BIO-INSPIRED FEATURE EXTRACTION AND ENHANCEMENT OF TARGETS MOVING AGAINST VISUAL CLUTTER DURING CLOSED LOOP PURSUIT

Kerry J. Halupka¹, Steven D. Wiederman¹, Benjamin S. Cazzolato², David C. O’Carroll¹

¹Adelaide Centre for Neuroscience Research, School of Medical Sciences & ²School of Mechanical Engineering, The University of Adelaide, Adelaide Australia

ABSTRACT

We developed a biologically inspired model for detection and pursuit of small targets against complex backgrounds and tested it in a closed-loop flight arena. A winner-takes-all network of local feature detectors based on insect Small Target Motion Detector (STMD) neurons was used to direct the gaze of a camera mimicking the viewpoint of the pursuer in a series of small steps (saccades) whilst fixating the background. The output of a direction-selective network of 2nd order local motion correlators was then used to enhance the relative salience of features in the direction of travel of the winning feature. The combination of saccadic fixation and robust target-ground discrimination provided by the STMD front-end, with an attention-like 2nd order salience enhancement provided very reliable capture of tiny targets even in visually challenging scenarios.

Index Terms— Target tracking, feature detection, salience, visual processing, biological image processing

1. INTRODUCTION

Identifying and tracking a small object moving against a complex and cluttered background from a moving platform is a complex task, yet one that the tiny insect brain solves with ease. Despite the limitation of lower visual resolution (~1-2°) than a typical camera or lens eye imposed by diffraction at each facet [1, 2], many insects use impressive feats of acrobatic flight to detect and pursue small prey or potential mates [2]. Insect vision is thus a source of inspiration for efficient models of image processing for target tracking.

Electrophysiological experiments have identified Small Target Motion Detector (STMD) neurons in the brain of insects that control pursuit behaviour [2]. These respond very selectively to small moving targets, with no response to larger features [2-4]. Furthermore, STMDs are extremely robust to targets against a variety of confounding background motion, such as that generated by the insect’s self motion [4, 5] - a situation that is the downfall of many artificial target detection algorithms [6].

In this paper we used closed-loop simulations of target pursuit against natural scenery to test a computational model inspired directly by STMDs and the pursuit behaviour of insects. Our simple model is computationally efficient compared with other methods such as Kalman filters [7]. We use correlation-based 1st and 2nd order motion detectors to effectively discriminate the target against cluttered backgrounds. We further test an additional 2nd order salience enhancement (attention) mechanism that limits distractions to the pursuit.

2. METHODS

2.1. Virtual World Scenario

We simulated closed-loop pursuit of targets in a Virtual Reality (VR) arena using the Simulink 3D toolbox (Mathworks Inc.). Fig. 1a shows targets (rendered as 7x7 mm cylindrical black objects) viewed by a pursuer against panoramic natural images exhibiting the 1/² relationship between spatial frequency and power [8, 9]. The background image includes complex, high contrast background clutter. The panoramic textures were rendered onto a 1.5 m diameter cylinder. Both pursuer and target motion were limited to translation and rotation within a single, horizontal plane. Video captured from a 39.9° X 63.6° pursuer viewport (Camera 2, Figure 1c) was then fed into subsequent image processing stages to control the pursuer’s trajectory. We generated target paths at a constant velocity. These included “saccadic” changes of target direction whenever the target approached within 20 cm of the cylinder wall, as observed in typical fly flight paths in confined spaces [10, 11]. Targets started 50 cm (angular size 0.8°x0.8°) from the pursuer and their location with respect to the background was randomized.

2.2. Elementary Motion Detection

Imagery rendered from the pursuer viewpoint (Fig. 1, Cam. 2) at a frame rate of 1000Hz was then processed by an array of local motion detectors, using luminance data from the green video channel. Subsequent processing (running in Simulink at 1/50 real time) simulated early visual processing
(spatiotemporal filtering) known from insect vision. All filters were matched to time constants observed in fly vision (for full details, see ref [12]).

Fig. 1. Virtual world Setup. (a) Virtual target pursuit occurs within a thin-walled cylinder. (b) Camera viewpoint 1 (Cam. 1) records the chase from above. (c) Grayscale (green channel) information from the pursuer viewpoint (Cam. 2) undergoes analysis through the detection algorithm.

Briefly, Gaussian blur (half-width $1.4^\circ$) simulates optical diffraction of individual facets [13]. Spatial sampling at $1^\circ$ intervals then approximates resolution of the fly compound eye [14]. We simulated temporal bandpass filtering of fly early visual processing using a physiologically matched difference of log-normal filter [12].

Figure 2 (a, b) illustrates two alternative biologically inspired models for processing motion of visual features [2]. We previously showed that a non-linear “Elementary Small Target Motion Detector” (ESTMD, Fig. 2a) mimics both the sharp selectivity for small targets, and the robustness of insect STMD neurons for rejecting background clutter [4, 15, 16]. This exploits an expected property of spatially circumscribed moving features: even against cluttered backgrounds, typical targets are most likely to have a leading edge opposite in sign to the trailing edge. The ESTMD multiplies partially rectified ON signals (luminance increments) with delayed OFF signals (decrements) from a common sample location (S1, Fig. 2a). Fast dynamic adaptation (gain reduction) within each OFF and ON detector rejects noise induced by local clutter, while potent lateral inhibition of neighboring ESTMDs promotes selectivity for features of limited angular extent (orthogonal to the motion direction) [15, 16].

The Hassenstein-Reichardt (HR) detector (Fig. 2b) is an alternative model for direction-selective ‘elementary motion detectors’ (EMDs). This correlates two spatially separated contrast signals (S1, S2) after a delay (via a low-pass filter) [17]. EMDs are inherently direction sensitive, but confer no selectivity for small targets. While neither model alone accounts for all observed properties of STMDs, we recently showed that cascading the two (Fig. 2c) maintains core STMD properties but with direction-selective outputs [18].

While the original ESTMD model is selective for dark targets [15], we confer sensitivity to either polarity by summing a 2nd ESTMD in which OFF is correlated with delayed ON from the same location.

Fig. 2. Correlation-based ‘elementary motion detectors’ (EMDs). (a) The non-directional ESTMD model for an elementary small target motion detector [15] uses strong centre-surround antagonism of part-rectified ON and OFF channels followed by the correlation of the delayed OFF with the undelayed ON signal. (b) The Hassenstein-Reichardt (HR) EMD correlates two spatially separated luminance signals (S1, S2) with S1 delayed before multiplication by the undelayed S2 signal [17]. Subtraction of a mirror symmetric sub-unit then yields opponent (negative) responses for motion in the anti-preferred direction. (c) Our cascaded ESTMD-EMD model includes summation of a second parallel ESTMD for bright contrast (OFF X delayed ON). Direction of target motion is computed via a cascaded HR-EMD on the output of the non-directional ESTMD subunits (S1, S2). This is used to mediate a predictive feed-forward facilitation of ESTMD outputs in the direction of travel of the ‘winning’ element.

2.3. Target pursuit and predictive facilitation

We simulated pursuits where target location was derived from the maximum output of the array of local ESTMDs, while the output of corresponding 2nd order HR-EMDs indicated direction for the ‘attended’ target. We adopted a “saccadic tracking” pursuit mode, as described from male houseflies, which update their heading based on an error
angle between target and the central axis of the pursuer's gaze [19, 20]. In our implementation, the pursuer only re-centres its heading towards a winning feature (i.e. maximum of the ESTMD stage) when it is within 5° of the viewport edge. This allows high pass filtering in early processing to promote ‘pop out’ of the target. During such flights, the pursuer’s motion can still trigger breakthrough responses to background features, potentially drawing ‘attention’ away from the target. Our recent electrophysiological recordings from dragonfly STMDs reveal an additional facilitation mechanism that could potentially prevent such breakthrough responses by amplifying the weak signal of tiny targets moving on long trajectories [21, 22]. We mimicked this by combining the location of the winning feature with the local strength of the corresponding ESTMD response fed into a facilitation grid via a low pass filter. On each time step this was shifted predictively in the direction of motion indicated by the sign of the EMD response at the same location, and by a simple constant spatial offset matched to the 100°/s optimum speed of the ESTMD [12]. This was then used to multiplicatively weight the output of ESTMDs on subsequent time steps.

3. RESULTS

Figure 3 shows data from a single pursuit in which the pursuer experienced variably challenging periods of target motion against background clutter. We calculated a discrimination metric based on distributions of pixel values in image as sampled by the early processing (Fig 3a, c) or in ESTMD outputs (Fig 3b). Although the target is nominally the darkest feature, it is not always so in the neural image. When distant from the pursuer (as in Fig 3a, c, left, sampled from the first vertical line indicated in Fig. 3 d, e), optical blur of adjacent features reduces effective target contrast for prolonged periods (e.g. red brackets in Fig 3 d,e) during which the background scene contains many features of equal or darker luminance (Fig. 3c,d). At such time points, some of these background features may generate equal or stronger ESTMD outputs (false positives, Fig 3b, left) so that the desired target is not the winner.

Despite the large degree of background clutter (remembering that the background itself generates local feature motion as the pursuer moves past it) the output of the ESTMD array at the known location of the target (Fig 3e) is most often greater than all other points (i.e. it is the winner). Nevertheless, false positives induced by background motion are sometimes large enough to lead to a failure of pursuit in the absence of facilitation, as illustrated by a 2nd example pursuit in Figure 4. After initial detection, the prolonged passage of the target across a patch of dark background leads to a weak ESTMD output at the target location (Fig 4a). Eventually, the greater response of other (“false positive”) ESTMDs leads to gaze fixation on the background and the target is lost from the viewport (indicated by the red inset in Fig. 4 a,b). Figure 4b shows relative target strength (target minus the strength of the next strongest feature). Negative values indicate a stronger feature in the background than the target.

The addition of the 2nd order predictive facilitation mechanism improves performance in a second pursuit commencing from the same point (dotted lines in Fig. 4a, b). While this second pursuit was also conducted in a closed-loop simulation, the target path and velocity was maintained at a constant 16cm/s in both trials, and the starting position against the background was identical. In fact, both pursuits were identical for approximately 0.6 seconds, in that neither responded to the target stimulus for that period of time. At
this point though, facilitation at the target location (Fig. 4d) built up enough to boost the target response over the threshold (Fig. 4f). The pursuer then shifts its gaze (i.e. course) to attend to the target, at the same point as the pursuer in the un-facilitated trial continued straight ahead (fixating the background) and thus lost the target off the edge of the viewpoint. In the latter part of the pursuit, the (now fixated) target traverses less challenging texture and ESTMD outputs rise to the saturation level (Fig 4a, b).

Figure 5 shows the aggregated effect of the two key stages of processing described here (i.e. the ESTMD itself, and output facilitation) over large number of repeated pursuits. We first analysed the effect of the ESTMD processing for improving target discrimination, by collating data as for figure 3 across 20 separate pursuits over 350 time steps in 20 separate pursuits and 2 different background images (Image B as illustrated in Fig. 1, Image A is a less complex image, ‘All’ from the set as per Figure 4 in ref [15]). Target and pursuer velocities were held constant at 16 cm/s. Crosses represent outliers. (b) Percentage of successful pursuits over 44 trails against Image B.

4. DISCUSSION
We have demonstrated the potential efficacy of biologically inspired feed-forward saliency enhancement, with little or no higher-order “intelligence” to improving the outcome of target pursuit against cluttered, complex backgrounds. While our facilitation is a simplified version of that presumed to occur during real insect pursuit, we have nonetheless shown that combination with a hybrid 2nd order motion detector significantly improves the success rate for pursuit of a targets at the resolution limits of the optics. Importantly, at key points leading to failure in the absence of such facilitation, the target subtends a very small angle – less than 1 effective pixel in the input image. This demonstrates the potential for this type of image processing for applications (e.g. µUAVs) where resolution and computational power are limited.

Our approach to relocate the facilitation locus is equivalent to a single (position) state, high-gain Kalman filter with saturation, and as such is not optimal [7,24,25]. An improvement would be to add a second (velocity) state with a pure forward-difference estimate to predict an optimal facilitation locus. We could then determine the target location and update the estimate of the target's two states (from a weighted combination of measured target and current prediction) to replicate a two-state Kalman filter with a predict/update stage. The ESTMD-EMD system as implemented here could thus form a front-end processor for high resolution imagery with 'high order' post-processing of target trajectories using more traditional (and computationally expensive) particle or Kalman filters [7, 24-26].
5. ACKNOWLEDGEMENT

This work was supported by the US Air Force Office of Scientific Research (FA2386-10-1-4114) and the Australian Research Council (DP130104572)

6. REFERENCES


