

A BIOMIMETIC VLSI ARCHITECTURE FOR SMALL TARGET TRACKING

Vivek Pant¹ and Charles M. Higgins^{1,2}

¹Department of Electrical and Computer Engineering

² Arizona Research Laboratories, Division of Neurobiology

The University of Arizona

Tucson, Arizona 85721

ABSTRACT

Tracking of a target in a cluttered environment requires extensive computational architecture. However, even a small housefly is adept at pursuing its prey. Biomimetic algorithms suggest a novel way of looking at this problem. In the lobula plate of a fly's brain, a neural circuit is hypothesized based on a tangential cell called the figure detection (FD) cell. The proposed small target fixation algorithm based on electrophysiological recordings does not take into account the translation of the pursuer during pursuit. We have modified the biological algorithm to include this aspect of tracking. In this paper, we present the elaborated biological algorithm for small target tracking, and an analog VLSI implementation of this algorithm.

1. INTRODUCTION

Target tracking in a cluttered environment from a moving platform is a computationally intensive task. Digital tracking systems that use pattern matching between successive frames are commonly used. However, the speed and computational power requirements for these systems are very high. An alternative solution lies in the implementation of biomimetic algorithms for target fixation and pursuit. Insect visual systems exhibit efficient tracking and pursuit behavior on a daily basis, whether it is for tracking of prey or a male's pursuit of a female of the same species [1]. This motivates us to understand the underlying neural algorithm that a fly utilizes for target detection and pursuit.

The problem of target tracking involves detection of a moving object. This sub-problem of finding the direction of motion has been studied and modeled extensively. A common type of such an elementary motion detector (EMD) is one based on asymmetric filtering of two adjacent sampling units, taking their product and then performing a time averaging operation on this output. This is known as a Hassenstein-Reichardt (HR) detector [2]. The discrimination of a small object from its background can then be done based on the optical flow-fields characterized by the direction of motion. One such algorithm was proposed by Reichardt et al [3], in which a direction selective output from the visual field was utilized to identify discontinuities in the optical flow-fields. It is believed that two functional classes of output elements of the visual ganglia are involved in object-background discrimination by relative motion in the fly: horizontal cells that respond to large textured patterns [4] and figure detection (FD) cells which are sensitive to small moving objects [5, 6] (see Figure 1). A postulated

This work was supported by the Office of Naval Research under Agreement N68936-00-2-0002.

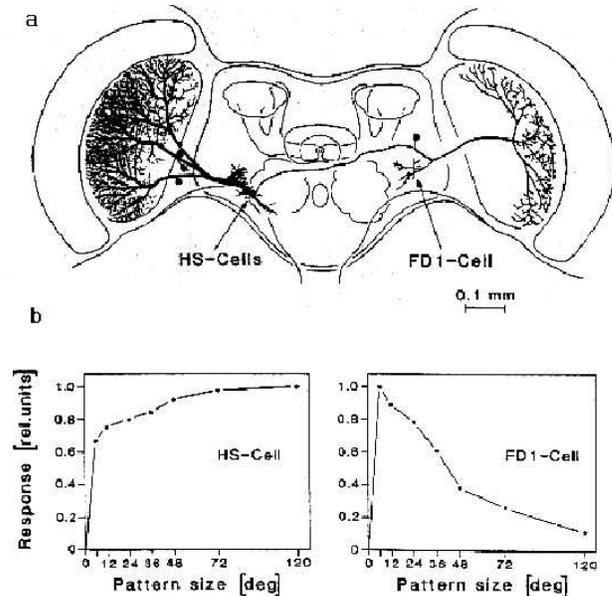


Fig. 1. Spatial integration properties of neural elements. (a) The position of the HS cell and the FD cell in the fly lobula plate. (b) Dependence of the mean response amplitude of a HS cell and a FD cell on the size of the stimulus pattern. The response of HS cell reaches its maximum for motion of large patterns, whereas the FD cell response is strongest for small moving patterns. Reproduced without permission from Egelhaaf and Borst [8].

neural circuit model based on the FD cells has been described by Reichardt, Egelhaaf and Guo [7]. This FD algorithm models the optomotor response (torque) generated by the fly while fixating on a target. Pursuit of a target involves identification of the target from its background, and the ability to keep the target in the view-field of the detector. This involves rotatory motion along with translation in the direction of the target. The FD algorithm can be used to discriminate a small object from its background. However, it does not take into account the complexity of the visual field when there is a simultaneous translation along with yaw maneuvering. We have modified the biological algorithm such that it also takes into account the translational motion of the pursuer towards the target while chasing. In this paper, we present a brief overview of this algorithm, and a description of the VLSI implementation of a target tracking FD sensor. The emulation of such a system has shown

the effectiveness of this algorithm in cluttered environments and results of the same are presented.

2. ALGORITHM

The biological algorithm implemented in this paper is based on spatiotemporal filtering of visual information. The visual input is sampled at several space coordinates by photoreceptors. This input when correlated with the visual input from an adjacent photoreceptor gives the directional motion output. A HR type detector is used here for determining the direction of motion. Each of the individual EMD's response can either be positive (preferred direction of motion) or negative (null direction). The output from this stage is split into two channels, positive v_i^+ and negative v_i^- respectively ($v_i^+ > 0$ and $v_i^- > 0$). It is then summed by two pooling units each on the left and right side of the visual field. A P^+ pool cell is activated by front-to-back motion, and a P^- pool cell by back-to-front motion. For N EMDs on the right side of the visual field, the pool cells perform the following operation:

$$P_r^+ = \sum_i^N v_r^+(i) \quad (1)$$

$$P_r^- = \sum_i^N v_r^-(i) \quad (2)$$

Similarly, P_l^+ and P_l^- sum the motion components on the left side of the visual field.

During pursuit, the translatory motion of the fly (robot) is coupled with the rotatory motion required to keep the target in the center of the view field. In our algorithm, we have accounted for this by combining the response from the pool-cells on both sides of the visual field.

1. When the angular velocity of the robot is smaller than a threshold value, the motion is approximated by a pure translatory motion. This is given as:

$$P_r^f(t) = P_r^+(t) + P_l^+(t) \quad (3)$$

$$P_r^r(t) = P_r^-(t) + P_l^-(t) \quad (4)$$

where P_r^f is the motion expected during forward translation, and P_r^r for translation in opposite direction.

2. When the angular velocity exceeds the threshold, the motion is approximated by a purely rotatory motion. This is given as:

$$P_r^f(t) = P_r^+(t) + P_l^-(t) \quad (5)$$

$$P_r^r(t) = P_r^-(t) + P_l^+(t) \quad (6)$$

where P_r^f is the motion expected during clockwise rotation, and P_r^r for rotation in opposite direction.

Consider a scenario where the object is moving along the preferred direction of the array of EMDs, and the background is static. As the fly (robot) moves its gaze towards the object (target), it sees an apparent motion of the background in the null direction. Thus, a few EMDs looking at the object give a positive output, while all the remaining looking at the background give a negative response. By separation of the two channels we achieve the first level of distinction between the object and the background. Next, the response from individual pixels is normalized over the entire

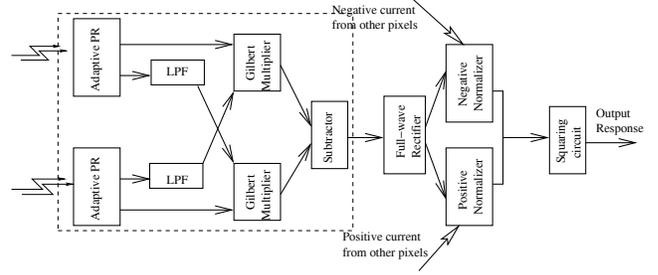


Fig. 2. Basic building blocks in a processing element (pixel) of the FD sensor. The direction of motion is computed using the HR detector shown in the box. The directional output from the HR sensor is split using a full wave rectifier circuit. This output is normalized and squared to produce a response, proportional to the torque response generated for a corresponding small object.

response pertaining to that channel, i.e. the positive response from a pixel is normalized by the entire positive response from all the photoreceptors, and the negative by the total negative response. In our case, we have a few pixels looking at a small target, and we get an average response by this normalization for the target. At the same time, the weak response from the background detected by rest of the sensors gets divided by a large number of otherwise idle sensors, given as:

$$y_i^+ = \frac{v_i^+}{P_r^f} \quad (7)$$

$$y_i^- = \frac{v_i^-}{P_r^r} \quad (8)$$

A non-linear expansion after normalization further suppresses the background response, and at the same time enhances the response due to a small target. This response is summed over the entire left/right visual field, such that for the right “eye”, the output is:

$$R_r(t) = \sum_{i=1}^N abs [y_i^+]^n - [|y_i^-|]^n \quad (9)$$

The exponent n is used to enhance the target response and suppress the background response. In our algorithm, n is a magnitude squaring factor ($n = 2$). The final torque output is the difference between the right and left eye, and is given as:

$$O(t) = R_r(t) - R_l(t) \quad (10)$$

3. VLSI IMPLEMENTATION

The building blocks identified in the previous section are replaced by their circuit equivalents to design a basic “pixel” (see Figure 2). This pixel computes the response over a small portion of the visual field. The output of several such pixels is used to compute the net global tracking response. The sub-circuits used in this architecture can be broadly put under two circuit blocks presented below.

3.1. HR detector

This circuit block consists of an adaptive photoreceptor, a lowpass filter, a multiplication stage, and finally a subtractor circuit. This

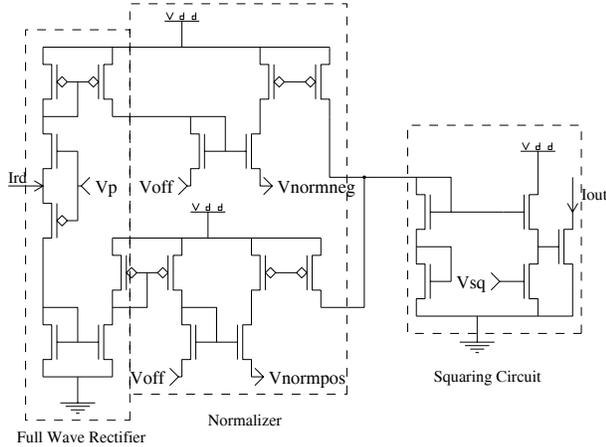


Fig. 3. Part of the pixel circuit for tracking computation. The full-wave rectification, positive and negative normalization, and squaring circuit stages are shown.

HR EMD has been implemented as a modified version of Harrison’s implementation of the HR detector [9].

The input stage is a modified version of Delbrück’s adaptive photoreceptor [10]. This photoreceptor can operate over a range of five orders of magnitude of light intensity, and has an output proportional to the logarithmic of its photoinduced current. This circuit has a bandpass characteristic, so it does not require an explicit highpass filter as in Harrison’s implementation. In our implementation we have used two outputs from this stage: a transient output V_{prou} and a long-time mean response V_{fb} . The photoreceptor output is then filtered by a lowpass filtering (LPF) stage. This is a standard $g_m - C$ filter implemented using P-type MOSFETs. The P-type filter stage allows the photoreceptor output to be in a wider common-mode voltage range. The HR detector is a correlation-based model and to achieve this correlation, a delayed and a non-delayed signal from the photoreceptor are multiplied. The lowpass filtered output V_{prfilt} acts like a delayed response from the photoreceptor. We have used a Gilbert multiplier circuit [11] to perform the multiplication. Again, a P-type implementation has been used to operate the circuit in its maximum input signal range. This circuit removes the mean response V_{fb} from V_{prou} and V_{prfilt} signals to get rid of the DC offset. The output from this multiplication stage is a current I_{mul} representing response of one half of the HR detector:

$$I_{mul} = I_b \cdot \frac{V_{prou} - V_{fb}}{2 \cdot V_T} \cdot \frac{V_{prfilt} - V_{fb}}{2 \cdot V_T} \quad (11)$$

where I_b is a bias current, and V_T is the thermal voltage. Figure 2 shows that a single pixel has two such sub-circuits to perform the complete motion algorithm. The output of these two mirror symmetric circuits is subtracted using a current mirror to get a motion output.

3.2. FD circuit

The second stage of this sensor is detection of the small target from its background and generating an output proportional to the torque response for steering the robot. Figure 3 shows the transistor level circuit used to implement this stage. A full-wave rectifier splits the

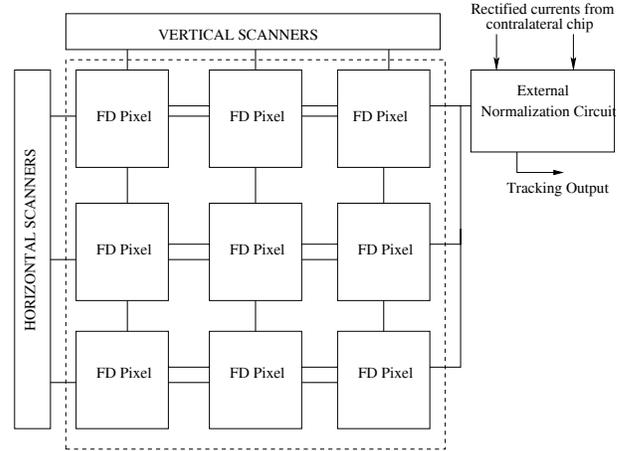


Fig. 4. Chip Architecture of the FD sensor chip. Arrays of FD pixels are juxtaposed, and interact to generate a tracking response. Vertical and horizontal scanner circuits are used to investigate individual pixel response for target localization. A normalization circuit common to the entire chip scales the current output from each pixel, as well as the interaction with another sensor acting as the second “eye”.

directional current output from the HR detector stage into positive and negative halves. Here, it should be noted that at any given instant, the output from a motion detector is either positive or negative, but never both. So, this rectification can be viewed as a routing mechanism for the current to its corresponding normalization circuit. A Gilbert normalizer circuit [12] is used to implement both positive and negative normalization. Each pixel acts as one of the n channels of the normalizer, and an external normalization circuit has the bias controls for the normalizer. The FD algorithm described in section 2 has two identical “eyes” that interact with each other during normalization. In our implementation we have utilized a channel, in both the normalizers, that receives its input from an external chip corresponding to the positive and negative half of the total HR detector response, summed over the entire chip. In addition, we have used another channel in the normalizer to take care of the condition when none of the channels are active (current through them is zero). This channel is controlled by an external bias to introduce a very small current such that the output current is the maximum through this channel in case of no activity. Translinear analysis of the normalizer circuit gives an output current for a positive channel of the form:

$$I_{norm}^+(i) = I_{bias} \cdot \frac{I_r^+(i)}{I_\beta + \sum_{j=1}^n I_r^+(j) + I_{left}^-} \quad (12)$$

where I_{bias} is the total bias current of the normalizer, $I_r^+(i)$ is the positive half of the HR detector output in i^{th} channel, I_{left}^- is the total negative current response of the HR detectors from the contralateral chip, and I_β is the small current to take into account the case of no activity.

As discussed, only one of the normalizer circuits is active at any given instant of time, so that the order of nonlinear expansion and subtraction of the two channels as shown in equation 9 can be reversed without changing the characteristics of the response. A squaring circuit nonlinearly expands the output from the normalization circuit. This current mode squaring circuit [13] gives an

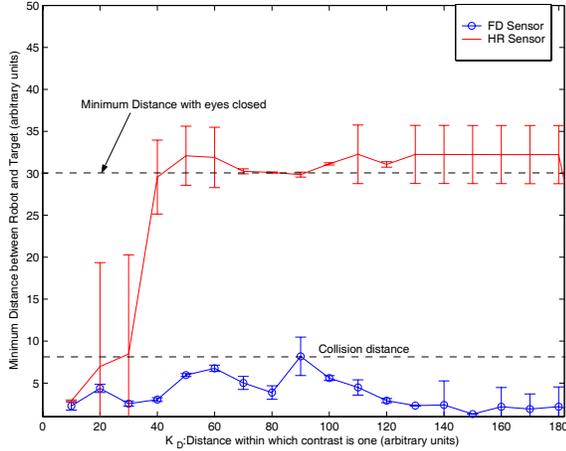


Fig. 5. System emulation with HR and FD sensors. Tracking can be done using a less sophisticated sensor like the HR detector only for less cluttered environment. However, the FD sensor is effective even in highly cluttered scenarios. The minimum distance between the robot and the target when the robot is moving without any visual inputs is also shown.

output current that is proportional to the torque response due to the small moving target, and by applying translinear analysis can be shown to be of the form:

$$I_{out} = \frac{(I_{norm})^2}{I_{sq}} \quad (13)$$

where I_{sq} is a bias current controlled by an external bias V_{sq} . The complete architectural layout of the FD sensor chip is shown in Figure 4.

4. RESULTS

The FD sensor chip has been fabricated through MOSIS in a standard $1.6 \mu\text{m}$ CMOS process. The resolution of this sensor is 13×6 pixels on a $2.2 \text{ mm} \times 2.2 \text{ mm}$ die. We have performed a system level emulation of this circuit, and used the output response to control the trajectory of a simulated robot as it chases a target. The effectiveness of this algorithm over a less sophisticated algorithm based on only direction of motion (the HR detector) is shown in Figure 5. Two identical robots were controlled by two different sensors, the HR sensor and the FD sensor. In the system emulation, the arena was a 2-D plane 300×300 space-units in size that was bounded by four walls with 20 fixed objects on each wall, with a moving target. The contrast of the fixed objects varied with distance as:

$$C_i = \min \left(1.0, \frac{K_D}{D_i} \right) \quad (14)$$

where D_i is the distance of an object from the robot, and K_D is the distance within which the contrast is true (not scaled). The robot with the HR detector circuit was not able to track a target when the contrast scaling distance K_D was increased, i.e. the background became more cluttered. The robot moved in the simulated arena as if there was no visual input. Under similar conditions, the robot equipped with FD sensor circuitry was able to detect and track the target, even for bright background objects. This ability of the FD

sensor motivates its potential of tracking a target in complex real world scenarios.

This biomimetic algorithm, therefore, presents a simple but robust model for tracking in cluttered scenarios. The implementation of the entire circuit has been done in the subthreshold region of MOSFET operation. This reduces the power dissipation of the entire chip. This architecture can be used to detect the direction of motion via the HR detector output, or to track a target utilizing the FD sensor, or to localize the target by inspecting individual pixel responses. Thus, this aVLSI architecture has all the requisite mechanisms on board for tracking in real world scenarios, as predicted by the system level emulations.

5. REFERENCES

- [1] M.F. Land and T.S. Collett, "Chasing behaviour of houseflies (*Fannia canicularis*): description and analysis," *J. Comp. Physiol.*, vol. 89, pp. 331–357, 1974.
- [2] W. Reichardt, "Autocorrelation, a principle for the evaluation of sensory information by the central nervous system," in *Sensory Communication*, WA Rosenblith, Ed., pp. 303–317. MIT Press, New York, 1961.
- [3] W. Reichardt, T. Poggio, and K. Hausen, "Figure-ground discrimination by relative movement in the visual system of the fly. II. Towards the neural circuitry," *Biol. Cybern.*, vol. 46, pp. 1–30, 1983.
- [4] K. Hausen and M. Egelhaaf, *Chapter 18: Neural mechanisms of visual course control in insects*, vol. 01, pp. 1–34, 1989.
- [5] M. Egelhaaf, "On the neuronal basis of figure-ground discrimination by relative motion in the visual system of the fly. II. Figure-detection cells, a new class of visual interneurons," *Biol. Cybern.*, vol. 52, pp. 195–209, 1985.
- [6] M. Egelhaaf, "On the neuronal basis of figure-ground discrimination by relative motion in the visual system of the fly. III. Possible input circuitries and behavioral significance of the FD-cells," *Biol. Cybern.*, vol. 52, pp. 267–280, 1985.
- [7] W. Reichardt, M. Egelhaaf, and A.K. Guo, "Processing of figure and background motion in the visual-system of the fly," *Biol. Cybern.*, vol. 61, pp. 327–345, 1989.
- [8] M. Egelhaaf and A. Borst, "A look into the cockpit of the fly - visual orientation, algorithms, and identified neurons," *J. Neurosci.*, vol. 13, pp. 4563–4574, 1993.
- [9] R. Harrison and C. Koch, "A robust analog VLSI Reichardt motion sensor," *Analog Integrated Circuits and Signal Processing*, vol. 24, pp. 213–229, 2000.
- [10] T. Delbrück, "Silicon retina with correlation-based, velocity-tuned pixels," *IEEE Trans. Neural Networks*, vol. 4, pp. 529–541, May 1993.
- [11] B. Gilbert, "A high-performance monolithic multiplier using active feedback," *IEEE J. Solid-State Circuits*, vol. SC-9, pp. 267–276, Dec. 1974.
- [12] B. Gilbert, "A monolithic 16-channel analog array normalizer," *IEEE J. Solid-State Circuits*, vol. 19, pp. 956–963, 1984.
- [13] C. Toumazou, F.J. Lidgey, and D.G. Haigh, *Analogue IC design: the current-mode approach*, Peter Peregrinus Ltd., 1990.