

# An elaborated model of fly small-target tracking

Charles M. Higgins, Vivek Pant

Electrical and Computer Engineering/ARL Division of Neurobiology, The University of Arizona, 1230 East Speedway Boulevard, Tucson, AZ 85721, USA

Received: 27 February 2004 / Accepted: 20 July 2004 / Published online: 30 November 2004

**Abstract.** Flies have the capability to visually track small moving targets, even across cluttered backgrounds. Previous computational models, based on figure detection (FD) cells identified in the fly, have suggested how this may be accomplished at a neuronal level based on information about relative motion between the target and the background. We experimented with the use of this “small-field system model” for the tracking of small moving targets by a simulated fly in a cluttered environment and discovered some functional limitations. As a result of these experiments, we propose elaborations of the original small-field system model to support stronger effects of background motion on small-field responses, proper accounting for more complex optical flow fields, and more direct guidance toward the target. We show that the elaborated model achieves much better tracking performance than the original model in complex visual environments and discuss the biological implications of our elaborations. The elaborated model may help to explain recent electrophysiological data on FD cells that seem to contradict the original model.

## 1 Introduction

It is well known that male flies visually track small moving targets for the purpose of mating or predation. Previous authors have suggested that this *figure-ground discrimination* is accomplished based on the relative motion of the target to the background (Land and Collett 1974), and the neuronal basis for this discrimination in the lobula plate of the fly has been extensively explored and modeled.

Originally derived from behavioral experiments (Reichardt and Poggio 1979), the model of Reichardt et al. (1983) postulated neuronally plausible mechanisms for figure-ground discrimination based on unidentified cells. Upon the discovery of a new class of lobula plate tangential neurons known as figure detection (FD) cells

(Egelhaaf 1985a,b) sensitive to small moving objects but not to wide-field motion, the model was further developed (Egelhaaf 1985c; Reichardt et al. 1989) to predict torque responses of fixed flies in the presence of rotational visual stimuli. The authors proposed that two parallel systems operate in the lobula plate of the fly: a large-field motion system, associated behaviorally with the compensatory optomotor response and neuronally with horizontal cells that respond most strongly to wide-field stimulation (Hausen and Egelhaaf 1989), and a small-field system (referred to as the SF system), employing FD cells, which is most sensitive to small moving objects. This small-field sensitivity was predicted to be the result of inhibition from wide-field neurons, and thus the model response to small-field motion decreases as the wide-field background motion increases. Lobula plate tangential neurons (CH cells) that respond to wide-field stimuli and are likely to inhibit the response of FD cells were later identified (Warzecha et al. 1993; Egelhaaf et al. 1993), supporting the model. However, more recently Kimmerle and Egelhaaf (2000) played optical flow patterns generated by a tethered flying fly back to a fixed fly preparation and found that the response of FD cells, while sensitive to the motion of small objects as expected, showed no strong dependence on background motion, implying that the SF-system model is incomplete.

Using the most advanced model of the SF system available, we have experimented in simulation with visual tracking of small moving targets in a cluttered visual environment. Other authors have performed related experiments (Huber and Bülthoff 1998; Korrel 2000), but none actually implemented the biological small-field model. In our experiments, the simulated insect flies forward, tracking a small moving target in a two-dimensional (2D) environment of fixed objects, and thus our visual stimuli are significantly different from those of Reichardt et al. (1989) in which the fly was fixed and the environment rotated around it. We found that, in order to achieve satisfactory visual tracking performance in cluttered environments, it was necessary to elaborate the model in three ways: first, to strengthen the effects of background motion on small-field responses; second, to account for the

Correspondence to: C. M. Higgins  
(e-mail: higgins@ece.arizona.edu  
Tel.: +1-520-6216604, Fax: +1-520-6212478)

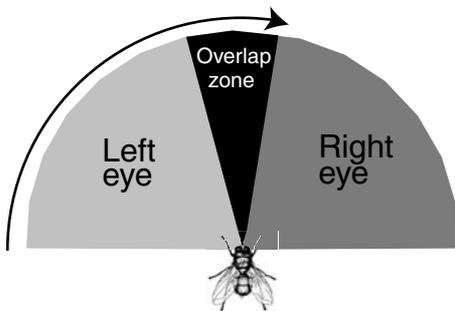
possibility of background stimuli that do not rotate; and third, to allow for proper tracking of targets moving toward the midline of the simulated fly. In this paper, we describe our elaborated version of the SF-system model, present the simulation results of a fly tracking a small moving object with the model in a cluttered environment, and suggest the biological implications of our modifications. An earlier version of this research was previously published in thesis form (Pant 2003).

## 2 Methods

Simulations of visual target tracking with a model fly were conducted in Matlab (The Mathworks, Natick, MA, USA). The fly was situated in a 2D arena  $300 \times 300$  arbitrary space units in size. The visual field of each eye of the fly was a one-dimensional (1D) array of 110 pixels. The position of each pixel corresponded uniformly to an angle from the center of the simulated fly, as illustrated in Fig. 1.

A moving “target” and fixed “background objects” were visible to the fly on a black background, each being a stereotypical object with a brightness in the fly’s visual field proportional to its distance from the light source (described in detail below). The horizontal size of each object in the visual field varied inversely with its distance from the fly. The target was distinguished from background objects only by the fact that the target moved in the arena while background objects were fixed in position. Aside from the target and the fixed background objects, no part of the arena was ever visible. The visual field of the fly was constructed geometrically at each timestep by taking into account the fly’s position relative to objects in the arena.

The fly’s state was specified by an  $x, y$  position in the arena and by an orientation  $\theta$  in which, analogously to a flying insect, it could translate. The fly was assigned a fixed translatory velocity  $v$  of 18 space-units/s, and the angular velocity of the fly was computed directly proportional to the motor response  $R_{SF}(t)$  produced by the SF-system model (17). Thus the fly’s state evolved over time as:



**Fig. 1.** Visual fields of the simulated fly. Angles to the left, indicated by the light gray area, are seen only by the left eye. Near the midline, the black area indicates a range of angles seen by both eyes. The arrow indicates the entire angular extent of the left eye. Angles to the right seen only by the right eye are indicated by the dark gray area. Each eye covered the angles from  $90^\circ$  to one side of the midline to  $9^\circ$  on the other side with 110 pixels. The two eyes thus had an overlapping zone of 20 pixels ( $18^\circ$  of visual angle) centered on the midline; the combined total field of view of the simulated fly covered  $180^\circ$

$$\dot{\theta}(t) = g \cdot R_{SF}(t), \quad (1)$$

$$\theta(t+1) = \theta(t) + \dot{\theta}(t) \cdot \tau, \quad (2)$$

$$x(t+1) = x(t) + v \cdot \cos(\theta(t)) \cdot \tau, \quad (3)$$

$$y(t+1) = y(t) + v \cdot \sin(\theta(t)) \cdot \tau, \quad (4)$$

where  $g$  is a control loop gain factor set to stabilize the performance of the fly at a value of 10 unless otherwise specified, and  $\tau$  is the timestep (time between frames) of the simulation (10 ms). In order to emulate a more realistic physical fly, the angular velocity of the simulated fly  $\dot{\theta}$  was limited to a maximum of  $\pi$  radians per second. Simulations were terminated when the  $x, y$  position of the simulated fly went outside of the arena, when the fly moved vertically ahead of the target, or when the distance between fly and target dropped below a threshold of 5 space-units, indicating a collision.

The fly began each simulation at the bottom center of the arena (coordinates 150, 30). The initial vertical position of the target was 60 space-units ahead of the fly, and its horizontal starting position was allowed to be one of nine equally spaced positions from 90 to 210 space-units to allow for a range of interception difficulties. In all simulations the target’s initial position was within the fly’s visual field. The target moved at a fixed vertical speed of 12 space-units/s and moved horizontally with a sinusoidally varying speed peaking at 24 space-units/s and with a frequency of 0.4 cycles per simulated second. Refer to Fig. 6 for an example interception.

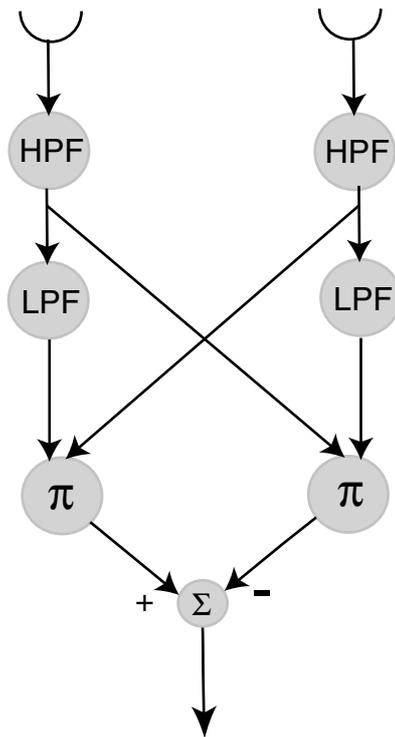
A simulated light source was located at the center of the arena. In order to allow variation of the strength of background clutter, the brightness  $B$  of objects ( $0 \leq B \leq 1$ ) varied inversely with their distance  $D$  from the light source as

$$B = \frac{1}{1 + D/D_{\text{half}}}, \quad (5)$$

where  $D_{\text{half}}$  is a constant that specifies the distance from the light source at which an object has 50% brightness. This formula is in agreement with modern computer graphics software, which generally considers the inverse square-law reduction of brightness predicted from a point light source to be too severe for synthesizing realistic images. Increasing  $D_{\text{half}}$  increases the brightness of both the target and the background objects, but it can also be shown that a larger value of  $D_{\text{half}}$  decreases the contrast of the target relative to background objects, thus making the target harder to detect.

Elementary motion detection was computed on the sequence of 1D visual images seen by the simulated fly using an array of 109 Hassenstein–Reichardt (HR) motion detectors (Hassenstein and Reichardt 1956; Van Santen and Sperling 1985) in each eye (Fig. 2). First-order high-pass and low-pass filters had time constants of 100 ms. As in the original simulations of the SF-system model (Reichardt et al. 1989), our elementary motion detection model does not employ contrast saturation (Egelhaaf and Borst 1989).

In order for the performance of the large-field model (25), the sum of a large number of motion detector outputs, to be comparable to the SF-system model, it was



**Fig. 2.** Hassenstein–Reichardt elementary motion detection model. Semicircles at top indicate photoreceptors. *HPF* and *LPF* indicate, respectively, high-pass and low-pass filters,  $\pi$  indicates a multiplication, and  $\Sigma$  a sum. This model produces an output whose sign indicates stimulus direction

necessary to introduce a relative gain of 1/100. This gain was chosen empirically to roughly match the tracking performance of the large-field and small-field models in the presence of a target and no background objects, in which situation they should be comparable.

To allow comparison between models, a numerical metric of tracking performance in each simulation was devised. This metric requires the computation in each frame  $i$  of the distance  $D_{\text{target}}(i)$  from the simulated fly to the target and the angle  $\theta_{\text{target}}(i)$  between the fly's heading direction and the direction toward the target. The metric is based on the root mean square (RMS) value of the product of these two signals:

$$M = M_{\text{ref}} - \sqrt{\frac{1}{N_f} \cdot \sum_{i=1}^{N_f} [D_{\text{target}}(i) \cdot \theta_{\text{target}}(i)]^2}, \quad (6)$$

where  $N_f$  is the number of frames in the simulation and  $M_{\text{ref}}$  is the RMS value obtained by using zero angular velocity throughout the experiment. This metric penalizes turning away from the target when the distance is great but allows for large relative target angles when distances are small, as is inevitable at the end of a successful interception. Negative values of the metric  $M$  indicate turning away from the target, zero indicates movement straight ahead, and positive values indicate turning toward the target, and thus good tracking performance.

For all experimental data shown, simulations were run with all nine possible initial horizontal target positions,

and the median metric over these simulations was taken. Since the difficulty of the interception varies with initial target position, the expected tracking performance is different for each position, and thus standard deviation is not shown.

### 3 Results

In order to perform closed-loop simulations of visual target tracking using the SF-system model, it was first necessary to consider how to use the model to control a simulated fly. In two ways, the torque data presented by Reichardt et al. (1983) are not the torque that would be generated by a fly in flight.

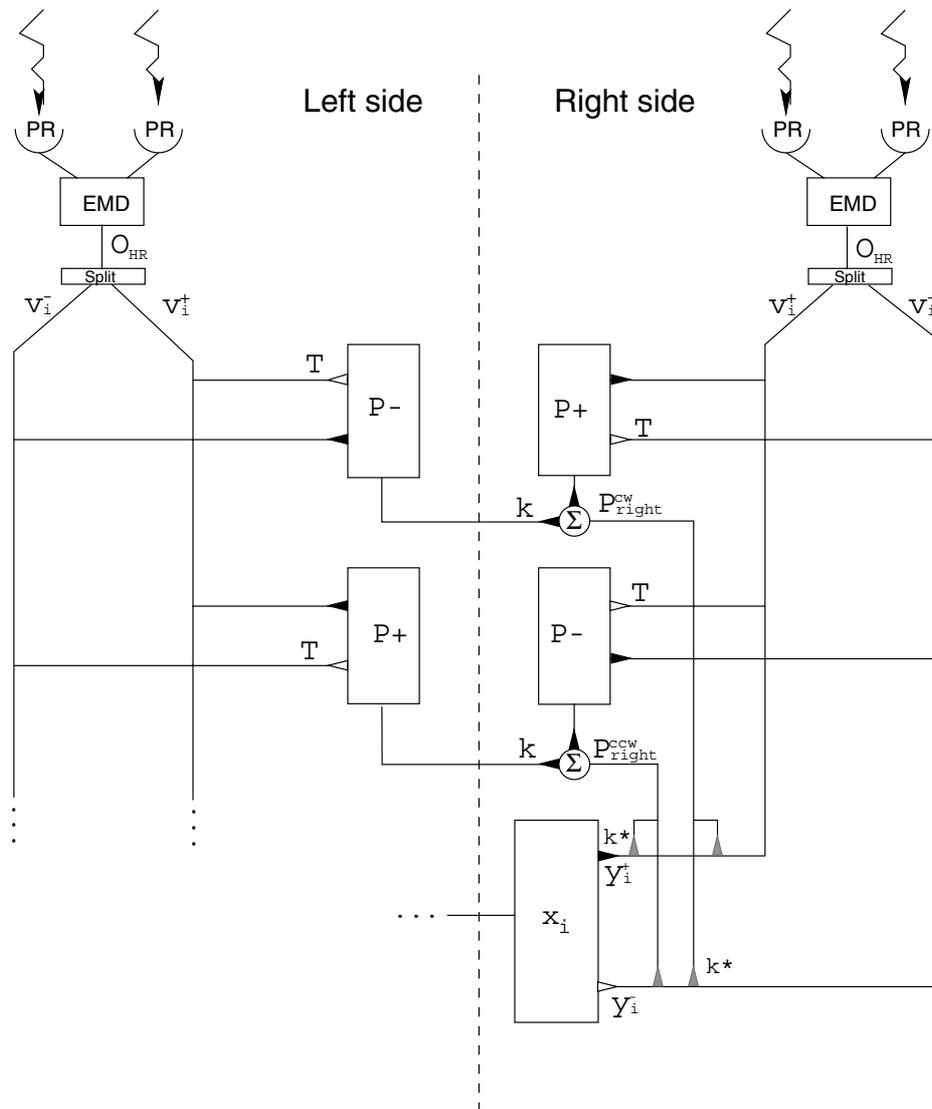
Firstly, in order to easily evaluate “online” the torque response of the fly, Reichardt et al. (1983) processed the raw torque data through an instrument that created a running average (a Princeton signal averager model 4202). This required that the authors add a running average to their model as well in order to be comparable to the torque data collected. However, this running average is not appropriately added to the control signal of our simulated fly since it does not represent the real-time torque response of the fly. Additionally, this long-term average would lead to a crippling delay in the control loop.

Secondly, analogously to Kimmerle et al. (2000), we use the output of the model to represent the angular velocity, not the torque (proportional to the derivative of the angular velocity), of our simulated fly. Since the sign of the model output indicates which side of the midline the target is on, use of the model output as torque would force the simulated fly to overshoot the target and track in a “zig-zag” fashion, since it would be impossible to immediately cancel the built-up angular velocity at the time the target crossed the midline. This is in contrast to the extremely stable tracking performance documented below. If our simulated fly were fixed as in the behavioral experiments of Reichardt et al. (1983) by zeroing the angle and angular velocity each timestep, the “torque” generated from use of the SF-system model in this way would indeed be exactly the model output, since the change in angular velocity (torque) at each timestep would be from zero to the model output.

In the following sections, the model of the SF system as proposed by Reichardt et al. (1989) is first mathematically described, then the elaborations necessary for small-target tracking are presented, and finally our tracking simulation results with the elaborated model are given.

#### 3.1 The small-field system model

In the model of Reichardt et al. (1989), the input to the SF system is provided by a retinotopic array of elementary motion detectors (EMDs), each implementing the Hassenstein–Reichardt motion detection model (Hassenstein and Reichardt 1956; Van Santen and Sperling 1985). EMDs in both eyes have a positive response to front-to-back motion from the midline of the insect. According to their preferred direction of motion, FD cells are excited



**Fig. 3.** Model of the SF system based on the FD cell. Excitatory and inhibitory synapses are indicated by *black* and *white triangles*, respectively. Shunting inhibition is indicated by *gray triangles*.  $\Sigma$  indicates a sum. Responses from neighboring photoreceptors (PR) are input to EMDs. EMD outputs are split into positive and negative components. These components are aggregated into directionally selective monocular pool cells ( $P^+$ ,  $P^-$ ) and then into CW and CCW binocular

lar pool cell responses ( $P_{\text{right}}^{\text{CW}}$  and  $P_{\text{right}}^{\text{CCW}}$ ). These directionally selective binocular pool cells interact via shunting inhibition with the individual motion detector output channels prior to their combination by unit  $x_i$ . For simplicity, only the right-side computation is diagrammed. The motor output is computed as the difference between the spatially summed  $x_i$  from the two sides. Modified from Reichardt et al. (1989)

by one of these types of motion detectors and inhibited by the other. The inhibition is brought about by synaptic transmission from *pool cells* that aggregate the response of the motion detectors over the entire visual field.

A block diagram of the model is presented in Fig. 3. Two adjacent photoreceptors provide visual input to individual EMDs. The motion detector output  $O_{\text{HR}}$  is split into front-to-back responses  $v_i^+ > 0$  and back-to-front responses  $v_i^- > 0$  such that  $O_{\text{HR}} = v_i^+ - v_i^-$ . Only one of  $v_i^+$  and  $v_i^-$  is nonzero at any given time. Two sets of monocular pool cells spatially sum the EMD responses and are thus direction selective. The  $P^+$  pool cells have positive responses to front-to-back motion and are inhibited by back-to-front motion. Similarly, the  $P^-$  pool cells respond positively to back-to-front motion. The inhibitory input to each pool cell is weighted by a factor  $0 < T < 1$ ,

and thus the monocular pool cell response is given by the following expressions:

$$P^+(t) = \sum_{i=1}^N [v_i^+(t) - T \cdot v_i^-(t)], \quad (7)$$

$$P^-(t) = \sum_{i=1}^N [v_i^-(t) - T \cdot v_i^+(t)], \quad (8)$$

where  $N$  is the number of EMDs associated with a particular eye. Monocular pool cells from both sides of the brain interact to form clockwise (CW) and counterclockwise (CCW) binocular pool cells with a relative contribution of ipsilateral and contralateral input of  $0 < k < 1$ :

$$P_{\text{right}}^{\text{CW}}(t) = P_{\text{right}}^+(t) + k \cdot P_{\text{left}}^-(t), \quad (9)$$

$$P_{\text{right}}^{\text{CCW}}(t) = P_{\text{right}}^-(t) + k \cdot P_{\text{left}}^+(t). \quad (10)$$

Left-side binocular pool cells  $P_{\text{left}}^{\text{CW}}$  and  $P_{\text{left}}^{\text{CCW}}$  are each computed from the same two monocular pool cells as above, but with the parameter  $k$  on the response from the contralateral eye. These binocular pool cell responses are comparable to the response of wide-field tangential cells in the lobula plate of the fly, and specifically the identified CH cell has been proposed as a strong candidate for one type of pool cell (Warzecha et al. 1993; Egelhaaf et al. 1993). The CW and CCW pool cell responses interact with individual EMD responses. The relative contribution of both the pool cells to individual EMDs is weighted by a factor  $0 < k^* < 1$  to maintain the direction selectivity. The response of the detector channels are normalized through shunting inhibition by the pool cell responses. Assuming a saturation of the output approximated by an exponent  $0 < q < 1$ , the normalized detector channels are given by:

$$y_{i,\text{right}}^+(t) = \frac{v_{i,\text{right}}^+(t)}{\beta + [P_{\text{right}}^{\text{CW}}(t) + k^* \cdot P_{\text{right}}^{\text{CCW}}(t)]^q}, \quad (11)$$

$$y_{i,\text{right}}^-(t) = \frac{v_{i,\text{right}}^-(t)}{\beta + [P_{\text{right}}^{\text{CCW}}(t) + k^* \cdot P_{\text{right}}^{\text{CW}}(t)]^q}, \quad (12)$$

where  $\beta$  is a constant related to shunting inhibition. To properly emulate shunting inhibition, only positive values of the sum of binocular pool cells are used in the normalization; negative values are set to zero. Left-side units  $y_{i,\text{left}}^+$  and  $y_{i,\text{left}}^-$  are computed similarly. The excitatory and inhibitory responses are nonlinearly combined:

$$x_{i,\text{right}}(t) = [y_{i,\text{right}}^+(t)]^n - [y_{i,\text{right}}^-(t)]^n, \quad (13)$$

$$x_{i,\text{left}}(t) = [y_{i,\text{left}}^+(t)]^n - [y_{i,\text{left}}^-(t)]^n, \quad (14)$$

where the exponent  $n$  is used to enhance large values. The responses from these units on either side of the brain are accumulated:

$$X_{\text{right}}(t) = \sum_{i=1}^N x_{i,\text{right}}(t), \quad (15)$$

$$X_{\text{left}}(t) = \sum_{i=1}^N x_{i,\text{left}}(t). \quad (16)$$

These  $X$  expressions respond comparably to identified FD cells, directionally sensitive to the motion of small objects anywhere in a wide visual field but not sensitive to wide-field motion. To truly represent the firing-rate response of a spiking FD cell, the response of the  $X$  units must of course be rectified (positive part only taken). To obtain the model instantaneous output, these values are then subtracted from each other:

$$R_{\text{SF}}(t) = X_{\text{left}}(t) - X_{\text{right}}(t). \quad (17)$$

The final model output, compared to torque responses recorded from a fly, is a combination of the real time response  $R_{\text{SF}}(t)$  and a running average of the response. The

running average was required due to the method of torque data collection, as discussed earlier.

The numerical parameters used in our simulations of the original SF-system model (7)–(17) were  $T = 0.3$ ,  $k = 0.7$ ,  $\beta = 0.1$ ,  $k^* = 0.3$ ,  $q = 0.5$ , and  $n = 3$ . With the exception of  $\beta$ , these parameters were chosen to match those used by Reichardt et al. (1989). Likely due to differences in our implementations of elementary motion detection,  $\beta$  had to be set at 0.1 rather than 0.6 as used in the original authors' simulations to allow the normalization operation to work as intended.

### 3.2 The elaborated SF-system model

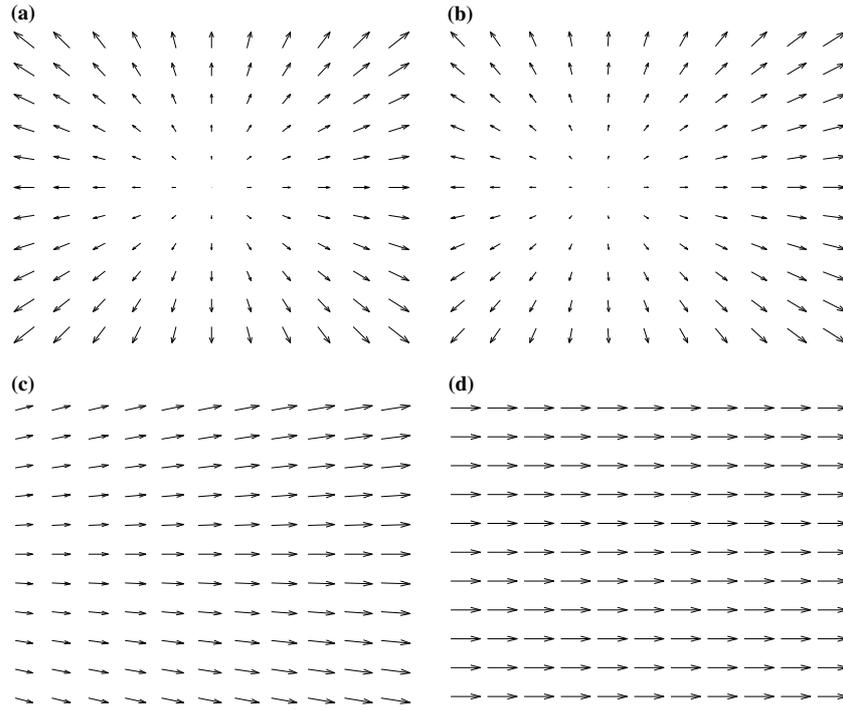
While experimenting with the use of the SF-system model for target tracking, we found it necessary to elaborate the original model in three ways. Firstly, we modified the model so that the parameter  $k^*$  strengthens the normalization operation rather than weakening it. Secondly, we incorporated the possibility of more complex background motion than in the original model. Finally, we modified the model so that the SF-system model would always guide the simulated fly toward the target, because in some cases the original model would guide the fly away. These elaborations are described in detail below.

**3.2.1 Enhanced normalization.** The parameter  $k^*$  in the model describes proposed interactions between binocular pool cells on a particular side of the brain. In the original model, the normalization of a particular small-field motion output (for example,  $v_{i,\text{right}}^+$ ) was by the sum of the pool cell that would be excited by that particular direction of motion ( $P_{\text{right}}^{\text{CW}}$ ) and  $k^*$  times the pool cell *inhibited* by that direction of motion ( $P_{\text{right}}^{\text{CCW}}$ ). Thus the larger the parameter  $k^*$ , the weaker the normalization operation, and for  $k^* = 1$  the normalization operation becomes virtually uniform for all  $y_i$  regardless of the pattern of visual motion. The mixing of complementary and contradictory information implied by  $k^*$  ensures that the ratio between strongly and weakly activated pool cells is always small, and thus that the resulting enhancement of target signal strength is minimized. For this reason, the original model is only weakly able to detect targets in cluttered backgrounds.

We propose that  $k^*$  be a *negative* parameter, indicating inhibition between binocular pool cells before the shunting operation. In this formulation,  $k^*$  enhances the normalization by combining information that is consistent with the motion being normalized. This modification increases the ratio between strongly and weakly activated pool cells and greatly strengthens the response to motion that is inconsistent with the estimated background motion.

A value of  $k^* = -0.3$  was used in simulations of the elaborated model with enhanced normalization. All other constants were as specified for the original model.

**3.2.2 Allowing for more complex background motion.** The SF-system model as proposed by Reichardt et al. (1989) modeled the target fixation behavior of a fly in a fixed



**Fig. 4.** Optical flow generated during pursuit. **a** Pure expanding optical flow generated by platform translation with no component of rotation. **b** Expanding optical flow with shifted FOE due to small

rotational component. **c** Large rotational component of platform motion while also translating. **d** Pure rotational motion, generating a linear flow field

position around which visual stimuli rotated. Therefore, it did not take into consideration the complexity of the optical flow field observed by a moving fly pursuing a target. During pursuit, the translatory motion of the fly is coupled with the rotational motion required to keep the target in the center of the view field. Translatory motion of the fly alone produces an expanding optical flow field with the focus of expansion (FOE) at the midline (Gibson 1950), as illustrated in Fig. 4a. When translating with an added component of rotational motion, the FOE is shifted to one side of the view field due to superposition of the linear optical flow generated by the rotation upon the expanding flow generated by translation. If the rotation rate is small (Fig. 4b), the shift of the FOE is small, and the field can be approximated as a pure expanding optical flow field. However, if the rotation rate is large (Fig. 4c), the FOE shifts completely out of view, and the optical flow is better approximated by a pure linear pattern (Fig. 4d).

The purpose of the shunting inhibition “normalization” in the SF-system model is to reduce the strength of small-field motion signals that are consistent with the estimated background motion. Only small-field motion that is *distinct in sign* from the estimated background motion pattern should lead to strong responses from the model. By synthesizing only pool cell signals that correspond to rotation ( $P_{\text{right}}^{\text{CW}}$  and  $P_{\text{right}}^{\text{CCW}}$ ), the possible estimates for background motion are limited to rotation of the fly. In the elaborated model, we further synthesize pool cells for optical flow expansion, which leads to front-to-back (positive) motion on both eyes and related pool cells for contraction:

$$P_{\text{right}}^{\text{exp}} = P_{\text{right}}^{+} + k \cdot P_{\text{left}}^{+}, \quad (18)$$

$$P_{\text{right}}^{\text{con}} = P_{\text{right}}^{-} + k \cdot P_{\text{left}}^{-}. \quad (19)$$

Expressions for the left-side  $P_{\text{left}}^{\text{exp}}$  and  $P_{\text{left}}^{\text{con}}$  are computed similarly, with the parameter  $k$  multiplied by the contralateral pool cell.

It is not desirable to simultaneously normalize by all possible background motion types, since this would lead to small responses in all cases. Thus we must choose an estimate of the background motion based upon our visual input. There are many ways to do this, but we propose that either an estimate of expansive or rotatory background motion be chosen based on the relative pool cell activations. In our elaborated model, when the sum of the pool cells for expansion exceeds that of rotation,

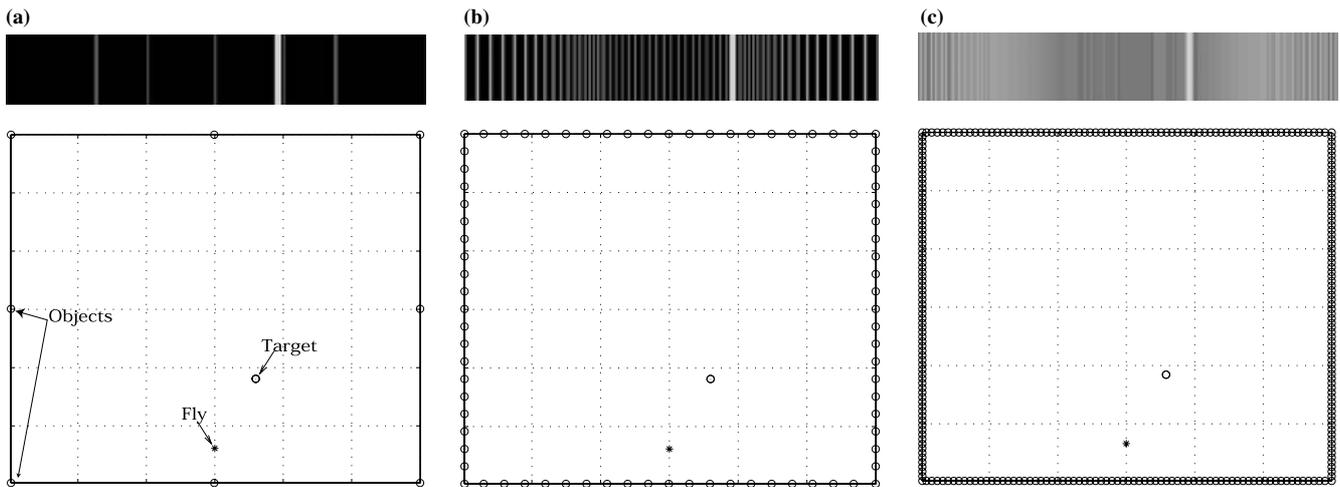
$$P_{\text{right}}^{\text{exp}} + P_{\text{right}}^{\text{con}} + P_{\text{left}}^{\text{exp}} + P_{\text{left}}^{\text{con}} > P_{\text{right}}^{\text{CCW}} + P_{\text{right}}^{\text{CW}} + P_{\text{left}}^{\text{CCW}} + P_{\text{left}}^{\text{CW}}, \quad (20)$$

the estimated background motion is approximated by a pure expansion (Fig. 4a). Under this condition, normalization of small-field outputs is by expansion and contraction pool cells:

$$y_{i,\text{right}}^{+}(t) = \frac{v_{i,\text{right}}^{+}(t)}{\beta + [P_{\text{right}}^{\text{exp}}(t) + k \cdot P_{\text{right}}^{\text{con}}(t)]^q}, \quad (21)$$

$$y_{i,\text{right}}^{-}(t) = \frac{v_{i,\text{right}}^{-}(t)}{\beta + [P_{\text{right}}^{\text{con}}(t) + k \cdot P_{\text{right}}^{\text{exp}}(t)]^q}. \quad (22)$$

Expressions for the left-side  $y_{i,\text{left}}^{+}$  and  $y_{i,\text{left}}^{-}$  are computed similarly. When the condition of (20) is not met, the optical



**Fig. 5.** Example visual fields seen by the simulated fly. Each panel shows at *top* the visual field of the simulated fly and at *bottom* the corresponding top view of the arena at the beginning of a simulation. **a** The case of eight objects present on the boundary of the arena. **b** The same information for 80 objects. **c** For 300 objects.  $D_{\text{half}}$  was set at

flow field is approximated as purely rotational (Fig. 4d), and the output of the normalization stage is represented as in the original model by (11)–(12). In this manner, the visual environment is used to decide which estimate is more appropriate.

**3.2.3 Guidance from nondirectional motion.** The original model suggested small-field units  $x_{i,\text{left}}$  and  $x_{i,\text{right}}$  [(13) and (14)] that would respond strongly and in a signed directional manner to small moving objects. These units were spatially summed and used to produce the motor output of the system. This directionality implies that the turn angle generated by the SF system as formulated for a target moving toward the midline will guide the simulated fly away from the target. For this reason, the original SF-system model is unable to track targets that make significant course changes during the interception (see Discussion).

We modified the model to initiate a turn toward a possible target in all cases. Rather than summing the  $x_{i,\text{left}}$  and  $x_{i,\text{right}}$  units directly into the model output, they are summed in absolute value, making them responsive to small moving targets in a nondirectional manner. Therefore, target motion in either direction to the left of the midline will lead to a turn toward the left. This change requires us to replace (15) and (16) in the original model with

$$X_{\text{right}}(t) = \sum_{i=1}^N |x_{i,\text{right}}(t)|, \quad (23)$$

$$X_{\text{left}}(t) = \sum_{i=1}^N |x_{i,\text{left}}(t)|. \quad (24)$$

The absolute value operation cannot be taken after the summation because this would allow a large number of small  $x_i$  outputs of one sign (corresponding to normalized background motion) to cancel a small number of  $x_i$  outputs of the opposite sign (corresponding to a target).

300. In the top view, *circles on the boundary* represent fixed objects, an *asterisk* represents the simulated fly, and a *circle inside the arena* represents the target. Note that the target brightness is higher than that of the background objects due to its smaller distance from the light source

Since the  $X_{\text{right}}$  and  $X_{\text{left}}$  units corresponded in the original model to identified FD cells, and since these units as formulated above respond to small-field motion without regard to direction, this modification must be reconciled with data on FD cells (see Discussion).

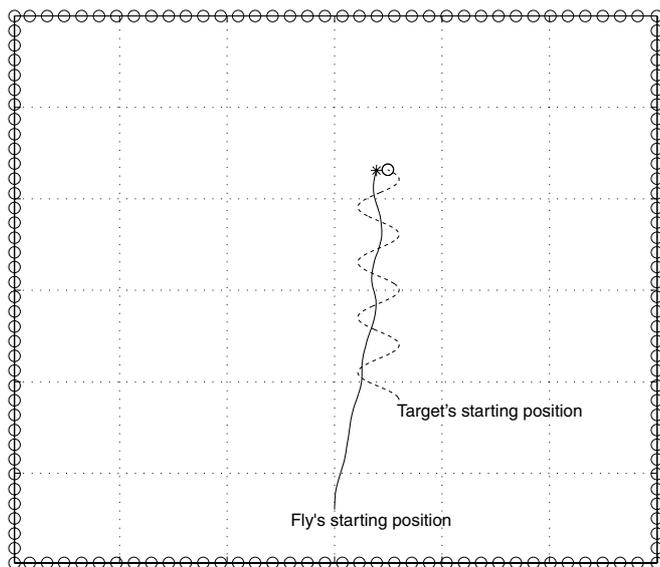
### 3.3 Simulations with the elaborated SF-system model

In order to investigate the tracking performance of the elaborated SF-system model, we simulated a fly moving through a 2D arena (see Methods for details). A number of fixed objects were placed on the boundary of the arena to generate a cluttered background. A sinusoidally moving target, emulating the motion of another insect, was also placed in the arena. Figure 5 illustrates some of the views seen by the simulated fly, and Fig. 6 shows an example of the tracking performance of the model.

In order to measure the effect of the background objects on tracking, we compare the tracking performance of the elaborated SF-system model to a naive tracking model that simply turns toward any motion that it sees. This system may be expressed in terms of the pool cells of the SF-system model as:

$$R_{\text{LF}}(t) = \left| P_{\text{left}}^+ - P_{\text{left}}^- \right| - \left| P_{\text{right}}^+ - P_{\text{right}}^- \right|. \quad (25)$$

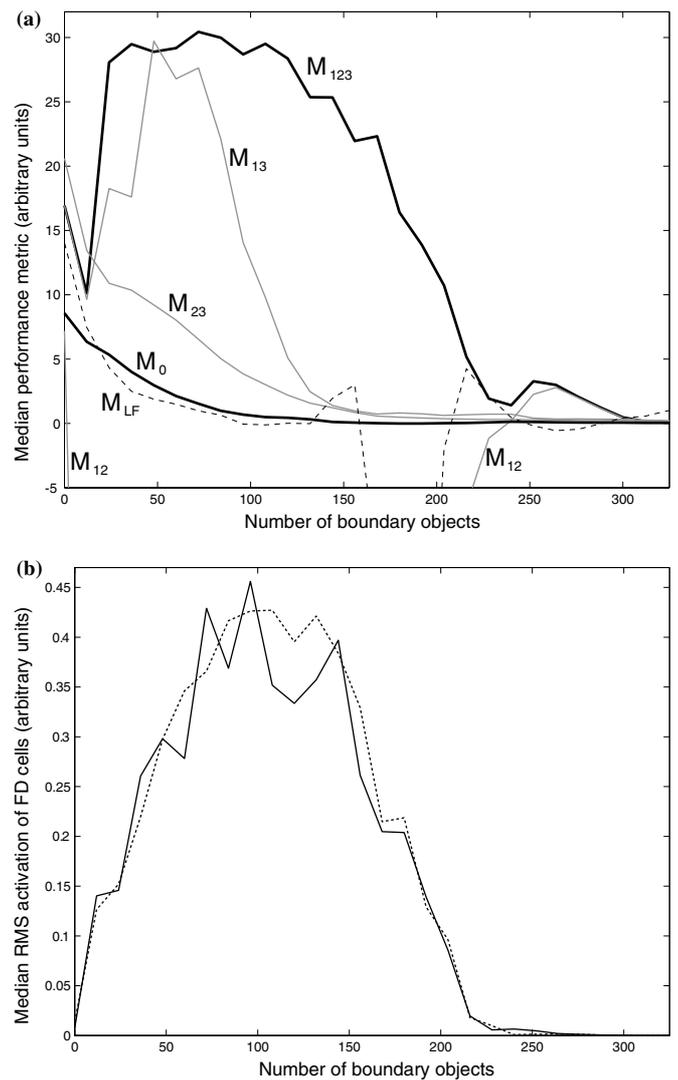
In the presence of a single moving target object and no background objects, this model will guide the simulated fly appropriately. In fact, due to the square-law contrast response of the HR motion detector (Van Santen and Sperling 1985), it is not unreasonable to expect that this model would even be able to reject low-contrast background objects to some extent. We refer to this tracking system as the LF system since it responds to large-field motion as well as small-field motion, although this system is much less sophisticated than the LF system of Reichardt et al. (1989).



**Fig. 6.** Path of the simulated fly as it tracks a moving object during a typical experiment. The *solid line* shows the path of the fly and the *dashed line* the path of the target. This simulation employed 150 background objects with  $D_{\text{half}}$  set at 300 and ended with a simulated collision

Tracking performance is measured for each simulation by the metric of (6), computed from the angle and distance between the simulated fly and the target over the course of the interception. A zero metric indicates that the fly has flown straight ahead without turning, as it might with no visual input at all. Positive metrics indicate a reduction of angle and distance to target relative to the “blind” case, and thus represent good performance. Negative metrics indicate turns away from the target, and thus worse performance than the “blind” case. Each simulation was run for a range of initial target positions (see Methods), including initial positions to the left of, in front of, and to the right of the fly initial position. The difficulty of the tracking experiment increases as the initial target position moves away from directly ahead of the fly in either direction due to the greater initial distance to target and the larger turn angle required to center the target in the visual field. Each data point shows the median tracking performance over the entire range of initial target positions, thus encompassing the overall tracking performance of the model.

To investigate the strength of the SF system in rejecting background clutter, we varied the number of objects on the walls of the arena. Tracking performance is plotted in Fig. 7a. While the fully elaborated model ( $M_{123}$ ) is able to track the target in the case of no background objects, as the number of background objects increases its performance actually improves. This results from the normalization operation. In the case of no background objects, the model’s estimate of the background motion comes entirely from the target, which is thus reduced in strength. However, as background objects appear, it is possible to estimate the background motion separately from the target, after which target motion contrary to the estimated background motion is strongly amplified. In the range from approximately 25 to 150 background



**Fig. 7.** Effect of variation of the number of boundary objects on tracking simulations. Simulations were conducted with  $D_{\text{half}} = 300$ . **a** Tracking performance of the SF and LF systems. For each point on the graph, a simulation was run for all nine possible initial target positions (see Methods), and for each simulation the metric  $M$  (6) was computed. Plotted is the median of the metric over all possible initial target positions. The *bold line at top* ( $M_{123}$ ) indicates the fully elaborated model. The *bold line at bottom* ( $M_0$ ) has all elaborations turned off, and thus is equivalent to the original model. The *gray lines* marked  $M_{23}$ ,  $M_{13}$ , and  $M_{12}$  show the effect of the removal of each elaboration individually. The *dashed line* ( $M_{LF}$ ) indicates the LF system response. **b** Response of simulated FD cells in the elaborated model. Plotted is the median RMS activation of simulated FD cells (in arbitrary units) over all possible initial target positions. The *solid line* indicates the left FD cell and the *dashed line* the right FD cell. The median response of both cells is quite similar, indicative of a range of initial target positions to the left and right of the fly. Both cells become activated as the number of background objects increases and then become inactive as estimation of the background motion becomes more difficult

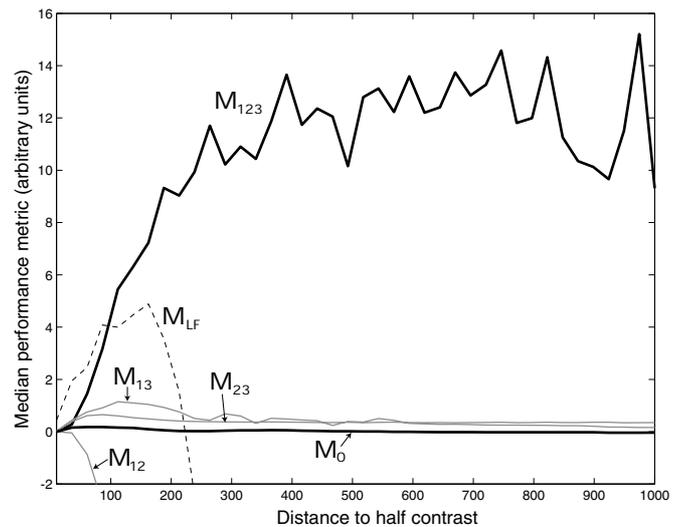
objects, the performance of the model holds roughly constant. As shown in Fig. 7b, the RMS activation of simulated FD cells [(15) and (16)] during the tracking experiment also holds roughly constant over this range. As the number of objects continues to increase beyond about 170, the background viewed by the simulated fly approaches a continuum with brightness graded by dis-

tance, and the tracking actually becomes easier since edges of individual objects no longer create strong motion signals. In fact, when the number of objects is increased beyond about 230, all models are able to track the target, although the contrast of the target relative to the background is very small, and thus control signals are very weak.

Figure 7a also shows the effects on the model of the removal of each elaboration individually and of the removal of all three. With  $k^*$  positive ( $M_{23}$ ), the model is still able to track the target, but with significantly less gain. With a fixed assumption of rotation ( $M_{13}$ ), performance is good until the number of background objects reaches about 80 and expansive motion of the background becomes strong, after which performance declines. With the use of directional motion for guidance ( $M_{12}$ ), the model turns strongly away from the target. With all three elaborations removed ( $M_0$ ), the gain of the model is so small that it fails to turn significantly toward the target. The LF system ( $M_{LF}$ ) is able to track the target only for a very small number of objects, after which it devolves into an optomotorlike response, fixated by the background. In some cases for a large number of background objects, this response actually leads to turns away from the target.

Using a relatively large number of background objects (200), we varied the parameter  $D_{\text{half}}$ , which affects not only the brightness of the target but the contrast of the target relative to the background as well. The results of these experiments are shown in Fig. 8. Since the strength of the EMD output is proportional to the square of contrast, the angular velocities generated by the model for low values of  $D_{\text{half}}$  are close to zero. As  $D_{\text{half}}$  increases, the performance of the fully elaborated model ( $M_{123}$ ) improves and saturates. At the highest value of  $D_{\text{half}}$  shown, the tracking problem is very hard due to a low contrast of the target relative to the background, and only the fully elaborated model is able to successfully track in this case. Figure 8 also shows the effects on the model of the removal of each elaboration individually and of the removal of all three. With  $k^*$  positive ( $M_{23}$ ), or with a permanent assumption of rotation ( $M_{13}$ ), the gain of the model output is so reduced that the fly is virtually unable to turn. As before, with the use of directional motion for guidance ( $M_{12}$ ), the model turns strongly away from the target. With all three elaborations removed ( $M_0$ ), the gain of the model is again so small that it fails to turn significantly toward the target. The LF system ( $M_{LF}$ ) is able to track the target only for very low values of  $D_{\text{half}}$  in which the background contrast is extremely small. When  $D_{\text{half}}$  is sufficient to generate significant background motion, the LF system completely fails to track the target.

Since the gain of the original model ( $M_0$ ) is so small relative to the elaborated model ( $M_{123}$ ), it is reasonable to ask if the elaborations have simply increased the effective control loop gain of the model. To investigate this possibility, we have increased the relative control loop gain of the original model by factors of up to 100 and plotted performance against  $D_{\text{half}}$  in Fig. 9. Only for very low values of  $D_{\text{half}}$  does the original model ever match the performance of the elaborated model, and in no case is the original model able to handle the low contrast between target and



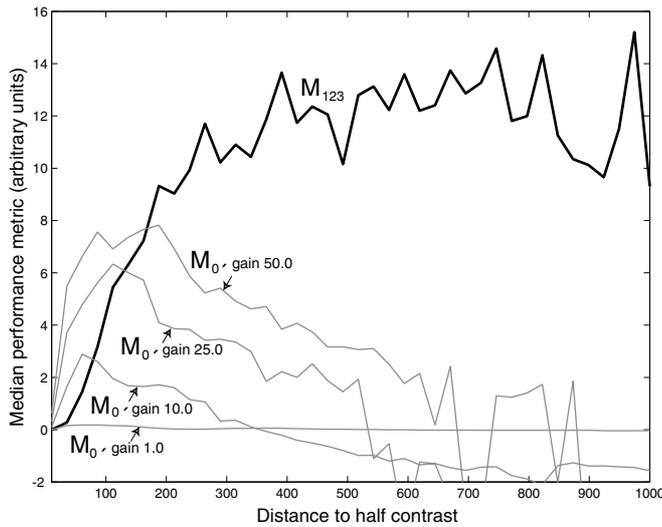
**Fig. 8.** Effect of the variation of  $D_{\text{half}}$ , the distance at which brightness reaches 50%, on tracking performance of the SF and LF systems. Simulations were conducted using 200 boundary objects. For each point on the graph, a simulation was run for all nine possible initial target positions (see Methods), and for each simulation the metric  $M$  (6) was computed. Plotted is the median of the metric over all possible initial target positions. Increasing  $D_{\text{half}}$  increases the brightness of the target, but also increases the brightness of background objects. The *bold line at top* ( $M_{123}$ ) indicates the fully elaborated model. The *bold line at bottom* ( $M_0$ ) has all elaborations turned off, and thus is equivalent to the original model. The *gray lines* marked  $M_{23}$ ,  $M_{13}$ , and  $M_{12}$  show the effect of the removal of each elaboration individually. The *dashed line* ( $M_{LF}$ ) indicates the LF system response

background implied by large values of  $D_{\text{half}}$ . For the most difficult case shown ( $D_{\text{half}} = 1000$ ), the performance of the original model becomes increasingly worse as control loop gain increases.

#### 4 Discussion

We have presented an elaborated version of the small-field system model of Reichardt et al. (1989) and described how the elaborated model may be used to allow a simulated fly to track a small moving target in a background cluttered with stationary distractor objects. The first elaboration addresses an alteration of the interaction of binocular pool cells to strengthen rather than weaken the effects of normalization. The second elaboration addresses proper accounting for expanding optical flow fields as well as the rotational fields included in the original model. The third and final elaboration alters the model to direct a turn toward the target regardless of the direction of its motion. Our simulation results have shown that the elaborated model is capable of tracking a small moving target even in situations in which the original model fails.

While these three elaborations were made to improve tracking performance, the elaborated model can still serve to predict the torque data of Reichardt et al. (1989). The alteration of binocular pool cell interaction increases the magnitude of model response but does not alter the selectivity to small moving targets. The elaborated model will automatically choose rotational flow fields for



**Fig. 9.** Effects of increasing the relative gain of the original model. For each point on the graph, a simulation was run for all nine possible initial target positions (see Methods), and for each simulation the metric  $M$  (6) was computed. Plotted is the median of the metric over all possible initial target positions. While  $D_{\text{half}}$  is swept over the same values as in Fig. 8, the effects of relative gains of 1.0, 10.0, 25.0, and 50.0 on the original model are shown. The **bold line** ( $M_{123}$ ) indicates the fully elaborated model. *Gray lines* ( $M_0$ ) indicate the original model with a variety of relative gains. The beginning of numerical instability is evident in the data for the two highest relative gains shown. With a relative gain of 100 (data not shown), the original model was numerically unstable for  $D_{\text{half}}$  larger than 500 and exhibited unpredictable performance metrics

normalization in the presence of rotational visual stimuli, thus employing the original model equations. The details of response caused by use of turns toward, rather than away from, the target will be smoothed out by the running average operation necessary to compare with the torque data collected.

The first elaboration suggests alteration of the sign of the parameter  $k^*$  such that binocular pool cells on the same side of the brain interact in an inhibitory fashion before shunting inhibition of the small-field motion outputs. In the original model,  $k^*$  was the only parameter an *increase* in the value of which would detract from tracking performance. A positive value of  $k^*$  throws away information by mixing contradictory signals, whereas a negative value makes best use of the redundancy inherent in having two binocular pool cells tuned for optical flow fields of opposite signs. This use of redundancy is analogous to that implied by the parameter  $T$  in the model that combines motion detectors of opposite signs and to the parameter  $k$  in the model that combines monocular pool cells.

The second elaboration involving proper accounting for expansive flow fields becomes especially important at the higher values of  $D_{\text{half}}$  shown in Fig. 8. In these cases, the background motion pattern is very strongly expansive. The original model, expecting rotational flow fields, experiences equal parts of CW and CCW rotation and thus normalizes all motion signals by the same quantity. It is thus unable to track the target separately from the background. This elaboration allows magnification

of motion signals that are inconsistent with an expansive pattern of background optical flow. Neuronally, this elaboration requires two more binocular pool cells on each side of the brain sensitive to expanding and contracting optical flow fields, which might be identified with a number of tangential cells in the lobula plate of the fly (Borst 1991). Although visual experience with contracting optical flow fields may be rare for a fly, the presence of such a unit with the use of negative  $k^*$  increases the strength of normalization in the presence of expansive visual stimuli. The pool cell unit for contraction is included for symmetry, but its presence is not necessary to the functioning of the model. The alteration of the shunting operation from expansive to rotational based on the relative activation of pool cells might be accomplished by shunting inhibition of the rotational normalizing pathway from the expansion and contraction pool cells, and vice versa. Such a shunting inhibitory connection could act as a “veto” to prevent normalization by the class of motion not being experienced.

Due to the formulation of the original model, a target moving toward the midline of the insect caused a turn away from the target. In the case of a target making unpredictable course changes, this turn could lead to loss of sight of the target. The third elaboration alters the expression for computing the torque output so that a target on the left always leads to a turn to the left, resulting in a nondirectional small-field motion unit. This use of nondirectional motion in the guidance of the system has implications for the model of identified neurons, since the expressions altered were those identified with FD cells. Since nondirectional FD cells have not been suggested by the electrophysiology, it is necessary to reformulate the expressions of (23) and (24) in terms of directional cells. This can be done by synthesizing directional FD model cells as in the original model

$$X_{\text{FD1}}(t) = \sum_{i=1}^N [y_i^+(t)]^n - [y_i^-(t)]^n \quad (26)$$

and postulating the existence of a directional FD cell that responds strongly to small targets moving in a single direction:

$$X_{\text{FD2}}(t) = 2 \cdot \sum_{i=1}^N [y_i^+(t)]^n. \quad (27)$$

The output can then be calculated in terms of the difference of these two cells:

$$X(t) = X_{\text{FD2}}(t) - X_{\text{FD1}}(t), \quad (28)$$

which, due to the mutual exclusion of  $y_i^+$  and  $y_i^-$ , can be shown to be mathematically identical to the calculation of (23) and (24). Thus the interaction of four different FD cells, rather than two as in the original model, is required to synthesize the output of the elaborated model.

Keeping the target centered in the visual field is the best approach to tracking when the target trajectory is unpredictable, which the authors consider to be the most likely

case when insects track other insects. To improve tracking performance in this case, our third elaboration alters the model such that it actuates a turn toward the target regardless of the direction of target motion. However, in the case of a target moving toward the midline of the insect that holds a steady course and thus follows a predictable linear path, the turn actuated by the model without the third elaboration (away from the target) actually leads to a better interception than with the fully elaborated model. This is due to the fact that information about the *future* position of the target is effectively used to reduce the interception time. It can be shown that, rather than turning to face the target immediately, the shortest time to interception in the case of a linearly moving target is obtained by holding the angular velocity of the target at zero, thus keeping it at a constant angular position in the visual field. The path of the simulated fly will intersect the target path ahead of the current position of the target. This is exactly the behavior generated by the model without the third elaboration. This interception strategy is well known in missile guidance literature as *proportional navigation* (Zarchan 2002), usually occurring between a fast-moving missile and a relatively slow-moving aircraft, whose flight path over small time scales may be approximated as linear. In the case of a linearly moving target, or a target moving much more slowly than the fly, the elaborated model may be used without the third elaboration.

Kimmerle and Egelhaaf (2000) replayed optical flow patterns generated by tethered flying flies to flies in a flight simulator and recorded from FD cells. The electrophysiological response of an FD cell was only weakly dependent on background motion, despite evidence that wide-field motion inhibits FD cell activation. In the present model as well as in the original, background motion inhibits the activation of FD cells. However, Fig. 7a shows that over a wide range of background complexities, from approximately 25 to 150 boundary objects, the performance of the elaborated model is maintained. Over this range, Fig. 7b shows that the RMS activation of FD cells in the elaborated model (26) is roughly maintained despite an increase in background motion. This is possible because of the first elaboration (negative  $k^*$ ), which suggests inhibition of binocular pool cells by other binocular pool cells, which then provide shunting inhibitory inputs to small-field units. This inhibition of inhibitory units was in fact suggested by Kimmerle and Egelhaaf (2000) as a possible explanation for the insensitivity of FD cells to background stimuli. Mathematically, this elaboration allows the normalizing term involving a weighted sum of binocular pool cells to approach zero. In this case, (21) and (22) simplify to

$$y_{i,\text{right}}^+(t) = \frac{v_{i,\text{right}}^+(t)}{\beta}, \quad (29)$$

$$y_{i,\text{right}}^-(t) = \frac{v_{i,\text{right}}^-(t)}{\beta}, \quad (30)$$

and the normalization is independent of the visual stimulus. Thus the first elaboration allows responses of motion inconsistent with the background to be emphasized without regard to the strength of background motion.

While the present work addresses the neural basis of small-target tracking in the lobula plate of the fly, other authors have detailed a separate circuit for visually based tracking in the *lobula* of the fly. Gronenberg and Strausfeld (1991) have identified a set of male-specific neurons implicated in a neuronal realization of the 1974 Land and Collett model that is proposed to underlie target tracking in the fly. This circuit suggests that a target is brought into the acute region of the male compound eye using non-directional motion detectors and held there using directional motion detectors. Presumably this tracking system shares the same underlying motion detection substrate as the present model, but the situations in which each system is uniquely valuable are presently unknown.

*Acknowledgements.* This work was supported by the US Office of Naval Research under agreement number N68936-00-2-0002.

## References

- Borst A (1991) Fly visual interneurons responsive to image expansion. *Zoologische Jahrbucher – Abteilung für Allgemeine Zoologie und Physiologie* 95:305–313
- Egelhaaf M (1985a) On the neuronal basis of figure-ground discrimination by relative motion in the visual system of the fly: I. Behavioral constraints imposed on the neuronal network and the role of the optomotor system. *Biol Cybern* 52:123–140
- Egelhaaf M (1985b) 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* 52:195–209
- Egelhaaf M (1985c) 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* 52:267–280
- Egelhaaf M, Borst A (1989) Transient and steady-state response properties of movement detectors. *J Opt Soc Am A* 6:116–127
- Egelhaaf M, Borst A, Warzecha A-K, Flecks S, Wildeman A (1993) Neural circuit tuning fly visual interneurons to motion of small objects: II. Input organization of inhibitory circuit elements revealed by electrophysiological and optical recording techniques. *J Neurophysiol* 69(2):340–351
- Gibson JJ (1950) *The perception of the visual world*. Houghton Mifflin, Boston
- Gronenberg W, Strausfeld NJ (1991) Descending pathways connecting the male-specific visual system of flies to the neck and flight motor. *J Comp Physiol A* 169:413–426
- Hassenstein B, Reichardt W (1956) Systemtheoretische analyse der Zeit-, Reihenfolgen- und Vorzeichenbewertung bei der Bewegungsperezeption des Rüsselkäfers *Chlorophanus*. *Zeitschrift für Naturforschung* 11b:513–524
- Hausen K, Egelhaaf M (1989) Neural mechanisms of visual course control in insects. In: Stavenga DG, Hardie RC (eds) *Facets of vision*, Chap 18. Springer, Berlin Heidelberg New York, pp 391–424
- Huber SA, Bühlhoff HH (1998) Simulation and robot implementation of visual orientation behaviors of flies. In: *From animals to animats, proceedings of the fifth conference on the simulation of adaptive behaviour*, vol 5. MIT Press, Cambridge, MA, pp 77–85

- Kimmerle B, Egelhaaf M (2000) Performance of fly visual interneurons during object fixation. *J Neurosci* 20:6256–6266
- Kimmerle B, Eickermann J, Egelhaaf M (2000) Object fixation by the blowfly during tethered flight in a simulated three-dimensional environment. *J Exp Biol* 203:1723–1732
- Korrel HWA (2000) Simulation of visual orientation behaviour of the fly. Master's thesis, Department of Computer Science, University of Maastricht
- Land MF, Collett TS (1974) Chasing behaviour of houseflies (*Fannia canicularis*): description and analysis. *J Comp Physiol* 89:331–357
- Pant V (2003) Modular neuromorphic VLSI architectures for visual motion and target tracking. Master's thesis, Department of Electrical and Computer Engineering, The University of Arizona, Tucson, AZ
- Reichardt W, Egelhaaf M, Guo AK (1989) Processing of figure and background motion in the visual-system of the fly. *Biol Cybern* 61:327–345
- Reichardt W, Poggio T (1979) Figure-ground discrimination by relative movement in the visual system of the fly: I. Experimental results. *Biol Cybern* 35:81–100
- Reichardt W, Poggio T, Hausen K (1983) Figure-ground discrimination by relative movement in the visual system of the fly: II. Towards the neural circuitry. *Biol Cybern* 46:1–30
- Van Santen JPH, Sperling G (1985) Elaborated Reichardt detectors. *J Opt Soc Am A* 2:300–320
- Warzecha A-K, Egelhaaf M, Borst A (1993) Neural circuit tuning fly visual interneurons to motion of small objects: I. Dissection of the circuit by pharmacological and photoinactivation techniques. *J Neurophysiol* 69(2):329–339
- Zarchan P (2002) Tactical and strategic missile guidance, 4th edn. American Institute of Aeronautics and Astronautics Press, Reston, VA