Space-variant active vision and visually guided robotics: Design and construction of a high-peformance miniature vehicle *

Gil Engel † Douglas N. Greve ‡ Joseph M. Lubin † Eric L. Schwartz †‡

†Vision Applications Inc.
280 Lincoln St.
Allston, MA USA 02134

‡Dept. Cognitive and
Neural Systems
Boston University
Boston, MA USA 02146

Abstract

In this paper, the design and implementation of a miniature visually guided robot vehicle are described. The design principle underlying this vehicle, which is called CORTEX-II, is the space-variant visual architecture of visual cortex. This allows a wide visual field and high resolution to be maintained in a small and relatively inexpensive system. With a total weight of under twenty kg and size of about .5 meters, CORTEX-II incorporates a three degree of freedom active vision system (pan, tilt and zoom) mounted on a 1 degree of freedom "neck", 160 Mflops (loosely coupled array of four TI320C40 parallel DSP's) and 48 megabytes of memory. Topics discussed here are modification of a stock RC vehicle platform, integration of the parallel DSP platform, active vision system, battery and power systems, communications, and a programming environment for development of the real-time parallel tasks that is transparently portable between simulation and robot run time contexts. Potential applications for systems of this type include traffic monitoring, autonomous driving research, security and surveillance, space-exploration, smart weapons systems, and other application domains in which small, low-cost, high performance visually guided robots are required.

1 Introduction

In this paper, we describe the design and construction of a miniature visually guided robot, using space-variant active vision design principles. In terms of size (.5 Meter), weight (twenty kg), and computational power (160 MFlops), this platform provides a significant advantage over most contemporary visually guided robots for applications in which it is necessary to jointly optimize size, weight and cost.

The term "space-variant vision" has been used (Yeshurun and Schwartz, 1987; Yeshurun and Schwartz, 1989; Schwartz, 1980) to describe sensor architectures in which pixel size increases from a central "foveal" region. One motivation for interest in this architecture is that it provides a large advantage in the "space-complexity" of a vision architecture, compared to a conventional space-invariant camera system. Research into log polar space-variant architectures has been pursued for more than a decade (Weiman and Chaikin, 1979; Sandini and Tagliasco, 1980; Schwartz, 1977; Baloch and Waxman, 1991). And multi-resolution, or pyramid architectures (Burt and Adelson, 1981) are closely related, particularly when a "truncated" or "foveating" pyramid is used (Burt, 1988).

Space-variant vision can reduce the pixel burden by factors of 100 (512x512 sensor regime) to 10,000 (human vision). [1] Since virtually all hardware requirements in a vision system scale with the pixel-bandwidth requirements (e.g. sensor cost, memory, CPU bandwidth, power) space-variant vision will almost certainly become a major architectural approach for the construction of high-performance, small, low-cost visually guided robots, as it has evolved to represent the current architecture of all higher vertebrate

---

*Supported by ARPA ANNT Program, ONR Contract #N00014-92-C-0119. Douglas Greve was partially supported by grants NSF IRI90-24877 and ONR N00014-92-J-1309. Thanks to Pehr Anderson for help with electronic design and fabrication, to Bruce Fischl for discussions on software engineering, and to Barry Levin for optical design and fabrication.

[1] The huge data compression outlined comes at the expense of resolution in the peripheral field, and is thus predicated on a vision model in which relatively localized areas of interest are serially "fixated." This presupposes the availability of active vision, "attention algorithms", motor control, etc.

1051-4651/94 $04.00 © 1994 IEEE

visual systems. In this paper, we outline some of the design considerations involved in constructing such a system.

2 Design Considerations

CORTEX-II is intended to navigate on city sidewalks or fields, recognize familiar landmarks, and perform optical character recognition (OCR) on signs and/or license plates. [2] Since there is little "off-the shelf" availability of many of the components needed to build systems of this type, it became necessary to solve a large set of problems, including the design of power and battery circuitry, the construction of a small low-cost, high-performance active vision actuator system, a wide dynamic range camera, and the modification of available vehicle platforms. These issues will now be described.

3 Mechanical and Electrical Design

3.1 Mechanical Elements

3.1.1 Vehicle Assembly

CORTEX-II is based on a "Thunder King" radio controlled (RC) vehicle made by Model Rectifier Corporation (MRC). This vehicle, designed for off-road racing competition, is a 1/10th scale truck, which measures 50x40 cm and comfortably supports a 3U height VME rack, battery systems, active vision camera and other electronic equipment, as shown in Figure 1.

With minor modifications, the platform is capable of carrying a pay-load of about 9 kg at a maximum speed of 12 kph. It can climb moderate inclines and travel on hard surfaces (eg. sidewalks, dirt, and short grass), limited by the wheel height, which is about 12 cm. The result is a convenient robotic platform for use on paved roads and other smooth terrain.

The prototype CORTEX-II shown in Figure 1 is a development platform which contains a maximal amount of hardware: in addition to an array of parallel C40 DSP processors(four C40 DSP's), CORTEX-II contains an embedded 486 PC and a 250 Megabyte miniature disk drive, along with a 48 Megabytes of DRAM. Future field deployable versions are expected to require a minimal number of C40 DSP's, no PC or disk drive and relatively little on board memory.

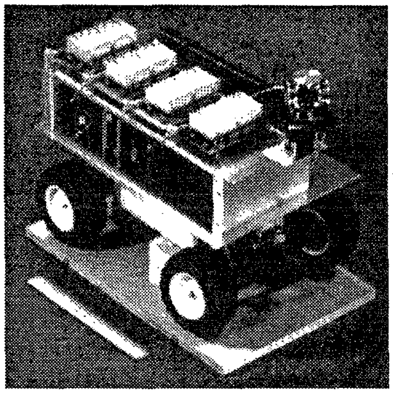

Figure 1: Cortex-2. A 12 inch rule is provided for scale in the lower left. The SPM camera system is mounted on the upper right corner of the vehicle. Wire wrap boards containing vehicle and actuator control latches, frame grabber and motor control are seen as vertical cards within the VME 3U card cage. The C40 DSP's and an Air-LAN wireless ethernet board are mounted vertically behind the card cage, and an embedded PC and disk drive are mounted under the card cage. A nickel metal hydride (NiMH) battery is seen under the card cage on the left side.

3.1.2 Camera Motor

Actuation of the camera is a critical detail of any active vision system. If the system is to be small, lightweight, and high performance, the availability of "off-the-shelf" actuators is problematic. We have developed a novel actuator, called a Spherical Pointing Motor, originally described in (Bederson et al., 1994a; Bederson et al., 1992). This device uses a system of three orthogonal coils to point a rare-earth rotor, upon which the sensor and optics are mounted.

Issues related to the control of this type of actuator are described in other work (Greve et al, in preparation). The dynamics of the SPM are comparable to that of the human eye (saccadic speed of about 600 degrees/sec), while the size and weight of the SPM camera system, which measures 5.8x5.8x5.8 cm and weighs 140 grams, is close to that of the human eye, especially in comparison with other contemporary active vision actuation systems. [3]

---

[2] In previous research, we have demonstrated the ability of a stationary space-variant active vision system to track moving vehicles, locate their license plates in real time, and perform OCR on the license plates. This earlier system, called CORTEX-I, has been described in (Ong et al., 1992; Bederson et al., 1992).

[3] An additional advantage of the SPM is that it is constructed from a few hundred feet of magnet wire, a small rare earth magnet, and a simple pin-bearing gimbal. No precision machining,

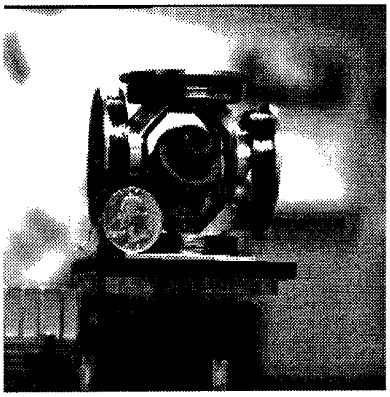

Figure 2: Spherical Pointing Motor active vision camera. A U.S. Quarter is shown for scale. Four of the five drive coils are visible here, and the lens and magnetic rotor are visible in the center. The bearing is a simple pin bearing gimbal design. The lens is an actuated zoom lens with a small coil on the lens acting as a solenoid in the magnetic field of the main magnet rotor.

4 Space-variant Image Synthesis and Optics

4.1 Space-variant Image Synthesis

Although we have experimented with custom VLSI (CMOS) sensor's which incorporate a "foveal" space-variant architecture at the silicon level, and other groups have fabricated successful CCD sensors of this type (van der Spiegel et al., 1989), our current effort is centered on the use of conventional "off-the-shelf" sensors, together with highly optimized software which converts the familiar video raster into a space-variant (e.g. log-polar) format. We favor this approach for three reasons:

1. Conventional sensors of very high optical and imaging quality are evolving at a rapid rate, and we find it advantageous to be able to exploit the huge amount of industrial effort in this area. It is extremely time-consuming to develop custom sensors, and the optical imaging quality of such sensors is likely to lag behind the current state of the art.
2. We have developed highly optimized software to perform the log polar conversion: for example, a single C40 DSP can perform this "image-warp" at more than 25 frames per second. The key idea for this optimized approach to image warping is to represent the look-up table for the image warp as a "run-length encoded" object. This is an obvious idea in terms of saving memory. Less obvious, however, is to effect the image warp in a run-length encoded fashion: incoming pixels from the camera are summed in a register which is not unloaded till the end of each "run". This provides a large CPU efficiency, which can approach an order of magnitude, depending on hardware. We have described this optimized image-warp, based on run-length encoded look-up tables, in detail in (Bederson et al., 1994b), and the method was first described by us in a recent patent application (Vision Applications Inc.,1992,pending).
3. By using a "computed" space-variant sensor approach, we are able to change the parameters of the map function for different applications, or, indeed, in real time.

4.2 Programming model

Software for the robot is compiled using 3L's Parallel C programming environment. Parallel C consists of Texas Instruments' ANSI compliant compiler, a set of library routines which provide support for various C40 features, a linker, a configurer, a debugger, a kernel for each of the C40s in the processor network and a server program which runs on a host PC. A PC-resident server loads the kernel and application code onto the C40s and provides the C40 kernels with run-time access to the PC's resources including the monitor and the file system. The PC is not required at run time, and our development system has an embedded PC merely to simplify loading the C40 kernel and application code.

4.3 Development Environment

The software development environment and code structure of CORTEX-II have been designed to achieve a high degree of platform independence. ANSI C compatible compilers are used on both SUN workstations and PCs to allow algorithms to be developed and tested transparently in both simulated and real environments. This is accomplished through careful code structuring, as well as the use of the object-oriented technique of late binding. Late binding allows a software module to generate messages, such as navigation commands, without knowledge of the module(s) that receive the message. This transparency be-

---

encoders, or other expensive components are required, and we estimate that the actuator alone could be built for less than 10 dollars.

tween SUN-based simulation and C40 based run-time execution has proved to be a major convenience for developing applications.

5 Conclusion

We have designed and built a development platform for visually-guided autonomous navigation with several unique features. Its architecture is based on a space-variant active-vision system whose mechanical performance exceeds that of the human eye. State-of-the-art DSP technology is used to achieve 160 MFLOP's of parallel-processing power. On-board hardware allows (untethered) radio contact with the vehicle for remote observation, debugging, and operation. Despite all these features, the vehicle weighs less than 20 kg, is about .5 M in size, and can achieve speeds in excess of 5 mph. The initial benchmark application of this vehicle is to drive autonomously on a city side-walk, search for license plate targets, and to approach the license-plate, fixate, and read (OCR) the license plate. The current status of this application will be illustrated with a video tape. It is expected that this particular task represents a broad set of equivalent target-acquisition, classification, and autonomous driving tasks. Related tasks which have application value in this context are in the areas of surveillance, autonomous driving, space and planetary exploration, smart weapons systems, and other applications in which small, low-cost high-performance visually guided robots are required.

References

Baloch, A. A. and Waxman, A. M. (1991). Visual learning: adaptive expections, and behavioural conditioning of the mobile robot mavin. *Neural Networks*, 4:271–302.

Bederson, B., Wallace, R., and Schwartz, E. (1994a). Two miniature pan-tilt devices. *IEEE International Conference on Robotics and Automation*, page In press.

Bederson, B., Wallace, R., and Schwartz, E. L. (1994b). Cortex-I: A miniaturized space variant vision system. In Mammone, R. J., editor, *Artificial Neural Networks for Speech and Vision*, pages 429–454. Chapman and Hall.

Bederson, B., Wallace, R. S., and Schwartz, E. L. (1992). A miniaturized active vision system. In *11th IAPR International Conference on Pattern Recognition*, volume B of *Specialty Conference on Pattern Recognition Hardware Architecture*, pages 58–62, The Hague, Netherlands.

Burt, P. and Adelson, T. (1981). A laplacian pyramid for data compression. *IEEE Transactions on Communications*, 8:1230–1245.

Burt, P. J. (1988). Algorithms and architectures for smart sensing. *Proc. DARPA Image Understanding Workshop*, pages 139–153.

Ong, P. W., Bederson, B., Wallace, R., and Schwartz, E. (1992). Space variant optical character recognition. *11th IAPR International Conference on Pattern Recognition*, D:504–508.

Sandini, G. and Tagliasco, V. (1980). An anthropomorphic retina-like structure for scene analysis. *Computer Graphics and Image Processing*, 14:365–372.

Schwartz, E. L. (1977). Spatial mapping in primate sensory projection: analytic structure and relevance to perception. *Biological Cybernetics*, 25:181–194.

Schwartz, E. L. (1980). Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding. *Vision Research*, 20:645–669.

van der Spiegel, J., Kreider, F., Claiys, C., Debusschere, I., Sandini, G., Dario, P., Fantini, F., Belluti, P., and Soncini, G. (1989). A foveated retina-like sensor using ccd technology. In Mead, C. and Ismail, M., editors, *Analog VLSI Implementations of Neural Networks*. Kluwer, Boston.

Weiman, C. F. and Chaikin, G. (1979). Logarithmic spiral grids for image-processing and display. *Computer Graphics and Image Processing*, 11:197–226.

Yeshurun, Y. and Schwartz, E. L. (1987). *Shape description with a space variant sensor: algorithms for scan-path, fusion and convergence over multiple scans*. Number Robotics Report No. 109. NYU Courant Institute.

Yeshurun, Y. and Schwartz, E. L. (1989). Shape description with a space-variant sensor: algorithms for scan-path, fusion and convergence over multiple scans. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11:1217–1222.