

# Peg-in-Hole Using 3D Workpiece Reconstruction and CNN-based Hole Detection

Michelangelo Nigro<sup>†</sup>, Monica Sileo<sup>†</sup>, Francesco Pierri<sup>†</sup>,  
 Katia Genovese<sup>†</sup>, Domenico D. Bloisi<sup>\*</sup>, Fabrizio Caccavale<sup>†</sup>

**Abstract**—This paper presents a method to cope with autonomous assembly tasks in the presence of uncertainties. To this aim, a Peg-in-Hole operation is considered, where the target workpiece position is unknown and the peg-hole clearance is small. Deep learning based hole detection and 3D surface reconstruction techniques are combined for accurate workpiece localization. In detail, the hole is detected by using a convolutional neural network (CNN), while the target workpiece surface is reconstructed via 3D-Digital Image Correlation (3D-DIC). Peg insertion is performed via admittance control that confers the suitable compliance to the peg. Experiments on a collaborative manipulator confirm that the proposed approach can be promising for achieving a better degree of autonomy for a class of robotic tasks in partially structured environments.

## I. INTRODUCTION

The evolution of manufacturing in the last decade demands for more autonomous, safe and effective robotic systems, capable of fast adaption to rapidly changing production requirements. Such a flexibility can be achieved by endowing robotic systems with advanced capabilities of executing complex tasks in partially structured environments. Thus, new approaches, integrating different methodologies and technologies, are to be investigated.

The Peg-in-Hole assembly task is a challenging task due to accuracy required both for detecting the hole and positioning the robot. The common approach to Peg-in-Hole includes two steps: the search, aimed at localizing the hole and aligning the peg to its axis, and the insertion. Regarding the insertion phase, the accumulated errors due to imperfect knowledge of the hole position and robot positioning errors, make the pure position control approach possible only in the presence of generous clearances. Thus, active or passive force control methods are to be adopted, in particular the approaches based on impedance control [1] are the most popular. Impedance approaches to the Peg-in-Hole can be found in [2], [3] for single-arm robots or in [4] for multi-arm systems.

Regarding the search phase, the existing approaches can be roughly classified into those based on visual sensor feedback and those based on exploration of the hole neighborhood. Among the first category, in [5] a high-speed camera is adopted to align the peg to the hole, while more recently [6] proposed a visual coaxial system. Whilst high performance can be achieved via visual methods, in practical scenarios they can suffer from light conditions, texture or reflection of the objects. Moreover, visual servoing methods are affected by calibration errors and by the fact that often the peg occludes the eye-in-hand camera field of view. Regarding the exploration of the hole neighborhood, it often requires

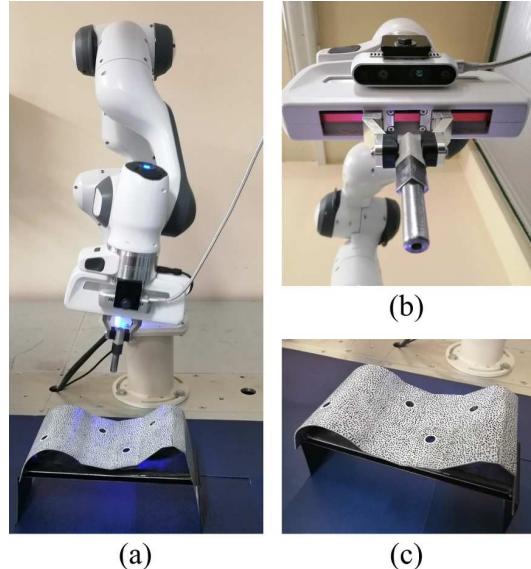


Fig. 1. Peg-in-Hole application scenario: a) the robot manipulator used in the experiments; b) visual sensor and the peg; c) workpiece.

the use of force-torque sensors, as in [2], where a shape recognition algorithm is adopted to extract the outline of the peg from the force-torque sensor data collected during the contact with the object surface. A common approach is to map the interaction moments onto the position and tilt of the hole, in order to guide the peg pose during the insertion [7].

The presence of force/torque sensors increases the overall system cost. Thus, approaches aimed at estimating the state of contact using only joint position sensors have been developed: e.g., in [8] a blind search method is proposed, consisting in tilting and covering the hole neighborhood by moving the peg along an assigned spiral path. Similarly, in [9], dealing with the insertion of a charging plug in its socket, a search path describing a Lissajous curve is adopted, while the wrist-mounted force/torque sensor is replaced by the joint torque sensors commonly mounted on collaborative robots.

Blind search methods are often time-consuming and require an accurate initial estimate of the hole position in order to converge. Hence, visual and force-torque data have been used together in [10], where a learning-based visual servoing is adopted to map the distance from the peg center to the hole, while a spiral search is in charge of the fine alignment. Deep learning has been recently used in [11], where a self supervised multi-modal representation of the sensory output is applied, and in [12], where a multi-layer perceptron network is trained on a data set including object position and interaction forces for a polyhedral pegs in contact with the holes.

In this paper, an autonomous strategy, able to cope with large errors on pose of the target workpiece, is proposed

<sup>†</sup>School of Engineering, <sup>\*</sup>Department of Mathematics, Computer Science, and Economics, University of Basilicata, 85100 Potenza, Italy.

Corresponding author:francesco.pierri@unibas.it

This research has been supported by the project ICOSAF (Integrated collaborative systems for Smart Factory - ARS01\_00861), funded by MIUR under PON R&I 2014-2020.



Fig. 2. Peg with the attached frame (left) and its dimensioning (right).

and experimentally tested for a Peg-in-Hole assembly task. A collaborative robot manipulator (Fig. 1(a)) equipped with an Intel Realsense D435 camera (Fig. 1(b)) is considered. The robot carries a steel peg (Fig. 1(b)), while the workpiece is characterized by non-flat steel surface that holds four unchamfered holes (Fig. 1(c)) with small peg-hole clearance. It is assumed that the workpiece is positioned in the robot workspace with positional uncertainty much greater than the size of the holes. A CNN based approach for object detection is combined with a 3D surface reconstruction method to accurately localize the holes. In detail, the holes are detected by using the YOLOv3 detector [13] provided by the Darknet framework [14]. The 3D surface reconstruction is obtained with 3D-Digital Image Correlation (3D-DIC), a non-contact full-field technique widely used for shape, deformation, and strain measurements in experimental mechanics [15]. From a series of image-pairs of contiguous surface areas (with partial overlap), 3D-DIC computes highly dense and regularly spaced 3D points over an extended field of view of the test part. Contrarily to single-camera eye-in-hand DIC systems [16], 3D-DIC is used to merge the reconstructed patches within the same reference system, without information on the end-effector pose. A continuous distribution of normal vectors is computed over the tessellated surface and then transferred to the robot reference frame together with the information on the 3D position of the hole. Finally, the peg insertion is performed by means of an admittance control [17], conferring mechanical impedance behavior to the peg, without resorting to a wrist-mounted force/torque sensor.

## II. SYSTEM DESCRIPTION

The considered setup consists of a robot manipulator equipped with a camera in eye-in-hand configuration. In particular, the Franka Emika Panda robot has been used, characterized by 7 revolute joints, each mounting a torque sensor, and equipped with an Intel Realsense D435 camera, mounted on the end-effector via a 3D printed support (see Fig. 1b). The camera holds two stereo infrared imagers, an RGB module and an infrared projector. The task is to insert the peg in Fig. 2, with a diameter of 12.2 mm, in one of the holes, characterized by a diameter of 12.4 mm and unknown tilt, of a workpiece manufactured by bending and forming a steel sheet into a geometry with either flat and curved regions (Fig. 3). It is assumed that the peg is symmetric and rigidly grasped by the robot gripper.

The coordinated frames of interest are the inertial frame,  $\Sigma$ , coincident with the robot base frame, the frame attached to the robot end-effector,  $\Sigma_e$ , the frame attached to the tip of the peg,  $\Sigma_p = \{O_p, n_p, s_p, a_p\}$  (Fig. 2) and  $\beta$  reference frames  $\Sigma_{h_i} = \{O_{h_i}, n_{h_i}, s_{h_i}, a_{h_i}\}$  ( $i = 1, \dots, \beta$ ), each attached to the center of the  $\beta$  holes (Fig. 3).

Due to the assumption of rigid grasp for the peg, the

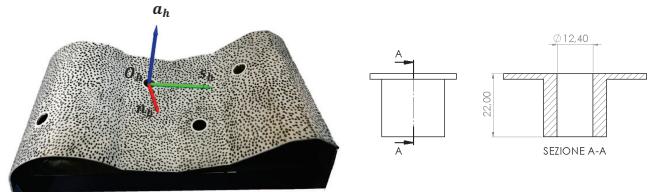


Fig. 3. Workpiece with a  $\Sigma_h$  frame (left) and hole dimensioning (right).

position  $p_p$  of  $O_p$  can be described as

$$p_p = p_e + R_e p_p^e, \quad (1)$$

where  $p_e$  and  $R_e$  are, respectively, the position and orientation of the robot end-effector, given by the standard direct kinematics,  $p_p^e$  is the constant relative position of  $O_p$  in  $\Sigma_e$ . The orientation of  $\Sigma_p$  is represented by the rotation matrix

$$R_p = R_e R_p^e, \quad (2)$$

where  $R_p^e$  is the constant relative rotation matrix of the peg frame with respect to end-effector frame.

The dynamics of a  $n$  degrees of freedom manipulator is given by

$$M(\mathbf{q})\ddot{\mathbf{q}} + C(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} + \mathbf{F}\dot{\mathbf{q}} + \mathbf{g}(\mathbf{q}) = \boldsymbol{\tau} + \boldsymbol{\tau}_e \quad (3)$$

where  $\mathbf{q} \in \mathbb{R}^n$  ( $\dot{\mathbf{q}}$  and  $\ddot{\mathbf{q}}$ ) is the vector of joint positions (velocities and accelerations),  $\boldsymbol{\tau} \in \mathbb{R}^n$  is the vector of joints torques,  $M(\mathbf{q}) \in \mathbb{R}^{n \times n}$  is the symmetric and positive definite inertia matrix,  $C(\mathbf{q}, \dot{\mathbf{q}}) \in \mathbb{R}^{n \times n}$  is the centripetal and Coriolis terms matrix,  $\mathbf{g}(\mathbf{q}) \in \mathbb{R}^n$  is the vector of gravity terms,  $\mathbf{F} \in \mathbb{R}^{n \times n}$  is the matrix of viscous friction terms. The term  $\boldsymbol{\tau}_e = \mathbf{J}^T(\mathbf{q})\mathbf{h} \in \mathbb{R}^n$  represents the torques induced at the joints by the contact wrench  $\mathbf{h} = [\mathbf{f}^T \ \boldsymbol{\mu}^T]^T \in \mathbb{R}^6$  exerted by the environment on the peg tip, where  $\mathbf{f} \in \mathbb{R}^3$  is the force and  $\boldsymbol{\mu} \in \mathbb{R}^3$  is the moment. The matrix  $\mathbf{J}(\mathbf{q}) \in \mathbb{R}^{n \times n}$  is the Jacobian of the robot, mapping the joint velocities in the linear and angular velocity of the frame  $\Sigma_p$ .

## III. PEG-IN-HOLE STRATEGY

In an assembly line, in the presence of small production volumes and complex 3-D parts, often the positioning of the target object is manually operated. Hence, it is assumed that the workpiece is located within the robot workspace, but uncertainties on both the holes' position and their tilt can be far larger than the task tolerance.

To tackle the above described problems, the following strategy (depicted in Fig. 4) is proposed:

- The robot moves its eye-in-hand camera over a semi-sphere spanning its workspace, in such a way to scan the workpiece surface. In each position, a pair of images is acquired with the previously calibrated eye-in-hand stereo-cameras system. In the current setup, the two Realsense IR cameras are used.
- The acquired images are the input for a CNN that detects the holes on the workpiece surface.
- Each pair of stereo-images from the acquired series is processed with 3D Digital Image Correlation (DIC) [15] to retrieve the 3D position of a highly dense set of points on a portion of the workpiece surface; a subsequent merging operation allows the transformation of each reconstructed point cloud from the local coordinate system of the master camera to a common global reference

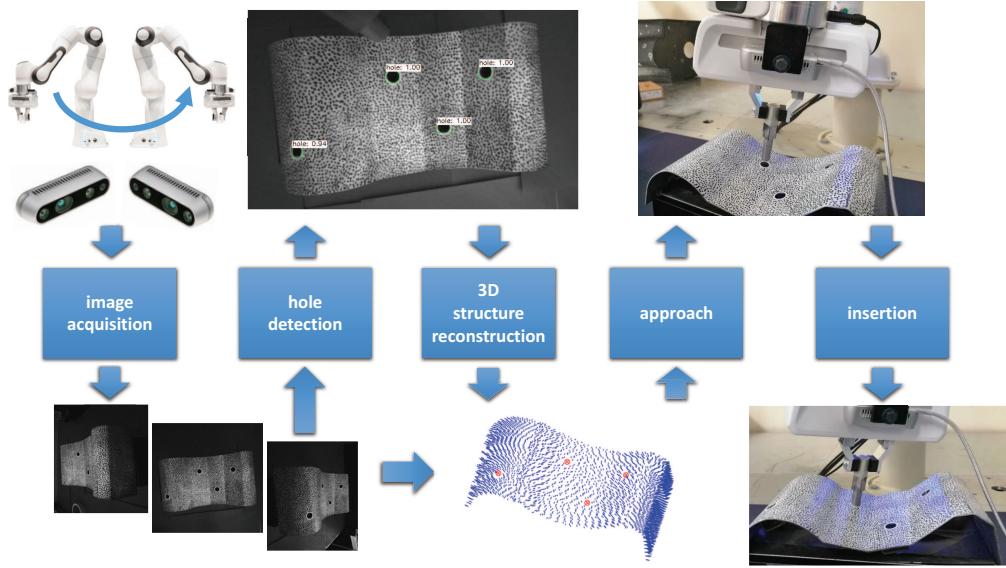


Fig. 4. The proposed peg in hole pipeline.

system. Stereo-triangulation is used to determine the 3D position of the centers of the holes. Finally, the distribution of the local normal vector is computed over the tessellated surface reconstructed with stereo-DIC.

- The robot, moves the peg close to the hole, aligning the approach unit vector with the hole axis (*approach* phase).
- The peg is inserted in the hole by moving the robot under an admittance control strategy, so as to confer the suitable compliance to the peg (*insertion* phase).

#### A. Holes detection

A supervised approach has been chosen to detect the holes on the surface of the workpiece. This choice is motivated by the fact that the surface is textured and this can lead to false positives when using unsupervised computer vision techniques (e.g., the Hough circle transform). Among the different supervised object detection architectures, the YOLOv3 [13] has been chosen, since it is faster than classifier-based systems but with similar accuracy. YOLOv3 makes predictions with a single network evaluation by considering object detection as a single regression problem. Given an image, YOLOv3 exploits a 53-layers CNN (called Darknet-53) to generate a vector of bounding boxes and class predictions in the output. To create the “hole detector”, 175 images of size  $640 \times 480$  and captured at different distances from the holes, have been annotated using the LabelImg tool. We considered 108 images as training set and the remaining 67 as test set. An Intel Xeon 3.7 GHz CPU 32 GB RAM with a NVIDIA Quadro P4000 8GB GPU has been used to carry out the training phase, which required about 10 hours to complete with batch and subdivision sizes of 64 and 16, respectively. The processing time for detecting the holes on a single frame is about 27ms. Holes are detected into each image pair, creating a binary mask to facilitate the subsequent greyscale thresholding and centroids computation.

#### B. 3D surface reconstruction

The workpiece surface is provided with a stochastic black speckle pattern on a white background, thus allowing the

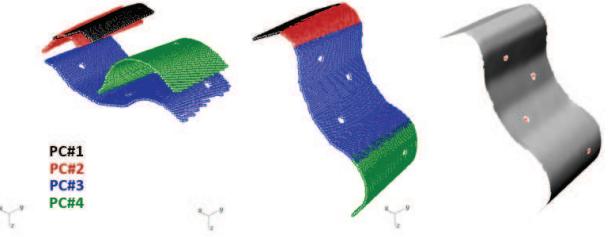


Fig. 5. 3D reconstruction and merging of the point clouds from four different contiguous positions of the eye-in-hand camera. The z axis is the optical axis of the master camera in the first position of the sequence.

implementation of an area-based image registration with a subset-based DIC approach. Briefly, for each image pairs captured by the stereo IR Realsense sensor at the maximum spatial resolution ( $1280 \times 720$  pixels), a regular dense grid of points (5 pixel pitch) is defined over the region of interest (ROI) of the reference image (master camera). The DIC algorithm then seeks for the best correspondence in the target image on the basis of the local distribution of the greyscale intensity over a subset of  $21 \times 21$  pixels around each data point. The sensor coordinates of the pairs of corresponding image points are hence used to reconstruct the position of the 3D workpiece point via triangulation. To this aim, the stereo-camera system was previously calibrated by using 30 images of a 2D checkerboard flat pattern according to the camera calibration method proposed in [18]. An average reprojection error of 0.036 and 0.042 pixels was found for the right and left camera, respectively. The target was reconstructed in the 30 positions over the measurement volume with an error of  $0.24 \pm 0.17$  mm. For each position of the scanning sequence, a point cloud of a portion of the workpiece surface is reconstructed in the reference system of the master camera (Fig.5). DIC is then used to match overlapping portions of the ROIs in image pairs from contiguous hand positions. Finally, the rigid transformation that overlaps with the minimum distance the corresponding points data in two contiguous point clouds is found through nonlinear optimization. These

transformations are used to move and merge the point clouds into a unique reference system (master camera in the first position of the sequence). The merging error, defined as the average Euclidean distance between corresponding points from four contiguous views, is  $0.26 \pm 0.18$  mm, that is only slightly larger than the reconstruction error. From the highly dense and regular set of points measured with DIC, a triangular mesh was automatically built via the Delaunay tessellation algorithm. A plane was calculated for each triplet of points of the mesh, thus allowing to retrieve the distribution of the local normal vector over the whole reconstructed surface with the same spatial resolution of the DIC points grid reconstruction (about 2 mm spacing). Fig. 5 shows the surface patches 3D-reconstruction and merging, and the obtained tessellated mesh with superimposed the position of the hole centers calculated via triangulation.

### C. Approach to the hole

The above procedure allows to compute the positions,  $\mathbf{p}_{h_i}^{c^*}$  ( $i = 1, \dots, \beta$ ), of the hole centers,  $O_{h_i}$ , and their normal unit vector,  $\mathbf{a}_{h_i}^{c^*}$ , in the reference frame  $\Sigma_{c^*}$  of the master camera in the first position of the sequence. In order to perform the Peg-in-Hole task, it is necessary to transform such vectors in the inertial frame. Let us define the  $(4 \times 4)$  homogeneous transformation matrices  $\mathbf{A}_{p^*}$ , performing the transformation between the peg frame,  $\Sigma_{p^*}$ , in the first position of the sequence and the inertial frame, and  $\mathbf{A}_c^p$ , performing the transformation between the master camera frame,  $\Sigma_c$ , and the peg frame,  $\Sigma_p$ , obtained via the calibration method. Hence, the generic hole center position in the inertial frame can be computed as

$$\begin{bmatrix} \mathbf{p}_h \\ 1 \end{bmatrix} = \mathbf{A}_{p^*} \mathbf{A}_{c^*}^{p^*} \begin{bmatrix} \mathbf{p}_{h_i}^{c^*} \\ 1 \end{bmatrix}, \quad (4)$$

where the subscript  $i$  is omitted for brevity; it is worth remarking that, since  $\mathbf{A}_c^p$  is constant, it is  $\mathbf{A}_{c^*}^{p^*} = \mathbf{A}_c^p$ . Similarly, the vector normal to the workpiece surface in  $O_h$  can be transformed in the inertial frame as

$$\mathbf{a}_h = \mathbf{R}_{p^*} \mathbf{R}_c^p \mathbf{a}_{h_i}^{c^*}, \quad (5)$$

where  $\mathbf{R}_c^p$  is the rotation matrix extracted by the homogeneous transformation matrix  $\mathbf{A}_c^p$ .

In order to approach the hole, the robot motion is planned to move the origin of  $\Sigma_p$  in the neighborhood of the workpiece surface and, at the same time, to align the axis  $\mathbf{a}_p$  to  $\mathbf{a}_h$ . To this aim, a closed-loop inverse kinematics algorithm with task priority [19] has been implemented, where two tasks are assigned: *peg alignment* (primary task with highest priority) and *position tracking* of  $O_p$  (secondary task with lower priority). The goal of the *peg alignment* task is to align the unit vector  $\mathbf{a}_p$  to  $\mathbf{a}_h$ , thus the task function is

$$\sigma_1 = (\mathbf{a}_h - \mathbf{a}_p)^T (\mathbf{a}_h - \mathbf{a}_p), \quad (6)$$

with corresponding Jacobian matrix

$$\mathbf{J}_1 = 2(\mathbf{a}_h - \mathbf{a}_p)^T \mathbf{S}(\mathbf{a}_p) \mathbf{J}_O(\mathbf{q}) \in \mathbb{R}^{1 \times n}, \quad (7)$$

where  $\mathbf{S}(\cdot)$  is the skew symmetric operator ([20], pp.106-107) and  $\mathbf{J}_O(\mathbf{q})$  is the orientation part of the Jacobian  $\mathbf{J}(\mathbf{q})$  defined in (3). The *position tracking* task is aimed at tracking a desired trajectory for the peg tip, from its current position to a target point  $P_1$ , whose coordinate in the reference frame  $\Sigma_h$  are  $\{0, 0, \Delta_1\}$ , i.e., a point belonging to the axis  $\mathbf{a}_h$  at a distance  $\Delta_1$  from  $O_h$ . The task function is the position of

$O_p$ ,  $\mathbf{p}_p$ , and the task Jacobian,  $\mathbf{J}_2$ , is the positional part of the Jacobian  $\mathbf{J}(\mathbf{q})$  defined in (3). In order to compute the commanded joint velocities, a Null-Space Behavioral control [19] is devised, in which the velocities of the secondary task are projected onto the null space of the primary task Jacobian, i.e.,

$$\dot{\mathbf{q}} = \mathbf{J}_1^\dagger (-k\sigma_1) + (\mathbf{I}_n - \mathbf{J}_1^\dagger \mathbf{J}_1) \mathbf{J}_2^\dagger (\dot{\mathbf{p}}_{p_d} + \mathbf{K}(\mathbf{p}_{p_d} - \mathbf{p}_p)), \quad (8)$$

where  $\mathbf{p}_{p_d}$  ( $\dot{\mathbf{p}}_{p_d}$ ) is the desired position (linear velocity) of  $O_p$ , while  $k$  and  $\mathbf{K} \in \mathbb{R}^{3 \times 3}$  are, respectively, a positive definite scalar and matrix gain. The matrix  $\mathbf{I}_n - \mathbf{J}_1^\dagger \mathbf{J}_1$  is a projector onto the null space of  $\mathbf{J}_1$ , with  $\mathbf{I}_n$  the  $n \times n$  identity matrix.

### D. Peg insertion

Once the approach phase has been completed, the desired pose  $\mathbf{x}_d = [\mathbf{p}_d^T \ \phi_d^T]^T$  for  $O_p$  can be computed. In particular, the orientation is kept constant while the desired position trajectory is a fifth degree polynomial from the approach position to a target point  $P_2$ , whose coordinate in the  $\Sigma_h$  are  $\{0, 0, -\Delta_2\}$ , where  $\Delta_2$  includes the height of the cylindrical part of the peg, called  $\delta$  in Fig.2, and a further translation along  $-\mathbf{a}_h$  that favors the insertion.

In order to limit the mechanical stresses and ensure compliance to the peg, an admittance control has been implemented at the peg tip level. It is assumed that the robot manipulator is not equipped with a wrist-mounted force/torque sensor, but, as usual in collaborative robots, with torque sensors at the joints. Thus, the wrench acting on the peg tip is estimated via an observer based on the generalized momentum

$$\boldsymbol{\nu} = \mathbf{M}(\mathbf{q}) \dot{\mathbf{q}}. \quad (9)$$

In detail, an estimate of the torques induced at the joints by the contact wrench can be computed as follows [21]

$$\hat{\boldsymbol{\tau}}_e = \mathbf{K}_o \left[ (\boldsymbol{\nu}(t) - \boldsymbol{\nu}(t_0)) + \int_{t_0}^t (\mathbf{C}^T(\mathbf{q}, \dot{\mathbf{q}}) \dot{\mathbf{q}} - \mathbf{F} \dot{\mathbf{q}} - \mathbf{g}(\mathbf{q}) + \boldsymbol{\tau} + \hat{\boldsymbol{\tau}}_e) ds \right], \quad (10)$$

where  $t$  and  $t_0$  are the current and initial time instant respectively,  $\mathbf{K}_o \in \mathbb{R}^n$  is a positive definite gain matrix. By reasonably assuming that  $\dot{\mathbf{q}}(t_0)$  is null, it implies that  $\boldsymbol{\nu}(t_0)$  is null as well. Then, the following dynamics for the estimation is obtained

$$\dot{\hat{\boldsymbol{\tau}}}_e + \mathbf{K}_o \hat{\boldsymbol{\tau}}_e = \mathbf{K}_o \boldsymbol{\tau}_e. \quad (11)$$

Equation (11) is a first-order low-pass dynamic system, therefore,  $\hat{\boldsymbol{\tau}}_e \rightarrow \boldsymbol{\tau}_e$  when  $t \rightarrow \infty$  for any positive definite gain matrix  $\mathbf{K}_o$ . Then, by considering the expression of  $\boldsymbol{\tau}_e$  in (3) an estimate of the external wrench is [22]

$$\hat{\mathbf{h}} = \mathbf{J}^{\dagger T}(\mathbf{q}) \hat{\boldsymbol{\tau}}_e. \quad (12)$$

Then, the reference pose for  $\Sigma_p$ ,  $\mathbf{x}_r = [\mathbf{p}_r^T \ \phi_r^T]^T$ , is computed via the admittance filter

$$\mathbf{M}_p \Delta \ddot{\mathbf{x}} + \mathbf{D}_p \Delta \dot{\mathbf{x}} + \mathbf{K}_p \Delta \mathbf{x} = \mathbf{T}_A^T(\phi_p) \hat{\mathbf{h}}, \quad (13)$$

where  $\mathbf{M}_p$ ,  $\mathbf{D}_p$  and  $\mathbf{K}_p$  are, respectively, the virtual inertia, damping and stiffness matrices imposed to the peg,

$\Delta\mathbf{x} = \mathbf{x}_d - \mathbf{x}_r$ , and

$$\mathbf{T}_A^T(\phi_p) = \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{O} & \mathbf{T}(\phi_p) \end{bmatrix}, \quad (14)$$

with  $\mathbf{T}(\phi_p)$  the matrix that maps the time derivative of the Euler angles  $\phi_p$ , representing the peg orientation, to the angular velocity ([20], pp.120-121). Finally, the joint velocities are computed as

$$\dot{\mathbf{q}} = [\mathbf{T}_A(\phi_p)\mathbf{J}]^\dagger (\dot{\mathbf{x}}_r + \Lambda(\mathbf{x}_r - \mathbf{x})), \quad (15)$$

where  $\Lambda \in \mathbb{R}^{6 \times 6}$  is a positive definite gain matrix.

#### IV. EXPERIMENTAL RESULTS

For the experiments, the robot has been commanded via the libfranka interface, running at 1 KHz, that provides joint positions, velocities and torques. The robot has been controlled in velocity-mode by sending the desired joint velocities output by the kinematic inversions (8) for the approach and (15) for the insertion. The dynamic parameters identified in [23] have been adopted to build in the wrench observer (10). Such parameters have been suitably modified in order to take into consideration the contribution to the inertia and gravity terms of the gripper and the peg.

In order to perform the 3D surface reconstruction, 4 pairs of images have been taken by the stereo infrared imagers of the Realsense camera via the librealsense2 library. Then, hole detection is performed on each image.

Table I reports the inverse kinematics gains as well as the observer and admittance filter parameters. It is worth noticing that the virtual stiffness  $\mathbf{K}_p$  has been tuned in such a way the peg is more stiff along the  $z$  axis and more compliant along the other axes in order to simplify the insertion in the hole. For practical implementation of the control scheme, the following adjustments have been made:

- The estimated contact wrench  $\hat{\mathbf{h}}$  output by the wrench observer (10) is filtered with a digital low-pass filter before sending to the admittance filter.
- To suppress non-existent small force and torque estimations owing to unmodeled dynamics and sensor noise, a *dead zone* on the admittance filter input has been implemented. Any value of force component below 3 N and any value of moment below 1 Nm estimated by the observer were neglected. Moreover, to achieve a continuous wrench signal, the same thresholds have been subtracted from higher estimations. This implies that the admittance filter receives as input a wrench slightly lower than the real interaction wrench.

Fig. 6 reports some snapshots taken during the experiment. Fig. 6(a) shows the random selected robot initial pose; in

TABLE I  
CONTROLLER AND OBSERVER GAINS.

| Gain                    | Value                               |
|-------------------------|-------------------------------------|
| $k$ eq. (8)             | 1                                   |
| $\mathbf{K}$ eq. (8)    | diag[150, 150, 150]                 |
| $\mathbf{K}_o$ eq. (10) | diag[10, 10, 10, 10, 15, 15, 15]    |
| $\mathbf{K}_p$ eq. (13) | diag[45, 45, 150, 0.15, 0.15, 0.15] |
| $\mathbf{D}_p$ eq. (13) | diag[500, 500, 1000, 25, 25, 25]    |
| $\mathbf{M}_p$ eq. (13) | diag[10, 10, 10, 0.5, 0.5, 0.5]     |
| $\Lambda$ eq. (15)      | diag[150, 150, 150, 20, 20, 20]     |

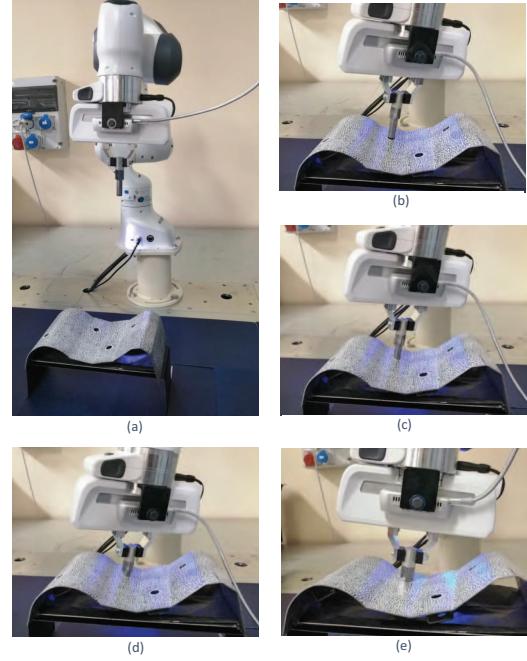


Fig. 6. Snapshots of the Peg-in-Hole assembly task: (a) robot initial pose; (b) approach phase; (c) beginning of the insertion phase; (d) insertion phase at the maximum contact wrench; (e) end of the insertion.

Fig. 6(b), the robot reaches the target point  $P_1$  at the end of the approach phase; in Figs. 6(c) and 6(d), the poses corresponding, respectively, to the beginning of the insertion phase (at about 7 seconds) and the maximum contact wrench are reported; finally, Fig. 6(e) shows the end of the insertion phase (at about 15 seconds).

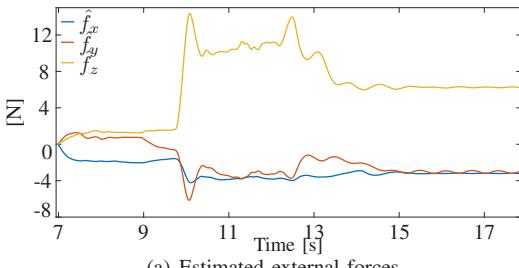
Fig. 7 reports the filtered external forces (Fig. 7(a)) and moments (Fig. 7(b)) estimated by the observer during the insertion. The maximum force is experienced along the  $z$  axis, coherently with the assigned virtual stiffness and the tilt of the considered hole, and thus, non-negligible moments around the  $x$  and  $y$  axis arise. Such forces and moments cause a deviation of the trajectories output by the admittance filter with respect to the desired ones (Fig. 8).

Finally, Fig. 9 shows the peg alignment task error: it can be noticed that the error converges to zero during the approach phase, then, during the insertion, when the task is not active, due to uncertainties on workpiece reconstruction and on camera calibration, the peg slightly modifies its tilt and the error grows up. The value of  $1.4 \cdot 10^{-4}$  at the end of the insertion represent the squared norm error between the actual hole tilt and the estimated one.

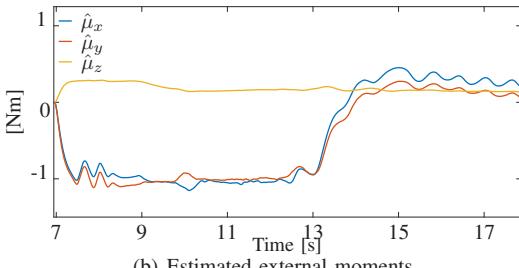
A video of the experiment can be found at <https://tinyurl.com/iros2020>

#### V. CONCLUSIONS AND FUTURE WORK

In this paper, an approach to achieve autonomous execution of robotic assembly tasks in partially structured environments is developed and experimentally demonstrated through a classical Peg-in-Hole task. The present work is only a preliminary approach to the problem and it can be further improved and complemented under several points of view. The performance of the 3D-DIC strongly depends on the quality of the pattern on the surface and the illumination. In practice, it is not always possible to ensure the presence

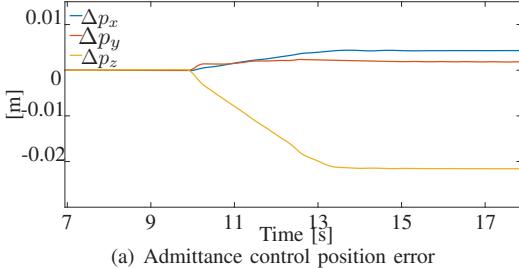


(a) Estimated external forces

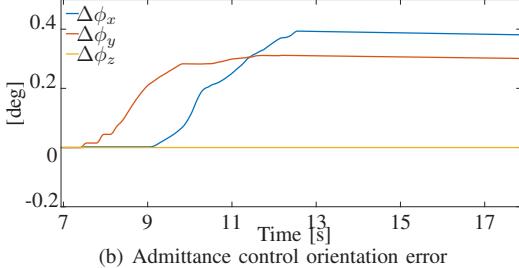


(b) Estimated external moments

Fig. 7. Estimated external wrench exerted by the environment on the robot.



(a) Admittance control position error



(b) Admittance control orientation error

Fig. 8. Error between the desired and reference pose: position (a) and orientation (b).

of a suitable pattern on the surface, but it is possible to use stereo-cameras with larger sensor resolution and specially designed projected patterns. As for the interaction control, future work will be devoted to test different formulations of the admittance control (e.g., by expressing the orientation errors via geometrically meaningful representations).

## REFERENCES

- [1] N. Hogan, "Impedance control: An approach to manipulation: Parts i-iii," *Journal of dynamic systems, measurement, and control*, vol. 107, no. 1, pp. 1–24, 1985.
- [2] Y.-L. Kim, B.-S. Kim, and J.-B. Song, "Hole detection algorithm for square peg-in-hole using force-based shape recognition," in *2012 IEEE International Conference on Automation Science and Engineering (CASE)*. IEEE, 2012, pp. 1074–1079.
- [3] H.-C. Song, Y.-L. Kim, and J.-B. Song, "Guidance algorithm for complex-shape peg-in-hole strategy based on geometrical information and force control," *Advanced Robotics*, vol. 30, no. 8, pp. 552–563, 2016.
- [4] F. Caccavale, C. Natale, B. Siciliano, and L. Villani, "Control of two industrial robots for parts mating," in *Proceedings of the 1998 IEEE International Conference on Control Applications (Cat. No. 98CH36104)*, vol. 1. IEEE, 1998, pp. 562–566.
- [5] S. Huang, K. Murakami, Y. Yamakawa, T. Senoo, and M. Ishikawa, "Fast peg-and-hole alignment using visual compliance," in *2013 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2013, pp. 286–292.
- [6] Z. Yang, W. Liu, H. Li, and Z. Li, "A coaxial vision assembly algorithm for un-centripetal holes on large-scale stereo workpiece using multiple-dof robot," in *2018 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE, 2018, pp. 1–6.
- [7] W. S. Newman, Y. Zhao, and Y.-H. Pao, "Interpretation of force and moment signals for compliant peg-in-hole assembly," in *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No. 01CH37164)*, vol. 1. IEEE, 2001, pp. 571–576.
- [8] H. Park, J. Park, D.-H. Lee, J.-H. Park, M.-H. Baeg, and J.-H. Bae, "Compliance-based robotic peg-in-hole assembly strategy without force feedback," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 8, pp. 6299–6309, 2017.
- [9] M. Jokesch, J. Suchý, A. Winkler, A. Fross, and U. Thomas, "Generic algorithm for peg-in-hole assembly tasks for pin alignments with impedance controlled robots," in *Robot 2015: Second Iberian Robotics Conference*. Springer, 2016, pp. 105–117.
- [10] J. C. Triyonoputro, W. Wan, and K. Harada, "Quickly inserting pegs into uncertain holes using multi-view images and deep network trained on synthetic data," *arXiv preprint arXiv:1902.09157*, 2019.
- [11] M. A. Lee, Y. Zhu, K. Srinivasan, P. Shah, S. Savarese, L. Fei-Fei, A. Garg, and J. Bohg, "Making sense of vision and touch: Self-supervised learning of multimodal representations for contact-rich tasks," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 8943–8950.
- [12] G. De Magistris, A. Munawar, T.-H. Pham, T. Inoue, P. Vinayavekhin, and R. Tachibana, "Experimental force-torque dataset for robot learning of multi-shape insertion," *arXiv preprint arXiv:1807.06749*, 2018.
- [13] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv*, 2018.
- [14] J. Redmon, "Darknet: Open source neural networks in c," <http://pjreddie.com/darknet/>, 2013–2016.
- [15] M. A. Sutton, J. J. Orteu, and H. Schreier, *Image correlation for shape, motion and deformation measurements: basic concepts, theory and applications*. Springer Science & Business Media, 2009.
- [16] M. Khrenov, H. Bruck, and S. Gupta, "A novel single camera robotic approach for three-dimensional digital image correlation with targetless extrinsic calibration and expanded view angles," *Experimental Techniques*, vol. 42, no. 6, pp. 563–574, 2018.
- [17] L. Villani and J. De Schutter, "Force control," in *Handbook of Robotics*, Springer-Verlag, B. Siciliano, and O. Khatib, Eds., 2008.
- [18] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.
- [19] F. Arrichiello, S. Chiaverini, G. Indiveri, and P. Pedone, "The null-space based behavioral control for mobile robots with velocity actuator saturations," *International Journal of Robotics Research*, vol. 29, no. 10, pp. 1317–1337, 2010.
- [20] B. Siciliano, L. Sciavicco, L. Villani, and G. Oriolo, *Robotics – Modelling, Planning and Control*. London, UK: Springer, 2009.
- [21] A. De Luca and R. Mattone, "Sensorless robot collision detection and hybrid force/motion control," in *2005 IEEE Int.Conf. on Robotics and Automation*, 2005, pp. 999–1004.
- [22] F. Ficuciello, A. Romano, L. Villani, and B. Siciliano, "Cartesian impedance control of redundant manipulators for human-robot co-manipulation," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 2120–2125.
- [23] C. Gaz, M. Cognetti, A. Oliva, P. R. Giordano, and A. De Luca, "Dynamic identification of the franka emika panda robot with retrieval of feasible parameters using penalty-based optimization," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4147–4154, 2019.

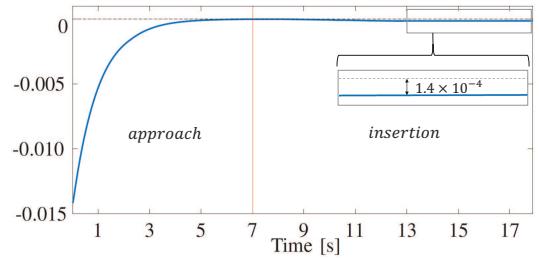


Fig. 9. Alignment peg task error

- 4240