Research article

Machine vision approach for robotic assembly

Mario Peña-Cabrera
Institute of Research on Applied Mathematics and Systems (IIMAS-UNAM), Mexico City, Mexico, and

Ismael Lopez-Juarez, Reyes Rios-Cabrera and Jorge Corona-Castuera
CIATEQ A.C., Advanced Technology Centre, Queretaro, Mexico

Abstract

Purpose – Outcome with a novel methodology for online recognition and classification of pieces in robotic assembly tasks and its application into an intelligent manufacturing cell.

Design/methodology/approach – The performance of industrial robots working in unstructured environments can be improved using visual perception and learning techniques. The object recognition is accomplished using an artificial neural network (ANN) architecture which receives a descriptive vector called CFD&POSE as the input. Experimental results were done within a manufacturing cell and assembly parts.

Findings – Find this vector represents an innovative methodology for classification and identification of pieces in robotic tasks, obtaining fast recognition and pose estimation information in real time. The vector compresses 3D object data from assembly parts and it is invariant to scale, rotation and orientation, and it also supports a wide range of illumination levels.

Research limitations/implications – Provides vision guidance in assembly tasks, current work addresses the use of ANN's for assembly and object recognition separately, future work is oriented to use the same neural controller for all different sensorial modes.

Practical implications – Intelligent manufacturing cells developed with multimodal sensor capabilities, might use this methodology for future industrial applications including robotics fixtureless assembly. The approach in combination with the fast learning capability of ART networks indicates the suitability for industrial robot applications as it is demonstrated through experimental results.

Originality/value – This paper introduces a novel method which uses collections of 2D images to obtain a very fast feature data – "current frame descriptor vector" – of an object by using image projections and canonical forms geometry grouping for invariant object recognition.

Keywords Robotics, Assembly, Neural nets

Paper type Research paper

1. Introduction

Robotics field has grown considerably with new technologies, industrial robots today, needs sensorial capabilities to achieve non-structured and more sophisticated tasks; vision systems as a sensorial mode for robots have a growing demand requiring more complex and faster image processing functions in order to implement more sophisticated industrial applications, like assembly automation.

In this sense, vision recognition systems must be capable of perceiving and detecting images and objects, as close as the human vision does; this fact has encouraged research activity to design artificial vision systems based on the neural morphology of the biological human vision system. Now scientists understand better about how computational neural structures and artificial vision systems must be designed following neural paradigms, mathematical models and computational architectures. When a system involves these aspects, it can be referred to as a "Neuro-Vision System" (Gupta and Knopf, 1993; Peña-Cabrera *et al.*, 2004) which can be defined as an artificial machine with ability to see our environment and provide visual formatted information for real time applications.

This document, reports research progress of a novel neural-based architecture method for invariant object recognition and applied to self-adapting industrial robots which can perform assembly tasks. It has been shown by psychological and clinical studies that visual object recognition involves a large activity area on the cerebral cortex when objects are seen the first time and the region's activity is reduced when familiar objects are perceived (Gupta and Knopf, 1993) New objects can also be learned quickly if certain clues are given to the learner. Following this psychological evidence a novel architecture called SIRIO (*Sistema Inteligente de Reconocimiento Invariante de Objetos*) was designed. The architecture is firstly trained with clues representing different objects that the robot is likely to encounter within the working space to form its initial knowledge base. This information then triggers the online learning subsystem based on an artificial neural network (ANN), the new image vector descriptors override initial clues, and the robot learns to identify familiar objects and to learn new ones.

The above ideas suggested that it was possible to get fast and reliable information from a simple but focused analysis of what an object might show. The very important aspects of the

The Emerald Research Register for this journal is available at
www.emeraldinsight.com/researchregister

The current issue and full text archive of this journal is available at
www.emeraldinsight.com/0144-5154.htm

Assembly Automation
25/3 (2005) 204–216
© Emerald Group Publishing Limited [ISSN 0144-5154]
[DOI 10.1108/01445150510610926]

scene (we have called "clues"), can be used later to retrieve memorized aspects of the object without having to recall detailed features. By using neural networks it is possible to learn manipulative skills, which can be used by an industrial manipulator (Lopez-Juarez and Howarth, 2000). In someway we humans do that process once an object has been seen and learned for the first time.

The article describes a methodology for online object recognition, based on ANN for identification and classification purposes. A robust algorithm for perimeter and centroid calculations, object functions and pose estimation is presented.

2. Background and related work and considerations

2.1 Related work

Intelligent manufacturing cells using robots with sensorial capabilities are being investigated using artificial intelligence techniques like ANN and Fuzzy Logic among others, since the mathematical and control models are simplified.

Acquiring information from multiple sensors in manufacturing systems provides robustness and self-adaptation capabilities, hence improving the performance in industrial robot applications. A few researchers have applied neural networks to assembly operations with manipulators and force feedback. Vijaykumar *et al.* (1994) used back propagation (BP) and reinforcement learning (RL) to control a zebra robot, its neural controller was based on the location error reduction beginning from a known location, Enric and del pobil (1997) employed self-organization map (SOM) and RL to control a zebra robot, the location of the destination piece was unknown, Howarth (1998) utilized BP and RL to control a SCARA robot, without knowing the location of assembly, Lopez-Juarez (2000) implemented FuzzyARTMAP to control a PUMA robot also with an unknown location. All of the above authors considered only constraint motion control during assembly; however, to complete the autonomy of the assembly system a machine vision system has also to be considered. Additionally, a new concept was introduced by Hoska (1988) called "Robotic Fixtureless Assembly" (RFA) that eliminates the need of using complex and rigid fixtures, which involves new technical challenges, but allows very potential solutions. Ngyuen and Mills (1996) have studied RFA of flexible parts with a dynamic model of two robots with a proposed algorithm, which does not require measurements of the part deflections. Plut and Bone (1996) and Plut and Bone (1997) presented a grasp planning strategy for RFA. The goal of RFA is to replace these fixtures with sensor-guided robots, which can work within RFA workcells. The development of such vision-guided robots equipped with programmable grippers might permit holding a wide range of part shapes without tool changing. Using ANN, an integrated intelligent vision-guided system can be achieved as it is shown by Langley and D'Eleuterio (2003). This job can be achieved by using two-dimensional (2D) computer vision in different manner so that 3D invariant object recognition and POSE calculation might be used for aligning parts in assembly tasks if an –"adequate descriptor vector" – is used and interfaced in real time to a robot. Many authors had come with descriptor vectors and image transformations, used as general methods for computer vision applications in order to extract invariant features from shapes. Aguado *et. al* (2002) developed a new formulation and methodology for including invariance in general form of the Hough transform, Chin-Hsiung *et al.* (2001) designed a technique for computing shape moments based on the quadtree representation of images, Best and McKay (1992) describe a method for registration of 3D shapes in minutes, Torralba and Oliva, 2002 present a method to infer the scale of the scene by recognizing properties of the scene structure for depth estimation, Freeman (1961) introduced the first approach for representing digital curves using chain codes, and showing classical methods for processing chains in Freeman (1974), Bribiesca (1999) developed a new chain code for shapes composed of regular cells, which has recently evolved even to represent 3D paths and knots.

Some authors use multiple cameras or multiple views to extract information, performs invariant object recognition and determine object's position and motion, Underwood *et al.*(1975) developed a visual learning system using multiple views which requires deterministic description of the object's surfaces like measurements and interconnections, Yong-Sheng Chen *et al.* (2001) propose a method to estimate the 3D ego-motion of an observer moving in a static environment, Murase and Nayar (1995) have worked in visual learning and recognition of 3D objects from appearance, Gonzalez-Galvan *et al.* (1997) developed a procedure for precision measure in 3D rigid-body positioning using camera-space manipulation for assembly. Dickmanns (1998) and Kollnig and Nagel (1997) have shown solutions to facilitate the use of vision for real world-interaction, Hager *et al.* (1995) and Papanikolopoulous and Khosla (1993) use markers on the object to simplify detection and tracking of cues.

Some other authors have contributed with techniques for invariant pattern classification, like classical methods as the universal axis of Lin, 1996 and invariant moments of Hu (1962) or artificial intelligence techniques, as used by Yüceer and Oflazer (1993) which describes an hybrid pattern classification system based on a pattern pre-processor and an ANN invariant to rotation, scaling and translation, Stavros and Lisboa (1992) developed a method to reduce and control the number of weights of a third order network using moment classifiers and You and Ford (1994) proposed a network for invariant object recognition of objects in binary images. Applications of guided vision used for assembly are well illustrated by Bone and Capson, 2003 which developed a vision-guide fixtureless assembly system using a 2D computer vision for robust grasping and a 3D computer vision to align parts prior to mating, and (Jörg *et al.*, 2000) designing a flexible robot-assembly system using a multi-sensory approach and force feedback in the assembly of moving components.

2.2 Original work

Moment invariants are the most popular descriptors for image regions and boundary segments, but computation of moments of a 2D image involves a significant amount of multiplications and additions in a direct method, fast algorithms have been proposed for these calculations, as it is the case of Philips (1993) for binary images. The computation of moments can be simplified since it contains only the information about the shape of the image as proposed by Chen (1990). In many real-time industry applications the speed of computation is very important, the 2D moment computation is intensive and involves parallel processing, which can become the bottleneck of the system when moments are used as major features. This paper introduces a novel method which uses collections of 2D images to obtain a very fast feature data – "current frame descriptor vector" – of an object by using image projections and canonical forms geometry grouping for invariant object recognition, producing 3D POSE information for different pre-defined assembly parts. A fast algorithm allows calculation of a boundary object function and centroid which defines and compress 3D object information, the algorithm uses a Weight Matrix Transformation introduced by Peña-Cabrera *et al.* (2004) to generate a CFD&POSE vector which gives object recognition and pose estimation information to the robot for grasping assembly components, which in conjunction with a FuzzyARTMAP ANN forms the system called SIRIO which recognize, learns and performs pose estimation of assembly components in the order of milliseconds, which constitutes a practical tool for real-world applications.

2.3 Visual behaviour

The problems of modelling and understanding visual behaviour and their semantics are often regarded as computationally ill-defined. Cognitive understanding cannot adequately explain why we associate particular meanings with observed behaviours. Interpretation of visual behaviour can rely on simple mappings from recognized patterns of motion to semantics, but human activity is complex, the same behaviour may have several different meanings depending upon the scene and task context. Behaviour interpretation often also requires real-time performance if it is to be correct in the relevant dynamic context, by real time, it is not necessary implied that all computation must be performed at full video frame-rate, as long as the interpretation of behaviour proceeds within some required time constraint (Gong and Buxton, 2002).

Considering that it is estimated that 60 per cent of sensory information in humans is provided by the visual pathway (Kronauer and Zeevi, 1985), and the biological vision concerning the pathway is a massively parallel architecture using basic hierarchical information processing (Uhr, 1980), it seems logical to look for an alternative approach with less computational power to better emulate the human visual system and it is given by connectionist models of the human cognitive process, we considered this idea to develop a machine vision system for robotic assembly.

The paper describes the robotic cell architecture with integrated force sensing and vision capability; however, the vision mode is emphasized throughout the paper.

2.4 Inspiring ideas and ART models

Knowledge can be built either empirically or by hand as suggested by Towell and Shavlik (1994). Empirical knowledge can be thought of as giving examples on how to react to certain stimuli without any explanation and hand-built knowledge, where the knowledge is acquired by only giving explanations but without examples. It was determined that in robotic systems, a suitable strategy should include a combination of both methods. Furthermore, this idea is supported by psychological evidence that suggests that theory and examples interact closely during human learning (Feldman, 1993).

Learning in natural cognitive systems, including our own, follows a sequential process as it is demonstrated in our daily life. Events are learnt incrementally, for instance, during childhood when we start making new friends, we also learn more faces and this process continues through life. This learning is also stable because the learning of new faces does not disrupt our previous knowledge. These premises are the core for the development of connectionist models of the human brain and are supported by Psychology, Biology and Computer Sciences. Psychological studies suggest the sequential learning of events at different stages or "storage levels" termed as sensory memory (SM), short term memory (STM) and long term memory (LTM).

There are different types of ANN, for this research a Fuzzy ARTMAP network is used. This network was chosen because of its incremental knowledge capabilities and stability, but mostly because of the fast recognition and geometrical classification responses.

The adaptive resonance theory (ART) is a well established associative brain and competitive model introduced as a theory of the human cognitive processing developed by Stephen Grossberg at Boston University. Grossberg resumed the situations mentioned above in what he called the *Stability-Plasticity Dilemma* suggesting that connectionist models should be able to adaptively switch between its plastic and stable modes. That is, a system should exhibit plasticity to accommodate new information regarding unfamiliar events. But also, it should remain in a stable condition if familiar or irrelevant information is being presented. He identified the problem as due to basic properties of associative learning and lateral inhibition. An analysis of this instability, together with data of categorisation, conditioning, and attention led to the introduction of the ART model that stabilises the memory of self-organising feature maps in response to an arbitrary stream of input patterns (Grossberg, 1976). The core principles of this theory and how STM and LTM interact during network processes of activation, associative learning and recall were published in the scientific literature back in the 1960s.

The theory has evolved in a series of real-time architectures for unsupervised learning, the ART-1 algorithm for binary input patterns (Carpenter and Grossberg, 1981). Supervised learning is also possible through ARTMAP (Carpenter and Grossberg, 1991) that uses two ART-1 modules that can be trained to learn the correspondence between input patterns and desired output classes. Different model variations have been developed to date based on the original ART-1 algorithm, ART-2, ART-2a, ART-3, Gaussian ART, EMAP, ViewNET, Fusion ARTMAP, LaminART just to mention but a few.

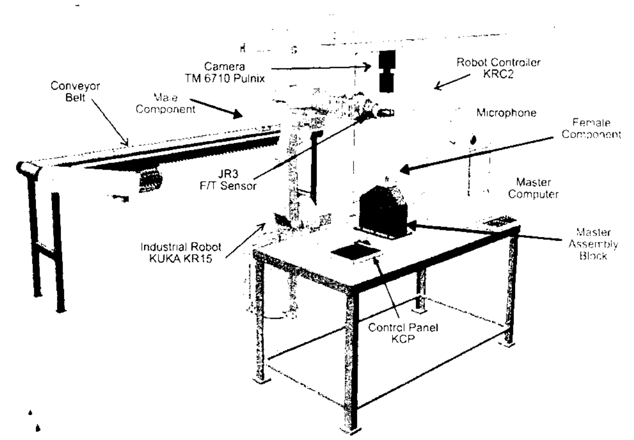
3. Manufacturing cell

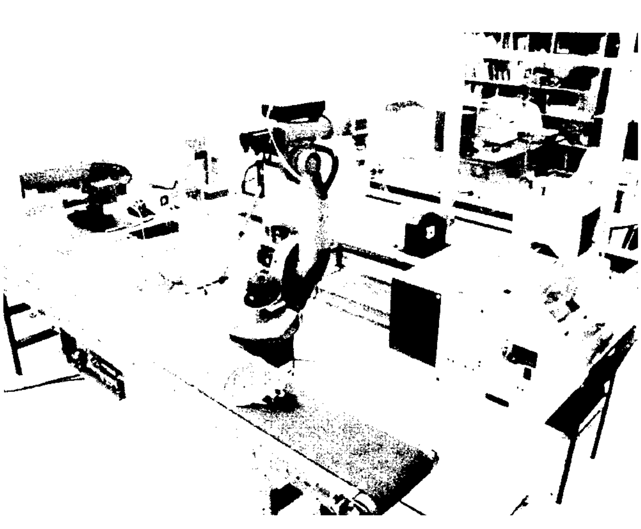
Different sensors have been used in manufacturing systems to achieve specific tasks such as robot guiding, soldering, sorting, quality control and inspection. Integration of new architectures and methods using sensorial modalities in manufacturing cells like vision, force-sensing and voice recognition becomes an open research field. Most automated systems integrators and designers had pushed hard to get faster and more accurate industrial robot systems but sensorial capabilities have not been developed completely to provide the required flexibility and autonomy for manufacturing tasks. Basic requirements within an industrial production environment have to be satisfied to guarantee an acceptable manufacturing process; some factors are the tool or work-piece position uncertainty, which is achieved by using expensive structured manufacturing cells. Other factors are the force-torque and interaction evaluation with the task environment. By using self-adaptive robots with sensorial capabilities and skill learning online, great flexibility and adaptability is given to manufactured processes, so the idea of giving machines capabilities like humans in learning and execution tasks becomes real (Wu *et al.*, 1996).

3.1 Workcell architecture

The workcell is formed basically by a 6 DOF KUKA KR15 industrial robot, KRC2 robot controller, KUKA Control Panel (KCP), PC Master Computer, JR3 F/T sensor attached to the robot's wrist, a ceiling mounted CCD camera and a conveyor belt as it is illustrated in Figure 1. The main units of the robot system are the KRC2 controller and the robot-arm itself. The KRC2 controller houses the components that control and power the robot-arm. The Master Computer host the DSP-based F/T sensor card and also communicates with the robot controller at lower level via serial port using the 3964R protocol. The vision system uses an auxiliary computer – not shown – in which algorithms for POSE determination (orientation and location) reside. POSE information about the components on the conveyor belt is provided by the Auxiliary Computer to the Master Computer, which in turn issues proper motion commands to the KRC2 controller for component grasping. Once the part (male component) is held by the robot, then the vision system also determines the female location at the Master Assembly Block and sends the female centroid information to the Master Computer in order to move the male component above the female component in readiness for assembly.

Figure 1 Manufacturing cell





3.2 Control architecture

Hardware control architecture is comprised by the elements depicted in previous section; Figure 2 shows the communication interaction among modules. Power supply and data are interconnected by way of two master cables, the controller houses the power supply and control components for the robot-arm and connects with vision and force-sensing controllers using a serial port. Programs are coded in high level language and robot and vision commands uses low level language.

3.3 Assembly

The success of assembly operations using industrial robots is currently based on the accuracy of the robot itself and the precise knowledge of the environment, i.e., information about the geometry of the assembly parts and their localisation in the workspace. Robot manipulators operate in real world situations with a high degree of uncertainty and require sensing systems to compensate from potential errors during operations. Uncertainties come from a wide variety of sources such as robot positioning errors, gear backlash, arm deflection, ageing of mechanisms and disturbances. Controlling all the above aspects would certainly be a very difficult task; therefore a simpler approach is preferred like using vision-guided robots for aligning parts in assembly tasks.

Several tests were carried out to assess the vision-guided assembly process using aluminium pegs with different cross-sectional geometry: circular, squared and radiused-square (termed radiused-square because it was a square peg with one corner rounded). These components are shown in Figure 3 as well as the peg-in-hole operation in Plate 1. The diameter of the circular peg was 25 mm and the side of the square peg was also 25 mm. The dimensions of the non-symmetric part, the radiused-square, was the same as the squared peg with one corner rounded to a radius of 12.5 mm. Clearances between pegs and mating pairs were 0.1 mm, chamfers were at 45° with 5 mm width. The assembly was ended when 3/4 of the body of

Figure 2 Control architecture

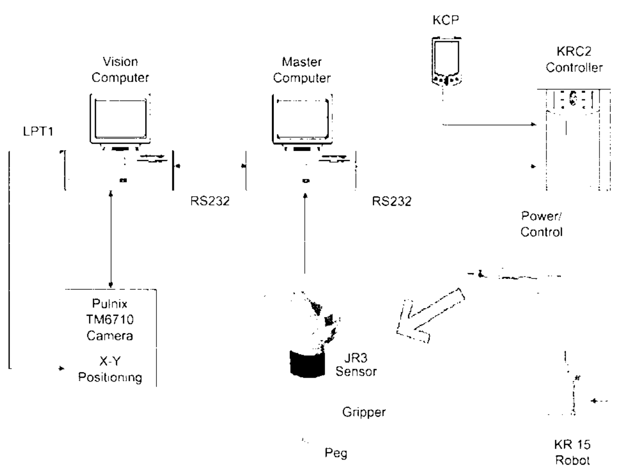


Figure 3 Assembly components

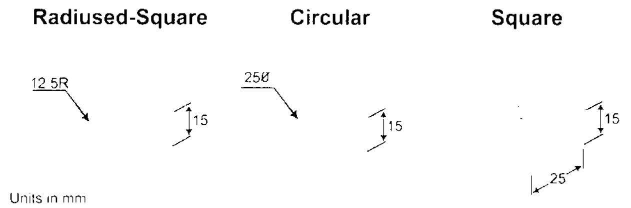
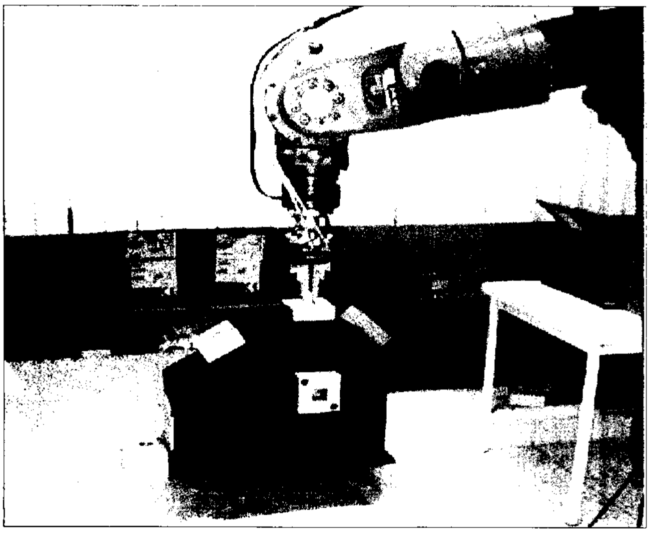


Plate 1 Peg-in-hole operation



the peg were inside the hole. This represented 140 motion steps in the -Z assembly direction.

In our experiments, the robot grasps pieces from a conveyor belt and performs an assembly task using a force-sensing system architecture called SIEM (*Sistema Inteligente de Ensamble Mecánico*), the vision system gets an image to recognize and calculates the object's pose estimation and sends the information to the robot.

4. Vision system

4.1 Vision workspace

The vision system was implemented with a high speed camera CCD/B&W, PULNIX 6710, with 640 × 480 resolution; camera movements over the X and Y axis were implemented with a computer controlled 2D positioning system as shown in Plate 2.

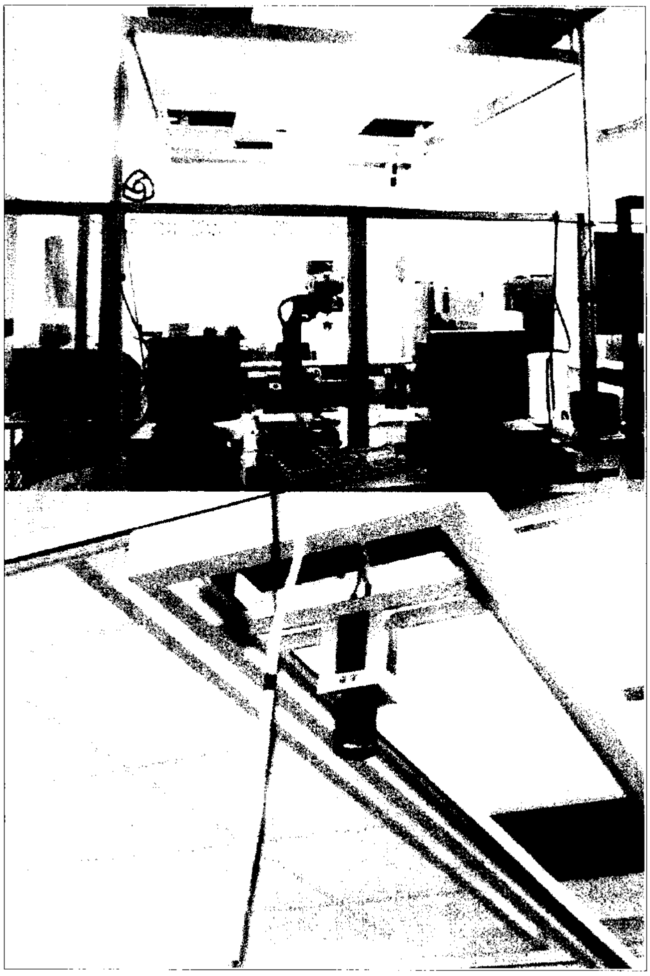
4.2 System integration

The vision system interaction schedule is carried out by a serial communication link with the robotic assembly module and a custom interface with the camera positioning system as it was shown in Figure 2.

The robotic assembly system sends commands to the vision system as follows:

(1) *$SENDINF#1 Send Information of Zone 1*: zone 1 is the place where the robot grasps the male components. The robot can locate different pieces and their characteristics.
(2) *$SENDINF#2 Send information of zone 2*: zone 2 is the place where the robot is performing the assembly task. The assembly system can request information about the female component such as position and shape.
(3) *$RESEND#X Resend information of zone X*: This command will be useful when the information received by the assembly system coming from the vision system is incorrect, due to an error in the check sum or any other error.

Plate 2 Vision workspace. Overview (top) and close up (bottom)



The communication protocol is shown in Figure 4.

The response from the vision system is a function of the request command from the assembly system, which coordinates the activities of the intelligent manufacturing cell (Castuera and López-Juárez, 2004).

4.3 Invariant object recognition

The proposed methodology for invariant object recognition is based on the use of canonic shapes within what we called the primitive knowledge base (PKB). Once having embedded this knowledge, the idea is to improve and refine it online, which compares favourably with Gestalt principles such as grouping, proximity, similarity and simplicity (Feldman, 1993). To

Figure 4 Communication protocol

# Zone	Command	Type	C-Sum

illustrate the methodology, it will be useful to consider the assembly components used during experiments. The 2D representation of the working pieces is shown in Figure 5.

These canonical shapes serve as "clues" inserted initially in the PKB which initialise the grouping process (clustering). The knowledge is acquired by presenting multiple instances of the object such as those shown in Figure 6 where and example of the circular shape and some of the possible views are illustrated. The following step is to code the object's information to get a descriptor vector, so that its description be invariant to location, scaling and rotation, the algorithm is explained in the following section.

Having such a descriptor vector, an ANN can be trained and it is expected to have incremental knowledge to conform the *descriptor vector families* which can be generated online with the same vision system.

5. Object recognition methodology

The following methodology is employed:
(1) Finding the region of interest (ROI);
(2) Calculate the histogram of the image;
(3) Search for pieces;
(4) Centroid calculation;
(5) Piece orientation;
(6) Calculate boundary object function (BOF);
(7) Descriptor vector generation and normalization (CFD&POSE); and
(8) Information processing in the neural network.

5.1 Finding the region of interest

It is desirable first to segment the region of the whole scene to have only the workpieces region of interest (ROI). There are two defined regions of interest in the manufacturing cell. The assembly workspace (zone 1) and the identification/grasping workspace (zone 2). The camera has to be positioned in the vision zone requested by the robot. The 2D positioning system, which uses feedback vision using an searching algorithm employing two LED's in order to reach the exact reference for the position of the vision system. The original image is 480×640 pixels, eight-bit greyscale resolution.

Figure 5 2D representation of assembly pieces

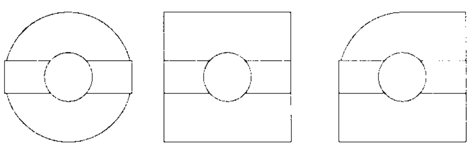


Figure 6 Object examples

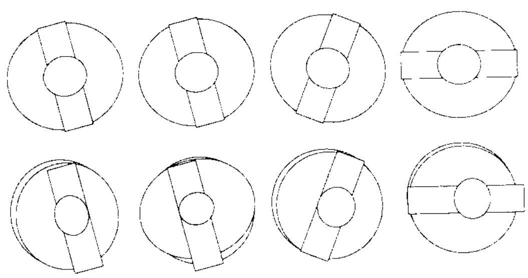


Image conditioning is carried out avoiding the processing of small objects and finding the initial position of the desired zone, (in this example zone 1). The quantized grey level value of the LEDs in the image is greater than or equal to a specific value GL, regardless of the amount of light in the zone (see Figure 7). With this process, most of the objects that can confuse the system are rejected. Then the ROI is first extracted by using the 2D histogram information and initial position reference.

To determine which are the more approximated white blobs within the image, it has to be considered the mark using the following criteria:
- Colour $Gl > 245$
- $25 \leq$ Perimeter ≤ 35 pixels (i.e., the measured size of the leds)
- The distance between the leds, must be constant at $50\,\mathrm{mm} \pm 3\,\mathrm{mm}$.

In the initial position search, only the objects that fulfil all mentioned characteristics are processed, all others are rejected. In this way, initial position is found and ROI defined as showed in Figure 7.

5.2 Image histogram process

An algorithm using 1D and 2D image histograms is used in order to provide the system of illumination invariance within some specific range. From these histograms, threshold values are used for image segmentation of the background and the pieces within the ROI eliminating the noise that may appear. This dynamic threshold value calculation allows independent light conditions operation of the system. The 1D histogram normally has the aspect shown in Figure 8.

Figure 7 Zone 1 vision workspace

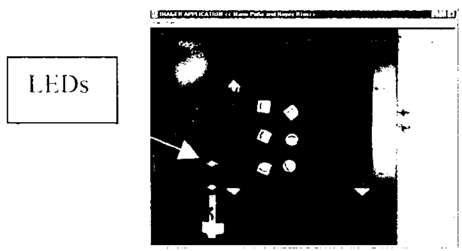
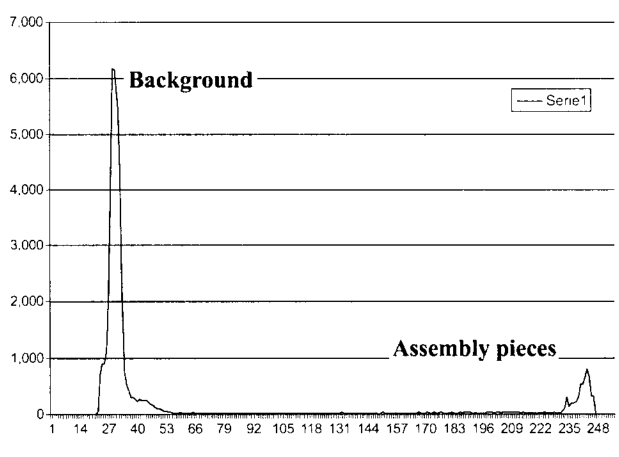


Figure 8 Histogram of the region of interest (ROI)



The two peaks in the histogram represent the background and the pieces in the image. After the histogram calculation, an image binarization is performed using a threshold operator.

5.3 Search for pieces

For searching purposes, the system calculates the perimeter obtaining:

- Number of points around a piece
- Group of points coordinates X and Y, corresponding to the perimeter of the piece measured clockwise
- Boundaries of the piece 2D Bounding Box (2D-BB)

The perimeter calculation for every piece in the ROI is performed after the binarization. Search is always accomplished from left to right and from top to bottom.

Once a white pixel is found, all the perimeter is calculated with a search function (see Figure 7). The next definitions are useful to understand the algorithm:(Figure 9)

A *nearer pixel to the boundary* is any pixel surrounded mostly by black pixels in connectivity eight.

A *farther pixel to the boundary* is any pixel that is not surrounded by black pixels in connectivity eight.

The *highest and lowest* coordinates are the ones that create a rectangle (Boundary Box).

The search algorithm executes the following procedures once it has found a white pixel:

(1) Searches for the nearer pixel to the boundary that has not been already located.
(2) Assigns the label of actual pixel to the nearer pixel to the boundary recently found.
(3) Paints the last pixel as a visited pixel.
(4) If the new coordinates are higher than the last higher coordinates, it is assigned the new values to the higher coordinates.
(5) If the new coordinates are lower than the last lower coordinates, it is assigned the new values to the lower coordinates.
(6) Steps 1-5 are repeated until the procedure returns to the initial point, or no other nearer pixel to the boundary is found.

This technique will surround any irregular shape, and will not process useless pixels of the image, therefore this is a fast algorithm that can perform online classification, and can be classified as linear:

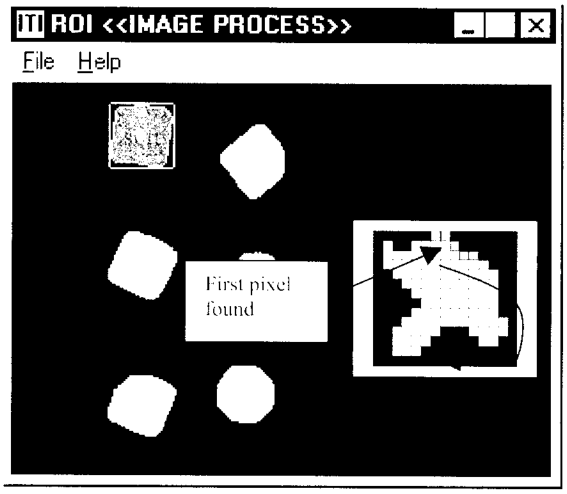
$$O(N*8*4)$$

where N is the size of the perimeter, and 8 and4 are the number of comparisons the algorithm needs to find the pixel farer to the boundary, the main difference with the traditional algorithm consist of making the sweep in an uncertain area which is always larger than the figure, this turns the algorithm into:

$$O(N*M)$$

$N*M$, is the size of the Boundary Box in use, and it does not obtain the coordinates of the perimeter in the desired order.

Figure 9 Perimeter calculation of a workpiece



5.4 Centroid calculation

The procedure proposed for centroid calculation is performed at the same time that the coordinates of the perimeter are calculated without using the $N*M$ pixels box, (*Boundary Box*).

The coordinates of the centroid (Xc, Yc) are calculated with the following procedure:

(1) If a new pixel is found and it has not been added, the value of i, j coordinates from pixel to left is added, until a new black or visited pixel is found.
(2) While a new pixel is found repeat step 1.

Figure 10 demonstrates how the sum is made from right to left as indicated by the black arrows.

The equation (1) is used for centroid calculation in binarized images:

$$X_c = \frac{\sum_{x,y} j}{A}, \; Y_c = \frac{\sum_{x,y} i}{A} \qquad (1)$$

where A is the area or number of pixels of the piece.

5.5 Piece orientation

The projected shadow by the pieces is used to obtain its orientation. Within the shadow, the largest straight line is used to calculate the orientation angle of the piece using the slope of this line, see Plate 3.

The negative image of the shadow is obtained becoming a white object, from which, the perimeter is calculated and also the two most distant points (x_1 y_1, x_2 y_2) are determined. These points define the largest straight line, the equation for the distance between two points is used to corroborate if is the

Figure 10 Centroid calculation

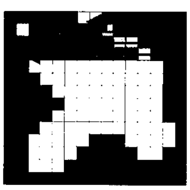
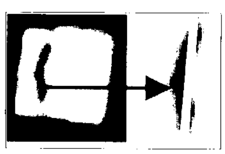


Plate 3 Shadow for the orientation



largest straight line, and also if it pass trough the centroid equation (2).

$$Y_C - y_1 = m(X_C - x_1) \tag{2}$$

slope is obtained using equation (3):

$$m = \frac{y_2 - y_2}{x_2 - x_1} \tag{3}$$

5.6 Boundary object function (BOF)

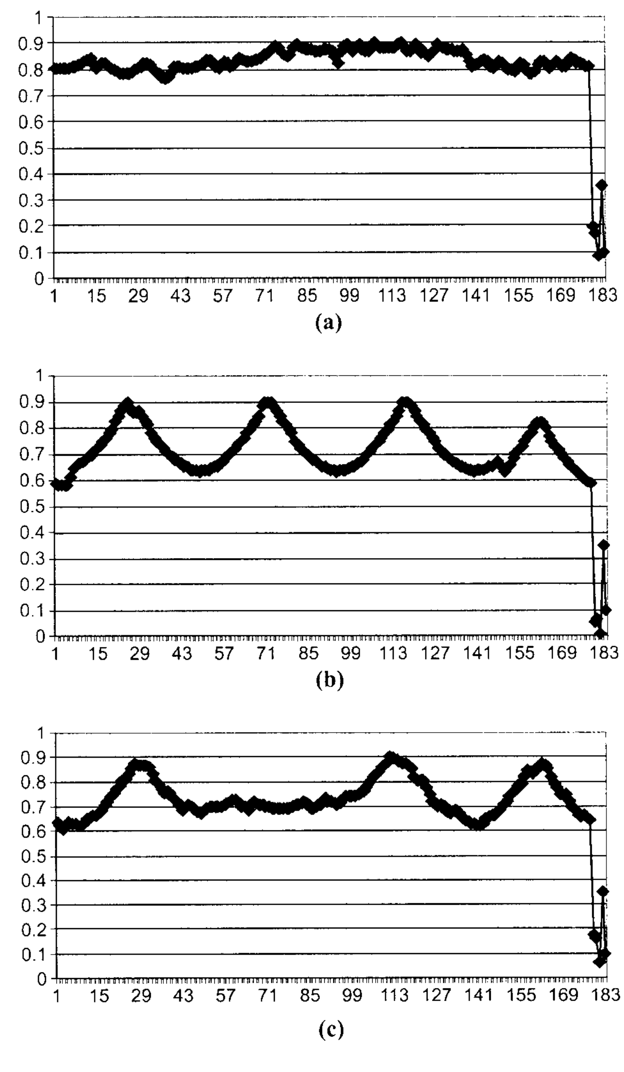
The boundary object function (BOF), is the function that describes a specific piece and it will vary according to the shape (see Figure 11).

The centroid, the coordinates of the perimeter and the distance from the centroid to the perimeter points are used to calculate the BOF.

With the coordinates P_1 (X_1, Y_1) and P_2 (X_2, Y_2), the equation (4) is applied:

$$d(P_1, P_2) = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \tag{4}$$

Figure 11 BOF (a) circle; (b) square; (c) radiused-square



5.7 Descriptive vector generation and normalization

Once the information has been processed, a descriptive vector is generated. This vector is the input for the neural network:

The descriptive vector is called CFD & POSE and it is conformed by:

$$[\text{CFD \& POSE}] = \begin{bmatrix} D_1 \\ D_2 \\ D_3 \\ D_n \\ X_c \\ Y_c \\ \phi \\ Z \\ ID \end{bmatrix} \tag{5}$$

where D_i are the distances from the centroid to the perimeter of the object. X_C, Y_C, are the coordinates of the centroid. ϕ, is the orientation angle. Z is the height of the object ID is a code number related to the geometry of the components.

5.8 Information processing in the neural network

The vision system extends the BOF data vectors to 180, plus five more data vectors, centroid (*Xc*, *Yc*), orientation, height and ID as showed in Table I.

And this is the input to the FuzzyARTMAP neural network.

6. Experimental results

The methodology was coded using Visual C++6.0 and a PC with PIII processor at 800 MHZ and 192 MB RAM.

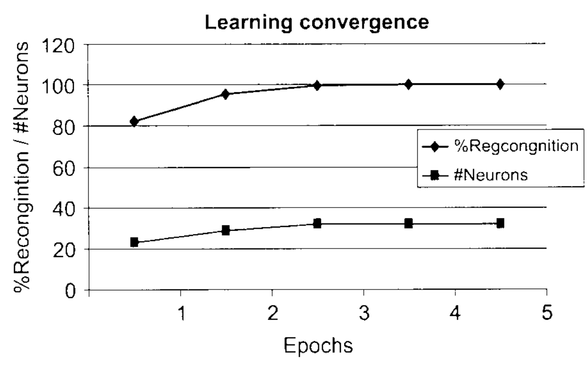
6.1 Object recognition

In order to test the robustness of the ANN, the Fuzzy ARTMAP Neural Network was trained first with 2,808 different patterns and its learning capability analyzed. Results regarding the percentage of recognition and the number of generated neurons are shown in Figure 12. The graph shows

Table I Comp

Data 1-180	Cen 182-180	Orient 183	Height 184	ID 185

Figure 12 Learning of the neural network



how the system learned all patterns in three epochs, creating only 32 neurons to classify 2,808 patterns.

The average time for training is 4.42 ms, and the average for testing is 1.0 ms.

Results reported in this article employed 216 patterns corresponding to 72 square, 72 circle and 72 radiused-square components of the same size. The orientation values were 0, 45, 90, 135, 180, 215, 270 and 315°.

With these training patterns set, the system was able to classified correctly 100 per cent of the pieces presented on- line even if they were not of the same size, orientation or locations and for different light conditions. The pieces used to train the neural network are shown in Figure 13 and 14 shows different graphs corresponding to different descriptor vectors for different positions, sizes and illumination conditions of these components.

Several tests with different geometry, positions and light conditions, were carried out online (Figure 15).

The normalization of the BOF is done using the *maximum value divisor of the distance* vector method. This method allows having very similar patterns as input vectors to the neural network, getting a significant improvement in the operation system. Figure 16 shows the generated similar patterns using totally different size, location and orientation conditions for working pieces.

In our experiments, the object recognition method used the above components to demonstrate the assembly process, however, the SIRIO system can recognize more complex components as it is shown in Figure 17, where four different animal shapes were used for testing. Same results for invariance and object recognition were obtained.

The CFD & POSE vectors were generated and Figure 18 shows their respective BOF functions. The information

Figure 13 Workpieces used to create the initial knowledge base

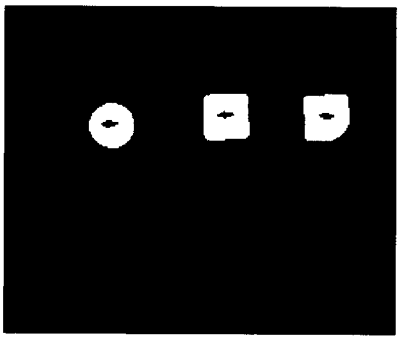


Figure 14 (a) circle, (b) square, (c) radiused-square

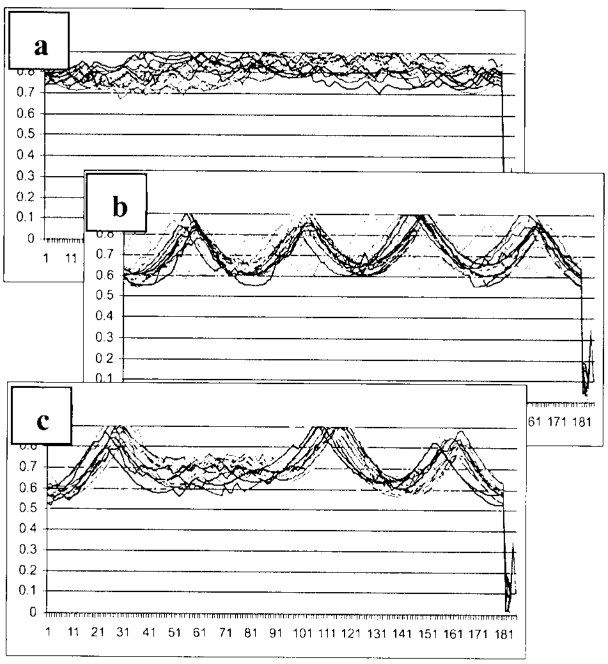


Figure 15 Pieces used to test the system

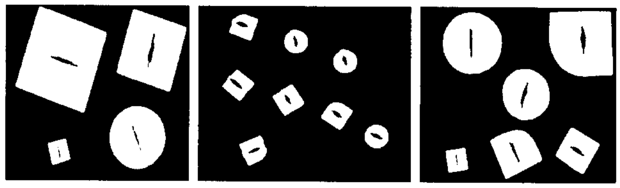
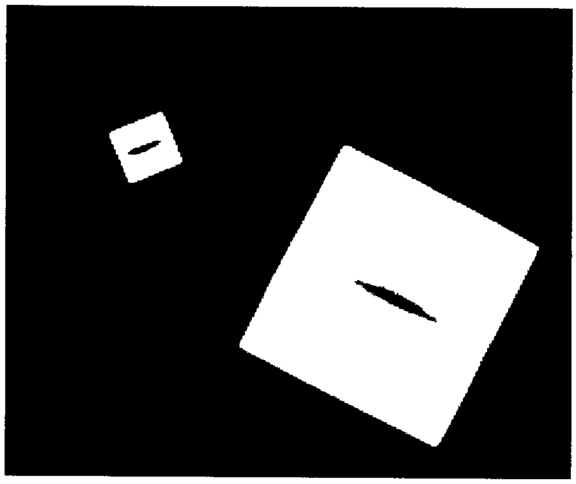
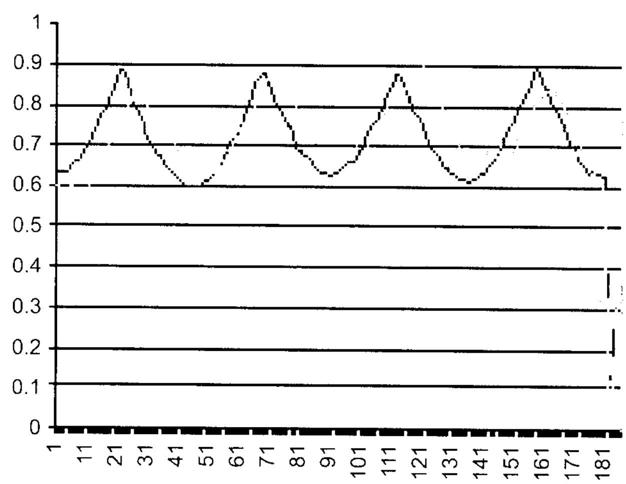


Figure 16 (a) squares (b) Similar patterns

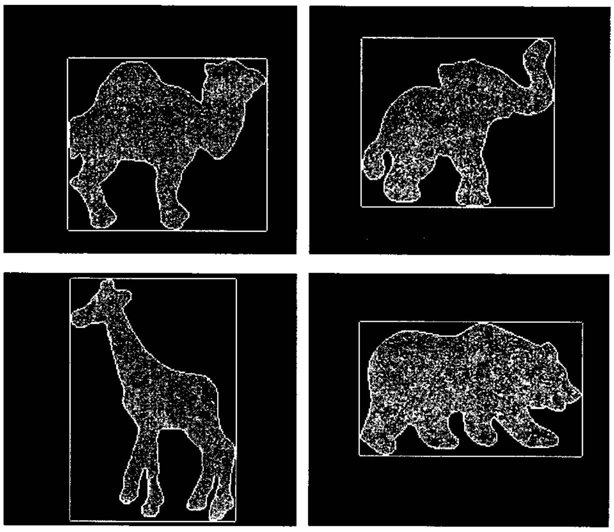


(a) Squares



(b) Similar Patterns

Figure 17 More complex components



contained within the vector allows the system to remember even when and where the object was used for the last time.

6.2 Assembly cycles in the manufacturing cell

Table II shows the results for 18 assembly cycles using the vision system SIRIO and the assembly system SIEM. This table contains information regarding the type of piece in use, presence or absence of chamfer, total operational time (object recognition, grasping, moving the part from pick up place to assembly point and assembly), the calculated error based on the centroid and rotation angle of the pieces for zone 1 and the offset error in zone 2. Finally, in the last column, the type of geometry recognized online by the neural network is provided.

The vision system provides the robot with the capability to approach two zones:

Zone 1: assembly workpiece (male) area of vision, where robot picks up the pieces after having POSE information of the object, then grasps and takes them to zone 2.

Zone 2: peg-in-hole assembly (female) area of vision, here the visually guided robot approaches the zone where the female component is located to achieve the assembly task and releasing the control of the operation to the SIEM assembly system.

The POSE 1 means the location estimation of a workpiece within the zone 1, and the POSE 2 means the location estimation within the zone 2 of the workpiece/counterpart.

Grasp testing (zone 1) was done for each geometry; every one was placed three times within the vision area, incrementing 10° its orientation and changing the locations in four defined poses. In the zone 2, the location of the female component was constant so the angle too, but the robot did not known the pose previously. The first nine assembly cycles were done with female chamfered components and the last nine with chamfer-less components.

The average time of the total cycle is 1:50.6 min and the minimum time is 1:46 min, the longest time is: 1:58 min.

The average of the error made in both zones is: 0.8625 mm, the minimum is: 0 mm while the maximum is 3.4 mm.

The average of the error angle is: 4.27°, the minimum is: 0° and the maximum is 9°.

Figure 19, shows 18 different *X* and *Y* points where the robot might reach the male component showed as error *X* (mm) and error *Y* (mm), and Figure 20 shows the angle error for orientation grasping purpose using the same notation.

In the Assembly area the robot gets vision guided capabilities to approach the zone to the centre of the workpiece/counterpart, Figure 21 shows 18 different *X* and *Y* points where the robot might reach the female and releases control to force/sensing system.

The 18 assembly cycles were done successfully. Figures 19-20 show that all the poses given by the SIRIO are inside the error limits in both areas: zone 1 and zone 2. This permitted to have a 100 per cent of success in the total assembly cycle operation.

7. Conclusions and future work

A novel methodology for fast object recognition and POSE estimation for assembly components in manufacturing cells

Figure 18 Descriptor vector for animals

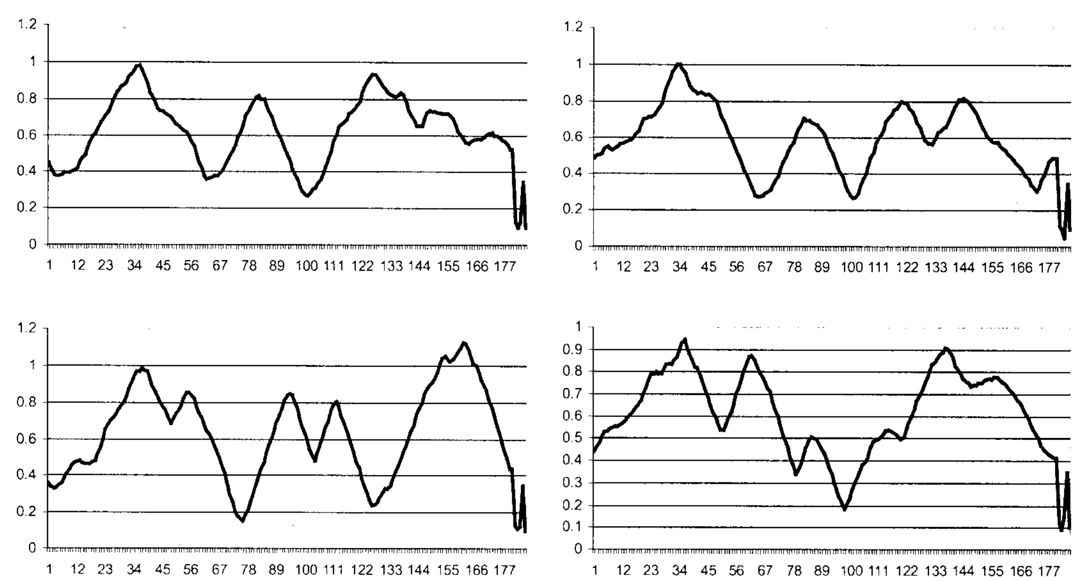


Table II 18 Assembly cycles using the vision system SIRIO and the assembly system SIEM

				Zone 1			Error zone 1			Zone 2		Error zone 2		
No.	Piece	Cham	Total time (min)	X (mm)	Y (mm)	RZ°	X (mm)	Y (mm)	RZ°	X (mm)	Y (mm)	X (mm)	Y (mm)	Neural classific.
1	Square	Yes	49	59.85	142.2	12	0.15	−0.2	−2	82.8	102	−2.2	−0.9	square
2	Square	Yes	01:58	58.6	44.8	22	1.4	0.2	−2	83.8	100.2	−1	0.9	square
3	Square	Yes	01:46	174.8	45.8	26	0.2	0.2	4	84.9	101.1	−0.1	0.2	square
4	Radius	Yes	01:49	177.7	143.1	41	1.3	0.9	−1	84.9	101.1	0.3	0.1	radius
5	Radius	Yes	01:47	58.6	142.2	48	1.4	−0.2	2	84.9	101.1	0.2	0.3	radius
6	Radius	Yes	01:54	59.5	42.8	63	0.5	2.2	−3	86	101.1	2.1	−0.1	radius
7	Circle	Yes	01:47	174.8	45.8	79	0.2	0.2	−9	84.9	101.1	1	−0.1	circle
8	Circle	Yes	01:48	176.7	143.1	80	2.3	0.9	0	86	101.1	1.8	−0.3	circle
9	Circle	Yes	01:46	56.6	143.1	98	3.4	−1.1	−8	84.9	101.1	1.1	0	circle
10	Square	No	01:53	58.6	42.8	94	1.4	2.2	6	84.9	101.1	0.4	−0.1	square
11	Square	No	01:51	173.8	44.8	116	1.2	1.2	−6	84.9	101.1	0.6	−0.3	square
12	Square	No	01:47	177.7	143.1	116	1.3	0.9	4	84.9	100.2	0.4	0.6	square
13	Radius	No	01:56	58.6	143.1	124	1.4	−1.1	6	83.8	101.1	−0.2	−0.2	radius
14	Radius	No	01:50	59.85	43.8	143	0.15	1.2	−3	82.8	100.2	−1.4	0.8	radius
15	Radius	No	01:53	173.8	43.8	155	1.2	2.2	−5	82.8	101.1	−1.4	−0.1	radius
16	Circle	No	01:46	177.7	143.1	157	1.3	0.9	3	84.9	100.2	0.4	0.8	circle
17	Circle	No	01:50	58.6	143.1	175	1.4	−1.1	−5	83.8	101.1	−0.5	−0.3	circle
18	Circle	No	01:51	58.6	44.8	172	1.4	0.2	8	83.8	100.2	−1.3	0.9	circle

Note: The time showed in fourth column is the total time to achieve a complete assembly cycle in the manufacturing cell, including object recognition time

has been described. Experimental results show the methodology. Issues regarding image processing, centroid and perimeter calculation are illustrated. The methodology was tested on a manufacturing cell with assembly components. Results show the feasibility of the method to send grasping and morphologic information (coordinates and classification characteristics) to a robot in real-time. A robust positioning system that corrected errors due to wheel sliding was implemented using visual feedback information. The overall methodology was implemented and integrated in a prototype cell showing real performance of industrial processes. Accurate recognition of assembly components and workpieces identification was successfully carried out by using a FuzzyARTMAP neural network model. The performance of this model was satisfactory with recognition times lower than 5 ms. and identification rates of 100 per cent.

Experimental measurements showed ± 3 mm of precision error in the information sent to the robot. The orientation angle of the pieces had up to ± 9° error, good enough for the robot to be able to handle the pieces.

The intelligent manufacturing cell is being developed with multimodal sensor capabilities, and uses the methodology presented in this work. Current work addresses the use of ANN's for assembly and object recognition separately, future work is oriented to use the same neural controller for all different sensorial modes. The SIRIO vision system

Figure 19 Positional error referenced to real centroid in male component

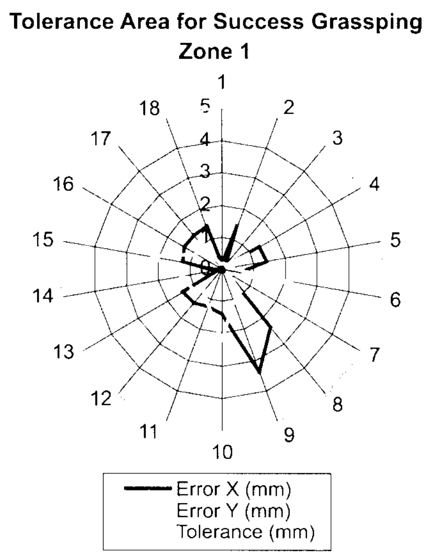


Figure 20 Rotational error for orientation grasp

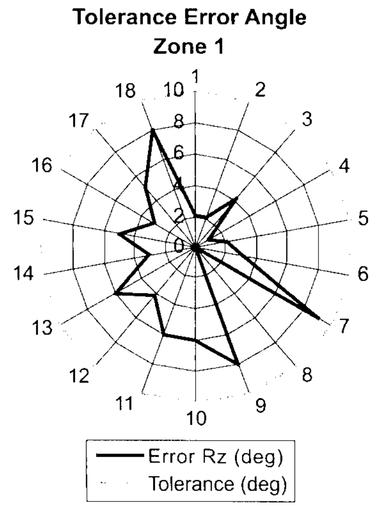
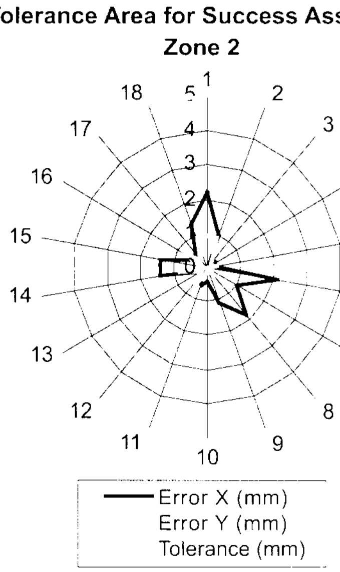


Figure 21 Positional error referenced to real centroid in female component



architecture is being improved for handling complex 3D objects in manufacturing applications.

8. Glossary

ANN = artificial neural network
ART = adaptive resonance theory
BB = boundary box
BOF = boundary object function
BP = back propagation
CFD = current frame descriptor
GL = grey level value
PKB = primitive knowledge data base
POSE = position and orientation estimation
RFA = robotics fixtureless assembly
RL = reinforcement learning
SIEM = assembly manufacturing intelligent system
SIRIO = invariant object recognition intelligent system
SOM = self organized map

References

Aguado, A.S., Montiel, E. and Nixon, M. (2002), "Invariant characterization of the Hough transform for pose estimation of arbitrary shapes", *Pattern Recognition*, Vol. 35, pp. 1083-97.

Best, P.J. and McKay, N.D. (1992), "A method for registration of 3-D shapes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14 No. 2.

Bone, G.M. and Capson, D. (2003), "Vision-guided fixtureless assembly of automotive components", *Robotics and Computer Integrated Manufacturing*, Vol. 19, pp. 79-87.

Bribiesca, E. (1999), "A new chain code", *Pattern Recognition*, Vol. 32, pp. 235-51.

Carpenter, G.A. and Grossberg, S. (1987), "A massively parallel architecture for a self-organizing neural pattern recognition machine", *Computer Vision, Graphics, and Image Processing*, Vol. 37, pp. 54-115.

Carpenter, G.A., Grossberg, S. and Reynolds, J.H. (1991), "ARTMAP: supervised real-time learning and classification of nonstationary data by self-organizing neural network", *Neural Networks*, Vol. 4, pp. 565-88.

Castuera, J.C. and López-Juárez, I. (2004), "Intelligent task level planning for robotic assembly: issues and experiments", paper presented at the Mexican International Conferences on Artificial Intelligence, Springer-Verlag, New York, NY.

Chen, K. (1990), "Efficient parallel algorithms for computation of two-dimensional image moments", *Pattern Recognition*, Vol. 23, pp. 109-19.

Chin-Hsiung, Wu *et al.*, (2001), "A new computation of shape moments via quadtree decomposition", *Pattern Recognition*, Vol. 34, pp. 1319-30.

Dickmanns, E. (1998), "Vehicles capable of dynamic vision: a new breed of technical beings?", *Artifical Intelligence*, Vol. 103, pp. 49-76.

Gupta, M.M. and Knopf, G. (1993), *Neuro-Vision Systems: a Tutorial*, a selected reprint, Volume IEEE, Neural Networks Council Sponsor, IEEE Press, New York, NY.

Gupta, M.M. and Knopf, G. (1993), *Neuro-Vision Systems: Part 6 Computational Architectures and Applications*, A selected reprint Volume IEEE Neural Networks Council Sponsor, IEEE Press, New York, NY.

Lopez-Juarez, I. and Howarth, M. (2000), "Learning, Manipulative Skills with ART", *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems.*

Peña-Cabrera, M. *et al.* (2004), "Un Proceso de Aprendizaje para Reconocimiento de Objetos en Línea en Tareas Robotizadas", 3ra Conferencia Iberoamericana en Sistemas, Cibernética e Informática CISCI, del 21 al 25 de julio del, Orlando, FL.

Vijaykumar, G., Franklin, J.A. and Benbrahim, H. (1994), "Acquiring robot skills via reinforcement learning", *IEEE Control Systems*, pp. 13-24.

Enric, Cervera and del pobil, A.P. (1997), "Programming and learning in real world manipulation tasks", *Proceedings of the IEEE/RSJ International Conference on Intelligent Robot and Systems*, pp. 471-6.

Feldman, R.S. (1993), *Understanding Psychology*, 3rd ed., McGraw-Hill, New York, NY.

Freeman, H. (1961), "On the encoding of arbitrary geometric configurations", *IRE Trans. Electron. Comput.*, Vol. EC-10, pp. 260-8.

Freeman, H. (1974), "Computer processing of line drawings images", *ACM Comput. Surveys*, Vol. 6, pp. 57-97.

Gong, S. and Buxton, H. (2002), "Understanding visual behaviour", *Image and Vision Computing*, Vol. 20 No. 12, pp. 825-6.

Gonzalez-Galvan, E.J. *et al.*, (1997), "Application of precision-enhancing measure in 3D rigid-body positioning using camera-space manipulation", *The International Journal of Robotics Research*, Vol. 16 No. 2, pp. 240-57.

Grossberg, S. (1976), "Adaptive pattern classification and universal recoding. II: feedback, expectation, olfaction and illusions", *Biological Cybernetics*, Vol. 23, pp. 187-202.

Hager, G. *et al.* (1995), "Calibration-free visual control using projective invariance", *Proceedings of the ICCV*, pp. 1009-15.

Hoska, D.R. (1988), "Fixtureless assembly manufacturing", *Manuf Eng.*, Vol. 100, pp. 49-54.

Assembly Automation
Volume 25 · Number 3 · 2005 · 204–216

Howarth, M. (1998), "An investigation of task level programming for robotic assembly", PhD thesis, The Nottingham Trent University, Nottingham.

Hu, M.K. (1962), "Visual pattern recognition by moment invariants", *IRE Trans. Inform. Theory*, Vol. IT-8, pp. 179-87.

Jörg, S. *et al.* (2000), "Flexible robot-assembly using a multi-sensory approach", *Proceedings of the IEEE, International Conference on Robotics and Automation*, San Francisco, CA, USA, pp. 3687-94.

Kollnig, H. and Nagel, H. (1997), "3D pose estimation by directly matching polyhedral models to gray value gradients", *International Journal of Computer Vision*, Vol. 23 No. 3, pp. 282-302.

Kronauer, R.E. and Zeevi, Y. (1985), "Reorganization and diversification of signals in vision", *IEEE Trans. Syst. Man, Cybern.*, Vol. SMC-15 No. 1, pp. 91-101.

Langley, C.S. and D'Eleuterio, G.M.T. (2003), "A memory efficient neural network for robotic pose estimation", *Proceedings of the 2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation*, Vol. 1, pp. 418-23.

Lopez-Juarez, I. (2000), "Online learning for robotic assembly using artificial neural networks and contact force sensing", PhD thesis, The Nottingham Trent University, Nottingham,.

Lin, J.C. (1996), "The family of universal axes", *Pattern Recognition*, Vol. 29 No. 3, pp. 477-85.

Murase, H. and Shree K., Nayar (1995), "Visual learning and recognition of 3-D objects from appearance", *International Journal of Computer Vision*, Vol. 14, pp. 5-24.

Ngyuen, W. and Mills, J.K. (1996), "Multirobot control for flexible fixtureless assembly of flexible sheet metal autobody parts", *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2340-5.

Papanikolopoulous, N.P. and Khosla, K. (1993), "Adaptive robotic visual tracking: theory and experiments", *IEEE Transactions on Automatic Controller*, Vol. 38 No. 3, pp. 429-45.

Peña-Cabrera, M. *et al.* (2004), "A learning approach for on-line object recognition in robotic tasks", paper presented at the Mexican International Conference on Computer Science ENC 2004, IEEE Computer Society Press, México.

Philips, W. (1993), "A new fast algorithm for moment computation", *Pattern Recognition*, Vol. 26, pp. 1619-21.

Plut, W.J. and Bone, G.M. (1996), "Limited mobility grasps for fixtureless assembly", *Proceedings of the IEEE International Conference on Robotics and Automation*, Minneapolis, MN, pp. 1465-70.

Plut, W.J. and Bone, G.M. (1997), "3-D flexible fixturing using multi-degree of freedom gripper for robotics fixtureless assembly", *Proceedings of the IEEE International Conference on Robotics and Automation*, Alburquerque, NM, pp. 379-84.

Stavros, J. and Lisboa, P. (1992), "Transltion, rotation, and scale invariant pattern recognition by high-order neural networks and moment classifiers", *IEEE Transactions on Neural Networks*, Vol. 3 No. 2.

Torralba, A. and Oliva, A. (2002), "Depth Estimation from Image Structure", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24 No. 9.

Towell, G.G. and Shavlik, J.W. (1994), "Knowledge-based artificial neural networks", *Artificial Intelligence*, Vol. 70 Nos 1/2, pp. 119-66.

Uhr, L. (1980), "Psychological motivation and underlying concepts", in Tanimoto, S. and Klinger, A. (Eds), *Structured Computer Vision*, pp. 1-30.

Underwood, S.A. *et al.*, (1975), "Visual learning from multiple views", *IEEE Transactions on Computers*, Vol. c-24 No. 6.

Wu, L., Oviatt, S.L. and Cohen, P.R. (1999), "Multimodal integration – A statical view", *IEEE Transactions on Multimedia*, Vol. 1 No. 4, pp. 334-41.

Yüceer, C. and Oflazer, K. (1993), "A rotation, scaling and translation invariant pattern classification system", *Pattern Recognition*, Vol. 26 No. 5, pp. 687-710.

Yong-Sheng Chen *et al.*, (2001), "Three dimensional ego-motion estimation from motion fields observed with multiple cameras", *Pattern Recognition*, Vol. 34, pp. 1573-83.

You, S.D. and Ford, G.E. (1994), "Network model for invariant object recognition", *Pattern Recognition Letters*, Vol. 15, pp. 761-7.