K. Hashimoto, "Fast sensor scheduling with communication costs for sensor networks," in *Proc. Amer. Control Conf.*, Jun. 2010, pp. 295–300.

[17] H.-Y. Kuo, H.-R. Su, S.-H. Lai, and C.-C. Wu, "3D object detection and pose estimation from depth image for robotic bin picking," in *Proc. IEEE Int. Conf. Autom. Sci. Eng. (CASE)*, Aug. 2014, pp. 1264–1269, doi: 10.1109/COASE.2014.6899489.

[18] D. Liu, S. Arai, Z. Feng, J. Miao, Y. Xu, J. Kinugawa, and K. Kosuge, "2D object localization based point pair feature for pose estimation," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2018, pp. 1119–1124.

[19] J. Mainprice, R. Hayne, and D. Berenson, "Predicting human reaching motion in collaborative tasks using inverse optimal control and iterative re-planning," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2015, pp. 885–892.

[20] S. Arai, A. L. Pettersson, and K. Hashimoto, "Fast prediction of a worker's reaching motion without a skeleton model (F-PREMO)," *IEEE Access*, vol. 8, pp. 90340–90350, 2020.

[21] S. Azimi, B. Lall, and T. K. Gandhi, "Performance evalution of 3D keypoint detectors and descriptors for plants health classification," in *Proc. 16th Int. Conf. Mach. Vis. Appl. (MVA)*, May 2019, pp. 1–6, doi: 10.23919/MVA.2019.8758002.

[22] S. Arai, N. Fukuchi, and K. Hashimoto, "FAst detection algorithm for 3D keypoints (FADA-3K)," *IEEE Access*, vol. 8, pp. 189556–189564, 2020.

[23] D. Liu, S. Arai, F. Tokuda, Y. Xu, J. Kinugawa, and K. Kosuge, "Deep-learning based robust edge detection for point pair feature-based pose estimation with multiple edge appearance models," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2019, pp. 2920–2925.

[24] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, "Deep learning for 3D point clouds: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4338–4364, Dec. 2021, doi: 10.1109/TPAMI.2020.3005434.

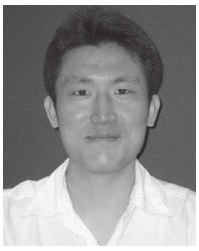
[25] Y. Xu, S. Arai, F. Tokuda, and K. Kosuge, "A convolutional neural network for point cloud instance segmentation in cluttered scene trained by synthetic data without color," *IEEE Access*, vol. 8, pp. 70262–70269, 2020, doi: 10.1109/ACCESS.2020.2978506.

[26] Y. Xu, S. Arai, D. Liu, F. Lin, and K. Kosuge, "FPCC: Fast point cloud clustering-based instance segmentation for industrial bin-picking," *Neurocomputing*, vol. 494, pp. 255–268, Jul. 2022.

[27] S.-W. Lee, C.-M. Hsu, M.-C. Lee, Y.-T. Fu, F. Atas, and A. Tsai, "Fast point cloud feature extraction for real-time SLAM," in *Proc. Int. Autom. Control Conf. (CACS)*, Nov. 2019, pp. 1–6, doi: 10.1109/CACS47674.2019.9024355.

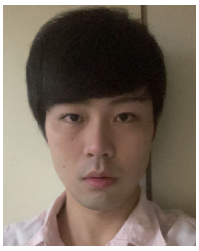
[28] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado, "A state of the art in structured light patterns for surface profilometry," *Pattern Recognit.*, vol. 43, no. 8, pp. 2666–2680, Aug. 2010.

[29] M. Gupta, A. Agrawal, A. Veeraraghavan, and S. G. Narasimhan, "Structured light 3D scanning in the presence of global illumination," in *Proc. CVPR*, Jun. 2011, pp. 713–720.
[30] J. L. Posdamer and M. D. Altschuler, "Surface measurement by space-encoded projected beam systems," *Comput. Graph. Image Process.*, vol. 18, no. 1, pp. 1–17, Jan. 1982.
[31] T. Chen, H. P. A. Lensch, C. Fuchs, and H.-P. Seidel, "Polarization and phase-shifting for 3D scanning of translucent objects," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[32] M. O'Toole, J. Mather, and K. N. Kutulakos, "3D shape and indirect appearance by structured light transport," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3246–3253.
[33] N. Chiba, S. Arai, and K. Hashimoto, "Feedback projection for 3D measurements under complex lighting conditions," in *Proc. Amer. Control Conf. (ACC)*, May 2017, pp. 4649–4656.
[34] D. Khadraoui, G. Motyl, P. Martinet, J. Gallice, and F. Chaumette, "Visual servoing in robotics scheme using a camera/laser-stripe sensor," *IEEE Trans. Robot. Autom.*, vol. 12, no. 5, pp. 743–750, Oct. 1996.
[35] J. Pages, C. Collewet, F. Chaumette, and J. Salvi, "Plane-to-plane positioning from image-based visual servoing and structured light," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2004, pp. 1004–1009.
[36] J. Pages, C. Collewet, F. Chaumette, and J. Salvi, "Robust decoupled visual servoing based on structured light," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Aug. 2005, pp. 2676–2681.
[37] J. Pagès, C. Collewet, F. Chaumette, and J. Salvi, "A camera-projector system for robot positioning by visual servoing," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop (CVPRW)*, Jun. 2006, p. 2.
[38] J. Pagès, C. Collewet, F. Chaumette, and J. Salvi, "Visual servoing by means of structured light: Visual-based robot task optimisation from controlled light pattern projection," Tech. Rep., 2008.
[39] P. C Ng and S. Henikoff, "SIFT: Predicting amino acid changes that affect protein function," *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3812–3814, Jul. 2003.
[40] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. 9th Eur. Conf. Comput. Vis.*, vol. 3951, May 2006, pp. 404–417.
[41] P. F. Alcantarilla and T. Solutions, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1281–1298, Sep. 2011.
[42] G. Silveira and E. Malis, "Direct visual servoing: vision-based estimation and control using only nonmetric information," *IEEE Trans. Robot.*, vol. 28, no. 4, pp. 974–980, Aug. 2012.
[43] A. Kaehler and G. Bradski, *Learning OpenCV 3: Computer Vision in C++ With the OpenCV Library*. Sebastopol, CA, USA: O'Reilly Media, 2016.



SHOGO ARAI (Member, IEEE) received the B.S. degree in aerospace engineering and the M.S. and Ph.D. degrees in information sciences from Tohoku University, Sendai, Japan, in 2005, 2007, and 2010, respectively. From 2010 to 2016, he was an Assistant Professor with the Intelligent Control Systems Laboratory, Tohoku University. In 2016, he joined the System Robotics Laboratory, Department of Robotics, Tohoku University, as an Associate Professor. Recently, he joined Tokyo University of Science, where he is currently an Associate Professor. His research focuses on the fields of robot vision, visual servoing, bin-picking, motion planning, deep learning, 3-D measurement, networked control systems, and multi-agent systems.

He received the Young Scientists Award from the Minister of Education, Culture, Sports, Science and Technology, in 2022; the Best Paper Award from FA Foundation, in 2019; the 32th Best Paper Award from the Robotics Society of Japan, in 2019; the Certificate of Merit for Best Presentation from the Japan Society of Mechanical Engineers, in 2019; the Excellent Paper Award from the Institute of Systems from Control and Information Engineers, in 2010; the Best Paper Award Finalist at IEEE International Conference on Mechatronics and Automation, in 2012; and the SI2020, SI2019, and SI2018 Excellent Presentation Award from the Society of Instrument and Control Engineers, in 2020, 2019, and 2018, respectively.



YOSHIHIRO MIYAMOTO received the B.S. degree in mechanical and aerospace engineering from Tohoku University, Sendai, Japan, in 2017, and the M.S. degree from the Department of Robotics, Tohoku University, in 2019. His research interests include the robot vision and visual servoing.

AKINARI KOBAYASHI received the B.S. degree in mechanical and aerospace engineering from Tohoku University, Sendai, Japan, in 2013, the M.S. degree from the Department of Bioengineering and Robotics, Tohoku University, in 2017, and the Ph.D. degree from the Department of Robotics, Tohoku University, in 2020.

He is currently a Postdoctoral Fellow with Centre for Garment Production Ltd., which has been established by collaborative research between Tohoku University and The University of Hong Kong. His research interests include robot hands and grippers, under-actuated mechanisms, and automation systems.

KAZUHIRO KOSUGE (Life Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees in control Engineering from Tokyo Institute of Technology, in 1978, 1980, and 1988, respectively.

He was a Research and Development Staff with the Production Engineering Department, Nippon Denso Company Ltd.; a Research Associate with Tokyo Institute of Technology; and an Associate Professor with Nagoya University. In 1995, he joined Tohoku University, as a Professor, where he was a Distinguished Professor, from 2018 to 2021. He is currently the Director of the Transformative AI and Robotics International Research Center and a Specially Appointed Professor with the Graduate School of Engineering, Tohoku University, Japan. He recently joined The University of Hong Kong as the Chair Professor of the Department of Electrical and Electronic Engineering.

Prof. Kosuge is a JSME Fellow, a SICE Fellow, a RSJ Fellow, a JSAE Fellow, and a member of the Engineering Academy of Japan. He received the Medal of Honor and the Medal with Purple Ribbon from the Government of Japan, in 2018–a national honor in recognition of his prominent contributions to academic and industrial advancements. He also received the IEEE RAS George Saridis Leadership Award in Robotics and Automation, in 2021, for his exceptional vision of innovative research and outstanding leadership in the robotics and automation community through technical activity management. He was the President of the IEEE Robotics and Automation Society, from 2010 to 2011; the IEEE Division X Director, from 2015 to 2016; and the IEEE Vice President of Technical Activities, in 2020.