HIGGINS LABORATORY TECHNICAL REPORT

Airborne Visual Navigation using Biomimetic VLSI Vision Chips

Charles M. Higgins
Electrical and Computer Engineering / ARL Division of Neurobiology
The University of Arizona
Tucson, AZ 85721 USA

October 1, 2002

AIRBORNE VISUAL NAVIGATION USING BIOMIMETIC VLSI VISION CHIPS

Charles M. Higgins

Electrical and Computer Engineering / ARL Division of Neurobiology
The University of Arizona
Tucson, AZ 85721 USA

ABSTRACT

Organisms such as bees and flies are superb at visually-guided navigation in real-world environments. We have implemented visual motion processing algorithms inspired by the visual systems of insects in custom analog/digital VLSI vision chips. In this paper we describe the implementation and operation of these algorithms, and discuss how these vision chips may be applied to practical problems of airborne visually-guided navigation.

1. INTRODUCTION

The motion processing algorithms evolved over the last few hundred million years by biological organisms from insects to primates are quite different from those developed in modern computer vision over the last few decades. Rather than using a *value code* in which the motion of a local image region is represented with a two-dimensional vector (optical flow), biological systems use a *place code* in which the activation of one out of a set of specially tuned units encodes motion. Rather than making use of units which are sensitive to the velocity of moving objects, biological systems use motion units which are tuned in spatial and temporal frequency.

One might naturally wonder if there are some advantages to these truly ancient computational paradigms over conventional computer vision algorithms, given the comparative success of biological organisms in real environments. While this is an ongoing research area, some relative advantages and disadvantages have become clear. Using motion units that can be adaptively tuned to targets or scene spectra of interest is clearly an advantage when trying to measure the motion of a specific object while the imaging platform is in motion. However, the accompanying disadvantage is the need to implement multiple parallel motion units with different tunings to avoid missing terrain features or unexpected obstacles. Because this kind of parallelism becomes very computationally intensive in a conventional discrete-time serial computing paradigm, we have implemented these algorithms in VLSI vision chips.

2. VISION CHIPS

There are a number of well understood computational algorithms which are used at an abstract mathematical level to model motion detection in biological organisms ([1, 2], but look also for [3]). These algorithms fall in the class of *motion energy* methods because they measure the spectral energy in a specified band of spatial and temporal frequency.

We have chosen to implement a version of the Adelson-Bergen algorithm [2] because it can be done quite compactly. Earlier versions of this implementation have been discussed in previous publications [4, 5], but the latest iteration has substantially revised circuitry and improved signal-to-noise ratio. Our version of this algorithm processes the inputs from two neighboring visual sampling points through temporal filters, additions, subtractions and nonlinearities to result in a local motion-sensitive signal.

These operations are completely implemented in analog VLSI circuitry to maximize power- and space- efficiency. An image is directly focused onto the silicon die. The visual front end is an adaptive photoreceptor [6] which transduces local light intensity into a voltage while adapting to the mean illumination, allowing operation over several orders of magnitude of light intensity without a change in voltage bias settings or lens iris opening. The high-level schematic of the implementation is shown and described in Figure 1. The final current output of this circuit represents the direction of a moving visual stimulus; zero current means no motion, positive current indicates leftward motion, and negative current indicates rightward motion.

This VLSI visual motion circuit was layed out and fabricated through the MOSIS service in a 1.5 μm process; chip layout is shown in Figure 2. Thirteen copies of this motion cell are present on the chip in a one-dimensional row. In order to achieve maximum signal-to-noise ratio, the current outputs of the entire row of pixels were summed by exploiting Kirchoff's current law; the summed response to moving stimuli is shown in Figure 3.

3. EXPERIMENTS IN VISUAL STABILIZATION

Biomimetic VLSI vision chips are well matched to airborne platforms due to their low power and weight requirements, and their potential for high-speed response. We are interested in applying these techniques to the problem of airborne visual navigation. A degenerate portion of the full three-dimensional navigation problem is simply the visual stabilization of an airborne platform.[1] To evaluate our visual motion sensors, we have suspended part of an airship system such that it can only move rotationally (only the yaw angle is free to change), as shown in Figure 4. Using two propellers to generate torque, the job of the electronics is to cancel the rotation of the platform using visual information alone. The two propellers always turn in opposite directions and their control signal is simply the summed outputs of the two vision chips (i.e. proportional control). To facilitate experimentation of several types, two vision chips have been mounted on the platform, both looking slightly downward, with a relative angle of approximately ninety degrees.

[1] A similar task has been demonstrated by Harrison and Koch [7] for a terrestrial platform.

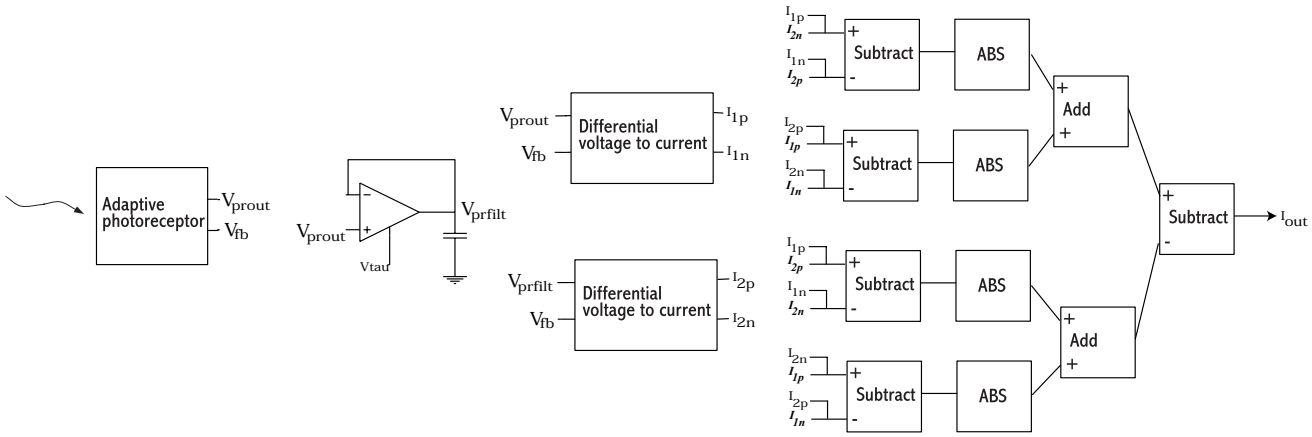
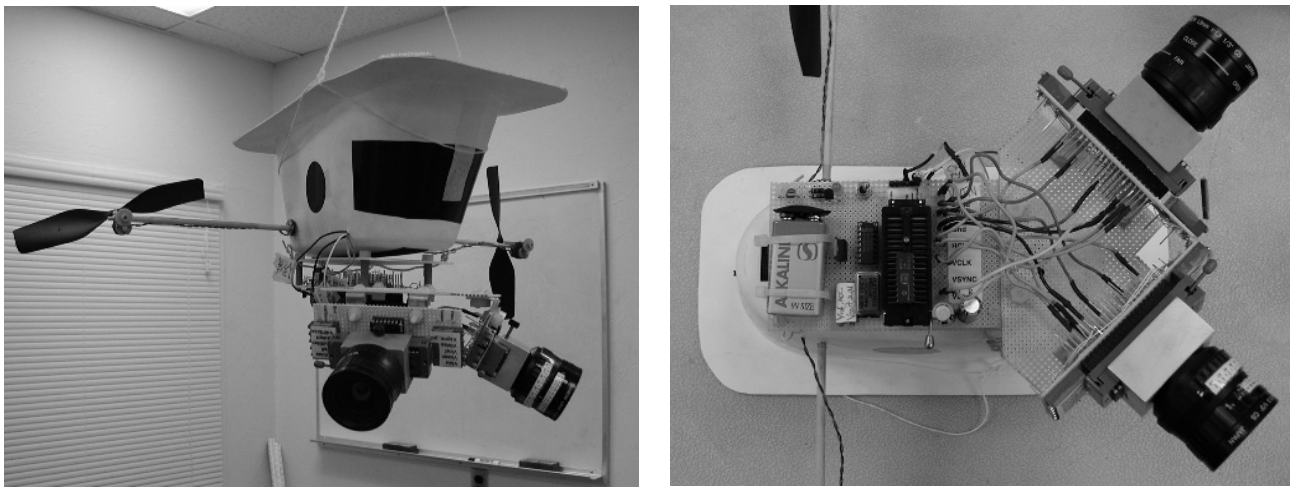


Figure 1: High-level schematic of a circuit implementing a version of the Adelson-Bergen visual motion algorithm. Signal flow is from left to right, starting with phototransduction by the adaptive photoreceptor which provides both an output voltage representing the local contrast (V_{rout}) and a second signal representing the long-term average of local illumination (V_{fb}). The contrast signal is low-pass filtered in voltage mode using a five-transistor transconductance amplifier (V_{prfilt}). These voltage signals are converted to differential currents using differential pairs. The rest of the algorithm involves only sums, absolute values, and differences of currents. Signal currents in italics come from the neighboring pixel to the left. Addition of currents is accomplished via Kirchoff's current law, and subtraction by using a current mirror to reverse the direction of one of the two currents. The absolute value block (ABS) is a three-transistor current-mode full-wave rectifier circuit. The total number of transistors in a pixel is 43.



(a) (b)

Figure 4: Photographs of experimental setup for visual stabilization of a suspended platform. (a) Airship "gondola" suspended from the ceiling with two vision chips used to control the rotors. Each vision chip is underneath a lens which serves to focus an image of the world on it. (b) Closeup of electronics. The analog output of the vision chips (at right) is fed into a microcontroller (near center) which converts these signals into pulse-width modulated digital signals to control the airship servos. The sensory and control electronics is completely powered by a 9-volt battery.

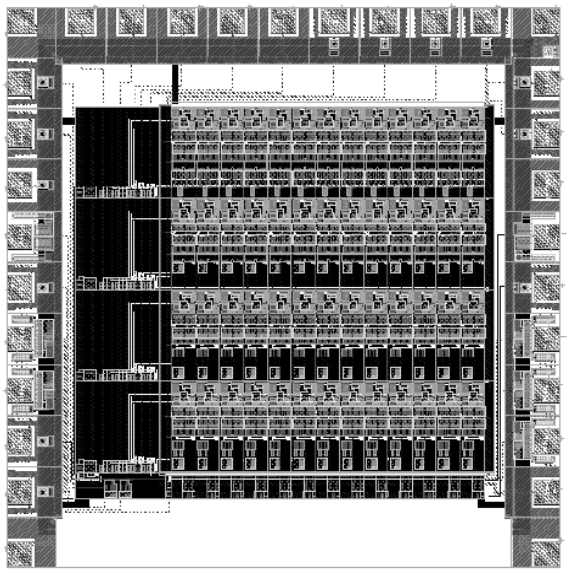


Figure 2: Layout of visual motion chip in a 1.2 μm CMOS process. Total size of the die is 2.1 $\times$ 2.1 mm^2. Four different motion algorithms (only one of which is described here) are tested on this chip, with 13 copies of each on four rows.

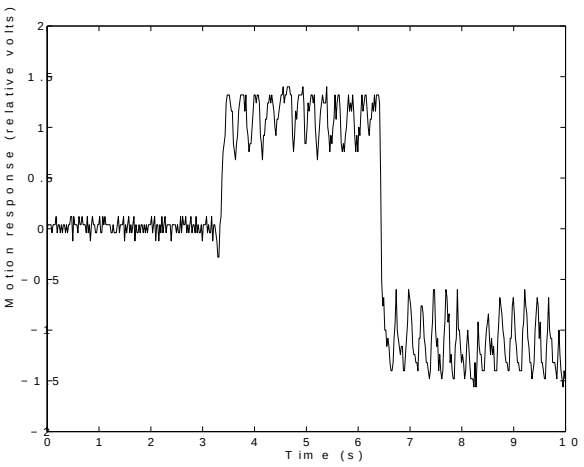


Figure 3: Summed response of a one-dimensional array of 13 motion sensors to a moving sinusoidal grating stimulus. The chip output is actually a current which is converted to a voltage outside the chip using a current sense amplifier with a 1 $M\Omega$ feedback resistor. The response plotted is the sense amplifier output voltage relative to the reference voltage. For the first 3 seconds, the stimulus is stationary and the chip responds only to fluorescent ambient illumination. The stimulus then moves left for approximately 4 seconds, and then changes direction to the right. For this stimulus, there is an extremely strong distinction between leftward, nonmoving, and rightward stimuli.

Each vision chip has a fairly narrow field of view due to the 8 mm focal length lenses used. The very low spatial resolution of each motion sensor makes it possible for the motion chip to respond only to fairly broad (medium spatial frequency) features of the environment.

Our first experiment is one in which we demonstrate the limitations of the vision system. We have handicapped the system in three ways. Firstly, as the platform rotates on its tether, for a large portion of its rotation the objects in the view of any single vision chip have little or no features at spatial frequencies which allow the chip to respond significantly (the chip is looking at a blank wall and a uniformly-upholstered couch). Secondly, the torque outputs are imbalanced. That is, a given negative control signal will result in more negative torque than the corresponding positive control signal, giving the whole system a bias towards spinning in one direction. Thirdly, we have turned off one of the vision chips, effectively "closing one eye".

These three handicaps together result in the stabilization performance shown in Figure 5(a). The platform stabilizes briefly at an angle where the vision chip can see spatial frequency of interest. However, the imbalanced control signals inevitably result in a rotation to angles where no significant motion is seen, leading to a rotation of the platform around to an angle where texture can again be seen.

If we now enable both vision chips, eliminating only the third handicap, the system has no trouble stabilizing itself as shown in Figure 5(b). This is due to the fact that one of the two vision chips is always getting a good indication of platform motion. However, at some rotation angles the stabilization performance is less precise than at other angles, as demonstrated in the inset panel in Figure 5(b).

4. AIRBORNE VISUAL NAVIGATION

Qualitatively similar yaw stabilization performance to that discussed in the suspended system above has been achieved in a free-flying airship.

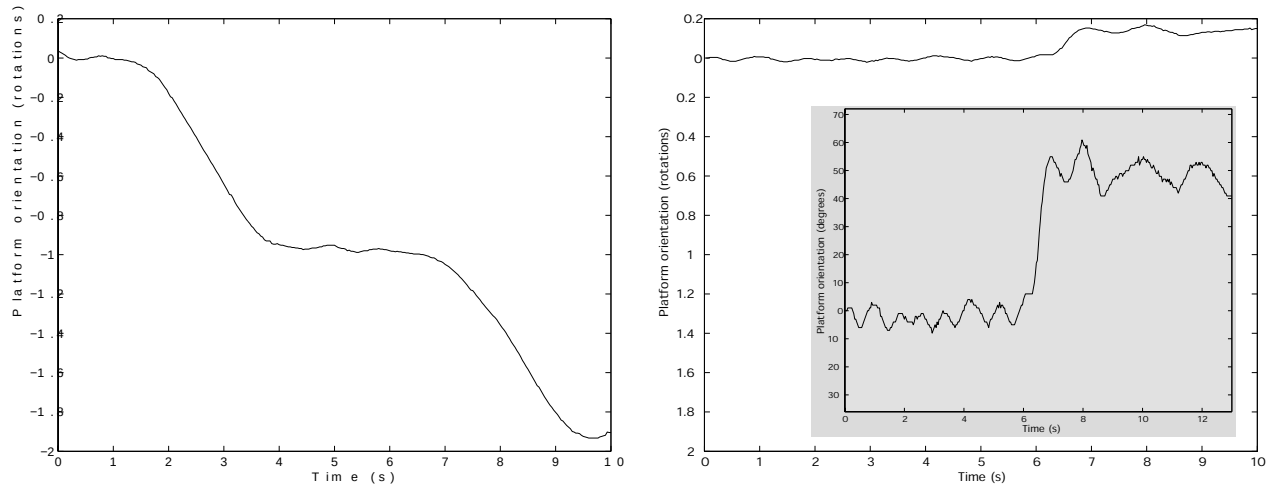
Experiments currently in progress investigate the ability of the same system to avoid obstacles in flight. The biologically-inspired algorithms are based on two sets of studies performed in insects. The first set of studies [8] showed that honeybees in forward flight match the global image velocity of the left and right eyes to steer a centered path between relatively close walls on both sides. The second set of studies [9, 10, 11] describes neurons in locusts and moths which warn of an approaching collision with an object or predator.

The first of these biological algorithms can be approximately emulated in our system by controlling the yaw angle of the system in forward flight to match the response of left and right motion detectors. This will allow the system to avoid running into walls and other extended objects. In addition, the second algorithm can be crudely implemented by initiating a turn when one or the other of the motion inputs exceeds a pre-set threshold.

5. FUTURE WORK

All of the rudimentary algorithms discussed above make use only of the global sum of motion detectors. By making use of the spatial resolution already present in the vision chips, better performance could be obtained.

Models of the locust obstacle avoidance system [11] can be closely emulated [12] by using motion detector outputs to estimate the size and angular expansion velocity of an approaching object.



(a) (b)

Figure 5: Experimental results for visual stabilization of a suspended platform. (a) Narrow field of view: single vision chip operation. The angular axis is scaled so that a value of unity indicates a 360 degree rotation. The platform stabilizes briefly, but then rotates back to the point where it can again see features of interest. (b) Wide field of view: two vision chips operating (main panel is same scale as panel a). At approximately 6.2 seconds, the platform was moved by hand to a different angle. Inset panel is a detail view, scaled in degrees. The system oscillates about a constant mean position; the oscillation is due both to the imbalanced control signals and to the fact that the proportional control loop gain was simply unity. At the first position where the system stabilizes, the amplitude of its oscillation about the mean is about 10 degrees. After the system has been manually repositioned, the oscillation amplitude is increased to about 15 degrees due to the sparser visual scene being observed by the vision chips.

This would allow synthesis of a signal which peaks at a given time-to-collision regardless of the particular object approaching.

The monocular "depth perception" models of Wicklein and Strausfeld [10] offer a more sophisicated route to estimating the relative danger of approaching objects and choosing an appropriate course which remains feasible with a biomimetic VLSI implementation.

6. ACKNOWLEDGEMENTS

This research was supported by the Office of Naval Research under cooperative agreement number N68936-00-2-0002. The author would like to gratefully acknowledge the contributions of Vivek Pant in building the electronic circuit boards used in this experiment, and other essential technical assistance from Robert Stickney, Anusha Muthu Natarajan and Jad Halimeh.

7. REFERENCES

[1] Jan P. H. Van Santen and George Sperling, "Elaborated Reichardt detectors," *Journal of the Optical Society of America A*, vol. 2, pp. 300–320, 1985.

[2] E.H. Adelson and J.R. Bergen, "Spatiotemporal energy models for the perception of motion," *J. Optical Society of America A*, vol. 2, no. 2, pp. 284–299, 1985.

[3] C. M. Higgins, J. K. Douglass, and N. J. Strausfeld, "The neural basis of elementary motion detection in Dipteran insects," In preparation for submission to the Journal of Neuroscience, 2002.

[4] C.M. Higgins and S.K. Korrapati, "An analog VLSI motion energy sensor based on the Adelson-Bergen algorithm," in *Proceedings of the International Symposium on Biologically-Inspired Systems*, Wollongong, NSW, Australia, December 2000.

[5] S.K. Korrapati, "An analog VLSI motion energy sensor and its applications in system-level robotic design," M.S. thesis, Department of Electrical and Computer Engineering, The University of Arizona, Tucson, Arizona, August 2001.

[6] T. Delbrück and C. Mead, "Analog VLSI phototransduction by continuous-time, adaptive, logarithmic photoreceptor circuits," Tech. Rep. 30, Department of Computation and Neural Systems, California Institute of Technology, 1993.

[7] R. R. Harrison and C. Koch, "A robust analog VLSI motion sensor," *Autonomous Robotics*, vol. 7, pp. 211–224, 1999.

[8] M. V. Srinivasan, S. W. Zhang, M. Lehrer, and T. S. Collett, "Honeybee navigation en route to the goal: visual flight control and odometry," *J. Exp. Biol.*, vol. 199, pp. 237–244, 1996.

[9] C. Rind and J. Simmons, "Signaling of object approach by the DCMD neuron of the locust," *J. Neurophysiology*, vol. 77, 1997.

[10] M. Wicklein and N. J. Strausfeld, "The organization and significance of neurons detecting change of depth in the hawk moth *Manduca Sexta*," *J. Comp. Neurology*, vol. 424, no. 2, pp. 356–376, 2000.

[11] F. Gabbiani, C. Mo, and G. Laurent, "Invariance of angular threshold computation in a wide-field looming-sensitive neuron," *J. Neuroscience*, vol. 21, no. 1, pp. 314–329, 2001.

[12] G. Indiveri, "Analog VLSI model of locust DCMD neuron response for computation of object approach," in *Neuromorphic Systems: Engineering silicon from neurobiology*, August 1997.