# DATA FUSION FROM MULTIPLE CAMERAS FOR AUTOMATIC DISASSEMBLY

**P. Gil , S. T. Puente , F. Torres, J. Pomares, F.A. Candelas**

Automatics, Robotics & Computer Vision Group (http://www.disclab.ua.es/gava/ing)
Dpto. Física, Ingeniería de Sistemas y Teoría de la Señal
Escuela Politécnica Superior - University of Alicante
Crtra. San Vicente, s/n, 03080 Alicante (Spain)
pgil@disc.ua.es

Abstract: In this paper we analyze, in some detail, the vision system architecture for disassembly applications. This work is carried out in the context of motion and stereo analysis. The methodology presented is useful for work in manufacturing conditions facing difficult situations like the occlusion of components.
The recognition and location of three-dimensional objects is important for automatic disassembly. A data fusion from a multiple-camera scheme has been proposed for extracting information from the scene. Data provided by some sensors is used to determine object recognition, location and orientation. *Copyright Ó 2001 IFAC*

Keywords: disassembly, robot vision, sensorial fusion, flexible automation

## 1. INTRODUCTION

It is useful to use a flexible manufacturing cell in disassembly work, if there is great customizability and adaptability in the case of different products and, if it is possible, in an automatic way and on the sort term. This is the only way that a certain cadence can therefore be guaranteed in this process and its automation is justified. (Torres et al, 2000)

A disassembly cell achieves great flexibility in a 2D-3D recognition and location system. This system has to be robust and versatile and must be able to carry out the tasks of recognition with the greatest of precision.

Object recognition by the human eye is simple and continuous. The human uses this recognition to interact with an unstructured environment, to move himself in avoiding obstacles, to use things and tools. If we want the robots to behave and interact in a similar way, we have to program them for the recognition of objects in a disassembly cell.

The problem of component recognition can be envisiged as a combination optimization problem, in which we look for the correspondence between component and image data or the correspondence between component and the set of properties and features of the objects. Afterwards, we shall describe sertain strategies for the recognition of assembled pieces.

Current research presents object recognition as a difficult problem to solve. In assembled systems, components are strongly connected to one another. Furthermore, such components are partially occluded and, therefore, non-visible to camera sensors.

Also, the recognition and location of objects from a camera image is difficult in real environments. The problem is even greater if the image is not composed of isolated components. For this reason, the acquisition system has to be able to distinguish and isolate certain characteristics in the image.

When there is a group of similar component in the scene, the occlusions are more important (Boshra and Ismail, 2000). In such a case, we need a classification of the characteristics of the components to help us to discriminate, when, for example, some component features are completely hidden, some component features are distorted because the components are partially visible,

and/or the components are hidden due to its three-dimensional shape.

The presence of shadows is another problem. Shadows can imply an added difficulty for the recognition of components in assembled systems. Shadows cause confusion in defining borders, the consequence of which is a confusion between the component's edges and its connections with other assembled components. This, therefore, makes it important to introduce information like color continuity and textures.

The remainder of paper is organized as follows. In Section 2 we describe recognition and location techniques. A vision system and the envirnoment in which it functions are discussed in Section 3. In Section 4, we present the performance of the proposed system. Finally, in section 5, an example of disassembly is presented.

## 2. RECOGNITION AND LOCATION TOOLS

The tools used are based on a system of three-dimensional object recognition and location, by means of computer vision techniques. It consists of identifying one or more objects in the image, in accordance with their database models. The object-models in the database may also be from three different sources: A CAD-CAM model, a prototype image of the component or a geometric shape.

### 2.1 Recognition

Three-dimensional object recognition is based on the formation of correspondences between image characteristics and object models. Typical associated stages for recognition are (Flynn and Jain, 1994):

- Indexing: which consists of the selection of a set of candidate models, which are used in the following stage. At this stage any models that are not present in the scene are rejected.
- Matching: the selection of a model from the candidates, as well as the set of correspondences between the image and model's characteristics, all of which composes a concrete hypothesis.
- Location: a rotation and a transfer are estimated, which, on application to the model's characteristics, allow us to recognize the object in the image.
- Verification: the final stage. This allows us to verify the goodness of fit of the recognition. One has to decide what characteristics of the sensor are going to be significant for the recognition of

objects in the image. Invariant descriptors are used when they represent object characteristics, independently of the point of view.

In disassembly applications it is necessary to recognize and analyze the different partially occluded components. In addition, the perfect location of all the components is also necessary. As such, we achieve the correct execution of a disassembly strategy.

Some strategies for the recognition of objects are now described:

- Projective invariants (Faugeras, 1993; Long Q, 1995; Roh and Kweon, 2000; Suk T and Flusser J, 2000): the current trend is to use projective geometry as a tool for recognition in computer vision, for two main reasons. First, projective geometry provides very useful projective invariant characteristics for recognition. Secondly, The pinhole camera model is basically projective, since neither the distance, the angles nor the parallelism are maintained. The invariant descriptors do not change with any transformation.
- Features vector: This is a straightforward technique. Some of its features are color, area, perimeter, shape, texture, and so on. This sort of strategy has too many limitations and it does not consider the space relationships.
- Other strategies: Photometry adjustments to the models; whose basic idea is the Hough transformation, which allows us to identify geometric primitives like, for example, lines in images. Correspondence of models with symbolic data structures. The detection of silhouettes and shapes, which consists of segmentation techniques based on CSS (Curvature Scale Space).

There are many strategies for seeking correspondences between real assembled component and data model in the database. The optimal solution consists of using a combination of all of the forgoing techniques.

### 2.2 Location

Three-dimensional object location techniques can be found in the literature, and are of two different categories, (Flynn and Jain, 1994) either two-dimensional or three-dimensional primitives.

Two-dimensional primitive techniques are very diverse: i.e., the extraction of color and texture features, using

DCT (Zhong and Jain, 2000); the extraction of object shapes from the histogram of edges and angles.

Three-dimensional location is established on the basis of structural 2D characteristic groups and their correspondences with candidate structures of 3D models. Therefore, the correspondence integrates the transformation perspective with the model and its image. Huttenlocher and Ullman (Huttenlocher and Ullman, 1987) show these sorts of location systems.

Three-dimensional information lost in an image is solved partially from structural relations between the different object characteristics, and the projective transformation constrains. In many cases, these restrictions do not give rise to only one solution.

Primitive three-dimensional techniques show important advantages due to the greater geometric information incorporated, facilitating the location process. Flynn and Jain (Flynn and Jain, 1991) describe, in their BONSAI system, an algorithm to estimate the position and the rotation transformation.

*2.3 Objectives*

The main objective of this research is to develop a disassembly system and an image analysis system. It is able to locate and recognize components in enviroment of a real disassembly system.
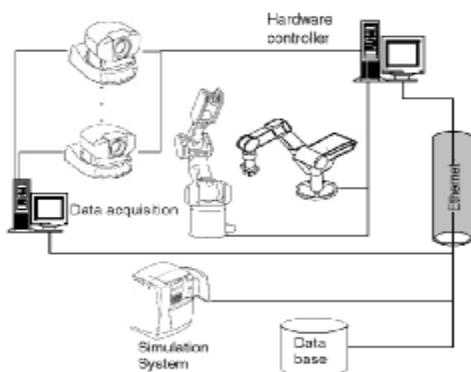


Fig. 1: System scheme.

This implies that the recognition and location system fulfills requirements, such as:

- The system has to be able to recognize all of the different assembled components.
- The behavior of the method has to be robust in the face of the problems mentioned previously and in the presence of adverse conditions like non-

structured light, shadows, variable light conditions, noise, occlusions, and so on.
- The recognition module has to provide a location in world coordinates that allows the robots to find the component to be disassembled.
- The entire system must operate in real time to accomplish a minimum of delay. On the other hand, the entire system has to be completely automated.

## 3. SYSTEM STRUCTURE

The system used in the application is made up of the elements shown in figure 1.

The architecture is distributed within a Local Ethernet Network of 100 Mbs. Each process unit carries out the specific tasks for which it is dedicated.

The simulation system is the heart of the application. It makes, not only a previous simulation of all the operations, but also the control of the on-line application. This system consists of a SGI Octane R12000. It defines the robot's trajectories using high level information from the data acquisition unit.

The hardware controller is dedicated to the generation of low level commands for the robot's trajectories, as well as controlling the parameters of the sensorial system (zoom, convergence, pan, tilt, etc).

The aims of the data acquisition unit are tasks of visual servoing by robots and recognition and location of the different components from the product to be disassembled. To do so, the controller compares the database information with data acquiered and processed a low level, from the images captured by the cameras.

The knowledge base is implemented in the database. All the elements are connected to the rest of units through the Ethernet Network.

*3.1 The cell used.*

Generic disassembly cell normally has the following components (Fig 2):

- One or several robots to conduct the different operations.
- One element to exchange tools for each robot. The recognition module has to provide a location in world coordinates that allows robots to find the component to be disassembled.

- An ample set of tools (screwdrivers, keys, cut elements, drill, clamps...).
- A work-table that can be customized. It afford the holding of a great amount of products of different kinds and sizes.
- Sensors: artificial vision, force, presence, rank.
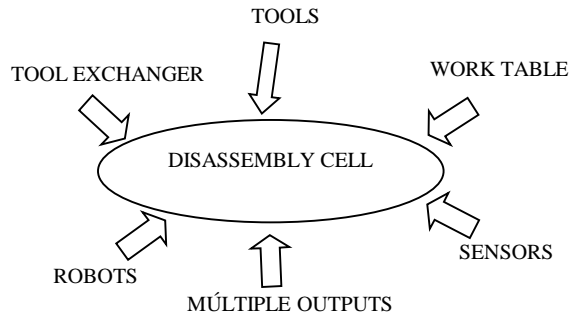- Multiple outputs for the different components of the product.



Fig. 2: Parts of a disassembly cell.

The automatic disassembly cell consists of the following components:

- PA-10 robot equipped with 7 degrees of freedom (Fig 3a). Robot PA-10 (MAN-ROBOT) is used for manipulating the components to be disassembled.



Fig. 3a: PA-10 Robot    Fig. 3b: Scorbot ER-IX
Robot

- Scorbot ER IX Robot equipped with five degrees of freedom (Fig 3b). The Scorbot robot (VISUAL-ROBOT) is dedicated to guidance tasks.
- Tools (clamps, etc) to separate the components from one another, once the connections have been eliminated. Tools to remove fasteners (screwdrivers).
- Computer vision sensors formed by an acquisition system.

*3.2. Computer vision sensors*

The recognition and location of components depend on the architecture of the system of image acquisition. If we wish to obtain a disassembly cell, it is important that it works in real time. Not only the optics of the camera but also its movement, vary automatically according to the sensorized enviroment's necessities and the variations that take place in it during the disassembly of the different components. This is due to the disappearance of the fasteners or of the orientation of components that are still assembled.

The camera system is structured like a dynamic cooperative mode (Brooks and Iyengar, 1998) (Fig. 4). It consists of a stereo-pair that allows us to capture three-dimensional information from the elements to be disassembled. (Point 1 of Fig. 4). The stereo-pair is made up of two EVI-D31 cameras. (Fig. 5). The active vision system can actively control its parameters, such as position, orientation, focus, zoom, vergence angle, tilt angle and roll angle.
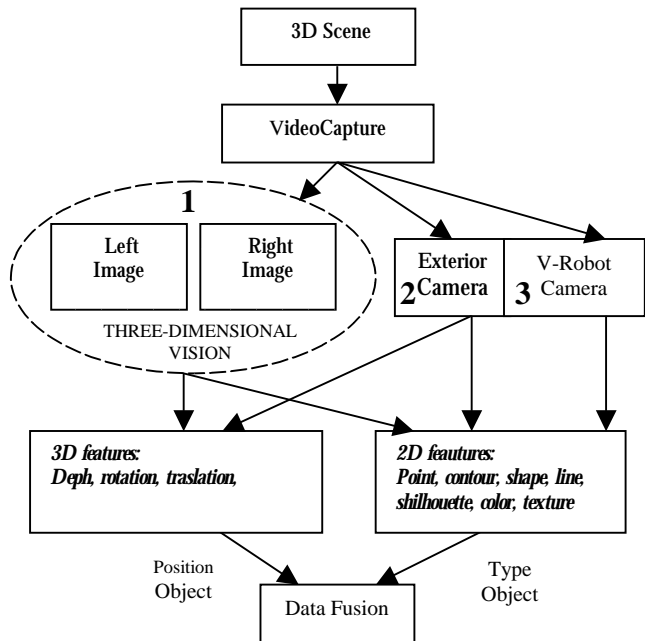


Fig. 4: Sensorization system

As additional acquisition elements, two extra cameras are used. A Sony CV-950 fixed camera is used as an external camera (Point 2 of Fig. 4), and. a JAI-M536, which is a remote micro-head camera system. (Fig. 6). The small size and weight of the JAI-M536 allows us to put it on the end of the VISUAL-ROBOT with an illumination system (Point 3 of Fig. 4).

Fig. 5: Stereo-pair



Fig. 6: JAI-M536 Camera

The acquisition system is completed with a pre-processing system and it is composed of an image acquisition and a processing board (IAPB) MATROX-GENESIS model which uses Matrox Imaging Library (MIL).

The sensorial fusion is made by combining the data from several cameras to improve the accuracy of the information. (Fig. 4).

## 4. METHODOLOGY

The system's operation can be described in three steps:

1.- Recognition and location with the stereo-pair and with the fix ed camera located outside.
2.- Orientation and position control of the MAN-ROBOT.
3.- Visual Servoing and tracking.

### 4.1. Information from multiples views

First, the stereo-pair and the external camera capture images to be disassembled. These images along with the available information in the database allows us to determine the first disassembly strategy.

Once the product to be disassembled is known, the components in the image are focused. That is to say, interesting regions in the image are limited according to the database information about the element to be disassembled, using mainly pre-processing and segmentation techniques with the help of the IAPB Matrox-Génesis.

When the system knows which component has to be disassembled, after recognizing and locating it, the acquisition system obtains information about the fasteners.

In such case, the system begins to apply pre-processing and segmentation techniques which allow us to identify the components suitably. At this stage we use the VISUAL-ROBOT camera to get a better approach to the fasteners.

### 4.2 MAN-ROBOT, control and orientation

Based on the data obtained from multiple views of the scene, and after extracting the features that we need, the movement strategies that the robot does to manipulate a component of the scene are considered.

From the data fusion, 3D-position of the MAN-ROBOT and the position of the component to be disassembled are estimated.

The objective of the VISUAL-ROBOT is to guide the MAN-ROBOT and it is the tool to control the correct execution of an operation (Tonko and Nagel, 2000) providing information that is impossible to obtain from the rest of the cameras.

### 4.3 Visual Servoing and tracking

Integration of a robotic arm and stereo vision system are quite often used in tracking systems for mobile objects in a three-dimensional space (Brooks and Iyengar, 1998). Our system uses this kind of technique to achieve control and tracking of the disassembly operation at every moment.

The manipulation operation is controlled in real time from the movement of the MAN-ROBOT and the movement of the current disassembly component (input). The control and tracking system provides Cartesian coordinates (output).

We know the position of the VISUAL-ROBOT camera in the world (3D-coordinates). This position is calculated by the combination of the information that it is extracted not only from the stereo-pair but also from the external camera.

Futhermore, we know the positions of the MAN-ROBOT and the disassembled component at every moment.

Using the relative VISUAL-ROBOT position in contrast to the relative MAN-ROBOT position, we correct the position of the former by catching the information that we needs in the manipulation (feedback).

## 5. DISASSEMBLY OF A REMOTE CONTROL

By means of the fusion of information from the different images shown in figure 7, the identification and recognition of the different components is carried out (Fig. 8). Using this information as well as the database, the components are identified and reconstructed.



Fig. 7: Images obtained by the pair and the external camera.

According to the different disassembly components, the information is updated by the system to improve the product's disassembly.
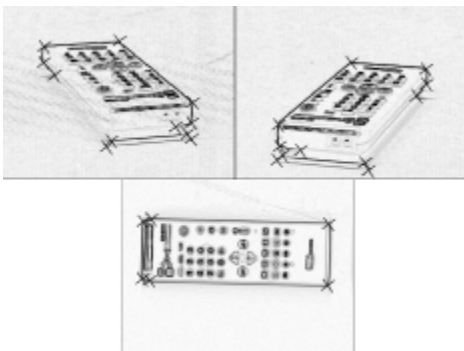


Fig. 8: Matching of the images in figure 7

That matching is repeated for each component and then the disassembly sequence of the product shown in at figure 9 is obtained.
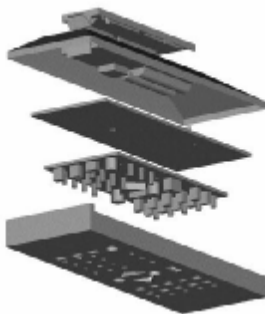


Fig. 9: Disassembly of a remote control

## REFERENCES

Boshra M., Ismail M.A. (2000). Recognition of occluded polyhedra from range images. *Pattern Recognition* **V. 33**, pp. 1351-1367.

Brooks R.R., Iyengar S.S.. (1998). Multi-Sensor Fusion: Fundamentals and Applications with Software *Ed. Prentice Hall PTR*.

Faugeras, O.D. (1993) Three-Dimensional Computer Vision: A Geometric Viewpoint. *MIT Press*.

Flynn, P.J., Jain, A.K. (1994) Handbook of Pattern Recognition and Image Processing: Computer Vision, *Academic Press, San Diego* **, V. 2**.

Flynn, P.J., Jain, A.K., (1991). BONSAI:3D object recognition using constrained search, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pp. 1066-1075.

Huttenlocher, D.P., Ullman, S. (1987). Object recognition using alignment, *Proc. Int. Conf. Comp. Vision, 1st, London*, pp. 102-111.

Kyong Sig Roh, In So Kweon. (2000). 3-D object recognition using a new invariant relationship by single-view. *Patter Recognition* **V. 33**, pp 741-754

Long Q. (1995). Invariants of Six Points and Projective Reconstruction From Three Uncalibrated Images. *IEEE Transactions of Pattern Analysis and Machine Intelligence*. **V. 17. No 1**. pp 34-46.

Suk T., Flusser J. (2000). Point-based projective invariants. *Pattern Recognition* **V. 33**, pp. 251-261.

Tonko M., Nagel H-H. (2000). Model-Based Stereo-Tracking o fNon-Polyhedral Objects for Automatic Disassembly Experiments. *International Journal of Computer Vision* **V. 37 I.1**, pp. 99–118.

Torres F., Puente S.T., Gil P., Candelas F.A., Ortiz F.G.. (2000). MOGEDA: Módelo Genérico de Desensamblado Automático. *Proc Jornadas de Automática, Sevilla (Spain)*.

Zhong Yu, Jain A.K. (2000). Object location using color, texture and shape. *Pattern Recognition* **V. 33** pp 671-684.