USING VISUAL AND FORCE INFORMATION IN ROBOT-ROBOT COOPERATION TO BUILD METALLIC STRUCTURES

Jorge Pomares, Pablo Gil, Juan Antonio Corrales, Gabriel J. García, Santiago T. Puente, Fernando Torres

Abstract:

In this paper, a cooperative robot-robot approach to construct metallic structures is presented. In order to develop this task, a visual-force control system is proposed. The visual information is composed of an eye-in-hand camera, and a time of flight 3D camera. Both robots are equipped by a force sensor at the end-effector. In order to allow a human cooperate with both robots, an inertial motion capture system and an indoor localization system are employed. This multisensorial approach allows the robots to cooperatively construct the metallic structure in a flexible way and sharing the workspace with a human operator.

Keywords: *visual servoing, force control, sensor fusion, estimation algorithms, robot vision.*

1. Introduction

The automatic assembly processes involve different disciplines such as assembly sequence generation, assembly interpretation, robot positioning techniques based on vision and other sensors and handling of objects of the assembly [7].

Sensors are an important subject within the machine vision for an intelligence manipulation of objects, in situations with a high degree of randomness in the environment. The sensors increase the ability of a robot to adapt to its working environment. Currently, visual sensory feedback techniques are widely considered by researches for manufacturing process automation. Over the last few years, these techniques have been used for inspection and handling of objects [11] for estimation of pose with range data and three-dimensional image processing [4] or with stereo vision [9]. Currently, the human robot interaction to help in the modelling and localization of objects [10], and the sensorial fusion and control techniques to pose and insert objects [14] in assembly processes are employed more and more.

The assembly system proposed in this paper has important advantages over the classic assembly systems, mainly due to its interaction between human and robot. In this system, the human will perform assistance tasks in the manipulation and positioning of objects. Another important aspect is the extensive use of sensors in the different phases of the task.

The implemented system is composed of several subsystems. Among them a visual-force control subsystem to guide the movement of the robot and control the manipulation of objects in each planned task is emphasized. On the one hand, the basic task of the visual information is to control the pose of the robot's end-effector using information extracted from images of the scene. On the other hand, the force information is used to control the handling and grasping of objects which are manipulated. The visual information is obtained from a camera mounted on a robot's end-effector, and the force data is obtained from a force sensor. The metallic structure to be assembled is manipulated with different tools which are interchanged automatically depending on the task that has been planned. Furthermore, the movement of a human who interacts with the robot at the same workspace is controlled and his positions are modelled with a RTLS (Real-time Location System) of radio frequency UWB (Ultra-WideBand) and with a full-body human motion capture suit. This suit is based on inertial sensors, a biomechanical model and sensor fusion algorithms. Finally, the proposed assembly system is complemented with a time of flight 3D-camera to help the visual control subsystem determine the localization of objects.

To show how each subsystem works in an assembly process, a complex metal structure has been built. The key in constructing this, it is to combine grip and insertion movements among several types of metal pieces using robotic and human manipulators jointly to perform collaborative tasks that facilitate the correct assembly with robustness.

This paper is organized as follows: The system architecture is presented in Section 2. Section 3 describes briefly the different phases of the system. These phases are presented in detail in the following sections. The visual servoing and the visual-force control approach employed to guide the robot are described in Sections 4 and 5 respectively. The robot-robot and human-robot cooperation during the task are shown in Sections 6 and 7. The final section presents the main conclusions reached.

2. System architecture

The system architecture is composed of two 7 d.o.f. Mitsubishi PA-10 robots which are able to work cooperatively. Both robots are equipped by a tool-interchanger to employ the required tools during the task (gripper, robotic hand, screwdriver, camera, etc.). Both robots are equipped with a force sensor.

An inertial human motion capture system (*GypsyGyro-18* from *Animazoo*) and an indoor localization system (*Ubisense*) based on Ultra-WideBand (UWB) pulses are used to localize precisely the human operator who collaborates in the assembly task. The motion capture system is composed of 18 small inertial sensors (gyroscopes) which measure the orientation (roll, pitch and yaw) of the operator's limbs. The UWB localization system is composed of 4 sensors which are situated at fixed positions in the workplace and a small tag which is carried by the human operator. This tag sends UWB pulses to the sensors which estimate the global position of the human.

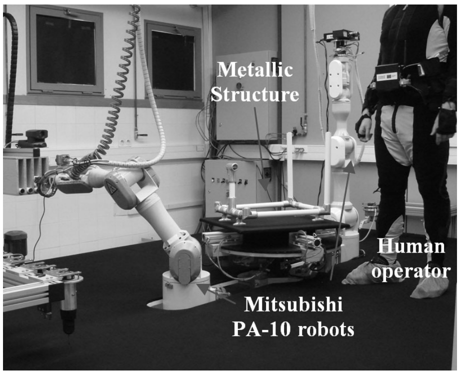

Fig. 1. System architecture.

3. Phases in the assembly system

The different phases which compose the assembly system are illustrated in Fig. 2. These phases are the following:

- *Phase 1*. Visual Servoing. This system is employed to guide the robot by using visual information.
- *Phase 2*. Visual-force control. This approach is employed during the insertion to control not only the robot position but also the robot interaction forces.
- *Phase 3*. Robot-robot cooperation. The two robots are required to work jointly in order to detect with a robot visual features of the insertion task performed by the other robot.
- *Phase 4*. Robot and human sharing the workspace. The system coordinates the robot behaviour between the human and the robot.

In the next sections these phases are described in detail.

4. Visual servoing

In this section, an approach to guide the robot using visual information is presented. To do this, it is necessary to track the desired trajectories by using a visual servoing system employing an eye-in-hand camera system.

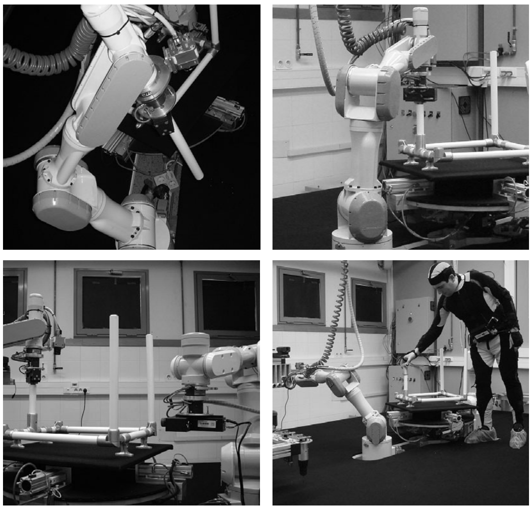

Fig. 2. Phases in the assembly system.

In a robotic task, the robot must frequently be positioned at a fixed location with respect to the objects in the scene. However, the position of these objects is not always controlled. So, it is not possible to previously assure the location of the end-effector of the robot to correctly accomplish the task. Visual servoing is a technique that allows positioning a robot with respect to an object using visual information [8].

Basically, the visual servoing approach consists of extracting visual data from an image acquired from a camera and comparing it with the visual data obtained at the desired position of the robot. By minimizing the error between the two images it is possible to control the robot to the desired position. Image-based visual servoing uses only the visual data obtained in an image to control the robot movement. The behaviour of these systems has been proved to be robust in local conditions (i.e., in conditions in which the initial position of the robot is very near to its final location) [2]. However, in large displacements, the errors in the computation of the intrinsic parameters of the camera have influence on the correct behaviour of the system [1]. Image-based visual servoing is adequate to position a robot from an initial point to a desired location, but it cannot control intermediate 3D positions of the end-effector.

A solution to this problem is to achieve the correct location following a desired path. The desired path, $T = \{{}^{k}s/k \in 1..N\}$ (with ${}^{k}s$ being the set of M points or visual features observed by the camera at instant k, ${}^{k}s = \{{}^{k}f_i/i \in 1..M\}$), is sampled and then these references are sent to the system as the desired references for each moment. In this way, the current and the final positions are very close together, and the system takes advantage of the good local behaviour of image-based visual servoing.

A visual servoing task can be described by an image function, e_t, which must be regulated to 0:

$$e_t = s - s^* \tag{1}$$

where s is a $M \times 1$ vector containing M visual features corresponding to the current state, while s^* denotes the visual features values in the desired state.

With L_s is represented the interaction matrix which relates the variations in the image with the variation in the velocity of the camera:

$$\dot{s} = L_s \cdot \dot{r} \tag{2}$$

where $\dot{s}$ are the time derivative of the image features and $\dot{r}$ indicates the velocity of the camera.

By imposing an exponential decrease of $e_t\,(\dot{e}_t = -\lambda_1 e_t)$ it is possible to obtain the following control action for a classical image-based visual servoing:

$$v_c = -\lambda_1 \hat{L}_s^{+}(s - s^*) \tag{3}$$

where $\hat{L}_s^{+}$ is the pseudoinverse of an approximation of the interaction matrix [8].

The method employed to track a previously defined path in the image space must be able to control the desired tracking velocity. The set of visual features observed at the initial camera position are represented with ${}^{1}s$. From this initial set of image features it is necessary to find an image configuration which provides the robot with the desired velocity, $|v_d|$. To do so, the system iterates over the set T. For each image configuration ${}^{k}s$ the corresponding camera velocity is determined considering an image-based visual servoing system (at this first stage $s = {}^{1}s$):

$${}^{k}v = -\lambda_1 \hat{L}_s^{+}(s - {}^{k}s) \tag{4}$$

This process continues until $|{}^{k}v|$ is greater than the desired velocity, $|v_d|$. At this moment, the set of features ${}^{k}s$ will be the desired features to be used by an image-based visual servoing system (see Equation (3)). However, the visual features, ${}^{j}s$, which provide the desired velocity are between ${}^{k}s$ and ${}^{k-1}s$. To obtain the correct image features the method described in [5] is employed.

Therefore, once the control law represented in Equation (4) is executed, the system searches again for a new image configuration which provides the desired velocity. This process continues until the complete trajectory is tracked.

5. Visual-force control

Now, we consider the task of tracking a path using visual and force information. The visual loop carries out the tracking of the desired trajectory in the image space. To do this, as it has been described in Section 4, the method to track trajectories in the image is employed:

$$v_c = -\lambda_1 \hat{L}_s^{+}(s - {}^{j}s) \tag{5}$$

where ${}^{j}s$ is the set of features in the path obtained by the system to maintain the desired velocity.

Previously to define the visual-force controller employed, the meaning of the force-image interaction matrix, L_{FI}, is described. To do this, considering F as the interaction forces obtained with respect to the robot end-effector and r as the end-effector location. The interaction matrix for the interaction forces, L_F, is defined in this way:

$$L_F = \frac{\partial F}{\partial r} \rightarrow L_s^{+} = (L_F^{T}\, L_F)^{-1} L_F^{T} = \frac{\partial r}{\partial F} \tag{6}$$

Through this last relationship and by applying (2) it is obtained:

$$\dot{s} = L_s \cdot \frac{\partial r}{\partial t} = L_s \cdot \frac{\partial r}{\partial F} \cdot \frac{\partial F}{\partial t} = L_s \cdot L_F^{+} \cdot \dot{F} \rightarrow \dot{s} = L_{FI} \cdot \dot{F} \tag{7}$$

where $\dot{F}$ are the time derivative of the interaction forces and $L_{FI} = L_s \cdot L_F^{+}$ is the interaction matrix. This matrix is estimated using exponentially weighted least-squares [6].

As it has been described in previous works [12], in order to guarantee the coherence between the visual and force information, it is necessary to modify the image trajectory through the interaction forces. Therefore, in an application in which it is necessary to maintain a constant force with the workspace, the image trajectory must be modified depending on the interaction forces. To do so, using the matrix LFI, the new desired features used by the controller during the contact will be:

$$s_d = {}^{j}s + L_{FI} \cdot (F - F_d) \tag{8}$$

Applying (8) in (3), the system is able to track a previously defined path in the image being compliant with the surface of the interaction object:

$$v_c = -\lambda_1 \hat{L}_s^{+}(s - s_d) \tag{9}$$

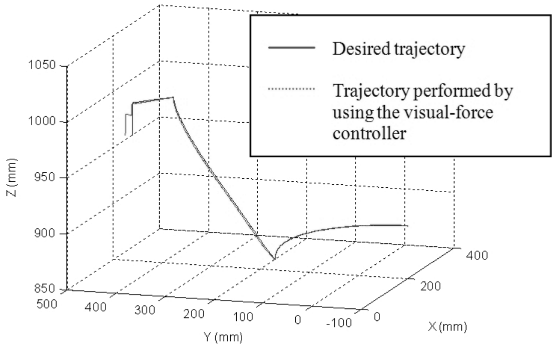

Fig. 3. 3D evolution of the end-effector in a bar insertion task.

Figure 3 shows the 3D path to perform one of the assemblies to construct the structure. The desired path has been modified taking into account the forces measured at the end-effector of the robot. In this way, the robot is able to correctly introduce the bar into the aluminium

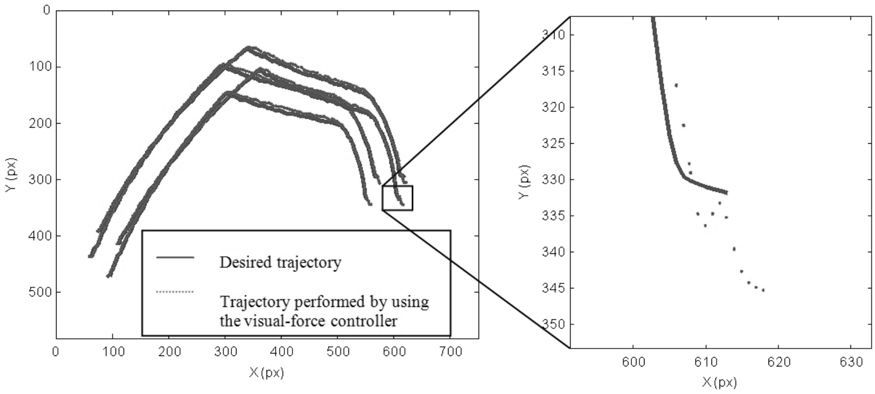

Fig. 4. On-line modification of the features in the image in an insertion task by using the visual-force controller.

holder. Figure 4 shows the desired image path and the path modified by the visual-force controller described in this section. The task can be accomplished thanks to the force-image interaction matrix which allows the robot to modify the desired image trajectory. The trajectory in the image space is recomputed on-line.

6. Robot-robot cooperation

Once the bar has been inserted, a screw must be inserted to join the new bar with the structure. Before the insertion of the screw, it is necessary to make coincident the hole in the structure and the hole in the tube. As it is previously described, one robot (robot 1) rotates the tube until its hole coincides with the structure hole.

In order to distribute tasks between the robots, a global planner is employed [15]. To perform these tasks, the global planner generates two tasks: "Detecting the hole" (T_1) and "Inserting the bolt" (T_2). The task T_1 is divided into two actions: "Location of the bar hole" (A_{11}) and "Rotating the bar to find the hole" (A_{12}). The task T_2 has only one action: "Inserting the bolt" (A_2). Once the actions to be performed are generated, the task planner has to distribute them among robots. Considering the tools available (both robots are equipped with a force sensor, the robot 1 has available a parallel gripper, a screwdriver and a vacuum; the robot 2 has available a Barret-hand and a range camera), the action A_{11} must be performed by the robot 2. To perform the action A_{12} both robots are required, the robot 1 to rotate the bar using the gripper and the robot 2 to locate the hole with the range camera. The action A_2 must be performed by the robot 1 because is the one that has the screwdriver.

The action A_{11} has to be performed previously to the tube insertion, because in other case the hole will not be visible to locate it. To locate the hole, its position is approximately known, using a CAD model of the workspace. With that information, the robot has to position the camera in front of the hole. According to the geometric restrictions, the trajectory planner determines the movements of the robot to maximize the visibility of the hole. Fig. 5 shows the sequence of images captured by the range camera along the movement of the robot. In that sequence the hole in the structure is located, maximizing its visibility. Initially, the bar is not visible at all in the image. With the movement of the camera the visibility of the structure is increased, improving the visibility of the hole that is the target of that action.

Once this action is done, the robot 1 has to insert the bar in the structure. After this, the bar must be orientated to achieve the correct visibility of the hole. While robot 2 holds the camera, the robot 1 must rotate the bar. These are the actions assigned by the task planner to each robot. If the hole is not visible, the bar must be orientated looking for the correct orientation of the bar to have the hole accessible for inserting the bolt. This last action is performed in a cooperative way, one robot is required to rotate the bar and other is used to control the range camera. Once the bar is properly oriented the robot changes the gripper for a screwdriver to insert the bolt in the hole [13].

7. Robot and human sharing the workspace

A human operator collaborates in the assembly task in order to add a T-connector at the end of each tube of the metallic structure. The operator will place the connectors because this is a difficult task to perform for the robots. Meanwhile, the two robots will place the tubes because they might be too heavy for the human. When the human approaches the metallic structure to perform this task, she/he may enter the workspace of the robots. Because of this fact, the system has to ensure the safety of the human operator by tracking precisely her/his location.

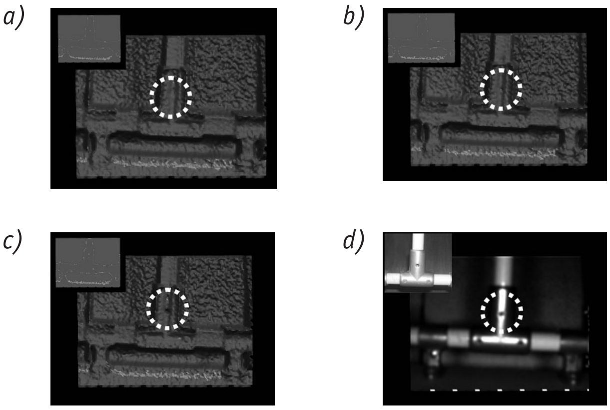

Fig. 5. Location of the bar hole. a) The hole is not visible range camera view, b) the hole start to be visible in the range camera view, c) the hole is visible in the range camera view, d) grey and real image of the hole.

An inertial motion capture system is used to avoid possible collisions between the human operator and the robots. This system is able to track all the movements of the full body of the human and it represents them on a 3D hierarchical skeleton (Fig. 6). Thereby, this system not only estimates the global position of the operator in the environment but it also determines the location of all the limbs of his/her body. Although this system registers very precisely the relative positions of the different parts of the skeleton, it accumulates an important error in the global displacement of the skeleton in the workplace. Therefore, an additional localization system is needed in order to correct this error.

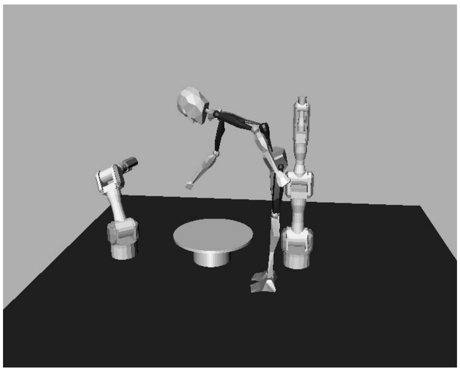

Fig. 6. 3D representation of the skeleton registered by the motion capture system. The other components of the environment (robots and turn-table) are also represented.

A UWB localization system is used to correct the global translational error of the motion capture system. The UWB localization system registers more precise global translation measurements but it has a smaller sampling rate (5-9 Hz instead of 30-120 Hz). The fusion of the global translation measurements from both tracking systems will combine their advantages: the motion capture system will keep a high sampling rate (30 Hz) while the UWB system will correct the accumulated translation error.

A fusion algorithm based on a standard Kalman filter [3] has been applied in order to combine the translation measurements from both trackers. The measurements from the motion capture system are introduced in the prediction step of the Kalman filter while the measurements from the UWB system are introduced in the correction step. Therefore, the prediction step will be executed with a higher frequency than the correction step. Each time a measurement from the UWB system is received, the correction step of the filter is executed and the transformation matrix between the coordinate systems of both trackers is re-calculated. This new transformation matrix is applied to the subsequent measurements from the motion capture system and thus their accumulated error is corrected. Between each pair of UWB measurements, several measurements from the motion capture system are registered. Thereby, the tracking system keeps a high sampling rate (30 Hz) which is appropriate for human motion detection. In Fig. 7 two types of movements for a human in the workspace are shown.

- A movement with linear displacement: This movement is used by human when he goes to the structure that is being built.
- A movement with rectangular displacement: This movement is used by human when he walks around the metallic structure in the workspace.

For each movement, the global translation of the human in the workspace is compute with the fusion of both systems: UWB and human motion capture. We can observe how the position obtained is better calculated when the fusion is employed.

The result of the fusion algorithm is a set of translation measurements which determine the global position of the human operator in the workplace. These measurements are applied to the relative measurements of the motion capture skeleton in order to obtain the global position of each limb of the human operator's body. The algorithm that controls the robots' movements will verify that the distance between each limb of the human and the end-effector of each robot is always greater than a specified threshold (1 m). When the human-robot distance is smaller than the safety threshold, the robot will stop its normal behaviour and will initiate a safety behaviour. The robot will remain still until the human-robot distance is again greater than the threshold. Thereby, collisions between the human and any of the robots are completely avoided and the human's safety is ensured.

8. Conclusions

In this paper a robotic system to assembly a metallic structure has been presented. An important aspect of the proposed application is the flexibility that provides the multisensorial system employed. These sensorial systems developed in our previous works are working in this application cooperatively in order to provide a high degree of flexibility. Furthermore, in order to successfully develop the task, it is necessary to work in the same workspace the human and the robot. To do so, in this paper an inertial motion capture system is used to avoid possible collisions between the human operator and the robots.

Furthermore, we have presented different ways to inspect different assembly tasks. We have used a time independent visual servoing system to guide robots. This system is independent of interruptions in the task which make lose the references to follow the trajectory. In addition, the visual servoing has been complemented with force control to correct the robot position.

On the other hand, we have also studied how the information provided from human capture system and UWB system determines the exact position of the human in the workspace to maintain the security distance between robot and human to avoid collisions. To do this, a method of data fusion based on Kalman filter has been used.

Finally, we have shown the utility to combine tasks of assembly and inspection between robots to performance some tasks. For example, the manipulation of a bar by a robot while another with a range camera detects the adapted position in the insertion task.

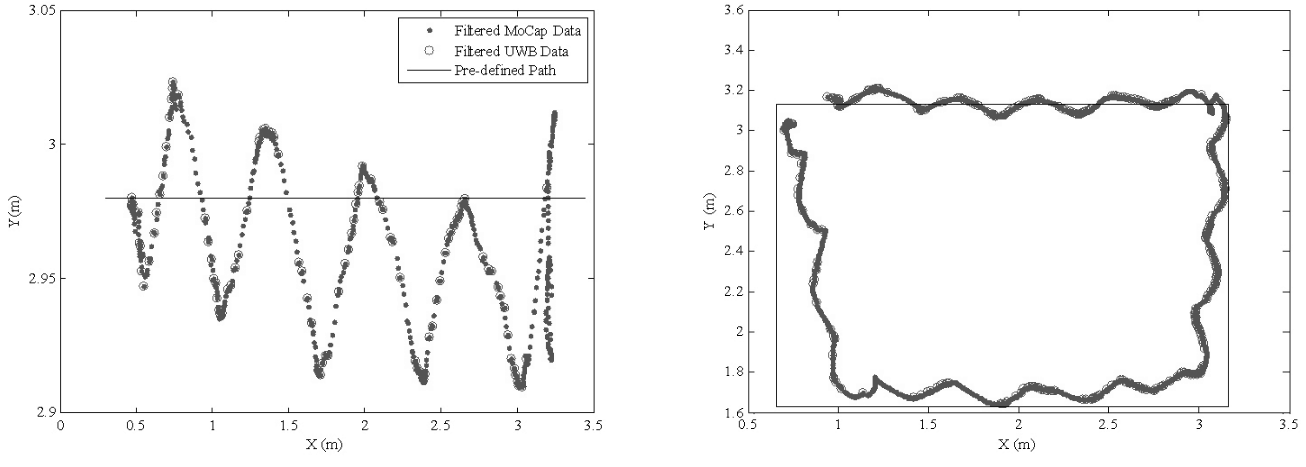

Fig. 7. Position estimates obtained with the fusion algorithm.

ACKNOWLEDGMENTS

This work was funded by the Spanish MEC project "Design, Implementation and Experimentation of Intelligent Manipulation Scenarios for Automatic Assembly and Disassembly Applications (DPI2005-06222)"and the FPU grant AP2005-1458.

AUTHORS

Jorge Pomares*, Pablo Gil, Juan Antonio Corrales, Gabriel J. García, Santiago T. Puente, Fernando Torres - Physics, Systems Engineering and Signal Theory Department, University of Alicante, PO Box 99, 03080, Alicante. Spain. Tel. +34 965903400. Fax. +34 965909750. E-mails: {jpomares, pablo.gil, jcorrales, gjgg, santiago.puente, fernando.torres}@ua.es

* Corresponding author

References

[1] Chaumette F. and Hutchinson S., "Visual Servo Control, Part I: Basic Approaches", *IEEE Robotics and Automation Magazine*, vol. 13, no. 4, 2006, pp. 82-90.

[2] Chaumette F., "Potential problems of convergence in visual servoing", *Int. Symposium on Mathematical Theory of Networks and Systems,* Padoue, Italy, 1998.

[3] Corrales, J.A., Candelas F.A., Torres F., "Hybrid Tracking of Human Operators using IMU/UWB Data Fusion by a Kalman Filter", In: 3rd *ACM/IEEE International Conference on Human-Robot Interaction*, Amsterdam, 2008.

[4] Dongming Z., Songtao L., "A 3D image processing method for manufacturing process automation", *Computers in Industry,* vol. 56, 2005, pp. 975-985.

[5] Garcia G.J., Pomares J., Torres, F., "A new time-independent image path tracker to guide robots using visual servoing", In: *12th IEEE International Conference on Emerging Technologies and Factory Automation*, Patras, Greece, 2007.

[6] Garcia, G.J., Pomares, J., Torres, F., "Robot guidance by estimating the force-image interaction matrix", *IFAC International Workshop on Intelligent Manufacturing Systems 2007*, Alicante, Spain, 2007.

[7] Gil P., Pomares J., Puente S.T., Diaz C., Candelas F., Torres F., "Flexible multi-sensorial system for automatic disassembly using cooperative robots", *Computer Integrated Manufacturing,* vol. 20, no. 8, 2007, pp. 757-772.

[8] Hutchinson S., Hager G.D., Corke P.I., "A tutorial on visual servo control", *IEEE Trans. Robotics and Automation*, vol. 12, no. 5, 1996, pp. 651-670.

[9] Kosmopoulos D., Varvarigou T., "Automated inspection of gaps on the automobile production line through stereo vision and specula reflection", *Computers in Industry,* vol. 46, 2001, pp. 49-63.

[10] Motai Y., "Salient feature extraction of industrial objects for an automated assembly system", *Computers in Industry,* vol. 56, 2005, pp. 943-957.

[11] Pauli J., Schmidt A., Sommer G., "Vision-based integrated system for object inspection and handling", *Robotics and Autonomous Systems,* vol. 37, 2001, pp. 297-309.

[12] Pomares J., Torres F., "Movement-flow based visual servoing and force control fusion for manipulation tasks in unstructured environments", *IEEE Transactions on Systems, Man, and Cybernetics-Part C*, vol. 35, no. 1, 2005, pp. 4-15.

[13] Puente S.T., Torres F., "Automatic screws removal in a disassembly process", In: *1st CLAWAR/EURON Workshop on Robots in Entertainment, Leisure and Hobby,* 2004.

[14] Son C., "Optimal control planning strategies with fuzzy entropy and sensor fusion for robotic part assembly tasks", *International Journal of Machine Tools and Manufacture,* vol. 42, 2002, pp. 1313-1335.

[15] Torres F., Puente S., Aracil R., "Disassembly planning based on precedence relations among assemblies", *The International Journal of Advanced Manufacturing Technology*. vol. 21, no. 5, 2003, pp. 317-327.