**Proceedings of the Canadian Society for Mechanical Engineering International Congress 2022**
**CSME Congress 2022**
**June 5-8, 2022, Edmonton, AB, Canada**

# EYE-IN-HAND/EYE-TO-HAND COOPERATED PEG-IN-HOLE MICROASSEMBLY OPERATIONS USING COORDINATE-BASED AUGMENTED REALITY

R. J. Chang, W. J. Chien
Department of Mechanical Engineering, National Cheng Kung University, Tainan, Taiwan, ROC

***Abstract***—A new eye-in-hand/eye-to-hand cooperated visual system in peg-in-hole microassembly work is assisted by coordinate-based augmented reality (AR) through three dimensional (3D)-hybrid visual calibration and servo. The 3D hybrid visual calibration is to combine real camera's intrinsic parameters and virtual camera's extrinsic parameters in step-wise visual calibration. The detecting and tracking of image feature points are undertaken by utilizing marker and estimator. By employing 3D model and ray casting, the 3D coordinates on the matching rod, including the center of mating hole, and corner points of supporting block correspondent to the 2D virtual image points are extracted. Regarding the image viewed by the real and virtual cameras in AR, a calibrated virtual camera is utilized to track the coordinates of real mating hole and supporting block. The effectiveness of coordinate-based AR technology by the eye-in-hand configuration is tested in the alignment between micropeg and mating hole for inserting a micropeg of diameter 80 $\mu m$ with length 1~1.4 *mm* into a mating rod with 100 $\mu m$ hole.

***Keywords-microassembly; eye-in-hand; eye-to-hand; augmented reality***

## I. INTRODUCTION

In the manufacturing cycle of microsystems, assembly is a crucial operation in production [1]. For achieving critical microassembly operations, a computer aided design (CAD) technique to realize virtual reality (VR) has been developed to enhance operational efficiency and effectiveness [2-7]. Recently, AR has been implemented and tested for the microscopic peg-in-hole assembly operations [8,9]. The advantages of utilizing AR in two-dimensional (2D) image-based microassembly [8] has been further exploited in the AR-assisted microassembly with 3D model-based visual servo [9]. The 3D model in the AR-assisted microscopic visual system can compensate the feature points of micro object when they are out of the FOV and/or lack of resolution. The AR-assisted algorithm was tested to give the same success rate as that by the pure visual-servo algorithm in peg-in-hole microassembly. In regarding to image-based visual servo, some efforts have been made to implement hybrid eye-in-hand/eye-to-hand camera configurations for complex tasks [10,11]. However, only eye-to-hand configuration has been implemented in microassmbly operations. The advantage and challenge of the hybrid camera configurations in microassembly operation needs to be investigated and exploited.

The importance of the present AR-assisted microassembly research is to contribute an eye-in-hand/eye-to-hand cooperated camera configuration as well as 3D-hybrid visual calibration and servo for achieving microassembly work. This paper is organized from system installation, virtual environment, object model, hybrid visual calibration, to AR-assisted visual servo. The operation of gripping micropeg,and finishing micropeg alignment as well as insertion are under eye-to-hand, and eye-in-hand servo, respectively. The AR is tested in the 3D model-based microassembly alignment and insertion under eye-in-hand camera. Finally, the effectiveness of the present approach in microassembly is concluded.

## II. SYSTEM INSTALLATION, VIRTUAL ENVIRONMENT, AND OBJECT MODEL

### *A. Installation of Microassembly System*

A peg-in-hole microassembly system including a micromanipulator system, working stages, and three CCD was installed. The micromanipulator system mainly consists of the stereolithography-printed 3D microgripper, piezo actuator, and plastic tube. The 3D microgripper has 3 gripping jaws driven by piezo actuator and the gripper openings can be adjusted by the end screw to fit different size of objects. The detail assembly of the micromanipulator is delineated in [12]. The test of 3D microgripper in gripping 80 $\mu m$ wire is shown in Fig. 1.

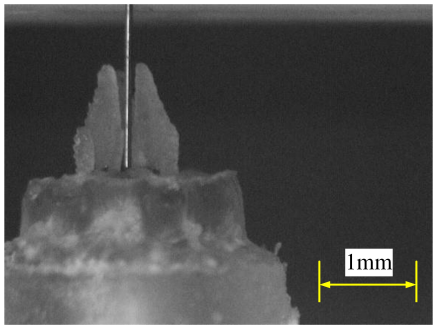


Figure 1. Testing of 3D-printed resin microgripper in gripping 80 $\mu m$ wire.

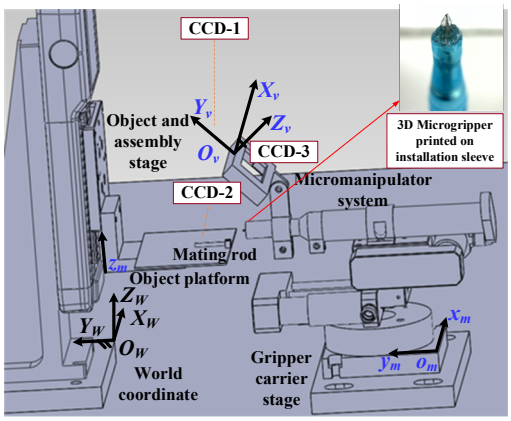


Figure 2. SolidWorks drawing of microassembly system with coordinate system of CCD-3.

In the present microassembly system, there are three cameras: CCD-*i*, *i*=1, 2, 3. The resolution of the two eye-to-hand CCD systems are 13 *μm/pixel* for CCD-1 and 12 *μm/pixel* for CCD-2. Regarding the eye-in-hand CCD-3, the resolution can be adjusted in the range 3-5 *μm/pixel*. The installation of CCD-3, micromanipulator system, gripper carrier stage, as well as object and assembly stage are shown in Fig. 2.

For the installation as shown in Fig. 2, the origin $\boldsymbol{O_w}$ and ($X_w$, $Y_w$, $Z_w$) is a world coordinate system. In the ideal installation design, the pointing directions by the optical axis of CCD-1, CCD-2 were along the negative $Z_W$ axis and negative $X_W$ axis, respectively. The CCD-3 was along the optical axis, which was parallel to the $Z_W$-$Y_W$ plane, and inclined at 45° with respect to the $Y_W$ axis and pointing in the downward direction. The gripper was designed to be installed parallel to the $x_m$-$y_m$ plane of the gripper carrier stage. The mating rod was installed such that its longitudinal central axis was parallel to $Y_W$ axis. Regarding the micropeg feeding on the object platform, the longitudinal central axis of the micropeg was parallel to the $X_W$-$Y_W$ plane and the axis formed an allowable small angle with the $Y_W$ axis. The initial alignment of micropeg was manually adjusted. Regarding the moving direction of the gripper carrier stage, the $x_m$ and $y_m$ axes were parallel to the $X_W$ and $Y_W$ axes, respectively. Regarding the moving direction of the object and assembly stage, the $z_m$ axis was parallel to the $Z_W$ axis. The object and assembly stage which carried mating rod was driven by stepping motor of SIGMA-KOKI, Japan. The stage provided 1$\mu m$ resolution with 35$mm$ stroke for vertical motion.

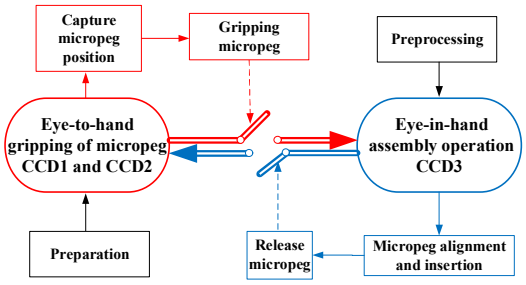


Figure 3. Microassembly work by switching operations between eye-to-hand and eye-in-hand configurations.

## B. *Switching between Eye-to-hand and Eye-in-hand Configurations in Microassembly Operations*

There are three phases in the microassembly process. The first phase was gripping a micropeg. The second phase was to transport micropeg approaching the mating hole. The third phase was micropeg alignment and insertion. In assembly work, global fixed cameras of CCD-1 and CCD-2 in the eye-to-hand configuration were limited to view the detail scene of micropeg and hole. On the other hand, it was not possible for local camera CCD-3 in the eye-in-hand configuration to view the whole workspace and approach micropeg on working stage. By taking the advantage of the eye-to-hand and eye-in-hand configurations in micro assembly operations, the eye-to-hand configuration was employed in the first gripping phase and the eye-in-hand configuration was utilized in the last alignment and assembly phase. An eye-in-hand/eye-to-hand cooperated configuration and its operations is proposed as shown in Fig. 3. The coordinates in the camera of CCD-1 and CCD-2 as well as the coordinates in CCD-3 and the parameters of CCD-3 in virtual environment are shown in Fig. 4.

## C. *Virtual Environment*

In considering the final eye-in-hand assembly to be operated under camera-3 and assisted by AR, only one virtual camera of CCD-3 is employed. Thus, the superscript in the coordinate system of virtual camera-3 is ignored for brevity. By referring to Fig. 4, the origin $\boldsymbol{E}_V$ is defined as the location of the camera-3. The point ($E_x$, $E_y$, $E_z$) denotes the translational location of virtual camera-3. In the camera space with coordinate systems ($X_V$, $Y_V$, $Z_V$) and image coordinates ($U$, $V$), the $X_V$-axis is defined as the same direction of $U$, the $Y_V$-axis is defined as the same direction of $V$, and $Z_V$-axis is defined in the direction opposite to the axis of camera-3, which points away from cameras' aiming. The virtual scene viewed by cameras is developed by utilizing OpenGL [13]. The poses of virtual camera are defined through the rotation of Euler angles with the rotational sequence: $\gamma \rightarrow \beta \rightarrow \alpha$. In the ideal installation of virtual camera, the rotational angles are $\alpha=45^{\circ}, \beta=\gamma=0^{\circ}$ for camera-3.

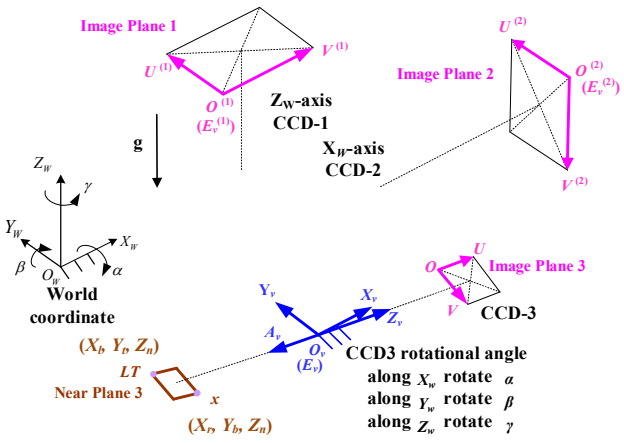


Figure 4. Schematic diagram of the coordinates in installing camera CCD-1, CCD-2, CCD-3 and the coordinates of CCD-3 in virtual environment.

In perspective projection by eye-in-hand camera-3, the virtual camera model which transforms world coordinate into screen coordinate is

$$\begin{bmatrix} x_s \\ y_s \\ z_s \\ 1 \end{bmatrix} = \frac{1}{w_c} \boldsymbol{V} \cdot \boldsymbol{P}_p \cdot \boldsymbol{M}_v \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}, \tag{1}$$

where $w_c$ is a perspective normalization constant, $\boldsymbol{V}$ is a viewport matrix, $\boldsymbol{P}_p$ is a perspective projection matrix, and $\boldsymbol{M}_v$ is a view matrix. These matrices are defined as follows [13]. The viewport matrix $\boldsymbol{V}$ which is constructed by the ($r_w$, $t_w$) and ($l_w$, $b_w$) on the image plane is

$$\boldsymbol{V} = \begin{bmatrix} \frac{r_w - l_w}{2} & 0 & 0 & \frac{r_w + l_w}{2} \\ 0 & \frac{t_w - b_w}{2} & 0 & \frac{t_w + b_w}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{2}$$

The projection matrix $\boldsymbol{P}_p$ which determines the range of FOV and symmetric view volume is given through the vertical field of view, *VFOV* and aspect ratio, *ar* as

$$\boldsymbol{P}_p = \begin{bmatrix} P_1 & 0 & 0 & 0 \\ 0 & P_2 & 0 & 0 \\ 0 & 0 & \frac{-f-n}{f-n} & \frac{-2fn}{f-n} \\ 0 & 0 & -1 & 0 \end{bmatrix}, \tag{3}$$

where $P_1 = 1/ar \cdot \tan(VFOV/2)$, $P_2 = 1/\tan(VFOV/2)$, $n$ and $f$ are the distance from the center of camera to the near plane and far plane respectively.

Regarding the view matrix $\boldsymbol{M}_v$, it is noted that the pose adjustment of virtual camera in OpenGL is based on the coordinate system of camera instead of world coordinate. The view matrix in camera coordinate is defined as $\boldsymbol{M}_v^o$, where the superscript denotes the origin $\boldsymbol{O_w}$ in Fig. 4. The view matrix can be derived through inverse as

$$\boldsymbol{M}_v = (\boldsymbol{M}_v^o)^{-1}, \tag{4}$$

where

$$\boldsymbol{M}_v^o = \begin{bmatrix} X_x^V & Y_x^V & Z_x^V & E_x \\ X_y^V & Y_y^V & Z_y^V & E_y \\ X_z^V & Y_z^V & Z_z^V & E_z \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{5}$$

### D. Three-dimensional Model and Feature Points

The commercial modeling software SolidWorks is used to create the CAD model of the microassembly system. In programming, the 3D models constructed by CAD software are exported to VRML (Virtual Reality Mockup Language) file before being loaded into the virtual environment by the software LabVIEW. The extraction of 3D coordinates on the virtual mating rod and supporting block is realized by raycasting technique [9].

## III. THREE-DIMENSIONAL HYBRID VISUAL CALIBRATION

### A. Real Calibration of Intrinsic Parameters

In the calibration of real camera parameters, the most adopted technique is Zhang's method [14]. Consider a camera model to be expressed as

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \boldsymbol{K} \begin{bmatrix} \boldsymbol{R} & | & \boldsymbol{t} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}, \tag{6}$$

where $z_c$ is an arbitrary scale factor, $\begin{bmatrix} \boldsymbol{R} & | & \boldsymbol{t} \end{bmatrix}$ is a camera extrinsic matrix, and $\boldsymbol{K}$ is a camera intrinsic matrix. The intrinsic parameters of $\boldsymbol{K}$ are given by

$$\boldsymbol{K} = \begin{bmatrix} \alpha_x & s & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{7}$$

where $(u_0, v_0)$ is the coordinates of the principal point, $\alpha_x$, $\alpha_y$ are the scale factors in image *u* and *v* axes, and *s* is a parameter describing the skewness of the two image axes. By defining a matrix $\boldsymbol{B}$=[$b_{ij}$] as

$$\boldsymbol{B} = \boldsymbol{K}^{-T} \boldsymbol{K}^{-1}, \tag{8}$$

then the intrinsic parameters can be expressed as functions of $b_{ij}$, and consequently, these parameters can be calibrated.

### B. Step-wise Non-linear Virtual Calibration

Consider that a pair of virtual and real line segments lies on an image plane. The two points at $(u_1, v_1)$ and $(u_2, v_2)$ are given as the endpoints of a real line. The two points at $(x_{s,1}, y_{s,1})$ and $(x_{s,2}, y_{s,2})$ are given as the endpoints of a virtual line. If a feature coordinate ($x_w$, $y_w$, $z_w$) on the surface of virtual model is known, the projected coordinate on image plane is given by (1). The normal distances from the endpoints of image line to the projected virtual line are derived as

$$h_{lj} = (Au_l + Bv_l + C)/M, l = 1,2. \tag{9}$$

where $A = y_{s,2} - y_{s,1}$, $B = x_{s,1} - x_{s,2}$, $C = x_{s,2}y_{s,1} - x_{s,1}y_{s,2}$, and $M = \sqrt{A^2 + B^2}$. Here, the parameters in (9) are function of $(x_{s,j}, y_{s,j})$, which are in terms of function of camera parameters by ignoring $(u_0, v_0)$:

$$\boldsymbol{\Psi} = (E_x, E_y, E_z, \alpha, \alpha_x, \alpha_y, s)^T. \tag{10}$$

The formulation of (10) indicates that at least four pairs of real and virtual line segments are required for calibration. In virtual calibration with data collected, one defines (11) to include different data points as

$$\boldsymbol{h}_j = \begin{bmatrix} h_{1j}(\boldsymbol{\Psi}) \\ h_{2j}(\boldsymbol{\Psi}) \end{bmatrix}. \tag{11}$$

The vector $\boldsymbol{h}_j$ can be augmented to form $\boldsymbol{H}$. It is noted that $\boldsymbol{H}$ = 0 formulates a set of $2N \times 1$ non-linear algebraic equations

with unknown camera parameters. The objective of calibration is to estimate camera parameters such that an error function, which is defined as the sum of square error between the normal distances of real and virtual lines, can be minimized. The iterative numerical scheme is:

$$\hat{\boldsymbol{\Psi}}_{k+1} = \hat{\boldsymbol{\Psi}}_k - (\boldsymbol{J}^T\boldsymbol{W}\boldsymbol{J})^{-1}\boldsymbol{J}^T\boldsymbol{W}\boldsymbol{H}(\hat{\boldsymbol{\Psi}}_k), \tag{12}$$

where $\boldsymbol{J}$ are Jacobian matrices of $2N \times 7$ which are established by taking the partial derivative of $\boldsymbol{H}$ with respect to $\boldsymbol{\Psi}$.

In practice, the linear calibration will be first derived by assuming $\hat{\beta} = \hat{\gamma} = 0^\circ$ for obtaining a closed-form formulation. Then, one can derive a linear relationship for solving camera parameters $\boldsymbol{\Psi}$= ($E_x$, $E_y$, $E_z$, $\alpha$). The estimated parameters are employed as the initial condition for the later non-linear virtual calibration.

For the step-wise nonlinear calibration, it will be iterated for three stages in sequence, starting from the first step with $\alpha$, to the second step to add $\beta$, and the third step to add $\gamma$. By employing this procedure, the finding of global minimum by (12) is guaranteed without diverging.

## IV. VISUAL SERVO IN REGISTRATION

### A. Interaction Matrix

An interaction matrix, or image Jacobian, which relates end effector velocity in the task space to the rate of change of feature parameter in image space [9], will be derived. Consider a feature point on the surface of virtual mating rod, the position of feature point is expressed as $\boldsymbol{X}_o = x_o\vec{i} + y_o\vec{j} + z_o\vec{k}$, which is referred to object frame. The transformation from world coordinate $(x_w, y_w, z_w)$ to object coordinate, which is fixed on the centroid of mating rod and parallel to the world coordinate, is first established through modal matrix $\boldsymbol{M}_G$. The $\boldsymbol{M}_G$ is constructed by the centroid with coordinate $\boldsymbol{G}(g_x, g_y, g_z)$ which is referred to the world frame. Next, the transformation from world coordinate to normalized device coordinate by camera-3 is established. Thus, one derives

$$x_n = \frac{1}{w_c}(P_1 x_o + P_1(g_x - E_x))$$

$$y_n = \frac{1}{w_c}(P_2\cos\alpha\, y_o + P_2\sin\alpha\, z_o + P_2(g_y\cos\alpha + g_z\sin\alpha - E_y\cos\alpha + E_z\sin\alpha))$$

$$\begin{aligned} z_n = \frac{1}{w_c}\Big(&-\frac{-f-n}{f-n}\sin\alpha\, y_o + \frac{-f-n}{f-n}\cos\alpha\, z_o + \\ &\frac{-2fn}{f-n} + \frac{-f-n}{f-n}(-E_z\cos\alpha + E_y\sin\alpha + g_z\cos\alpha - g_y\sin\alpha)\Big), \end{aligned} \tag{13}$$

where $w_c = y_o\sin\alpha - z_o\cos\alpha + E_z\cos\alpha - E_y\sin\alpha - g_z\cos\alpha + g_y\sin\alpha$

In considering real-time tracking servo, the virtual camera is assumed to have been calibrated. By setting the angles to be constant and taking the time derivative of (13), one obtains

$$\begin{aligned} \dot{x}_n &= [(x_c\cos\alpha)\dot{z}_o + (-P_1 w_c)\dot{E}_x + (x_c\sin\alpha)\dot{E}_y]/(w_c^2) \\ \dot{y}_n &= [(w_c P_2\sin\alpha + y_c\cos\alpha)\dot{z}_o + (-w_c P_2\cos\alpha + y_c\sin\alpha)\dot{E}_y]/(w_c^2). \end{aligned} \tag{14}$$

Regarding a specific feature point, the velocity is expressed as $\dot{\boldsymbol{X}}_W = \dot{x}_w\vec{i} + \dot{y}_w\vec{j} + \dot{z}_w\vec{k}$. Since the real mating rod is a rigid body, the velocity of feature point is expressed through the translational centroid velocity $\boldsymbol{V}_G^t = v_x\vec{i} + v_y\vec{j} + v_z\vec{k}$ and angular velocity of the object, $\boldsymbol{\Omega} = \omega_x\vec{i} + \omega_y\vec{i} + \omega_z\vec{i}$. If the mating rod only has translational centroid motion along $Z_w$ axis, the relationship between the velocity in normalized device coordinate and translational centroid velocity of virtual mating rod as well as translational velocity of virtual camera is given by

$$\begin{bmatrix} \dot{x}_n \\ \dot{y}_n \end{bmatrix} = \frac{1}{w_c^2}\begin{bmatrix} x_c\cos\alpha & -P_1 w_c & x_c\sin\alpha \\ w_c P_2\sin\alpha + y_c\cos\alpha & 0 & -w_c P_2\cos\alpha + y_c\sin\alpha \end{bmatrix}\begin{bmatrix} v_z \\ \dot{E}_x \\ \dot{E}_y \end{bmatrix}. \tag{15}$$

By defining interaction matrices $\boldsymbol{L}_o(t)$, (15) is expressed as

$$\dot{\boldsymbol{s}}_o(t) = \boldsymbol{L}_o(t)\cdot\boldsymbol{V}_C(t), \tag{16}$$

### B. Virtual Tracking and Regulation

In the alignment servo, the visual servo is to drive the virtual mating rod and supporting block to follow the real one as observed by CCD-3. By defining an error $e(t)$ between the virtual projected point $\boldsymbol{s}_0(t)$ and the correspondent real image point $\boldsymbol{s}^*(t)$ as

$$\boldsymbol{e}(t) = \boldsymbol{s}_o(t) - \boldsymbol{s}^*(t), \tag{17}$$

a servo control is designed as a sequential tracking control along with set points regulation. When a tracking command is sent to working stages, the virtual servo is activated. The actuating signal is given as

$$\boldsymbol{V}_C(t) = -\lambda\boldsymbol{L}_o^+(t)\cdot\boldsymbol{e}(t),\ \lambda > 0, \tag{18}$$

where $\boldsymbol{L}_o^+(t)$ are the pseudo-inverse of $\boldsymbol{L}_o(t)$. The regulation is achieved through adjusting the centroid position of the mating rod together with mapping transformation. Regarding the control law (18), the selection of $\lambda$=1 gives the best tracking performance [9]. In simulating the tracking response, an average tracking error for the sum of distance error of $N$ feature points is defined as

$$e_{av}(t) = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(\boldsymbol{s}_{o,i}(t) - \boldsymbol{s}_i^*(t))^T(s_{o,i}(t) - s_i^*(t))}. \tag{19}$$

In the tracking control, the error correction will not stop until a specified error bound on $e_{av}$ is satisfied.

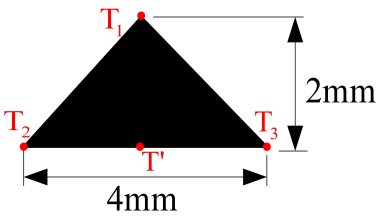


Figure 5. Triangular plate marker installed on the sleeve surface of gripper manipulator.

## V. MARKER AND ESTIMATION OF FEATURE COORDINATE

### A. Estimation of Gripper Tip Coordinate

The position estimation of gripper tip is carried out through a triangular plate which is pointed forward and fixed on the sleeve surface of micromanipulator. The triangular plat with the known feature positions $T_1$, $T_2$, $T_3$ and the unknown position $T'$, which is located at the middle between $T_1$ and $T_3$, is shown in Fig. 5. By referring to the CCD-3 in Figs. 2 and 4, if the world coordinate of $T'$ is estimated as $(x_{T'}, y_{T'}, z_{T'})$, $\delta$ is a distance of the tip position along the axis of gripper manipulator, and the radius of sleeve tip is $r$, then the position of gripper tip is estimated as $(x_{T'}, y_{T'} + \delta, z_{T'} + r)$, by ignoring the thickness of the triangular plate. Here, it is noted that estimating the coordinate of $T'$ is through the calibrated camera parameters.

### B. Feature Points of Mating Rod and Supporting Block

A mating rod and a supporting block are installed and shown in Fig. 6. The real feature points on the corner of supporting block and on the circular edge of mating rod are detected through utilizing characteristic lines, scanning, and identification.

## VI. COORDINATE-BASED AUGMENTED REALITY IN MICROASSEMBLY

In the peg-in-hole assembly test, cylindrical micropegs, which are made of Ti-Ni wire, have diameter 80 $\mu m$ and length 1~1.4 $mm$. A mating cylindrical rod has a mating hole of diameter 100 $\mu m$. The assembling clearance ratio is 0.2. The AR technology is mainly employed in the last phase of alignment and insertion between micropeg and mating hole.

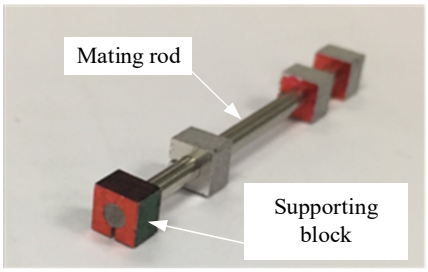


Figure 6. Mating rod installed on the supporting block.

### A. Initial Calibration, Validation, and Estimation

(1) Approaching and focusing

Before executing the final phase of micropeg alignment and insertion, the microgripper approached to the mating rod and an autofocusing operation [12] was initiated for capturing the image of mating rod and supporting block.

(2) Estimation of the intrinsic and extrinsic parameters

At first, a chessboard of 10$mm$ ×17$mm$ with 1$mm$ grid length on a plane was used to estimate intrinsic parameters. After calibration, the error of each feature point was tested to give about 0.8 *pixels*. Next, the intrinsic parameters were employed for the subsequent linear and nonlinear calibrations. In the visual calibration, a virtual image including the mating rod and supporting block was employed. The feature points consisted of 6 corner points of supporting block and 2 horizontal points on the circular edge of mating rod. After final calibration, the average error of feature points by (19) gave 2 *pixels*.

(3) Estimating center coordinates of gripper tip and matching hole in AR

When the microgripper was at the approaching position, the coordinates of feature points were obtained from an AR image. An AR image with real image of gripper as well as virtual image of mating rod and supporting block was shown in Fig. 7. The center coordinate of mating hole from virtual model was (0, -20, 0.04419). The feature points of triangular plat, which was installed on the sleeve, was estimated through calibrated camera parameters. The position of $T'$ was estimated to give $T'$ = (0.903677, -26.33871, -1.136224). Since the radius of sleeve tip was *1.0 mm* and the axial distance from $T'$ to gripper tip was 3.6 *mm*, the position of gripper tip was estimated as (0.903677, -22.73871, -0.136224) in unit *mm*.

### B. AR-assisted Alignment and Assembly

For the microassembly work, a micropeg on the object and assembly stage was first gripped by microgripper under eye-to-hand CCD-1 and CCD-2. Then, the micropeg held by micro gripper was moved to the same approaching position as that of performing calibration.

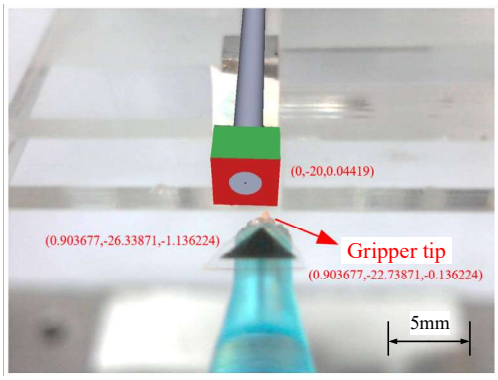


Figure 7. An AR image with real image of gripper as well as virtual image of mating rod and supporting block.

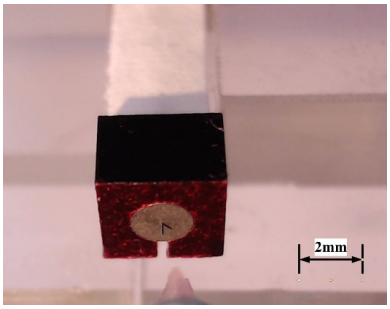

Figure 8. Finish assembly operation.

A calibrated CCD-3 along with the correspondent virtual camera for tracking the mating rod was switched on and the visual servo was activated. The coordinate of the center position of the mating hole was clearly and accurately displayed for assembly operation. In dynamic tracking process, the steady-state average error was about 3 *pixels* in registration. In the alignment operation, the real tip of micropeg and the virtual center of mating hole in the coordinate of $X_W$-axis and $Z_W$-axis were aligned in sequence. After alignment, the micropeg was operated to move and insert into the mating hole along the $Y_W$ axis. The micropeg which had been inserted into the mating hole was shown in Fig. 8. The success rate of the present AR-assisted microassembly mainly depends on the marker, position estimator, dynamic registration, microgripper, and the resolution of eye-in-hand CCD. The accuracy in estimating the tip coordinate of gripping micropeg relies on the accurate gripping pose by 3D microgripper. The resolution of CCD-3 was most essential to assure successful insertion. When the resolution was adjusted to the lowest 5 *μm/pixel*, the success rate was less than one half since the average registration error was 15 *μm* which was close to clearance of 20 *μm*. However, when the resolution was adjusted to the highest 3 *μm/pixel*, it provided accurate matching coordinates which ensured 100% successful operation of peg-in-hole microassembly. Regarding the eye-in-hand camera configuration, compared with the eye-to-hand camera in existing microassembly literature, it was flexible to perform alignment and assembly without the concern of viewing and focusing by fixed camera.

## VII. CONCLUSION

A 3D-hybrid visual calibration and visual servo in coordinate-based AR has been implemented and tested under the eye-in-hand/eye-to-hand cooperated camera system. In the microassembly test, a micropeg of diameter 80 *μm* and length 1~1.4 *mm* was aligned and assembled to a hole of diameter 100 *μm*. The coordinate-based AR technology was successfully employed in the final phase of alignment and insertion between micropeg and mating hole through eye-in-hand visual servo. In the present approach, the coordinates of feature points are available for alignment and insertion operations. The accuracy in estimating the tip coordinate of gripping micropeg relies on the accurate gripping pose by 3D microgripper. In the final phase under visual servo, an average error in registration is about 3 *pixels* and consequently, the resolution of CCD-3 is most essential to assure successful insertion. The present results reveal that the 3D model-based AR microassembly system effectively facilitates human operator to accomplish the task of peg-in-hole microassembly under high resolution of CCD-3. In addition, the eye-in-hand camera configuration is flexible to perform alignment and assembly without the concern of viewing and focusing by fixed camera in the eye-to-hand configuration.

## REFERENCES

[1] R. Li and H. Qiao, "A survey of methods and strategies for high-precision robotic grasping and assembly tasks- Some new trends," IEEE/ASME Trans. Mechatronics, vol. 24, Issue 6, pp. 2718-2732, 2019.

[2] A. Ferreira, C. Cassier, and S. Hirai, "Automatic microassembly system assisted by vision servoing and virtual reality," IEEE/ASME Trans. Mechatronics, vol. 9, no. 2, pp. 321-333, 2004.

[3] K. B. Yesin and B. J. Nelson, "A CAD model based tracking system for visually guided microassembly," Robotica, vol. 23, pp. 409-418, 2005.

[4] B. Tamadazte and E. Marchands, "CAD model based tracking and 3D visual-based control for MEMS microassembly," Int. J. Robotics Resesrch, vol. 29, no. 11, pp. 1416-1434, 2010.

[5] R. J. Chang, C. Y. Lin, and P. S. Lin, "Visual-based automation of peg-in-hole microassembly process," ASME J. Manufacturing Science and Engineering, vol. 133, 041015, pp. 1-12, 2011.

[6] A. V. Kudryavtsev, G. J. Laurent, C. Clévy, B. Tamadazte, and P. Lutz, "Analysis of CAD model-based visual tracking for microassembly using a new block set for Matlab/Simulink," Int. J. Optomechatronics, vol. 9, pp. 295-309, 2015.

[7] R. J. Chang and J. C., Jau, "Error measurement and calibration in developing virtual-reality-assisted microassembly system," Int. J. Automation Technology, vol. 9, no. 6, pp. 619-628, 2015.

[8] R. J. Chang and J. C. Jau, "Augmented reality in peg-in-hole microassembly operations," Int. J. Automation Technology, vol. 10, no. 3, pp. 438-446, 2016.

[9] R. J. Chang and J. F. Liu, "Model-based coarse-fine virtual calibration and visual servo for augmented reality-assisted peg-in-hole microassembly," ASME J. Micro- and Nano-Manufacturing, vol. 6, pp. 1-12, 04102, 2018.

[10] G. Flandin, F. Chaumette, and E. Marchand, "Eye-in-hand/Eye-to-hand cooperation for visual servoing," Proc. IEEE Int. Conf. Robotics and Automation, vol. 3, pp. 2741–2746, 2000.

[11] V. Lippiello, B. Siciliano, and L. Villani, "Eye-in-hand/Eye-to-hand multi-camera visual servoing," Proc. 44th IEEE Conf. Decision and Control, pp. 5354-5359, Spain, 2005.

[12] R. J. Chang and Y. C. Chien, "Polymer Microgripper with Autofocusing and Visual Tracking Operations to Grip Particle Moving in Liquid," Actuators, vol. 7, no. 2, 27, 2018.

[13] R. S., Wright, N., Haemel, and G., Sellers, OpenGL SuperBible: Comprehensive Tutorial and Reference, Pearson Education, 2010.

[14] Z., Zhang, "A flexible new technique for camera calibration," IEEE Pattern Analysis and Machine Intelligence, vol. 22, no. 11, pp. 1330-1334, 2000.