

1993-01

# An Active Pattern Recognition Architecture for Mobile Robots

---

<https://hdl.handle.net/2144/1982>

*"Downloaded from OpenBU. Boston University's institutional repository."*

**AN ACTIVE PATTERN RECOGNITION ARCHITECTURE  
FOR MOBILE ROBOTS**

J. Mario Aguilar and Jose L. Contreras-Vidal

**January 1993**

**Technical Report CAS/CNS-93-005**

Permission to copy without fee all or part of this material is granted provided that: 1. the copies are not made or distributed for direct commercial advantage, 2. the report title, author, document number, and release date appear, and notice is given that copying is by permission of the BOSTON UNIVERSITY CENTER FOR ADAPTIVE SYSTEMS AND DEPARTMENT OF COGNITIVE AND NEURAL SYSTEMS. To copy otherwise, or to republish, requires a fee and/or special permission.

Copyright © 1993

Boston University Center for Adaptive Systems and  
Department of Cognitive and Neural Systems  
111 Cummington Street  
Boston, MA 02215

# AN ACTIVE PATTERN RECOGNITION ARCHITECTURE FOR MOBILE ROBOTICS

J. Mario Aguilar<sup>1</sup> and José L. Contreras-Vidal<sup>2</sup>

Cognitive and Neural Systems Department

111 Cummington St., Rm. 240

Boston University

Boston, MA 02215

## Abstract

An active, attentionally-modulated recognition architecture is proposed for object recognition and scene analysis. The proposed architecture forms part of navigation and trajectory planning modules for mobile robots. Key characteristics of the system include movement planning and execution based on environmental factors and internal goal definitions. Real-time implementation of the system is based on space-variant representation of the visual field, as well as an optimal visual processing scheme utilizing separate and parallel channels for the extraction of boundaries and stimulus qualities. A spatial and temporal grouping module (VWM) allows for scene scanning, multi-object segmentation, and featural/object priming. VWM is used to modulate a trajectory formation module capable of redirecting the focus of spatial attention. Finally, an object recognition module based on adaptive resonance theory is interfaced through VWM to the visual processing module. The system is capable of using information from different modalities to disambiguate sensory input.

## Introduction

Trajectory planning and generation, and real-time monitoring of the environment are critical components of systems for mobile robots. Computing resources and time constraints require that any algorithm be efficient and capable of real-time operation. Unfortunately, current approaches cannot accomplish this because the robot is treated as a passive entity striving to make sense of the information impinging upon its sensors. In this proposal however, we propose an active role and identify several characteristics that an autonomous robot should include in its control architecture. First, the sensory system should be active as to react to novel environments, alter the characteristics of the sensory apparatus to improve the quality of the input, and look intelligently for useful targets. Second, the system should combine information from different modalities to decrease uncertainty in the measurements. Third, intelligent robots should make use of previous information and context to speed-up sensory processing and planning. In this paper, we propose an active navigation architecture for mobile robotic applications. The system uses space-variant processing to decrease the computational requirements; fast visual processing algorithms, and continuous sensory and goal (priming) feedback. The model is both stimulus and attentionally driven. By the former process, the system reacts to new environments and can trigger learning. The latter process allows for top-down modulation (e.g. target specific information) of both the recognition and pre-attentive visual segmentation of the scene.

In the next section, an optimal stimulus representation is presented; then, an image enhancement and segmentation module which utilizes separate and parallel channels for the extraction of boundaries and stimulus qualities respectively will be discussed. Next, a spatial and temporal grouping module (Visual Working Memory, VWM) which allows for scene scanning by dynamically changing the temporal and spatial focus of attention within a given scene is introduced. This stage is used to modulate a trajectory formation module capable of redirecting the focus of spatial atten-

---

1. Supported in part by DARPA (AFOSR 90-0083) and Office of Naval Research (ONR N00014-92-J-1309).

To whom correspondence should be addressed.

2. On leave from Instituto Tecnológico y de Estudios Superiores de Monterrey. Supported by a fellowship from CONACYT # 63462.

tion. Finally, an object recognition module based on adaptive resonance theory is interfaced through VWM to the segmentation module (Carpenter & Grossberg, 1987). The system is capable of using information from different modalities to disambiguate sensory input.

### World Representation

The test-bed model utilizes information from two modalities: visual and range data. Combining data from different sensors can improve the signal-to-noise ratio of the measurements, eliminating uncertainty introduced by sensor transduction, misalignment, or miscalibration. The system uses vector associative maps (Gaudio and Grossberg, 1991) to combine range data from vision and ultrasonic rangefinders (Contreras-Vidal et al. 1992).

Visual data is pre-processed using a space variant representation of the image to provide not only with attentional focusing but also to reduce the computational requirements, without losing global information (Schwartz, 1980). The transformation is obtained by a mapping consisting of an exponential decay in the resolution of the image; therefore reducing the size of the image. Also, the transformation defines a relative region around the center of the image (i.e. fovea) at which full resolution is maintained thus leading to spatial focusing.

In addition, the space-variant representation avoids a subtle problem faced by image processing systems, mainly, the number and scale of filters to be utilized. In a system capable of actively changing the focusing site, filters with a single scale can functionally compute the same transformation as that of a bank of filters. Here again, computational complexity is reduced by requiring only one convolution with the image.

### Gabor Filtering and Spatial Segmentation

A preattentive visual system, based on the Boundary Contour System/Feature Contour System (Grossberg and Mingolla, 1985) with Gabor-based receptive fields is used to enhance and segment the visual stimulus. The system uses attentional and priming information (e.g. target dependent) to modulate its response in real-time. This system consists of two parallel pathways which extract high frequency (HF) information (e.g. boundaries, edges, or lines) and low frequency (LF) information (e.g. brightness or color) respectively. Both channels have noise suppression properties. In particular, the HF channel reconstructs, emphasizes, and extracts boundaries from the image. The LF channel performs filtering of "salt-and-pepper" and wide-band noise. This system contrast-enhances the image and suppresses the noise without blurring the image by fusing the HF channel with the LF channel.

A key feature of the system is that filtering scale is time varying and modulated by attentional factors. This mechanism allows for an ordered process by which the system can first detect the presence of a stimulus (large scale), second, extract its form (medium scale), and third, perform detailed feature extraction (small scale).

These modules are computed using a Fast Fourier Transform based implementation of the original BCS/FCS algorithm. This is accomplished in an iterative manner to speed-up the image processing and segmentation stage (Contreras-Vidal and Aguilar, 1993).

### Visual Working Memory

The preattentive visual system is interfaced with an object recognition system through a visual working memory. This module consists of a 2D, feedback, distant dependent, on-center, off-surround, shunting network. It receives inputs from the preattentive system (bottom-up), the object recognition system (top-down), and from other sensory modalities (data fusion). It combines spatial and temporal grouping of input patterns to segment each component of the image. The dynamics are based on neurotransmitter gates (a type of medium-term memory) to allow for scene scanning by dynamically changing the spatial focus of attention within a given scene.

### Object Recognition System

This stage is interfaced with an object recognition system, namely ART-2D, through the visual working memory (VWM) described above. The ART-2D system allows for supervised or unsupervised stable category learning of ob-

jects. Its spatial specificity is accomplished through the use of several processing scales. Larger scales provide shift-invariance at the risk of losing detail. Smaller scales are sensitive to small displacements of the visual field, but they discriminate subtle form differences. In a practical implementation, the system switches from larger to smaller scales in a discrete fashion based on the current match level between input and templates.

### **Target Selection and Trajectory Generation**

Selective sampling of the scene enables one to focus upon the most important aspects of the stimuli regardless of whether it is accomplished by moving the sensor on the robot or by moving the robot itself. In the former, active scanning is utilized to obtain a world view or to search for targets given a static world position. Through this process the system can generate a map of its surroundings in order to establish an action plan for target acquisition. In the latter case, explicit movements of a robot with fixed sensors could provide the means for detecting, tracking, monitoring, and executing the plan for target acquisition. In this article we present an implementation of the first proposal and suggest complementing it with the second approach. For example, the robot could monitor its surroundings for sudden changes in the environment, as well as plan subsequent actions while approaching a specified target.

By its nature, focusing our resources requires a mechanism for selecting appropriate targets in the visual field, either for search or refinement of our current knowledge of the world. In the present architecture we propose the use of both data-driven and knowledge-driven information. The characteristics of the bottom-up information can allow to determine regions of interest (i.e. areas of high contrast or curvature). This module is implemented by utilizing a contrast-enhancement layer and selecting the highest peaks in the image. The process of target selection is also modulated by knowledge-driven information from both the boundary detection module and the visual working memory. These in turn are modulated by the object recognition stage. The contribution of the boundary detection module is in the form of emphasizing objects and thus biasing those peaks associated with distinct objects. These processes allow for constant scanning of the scene in the absence of priming signals. Working memory can increase activity in those peaks associated with a primed object which occurs in the visual field or by the current state of the matching process.

This module consists of a set of topographical maps for saccade generation. Information in these maps provides for velocity and directional instantiating through separate quantities (a type of energy-pattern factorization) (Bullock and Grossberg, 1988). Interaction among these maps allows for computation of the position for the next targeting movement and generating a trajectory. The system is capable of maintaining a trace of previous directions and hence is able to form maps of the environment and learn them. These processes allow for constant scanning of the scene in the absence of priming signals.

### **Block Diagram and Simulations**

Figure 1 shows the block diagram of the whole system. We have defined two types of inputs which are combined in VWM. Range data is preprocessed and sent directly to it. Visual inputs in the space variant form are contrast enhanced and sent to the filtering/segmentation and target selection stages. Information from feature extraction and boundary extraction are finally relayed to VWM. Signals in the target selection and trajectory formation block are modulated by definition of boundaries and state of the VWM/ORS system. This trajectory is instantiated as movements of the robot which consequently provide a new set of inputs. Thus, the robot's perception of stimuli can guide actions for information retrieval which, in turn, provide renewed stimuli to its sensors.

Figure 2 shows a sample simulation for the acquisition of a target. In this simulation, a given robot's view of the environment is shown. Each view is processed through the space-variant mapping to reduce the overall dimension of the image. The resulting image is further processed to extract useful information from the scene (e.g. the world "Sale"). If at the current region of attention, no target is found, the system is driven to a different point in space (i.e. camera moves) by selecting a new visual target within the visual field. As soon as a match occurs between the primed information and an object in the central region (fifth image in the sequence), scanning ceases. The last image to the right of the sequence shows the information in VWM at the end of the sequence.

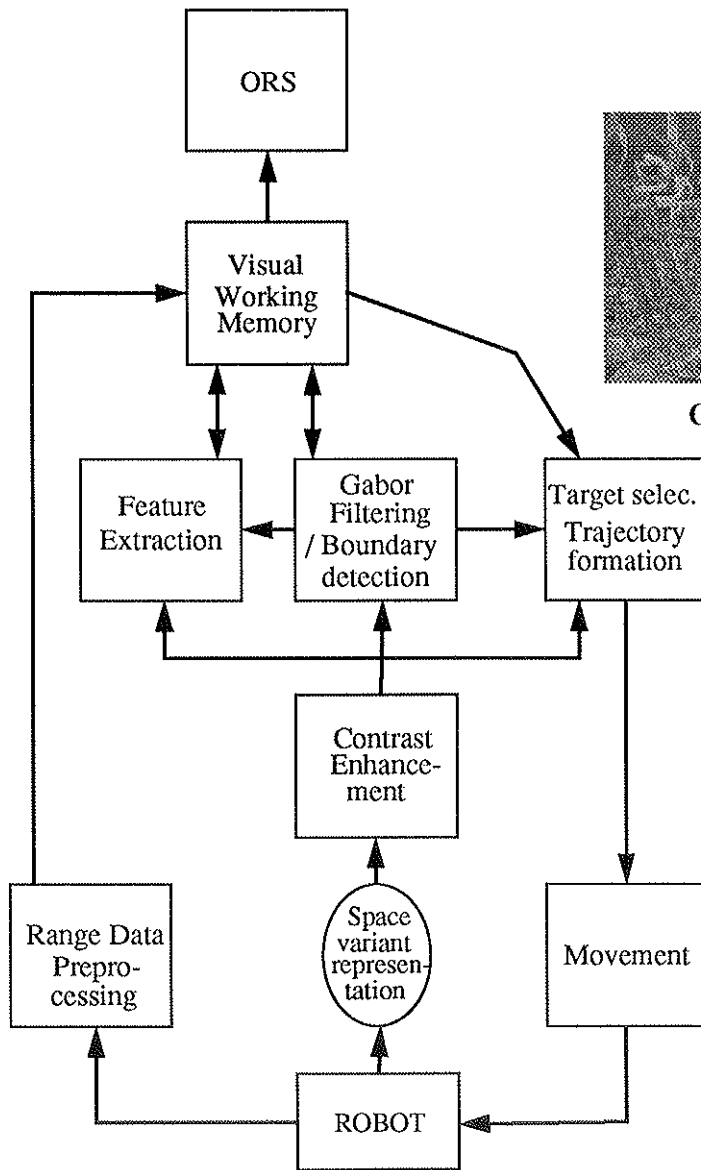


Figure 1: Block diagram of proposed system.

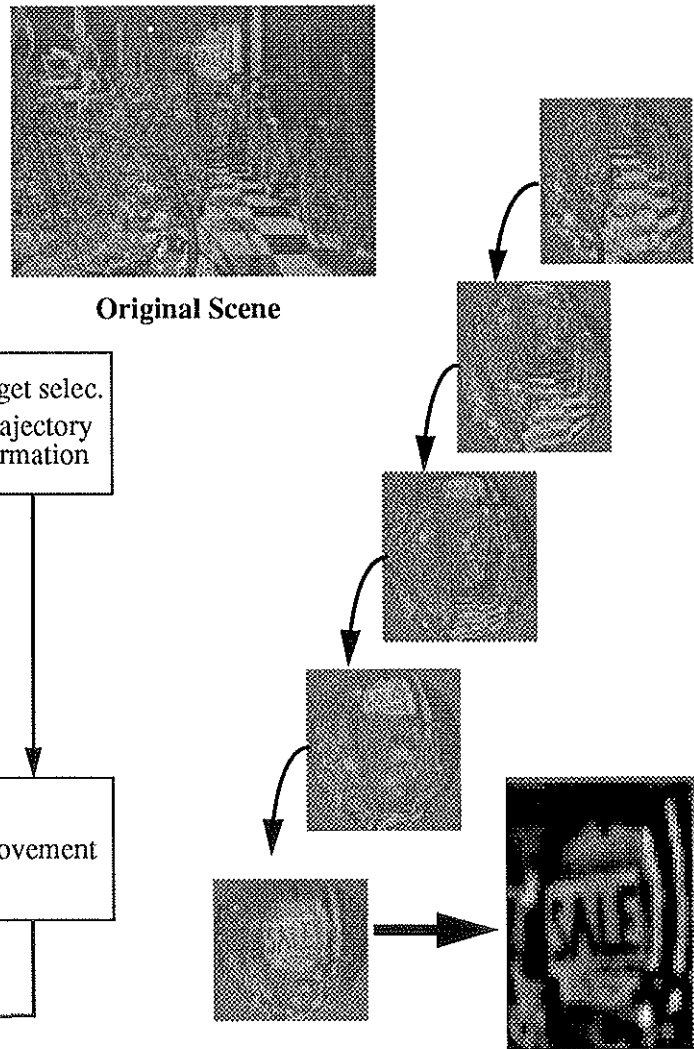


Figure 2: Sample simulation in which the target to be acquired is the SALE sign.

### References

- Bullock, D. and Grossberg, S. (1988). *Psychological Review*, **95**, 49-90.
- Carpenter, G. and Grossberg, S. (1987). *Applied Optics*, **26**, 4919-4930.
- Contreras-Vidal, J.L. and Aguilar, M. (1993). Submitted to ICANN'93, Amsterdam, The Netherlands.
- Contreras-Vidal, J.L., Aguilar, J.M., Lopez-Coronado, J.L., and Zalama, E. (1992). *Proceedings of the SPIE Conference: Applications of Neural Networks X*. Orlando, FL.
- Gaudio, P. and Grossberg, S. (1991). *Neural Networks*, **4**(2), 147-184.
- Grossberg, S. and Mingolla, E. (1985). *Psychological Review*, **92**, 173-211.
- Schwartz, E.L. (1980). *Vision Research*, **20**, 645-669.