Proceedings of the 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems Yokohama, Japan July 26-30, 1993

AN APPROACH TO DISASSEMBLY PROBLEMS IN ROBOTICS

Paolo Dario and Michele Rucci

ARTS Lab
Scuola Superiore S. Anna
Pisa, Italy

Abstract

~~*Disassembly and recycling are becoming important factors in a society that, like the present one, is more and more concerned about the environmental impact, rather than only the economic aspects, of manufacturing. The traditional approach to automation and robotics has always been focused on assembly problems, and has almost neglected the problem of managing the manufactured products at the end of their life cycle. Thus, the problem of disassembly offers tremendous opportunities for robotics application and also for theoretical and speculative robotic research.*~~

~~In this paper we~~ The authors present motivations for the long term research program on the various aspects of disassembly ~~that we have~~ undertaken in ~~our~~ their laboratory, and outline a theoretical framework for approaching disassembly problems. ~~Finally,~~ an application is described which is derived from such a framework in the context of a specific case-study of disassembly.

1. Introduction

The concept of assembly has become important in parallel with the beginning of the Industrial Revolution in the last half of the XVIII century. The financial driving forces for manufacturing goods in large quantities and at low cost have encouraged the development of automated machines and, ultimately, processes for assembly tasks. In the last decades, the extrapolation and the generalization of some of the concepts and technologies associated with industrial automation have contributed to the growth of modern Robotics and of some aspects of Artificial Intelligence.

Recently, different priorities have emerged, that "complement" those associated with assembly. In contrast to the past ones, these new priorities are related to the concept of *environmental conscious manufacturing*, and, even more, to the notion of *disassembly*, and to such tasks as dismantling and recycling.

The economic implications of the problem of disassembly, in a world like the present one which is increasingly more sensitive to economic wastes and to the possibilities of recovering raw materials, are far too evident. In fact one could even speculate that a systematic development and exploitation of disassembly methodologies cou[illegible] the future to the birth of a sort of *"inverse" factory* that, by operating in symbiosis with the more traditional manufacturing factories, will close the open loop production cycle (design - assembly - use) and might eventually affect some macro-economic aspects of the industrial world. As depicted in Fig.1, the inverse factory could lead to the full implementation of the concept of the so called *"integrated design"*, that is the design of products that are conceived not only to be assembled and used, but also to be totally disassembled and recycled.

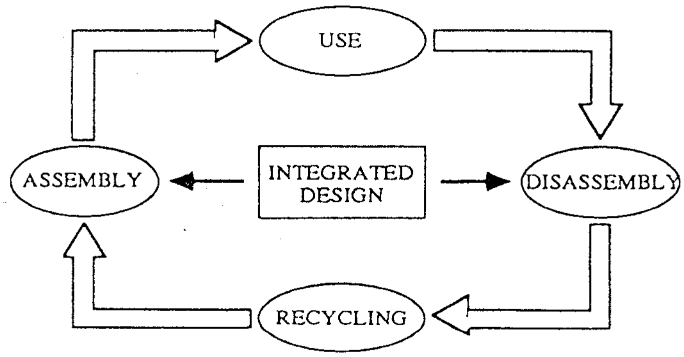

Fig.1: Concept of integrated design and the loop closed by disassembly and recycling.

0-7803-0823-9/93/$3.00 (C) 1993 IEEE

This concept (whose implications on the present industrial structure are so profound that its implementation might bring to a sort of "second" industrial revolution) implies the analysis of problems of various kind:

(a) financial,
(b) logistic,
(c) technical,
(d) theoretical.

In this paper we address preliminarily some issues related to problems of type c) and d). The basic difficulties which arise in the development of disassembly systems are described in Section 2. In Section 3 we present a theoretical framework that we intend to use as a guideline for future work. Finally, in Section 4 a preliminary implementation of such theoretical concepts is illustrated for a specific case-study.

2. The theoretical problem of disassembly

A crucial difference occurring between disassembly and assembly tasks is that in the former ones objects have had a past "life", which could have modified them unpredictably both in their appearance and in their internal structure. As a consequence, in general, a disassembly system should deal with environments characterized by a much higher degree of uncertainty than for assembly tasks. Integrated design is a powerful tool for reducing such uncertainty: if the features of a specific product are properly designed, a dedicated disassembly system can actually work in a partially structured environment. However, integrated design is not always possible for various reasons. For example, integrated design efforts are useless with those goods (such as common garbage) which typically undergo large modifications during their life.

In our laboratory we have recently focused our attention on the very general problem of disassembly, i.e. without any specific *a priori* knowledge of the environment. In this case, disassembly involves a large number of extremely complex problems, many of which are still unsolved. To the best of our knowledge, no systematic approach has been proposed so far to the general problem of disassembly.

A general disassembly system should be able to single out the parts of the objects (or the materials) of interest dealing with the simultaneous presence of many objects, each with its unknown shape. Thus the system should be able to correctly understand the world and to interact with it efficiently. Unfortunately, the state-of-the-art in those disciplines related to perception is not advanced enough to allow the development of general disassembly systems. Researchers involved in the field of perception, in particular visual perception, have put aside the final goal of recognition in order to focus on the analysis of the separate modalities by which sensory signals are processed.

Nevertheless, some applications have been proposed which for the purpose of our work can be included in the class of "disassembly problems" [1], [2]. Some of these systems have a large *a priori* knowledge of the environment, which let them operate in the presence of a certain degree of uncertainty. For example, the systems described in [1] concern the sorting of parcels and the cleaning of refectory trays.

3. The proposed theoretical framework

Our approach to the general theoretical problems of disassembly is based on several considerations which are described in the following. Most of such considerations are related to the state of art of the disciplines involved and their limitations, others have emerged from our own research activity at the ARTS Lab. The main notions that must be considered concern the following basic system requirements:

(i) inter-sensorial data fusion,
(ii) intra-sensorial data fusion,
(iii) learning and adaptability,
(iv) simultaneous use of Artificial Neural Networks and symbolic techniques.

Inter-sensorial data fusion concern the fusion of data available with different sensorial modalities: the simultaneous use of more than a single sensorial modality could result in a great simplification of perception goals. Things that are difficult to understand with one sense can be obvious with others and all the senses should cooperate to the final understanding of the scene. In this context very important are also sensorimotor procedures aimed at gathering further information from the scene by actively modifying it. Recently the use of sensorimotor procedures has become important in combination with the active perception paradigm [3] and with its sensorial diversification such as active vision [4], active touch [5], and so forth.

Our interest for intra-sensorial data fusion derives from the fact that typically, due to the high specialization of processing algorithms, only a small part of the total amount of information present at the input can be efficiently elaborated [6]. The remaining part of the information not only is neglected, but acts as noise. Furthermore, these algorithms are not conceived for working in cooperation with others, so that their integration into a single general-purpose algorithm is extremely difficult.

A very interesting consideration concerns the advantages of allowing the system to learn during its operative life. In general "learning" implies an increment of the stored knowledge, that is the inclusion in the memory of new models, or parts of models, or the modification of those already acquired. A different aspect of learning is the discover of new exploratory strategy which can be used to understand the scene. Both are appealing when working in unstructured environments, in presence of a high degree of uncertainty. In fact, by means of learning procedures the system can adapt itself to deal with variable conditions. In order to make learning possible, the set of representations of the world (the memory of the system) should be organized in a simple and organic manner. In particular a pseudo-symbolic representation, with a one-to-one correspondence among models and percepts is suitable for easy update. A related problem regards the amount of information that should be included in the system at the beginning, which is opposed to what will be learnt after. It should be possible to include in the memory of the system notions concerning the various parts which build up a standard object. In this way the separation of the parts is simplified afterwards. Of course a great flexibility is required, so as to treat objects with lacking parts and altered features.

A final consideration relates to the sort of scepticism that has surrounded the field of Artificial Intelligence in the last decade [2]: on one side, symbolic processing has shown to be suitable only for very high level tasks and to fail in low-level perception tasks; on the other side, Artificial Neural Networks (ANNs) techniques [7], as they have been generally used up to now, have proved to be successful only in applications were it is possible to find out a set of prototypical patterns. We believe that both these techniques can be useful if applied at different levels of processing. In particular the cooperative/competitive aspects of ANN is important in overcoming the aforementioned data fusion problems.

In order to address disassembly problems consistently with the state of the art of perception, we propose - based on the above considerations - a very general theoretical framework which we want to use as a guideline for future work. As will be shown in the following, this framework can be used to devise learning robotics systems capable of fusing inter and intra-sensorial data, which can be implemented by both ANNs and symbolic techniques.

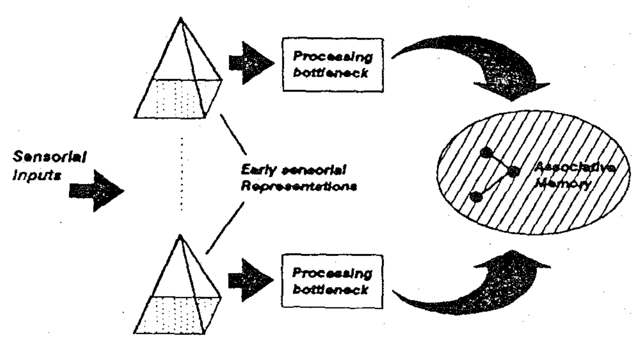

Fig.2: The proposed theoretical framework. Different sensorial information from the pyramids is elaborated recursively through the processing bottlenecks in order to produce memory activation.

An important contribution to the proposed framework derives from a possible theoretical scheme of the human visual system which has been recently developed in the field of visual perception [8]. The system is depicted in Fig. 2. The scheme includes a set of multiresolution multifeature pyramids [9] and an associative memory in addition to a high level task-dependent planning module (not shown in the figure). The pyramids are as many as the sensorial modalities accepted by the system and they process the input signals so as to produce a representation of the data with different levels of resolution and with different features. The associative memory is a massive parallel system including models of the world linked in associative or inhibitory manner with the others. A model is composed of the representations of its significant parts and features. Each sensorial channel elaborates at every time a small fixed amount of information (the processing bottleneck) and produces activation in the associative memory. Due to the bottleneck, a trade-off is established between level of resolution and spatial and/or temporal width of the examined sampled signal. Perception is accomplished by means of many iterations each one including sampling of information and memory activation.

Data fusion is accomplished at two levels: in the multifeature pyramids data can be organized by letting different intra-sensorial features cooperate; in the activation of the memory, due to the many steps imposed by the processing bottlenecks, inter-sensorial and intra-sensorial data fusion is carried out. The structure of the memory is also suitable for future updating of the stored knowledge, so as to implement learning.

A crucial point concerns the control of successive samplings of information from the pyramids, that is, where (at which level of resolution and at which spatial position) at a given time should processing be focused on. The mechanisms by which this is accomplished are equivalent to attention mechanisms in humans and they depend both on bottom-up and top-down processing. Bottom-up mechanisms act in a quite reflexive manner, drawing attention to salient input cues [10]. High level mechanisms are based on the activation of the associative memory: for example, the activation of a part of an object model can force the system to look for the other parts of this model in order to assess if the object is really present in the input scene. An exploratory strategy of the scene (which in our terms is an exploratory strategy of the pyramids) is carried out by the planning module with symbolic reasoning based on both the salient cues and the memory activation, mediated by the task at hand. This results in an adaptive (not fixed) exploration path which changes both with the examined scene and with the history of the system.

4. An applicative example: the shopping-bag case

The proposed framework for perception has been implemented and tested preliminarily for the case of a specific application. The problem we considered is the classification of common objects that could usually be found in a shopping bag case. Due to the virtually illimitate number of sizes and shapes that objects can possess, this is an extremely complex problem and is well suited for testing our approach to disassembly. Actually, the shopping-bag case is extremely significant in the perspective of investigating disassembly since most of the common garbage is composed of shopping products, after use. The scenario of the experiment is shown in Fig. 3:

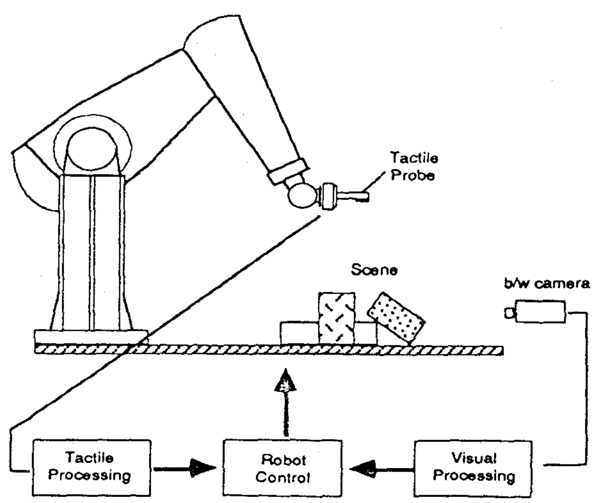

Fig.3: The shopping-bag problem scenario.

Several objects are located in front of a robotic PUMA 562 manipulator, in the visual field of a b/w TV camera. The end effector of the arm contains a multifunctional tactile probe developed in our laboratory [11], which is a composite fingertip incorporating multiple sensing capabilities (see Fig.4). In its present version the probe possesses a 16x16 piezoresistive array sensor, a three element dynamic sensor for determining surface roughness, and a thermal sensor. A black coat is laid on the table under the objects and illumination conditions are such as to avoid strong shades and sharp reflexes. A calibration procedure which makes use of six fixed points [12] is used in order to find the correspondence between arm and camera reference systems.

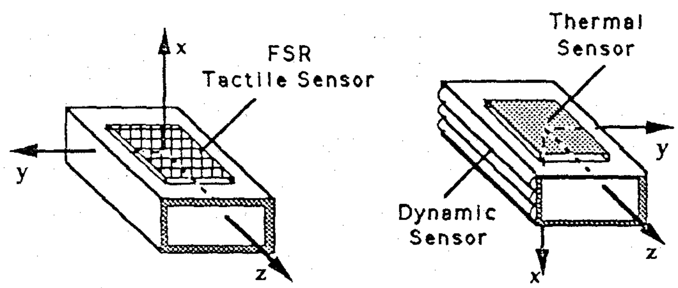

Fig.4: The multifunctional tactile probe. Static and dynamic tactile signals are produced along with thermal signals.

Let us consider first the case of a single object in the scene; as a first application, we have focused our attention on the inter-sensorial data fusion and on the ANNs-symbolic techniques integration requirements. A possible way to generalize these results to the multiple-objects case is described later.

We have recently developed a system architecture for the visual recognition of stationary objects [13]. This system is able to classify an examined object, which could have different orientations, as a member of a set of a priori defined class. The technique is applicable when the object is centered on the observed image and if it is the only object present in the scene. In the present application we have extended this system architecture so as to include also tactile information and we have implemented it by means of artificial neural networks techniques. The system architecture is shown in Fig.5.

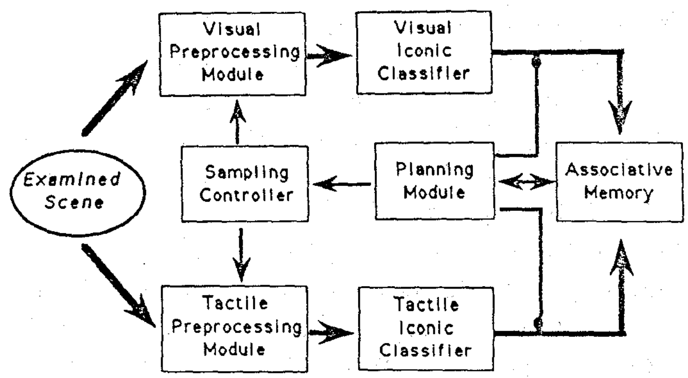

Fig.5: The system architecture derived from the theoretical framework for the shopping-bag case study.

The architecture consists of two pyramids, a visual pyramid and a tactile pyramid, both including the representation of a single feature: edges in the visual pyramid and roughness in the tactile one. Actually, the pyramids are not calculated explicitly, but their parts of interest at a given time are produced by examining the scene by means of two preprocessing modules: a Visual Preprocessing Module (VPM) and a Tactile Preprocessing Module (TPM). Both the modules gather information from different parts of the scene and at different levels of resolution, producing at every time a fixed dimension image (*visual icon*) in the visual channel and a fixed dimension vector (*tactile icon*) in the tactile channel. The VPM extracts the edges of the analyzed part of the scene and reduces the resolution of the obtained image; a Fast Fourier Transform of the signal is carried out in the TPM, and the resulting data are resampled with a wider step.

Each object to recognize is represented in the associative memory with a set of units acting as *grandmother cells* [14], organized in two layers. In the first layer (the feature layer), each unit represents a part of the object or a tactile feature. In the second layer (hypothesis layer) a composite object is represented with a single unit connected with the previous ones in an excitatory manner. A memory representation of an object is shown in Fig.6.

Each visual unit in the feature layer is activated when the information sampled by the VPM is centered on the part that the unit represents and the size of the sampling area is such to enclose

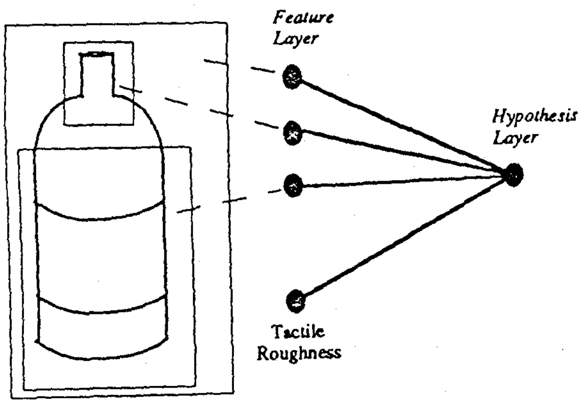

Fig.6: The representation of a bottle in the associative memory is composed of four units in the feature layer and of a single unit in the hypothesis layer. The visual feature units are activated when the corresponding part is examined, while the tactile unit is activated when the expected roughness is found.

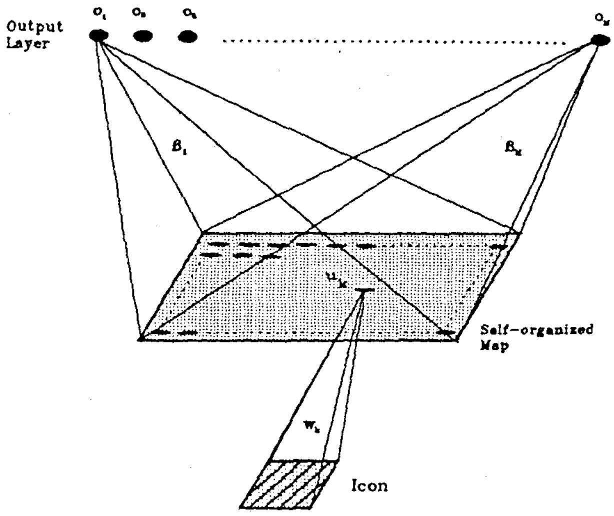

Fig.7: Scheme of the Visual Iconic Classifier. The input icon is properly classified so as to produce activation for the feature layer units of the memory (see text for details).

completely that feature. In a similar way a tactile unit in the feature layer is activated when the previous conditions are met and the roughness of the object in that point is close to the expected roughness. The units in the feature layer act as accumulators storing and cumulating the activation provided by the processing channels. The units in the hypothesis layer are linked in a Winner-Take-All fashion [15] so that at every instant only one of them is activated. The activation of one unit in the hypothesis layer is equivalent to the formulation of an hypothesis on the identity of the observed object. Recognition is intended as the eventual determination of a winning unit.

The activation of the units in the memory is provided by the two iconic classifiers (see Fig. 5) that process the incoming icons. As illustrated in Fig. 7, the Visual Iconic Classifier (VIC) is a three layer ANN whose outputs are the inputs for the units in the feature layer of the memory.

The first layer of the net is trained as a self-organized topological feature map [16] [17], so that similar input patterns activate units close to each other in the bidimensional topology of the layer. Only the weights of the first layer are evaluated with training, while the second layer weights are determined deterministically so as to code the activation of the topological feature map to the feature units of the memory. The learning procedure of the self-organized topological feature map is an unsupervised training process which requires the recurrent presentation of the patterns of the training set to the net. The weight vectors of the maximally responding unit and of the units in a neighborhood of it are changed toward the input vector. The size of this neighborhood and the parameters regulating the weight modifications decrease during training. The main difference between the first layer of the VIC and a topological feature map is the lack of intra-layer connections in the output layer, which is responsible for the absence of blobs of activation.

The Tactile Iconic Classifier (TIC) is a four layer feed-forward ANN which has been trained with the back-propagation algorithm. The back-propagation algorithm [18] has been preferred over the bidimensional self-organized topological feature map as in the visual channel because performances of the network were already very good. The network has been trained to classify different roughness of the objects; then each output is linked to a corresponding tactile unit of the feature layer.

The connections between the output units of the VIC and of the TIC and the feature layer units have binary weights which change during the recognition process. Their values are set by the planning module so that only the feature layer units corresponding to expected features are allowed to receive activation; that is, when the system is planning to examine a certain feature of the object, only the connections with the unit representing that feature are set to one, while all the others are set to zero.

The operation of recognition begins by examining the overall object at very low resolution. In this way a set of hypotheses on the identity of the observed object is formulated and the visual and tactile features of the winning hypothesis are sequentially examined. Recognition is achieved when the external activation (that is the activation provided by the feature units) of the winning hypothesis unit is greater than a predetermined threshold. If an initially winning unit is not confirmed by successive samplings of information the planning module resets all the units of the feature layer connected with that hypothesis unit so that the winning hypothesis changes.

The location within the input scene where to focus attention by allocating visual or tactile sampling of data is carried out by a Sampling Controller (SC). The SC is composed of a set of feed-forward networks trained with the back-propagation algorithm. There is a net for each unit in the feature layer, but at a given time only one is activated by the Planning Module. Each network accepts as input the first visual icon (i.e. the overall object analysed at lowest resolution) and provides the coordinates and the extension of the selected area of interest in the camera image.

In the general shopping-bag case, objects are not centered in the image, nor an image contains a single object. The first problem could be solved by accurate sampling from the pyramids so as to center the gathering of data on the real object position. A low level attention mechanism based on the concentration of the edges can easily provide this position by pointing out the presence of a target in an area of the visual field. This attention mechanism can estimate also the size of the target, i.e. the resolution at which the object should be examined.

Dealing with several object at the same time is more difficult. In principle, several techniques can be used, including active movements of the point of view (that is the camera could be moved until a point of view is found were the objects are separated) are possible. Our approach to this problem involves the use of "pushing" procedures [19], that is the **active** interaction of the robot with the scene by moving the objects. If recognition is not performed by vision only, i.e. if none of the hypothesis is activated, the robotic arm pushes the presumed agglomerate of objects and the optical flow field is examined in order to assess if more than one object is present and to separate them. Many of the algorithms relative to optical flow are limited to rigid objects. Tactile information could be used to verify if the hypothesis is valid.

5. Conclusions

We have outlined that many factors which contribute, in our opinion, to make disassembly an extremely promising field of investigation for robotics research. Nevertheless, tremendous difficulties are found in the development of general disassembly systems. In this paper, which is the first of a series which will describe our research efforts on various aspects of disassembly, we have discussed main motivations and problems, as well as our theoretical and applicative approach. At present our research is moving along several directions: on one side, we are including new sensory inputs in the architecture described in Section 4, such as different tactile (tactile pattern, temperature and hardness) and visual (regions and motion) features so as to meet also the intra-sensorial data fusion requirement; additional sensing inputs, such as senses of smell and hearing, will also be added. Furthermore, the system is being extended to the general case of many overlapping objects present at the same time in the scene. According to this goal we are implementing some low level attention mechanisms, and some new explorations strategies as well as on line learning procedures by means of ANN techniques. Finally, exploratory procedures based on active vision and touch, and on active manipulation are planned to be incorporated in the future.

Acknowledgements

This work has been partly funded by the Special Project on Robotics of the Italian National Research Council and by MURST. One of the authors (M. Rucci) has been supported by a fellowship from *Istituto per la Ricostruzione Industriale*. The contribution of Claudia Guadagnini and Cecilia Laschi is also gratefully acknowledged.

References

[1] D.Viescher "Cooperating robot with visual and tactile skills", Proc. IEEE Int. Conf. on Robotics and Automation pp. 2018-2025, Nice,1992.

[2] R. Bajcsy and C. Tsikos "Assembly via disassembly: a case in machine perceptual development", Proc. 5th Int. Symp. Robotics Research, Tokyo, August 1989.

[3] R. Bajcsy "Active Perception", Proceedings of the IEEE, 76, pp. 996-1005, 1988.

[4] M. Tistarelli and G. Sandini, "Dynamic Aspects in Active Vision", CVGIP: Image Understanding, 56, 1, pp. 108-129, 1992.

[5] P. Dario, A.M. Sabatini, B. Allotta, M. Bergamasco and G. Buttazzo "Object Characterization and Sorting by Active Touch", Proc. of IROS '91, pp. 1353-1356, Osaka, Japan, 1991.

[6] S. Grossberg, E. Mingolla and D. Todorovic "A Neural Network Architecture for Preattentive Vision", IEEE Trans. Biom. Eng., 36, 1, 1989.

[7] A. Maren, C. Harston and R.Pap *Handbook of Neural Computing Aplications*, Academic Press Inc., 1990.

[8] K. Nakayama "The iconic bottleneck and the tenuous link between early visual processing and perception" in Vision, Coding and Efficiency, C. Blakemore (Ed.), University Press, pp. 411-422, 1990.

[9] A. Rosenfeld, (Ed.) *Multiresolution Image Processing and Analysis*, Springer-Verlag, New York, 1984.

[10] P.J. Burt "Smart Sensing within a Pyramid Vision Machine", Proceedings of IEEE, 76, pp. 1006-1015, 1988.

[11] P.Dario, A.M.Sabatini, B.Allotta, M.Bergamasco and G.Buttazzo "A Fingertip Sensor with Proximity, Tactile and Force Sensing Capabilities", Proc. of IROS '90, pp. 883-889, Tsuchiura, Japan, 1990.

[12] F. Fanizza "Camera Calibration for Robotic Control", Technical Report ARTS Lab 92-07, September 1992.

[13] M. Rucci "Neural Networks and Dynamic Vision", Perception, 21 sup. 2, pp. 103, September 1992.

[14] H.B. Barlow "Single units and sensation: a neuron doctrine for perceptual psychology?", Perception, pp. 371-394, 1972.

[15] J. Feldman, and D. Ballard "Connectionist models and their properties", Cognitive Science, 6, pp. 205-254, 1982.

[16] T. Kohonen "Self-Organized Formation of a Topologically Correct Feature Maps", Biological Cybernetics 43, 59-69, 1989.

[17] T. Kohonen *Self-Organization and Associative Memory* (2-nd Ed.) Springer Verlag, Berlin, 1988.

[18] D.E. Rumelhart, G.E. Hinton, R.J. Williams "Learning Internal Representation by Error Propagation", in *Parallel Distributed Processing*, vol. 1, pp. 318-362, Cambridge, MA: MIT Press, 1986.

[19] Z. Balorda "Reducing uncertainty of objects by robot pushing" Proc. IEEE Int. Conf. on Robotics and Automation, Sacramento, 1990.