Lanzetta, M.; Dini, G.: *An Integrated Vision-Force System for Peg-in-hole Assembly Operations*
ICME '98, Proceedings of the CIRP International Seminar on «Intelligent Computation in Manufacturing Engineering», Ed.R.Teti, Capri (Naples) Italy, July 1st-3rd, 1998, p.615-621

AN INTEGRATED VISION-FORCE SYSTEM FOR PEG-IN-HOLE ASSEMBLY OPERATIONS

M. Lanzetta, G. Dini

Department of Production Engineering, University of Pisa
Via Bonanno Pisano, 25/B, 56126 Pisa, Italy

Abstract: In this article the integration of an artificial vision system and a 6-component force-torque sensor for high accuracy insertions is proposed. The former has the advantage of intervening before the contact of the two parts and providing an initial correction to allow the force sensor action; the latter is charged of completing the operation through the control of a neural network.
An experimental facility has been developed for performance assessment. Two monochromatic CCD cameras are positioned with the optical axes perpendicular to each other, outside the working area in order not to interfere with the assembly operations.
During tests it has been shown that the artificial vision system yields a fast, reliable and accurate measurement, by direct comparison with a co-ordinate-measuring machine. Results show that the use of an integrated vision-force sensor represents a complete solution to the peg-in-hole problem. Furthermore the 3-D localisation algorithm implemented on the vision system and the problems concerned with the system engineering are dealt with.

Keywords: automated assembly, peg-in-hole, vision and force sensor, artificial neural networks

1. Introduction

In automated assembly some operations such as part mating, screwing, pressing, etc., may need a high accuracy and consequently a constant monitoring of one or more process parameters. One of the most critical assembly operations, under this viewpoint, is represented by the so-called *peg-in-hole* problem (Sfantsikopoulos *et al.*, 1994), that is the assembly of two parts with a tight tolerance, involving about 35% of automated assembly operations. This problem is due to the uncertainty of the relative position of the peg and the hole caused by positioning errors of the robot or the pallet, grasping errors of the gripper, etc. The approaches to solve this problem are twofold:

1. Improving the robot and fixture accuracy. These solutions are usually expensive and are not always possible.
2. Passive or active compliance techniques are successfully used, but several problems are still unsolved.

As far as active compliance techniques are concerned, several kinds of sensors have been used to monitor and control the operation, such as force sensors (Little, 1992), tactile sensors (Iversen, 1993), optical sensors (Janocha and Menzel, 1989; Schweigert, 1992), and vision systems (Iversen, 1994; Snyder, 1996). In particular force sensors appear as the most efficient solution due to their ability to recognise jamming phenomena and to supply reliable data to correct the robot arm position.

However, the use of a single sensor might not be sufficient mainly for two reasons: the sensor does not provide the necessary accuracy or it has some intrinsic limitations that prevent a complete control over the whole process. In particular, in peg-in-hole operations with high initial positioning errors, a force sensor alone is not able to provide a correct output if the contact between the two parts occurs outside the hole. For this task, a vision system or an optical sensor can be used. Conversely, these latter are not able to retrieve information after the initial insertion because the peg and the hole surface and edges are not visible.

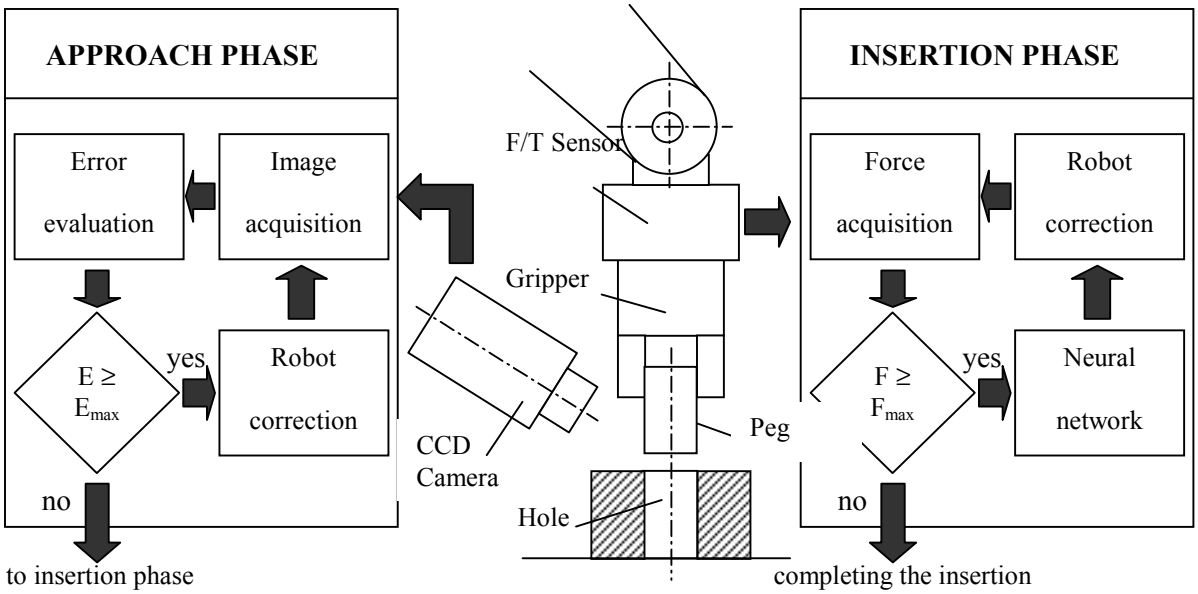

Fig. 1 Schematic view of the vision-force assembly system

The co-operation of two sensors (Janocha and Menzel, 1989; Fujita, 1988) like a vision and a force sensor is investigated in this paper. The former has the advantage of intervening before the contact of the two parts and providing an initial correction to allow the force sensor action; the latter is charged of completing the operation. The main purpose of this work is twofold: to contribute in the development of multi-sensory devices for assembly by developing a particular system in which the sensor integration can be exploited both during the *on-line* insertion *phase* and in the system *training phase* (Dini, 1997, 1998).

2. The vision-force system integration

In Fig. 1, the proposed system configuration is displayed with the flow-chart adopted in the on-line phase. The insertion task is performed in two steps:

1. *The approach phase*: when the peg is outside the hole a stereo vision system estimates the position and inclination error; if this error is greater than the addition of the peg and hole chamfer width, the vision system provides a first correction to the robot arm; otherwise the insertion phase is directly executed.
2. *The insertion phase*: when the peg and the hole surfaces are in contact, a 6-component F/T sensor mounted on the wrist and controlled by a neural network is adopted. A 3-layer feed-forward network, trained with a back-propagation technique has been used. The necessary correction is generated in function of the measured force components; the implementation details can be found in (Dini 1997, 1998).

A major benefit coming from the integration is that the vision system output can be exploited also for the training of the neural network. It is well known that a critical aspect in the use of neural networks is

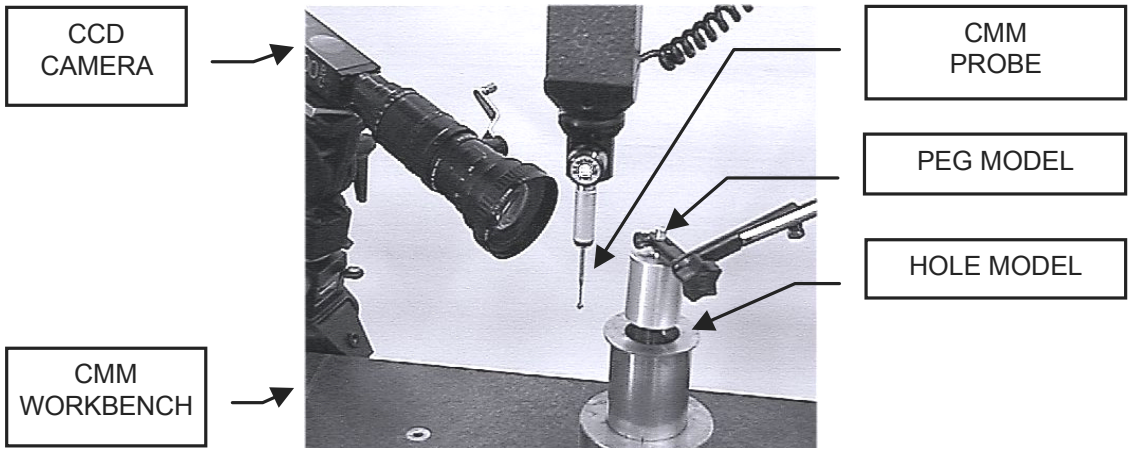

Fig. 2 Detail of the experimental facility

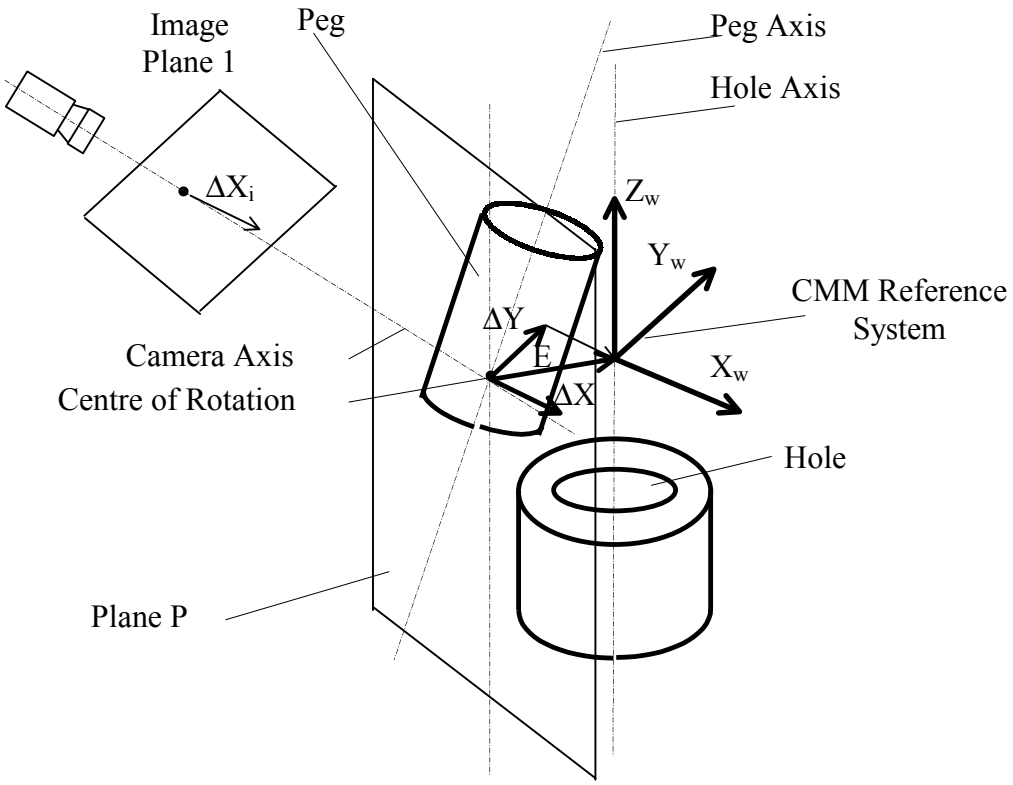

Fig. 3 Vision system configuration and definitions

the collection of samples, represented in this case by the force components and the relative peg-hole positions. The collection of position data is crucial as it can be affected by the repeatability error of the robot arm and must be estimated more accurately. In the proposed approach, this task is carried out by the vision system.

In the remainder of this article, the vision system configuration and algorithms will be described.

3. Experimental set-up

For image processing and acquisition, the system developed at the Department of Production Engineering has been used. It consists of a commercial acquisition and processing card with a user-friendly interface. The developed program runs on a PC. The system has a standard CCIR resolution (756×567 pixels).

The configuration consists of a stereo vision system with two CCD cameras approximately perpendicular to each other. Considering the problem symmetry, in Fig. 3 only one camera is displayed for clarity reasons. It is not necessary that the two cameras make an angle of exactly 90° (which is very difficult to obtain in practical applications), as the peg position error E is calculated by a vector addition of the two components ΔX e ΔY acquired respectively by the camera 1 (Fig. 3) and the camera 2 (not displayed). The camera axis is tilted with respect to the peg and hole axes to allow a contemporary view of the peg and the hole. With such configuration, in this simple case examined, the edges appear respectively as straight lines and ellipses (Fig. 5.b).

To assess the vision system performance, a more accurate measuring system has been used, which is addressed to as the absolute or *world* reference since all other measurements (vision system, robot, pallet, etc.) are referred to it. In the described application the output of the vision system has been compared with the measurements obtained by a CMM (Fig. 2), which has an accuracy of less than 4 μm.

The CMM reference system (X_w, Y_w, Z_w) is positioned in the correct final peg position with the Z_w axis coincident with the hole axis, with no rotation constraints about this axis, and with the origin coincident with the vertical co-ordinate of the peg «centre of rotation», (thus $\Delta Z = \Delta Z_w = 0$). In the most general case $\Delta X \neq \Delta X_w$ and $\Delta Y \neq \Delta Y_w$ but, of course,

$$\Delta X \oplus \Delta Y = \Delta X_w \oplus \Delta Y_w = E$$

The centre of rotation is considered to estimate the peg position as the intersection of the peg axis with a plane parallel to the hole plane. This point is assumed as the centre of rotation for the relative error correction and can be anywhere inside or outside the peg. The most suitable position is close to the bottom face so that rotation corrections do not increase the peg-hole relative distance.

Plane P is displayed in Fig. 3 for construction and clarity. It is parallel to the hole axis and contains the centre of rotation and the vector ΔX, which is parallel to the camera image plane 1 and to its projection ΔX_i

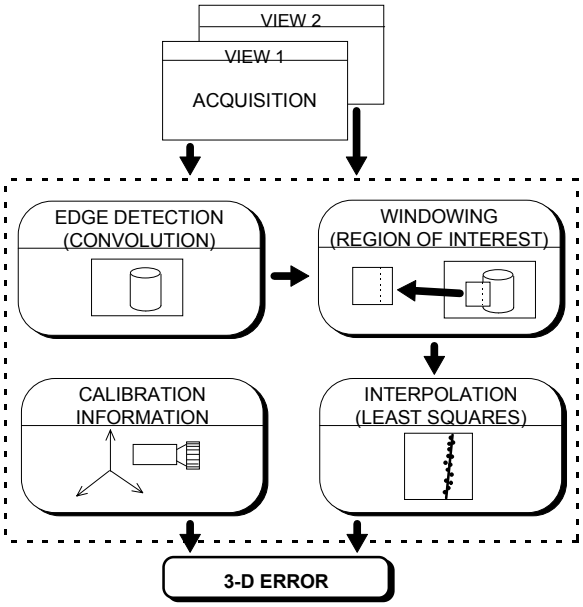

Fig. 4 Flow-chart of the vision algorithm

(magnified). Concerning the camera 2, ΔY and ΔY_i (not displayed), constructions are the same.

An experimental facility (Fig. 2) has been developed including the following elements: a peg with a support and a hole. The peg support is placed on an X-Y table. The peg inclination can be also changed.

Dealing with monochromatic images, a fluorescent lamp, which is widespread in industrial environment for its low cost, has been used.

4. The algorithm

The algorithm developed for the peg and the hole localisation is summarised in Fig. 4. The essential features of this approach are:

- Analysis of a couple of images in parallel. The information retrieved from each image is processed independently;
- 2-D technique application. In the present case the displacement in the camera axis direction are negligible, therefore the hypotheses of «weak perspective» can be assumed (Lanzetta, 1996);
- Interpolation. The position accuracy is increased finding the best fit on geometric primitives of detected points belonging to edges.

Assuming a Gaussian distribution, the magnitude of the benefit coming from interpolation is approximately $1/\sqrt{N}$, where N is the number of samples, that is the number of pixels belonging to edges.

In Fig. 5.a, a sample image is displayed. The signal magnitude (profile), along a horizontal row (Fig. 5.c) shows the highest gradient (in the selected circle) corresponding to the peg edge position, which can be easily detected by filtering the image (Baxes, 1994). Considering the small orientation changes, the following 3×3 kernel has been used for the convolution:

$$\text{vert\_ edge} = \begin{pmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{pmatrix}$$

The negative of the resulting image is displayed in (Fig. 5.b).

To achieve the best algorithm speed, a windowing operation can be performed.

Dealing with straight lines, in the examined case the image centre has been used, in order to reduce the effect of optical distortions.

The points belonging to the peg and hole edges (Fig. 5.b) are interpolated according to the least squares and the position error ΔX_i in image co-ordinates is estimated. This method can be extended to parts of any shape, by separating their edge into simpler geometric primitives, such as straight lines and arcs of ellipse and by selecting, within the image, appropriate reference lines belonging to the peg or to the hole edge. For instance, during tests the peg error has been considered only and of course, the same method can be applied to calculate separately the hole error too. The selection of segments or arcs belonging to the reference contour is a critical operation and the algorithm reliability and accuracy highly depends on it.

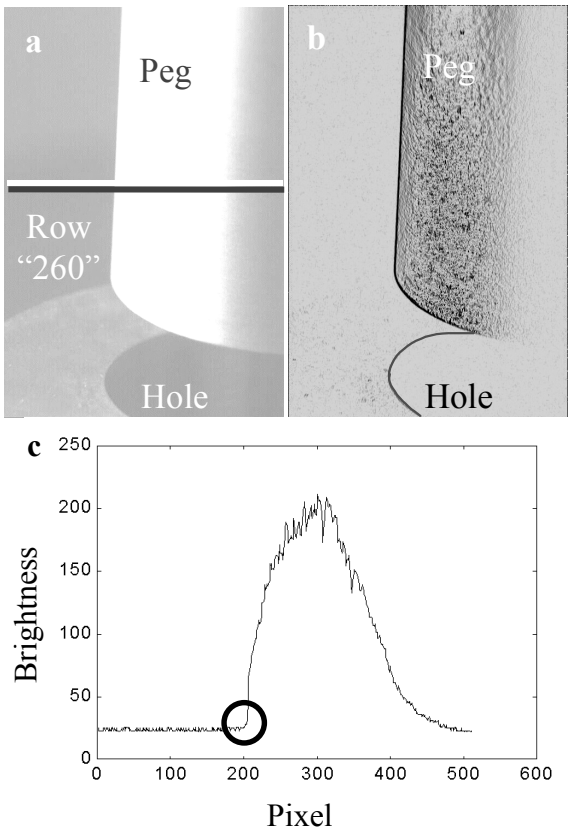

Fig. 5.a Original image; b filtered image (negative); c horizontal «profile» (row "260")

For 3-D position recovery, the image co-ordinates are converted into space co-ordinates using calibration data, by finding the correspondence $\Delta X \Leftrightarrow \Delta X_i$ and $\Delta Y \Leftrightarrow \Delta Y_i$.

5. Experimental tests

The main purpose of experimental tests is to assess the vision system accuracy, which depends on the theoretical spatial resolution; with the current configuration it is

$$\text{spatial\_res} = \text{peg\_width}/\text{image\_width} \cong 40/567 \cong 70 \ \mu\text{m/pixel}$$

This theoretical value can be affected by the following error sources:

1. *the system repeatability*, which can also be reduced with multiple acquisitions for Gaussian error distributions as described in chapter 4.;
2. *the influence between the two components*, due to perspective. The position detected by each camera is affected by an error that significantly depends on the camera-object distance and, for the effect of perspective, it is inversely proportional to that distance: the higher the distance, the lower the error. Of course, to increase the distance a suitable and more expensive camera objective should be used. With a suitable system configuration it is always possible to reduce this effect and fulfil the «weak perspective» hypothesis (Lanzetta, 1996). The magnitude of the perspective effect, expressed as the ratio between the maximum peg positioning error versus the camera-peg distance should be smaller than the resolution.

Considering the system symmetry only one camera has been used in experiments. The 3-D peg (hole) position is obtained with a vector addition of the two components detected by each camera (Fig. 3).

To evaluate the system performance, two different tests have been carried out with the following purposes: (1) to detect the peg and hole position and (2) their inclination.

In the examined case, a set of 72 points has been collected in a *work area* of ±2 mm along each axis, around the correct position for insertion, with steps of about 10 μm (steps decrease approaching the correct peg position).

Sample acquisition is carried out through the following steps:

1. the hole position is measured by the CMM to find the Z_w axis of the reference system;
2. the peg is put in the correct position for insertion; the centre of rotation (displayed in Fig. 3) is chosen using a conventional plane perpendicular to the Z_w axis (and hence parallel to the hole axis);
3. the selected relative errors are provided through micrometer adjustments of the peg support;
4. the actual peg position is measured through the CMM in the (X_w, Y_w, Z_w) reference system; the difference between the actual and the initial hole position represents the actual error;
5. three images are acquired and the resulting position and inclinations according to the vision system are calculated;
6. the procedure loops starting from step three for other acquisitions.

6. Results

In Fig. 6 a graph is displayed that summarises the main experimental results. To better understand the use of this graph, two sample positions are displayed: A and B. On the main axes the following values can be read: the world co-ordinates (X_A, Y_A) and (X_B, Y_B); their relative distance (ΔX, ΔY) and the corresponding measured distance in pixel ΔX_i.

At each position, the two co-ordinates lay on a vertical straight line.

Only a correction to the X co-ordinate can be provided from this graph, since the Y co-ordinates are not univocal, as explained later.

Observing the peg position in the sample set displayed in Fig. 6, the following considerations can be deduced on the vision system precision:

1. *Problem linearity*: the problem has shown a strong linearity both for ΔX_w and ΔY_w (Fig. 6) which derives from the hypotheses of «weak perspective». In this sense, all over the work area the agreement between image co-ordinates and peg position is obtained through a calibration factor as described later, therefore the vision system accuracy is dominated by the subsequent effects.
2. *Direct proportionality*: with a rotation about the Z_w axis of the CMM (absolute) reference system to have the X_w axis parallel to the image plane of camera 1 (Fig. 3) a direct proportionality between the peg position in pixels and its error ΔX_w is obtained. On the contrary, the error ΔY_w is not univocal from view 1, as shown in Fig. 6, where the Y_w co-ordinates do not lay on a single straight line like the X_w co-ordinates.
3. *Vision system accuracy*: the vision system accuracy can be expressed as the maximum error (distance) of samples from the interpolated minimum square straight line. Of course, the system accuracy is configuration- and device-dependent. In the used system, it is better than 40 μm, that is less than 10^{-5} % of the peg diameter.

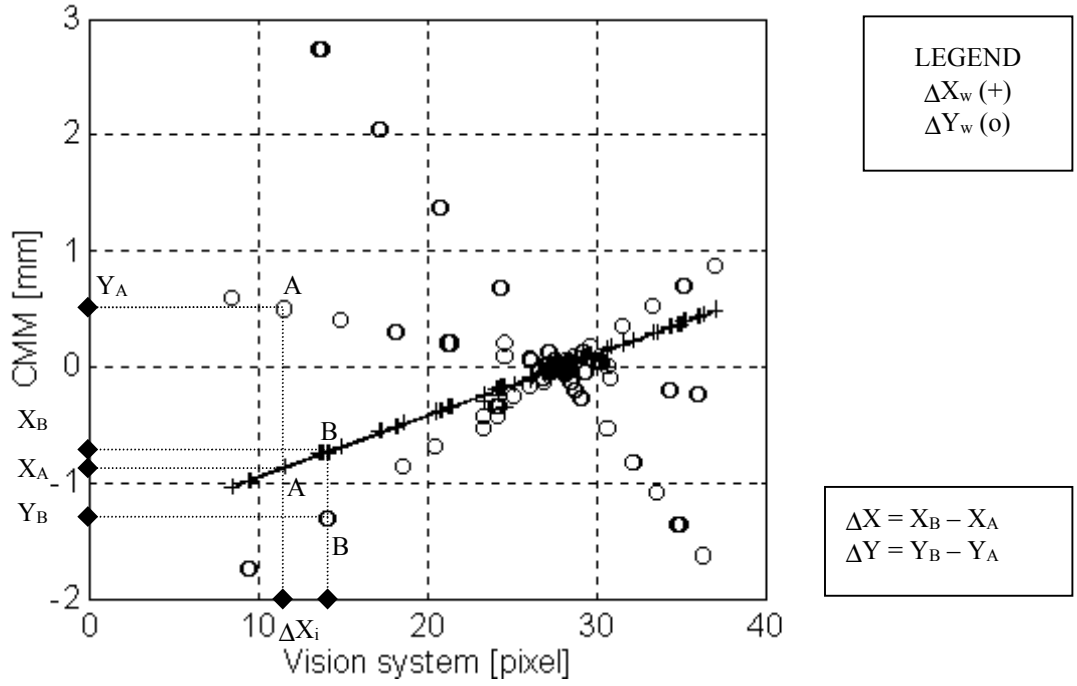

Fig. 6 Experimental results (72 samples)

4. *Vision system repeatability*: during tests, subpixel accuracy has been achieved as already observed in chapter 5. The vision system repeatability is about 0.05 pixels. This implies that the system accuracy can be further increased by multiple acquisition, three at each position in this particular case.
5. *Components decoupling* since the problem is three-dimensional, there is an error due to the perspective effect. The error detection is, of course, more accurate for displacements that are parallel to the image (CCD sensor) plane, in a sense according to how "weak" is the perspective. If this error is not negligible for the system configuration, correction should be performed in several steps, in order to achieve a better positioning and get a better error estimation. The correction accuracy increases for lower corrections. In particular, a correction to the greatest between ΔX_w and ΔY_w should be provided first.
6. *Vision system calibration*: considering the problem linearity, the vision system calibration is reduced to the calculation of a conversion parameter yielding the best interpolation of experimental data. In the obtained graph (Fig. 6), the measured peg error (in pixels) corresponds to the necessary correction (in mm).

The correction to be provided in the first steps of the approach phase is affected by an error of higher magnitude. It has been shown that the force system described in (Dini 1997, 1998) allows insertion in two steps in the 70 – 90 % of cases.

Regarding the peg angle estimation, a repeatability and precision test has been performed through the acquisition of a sample of 194 peg positions uniformly distributed within the work area at constant inclination. The main results of this analysis can be summarised as follows:

- considering the triplets of acquisitions at the same position (see chapter 5.), the peg inclination measured by the vision system is affected by a repeatability error of about 2-3°;
- the system precision is affected by an average error of about 12° (6.5° - 16.2°). This is because the edge position is more sensible to the angle of view at different positions and therefore the observed peg generatrix changes too.

These errors are obviously excessive and for this reason an angular correction made by the vision system is not feasible. Nevertheless, this correction is not critical, since only the contact between the two parts is the main goal of the approach phase. In addition, it should be considered that only one view of the peg has been used and that error can significantly be increased by estimating its inclination from two views.

The peg inclination is crucial for insertion in order to avoid jamming and it depends on the peg and hole tolerance and dimensions. However, beside accuracy, vision systems have an intrinsic limitation that prevents from using them to complete insertion. Let us consider, for instance, the case of misalignment between peg and hole caused by incorrect hole execution. In these situations, the vision system is unable to provide a correction and the insertion cannot be completed even if the peg inclination is correct. These situations can be neatly solved by the force sensor (Dini, 1997, 1998).

7. System engineering

No particular problems have been found during experimental tests, hence the system engineering requires the usual care to be paid in the development of a vision application for an industrial environment. For the application of the described algorithm high magnification lenses should be adopted in order to achieve the required accuracy according to the «weak perspective» models. The vision system requires a minimum of two cameras in the configuration of Fig. 3, with the optical axes approximately perpendicular to each other and tilted. According to budget and accuracy constraints, multiple cameras focusing on the different parts being assembled can be adopted.

In experimental tests, a CMM has been used to assess the system accuracy and for calibration purposes as well, however in practical application a different absolute measuring system is needed. The system calibration can be performed with standard techniques (e. g. using physical models whose shape and dimension is known *a priori*).

The system linearity could be exploited in order to achieve a better accuracy even with a less accurate measuring system, for instance the robot itself: this aspect could be a subject for further investigation.

Lighting is not critical, apart from avoiding reflections on the curved metal surface, by using a diffuse lighting. A high power light source is to be preferred in order to increase the «signal to noise» ratio and to reduce the effect of external lighting by closing the camera shutter, for more general applications. The presence of a background different from a black one could be eliminated by image subtraction.

The main limitations of the developed system are due to optical aberration and out-of-focus problems that can be easily solved by a proper hardware selection or by software means.

With special purpose hardware, such as the described system, all the processing can be performed in less than 80 ms, which is compatible with industrial needs, even in the case multiple acquisitions are required.

8. Conclusions

A vision-force approach to the peg-in-hole problem has been examined. In the described system, the relative peg-hole position and inclination can be detected by a vision system to provide an initial correction and in order to have the insertion operation completed with the aid of a force sensor. The complementary intrinsic limitations of both systems represent a good reason for the integration. In addition, the co-operation of the two systems can be exploited in the training phase as well.

The accuracy of the developed vision system is smaller than 40 μm for the peg position and about 12° for the peg inclination. The inclination correction is not critical as in the approach phase, only the contact between the two parts is required. A higher accuracy could be achieved by increasing the sensor resolution or the lens magnification but it has no practical interest in the examined application.

References

Baxes, G. A. (1994). *Digital Image Processing. Principles and Applications*, Wiley & Sons, New York.

Dini, G. (1997). Active compliance in high precision assembly by neural networks, *Proc. of 3rd Aitem Conf.*, Salerno, Sept. 17-19, pp. 259-266.

Dini, G. (1998). Application of Neural Networks in Robotized Peg-in-Hole Insertions Using a Pre-Process Learning Method, accepted for the publication on *Int. Jou. for Manufacturing Science and Production*, Freund Publishing.

Fujita, N. (1988). Assembly of Blocks by Autonomous Assembly Robot with Intelligence (ARI), *Annals of the CIRP*, Vol. 37, No. 1, pp. 33-36.

Iversen, W.R. (1993). Tactile Sensing, 1990s Style, *Assembly*, February-March, pp. 23-26.

Iversen, W.R. (1994). Focus On Machine Vision, *Assembly*, April, pp. 16-19.

Janocha, H., Menzel, E. (1989). Assembly of PIH-Devices with Robots, *Annals of the CIRP*, Vol. 38, No. 1, pp. 29-32.

Lanzetta, M. (1996) Stereo Vision with Neural Network (in Italian), Automazione e Strumentazione, Year XLIV, No. 7, July/August, pp. 107-116.

Little, R. (1992). Force/torque sensing in robotic manufacturing, *Sensors*, Nov., Vol. 9, No. 12.

Schweigert, U. (1992). Sensor-guided assembly, *Sensor review*, Vol. 12, No. 4, pp. 23-27.

Sfantsikopoulos, M.M., Diplaris, S.C., Papazoglou, P.N. (1994). An Accuracy Analysis of the Peg-and-hole Assembly Problem, *Int. Jou. Mach. Tools Manufact.*, Vol. 34, No. 5, pp. 617-623.

Snyder, J. (1996). Seeing the Light, *Assembly*, August, pp. 28-32.