A probabilistic tracker for a bio-inspired target detection algorithm

John V. James, Benjamin S. Cazzolato and **Steven Grainger**
School of Mechanical Engineering
University of Adelaide, Australia
{john.james, benjamin.cazzolato, steven.grainger}@adelaide.edu.au

Steven D. Wiederman
Adelaide Medical School
University of Adelaide, Australia
steven.wiederman@adelaide.edu.au

Abstract

A set of neurons in the dragonfly brain respond selectively to small moving targets. It is proposed that these small target motion detectors neurons (STMDs) subserve the dragonfly's aerial pursuits of prey. Behavioural studies have shown that dragonflies use internal models of prey position and trajectory during pursuits. How this target state estimation is carried out in the brain is poorly understood, however physiological studies suggest two properties of the dragonfly's tracking system. Firstly, 'selective attention' whereby a competitive process ensures responses to all but one visible target are suppressed. Secondly, a spatially inhomogeneous 'predictive gain modulation' whereby responses to a target's estimated, continuous trajectory are enhanced. Previous research presented a biologically-inspired tracker based on 'elementary small target motion detectors' (ESTMD), 'matched filter' units proposed to underlie physiologically observed STMD responses. However, previous studies did not include a comparative evaluation of alternative tracking methods for use with the ESTMD model. Here we present a grid-based estimator for use with the ESTMD detection model which incorporates an input-selection method analogous to selective attention. We compare our novel approach to, among others, the biologically-inspired tracker and find that our tracker shows better performance in our test scenario.

1 Introduction

Aerial predators must be able to successfully detect and pursue their prey in order to survive. Where such pursuit is visually guided, the predator must be able to rapidly detect, localise and track prey amongst the complex visual environments encountered in natural settings. Species of effective aerial predators, the dragonflies, engage in highly dynamic, visually guided pursuits of prey with high rates of success [Olberg *et al.*, 2000]. These pursuits are controlled by a tiny brain with around two millions neurons, suggesting that the dragonfly's target detection algorithm is both effective and efficient. Investigations into the neuronal basis of target detection by dragonflies and other insect species have revealed neurons which respond selectively to small moving targets, called small target motion detectors (STMDs) [O'Carroll, 1993; Nordström *et al.*, 2006]. These neuronal responses are tuned to target velocity and size, whilst sensitive to contrast, and thus appear to represent an ambiguous measurement of these variables. In dragonflies, it is not presently feasible to directly investigate the correlations between STMD activity and behaviour because current methods for recording neuronal activity require immobilisation of the animal. It is thus unclear how dragonflies are able to overcome the apparent shortcomings of STMDs as target detectors to infer the position of prey sufficiently accurately to enable successful capture. Understanding how this is achieved may facilitate the application of efficient detectors inspired by STMDs in applications including autonomous robotics.

Using a model of dragonfly target detecting neurons, the elementary small motion detector (ESTMD) model, [Wiederman *et al.*, 2008] ourselves and other researchers have investigated bio-inspired [Bagheri *et al.*, 2017a; 2017b] and statistics-based tracking methods [Wang *et al.*, 2019] for differentiating targets from background clutter. Moreover, the combination of ESTMD and bio-inspired tracking has been evaluated against other object detection algorithms [Bagheri *et al.*, 2017b]. However, these studies have not compared how well alternative tracking methods, or the absence of tracking, influence closed-loop tracking tasks based on ESTMD outputs. In this paper we present a comparative evaluation of sev-

eral tracking methods, including the bio-inspired tracker described by Bagheri et al. We quantify how well each of these methods performs given the presence of distractors and different delays in effecting changes in pursuer angle. We find that that the tracker proposed in this paper achieves greater performance than the alternatives tested.

2 Related work

Discrete state estimation may be framed as a problem of finding, for every time step k, an optimal state estimate x_k given $z_{1:k}$, the history of measurements up to time step k. For linear systems meeting certain criteria, a Kalman filter is an optimal estimator. Various estimators have been developed for non-linear systems such as: the extended Kalman filter, unscented Kalman filter [Wan and van der Merwe, 2000], grid-based estimators and particle filters [Arulampalam *et al.*, 2002]. Grid-based estimators and particle filters each use finite samples to approximate the continuous state probability distribution. Grid-based estimators divide the range of continuous states into discrete points and maintain a progressively updated set of weights representing the conditional probabilities of these discretised states. Particle filters use finite random samples of the state space, weighted according to the estimated probability of each state, with the continuous probability distribution function represented implicitly via the weights. Whichever approach is used for estimation, a model relating the measurements to the estimated states must be decided upon, and deriving this will not always be straight forward.

The estimation method used by the dragonfly is unknown, however some properties of its target tracking system have been revealed through electrophysiological and behavioural studies. Electrophysiology is a body of techniques by which the electrical responses of neurons to stimuli in living, and even behaving, animals can be measured quantitatively. Such techniques enabled the identification of several distinct classes of STMD. Different STMDs produce responses to targets present in differently sized and positioned regions of the visual field, these regions being called the neuron's receptive field [Barnett *et al.*, 2007]. A dragonfly STMD with a receptive field spanning an entire visual hemisphere has shown progressive enhancement of responses to targets following continuous paths [Dunbier *et al.*, 2012]. This enhancement is subserved by a predictive focus of gain modulation whereby responses to a target in parts of the visual field ahead of the target's past trajectory are enhanced [Wiederman *et al.*, 2017]. It is spatially localised, but not directionally selective and is accompanied by a suppression of responses elsewhere [Wiederman *et al.*, 2017; Fabian *et al.*, 2019]. The same neuron has also been found to respond selectively to one of several targets present in its receptive field, indicating selective attention [Wiederman and O'Carroll, 2013]. The combination of selective attention and the predictive properties of facilitation suggest a role for STMDs in enabling an estimation of the future states of the target. Such predictive state estimation is suggested by behavioural studies whereby pursuits of simulated prey appeared to rely upon internal models of the prey's position [Mischiati *et al.*, 2015]. The responses of visual neurons in other insects has been shown to be influenced by behavioural state and by motor commands [Fujiwara *et al.*, 2017], suggesting that proprioceptive inputs and efference copies of motor commands may be available for the dragonfly's target state estimation system.

Here we present an empirically derived observation model for relating the ESTMD outputs to the speed of the target. Furthermore, a direct comparison of the biologically inspired facilitation tracker to other trackers based on the same ESTMD outputs. Finally, we provide an evaluation of the influence of saccade duration on tracking success when using the ESTMD model for target detection.

3 Methods

To evaluate the performance of the trackers, we performed closed-loop simulations of a target moving through a natural background in the presence of moving distractors. Based on the outputs of the ESTMD model, each tracker guided panning movements of the observer with the objective of keeping the target inside the visual field. Performance was measured by comparing how long the target remained in the visual field across different test scenarios. This pursuit-based method for measuring performance enabled comparisons with previous evaluations of the facilitation-based tracker, which used closed-loop pursuits. Evaluating performance in this way required challenging tracking scenarios which would produce failures.

3.1 Tracking scenario

Backgrounds were generated by projecting cylindrical panoramic images of natural scenes on to a plane and then drawing a target and distractors on to it. The panoramic images covered a variety of natural and man-made environments [Brinkworth and O'Carroll, 2008]. Targets were drawn as a square with nil luminance and sides spanning approximately 1° of the visual field. The target moved at constant altitude following a predetermined azimuthal trajectory. These trajectories were randomly generated by beginning at an angular velocity of $120\,^\circ\,\mathrm{s}^{-1}$, moving rightward across the visual field, and then every 500 ms selecting a new random

angular velocity from a normal distribution with mean $0\,^\circ\,s^{-1}$ and standard deviation $80\,^\circ\,s^{-1}$. Targets accelerate towards the new velocity set-point with an angular acceleration of $80\,^\circ\,s^{-2}$. At the beginning of a trial, targets entered the visual field from the left-hand side. This gave rise to trajectories with a variety of angular velocities which commonly featured slow target motion. Periods without significant target motion emphasised the importance of target tracking to ignore distractors.

The ESTMD model is responsive to features with a spatiotemporal profile of a small moving target [Wiederman *et al.*, 2008; Wiederman and O'Carroll, 2011]. Although natural scenes feature diverse spatial features, only a subset of these will produce relevant distracting clutter for the ESTMD model during translation or rotation of the observer. Previous studies have shown that these features are surprisingly rare [Wiederman and O'Carroll, 2011]. To reliably introduce relevant clutter into the scene to create challenging tracking scenarios, we drew distracting features with randomised luminance and the same spatial profile as the target in random positions in 3-D space ahead of the observer. These distractors were placed after the first turn of the observer, and were replaced after each subsequent turn. Between turns, the observer moved directly forward through the 3-D environment such that distractors moved progressively through the visual field. The naturalistic backdrop was assumed to be distant to the observer and therefore unchanged by translation of the observer. These images were then blurred (Gaussian filter, 1.4° full width at half maximum) and subsampled (approximately 1° intervals) to reflect the limited acuity and resolution of compound eyes. The resulting images were processed using the ESTMD model to produce inputs for the trackers.

3.2 ESTMD model

The ESTMD model combines linear spatial and temporal filters with several non-linear operations to produce a non-linear matched filter for small moving targets. Input luminance values are filtered with a digital linear temporal bandpass filter representing the temporal aspects of early visual processing, including the photoreceptors and large monopolar cells. The 9^{th} order filter [Bagheri *et al.*, 2015] has the form:

$$y_k = \sum_{i=1}^{9} b_i L_{k-i} - \sum_{j=1}^{9} a_j y_{k-j} \tag{1}$$

where y_k are the filter outputs, L_k are the input luminance signals and the filter coefficients a and$b$ are given in Table 1. Centre-surround antagonism in the lamina is represented by a linear spatial filter with kernel:

$$\begin{bmatrix} -\frac{1}{9} & -\frac{1}{9} & -\frac{1}{9} \\ -\frac{1}{9} & \frac{8}{9} & -\frac{1}{9} \\ -\frac{1}{9} & -\frac{1}{9} & -\frac{1}{9} \end{bmatrix} \tag{2}$$

The signal is then half-wave rectified into two separate channels: one for positive values (ON) or negative values (OFF). This reflects separation of responses to luminance increments and decrements into separate neuronal pathways in the insect brain. Non-linear adaptation on these separate channels is represented by filtering each channel with a non-linear temporal filter given by:

$$Y_k = u_k - \gamma_{k-1} \tag{3}$$

where u_k is the current input, Y_k is the current output, γ_k is a state variable of the filter given by:

$$\gamma_k = \frac{t_s}{t_s + \tau_k} u_k + \frac{\tau_k}{t_s + \tau_k} \gamma_{k-1} \tag{4}$$

where $t_s = 1$ ms is the model sample time and τ_k is given by:

$$\tau_k = \begin{cases} 10 \text{ ms} & , \quad u_k - \gamma_{k-1} \geq 0 \\ 100 \text{ ms} & , \quad \text{otherwise} \end{cases} \tag{5}$$

The values of τ_k reflect the time course of physiologically observed adaptation [Wiederman *et al.*, 2008]. After the non-linear filter, the signal is thresholded to a minimum of zero such that the signal is always non-negative. To match the size selective responses of STMD neurons, the signal is filtered with a strong linear spatial filter with kernel:

$$\begin{bmatrix} -1 & -1 & -1 & -1 & -1 \\ -1 & 0 & 0 & 0 & -1 \\ -1 & 0 & 2 & 0 & -1 \\ -1 & 0 & 0 & 0 & -1 \\ -1 & -1 & -1 & -1 & -1 \end{bmatrix} \tag{6}$$

The signal is again thresholded to zero. To reproduce the velocity tuning of STMD neurons, one of the channels is then delayed using a low-pass filter given by [Bagheri *et al.*, 2015]:

$$H(z) = \frac{\frac{1}{51} z + \frac{1}{51}}{z - \frac{49}{51}} \tag{7}$$

For a low luminance target selective variant of the ESTMD model, the OFF channel is delayed and vice versa for a high luminance target. We tested a variant which is dark target selective against a variant which combines the outputs of both a light target selective and dark target selective variant but found little difference between the two and so have not separately presented these results.

3.3 Probabilistic tracker

Our probabilistic tracker is based on an approximate grid-based estimator. That is, the continuous position states, x_k, and speed states, v_k, of the target were approximated by a finite number of discrete points x_k^i and v_k^j. The direction of movement was not estimated. The position state was approximated by a sub-pixel grid

Table 1: Early visual processing filter coefficients

	1	2	3	4	5	6	7	8	9
a	0.0001	-0.0012	0.0063	-0.0222	0.0609	-0.1013	0.2363	-0.3313	0.1524
b	-5.1664	12.2955	-17.9486	17.9264	-12.8058	6.5661	-2.3291	0.5166	-0.0542

formed by the union the centres of the pixels and a set of points halfway between each of the pixel centres vertically, horizontally and diagonally. The resulting grid for a $H \times V$ pixel image contained $4HV$–$2(H+V)+1$ grid points. Each discrete position state x_k^i was assigned a weight, $m_{k|k-1}^i$, representing the conditional probability of that state existing given the history of measurements, with the weight defined by:

$$m_{k|k-1}^i = p(x_k = x_k^i | \boldsymbol{z}_{1:k-1}) \tag{8}$$

where x_k is the position state and $\boldsymbol{z}_{1:k-1}$ is the history of measurements up to time $k-1$. The speed state weight $n_{k|k-1}^i$ is defined similarly:

$$n_{k|k-1}^i = p(v_k = v_k^i | \boldsymbol{z}_{1:k-1}) \tag{9}$$

where v_k is the speed state and v_k^i are discrete velocity points from nil to $800\,^\circ\,\mathrm{s}^{-1}$ at $0.48\,^\circ\,\mathrm{s}^{-1}$ intervals. In addition to the target states, the tracker maintained a binary flag, A_k, representing whether the target is currently acquired. The tracker was updated every four frames. To find the measurement for a time step, z_k, the ESTMD outputs for the current frame and the preceding three were summed into a single frame and then a single output was selected, with the rules for selection differing based on whether the target is acquired. To initialise the tracker, the speed and position weights were set to be uniform and the target was assumed to not be acquired. The tracker implemented the following procedure:

1. Select an ESTMD output to use for updating the tracker. The selection methodology is described below. If no outputs meet the selection criteria then the tracker is updated independently of the measurements . The selected output, if any, is denoted by $\boldsymbol{z}_k$, comprising a pixel position and output magnitude denoted by $|\boldsymbol{z}_k|$.
2. Evaluate n_k^i (speed prediction) using:

$$n_{k|k-1}^i \triangleq \sum_{j=1}^{N_v} n_{k-1|k-1}^i p(v_k^i | v_{k-1}^j) \tag{10}$$

   where N_v is the number of discrete velocity values and $p(v_k^i|v_{k-1}^j)$ is given by a normalised 2-D Gaussian centred on v_{k-1}^j with standard deviation $0.48^\circ\,\mathrm{s}^{-1}$ evaluated at v_k^i.
3. If an output was selected and $|\boldsymbol{z}_k|$ is above the threshold of 5.2×10^{-3}, evaluate $n_{k|k}^i$ (speed update) using:

$$n_{k|k}^i = \frac{n_{k|k-1}^i p(\boldsymbol{z}_k | v_k^i)}{\sum_{j=1}^{N_v} n_{k|k-1}^j p(\boldsymbol{z}_k | v_k^j)} \tag{11}$$

   where $p(\boldsymbol{z}_k|v_k^i)$ is evaluated using the observation model described below. Otherwise, evaluate $n_{k|k}^i$ with:

$$n_{k|k}^i = \frac{n_{k|k-1}^i}{\sum_{j=1}^{N_v} n_{k|k-1}^j} \tag{12}$$

4. Evaluate $m_{k|k-1}^i$ (position prediction) by convolving $m_{k-1|k-1}^i$ with the position prediction kernel described below.
5. If the selected measurement $\boldsymbol{z}_k$ is above the threshold of 5.2×10^{-3}, evaluate $m_{k|k}^i$ (position update) using:

$$m_{k|k}^i = \frac{m_{k|k-1}^i p(\boldsymbol{z}_k | x_k^i)}{\sum_{j=1}^{N_x} m_{k|k-1}^j p(\boldsymbol{z}_k | x_k^j)} \tag{13}$$

   where $p(\boldsymbol{z}_k|x_k^i)$ is a normalised 2-D Gaussian centred on x_k^i with standard deviation 0.5° evaluated at the position of the measurement $\boldsymbol{z}_k$ and N_x is the number of grid points. Otherwise, evaluate $m_{k|k}^i$ with:

$$m_{k|k}^i = \frac{m_{k|k-1}^i}{\sum_{j=1}^{N_x} m_{k|k-1}^j} \tag{14}$$

6. Evaluate $\phi_k = \max(m_{k|k}^i n_{k|k}^j)$, which is used as a measure of the certainty about the target states at time step k.
7. Set the acquisition flag state, A_k, according to:

$$A_k = \begin{cases} \text{true} & , \quad \phi_k > 1 \times 10^{-5} \\ \text{false} & , \quad \phi_k \leq 1 \times 10^{-5} \text{ and } \max(m_{k|k}^i) < 4 \times 10^{-3} \\ A_{k-1} & , \quad \text{otherwise} \end{cases} \tag{15}$$

8. If all of the following conditions are met then execute a turn with magnitude determined as described below:
   (a) $\max(m_{k|k}^i)$ is more than 5° from the horizontal centre of the visual field;
   (b) $\max(m_{k|k}^i) > 5 \times 10^{-8}$;

(c) a saccade is not in progress and at least 40 ms have elapsed since the end of the most recent saccade; and

(d) A_k is true.

9. If a turn was executed, remap the values of $m^i_{k|k}$ by shifting the weights azimuthally such that the position of $\max(m^i_{k|k})$ is centralised horizontally in the visual field. The set of grid points which do not map to other grid points, i.e. because they are close to one side of the visual field, are given a uniform value calculated to achieve the result that $\sum_{i=1}^{N_x} m^i_{k|k} = 1$ after the remapping.

Pursuit logic

The same pursuit logic is used for all trackers based on the estimated target position. Let θ be the estimated azimuthal position of the target relative to the centre of the visual field. A turn is executed if $|\theta| > 5°$. Let $\Delta\theta$ be the angle of the turn such that the viewing angle after the turn, θ' is given by:

$$\theta' = \theta + \Delta\theta \tag{16}$$

When turning towards a target, two possible strategies are to turn in order to centre the visual field at the position of the target at the time the turn completes, or to centre at the position of the target at the time the turn is initiated. We tested both alternatives. The angle of a turn to the target's position at the start of the turn is simply given by:

$$\Delta\theta = -\theta \tag{17}$$

The angle of a turn to the target's expected position at the end of the turn is given by:

$$\Delta\theta = -\theta + \text{sign}(\theta) \times \Big((T_S + T_T) \times 60 + 3 \Big) \tag{18}$$

where T_S is the saccade duration, T_T is the period of time for which model outputs are ignored following a turn and the $\text{sign}(\theta)$ function is defined as:

$$\text{sign}(\theta) = \begin{cases} 1 & , \quad \theta > 0 \\ -1 & , \quad \theta < 0 \\ 0 & , \quad \theta = 0 \end{cases} \tag{19}$$

Observation model

The outputs of the ESTMD depend upon local contrast, the target velocity and its size. Accordingly, the ESTMD outputs are ambiguous measurements of the target states. To construct a suitable observation model, we generated imagery using one of the natural backgrounds with a 40° horizontal and 20° vertical field of view. We then rotated the background by 15° at $60\,°\,\text{s}^{-1}$ before halting the background motion and introducing three $1° \times 1°$ nil luminance square targets at different altitudes at the left-hand side of the visual field. The targets moved at the same speed from left to right at constant altitude. We simultaneously drew the same targets on to a white background. We summed the outputs of the model across four frames and identified the locations where this summed output exceeded 0.01 for targets drawn on the white background. We treated those locations as being target-induced responses, and recorded the outputs generated by the naturalistic background at those locations. The targets moved across the field of view until exiting at the right-hand side, at which point we again rotated the background by 15° at $60°\,\text{s}^{-1}$ before again drawing targets. This was repeated 24 times, that is until the whole 360° background had been scanned. From these outputs, we constructed histograms of the responses of the model for the given target velocity using response magnitude bins defined by edges from 0 to 0.5 at 0.01 spacing. The final bin included all responses greater than 0.5. Repeating this exercise for target velocities between $20\,°\,\text{s}^{-1}$ and $800\,°\,\text{s}^{-1}$ produced a set of histograms of responses for each velocity. From these we calculate the probability mass function $p(\boldsymbol{z}_k|v^i_k)$, giving the probability of observing an ESTMD output $\boldsymbol{z}_k$ if the target is moving at speed v^i_k.

Input selection

To emulate a selection mechanism whereby only the most salient output is used for updating the state estimate, at most one output per frame is used to update the tracker. Here, we define the notion of salience as being the extent to which the output agrees with the current estimate of the states. We measure this with α_k, defined as:

$$\alpha_k(\mathbf{p}) = \sum_{i \in I_x} \sum_{j}^{N_v} \beta(\boldsymbol{p}, x^i_{k-1}, v^j_k) m^i_{k-1|k-1} n^j_{k|k-1} \tag{20}$$

where: $\boldsymbol{p}$ is a pixel of the input image; $\beta(\boldsymbol{p}, x^i_{k-1}, v^j_k)$ is a mask representing whether or not the target could have moved to position $\boldsymbol{p}$ from x^i_{k-1} with the speed v^j_k during the update interval; and, I_x are the indices of the set of sub-pixel grid points which fall within a 5×5 patch centred on $\boldsymbol{p}$. To determine β for each pixel $\boldsymbol{p}$, we calculate the minimum and maximum distances from the centre of pixel $\boldsymbol{p}$ to the edges of nearby pixels. This provides a set of minimum and maximum distances which when divided by the update period of the tracker gives a set of speed intervals I_p which could have resulted in the target moving from x^i_{k-1} to $\boldsymbol{p}$. The mask $\beta(\boldsymbol{p}, x^i_{k-1}, v^j_k)$ is then defined by:

$$\beta(\boldsymbol{p}, x^i_{k-1}, v^j_k) = \begin{cases} 1 & , \quad \text{if } v^j_k \in I_p \\ 0 & , \quad \text{otherwise} \end{cases} \tag{21}$$

The same set of masks is used for every $\boldsymbol{p}$. Conceptually, α_k represents the probability that at the previous

timestep, the target was in a position which could have moved to the pixel under consideration with a velocity that produced the observed magnitude of response.

Update of position weights

The position weights, m_k^i, are updated by convolving m_{k-1}^i with a 2-D kernel K which has its values defined by the velocity weights. Because direction is not estimated, the kernel is symmetrical and the same is used for each grid point. The kernel K is defined by:

$$K_{f,g} = k_x p(v_k^i = v_{req} | \mathbf{z}_{1:k-1}) \tag{22}$$

where f and g are integers in the range $[-2, 2]$, k_x is a normalisation constant, v_{req} is the speed from v^j that most closely matches the speed required to move to point $[f, g]$ from the centre of the kernel during the update period.

Fixed speed (FS) variant

It was apparent from our early investigations that the combination of the statistics of input target velocities and the observation model had the effect that the speed weights would converge to a distribution similar to that shown in Figure 1. As such, we separately evaluated a variant of the probabilistic tracker which did not update the speed weights n_k^j based on measurements. Instead, the speed weights shown in Figure 1 were used throughout. The two variants will henceforth be referred to as the fixed speed (FS) and non-fixed speed (NFS) variants.

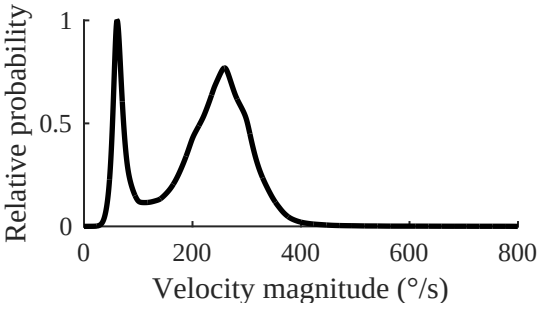

Figure 1: Fixed speed weights used in the FS variant of the probabilistic tracker

3.4 Facilitation tracker

The facilitation tracker aims to approximate the spatial enhancement of responses to targets moving on continuous paths observed in dragonfly neurons. We adapted the publicly available implementation of the tracker from [Bagheri *et al.*, 2017a] to suit our simulation environment and used parameters which were reported by Bagheri et al. to perform well. Briefly, the tracker maintains a gain value for every pixel in the image. When a target is detected, the gain values are increased at the target's projected future position. Gain values elsewhere gradually decay. The tracker is updated every millisecond, except during turns or the 40 ms following a turn. To estimate the position of the target, the output of the ESTMD model is multiplied pixelwise by the gain values and then the location of the maximum output is taken to be the target position. We used the parameters in Table 2.

Table 2: Facilitation tracker parameter settings.

Parameter	Value
Facilitation kernel size	7°
Facilitation time constant	200 ms
Elementary motion detector time constant	40 ms
Output threshold	0.001
Facilitation gain	625

3.5 Constant gain field (CG)

If the target is kept close to the centre of the visual field, then the facilitation tracker could be expected to result in high gain in the centre of the visual field and lower gain elsewhere. To determine whether dynamically updating the gain field offers a performance advantage over simply assuming constant gain in the centre, we define a gain field $k_g(x, y)$ as:

$$k_g(x,y) = \begin{cases} 625 & , \quad |x_h - \bar{x}_h| < 5 \\ 1 & , \quad \text{otherwise} \end{cases} \tag{23}$$

where x_h is the horizontal position of the pixel and $\bar{x}_h$ is the horizontal centre of the visual field. The modulated outputs I_G are then calculated with:

$$I_G(x, y, k) = k_G(x, y) I(x, y, k) \tag{24}$$

Turns were executed if the location of the maximum was more than 5° from the centre of the visual field and at least 30 ms had elapsed since the last turn. The magnitude of turns was calculated in the same way as for the other trackers.

3.6 Saccade trajectories

Behavioural measurements of saccade magnitudes and durations in dragonflies have shown that dragonflies execute saccades over a period of approximately 50 ms [Lin and Leonardo, 2017]. This is a significant duration given the temporal characteristics of the ESTMD filter. During saccades, the dragonfly's head achieved angular velocities above $1400\,^\circ\,\mathrm{s}^{-1}$ requiring peak angular accelerations exceeding $5.6 \times 10^4\,^\circ\,\mathrm{s}^{-2}$. The duration of behaviourally observed saccades were mostly uniform despite differences in the total angular displacement of the saccades. Previous studies [Bagheri *et al.*, 2015] used instantaneous, and therefore unrealistic, changes of the

viewer angle. To emulate realistic saccades, we generated turns of fixed duration having a triangular velocity profile with the maximum velocity determined so as to achieve the desired angular displacement:

$$v_{max} = \frac{2(\theta_f - \theta_i)}{\Delta t} \tag{25}$$

where v_{max} is the maximum velocity of the saccade, θ_i and θ_f are respectively the start and end angles of the saccade and Δt is the duration of the saccade.

3.7 Performance evaluation

At every timestep, the azimuthal angle between the centre of the visual field and the target was calculated to determine whether the target was visible or not. If the target was outside the visual field for 500 ms consecutively, the simulation was terminated. Previous studies involving the facilitation tracker used a horizontal field of view of 94° and evaluated tracking based on whether a simulated pursuit was eventually successful whether or not the target left the field of view during pursuit [Bagheri *et al.*, 2015]. Performance was evaluated in two different ways: firstly, comparing how many time steps elapsed before the target first exited the visual field; and, secondly comparing the time at which the simulation was terminated. The latter method allows for more direct comparison with the results of Bagheri et al. The maximum performance result was 30 s of successful tracking, indicating that the tracker had kept the target in the visual field for the entire duration of the target's randomised trajectory.

4 Results

We compared the performance of the trackers on the same sets of tracking scenarios and quantified how the presence of distractors, the length of saccades and the movement of the target affected performance. All of the results presented here are based on at least 30 target trajectories for each of the 15 backgrounds used.

4.1 Effect of distractor strength

We evaluated the effect of adding distractors with a different simulated forward velocity on tracking performance. The introduction of distractors significantly increased the difficulty of the target following task, and resulted in lower performance for all trackers: Figure 2. However, the probabilistic trackers showed greater robustness in the presence of distractors, indicated by the corresponding curves in Figure 2 being further to the right than for the other trackers. Interestingly, the FS variant of the probabilistic tracker performed better than the NFS variant for forward velocities of $1\,\mathrm{m\,s^{-1}}$ and $3\,\mathrm{m\,s^{-1}}$: Figure 2A & B. At a forward velocity of $6\,\mathrm{m\,s^{-1}}$, the probabilistic NFS variant performed better than the FS variant: Figure 2C. This result may be due to poor agreement between the FS variant's speed weights and the slow apparent motion of the distractors for the lower forward velocities. For slow distractors, the facilitation tracker performed well and surpassed the CG tracker, indicating that the dynamic adjustment of the gain field was beneficial in this case: Figure 2A. However, for fast moving distractors, the facilitation tracker produced worse performance than no tracking: Figure 2B - D. Figure 2D shows the performance of the trackers if evaluated on the basis that failure does not occur until the target has been outside the visual field for at least 500 ms.

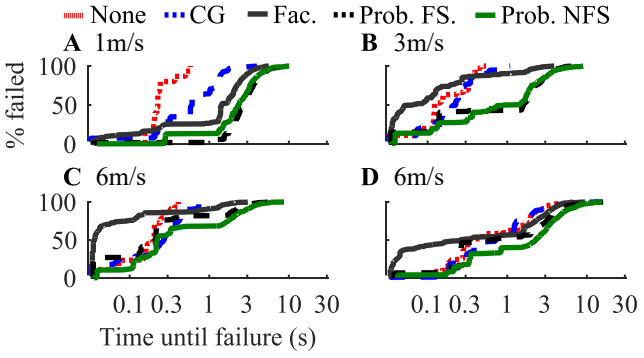

Figure 2: Effect of distractors on tracker performance. Lines show, for each tracker, the proportion of trials which had failed by the time indicated in the abscissa. Lines terminate at the time by which all trials had failed. The observer moves towards the distractors at the velocity shown in the label. A-C) Failure is taken to occur when the target leaves the field of view for the first time. The visual field is 60° horizontal × 30° vertical. D) Failure is taken to occur when the target has been offscreen for 500 ms consecutively. The visual field is 90° horizontal x 45° vertical.

4.2 Combination of saccade delays and distractors

We evaluated whether the length of saccades affected performance in the presence of distractors. Greater saccade delays compounded the reduction in performance due to distractors, shown by the downward slopes of the curves in Figure 3. When attempting to predict the location of the target after a saccade, the magnitude of the required turn angle increases linearly with the saccade duration. When using a predictive turn with 100 ms saccade duration, the smallest turn that could be executed was 11°, such that a turn in the wrong direction would was likely to cause the target to leave the visual

field. Predictive turns resulted in lower performance for a saccade duration of 100 ms: compare Figure 3A and B. Without a predictive turn, similar performance was achieved at 50 ms and 100 ms: Figure 3A. Predictive turns resulted in greater performance for saccade durations of 25 ms and 50 ms: compare Figure 3A and B. The FS variant of the probabilistic tracker showed greater performance than the NFS variant for a saccade duration of 25 ms and forward velocity of 3 ms: Figure 3B and C. Because distractors were not drawn during saccades and new distractors were placed at the end of each saccade, all other things being equal a longer saccade duration has the effect that distractors will be visible for a shorter period of time before another turn is required to follow the target. Distractors will exhibit gradually increasing apparent motion over time as the observer moves closer to the distractors. In light of the observation in the preceding sub-section that the probabilistic FS variant achieves better performance for slow distractor motion, the advantage for the probabilistic FS tracker at 25 ms and 3 m s^{-1} reflects the slower apparent distractor motion.

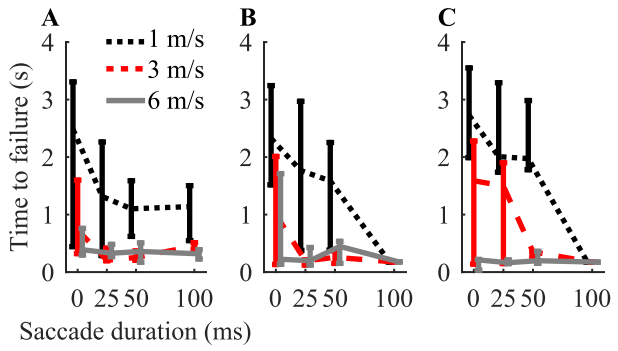

Figure 3: Effect of saccade duration with distractors. Failure was taken to occur when the target first left the field of view. Lines connect the median times to failure for each forward velocity value and vertical bars show the interquartile range. The same saccade durations were used for each variant but the lines have been offset horizontally for clarity. A) Performance for the probabilistic tracker NFS with a field of view of 60° × 30° and not using a predictive turn during pursuit. B) As in A but using a predictive turn. C) As in B but using the probabilistic FS tracker

To compare the probabilistic and facilitation tracker in a way which more closely matches prior evaluation by Bagheri et al., we also compared these two by evaluating failure based on when the target was outside the visual field for more than 500 ms consecutively. Under this analysis, the probabilistic tracker performed substantially better than the facilitation tracker for a forward velocity of 1 m s^{-1} and when using instant turns: Figure 4A&B. For a forward velocity of 3 m s^{-1}, the facilitation tracker displayed better performance than the probabilistic for 50 ms or 100 ms saccade durations. Interestingly, a 50 ms saccade matches the behaviourally observed dragonfly saccade duration and 3 m s^{-1} is closest to behaviourally observed maximum pursuit speeds of 4 m s^{-1} [Lin and Leonardo, 2017], and as such these conditions are the most naturalistic. While in each case tracking was lost well before the end of the trial, successfully pursuing the target for more than 1 s is sufficient for a typical dragonfly prey pursuit which lasts 348 +/- 110 ms [Lin and Leonardo, 2017].

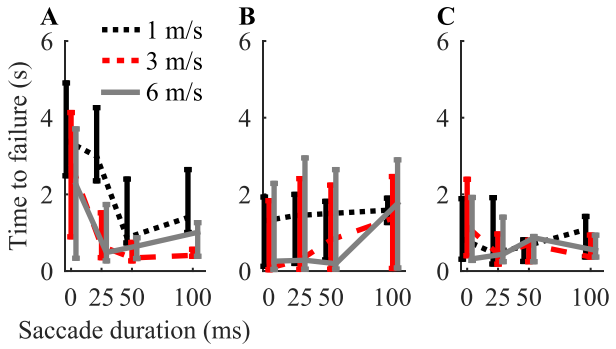

Figure 4: Time to failure with different saccade durations and forward velocities. Time to failure is measured by when the target has been outside the visual field for more than 500 ms consecutively. Lines connect the median performance at different target velocities. Vertical bars show the interquartile range. The same saccade durations were used for each line but they have been horizontally offset for clarity. A) The probabilistic tracker NFS variant. B) The facilitation tracker. C) No tracking.

4.3 Velocity of target prior to failure

As the ESTMD model is a motion-based detector, if the target moves sufficiently slowly then it is essentially invisible to the detector. When distractors are present, a slowly moving target may produce weaker responses than relatively swift distractors. To compare how effectively each tracker handled this complication, we quantified the median target velocity in the 30 ms prior to failure. With slow moving distractors, the majority of failures for the probabilistic NFS and FS variants and also the facilitation tracker occurred when the target was moving slower than 80 ° s^{-1}, indicating that tracking was generally successful when the target was moving quickly despite the presence of distractors: Figure 5A.

With a forward velocity of $3\,\mathrm{m\,s^{-1}}$, the two probabilistic trackers and the facilitation tracker showed increased failures at high target velocities: Figure 5B and C. However, the facilitation tracker showed a much greater shift towards failures at high target velocities than the probabilistic trackers. At $6\,\mathrm{m\,s^{-1}}$, the majority of failures for all trackers were occurring at high target velocities, which is consistent with failure occurring very early in the trial as the target always began the trial moving at $120\,^\circ\,\mathrm{s^{-1}}$. These results indicate that the probabilistic trackers can robustly handle weak distractors when there is a strong target signature and that these trackers suffer less of a degradation in performance for stronger distractors than the facilitation tracker.

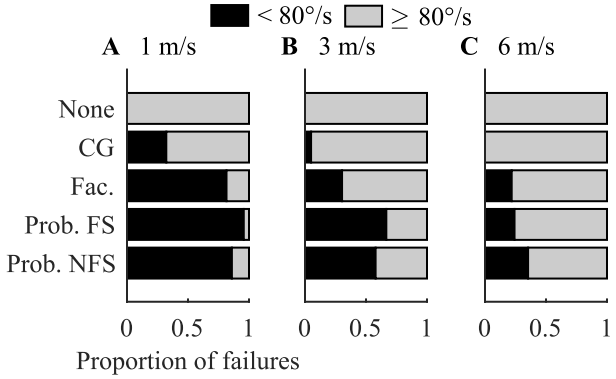

Figure 5: The probability of a failure occurring as a function of the median target speed in the last 30 ms. Failure was taken to occur when the target first left the screen. The visual field was 60° horizontal by 30° vertical. The stacked bars indicate the proportion of failures associated with low ($< 80\,^\circ\,\mathrm{s^{-1}}$) or high ($\geq 80\,^\circ\,\mathrm{s^{-1}}$) median target velocity prior to failure. Instant turns were used (zero saccade duration). The observer moved towards the distractors at the velocity shown in the labels.

5 Discussion

Recent investigations into predictive gain modulation in dragonfly STMDs have shown that although the magnitude of modulation depends upon the velocity of the target, which itself correlates with the strength of the STMD response, the rate at which this gain spreads through the visual field is velocity independent [Fabian *et al.*, 2019]. One interpretation of this result is that if this predictive gain forms part of the dragonfly's state estimation mechanism, then that mechanism effectively assumes a constant target velocity. That is, if the gain value for a point in space is reflective of the likelihood that the target is expected to be in the visual field, then a constant rate of spread of that likelihood is reflective of a constant target velocity. Our result that the probabilistic FS variant performed as well as, or in some cases better, than the NFS variant forms an interesting parallel with these physiological findings.

The apparently poor performance of the facilitation tracker in the presence of strong distractors in this study may be explained by two factors. Firstly, we did not search for optimal parameters for the tracker, but rather used parameters which were reported to work well in [Bagheri *et al.*, 2015]. Our tracking scenario differs in many respects from the tracking scenario used by Bagheri et al., most notably in the way that visual clutter was generated, and so it would be unsurprising if the two scenarios do not share the same set of optimal parameters. Secondly, in our scenario the target always entered from the left side of the visual field. As a result, the gain values of the facilitation tracker would, at least in the early part of the trial, be relatively high in the left side of the visual field, leading to greater susceptibility to distractors in that region. It may be that if the target had first appeared in the centre of the visual field, different results would be observed; however, motion of the target from the periphery to the visual midline has been shown to induce prey pursuits [Lin and Leonardo, 2017] and so this is a behaviourally relevant scenario.

6 Conclusion

We evaluated a tracker based on a grid-based estimator against several alternative tracking methods including a bio-inspired tracker which had previously been evaluated using ESTMD outputs. The probabilistic tracker was able to reject distractors more effectively than the alternative trackers when the observer was able to execute instantaneous changes of angle, or saccades. Longer saccade durations significantly impaired the performance of the tracker. A variant of the probabilistic tracker using a fixed distribution of speed was more resilient to slow moving distractors. Our results show that using a suitable observation model allowed for effective tracking of a target amidst distractors using the outputs of the ESTMD model without estimating the direction of target motion.

Acknowledgments

This work was supported with supercomputing resources provided by the Phoenix HPC service at the University of Adelaide. This research received funding from Australian Research Council Future Fellowship (FF180100466). The authors wish to thank the Adelaide Botanic Gardens for supporting this research. This research was supported by an Australian Government Research Training Program (RTP) Scholarship.

References

[Arulampalam *et al.*, 2002] M. Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174-188, February 2002.

[Bagheri *et al.*, 2015] Zahra M. Bagheri, Steven D. Wiederman, Benjamin S. Cazzolato, Steven Grainger, and David C. O'Carroll. Properties of neuronal facilitation that improve target tracking in natural pursuit simulations *Journal of the Royal Society Interface*, 12:20150083, July 2015.

[Bagheri *et al.*, 2017a] Zahra M. Bagheri, Benjamin S. Cazzolato, Steven Grainger, David C. O'Carroll, and Steven D. Wiederman. An autonomous robot inspired by insect neurophysiology pursues moving features in natural environments. *Journal of Neural Engineering*, 14:046030, August 2017.

[Bagheri *et al.*, 2017b] Zahra M. Bagheri, Steven D. Wiederman, Benjamin S. Cazzolato, and David C. O'Carroll. Performance of an insect-inspired target tracker in natural conditions. *Bioinspiration and Biomimetics*, 12:025006, February 2017.

[Barnett *et al.*, 2007] Paul D. Barnett, Karin Nordström,and David C. O'Carroll. Retinotopic organization of small-field-target-detecting neurons in the insect visual system. *Current Biology*, 17:569-578, February 2006.

[Brinkworth and O'Carroll, 2008] Russell S. A. Brinkworth, and David C. O'Carroll. Robust models for optic flow coding in natural scenes inspired by insect biology. *PLoS Computational Biology*, 5(11):e1000555, November 2009.

[Dunbier *et al.*, 2012] James R. Dunbier, Steven D. Wiederman, Patrick A. Shoemaker, and David C. O'Carroll. Facilitation of dragonfly target-detecting neurons by slow moving features on continuous paths. *Frontiers in Neural Circuits*, 6(79):1-11, October 2012.

[Fabian *et al.*, 2019] Joseph M. Fabian, James R. Dunbier, David C. O'Carroll, and Steven D. Wiederman. Properties of predictive gain modulation in a dragonfly visual neuron. *Journal of Experimental Biology*, 222:jeb207316, September 2019.

[Fujiwara *et al.*, 2017] Terufumi Fujiwara, Tomás L. Cruz, James P. Bohnslav, and M. Eugenia Chiappe. A faithful internal representation of walking movements in the Drosophila visual system *Nature Neuroscience*, 20(1):72-82, January 2017.

[Lin and Leonardo, 2017] Huai-Ti Lin and Anthony Leonardo. Heuristic rules underlying dragonfly prey selection and interception. *Current Biology*, 27:1124-1137, April 2017.

[Mischiati *et al.*, 2015] Matteo Mischiati, Huai-Ti Lin, Paul Herold, Elliot Imler, Robert Olberg, and Anthony Leonardo. Internal models direct dragonfly interception steering. *Nature*, 517:393-U410, January 2015.

[Nordström *et al.*, 2006] Karin Nordström, Paul D. Barnett, and David C. O'Carroll. Insect detection of small targets moving in visual clutter. *PLoS Biology*, 4(3):378-386, February 2006.

[O'Carroll, 1993] David C. O'Carroll. Feature-detecting neurons in dragonflies. *Nature*, 362:541-543, April 1993.

[Olberg *et al.*, 2000] Robert M. Olberg, A.H. Worthington, and K.R. Venator. Prey pursuit and interception in dragonflies. *Journal of Comparative Physiology A*, 186:155-162, February 2000.

[Wang *et al.*, 2019] Hongxin Wang, Jigen Peng, Xuqiang Zheng, and Shigang Yue. A robust visual system for small target motion detection against cluttered moving backgrounds. *IEEE Transactions on Neural Networks and Learning Systems*, Early Access.

[Wan and van der Merwe, 2000] Eric A. Wan and Rudolph van der Merwe. The unscented Kalman filter for nonlinear estimation. *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium*, 4 October 2000, Lake Louise, Alberta, Canada, Canada.

[Wiederman *et al.*, 2008] Steven D. Wiederman, Patrick A. Shoemaker, and David C. O'Carroll. A model for the detection of moving targets in visual clutter inspired by insect physiology. *PLoS One*, 3(7):1-11, July 2008.

[Wiederman and O'Carroll, 2011] Steven D. Wiederman, and David C. O'Carroll. Discrimination of features in natural Scenes by a dragonfly neuron. *The Journal of Neuroscience*, 31(19):7141–7144, May 2011.

[Wiederman and O'Carroll, 2013] Steven D. Wiederman, and David C. O'Carroll. Selective attention in an insect visual neuron. *Current Biology*, 23:156-161, January 2013.

[Wiederman *et al.*, 2017] Steven D. Wiederman, Joseph M. Fabian, James R. Dunbier, and David C. O'Carroll. A predictive focus of gain modulation encodes target trajectories in insect vision *eLife*, 6:e26478, July 2017.