

Multimodal, Multilevel Selective Attention

Micheal Hewett

Department of Computer Sciences
University of Texas at Austin
Austin, Texas 78712-1188
hewett@cs.utexas.edu

Perception and Attention

Early knowledge based systems did not incorporate high-bandwidth I/O due to performance limitations of computers of that era. Today, intelligent agents and robots running on much more powerful computers can incorporate vision, sound, network, sonar and other modes of input. These additional inputs provide much more information about the environment, but bring additional problems related to control of perception.

Perceptual input streams (called *modes* in the psychology literature) can have greatly varying bandwidth. In people, the sense of touch has a low bandwidth, while the sense of vision has a very high bandwidth. The human brain can not completely process all of the information from one high bandwidth mode, much less simultaneously process all the information available from all modes. To control the amount of perceptual input processed, humans use *selective attention* (Treisman 1993).

Computer vision is a difficult problem which, if solved, could provide robots with a large amount of useful information. However, visual input can not be processed efficiently without using a top-down attention mechanism (Tsotsos 1987). Several computational models of visual selective attention have been developed (e.g. (Reece 1992)).¹

Generalizing Selective Attention

Given the trends in computer hardware and robot technology, attention mechanisms are going to become even more important than they are now. Unfortunately, visual attention models can not be directly applied to other input modes. For example, (Ahmad 1991) utilizes a network containing three nodes for each of the pixels in a 256 by 256 visual array. This network would be useless for acoustic or textual input.

The goal of my thesis is to develop a generalized selective attention mechanism that handles all modes of input and ranges of bandwidths. Since it is doubtful that single attention mechanism can efficiently handle all types of input, the structure of the generalized mechanism is that of a programmable router. Attention con-

trol is routed from the cognitive component to relevant attention modules, each of which is specific to a mode. In return, data is routed from the attention modules to the cognitive component. All of the modules conform to a standard interface. The challenge is to make the interface general, yet transparent enough so that information is not lost as the data moves from its source to the cognitive component, and as attention commands move in the opposite direction.

Attention commands are routed to any applicable input mode. For example, a goal of finding a **wall** would send attention commands to the visual and sonar modes, while a goal of finding **people** would be routed to visual and acoustic modes. The relative priority of input modes would be governed by intermodal attention control.

Applying Multimodal Attention

We are applying multimodal, multilevel selective attention in an autonomous wheelchair. The wheelchair contains infrared, sonar, visual and user interface sensors, each of which has a different bandwidth and data type. Since the primary goal of the project is assistance in navigation and movement, the sensor inputs are extremely important. We will use the generalized attention mechanism to control perception.

References

- Ahmad, S. 1991. *VISIT: An Efficient Computational Model of Human Visual Attention*. Ph.D. Dissertation, University of Illinois at Urbana-Champaign, Urbana-Champaign, Illinois.
- Reece, D. B. 1992. *Selective Perception for Robot Driving*. Ph.D. Dissertation, Carnegie-Mellon University, Pittsburgh, Pennsylvania.
- Treisman, A. 1993. Features and objects. In Baddeley, A., and Weiskrantz, L., eds., *Attention: Selection, Awareness and Control*. Clarendon Press.
- Tsotsos, J. K. 1987. A 'complexity level' analysis of immediate vision. *International Journal of Computer Vision* 1(4):303-320.

¹Copyright © 1998, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.