2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

Fast Peg-and-Hole Alignment Using Visual Compliance

Shouren Huang, Kenichi Murakami, Yuji Yamakawa, Taku Senoo and Masatoshi Ishikawa

***Abstract*— This paper presents a visual compliance strategy to deal with the problem of fast peg-and-hole alignment with large position and attitude uncertainty. With the use of visual compliance and adoption of a light-weight 3-DOF active peg, decoupled alignment for position and attitude is realized. The active peg is capable of high-speed motion and with less dynamic defects than a traditional robot arm. Two high-speed cameras, one configured as eye-in-hand and the other as eye-to-hand are adopted to provide with the task-space feedback. Visual constraints for effecting the visual compliant motion are analyzed. Alignment experiments show that peg-and-hole alignment with the proposed approach could be successfully realized with robust convergence, and on average, the alignment could be realized within 0.7 s in our experimental setting.**

I. Introduction

Peg-and-hole alignment is a well-addressed topic for autonomous assembly control. There are two issues for realizing a general peg-and-hole alignment, the position and the attitude. Consider a cylindrical peg and a cylindrical hole in the work space $\mathcal{T} = SE(3)$, let φ and ψ represent the heading angle and the pitch angle as shown in Fig.1, and if we set the position and attitude of the peg and the hole to be $\boldsymbol{P}(x, y, z), \boldsymbol{H}(x, y, z)$ and $\boldsymbol{P}(\varphi, \psi), \boldsymbol{H}(\varphi, \psi)$ respectively, the alignment problem is the minimization: $\min\{\mid \boldsymbol{P}(x, y, z) - \boldsymbol{H}(x, y, z) \mid + \mid \boldsymbol{P}(\varphi, \psi) - \boldsymbol{H}(\varphi, \psi) \mid\}$. The coupling between the position alignment and attitude alignment exists as a challenging issue for the fast convergence of the alignment process.

Traditionally, peg-and-hole alignment is usually realized by using mechanical compliance (force control) after realizing contact regulation (position control). Force control can be done by passive approaches such as the Remote Center of Compliance (RCC) [1], or active approaches with force/torque sensors [2].

Peg-and-hole alignment realized by visual feedback control also holds many advantages, such as flexibility, compensation for system and workspace uncertainty, and good tolerance to calibration error. Lately several works addressed the problem of micro peg-and-hole alignment [3],[4]. Since micro-assembly requires high accuracy within a small workspace, a global, fixed configuration of cameras is favorable. But in macro assembly with big workspaces and more dynamic uncertainties, the eye-in-hand configuration holds the advantages of higher flexibility, higher accuracy and occlusion avoidance. Yoshimi *et al.* [5] proposed an eye-in-hand approach for peg-and-hole alignment. They showed that rotational invariance could be incorporated into the strategy of peg-and-hole alignment, and an image Jacobian estimation method enables the system to be free of calibration. Their approach holds the drawback that the alignment process was complicated and time consuming. Furthermore, they mainly focused on the position alignment; the attitude alignment between peg and hole was not studied.

S. Huang, K. Murakami, Y. Yamakawa, T. Senoo and M. Ishikawa are with the Dept. of Creative Informatics, Graduate School of Information Science and Technology, Univ. of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. Huang_shouren@ipc.i.u-tokyo.ac.jp

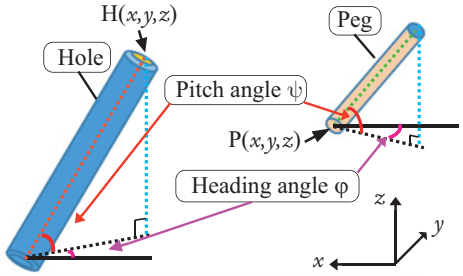


Fig. 1. Peg-and-hole alignment problem with position and attitude uncertainty

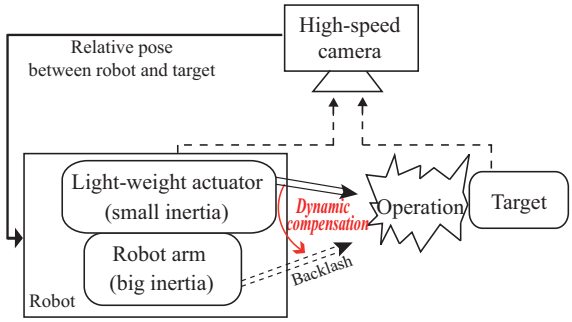


Fig. 2. Concept

Morel *et al.* [6] proposed a strategy for peg-in-hole manipulation with the combination of visual tracking and force compliance control. Their control scheme involves a position based impedance controller with force feedback, and a visual feedback loop to provide the reference trajectory to the impedance controller. The impedance controller and the vision-based controller can be designed separately, and the latter alone can represent the classical image-based visual servoing method for the peg-and-hole alignment problem. However, as the experimental results shown in [6], the modeling error and calibration error would affect the alignment results greatly, and thus the force feedback is needed to compensate for the forces undesirably generated by the 2D visual servoing.

Inspired by the so-called visual compliance control [7], this paper focus on the realization of fast macro peg-and-hole alignment with position and attitude uncertainty through

visual feedback only, without the position compensation from force-feedback control, and of course without the computation of the insertion trajectory in advance. Castano *et al.* [7] proposed the task-level visual compliance control with a hybrid vision/position control structure. Visual compliance is analogous to physical compliance, as the robot's end-effector maintains contact with a visual constraint surface, and visual compliant motion moves the end-effector along a projection ray that passes through the focal center of the camera.

In this study, one of our goal is to realize a high-speed manipulation. For high-speed operation, dynamic defects such as backlash due to large inertia, would reduce the performance of the manipulation. As shown in Fig.2, the concept of our approach is to compensate for the robot arm's dynamics (or the dynamical effects for interacting with the target) through two aspects. The first aspect is to add a light-weight high-speed actuator to the end of the robot arm. The second aspect is using high-speed cameras to provide with task-space feedback information of the relative pose between robot and target. This concept is useful in cases that the dynamics model of the robot arm is not available while high-speed manipulation is expected. As for the high-speed actuator, the high-speed hand [8] developed by our laboratory has been adopted in several applications and good performance for high-speed dynamic manipulation has been demonstrated [9].

In this paper, as one of the application of our robot control concept shown in Fig.2, we exploit a visual compliant motion strategy for the fast peg-and-hole alignment problem (with large initial errors) through adoption of a high-speed 3-DOF active peg cooperating with the robot arm. The rest of the paper is organized as follows: Section II illustrates system design and alignment methodology of the proposed approach. Section III addresses the visual compliant motion control of the robot arm and the active peg. Experiments are conducted in Section IV.

II. The Proposed Approach

A. System design

In accordance with our concept shown in Fig.2, we designed the system for peg-and-hole alignment study as shown in Fig.3. The system consists of a 4-DOF high-speed robot arm, an active peg and two high-speed cameras. The maximum velocity of the robot arm's end-effector can reach 27.22 m/s. One high-speed Eosens vision system is configured as eye-in-hand and another as eye-to-hand, both with an image resolution of 720×720 $pixels$ and a feedback rate of 1000 Hz. An angle of about 20° is configured between the heading direction of the peg and the eye-in-hand camera on the horizontal plane. The two high-speed cameras are configured to observe both the hole and the peg, thus the system falls into the endpoint closed-loop (ECL) category [12]. The light-weight active peg is realized by a high-speed 3-DOF (two rotational and one prismatic joints) finger with the peg attached to it, and with a weight of about 0.17 kg. The two rotational joints are for attitude alignment, and the prismatic joint is for the insertion after alignment. The finger can close its joints at $1800°/s$.

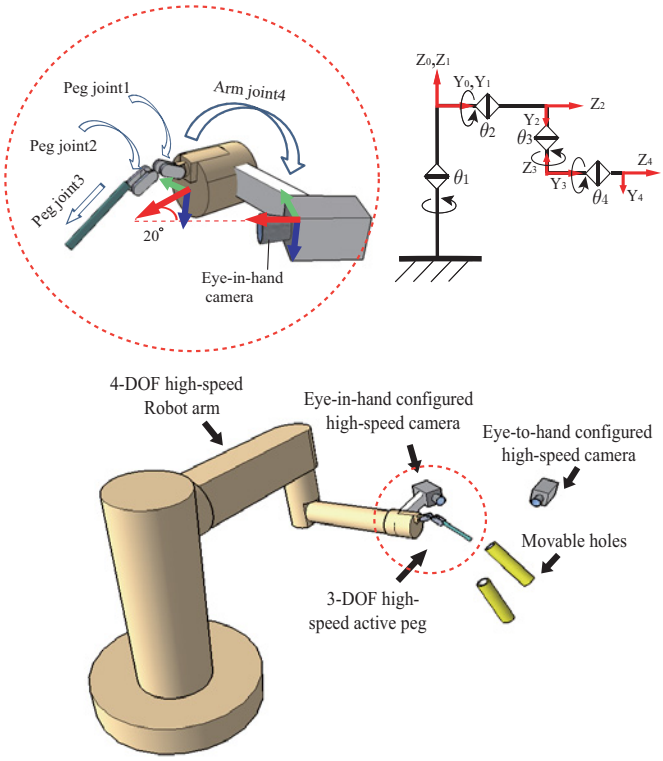


Fig. 3. System configuration

For the better clarify, hereafter we will refer to the eye-in-hand camera's frame as Σ_{ci} with its image plane ξ, and the eye-to-hand camera's frame as Σ_{ct} with the image plane ζ.

B. Alignment methodology

Consider a static hole with position and attitude uncertainty, the robot should guide the peg to realize the alignment with the hole's position and attitude. As for the conventional methods of peg-and-hole alignment, since the peg is a fixed tool on the robot arm, it's not easy to realize a fast and accurate alignment. Not only do the dynamics of the robot arm cause unwanted effects, but also, the adjustment of the position and attitude are coupled, which is not good for fast convergence.

In this paper, we present a visual compliance strategy with adoption of a 3-DOF high-speed active peg to deal with this problem, and we intend to realize a fast peg-and-hole alignment manipulation. There are three visual constraints to effect the corresponding visual compliant motions:

- Co-point constraint in images (eye-in/to-hand camera) for the peg and the hole. This visual compliant motion is effected on the robot arm. Hereafter we define the word "Co-point" as: two points located on the same image plane with a sufficient small distance. As shown in Fig.4(a), the hole's endpoint $\boldsymbol{H}(x, y, z)$ and the peg's endpoint $\boldsymbol{P}(x, y, z)$ should be controlled to be aligned on $\boldsymbol{\xi}_{end} \in \mathbb{Z}^2$ in image plane ξ. Since during the first

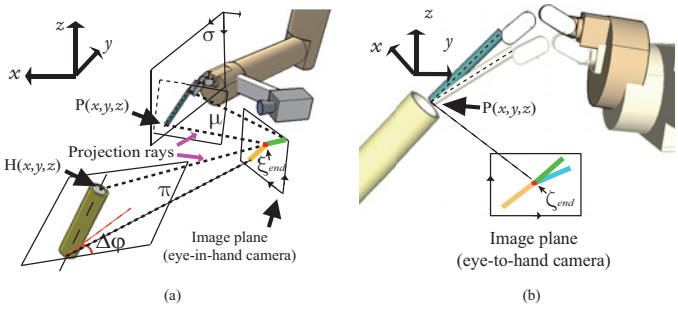


Fig. 4. Visual constraints for effecting visual compliant motion

stage of the alignment, the peg will be motionless, and this constraint is actually to effect the visual compliant motion for the robot arm along the projection ray formed by $\boldsymbol{H}(x,y,z)$ and $\boldsymbol{\xi}_{end}$.

- Line-parallel constraint in images (eye-in-hand camera) for the peg and the hole, which corresponds to the plane parallel constraint in Cartesian space. This visual compliant motion is effected on the robot arm. As shown in Fig.4(a), the active peg moves within the plane of σ. The center line of the hole in Cartesian and in image form the plane of π, and the center line of the peg in Cartesian and in image form the plane of μ. In order to keep μ and π to be parallel, the center line of the hole and the peg should be parallel in the image plane. This constraint actually realizes the alignment in the direction of the heading angle φ as illustrated in Fig.1. The robot arm's joint4 (Fig.3) will be controlled by this visual constraint to effect the visual compliant rotational motion.
- Co-point constraint in images (eye-to-hand camera) for the peg and the hole. This visual compliant motion is effected on the active peg. As shown in Fig.4(b), once the hole's endpoint $\boldsymbol{H}(x,y,z)$ and the peg's endpoint $\boldsymbol{P}(x,y,z)$ reach the same position $\boldsymbol{\zeta}_{end} \in \mathbb{Z}^2$ in the eye-to-hand camera's image plane ζ, the active peg should be controlled to keep on the position $\boldsymbol{\zeta}_{end}$ and looks like "motionless" while the robot arm is still moving under the other two visual constraints. This is to accommodate for the hole's pitch angle ψ (Fig.1).

III. Visual Compliant Motion Control

For clarifying how the visual compliant motion control is realized for the robot arm and the active peg, we will look into them one by one in the order illustrated above. Finally, the alignment algorithm for combining all of these motions will be addressed.

A. Co-point constraint (eye-in/to-hand camera)

In [7], a hybrid Jacobian matrix $\boldsymbol{\mathcal{J}}_{vc}$ is adopted, with the first two rows corresponding to vision-based control, and the third row corresponding to position based control. In this study, we adopt vision-based control only.

Let $\dot{\boldsymbol{r}}_e$ represent the velocity screw of the end-effector relative to the end-effector's frame Σ_e, and $\dot{\boldsymbol{r}}_{ci}$ be the velocity screw of the eye-in-hand camera respected to the camera's frame Σ_{ci}. ${}^e\boldsymbol{W}_{ci}$ is the transformation matrix between $\dot{\boldsymbol{r}}_e$ and $\dot{\boldsymbol{r}}_{ci}$. From the differential motion relationship between two frames [10], we have

$$\dot{\boldsymbol{r}}_e = {}^e\boldsymbol{W}_{ci}\dot{\boldsymbol{r}}_{ci} \tag{1}$$

with

$${}^e\boldsymbol{W}_{ci} = \begin{bmatrix} {}^e\boldsymbol{R}_c & S({}^e\boldsymbol{t}_c)\,{}^e\boldsymbol{R}_c \\ 0 & {}^e\boldsymbol{R}_c \end{bmatrix} \tag{2}$$

where ${}^e\boldsymbol{R}_c, {}^e\boldsymbol{t}_c$ are the rotational matrix and translational vector between the end-effector frame and the camera frame Σ_{ci}, and suppose they have been calibrated ahead. $S({}^e\boldsymbol{t}_c)$ represents the skew-symmetric matrix associated with ${}^e\boldsymbol{t}_c$.

Let $\boldsymbol{\xi} = (u,v)^T$ be a point in image plane ξ, projected from the point $[X,Y,Z]^T$ in the frame Σ_{ci}. Let $\dot{\boldsymbol{r}}_{ci} = [T_x, T_y, T_z, \omega_x, \omega_y, \omega_z]^T$, then we have

$$\dot{\boldsymbol{\xi}} = \boldsymbol{\mathcal{J}}_{img}\dot{\boldsymbol{r}}_{ci} \tag{3}$$

where $\boldsymbol{\mathcal{J}}_{img}$ is the classical image Jacobian matrix [11],

$$\boldsymbol{\mathcal{J}}_{img} = \begin{bmatrix} \frac{f}{Z} & 0 & -\frac{u}{Z} & -\frac{uv}{f} & \frac{u^2+f^2}{f} & -v \\ 0 & \frac{f}{Z} & -\frac{v}{Z} & -\frac{v^2+f^2}{f} & \frac{uv}{f} & u \end{bmatrix} \tag{4}$$

where f is the camera's focal length, and Z is depth information.

Follow the method of the partitioned approach for image-based visual servoing approach [11], we change Eqn.(3) to be

$$\dot{\boldsymbol{\xi}} = \boldsymbol{\mathcal{J}}_{tz}T_z + \boldsymbol{\mathcal{J}}_{ref}\dot{\boldsymbol{r}}_{ref} \tag{5}$$

where $\dot{\boldsymbol{r}}_{ref} = [T_x, T_y, \omega_x, \omega_y, \omega_z]^T$, $\boldsymbol{\mathcal{J}}_{tz}$ is the third column of $\boldsymbol{\mathcal{J}}_{img}$, and $\boldsymbol{\mathcal{J}}_{ref}$ is formed by the rest five columns of $\boldsymbol{\mathcal{J}}_{img}$. Then we have

$$\dot{\boldsymbol{r}}_{ref} = \boldsymbol{\mathcal{J}}^+_{ref}(\dot{\boldsymbol{\xi}} - \boldsymbol{\mathcal{J}}_{tz}T_z) \tag{6}$$

where, $\boldsymbol{\mathcal{J}}^+_{ref}$ represents the pseudo-inverse for $\boldsymbol{\mathcal{J}}_{ref}$. The camera's $z-$axis translational control is given by

$$T_z = \gamma[\alpha(mdis(\boldsymbol{\zeta}^h_{end} - \boldsymbol{\zeta}^p_{end})) + (1-\alpha)(\zeta^h_{att} - \zeta^p_{att})] \tag{7}$$

where γ is a scalar gain factor, $\boldsymbol{\zeta}^h_{end} = (u'_h, v'_h)^T, \in \mathbb{Z}^2$ refers to the hole's endpoint, and $\boldsymbol{\zeta}^p_{end} = (u'_p, v'_p)^T, \in \mathbb{Z}^2$ represents the peg's endpoint, $\zeta^p_{att} \in \mathbb{Z}^1$ is for the peg's attitude and $\zeta^h_{att} \in \mathbb{Z}^1$ for the hole's attitude, α $(0 < \alpha < 1)$ is a scalar normalizing factor. $mdis(\cdot)$ represents the Manhattan distance calculation defined as

$$mdis(\boldsymbol{\zeta}^h_{end} - \boldsymbol{\zeta}^p_{end}) = \mid u'_h - u'_p \mid + \mid v'_h - v'_p \mid \tag{8}$$

Note that the image features related to Eqn.(7) are all from the eye-to-hand camera's image plane ζ. Here we only care about the one direction regulation to avoid the local minimum of T_z, which means $\alpha(mdis(\boldsymbol{\zeta}^h_{end} - \boldsymbol{\zeta}^p_{end}))$ and $(1-\alpha)(\zeta^h_{att} - \zeta^p_{att})$ have the same sign. The attitude for the peg ζ^p_{att} is defined as

$$\zeta^p_{att} = \delta v'/\delta u' \tag{9}$$

with $\delta v', \delta u'$ the image coordinate differences between the peg's two endpoints in image plane ζ. ζ^h_{att} is defined by the same method.

With T_z and $\dot{\boldsymbol{r}}_{ref}$ we then reconstruct the camera's motion vector $\dot{\boldsymbol{r}}_{ci} = [T_x, T_y, T_z, \omega_x, \omega_y, \omega_z]^T$, and with Eqn.(1) we can obtain the end-effector's velocity screw $\dot{\boldsymbol{r}}_e$. After that, with the robot arm's forward kinematics and its Jacobian matrix, we can further calculate the corresponding joint angles.

It should be pointed out, in order to construct the image Jacobian $\boldsymbol{\mathcal{J}}_{img}$, the depth Z should be known. Here, we roughly estimate it from the eye-to-hand camera's images by

$$\widehat{Z} = L_p \frac{\mid \boldsymbol{\zeta}^p_{end} - \boldsymbol{\zeta}^h_{end} \mid}{\zeta^p_l} + K_{cp} \tag{10}$$

where L_p is the peg's length which is known, $\zeta^p_l \in \mathbb{Z}^1$ is the length of the peg in ζ, and $\mid \boldsymbol{\zeta}^p_{end} - \boldsymbol{\zeta}^h_{end} \mid$ represents the distance from the peg to the hole. K_{cp} is the compensation part due to the offset from the peg to the eye-in-hand camera.

B. Line-parallel constraint (eye-in-hand camera)

The robot arm's joint4 (as shown in Fig.3) is adopted to fit for an arbitrary heading angle of the hole by visual compliant motion. Hereafter, we will refer it as the attitude accommodating joint. In fact, in order to realize line parallel constraint motion considering the peg's motion, the easiest way is to configure the eye-in-hand camera in such a way that the vertical plane passing through the optical axis of the camera is coincide with the plane σ (Fig.4(a)). Thus we simply adjust the hole's attitude in images to be vertical to keep the constraint, since the peg's attitude in images is always the same whether it moves or not. This actually requires the camera to be set in the center of the end-effector with strict accuracy relative to the peg's pose, whereas it is usually difficult to realize due to limited physical space and assembly error. In our case, the eye-in-hand camera and the peg are configured with an angle, and we adopt a new image feature $\xi_{head} \in \mathbb{Z}^1$ to realize the parallel constraint regulation. ξ_{head} is defined as the ratio between hole's length $\xi^h_l \in \mathbb{Z}^1$ and its attitude value $\xi^h_{att} \in \mathbb{Z}^1$ in images

$$\xi_{head} = \xi^h_l / \xi^h_{att} \tag{11}$$

where ξ^h_l is defined by the Manhattan distance between the hole's two endpoints in images, and ξ^h_{att} is defined as

$$\xi^h_{att} = \delta v / \delta u \tag{12}$$

with $\delta v, \delta u$ the image coordinate differences between the hole's two endpoints in ξ.

Let us check the new image feature ξ_{head} under different pose of the hole. With the condition that the 3D position and the pitch angle's alignment between peg and hole have been realized, we move the attitude accommodating joint. As Fig.5 shows, with the attitude accommodating joint's rotation, the hole's attitude changes in images and a different configuration of the hole will need a different rotation angle for alignment. When the attitude of the peg and that of the hole reach the same value, we write down the corresponding value of the new image feature. We could see that in three situations, the value of ξ_{head} is 74, 73 and 73 (the unit has been normalized from $pixels$), and even if we do more tests with different configurations of the hole, we still find that ξ_{head} varies around 73 with small errors. Although here we will omit the strict deviation of this approximately invariant parameter, we adopt ξ_{head} as the image feature for the regulation of the attitude accommodating joint to realize the visual complaint motion. However calibration for ξ_{head} is needed. During the experiments, we will see that this new image feature could work very well for the alignment. The control law for this visual compliant motion is

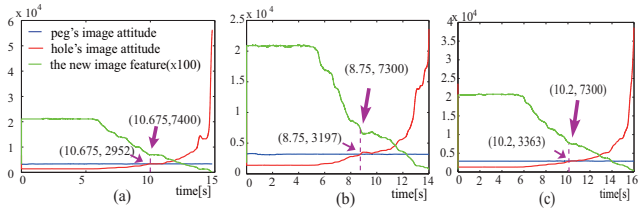


Fig. 5. Tests of ξ_{head} under different configurations of the hole

$$\tau_{head} = -K^p_{head}(\xi_{head} - \xi^*_{head}) - K^d_{head}\dot{q} + g \tag{13}$$

where, K^p_{head}, K^d_{head} are positive gain factors, τ_{head} is the input torque for the attitude accommodating joint, ξ^*_{head} is the calibrated value of ξ_{head} for alignment, $\dot{q}$ is the joint velocity, and g is the gravity compensation mainly due to the camera (the active peg is light-weight). Note that, the value of ξ^*_{head} is mainly affected by the external parameters of the eye-in-hand camera.

C. Co-point constraint (eye-to-hand camera)

As have been defined above, the attitude of peg and hole in the eye-to-hand camera's images are ζ^p_{att}, ζ^h_{att} respectively. These two are mainly reflecting the pitch angle ψ (Fig.1). Suppose the attitude angle of the peg is initially smaller than the hole's, under the robot arm's visual compliant motion, the hole's endpoint $\boldsymbol{H}(x, y, z)$ and the peg's endpoint $\boldsymbol{P}(x, y, z)$ reach the same point $\boldsymbol{\zeta}_{end} = (u'_0, v'_0)^T$ in image plane ζ. Since the robot arm is still moving, the peg's endpoint will generate a small displacement $\Delta\boldsymbol{\zeta}_{end} = (\Delta u', \Delta v')^T$ from $\boldsymbol{\zeta}_{end}$ in images. With the resolved-rate control, the active peg is controlled to eliminate the displacement, and as a result, the pitch angle of the peg becomes larger until the regulation finishes, and the adjusting process acts like the physical compliant motion. The control law for the active peg is

$$\boldsymbol{\tau}_{peg} = -K^p_{peg}\boldsymbol{J}^T(\frac{L_p}{\zeta^p_l}\Delta\boldsymbol{\zeta}_{end}) - K^d_{peg}\dot{\boldsymbol{q}}_p + \boldsymbol{g} \tag{14}$$

where, K^p_{peg}, K^d_{peg} are positive definite coefficient matrices, $\boldsymbol{\tau}_{peg}$ are input torques for the active peg's two rotational joints, and $\dot{\boldsymbol{q}}_p$ is the peg's joint velocity, L_p/ζ^p_l acts as a scale factor, $\boldsymbol{g}$ is the gravity compensation, and $\boldsymbol{J}^T$ is the transpose of the active peg's Jacobian matrix. More details for the task-space PD control could be found in [13].

It should be noted that this simple approach is fit for our high-speed visual feedback and the high-speed active peg which has big friction forces for the joints, thus the dynamic effects from the robot arm could be transferred to the endpoint of the peg, and visual compliant motion can be realized.

D. Peg-and-hole alignment

The peg-and-hole alignment is realized with the combination of the three visual compliant motions described above. The algorithm flow is shown in Fig.6. To summarize, we have four convergence criteria from the two cameras for judging whether the alignment realized or not. They are: $\boldsymbol{err}^{\xi}_{end} = \boldsymbol{\xi}^{p}_{end} - \boldsymbol{\xi}^{h}_{end}$, the error between the peg and the hole's endpoint in image plane ξ; $err^{\xi}_{head} = \xi_{head} - \xi^{*}_{head}$, the error input for the line-parallel compliant motion; $err^{\zeta}_{att} = \zeta^{p}_{att} - \zeta^{h}_{att}$, the attitude error between peg and hole in image plane ζ; and $err^{\zeta}_{d} = \mid \boldsymbol{\zeta}^{p}_{end} - \boldsymbol{\zeta}^{h}_{end} \mid$, the distance between peg and hole's endpoint. The convergence time for alignment is defined as the time between the start of $phase$ 1 and the end of $phase$ 2 (Fig.6).

During the $phase$ 2 (Fig.6), since the active peg will be regulated to realize visual compliant motion, $\boldsymbol{\xi}^{p}_{end}$ would change, whereas the robot arm takes $\boldsymbol{\xi}^{p}_{end}$ as the target position for regulation. Since we want to realize a fast manipulation for the robot arm, a group of relatively large proportional gain factors is adopted. The high-speed active peg's motion may bring sudden changes of $\boldsymbol{\xi}^{p}_{end}$, which would cause unwanted effects for the robot arm's dynamic performance. In order to deal with such problem, as well as to prevent the sudden impact from the right start of the regulation due to the initial large visual errors, a shunting model [14] filter is adopted to modify the visual error $\boldsymbol{err}^{\xi}_{end} = \boldsymbol{\xi}^{p}_{end} - \boldsymbol{\xi}^{h}_{end}$. The shunting model is a neural-dynamics model and a typical shunting equation can be described as

$$\frac{dx_i}{dt} = -Ax_i + (B - x_i)S^{e}_{i}(t) - (D + x_i)S^{i}_{i}(t) \quad (15)$$

where $x_i \in \Re^1$ is the neural activity of ith-neuron. A, B and D are nonnegative constants describing the passive decay rate, the upper and lower bounds of x_i respectively. S^{e}_{i} and S^{i}_{i} are the excitatory input and the inhibitory input to the neuron respectively. The shunting model has the properties of automatic gain control, the state response bounded to the finite region $[-D, B]$, and smooth outputs even with inputs having sudden stimulus. Let the upper and lower bounds $B = D$, then the steady-state solution of Eqn.(15) is given by

$$x_i = x_i \frac{B}{A + \mid x_i \mid}, B = D \quad (16)$$

Then we have

$$\boldsymbol{err}^{\xi}_{end} = \boldsymbol{err}^{\xi}_{end} \frac{B}{A + \mid \boldsymbol{err}^{\xi}_{end} \mid} \quad (17)$$

Note that the parameter B decides the bounds, and A affect the duration time to reach the steady state.

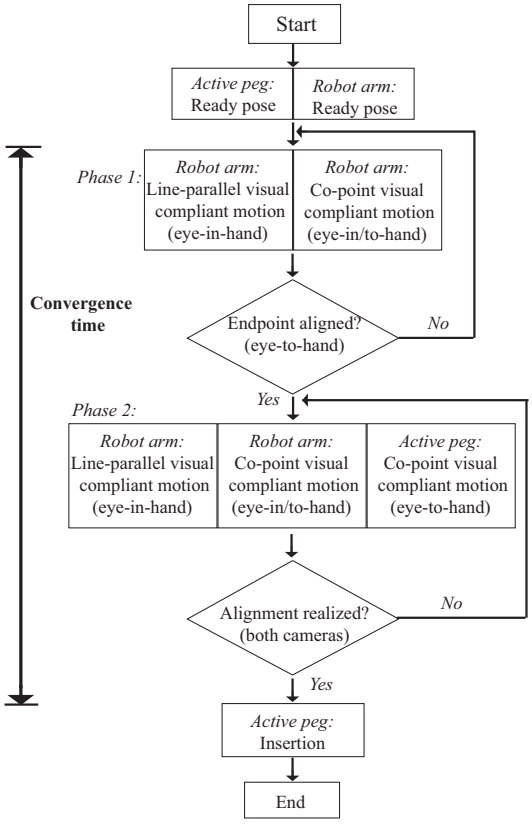


Fig. 6. Algorithm flow of peg-and-hole alignment

IV. Experiment

A. Experimental setting

The active peg's outside diameter was 10 mm, and the hole's inner diameter was 14 mm. We set the stroke of the insertion for the peg into the hole to be 20 mm. For both peg and hole, two marks made of light-reflecting material were fixed on the end part to indicate the peg and the hole's endpoints. We set the length for hole and peg to be 40 mm and 24 mm respectively. The image processing algorithm calculated the moment feature for each mark to generate peg's and hole's image position ($\boldsymbol{\xi}^{p}_{end}$, $\boldsymbol{\xi}^{h}_{end}$ in image plane ξ; $\boldsymbol{\zeta}^{p}_{end}$, $\boldsymbol{\zeta}^{h}_{end}$ in image plane ζ), attitude angle (ξ^{h}_{att} in ξ; ζ^{p}_{att}, ζ^{h}_{att} in ζ), and the length information (ξ^{h}_{l} in ξ; ζ^{p}_{l} in ζ).

The hole was fixed on a pan-tilt platform with 50 mm offset from the platform's center, thus the hole's position $\boldsymbol{H}(x, y, z)$ and its attitude $\boldsymbol{H}(\varphi, \psi)$ could be set to different configurations. In the experiments, three different configurations were set randomly for the hole, and then three times' alignments were realized continuously. For the first alignment, from 0 - 5.0 s is for moving the active peg and the hole to the preparation pose. For the second and third alignment, the preparation time takes 0.7 s. The insertion takes 0.5 s for each alignment (if alignment succeeds).

B. Experimental result

Figure 8 and 9 show the process of the peg-and-hole alignment for three different configurations of the hole

TABLE I

AVERAGE ALIGNMENT PERFORMANCE.

Pose	*Converge time* (s)	*Success rate* (%)
1	0.39	85
2	0.68	90
3	0.53	85

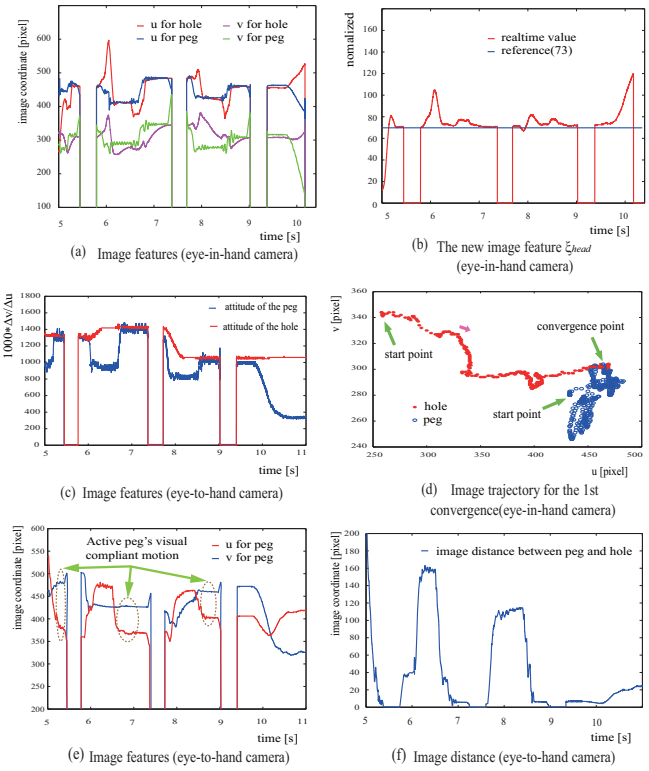
with two different point of view. The position and attitude alignment for the three times were converged at 0.366 s, 0.714 s, 0.456 s respectively (excluding the preparation time and insertion time), and then the insertions were triggered.

As shown in Fig.7(a), from the view of the eye-in-hand camera, the peg's position realized alignment with the hole's position for the three times, and as Fig.7(b) shows, for the regulation of the heading attitude by the image feature ξ_{head}, also reached the reference value for each time. From the view of the eye-to-hand camera, as Fig.7(c) shows, the attitude regulation for the pitch angle realized convergence for each time; in Fig.7(f), the image distances between the peg and hole's endpoint when alignment converged for three times were almost the same. From Fig.7(e), we can see that while the active peg conducted the visual compliant motion, the peg's endpoint could be kept within a small range although affected by the arm's motion and its dynamics. Finally, Fig.7(d) shows the image trajectory for the hole and the peg's endpoint during the first alignment. Since the insertion of the peg into the hole will cause the peg's endpoint (mark) lost in the cameras, there are three jumps of the image features to zero. Table I shows the average convergence time, success rate of the hole's three different pose for 20 trials.

In the experiment, for three configurations of the hole, the success rate of the alignment (insertion of the peg into the hole successfully conducted) was about 85%, with an average time of about 0.7 s (excluding the preparation time). The failures have happened mainly due to the position misalignment, which resulted from the noises of the images and the dynamics of the robot arm to our analysis. Since in the experiment, we set the peg's outside diameter and the hole's inner diameter to be 10 mm, 14 mm respectively, the tolerance error for the insertion is ± 2 mm (Suppose the attitude have been aligned well). The videos of the experiment can be found on the website [15].

C. Discussion

As addressed in the experimental result, the misalignment had partially resulted from the dynamics of the robot arm. In other words, the results were affected by the robot arm's dynamics in the directions along the two axes of the image plane ξ during the $phase$ 2 (Fig.6). As a matter of fact, currently the active peg can only conduct 2D motion due to the limited DOFs, and the dynamic compensation as proposed in our concept (Fig.2) could only be realized in the direction along the projection ray formed by $\boldsymbol{H}(x, y, z)$ and $\boldsymbol{\xi}_{end}$ (Fig.4(a)). In order to compensate for the other directions' dynamic effects brought by the robot arm, corresponding DOFs need to be added.



(a) Image features (eye-in-hand camera)
(b) The new image feature ξ_{head} (eye-in-hand camera)
(c) Image features (eye-to-hand camera)
(d) Image trajectory for the 1st convergence(eye-in-hand camera)
(e) Image features (eye-to-hand camera)
(f) Image distance (eye-to-hand camera)

Notes:

1) 0 ~ 5.0[s]: preparation ; 5.0 ~ 5.366 [s]: 1st alignment; 5.366 ~ 5.866 [s]: insertion
5.866 ~ 6.566[s]: preparation ; 6.566 ~ 7.280 [s]: 2nd alignment ; 7.280 ~ 7.780 [s]: insertion
7.780 ~ 8.480[s]: preparation ; 8.480 ~ 8.936 [s]: 3rd alignment ; 8.936 ~ 9.436[s]: insertion

2) The insertion will cause image features' value to zero.

Fig. 7. Image features during alignment process

V. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a visual compliance approach for macro peg-and-hole alignment with two high-speed cameras and a high-speed 3-DOF active peg. Large position and attitude uncertainty of the hole was addressed. By exploiting the visual constraints for the alignment process, visual compliant motion could be effected to realize a robust and fast convergence. With the eye-in-hand and eye-to-hand configured high-speed vision sensors, task-space regulation was realized. As a result, the alignment of peg and hole with position and attitude uncertainty could be realized with robust convergence in most cases. For most cases of the alignment, the convergence could be realized within 0.7 s under our experimental setting, with a success rate about 85%. Since the failures mainly resulted from the noises of our simple image processing algorithms as well as the robot arm's dynamics, a much robust image feature extracting algorithm may improve the success rate.

As for the application, we think the visual servoing approach for peg-and-hole alignment with position and attitude uncertainty is directly applicable for the cases with rough accuracy requirement, and also can be applied for high accuracy applications as a pretreatment step to reduce jamming or wedging conditions that usually happen in traditional approaches.

Although we can say the robot arm's dynamics have been

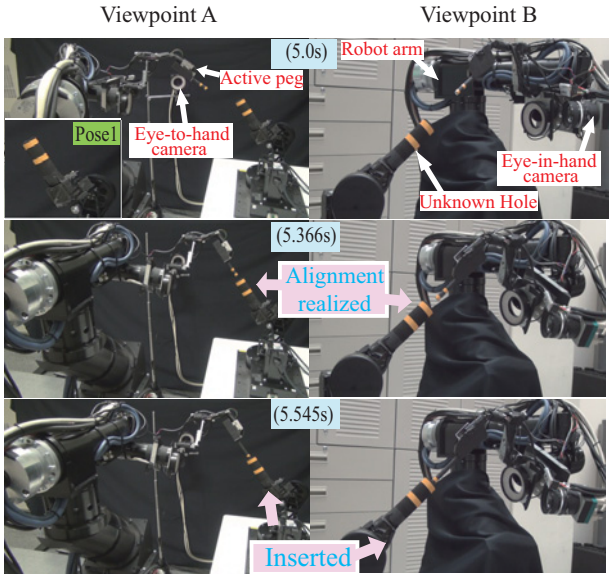
compensated by the visual compliant motion of the active peg in the limited direction, since currently the active peg can only conduct constrained 2D motion, there are lacking of DOFs in other directions for further compensation. In the future work, we would like to develop an active peg with more DOFs to realize a more fast and accurate manipulation.

VI. Acknowledgments

We would like to thanks Dr. Carson Reynolds and Dr. Niklas Bergström for suggesting many improvements. Especially, we would like to present our deep condolences to the passing away of our highly respected Dr. Carson Reynolds, and we will miss him forever.

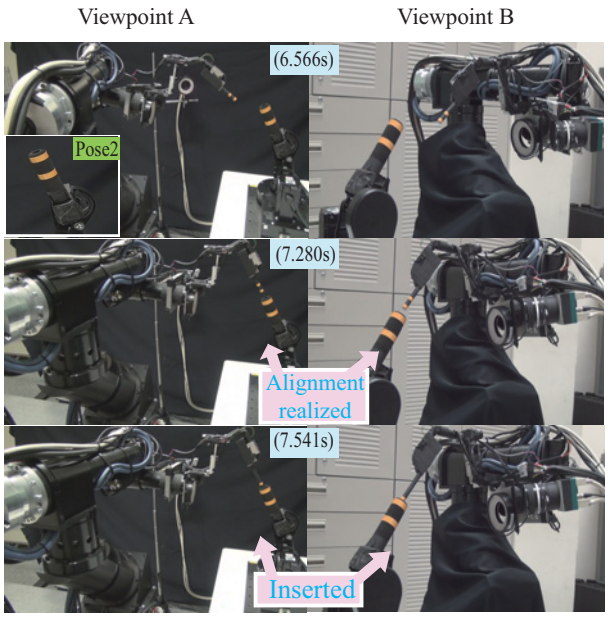
References

[1] Whitney, D. E., Quasi-static assembly of compliantly supported rigid parts, *ASME Journal of Dynamic Systems Measurement and Control*, vol.104, no. 1, pp. 65-77, 1982.
[2] Kim, I.; Lim, D., Active Peg-in-hole of Chamferless Parts using Force/Moment Sensor, *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 948-953, Oct. 1999.
[3] Wang, J. et al., Microassembly of micro peg-and-hole using uncalibrated visual servoing method, *Precision Engineering*, Vol. 32, pp. 173-181, 2008.
[4] Chang, R. J. ; Lin, C. Y. and Lin, P. S., Visual-Based Automation of Peg-in-Hole Microassembly Process, *Journal of Manufacturing Science and Engineering*, Vol. 133, pp. 041015-1-041015-12, Aug. 2011.
[5] Yoshimi, B. H. ; Allen, P. K., Active, uncalibrated visual servoing, *Proc. IEEE Int. Conf. on Robotics and Automation*, pp. 156-161, 1994.
[6] Morel, G.; Malis, E.; Boudet, S., Impedance based combination of visual and force control, *Proc. IEEE Int. Conf. on Robotics and Automation*, vol. 10, no.3, pp. 1743174, 1998.
[7] Castano, A.; Hutchinson, S., Visual compliance: task-directed visual servo control, *IEEE Trans. on Robotics and Automation*, vol. 10, no.3, pp. 334-342, June 1994.
[8] A. Namiki; Y. Imai; M. Ishikawa; and M. Kaneko, Development of a High-speed Multifingered Hand System and Its Application to Catching, *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 2666-2671, Oct. 2003.
[9] http://www.k2.t.u-tokyo.ac.jp/fusion/HighspeedHand/
[10] Spong, M. W.; Hutchinson, S.; Vidyasagar, M., Robot modeling and control, *John Wiley & Sons, Inc.*, 2006.
[11] Corke, P. I. ; Hutchinson, S. A., A new partitioned approach to image-based visual servoing control, *IEEE Trans. on Robotics and Automation*, vol. 17, no.4,pp. 507-515, Aug. 2001.
[12] S. Hutchinson, G. Hager, and P. Corke, A tutorial on visual servo control, *IEEE Trans. Robot. Automa.*, vol. 12, no. 5, pp. 651-670, Oct. 1996.
[13] Cheah, C.C., Task-space PD control of robot manipulators: Unified analysis and duality property, *I. J. Robotics Res.*, vol. 27, no. 10, pp. 1152-1170, 2008.
[14] Grossberg S., Nonlinear neural networks: Principles, mechanisms, and architecture, *Neural Networks*, vol. 1, pp. 17-61, 1988.
[15] http://www.k2.t.u-tokyo.ac.jp/fusion/PegHoleAlignment/

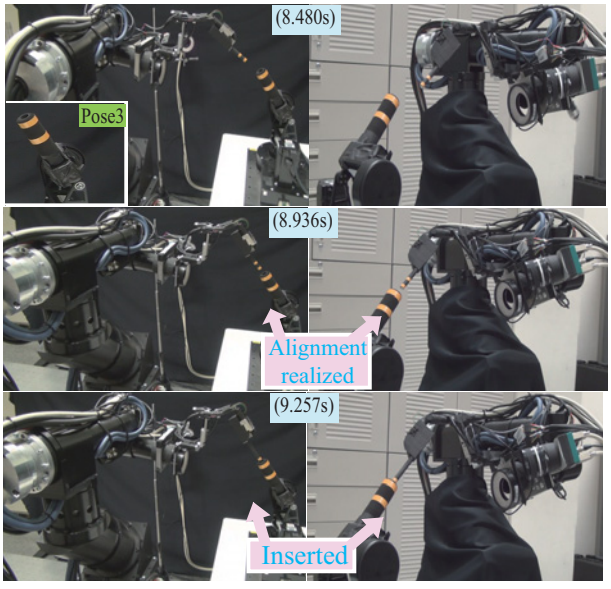


(a) Alignment - 1 (Hole's pose1)

Fig. 8. Continuous images for three times' alignment from two viewpoints



(b) Alignment - 2 (Hole's pose2)



(c) Alignment - 3 (Hole's pose3)

Fig. 9. Continuous images for three times' alignment from two viewpoints (continue)