Promising Directions in Active Vision

Written by the attendees of the NSF Active Vision Workshop
University of Chicago, August 5-7, 1991

Edited by Michael J. Swain and Markus Stricker

Sponsored by the National Science Foundation,
IBM and Hughes Research Laboratories

University of Chicago Technical Report CS 91-27
November, 1991

For additional copies, write:

Department of Computer Science
University of Chicago
1100 E. 58th Street
Chicago, Illinois 60637
U.S.A.

Contributors:

A. Lynn Abbott	Virginia Polytechnic Institute
Narendra Ahuja	University of Illinois
Peter K. Allen	Columbia University
John (Yiannis) Aloimonos	University of Maryland
Minoru Asada	Osaka University
Ruzena Bajcsy	University of Pennsylvania
Dana H. Ballard	University of Rochester
Ben Bederson	Vision Applications Inc.
Ruud M. Bolle	IBM
Peter Bonasso	The MITRE Corporation
Christopher M. Brown	University of Rochester
Peter J. Burt	SRI/David Sarnoff Research Center
David Chapman	Teleos Research
James J. Clark	Harvard University
Jonathon H. Connell	IBM
Paul R. Cooper	Northwestern University
Jill D. Crisman	Northeastern University
James L. Crowley	LIFIA (France)
Michael J. Daily	Hughes Research Laboratories
Jan-Olaf Eklundh	Royal Institute of Technology, Sweden
Frank P. Ferrie	McGill University
R. James Firby	University of Chicago
Martin Herman	National Institute of Science and Technology
Philip Kahn	Advanced Decision Systems
Eric Krotkov	Carnegie Mellon University
Niels da Vitoria Lobo	University of Toronto
Howard Moraff	National Science Foundation
Randal C. Nelson	University of Rochester
H. Keith Nishihara	Teleos Research
Thomas J. Olson	University of Virginia
Daniel Raviv	Florida Atlantic University
Giulio Sandini	University of Genoa
Eric L. Schwartz	New York University / Vision Applications Inc.
Markus Stricker	University of Chicago
Michael J. Swain	University of Chicago
John K. Tsotsos	University of Toronto
Richard Wallace	Vision Applications Inc.

Summary

Active vision systems have mechanisms that can actively control camera parameters such as position, orientation, focus, zoom, aperture and vergence (in a two camera system) in response to the requirements of the task and external stimuli. They may also have features such as spatially variant (foveal) sensors. More broadly, active vision encompasses *attention*, selective sensing in space, resolution and time, whether it is achieved by modifying physical camera parameters or the way data is processed after leaving the camera.

In the active vision paradigm, the basic components of the visual system are visual behaviors tightly integrated with the actions they support; these behaviors may not require elaborate categorical representations of the 3-D world. Because the cost of generating and updating a complete, detailed model of our everyday environment is too high, this approach to vision is vital for achieving robust, real-time perception in the real world. In addition, active control of imaging parameters has been shown to simplify scene interpretation by eliminating the ambiguity present in single images.

This document describes promising directions for research in active vision and possible applications of this research. It also discusses progress in experimental equipment for supporting this research and potential applications. Important research areas in active vision include attention, foveal sensing, gaze control, eye-hand coordination, and integration with robot architectures:

Attention Selective processing of regions with restricted location, motion, or depth, is necessary for achieving real-time performance with limited resources. Although the design of the attention system is potentially highly complex, and will affect the design of visual processing at all levels, little is known in detail about what this system should look like.

Foveal Sensing A spatially-variant (foveal) sensor permits high resolution at the location of interest without the cost of uniformly high resolution. Issues that need exploring include the properties of unconventional "log-polar" coordinate systems specifically suited to foveal sensing, and the determination of visual features which

can be informative at low resolution to allow reliable selection of targets in the periphery.

Gaze Control The alteration of imaging parameters to aid in the performance of visual tasks, or *gaze control*, is useful for many tasks, including image stabilization, overcoming a limited field of view, figure-ground separation and range estimation. Gaze control is divided into two primary categories: Gaze stabilization and gaze change. The former consists of controlling the camera to maintain clear images of some world point that may be in motion with respect to the camera. The latter category is used to transfer stabilized gaze to new fixation points, similar to human saccadic eye movements.

Gaze Stabilization Timely extraction of the error signal is an important problem in this area. Several interdependent layers of control are needed to deal with the many different conditions under which stabilization or tracking is performed. Stabilization can involve inputs from different vision algorithms and sensors. The integration of these is an important, difficult problem. Work is also needed in the design of robust, adaptable domain and plant models. These models are needed to improve tracking of objects with various patterns of movement and robustness to errors in calibration of the camera-motor system. At a higher level, gaze stabilization must be integrated with methods of control changes in gaze.

Gaze Change Changes in gaze are often motivated by the task; learning the pattern of eye movements to accomplish a given task is an important but relatively unexplored problem. Tasks may have conflicting requirements, as in the case where different tasks would direct the fovea to different points in visual space. As systems with multiple behaviors are built, this scheduling problem will become increasingly paramount.

Eye-Hand Coordination Sensors on a manipulator can augment vision with additional information. Manipulation behavior also provides a performance metric for the vision system and geometric feedback.

Eye-hand systems offer the unique ability to determine material, geometric, and haptic characteristics of objects that are inaccessible to vision systems alone. We need representations and strategies for integrating visual information with hand sensory information.

We must build systems that function robustly without accurate calibration. This is particularly important for eye-hand systems, since both subsystems have distinct geometric parameters and the subsystems must work together. An eye-hand system offers a unique potential for self-calibration, by sending motion instructions to the manipulator and observing their effects.

We see the single biggest obstacle to progress in this area as the cost of hand systems with significant capabilities. We would like to see a cost effective hand system be developed so that it becomes more affordable for many laboratories.

Integration with Robot Architectures Active vision research should be conducted in the context of ongoing research in robot architectures. This is important for two reasons. First, many of the relevant issues are being addressed by people doing research in robot architectures, and active vision investigators can benefit from these experiences. Secondly, researchers developing visual routines need to maintain a view toward the purposes for which robot architectures are being investigated. Important issues to be researched in this context include:

- What set of primitive visual operations are necessary for robot tasks in general?
- What are the appropriate levels of primitive operations?
- What is the right combination of pre-attentive and attentive processing?

Within the document we have made suggestions for benchmarks within the areas of research listed above. More importantly, we believe that the task of developing meaningful benchmarks applicable to many different types of active vision systems is itself a difficult research problem deserving support.

Hardware Support Simulation is inadequate for verifying and demonstrating active vision algorithms. To facilitate experimental research in active vision, we recommend the funding of design, development, and distribution of an inexpensive integrated camera platform with controllable optics. Funding for novel hardware concepts should also be available. For example, emphasis needs to be placed on investigating novel actuator technologies to make eye platforms that are smaller, faster, and more accurate.

Advances in image processing hardware and sensor technology have assisted progress in the active vision field. Fast image acquisition and low-level image processing are currently within the financial reach of most vision labs. The design of special purpose accelerator boards designed by vision researchers is becoming feasible and should be encouraged. Research in foveal sensor arrays is of interest, as well as digital technology for real-time simulation of foveal arrays by remapping the output of a uniform sensor.

While image processing is often well described by SIMD models, MIMD architectures may well be necessary to support the necessary concurrency of cooperating high-level vision, control, reasoning and planning tasks. Therefore, programming languages and software environments that make programming heterogeneous SIMD/MIMD systems easier would facilitate active vision research.

Applications Active vision techniques can potentially enable large gains in cost and performance of computer vision systems in many application areas. In the document we give examples of these gains in the areas of unmanned vehicles, industrial processes, aids for the handicapped, and service robots.

Contents

1 Introduction

A growing number of researchers have found that the ability to actively control the camera parameters, and more generally the ability to combine vision with behavior, is vital to achieving robust, real-time perception for a robot interacting with a complex, dynamic world. In this paradigm called *active vision*, vision is most readily understood in the context of the visual behaviors that the system is engaged in; these behaviors may not require elaborate categorical representations of the 3-D world.

This document describes promising directions for research in active vision and possible applications of this research. It also discusses progress in experimental equipment for supporting this research. The document was written by the attendees of the Active Vision Workshop held at the University of Chicago, August 5-7, 1991. This workshop brought together an international collection of leaders within this emerging research community.

1.1 Why Active Vision?

Active vision systems have mechanisms that can actively control camera parameters such as orientation, focus, zoom, aperture and vergence (in a two camera system) in response to the requirements of the task and external stimuli. They may also have anthropomorphic features such as spatially variant (foveal) sensors. More broadly, active vision encompasses *attention*, selectively sensing in space, resolution and time, whether it is achieved by modifying physical camera parameters or the way data is processed after leaving the camera.

Active systems can dramatically simplify the computations of early vision; enabling areas of interest to be examined at the desired resolution without the costs of uniformly high resolution sensing, simplifying segmentation of an object of interest from its background, using controlled motions to disambiguate solutions that are otherwise underconstrained, examining unseen areas of the scene, and simplifying the transition from image to world coordinates.

The tight coupling between perception and action proposed in the active vision paradigm does not end with camera movements. In this paradigm, visual processing is tied closely with the activities it supports (navigation, manipulation, signaling danger or opportunity, etc.) allowing simplified control algorithms and scene representations, quick response time, and increased success at supporting the goals of the activities. In addition, integration provides new metrics to judge vision algorithms that are better tuned to applications of computer vision.

Current interest in active vision research has been made possible by the recent advent of lightweight cameras and off-the-shelf computer-controllable motors, and advances in real-time image processing equipment. The spread of the experimental equipment for active vision research is a good indicator of the growing popularity of the field. For example, the first computer-controlled binocular camera platform that could perform rapid changes in gaze was built at the University of Pennsylvania in the early 1980's. Now at least twenty other labs have acquired or will soon be acquiring such vision platforms, including: Harvard University, Hughes Research Laboratories, INRIA (France), LIFIA (France), McGill University, MIT, MITRE Corporation, New York University, Osaka University (Japan), Oxford University, Royal Institute of Technology (Sweden), Teleos Research, University of Chicago, University of Illinois, University of Rochester, University of Maryland, University of Massachusetts, University of Toronto, and York University (Canada).

1.2 Outline

Attention is discussed in Section 2.1. The concept of selective sensing, introduced here, motivates many of the topics discussed in later chapters. Foveal sensing is discussed in Section 2.2. It is essential with foveal sensor and often useful with a uniformly sampled sensor to maintain the camera axis pointing at the location of interest. This problem is called *gaze control*, and is the subject of Section 2.3. Many locations in the world need to be examined, either because of events which occur outside the agent, or because of the goals of the agent itself. How to shift attention to other areas of the scene and schedule the allocation of resources such as the fovea is the subject of Section 2.4.

One of the most important tasks that the human visual system engages in is eye-hand coordination. This task exhibits the locality that makes it a natural application of spatially-variant sensing. In addition, there are many other reasons why this is an ideal domain in which to study active vision. Because of the multitude of tasks that can be attempted with an eye-hand system, the domain is natural for investigating the computational advantages and robustness of task-based algorithms. Through the use of the manipulator, the agent can control more than just the camera parameters in order to simplify the interpretation of