THE UNIVERSITY of ADELAIDE

An Insect-Inspired Target Tracking Mechanism for Autonomous Vehicles

Zahra Bagheri

School of Mechanical Engineering and School of Medicine

The University of Adelaide

Adelaide, Australia

July 2017

A thesis submitted in fulfilment of the requirements

for the degree of Ph.D. in Mechanical Engineering

Thesis Summary

Target tracking is a complicated task from an engineering perspective, especially where targets are small and seen against complex natural environments. Due to the high demand for robust target tracking algorithms a great deal of research has focused on this area. However, most engineering solutions developed for this purpose are often unreliable in real world conditions or too computationally expensive to be used in real-time applications.

While engineering methods try to solve the problem of target detection and tracking by using high resolution input images, fast processors, with typically computationally expensive methods, a quick glance at nature provides evidence that practical real world solutions for target tracking exist. Many animals track targets for predation, territorial or mating purposes and with millions of years of evolution behind them, it seems reasonable to assume that these solutions are highly efficient. For instance, despite their low resolution compound eyes and tiny brains, many flying insects have evolved superb abilities to track targets in visual clutter even in the presence of other distracting stimuli, such as swarms of prey and conspecifics. The accessibility of the dragonfly for stable electrophysiological recordings makes this insect an ideal and tractable model system for investigating the neuronal correlates for complex tasks such as target pursuit.

Studies on dragonflies identified and characterized a set of neurons likely to mediate target detection and pursuit referred to as ‘small target motion detector’ (STMD) neurons. These neurons are selective for tiny targets, are velocity-tuned, contrast-sensitive and respond robustly to targets even against the motion of background. These neurons have shown several high-order properties which can contribute to the dragonfly’s ability to robustly pursue prey with over a 97% success rate. These include the recent electrophysiological observations of response ‘facilitation’ (a slow build-up of response to targets that move on

long, continuous trajectories) and ‘selective attention’, a competitive mechanism that selects one target from alternatives.

In this thesis, I adopted a bio-inspired approach to develop a solution for the problem of target tracking and pursuit. Directly inspired by recent physiological breakthroughs in understanding the insect brain, I developed a closed-loop target tracking system that uses an active saccadic gaze fixation strategy inspired by insect pursuit. First, I tested this model in virtual world simulations using MATLAB/Simulink. The results of these simulations show robust performance of this insect-inspired model, achieving high prey capture success even within complex background clutter, low contrast and high relative speed of pursued prey. Additionally, these results show that inclusion of facilitation not only substantially improves success for even short-duration pursuits, it also enhances the ability to ‘attend’ to one target in the presence of distracters.

This inspect-inspired system has a relatively simple image processing strategy compared to state-of-the-art trackers developed recently for computer vision applications. Traditional machine vision approaches incorporate elaborations to handle challenges and non-idealities in the natural environments such as local flicker and illumination changes, and non-smooth and non-linear target trajectories. Therefore, the question arises as whether this insect inspired tracker can match their performance when given similar challenges? I investigated this question by testing both the efficacy and efficiency of this insect-inspired model in open-loop, using a widely-used set of videos recorded under natural conditions. I directly compared the performance of this model with several state-of-the-art engineering algorithms using the same hardware, software environment and stimuli. This insect-inspired model exhibits robust performance in tracking small moving targets even in very challenging natural scenarios, outperforming the best of the engineered approaches. Furthermore, it operates more efficiently compared to the other approaches, in some cases dramatically so.

Computer vision literature traditionally test target tracking algorithms only in open-loop. However, one of the main purposes for developing these algorithms is implementation in real-time robotic applications. Therefore, it is still unclear how these algorithms might perform in closed-loop real-world applications where inclusion of sensors and actuators on a physical robot results in additional latency which can affect the stability of the feedback process. Additionally, studies show that animals interact with the target by changing eye or body movements, which then modulate the visual inputs underlying the detection and selection task (via closed-loop feedback). This active vision system may be a key to exploiting visual information by the simple insect brain for complex tasks such as target tracking. Therefore, I implemented this insect-inspired model along with insect active vision in a robotic platform. I tested this robotic implementation both in indoor and outdoor environments against different challenges which exist in real-world conditions such as vibration, illumination variation, and distracting stimuli. The experimental results show that the robotic implementation is capable of handling these challenges and robustly pursuing a target even in highly challenging scenarios.

Thesis Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of the published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Zahra Bágheri

Acknowledgments

I would like to acknowledge the efforts of all the people who have made a contribution towards this thesis. I would cherish memories of all of them for the rest of my life. Many thanks to my mom and dad, Marzieh and Masoud, who have been so understanding throughout my study and always encouraged me to pursue my dreams. I am thankful to their unconditional support of all my undertakings, academic and otherwise.

I would also like to thank my wonderful supervisors and mentors Ben Cazzolato, Steven Wiederman, David O'Carroll and Steven Grainger for their great guidance, insightful inputs, encouragement, endless support and friendship. I could not possibly wish for a better supervisory panel.

I would also like to give thanks to my friends Nima Sedaghatizadeh, Maryam Yazdani, Farzin Ghanadi, Pouria Aryan, Mehdi Jafarian, Pegah Haseli, Nataliia Sergiienko, Jingjing Ye, Sara Marzban and many more that I cannot list all of them who made my journey joyful and helped me through the dark moments of my study. I also would like to acknowledge Farid Christo for his tremendous emotional support and encouragement.

Many thanks to members of the physiology lab James Dunbier, Joseph Fabian, Samuel Polacek, Bernard Evans, Elisa Rigosi and Benjamin Harvey for welcoming me to the lab and helping me to understand the physiological side of the work.

I am also indebted to the computing officer in Mechanical Engineering Billy Constantine and the "electronics guys", Philip Schmidt, Norio Itsumi, Derek Franklin and Lydia Zhang for all their help with the design and construction of the electronics and experimental apparatus. Moreover, I would like to express my gratitude to Professor Ian Reid for his technical support and insightful feedback. This research has been undertaken with the

support of the Australian Research Council and the Swedish Research Council which is gratefully acknowledged.

Common Acronyms

Centrifugal small target motion detector 1 (CSTMD1)

Elementary small target motion detector (ESTMD)

Elementary motion detector (EMD)

Insect-inspired tracker (IIT)

Large monopolar cells (LMC)

Rectifying transient cell (RTC)

Small field small target motion detector (SF-STMD)

Small target motion detector (STMD)

Author's Comments

All publications within this thesis are in the exact form of the original article as published or as prepared for submission in cases where the article is not yet published, with the exception of the following considerations:

1. Typesetting and referencing format has been altered so that there is a consistent format throughout the entire thesis.
2. Figures have been inserted into the text at appropriate places, which may differ from the final published version of the papers.
3. Figures are cross referenced throughout the text as they are in the published or prepared versions of papers, but are captioned with reference to the chapter and figure number, e.g. Figure 1 in Chapter 2 is captioned as Figure 2.1.

Table of Contents

Chapter 1. Introduction

Historically, development of target tracking algorithms dates back to the era before the ubiquity of personal computers. Some of the earliest works on target tracking appeared in the 1950s (Wax, 1955), 1960s (Kalman, 1960; Sittler 1964), and 1970s (Singer and Stein, 1971). The initial works in this field were mostly proposed to automatically track objects by means of radar systems. The increasing need for automatic tracking systems and the ability of visual systems to provide rich information of the real-world such as shape, colour and texture, shifted the focus of attention towards visual tracking methods.

Due to the increasing demand for automation, developing a robust tracking algorithm has been the focus of much research during the last decade. The potential applications for such visual target tracking systems include autonomous vehicle navigation, surveillance systems, wildlife study, human assistance mobile robots, surgical robots and bionic vision. All these applications and many others identify a common requirement for technology that can successfully extract features of interest, track them robustly within complex environments through long trajectories and do so even in the presence of other distractions.

However, detecting and tracking a moving object against a cluttered background is one of the most challenging tasks for both natural and artificial visual systems. The pursuer must overcome different challenges during the pursuit to be successful:

1) **Ego motion and background motion:** The pursuer motion through the world causes background motion. This wide field optic flow is fundamental to detection, control and successful completion of pursuit.
2) **Obstacle avoidance:** Obstacles can cause collision or visual occlusions during the pursuit, therefore, obstacle avoidance is necessary for a successful pursuit.

3) **Target detection:** This task is complicated due to illumination changes, background clutter and texture, rapid changes in target appearance, partial or full occlusion, sudden changes in target trajectory, and presence of distractors within the environment.
4) **Visual attention:** Once the target is detected, the pursuer needs to maintain attention on it and ignore other salient distracters.
5) **Closed-loop pursuit:** Once all the other conditions have been met the pursuer must control its translational and rotational speed to close the distance between itself and the target.

Many algorithms have been developed over the last decade to address the problem of object detection and tracking for different scenarios. Most of these methods use assumptions to simplify the situations and make the tracking problem tractable. For instance, smoothness of motion, minimal amount of occlusion, illumination constancy, high contrast with respect to background, etc., are the common requirements in many of the developed algorithms (Yilmaz et al., 2006). Consequently, most of these methods collapse when it comes to tracking objects in real world situations, within a distracting environment or in the absence of relative background motion. Moreover, most of these methods and techniques involve complex and time consuming computational mechanisms which require significant processing capacity that makes them impractical in many applications. This identifies a clear need for an alternative and more efficient approach to solving at least a subset of the target tracking problem.

Insects such as dragonflies have a low spatial acuity visual system, and a small size, light-weight and low-power neuronal architecture. Nonetheless, they show remarkable visual guided behaviour in chasing other insects, e.g. for predation, territorial or mating behaviour, even against complex moving backgrounds (Collett and Land, 1975; Wehrhahn, 1979) or in

the presence of distractors (Corbet, 1999; Wiederman and O'Caroll, 2013). All these features make insects an ideal group to draw inspiration from in the context of target detection and pursuit. Therefore, within this thesis I am aiming at developing an insect-inspired target tracking mechanism based on the state-of-the-art knowledge in the field of insect neurobiology and behaviour.

In the following sections, first I review engineering approaches for target tracking and then introduce the state of knowledge in insect behaviour and neural mechanisms involved in their target detection and pursuit. At the end of this chapter I explore the bio-inspired motion detection algorithms**.** These underlie the central focus for my thesis.

1.1 Target Detection and Tracking Models in Engineering

Traditionally, computer vision techniques approach the problem of target tracking either from a detection or tracking perspective (Ren et al., 2003). Detection considers the video frames as independent and localizes all objects that correspond to an object model. Tracking uses temporal coherence in consecutive frames to estimate the object motion and generate its trajectory. Machine learning is often employed in both of these approaches. Detectors use machine learning to build better models that cover various appearances of the object. Trackers use machine learning to adapt to changes of the object appearance. In recent years, some research has shown that the integration of image-based target detection and tracking improves the robustness of the overall system (Wang et al., 2008; Kalal et al., 2012).

1.1.1 Detection

Object detectors are used to identify the location of the objects in an input image and play a key role in tracking, especially when the target is partially or fully occluded or moves in and out of the camera field of view. Object detectors do not make any assumptions about the number of objects nor their location in the image. The objects are described by a model that

is built in a training phase which usually remains fixed during run-time. At the core of detectors there is a binary classifier, which classifies patches in an input image, and the goal is to find a decision boundary that separates the target from the negative examples. Figure 1.1 illustrates a typical detection system. In the following I review the most common detection approaches in computer vision literature.

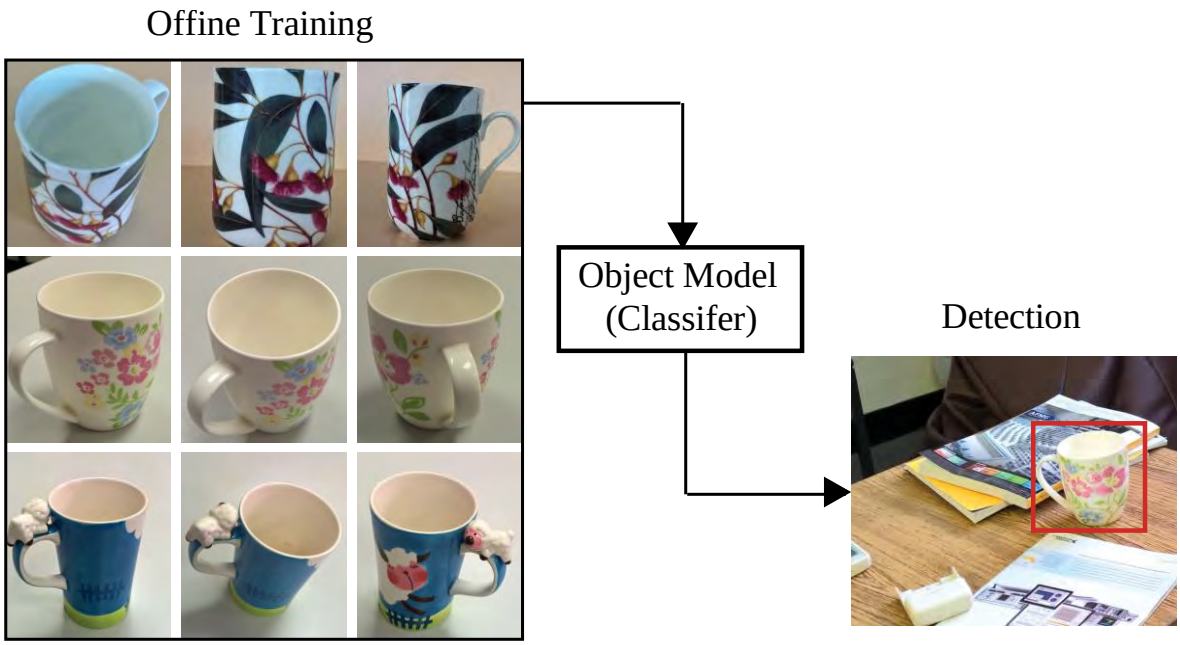

Figure 1.1. Diagram of a typical detection system showing the three stages of detection.

1.1.1.1 Features used in object detection

Target features play a key role in object detection as they provide the information about the object. The main challenge in visual detection resides in the changes in object appearance caused by viewpoint, occlusion, illumination, texture, etc. Therefore, different object features have been used in order to extract the most unique feature of the object and make it distinguishable in the feature space. The most commonly used features for description of object appearance include:

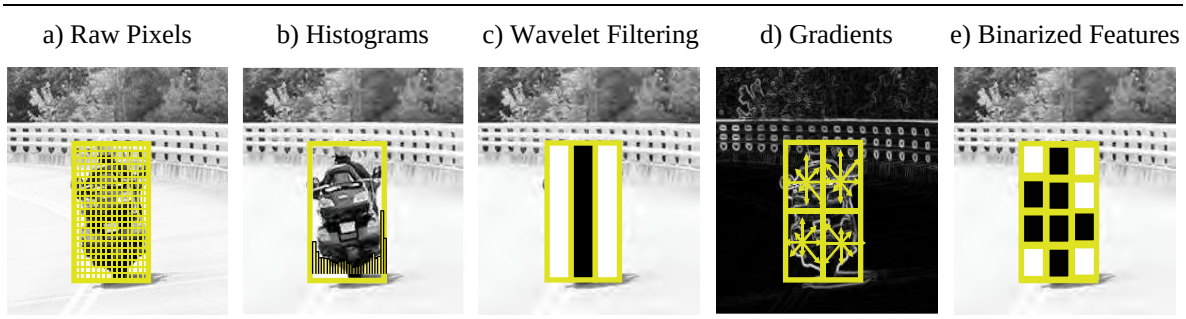

Figure 1.2. Features that are commonly used to represent appearance in object detection. The original photo was downloaded from legis.wisconsin.gov.

- **Raw pixels.** Raw pixel (Figure 1.2a) represents the objects by the intensity values of pixels. This representation method has been widely used in visual tracking due to its simplicity and efficiency (Silveria and Malis, 2007; Ho et al., 2004; Ross et al., 2008; Li et al., 2007; Wen et al., 2009; Hu et al., 2010; Li et al., 2008). However, these methods suffer from high dimensionality and lack of robustness to appearance variations.

- **Histograms.** Histogram or a combination of different histograms to represent the object appearance by distribution of colours, grey-scale values, texture, edge orientations, etc. (Figure 1.2b). The main advantage of these methods is their effectiveness and efficiency in capturing the distribution of target features. However, the histogram approach only takes into account the value of pixels rather than the location information. Therefore, one of the major drawbacks of this approach is the loss of spatial information. Moreover, this approach can easily lose the target in the background clutter with a similar histogram.

- **Wavelet filtering.** This method of object representation use wavelet transforms to filter the object regions in different scales or directions and detect predefined intensity patterns in an image patch (Figure 1.2c). One of the most popular types of wavelet filters are Haar-like filters (Papageorgiou et al., 1998). Haar-like features are

relatively robust to noise and lighting changes (Chen et al., 2008). One significant problem of Haar wavelet filters are their extremely high dimensionality in space. This problem is usually alleviated by selecting only important features.

- **Gradients.** Gradients represent a significant cue in many object recognition systems. Gradient features (Figure 1.2d) are invariant to changes in illumination and shadowing, and provide better shape cues than grey level intensity or colour patterns (Wang, 2011). The gradient-based features include edge orientation histograms (EOHs) (Gerónimo et al., 2007a; Geronimo et al., 2007b; Levi and Y. Weiss, 2004), histograms of oriented gradients (HOGs) (Dalal, 2006; Dalal and Triggs, 2005) multi-level edge energy features (Maji et al., 2008), shapelets (Sabzmeydani and Mori, 2007), and edge density (Phung and Bouzerdoum, 2007). HOGs (Dalal, 2006; Dalal and Triggs, 2005) are the most successful features for object detection, particularly for human detection and footwear-based gender recognition (Yuan et al., 2010). Since the first development of HOGs different modifications have been proposed to improve the accuracy and computational efficiency of HOGs (Wang et al., 2009; Watanabe, 2009).

- **Binarized features.** These type of features convert real-valued intensity patterns to binary codes (Figure 1.2e). The Local Binary Pattern (LBP) (Ojala et al., 2002) is an efficient texture operator commonly used in object detection. The LBP generates a binary code for a pixel's nearest neighbours. With some improvement on discriminability of the binary code, the LBP was successfully applied to face detection (Hadid et al., 2004). Robustness to illumination variations and computational efficiency are the main reasons behind the popularity of the LBP.

1.1.1.2 Detection of Object Instance

In this section I explore the approaches which detect the objects by their instances, such as a specific book cover. For certain classes of objects (such as planar or textured objects), the detection of object instances has reached a level of maturity. However, objects that are non-rigid and non-textured still remain challenging for detection. The methods for detection of object instances can be categorized as (i) global, (ii) local, and (iii) learning approaches.

(i) **Global appearance models** are top-down approaches which represent the observation of the object as a whole. These methods encode the object appearance by a collection of examples and search the image to find the closest match to the object model (Murase and Nayar, 1995; Hinterstoisser et al., 2009; Hinterstoisser et al., 2010). Global appearance models are appealing since they provide much information about the object. However, they are sensitive to viewpoint, occlusions and background clutter.

(ii) **Local appearance methods** calculate local patches around the spatio-temporal points of interest and then create a final representation by combining the related patches (Lowe, 2004; Obdrzalek and Matas, 2005; Lepetit et al., 2005; Taylor and Drummond, 2009; Pilet and Saito, 2010). Unlike the global models, local approaches are resistant to occlusions. However, extraction of sufficient relevant interest points usually requires excessive pre-processing.

(iii) **Learning approaches.** Both global and local appearance models usually have two processing stages; training and testing. The training state is essential and often requires a large number of human-annotated examples. Consequently, the application of these methods is restricted to scenarios when the object appearance is known in advance. Recent models try to process the training during the run-time. A common strategy is to add new target appearances into the object

classifier (Hinterstoisser et al., 2010; Calonder et al., 2008; Hinterstoisser et al., 2009; Pilet and Saito, 2010) but the challenge remains as how often the classifier should be updated and learn new appearances.

1.1.2 Tracking

Tracking, also known as frame-to-frame tracking, is the task of estimating object motion between consecutive frames (Figure 1.3) (Yilmaz et al, 2006). Trackers identify the object trajectory by inferring a temporal sequence of its states (e.g. location, scale, speed, pose). The implicit assumption of these algorithms is that the location of the object in the previous frame is known. However, this assumption is not valid when the target is occluded or leaves the camera view for a while. Here I use the object state to classify tracking algorithms into five categories (Figure 1.4).

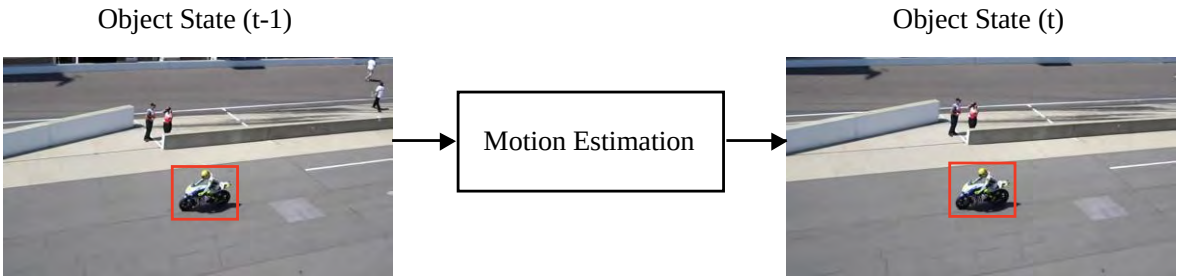

Figure 1.3. Illustration of a typical tracking system. The original photos are from ALOV300 dataset (Smeulders et al., 2014).

1. **Points** (Figure 1.4a) are usually suitable to represent small objects where their scale does not change dramatically. Point trackers estimate only translation of the object. Many successful tracking methods have been developed based on point trackers (Veenman et al., 2001; Sahfique and Shah, 2005; Zimmermann et al., 2009; Takacs et al., 2010). The work of Sahfique and Shah (2005) shows a high level of accuracy despite a significant level of noise in the tested videos. Although some proposed point tracking methods can cope with occlusions and foreground clutter, these

methods have not effectively addressed the effect of illumination changes (Cannons, 2008).

2. **Geometric shapes** such as rectangles, ellipses or other primitive geometric shapes (Figure 1.4b) (Jepson et al., 2003; Comaniciu et al., 2003; Dowson and Bowden, 2005; Rahimi et al., 2008; Kalal et al., 2012) can model the target motion by translation, affine, or projective transformation. Geometric shape representation is very popular since it can be used for general purpose tracking and real-time applications. One limitation of geometric shape models is that parts of the background might reside inside the defined shape which can lead to tracking failure.

3. **3D models** are used to represent rigid objects, for which the 3D geometry is known (Figure 1.4c) (Vacchetti et al, 2004; Leng and Wang, 2004; Lepetit et al., 2005; Klein and Murray, 2007). These 3D models can be obtained through CAD, range-finders, or using manual methods. 3D trackers estimate location, scale and pose of the object and have been applied to various objects including human faces (Vacchetti et al., 2004). However, 3D trackers have mostly been examined only under simple and controlled environments. Hence, their performance under real world conditions and cluttered environments is, as yet, largely unknown.

4. **Contours** are used to represent the boundary of non-rigid objects such as animals and the human body (Figure 1.4d). Contour trackers have been significantly improved since their original inception. Different contour trackers have been proposed (Paragios and Deriche, 2000; Yilmaz et al., 2004, Bibby and Reid, 2008; Bibby and Reid, 2010) to address some of the issues related to object tracking, such as automatic initialization and occlusion. Although these approaches have successfully solved some issues, none are truly robust to background clutter.

5. **Articulated models** are used to represent the motion of non-rigid objects consisting of several rigid parts connected to each other with joints (Figure 1.4e) (Wang et al., 2003, Ramanan et al., 2007, Buehler et al., 2008). The relationship between the parts are determined by kinematic motion models. These methods mainly have been used for tracking specific targets such as humans. Due to the complexity of these models they have remained unattractive for general purpose tracking.

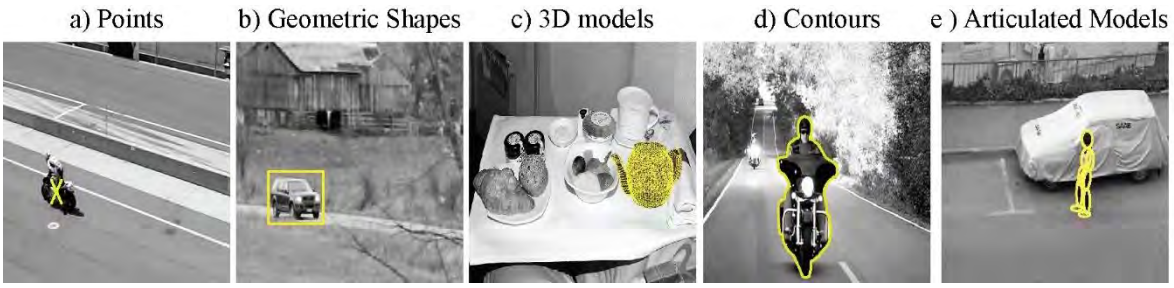

Figure 1.4. Classification of trackers based the representation of the object states: a) points, b) geometric shapes, c) 3D models, d) contours, and e) articulated models. The original photo for a is from Smeulders et al. (2014), b, d and e from Wu et al. (2015), and d from s104.photobucket.com.

1.1.2.1 Representation of Motion

Vision based tracking uses information extracted from the video and prior knowledge about the target states to estimate the object motion and fit the object model to the current frame. The Kalman filter and its variants (e.g. Extended Kalman filter and Unscented Kalman filter) and particle filters are the most commonly used models for this purpose.

Kalman filter (Kalman, 1960) is a prediction and correction tool which uses the states of the previous time step and observable measurements to compute a statistically optimal estimate for the hidden states of a system. Although Kalman filters are a powerful estimation tool, they have limitations. The mathematical model of the Kalman filter assumes that the dynamic model is linear but some systems are not well-described by linear equations. Another limitation of the Kalman filter arises from modelling the measurement uncertainties

by white Gaussian noise processes. There are many instances where this simplified model is not appropriate such as tracking a target throughout a cluttered environment, where the measurement distribution might not be a unimodal Gaussian.

Extended Kalman filter (EKF) (Bar-Shalom, 1987) is a variation of Kalman filter which was developed to provide prediction and correction for non-linear models. In the Extended Kalman filter framework a Taylor series expansion is used as a linear approximation of non-linear models. The strength of the EKF lies in its simplicity and computational efficiency. Nonetheless, unlike the Kalman filter, the Extended Kalman filter in general is not an optimal estimator. In addition, due to the Extended Kalman filter's sensitivity to linearization errors and covariance calculations, the filter may quickly diverge.

Unscented Kalman filter (UKF) (Julier et al., 1995) is another popular non-linear variation of Kalman filter. The UKF utilizes deterministic sampling methods to represent the measurement and state variables. The UKF tends to be more robust and more accurate than the EKF in its estimation of error. However, neither the EKF nor the UKF solve the cases where white Gaussian noise cannot be used as an estimation descriptor of measurement uncertainties.

Particle filters or Sequential Monte Carlo filters (Isard & Blake, 1996) maintain a probability distribution over the state of the object being tracked by using a set of weighted samples, or particles. Each particle represents a possible instantiation of the state of the object. In other words, each particle is a guess representing one possible location (or other states) of the object being tracked. The denser the particles at one location or state, the more likely it represents the target.

The main advantage of particle filters over Kalman filters (and variations) are their applicability to nonlinear models and non-Gaussian noise processes. Although with a

sufficient number of samples, particle filters are more accurate than both the EKF and UKF, when the simulated sample is not sufficiently large, they might suffer from sample impoverishment. Additionally, the number of required samples and therefore the computational load for a particle filter increases exponentially with the number of states which can cause problems for real-time applications.

1.1.2.2 Tracking models

Tracking models can be categorized in two main classes based on the type of information they represent: (i) generative, and (ii) discriminative. Generative models represent the appearance of the object ignoring the environment where the object moves. Discriminative models focus on differences between the object and the environment. Both of these models are either static (remain fixed during tracking), or adaptive (accept new information during tracking).

State-of-the-art trackers are often adaptive, i.e. they update the object model during run-time, which allows them to handle changes in object appearance, illumination or environment. One of the main problems with adaptive algorithms is that errors from the update accumulate over time and the tracker slowly slips away from the object (drift problem). Drift is different from tracking failure, which is a sudden incorrect estimation of the object state. Tracking failures typically happen when the object dramatically changes appearance, gets fully occluded or moves out of the cameras field of view.

Generative Trackers

Generative trackers (Schweitzer et al., 2002; Reddy & Chatterji, 2002; Comaniciu et al., 2003; Porikli et al., 2006; Ross et al., 2008) model the appearance of the object target and search for image regions that best match with this model. Template trackers are the most common form of generative trackers which represent the object appearance by a template

(e.g. an image patch). While it is critical to construct an appearance model which is capable of handling various challenges involved in tracking, the required computational complexity is often time consuming. To increase the speed of tracking, some models use integral images (Schweitzer et al., 2002), or the frequency domain (Reddy & Chatterji, 2002). Another approach to speed up the process of the tracking is to restrict the search area. Although restricted exploration of the neighbouring region increases the efficiency of the template tracker, the tracker may lose the target if the frame-to-frame motion is larger than expected. Therefore, many trackers use the motion models which were discussed in Section 1.2.2.1 to reduce the search area and increase the robustness of the tracking.

One of the main challenges in template tracking is finding a trade-off between static and adaptive tracking. A single static template cannot provide sufficient models for all the appearances of the object. On the other hand, adaptation of the template can lead to drift and eventually tracking failure. Recent advances (Matthews et al., 2004; Dowson & Bowden, 2005; Rahimi et al., 2008; Kalal et al., 2012) in generative tracking have shown that drift can be reduced by reusing previous observations of the target during tracking, and that the resistance to partial occlusions can be achieved by decomposing the template into independent parts. However, the generative trackers only use models of the object appearances and discard useful information from the background. Consequently, they can easily fail in background clutter.

Discriminative trackers

Unlike generative trackers, discriminative trackers use information from both the target and the background to form a binary classifier that distinguishes the object from its background. The static discriminative tracker proposed by Avidan (2004) was one of the earliest work in this field. He demonstrated successful performance of his proposed tracker on the task of

vehicle tracking. However, static discriminative trackers require an extensive offline training dataset to provide all appearances of the object and background.

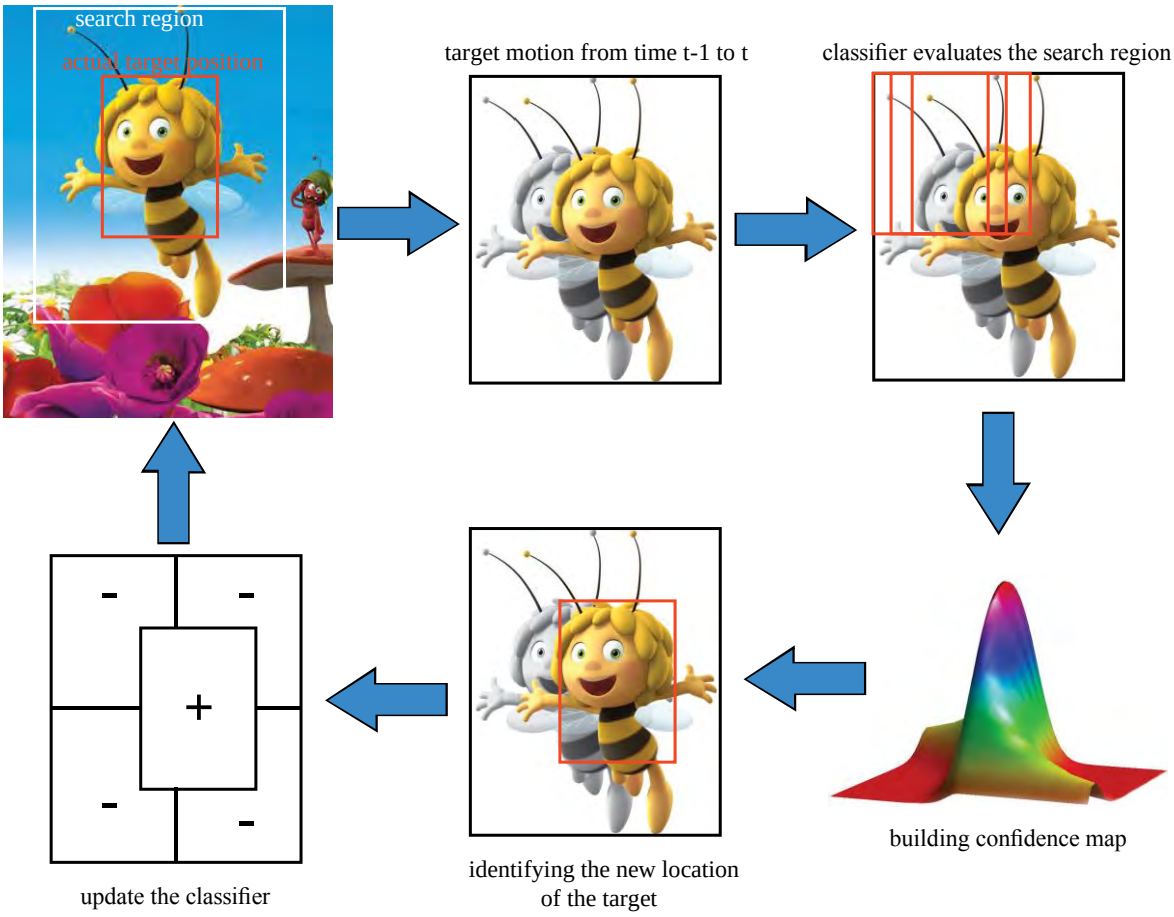

Figure 1.5. Block diagram of a typical adaptive discriminative tracker. The original image is from 'Maya the bee' movie which was downloaded from animationmagazine.net.

Adaptive discriminative trackers (Collins et al, 2005; Avidan, 2007; Grabner & Bischof, 2006; Babenko et al., 2011; Hare et al., 2011; Kalal et al., 2012) update the classifier during run-time. A typical procedure for these trackers (Figure 1.5) is that the tracker builds a simple classifier in the first frame. In each frame, the tracker applies the classifier on the surrounding area of the previous location of the object (e.g. by calculating the confidence map). Afterwards, the tracker identifies the new location of the object and performs an update. The classifier uses this new location to generate new positive and negative labels and then deploys a new update.

Adaptive discriminative trackers are capable of tracking a wide range of objects immediately after initialization. The state-of-the-art adaptive discriminative trackers (Henriques et al., 2015; Kalal et al., 2012; Hare et al., 2011) have shown successful tracking of objects with significant appearance changes and within cluttered environments. A key element in these types of trackers is the speed of the classifier's adaptation. A rapid adaptation increases the impact of new appearances on the classifier so the model can handle abrupt changes in target appearance. On the other hand, if the object is not visible for a long period (e.g. occluded or out of view) the classifier will eventually forget all of the object's information and may never recover.

1.1.3 Machine Learning

Recent advances in machine learning provide powerful tools for modern vision techniques which are required to efficiently learn from large quantities of data. These tools basically classify patches in the input image to find a decision boundary that separates the positive examples from the negative ones. Both detecting and tracking algorithms use machine learning methods to improve the labelled samples and consequently robustness of target tracking. In this section I review the machine learning strategies which are commonly used in target detection and tracking. These techniques can be categorised into two major groups; supervised learning and semi-supervised learning (Figure 1.6).

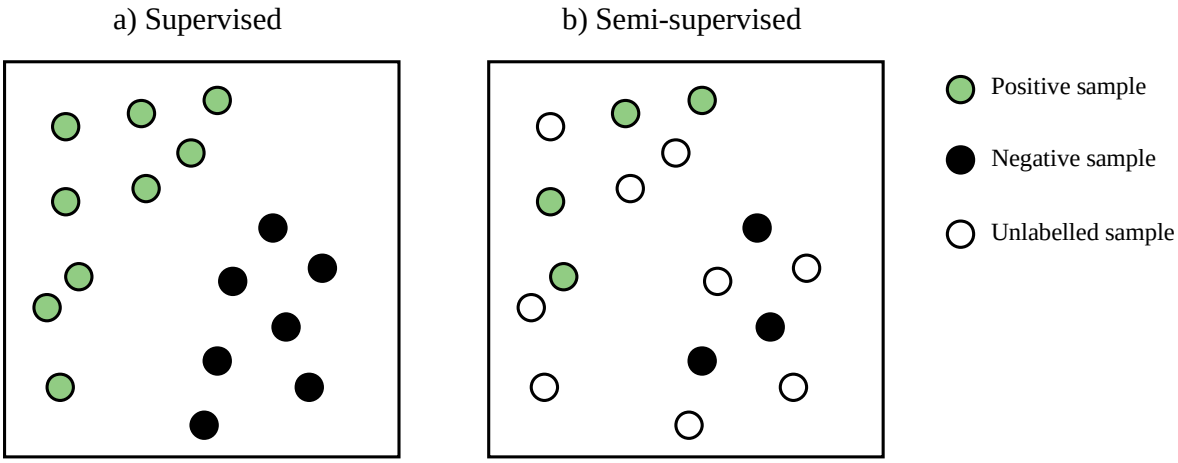

Figure 1.6. Two main categories of machine learning: a) Supervised learning which only takes the labelled samples into account; b) Semi-supervised learning which exploit both labelled and unlabelled training examples.

1.1.3.1 Supervised Learning

In supervised learning the training dataset only contains labelled data (Figure 1.6a). The objective is that the algorithm correctly determines the class labels for unseen instances (classification). The name 'supervised' implies that the learner is provided with the necessary labelled data. Detectors are traditionally trained using supervised learning. Although supervised learning is not suitable for tracking of unknown objects, it can be applied to scenarios where the class of the object is known in advance.

Boosting (Schapire, 1990; Freund, 1995) is a supervised learning strategy which is formed based on the idea that a strong classifier can be obtained by combining a set of weak classifiers (Figure 1.7). A weak classifier is only slightly correlated with the true classification. The boosting algorithm first generates a set of poor classifiers by calculating a distribution of weights over training data. Then the algorithm performs an update which increases the weight of misclassified examples and decreases the weights of the correctly classified samples. The algorithm iterates over these steps to form the strong classifier (Figure 1.7).

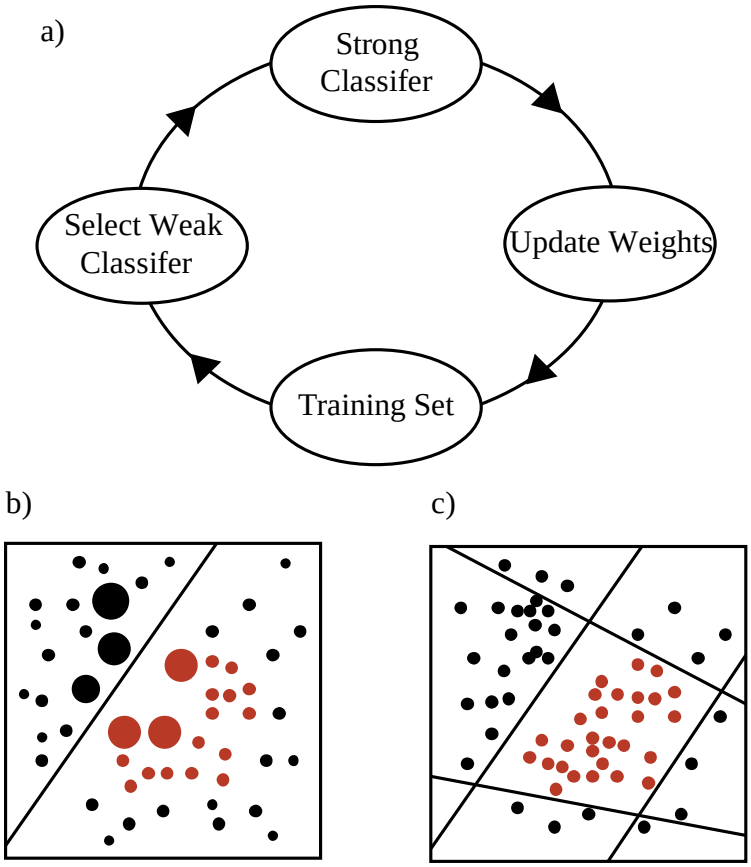

Figure 1.7. Boosting: a) the block diagram; b) a weak classifier and a weighted training set; c) a strong classifier.

The original boosting method was first proposed by Schapire (1990) and then Freund (1995) generalized the algorithm to combine an arbitrary number of weak classifiers. However, it was in 1997 when Freund and Schapire developed a practical and adaptive version of boosting named AdaBoost (Freund and Schapire, 1997). AdaBoost has been widely used in many applications including feature selection and extraction (Viola et al., 2003; Viola et al., 2005; Opelt et al., 2006).

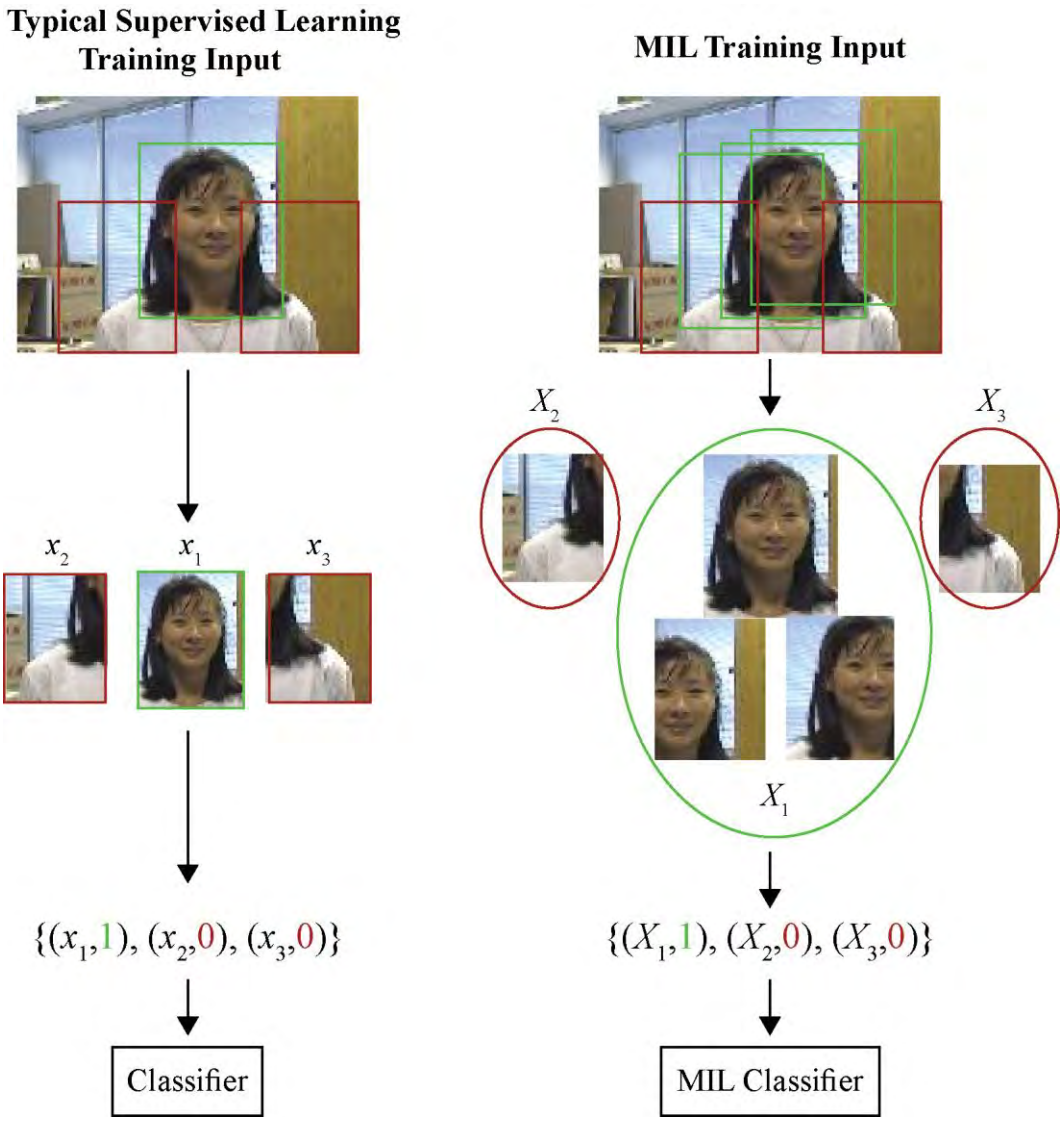

Figure 1.8. Difference between supervised learning classifier and Multiple Instance Learning (MIL) classification. The MIL algorithm labels a group of instances while other supervised learning algorithms typically label individual samples. The original photo is from Wu et al., (2015).

AdaBoost is a fast, simple and easy to use algorithm. It does not require any prior knowledge about the weak learner and so can be flexibly combined with any method for finding weak hypotheses. Nonetheless, Adaboost can fail in the cases where the given data is insufficient or the hypotheses are too weak. Additionally, boosting methods are generally sensitive to noise (Banfield et al., 2007).

Multiple Instance Learning (MIL) is a variation of supervised learning which is proposed to deal with uncertainty of sample labels. In comparison to other supervised learning methods, in MIL the labels are only assigned to a group of instances and there are no labels on the individual samples (Figure 1.8). In MIL algorithms a group has a positive label if at least one sample in that group is positive. Samples in negative labelled groups belong to the negative class, so there is no uncertainty about their label.

The MIL technique has been successfully implemented in different computer vision areas, such as object detection (Zhang et al., 2005; Dollár et al., 2008), object recognition (Andrews et al., 2003; Galleguillos et al., 2008; Vijayanarasimhan, 2008) and object tracking (Babenko et al., 2009; Babenko et al., 2011; Zeisl et al., 2010). One of the main issues of the MIL classifiers is that there might be samples within one group that do not convey any information about its group, or they can be more related to other classes. Consequently, the problem can become harder than even noisy supervised learning classifiers.

1.1.3.2 Semi-supervised learning

Semi-supervised learning exploits both labelled and unlabelled training examples to provide a high performance classifier. Figure 1.9 illustrates a simple form of semi-supervised learning strategy. The goal is to train a classifier from two labelled samples and a collection of unlabelled samples. The semi-supervised learning classifier clusters the unlabelled samples (Figure 1.9 b) and use the labelled samples to assign a class to each cluster (Figure 1.9.c). This classification is based on the assumption that if points are in the same cluster, they are likely to be of the same class (cluster assumption). Some popular semi-supervised learning models include Expectation-Maximization, self-training and co-training.

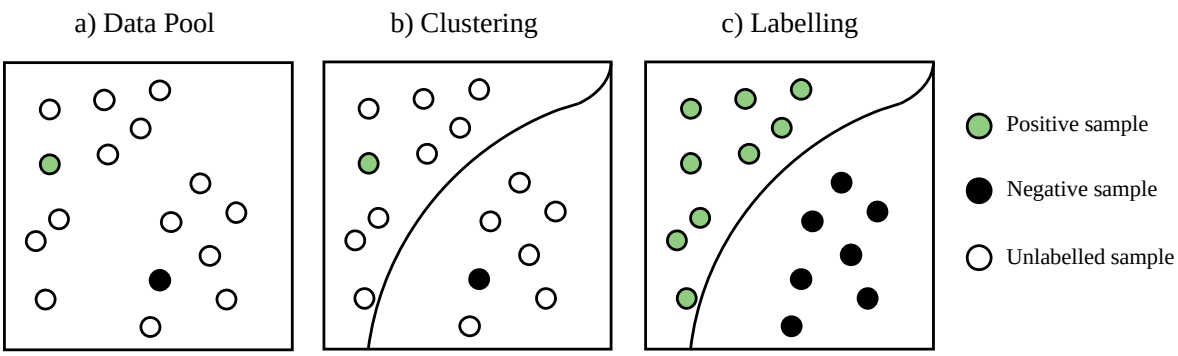

Figure 1.9. A simple semi-supervised strategy: a) uses both labelled and unlabelled samples and b) performs clustering on the data; and c) use the labelled samples to assign a class to each cluster.

Expectation-Maximization (EM) (Dempste et al., 1977) estimates model parameters given unlabelled data by iteration over two steps: (i) The E-step where the algorithm first trains a classifier using the labelled data, and assigns probabilistic weights to unlabelled data by calculating the expectation of the missing class labels. (ii) The M-step where it finds the classifier parameters that locally maximises the likelihood of both labelled and unlabelled data. Then it trains the classifier using new labelled data. EM was successfully applied to document classification (Nigam et al., 2000) and learning of object categories (Fergus et al., 2003).

The main advantages of the EM algorithm are its simplicity and ease of implementation (Couvreur, 1997). Therefore, it can be implemented using parallel computation. Also its memory requirements tend to be modest compared to other methods (Couvreur, 1997). Moreover, the EM algorithm is numerically very stable. However, it has extremely slow linear convergence is some cases.

Self-training starts by training a classifier with the small amount of labelled data. Then the classifier is used to classify the unlabelled datasets. Typically, the unlabelled examples with most confidence are added to the training set. The classifier is re-trained and the procedure repeated. The classifier is re-trained with the new data and the process is repeated. Self-

learning has been successfully applied to object detection (Rosenberg et al., 2005) and object tracking (Avidan, 2007; Collins et al., 2005; Lim et al., 2004). The major advantage of self-training is its simplicity. However, since the classifier uses its own predictions to teach itself, an early classification mistake will lead to generating incorrectly labelled data. Some algorithms (Sillito & Fisher, 2008; Uhlmann et al., 2014) try to alleviate this problem by defining a threshold for the prediction confidence.

Co-training (Blum & Mitchell, 1998) is a learning method similar to self-training with a critical difference. Co-training uses two independent classifiers to discriminate the cases that are ambiguous for the other classifier, therefore, they can mutually train each other in an iterative process (Figure 1.10). To create such independent classifiers, co-training assumes that (i) features can be divided into two independent sets and (ii) each sub-feature set is sufficient to train a good classifier.

Co-training has been used for car detection in surveillance (Levin et al, 2003), moving object recognition (Javed et al, 2005) and tracking (Tang et al, 2007; Yu et al., 2008). The method performs well only if the two sub-features actually meet the independence assumption (Nigam and Ghani, 2000). However, since the image patches (samples) are extracted from a single modality, they may be dependent and consequently violate the fundamental assumptions of co-training.

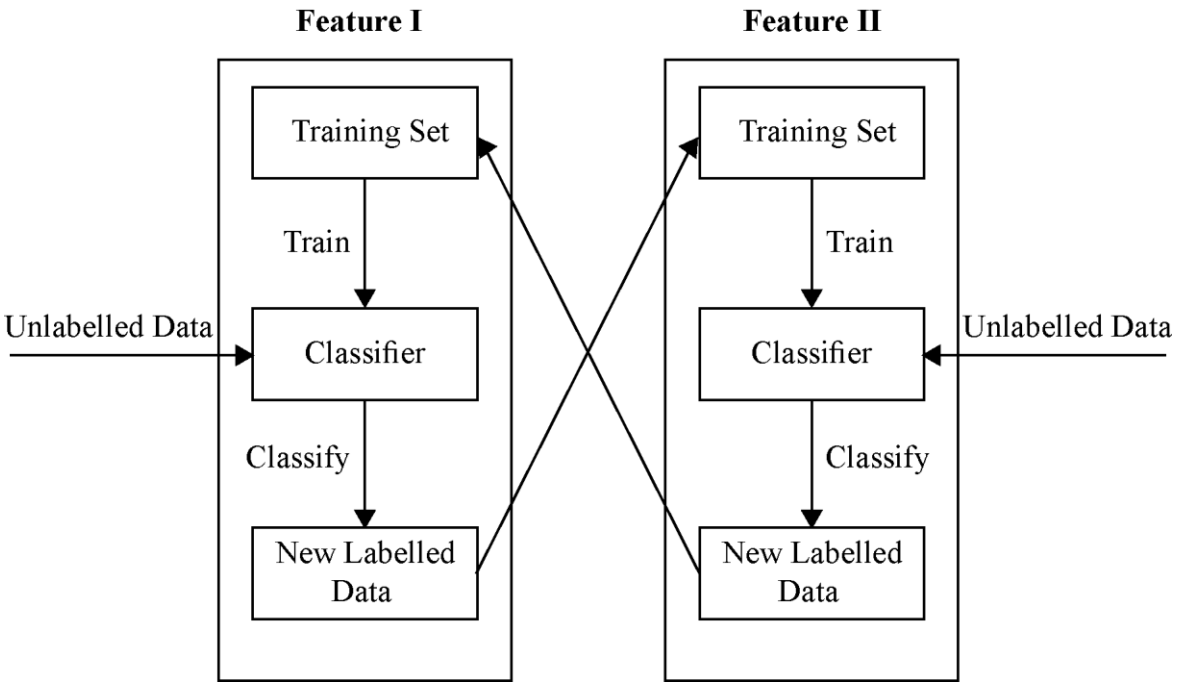

Figure 1.10. Co-training illustration. The training is initialised by the training of two separate classifiers using the labelled examples. Both classifiers are then evaluated on unlabelled data. The confidently labelled samples from the first classifier are used to augment the training set of the second classifier and vice versa in an iterative process.

1.1.4 Summary

Although the computer vision methods I reviewed here all have their own advantages and disadvantages, the main problem with the current computer vision methods can be summarised as follows. Object detection models reached a level of maturity for scenarios where a sufficient number of training examples can be generated. However, the underlying assumption of object detectors is a separation of training and run-time phase. This limits their applicability to objects that can be modelled in advance. Methods that enable efficient online update of detectors have been proposed, but the problem of when to update these detectors is still an unsolved issue.

Tracking methods are becoming increasingly complex to handle increasingly challenging environments and appearance changes. The main assumption of trackers is that the object

state in the previous frame is known. However, this is an invalid assumption in unconstrained environments due to the challenges which exist in real-world conditions such as occlusion, direct/scattered sunlight and camouflage in the background clutter. Therefore, the tracker eventually fails, especially in long term tracking.

Machine learning techniques play a key role in the state-of-the-art target detection and tracking methods. Machine learning allows building of models that cover various appearances of the object and adapt to changes of the object appearance. I reviewed two classes of learning approaches often used for target detection and tracking: supervised and semi-supervised. In the supervised approaches the learning is often realized by standard variants of boosting, which do not perform well and should be combined with bootstrapping to handle large training sets. In the semi-supervised approaches the improvements are marginal since it is hard to find independent features that would efficiently drive the learning process.

Since the main motivation for developing target tracking algorithms is their application in robotic applications, the processing speed plays a crucial role in this context. State-of-the-art target tracking models try to increase the robustness of tracking by combining detection and tracking with the addition of machine learning. However, this combination substantially increases the computational complexity and therefore, the processing time for these models. In Chapter 4, I directly compare the efficiency and efficacy of my insect-inspired model with some of these computer vision models to investigate whether insect-inspired approaches provide suitable alternatives for computer vision algorithms.

1.2 Motion Detection pathway in Flying Insects

Despite the enormous effort and significant progress in the field of visual target tracking, the lack of an efficient and robust algorithm capable of tracking objects in the most complex

environments is still evident and target tracking is still a highly active area in computer vision literature. However, a quick glance at nature shows that target tracking has been resolved even in seemingly simple biological systems such as insects. With millions of years of evolution behind them, it seems reasonable to assume that these biological solutions are highly efficient. All these observations motivated me to investigate how biological systems such as insects solve the problem of target tracking differently from computer vision algorithms.

In this section I briefly review flying insect visual pathways to provide the required background information necessary for this research and explore how these biological systems address challenges of target tracking compared to computer vision approaches.

1.2.1 Structure of a compound eye

The main visual organs of many insects are their compound eyes which are composed of thousands of closely-packed units called ommatidium (Figure 1.11). Ommatidium is the structural and functional unit of vision, where its components include the corneal lens (facets), a crystalline cone, a rhabdom and a more or less complicated pigment screen.

Externally, each ommatidium is marked by a convex corneal lens which gathers the light and directs it to the crystalline cone. Together, the lens and the crystalline cone form a dioptric apparatus that refracts incoming light down into a receptor region containing visual pigment. The light-sensitive part of an ommatidium is called the rhabdom. Within each ommatidium the rhabdom contains eight or more photoreceptors comprised of the light-absorbing visual pigments. These pigments absorb certain wavelengths of incident light and generate nerve impulses through a photochemical process.

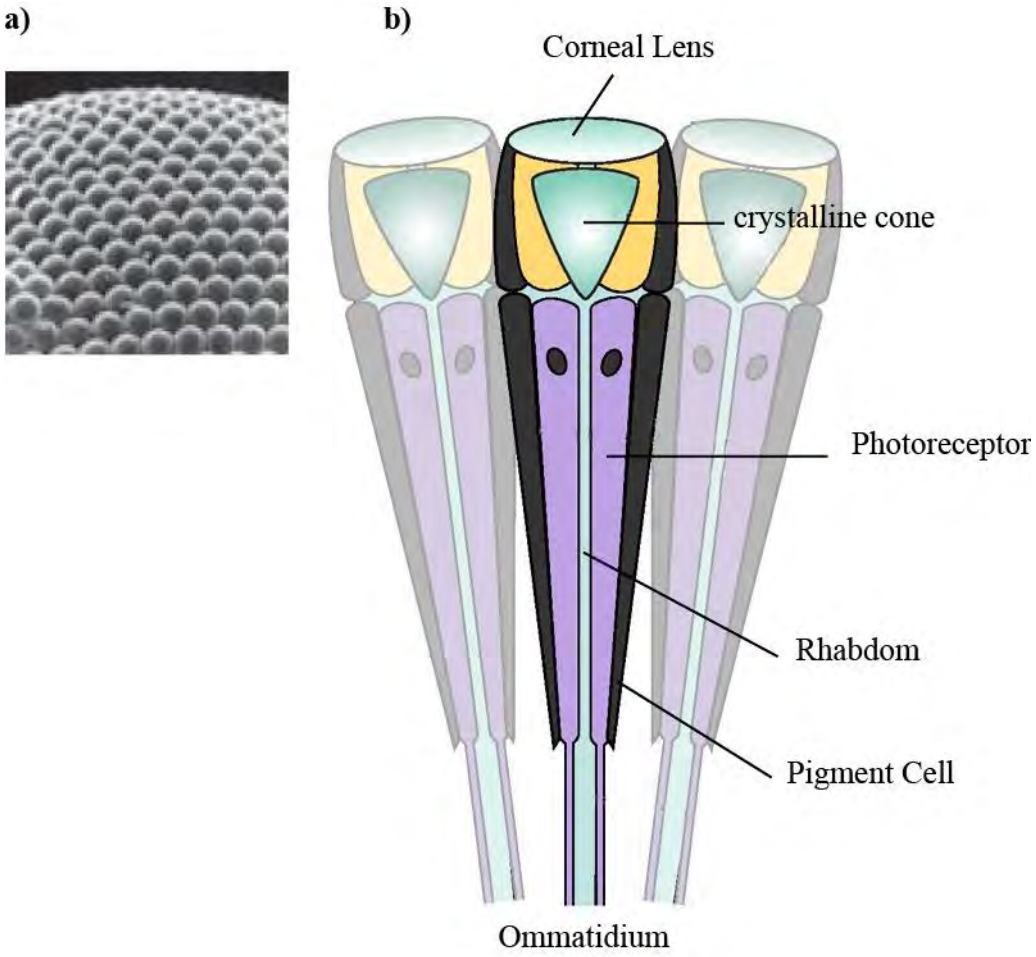

Figure 1.11. Compound eye of flying insects; a) surface of compound eye, showing the facet lenses. b) Schematic illustration of the compound eye of an insect modified from Figure 1, Srinivasan (2011).

1.2.1.1 Optics

Optics of the compound eye process the light intensity of the surrounding environment via thousands of regularly arranged facet lenses (Figure 1.11 a). Each facet points to a slightly different part of the visual field and accepts light from a narrow angle (Nilsson, 1989; Strausfeld, 1989). Collectively, the facets in the two eyes capture a near-panoramic view of the environment, with considerable binocular overlap.

The ability of an eye to resolve detail depends on two factors. The first one is the fineness of the moasic of receptive elements that sample the image, which is usually described in

terms of the spatial sampling frequency of the mosaic (Land, 1989). The second one is the optical quality of the image which mainly is determined by the spatial cut-off frequency of the optics (Land, 1989).

In a diffraction-limited eye the interreceptor angle is inversely related to the size of the aperture (Land, 1989). Therefore, the bigger the lens the finer the mosaic. Consequently, compound eyes are naturally low resolution structures due to the small diameter of the individual facets. A higher resolution of compound eye as a whole requires an increase in both size and number of ommatidia. A compound eye with the same distribution of resolution as human eye would result in an eye with a diameter of 1 m (Kirschfeld and Wenk, 1976). To overcome this limitation many insects are equipped with a region of eye with modest overall resolution. This small region is called the "acute zone" and serves for the detection of other small insects at greater distances (Horridge, 1978; Land and Eckert, 1985). Often this region is only present in the male for mating purposes (hoverflies, drone bees, mayflies) (Figure 1.12). However, in predatory insects such as dragonflies the acute zone is found in both sexes (Horridge, 1978; Land, 1997; Land 1989). A feature of acute zones is that facets are usually bigger than elsewhere in the eye (Figure 1.12).

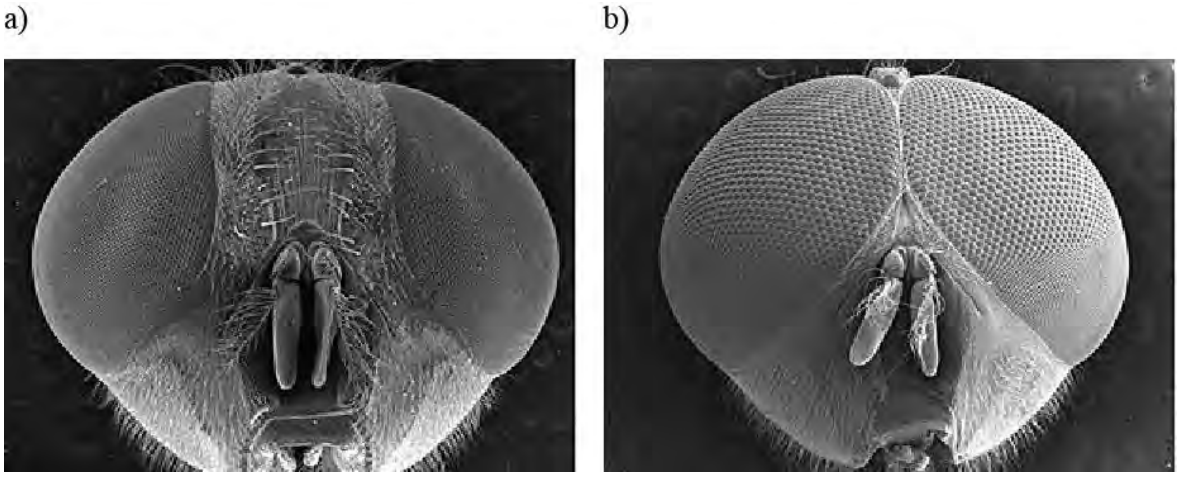

Figure 1.12. Sexual dimorphism modified from Figure 1 and Figure 2 Sukontason et al. (2008). The compound eye of a) female (left) and b) male (right) C. megacephala. The acute zone in dorsal eye of male has larger facets which provides a higher resolution for mating purposes.

1.2.1.2 Photoreceptors

Object luminance varies significantly within natural environments due to constant change in sunlight, shadows and reflection from surfaces. Therefore, photoreceptors should be able to adapt to changes in light level of at least 2-3 $\log_{10}$ units in a very short time (van Hateren, 1997).

The photoreceptors which are involved in the motion processing pathway are sensitive to UV or green-blue light (Hardie, 1985). Photoreceptors code for contrast (Barlow, 1961; Laughlin et al., 1987) and change their contrast gain dynamically dependent on the background to reduce noise and improve the signal-to-noise ratio (SNR) (Juusola et al., 1994). Generally, photoreceptors act as low-pass filters that change to band-pass filters with increasing light adaptation (Jarvilehto and Zettler, 1971). The study by Brinkworth et al. (2008) shows that the temporal processing in photoreceptors improves the spatial discriminability of the target by approximately 70%. Photoreceptors transform the processed information to the visual interneurons in the lamina by producing graded membrane potentials (Yarfitz and Hurley, 1994, Hardie, 2001).

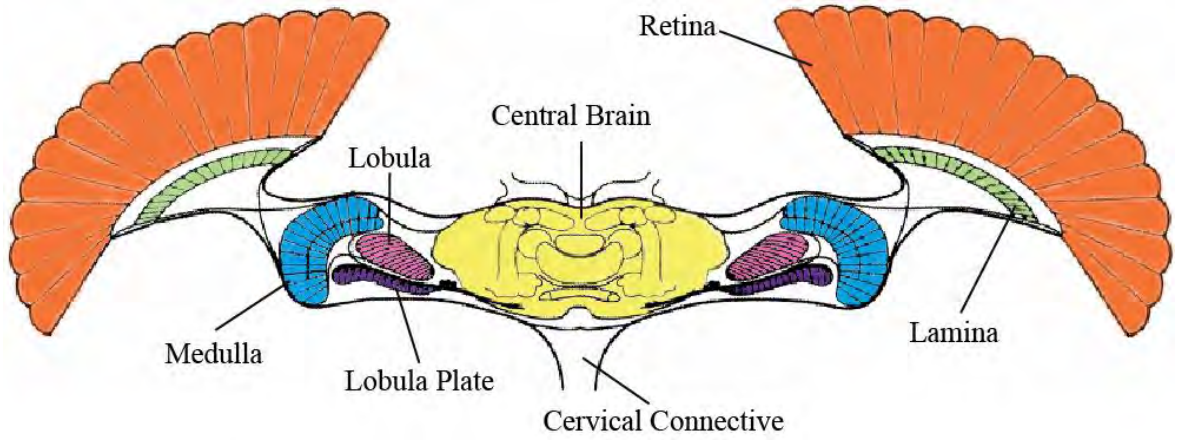

Figure 1.13. Schematic of horizontal cross section of fly head modified from Figure 2 Hausen (1982) showing the visual processing regions including retina, lamina, medulla, lobula and lobula plate.

1.2.2 Lamina

The second optic ganglion, the lamina (Figure 1.13), is the primary site of redundancy reduction in the insect visual pathway. Each retinotopically arranged group of neurons in the lamina, which is referred to as neuro-ommatidia, corresponds to a sampling point in visual space. This columnar organisation is maintained through the subsequent visual stages (Laughlin, 1984; Shaw, 1984).

The large monopolar cells (LMC) of lamina are directly postsynaptic to the photoreceptors and remove redundant information by spatio-temporal high-pass filtering (Srinivasan et al., 1982). Studies show that the LMC dynamically changes its filtration characteristics (both in time and space) depending on the current visual conditions (Laughlin et al., 1987; Juusola et al., 1995; Srinivasan et al., 1990). While under dark conditions (i.e. low SNR) the LMC act more like low-pass filters becoming more transient and high-pass in nature as the luminance increases (Juusola et al., 1995). These transient responses enhance object boundaries and consequently improve target discriminability (Srinivasan et al., 1990). Spatial filtering in the LMCs is believed to be the result of inhibitory interactions between nearest neighbor receptors (Srinivasan et al., 1982).

1.2.3 Rectifying Transient Cells

The downstream neurons that have been identified in the 2nd optic neuropil (medulla) of locust (Osorio, 1991; O'Carroll et al., 1992), and blowfly (Jansonius and van Hateren, 1991; Wiederman et al., 2008), separate transient ON and OFF phases via partial rectification. Wiederman et al. (2008) refer to these neurons as rectifying transient cells (RTCs). Each sub-pathway adapts independently to luminance changes, dependent on the polarity (increment or decrement) of the change (Jansonius and van Hateren, 1991). The state of adaptation is fast if depolarizing, and slow when repolarizing. Consequently, RTCs quickly adapt to repeated inputs (background variance), while selectively responding to 'novel' stimuli (targets) (Wiederman et al., 2013). These processing properties of RTCs are well suited as additional input processing stages for the insect target motion detection pathway (Wiederman et al., 2013).

1.2.4 Higher Order Pathway

In previous sections I explained how optics and early visual processing neurons play an important role in the enhancement of a visual signal. However, the motion specialized neurons do not appear until the second and third optic ganglia of the insect brain, the medulla and lobula. In this section I briefly explore the motion detection neurons but the main focus is "small target motion detection" neurons.

1.2.4.1 Lobula Plate Tangential Cells

Lobula plate tangential cells (LPTC) respond to wide-field motion. These cells are sensitive to directional visual motion in areas of the visual field and often correspond with the three rotational elements (pitch, yaw and roll) as well as translational, progressive self-motion (Borst and Haag, 2002; Krapp and Hengstenberg, 1996).

1.2.4.2 Lobula Giant Motion Detector and Descending Contralateral Motion Detector

The lobula giant motion detector (LGMD) system within the lobula are respondent to 'looming' motion (Strausfeld and Nassel, 1980; Rind and Simmons, 1992; Simmons and Rind 1997; Gabbiani et al., 2002). Looming neurons are sensitive to the motion of objects growing larger as they approach. These neurons are likely to be part of a circuit which triggers escape/avoidance behaviour when a looming visual stimulus reaches a certain angular size (Hatsopoulos et al., 1995; Santer et al., 2012).

1.2.4.3 Feature Detecting (FD) Neurons

Feature detecting (FD) neurons of lobula plate, are motion opponent neurons and respond to medium size (rather than wide-field) motion (Egelhaf, 1985). FD neurons respond optimally to gratings with limited spatial extent (>10º) travelling in the preferred direction (Egelhaaf, 1985).

1.2.4.4 Small Target Motion Detector Neurons

Small target motion detector (STMD) neurons are a class of lobula neuron which are likely to play an important role in target detection and discrimination. These neurons are size and velocity tuned and contrast sensitive which robustly and specifically respond to small moving objects even in presence of background clutter and motion (O'Carroll, 1993; Nordstrom et al., 2006; Nordström and O'Carroll, 2009; O'Carroll and Wiederman, 2014). This response behaviour of STMDs is even more impressive when we note that the features that STMDs are tuned to are on the same scale as the spatial sampling resolution of the eye itself, therefore, optical blur may make the feature contrast very low (Nordstrom et al., 2006; O'Carroll & Wiederman, 2014).

Significant variations have been observed in the receptive field size, direction selectivity and response modality of the STMDs (Nordstrom & O'Carroll, 2009). The receptive field of

STMDs varies between only a few degrees (small-field STMDs) to one hemifield or both hemispheres (large-field STMDs). The small-field STMDs are retinotopically organised (Barnett et al., 2007) and it is hypothesized that their outputs are likely processed by large-field STMDs (Geurten et al., 2007; Nordstrom & O'Carroll, 2009).

One type of STMD neuron from the lateral mid-brain of the dragonfly, the 'centrifugal small target motion detector 1' (CSTMD1), has shown several high-order properties such as facilitation and selective attention which can contribute to the dragonfly's ability to robustly pursue prey with over a 97% success rate (Olberg et al, 2000).

Facilitation. In general, facilitation can be defined as the enhancement of a neuron's response to a stimulus as a result of a prior stimulation. Recent studies (Nordstrom et al., 2011; Dunbier et al., 2011; Dunbier et al., 2012) have shown strong evidence for facilitation in CSTMD1 neurons. The onset response of CSTMD1 builds up slowly to objects moved continuously through the receptive field and reaches a steady state over a time course of hundreds of milliseconds (Nordstrom et al., 2011). Despite the sluggish onset response, CSTMD1 have shown fast decay to object disappearance (Nordstrom et al., 2011). Nordstrom et al. (2011) argued that this asymmetry is a strong evidence against just a simple low-pass filter mechanism in higher order neurons that integrate local motion detectors. They suggested that the asymmetric time course of the CSTMD1 response is due to a slow facilitation in responses. Further research supports this, showing that the CSTMD1 resets to a naïve, non-facilitated state when there are large breaks in the trajectory (Dunbier et al., 2011; Dunbier et al., 2012). Furthermore, even with this slow build-up in response, the underlying velocity tuning of the neuron is similar to other STMD neurons (60-190 deg.s^{-1}) (Geurten et al., 2007). One possible benefit of a such a facilitation mechanism is that it would potentially reject noise within noisy (cluttered) environments, permitting the very high amplification required to respond to very small or low contrast targets.

Selective Attention. Attention refers to the processes by which an organism selects a subset of available information for particular focus (Treisman, 1969; Driver, 2001). There are two major theories for mechanisms by which object selection occurs. The first one is a bottom-up theory which suggests that early stages of visual processing provide 'visual salience' to some stimuli to stand out from the crowd (Itti and Koch, 2000). The other component of attention involves an endogenous, top-down process. In this process the animal directs attention to a small subregion of the visual field and deliberately suppresses the relative salience of other areas, even if the features there are inherently more salient from a signal detection standpoint (Treisman and Gelade, 1980).

Neuroanatomy of CSTMD1 neurons suggest a potential role in attention as targets move from one visual hemisphere to the other (Bolzon et al., 2009; Geurten et al, 2007). Wiederman and O'Carroll (2013) have tested this possible role by comparing CSTMD1's response to single and paired stimuli. They showed that this neuron competitively selects one target in the presence of distracters responding as if only one target was presented. This competitive selection was observed irrespective of target size, contrast, or separation. Facilitation can potentially play an important role in this competitive selection by enhancing saliency of one target over the other.

Such a competitive selection is essential for a control system for target pursuit allowing tracking of individual targets in the presence of distracters, without changing the gain of the control loop. Despite recent breakthroughs in insect physiology it is still unclear where CSTMD1 sits within such a target pursuit control system or what is the underlying hierarchy of mechanisms of competitive selection. The CSTMD1 could reflect the output of a bottom-up attention mechanism emerging from a competitive process occurring at a lower level in the STMD pathway. However, there is no evidence which rules out a top-down, endogenous

attention process. Current physiological experiments try to investigate the possibilities of these two mechanisms in flying insects.

1.3 Target Detection and Pursuit in Insects

Insects adopt a number of behaviours that enhance their ability to detect and pursue targets in their environments. These behaviours synergise with anatomical and optical adaptations to enhance detection of targets. In this section I briefly overview these behaviours.

1.3.1 Target Detection

Three main detection modes are identified in flying insects: perching, hovering and hawking. Perching or initiating predation from a fixed perch is the most common detection behaviour among flying insects. This 'sit and wait' detection strategy allows the motion of a relevant target to 'pop out' against a stationary background (Srinivasan, 1998). A 97% of successful capture rate is reported for perching dragonflies (Olberg et al., 2000).

Hovering species, including hoverflies and bee-flies, hover at a fixed spot within their territory, maintaining position by minimizing optic flow. Hovering is more visually challenging than perching, however, it reduces the reaction time for the animal since it is already in flight.

Hawking is the most challenging mode of target detection which involves detection and pursuit of conspecifics and prey whilst patrolling over large areas of land or water (Corbet, 1999; Sherk, 1978). The main challenge in hawking mode is identification of target motion in the presence of the pursuer's ego-motion. In several hawking insects, the difficulty of target detection against the background is alleviated by two strategies. Firstly, these insects often detect targets in the fronto-dorsal eye region, i.e., against the sky, which provides a vividly clear background for detection of small moving prey (Syder et al., 1977; Zeil, 1983; Corbet, 1999). Secondly, many species keep the target centered in the frontal eye region,

where the expansive ego-motion during their forward motion through the world is minimum (Land, 1981; Land, 1997).

1.3.2 Target Discrimination

Once a target is detected, the animal needs to identify whether it is a viable prey, a conspecific competitor, a potential mate, a predator, or just an irrelevant distant object. Object grouping can be thought of as the ability to classify items that are 'similar' according to their shared characteristics, even though they are distinguishable from one another. These are the tasks similar to the ones that computer vision literature try to achieve via machine learning (Section 1.1.3). Studies have revealed that monkeys and other primates are able to categorise complex visual images, such as photographs of human faces, trees and other animals (Davenport and Rogers, 1971; Vogels, 1999; Freedman et al., 2001). Pigeons also have the capacity to group objects into a number of different categories, such as people, other pigeons, trees, water, landscapes and so on (Mallott and Siddall, 1972; Herrnstein, 1984; Roitblat, 1987; Huber et al., 2000).

The traditional view of target discrimination in insects is that they lack cognitive ability, and the insect brain is just a simple 'hard-wired' circuit (Giurfa and Manzel, 1997) which evokes a fixed motor pattern in response to an external stimulus. For example, a male drosophila fruit fly will mate with a female fly once pheromonal, visual and mechanosensory cues coincide in the right pattern and above a certain threshold (Greensan and Ferveur, 2000).

However, recent studies have observed remarkable visual learning and discrimination abilities in honeybees. Indeed, bees are capable of discriminating complex forest scenes (Dyer et al., 2008), categorizing different flower shapes (Zhang et al., 2004), human faces (Dyer et al., 2005; Dyer and Voung, 2008, Avarguès-Weber et al. 2010), and more

surprisingly they even can be trained to discriminate Monet paintings from Picasso ones (Wu et al., 2013).

Although numerous studies have demonstrated that honeybees use a range of stimulus features such as colours (Chittika et al., 1993), shapes (Lehrer et al., 1995), symmetry (Horridge and Zhang, 1995; Giurfa et al., 1996; Horridge, 1996) and orientation of objects (Wehner, 1971; van Hateren et al., 1990), the cues that honeybees use to solve complex visual categorization are still under investigation. Some studies stay with the traditional view and believe that bee vision relies only on low-level feature detectors and elemental cues with little or no plasticity for learning (Horridge, 2000; Horridge, 2005; Horridge 2009). However, this simple elemental processing cannot explain the ability of bees to apply previously acquired information in solving novel tasks, categorize new stimuli that significantly differ in low-level cues, and transfer abstract concepts to novel domains. Therefore, other views suggest that for simple visual tasks, honeybees may rely on elemental processing. However, as the complexity of the task increases, honeybees can learn to switch to non-elemental processing (e.g. configural type processing and rule-learning), and use top-down information to solve novel tasks (Giurfa et al., 2003; Stach et al., 2004; Stach and Giurfa, 2005; Giurfa, 2007; Avargue`s-Weber et al., 2010; Dyer, 2012).

Although visual object categorization has only been studied in honeybees, it is likely that other insects such as dragonflies employ grouping and learning for complex tasks such as target detection and tracking.

1.3.3 Pursuit Strategy

During a pursuit, an insect has to control its forward velocity and distance to the target while fixating the target in the frontal visual field. Two different gaze control strategies have been observed among flying insects (Figure 1.14), referred to as tracking and interception (Collet

and Land, 1978). Many insects use a tracking strategy where the pursuer steers to minimize the deviation of the pursued target from the pursuer's visual midline (Fig 1.14). Tracking results in spiraling flights that will result in a successful pursuit if pursuer is faster than the target.

Interception (Collet and Land, 1978) is a pursuit strategy which is observed in hoverflies and dragonflies (Collet and Land, 1978; Olberg et al., 2000) (Figure 1.14b). Interception is a process which requires prediction and planning which is functionally similar to the reaching process in primates. This high-performance control of behaviour requires internal models of sensorimotor system (Franklin, and Wolpert, 2011). Studies of humans and non-human primates have identified three types of internal models involved in sensorimotor control: 1) physical models to predict properties of the world (Zago et al., 2004; Flanagan et al., 2001); 2) inverse models to generate the motor commands needed to attain desired sensory states (Kawato, 1999); 3) forward models to predict the sensory consequences of self-movement (Wolpert et al., 1995; Mehta and Schaal, 2002). However, up until recently, it was unknown whether insects rely on internal models to guide actions.

Mischiati et al. (2015) investigated the existence of internal models in dragonfly by tracking the position and orientation of the dragonfly's head and body during flight. They provided a compelling case that interception steering relies on both predictive and reactive control which is driven by forward and inverse models of dragonfly body dynamics and by models of prey motion with visual feedback.

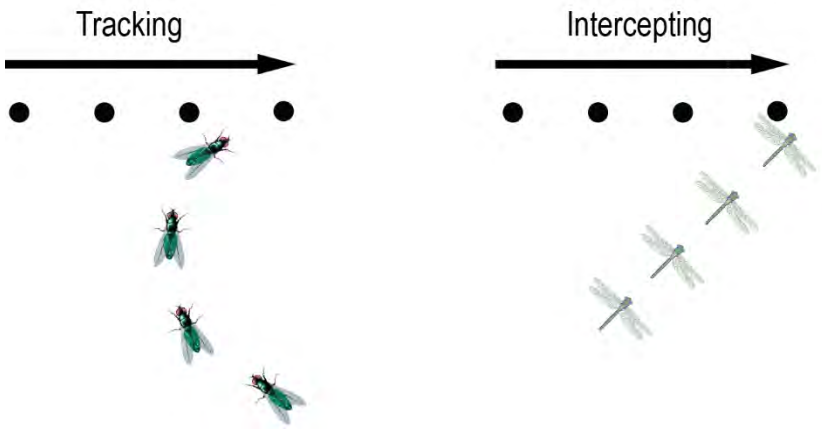

Figure 1.14. Two main flying insects' pursuit strategies modified from Figure 1 Olberg et al. (2000). The black circles represent the target.

1.4 Biologically-Inspired Motion Detection Models

As I described in the introduction of this chapter, the pursuer faces two general form of motion. The first one is induced by the pursuer ego-motion causing rotary and translator movement of the entire visual field. The second one is the movement of objects within visual surround. The computation basis of motion detection in biological system has been the focus of many studies and two categories of motion detection mechanism have been proposed in the literature. The first category which has been proposed to explain motion vision is 'feature tracking' mechanisms (Braddick 1980; Ullman 1983). As I explained earlier 'feature tracking' mechanisms are extensively used in computer vision. The main disadvantage of such schemes is identification of the location of features. The other category of motion detection models is 'intensity based' which employs information on local spatiotemporal changes of intensity to measure velocity. Two mathematically distinct families of 'intensity based' models are the gradient scheme (Limb and Murphy, 1975; Mar and Ulman, 1981; Srinivasan, 1990) and correlation schemes (Hassenstein and Reichardt, 1956).

The gradient detector calculates the velocity signal by dividing the temporal derivative of local luminance $\boldsymbol{\partial I(x,t)/\partial t}$ by its spatial derivative $\boldsymbol{\partial I(x,t)/\partial x}$ (Figure 1.15a). The gradient detector provides a signal that is proportional to the image velocity at each point and does not depend on pattern properties. Another subclass of the intensity-based models is 'correlation type' detectors, which apply spatiotemporal correlations to extract motion signals (Reichardt, 1957). The correlation based models has been very successful in describing motion sensitivity in animal vision from insects to primates (Borst and Egelhaaf 1989). In the following sections, I explore the models that are most relevant to the content of this thesis in further detail.

1.4.1 The Hassenstein-Reichardt Detector

The Hassenstein-Reichardt elementary motion detector (HR-EMD) is the simplest form of EMD which is inspired by the study of behavioural turning response (the optomotor response) of the beetle Chlorophanus to the movement of the visual surround. The HR-EMD model (Figure 1.15b) is composed of two spatially separated input channels such as photoreceptors (Hassenstein and Reichardt, 1956). The HR-EMD multiplies these signals, with one signal delayed relative to the other (Hassenstein and Reichardt, 1956). The output of HR-EMD is a direction selective signal which will be more positive when a pattern moves in the preferred direction and more negative in the reverse direction. The spatial separation and correlation provides sensitivity to moving stimuli without computing derivatives (a process that would amplify noise).

Various versions of this basic EMD with different filtering schemes have been proposed to make the model robust to diverse static environmental visual conditions. However, some of the elaborations to the EMD that enable the model to operate under more demanding visual conditions reduce its efficiency at the basic processes for which it is best known (Frye, 2015).

Although the predictions of EMD models are consistent with a number of behavioural and electrophysiological results, it remains as an ambiguous motion sensor. The output of EMD is an ambiguous function of velocity of a moving stimulus, although it has a strong dependence on spatial structure and contrast of the stimulus. These shortcomings, as well as the results of additional behavioural experiments that appear inconsistent with model predictions suggest that other mechanisms of motion detection may be involved in the insect brain (Srinivasan et al. 1993).

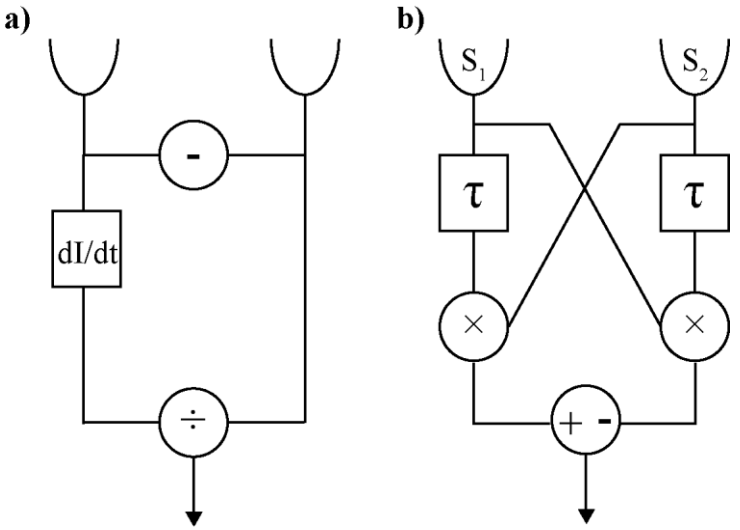

Figure 1.15. a) The gradient detector calculates velocity by dividing the derivative of luminance over time, by the derivative of luminance over space. b) HR-EMD. The correlator composed of two sub-units. Each sub-unit consists of two spatially separated inputs, with one of the signals delayed relative to the other before multiplication between two arms.

1.4.2 Elementary Small Target Motion Detector

A modified correlator model has predicted a different set of computations for an especially challenging form of visual perception, small object detection. The recent work of Wiederman et al. (2008, 2007) provides a neuromorphic model for target discrimination in visual clutter. This model suggests that STMD neurons with large receptive fields could be the product of summation across a retinotoptic array of putative 'elementary STMDs'

(ESTMDS). The Wiederman et al. model includes nonlinear filtering based on fly optics, the photoreceptors, LMCs, and RTC (Figure 1.16). Their model provides a good match for size and velocity tuning, and contrast sensitivity of STMDs. Subsequently, Halupka et al. (2011, 2013) developed a discrete closed-loop version of this model which displayed very similar tuning characteristics to the continuous version.

However, the ESTMD model is not a true motion detector, but rather a spatiotemporal filter for a small, dark target whether moving continuously or just a single flicker. Nonetheless, it is unlikely that STMD neurons respond to a rapid variation in brightness. The recent observation of facilitatory behaviour of CSTMD1 neurons (Nordström et al., 2011; Dunbier et al., 2011; Dunbier et al., 2012) as well as their selective attention (Wiederman and O'Carroll, 2013) have raised interesting new questions regarding the role that CSTMD1 plays in the small target motion detection pathway. These findings suggest that the dragonfly displays a dynamically shifting centre of salience dependant on the previous stimulation. Therefore, it is respondent to persistent stimuli such as another insect which moves in a long continuous trajectory rather than short transient stimuli (i.e. local flicker).

Moreover, directional information is a key requirement to predict the future path of a target. While many STMD neurons have shown selectivity for the direction of target movement (O'Carroll, 1993; Barnett et al., 2007; Nordström and O'Carroll, 2006), the ESTMD model is a non-directional model. Additionally, while the ESTMD model is only selective for dark stimuli, the target polarity and contrast can change significantly dependant on the background and lighting throughout the pursuit. Therefore, a mechanism is required to track targets independent of their contrast polarity. All these suggest that robust target tracking behaviour observed in dragonfly requires additional computational components to address the limitations of the ESTMD model.

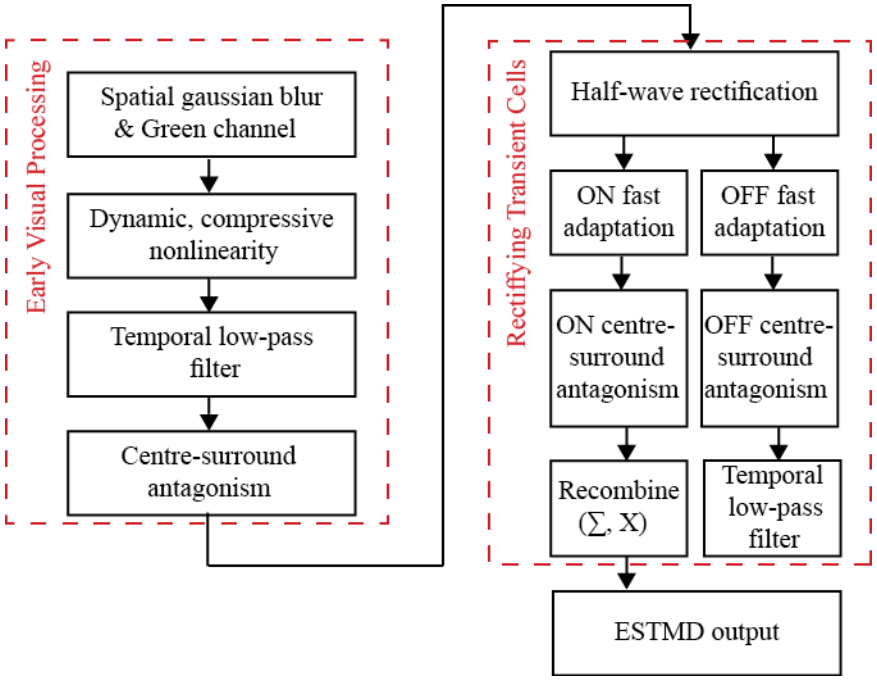

Figure 1.16. Overview of ESTMD model modified from Figure 2 Wiederman et al. (2007).

1.4.3 Summary

Our research group has been investigating the neuronal mechanism underlying target detection and tracking in flying insects through electrophysiological experiments and computational modelling for several years. Although several motion detection models have been developed based on the insect visual system (as discussed above), the model that I present in this thesis is an extension of an earlier model developed in our own research group (Halupka et al., 2011; Halupka et al., 2013). As I mentioned previously, the Halupka et al. (2011, 2013) model is a discrete version of the ESTMD model (Wiederman et al., 2008), which has been implemented in closed-loop using a virtual reality environment. However, like the original ESTMD model, the Halupka et al. (2011, 2013) model lacks certain characteristics of the STMD neurons as I discussed in Section 1.4.2, including a predictive facilitation mechanism and selectivity for small moving targets irrespective of their contrast polarity.

1.5 Hardware Applications of Insect-Inspired Motion Detection

In recent years, the insect-inspired motion detection models and their applications to bio-inspired robot sensors and analog-VLSI chip design have been the focus of much research. Franceschini et al. (1992) was one of the first studies which employed the principles of 'elementary motion detector' to develop ground-based robots that would navigate autonomously and avoid collisions with obstacles. Following this work different studies attempted to implement insect-inspired vision models on hardware platforms for various applications which can be summarized as three different types:

1) Bio-inspired circuits embedded in the control structure of mobile robots. Examples include a model of locust Lobula Giant Movement Detector (LGMD) for collision detection (Blanchard et al., 2000), fly-inspired obstacle avoidance (Zufferey and Floreano, 2006), and safe navigation through narrow corridors (Coombs and Roberts, 1992; Santos-Victor et al., 2012; Conroy et al., 2009).

2) Bioinspired chips such as the neuromorphic chips developed based on fly vision (Harrison, 2000), neuromorphic eyes for mini-unmanned aerial vehicles (UAVs) (Ruffier and Franceschini, 2003), and VLSI retinal circuits (Liu, 2000).

3) Bio-inspired behavioural strategies such as docking (Zhang et al., 2013; Zhang et al., 2014; Xie et al., 2013; Kendoul, 2014), executing smooth landings (Thurrowgood et al., 2014; Srinivasan et al., 2001), or interception (Strydom et al., 2015).

In all these insect-eye-inspired designs, the goal has been to make fast, robust, lightweight and low-power vision systems.

1.6 Thesis Aims and Scope

Although very different in detail, target tracking in both computer vision algorithms and biological systems have three main steps; detection, selection and tracking. By looking at computer vision literature, it appears to me that computer vision models are often loosely based on feedforward hierarchies of the ventral stream in the visual cortex of primates (Figure 1.17). The ventral stream begins with the primary visual cortex (V1), goes through the secondary visual cortex (V2), then through visual area V4, and to the inferior temporal cortex (IT cortex). Neurons in lower visual areas have small receptive fields and are sensitive to basic visual features such as edges and lines. These neurons send signals to neurons at the next stage, which code for more complex features. By the V4 visual area, the neurons are selective for basic shapes, and by IT they respond in a viewpoint-invariant manner to full objects (e.g. human face). Computer vision models rely on the established features of V1 (through V1-style Gabor filtering or Haar wavelets) and then directly apply sophisticated machine learning techniques to detect what object categories are likely to be in the image. These computer vision models bypassed thinking about intermediate representations (i.e. V2 and V4) altogether in their implementations. However, the recent studies show that the remarkable robustness in biological visual systems is likely achieved gradually over different hierarchical levels (Kermani Kolankeh et al., 2015). Moreover, studies show that the visual system is not purely hierarchical (Herzog and Clarke, 2014; Ghodrati et al., 2014; Tschechne and Neumann, 2014). Feedback connections are believed to play an important role in visual processing, as they improve local activations with contextual information that is represented at higher visual areas (Tschechne and Neumann, 2014). When objects are presented on natural backgrounds reaction times in humans significantly increases compared to plain backgrounds, suggesting that some further feedback processes should happen when

objects have cluttered natural backgrounds (Ghodrati et al., 2014). These might explain why computer vision methods are still far behind the robustness of biological visual systems.

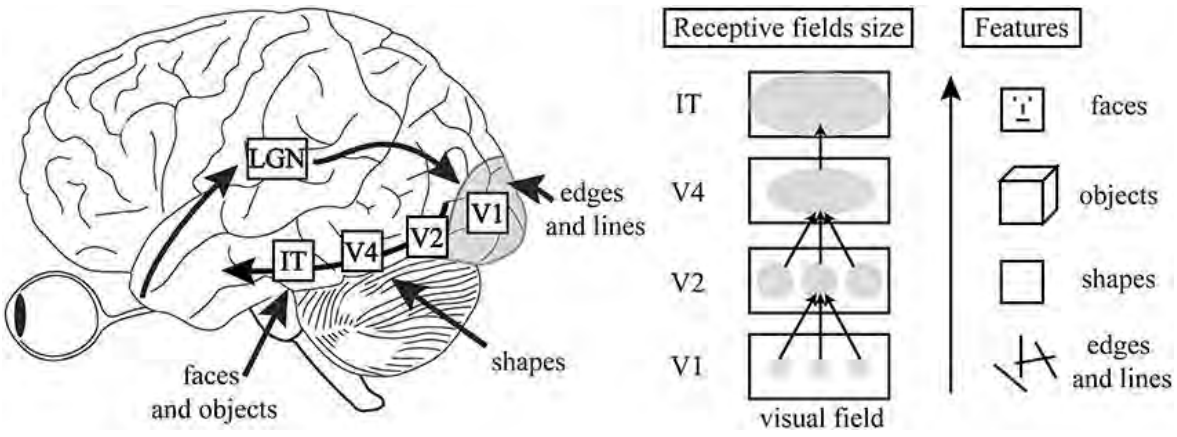

Figure 1.17. Ventral stream in primate visual cortex reproduced from Figure 1 of Herzog and Clark (2014). Visual information processing starts at the retina, proceeds to the lateral geniculate nucleus (LGN), then to the primary visual cortex (V1), the secondary visual cortex (V2), visual area V4, and inferior temporal cortex (IT cortex). The receptive field of neurons gradually increases in the higher areas, integrating information over larger and larger regions of the visual field. Low order neurons, such as V1, code for basic features such as edges and lines while higher order neurons such as V4 and IT are selective for basic shapes and full objects respectively.

Due to the complexity of recordings from primates' brains there is little known about the computational process of intermediate level neurons (V2 and V4) and the feedback processes involved in target detection and tracking. However, target tracking is not limited to primates and amazingly robust target tracking behaviour has been observed even in seemingly simple animals. For example, many species of flying insects, such as dragonflies, detect and chase prey or conspecifics within a visually cluttered surround for predation, territorial or mating behaviour. Remarkably, despite their limited neuronal architecture (Strausfeld, 1976) and low-resolution visual system (~1°), the dragonfly is capable of performing this task even in the presence of other distracting stimuli, such as swarms of prey and conspecifics (Corbet,

1999; Wiederman and O'Carroll, 2013). This behaviour requires an underlying neuronal network capable of processing algorithms for target detection (against cluttered backgrounds), selection (amidst many distracters) and interaction (varying gaze and pursuit strategies). Amazingly, the relatively simple brain of the dragonfly employs neuronal algorithms similar to the ones developed in humans and other primates (Clark et al., 2014; Mischiati et al., 2015; Wiederman and O'Carroll, 2013). Moreover, insects exhibit a remarkable behavioural plasticity as numerous species learn and memorize different sorts of sensory cues as predictors of reward (Matsumoto and Mizunami, 2000; Giurfa, 2007; Menzel, 1999; Dupuy et al., 2006) or punishment (Vergoz et al., 2007; Davis, 2005). They form memories of such experiences that can be retrieved at different times after learning, from the short-term to the long-term range (Giurfa, 2013) which is comparable in many aspects to vertebrates. For examples, bees can extract general properties of a stimulus and apply them to distinguish between other stimuli which they have never experienced before (Wu et al., 2013; Zhang et al., 2004). However, the accessibility of the insects for stable, electrophysiological recordings has made them an ideal and tractable model system for investigating the neuronal correlates for this detection, selection and interaction behaviour.

Fortunately, as a result of the recent breakthroughs in understanding insect vision we are now at a point where modelling and implementing similar strategies in an autonomous system is a practical possibility. This project therefore aims to adopt an insect-inspired approach to target tracking and pursuit, based largely on recent physiological research on the insect visual system. At the same time, re-engineering such a mechanism in robot hardware and software provides useful insights into how the neural systems might work. Therefore, the main objective of this thesis is to develop a closed-loop insect-inspired target tracking model based on 'elementary small target detection' operation and recent observations of facilitation and selective attention. Additionally, I will investigate whether

insect-inspired algorithms provide a suitable alternative for computer vision and robotic applications.

Moreover, while many studies focused on the role of motion adaptation (a reduction in sensitivity seen after the system is exposed to moving imagery) in information processing in flying insects (Maddess and Laughlin, 1985; Clifford and Langley, 1996; Harris et al., 2000), there are very few studies that test the effect of facilitation (enhancement in the response gain due to prior stimulations) in complicated information processing such as selection. Therefore, investigating the role of facilitation in target detection, selection and pursuit is another focus of my thesis.

Although the results of electrophysiological experiments and computational modelling provide insight into insect neurophysiology, our understanding of target tracking sensorimotor mechanisms is still very limited. An important question within this context is how animal saccadic movement or environmental factors change neural responses underlying the detection and selection task. To answer these questions, experiments require directly linking neural circuits and behaviour, however, during physiological recordings the insect is restrained with wax and can only experience imposed, open-loop stimuli. To model sensorimotor systems, it is necessary to accurately represent the physical interaction of the animal and the environment which is very complex to model in simulations. However, robots provide a suitable alternative to model such sensorimotor mechanisms. Therefore, another main goal of this thesis is to implement the insect-inspired target tracking mechanism on a robotic platform to test it in unstructured natural environments under demanding conditions similar to what insects experience during pursuit.

1.7 Thesis Outline

The rest of this thesis is structured as follows. In Chapter 2 and 3, I introduce an elaborated version of the ESTMD model, with inclusion of a model for the recently observed facilitation mechanism. I implement this elaborated model in a closed loop target tracking system that uses an active saccadic gaze fixation strategy inspired by insect pursuit. I test this model in virtual world simulations against heavily cluttered natural scenes using MATLAB/Simulink. Additionally, I use this model to investigate the role of some neuronal properties in closed-loop target tracking and pursuit.

In Chapter 3, I test both the efficacy and efficiency of this insect-inspired model in open-loop, using a widely-used set of videos recorded under natural conditions. I directly compare the performance of this model with several state-of-the-art engineering algorithms using the same hardware, software environment and stimuli.

Computer vision literature traditionally tests target tracking algorithms only in open-loop. However, one of the main purposes for developing these algorithms is implementation in real-time robotic applications. Therefore, it is still unclear how these algorithms might perform in closed-loop, real-world applications, where inclusion of sensors, actuators and physical robot dynamics results in additional latency which can affect the stability of the feedback process. Additionally, studies show that animals interact with the target by changing eye or body movements, which then modulate the visual inputs underlying the detection and selection task (via closed-loop feedback). This active vision system may be a key to exploiting visual information by the simple insect brain for complex tasks such as target tracking. Therefore, in Chapter 4, I implement this insect-inspired model along with insect active vision in a ground-based robotic platform (Figure 1.18). The choice of a ground robot would constrain the pursuit to a 2-dimensional environment. However, it maintains the

focus on testing the key algorithmic questions and model performance under real world challenges (e.g. lighting changes, vibration, presence of distracters, hardware limitations) rather than engineering problems associated with UAVs. I test this robotic implementation both in indoor and outdoor environments against different challenges which exist in real-world conditions including vibration, illumination variation, and distracting stimuli.

In the final chapter, the key conclusions that have been documented throughout the thesis will be presented. It must be noted that the investigations discussed in this thesis are only the beginning of the development of a robust insect-inspired model. Consequently, a significant amount of potential work still exists for future investigations which is also explained in the final chapter.

Figure 1.18. The insect-inspired target tracking model is implemented on a ground-based robot to autonomously track targets within natural environments.

1.8 Publications Arising from This Thesis

Bagheri Z. M., Cazzolato B. S., Grainger S., O'Carroll D. C., & Wiederman S. D. (In Press). An autonomous robot inspired by insect neurophysiology pursues moving features in natural environments. Journal of Neural Engineering.

Bagheri Z. M., Wiederman S. D., Cazzolato B.S., Grainger S., O'Carroll D.C. (2017). Performance of an insect-inspired target tracker in natural conditions. Bioinspiration & Biomimetics, 12(2), 025006.

Bagheri Z. M., Wiederman S. D., Cazzolato B. S., Grainger S., & O'Carroll D. C. (2015). Properties of neuronal facilitation that improve target tracking in natural pursuit simulations. Journal of the Royal Society Interface, 12(108), 20150083.

Bagheri Z.M., Wiederman S. D., Cazzolato B., Grainger S., & O'Carroll D. C. (2015). Robustness and real-time performance of an insect inspired target tracking algorithm under natural conditions. In IEEE Symposium Series on Computational Intelligence, Cape Town, South Africa97-102.

Bagheri Z., Wiederman S. D., Cazzolato B. S., Grainger S., & O'Carroll D. C. (2014). Performance assessment of an insect-inspired target tracking model in background clutter. In 13th International Conference on Control Automation Robotics & Vision, Singapore, IEEE, 822-826.

Bagheri Z. M., Wiederman S. D., Cazzolato B. S., Grainger S., & O'Carroll D. C. (2014). A biologically inspired facilitation mechanism enhances the detection and pursuit of targets of varying contrast. In 16th International Conference on Digital Image Computing: Techniques and Applications, Wollongong, Australia IEEE, 1-5.

Bagheri Z., Wiederman S.D., Cazzolato B.S., Grainger S., O'Carroll D.C. (2014). An insect inspired target tracking mechanism for autonomous vehicles. In 11th International Conference on Informatics in Control, Automation and Robotics (ICINCO), Vienna, Austria 30-38 **(Won 'the best PhD student' award).**

References

Andrews S., Tsochantaridis I., and Hofmann T. (2002). Support vector machines for multiple-instance learning. In Advances in Neural Information Processing Systems Conference, 577-584.

Avarguès-Weber, A., Portelli, G., Benard, J., Dyer, A., & Giurfa, M. (2010). Configural processing enables discrimination and categorization of face-like stimuli in honeybees. Journal of Experimental Biology, 213(4), 593-601.

Avidan S. (2007). Ensemble tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(2), 261-271.

Avidan S. (2004). Support vector tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(8), 1064-1072.

Babenko B., Yang M. H., & Belongie S. (2011). Robust object tracking with online multiple instance learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(8), 1619-1632.

Babenko B., Yang M. H., & Belongie S. (2009). Visual tracking with online multiple instance learning. In IEEE Conference on Computer Vision and Pattern Recognition, 983-990.

Banfield R. E., Hall L. O., Bowyer K. W., & Kegelmeyer W. P. (2007). A comparison of decision tree ensemble creation techniques. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(1), 173-180

Barlow H. B. (1961). Possible principles underlying the transformations of sensory messages. Sensory Communication, Ed W. Rosenblith Ch13, M.I.T. Press, Cambridge, Mass, 217-234.

Barnett P. D., Nordström K., & O'Carroll D. C. (2007). Retinotopic organization of small-field-target-detecting neurons in the insect visual system. Current Biology, 17(7), 569-578.

Bar-Shalom, Y. (1987). Tracking and data association. Academic Press Professional, Inc.

Bibby C., & Reid I. (2010). Real-time tracking of multiple occluding objects using level sets. In IEEE Conference on Computer Vision and Pattern Recognition, 1307-1314.

Bibby C., & Reid I. (2008). Robust real-time visual tracking using pixel-wise posteriors. In European Conference on Computer Vision, 831-844.

Blanchard M., Rind F. C., & Verschure P. F. (2000). Collision avoidance using a model of the locust LGMD neuron. Robotics and Autonomous Systems, 30(1), 17-38.

Blum A., & Mitchell T. (1998). Combining labeled and unlabeled data with co-training. In Proceedings of the 11th Annual Conference on Computational Learning Theory, 92-100.

Bolzon D. M., Nordström K., & O'Carroll D. C. (2009). Local and large-range inhibition in feature detection. The Journal of Neuroscience, 29(45), 14143-14150.

Borst A., & Egelhaaf M. (1989). Principles of visual motion detection. Trends in Neurosciences, 12(8), 297-306.

Borst A., & Haag J. (2002). Neural networks in the cockpit of the fly. Journal of Comparative Physiology A, 188(6), 419-437.

Braddick O. J., Ruddock K. H., Morgan M. J., & Marr D. (1980). Low-level and high-level processes in apparent motion [and discussion]. Philosophical Transactions of the Royal Society B: Biological Sciences, 290(1038), 137-151.

Brinkworth R. S., Mah E. L., Gray J. P., & O'Carroll D. C. (2008). Photoreceptor processing improves salience facilitating small target detection in cluttered scenes. Journal of Vision, 8(11), 1-17.

Buehler P., Everingham M., Huttenlocher D. P., & Zisserman A. (2008). Long term arm and hand tracking for continuous sign language TV broadcasts. In Proceedings of the 19th British Machine Vision Conference, 1105-1114.

Calonder M., Lepetit V., & Fua P. (2008). Keypoint signatures for fast learning and recognition. In European Conference on Computer Vision, 58-71.

Chen Q., Georganas N. D., & Petriu E. M. (2008). Hand gesture recognition using Haar-like features and a stochastic context-free grammar. IEEE Transactions on Instrumentation and Measurement, 57(8), 1562-1571.

Chittka, L., Vorobyev, M., Shmida, A. and Menzel, R. (1993). Bee colour vision – the optimal system for discrimination of flower colour with three spectral photoreceptor types. In Sensory Systems of Arthropods (ed. K. Wiese), pp. 211-218. Basel, Switzerland: Birkhaeuser Verlag Press.

Clifford C. W. G., & Langley K. (1996). A model of temporal adaptation in fly motion vision. Vision Research, 36(16), 2595-2608.

Collett T. S., & Land M. F. (1978). How hoverflies compute interception courses. Journal of Comparative Physiology, 125(3), 191-204.

Collett T. S., & Land M. F. (1975). Visual control of flight behavior in the hoverfly Syritta pipiens L. Journal of Comparative Physiology, 99(1), 1-66.

Collins R. T., Liu Y., & Leordeanu M. (2005). Online selection of discriminative tracking features. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(10), 1631-1643.

Comaniciu D., Ramesh V., & Meer P. (2003). Kernel-based object tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(5), 564-577.

Conroy J., Gremillion G., Ranganathan B., & Humbert J. S. (2009). Implementation of wide-field integration of optic flow for autonomous quadrotor navigation. Autonomous robots, 27(3), 189-198.

Coombs D., & Roberts K. (1992). 'Bee-bot': using peripheral optical flow to avoid obstacles. In Applications in Optical Science and Engineering, 714-721.

Corbet, P. S. (1999). Dragonflies: behaviour and ecology of Odonata. Harley Books.

Couvreur C. (1997). The EM algorithm: A guided tour. In Computer Intensive Methods in Control and Signal Processing, Birkhäuser Boston, 209-222.

Clark D. A., Fitzgerald J. E., Ales J. M., Gohl D. M., Silies M. A., Norcia A. M., & Clandinin T. R. (2014). Flies and humans share a motion estimation strategy that exploits natural scene statistics. Nature Neuroscience, 17(2), 296-303.

Dalal N. (2006). Finding people in images and videos. Doctoral dissertation, Institut National Polytechnique de Grenoble-INPG.

Dalal N., & Triggs B. (2005). Histograms of oriented gradients for human detection. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1, 886-893.

Davenport R. K., & Rogers C. M. (1971). Perception of photographs by apes. Behaviour, 39, 318-320.

Davis R. L. (2005). Olfactory memory formation in Drosophila: from molecular to systems neuroscience. Annual Reviews of Neuroscience, 28, 275-302.

Dempster A. P., Laird N. M., & Rubin D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (methodological), Ser.B39 (1), 1–38.

Dollár P., Babenko B., Belongie S., Perona P., & Tu Z. (2008). Multiple component learning for object detection. In European Conference on Computer Vision, 211-224.

Dowson N. D., & Bowden R. (2005). Simultaneous modeling and tracking (SMAT) of feature sets. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2, 99-105.

Driver J. (2001). A selective review of selective attention research from the past century. British Journal of Psychology, 92(1), 53-78.

Dunbier J. R., Wiederman S. D., Shoemaker P. A. & O'Carroll D. C. (2012). Facilitation of dragonfly target-detecting neurons by slow moving features on continuous paths. Frontiers in Neural Circuits, 6, 79.

Dunbier J. R., Wiederman S. D., Shoemaker P. A., & O'Carroll, D. C. (2011). Modelling the temporal response properties of an insect small target motion detector. In Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing, 125-130.

Dupuy, F., Sandoz, J. C., Giurfa, M., & Josens, R. (2006). Individual olfactory learning in Camponotus ants. Animal Behaviour, 72(5), 1081-1091.

Dyer A. G. (2012). The mysterious cognitive abilities of bees: why models of visual processing need to consider experience and individual differences in animal performance. Journal of Experimental Biology, 215(3), 387-395.

Dyer A. G., & Vuong Q. C. (2008). Insect brains use image interpolation mechanisms to recognise rotated objects. PLoS One, 3(12), e4086.

Dyer A. G., Rosa M. G., & Reser D. H. (2008). Honeybees can recognise images of complex natural scenes for use as potential landmarks. Journal of Experimental Biology, 211(8), 1180-1186.

Dyer A. G., Neumeyer C., & Chittka L. (2005). Honeybee (Apis mellifera) vision can discriminate between and recognise images of human faces. Journal of Experimental Biology, 208(24), 4709-4714.

Egelhaaf M. (1985). On the neuronal basis of figure-ground discrimination by relative motion in the visual system of the fly. Biological Cybernetics, 52(2), 123-140.

Fergus R., Perona P., & Zisserman A. (2003). Object class recognition by unsupervised scale-invariant learning. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2, 264-271.

Flanagan J. R., King S., Wolpert D. M., & Johansson R. S. (2001). Sensorimotor prediction and memory in object manipulation. Canadian Journal of Experimental Psychology, 55(2), 87.

Franceschini N., Pichon J. M., Blanes C., & Brady J. M. (1992). From insect vision to robot vision [and discussion]. Philosophical Transactions of the Royal Society B: Biological Sciences, 337(1281), 283-294.

Franklin D. W., & Wolpert D. M. (2011). Computational mechanisms of sensorimotor control. Neuron, 72(3), 425-442.

Freedman D. J., Riesenhuber M., Poggio T., & Miller E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. Science, 291(5502), 312-316.

Freund Y. (1995). Boosting a weak learning algorithm by majority. Information and Computation, 2(121), 256-285.

Freund Y. & R E Schapire R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1), 119–139.

Frye M. (2015). Elementary motion detectors. Current Biology, 25(6), R215.

Gabbiani F., Krapp H. G., Koch C., & Laurent G. (2002). Multiplicative computation in a visual neuron sensitive to looming. Nature, 420(6913), 320-324.

Galleguillos C., Babenko B., Rabinovich A., & Belongie S. (2008). Weakly supervised object localization with stable segmentations. In European Conference on Computer Vision, 193-207.

Gerónimo D., López A., Ponsa D., & Sappa, A. D. (2007 a). Haar wavelets and edge orientation histograms for on–board pedestrian detection. In Iberian Conference on Pattern Recognition and Image Analysis, 418-425.

Gerónimo D., Sappa A., López A., & Ponsa D. (2007b). Adaptive image sampling and windows classification for on-board pedestrian detection. In Proceedings of the International Conference on Computer Vision Systems, 39.

Geurten B. R., Nordström K., Sprayberry J. D., Bolzon D. M., & O'Carroll D. C. (2007). Neural mechanisms underlying target detection in a dragonfly centrifugal neuron. Journal of Experimental Biology, 210(18), 3277-3284.

Ghodrati M., Farzmahdi A., Rajaei K., Ebrahimpour R., & Khaligh-Razavi S. M. (2014). Feedforward object-vision models only tolerate small image variations compared to human. Frontiers in computational neuroscience, 8, 74.

Giurfa M. (2013). Cognition with few neurons: higher-order learning in insects. Trends in neurosciences, 36(5), 285-294.

Giurfa, M. (2007). Behavioral and neural analysis of associative learning in the honeybee: a taste from the magic well. Journal of Comparative Physiology A, 193(8), 801-824.

Giurfa M., & Menzel R. (1997). Insect visual perception: complex abilities of simple nervous systems. Current Opinion in Neurobiology, 7(4), 505-513.

Giurfa M., Schubert M., Reisenman C., Gerber B., & Lachnit H. (2003). The effect of cumulative experience on the use of elemental and configural visual discrimination strategies in honeybees. Behavioural Brain Research, 145(1), 161-169.

Giurfa, M., Eichmann, B., & Menzel, R. (1996). Symmetry perception in an insect. Nature, 382, 485-461.

Grabner H., & Bischof H. (2006). On-line boosting and vision. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1, 260-267.

Greenspan R. J., & Ferveur J. F. (2000). Courtship in drosophila. Annual Review of Genetics, 34(1), 205-232.

Hadid A., Pietikainen M., & Ahonen T. (2004). A discriminative feature space for detecting and recognizing faces. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2, 797-804.

Halupka K. J., Wiederman S. D., Cazzolato B. S., & O'Carroll D. C. (2013). Bio-inspired feature extraction and enhancement of targets moving against visual clutter during closed loop pursuit. In IEEE International Conference on Image Processing, 4098-4102.

Halupka K. J., Wiederman S. D., Cazzolato B. S., & O'Carroll D. C. (2011). Discrete implementation of biologically inspired image processing for target detection. In IEEE Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 143-148.

Hardie R. C. (2001). Phototransduction in Drosophila melanogaster. Journal of Experimental Biology, 204(20), 3403-3409.

Hardie R. C. (1985). Functional organization of the fly retina. In Progress in Sensory Physiology, 1-79.

Hare S., Saffari A., & Torr P. H. (2011). Struck: Structured output tracking with kernels. In International Conference on Computer Vision, 263-270.

Harris R. A., O'Carroll D. C., & Laughlin S. B. (2000). Contrast gain reduction in fly motion adaptation. Neuron, 28(2), 595-606.

Harrison R. R. (2000). An analog VLSI motion sensor based on the fly visual system (Doctoral dissertation, California Institute of Technology).

Hassenstein V. B., & Reichardt W. (1956). System theoretical analysis of time, sequence and sign analysis of the motion perception of the snout-beetle Chlorophanus. Z Naturforsch B, 11, 513-524.

Hatsopoulos N., Gabbiani F., & Laurent G. (1995). Elementary computation of object approach by a wide-field visual neuron. Science, 270(5238), 1000.

Hausen K. (1982). Motion sensitive interneurons in the optomotor system of the fly. Biological Cybernetics, 45(2), 143-156.

Henriques J. F., Caseiro R., Martins P., & Batista J. (2015). High-speed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(3), 583-596.

Herrnstein R. J. (1984). Objects, categories and discrimination stimuli. In Animal Cognition (ed. H. L. Roitblat, T. G. Bever, and H. S. Terrace), pp. 233-261. Hillsdale, NJ, USA: Erlbaum

Herzog M. H., & Clarke A. M. (2014). Why vision is not both hierarchical and feedforward. Frontiers in Computational Neuroscience, 8, 135.

Hinterstoisser S., Lepetit V., Ilic S., Fua P., & Navab N. (2010). Dominant orientation templates for real-time detection of texture-less objects. In IEEE Conference on Computer Vision and Pattern Recognition,10, 2257-2264.

Hinterstoisser S., Kutter O., Navab N., Fua P., & Lepetit V. (2009). Real-time learning of accurate patch rectification. In IEEE Conference on Computer Vision and Pattern Recognition, 2945-2952.

Ho J., Lee K. C., Yang M. H., & Kriegman D. (2004). Visual tracking using learned linear subspaces. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 782-789.

Horridge A. (2009). Generalization in visual recognition by the honeybee (Apis mellifera): A review and explanation. Journal of Insect Physiology, 55(6), 499-511.

Horridge A. (2007). The preferences of the honeybee (Apis mellifera) for different visual cues during the learning process. Journal of insect physiology, 53(9), 877-889.

Horridge A. (2005). What the honeybee sees: a review of the recognition system of Apis mellifera. Physiological Entomology, 30(1), 2-13.

Horridge A. (2000). Seven experiments on pattern vision of the honeybee, with a model. Vision Research, 40(19), 2589-2603.

Horridge G. A. (1996). The honeybee (Apis mellifera) detects bilateral symmetry and discriminates its axis. Journal of Insect Physiology, 42(8), 755-764.

Horridge G. A. (1978). The separation of visual axes in apposition compound eyes. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 285(1003), 1-59.

Horridge G. A., & Zhang S. W. (1995). Pattern vision in honeybees (Apis mellifera): flower-like patterns with no predominant orientation. Journal of Insect Physiology, 41(8), 681-688.

Hu W., Li, X., Zhang X., Shi X., Maybank S., & Zhang Z. (2011). Incremental tensor subspace learning and its applications to foreground segmentation and tracking. International Journal of Computer Vision, 91(3), 303-327.

Huber, L., roje, N. F., Loidolt, M., Aust, U., & Grass, D. (2000). Natural categorization through multiple feature learning in pigeons. The Quarterly Journal of Experimental Psychology: Section B, 53(4), 341-357.

Isard M., & Blake A. (1996). Contour tracking by stochastic propagation of conditional density. In European Conference on Computer Vision, 343-356.

Itti L., & Koch C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research, 40(10), 1489-1506.

Jansonius N. M., & Van Hateren J. H. (1991). Fast temporal adaptation of on-off units in the first optic chiasm of the blowfly. Journal of Comparative Physiology A, 168(6), 631-637.

Järvilehto M., & Zettler F. (1971). Localized intracellular potentials from pre-and postsynaptic components in the external plexiform layer of an insect retina. Zeitschrift Für Vergleichende Physiologie, 75(4), 422-440.

Javed O., Ali S., & Shah M. (2005). Online detection and classification of moving objects using progressively improving detectors. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 696-701.

Jepson A. D., Fleet D. J., & El-Maraghi T. F. (2003). Robust online appearance models for visual tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(10), 1296-1311.

Julier S. J., Uhlmann J. K., & Durrant-Whyte H. F. (1995). A new approach for filtering nonlinear systems. In Proceedings of the American Control Conference, 3 ,1628-1632.

Juusola M., Weckstrom M., Uusitalo R. O., Korenberg M. J., & French A. S. (1995). Nonlinear models of the first synapse in the light-adapted fly retina. Journal of Neurophysiology, 74(6), 2538-2547.

Juusola M., Kouvalainen E., Järvilehto M., & Weckström M. (1994). Contrast gain, signal-to-noise ratio, and linearity in light-adapted blowfly photoreceptors. The Journal of General Physiology, 104(3), 593-621.

Kalal Z., Mikolajczyk K., & Matas J. (2012). Tracking-learning-detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(7), 1409-1422.

Kalman R. E. (1960). A new approach to linear filtering and prediction problems. Journal of Basic Engineering, 82(1), 35-45.

Kawato M. (1999). Internal models for motor control and trajectory planning. Current Opinion in Neurobiology, 9(6), 718-727.

Kendoul F. (2014). Four-dimensional guidance and control of movement using time-to-contact: Application to automated docking and landing of unmanned rotorcraft systems. The International Journal of Robotics Research, 33(2), 237-267.

Kermani Kolankeh A., Teichmann M., & Hamker F. H. (2015). Competition improves robustness against loss of information. Frontiers in Computational Neuroscience, 9, 35.

Kirschfeld K., & Wenk P. (1976). The dorsal compound eye of simuliid flies. Zeitschrift für Naturforschung C, 31(11-12), 764-765.

Klein G., & Murray D. (2007). Parallel tracking and mapping for small AR workspaces. In 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, 225-234.

Krapp H. G., & Hengstenberg R. (1996). Estimation of self-motion by optic flow processing in single visual interneurons. Nature, 384(6608), 463-466.

Land M. F. (1997). Visual acuity in insects. Annual Review of Entomology,42(1), 147-177.

Land M. F. (1989). Variations in the structure and design of compound eyes. In Facets of vision, Springer Berlin Heidelberg, 90-111.

Land M. F. (1981). Optics and vision in invertebrates. Handbook of Sensory Physiology.

Land M. F. & Eckert H. (1985). Maps of the acute zones of fly eyes. Journal of Comparative Physiology A, 156(4), 525-538.

Laughlin S. (1984). The roles of parallel channels in early visual processing by the arthropod compound eye. In Photoreception and Vision in Invertebrates, Springer US, 457-481.

Laughlin S. B., Howard, J., & Blakeslee, B. (1987). Synaptic limitations to contrast coding in the retina of the blowfly Calliphora. Proceedings of the Royal Society of London B: Biological Sciences, 231(1265), 437-467.

Lehrer M., Horridge G. A., Zhang S. W., & Gadagkar R. (1995). Shape vision in bees: innate preference for flower-like patterns. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 347(1320), 123-137.

Leng J., & Wang H. (2004). Tracking as recognition: A stable 3D tracking framework. In 8th IEEE Conference on Control, Automation, Robotics and Vision, 3, 2303-2307.

Lepetit V., Lagger P., & Fua P. (2005). Randomized trees for real-time keypoint recognition. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2, 775-781.

Levi K., & Weiss Y. (2004). Learning object detection from a small number of examples: the importance of good features. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2, 53–60.

Levin A., Viola P., & Freund Y. (2003). Unsupervised improvement of visual detectors using co-training. In Proceedings 9th IEEE International Conference on Computer Vision, 626-633.

Li Y. (2004). On incremental and robust subspace learning. Pattern Recognition, 37(7), 1509-1518.

Li X., Hu W., Zhang Z., Zhang X., Zhu M., & Cheng J. (2008). Visual tracking via incremental log-Euclidean Riemannian subspace learning. In IEEE Conference on Computer Vision and Pattern,1-8.

Li X., Hu W., Zhang Z., Zhang X., & Luo G. (2007). Robust visual tracking based on incremental tensor subspace learning. In IEEE 11th International Conference on Computer Vision, 1-8.

Lim J., Ross D. A., Lin R. S., & Yang M. H. (2004). Incremental learning for visual tracking. In Advances in Neural Information Processing Systems, 793-800.

Limb J. O., & Murphy J. A. (1975). Estimating the velocity of moving images in television signals. Computer Graphics and Image Processing, 4(4), 311-327.

Lowe D. G. (2004). Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2), 91-110.

Liu S. C. (2000). A neuromorphic aVLSI model of global motion processing in the fly. IEEE Transactions on circuits and systems II: analog and digital signal processing, 47(12), 1458-1467.

Maddess T., & Laughlin S. B. (1985). Adaptation of the motion-sensitive neuron H1 is generated locally and governed by contrast frequency. Proceedings of the Royal Society of London B: Biological Sciences, 225(1239), 251-275.

Maji S., Berg A. C., & Malik J. (2008). Classification using intersection kernel support vector machines is efficient. In IEEE Conference on Computer Vision and Pattern Recognition, 8 pages.

Malott R. W., & Siddall J. W. (1972). Acquisition of the people concept in pigeons. Psychological Reports, 31(1), 3-13.

Marr D., & Ullman S. (1981). Directional selectivity and its use in early visual processing. Proceedings of the Royal Society of London B: Biological Sciences, 211(1183), 151-180.

Matsumoto Y., & Mizunami M. (2000). Olfactory learning in the cricket Gryllus bimaculatus. Journal of Experimental Biology, 203(17), 2581-2588.

Matthews I., Ishikawa T., & Baker S. (2004). The template update problem. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(6), 810-815.

Mehta B., & Schaal S. (2002). Forward models in visuomotor control. Journal of Neurophysiology, 88(2), 942-953.

Menzel, R. (1999). Memory dynamics in the honeybee. Journal of Comparative Physiology A, 185(4), 323-340.

Mischiati M., Lin H. T., Herold P., Imler E., Olberg R., & Leonardo A. (2015). Internal models direct dragonfly interception steering. Nature, 517(7534), 333-338.

Murase H., & Nayar S. K. (1995). Visual learning and recognition of 3-D objects from appearance. International Journal of Computer Vision, 14(1), 5-24.

Nigam K., & Ghani R. (2000). Analyzing the effectiveness and applicability of co-training. In Proceedings of the 9th International Conference on Information and Knowledge Management, 86-93.

Nigam K., McCallum A. K., Thrun S., & Mitchell T. (2000). Text classification from labeled and unlabeled documents using EM. Machine Learning, 39(2-3), 103-134.

Nilsson D. E. (1989). Optics and evolution of the compound eye. In Facets of vision, Springer Berlin Heidelberg, 30-73.

Nordström K., & O'Carroll D. C. (2009). Feature detection and the hypercomplex property in insects. Trends in Neurosciences, 32(7), 383-391.

Nordström K., Bolzon D. M., & O'Carroll D. C. (2011). Spatial facilitation by a high performance dragonfly target-detecting neuron. Biology Letters, 7(4), 588-592.

Nordström K., Barnett P. D., & O'Carroll D. C. (2006). Insect detection of small targets moving in visual clutter. PLOS Biology, 4(3), 378-386.

Obdrzalek S., & Matas J. (2005). Sub-linear indexing for large scale object recognition. In British Machine Vision Conference, 10 pages.

O'Carroll D. C. (1993). Feature-detecting neurons in dragonflies. Nature, 362, 541-543.

O'Carroll D. C., & Wiederman S. D. (2014). Contrast sensitivity and the detection of moving patterns and features. Philosophical Transactions of Royal Society B, 369(1636), 20130043.

O'Carroll D. C., Osorio D., James A. C., & Bush T. (1992). Local feedback mediated via amacrine cells in the insect optic lobe. Journal of Comparative Physiology A, 171(4), 447-455.

Ojala T., Pietikainen M., & Maenpaa T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7), 971-987.

Olberg R. M., Worthington A. H., & Venator K. R. (2000). Prey pursuit and interception in dragonflies. Journal of Comparative Physiology A, 186(2), 155-162.

Opelt A., Pinz A., Fussenegger M., & Auer P. (2006). Generic object recognition with boosting. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(3), 416-431.

Osorio D. (1991). Mechanisms of early visual processing in the medulla of the locust optic lobe: How self-inhibition, spatial-pooling, and signal rectification contribute to the properties of transient cells. Visual Neuroscience, 7(4), 345-355.

Papageorgiou C. P., Oren M., & Poggio T. (1998). A general framework for object detection. In 6th International Conference on Computer Vision, 555-562.

Phung S. & Bouzerdoum A. (2007). A new image feature for fast detection of people in images. International Journal of Information and Systems Sciences, 3(3), 383–391.

Pilet J., & Saito H. (2010). Virtually augmenting hundreds of real pictures: An approach based on learning, retrieval, and tracking. In IEEE Virtual Reality Conference, 71-78.

Porikli F., Tuzel O., & Meer P. (2006). Covariance tracking using model update based on Lie algebra. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 728-735.

Rahimi A., Morency L. P., & Darrell T. (2008). Reducing drift in differential tracking. Computer Vision and Image Understanding, 109(2), 97-111.

Ramanan D., Forsyth D. A., & Zisserman A. (2007). Tracking people by learning their appearance. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(1), 65-81.

Reddy B. S., & Chatterji B. N. (1996). An FFT-based technique for translation, rotation, and scale-invariant image registration. IEEE Transactions on Image Processing, 5(8), 1266-1271.

Reichardt W (1957) Autokorrelations-Auswertung als Funktionsprinzip des Zentralnervensystems. Z Naturforsch 12b, 448-457.

Ren Y., Chua C. S., & Ho Y. K. (2003). Motion detection with nonstationary background. Machine Vision and Applications, 13(5-6), 332-343.

Rind F. C., & Simmons P. J. (1992). Orthopteran DCMD neuron: A reevaluation of responses to moving objects. I. Selective responses to approaching objects. Journal of Neurophysiology, 68(5), 1654-1666.

Roitblat H. L. (1987). Language, concept and intelligence. In Introduction to Comparative Cognition, pp. 275-317. New York: W. H. Freeman and Company.

Ross D. A., Lim J., Lin R. S., & Yang M. H. (2008). Incremental learning for robust visual tracking. International Journal of Computer Vision, 77(1-3), 125-141.

Ruffier F., & Franceschini N. (2003). OCTAVE: a bioinspired visuo-motor control system for the guidance of micro-air-vehicles. In Microtechnologies for the New Millennium, 1-12.

Sabzmeydani P., & Mori G. (2007). Detecting pedestrians by learning shapelet features. In IEEE Conference on Computer Vision and Pattern Recognition,8 pages.

Santer R. D., Rind F. C., & Simmons P. J. (2012). Predator versus prey: Locust looming-detector neuron and behavioural responses to stimuli representing attacking bird predators. PloS One, 7(11), e50146.

Santos-Victor J., Sandini G., Curotto F., & Garibaldi S. (1995). Divergent stereo in autonomous navigation: From bees to robots. International Journal of Computer Vision, 14(2), 159-177.

Schapire R. E. (1990). The strength of weak learnability. Machine learning, 5(2), 197-227.

Schweitzer H., Bell J. W., & Wu F. (2002). Very fast template matching. In European Conference on Computer Vision, 358-372.

Shafique K., & Shah M. (2005). A noniterative greedy algorithm for multiframe point correspondence. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(1), 51-65.

Shaw S. R. (1984). Early visual processing in insects. Journal of Experimental Biology, 112(1), 225-251.

Sherk T. E. (1978). Development of the compound eyes of dragonflies (Odonata). III. Adult compound eyes. Journal of Experimental Zoology, 203(1), 61-79.

Sillito R. R., & Fisher R. B. (2008). Semi-supervised learning for anomalous trajectory detection. In British Machine Vision Conference, 1, 1035-1067.

Silveira G., & Malis E. (2007). Real-time visual tracking under arbitrary illumination changes. In IEEE Conference on Computer Vision and Pattern Recognition, 1-6.

Simmons P. J., & Rind F. C. (1997). Responses to object approach by a wide field visual neurone, the LGMD2 of the locust: Characterization and image cues. Journal of Comparative Physiology A, 180(3), 203-214.

Singer R. A., & Stein J. J. (1971). An optimal tracking filter for processing sensor data of imprecisely determined origin in surveillance systems. In IEEE Conference on Decision and Control, 10, 171-175.

Sittler R. W. (1964). An optimal data association problem in surveillance theory. IEEE Transactions on Military Electronics, 8(2), 125-139.

Smeulders A. W., Chu D. M., Cucchiara R., Calderara S., Dehghan A., & Shah, M. (2014). Visual tracking: An experimental survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(7), 1442-1468.

Snyder A. W., Laughlin S. B., & Stavenga D. G. (1977). Information capacity of eyes. Vision Research, 17(10), 1163-1175.

Sukontason K. L., Chaiwong T., Piangjai S., Upakut S., Moophayak K., & Sukontason K. (2008). Ommatidia of blow fly, house fly, and flesh fly: Implication of their vision efficiency. Parasitology Research, 103(1), 123-131.

Srinivasan M. V. (2011). Honeybees as a model for the study of visually guided flight, navigation, and biologically inspired robotics. Physiological Reviews, 91(2), 413-460.

Srinivasan, M. V. (1990). Generalized gradient schemes for the measurement of two-dimensional image motion. Biological Cybernetics, 63(6), 421-431.

Srinivasan M. V. (1998). Insects as Gibsonian animals. Ecological Psychology, 10(3-4), 251-270.

Srinivasan M. V., Zhang S., & Chahl J. S. (2001). Landing strategies in honeybees, and possible applications to autonomous airborne vehicles. The Biological Bulletin, 200(2), 216-221.

Srinivasan M. V., Zhang S. W., & Chandrashekara K. (1993). Evidence for two distinct movement-detecting mechanisms in insect vision. Naturwissenschaften, 80(1), 38-41.

Srinivasan M. V., Pinter R. B., & Osorio D. (1990). Matched filtering in the visual system of the fly: Large monopolar cells of the lamina are optimized to detect moving edges and blobs. Proceedings of the Royal Society of London B: Biological Sciences, 240(1298), 279-293.

Srinivasan M. V., Laughlin S. B., & Dubs A. (1982). Predictive coding: A fresh view of inhibition in the retina. Proceedings of the Royal Society of London B: Biological Sciences, 216(1205), 427-459.

Stach S., & Giurfa M. (2005). The influence of training length on generalization of visual feature assemblies in honeybees. Behavioural Brain Research, 161(1), 8-17.

Stach, S., Benard, J., & Giurfa, M. (2004). Local-feature assembling in visual pattern recognition and generalization in honeybees. Nature, 429(6993), 758-761.

Strausfeld N. J. (1989). Beneath the compound eye: Neuroanatomical analysis and physiological correlates in the study of insect vision. In Facets of vision, Springer Berlin Heidelberg, 317-359.

Strausfeld N. J. (1976). Some quantitative aspects of the fly's brain. Atlas of an Insect Brain, Heidelberg, Germany: Springer-Verlag, 49-56.

Strausfeld N. J., & Nassel D. R. (1981). Neuroarchitectures serving compound eyes of Crustacea and insects. Handbook of Sensory Physiology, Springer-Verlag, 1–132.

Strydom R., Singh S. P., & Srinivasan M. V. (2015). Biologically inspired interception: A comparison of pursuit and constant bearing strategies in the presence of sensorimotor delay. In IEEE International Conference on Robotics and Biomimetics (ROBIO), 2442-2448.

Takacs G., Chandrasekhar V., Tsai S., Chen D., Grzeszczuk R., & Girod B. (2010). Unified real-time tracking and recognition with rotation-invariant fast features. In IEEE Conference on Computer Vision and Pattern Recognition, 934-941.

Tang F., Brennan S., Zhao Q., & Tao H. (2007). Co-tracking using semi-supervised support vector machines. In IEEE 11th International Conference on Computer Vision, 8 pages.

Taylor S., & Drummond T. (2009,). Multiple Target Localisation at over 100 FPS. In British Machine Vision Conference, 11 pages.

Thurrowgood, S., Moore, R. J., Soccol, D., Knight, M., & Srinivasan, M. V. (2014). A Biologically Inspired, Vision-based Guidance System for Automatic Landing of a Fixed-wing Aircraft. Journal of Field Robotics, 31(4), 699-727.

Treisman A. M. (1969). Strategies and models of selective attention. Psychological Review, 76(3), 282.

Treisman A. M., & Gelade G. (1980). A feature-integration theory of attention. Cognitive Psychology, 12(1), 97-136.

Tschechne, S., & Neumann, H. (2014). Hierarchical representation of shapes in visual cortex—from localized features to figural shape segregation. Frontiers in computational neuroscience, 8, 93.

Uhlmann S., Kiranyaz S., & Gabbouj M. (2014). Semi-supervised learning for ill-posed polarimetric SAR classification. Remote Sensing, 6(6), 4801-4830.

Ullman S (1983). The measurement of visual motion- computational considerations and some neurophysiological implications. Trends in Neuroscience 6,177-179.

Vacchetti L., Lepetit V., & Fua P. (2004). Stable real-time 3D tracking using online and offline information. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(10), 1385-1391.

Van Hateren J. H. (1997). Processing of natural time series of intensities by the visual system of the blowfly. Vision Research, 37(23), 3407-3416.

Van Hateren J. H., Srinivasan M. V., & Wait P. B. (1990). Pattern recognition in bees: orientation discrimination. Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 167(5), 649-654.

Veenman C. J., Reinders M. J., & Backer E. (2001). Resolving motion correspondence for densely moving points. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(1), 54-72.

Vergoz V., Roussel E., Sandoz J. C., & Giurfa M. (2007). Aversive learning in honeybees revealed by the olfactory conditioning of the sting extension reflex. PLoS One, 2(3), e288.

Vijayanarasimhan S., & Grauman K. (2008). Keywords to visual categories: Multiple-instance learning forweakly supervised object categorization. In IEEE Conference on Computer Vision and Pattern Recognition, 8 pages.

Viola P., Jones M. J., & Snow D. (2005). Detecting pedestrians using patterns of motion and appearance. International Journal of Computer Vision, 63(2), 153-161.

Vogels R. (1999). Categorization of complex visual images by rhesus monkeys. Part 1: behavioural study. European Journal of Neuroscience, 11(4), 1223-1238.

Wang S. (2011). A review of gradient-based and edge-based feature extraction methods for object detection. In IEEE 11th International Conference on Computer and Information Technology, 277-282.

Wang X., Hua G., & Han T. X. (2010). Discriminative tracking by metric learning. In European Conference on Computer Vision, 200-214.

Wang X., Han T. X., & Yan S. (2009). An HOG-LBP human detector with partial occlusion handling. In IEEE 12th International Conference on Computer Vision, 32-39.

Wang W., Yang J., & Gao W. (2008). Modeling background and segmenting moving objects from compressed video. IEEE Transactions on Circuits and Systems for Video Technology, 18(5), 670-681.

Wang L., Hu W., & Tan T. (2003). Recent developments in human motion analysis. Pattern Recognition, 36(3), 585-601.

Watanabe T., Ito S., & Yokoi K. (2009). Co-occurrence histograms of oriented gradients for pedestrian detection. In Pacific-Rim Symposium on Image and Video Technology, 37-47.

Wax N. (1955). Signal-to-noise improvement and the statistics of track populations. Journal of Applied Physics, 26(5), 586-595.

Wehner R. (1971). The generalization of directional visual stimuli in the honey bee, Apis mellifera. Journal of Insect Physiology, 17(8), 1579-1591.

Wehrhahn C. (1979). Sex-specific differences in the chasing behaviour of houseflies (Musca). Biological Cybernetics, 32(4), 239-241.

Wen J., Gao X., Li X., & Tao D. (2009). Incremental learning of weighted tensor subspace for visual tracking. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, 3788–3793.

Wiederman S. D., & O'Carroll D. C. (2013). Selective attention in an insect visual neuron. Current Biology, 23(2), 156-161.

Wiederman S. D., Shoemaker P. A., & O'Carroll D. C. (2013). Correlation between OFF and ON channels underlies dark target selectivity in an insect visual system. The Journal of Neuroscience, 33(32), 13225-13232.

Wiederman S. D., Shoemaker P. A., & O'Carroll D. C. (2008). A model for the detection of moving targets in visual clutter inspired by insect physiology. PloS One, 3(7), e2784.

Wiederman S., Shoemaker P. A., & O'Carroll D. C. (2007). Biologically inspired small target detection mechanisms. In 3rd IEEE International Conference on Intelligent Sensors, Sensor Networks and Information, 269-273.

Wolpert D. M., Ghahramani Z., & Jordan M. I. (1995). An internal model for sensorimotor integration. Science, 269(5232), 1880-1882.

Wu Y., Lim J., & Yang M. H. (2015). Object tracking benchmark. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(9), 1834-1848.

Wu, W., Moreno, A. M., Tangen, J. M., & Reinhard, J. (2013). Honeybees can discriminate between Monet and Picasso paintings. Journal of Comparative Physiology A, 199, 45-55.

Xie P., Ma O., & Zhang Z. (2013). A bio-inspired approach for UAV landing and perching. AIAA Guidance, Navigation, and Control (GNC) Conference, Guidance, Navigation, and Control and Co-located Conferences, (AIAA 2013-5108).

Yarfitz S., & Hurley J. B. (1994). Transduction mechanisms of vertebrate and invertebrate photoreceptors. Journal of Biological Chemistry, 269, 14329.

Yilmaz A., Javed O., & Shah M. (2006). Object tracking: A survey. ACM Computing Surveys (CSUR), 38(4), 1-45.

Yilmaz A., Li X., & Shah M. (2004). Contour-based object tracking with occlusion handling in video acquired using mobile cameras. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(11), 1531-1536.

Yu Q., Dinh T. B., & Medioni G. (2008). Online tracking and reacquisition using co-trained generative and discriminative trackers. In European Conference on Computer Vision, 678-691.

Yuan Y., Pang Y., & Li X. (2010). Footwear for gender recognition. IEEE Transactions on Circuits and Systems for Video Technology, 20(1), 131-135.

Zago M., Bosco G., Maffei V., Iosa M., Ivanenko Y. P., & Lacquaniti F. (2004). Internal models of target motion: expected dynamics overrides measured kinematics in timing manual interceptions. Journal of Neurophysiology, 91(4), 1620-1634.

Zeil J. (1983). Sexual dimorphism in the visual system of flies: the compound eyes and neural superposition in Bibionidae (Diptera). Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 150(3), 379-393.

Zeisl B., Leistner C., Saffari A., & Bischof H. (2010). On-line semi-supervised multiple-instance boosting. In IEEE Conference on Computer Vision and Pattern Recognition,1879-1879.

Zhang Z., Zhang S., Xie P., & Ma O. (2014). Bioinspired 4D trajectory generation for a UAS rapid point-to-point movement. Journal of Bionic Engineering, 11(1), 72-81.

Zhang Z., Xie P., & Ma O. (2013). Bio-inspired trajectory generation for UAV perching. In IEEE/ASME International Conference on Advanced Intelligent Mechatronics, pp. 997-1002.

Zhang S., Srinivasan M. V., Zhu H., & Wong J. (2004). Grouping of visual objects by honeybees. Journal of Experimental Biology, 207(19), 3289-3298.

Zhang C., Platt J. C., & Viola P. A. (2005). Multiple instance boosting for object detection. In Advances in Neural Information Processing Systems,1417-1424.

Zimmermann K., Matas J., & Svoboda T. (2009). Tracking by an optimal sequence of linear predictors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(4), 677-692.

Zufferey J. C., & Floreano D. (2006). Fly-inspired visual steering of an ultralight indoor aircraft. IEEE Transactions on Robotics, 22(1), 137-146.

Chapter 2. Properties of Neuronal Facilitation that Improve Target Tracking in Natural Pursuit Simulations

Wiederman et al. (2008) have developed a neuromorphic model for target discrimination in visual clutter based on the specificity of the target rather than the segregation of target and background motion (Section 1.4.2). However, this model is not a true motion detector. It would respond to the motion of a dark target as well as a localised black single flicker. This suggests that some other form of higher order neuronal mechanism is required for robust target tracking. Recent electrophysiological experiments from one type of STMD neurons (CSTMD1) revealed two interesting higher-order properties for these neurons; 'facilitation' (Nordström et al., 2011; Dunbier et al., 2011; Dunbier et al., 2012) and 'selective attention' (Wiederman and O'Carroll, 2013a) (see Section 1.2.4.4). I hypothesized that these properties play a key role in dragonfly's robust target tracking behaviour. Therefore, in this chapter I present an elaborated closed-loop model of this target-detection pathway (set in a virtual reality environment) which includes these recent neuronal properties. As I mentioned earlier (Section 1.4.3) this closed-loop model is based on the preliminary model developed in our research group (Halupka et al., 2011, Halupka et al., 2013). Using this model, I test the role of facilitation in selective attention, as well as target tracking and pursuit in flying insects. Although these modelling efforts focus on insect physiology, these results may be generalisable to our understanding of such principles in biological organisms, as well as for translational applications (e.g. artificial vision systems).

The preliminary results of this work were published in:

"Bagheri Z. M., Wiederman S. D., Cazzolato B. S., Grainger S., & O'Carroll D. C. (2014). A biologically inspired facilitation mechanism enhances the detection and pursuit of targets

of varying contrast. In 16th International Conference on Digital Image Computing: Techniques and Applications, IEEE, 1-5."

and is presented in Appendix B. The supplementary material for the current chapter is provided in Appendix A.

Statement of Authorship

Title of Paper	Properties of Neuronal Facilitation that Improve Target Tracking in Natural Pursuit Simulations
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Journal of The Royal Society Interface, 12(108), 20150083, DOI: 10.1098/rsif.2015.0083

Principal Author

Name of Principal Author (Candidate)	Zahra Bagheri
Contribution to the Paper	Developed the code, designed the experiments, ran the model simulations, analysed and interpreted the data and drafted the manuscript with editing contributions from the other authors.
Overall percentage (%)	50
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date: 6/7/2017

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Steven Wiederman
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, assisted with analysis and interpretation of the data, provided significant editing contribution (20%).
Signature	Date: 6/7/2017

Name of Co-Author	Benjamin Cazzolato
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, assisted with analysis and interpretation of the data, provided editing contribution (10%).
Signature	Date: 6/7/17

Name of Co-Author	Steven Grainger		
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, read and approved the draft (5%).		
Signature		Date	6/7/17

Name of Co-Author	David O'Carroll		
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, assisted with analysis and interpretation of the data, provided significant editing contribution (15%).		
Signature		Date	15-NOV-2016

Properties of Neuronal Facilitation that Improve Target Tracking in Natural Pursuit Simulations

Zahra M. Bagheri[1,2], Steven D. Wiederman[1], Benjamin S. Cazzolato[2], Steven Grainger[2], David C. O'Carroll[1,3]

[1] Adelaide Centre for Neuroscience Research, The University of Adelaide, Adelaide, SA, 5005, Australia

[2] School of Mechanical Engineering, The University of Adelaide, Adelaide, SA, 5005 Australia

[3] Department of Biology, Lund University, Sölvegatan 35, S-22362 Lund, Sweden

Published: Bagheri Z. M., Wiederman S. D., Cazzolato B. S., Grainger S., & O'Carroll D. C. (2015). Properties of neuronal facilitation that improve target tracking in natural pursuit simulations. Journal of The Royal Society Interface, 12(108), 20150083.

2.1 Abstract

Although flying insects have limited visual acuity (approx. 1°) and relatively small brains, many species pursue tiny targets against cluttered backgrounds with high success. Our previous computational model, inspired by electrophysiological recordings from insect 'small target motion detector' (STMD) neurons, did not account for several key properties described from the biological system. These include the recent observations of response 'facilitation' (a slow build-up of response to targets that move on long, continuous trajectories) and 'selective attention', a competitive mechanism that selects one target from alternatives. Here, we present an elaborated STMD-inspired model, implemented in a closed loop target-tracking system that uses an active saccadic gaze fixation strategy inspired by insect pursuit. We test this system against heavily cluttered natural scenes. Inclusion of facilitation not only substantially improves success for even short-duration pursuits, but it also enhances the ability to 'attend' to one target in the presence of distracters. Our model predicts optimal facilitation parameters that are static in space and dynamic in time, changing with respect to the amount of background clutter and the intended purpose of the pursuit. Our results provide insights into insect neurophysiology and show the potential of this algorithm for implementation in artificial visual systems and robotic applications.

2.2 Introduction

Many animals have evolved the ability to visually detect moving targets, often selecting a single target from amidst many distracters. Furthermore, these animals may interact with the target by initiating motor commands (e.g. eye or body movements), which then modulate visual inputs underlying the detection and selection task (via closed-loop feedback) (Land, 1999). While such 'active vision' is almost ubiquitous in guiding complex animal behaviour, it remains uncommon in artificial vision systems. However, active vision may be key to

exploiting the highly nonlinear pre-processing of visual information by the simple insect brain for complex tasks. For example, dragonflies detect, select and then chase prey or conspecifics within a visually cluttered surround for predation, territorial or mating behaviour (Corbet, 1999). Remarkably, despite limited visual resolution (approx. 1°), they perform this task even in the presence of other distracting stimuli, such as swarms of prey and conspecifics (Corbet, 1999; Wiederman and O'Carroll, 2013). Recent studies show that dragonflies rely on both predictive and reactive control for accurate target tracking (Mischiati et al., 2014), similar to those involved in hand-reaching by primates (Wolpert et al., 1995).

The accessibility of the dragonfly for stable electrophysiological recordings makes this insect an ideal and tractable model system for investigating the neuronal correlates for complex tasks such as target pursuit. Our laboratory has identified and characterized a set of neurons likely to mediate target detection and pursuit. These 'small target motion detector' (STMD) neurons of the insect lobula (third optic neuropil) are selective for tiny targets, on the same scale as the optical resolution of the eye. STMDs are velocity-tuned, contrast-sensitive and respond robustly to targets even against the motion of high-contrast background features (O'Carroll, 1993; Nordström et al., 2006; Nordström and O'Carroll, 2009; O'Carroll and Wiederman, 2014). Individual STMDs may have very small receptive fields, corresponding to just a few dozen facets of the compound eye, or view an entire visual hemisphere, suggesting a complex hierarchy in their contributions to underlying control systems for target pursuit (O'Carroll, 1993; Nordström et al., 2006; Nordström and O'Carroll, 2009; Barnett et al., 2007).

Our recent electrophysiological data lend strong support for an underlying algorithm for local target discrimination based on an 'elementary-STMD' (ESTMD) operation at each point in the image to provide a matched spatio-temporal filter for small moving targets

embedded within natural scenery (Wiederman et al., 2008). The ESTMD model reliably predicts several properties of STMD neurons, including their spatio-temporal tuning, their rejection of background motion (Wiederman and O'Carroll, 2011) and even the selectivity for dark targets seen in some STMDs (Wiederman et al., 2013). However, this remains a model only for the elementary operation of local target discrimination: it makes no attempt to account for how information integrated across a large visual field is used to control visual gaze or target pursuit. Moreover, the ESTMD model does not explain several recently described features of STMD physiology and insect behaviour required by such a control system:

(1) The ESTMD itself is a non-directional feature detector, yet directional information is a key requirement to predict the future path of a target. Indeed, many STMD neurons are selective for the direction of target movement (O'Carroll, 1993; Barnett et al., 2007; Nordström and O'Carroll, 2006).

(2) The ESTMD model and some STMD neurons are selective for the polarity of target contrast (Wiederman et al., 2013). Yet insects pursue targets against variable backgrounds, thus requiring a mechanism to track targets independent of their relative luminance.

(3) ESTMDs are local detectors, responding only to matched features within a limited area of space viewed by the central detector and its near neighbours. Recent physiological data provide us with evidence for additional nonlinear interactions between ESTMDs viewing nearby parts of the same scene. In CSTMD1, an identified dragonfly STMD, responses build over several hundreds of milliseconds only if targets move along continuous trajectories (Nordström et al., 2011). This ‘facilitation’ resets to a naive state when there are large breaks in the trajectory (Dunbier et al., 2011; Dunbier et al., 2012). We hypothesize that such encoding of the trajectory improves extraction of target signals by noise-constrained neurons

within noisy (cluttered) environments and may thus contribute to the dragonfly's ability to robustly pursue prey (at more than 97% success rate) (Olberg et al., 2000).

(4) Presented with several salient targets at different locations, several ESTMDs within a wide-field array would respond independently if features match their spatio-temporal tuning. Successful target pursuit requires a mechanism to select one target from among alternatives (e.g. a winner-takes-all network). Indeed, CSTMD1 neurons were recently shown to competitively select one target in the presence of distracters (Wiederman and O'Carroll, 2013a), responding as if only one target was presented. This was observed, irrespective of the target's size, contrast or separation (Wiederman and O'Carroll, 2013a). We hypothesize that local facilitation might play a role in this competitive selection by enhancing discriminability of one target over the other.

To test these hypotheses and address the limitations of the ESTMD, we present here an elaborated model for a control and pursuit system inspired directly by these latest physiological findings. This is based on inputs from ESTMDs (Wiederman et al., 2008), but incorporates additional processing to permit (i) prediction of target direction (ii) robust responses independent of the luminance contrast and (iii) a competitive selection mechanism that exploits response facilitation. We incorporate this front-end processing into a closed-loop simulation of dragonfly prey pursuit that incorporates an active saccadic gaze fixation strategy inspired by studies of insect pursuit behaviour (Olberg et al., 2000; Wehrhahn et al., 1982; Land MF and Collett, 1974). We then use this system to explore the effect of varying the spatial and temporal scale of response facilitation on pursuit success and target discriminability under different environmental conditions.

Despite a relatively simple image processing strategy compared with traditional machine vision approaches, this system proves to be remarkably robust, achieving high prey capture success even with complex background clutter, low contrast and high relative speed of

pursued prey. Hence, our results should be of interest to robotics engineers looking for computationally simple yet robust systems for figure/ground segregation. Moreover, while our results are consistent with emergent hypotheses for the role of hierarchical elements of the insect STMD system, we also identify several key principles for optimal performance of such a system that are directly testable in future physiological experiments.

2.3 Methods

Appendix A, Figure A.1 shows an overview of the target pursuit model implemented in MATLAB/SIMULINK. The model is composed of five subsystems: (i) a virtual reality environment to model insect position (predator and prey) and environmental parameters, (ii) an early visual processing stage, (iii) a target matched filtering ('ESTMD') stage, (iv) a position selection and facilitation mechanism, and (v) a saccadic pursuit algorithm based on insect behaviour.

2.3.1 Virtual-Reality Environment

A Virtual Reality (VR) environment (SIMULINK 3D animation toolbox, Mathworks Inc.) was composed of a cylindrical arena (radius 6 m), rendered with textures derived from natural panoramic image data from four natural scenes (see (Brinkworth and O'Carroll, 2009) for details on acquisition of the original HDR image data). Within this arena we generated randomized 'prey' paths (100 times for each condition tested) with biologically plausible saccadic turns (Schilstra C and van Hateren, 1999) initiated when the target approached within 50 cm of the cylindrical wall. At each saccade, the new heading was constrained in a range where targets turned between 30° and 150° in the horizontal plane (i.e. as viewed from above) and to a heading in the vertical dimension that ensured that it would not be lost from the limits of the camera viewport. Response transients from initial filter conditions were disregarded by introducing the 40 mm-sized target after a delay of 40

ms. Initial target location was 4 m away from the pursuer (angular size approx. 0.6°). Video was sampled at 1000 Hz from a 40°x98°-sized viewport to represent the visual field of the pursuer, which moved at a velocity of 8 ms^{-1}.

2.3.1.1 Simulations with one target

We tested target velocities between 50% and 200% of the pursuer velocity moving against one of four natural images. All images (Image A-D, Appendix A, Figure A.1b) had varying clutter, however, all contained $1/f$ power spectra; a statistical property of natural scenes (Field, 1987). The average clutter value (see Appendix A, Text A.3 for details of the quantifying method) and intensity of each image is presented in Appendix A, Table A1. We simulated a range of target intensities set specifically for each image (Appendix A, Table A1).

2.3.1.2 Simulations with two targets

To investigate competitive selection, we simulated a black target against Image B (Appendix A, Figure A.1b) with a second distracter target introduced 100 ms later. At the time of appearance of the second target, both targets had the same size, luminance, and distance from the pursuer. Both targets travelled in mirrored paths (Appendix A, Figure A.1c) at a velocity of 6 ms^{-1}.

2.3.2 Early Visual Processing

The optics of flying insects are limited by diffraction and other forms of optical interference within the facet lenses (Stavenga, 2003). This optical blur was modelled with a Gaussian low-pass filter (Appendix A, Text A.1). The green spectral sensitivity of the insect motion pathway was simulated by processing only the green channel from the original RGB images (Srinivasan and Guy, 1990).

In biological vision, redundant information is removed with neuronal adaptation (temporal high pass filtering) and centre-surround antagonism (spatial high pass filtering) observed in physiological recordings of photoreceptors and the first-order interneurons, the large monopolar cells (LMCs). Subsequent stages of our model simulate these properties (Appendix A, Text A.1).

2.3.3 Target Matched Filtering (ESTMD stage)

2.3.3.1 Rectifying transient cell

Rectifying transient cells (RTCs) within the insect second optic neuropil (medulla) exhibit partial rectification properties well suited as additional input processing stages for an STMD pathway (Wiederman et al., 2008). Similar processing properties were implemented in our model by modelling RTC-like independent adaptation to light increments or decrements (Osorio, 1991; Jansonius and van Hateren, 1991) with strong spatial centre-surround antagonism (Appendix A, Text A.2).

2.3.3.2 Independence to target polarity

At a single location, small targets are characterized by an initial rise (or fall) in brightness, and after a short delay are followed by a corresponding fall (or rise), irrespective of the direction of travel. This property of small features is exploited in the original ESTMD model (Wiederman et al., 2008) to provide selectivity for dark objects, by multiplying each ON channel with a delayed version of the OFF channel. Sensitivity to targets independent of their polarity was provided by multiplying each contrast channel (ON or OFF) with a delayed version of the opposite polarity (via a low-pass filter, τ=25 ms) and then summing the outputs.

2.3.4 Integration and Facilitation of ESTMD Outputs

Despite its elaboration to permit detection of both contrast polarities, our ESTMD model only provides local target discrimination. To generate an input to the subsequent control

system for target pursuit, we therefore added additional stages to integrate target motion across the full visual field of the camera. Our target selection and integration stages were inspired by the observed hierarchy of insect STMD neurons and the recent evidence for facilitation within their receptive fields (Nordström et al., 2011; Dunbier et al., 2011; Dunbier at al., 2012).

2.3.4.1 Target selection

Wide-field integration in our model begins with neuron-like soft saturation of ESTMD outputs, modelled as a hyperbolic tangent function, ensuring all signals lie between 0 and 1. A simple competitive selection mechanism is then added to the target detection algorithm by choosing the maximum of the output values across the full visual field of the input camera (equivalent to the receptive field of a wide-field STMD neuron). In our model, the location of this maximum is assumed to be the target location. In the insect STMD system, such local position information could be provided by the retinotopic array of small-field STMDs (SF-STMDs) which integrate local outputs of a small number of underlying ESTMDs, as observed in the hoverfly STMD pathway (Barnett et al., 2007).

2.3.4.2 Direction selectivity and local facilitation

We implemented facilitation as observed in dragonfly CSTMD1 neurons (Figure 2.1a,b) (Nordström et al., 2011; Dunbier et al., 2011; Dunbier et al., 2012) with a Gaussian-weighted 'map' dependent on the location of the winning feature, but shifted by the target velocity vector. The ESTMD output was multiplied with a low-pass filtered version of this facilitation map with a time constant that controls duration of the enhancement around the predicted location of the winning feature. This was varied in the range 40 to 2000 ms (13 values) thus spanning the typical facilitation time course (approx. 200 ms) observed in dragonfly STMDs (Dunbier et al., 2012). We varied the size of this map via two-dimensional Gaussian kernels

from 3° to 15° (half-width at half maximum, six values), thus spanning the observed size (approx. 7°) of the receptive fields of the SF-STMD neurons (Barnett et al., 2007).

Predicting the future target location required estimation of a velocity vector (Appendix A, Figure A.2) calculated using a traditional bio-inspired direction selective model; the Hassenstein-Reichardt elementary motion detector (HR-EMD) (Hassenstein and Reichardt, 1956). In this case the EMD was applied as a second-order motion detector on the ESTMD outputs which correlates adjacent inputs after a delay (via a low-pass filter, τ=40 ms) resulting in a direction selective output tuned to the velocity of small objects (Wiederman and O'Carroll, 2013b). The spatial shift of the facilitated area was determined by segmenting the output of the HR-EMD into three equal intervals (estimating the range of target velocity).

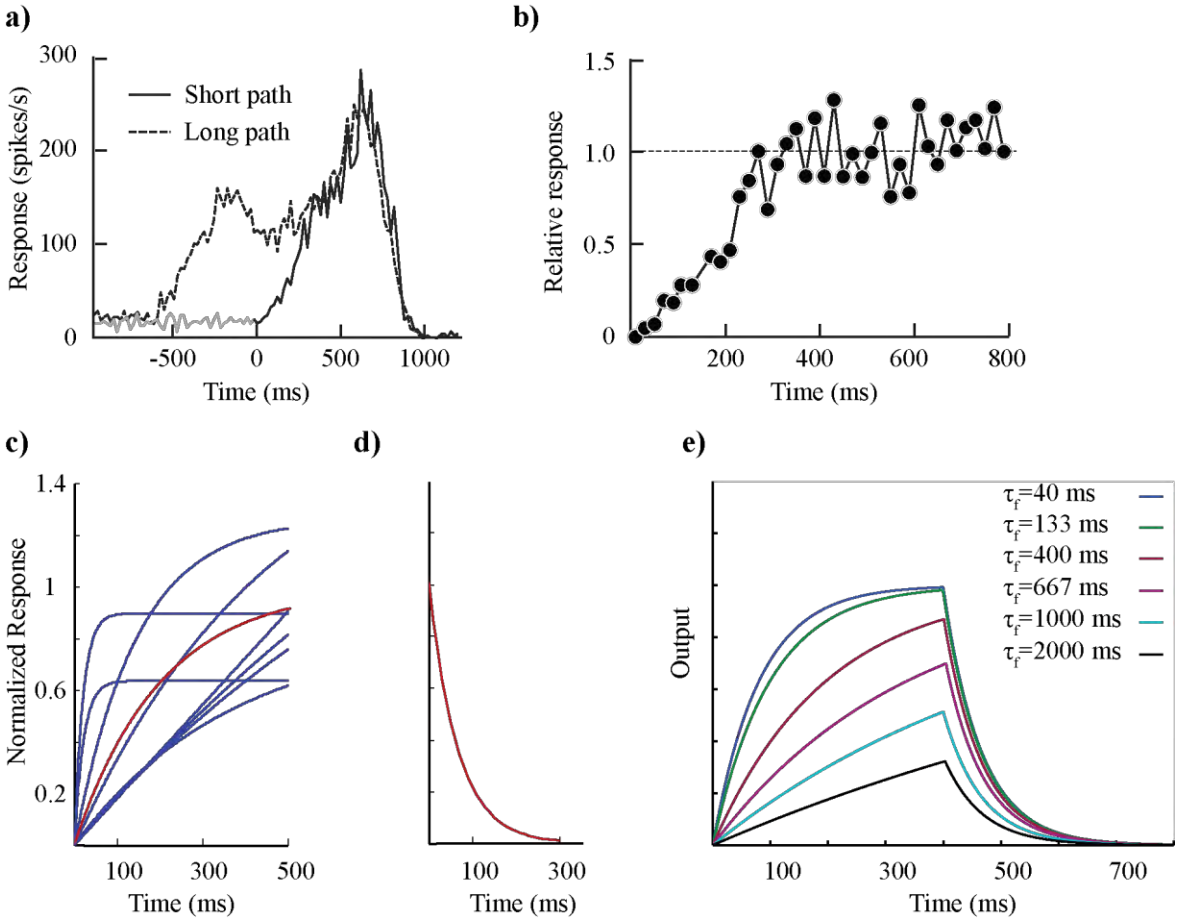

Figure 2.1. Response facilitation in a dragonfly target selective neuron, CSTMD1. a) When targets appear and commence a short path within the receptive field (solid line) responses rise above spontaneous levels within 50 ms then continue to rise for the next 450 ms. If the same target moves along a longer path (dashed line) facilitation leads to stronger responses compared with those of the short path stimulus. b) The facilitation time course is observed by subtracting the spontaneous level (grey line in a) and then dividing the short path response by the long path equivalent. c) The response time course of individual CSTMD1s to the onset of target motion are variable (blue lines) with mean response (red line) increasing over several hundreds of milliseconds. d) Individual responses of CSTMD1 to target offset are less variable (data not shown for clarity), with mean response (red line) rapidly decreasing. e) The step response of the computational model reproduces both onset and offset properties observed in the physiological systems. The time course of the model onset response changes depending on the facilitation time constant. Data in a and b are adapted from (Nordström et al., 2011).

2.3.4.3 Facilitation time course

We previously described CSTMD1 responses to the presentation of a moving target with a slow onset time course and a fast offset decay (Nordström et al., 2011). Figure 2.1c shows curve fits modified from individual CSTMD1 responses that we recorded to target onset (as described in Dunbier et al. (2012)). We previously suggested that this asymmetry reveals that the neuron is not merely 'sluggish', but rather that the slow facilitation in responses enhances encoding of the target trajectory. Further research supports this, showing that local discontinuities in target trajectory reset the response to a naive, non-facilitated state (Dunbier et al., 2011; Dunbier at al., 2012). Furthermore, even with this slow build-up in response, the underlying velocity tuning of the neuron is similar to other STMD neurons (Geurten et al., 2007).

To test whether our model of facilitation emulates this asymmetry in onset/offset time course under similar open-loop stimulus conditions, we first simulated an immobilized pursuer viewing a grey (50%) target that moved horizontally along the circumference of the arena against a white background. Our model (Figure 2.1e) reproduces the slow onset and rapid decay of the average CSTMD1 time course (Figure 2.1c, d red lines). In CSTMD1, the offset time course is consistently fast from neuron to neuron ($\bar{\tau} = 65$ ms). Although CSTMD1 shows an average 200 ms time constant in the response onset, individual examples also show large variability (Figure 2.1c, blue lines). The variation in onset kinetics is readily simulated by our alteration of the facilitation time constant, but does not affect the offset time course (Figure 2.1e).

2.3.5 Saccadic Pursuit Algorithm

Flying insects implement different pursuit strategies by maintaining the target at a specific angular position on the eye. For example, a male housefly will chase another fly by fixating it frontally (i.e., at azimuth 0°), thus 'tracking' the target in looping paths (Wehrhahn et al.,

1982; Land and Collet, 1974). On the other hand, an aerial predatory dragonfly will intercept prey, with a recent study indicating that their head movements maintain the target in a frontal region (approx. 5°) (Mischiati et al., 2014). Similar to these pursuit strategies, we implemented a hybrid 'tracking', initiating a frontal fixation saccade whenever the winning feature of the ESTMD output (Appendix A, Figure A.1a) moved more than 5° from the centre of the field of view. This strategy keeps the target close to the pole of expansion in the flow-field generated by the pursuer's own motion through the world, i.e. where local background image speeds are lowest. This discontinuous position-servo approach then allows the nonlinear spatiotemporal filtering inherent to the ESTMD pathway to enhance target 'pop out' against a highly cluttered background during the inter-saccade period.

STMD neurons are tuned to small objects, with peak responses to an optimum target size of 1.6°×1.6° (Wiederman et al., 2008). In the biological system, it is presumed that other neurons, such as the 'looming' system observed in the locusts (Rind and Simmons, 1992; Rind and Bramwell, 1996), could be recruited to finalize prey capture as the target nears and thus becomes larger than optimal for an STMD. For the sake of clarity, we therefore limited our modelling efforts to the STMD pathway and declared a successful capture when the pursuer came within 1.2 m of the frontally fixated prey in less than 2 s. To exclude 'fortuitous' last minute saccades towards background features in the vicinity of the intended target, we arbitrarily included the additional criterion that the target had to be the winner in the output of the ESTMD array for more than 50% of the last 6 ms of tracking.

2.4 Results

2.4.1 Testing Facilitation with One Target

2.4.1.1 Effect of facilitation on target discriminability in individual pursuits

While analysis of biological neurons is limited to studying facilitation as an 'always on' feature of the underlying detection pathway, our model allows us to simulate the fate of pursuit flights from identical starting points but with the facilitation mechanism turned on or off. This allows us to explore effects of facilitation both locally (target discriminability) and globally (average effect on pursuit success).

Figure 2.2 shows three individual examples of pursuit simulations where targets pass across challenging parts of the background. In all three, the non-facilitated pursuit fails, whilst facilitation results in successful target capture. Figure 2.2a,b shows two estimates of target contrast during the pursuit. The first is based on the input imagery: i.e., a simple measure of the signal to noise ratio based on the luminance difference between the target and its near background. The second is a weighted signal to noise ratio (*WSNR*) calculated (Appendix A, Text A.4) from target and background luminance, as well as target angular size (i.e., distance between target and pursuer). This takes account of the optical blur that demodulates the target at large distances owing to its small angular size. In all these examples, the *WSNR* gradually improves throughout the duration of the pursuit as the pursuer nears the target. This is because its angular size increases from the initial value (0.6°) - well below the 1.4° optical blur introduced by the optical sampling. Figure 2.2c shows target discriminability (Appendix A, Text A.5) in the ESTMD output for both facilitated and non-facilitated simulations. Discriminability values below 1 (dashed line) are due to detection failure (discriminability=0) or detection of stronger false positives (0<discriminability<1). In Example 1, the non-facilitated pursuit fails rapidly owing to low target discriminability,

whilst addition of facilitation leads to successful target capture. In Examples 2 and 3, the pursuit failure in non-facilitated simulations is due to detection of strong false positives generated by highly contrasting features of the background scene. In these examples, facilitation boosts the local response in the vicinity of the previously tracked target, maintaining focus on the target even though it is inherently less salient than such distracters.

2.4.1.2 Effect of facilitation kinetics on pursuit success in natural conditions

Although we observe clearly improved target discriminability in selected individual pursuits, facilitation might conceivably have a negative influence on target discrimination in some situations, e.g. by enhancing responses to background features. We therefore examined the pursuit success (%) with the addition of facilitation in different environmental scenarios. For this purpose we varied target to pursuer velocity ratio ($|V_t|/|V_p|$) and target intensity (Appendix A, Table A1) against different images (Appendix A, Figure A.1b). Additionally, we tested the effect of duration of enhancement (facilitation time constant) in these simulations.

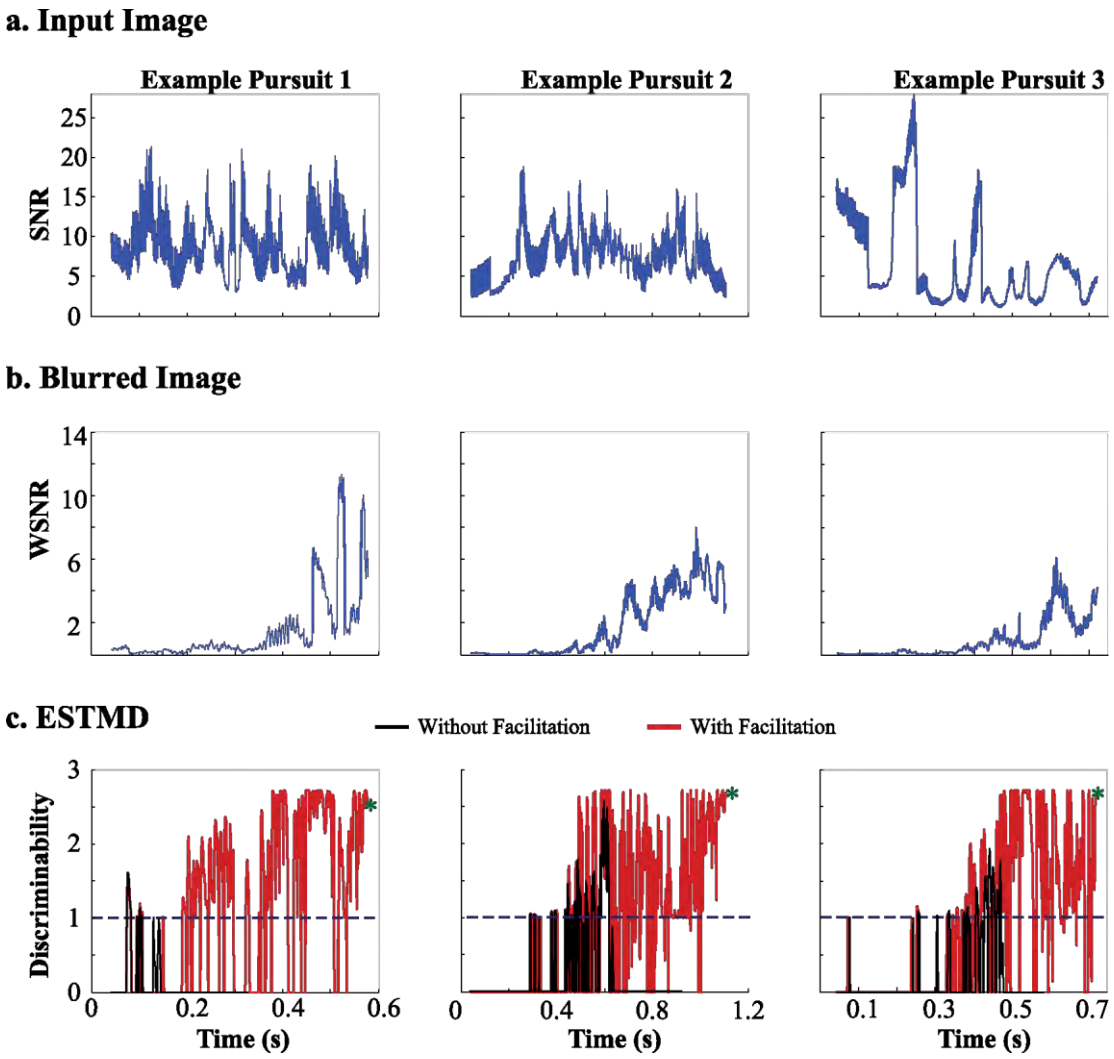

Figure 2.2. Examples of successful facilitated versus unsuccessful non-facilitated pursuits. a) Target contrast at the input image stage is calculated with signal-to-noise ratio (SNR). b) To account for target angular size, an elaborated target contrast measure (after optical blurring) uses a weighted signal-to-noise ratio metric (WSNR). The WSNR value improves throughout the duration of the example pursuits as the pursuer nears the target. c) In these examples, target discriminability at the ESTMD output stage is markedly improved with the addition of facilitation. The values above the dashed line indicate that the target is the 'winner', and the green asterisk indicates the successful capture.

Figure 2.3 shows pursuit success (averaged over four target intensities) varied across the four images. The more cluttered images, Image A and Image D, yield maximum success rates of approximately 50%. Both include features that evoke false positives, either naturalistic (e.g. foliage in Image A) or man-made (e.g. windows in Image D). Image B is

sparse and Image C contains predominantly straight edges. Neither evoke many false positives, resulting in maximum success rates exceeding 75%.

As target velocity exceeds that of the pursuer, pursuit success decreases because the pursuer is simply not fast enough to catch the target (Figure 2.3). However, given the randomized nature of its path, targets sometimes move towards the pursuer at some point during the simulation so pursuit success is greater than zero. Moreover, pursuit success exhibits tuning to optimal velocity ratios because if a target moves too slowly or too quickly, it falls out of the velocity tuning range that is an inherent property of the ESTMD model (Wiederman et al., 2008).

The optimum facilitation time constant (dashed line) changes in response to variation of the target-pursuer velocity ratio. As target velocity increases, the optimum time constant decreases. This reflects the fact that faster targets require a facilitated area of enhancement to rapidly ‘keep up’ and have less ‘sluggish’ kinetics (i.e. lower time constant). Hence the ideal time constant is likely to depend on the goal of the pursuit system (e.g. slower-moving prey versus fast-moving conspecifics).

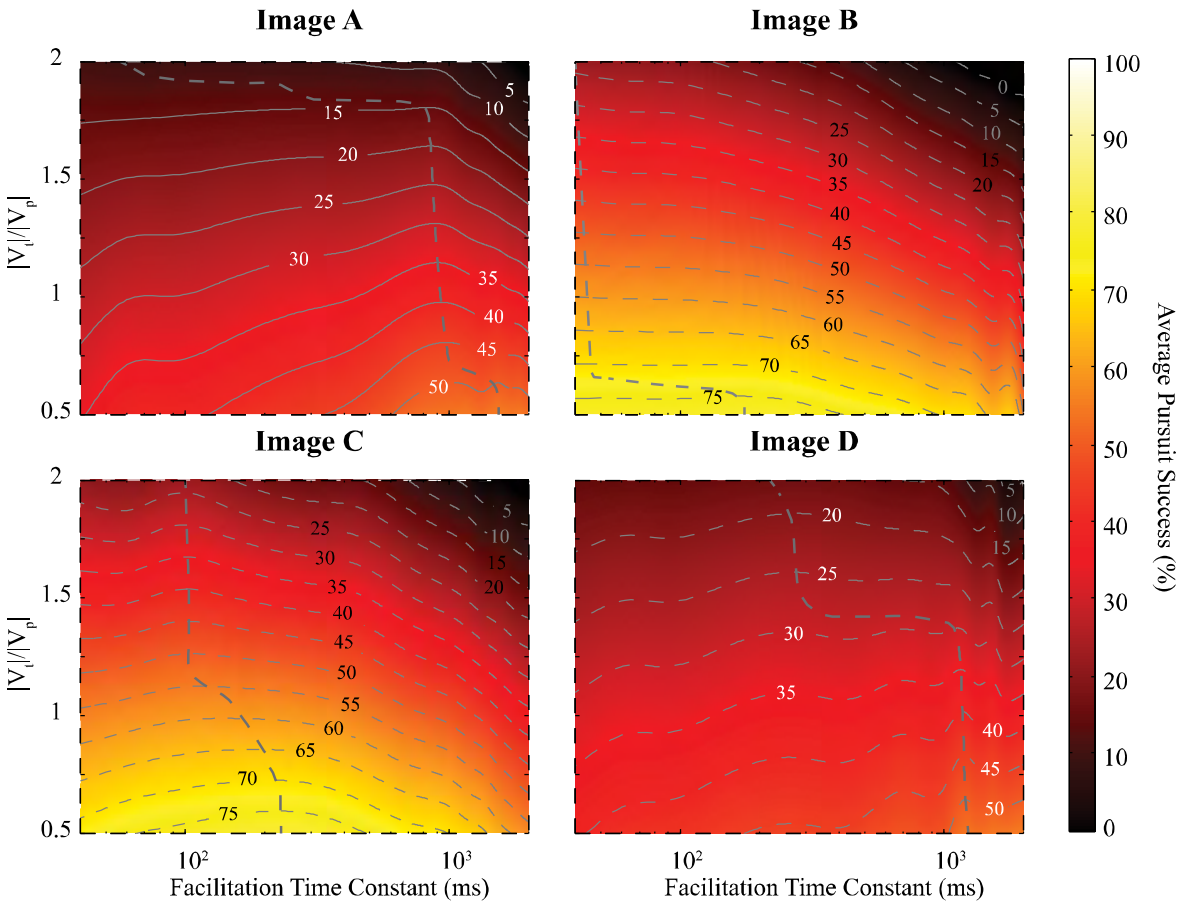

Figure 2.3. Average pursuit success at varying velocity ratios and with changes in facilitation duration for four different background images. This reveals higher pursuit success (%) for less cluttered scenes (Image B, Image C). Average pursuit success increases as the target-pursuer velocity ratio ($|V_t|/|V_p|$) decreases. These results reveal that there is an optimum facilitation time constant (dashed lines) which varies dependent on both target velocity and the background scene.

2.4.1.3 Model robustness and facilitation kinetics in clutter

In addition to the consistent influence of target speed, the optimum facilitation time constant changes across images. Pursuit success at any particular target speed improves with longer time constant in Images A and Image D, compared with the qualitatively less cluttered Images B and C (Figure 2.3). These data also suggest a relationship between the kinetics of facilitation and the average amount of background clutter. However, this is complicated by the fact that targets may sometimes be viewed against locally less cluttered parts of generally high-clutter scenes and vice versa. To quantify this relationship and investigate how it may

evolve during a pursuit, we developed a robustness metric (Appendix A, Text S6). This represents the real performance of the model as a percentage of the ideal performance. The minimum criterion for an acceptable performance is a successful pursuit; however, a higher target discriminability throughout the pursuit indicates more robust performance. Therefore, our robustness metric calculates success rate, with individual successes weighted by their average target discriminability.

We examined the evolution of pursuit robustness in response to changes in background clutter (Appendix A, Text A.3) and facilitation kinetics. Pursuit durations varied across all images with an overall mean of 0.6561±0.0031 s (Appendix A, Figure A.3). Figure 2.4 shows pursuits segmented into three equal periods (determined for each individual pursuit depending on its duration), as well as the whole pursuit (|Vt|/|Vp|=3/4). Robustness is initially less than 15% (Figure 2.4a), however, it increases as the pursuit progresses (Figure 2.4b,c), owing to the growth of target angular size and the build-up of facilitation. These data confirm that as background clutter increases, target discrimination decreases. More interestingly, the optimum facilitation time constant (dashed line) is oriented towards longer time constants as background clutter increases. This trend can be explained by more frequent camouflage of the target in more cluttered backgrounds. Consequently, a longer time constant enhances the area of target disappearance for a longer duration, thus improving target discrimination when it reappears. This is an analogue of the 'expectation' human observers have for the reappearance of a target, following disappearance behind an occluding object (Doherty et al., 2005). However, when the background is less cluttered, facilitation with a long time constant lags behind a visible target, with the potential to enhance false positives. Consequently, faster facilitation performs better for smaller clutter values.

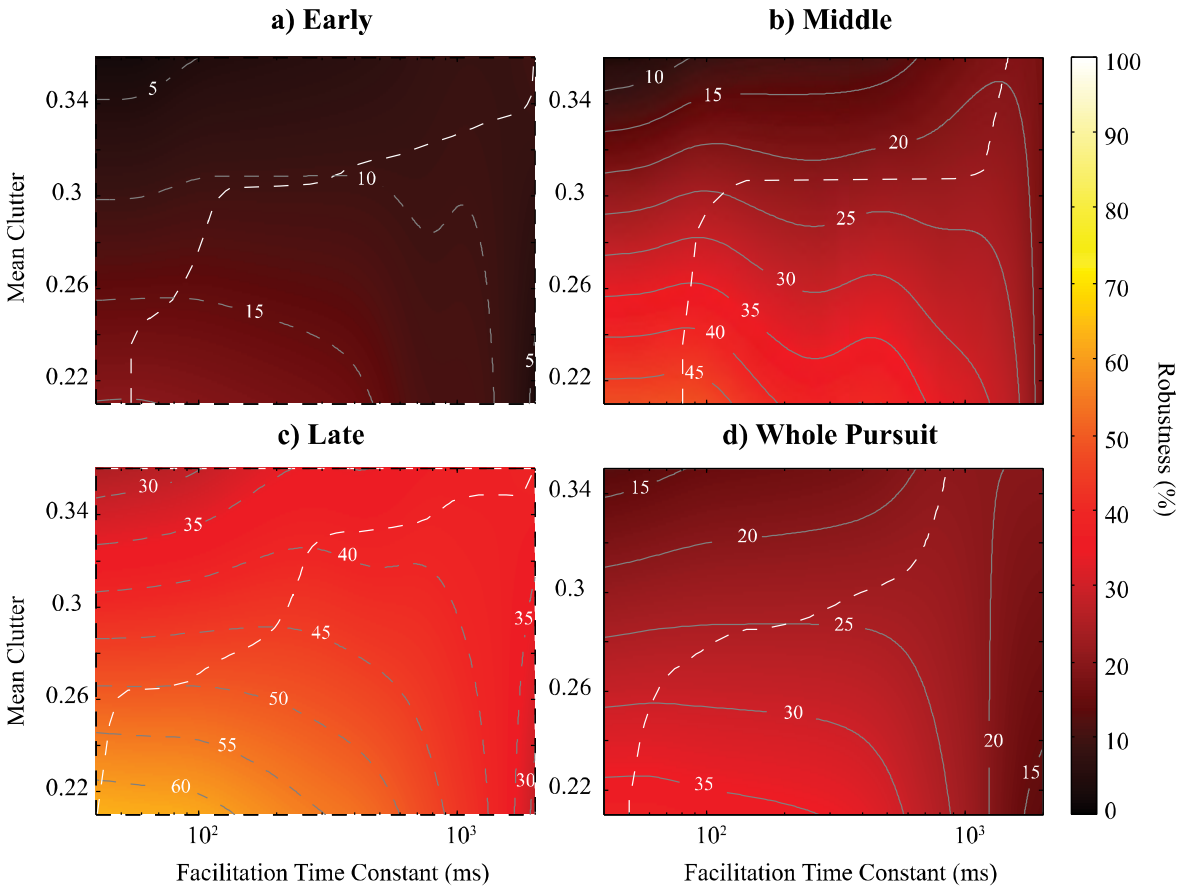

Figure 2.4. (a-d) A robustness metric is calculated (relating both target discriminability and pursuit success, see Appendix A, Text A.6) for three equal periods of pursuit (Early, Middle, and Late) as well as for the entire duration of the pursuit. Robustness improves throughout the pursuit owing to the growth of target angular size and the build-up of facilitation. Overall, higher local clutter has a detrimental effect on pursuit robustness. The oriented nature of the optima (dashed line) reveals that more reliable model performance in increased clutter is obtained with longer facilitation time constants.

2.4.1.4 The optimum facilitation kernel size

Our rationale for using a Gaussian-weighted patch as the basis for facilitation is that position information must be represented in the insect brain via retinotopically organised local feature detectors (Barnett et al., 2007). It is most likely that facilitation operates at the level of these units. In hoverflies, retinotopically organised SF-STMD neurons have approximately Gaussian receptive fields with a half-width approximately 7° (Barnett et al., 2007). It is interesting to consider whether there is a clear optimum scale for this operation. Figure 2.5

displays the effect of varying facilitation kernel size on pursuit success for different time constants, averaged over target intensities. Small sizes of the kernel (half-width less than 7°) diminishes the model's ability to successfully track the target. Because the velocity vector used to shift the facilitation map (Section 2.3.4.2) is only an estimate of the target velocity, too small a facilitation patch might not accurately enhance the area around the target, particularly when the target trajectory is unpredictable (e.g. during a prey saccade). Although larger kernels increase the probability of enhancing the correct location of the target, they also boost false positives in a larger area of the background. This likely explains declining pursuit success for kernel sizes above 9° (Figure 2.5). Pursuit success reaches its maximum at the size of 7°-9° for all images, remarkably similar to the size of insect SF-STMD receptive fields (Barnett et al., 2007).

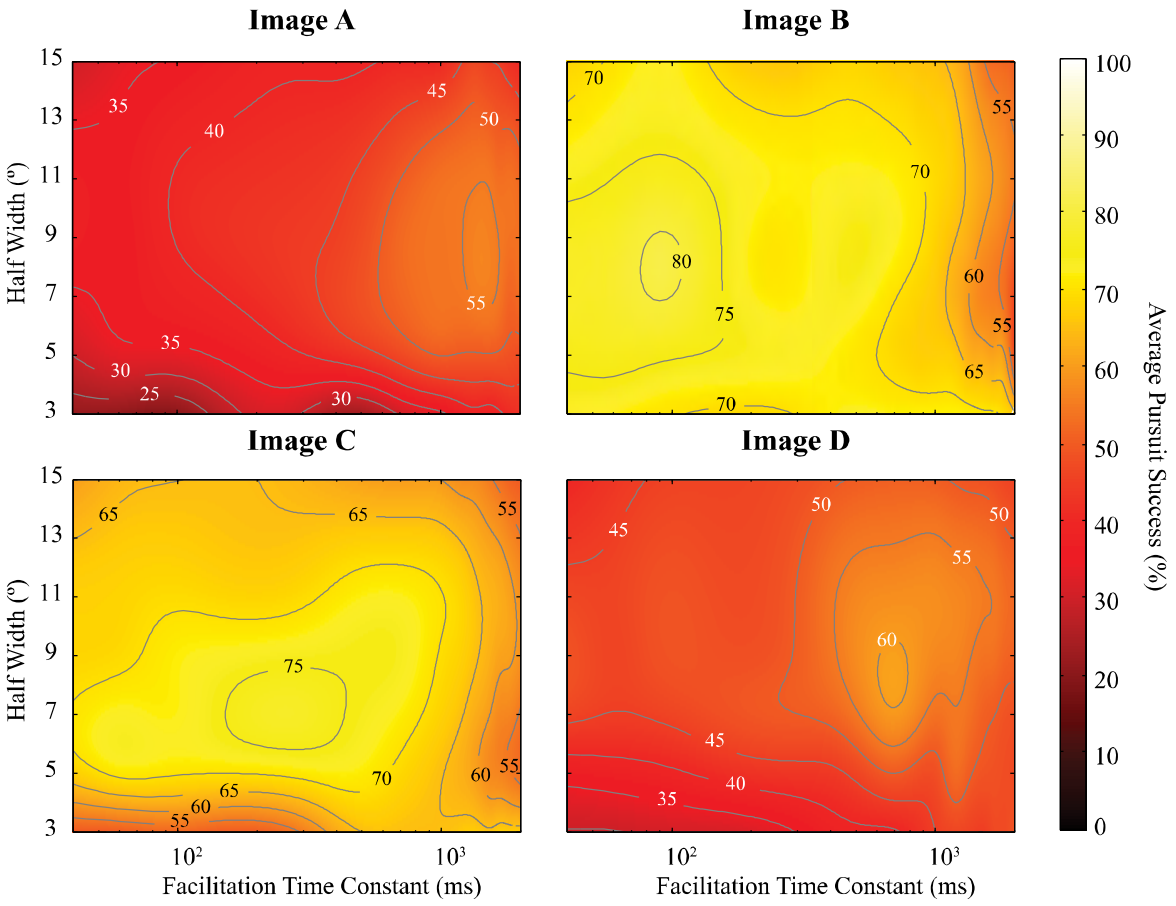

Figure 2.5. Average pursuit success for the background images (Appendix A, Figure A.1b) is plotted with respect to both the facilitation time constant and facilitation kernel size (half-width). More cluttered images have optima at longer facilitation time constants; however, spatial facilitation is optimal at 7°-9° irrespective of image.

2.4.1.5 Facilitated versus non-facilitated model

The preceding sections explored optimal parameters for the facilitation mechanism, but do not tell us about the resulting gain in performance. We therefore ran pursuits using optimal parameters for each image and then re-ran them from identical starting points with facilitation turned off, to quantify the *absolute* improvement in pursuit success (Figure 2.6). With less cluttered images (Image B and Image C), high contrast targets led to very high capture success even without facilitation, leaving little headway for improvement. We therefore also ran simulations across a larger range of target intensities than in the earlier simulations. The success rate in non-facilitated simulations (contour lines) improves as

target intensity is increased or decreased away from the mean background (dashed line). The difference between facilitated and non-facilitated model is shown with the colour map. The largest improvement from facilitation (hot colours) happens when the non-facilitated model is only successful in 30-60% of simulations. When the target contrast is too low, facilitation has no effect owing to the complete failure of target detection (grey area in the plot).

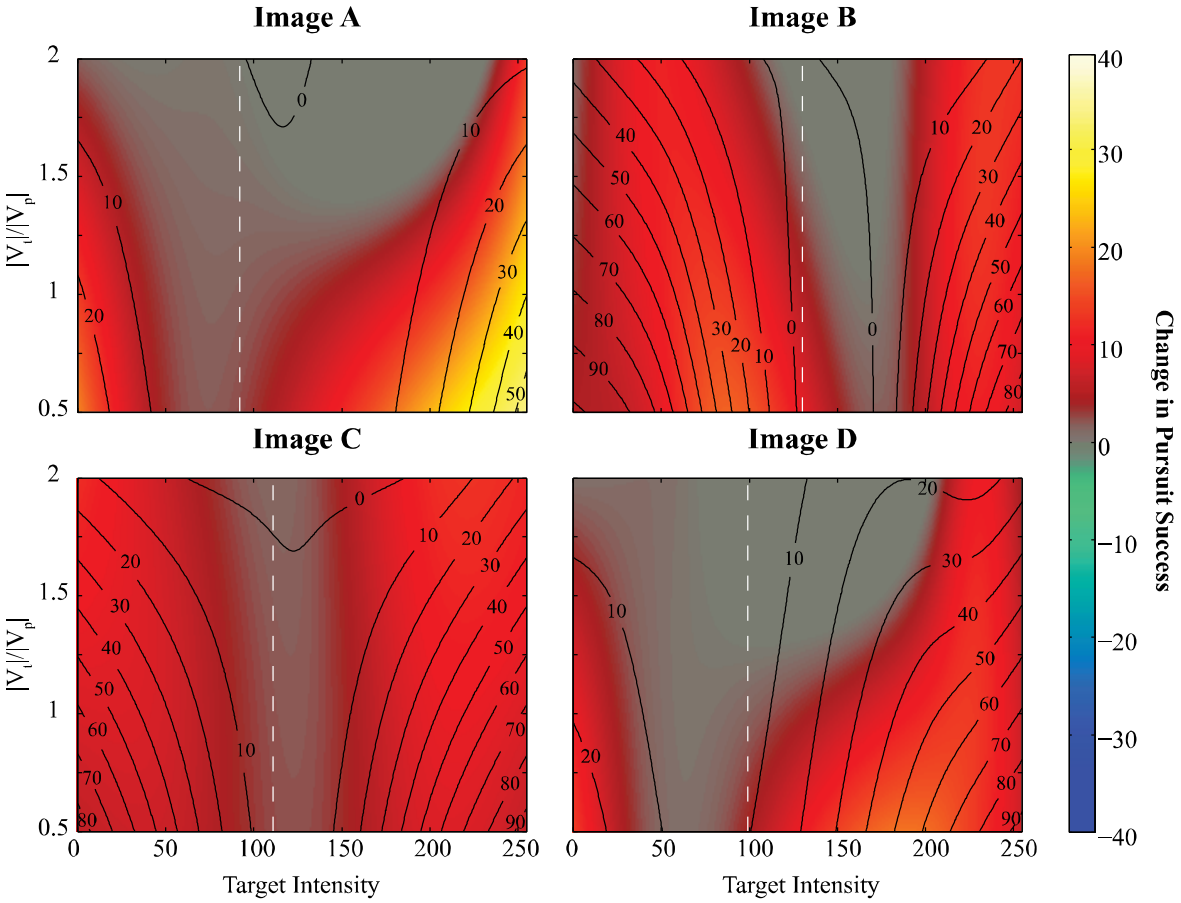

Figure 2.6. Optimized facilitation improves pursuit success rate over a range of target intensities and target-pursuer velocity ratios (|Vt|/|Vp|). Contour lines indicate pursuit success without facilitation, whilst the colour map portrays the difference in success owing to the addition of facilitation. Over all conditions, facilitation either has no effect or improves pursuit success (hot colours). White dashed lines indicate the mean background intensity.

2.4.2 Testing Competitive Selection with Two Targets

Dragonflies can feed among swarms of prey, requiring an attention mechanism to select one target. A likely neuronal correlate is the competitive selection previously observed in

CSTMD1 electrophysiological recordings (Wiederman and O'Carroll, 2013a). These neurons respond to only one of the competing stimuli at any point in time, although they may switch from one to the other. To examine the interplay between facilitation and competitive selection in our model, we introduced a distracter target to the simulations. We determined whether facilitation results in 'locking' on to one target, thus reducing the number of switches between two targets.

2.4.2.1 Examples of effect of facilitation on competitive selection

Figure 2.7 shows discriminability of the targets in the ESTMD output and their WSNR value (Appendix A, Text A.4) for simulations both with and without facilitation. As intuitively expected for two identical targets, in the non-facilitated case, both targets have similar discriminability and thus compete as 'winners' in the ESTMD output (Figure 2.7a). At any instant, the winner depends only on the local properties of the background, leading to a competitive switching between the two. As a consequence, any given saccade is more likely to result from a switch in 'attention' between the two targets (i.e. change in the current winner indicated by a discriminability value greater than 1) than a 'normal' fixation saccade to re-centre the tracked target in the frontal visual field. However, the addition of facilitation reduces both these attentional switches (Figure 2.7a-c) and resulting switching saccades (Figure 2.7a-c, small coloured markers). Switches are reduced even further by increasing facilitation time constant. In the example shown in Figure 2.7c, the longer time constant leads to the model focusing exclusively on the first target after 170 ms, with only normal fixation saccades after this time. In the same simulation with a shorter facilitation time constant (Figure 2.7b) a lock-on to the second target occurs after 240 ms, following several prior switches between the two. Considering the WSNR value, in the facilitated simulations, even when the winning target moves across an area of the background that causes its contrast to decrease below that of the alternative, the model does not necessarily switch to the

inherently more ‘salient’ target. An example is shown in Figure 2.7c, where the discriminability of target 1 is boosted by facilitation to a winning level (greater than1) even though target 2 has a higher WSNR at the corresponding time frame (the dashed box). These examples support the idea that facilitation plays a potential role in the selective attention observed in the STMD neurons (Wiederman and O’Carroll, 2013a).

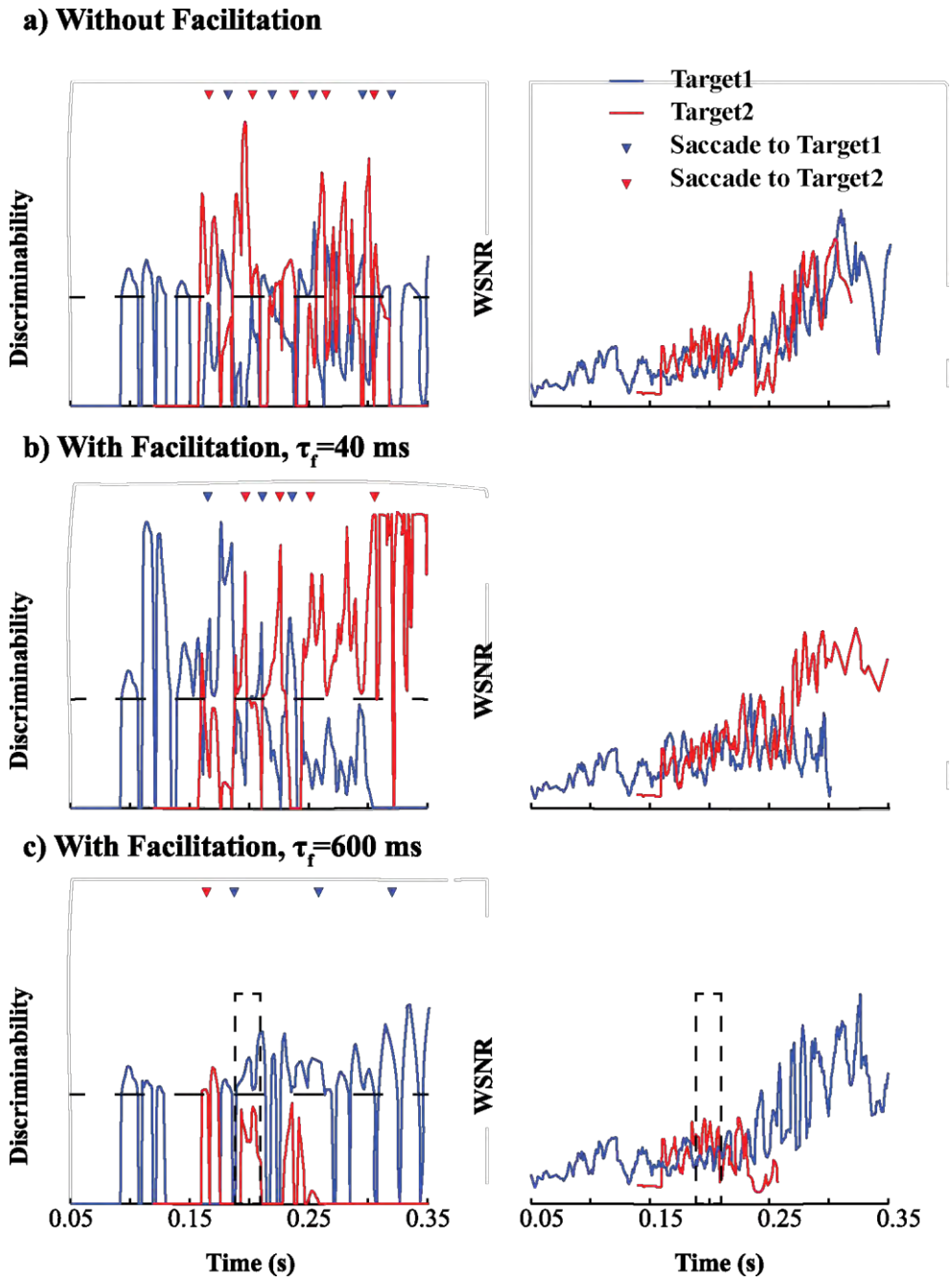

Figure 2.7. Example of facilitation effect on switches between two targets. a) In the non-facilitated simulation, targets have similar discriminability and compete closely as winner. b) The results of the same simulation with addition of facilitation (τ_f=40 ms). c) The same simulation with addition of a more sluggish facilitation (τ_f=600 ms). The small triangular markers show the saccades towards the target which has the same representing colour. This example shows that the facilitation mechanism effectively reduces the number of both attentional switches and switching saccades.

2.4.2.2 Effect of facilitation on saccade frequency

To further quantify the effect of facilitation on competitive selection, we calculated the frequency of both fixation saccades (saccades toward the same target) and switching saccades (from one target to the other). Figure 2.8 clearly shows that the frequency of switching saccades decreases with longer facilitation time constants. With longer time spent with individual targets as the winner, this leads to an increase in the frequency of fixation saccades (Figure 2.8a, open symbols). Interestingly, we see this increase in fixation saccades even for relatively short facilitation time constants (less than100 ms), even though these do not lead to a significant reduction in the corresponding frequency of switching saccades. This reflects the fact that the boost in local 'salience' to one or other (or both) targets resulting from even short facilitation time constants leads to a decrease in the time during a pursuit when neither target is the winner (Figure 2.8b). As expected, for long time constants, when fixation saccades are dominant, average saccadic turn angles approach the angle defined for initiating a frontal fixation saccade (5°, Figure 2.8c). Although the appearance of the second target is always as a mirror image of the first target and thus initially at a small angular separation (between 3° and 10°), as pursuits continue the different paths of the two targets leads to an increase in their angular separation. Consequently, the average saccadic turn angle increases dramatically for shorter time constants where switches between the two targets dominate saccades.

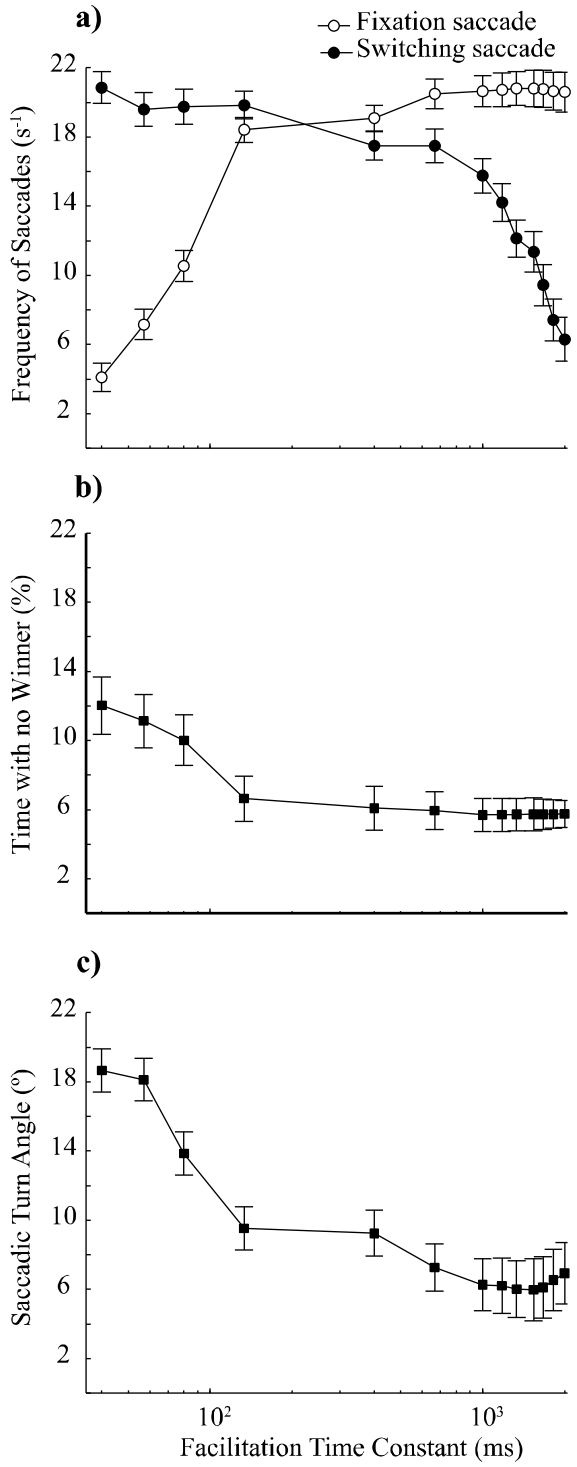

Figure 2.8. Effect of facilitation time constant on the saccades in pursuit of two targets. a) The increase in facilitation time constant leads to a decrease in the frequency of switching saccades and increase in the frequency of fixation saccades. b) The boost in local discriminability of the targets resulting from facilitation leads to a decrease in the percentage of the time during which neither target is the winner. c) As the facilitation time constant increases, the average saccadic turn angle converges to the value of pre-defined re-centring angle (5°) resulting from domination of fixation saccades.

2.5 Discussion

Our data clearly show that an elaborated version of the ESTMD model incorporating summation of feature detectors for both dark and light-contrasting targets provides robust detection of varying target intensities against a wide range of backgrounds. This rivals the remarkable sensitivity for low contrast targets of the insect visual system upon which it is based (O'Carroll and Wiederman, 2014).

2.5.1 Facilitation Time Constant

The newly discovered facilitation in CSTMD1 neurons is suspected to play a role in enhancing sensitivity for targets moving along long trajectories, possibly contributing to the high capture rate in dragonflies. Supporting this hypothesis, inclusion of facilitation in our closed-loop model substantially improves pursuit success. This particularly interesting improvement was even observed for very slow facilitation time constants despite the average duration of successful pursuits remaining relatively short. In this regard, our model optimization parallels analysis of both physiology and behaviour in dragonflies. Dragonfly pursuit flights are typically very short, with an average of 184 ms (Olberg et al., 2000) (although we note that this was calculated from the time the dragonfly *commenced* pursuit, so it is likely that the underlying neurons were already encoding target motion for some time prior). Yet the physiologically measured facilitation time reported in our earlier work is on the order of 300-500 ms ((Nordström et al., 2011; Dunbier et al., 2011; Dunbier at al., 2012), and Figure 2.1a-c).

One might intuitively expect that time constants governing biological image processing should be at least as fast as the behaviour for which they are employed. In fact, the same expectation is a fundamental basis of control theory. Reducing the phase delay to reach the steady-state mode in the shortest possible time is one of the main concerns in the design

process of closed-loop systems (Ogata and Yanjuan, 1970; Dorf, 1995). However, our observation – in both the biological system and in our model of it – suggests that this need not always be the case. The reduced effectiveness of facilitation with very short time constants most likely reflects numerous contributing factors including uncertainty of the actual target location at any given instant and periods during which the target is camouflaged in the background clutter. Having a sluggish time constant for facilitation allows 'persistence' in enhancement of the estimated vicinity of a temporarily invisible target. But as a consequence, the fully facilitated 'steady state' response of neurons observed in the laboratory after 500 ms of motion against a blank background may be rarely experienced in nature when the entire pursuit may be shorter (Olberg et al., 2000). On the other hand, prey pursuit is not the only target detection and pursuit task that many insects engage in, e.g. dragonflies also pursue fast moving conspecifics for several seconds in highly cluttered environments. Targets would be frontally fixated for much of such pursuits, providing ample time for STMD neurons to become fully facilitated.

As mentioned earlier, the actual optimum in this relationship between facilitation kinetics and pursuit behaviour likely depends on many factors and should be dynamic, changing based on the amount of background clutter and the target velocity. Therefore, it is possible that facilitation time course may reflect different 'modes' of behaviour adopted by different species. Some dragonflies capture small moving prey above water, where low vegetation in the ecosystem provides a less cluttered environment (Corbet, 1999). The rapid prey pursuit flights analysed by Olberg et al. (2000) were from perching dragonfly species that view prey 45-90° above the horizon at their perch location, allowing them to use the sky as a clear background for detection of small moving prey (Corbet, 1999). However, our physiological recordings are taken from a hawking dragonfly (*Hemicordulia*) which feeds while continuously flying and is frequently observed catching prey in complex visual clutter. We

predict that the optimum facilitation mechanism might thus differ substantially between species that adopt such varied behaviour. It awaits further experiments from other dragonfly species to test this hypothesis.

The other question thus arises as to whether facilitation observed in CSTMD1 neurons indeed exploits dynamic kinetics given different environments or target speeds (as our model predicts), or whether the insect has simply evolved a static facilitation time constant for its natural habitats. Because species like *Hemicordulia* must deal with a variety of different background scenes, an intriguing possibility is that a dynamic facilitatory time constant would allow for variation in both the background clutter as well as for the purpose of the pursuit (e.g. prey or conspecific). The variation in onset time course observed in individual CSTMD1 recordings (Figure 2.1c) may indicate a dynamically adaptable facilitation mechanism, possibly modulated by factors such as preceding visual stimuli (clutter) and behavioural states (e.g. attention). Further physiological and behavioural experiments will be required to address these questions more directly.

2.5.2 Spatial Mechanisms of Facilitation

Our implementation of facilitation involved an element with a Gaussian receptive field property inspired by recordings of small-field elements of the insect STMD pathway. We hypothesized that such elements represent a level at which target location is encoded by this pathway, but our simulations were able to explore a range of widths for the 'receptive field' of these facilitating elements. Impressively, our results reveal an optimum kernel size for facilitation close to the size of SF-STMD receptive field (approx. 7°) observed in hoverflies (Barnett et al., 2007). The fact that this receptive field size is intermediate between the size of the local elements (ESTMDs) that are actually responsible for local target motion detection and the larger receptive fields of STMD neurons like CSTMD1, suggests a hierarchy in the organisation of STMD neurons. This intermediate scale for SF-STMDs

within this hierarchy most likely results from the trade-off between uncertainty in the estimation of target location and in the value of long-lasting facilitation in maintaining an improved sensitivity for targets in clutter. This would be particularly important when features are temporarily obscured by their low contrast against the background or when passing behind occluding foreground features. Although a larger facilitation kernel may increase the probability of enhancing the appropriate target despite uncertainty in its precise location, it also increases the chance of detecting false positives in the background. We have not yet tested this hypothesis for occluding features, but our model architecture is ideally suited to elaborated simulations of pursuit in more structured three-dimensional scenarios.

2.5.3 Facilitation or Attention?

Our data support a possible role for facilitation in selective attention, even though we did not implement it as an attention mechanism per se. However, we noted that the addition of facilitation leads to both a decrease in the proportion of time during which neither target is the winner, and in the frequency of switches between fixating the two alternatives. We also see clear examples (Figure 2.7) where the previously fixated target remains the winner despite not always being the inherently more salient of the two – a classic hallmark of attention (Ipata et al., 2006; Sawaki et al., 2012). In this respect our data mirror the response of the dragonfly CSTMD1 neuron, which displays selective attention in response to two targets moving simultaneously in its receptive field (Wiederman and O'Carroll, 2013a). In CSTMD1, there are also clear instances where the initially fixated target remains the 'winner' even when the alternative would have produced a stronger response had it been presented alone (Wiederman and O'Carroll, 2013a). In our model, both the improvement in the relative frequency of normal fixation saccades towards the selected target (as opposed to switches) and the decrease in the average saccadic turn angle saturate as the facilitation time

constants approach the order of 100 to 300 ms. Once again, this is a remarkably close fit to the observed time constant in CSTMD1 (Nordström et al., 2011; Dunbier et al., 2012).

Selective attention in insects and other animals undoubtedly involves additional processes to the relatively simple selection mechanism we implemented here. Nevertheless, our results support a potentially important role for a 'bottom-up' competitive process in attention as an emergent property of lower level processing in the STMD pathway. However, given the relatively simple winner-takes-all mechanism that we implemented for target selection, we cannot reject the possibility that target selection in biological STMDs might not also reflect a top-down attention process. Testing such mechanisms in physiological recordings is severely restricted in the richness of stimuli that can be presented, because the animal is restrained with wax and can only experience stimuli imposed upon it in open-loop. A major advantage of our computational model is that it allows us to investigate future questions that are difficult or impossible to conceive in physiological recordings. For example, of the effect of facilitation on both attentional switches and the physical saccades. Given its ability to reproduce so much of the lower-level behaviour of the physiological system, our model thus provides a promising platform to further explore the possible 'higher order' network interactions that may be involved in target selection. As we have demonstrated here, a further advantage of our modelling approach is that it also generates hypotheses that require further physiological experiments that are feasible in open loop STMD recordings, such as the possibility that facilitation time constants may be influenced by the statistics of the background clutter against which stimuli are presented.

References

Barnett P. D., Nordström K., & O'Carroll D. C. (2007). Retinotopic organization of small-field-target-detecting neurons in the insect visual system. Current Biology, 17(7), 569-578.

Brinkworth R. S., & O'Carroll D. C. (2009). Robust models for optic flow coding in natural scenes inspired by insect biology. PLoS Computational Biology, 5(11), e1000555.

Corbet, P. S. (1999). Dragonflies: behaviour and ecology of Odonata. Harley Books.

Doherty J. R., Rao A., Mesulam M. M., & Nobre A. C. (2005). Synergistic effect of combined temporal and spatial expectations on visual attention. The Journal of Neuroscience, 25(36), 8259-8266.

Dorf R. C. (1995). Modern Control Systems. Addison-Wesley Longman Publishing Co., Boston.

Dunbier J. R., Wiederman S. D., Shoemaker P. A. & O'Carroll D. C. (2012). Facilitation of dragonfly target-detecting neurons by slow moving features on continuous paths. Frontiers in Neural Circuits, 6, 79.

Dunbier J. R., Wiederman S. D., Shoemaker P. A., & O'Carroll, D. C. (2011). Modelling the temporal response properties of an insect small target motion detector. In Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing,125-130.

Field D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. Journal of the Optical Society of America A, 4(12), 2379-2394.

Geurten B. R., Nordström K., Sprayberry J. D., Bolzon D. M., & O'Carroll D. C. (2007). Neural mechanisms underlying target detection in a dragonfly centrifugal neuron. Journal of Experimental Biology, 210(18), 3277-3284.

Halupka K. J., Wiederman S. D., Cazzolato B. S., & O'Carroll D. C. (2013). Bio-inspired feature extraction and enhancement of targets moving against visual clutter during closed loop pursuit. In IEEE International Conference on Image Processing, 4098-4102.

Halupka K. J., Wiederman S. D., Cazzolato B. S., & O'Carroll D. C. (2011). Discrete implementation of biologically inspired image processing for target detection. In IEEE Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 143-148.

Hassenstein, B. & Reichardt W. (1956). Systemtheoretische analyse der zeit-, reihenfolgen- und vorzeichenauswertung bei der bewegungsperzeption des rüsselkäfers chlorophanus. Zeitschrift für Naturforschung B, 11(9-10), 513-524.

Ipata A. E., Gee A. L., Goldberg M. E., & Bisley J. W. (2006). Activity in the lateral intraparietal area predicts the goal and latency of saccades in a free-viewing visual search task. The Journal of Neuroscience, 26(14), 3656-3661.

Jansonius N. M., & Van Hateren J. H. (1991). Fast temporal adaptation of on-off units in the first optic chiasm of the blowfly. Journal of Comparative Physiology A, 168(6), 631-637.

Land M. F. (1999). Motion and vision: why animals move their eyes. Journal of Comparative Physiology A, 185(4), 341-352.

Land M. F., & Collett T. S. (1974). Chasing behaviour of houseflies (Fannia canicularis). Journal of Comparative Physiology, 89(4), 331-357.

Mischiati M., Lin H. T., Herold P., Imler E., Olberg R., & Leonardo A. (2015). Internal models direct dragonfly interception steering. Nature, 517(7534), 333-338.

Nordström K., & O'Carroll D. C. (2009). Feature detection and the hypercomplex property in insects. Trends in Neurosciences, 32(7), 383-391.

Nordström K., & O'Carroll D. C. (2006). Small object detection neurons in female hoverflies. Proceedings of the Royal Society of London B: Biological Sciences, 273(1591), 1211-1216.

Nordström K., Bolzon D. M., & O'Carroll D. C. (2011). Spatial facilitation by a high-performance dragonfly target-detecting neuron. Biology Letters, 7(4), 588-592.

Nordström K., Barnett P. D., & O'Carroll D. C. (2006). Insect detection of small targets moving in visual clutter. PLoS Biology, 4(3), 4:378-386.

O'Carroll D.C. (1993). Feature-detecting neurons in dragonflies. Nature, 362, 541-543.

O'Carroll D. C., & Wiederman S. D. (2014). Contrast sensitivity and the detection of moving patterns and features. Philosophical Transactions of Royal Society B, 369(1636), 20130043.

Ogata K. & Yanjuan Y. (1970). Modern Control Engineering. Prentice Hall, New York.

Olberg R. M., Worthington A. H., & Venator K. R. (2000). Prey pursuit and interception in dragonflies. Journal of Comparative Physiology A, 186(2), 155-162.

Osorio D. (1991). Mechanisms of early visual processing in the medulla of the locust optic lobe: How self-inhibition, spatial-pooling, and signal rectification contribute to the properties of transient cells. Visual Neuroscience, 7(4), 345-355.

Rind F. C., & Bramwell D. I. (1996). Neural network based on the input organization of an identified neuron signaling impending collision. Journal of Neurophysiology, 75(3), 967-985.

Rind F. C., & Simmons P. J. (1992). Orthopteran DCMD neuron: A reevaluation of responses to moving objects. I. Selective responses to approaching objects. Journal of Neurophysiology, 68(5), 1654-1666.

Sawaki R., Geng J. J., & Luck S. J. (2012). A common neural mechanism for preventing and terminating the allocation of attention. The Journal of Neuroscience, 32(31), 10725-10736.

Schilstra C., & van Hateren J. (1999). Blowfly flight and optic flow. I. Thorax kinematics and flight dynamics. Journal of Experimental Biology, 202(11), 1481-1490.

Srinivasan M. V., & Guy R. G. (1990). Spectral properties of movement perception in the dronefly Eristalis. Journal of Comparative Physiology A, 166(3), 287-295.

Stavenga D. (2003). Angular and spectral sensitivity of fly photoreceptors. I. Integrated facet lens and rhabdomere optics. Journal of Comparative Physiology A, 189(1), 1-17.

Wehrhahn C., Poggio T., & Bülthoff H. (1982). Tracking and chasing in houseflies (Musca). Biological Cybernetics, 45(2), 123-130.

Wiederman S. D., & O'Carroll D. C. (2013a). Selective attention in an insect visual neuron. Current Biology, 23(2), 156-161.

Wiederman S. D., & O'Carroll D. C. (2013b). Biomimetic target detection: Modeling 2nd order correlation of OFF and ON channels. In IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP), 16-21.

Wiederman S. D., & O'Carroll D. C. (2011). Discrimination of features in natural scenes by a dragonfly neuron. The Journal of Neuroscience, 31(19), 7141-7144.

Wiederman S. D., Shoemaker P. A., & O'Carroll D. C. (2013). Correlation between OFF and ON channels underlies dark target selectivity in an insect visual system. The Journal of Neuroscience, 33(32), 13225-13232.

Wiederman S. D., Shoemaker P. A., & O'Carroll D. C. (2008). A model for the detection of moving targets in visual clutter inspired by insect physiology. PloS One, 3(7), e2784.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. Science, 269(5232), 1880-1882.

Chapter 3. Effect of Facilitation on the Efficiency and Efficacy of Target Tracking

In the previous chapter I investigated the role of facilitation and its parameters on closed-loop target tracking and pursuit. My results in Chapter 2 show that more sluggish facilitation kinetics can be beneficial in more cluttered backgrounds. It also enhances the ability to 'attend' to one target in the presence of distracters. However, it is still unclear how this facilitation time constant affects the duration of the pursuit. Although there is a large degree of overlap between the publication in the current chapter and the former one (Chapter 2), here I investigate the effect of facilitation both on efficiency and efficacy of target tracking. I propose a new metric to quantify the trade-off between efficiency and efficacy of the model which represent both the ability of the model and energy expended by a pursuer in capturing a prey successfully. Further details of this metric are provided in Appendix F.

Statement of Authorship

Title of Paper	Performance Assessment of an Insect-Inspired Target Tracking Model in Background Clutter.
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	13th International Conference on Control Automation Robotics & Vision (ICARCV), 2014, Singapore, pp. 822-826. DOI: 10.1109/ICARCV.2014.7064410

Principal Author

Name of Principal Author (Candidate)	Zahra Bagheri
Contribution to the Paper	Developed the code, designed the experiments, ran the model simulations, analysed and interpreted the data and drafted the manuscript with editing contributions from the other authors.
Overall percentage (%)	55
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 6/7/2017

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Steven Wiederman
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, assisted with analysis and interpretation of the data, provided significant editing contribution (20%).
Signature	Date 6/7/2017

Name of Co-Author	Benjamin Cazzolato
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, assisted with analysis and interpretation of the data, provided editing contribution (10%).
Signature	Date 6/7/17

Name of Co-Author	Steven Grainger		
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, read and approved the draft (5%).		
Signature		Date	6/7/17

Name of Co-Author	David O'Carroll		
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, assisted with analysis and interpretation of the data, provided editing contribution (10%).		
Signature		Date	15-NOV-2016

Performance Assessment of an Insect-Inspired Target Tracking Model in Background Clutter

Zahra Bagheri, Steven D. Wiederman, Benjamin S. Cazzolato, Steven Grainger,
The University of Adelaide, Adelaide, Australia.

David C. O'Carroll,
Department of Biology, Lund University, Lund, Sweden

Published: Bagheri Z., Wiederman S. D., Cazzolato B. S., Grainger S., & O'Carroll D. C. (2014). Performance assessment of an insect-inspired target tracking model in background clutter. In 13th International Conference on Control Automation Robotics & Vision, IEEE, 822-826.

3.1 Abstract

Biological visual systems provide excellent examples of robust target detection and tracking mechanisms capable of performing in a wide range of environments. Consequently, they have been sources of inspiration for many artificial vision algorithms. However, testing the robustness of target detection and tracking algorithms is a challenging task due to the diversity of environments for applications of these algorithms. Correlation between image quality metrics and model performance is one way to deal with this problem. Previously we developed a target detection model inspired by physiology of insects and implemented it in a closed loop target tracking algorithm. In the current paper we vary the kinetics of a salience-enhancing element of our algorithm and test its effect on the robustness of our model against different natural images to find the relationship between model performance and background clutter.

3.2 Introduction

A challenging problem for autonomous and robotic applications is the development of robust artificial vision systems that can detect and pursue moving targets within cluttered natural environments. Insects, such as dragonflies, have evolved a solution to this problem and are capable of tracking small prey or conspecifics in cluttered, natural environments. Dragonflies, despite their low spatial acuity (~1°) and tiny brain, have a high successful capture rate (97%) (Olberg et al., 2000). Dragonflies are thus an ideal animal model to draw inspiration for target-tracking algorithms, and have motivated extensive investigation into the neuronal system that underlies pursuit behaviour.

'Small target motion detector' (STMD) neurons in the dragonfly's lobula are size selective, velocity tuned and contrast sensitive (O'Carroll, 1993). They respond robustly to small targets moving against cluttered backgrounds (Nordström et al., 2006; Nordström and

O'Carroll, 2009). We developed a target detection computational model (MATLAB/Simulink), inspired directly from electrophysiological recordings from STMD neurons (Wiederman et al., 2008; Wiederman et al., 2010, Wiederman and O'Carroll, 2011). This 'elementary small target motion detector' (ESTMD) model effectively provides a highly nonlinear matched spatiotemporal filter for the detection of small moving targets in natural scenery. This model simulates the properties of STMD neurons. Unlike most engineering algorithms, the ESTMD model does not rely on relative motion between target and background for target discrimination (Wiederman et al., 2008). We recently developed a discrete-time implementation of this model and implemented it in a closed-loop control algorithm to simulate the pursuit of small targets within a virtual-reality arena (Halupka et al., 2011).

Recent studies on dragonflies reveal that one type of STMD neuron, CSTMD1, exhibits a facilitatory mechanism in tracking targets. Electrophysiological recordings show that the spiking activity of CSTMD1 builds over time in response to targets traversing in long, continuous trajectories (Nordström et al., 2011). Responses to stimuli moving in interrupted paths show that they reset to a naive state when there are breaks (~7°) in the trajectory path (Dunbier et al., 2012). This facilitatory mechanism can enhance the response to weak stimuli and direct attention to the estimated reappearance location of the object. This increases the robustness of pursuit even if the target is temporarily occluded.

We recently showed that inclusion of a simple form of slow facilitation in a dark-target selective ESTMD model, based on known physiological properties of STMDs, enhances detection and pursuit of dark contrasting targets (Halupka et al., 2013; Wiederman and O'Carroll, 2013). However, the robustness of a target-tracking algorithm requires extensive examination under different scenarios. One way to avoid this problem is to correlate the image features with the model performance and simplify the prediction of model behaviour.

Here we test the efficiency and efficacy of our bio-inspired model when pursuing targets of varying contrast against different natural images and measure the relationship between robustness and clutter of the background images.

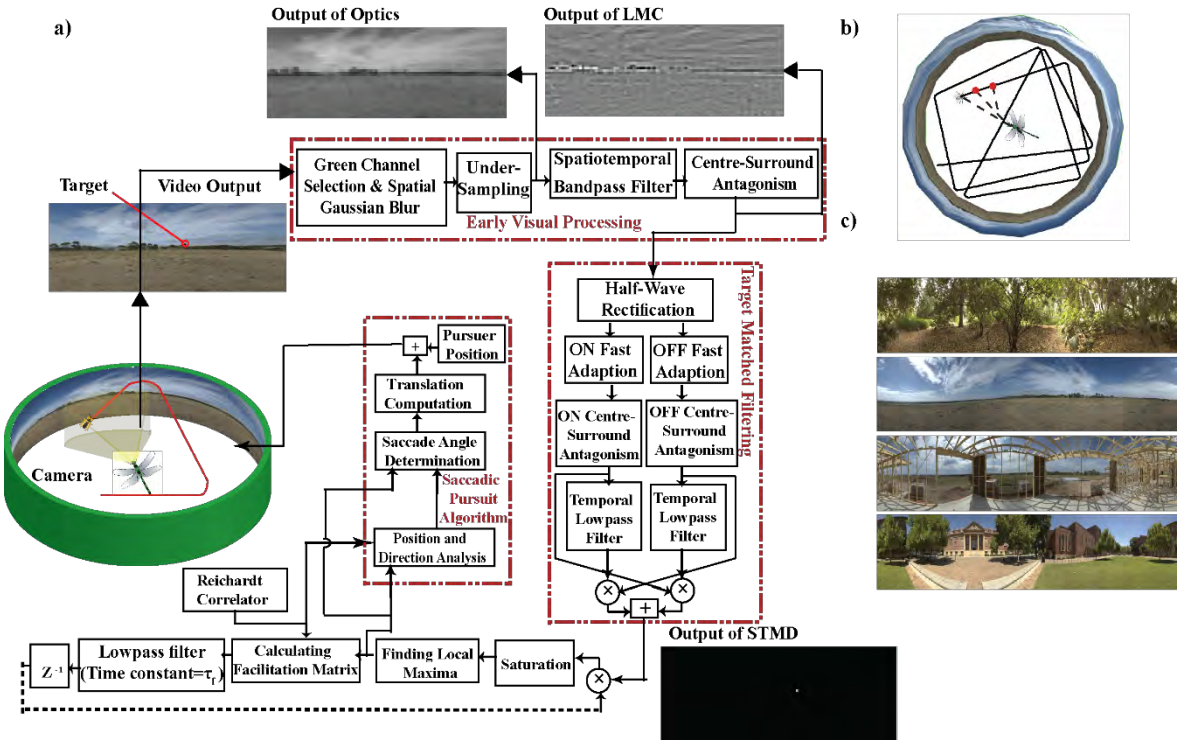

Figure 3.1. a) Overview of the closed-loop block diagram of the computational model for simulation and the output of each stage. We varied the facilitation low-pass filter time constant (τ_f) to find its effect on model performance. b) Plan view of a model of random target trajectory. c) Background Images: from top to bottom shows the rendered images A, B, C, and D respectively.

3.3 Methods

Figure 3.1 shows an overview of the computational model implemented in MATLAB/Simulink, which simulates the pursuit of small targets moving against cluttered backgrounds.

3.3.1 Virtual World Arena

We used Simulink 3D animation toolbox (Mathworks Inc.) to build a Virtual Reality (VR) arena as the front-end for our bio-inspired target detection and pursuit control algorithm. The pursuer chased the target within a cylindrical arena (of radius 6 m) rendered with natural images (see Brinkworth and O'Carroll (2008) for details of image acquisition). We tested the model with targets moving in randomized 3-dimensional paths with biologically plausible constraints on saccadic turn angle (Schilstra and van Hateren, 1999). If prey approached to within 50 cm of the cylinder wall, we initiated a random turn away from the VR boundary (Figure 3.1b).

Four target intensities from black to white were examined against each image to vary target contrast in different simulations. We simulated pursuits against four different panoramic natural scenes (Images A, B, C, D in Figure 3.1c). Although all of these images had 1/f power spectra, a statistical property of natural images (Field, 1987), they varied in the amount of background 'clutter' in the scene (Figure 3.1c). The values of target intensities along with mean intensity of each background are listed in Table 3.1.

We tested five different target-pursuer velocity ratios (|Vt|/|Vp|, where the subscripts t and p represent the target and pursuer respectively) ranging from 0.5 to 2 with the 'pursuer' moving at a constant velocity of 8 ms^{-1}. The start location of the 40 mm sized 'target' was at least 4 m away from the pursuer which yields an initiated target angular size of less than 0.6°. Pursuit simulations were repeated 50 times. Video was sampled from a 40°x98° sized viewport to represent the visual field of the pursuer and thus served as inputs to the detection, facilitation and pursuit algorithm.

Table 3.1. Target intensities used for simulations against different backgrounds and the mean intensity of each background.

Background	**Target Intensities (green channel, 8-bit)**				**Mean of Background (green channel, 8-bit)**
	Int 1	**Int 2**	**Int 3**	**Int 4**	
Image A	0	25	204	255	92
Image B	51	77	230	255	130
Image C	0	51	204	255	110
Image D	0	25	204	230	98

3.3.2 Insect_Inspired Target Discrimination Model

To emulate green sensitivity of the insect eye (Srinivasan and Guy, 1990) and optical blur of the insect compound eye (Stavenga, 2003), this model selects only the green channel of the RGB input image and applies a Gaussian blur (full-width at half maximum of 1.4°). The blurred image is sampled at 1° separation, which represents the resolution of the fly eye (Straw et al., 2006). The output of this stage (Figure 3.1a), is considered as the 'model input' to the target detection algorithm in further analyses.

The next stages of the insect visual pathway are the photoreceptors and then Large Monopolar Cells (LMCs). It is known that the photoreceptor responses are temporally limited (Laughlin and Weckström, 1993) and LMCs remove redundant information by acting as spatiotemporal contrast detectors (bandpass filters) (Coombe et al., 1989). In the ESTMD algorithm, the temporal properties of photoreceptor and LMC are modelled with a discrete version (Halupka et al., 2011) of a log-normal function (James, 1990) which provides a good match to the temporal impulse response. Then spatial high-pass filtering representing centre-surround antagonism is applied on the output image.

The output of early visual processing (Figure 3.1) is half-wave rectified to imitate the independent ON and OFF channels in insect vision (Wiederman et al., 2008). Each independent channel is processed via a fast adaptive mechanism. The fast adaptive mechanism is modelled by using a fast low-pass filter (τ=3 ms) when the input signal increases, and a slow low-pass filter (τ=70 ms) when it decreases. This adaptation process serves to inhibit repeating bursty inputs, such as noise. Both of the ON and OFF channels then undergo further strong centre-surround antagonism, selectively tuning the model to targets with small angular extent (orthogonal to the direction of travel). Sensitivity to both dark and light targets is provided by delaying and multiplying each contrast channel (ON or OFF) with a delayed version of the opposite polarity (delayed using a low-pass filter, τ=25 ms). This also conveys selectivity for objects that are small in the dimension matching the direction of travel, since a small feature will usually be characterized by an initial rise (or fall) in brightness at each point that it passes across, followed a corresponding fall (or rise) after a short delay. The output image undergoes non-linear saturation using a hyperbolic tangent function. This serves to ensure all signals lie between 0 and 1. Then the maximum is determined as the target.

3.3.3 Target Tracking Algorithm

The pursuit strategy implemented was ‘saccadic tracking’ as observed in male houseflies (Wehrhahn et al., 1982; Land and Collett, 1974). Saccadic turn angles are calculated in order to keep the target in the central axis of the pursuer’s gaze. Re-centring towards the target was initiated when the winning feature in ESTMD output moved 5°. This strategy promotes target ‘pop-out’ by permitting the spatiotemporal filters to ‘fade away’ (high-pass) the more distant background. A pursuit was considered successful only if the pursuer was within 1.2 m proximity of the target (the initial distance was at least 4 m) in less than 2 s and the target

was the winner in the output for more than 50% of the last 6 ms of the tracking, thus excluding detections from fortuitous saccades.

3.3.4 Reichardt Correlator

A Reichardt correlator is a biologically inspired model for 'elementary motion detectors' (EMDs) which produces a directionally selective response to motion (Hassenstein and Reichardt, 1956). The output of an EMD uses two spatially separated contrast signals and correlates them after a delay (via a low-pass filter). Here, we cascaded ESTMD with an EMD to maintain the core STMD properties but with direction-selective outputs (Wiederman and O'Carroll, 2013).

3.3.5 Facilitation Mechanism

We previously implemented a facilitation mechanism inspired by physiological experiments (Halupka et al., 2013). This mechanism enhanced a region near to the current location of maximum model output. The region was calculated for a single state (position) with a spatial offset predicting the next location of the target. Nonetheless, in the current version, the output of the EMD is thresholded and used to estimate a target velocity range. The output of ESTMD-EMD cascade is then used to estimate the next location of the target, building a weighted 'map' dependent on the location of the winning feature but offset in the direction of target movement. We multiplied the ESTMD output with a low-pass filtered version of this 'facilitation map' to model responses observed in the dragonfly neuron, CSTMD1 (Nordström et al., 2011; Dunbier et al., 2012). The role of the low-pass filter time constant here is to control the kinetics with which the facilitation matrix enhances the area around the winning feature. We varied this low-pass filter time constant in the range from 40 to 2000 ms (nine different values) to examine its effect on the model efficiency and efficacy.

3.3.6 Evaluation

3.3.6.1 Clutter Measure

To measure background clutter, we used the metric developed by Silk (1995). This method measures image clutter by convolving the image with an average kernel, which was chosen to be the same approximate size as the target (Ralph et al., 2006). The clutter value is calculated by the following formula:

$$CM = {}^{1}/_{N} \left(\sum_i \sum_j \left(b_{i,j} - \bar{B}_{i,j} \right)^2 \right)^{1/2} \tag{3.1}$$

where b is the value of the i,j pixel, $\bar{B}$ is the mean of the box centered at pixel i,j and N is the number of boxes convolved over the whole image.

3.3.6.2 Target Discriminability Measure

To determine the discriminability of the target at the output stage, we defined the following metric:

$$Discrim = \frac{M - N}{M} e^{(I_t - I_{max} - \sigma_{I_b})} \tag{3.2}$$

where M is the total number of background pixels, N represents the number of background pixels with equal or higher values than the target, I_t is the target intensity, I_{max} is the maximum intensity of the background pixels, and σ_{I_b} is the deviation of background pixels with higher value than the target value, given by:

$$\sigma_{I_b} = \sqrt{\sum_{I_i = I_t}^{I_{max}} n_i (I_t - I_i)^2} \tag{3.3}$$

where n_i is the number of pixels with the intensity of I_i. Based on this metric, the maximum possible discriminability is e, and whenever the target is the winner at the output, the discriminability value is greater than 1.

3.4 Results

We ran 40,000 simulations with and without facilitation against four different natural images to determine whether the inclusion of facilitation enhances robustness and efficiency of the model. For this purpose, we varied model parameters, such as: facilitation low-pass filter time constant (nine values), target velocity (five values), and target intensity (four values) in these simulations.

3.4.1 Effect of Facilitation Time Constant on Efficiency and Efficacy of the Model

We tested the effect of varying the facilitation time constant on model efficiency (pursuit duration) and efficacy (pursuit success) for each of the background images. These metrics represent both the ability of the model and energy expended by a pursuer in capturing a prey successfully. The distributions of capture times of the target moving at a velocity of 6 ms^{-1} for four facilitation time constants are illustrated in Figure 3.2. This figure shows that the distribution of capture time has distinct behaviour in response to variation of time constant in different backgrounds. For instance, in the heavily cluttered Image A longer time constants (667 ms and 1538 ms) lead to increased frequency of faster pursuit success. In Image B this happens with shorter time constants.

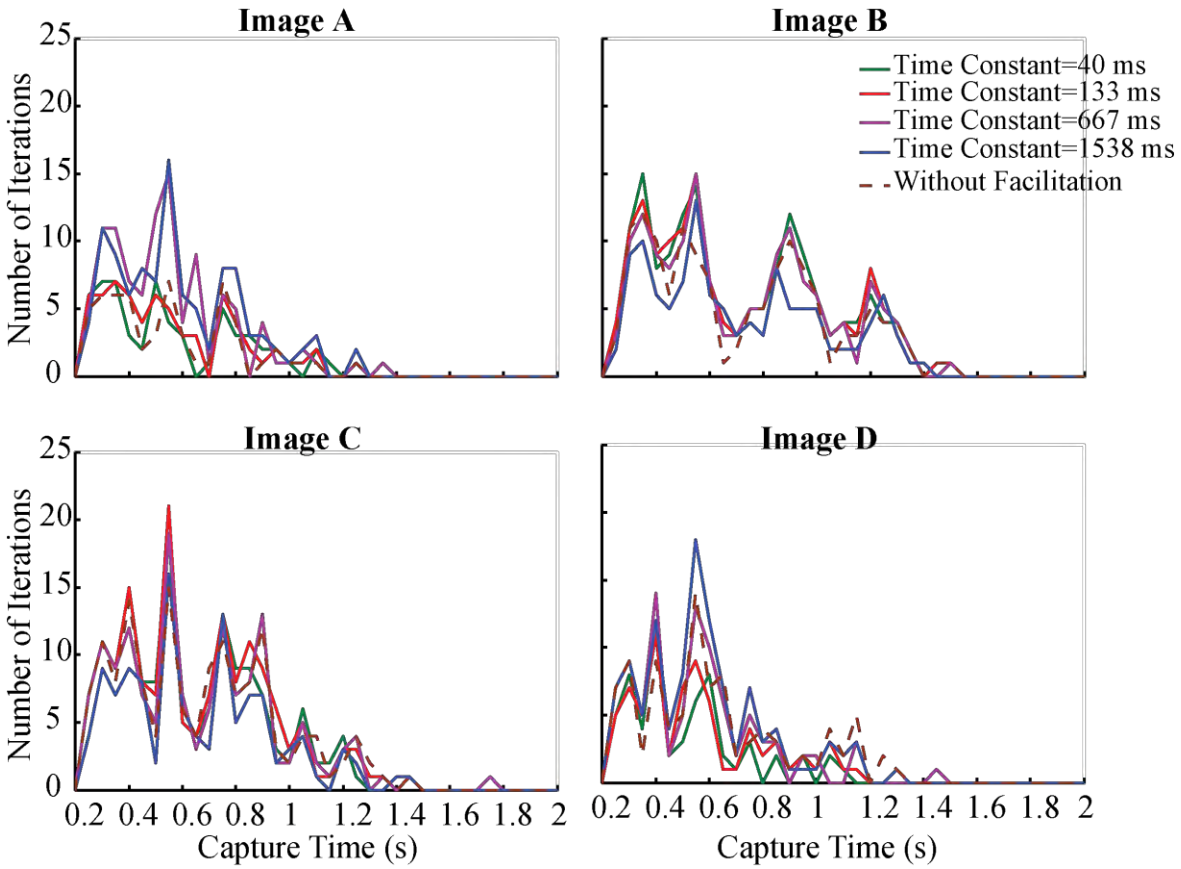

Figure 3.2. Distribution of capture time for simulations with target-pursuer velocity ratio ($|V_t|/|V_p|$) of 3/4.

A metric to represent pursuit success (combining both efficiency and efficacy) at different time constants, was calculated as the first moment of area around the axis perpendicular to x=2 and y=0 (which we refer to this axis as z') in Figure 3.2. For this purpose, the x-axis is normalized by the maximum simulation time (2 s) and the y-axis is divided by the total number of simulations for each time constant (200 simulations), giving a performance metric of

$$I_{z'} = \sum \left(\sqrt{\left(\frac{2-x}{2}\right)^2 + \left(\frac{y}{200}\right)^2}\, \Delta A \right) \tag{3.4}$$

where $I_{z'}$ is the first moment of area around z', and ΔA represents the area of each small element under the curve.

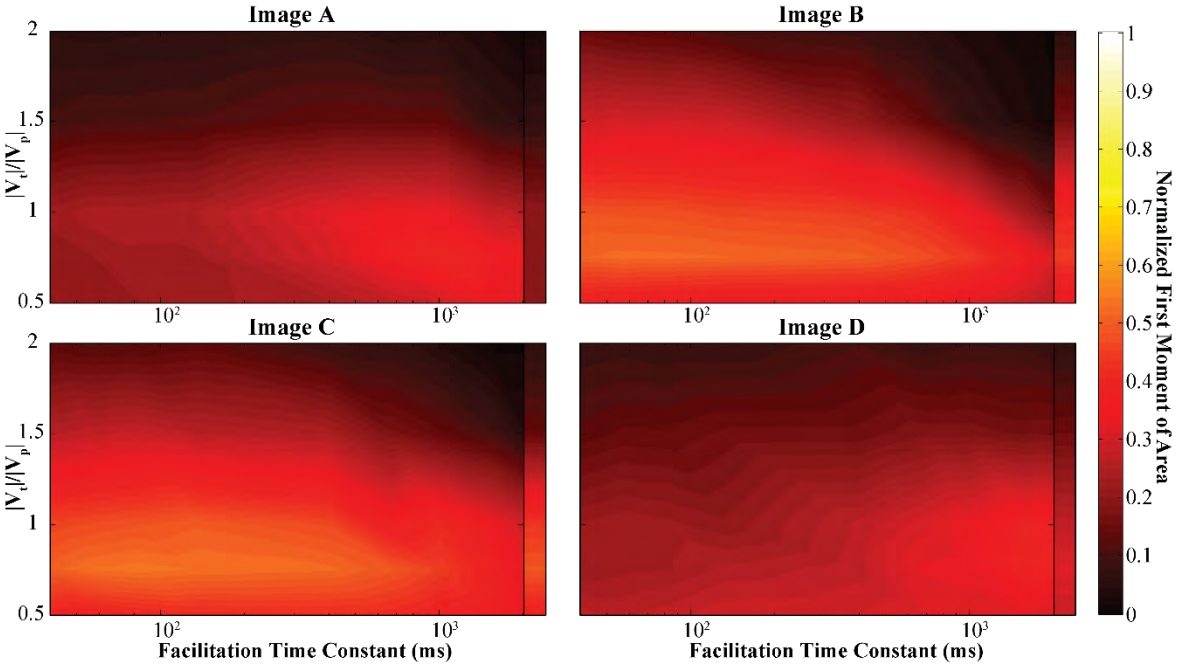

Figure 3.3. Calculated first moment of area of the capture time distribution around z' with respect to time constant and target-pursuer velocity ratio ($|V_t|/|V_p|$). The column on the right side of each plot shows the result of simulations without facilitation. Images A to D are shown in Figure 3.1c.

Figure 3.3 shows the values of $I_{z'}$ (pursuit efficiency and efficacy) at varying target-to-pursuer velocity ratios and facilitation low-pass filter time constant for all four images. The column in the right-hand side of each plot shows the results of simulations without facilitation.

Results reveal that model performance varied across different background images. As expected, success rate decreases dramatically as target velocity exceeds that of the pursuer. However, in all cases the addition of facilitation increases the performance of the model. Interestingly, each background image has a different optimum for the facilitation time constant. We hypothesized that this was due to a relationship between the facilitation time constant (τ_f in Figure 3.1) and the amount of background clutter in an image.

3.4.2 Optimum facilitation time constant across clutter

We used the results of 7,200 simulations with target-pursuer velocity ratio ($|Vt|/|Vp|$) of 3/4 to examine whether the optimum facilitation time constant is related to the degree of background clutter. We used the following formula to define robustness of the model facilitation for each particular mean background clutter and time constant value:

$$Robustness = \frac{\sum_{k=1}^{K} \overline{Discrim_k}}{\sum_{j=1}^{J} Max\ possible\ Discrim} \times 100 \qquad (3.5)$$

where J represents the total number of simulations of each dataset, K is the number of successful simulations of each data set, $\overline{Discrim_k}$ is the average discriminability during the pursuit calculated by (Eq. 3.2).

Figure 3.4 presents the calculated robustness of the pursuit at varying facilitation time constants and background clutter values. This figure clearly shows that the optimum facilitation time constant shifts towards longer time constants as clutter of the background increases. This trend makes intuitive sense since in more cluttered backgrounds the target is camouflaged more often. Consequently, a longer time constant enables the facilitation mechanism to enhance the area of target disappearance for a longer time, permitting the discrimination of the target again when it reappears - a simple form of predictive salience enhancement. The results show that although the robustness of the model decreases as the background clutter increases, the addition of facilitation can effectively increase the robustness of the model across different background images.

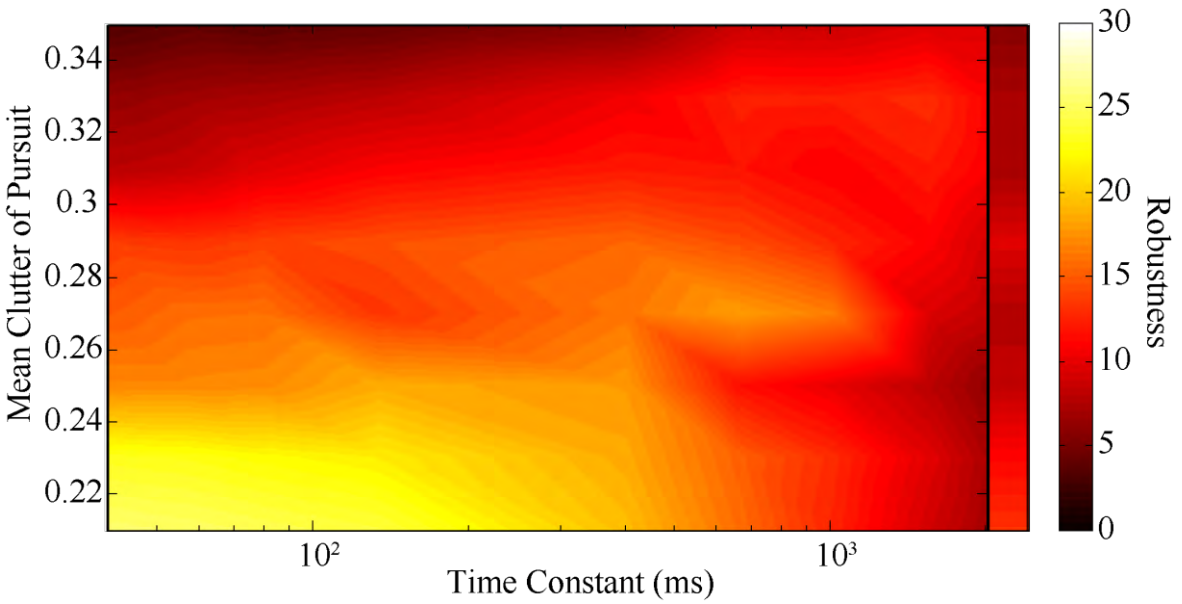

Figure 3.4. Robustness versus facilitation time constant and average clutter of the pursuit. The column on the right hand side of the plot shows the results of simulations without facilitation.

3.5 Discussion

The insect inspired ESTMD model provides a highly nonlinear spatiotemporal 'target matched' filtering which can detect small moving objects robustly, even against natural cluttered backgrounds. The model is improved with the addition of a recently described physiological phenomenon of facilitation - in effect, a bio-inspired estimation technique.

Here, we tested the robustness of the model with respect to variation of background clutter. Our results show that in more cluttered backgrounds the success of detection and pursuit of targets decreases, due to difficulty of pursuit. Nevertheless, by altering the facilitation time constant the model performs reliably in highly cluttered environments. Although we have modelled a simple facilitation mechanism, we have showed that it can effectively improve both efficacy and efficiency of the target pursuit against cluttered, complex backgrounds.

Acknowledgments

This research was supported by the Australian Research Council's Discovery Projects funding scheme (project number DP130104572).

References

Brinkworth R. S., & O'Carroll D. C. (2009). Robust models for optic flow coding in natural scenes inspired by insect biology. PLoS Computational Biology, 5(11), e1000555.

Coombe P. E., Srinivasan M. V., & Guy R. G. (1989). Are the large monopolar cells of the insect lamina on the optomotor pathway? Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 166(1), 23-35.

Dunbier J. R., Wiederman S. D., Shoemaker P. A. & O'Carroll D. C. (2012). Facilitation of dragonfly target-detecting neurons by slow moving features on continuous paths. Frontiers in Neural Circuits, 6, 79.

Field D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. Journal of the Optical Society of America A, 4(12), 2379-2394.

Halupka K. J., Wiederman S. D., Cazzolato B. S., & O'Carroll D. C. (2013). Bio-inspired feature extraction and enhancement of targets moving against visual clutter during closed loop pursuit. In IEEE International Conference on Image Processing, 4098-4102.

Halupka K. J., Wiederman S. D., Cazzolato B. S., & O'Carroll D. C. (2011). Discrete implementation of biologically inspired image processing for target detection. In IEEE Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 143-148.

Hassenstein, B. & Reichardt W. (1956). Systemtheoretische analyse der zeit-, reihenfolgen- und vorzeichenauswertung bei der bewegungsperzeption des rüsselkäfers chlorophanus. Zeitschrift für Naturforschung B, 11(9-10), 513-524.

James A.C. (1990). White-noise studies in the fly lamina. Doctoral dissertation, Australian National University, Canberra, Australia.

Land M. F., & Collett T. S. (1974). Chasing behaviour of houseflies (Fannia canicularis). Journal of Comparative Physiology, 89(4), 331-357.

Laughlin S. B., & Weckström M. (1993). Fast and slow photoreceptors—a comparative study of the functional diversity of coding and conductances in the Diptera. Journal of Comparative Physiology A, 172(5), 593-609.

Nordström K., & O'Carroll D. C. (2009). Feature detection and the hypercomplex property in insects. Trends in Neurosciences, 32(7), 383-391.

Nordström K., Bolzon D. M., & O'Carroll D. C. (2011). Spatial facilitation by a high-performance dragonfly target-detecting neuron. Biology Letters, 7(4), 588-592.

Nordström K., Barnett P. D., & O'Carroll D. C. (2006). Insect detection of small targets moving in visual clutter. PLoS Biology, 4(3), 4:378-386.

O'Carroll D.C. (1993). Feature-detecting neurons in dragonflies. Nature, 362, 541-543.

Olberg R. M., Worthington A. H., & Venator K. R. (2000). Prey pursuit and interception in dragonflies. Journal of Comparative Physiology A, 186(2), 155-162.

Ralph S. K., Irvine J., Snorrason M., & Vanstone S. (2006). An image metric-based ATR performance prediction testbed. In 35th IEEE Applied Imagery and Pattern Recognition Workshop, pp. 33-40.

Schilstra C., & van Hateren J. (1999). Blowfly flight and optic flow. I. Thorax kinematics and flight dynamics. Journal of Experimental Biology, 202(11), 1481-1490.

Silk J. D. (1995). Statistical variance analysis of clutter scenes and applications to a target acquisition test. Institute for Defence Analysis, Alexandria, VA.

Srinivasan M. V., & Guy R. G. (1990). Spectral properties of movement perception in the dronefly Eristalis. Journal of Comparative Physiology A, 166(3), 287-295.

Stavenga D. (2003). Angular and spectral sensitivity of fly photoreceptors. I. Integrated facet lens and rhabdomere optics. Journal of Comparative Physiology A, 189(1), 1-17.

Straw A. D., Warrant E. J., & O'Carroll D. C. (2006). A bright zone in male hoverfly (Eristalis tenax) eyes and associated faster motion detection and increased contrast sensitivity. Journal of Experimental Biology, 209(21), 4339-4354.

Wehrhahn C., Poggio T., & Bülthoff H. (1982). Tracking and chasing in houseflies (Musca). Biological Cybernetics, 45(2), 123-130.

Wiederman S. D., & O'Carroll D. C. (2013). Biomimetic target detection: Modeling 2nd order correlation of OFF and ON channels. In IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP), 16-21.

Wiederman S. D., & O'Carroll D. C. (2011). Discrimination of features in natural scenes by a dragonfly neuron. The Journal of Neuroscience, 31(19), 7141-7144.

Wiederman S. D., Shoemaker P. A., & O'Carroll D. C. (2013). Correlation between OFF and ON channels underlies dark target selectivity in an insect visual system. The Journal of Neuroscience, 33(32), 13225-13232.

Wiederman S., Brinkworth R. S., & O'Carroll D. C. (2010). Performance of a bio-inspired model for the robust detection of moving targets in high dynamic range natural scenes. Journal of Computational and Theoretical Nanoscience, 7(5), 911-920.

Wiederman S. D., Shoemaker P. A., & O'Carroll D. C. (2008). A model for the detection of moving targets in visual clutter inspired by insect physiology. PloS One, 3(7), e2784.

Chapter 4. Performance of an Insect-Inspired Target Tracker in Natural Conditions

In previous chapters I examined the performance of the insect-inspired target tracking model in closed-loop natural simulations. Our rationale behind the efficiency of the insect-inspired algorithms is the insect's miniature brain which evolved over millions of years to consume tiny amounts of power compared with even the most efficient digital processors. Despite this apparent efficiency however, there are very few studies that directly compare insect-inspired systems with engineering algorithms. Just because a brain is small, it is not necessarily simple – neuronal networks often implement complex and highly non-linear processing. The question remains whether insect-inspired algorithms are really useful alternatives for computer vision and robotic applications? In this chapter I address this question. I directly compare the insect-inspired tracker with state-of-the-art engineering solutions to test their computational efficiency and efficacy. The preliminary results of this work were published in:

"Bagheri Z. M., Wiederman S. D., Cazzolato B. S., Grainger S., & O'Carroll D. C. (2015). Robustness and Real-Time Performance of an Insect Inspired Target Tracking Algorithm Under Natural Conditions. In IEEE Symposium Series on Computational Intelligence, 97-102."

which is presented in Appendix C. The supplementary material for the current chapter is provided in Appendix D and the STNS dataset is presented in Appendix G.

Statement of Authorship

Title of Paper	Performance of an Insect-Inspired Target Tracker in Natural Conditions
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Bagheri, Z. M., Wiederman, S. D., Cazzolato, B. S., Grainger, S., & O'Carroll, D. C. (2017). Performance of an insect-inspired target tracker in natural conditions. Bioinspiration & Biomimetics, 12(2), 025006.

Principal Author

Name of Principal Author (Candidate)	Zahra Bagheri
Contribution to the Paper	Developed the code, designed the experiments, gathered videos, generated groundtruth, ran the model simulations, analysed and interpreted the data, drafted the manuscript with editing contributions from the other authors.
Overall percentage (%)	65
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date: 6/7/2017

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Steven Wiederman
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, assisted with analysis and interpretation of the data, provided significant editing contribution (15%).
Signature	Date: 6/7/2017

Name of Co-Author	Benjamin Cazzolato
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, assisted with analysis and interpretation of the data, provided editing contribution (10%).
Signature	Date: 6/7/17

Name of Co-Author	Steven Grainger		
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, read and approved the draft (5%).		
Signature		Date	6/7/17

Name of Co-Author	David O'Carroll		
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, provided editing contribution (5%).		
Signature		Date	3-JULY-2017

Performance of an Insect-Inspired Target Tracker in Natural Conditions

Zahra M. Bagheri[1,2], Steven D. Wiederman[1], Benjamin S. Cazzolato[2], Steven Grainger[2], David C. O'Carroll[1,3]

[1] Adelaide Medical School, The University of Adelaide, Adelaide, SA, 5005, Australia

[2] School of Mechanical Engineering, The University of Adelaide, Adelaide, SA, 5005 Australia

[3] Department of Biology, Lund University, Sölvegatan 35, S-22362 Lund, Sweden

Published: Bagheri Z., Wiederman S. D., Cazzolato B. S., Grainger S., & O'Carroll D. C. (2014). Performance of an Insect-Inspired Target Tracker in Natural Conditions. Bioinspiration & biomimetics, 12(2), 025006.

4.1 Abstract

Robust and efficient target-tracking algorithms embedded on moving platforms, are a requirement for many computer vision and robotic applications. However, deployment of a real-time system is challenging, even with the computational power of modern hardware. As inspiration, we look to biological solutions - lightweight and low-powered flying insects. For example, dragonflies pursue prey and mates within cluttered, natural environments, deftly selecting their target amidst swarms. In our laboratory, we study the physiology and morphology of dragonfly 'small target motion detector' neurons likely to underlie this pursuit behaviour. Here we describe our insect-inspired tracking (IIT) model derived from these data and compare its efficacy and efficiency with state-of-the-art engineering models. For model inputs, we use both publicly available video sequences, as well as our own task-specific dataset (small targets embedded within natural scenes). In the context of the tracking problem, we describe differences in object statistics within the video sequences. For the general dataset, our model often locks on to small components of larger objects, tracking these moving features. When input imagery includes small moving targets, for which our highly nonlinear filtering is matched, the robustness outperforms state-of-the-art trackers. In all scenarios, our insect-inspired tracker runs at least twice the speed of the comparison algorithms.

Index Terms—Visual target tracking, bio-inspired vision, real-time.

4.2 Introduction

Real-time target tracking is an important component of computer vision and robotic applications, employed in the fields of surveillance, human-computer interaction, intelligent transportation systems and human assistance mobile robots. However, this task is complicated by the diverse requirements that must be addressed in one computationally effective algorithm. Targets must be tracked with overall illumination changes, background clutter, rapid changes in target appearance, partial or full occlusion, non-smooth target trajectory and ego-motion.

Every tracker requires a description of the target, based on features such as gradient (Bay et al., 2006; Felzenszwalb et al., 2008; Gall and Lempitsky, 2013), colour (Abdel-Hakim and Farag, 2006; Burghouts and Geusebroek, 2009), texture (Ojala et al., 2002; Chen et al., 2010), spatiotemporal pattern (Scovanner et al., 2007; Zhao and Pietikainen, 2007), or a combination of these. However, irrespective of the descriptor quality, adaptive mechanisms must be employed to account for variation of the target's appearance throughout the pursuit. These adaptive, online algorithms can be formulated in two different categories; generative and discriminative.

Generative algorithms search for a target location which best matches the appearance model (Black and Jepson, 1998; Comaniciu et al., 2003; Ross et al., 2008; Porikli et al., 2006). One limitation is that they require numerous samples over successive frames. With only a few samples at the outset, most generative trackers assume target appearance does not change significantly during the training period. Discriminative trackers use both target and background information to build a binary classifier (Kalal et al., 2012; Babenko et al., 2011; Hare et al., 2011; Zhang et al., 2012). The classifier searches a local region constrained by target motion to determine a decision boundary for separating target from background.

Whilst discriminative methods tend to be noise sensitive, generative methods can fail within cluttered backgrounds (Yang et al., 2011).

Despite the high efficiency of online trackers, each update during run-time can introduce error in the target model. This cumulative error usually arises from uncertainty in object location or target occlusion, with drift resulting in tracking failure. State-of-the-art trackers use techniques such as robust loss functions (Leistner et al., 2009; Masnadi-Shirazi et al., 2010), semi-supervised learning (Chapelle et al., 2006; Zhu and Goldberg, 2009; Grabner et al., 2008; Saffari et al., 2010), multiple-instance learning (Babenko et al., 2011; Zeisl et al., 2010), and co-training (Blum and Mitchell., 1998; Javed et al., 2005; Levin et al., 2003; Kalal et al., 2012) to improve labelled samples and reduce drift during run-time.

Following object representation, object tracking involves a search process for inferring target trajectory from uncertain and ambiguous observations of states such as, position, velocity, scale, and orientation. The Kalman filter (Bar-Shalon and Fortmann, 1988) and its variations such as the Extended Kalman filter (EKF) (Bar-Shalon and Fortmann, 1988) and the Unscented Kalman filter (UKF) (Li et al., 2004) are extensively used in target tracking to find the optimal solution for target states. These methods model observation uncertainties by Gaussian processes which may not always be appropriate. For example, measurement distributions for target tracking within cluttered environments may not be unimodal Gaussian. Particle or Sequential Monte Carlo filters (Kitagawa, 1987) have been proposed to address this problem, by maintaining a probability distribution over the state of the object being tracked with a set of weighted samples. With a sufficient number of particle samples, these filters account for nonlinear target motion and non-Gaussian noise. However, as particle number increases exponentially with the number of states, computational load is a concern.

Considering the complexity of the target detection and tracking task, it is intriguing to observe the accuracy, efficiency and adaptability of biological visual systems. Robust target tracking behaviour is seen in seemingly simple animals, such as insects, with brain sizes measured in millimetres. The dragonfly selects and chases prey or conspecifics within a cluttered surround even in the presence of distracting stimuli (Corbet, 1999; Wiederman and O'Carroll, 2013) with a success rate over 97% (Olberg et al., 2000). This task is performed despite their limited visual acuity (~0.5°) and relatively small size, light-weight and low-power neuronal architecture.

We determined key properties of this system using intracellular, electrophysiological techniques to record from 'small target motion detector' (STMD) neurons. STMDs are size and velocity tuned and are sensitive to target contrast. A subset of STMDs respond even without relative motion between the target and a cluttered background (O'Carroll, 1993; Nordström et al., 2006; Nordström and O'Carroll, 2009; O'Carroll and Wiederman, 2014, O'Carroll et al., 2011; Wiederman and O'Carroll, 2011). Inspired directly by these physiological data, we developed an algorithm for local target discrimination based on an 'Elementary-STMD' (ESTMD) operation at each point in the image (Wiederman et al., 2008). This nonlinear model provides a matched spatiotemporal filter for small moving targets embedded within natural scenery (Wiederman et al., 2008). Recently, we elaborated this model (Halupka et al., 2013; Bagheri et al., 2014a; Bagheri et al., 2014b; Bagheri et al., 2015) to include a property observed in CSTMD1 (an identified STMD) termed 'facilitation' (Nordström et al., 2011; Dunbier et al., 2011; Dunbier et al., 2012), which accounts for the slow build-up in neuronal responses to targets that move in long continuous trajectories. We implemented this model in a closed-loop target tracking system within a virtual reality (VR) environment. We included an active saccadic gaze fixation strategy inspired by observations of insect pursuits (Halupka et al., 2011, 2013; Bagheri et al., 2014a, 2014b, 2015). We have

shown that facilitation not only substantially improves success for short-duration pursuits, it enhances 'attention' to one target in the presence of distracters (Bagheri et al., 2015). Facilitation may thus contribute to selective attention observed in the CSTMD1 neuron, which tracks a single target in the presence of a distracter (Wiederman and O'Carroll, 2013). This model shows robust performance with high prey capture success even within complex background clutter, low contrast and high relative speed of pursued prey (Bagheri et al., 2015).

Having optimized model tuning, in this paper we turn to quantifying the effectiveness and efficiency of our insect-inspired approach. Firstly, we compare robustness with other trackers, testing them with natural challenges and non-idealities in the input imagery, such as local flicker and illumination changes, and non-smooth and non-linear target trajectories. Furthermore, the inspect-inspired tracker utilizes a number of highly non-linear processing stages. To investigate whether these come at a cost of processing efficiency, we compare processing time of our algorithm with other computer vision approaches. We test efficacy and efficiency with a widely-used set of videos recorded under natural conditions. We directly compare the performance of our model with several state-of-the-art algorithms using the same hardware, software environment and video inputs.

Even though the insect-inspired model is intentionally tuned to small moving objects (on the same scale as the resolution of the flying insect compound eye), we find that it often performs favourably in these environments. When tracking small moving targets in natural scenes, our model exhibits robust performance, outperforming the best of the tracking algorithms. In all scenarios, our model operates more efficiently than the other trackers. Given the specificity of the task (small target detection), we demonstrate the feasibility of applying this bioinspired model to real-time robotic and computer vision applications. Furthermore, these

results provide insight into how our model could be applied to more generalised, object-tracking tasks.

4.3 Methods

4.3.1 Computational Model

Figure 4.1 shows the insect-inspired tracker (IIT) overview, implemented in MATLAB. Figure 4.2 shows example output at model stages. The detection and tracking model is composed of three subsystems: (1) Early visual processing, (2) Target-matched filtering ESTMD (Elementary Small Target Motion Detector) stage and (3) Integration and facilitation.

4.3.1.1 Early visual processing

The compound eye of flying insects is limited by diffraction at thousands of facet lenses (Stavenga, 2003). The resulting optical blur is modelled by a Gaussian function (full-width at half maximum of 1.4°) (Stavenga, 2003):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{(-(x-\mu)^2/(2\sigma^2))}, \tag{4.1}$$

where μ and σ are the mean and standard deviation respectively. Noting that the full-width at half maximum is desired to be 1.4°, the standard deviation of the Gaussian is therefore:

$$\sigma = \frac{1.4}{2\sqrt{2\ln 2}}. \tag{4.2}$$

We model an inter-receptor angle between ommatidial units of 1° (Straw et al., 2006) and green spectral sensitivity by selecting the green channel of the RGB image (Srinivasan and Guy, 1990).

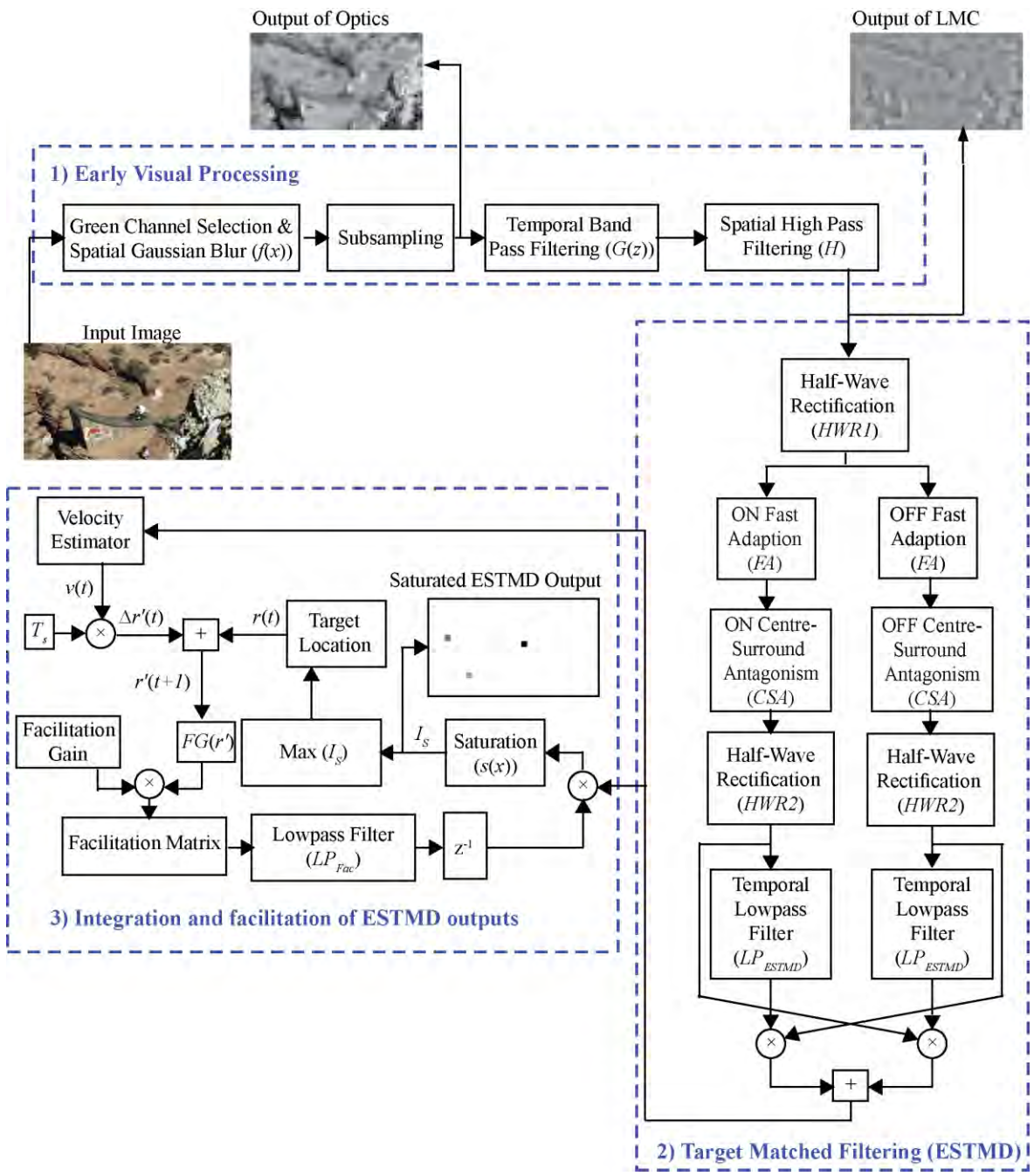

Figure 4.1. Model overview of the insect-inspired tracker (IIT). The model includes: 1) Early visual processing which simulates the spectral sensitivity, optical blur and low resolution sampling of the insect visual system. 2) Target matched filtering or Elementary Small Target Motion Detector (ESTMD) provides selectivity for small moving targets: separation of OFF and ON channels, fast temporal adaptation, strong surround antagonism and temporal correlation between opposite contrast polarity channels. 3) The facilitation mechanism which is observed in dragonfly CSTMD1 neurons was modelled by building a weighted map (FG(r'), supplementary material, Fig. S1) based on the predicted location of the target in the next sampling time (r'(t+1)). The predicted target location was calculated by shifting the location of the winning feature (r(t))

with an estimation of the target velocity vector (v(t)) provided by the Hassenstein-Reichardt elementary motion detector which was multiplied with sampling time (Ts). The facilitation mechanism multiplies the winning output of ESTMDs (maximum) with a delayed (z-1) version of a weighted map based on the current location of the winning feature but offset in the direction of the target's movement. The time constant of the facilitation low-pass filter controls the duration of the enhancement around the winning feature.

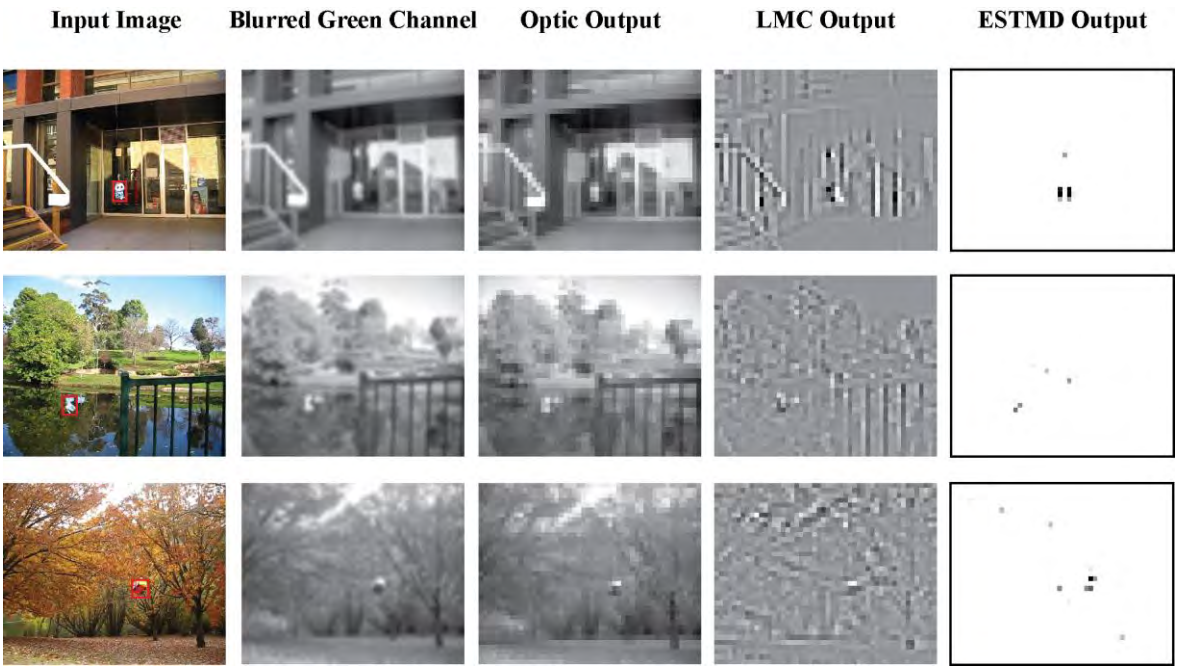

Figure 4.2. Output of model stages, including optics, large monopolar cells (LMC) and elementary small target motion detector (ESTMD). The red rectangle in the input image shows the bounding box of the target.

The large monopolar cells (LMCs) in the insect lamina remove redundant information by using neuronal adaptation (temporal high pass filtering) and centre-surround antagonism (spatial high pass filtering). The band-pass temporal properties of early visual processing (combining photoreceptors and LMCs) were simulated with a discrete log-normal transfer function (Halupka et al., 2011):

$$G(z) = \frac{\sum_{i=1}^{8} \alpha_i z^{(i-1)}}{z^8 + \sum_{j=1}^{8} \beta_j z^{(j-1)}}. \qquad (4.\ 3)$$

where $G(z)$ is the transfer function of the temporal filter in the z-domain (sampling time T_s=1 ms), i, j represent time index, α_i and β_j are the numerator and denominator coefficients of the

filter which are given in Table 4.1. The weak centre-surround antagonism was modelled by convolving the image with a kernel which subtracts 10% of the centre pixel from the nearest neighbouring pixels:

$$H = \begin{bmatrix} -1/9 & -1/9 & -1/9 \\ -1/9 & 8/9 & -1/9 \\ -1/9 & -1/9 & -1/9 \end{bmatrix}. \quad (4.4)$$

Table 4.1. Coefficients of the discrete log-normal function shown in Eq (4.3).

Numerator Coefficients		Denominator Coefficients	
α_1	-0.15240	β_1	0.06510
α_2	0.17890	β_2	-0.54180
α_3	-0.05740	β_3	2.14480
α_4	0.04390	β_4	-5.30600
α_5	-0.01700	β_5	9.00040
α_6	0.00520	β_6	-10.71100
α_7	-0.00110	β_7	8.68500
α_8	0.0001	β_8	-4.33300

4.3.1.2 Target matched filtering (ESTMD stage)

Rectifying transient cells (RTCs) of the insect medulla exhibit independent adaptation to light increments (ON channel) and decrements (OFF channel) (Osorio, 1991; Jansonius and J. van Hateren, 1991). The separation of the ON and OFF channel was modelled by half-wave rectification (*HWR1*):

$$HWR1 = \begin{cases} ON = \begin{cases} x \; if \; x > 0 \\ 0 \; if \; x \leq 0 \end{cases} \\ OFF = \begin{cases} -x \; if \; x < 0 \\ 0 \; if \; x \geq 0 \end{cases} \end{cases}. \quad (4.5)$$

The independent ON and OFF channels are processed through a fast-adaptive mechanism, with the state of adaptation determined by a nonlinear filter which switches its time constant dependent on whether the signal is increasing or decreasing (Wiederman et al., 2008;

Halupka et al., 2011) (Figure 4.3). Matched to the observed physiological properties, time constants were 'fast' (τ=3 ms) when channel input is increasing (depolarising) and 'slow' (τ=70 ms) when decreasing (hyperpolarising):

$$GC(I) = \begin{cases} \tau = 3 \; if \; I(t) - I(t-1) > 0 \\ \tau = 70 \; if \; I(t) - I(t-1) < 0 \end{cases}. \qquad (4.6)$$

where *GC* is the 'Gradient Check' function in Figure 4.3, τ is the time constant of the filter (first order), and *I* is the intensity of the pixel in the half-wave rectified channels ('*u*').

For each independent ON and OFF channel, this adaptation state subtractively inhibits the unaltered 'pass-through' signal. Therefore, in the presence of textual fluctuations, a novel ON or OFF contrast boundary is required to 'break-through' the adapted channel. Additionally, strong spatial centre-surround antagonism (CSA) was applied to each independent channel. Target size tuning is achieved by varying the gain and spatial extent of this centre-surround antagonism (Figure 4.4). A second half-wave rectification was applied to the output of the strong centre- surround antagonism to eliminate the negative values:

$$HWR2 = \begin{cases} x \; if \; x > 0 \\ 0 \; if \; x \leq 0 \end{cases}. \qquad (4.7)$$

At each location in space, small, moving targets are characterized by an initial rise (or fall) in brightness, and after a short delay are followed by a corresponding fall (or rise). This property of small features is matched by multiplying each contrast channel (ON or OFF) with a version of the opposite polarity delayed via a discrete first order low-pass filter (τ=25 ms, T_s=1 ms):

$$LP_{ESTMD} = \frac{z+1}{51z - 49}, \qquad (4.8)$$

and summing the output (Bagheri et al., 2015). This processing provides sensitivity to both contrasting target polarities (dark or light) (Bagheri et al., 2015).

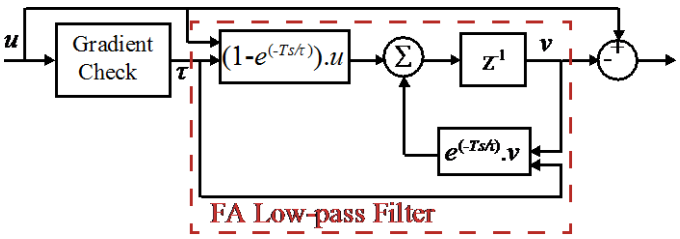

Figure 4.3. Fast adaptation algorithm (FA). The half-wave rectified channels '*u*', is fed into the 'Gradient Check' block (Eq. 4.6), which determines whether the luminance of each individual pixel is rising or falling, and outputs a matrix of appropriate 'τ' (time constant) values. Both '*u*' and 'τ' are fed into the FA low-pass filter block, which filters each pixel of the image according to the time constant value. The filtered image '*v*' is subtracted from the original image '*u*'. This computation emulates fast temporal adaptation, reducing responses to textural variations in the image.

Kernel S

-1	-1	-1	-1	-1
-1	-1	-1	-1	-1
-1	-1	2	-1	-1
-1	-1	-1	-1	-1
-1	-1	-1	-1	-1

Kernel M

-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

Kernel L

-1	-1	-1	-1	-1
-1	2	2	2	-1
-1	2	2	2	-1
-1	2	2	2	-1
-1	-1	-1	-1	-1

Figure 4.4. The kernels that are used in centre-surround antagonism (CSA) to change model size-tuning.

4.3.1.3 Integration and facilitation

A hyperbolic tangent function was used to model the neuron-like soft saturation of ESTMD outputs, ensuring a signal range between 0 and 1:

$$S(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \ . \qquad (4.9)$$

A simple competitive selection mechanism is added to the target detection algorithm by choosing the maximum of all ESTMD output values sampling the visual field. The location of this maximum in the ESTMD output is considered as the target location. The slow build-up of facilitation as observed in dragonfly CSTMD1 neurons (Nordström et al., 2011; Dunbier et al., 2011; Dunbier et al., 2012) permits the extraction of the target signal from noisy (cluttered) environments (Bagheri et al., 2015). Previously, we modelled this facilitation mechanism with a Gaussian weighted 'map', located relative to the winning

feature, shifted to account for the target's velocity (Bagheri et al., 2015). Here we implement facilitation with a retinotopic array of small-field STMDs, each integrating ESTMD output (~10°x10° region) with weights defined by a grid of 2D Gaussian kernels (half-width at half maximum of 5°) with centres separated by 5° (50% overlap) (Appendix D, Figure D1). To mimic the slow build-up of the response of CSTMD1 neurons the ESTMD output was multiplied with a low-pass filtered version of this facilitation map (T_s=1 ms).

$$LP_{Fac} = \frac{z+1}{\left(1+\frac{40}{w}\right)z + (1-\frac{40}{w})}, \qquad (4.10)$$

by changing w we varied the time constant of this discrete low-pass filter (facilitation time constant), thus controlling the duration of enhancement around the predicted location of the winning feature. We previously showed that the optimal facilitation time constant is dependent on the amount of background clutter and the target's velocity (Bagheri et al., 2015). In physiological recordings, we observe variability in individual response time-courses (mean ~200 ms), suggesting variability in facilitation time constants and modulatory factors that may dynamically change this parameter are currently under investigation. Here, we approximated this variability by determining the optimum facilitation time constant for each data sequence, testing across the range 40 to 2000 ms (9 values).

The directional component of the velocity vector was provided using a traditional bio-inspired direction selective model; the Hassenstein-Reichardt elementary motion detector (HR-EMD) (Hassenstein and Reichardt, 1956). The HR-EMD (Figure 4.5a) was applied to the ESTMD outputs, correlating two adjacent inputs, one after a delay (via a low-pass filter, τ=40 ms, T_s=1 ms):

$$LP_{HR-EMD} = \frac{z+1}{9z-7}. \qquad (4.11)$$

This results in a direction-selective output that is tuned to the velocity of small objects (Wiederman and O'Carroll, 2013b). The output of HR-EMD will be positive when the target moves from S1 to S2 (Fig. 5a), and negative in the reverse direction. The HR-EMD was applied in both horizontal and vertical directions to estimate whether the target is moving right/up (positive) or left/down (negative). The spatial shift of the facilitation kernel was determined by binning the magnitude of the output of the HR-EMD into three equal intervals, to estimate whether the speed of the target is slow, medium or fast (Bagheri et al., 2015). Figure 4.5b shows how this facilitation map builds up throughout tracking.

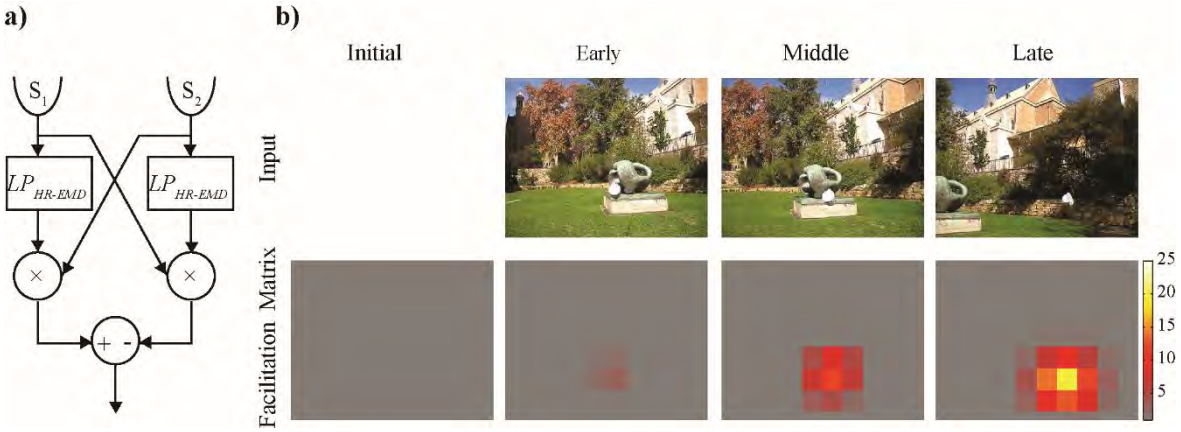

Figure 4.5. a) The Hassenstein-Reichardt (HR) EMD uses two spatially separated contrast signals (S1, S2) and correlates them after a delay (LP, low-pass filter) resulting in a direction selective output (Hassenstein and Reichardt, 1956). Subtracting the two mirrored symmetric sub-units yields positive response for the preferred motion direction (in this case left to right) and a negative one in the opposite direction (right to left). b) The facilitation map builds up slowly in response to targets that move along long, continuous trajectories. Therefore, as the tracking progresses, the facilitation map builds in strength around the selected target.

4.3.2 Input Imagery

4.3.2.1 High contrast target

The ESTMD model is size-tuned, due to spatial centre-surround antagonism and temporal cross correlation between local ON and OFF pathways. This forms a ‘matched filter’ for both the spatial and temporal characteristics of small, moving features. Additionally, the

model has inherent velocity tuning as observed in correlation-type motion detection mechanisms (i.e. HR-EMD). The position of the peak velocity response is dependent on the ON/OFF delay filter time constant as seen in Figure 4.1. We quantified this model tuning by presenting high contrast (black on white) targets of varying size and velocity.

4.3.2.2 Large targets within natural scenes

The IIT model was designed to detect small moving targets, in the order of a few degrees. However, because large objects may be composed of several small parts, we examined what the IIT model would do when presented with a range of object sizes. To test such scenarios, we used the CVPR2013 Online Object Tracking Benchmark (OOTB) (Wu et al., 2013). This is a popular and comprehensive benchmark dataset of 50 sequences, specifically designed for evaluating performance. The field of view (FOV) for these sequences was not available, therefore our 1° subsampling was implemented assuming capture by a 35 mm (equivalent) camera with a normal 50 mm lens (average diagonal FOV~55°). Figure 4.6a shows a 2-D histogram of target size and velocity within OOTB. This reveals that only a small proportion of targets within the OOTB dataset is within the tuning range of IIT (see Section 4.4.1).

4.3.2.3 Small targets within natural scenes

To match our problem definition, i.e. tracking small, moving targets in natural scenes, we recorded 25 additional video sequences (STNS Dataset). These sequences included heavy background clutter and camera motion. The statistical properties of the targets within these video sequences are presented in Figure 4.6b. The range of target size and velocity presented in the STNS dataset is one typically required for applications such as airborne surveillance. Datasets varied from 71 to 3872 frames, with an average of 760. These video sequences are available online, including the manually generated ground truth for each frame (https://figshare.com/articles/STNS_Dataset/4496768).

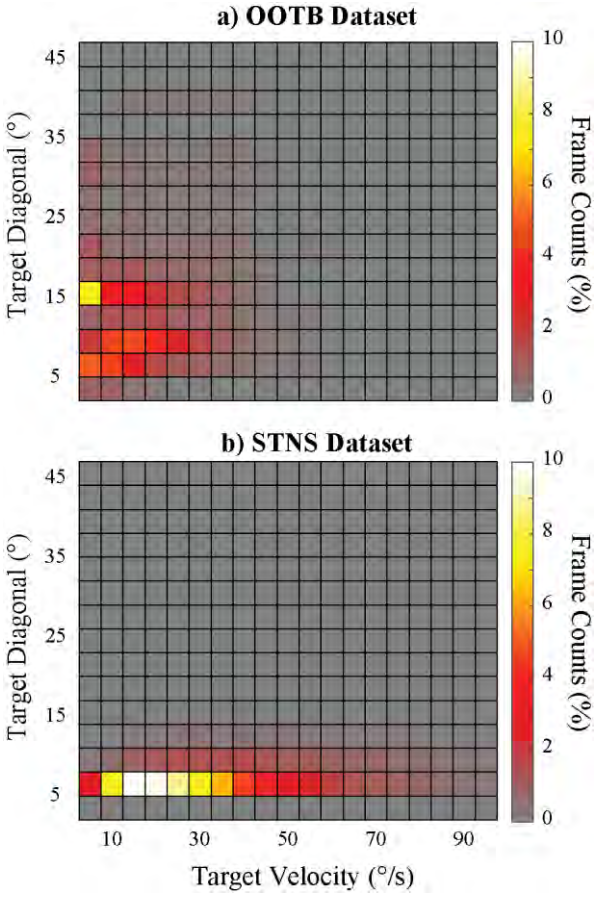

Figure 4.6. 2D histograms of dataset statistics showing the target size and velocity range in the tested datasets. The OOTB dataset has a broader range of objects at larger sizes, that move at slower speeds. Our problem definition is the tracking of small, moving targets as created in the STNS dataset (~5°, ~20-70°/s).

The datasets we used here contained a range (and type) of background motion including translation, rotation, and vibration. Although due to the lack of depth perception we could not quantify the motion in the direction perpendicular to the image plane, the camera motion in the image plane in these sequences ranges from a stationary camera (0 °/s) to an average camera velocity of 22±8 °/s.

4.3.3 Benchmarking Algorithms

To establish the efficacy and computational efficiency of our insect-inspired tracker (IIT) model, we compared its performance with three recent models DSST (Danelljan et al., 2014), KCF (Henriques et al., 2015), and MUSTer (Hong et al., 2015) as well as six 'classic',

highly-cited algorithms. We chose MATLAB implementations of these algorithms (provided by their authors), to maximise the fairness of comparison. The Matlab code for the IIT model is downloadable via https://figshare.com/s/377380f3def1ad7b9d44.

1- **Discriminative Scale Space Tracker (DSST)** (Danelljan et al., 2014) proposes a separate 1-dimensional correlation filter to estimate the target scale. This model uses the original feature space as the object representation.

2- **Kernelised correlation filter (KCF)** tracker (Henriques et al., 2015) uses multiple channels by summing over the results from all the channels in the Fourier domain. This model boosts its performance by using a Gaussian kernel and histogram-of-oriented-gradients (HOG) features.

3- **MUlti-Store Tracker (MUSTer)** (Hong et al., 2015) proposes a cooperative tracking framework inspired by a biological memory model called the Atkinson Shiffrin Memory Model (ASMM) (Atkinson and Shiffrin, 1968). ASMM is inspired by short-term and long-term memory in the human brain. Short-term memory which updates aggressively and forgets information quickly, stores local and temporal information. However, long-term memory which updates conservatively and maintains information for a long time, retains general and reliable information.

4- **Compressive Tracking (CT)** (Zhang et al., 2012) proposes an appearance model based on features extracted in the compressed domain. This tracker uses a sparse measurement matrix to extract the features for the appearance model.

5- **Incremental Visual Tracker (IVT)** (Ross et al., 2008) proposes an adaptive appearance model. This tracker calculates and stores the Eigen images of the latest target observation with incremental principal component analysis (IPCA) while slowly deletes the old observations.

6- **L1-minimization Tracker (L1T)** (Mei et al., 2011) employs sparse representation by

L1 to provide an occlusion insensitive method. This tracker uses a particle filter to find target windows. Then it defines sparse representation by using the intensity of sample windows close to the target location.

7- **Locally Orderless Tracker (LOT)** (Oron et al., 2012) divides the initial bounding box into super pixels. Each super pixel is represented by its centre of mass and average HSV-value. This tracker employs a parameterized Earth Mover Distance (Rubner et al., 2000) between the super pixels of the candidate and the target windows to calculate the likelihood of each particle sample.

8- **Super Pixel Tracker (SPT)** (Wang et al., 2011) clusters super pixels based on their histograms to form a discriminative appearance model for distinguishing the object from the background.

9- **Tracking, Learning and Detection (TLD**) (Kalal et al., 2012) combines a discriminative learning method with a detector and a Median-Flow tracker. The learning process uses a pair of false positives and false negatives experts to estimate the errors of the detector and to avoid these errors in future observations.

All models were tested in MATLAB on the same PC with a 4-core Intel i7 3770 CPU (3.4 GHz) and 16 GB RAM. The location of a target bounding box in the initial frame was provided for the benchmark algorithms. Likewise, in the first frame, we biased our IIT model toward the initial location of the target by allowing the facilitation to build up in the target region for 200 ms prior to the start of the experiment. This was implemented by feeding the target location in the first frame as the future location of the target (*r'(t+1)* in Figure 4.1) to the facilitation mechanism.

4.4 Results

Here, we present our measures of robustness and processing speed for the group of trackers. The protocol is to initialize the tracker in the first frame and then track the object of interest

until the last. The resultant trajectory is compared to ground truth using metrics as specified in each experiment.

4.4.1 Size and Velocity Tuning

An important feature of our model is its size and velocity tuning. The addition of facilitation increases target discriminability, but also modifies model size and velocity tuning. We quantified this tuning (Figure 4.7) to investigate the relationship between model responses and target properties within the input imagery. We varied facilitation time constants in addition to a non-facilitated model.

The model displays maximum responses at a target size of ~3-4°, diminishing above 10º. The model does not respond to targets slower than 20º/s and the optimum velocity increases as target size increases. This is due to the increased spatial separation between leading and trailing edges (in the direction of travel), which requires a faster transit speed to match the correlation delay between OFF and ON channels (confounding target width and velocity, as observed in physiology). This relationship between velocity and size might be beneficial in closed-loop pursuit as the insect approaches its target.

Figure 4.7 shows facilitation changes the size and velocity tuning of the model. While the non-facilitated model responds to target sizes of up to 14º, facilitation broadens the tuning range to 27º at high target velocities (V>200º/s, facilitation time constant τ=40 ms). The shorter the facilitation time constant is, the faster facilitation builds up to its maximum (Bagheri et al., 2015), therefore, model responses increase as facilitation time constant decreases. However, the choice of optimum facilitation time constant is complicated since it depends on different factors such as target velocity and background image statistics (Bagheri et al., 2015).

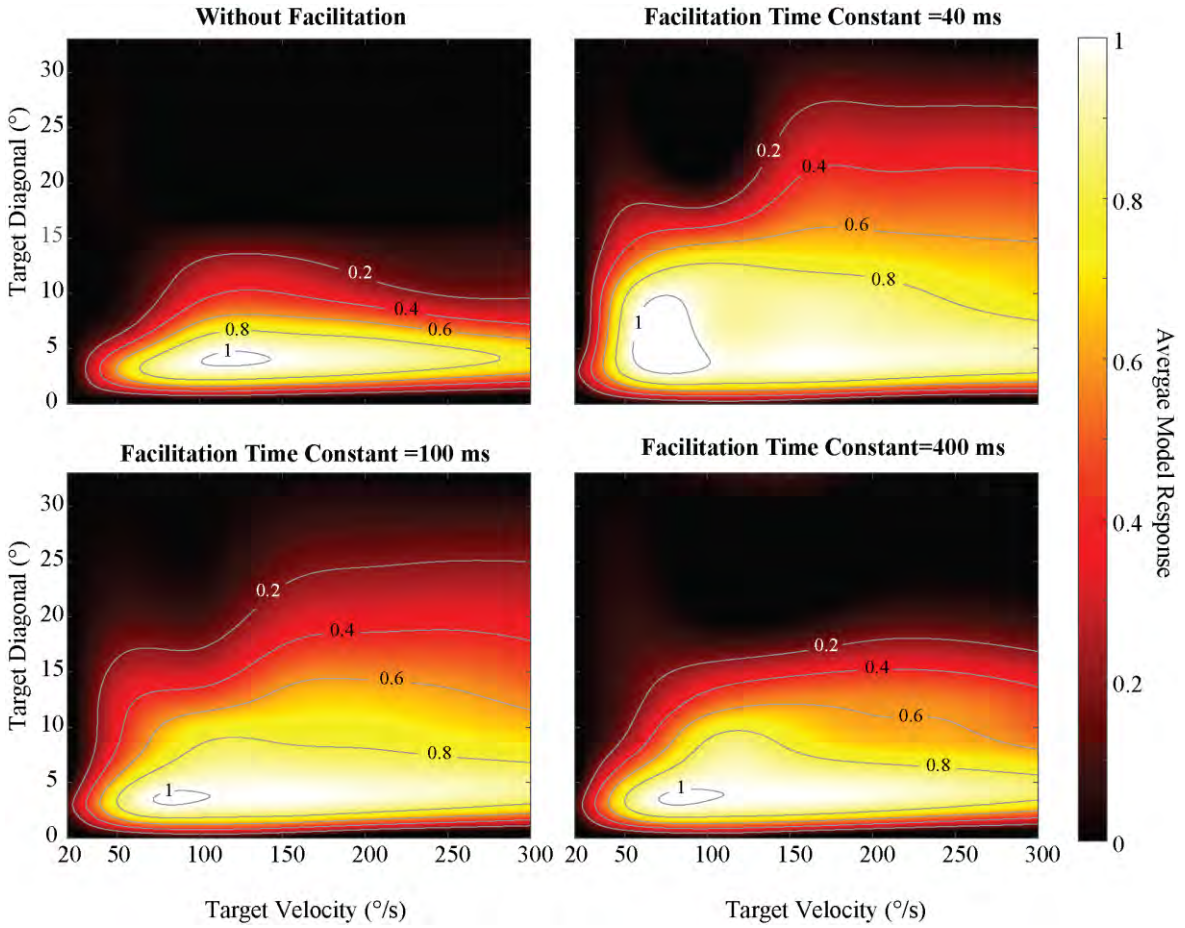

Figure 4.7. Size and velocity tuning of the model. The average model response was calculated by varying the size and velocity of a black target moving against a white background for both facilitated (with different time constant) and non-facilitated cases. The maximum response in all cases is for a target of ~3-4° moving at velocity of ~ 70-100 (°/s). However, the optimum velocity increases as the target size increases.

4.4.2 Size Tuning

Figure 4.8 shows tracking snapshots for each sequence, representative of early, middle and late stages of the tracking. In OOBT sequences, our IIT model selects sub-features of the large object within its tuning size, and retains that selection throughout the tracking. For example, in the *Couple* sequence the model locks on to the shoes of the pedestrians, in the *Mountain Bike* sequence it focuses on the bike seat and in the *Jogging* sequence it follows the head of the jogger. However, in the sequences belonging to the STNS dataset (*Key, Train, Pony2, Owl2*), where the target is already within the size tuning, it tracks the object itself.

Figure 4.8. Snapshots of IIT tracking results in three different frames representing early, middle and late stages for each sequence. The red square marks the location of the winning feature in the output of the IIT model.

The IIT model was designed to mimic size tuning observed in STMD neurons. However, recent physiological experiments have observed additional STMD types with peak responses for larger objects at ~10° (Wiederman et al., 2013). This raises the intriguing possibility of parallel pathways encoding different, broadly-tuned size ranges, which might be combined later in the visual pathway for a precise estimate of size (Evans et al., In press). This would be analogous to a human's capability to encode millions of colour wavelengths, with only three broadly tuned photoreceptor classes. To explore how individual, size-tuned pathways

would respond to input imagery, we varied the model's size tuning by changing the strong spatial centre-surround antagonism at the ESTMD stage (Figure 4.4). *Kernel S* (small) provides a stronger surround antagonism compared to *Kernel M* (medium, the default kernel), hence, the model is selective to smaller features. *Kernel L* (large) tunes the model to larger objects. Figure 4.9 shows how the size tuning range of the model changes with different kernels and Figure 4.10 provides examples of the sub-features tracked with different kernels. For instance, in the *Jogging* sequence, the model locks on to the head (3.6°) with *Kernel M*. However, the model is able to select other features, such as, a shoe (*Kernel S*, ~1°) and bike rider's body (*Kernel L*, ~7°).

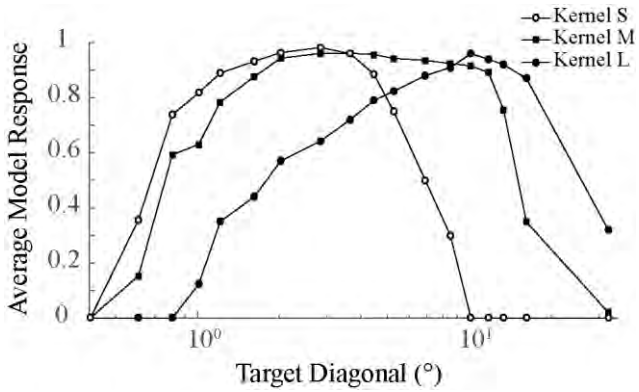

Figure 4.9. The size tuning range of the model changes with the choice of kernel for strong spatial centre-surround antagonism at the ESTMD stage. With Kernel S the model is selective for small targets and Kernel L tunes the model to larger objects. Kernel M provides a trade-off between selectivity to large features and small features. The average response of the model was measured for a black target moving against white background at velocity of 100 °/s with a facilitation time constant of 40 ms.

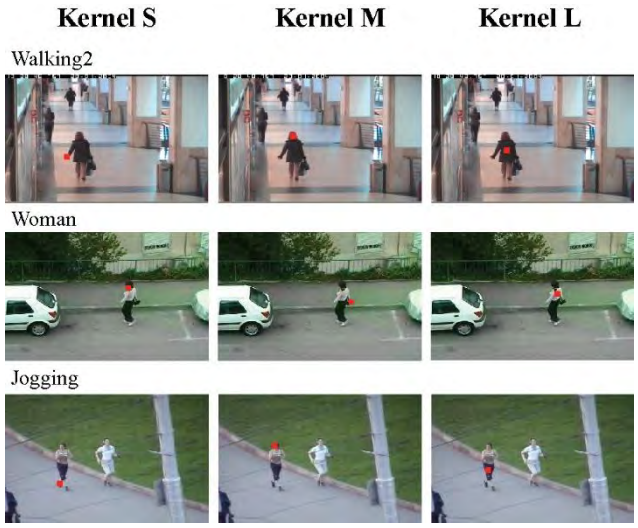

Figure 4.10. Different kernels for centre-surround antagonism change size tuning of the model and consequently sub-features of the target that the model selects for tracking. The red square marks the location of the winning feature in the output of the IIT model.

We measured success rate as the frame percentage where the target is correctly identified by the tracker. Target size and orientation could be useful information in some applications. However, for our purpose, targets are a small square (i.e. equal to its subsampling resolution) and we limit our success metric to correctly locating target position in each frame. For the IIT algorithm, if the location of the winning feature was within the ground truth box, it was considered a successful detection of the target.

Figure 4.11 shows box-and-whiskers plots summarizing how the size tuning of the model influences the success rate. The central mark (white circle) is median success rate, edges are the 25th and 75th percentiles, and whiskers are the non-outlier data range.

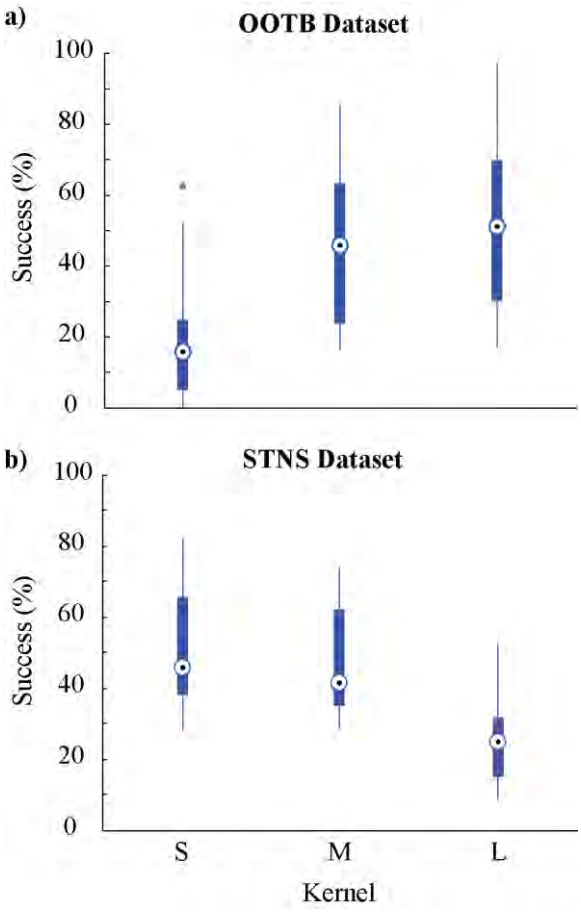

Figure 4.11. Effect of size tuning on model performance. Tuning to larger objects increases the model success rate in the OOTB dataset, however, it diminishes model performance in the STNS dataset. This is because the OOTB dataset is biased to large objects, whereas the STNS dataset has small moving targets.

For the OOTB dataset, which has generally larger targets, the success rate of the model increases with a kernel selected for larger objects (*Kernel L*). In contrast, for the STNS dataset which uses small targets deliberately designed to match the relative size of typical prey pursued by predatory dragonflies, *Kernel S* leads to a more robust model performance. *Kernel M* provides a trade-off between selectivity to large features and small features. Consequently, the success of the IIT model with *Kernel M* is between those for *Kernel S* and *Kernel L*. For the remaining analysis, we use the result of experiments with *Kernel M*.

4.4.3 Facilitation Time Constant

Our previously published work (Bagheri et al., 2015) show that the optimum facilitation time constant varies systematically with image statistics and target velocity, suggesting that a

dynamic modulation of time constant would improve target detection. However, in the current version of the model rather than implementing a dynamic control system for the time constant, we simply used several variants. We also tested the model with a global optimum facilitation time constant (τ_f=400 ms). Figure 4.12 shows the difference between median of these two implementations is not significant (less than 1.5%). However, using a global optimal time constant (IIT) diminishes the 25th and 75th percentile by ~8% compared to the implementation with suboptimal time constant for each data sequence (IIT*). In the following sections, we used the results of experiments with optimum time constant for each data sequence.

4.4.4 Benchmarking Success Plot

In computer vision literature, target detection is typically represented with a bounding box, with a common measure for success being a 50% overlap between the ground truth box and target bounding box (Wu et al., 2013; Smeulders et al., 2014). Here we scored each frame as a success if the bounding box centre was within the ground truth box. Although this provides a higher success rate than the common success measure, since our model represents the target with a small square, we used this metric to provide a fairer comparison between our model and the other engineering models.

Figure 4.12a shows OOTB dataset results which mainly consists of large targets (see Methods 4.3.2.2). The median of IIT*, CT, IVT, LOT and SPT are similar though all fall far behind the state-of-the-art trackers such as DSST, KCF and MUSTer. This is unsurprising, given our model is designed to track small moving targets in natural scenes, whilst the OOBT dataset is composed of many larger objects that are, in effect, false positives for our model design. Of interest, is that in the OOBT dataset, the IIT model often tracks smaller components of a large object. Whether the combination of parallel size-tuned pathways could be utilized to track OOBT objects is currently being investigated.

In comparison to the public dataset, our STNS dataset explores another set of challenges in target tracking; small moving targets that are frequently camouflaged against the background clutter. In this more challenging set of scenarios, the median performance of all trackers (Figure 4.12b) are lower than the OOTB dataset. Although the 75th percentile of our model is not as high as other trackers, it has the highest median and 25th percentile showing its more robust and reliable performance in the most challenging of these scenarios.

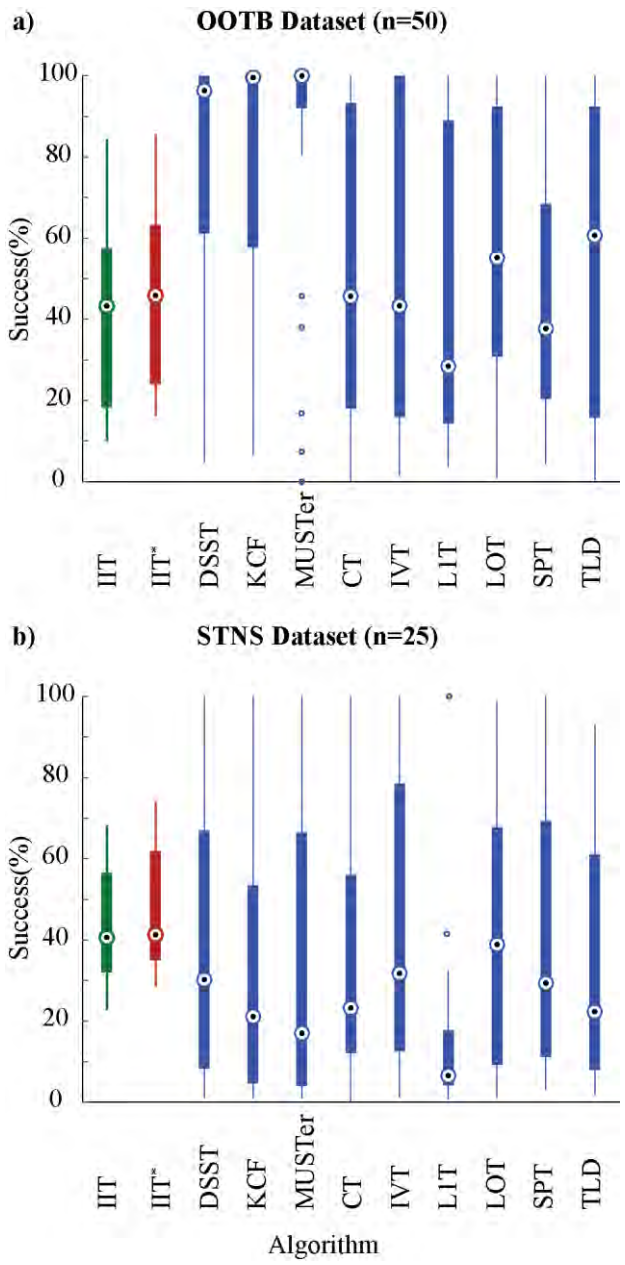

Figure 4.12. Box and whiskers plots for successful target tracking of different algorithms for a) OOTB dataset, b) our own dataset (STNS dataset). The white circles represent the median success rate, and the box shows the

25th and 75th percentile. The IIT and IIT* present the results with a global optimum time constant and optimum time constant for each sequence, respectively. The success of IIT model is the result of experiments with *Kernel M.*

4.4.5 Precision Plot

The Precision plot is an evaluation method recently adopted to measure the robustness of tracking (Wu et al., 2013; Danelljan et al., 2014; Henriques et al., 2015; Homg et al., 2015; Babenko et al., 2011). It shows the percentage of the frames where the Euclidean distance between the centre of the tracked target and the ground truth is within a given 'location error' threshold (Figure 4.13). A higher precision at low thresholds means the tracker is more accurate. Precision at a 20 pixel threshold is widely used as a performance benchmark in the literature (Wu et al., 2015; Henriques et al., 2015; Babenko et al., 2011).

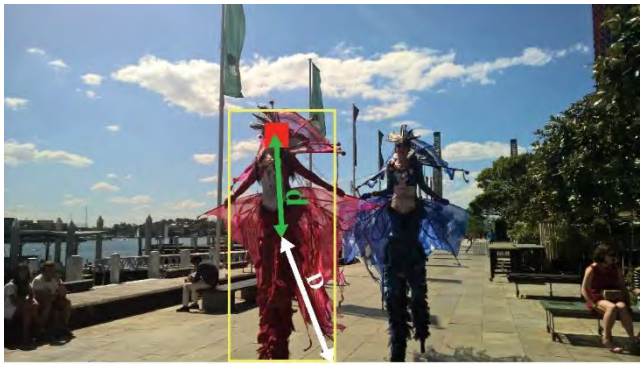

Figure 4.13. Schematic of the mechanism that is used for calculation of precision. The green arrow (d) shows the 'location error' between the centre of the ground truth and winning feature. The red arrow (D) shows half of the diagonal of the ground truth rectangle which is used for normalization. Therefore, the normalized location error is equal to d/D.

Figure 4.14a shows the precision plot for all trackers both for OOTB and STNS datasets. In the OOTB dataset our IIT model has a low precision until the threshold of 11 pixels. Nonetheless, its precision catches up to SPT, CT, L1T and IVT, LOT and TLD around the threshold of 16 pixels. However, its precision never nears DSST, KCF or MUSTer. In contrast, in the experiments with the STNS dataset, the precision of our model exceeds the

precision of all trackers at a threshold of 30 pixels.

The poor precision of our model below the threshold of 16 pixels with the OOTB dataset is due to two factors. Firstly, large objects are composed of small contrasting parts, allowing our model to lock on to these sub-features of the larger object. The result is robust target tracking, but with the location of the tracked feature typically offset from the centre of the object.

The second factor is the subsampling of the input image. In Figure 4.14a the grey rectangular area shows the location error thresholds below the average subsampling ratio that is applied to the images. This shows our model precision improves as it gets closer to the boundary of the grey region. Although subsampling reduces the redundant information in the scene and increases the processing speed, it also reduces model spatial resolution – a situation that mimics the low spatial acuity of dragonfly compound eyes. However, whether a high spatial acuity is necessary or not is generally application dependent. For example, in applications such as face tracking and gesture recognition in a crowded scene where the target representation is very important, an accurate detection of the target is desirable. Nonetheless, in aerial video surveillance where the target motion and ego-motion of the camera are the more important components, high spatial acuity is not necessary as long as the tracker detects the correct location of the target.

Although the precision at a 20 pixel threshold has been used for benchmarking in the literature, this somewhat arbitrary benchmark threshold does not tell the whole story. A 20 pixel offset is a relatively large error for tracking a small object (e.g. on the scale of pixels themselves), yet may not be at all significant for identifying a large object. Therefore, to account for the huge range of target sizes in these image sequences we also normalized the location error of the target in each frame by half of the diagonal of the ground truth rectangle within that frame (Figure 4.13). Figure 4.14b shows the results of this normalization. Similar

to the precision plot in the OOTB dataset our model provides an average level of precision compared to other algorithms. However, in the STNS dataset, over the entire upper 1/2 of this normalized error threshold (NT) our model is substantially more precise, revealing the successful identification of the correct small target location.

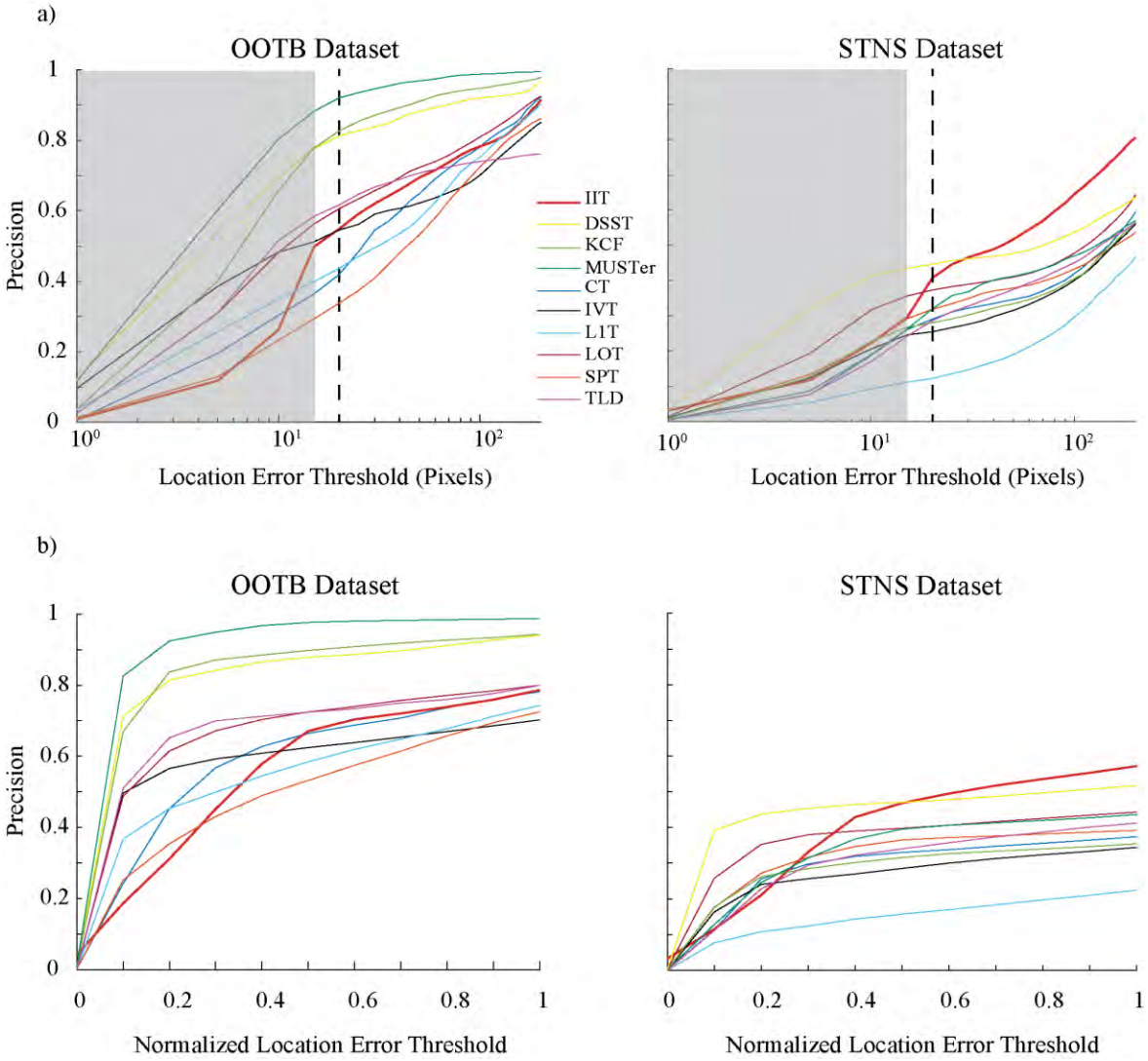

Figure 4.14. a) Precision plot for both OOTB and STNS datasets. The grey area shows the location error thresholds below the average subsampling ratio that is applied to the images in the IIT model. b) The precision of the trackers is normalized by half of the diagonal of the ground truth rectangle.

4.4.6 Overall Performance

Table 4.2 provides a descriptive summary of algorithm performance with the OOTB and STNS datasets. The average success represents the performance across all videos in each

dataset (n_{OOTB}=50, n_{IIT}=25). In addition to the average success rate of the sequences, we also calculated the weighted success which shows the percentage of the successful frames out of all the tested frames in each dataset (F_{OOTB}=29518, F_{STNS}=27514). This normalization accounted for the difficulty of 'long term' tracking, where it is easier for the trackers to lock on to the target in a short sequence than a long one.

Since our STNS dataset mainly includes small moving targets within heavily cluttered environments, the target is frequently well camouflaged against the background clutter. Therefore, it is more difficult for trackers to lock on to the target. This might explain why the weighted success of all the engineering algorithms for the STNS dataset are lower than their average success (Table 4.2). However, in both datasets, the weighted success of our model is higher than its average success. Our facilitation mechanism (based on the recently observed facilitatory behaviour of target-detecting neurons (Nordström et al., 2011; Dunbier et al., 2011; Dunbier et al., 2012)) builds up slowly in response to targets that move in long continuous trajectories, and thus improves target detection as tracking progresses (Bagheri et al., 2015).

Table 4.2. Summary of experimental results on both OOTB dataset and STNS dataset.

	Performance Measure	**Algorithm**									
		IIT	**DSST**	**KCF**	**MUSTer**	**CT**	**IVT**	**L1T**	**LOT**	**SPT**	**TLD**
OOTB Dataset	Average Success (%)	52.0	78.3	78.0	89.4	51.0	52.9	45.1	57.1	44.6	55.0
	Weighted Success (%)	60.0	85.0	87.5	94.3	62.8	60.6	51.5	69.3	47.8	70.0
	Precision (20 px)	0.56	0.81	0.83	0.92	0.42	0.56	0.44	0.61	0.34	0.62
	Normalized Precision (NT=1)	0.78	0.94	0.94	0.99	0.78	0.70	0.74	0.80	0.73	0.80
STNS Dataset	Average Success (%)	47.6	41.5	33.2	36.4	37.3	42.7	15.0	40.5	41.9	35.0
	Weighted Success (%)	52.1	34.5	29.9	36.8	32.7	35.8	14.2	40.0	35.9	32.2
	Precision (20 px)	0.41	0.44	0.27	0.32	0.29	0.25	0.12	0.37	0.32	0.29
	Normalized Precision (NT=1)	0.57	0.51	0.35	0.44	0.37	0.34	0.22	0.44	0.39	0.41

Overall, in the experiments with the OOTB dataset, while it does well, the performance of our IIT model does not exceed those of classical trackers such as CT, IVT, L1T, LOT, SPT and TLD. However, in the STNS dataset, our model outperforms all other trackers in terms of average success, weighted success, and normalized precision.

4.4.7 Processing Speed

In addition to tracking performance, processing efficiency is a critical concern in target tracking applications. Indeed, many trackers are considered impractical in real-time scenarios due to the long processing duration required using typical hardware (Yang et al., 2011; Zhang et al., 2012; Yilmaz et al., 2006). Figure 4.15 shows the calculated processing speed for all algorithms using the author-provided MATLAB code. The processing speed is consistent with values reported in the original papers of these algorithms. The most computationally expensive process that applies to images in the IIT model is the Gaussian optical blur in the early visual processing stage. However, previous studies have shown that a Gaussian blur can directly be obtained via optics (Brückner et al., 2006; Colonnier et al., 2015). Likewise, our software implementation of the optical blur built in to biological eyes could be readily implemented by using de-focusing optics in robotic applications. Therefore, we tested the processing speed of our IIT algorithm both with and without optical blur. Even with the processing time required for optical blur the speed of the IIT has the second best median among all trackers. However, when we eliminate the processing time for optical blur, the IIT model exceeds all other trackers, with an average performance approximately 2 times faster than the fastest engineering algorithm (KCF) (note the logarithmic scale in Figure 4.15).

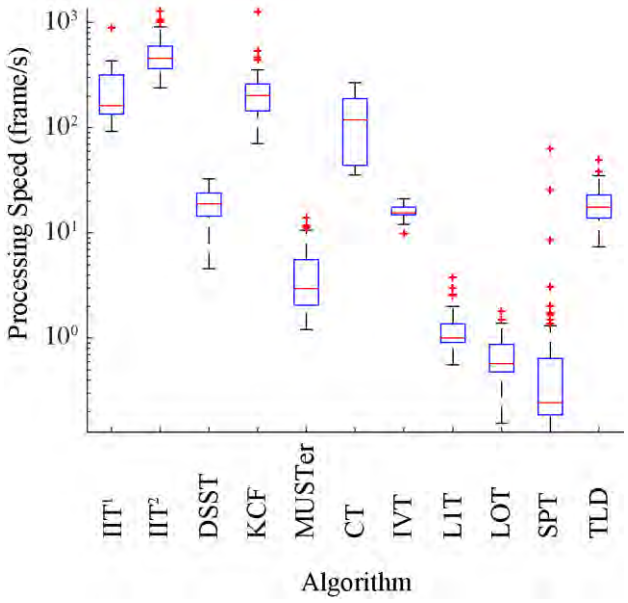

Figure 4.15. Processing speed of trackers. IIT^1 and IIT^2 represent the model with and without optical blur processing time, respectively.

4.4.8 Complexity of Algorithms

Although we presented the processing speed of algorithms for the tested sequences, the size of the input image can affect the processing speed of the models. Therefore, to identify how the processing time of these algorithms changes as the problem size increases, we determined the Big O notation (Bach and Shallit, 1996) for our IIT model as well as some of the fastest processing engineering models. The Big O notation characterizes functions according to their growth rates. For the IIT model there are two factors that can change the processing speed and therefore the complexity of the model. The first one is the size of the input image, which mainly affects the processing speed due to the Gaussian optical blur in the early visual processing stage. However, as we mentioned in Section 3.7, this optical blur can be achieved via hardware rather than software. The other factor is the Field of View (FOV) which changes the processing speed after optical blur. Table 3 summarizes the results of the complexity analysis. Although the complexity of the IIT model is quadratic in response to the changes of input image size, the processing time changes linearly in response

to variation of FOV. We should mention that we made no attempt to optimize our MATLAB code (e.g. it does not use parallel computation).

Table 4.3. The computational complexity of algorithms. IIT^1 and IIT^2 represent the complexity of IIT model in terms of change in the size of input image and size of FOV, respectively.

Algorithm	IIT^1	IIT^2	KCF	CT	DSST	TLD
Complexity	$O(n^2)$	O(n)	$O(n^3)$	O(n)	O(n)	$O(n^2)$

4.5 Conclusion

We have demonstrated the robustness and efficiency of a target-tracking algorithm inspired directly by insect neurophysiology. As with the dragonfly, our model is particularly suited to tracking small moving targets, rather than some of the larger objects within the public OOTB dataset. Model performance for this task is improved with different tuning kernel weights (Figure 4.4 and Figure 4.11) and in future research we will test the effect of integrating parallel, size-tuned sub-systems. This elaboration may improve precision and yet still retain the real-time performance inherent in our model approach. Although there is no definitive evidence for such parallel processing in insect vision, this processing is well-supported in human psychophysics experiments (Graham and Nachmias, 1971). How an insect integrates the information of these small moving features across the visual field to detect larger objects is a question physiologists are currently investigating. Nevertheless, in terms of processing speed, our model is one of the best among all the tested trackers, mimicking the remarkable efficiency of the insect visual system upon which it is based. As such, it may be well suited to applications where efficiency is paramount.

This is the first time that an insect-inspired target tracking algorithm has been directly tested against state-of–the-art engineered systems. One prior study (Shoemaker et al., 2011) compared an insect-inspired optic flow processing algorithm with computational optic flow

estimates based on Lucas-Kanade equations, both in closed loop simulation. As with our own model, their insect inspired system (Shoemaker et al., 2011) was computationally highly efficient and with sufficient elaboration was shown to be robust for certain closed-loop control robotic applications. However, in absolute performance terms, it never surpassed the engineered solution. Perhaps the conclusion of our current work is that despite the relatively simple feed-forward mechanism we implemented, the performance of our system exceeds all other trackers in the scenarios where the target is within its tuning range (i.e., when tracking small objects on the scale of the camera resolution). Nonetheless, our results show that this performance does not come at the cost of additional processing time and the model processes the frames at high speed, especially when the optical blur process is obtained via defocusing optics rather than software implementation.

Our previous modelling efforts suggest that the temporal optima of facilitation mechanism varies with respect to the amount of background clutter and target velocity (Bagheri et al., 2015). One of the limitations of the current work is that it uses a static facilitation time constant. Therefore, we simply used several variants of facilitation time constant to find the optimum for each sequence. However, for future robotic applications, an interesting extension of the model would be a system which dynamically modulates the facilitation time constant during run-time. For instant, a fast processing system such as fuzzy control can be exploited to estimate the facilitation time constant using target velocity (as estimated by the velocity estimator) and background clutter (e.g. using a simple metric for visual clutter such as the one developed by Silk (1995)). Since a fuzzy control system uses a set of logical rules to estimate the output, it is well suited for low-cost implementations and would cause very little overhead in terms of execution time.

Here, we tested our algorithm in open-loop, however, animals interact with the target by changing eye or body movements, which then modulate the visual inputs underlying the detection and selection task (via closed-loop feedback). This active vision system may be a key to further exploiting visual information by the simple insect brain for complex tasks such as target tracking. Future research will attempt to implement this model along with insect active vision gaze-control strategies in a robotic platform to test the performance of them together under real-world conditions.

Acknowledgements

This research was supported under Australian Research Council's Discovery Projects (project number DP130104572) and Discovery Early Career Researcher Award (project number DE150100548) funding scheme, and the Swedish Research Council (VR 2014-4904). We thank the manager of the Botanic Gardens in Adelaide for allowing insect collection and video capture.

References

Abdel-Hakim A. E., & Farag A. A. (2006). CSIFT: A SIFT descriptor with color invariant characteristics. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2, 1978-1983.

Atkinson R. C., & Shiffrin R. M. (1968). Human memory: A proposed system and its control processes. Psychology of Learning and Motivation, 2, 89-195.

Babenko B., Yang M. H., & Belongie S. (2011). Robust object tracking with online multiple instance learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(8), 1619-1632.

Bach E., & Shallit J. O. (1996). Algorithmic Number Theory: Efficient Algorithms (Vol. 1). MIT press.

Bagheri Z. M., Wiederman S. D., Cazzolato B. S., Grainger S., & O'Carroll D. C. (2015). Properties of neuronal facilitation that improve target tracking in natural pursuit simulations. Journal of The Royal Society Interface, 12(108), 20150083.

Bagheri Z. M., Wiederman S. D., Cazzolato B. S., Grainger S., & O'Carroll D. C. (2014a). A biologically inspired facilitation mechanism enhances the detection and pursuit of targets of varying contrast. In International Conference on Digital Image Computing: Techniques and Applications (DlCTA), 1-5.

Bagheri Z., Wiederman S. D., Cazzolato B. S., Grainger S., & O'Carroll D. C. (2014b). Performance assessment of an insect-inspired target tracking model in background clutter. In International Conference on Control Automation Robotics & Vision (ICARCV), 822-826.

Bar-Shalon Y. and Fortmann T. (1988). Tracking and data Association. Academic Press Professional, Inc. San Diego, CA, USA.

Bay H., Tuytelaars T., & Van Gool L. (2006). Surf: Speeded up robust features. In European Conference on Computer Vision, 404-417.

Black M. J., & Jepson A. D. (1998). Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. International Journal of Computer Vision, 26(1), 63-84.

Blum A., & Mitchell T. (1998). Combining labeled and unlabeled data with co-training. In Proceedings of the 11th Annual Conference on Computational Learning Theory, 92-100.

Brückner A., Duparré J., Bräuer A., & Tünnermann A. (2006). Artificial compound eye applying hyperacuity. Optics express, 14(25), 12076-12084.

Burghouts G. J., & Geusebroek J. M. (2009). Performance evaluation of local colour invariants. Computer Vision and Image Understanding, 113(1), 48-62.

Chapelle O., Schölkopf B. & Zien A., (2006). Semi-supervised learning. MIT Press, Cambridge, Massachusetts, USA, 2006.

Chen J., Shan S., He C., Zhao G., Pietikainen M., Chen X., & Gao W. (2010). WLD: A robust local image descriptor. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9), 1705-1720.

Comaniciu D., Ramesh V., & Meer P. (2003). Kernel-based object tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(5), 564-577.

Colonnier F., Manecy A., Juston R., Mallot H., Leitel R., Floreano D., & Viollet S. (2015). A small-scale hyperacute compound eye featuring active eye tremor: application to visual stabilization, target tracking, and short-range odometry. Bioinspiration & Biomimetics, 10(2), 026002.

Corbet P. S. (1999). Dragonflies: Behaviour and ecology of Odonata. Cornell University Press, Ithaca, New York.

Danelljan M., Häger G., Khan F., & Felsberg M. (2014). Accurate scale estimation for robust visual tracking. In British Machine Vision Conference, 1-5.

Dunbier J. R., Wiederman S. D., Shoemaker P. A., & O'Carroll D. C. (2011). Modelling the temporal response properties of an insect small target motion detector. In IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 125-130.

Dunbier J. R., Wiederman S. D., Shoemaker P. A., & O'Carroll D. C. (2012). Facilitation of dragonfly target-detecting neurons by slow moving features on continuous paths. Frontiers in Neural Circuits, 6, 79.

Evans B. J. E., O'Carroll D. C., & Wiederman S. D., (In Press). Salience invariance with divisive normalization in higher order insect neurons. 6th European Workshop on Visual Information Processing, EUVIP.

Felzenszwalb P., McAllester D., & Ramanan D. (2008). A discriminatively trained, multiscale, deformable part model. In IEEE Conference on Computer Vision and Pattern Recognition, 1-8.

Gall J., & Lempitsky V. (2013). Class-specific hough forests for object detection. In Decision Forests for Computer Vision and Medical Image Analysis, 143-157.

Grabner H., Leistner C., & Bischof H. (2008). Semi-supervised on-line boosting for robust tracking. In European Conference on Computer Vision, 234-247.

Graham N., & Nachmias J. (1971). Detection of grating patterns containing two spatial frequencies: A comparison of single-channel and multiple-channels models. Vision Research, 11(3), 251-259.

Halupka K. J., Wiederman S. D., Cazzolato B. S., & O'Carroll D. C. (2013). Bio-inspired feature extraction and enhancement of targets moving against visual clutter during closed loop pursuit. In IEEE International Conference on Image Processing (ICIP), 4098-4102.

Halupka K. J., Wiederman S. D., Cazzolato B. S., & O'Carroll, D. C. (2011, December). Discrete implementation of biologically inspired image processing for target detection. In IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 143-148.

Hare S., Saffari A., & Torr P. H. (2011). Struck: Structured output tracking with kernels. In IEEE International Conference on Computer Vision (ICCV), 263-270.

Hassenstein, B. & Reichardt W. (1956). Systemtheoretische analyse der zeit-, reihenfolgen- und vorzeichenauswertung bei der bewegungsperzeption des rüsselkäfers chlorophanus. Zeitschrift für Naturforschung B, 11(9-10), 513-524.

Henriques J. F., Caseiro R., Martins P., & Batista J. (2015). High-speed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(3), 583-596.

Hong Z., Chen Z., Wang C., Mei X., Prokhorov D., & Tao D. (2015). Multi-store tracker (muster): A cognitive psychology inspired approach to object tracking. IEEE Conference on Computer Vision and Pattern Recognition, 749-758.

Jansonius N. M., & Van Hateren J. H. (1991). Fast temporal adaptation of on-off units in the first optic chiasm of the blowfly. Journal of Comparative Physiology A, 168(6), 631-637.

Javed O., Ali S., & Shah M. (2005). Online detection and classification of moving objects using progressively improving detectors. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 696-701.

Kalal Z., Mikolajczyk K., & Matas J. (2012). Tracking-learning-detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(7), 1409-1422.

Kitagawa G. (1987). Non-Gaussian state-space modelling of nonstationary time series. Journal of the American statistical association, 82(400), 1032-1041.

Leistner C., Saffari A., Roth P. M., & Bischof H. (2009). On robustness of on-line boosting-a competitive study. In IEEE 12th Computer Vision Workshops, 1362-1369.

Levin A., Viola P., & Freund Y. (2003). Unsupervised improvement of visual detectors using co-training. In Proceedings of 9th IEEE International Conference on Computer Vision, 626-633.

Li P., Zhang T., & Ma B. (2004). Unscented Kalman filter for visual curve tracking. Image and Vision Computing, 22(2), 157-164.

Masnadi-Shirazi H., Mahadevan V., & Vasconcelos N. (2010). On the design of robust classifiers for computer vision. In IEEE Conference on Computer Vision and Pattern Recognition, 779-786.

Mei X., Ling H., Wu Y., Blasch E., & Bai L. (2011). Minimum error bounded efficient ℓ 1 tracker with occlusion detection. IEEE International Conference on Computer vision and pattern recognition (CVPR),1257-1264.

Nordström K., & O'Carroll D. C. (2009). Feature detection and the hypercomplex property in insects. Trends in Neurosciences, 32(7), 383-391.

Nordström K., Bolzon D. M., & O'Carroll D. C. (2011). Spatial facilitation by a high-performance dragonfly target-detecting neuron. Biology Letters, 7(4), 588-592.

Nordström K., Barnett P. D., & O'Carroll D. C. (2006). Insect detection of small targets moving in visual clutter. PLoS Biology, 4(3), e54.

O'Carroll, D. (1993). Feature-detecting neurons in dragonflies. Nature, 362, 541-543.

O'Carroll D. C., & Wiederman S. D. (2014). Contrast sensitivity and the detection of moving patterns and features. Philosophical Transactions of Royal Society B, 369(1636), 20130043.

O'Carroll D. C., Barnett P. D., & Nordström K. (2011). Local and global responses of insect motion detectors to the spatial structure of natural scenes. Journal of Vision, 11(14), 20-20.

Ojala T., Pietikainen M., & Maenpaa T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7), 971-987.

Olberg R. M., Worthington A. H., & Venator K. R. (2000). Prey pursuit and interception in dragonflies. Journal of Comparative Physiology A, 186(2), 155-162.

Oron S., Bar-Hillel A., Levi D., & Avidan S. (2012). Locally orderless tracking. IEEE International Conference on Computer vision and pattern recognition (CVPR), 1940-1947.

Osorio D. (1991). Mechanisms of early visual processing in the medulla of the locust optic lobe: How self-inhibition, spatial-pooling, and signal rectification contribute to the properties of transient cells. Visual Neuroscience, 7(4), 345-355.

Porikli F., Tuzel O., & Meer P., (2006). Covariance tracking using model update based on lie algebra. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 728-735.

Ross D. A., Lim J., Lin R. S., & Yang M. H. (2008). Incremental learning for robust visual tracking. International Journal of Computer Vision, 77(1-3), 125-141.

Rubner Y., Tomasi C., & Guibas L. J. (2000). The earth mover's distance as a metric for image retrieval. International Journal of Computer Vision, 40(2), 99-121.

Saffari A., Godec M., Pock T., Leistner C., & Bischof H. (2010). Online multi-class LPBoost. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3570-3577.

Scovanner P., Ali, S., & Shah, M. (2007). A 3-dimensional SIFT descriptor and its application to action recognition. In Proceedings of the 15th ACM International Conference on Multimedia, 357-360.

Shoemaker P. A., Hyslop A. M., & Humbert J. S. (2011). Optic flow estimation on trajectories generated by bio-inspired closed-loop flight. Biological Cybernetics, 104(4-5), 339-350.

Silk J. D. (1995) Statistical variance analysis of clutter scenes and applications to a target acquisition test. Institute for Defense Analysis, Alexandria, Virginia, 2950.

Smeulders A. W., Chu D. M., Cucchiara R., Calderara S., Dehghan A., & Shah, M. (2014). Visual tracking: An experimental survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(7), 1442-1468.

Srinivasan M. V., & Guy R. G. (1990). Spectral properties of movement perception in the dronefly Eristalis. Journal of Comparative Physiology A, 166(3), 287-295.

Stavenga D. (2003). Angular and spectral sensitivity of fly photoreceptors. I. Integrated facet lens and rhabdomere optics. Journal of Comparative Physiology A, 189(1), 1-17.

Straw A. D., Warrant E. J., & O'Carroll D. C. (2006). A bright zone in male hoverfly (Eristalis tenax) eyes and associated faster motion detection and increased contrast sensitivity. Journal of Experimental Biology, 209(21), 4339-4354.

Wang S., Lu H., Yang F., & Yang M. H. (2011). Superpixel tracking. In IEEE International Conference on Computer Vision (ICCV), 1323-1330.

Wiederman S. D., & O'Carroll D. C. (2013a). Selective attention in an insect visual neuron. Current Biology, 23(2), 156-161.

Wiederman S. D., & O'Carroll D. C. (2013b). Biomimetic target detection: Modeling 2nd order correlation of OFF and ON channels. In IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP), pp. 16-21.

Wiederman S. D., & O'Carroll D. C. (2011). Discrimination of features in natural scenes by a dragonfly neuron. The Journal of Neuroscience, 31(19), 7141-7144.

Wiederman S. D., Shoemaker P. A., & O'Carroll D. C. (2013). Correlation between OFF and ON channels underlies dark target selectivity in an insect visual system. The Journal of Neuroscience, 33(32), 13225-13232.

Wiederman S. D., Shoemaker P. A., & O'Carroll D. C. (2008). A model for the detection of moving targets in visual clutter inspired by insect physiology. PloS One, 3(7), e2784.

Wu, Y., Lim, J., & Yang, M. H. (2013). Online object tracking: A benchmark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2411-2418.

Yang H., Shao L., Zheng F., Wang L., & Song Z. (2011). Recent advances and trends in visual tracking: A review. Neurocomputing, 74(18), 3823-3831.

Yilmaz A., Javed O., & Shah M. (2006). Object tracking: A survey. ACM Computing Surveys (CSUR), 38(4), 1-45.

Zeisl, B., Leistner C., Saffari A., & Bischof H. (2010). On-line semi-supervised multiple-instance boosting. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1879-1879.

Zhang K., Zhang L., & Yang M. H. (2012). Real-time compressive tracking. In European Conference on Computer Vision, 864-877.

Zhao G., & Pietikainen M. (2007). Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(6), 915-928.

Zhu X., & Goldberg A. B. (2009). Introduction to semi-supervised learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, 3(1), 1-130.

Chapter 5. An autonomous robot inspired by insect neurophysiology pursues moving features in natural environments

The ability of animals to deal successfully with complex environments is often the main reason to adopt a biologically-inspired approach to engineering problems. This suggests that the model designed to emulate biological systems should be tested in natural conditions. In the previous chapter I have tested the insect-inspired model in open-loop using videos of natural environments. However, since the main purpose of target tracking algorithms is their implementation in robotic applications, an open-loop evaluation is not a proper indication of model robustness under natural conditions. Closed-loop target tracking is a complicated task which requires sensorimotor control and internal models. Moreover, sensory and actuators latencies as well as the physical robot dynamics might result in additional effects on the stability of the feedback process which can lead to tracking failure. Therefore, in this chapter I implement the insect-inspired model on a robotic platform to examine its performance under real-world conditions. In addition to its engineering side, this hardware implementation allows the exploration of the potential effect of the natural habitat on neuronal behaviour. The supplementary material for this chapter is presented in Appendix E.

Statement of Authorship

Title of Paper	An autonomous robot inspired by insect neurophysiology pursues moving features in natural environments
Publication Status	☐ Published ☑ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Bagheri Z. M., Cazzolato B. S., Grainger S., O'Carroll D. C., & Wiederman S. D. (In Press). An autonomous robot inspired by insect neurophysiology pursues moving features in natural environments. Journal of Neural Engineering.

Principal Author

Name of Principal Author (Candidate)	Zahra Bagheri
Contribution to the Paper	Developed the code for hardware implementation, designed the experiments, conducted the experiments, generated groundtruth, ran simulations, analysed and interpreted the data, wrote the manuscript with editing contributions from the other authors.
Overall percentage (%)	65
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date: 6/7/2017

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Benjamin Cazzolato
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, assisted with analysis and interpretation of the data, provided significant editing contribution (15%).
Signature	Date: 6/7/17

Name of Co-Author	Steven Grainger
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, provided editing contribution (5%).
Signature	Date: 6/7/17

Name of Co-Author	David O'Carroll		
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, provided editing contribution (5%).		
Signature		Date	3-JULY-2017.

Name of Co-Author	Steven Wiederman		
Contribution to the Paper	Participated in the initial conceptualisation and experimental and design, assisted with analysis and interpretation of the data, provided significant editing contribution (10%).		
Signature		Date	6/7/2017

An autonomous robot inspired by insect neurophysiology pursues moving features in natural environments

Zahra M. Bagheri[1,2], Benjamin S. Cazzolato[2],

Steven Grainger[2], David C. O'Carroll[3] and Steven D. Wiederman[1]

[1] Adelaide Medical School, The University of Adelaide, Adelaide, SA 5005, Australia

[2] School of Mechanical Engineering, The University of Adelaide, Adelaide, SA, Australia

[3] Department of Biology, Lund University, Sölvegatan 35, S-22362 Lund, Sweden

Bagheri Z., Cazzolato B. S., Grainger S., O'Carroll D. C., & Wiederman S. D. (In Press). An autonomous robot inspired by insect neurophysiology pursues moving features in natural environments. Journal of Neural Engineering.

5.1 Abstract

Objective. Many computer vision and robotic applications require the implementation of robust and efficient target-tracking algorithms on a moving platform. However, deployment of a real-time system is challenging, even with the computational power of modern hardware. Lightweight and low-powered flying insects, such as dragonflies, track prey or conspecifics within cluttered natural environments, illustrating an efficient biological solution to the target-tracking problem. *Approach*. We used our recent recordings from 'small target motion detector' neurons in the dragonfly brain to inspire the development of a closed-loop target detection and tracking algorithm. This model exploits *facilitation*, a slow build-up of response to targets which move along long, continuous trajectories, as seen in our electrophysiological data. To test performance in real-world conditions, we implemented this model on a robotic platform that uses active pursuit strategies based on insect behaviour. *Main results*. Our robot performs robustly in closed-loop pursuit of targets, despite a range of challenging conditions used in our experiments; low contrast targets, heavily cluttered environments and the presence of distracters. We show that the facilitation stage boosts responses to targets moving along continuous trajectories, improving contrast sensitivity and detection of small moving targets against textured backgrounds. Moreover, the temporal properties of facilitation play a useful role in handling vibration of the robotic platform. We also show that the adoption of feed-forward models which predict the sensory consequences of self-movement can significantly improve target detection during saccadic movements. *Significance.* Our results provide insight into the neuronal mechanisms that underlie biological target detection and selection (from a moving platform), as well as highlight the effectiveness of our bio-inspired algorithm in an artificial visual system.

5.2 Introduction

In recent years, there has been developing interest in the use of mobile robots for applications in industry, health and medical services, and entertainment products. Autonomous robots gather information about their surrounding environment via sensors (e.g. optical, ultrasonic, or thermal sensors), process this information and initiate motor commands to complete specific tasks with a self-determined behaviour. Biological systems employ similar sensory-motor control and autonomy to perform their daily activities. Thus, there is common ground in robotics and biology in understanding how such systems function and reverse engineering biological systems can provide blueprints for robotics applications.

Detecting and tracking a moving object against a cluttered background is among the most challenging tasks for both natural and artificial vision systems. Recent work has drawn inspiration from biological visual systems for the development of robust target tracking algorithms. For example, inspired by bird and fish behaviours, Zheng and Meng (2008) developed a population-based search algorithm, called particle swarm optimization (PSO) and implemented it in an object tracking algorithm. Zhang et al. (2010) proposed a model of target appearance for visual tracking that was inspired by the hierarchical models of object recognition in visual cortex (Riesenhuber and Poggio, 1999). Mahadevan and Vasconcelos (2013) developed a bio-inspired tracker combining bottom-up centre-surround discriminations and a target-tuned top-down saliency detector. Inspired by the fly's visual micro-scanning movements, Colonnier et al. (2015) developed a small-scale artificial compound eye, which estimates displacement by measuring angular positions of contrasting features. The researchers mounted the eye on a tethered robot and tracked contrasting objects (hands) moving over a textured background. More recently, Cai et al. (2016) presented a biologically inspired target tracking model which partially mimics ventral stream processing in the primate brain.

These biologically inspired target tracking studies mainly focus on models of primate vision, however, insects provide an ideal group to draw inspiration from within the context of target tracking. Many species of flying insects, such as dragonflies, are capable of detecting, selecting and chasing tiny prey or conspecifics. This capacity is all the more humbling for robotics engineers considering the insects limited visual resolution (~0.5°) and relatively small size (brain less than 2 mm wide), light-weight and low-power neuronal architecture (Webb et al., 2004). Remarkably, the dragonfly performs this task within a visually cluttered surround, in the presence of distracters (Corbet, 1999; Wiederman and O'Carroll, 2013a) and with a capture rate greater than 97% (Olberg et al., 2000). Such performance motivates the design of an insect-inspired, visual target tracking algorithm for autonomous robot control.

Using intracellular, electrophysiological techniques to record neuronal activity within the insect optic lobe, our laboratory has identified and characterized a set of neurons we refer to as 'small target motion detectors' (STMD) that likely mediate target detection and pursuit. These neurons are tuned to the size and velocity of targets, are sensitive to their contrast, yet can respond robustly to targets even without relative motion between them and a cluttered background (O'Carroll, 1993; Nordström et al., 2006; Nordström and O'Carroll, 2009; O'Carroll and Wiederman, 2014, O'Carroll et al., 2011; Wiederman and O'Carroll, 2011). Inspired directly by these physiological data, we have developed a nonlinear 'Elementary-STMD' (ESTMD) model for local target discrimination (Wiederman et al., 2008) and have implemented this model in a closed-loop target tracking system using a virtual reality (VR) environment (Halupka et al., 2011; Bagheri et al., 2014a; Bagheri et al., 2014b; Bagheri et al., 2015b). We elaborated (Bagheri et al., 2014a; Bagheri et al., 2015a) this closed-loop model to account for recent observations of 'facilitation' in STMD neurons (Nordström et al., 2011; Dunbier et al., 2011; Dunbier et al., 2012). Facilitation involves the spiking

response of these neurons building over several hundred milliseconds as targets move along continuous trajectories and resets to their naive state when there are large spatial or temporal breaks in the trajectory (Dunbier et al., 2011; Dunbier et al., 2012).

Using closed-loop simulations against cluttered natural scenes, we predicted that the optimum temporal properties of facilitation is dependent on the degree of background clutter and the purpose of the pursuit (i.e. predation or mating) (Bagheri et al., 2015a). We also showed that facilitation not only improves pursuit success, it enhances the ability to 'attend' to one target in the presence of distracters (Bagheri et al., 2015a). Simulations reveal robust performance, achieving high prey capture success rates even within complex backgrounds, for low contrast targets, or where the relative speed of pursued prey is high (Bagheri et al., 2015a). We recently benchmarked our model against several state-of-the-art trackers and although less computationally expensive, it matched or outperformed them, particularly when tracking small moving targets in natural scenes (Bagheri et al., 2015b; Bagheri et al., 2017).

Although our model is robust in simulation, the performance in response to uncertainties inherent within real environments (e.g. illumination changes, occlusions, and vibration) are as yet unknown. Moreover, robotic systems are limited by the sampling rate of their sensors, processing and actuators. It is unclear how our algorithm performs on a hardware platform, where inclusion of sensors and physical robot dynamics results in additional latency which may affect the stability of the feedback process.

In addition to engineering applications, bio-inspired robots can provide insight into the underlying, biological, sensorimotor system. For example, dragonflies must deal with turbulent air and vibration during the flight, whilst focusing on the target (Krapp and Wicklein, 2008; Collett, 1980). Furthermore, animals use eye or body movements to modulate the visual inputs (via closed-loop feedback) (Land, 2015). An important question

is how animal saccadic movement or environmental factors change neural responses underlying the detection and selection task. Our ability to investigate these questions in the biological system is limited, due to the open-loop nature of experiments (i.e. animal must be restrained). However, robots provide a suitable alternative to model such sensorimotor mechanisms all whilst embedded within real-world environments.

Here we present an implementation of our recently developed target-pursuit model on an autonomous, robot platform. We test the robot in both controlled conditions using an indoor environment projected with natural images, and in unstructured outdoor environments. We examine the effect of environmental parameters, facilitation and saccadic movement on robot performance. Our results show that even under demanding conditions (e.g. complex background clutter, illumination variation, presence of distracters, and vibration) our robotic implementation performs robustly with a success rate similar to that observed in simulations. Moreover, we identify several key principles for optimal performance of such a system under real-world conditions.

5.3 Methods

Figure 5.1 shows an overview of the hardware implementation of the insect-inspired tracker on a Husky A200 (Clearpath Robotics™) platform using ROS, C++ and OpenCV. A Blackfly camera (Point Grey Research Inc.) was mounted on the robot to capture video of the natural environment. Further details of hardware are provided in Section 5.3.3. Camera output serves as input to the insect-inspired target tracking model which calculates target location. The pursuit algorithm uses target location to calculate a 'saccadic' turn (short and fast yaw turns to change direction of gaze) angle as seen in insect behaviour (Wehrhahn et al., 1982; Land and Collett, 1974; Mischiati et al., 2015). The Matlab code for the insect-

inspired target tracking model is downloadable via https://figshare.com/s/377380f3def1ad7b9d44 (Bagheri et al., 2017).

5.3.1 Insect-Inspired Target Tracking Model

The insect-inspired target tracking model is composed of three subsystems: (1) early visual processing (2) target matched filtering (ESTMD) (3) position selection and facilitation mechanism. Detailed model equations are presented in the Appendix.

5.3.1.1 Early visual processing

The optics of flying insects are limited by diffraction and other forms of optical interference within the facet lenses (Stavenga, 2003). This optical blur was modelled with a Gaussian low-pass filter (full-width at half maximum of 1.4°), which is similar to the optical sampling of typical day-active insects (Stavenga, 2003). The average inter-receptor angle ($\Delta\varphi$) between photoreceptors can vary from tens of degrees in Collembola to 0.24° in the acute zone of the dragonfly Anax junius (Land, 1997). We sub-sampled the captured image at 1° intervals as an approximate match for the resolution of day-active flies (Straw et al., 2006), balancing acuity with computational efficiency of the algorithm. The green spectral sensitivity of the motion pathway in flying insects was simulated by processing only the green channel of the RGB imagery (Srinivasan and Guy, 1990). This pre-processing is considered as the 'model input' to the target tracking algorithm in further analyses.

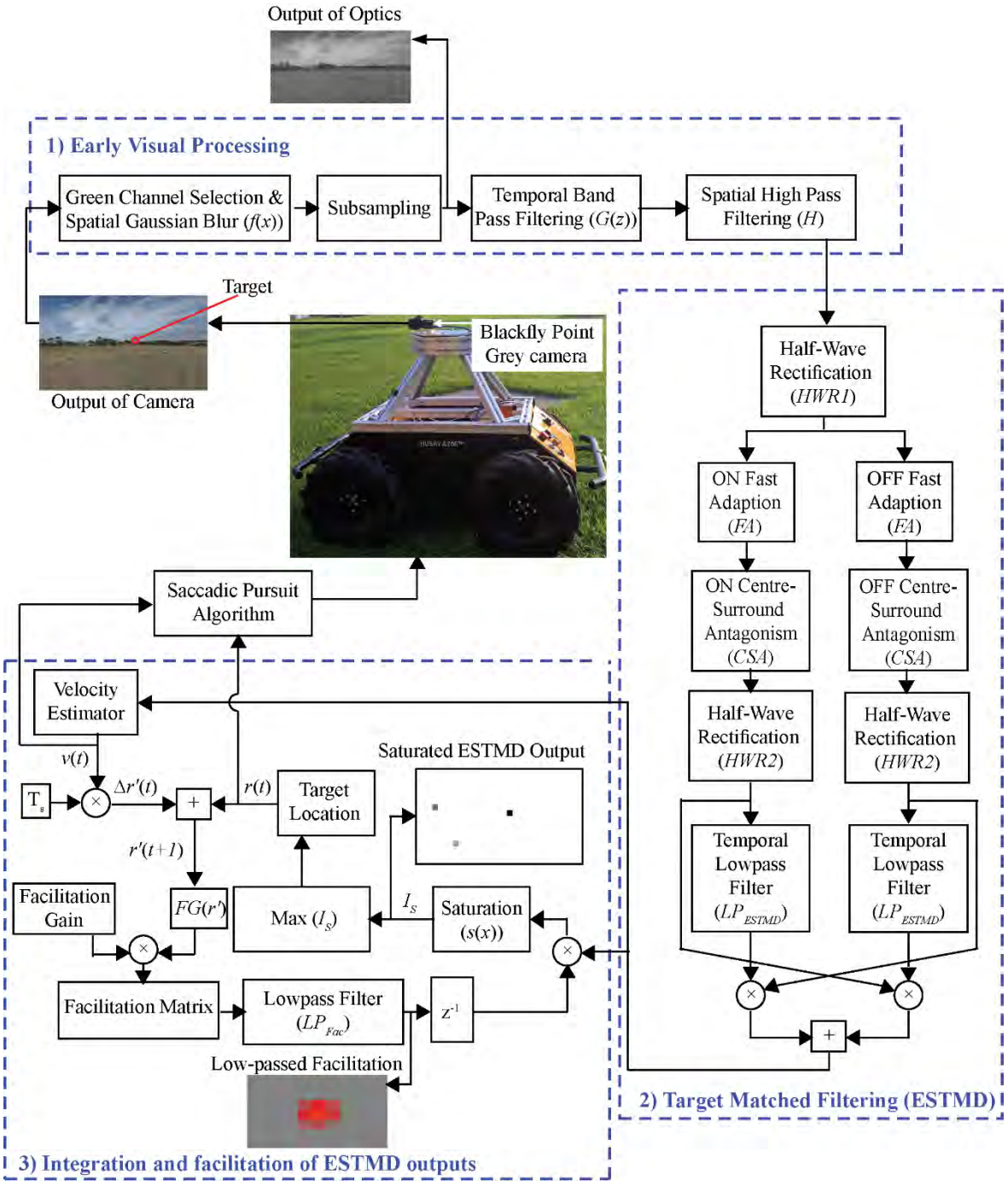

Figure 5.1. Overview of the hardware implementation of the insect-inspired tracking model. The robotic platform is a Clearpath Husky robot. A Point Grey Blackfly camera was mounted on the robot to capture input imagery. The camera output is processed by the insect-inspired, target-tracking model which calculates the target location**.** The insect-inspired target tracking model includes 1) Early visual processing which mimics the response characteristics of optics, photoreceptors and large monopolar cells (LMC) in flying insects. 2) The ESTMD stage which includes rectification, fast adaptation and centre-surround antagonism provides a matched spatiotemporal filter tuned for small moving targets. 3) The recently observed 'facilitation mechanism' is

modelled by building a weighted map (lowpass filter, time constant=τ_f) based on the predicted location of the target in the next sampling time ($r'(t+1)$). The predicted target location was calculated by shifting the location of the winning feature ($r(t)$) with an estimation of the target velocity vector ($v(t)$) provided by the Hassenstein-Reichardt elementary motion detector which was multiplied with sampling time (T_s). The output of ESTMDs is multiplied with a low-pass version of a weighted map. The time constant of the facilitation low-pass filter controls the duration of the enhancement around the winning feature. The pursuit algorithm calculates the saccadic turn angle based on the detected target location and direction of target motion.

Simulating biological vision (Srinivasan et al.,1982), redundant information was removed with neuronal adaptation (temporal high pass filtering) and centre-surround antagonism (spatial high pass filtering). We simulated the temporal properties of photoreceptors and the 1st-order interneurons, the large monopolar cells (LMCs), with a discrete log-normal function (*G(z)*) (Halupka et al., 2011). The filter properties were matched to the temporal impulse response observed in LMC recordings (James, 1990). Centre-surround antagonism as observed in physiological recordings was modelled by subtracting 10% of the centre pixel from the neighbouring pixels (*H*) which provides a zero DC spatial component.

5.3.1.2 Target matched filtering (ESTMD stage)

Rectifying transient cells (RTCs) within the insect 2nd optic neuropil (medulla) exhibit processing properties well suited as additional input processing stages for a small target motion detection pathway (Wiederman et al., 2008). RTCs exhibit independent adaptation to light increment (ON channel) or decrement (OFF channel) (Osorio, 1991; Jansonius and van Hateren, 1991). The separation of the ON and OFF channel was modelled by half-wave rectification (*HWR1*). Each channel was processed through a fast adaptive mechanism, with the state of adaptation determined by a nonlinear filter that switches its time constant. Time constants were 'fast' (τ_{FA}=3 ms) when channel input is increasing and 'slow' (τ_{FA}=70 ms) when decreasing. This adaptation causes subtractive inhibition of the unaltered 'pass-through' signal. Additionally, we implemented strong spatial centre-surround antagonism,

with each channel surround inhibits its next-nearest neighbours (see Wiederman et al. (2008) for details). The temporal adaptation reduces responses to background texture, while strong surround antagonism conveys selectivity for local edge features (i.e. features that are small in the dimension orthogonal to the direction of travel). A second half-wave rectification (*HWR2*) was applied to the output of the strong centre- surround antagonism to eliminate the negative values (a thresholding nonlinearity observed in spiking responses).

In the direction of travel, small targets are characterized by an initial rise (or fall) in brightness, and after a short delay are followed by a corresponding fall (or rise). This property of small features was exploited by multiplying each contrast channel (ON or OFF) with a delayed version of the opposite channel (delayed using a low-pass filter (LP_{ESTMD}), τ_{ESTMD}=25 ms) and then summing the outputs. This also confers sensitivity to targets independent of the polarity of their contrast against the background.

5.3.1.3 Integration and facilitation

Neuron-like soft saturation of ESTMD outputs was modelled with a hyperbolic tangent function ($S(x)$), ensuring all signals lie between 0 and 1. The target location was calculated by implementing a simple competitive selection mechanism which chooses the maximum of the output values across the visual field. In the insect target-detecting system, there is a retinotopic array of small-field STMDs (SF-STMDs) which we hypothesise integrates local outputs of a number of underlying ESTMDs (~10°x10° region) (Barnett et al., 2007, O'Carroll, 1993).

A facilitation mechanism as seen in biological STMDs (Nordström et al., 2011; Dunbier et al., 2011; 2012) was implemented by building a weighted 'map' dependent on the location of the winning feature but shifted by a target velocity vector (an estimation of future target location). In the insect system, the visual information could be facilitated by a retinotopic array of SF-STMDs with overlapping receptive fields. We modelled the spatial extent of the

weighted map by a grid of 2D Gaussian kernels (half-width at half maximum of 5°) with centres separated by 5° (*FG(r')*), providing a near optimum spatial size for the facilitation mechanism (Bagheri et al., 2015a). This is equal to 50% overlap between receptive fields of SF-STMD neurons. The ESTMD output was multiplied with a low-pass filtered (LP_{Fac}) version of this facilitation map. The facilitation time constant (τ_f) controls the duration of the enhancement around the location of the winning feature. Four different time constants (varied in the range 40 to 2000 ms) spanning the typical facilitation time course (~200 ms) observed in dragonfly STMDs (Dunbier et al., 2012) were tested in the experiments.

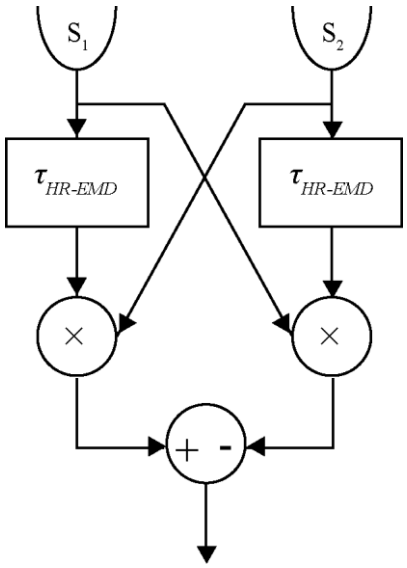

Figure 5.2. The HR-EMD were used to estimate the velocity of the target. HR-EMD employes two spatially separated signals (S_1, S_2) and correlates them after a delay (via a low-pass filter, $\tau_{HR\text{-}EMD}$=40 ms) resulting in a direction selective output (Hassenstein and Reichardt, 1956). Subtracting the two mirror-symmetric sub-units yields positive response for the preferred motion direction (in this case left to right) and a negative one in the opposite direction (right to left).

To provide an estimate for the future location of the target, we used the Hassenstein-Reichardt elementary motion detector (HR-EMD) (Hassenstein and Reichardt, 1956). The HR-EMD uses two spatially separated contrast signals and correlates them after a delay (Figure 5.2). In our model, the HR-EMD was applied as a 2nd-order motion detector on the

ESTMD (1st order motion detector) outputs (via a low-pass filter, $\tau_{HR\text{-}EMD}$=40 ms) resulting in a direction selective output. Although the HR-EMD confounds spatial attributes of the target (e.g. size in the direction of travel) it is also tuned to the velocity of the pre-filtered 'small targets' (Wiederman and O'Carroll, 2013b). The output of HR-EMD is positive when the target moves from S_1 to S_2 (Figure 5.2), and negative in the reverse direction. The HR-EMD was applied in both horizontal and vertical directions to estimate whether the target is moving right/up (positive) or left/down (negative). The spatial component of the target velocity vector was determined by binning the magnitude of the output of the HR-EMD into three equal intervals, to estimate whether the speed of the target is slow, medium or fast (Bagheri et al., 2015), a strategy similar to that observed behaviourally in crabs (Nalbach, 1989).

5.3.2 Saccadic Pursuit Algorithm

Flying insects use various pursuit strategies to control their forward velocity and distance, whilst fixating the target in the frontal visual field. For example, a male housefly uses a 0° 'tracking' strategy to chase another fly, resulting in complex looping pursuit paths (Wehrhahn et al., 1982; Land and Collett, 1974). An aerial predator, such as a perching dragonfly, uses an 'interception' strategy that maintains the prey at a fixed relative bearing (Mischiati et al., 2015). Inspired by these strategies, we implemented a hybrid pursuit mode (Bagheri et al., 2015a), where the robot initiates a frontal fixation saccade whenever the winning feature of the ESTMD output moved more than 5° from the centre of the field of view. This strategy keeps the target close to the pole of expansion in the flow-field generated by the pursuer's own progressive motion through the world, i.e., where local background image speeds are lowest. The fast adaptation mechanisms in the earlier visual processing then enhance target 'pop out' against a highly cluttered background during the inter-saccade period.

5.3.3 Experimental Setup

The robotic platform was a Husky A200 (Clearpath Robotics™) unmanned ground vehicle which uses an open source serial protocol. The development framework for this robot was Robot Operating System (ROS). A Blackfly camera (BFLY-U3-13S2C-CS, Point Grey Research Inc) with a CS mount 1/3" sensor (53°x100° sized viewport) was mounted on the robot to capture videos of the environment (Figure 5.1). Due to technical limitations of the camera, video was sampled at 20 Hz to represent the visual field of the robot, which moved at a velocity of 0.1 ms^{-1}. Although real-time, our autonomous system operated in a 'slowed down' environment (limited by the camera frame rate) with tracking of the output target location at a corresponding 20 Hz. Different size and colour foam balls were used as targets (see Table I for detailed parameters). The balls were fixed to a thin transparent line, wrapped around a motorized driving pulley (Figure 5.3). Two motor speeds were tested, resulting in target velocities of either 0.06 ms^{-1} ('*slow*') or 0.12 ms^{-1} ('*fast*'). The target track was mounted at a height which varied between 50-150 cm above the ground. Five idler turning pulleys changed the direction of the target path (Figure 5.3). Six different paths were tested for each set of environmental and model parameters. The pursuit was scored as a success if the robot passed within 1 m of the frontally fixated target, before the target completed one cycle.

Experiments were conducted both in an indoor environment under controlled conditions and in unstructured outdoor environments. For indoor experiments, images or videos of natural scenes (Figure 5.4a, 5.4b, Movie 1, and Movie 2 of the supplementary material) were projected (2 projectors) onto a wall as a backdrop for the target (image statistics, Appendix E, Table E1.) Outdoor experiments were conducted in four different locations at different times during the day throughout a month. Figure 5.4c shows images of these locations taken by the mounted robot camera.

Table 5.1. Target size and colours used for different experiments.

Target	**Colour**	Black		Dark Grey		Medium Grey		Light Grey		White	
	Diameter (mm)	65	100	65	100	65	100	65	100	65	100
Experiment	**Indoor**	✓	✓	✓		✓					
	Vibration	✓	✓								
	Outdoor		✓		✓		✓		✓		✓

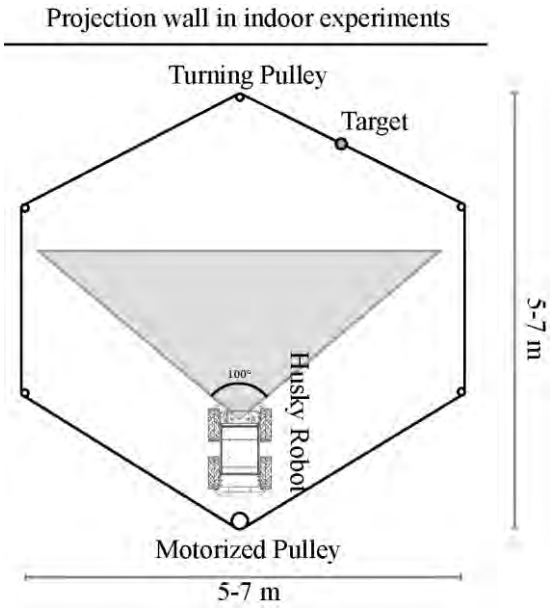

Figure 5.3. Experimental setup. A monofilament fishing line, wrapped around a motorized driving pulley moves the target along the track.

.

Figure 5.4. Backgrounds for experiments a) images and b) snapshots of videos of natural environments projected onto the wall in indoor experiments. c) Video output of robot showing the environmental conditions in each outdoor experimental location.

5.4 Results

5.4.1 Indoor Experiments

5.4.1.1 Effect of facilitation kinetics on pursuit success

Previous modelling efforts highlighted that optimum temporal parameters for facilitation (i.e. the duration of the enhanced region around a selected target) should be dynamically modulated based on the amount of background clutter and the target velocity (Appendix E, Figure E1 and Bagheri et al., 2015a). To validate simulation results and test our hardware implementation, we conducted indoor robotic experiments by projecting the same natural images onto the wall (Figure 5.4a).

Figure 5.5 shows the pursuit success rate averaged over the four targets used in these experiments (varying size and colour , Table 5.1). As in our previous simulations (Figure E1), images that include either naturalistic ('*Botanic*') or urban ('*Forecourt*') clutter, were more likely to evoke false positives. In contrast, sparse images ('*Field*') or those composed predominantly of straight edges ('*House*'), elicited fewer false positives. Consequently, *Field* and *House* have higher maximum average success rates than *Botanic* and *Forecourt*.

Although projecting images on the wall results in lower contrast background features, we observed similar results to those described in our previous virtual-reality simulations (Bagheri et al., 2015a). In all but the sparsest scene (*Field*), as target velocity increased the optimal time constant decreases (Figure 5.5, cf. red and green lines). A lower time constant is required to 'keep up' with the faster moving target. Additionally, the robot experiments show that the optimal facilitation time constant changes across background images. At either target speed, pursuit success improves with a longer time constant in cluttered images (*Botanic* and *Forecourt*), compared to sparser images (*Field* and *House*). This reveals that a longer facilitation time constant enhances the region around a camouflaged target for a longer duration, thus permitting reaquisition.

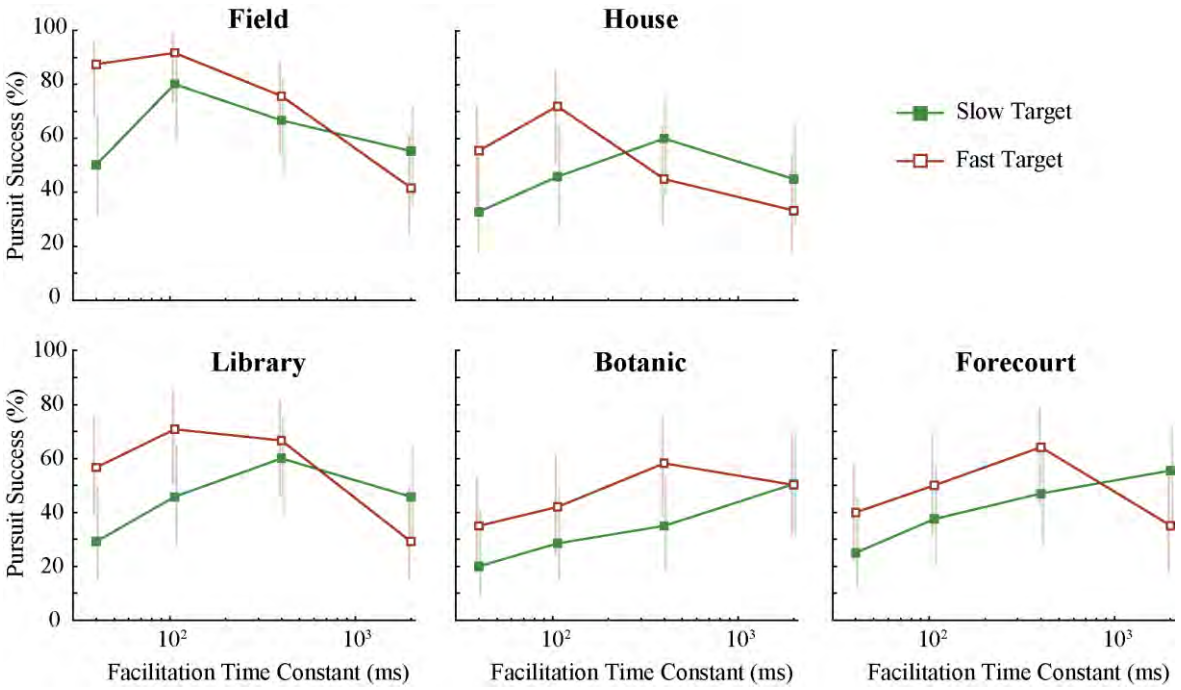

Figure 5.5. Pursuit success of indoor experiments against projected natural scenes for two target velocities (0.06 ms^{-1}, 0.12 ms^{-1}) with changes in facilitation time constant (±95% CI, adjusted Wald, n=24) across target contrasts ranging from high to low (Table 5.1). As expected, the robot has a higher pursuit success when tracking targets against less cluttered scenes (e.g. Field). These results show that the optimum facilitation time constant varies, dependent on both target velocity and the background scene.

5.4.1.2 Sensitivty to Vibration

A challenge for visual target-tracking from a mobile platform is dealing with motion blur and uncertainty in target location arising from environmental forces (e.g. wind) and vibration (Irani et al., 1992). To quantify the effect of vibration, we tested the robot on two artificially created, uneven surfaces (Figure 5.6a) and compared the results with a flat surface. The first (‘*Surface 1*’) includes numerous small obstacles (up to 30 mm) taped to the floor to generate brief bumps, whilst *Surface 2* includes both bumps and 15 mm to 30 mm stones randomly scattered.

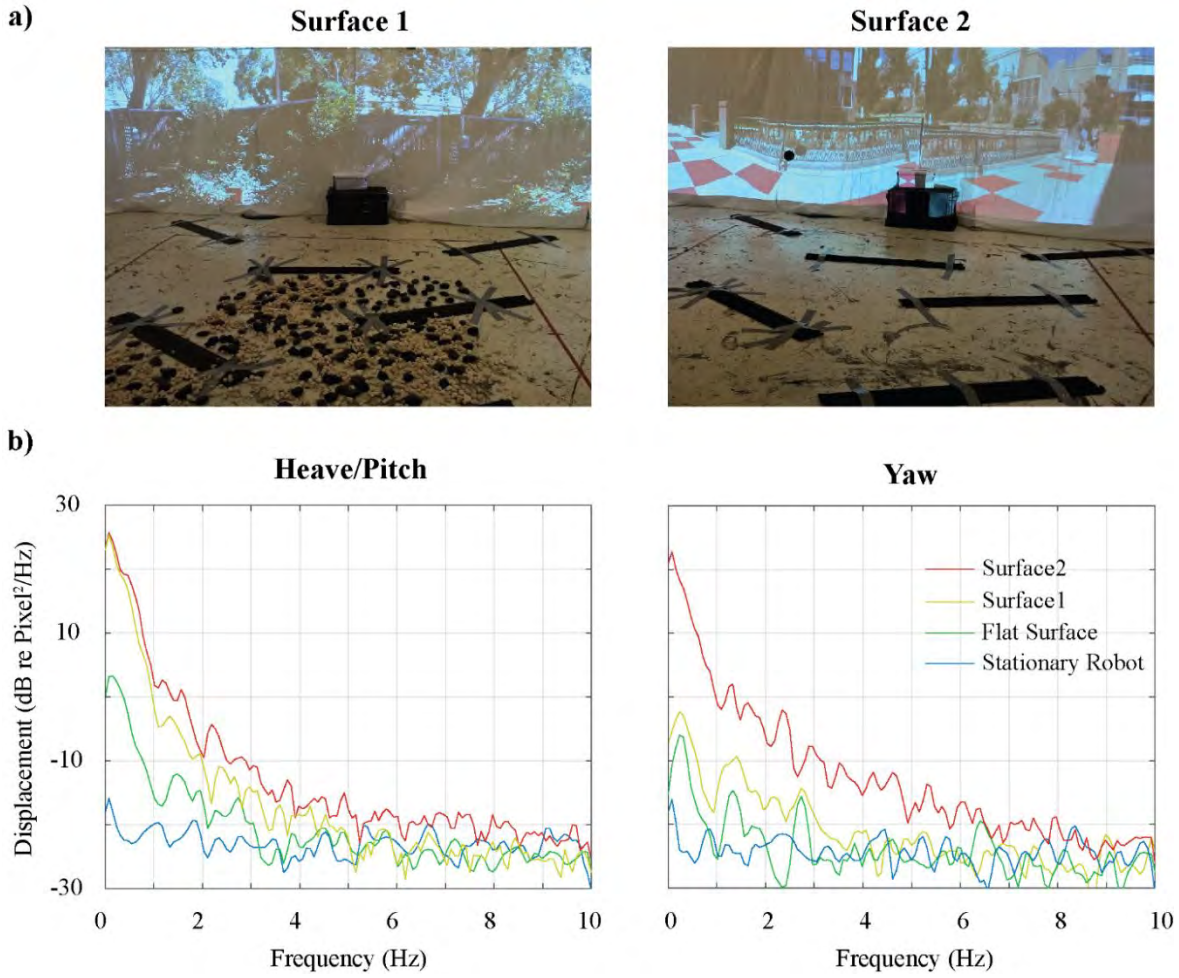

Figure 5.6. Vibration sensitivity. a) Two artificially created uneven surfaces that were used to quantify the effect of vibration on model performance. b) Spectral density of target displacement in the output of camera, resulting from heave/pitch and yaw vibration of the robot.

Figure 5.6b shows the spectral density of the surface-induced target displacement (robot heave/pitch and yaw) in the output of the camera. The target location in the input image was determined manually which results in low amplitude noise (stationary robot). The *Flat Surface* condition provides a baseline measure of vibration and disturbances arising from robot components (torque timing belt and wheel tread). *Surface 1* causes high amplitude transient events in heave/pitch but does not have any significant effect on yaw. *Surface 2* induced significant mid-frequency (1-3 Hz) vibration on both heave/pitch and yaw motion.

Figure 5.7 shows that pursuit success varies for different surfaces. *Surface1* slightly reduces the robot efficacy compared to *Flat Surface* (~ 8% at optimum time constant) due to the low frequency disturbances. However, as the difficulty of the surface increases (*Surface 2*) the effect of vibration becomes more prominent. The motion blur and uncertainties in target location caused from significant yaw events results in up to 25% reduction of pursuit success compared to *Flat Surface*. Interestingly, the optimum facilitation time constant decreases for *Surface 2*, meaning that a faster facilitation mechanism is required to keep up with the frequent changes of target location in the camera output.

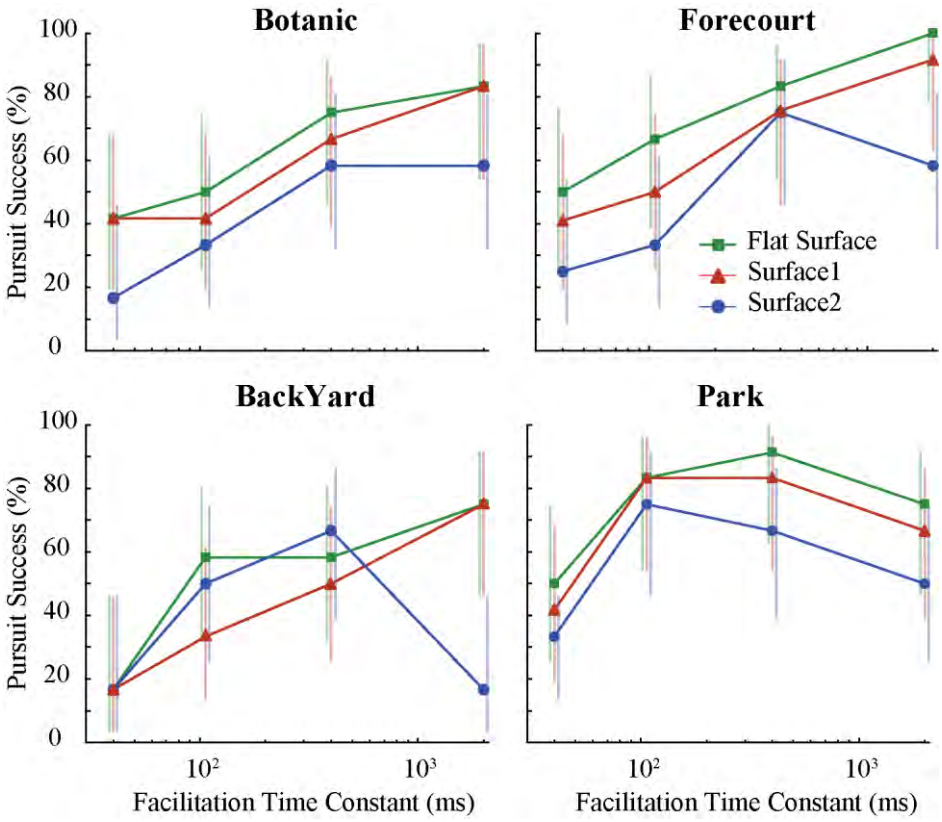

Figure 5.7. Effect of vibration on robot performance. Presence of uneven surfaces degrades performance (±95%CI, adjusted Wald, n=12). For *Surface 2*, a shift towards shorter time constants permits the model to withstand sudden changes in target location arising from robot vibration.

5.4.2 Outdoor experiments

We were able to control environmental parameters (i.e. illumination, clutter, vibration) in our indoor experiments and test their effect on robot performance. However, natural environments are very dynamic and the pursuer must deal with these challenges. To examine model performance in unstructured environments, we set our robot to run autonomously in several outdoor locations. An example of the robot footage as well as a video output of robot can be seen in supplementary material (Supplementary Movie 3 and Movie 4, respectively). Examples of robot trajectories in the experiments are shown in Appendix E, Figure E2.

5.4.2.1 Contrast Sensitivity in Open-loop Simulations

Illumination has a significant impact on the appearance of surfaces, as specular reflections and shadows change. This imposes an additional challenge when testing the influence of model parameters on target-tracking success in unstructured environments and elucidating causes of pursuit failure. One way to deal with this problem is to investigate correlations between target contrast and model performance.

To measure the contrast sensitivity of our model, we simulated an immobilized pursuer viewing targets of varying contrast drifting horizontally against a white background (for 550 ms). The data (Figure 5.8a) show that the addition of facilitation substantially increases model contrast gain. With a fast facilitation mechanism (τ_f=40 ms), the target contrast required to evoke 50% maximal response (C50) is very low (~0.15). However, as the time constant increases the sensitivity to low contrast targets decreases, such that C50 for the most sluggish facilitation time constant (τ_f=2000 ms) is 0.38. The C50 increases to approximately 0.42 for the non-facilitated model and the model never reaches the maximum response. In all cases, the model does not respond to targets with contrasts equal or lower than 0.1.

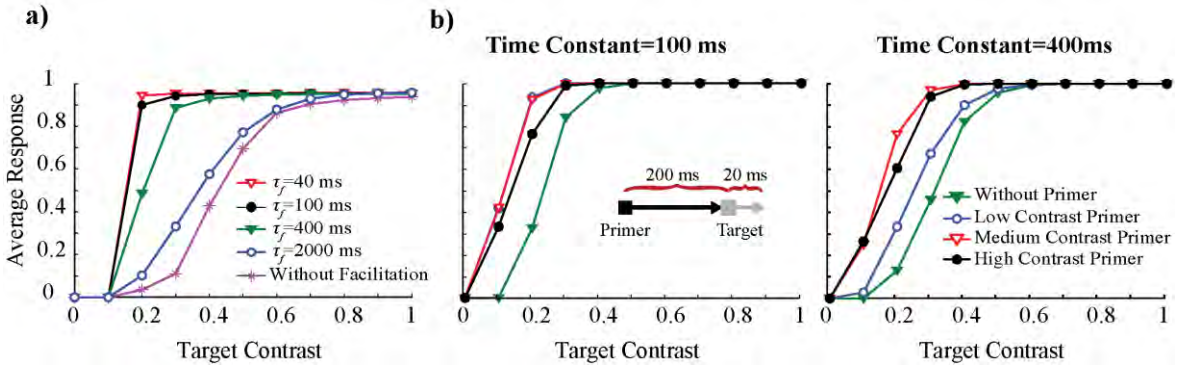

Figure 5.8. Contrast sensitivity of the model a) Open-loop model response increases as target contrast increases (purple line). The addition of facilitation (at various time constants) increases contrast sensitivity. However, there is a hard threshold at a target with contrast of 0.1. b) Targets of varying contrast were preceded by either a low (blue), medium (red) or high (black) contrast primer. Facilitation increases model contrast sensitivity and threshold.

The addition of a facilitation mechanism clearly increases the contrast sensitivity, but is the strength of this enhancement dependent on the strength of the facilitating (priming) target? To answer this question, we started the simulation with a stimulus of a low (0.2), medium (0.5) or high (1) contrast (primer) and varied the stimulus contrast after 200 ms (Target in Figure 5.8b). Figure 5.8b shows the average model responses to the tested target contrasts for a 20 ms interval immediately after priming.

Facilitation mechanisms with a short time constant build up faster and elicit higher responses. In the presence of a primer, facilitation builds prior to the contrast variation, thus improving the minimum contrast sensitivity (the model responds to a target contrast of 0.1). Interestingly, the low and medium contrast primer elicit a higher response than the high contrast primer, as also observed in physiological experiments (personal communication). This peculiarity of the contrast sensitivity functions observed in both modelling and physiology, emerges from the combined effects of fast temporal adaptation (ESTMD stage) and facilitation.

5.4.2.2 Contrast Sensitivity in Closed-loop Robotic Experiments

Targets in natural scenes vary in contrast which may induce an attentional 'switch' to a false positive in the scene, consequently resulting in pursuit failure. Figure 5.9a shows box-and-whiskers plots summarizing the average target contrast during pursuits (C_T, n=120). The central mark (red line) is median of average contrast, edges are the 25th and 75th percentiles, and whiskers are the non-outlier data range. Using our robotic platform, we quantified the effect of target contrast in closed-loop pursuits. Average target contrast during pursuits (C_T, Figure 5.9a) at the model input stage (following optical blur) were segmented into 5 bins and plotted against pursuit success (Figure 5.9b). Unsurprisingly, high contrast targets result in high capture success (~100% at 0.4-0.5 C_T). When target contrast is low, changes in facilitation time constant have little effect, due to detection failure. The largest effect of facilitation time constant is when target contrast is greater than 0.3. However, unlike the results of our contrast sensitivity (Section 5.4.2.1) a shorter time constant is not necessarily beneficial. This reflects the effect of target velocity and background image statistics on the optimum facilitation time constant (Section 5.4.1.1).

The robot can succeed during pursuit of low contrast targets ($C_T < 0.1$), albeit at a low rate (~10%). Because target contrast changes during pursuits (dependent on background and overall illumination), it is feasible that facilitation builds in response to a high contrast target and 'locking-on' to a target that decreases in contrast later in the pursuit (Section 5.4.2.1, Figure 5.8b).

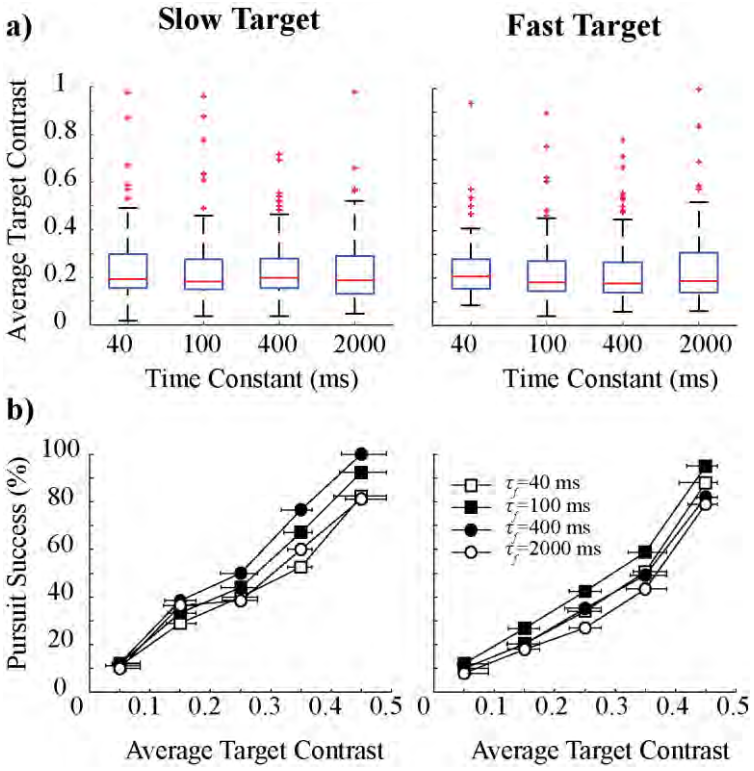

Figure 5.9. Effect of contrast on closed-loop robotic target tracking. a) The range of average target contrast during experiments (C_T, n=120). Although illumination, and therefore target contrast, changes during experiments, the median of average contrast is similar for different experimental parameters (i.e. time constant and target velocity). b) The average target contrast during pursuits were organized into 5 bins with 0.1 contrast intervals and the percentage of successful pursuits in each bin was calculated. The error bars show the deviation of contrast from the centre of each bin.

5.4.2.3 Overall Performance

Figure 5.10 shows the 2D histograms of input imagery statistics in outdoor robotic experiments and Table 5.2 summarizes the overall performance of our robot with the optimum time constant for each target velocity (*slow* and *fast*). Clutter was measured using the method developed by Silk (1995). Despite the very challenging conditions in our experiments (low contrast in highly cluttered backgrounds) our robot is capable of detecting the target in ~42% and ~36 % of the frames for the *slow* and *fast* target respectively. The more successful detection of the slower target is due to the size and velocity tuning properties of our algorithm, which is optimally tuned to target sizes of ~3-4° moving at velocity of ~

70-100 °/s (Bagheri et al., 2017). The maximum pursuit success of our autonomous robot is 56.7%, similar to our previously reported simulation results against images with comparable statistics (i.e. *Library*, *Forecourt*, and *Botanic*, see Bagheri et al. (2015a), Appendix E Table E1, and Figure E1). Impressively, this close similarity in the results is despite the challenges that exist in real-world conditions compared to the idealized simulation signal. This illustrates the robustness of this model when dealing with real-world challenges such as illumination variation, occlusions, presence of distracters and direct sunlight.

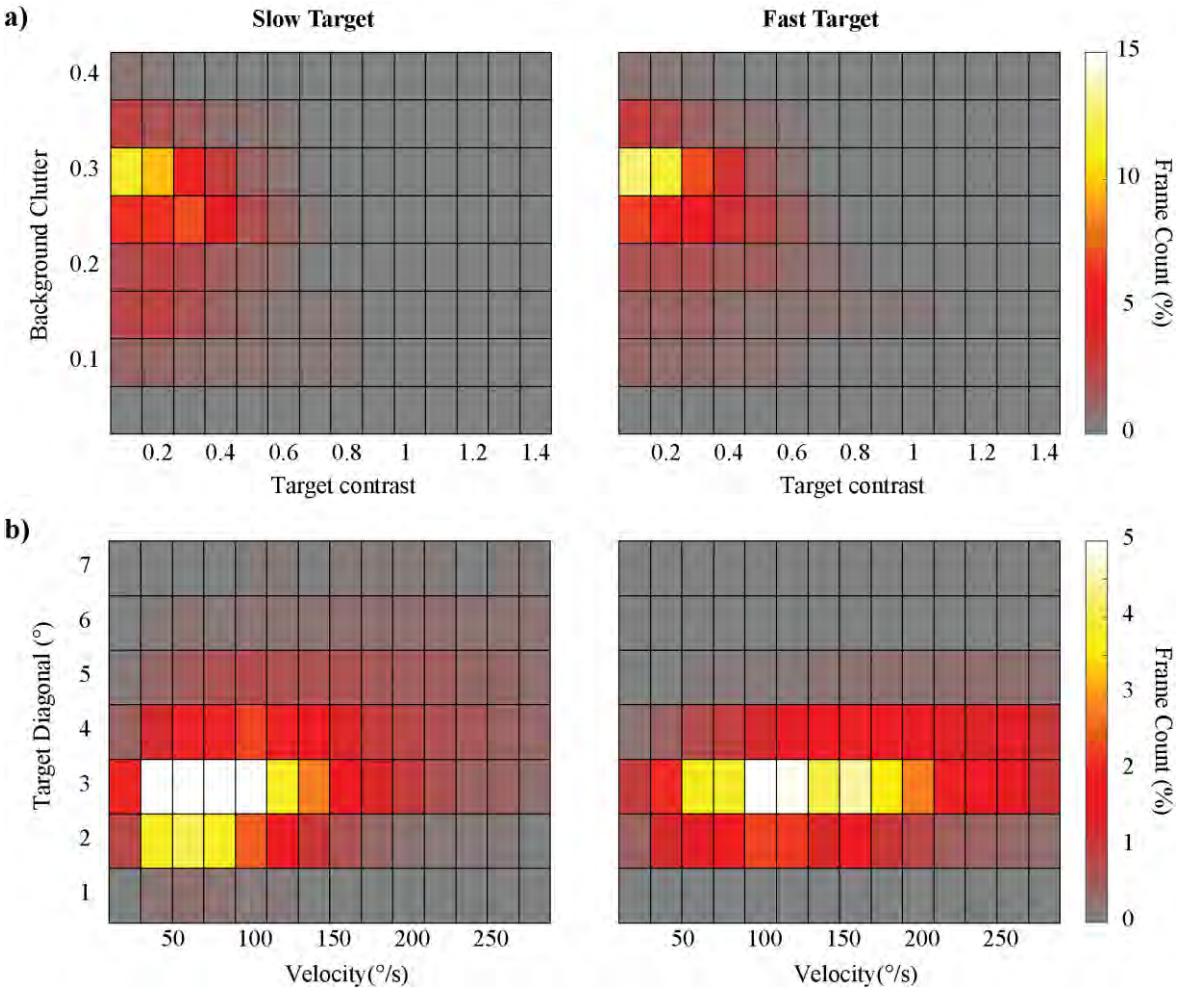

Figure 5.10. 2D histogram of input imagery statistics in outdoor experiments showing a) background clutter and target contrast b) the target size and velocity in the video output of the camera. Our size and velocity tuned model responds optimally to targets of ~3-4° moving at a velocity of ~ 70-100 (°/s) (Bagheri et al., 2017).

Table 5.2. Summary of experimental results for the optimum facilitation time constants.

Measure	Target Speed	
	Slow	**Fast**
Average Target Contrast	0.24	0.26
Average Background Clutter	0.28	0.27
Total Number of Frames	147828	110103
Successful Frame (%)	41.9	36.4
Pursuit Success (% ±95%CI, adjusted Wald, n=120)	56.7±8.7	44.2±8.7
Average Successful Pursuit Time (s)	52	64

5.4.2.4 Effect of Internal Models on Target Detection

During flight, insects make saccadic head movements to keep their target at a specific angular position on the eye (Wehrhahn, et al., 1982; Land and Collett, 1974, Mischiati et al., 2015). However, these saccadic movements cause frequent and substantial displacement of the retinal image. Thus, the movements require an anticipatory shift of visual attention from the pre-saccadic to post-saccadic locations. Such prediction and planning, essential to the high-performance control of behaviour, require internal models. It was recently discovered that flying insects rely on such internal models to guide actions (Mischiati et al., 2015; Kim et al., 2015).

We tested the effect of internal models on target tracking by feeding video captured by the robot camera (n=20) into an open-loop version of our target tracking model. The closed-loop robotic implementation allows predictive feed-forward relocation of the facilitation map to the post-saccadic location based on motor outputs. However, such information is unknown to the open-loop model, thus requiring both detection of the target and establishment of facilitation at the new post-saccadic location. Figure 5.11a shows that the median of successful detection in the open-loop system drops to 48% compared to 59% in the closed-

loop scenario, thus showing the essential role of internal models in target detection in tracking.

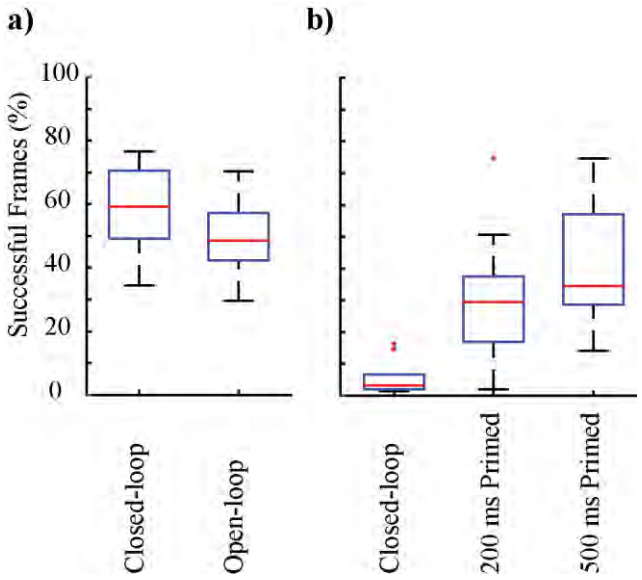

Figure 5.11. a) Effect of internal models on target detection. The closed-loop model predicts the new location of facilitation mechanism during saccades based on motor outputs while the open-loop model does not include the internal model of the sensorimotor system. b) Facilitation mechanism enhances the contrast sensitivity and therefore target detection in the presence of background clutter and distracters. The 200 ms and 500 ms primed facilitation mechanism increases successful detection of the target compared to un-primed closed-loop experiments.

5.4.2.5 Facilitation and Attention

Previously, we proposed a role for facilitation in the selective attention (Bagheri et al., 2015b) observed in dragonfly CSTMD1 neurons (Wiederman and O'Carroll, 2013a). This selection could be induced by a facilitation mechanism increasing contrast sensitivity around one target (Section 5.4.2.1), concomitant with surround suppression. We tested the effect of increased contrast sensitivity on target detection and tracking in natural environments and in the presence of distracters by using the model in open-loop. We used video imagery captured from the robot camera (n=10) in which the model struggles to detect the target in the early stages but locks on to it later during the pursuit (e.g. supplementary material, Movie 4).

These videos served as input to the open-loop simulations. Prior to the start of these simulations, we allowed the facilitation to build up in the initial target location for either 200 or 500 ms (primed). This was implemented by artificially feeding the target location in the first frame as the future location of the target ($r'(t+1)$ in Figure 5.1) to the facilitation mechanism. We measured the successful target detection within the frames prior to the model lock-on to the target. Figure 5.11b shows the difference between the primed (open-loop) and un-primed (closed-loop robotic experiments) model. The 200 ms and 500 ms primed facilitation increase the median of successful target detection to 29% and 34% compared to 3% in un-primed scenarios. These results show that presence of facilitation can direct the attention to target location irrespective of the presence of other distracters.

5.5 Discussion

Our recent benchmarking study (Bagheri et al., 2017) demonstrated that when tracking small moving targets in natural scenes (using open-loop simulations), our insect-inspired model exhibits robust and efficient performance, outperforming state-of-the-art tracking algorithms. The success of these simulations led us to test the efficacy and robustness of our autonomous hardware platform, within a complex, natural environment. Our data show that this system can effectively handle noise from a variety of sources (e.g. vibration, illumination) and can successfully track targets despite the challenging experimental conditions, such as, low target contrast, heavily cluttered environments and the presence of distracters.

5.5.1 Facilitation Time Constant

Using closed-loop simulations we previously showed that the optimum choice of facilitation time constant depends on both target velocity and background clutter (Bagheri et al., 2015a). The results of our closed-loop robotic experiments suggest a similar relationship. Here we

used several variants of facilitation time constant to find the optimum. In future robotic development, we aim to dynamically estimate the facilitation time constant, using real-time estimates of target velocity and background clutter. Given the efficiency of our model (Bagheri et al., 2017) such a dynamic modulation may increase the robustness of pursuit without affecting the stability of the feedback process. However, from a physiological perspective, determining whether STMD neurons use such a dynamic modulation requires further experimentation.

5.5.2 Velocity Estimation

We predicted the target's future location using an estimate of current target velocity, because our most recent physiological data indicates such a velocity dependency (unpublished observations). Rather than calculate a continuous velocity range, we used an HR-EMD model to categorize the target velocity into either 'slow', 'medium', or 'fast speed'. Similar banding into separate channels (temporal frequency) has been observed to underlie velocity estimation in crabs (Nalbach, 1989). Although the output of HR-EMD is a function of the velocity of the moving stimulus, this relationship is not monotonic and has a strong dependence on spatial structure and contrast of the stimulus. These shortcomings have led to elaborations of the original HR-EMD (Zanker et al., 1999; Rajesh et al., 2005; Brinkworth and O'Carroll, 2009). We hypothesised that our model's coarse spatial size of the facilitation matrix (15° by 15° area) would accommodate errors resulting from the texture dependency of the HR-EMD. However, in future work we will incorporate improved velocity estimator methods, derived from our developing understanding of the physiological system. Such methods may include elaborated HR-EMD or time-of-travel models (Vanhoutte et al., 2017; Roubieu et al., 2013; Viollet et al., 2014) and we will test their impact on the target tracking model

5.5.3 Effect of Vibration on Target Tracking

Whether flying, walking, or swimming, animals maneuvering in natural environments must deal with forces and vibration from uneven terrian or turbulence (in air or water). Vibration can impair the acquisition of information (e.g., by the eyes), the outcome of motor command of information (e.g., head or body movements) or the complex central processes that relate input sensory information to the output motor command.

The results show that our model can robustly track targets even in the presence of vibration. As vibration-induced target displacement increases (e.g. *Surface 2* in Figure 5.6a), a shorter facilitation time constant (on the order of 100-400 ms) improves the robustness of tracking. This match for the physiologically measured facilitation time (on the order of 300-500 ms) reported in our earlier work (Nordström et al., 2011; Dunbier et al., 2011; Dunbier et al., 2012) might reflect the evolution of a facilitation time constant in flying insects that allows them to deal with turbulent air and movement of the head induced by wing movements, while tracking prey or conspecifics.

The effect of vibration on vision may be decreased by reducing its transmission to the head / compound eye. Studies of fly behavior show that they control their direction of flight along with gaze through short, fast saccadic movements where head and body turn independently (Van Hateren and Schilstra, 1999). This uncoupling of the eye from its support enables the insect to maintain the orientation of the gaze even when disturbances occur which affect its body. Moreover it reduces the temporal blurring effects and may promote ‘popout’ of a target against a background as a result of the high-pass filtering at early stages of visual processing. Future robotic efforts will investigate the effect of such ‘active vision’ on handling vibration.

5.5.4 Internal Models

Most animals with good vision show a pattern of stable fixations with fast saccades that shift gaze direction (Land, 1999; Land and Collett, 1997; Findlay and Gilchrist, 2003). An essential component of successful target tracking is the ability to distinguish self-induced motion (such as rotation of the visual field caused by saccadic movement) from those imposed by the environment (e.g. vibration). When an insect drifts from its flight trajectory (due to environmental factors), the optomotor response maintains aerodynamic stability and compensates the animal's deviation from its initial flight trajectory (governed by an inner-loop control system). However, performing voluntary manoeuvres by the insect requires interaction between the control of visual reflexes (inner-loop) and other visually guided behaviours (outer-loop), otherwise the animal would be trapped by its inner-loop control system (von Holst and Mittelstaedt, 1950). The question is how the inner-loop optomotor pathway and an outer-loop pathway involved in chasing behaviour may interact? Although different interactions between inner-loop and outer-loop are proposed, recent studies (Mischiati et al., 2015; Kim et al., 2015) support a method postulated by von Holst and Mittelstaedt (1950). von Holst and Mittelstaedt (1950) proposed that with each motor command to initiate a voluntary locomotor turn, a copy of the motor command is sent to the visual system (efference copy), a concept similar to forward models in human motor control.

An important feature of forward modelling is that system outputs modulate sensory processing. Therefore, the visual system is not a feed-forward model driven by the sensory input alone, rather motor outputs are required to predict expected visual input. Our results show a significant improvement in target detection and tracking when an internal model is compared to a feedforward model. However, implementation of internal models remains an uncommon approach in artificial vision systems.

5.5.5 Dynamics of the Robot

Dragonflies are light, swift and agile animals with flights speeds up to 6.8 m/s (Dean, 2003). Robust detection of targets at such high velocities requires a temporal resolution close to what is observed in flying insect photoreceptors (100-300 Hz) (Niven et al., 2007). In the current study, robot and target velocities were limited due to the low temporal frequency of the camera. That is, the slow-moving, ground-based robotic platform does not deal with the same level of dynamic energy and motion control difficulties as experienced by dragonflies. However, the ground-based platform allowed us to address algorithmic questions and avoid the complexities associated with unmanned aerial vehicles (UAVs).

Given the robust performance of our model under real-world conditions, its high processing speed and low computational complexity (Bagheri et al., 2017), we will now turn to a UAV platform with a high-speed camera and field-programmable gate array (FPGA). To account for the faster dynamics of both predator and prey, we will modify (and test) tuning parameters accordingly. For integration of the stability and guidance of the UAV, an inner-loop control system can be exploited to constantly maintain the UAV's attitude in conjunction with its aerodynamic stability by compensating for any deviations in roll, pitch, and yaw. However, during a saccadic movement the outer-loop simply can change the set point of the inner-loop control system allowing performance of intended flight manoeuvres while keeping the inner-loop control active.

Implementation of the model on a UAV with high-speed camera should permit tracking of targets at high velocities. This would allow us to increase the level of noise in the input imagery (due to increase in dynamic energy of the robot) and test question such as the effect of noise reduction in the early visual processing on the model performance. Here, our choice of the time constants for filters replicated parameters observed in day-active insects, however with our robotics platform we will be able to implement a wider range of time

constants, such as those observed in nocturnal animals. Thus, our future robotic efforts will attempt to develop a robust fast-moving, UAV platform as well as permit us to address further comparative, physiological questions.

5.6 Conclusion

While the results of our recently developed insect-inspired target tracking model (Bagheri et al., 2015a) provide insight into insect neurophysiology, our understanding of animal sensorimotor systems is still limited. Experiments require directly linking neural circuits and behaviour, however, during physiological recordings our insect is restrained with wax and can only experience imposed, open-loop stimuli. To model sensorimotor systems, it is necessary to accurately represent the physical interaction of the animal and the environment which is very complex to model in simulations. To the best of our knowledge, this is the first time that a target tracking model inspired by insect neurophysiology has been implemented on an autonomous robot and tested under real-world conditions. We not only reveal robust model performance, but also provide insight into how insects' neuronal systems may handle varying challenges during target tracking and pursuit. That is, our hardware implementation provides a platform for better understanding the sensorimotor system of the insect as well as a prototype for engineering applications.

Acknowledgments

This research was supported under Australian Research Council's Discovery Projects (DP130104572), Discovery Early Career Researcher Award (DE150100548) funding scheme, and the Swedish Research Council (VR 2014-4904). We thank the manager of the Waite Campus of the University of Adelaide for allowing the robotic experiments.

References

Bagheri, Z. M., Wiederman, S. D., Cazzolato, B. S., Grainger, S., & O'Carroll, D. C. (2017). Performance of an insect-inspired target tracker in natural conditions. Bioinspiration & Biomimetics, 12(2), 025006.

Bagheri Z. M., Wiederman S. D., Cazzolato B. S., Grainger S., & O'Carroll D. C. (2015a). Properties of neuronal facilitation that improve target tracking in natural pursuit simulations. Journal of The Royal Society Interface, 12(108), 20150083.

Bagheri Z. M., Wiederman, S. D., Cazzolato, B., Grainger, S., & O'Carroll, D. C. (2015b). Robustness and real-time performance of an insect inspired target tracking algorithm under natural conditions. In Computational Intelligence, 2015 IEEE Symposium Series on (pp. 97-102). IEEE.

Bagheri Z. M., Wiederman S. D., Cazzolato B. S., Grainger S., & O'Carroll D. C. (2014a). A biologically inspired facilitation mechanism enhances the detection and pursuit of targets of varying contrast. In International Conference on Digital Image Computing: Techniques and Applications (DlCTA), 1-5.

Bagheri Z., Wiederman S. D., Cazzolato B. S., Grainger S., & O'Carroll D. C. (2014b). Performance assessment of an insect-inspired target tracking model in background clutter. In International Conference on Control Automation Robotics & Vision (ICARCV), 822-826.

Barnett P. D., Nordström K., & O'Carroll D. C. (2007) Retinotopic organization of small-field-target-detecting neurons in the insect visual system. Current Biology, 17, 569-578.

Brinkworth R. S., & O'Carroll D. C. (2009). Robust models for optic flow coding in natural scenes inspired by insect biology. PLoS Comput Biol, 5(11), e1000555.

Cai B., Xu X., Xing X., Jia K., Miao J., & Tao D. (2016). BIT: Biologically inspired tracker. IEEE Transactions on Image Processing, 25(3), 1327-1339.

Collett T. S. (1980). Angular tracking and the optomotor response an analysis of visual reflex interaction in a hoverfly. Journal of Comparative Physiology, 140(2), 145-158.

Colonnier F., Manecy A., Juston R., Mallot H., Leitel R., Floreano D., & Viollet S. (2015). A small-scale hyperacute compound eye featuring active eye tremor: application to visual stabilization, target tracking, and short-range odometry. Bioinspiration & biomimetics, 10(2), 026002.

Corbet P. S. (1999). Dragonflies: Behaviour and ecology of Odonata. Cornell University Press, Ithaca, New York.

Dean T. J. (2003). Fastest flyer. Book of insect records, chap. 1. University of Florida.

Dunbier J. R., Wiederman S. D., Shoemaker P. A., & O'Carroll D. C. (2012). Facilitation of dragonfly target-detecting neurons by slow moving features on continuous paths. Frontiers in Neural Circuits, 6, 79.

Dunbier J. R., Wiederman S. D., Shoemaker P. A., & O'Carroll D. C. (2011). Modelling the temporal response properties of an insect small target motion detector. In IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 125-130.

Findlay J. M. and Gilchrist I. D. (2003). Active Vision. Oxford: Oxford University Press.

Halupka K. J., Wiederman S. D., Cazzolato B. S., & O'Carroll D. C. (2013). Bio-inspired feature extraction and enhancement of targets moving against visual clutter during closed loop pursuit. In IEEE International Conference on Image Processing (ICIP), 4098-4102.

Halupka K. J., Wiederman S. D., Cazzolato B. S., & O'Carroll, D. C. (2011, December). Discrete implementation of biologically inspired image processing for target detection. In IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 143-148.

Hassenstein B. & Reichardt W. (1956). Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus. Zeitschrift für Naturforschung B, 11(9-10), 513-524.

James A. C. (1990). White-noise studies in the fly lamina. PhD Thesis, The Australian National University.

Irani M., Rousso B., & Peleg S. (1992). Detecting and tracking multiple moving objects using temporal integration. In European Conference on Computer Vision, 282-287.

Jansonius N. M., & Van Hateren J. H. (1991). Fast temporal adaptation of on-off units in the first optic chiasm of the blowfly. Journal of Comparative Physiology A, 168(6), 631-637.

Kim A. J., Fitzgerald J. K., & Maimon G. (2015). Cellular evidence for efference copy in Drosophila visuomotor processing. Nature neuroscience, 18(9), 1247-1255.

Krapp H. G., & Wicklein M. (2008). Central processing of visual information in insects. The Senses: A Comprehensive Reference, Volume 1, Academic Press, New York, pp. 131–203.

Land M. F. (2015). Eye movements of vertebrates and their relation to eye form and function. Journal of Comparative Physiology A, 201(2), 195-214.

Land M. F. (1999). Motion and vision: why animals move their eyes. Journal of Comparative Physiology A, 185(4), 341-352.

Land M. F. (1997). Visual acuity in insects. Annual Review of Entomology, 42(1), 147-177.

Land M. F., & Collett T. S. (1974). Chasing behaviour of houseflies (Fannia canicularis). Journal of Comparative Physiology, 89(4), 331-357.

Mahadevan V., & Vasconcelos N. (2013). Biologically inspired object tracking using center-surround saliency mechanisms. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(3), 541-554.

Mischiati M., Lin H. T., Herold P., Imler E., Olberg R., & Leonardo A. (2015). Internal models direct dragonfly interception steering. Nature, 517(7534), 333-338.

Nalbach H. O. (1989). Three temporal frequency channels constitute the dynamics of the optokinetic system of the crab, Carcinus maenas (L.). Biological Cybernetics, 61(1), 59-70.

Niven J. E., Anderson J. C., & Laughlin S. B. (2007). Fly photoreceptors demonstrate energy-information trade-offs in neural coding. PLoS Biology, 5(4), e116.

Nordström K., & O'Carroll D. C. (2009). Feature detection and the hypercomplex property in insects. Trends in Neurosciences, 32(7), 383-391.

Nordström K., Bolzon D. M., & O'Carroll D. C. (2011). Spatial facilitation by a high-performance dragonfly target-detecting neuron. Biology Letters, 7(4), 588-592.

Nordström K., Barnett P. D., & O'Carroll D. C. (2006). Insect detection of small targets moving in visual clutter. PLoS Biology, 4(3), e54.

O'Carroll, D. (1993). Feature-detecting neurons in dragonflies. Nature, 362, 541-543.

O'Carroll D. C., & Wiederman S. D. (2014). Contrast sensitivity and the detection of moving patterns and features. Philosophical Transactions of Royal Society B, 369(1636), 20130043.

O'Carroll D. C., Barnett P. D., & Nordström K. (2011). Local and global responses of insect motion detectors to the spatial structure of natural scenes. Journal of Vision, 11(14), 20-20.

Olberg R. M., Worthington A. H., & Venator K. R. (2000). Prey pursuit and interception in dragonflies. Journal of Comparative Physiology A, 186(2), 155-162.

Osorio D. (1991). Mechanisms of early visual processing in the medulla of the locust optic lobe: how self-inhibition, spatial-pooling, and signal rectification contribute to the properties of transient cells. Visual Neuroscience, 7(4), 345-355.

Rajesh S., O'Carroll D., & Abbott D. (2005). Man-made velocity estimators based on insect vision. Smart Materials and Structures, 14(2), 413.

Riesenhuber M., & Poggio T. (1999). Hierarchical models of object recognition in cortex. Nature Neuroscience, 2(11), 1019-1025.

Roubieu F. L., Expert F., Sabiron G., & Ruffier F. (2013). Two-directional 1-g visual motion sensor inspired by the fly's eye. IEEE Sensors Journal, 13(3), 1025-1035.

Silk J. D. (1995) Statistical variance analysis of clutter scenes and applications to a target acquisition test. Institute for Defense Analysis, Alexandria, Virginia, 2950.

Srinivasan M. V., Laughlin S. B., & Dubs A. (1982). Predictive coding: a fresh view of inhibition in the retina. Proceedings of the Royal Society of London B: Biological Sciences, 216(1205), 427-459.

Stavenga D. (2003). Angular and spectral sensitivity of fly photoreceptors. I. Integrated facet lens and rhabdomere optics. Journal of Comparative Physiology A, 189(1), 1-17.

Straw, A. D., Warrant, E. J., & O'Carroll, D. C. (2006). A bright zone in male hoverfly (Eristalistenax) eyes and associated faster motion detection and increased contrast sensitivity. Journal of Experimental Biology, 209(21), 4339-4354.

Van Hateren J. H., & Schilstra C. (1999). Blowfly flight and optic flow. II. Head movements during flight. Journal of Experimental Biology, 202(11), 1491-1500.

von Holst, E. and Mittelstaedt, H. (1950). Das Reafferenzprinzip. Wechselwirkungen zwischen Zentralnervensystem und Peripherie. Naturwissenschaften 37, 464–476.

Webb B., Harrison R. R., & Willis M. A. (2004). Sensorimotor control of navigation in arthropod and artificial systems. Arthropod Structure & Development, 33(3), 301-329.

Wehrhahn C., Poggio T., & Bülthoff H. (1982). Tracking and chasing in houseflies (Musca). Biological Cybernetics, 45(2), 123-130.

Wiederman S. D., & O'Carroll D. C. (2013a). Selective attention in an insect visual neuron. Current Biology, 23(2), 156-161.

Wiederman S. D., & O'Carroll DC (2013b) Biomimetic target detection: Modeling 2nd order correlation of OFF and ON channels. IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP), 16-21.

Wiederman S. D., & O'Carroll D. C. (2011). Discrimination of features in natural scenes by a dragonfly neuron. The Journal of Neuroscience, 31(19), 7141-7144.

Wiederman S. D., Shoemaker P. A., & O'Carroll D. C. (2008). A model for the detection of moving targets in visual clutter inspired by insect physiology. PloS One, 3(7), e2784.

Zhang, S., Yao, H., & Liu, S. (2010). Robust visual tracking using feature-based visual attention. In IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 1150-1153.

Zheng Y., & Meng Y. (2008). Swarm intelligence based dynamic object tracking. In IEEE World Congress on Computational Intelligence Evolutionary Computation, pp. 405-412.

Chapter 6. Conclusions and Future Work

In this thesis I have presented an insect-inspired model for tracking small moving targets within visually cluttered surrounds. I have tested the robustness and efficiency of this target-tracking algorithm both in simulations and robotic experiments. The data clearly show that this model provides robust detection of varying target contrast against a wide range of backgrounds. My direct comparison between this model and state-of-the-art engineering trackers show that when tracking small moving targets in natural scenes, this model exhibits robust performance, outperforming the best of the current tracking algorithms. However, the results show that this performance does not come at the cost of additional processing time. The main reason behind the fast processing speed of my model is reducing the resolution of the input image to 1°. However, the question is whether reducing the image resolution results in cutting out the necessary information and therefore, decreasing the accuracy of the computational model? Or using high resolution images only overloads the computational model with redundant information and slows down the processing speed? How much information is necessary? The fact that many animals only have a small high resolution area in their eyes (e.g. fovea, acute zone) and low resolution vision elsewhere within their field of view might provide the proof of the concept that high-resolution vision is not necessarily optimal. Although I have not tested the effect of resolution of input image on the accuracy of the model it would be an interesting question for future research.

As I mentioned in the introduction of this thesis, I used a ground-based robotic platform for hardware implementation to maintain the focus of my research on the key algorithmic questions rather than engineering problems associated with UAVs. However, target tracking in a 3-dimensional world introduces additional challenges in motion control and pursuit strategy. Furthermore, the tight weight budget in UAVs imposes additional constraints on

the choice of processors and camera. Recent advances in development cameras integrated with field-programmable gate array (FPGA) provide a light-weight and high-performance solution for this purpose. Translation of this insect-inspired model to a UAV platform is currently under development in our research group.

Besides the potential of this algorithm for artificial visual systems and robotic applications, it allowed us to gain insight into insect neurophysiology. The concluding remarks of the current work and potential future works are summarized in the following sections.

6.1 Facilitation

Over the course of this study the understanding of CSTMD1's facilitatory behaviour has increased through both my simulations and experiments. The results presented here include the effect of facilitation on target detection and tracking within natural environments. These results suggest that inclusion of facilitation in our closed-loop model substantially improves pursuit success. However, my results show that the choice of facilitation time constant depends on various factors such as target velocity and background clutter. These results are contrary to a fundamental basis of control theory where reducing the phase delay to reach the steady-state mode is desirable. However, natural conditions are highly dynamic and perhaps steady-state mode does not exist for an insect during pursuit. From these observations arises the question as to whether a dynamic facilitation time constant would improve target tracking and pursuit success. From modelling and robotic experiment perspective, such a dynamic modulation can be exploited with a fuzzy control system. However, whether CSTMD1 neurons use a dynamic mechanism requires further physiological experiments to address these questions more directly.

The recent electrophysiological recordings of CSTMD1 neurons within our laboratory reveals enhancement in front of the moving target, and suppression in the surround.

Although, the model of facilitation that I implemented here emulates enhancement in the direction of target motion and the asymmetry in onset/offset time course as observed in CSTMD1 neurons, it does not explain the inhibitory response of these neurons. However, this inhibition might play a key role in suppressing false positives arising from background clutter, thus improving target detection. Therefore, further modelling is required to investigate the effect of inhibition on target tracking and pursuit.

6.2 Parallel Computation

STMD neurons are tuned to small objects, with peak responses to an optimum target size of 1.6°×1.6° (Wiederman et al., 2008). The model I presented here was designed to mimic size tuning observed in STMD neurons. However, during the pursuit the image size of the target in the retina increases as the insect approaches its target. Therefore, further neuronal processing is required to detect a large object and finalize prey capture.

My modelling effort to vary the size selectivity of the ESTMD model (Chapter 4) shows the possibility of parallel pathways encoding different, broadly-tuned size ranges, which might be combined later in the visual pathway. Recent observation of unidentified STMD type neurons with peak responses for larger objects at ~10° (Wiederman et al., 2013) also provides support for this hypothesis. Parallel processing is well-supported in human psychophysics experiments (Graham and Nachmias, 1971). Moreover, parallelism has been extensively employed in engineering to provide high-performance computing and reduce power consumption. It is possible that the same parallel computation is employed by the low-powered brain of insects to process highly complex visual inputs. Another intriguing possibility is that other neurons, such as the 'looming' system observed in the locusts (Rind and Simmons, 1992; Rind and Bramwell, 1996), could be recruited when the target is larger

than optimal for an STMD. However, further electrophysiological and neuroanatomical experiments are required to investigate how insects track large objects.

6.3 Attention

The study by Wiederman and O'Carroll (2013) shows that CSTMD1 neurons competitively select one target in the presence of distracters. Their results show that competition is key because the attended target changes between trials. They suggested that the variability in the attended target is either due to the modulation of the underlying target salience over trials, or a higher-order mechanism of bias (Desimone, 1998). Within this concept, the facilitatory behaviour of CSTMD1 neurons suggest an interesting role for the first possibility. The results I present in this thesis support a possible role for facilitation in selective attention. These results could be indicative of a bottom-up attention mechanism emerging from a competitive process occurring at a lower level in STMD pathway.

This selective attention would be an essential component of a control system for target pursuit. The selection mechanism I implemented here is just a simple form of 'winner-take-all-network'. However, such a sophisticated trajectory prediction in insects and other animals undoubtedly involves additional processes. Therefore, an interesting extension to the current tracker would be a model of selective attention (such as Shoemaker et al. (2013) model) which can increase the robustness of the model in the presence of distractors.

6.4 Active Vision

Many animals use eye, head or body movement to shift the direction of gaze. Active gaze fixation and stabilization may provide the key to an efficient target tracking system. Active gaze control can reduce the computational process and resources required to extract relevant visual information from the environment by fixing gaze on a given moving feature and enabling areas of interest to be examined at the desired resolution without the cost of

uniformly high resolution sensing (Cannons, 2008). Inspired by insect gaze fixation strategies we developed a novel fixation algorithm. Instead of just trying to keep the target perfectly centred on its field of view, our system locks on to the background and lets the target move against it. This reduces distractions from the background and gives time for underlying nonlinear spatiotemporal filtering inherent to the ESTMD pathway to work. It then makes small movements of its gaze and rotates towards the target to keep the target roughly frontal.

One of the weaknesses of the robotic implementation I used here is that the camera was mechanically coupled to the robot. Therefore, strong disturbances such as an uneven surface could easily perturb the body, and hence the camera. In freely flying insects, active gaze stabilization mechanisms reduce the effect of disturbances such as vibrations, wind gusts, or body jerks (Chan et al., 1998; Hengstenberg, 1988; Miles and Wallman, 1993; Schilstra and Van Hateren, 1998; Zeil eta al., 2008). A similar mechanism is employed by vertebrates via vestibulo–ocular reflexes (VOR) to hold the gaze still in space when the head turns (Huterer and Cullen, 2002). On a robotic platform, a pan-tilt-zoom (PTZ) mechanism could be used to implement such uncoupling of the camera from its supports. This PTZ mechanism is currently under development in our group.

6.5 Final Remark

I should mention that the model I presented here lacks certain aspects of insect vision as well as computer vision models that I discussed in detail in Chapter 1. This model does not account for the learning and classification process involved in target selection and tracking. Despite this shortcoming, this model shows successful detection and tracking of moving targets in very challenging scenarios. This recalls the question that I discussed in the introduction of this thesis (Section 1.3.2); what kind of cues insects use for target

discrimination? Do they simply have a 'hard-wired' system which only tracks the targets that coincide with the correct size and velocity tuning range? Or are they capable of learning and using top-down information to classify moving objects? My results show that the pursuit success of the model significantly decreases as the clutter of the background increases. However, insects such as the dragonfly have shown remarkable pursuit success (greater than 97%) in heavily cluttered environments (Olberg et al, 2000). Therefore, there is a great possibility that insects use learning and classification processes in these more challenging scenarios to discriminate targets from similar distractors that exist in the environment. Moreover, the version of the insect-inspired model that I presented here does not include the complex feedback processes that exist in biological visual systems (Section 1.6). Investigating the computational mechanism of these feedback processes is the future direction of our lab.

The investigation of biological target detection and tracking is a vast scientific area which is experiencing a significant growth and breakthroughs from both an electrophysiological and a modelling perspective. This thesis is only the beginning of an exciting field of research to find solutions for engineering problems and answers for physiological questions. Further understanding of such a remarkable system will emerge from future endeavours.

References

Cannons K. (2008). A review of visual tracking. Dept. Comput. Sci. Eng., York Univ., Toronto, Canada, Tech. Rep. CSE-2008-07.

Chan W. P., Prete F., & Dickinson M. H. (1998). Visual input to the efferent control system of a fly's gyroscope. Science, 280(5361), 289-292.

Desimone R. (1996). Neural mechanisms for visual memory and their role in attention. Proceedings of the National Academy of Sciences, 93(24), 13494-13499.

Graham N., & Nachmias J (1971). Detection of grating patterns containing two spatial frequencies: A comparison of single-channel and multiple-channels models. Vision Research, 11(3), 251-259.

Hengstenberg R. (1988). Mechanosensory control of compensatory head roll during flight in the blowfly Calliphora Erythrocephala Meig. Journal of Comparative Physiology A, 163(2), 151-165.

Huterer M., & Cullen K. E. (2002). Vestibuloocular reflex dynamics during high-frequency and high-acceleration rotations of the head on body in rhesus monkey. Journal of Neurophysiology, 88(1), 13-28.

Miles F. A., & Wallman J. (1993). Visual motion and its role in the stabilization of gaze. Amsterdam: Elsevier. 417 pp.

Olberg R. M., Worthington A. H., & Venator K. R. (2000). Prey pursuit and interception in dragonflies. Journal of Comparative Physiology A, 186(2), 155-162.

Rind F. C., & Bramwell D. I. (1996). Neural network based on the input organization of an identified neuron signalling impending collision. Journal of Neurophysiology, 75(3), 967-985.

Rind F. C., & Simmons P. J. (1992). Orthopteran DCMD neuron: A re-evaluation of responses to moving objects. I. Selective responses to approaching objects. Journal of Neurophysiology, 68(5), 1654-1666.

Schilstra C., & Van Hateren J. H. (1998). Stabilizing gaze in flying blowflies. Nature, 395(6703), 654-654.

Shoemaker P. A., Wiederman S. D., & O'Carroll D. C. (2013). Can a competitive neural network explain selective attention in insect target tracking neurons? In 6th International IEEE/EMBS Conference on Neural Engineering (NER), 903-906.

Wiederman S. D., & O'Carroll D. C. (2013). Selective attention in an insect visual neuron. Current Biology, 23(2), 156-161.

Wiederman S. D., Shoemaker P. A., & O'Carroll D. C. (2013). Correlation between OFF and ON channels underlies dark target selectivity in an insect visual system. The Journal of Neuroscience, 33(32), 13225-13232.

Wiederman S. D., Shoemaker P. A., & O'Carroll D. C. (2008). A model for the detection of moving targets in visual clutter inspired by insect physiology. PloS One, 3(7), e2784.

Zeil J., Boeddeker N., & Hemmi J. M. (2008). Vision and the organization of behaviour. Current Biology, 18(8), R320-R323.

Appendix A. Supplementary Material for Chapter 2

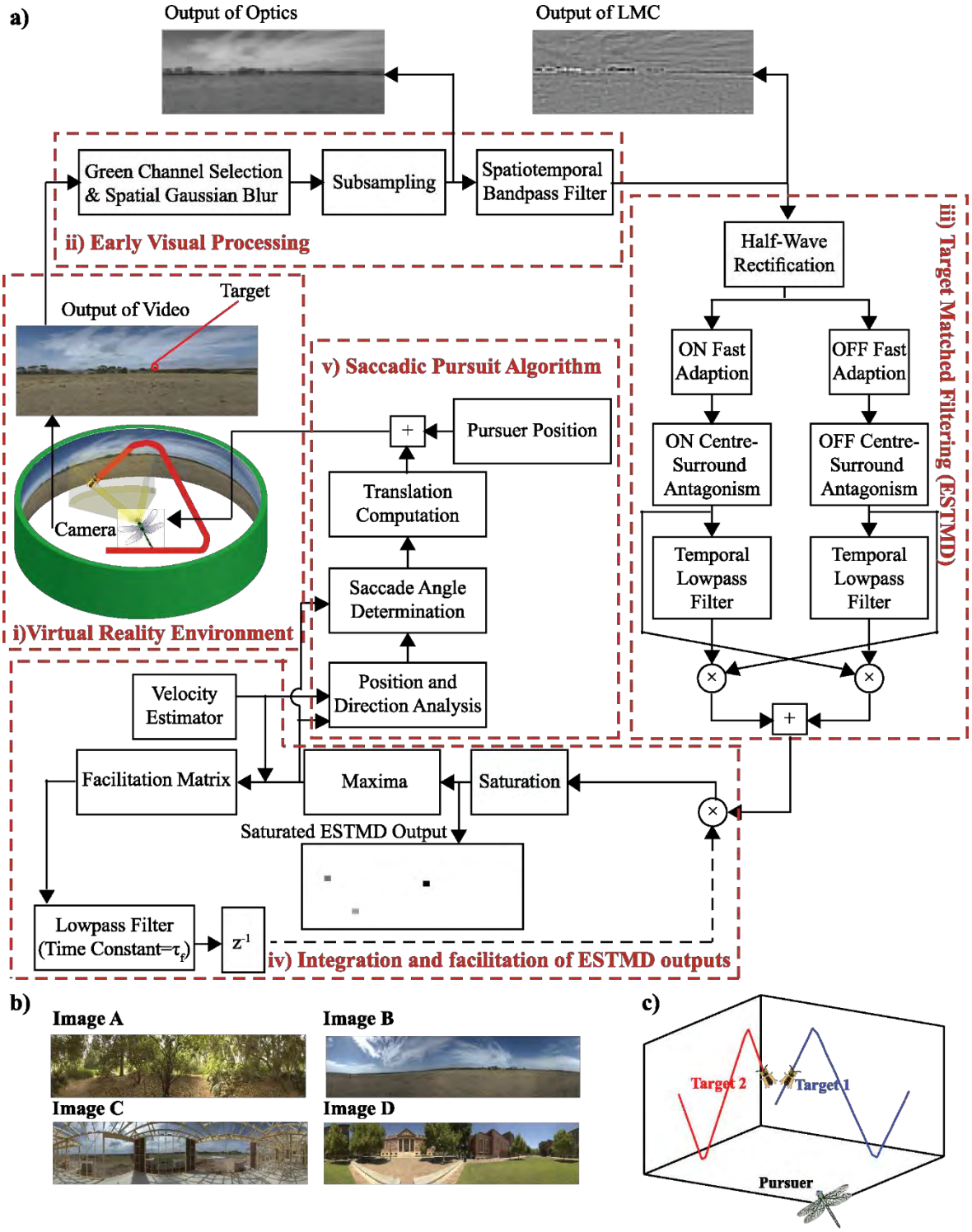

Figure A.1. Overview of the bio-inspired, target-pursuit model and simulation environment a) The closed loop model includes a virtual reality environment used to navigate both pursuer and target (i). The early visual processing stage simulates insect spectral sensitivity, optical blur and low resolution sampling of the insect

visual system (ii). Selectivity for small moving targets is generated with the ESTMD stage (iii). The 'facilitation mechanism' is modelled by multiplying the output of ESTMDs with a delayed version of a weighted map (lowpass filter, time constant=τ_f) based on the current location of the winning feature but offset in the direction of the target's movement (iv). The saccadic pursuit algorithm (v) calculates the pursuer turn angle inspired by insect's pursuit strategy. b) Four natural images used as backgrounds in the pursuit simulations. c) When two targets are simulated to test competitive selection, both move in symmetrical paths with the second distracter target introduced 100 ms after the first target.

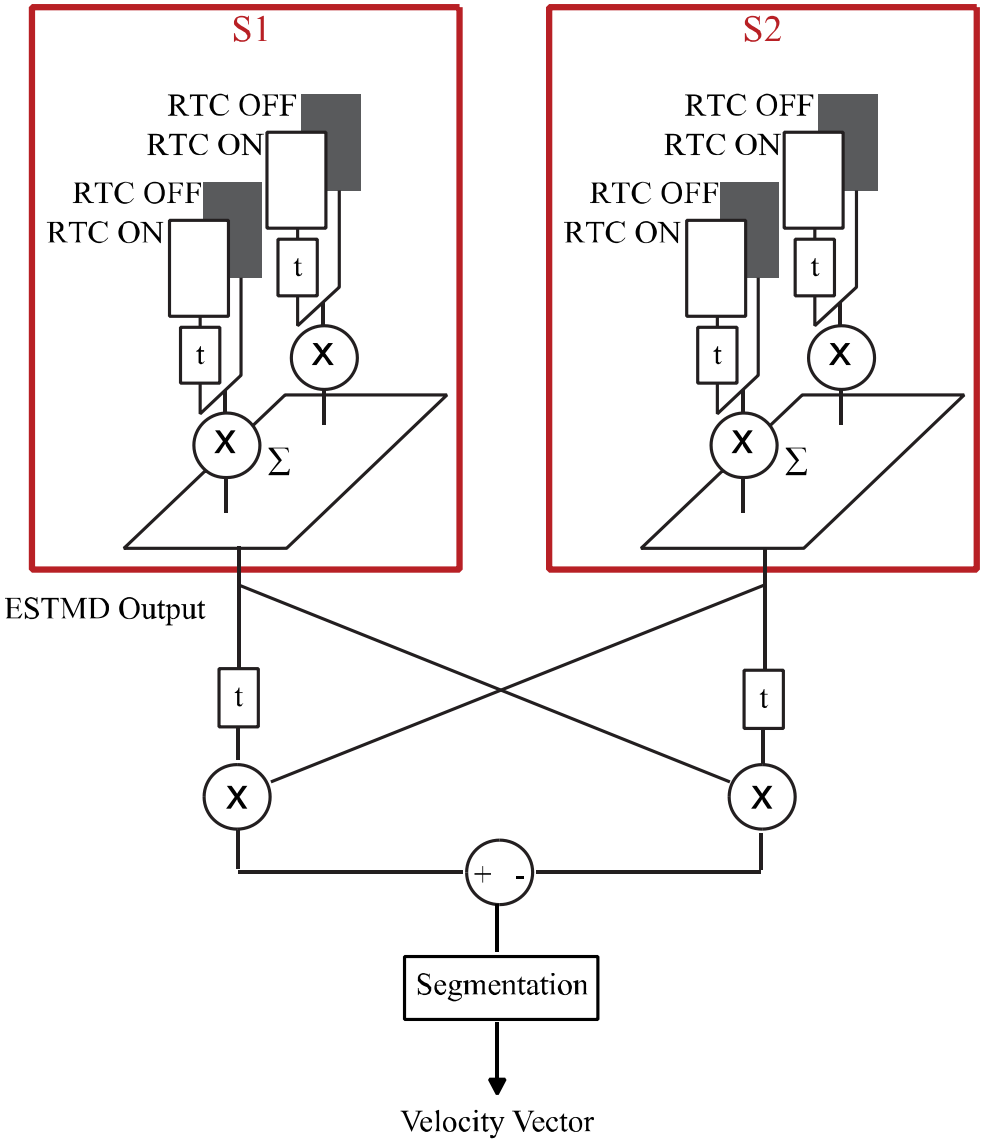

Figure A.2. Velocity Estimator. To confer sensitivity to targets independent of the polarity of their contrast against the background, we added an equivalent operation in parallel, multiplying each contrast channel (ON or OFF) with a delayed version of the opposite polarity (delayed using a low-pass filter, τ=25 ms) and then summing the outputs. The Hassenstein-Reichardt (HR) EMD correlates two spatially separated contrast signals (S1, S2) and correlates them after a delay (via a low-pass filter) resulting in a direction selective output [1]. Subtracting the two mirrored symmetric sub-unit yields positive response for the preferred motion direction (in this case left to right) and a negative one in the opposite direction (right to left). The output of a HR-EMD to a discrete target of constant width in the direction of motion increases with target velocity up to an optimum dependent on the delay filter time constant [2]. Due to this known relationship, the HR-EMD may also be used to approximate the magnitude velocity of a moving object [3]. However, the angular size of the target changes

during the pursuit, therefore, the output of the HR-EMD is segmented into three intervals to estimate the range of target angular velocity magnitude. This velocity vector is used to shift the facilitation matrix in the direction of travel of the 'winning' feature.

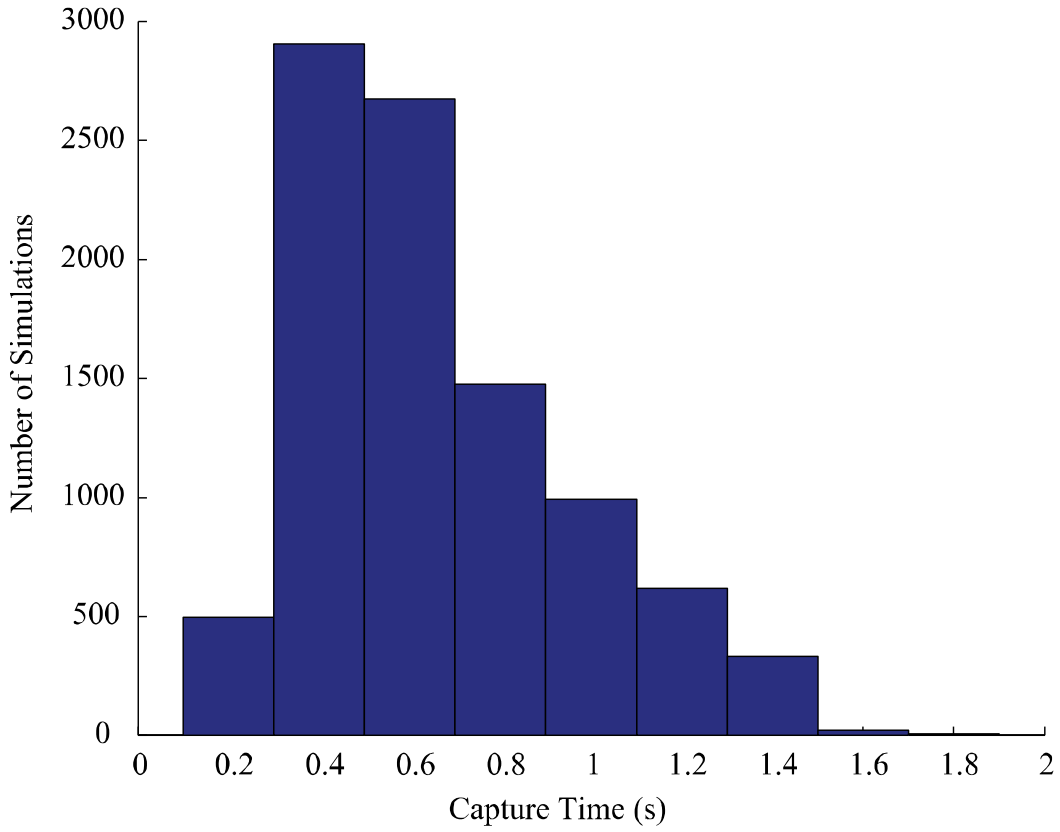

Figure A.3. Distribution of capture time for the simulations with |Vt|/|Vp|=3/4

Table A.1. Target intensities used for simulations against different backgrounds and the mean intensity of each background. For each image, two target intensities were below and two above mean luminance. The maximum or minimum target luminance was adjusted based on an initial set of simulations over a larger range of target intensities for image. This allowed us to select targets with a low enough average contrast to reduce prey capture success well below 100% in non-facilitated trials and thus provide a basis for comparison of the improvement following facilitation for the most easily detected targets for each image. The two additional lower contrasts were selected to give approximately the same success rate in non-facilitated pursuits for different images.

Background	**Target Intensities (green channel, 8-bit)**	**Mean of Background (green channel, 8-bit)**
Image A	0, 25, 204, 255	92
Image B	51,77, 230, 255	130
Image C	0, 51, 204, 255	110
Image D	0, 25, 204, 230	98

Text A.1. Early Visual Processing

The optics of flying insects are limited by diffraction and other forms of optical interference within the facet lenses [4]. This optical blur was modelled with a Gaussian lowpass filter (full-width at half maximum of 1.4°), which is similar to the optical sampling of typical day-active insects [4]. Further sub-sampling of the captured image (at 1° intervals) represented the average inter-receptor angle between photoreceptors [5]. The green spectral sensitivity of the insect motion pathway was simulated by processing only the green channel of the RGB imagery [6].

In biological vision, redundant information is removed with neuronal adaptation (temporal high pass filtering) and centre-surround antagonism (spatial high pass filtering), both of which follow the inherent low-pass sampling properties of the photoreceptors. We simulated the temporal properties of photoreceptors and the 1st order interneurons, the large monopolar

cells (LMCs), with a discrete log-normal function [7]. The filter properties were matched to the temporal impulse response observed in LMC recordings [8]. Weak centre-surround antagonism as observed in physiological recordings, was modelled by subtracting (10% of the centre pixel) from the nearest neighbouring pixels.

Text A.2. Rectifying transient cells (RTCs)

Rectifying transient cells (RTCs) within the insect 2nd optic neuropil (medulla) exhibit processing properties well suited as additional input processing stages for a small target motion detection pathway [9]. RTCs exhibit independent adaptation to light increment (ON channel) or decrement (OFF channel) [10, 11]. To simulate the temporal processing of RTCs, we separated ON and OFF channels via temporal high pass filtering (τ=40 ms) and half-wave rectification. Each channel was processed through a fast adaptive mechanism, with the state of adaptation determined by a nonlinear filter that switches its time constant [7, 9]. This adaptation state causes subtractive inhibition of the unaltered 'pass-through' signal. Additionally, we implemented strong spatial centre-surround antagonism, with each channel surround inhibiting its next-nearest neighbours. This strong surround antagonism conveys selectivity for local edge features.

Text A.3. Input: Clutter Measure

To measure background clutter, we used the metric developed by Silk [12]. This method measures image clutter by convolving the image with an average kernel, which was chosen to be approximately the same size as the target [13]. The clutter value, *CM,* is calculated by the following formula:

$$CM = {}^{1}/_{N} \left(\sum_{i} \sum_{j} \left(b_{i,j} - \bar{B}_{i,j} \right)^{2} \right)^{{}^{1}/_{2}} \quad \text{(S1)}$$

where *b* is the value of the *i,j* pixel before convolution, $\bar{B}$ is the mean of the box centered at pixel *i,j* and *N* is the number of boxes convolved over the whole image.

Text A.4. Input: Target Contrast Measure

To measure how the target differs from the background at each time step of the pursuit in the input (after optical blur), we used the magnitude of a weighted signal-to-noise ratio (*WSNR*):

$$WSNR = w(d)\left|\frac{I_t - \bar{I}_b}{\sigma_b}\right|, \tag{S2}$$

The weighing function *w*(*d*) represents the effect of target angular size on its detectability against background. I_t represents the target intensity, $\bar{I}_b$ and σ_b respectively are the mean value and standard deviation of the next nearest neighbors of the background to the target point, *d* is the distance between target and the pursuer, and *w*(*d*) is defined as:

$$w(d) = \frac{1}{2} + \frac{1}{2}\tanh(a + \frac{b}{d}), \tag{S3}$$

where *a* and *b* are chosen based on the target size selectivity of the model.

Text A.5. Output: Target Discriminability Measure

To determine the discriminability of the target at the output stage, we defined a metric which measures the difference between target value and false positives:

$$Discrim = \frac{M-N}{M} e^{(I_t - I_{max} - \sigma_{I_b})}, \tag{S4}$$

where *M* is the total number of background pixels, *N* represents the number of background pixels with equal or higher values than the target (stronger false positives), I_t is the target intensity, I_{max} is the maximum intensity of the background pixels, and σ_{I_b} is the deviation of stronger false positives from target value, given by:

$$\sigma_{I_b} = \sqrt{\sum_{I_i=I_t}^{I_{max}} n_i (I_t - I_i)^2}. \tag{S5}$$

where n_i is the number of pixels with the intensity of I_i. Based on this metric, the maximum possible discriminability (e) happens when no false positive is detected and I_t is at its maximum (1), and whenever the target is the winner at the output, the discriminability value is greater than 1. Due to half-wave-rectifications in the ESTMD stage the background in the output of ESTMD model is black (0 value) and the target and any false positives has a value greater than 0. Therefore, it would work for both dark and light targets.

Text A.6. Robustness Metric

The robustness metric represents the real performance of the model as a percentage of the ideal performance. The minimum criterion for an acceptable performance of the model is a successful pursuit. However, a higher target discriminability during the pursuit indicates a more robust performance. Therefore, for a sample size with N successful pursuits, we measured the performance of the model as:

$$\sum_{i=1}^{N} \overline{Discrim_i} \tag{S6}$$

where $\overline{Discrim_i}$ is the average discriminability in corresponding duration of the pursuit. The ideal performance of the model occurs when all the simulations within the data set is completed successfully with a maximum target discriminability (e) throughout the pursuit. Using the Equation S6 the ideal performance of the model for a sample size of M would be $M \times e$. Therefore, the *robustness* metric for a sample set is:

$$Robustness = \frac{\sum_{i=1}^{N} \overline{Discrim_i}}{M \times e} \tag{S7}$$

1. Hassenstein B and Reichardt W (1956) Analyse der zeit-, reihenfolgen-und vorzeichenauswertung bei der bewegungsperzeption des rüsselkäfers Chlorophanus. Z. Naturforsch, 11b: 513–524.

2. Dunbier JR, Wiederman SD, Shoemaker PA and O'Carroll DC (2011) Modelling the temporal response properties of an insect small target motion detector. ISSNIP, 125-130.

3. Dror R, O'Carroll DC, and Laughin S (2001) Accuracy of velocity estimation by Reichardt correlators. JOSA A, 18:241-252.

4. Stavenga DG (2003) Angular and spectral sensitivity of fly photoreceptors. I. Integrated facet lens and Rhabdomere optics. J. Comp. Physiol. A, 189:1-17.

5. Straw AD, Warrant EJ and O'Carroll DC (2006) A bright zone in male hoverfly (Eristalis tenax) eyes and associated faster motion detection and increased contrast sensitivity. J. Exp. Biol., 209:4339-4354.

6. Srinivasan MV and Guy RG (1990) Spectral properties of movement perception in the dronefly Eristalis. J. Comp. Physiol. A, 166:287-295.

7. Halupka KJ, Wiederman SD, Cazzolato BS and O'Carroll DC (2011) Discrete implementation of biologically inspired image processing for target detection. ISSNIP, 143-148.

8. James AC (1990) White-noise studies in the fly lamina. PhD Thesis, The Australian National University.

9. Wiederman SD, Shoemaker PA, and O'Carroll DC (2008) A model for the detection of moving targets in visual clutter inspired by insect physiology. PLoS ONE, 3:1-11.

10. Osorio D (1991) Mechanisms of early visual processing in the medulla of the locust optic lobe-how self-inhibition, spatial-pooling, and signal rectification contribute to the properties of transient cells. Visual Neurosci., 7:345-355.

11. Jansonius N and van Hateren J (1991) Fast temporal adaptation of on-off units in the first optic chiasm of the blowfly. J. Comp. Physiol. A, 168:631–637.

12. Silk JD (1995) Statistical variance analysis of clutter scenes and applications to a target acquisition test. Institute for Defense Analysis, Alexandria, VA 2950-1995.

13. Ralph SK, Irvine J, Snorrason M and Vanstone S (2006) An image metric-based ATR performance prediction testbed. AIPR Workshop 33-33.

Statement of Authorship

Title of Paper	A Biologically Inspired Facilitation Mechanism Enhances the Detection and Pursuit of Targets of Varying Contrast
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Digital Image Computing: Techniques and Applications (DICTA), 2014, Wollongong, NSW, pp. 1-5, DOI: 10.1109/DICTA.2014.7008082.

Principal Author

Name of Principal Author (Candidate)	Zahra Bagheri
Contribution to the Paper	Developed the code, designed the experiments, ran the model simulations, analysed and interpreted the data and drafted the manuscript with editing contributions from the other authors.
Overall percentage (%)	55
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 6/7/2017

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Steven Wiederman
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, assisted with analysis and interpretation of the data, provided significant editing contribution (20%).
Signature	Date 6/7/2017

Name of Co-Author	Benjamin Cazzolato
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, assisted with analysis and interpretation of the data, provided editing contribution (10%).
Signature	Date 6/7/17

Name of Co-Author	Steven Grainger		
Contribution to the Paper	Participated in the initial conceptualisation and experimental designs, read and approved the draft (5%).		
Signature		Date	6/7/17

Name of Co-Author	David O'Carroll		
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, assisted with analysis and interpretation of the data, provided editing contribution (10%).		
Signature		Date	15-NOV-2016

Appendix B. A Biologically Inspired Facilitation Mechanism Enhances the Detection and Pursuit of Targets of Varying Contrast

Zahra M. Bagheri, Steven D. Wiederman, Benjamin S. Cazzolato, Steven Grainger

The University of Adelaide, Adelaide, Australia

David C. O'Carroll

Department of Biology**,** Lund University**,** Lund, Sweden

Abstract— Many species of flying insects detect and chase prey or conspecifics within a visually cluttered surround, e.g. for predation, territorial or mating behaviour. We modelled such detection and pursuit for small moving targets, and tested it within a closed-loop, virtual reality flight arena. Our model is inspired directly by electrophysiological recordings from 'small target motion detector' (STMD) neurons in the insect brain that are likely to underlie this behavioral task. The front-end uses a variant of a biologically inspired 'elementary' small target motion detector (ESTMD), elaborated to detect targets in natural scenes of both contrast polarities (i.e. both dark and light targets). We also include an additional model for the recently identified physiological 'facilitation' mechanism believed to form the basis for selective attention in insect STMDs, and quantify the improvement this provides for pursuit success and target discriminability over a range of target contrasts.

Keywords— Target tracking, feature detection, biological image processing, visual processing, salience.

I. INTRODUCTION

The dragonfly is an aerobatic predator capable of visually detecting small, moving prey, often against a cluttered background. Remarkably, the dragonfly performs this task even in the presence of distracting stimuli, such as swarms of prey and conspecifics [1,2]. From an engineering perspective, our ability to build an artificial vision system that can emulate such feature discrimination from a moving platform is poor [3], particularly compared to the small size, light-weight and low-power neuronal architecture that underlies insect behaviour [4].

Our approach to developing an engineered solution to this problem is to model the neuronal pathway likely to underlie pursuit behaviour. We use intracellular, electrophysiological techniques to record from 'small target motion detector' (STMD) neurons, in response to the presentation of various visual stimuli. We have shown that these neurons are size selective, velocity tuned, contrast sensitive and respond robustly to targets, even without relative motion between them and a cluttered background [5-7]. These physiological data have inspired the development of a computational model (MATLAB / Simulink) that effectively provides a matched spatiotemporal filter for the detection of moving targets in natural scenery [8-10]. We refer to this model as the 'elementary small target motion detector' (ESTMD) and have recently elaborated it to include a virtual reality front-end and to simulate closed-loop pursuits based on known chase behaviours of flying insects [11].

One type of dragonfly STMD neuron, CSTMD1, exhibits a higher-order property we term 'facilitation'. The spiking activity of CSTMD1 builds over time in response to targets that move through long, continuous trajectories [12]. This enhancement in response to successive stimulation (i.e. facilitation) encodes the trajectory of the target and is reset when there are local breaks in the trajectory path [13]. We hypothesize that such a neuronal facilitatory mechanism underlies the robust pursuits of prey (over 97% success rate) observed in the dragonfly [14].

We recently showed that inclusion of a simple form of slow facilitation in a dark-target selective ESTMD model based on known physiological properties of STMDs [15, 16], enhances detection and pursuit of dark contrasting targets. However, in a natural environment, targets will have different changing target contrast polarity (light or dark) dependent on the corresponding background at any moment during the pursuit. In this paper we developed new metrics to test robustness under these conditions for a model modified to respond to both contrast polarities. We tested whether the inclusion of facilitation differentially improves capture success across a range of target luminances, and examined trade-offs between the time-constant of the facilitation mechanism and model performance over a range of relative pursuer and target speeds.

II. METHODS

A. Virtual-Reality Front-End

Figure B.1 illustrates the Virtual Reality (VR) arena used as the front-end for our bio-inspired target detection and pursuit control algorithm which was implemented in the Simulink 3D

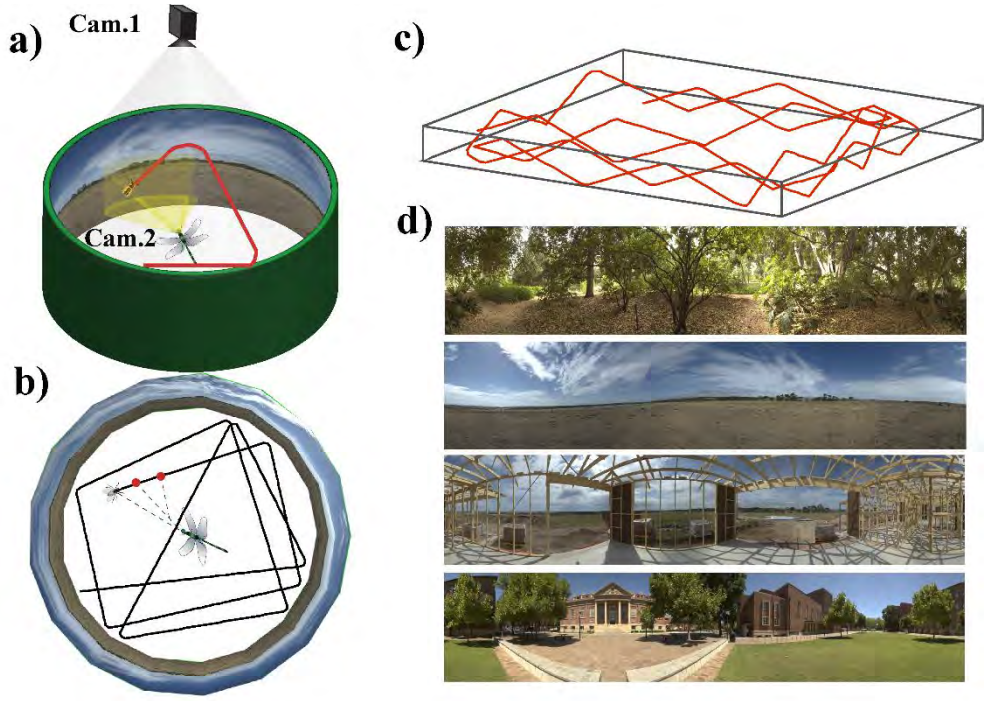

Figure B.1. a) Virtual Reality Environment b) a model of random target trajectory in 2D c) target trajectory in 3D d) From top to bottom shows the rendered images A, B, C, and D respectively.

animation toolbox (Mathworks Inc.). Within a cylindrical arena (of radius 6 m), we generated randomized paths with biologically plausible constraints on 'saccadic' turn angles [17]. We tested five different velocities ranging from 4 ms-1 to 16 ms-1 for a 40 mm sized 'prey', with the 'pursuer' moving at a velocity of 8 ms-1. The target start location was at least 4 m away from the pursuer (angular size<0.6°). We varied target contrast in different simulations by altering target luminance across a range of intensities, i.e. RGB values from 0 (black) to 255 (white).

Pursuit simulations were repeated 50 times with randomized three dimensional paths. We simulated pursuits in four different panoramic natural scenes (Images A, B, C, D in Figure B.1 d) rendered onto the cylindrical wall of the VR world. If prey approached to close proximity (50 cm) of this wall, we initiated a 'saccade' away from the VR boundary (Figure B.1 b). Although all of these images had 1/f power spectra, a statistical property of natural images [18], they varied in the amount of background 'clutter' in the scene (Figure B.1d).

Video was sampled from a 40°x98° sized viewport to represent the visual field of the predator and thus served as inputs to the detection, facilitation and pursuit algorithm.

B. Elementary Small Target Motion Detection

We modeled the optical blur of the compound eye (limited primarily by diffraction) with a Gaussian function of full-width at half maximum of 1.4° [19]. Insect motion sensitive pathways are primarily sensitive to green light, so we selected only the green channel of the RGB input images [20]. The filtered image was then sampled at 1°, based on the measured inter-receptor angle of the fly visual system [21]. The output of this stage is considered as the 'model input' to the target detection algorithm in further analyses.

Early visual processing (photoreceptors and 1st order interneurons) were simulated with spatiotemporal bandpass filtering, matched to properties observed in insect vision [22]. A 2nd model for an 'elementary' small target motion detector (ESTMD) was then implemented, based on the response properties of rectifying, transient cells as observed in several insect species [23,24]. Transient ON and OFF contrasts are separated via temporal high pass filtering (τ=40 ms) and half-wave rectification. These independent ON and OFF channels are processed through a fast adaptive mechanism, with the state of adaptation determined by a nonlinear filter. This filter switches its time constant. Time constants are 'fast' (τ=3 ms) when channel input is increasing and 'slow' (τ=70 ms) when decreasing. This adaptation state causes subtractive inhibition of the unaltered 'pass-through' signal. The result of this complex, nonlinear filtering is the signalling of 'novel' transient contrast changes (of a particular channel phase, ON or OFF) with the suppression of fluctuating textural variations.

Both of the ON and OFF channels then undergo further strong center-surround antagonism, selectively tuning the model to targets with small angular extent (orthogonal to the direction

of travel). Sensitivity to both dark and light targets is provided by delaying and multiplying each contrast channel (ON or OFF) with a delayed version of the opposite polarity (delayed using a low-pass filter, τ=25 ms). This also conveys selectivity for objects that are small in the dimension matching the direction of travel, since a small feature will usually be characterized by an initial rise (or fall) in brightness at each point that it passes across, followed a corresponding fall (or rise) after a short delay. The output image undergoes non-linear saturation using a hyperbolic tangent function. This serves to ensure all signals lie between 0 and 1. Then the maximum is determined as the target.

An alternative model for direction-selective 'elementary motion detectors' (EMDs) is the Hassenstein-Reichardt (HR) detector [25]. Two spatially separated contrast signals are correlated after a delay (via a low-pass filter). 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 maintains core STMD properties but with direction-selective outputs [26]. As outlined in Figure B.2 we cascaded the outputs of an ESTMD model with an EMD (the 'Reichardt Correlator') to provide a measure of the direction in which targets moved.

C. Target Pursuit

In the model variant without facilitation, closed loop pursuit was directed towards the target location determined from the maximum output of the array of local ESTMDs. The pursuit strategy implemented was 'saccadic tracking' as observed from male houseflies, where heading is calculated from the error angle between target and the central axis of the pursuer's gaze [27,28]. We implemented re-centring towards the target, only when it moved 5°. This strategy promotes target 'pop-out' by permitting the spatiotemporal filters to 'fade away' (high-pass) the more distant background. A successful pursuit was defined as a target

captured in less than 2 s within 1.2 m distance from the pursuer and the target had to be the winner in the output for more than 50% of the last 6 ms of the tracking.

D. Facilitation

We implemented facilitation by building a weighted 'map' dependent on the location of the winning feature but shifted by the target velocity vector (provided by the cascaded ESTMD-EMD). We multiplied the ESTMD model output with a low-pass filtered version of this 'facilitation map'. The role of the low pass filter time constant here is to control the kinetics with which the facilitation matrix enhances the area around the winning feature. We tested nine different time constants (40-2000 ms) to obtain an optimum facilitation and investigated the effect of this lowpass filter time constant on the model performance.

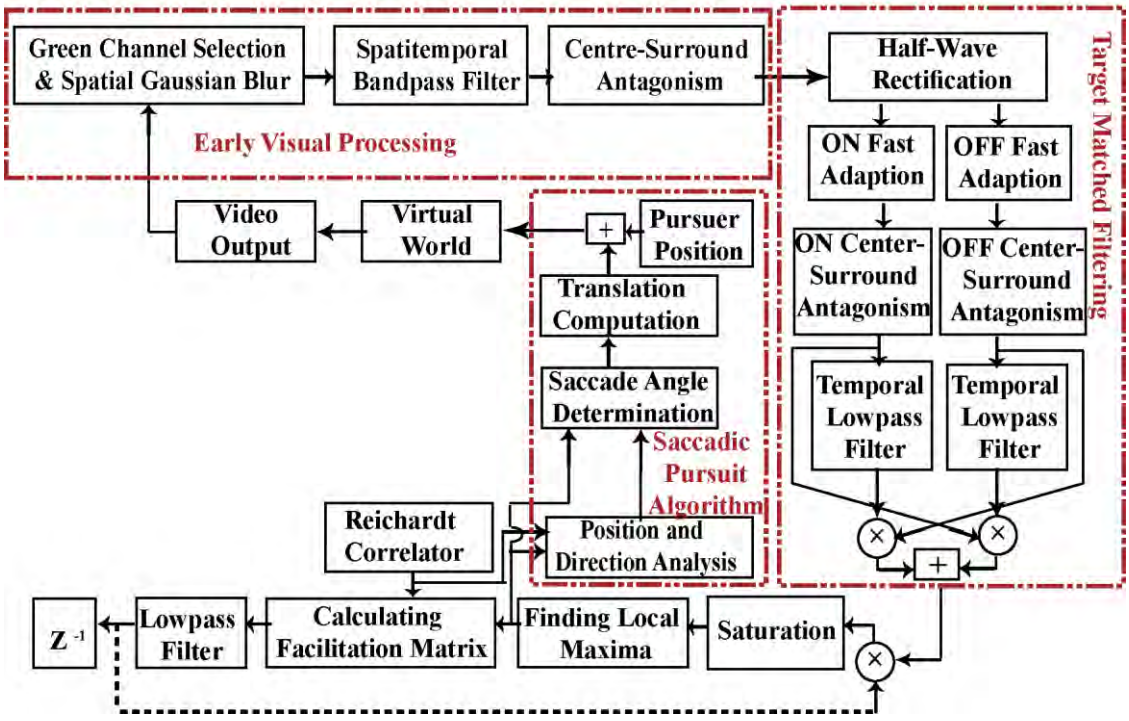

Figure B.2. The block diagram of the closed-loop simulation in MATLAB/Simulink includes; (1) Early visual processing (2) Target matched filtering (3) Added directionality from a Reichardt correlation of the ESTMD output (4) Saccadic Pursuit algorithm.

E. Evaluation

1) Input: Target Contrast Measure

To measure how our target differs from background at each time step of the pursuit in the input (after optical blur), we used magnitude of a weighted SNR (*WSNR*). The weighing function *w(d)* represents the effect of target angular size on its detectability against background:

$$WSNR = w(d)\left|\frac{I_t - \bar{I}_b}{\sigma_b}\right|, \quad \text{(B1)}$$

where I_t represents the target intensity, $\bar{I}_b$ is the mean value of the next nearest neighbours of the background to the target point, *d* is the distance between target and the pursuer, and *w(d)* is defined as:

$$w(d) = \frac{1}{2} + \frac{1}{2}\tanh(a + \frac{b}{d}), \quad \text{(B2)}$$

and *a* and *b* are chosen based on the target size selectivity of the model.

2) Output: Target Discriminability Measure

To determine the discriminability of the target at the output stage, we defined the following metric:

$$Discrim = \frac{M-N}{M} e^{(I_t - I_{max} - \sigma_{I_b})}, \quad \text{(B3)}$$

where *M* is the total number of background pixels, *N* represents the number of background pixels with equal or higher values than the target, I_t is the target intensity, I_{max} is the maximum intensity of the background pixels, and σ_{I_b} is the deviation of background pixels with higher value than the target value, given by:

$$\sigma_{I_b} = \sqrt{\sum_{I_i=I_t}^{I_{max}} n_i (I_t - I_i)^2}. \quad \text{(B4)}$$

where n_i is the number of pixels with the intensity of I_i. Based on this metric, the maximum possible discriminability is e, and whenever the target is the winner at the output, the discriminability value is greater than 1.

III. RESULTS

We ran multiple simulations to determine whether the ESTMD model could successfully discriminate targets of either contrast polarity in various natural scenes. Moreover, to optimize the facilitation mechanism, we examined the effect of time constant on the model performance (Figure B.2) for a variety of target velocities. Additionally, we tested the effectiveness of running closed-loop simulations with and without the facilitation mechanism (for the same path and initial condition).

Figure B.3 shows the success of target pursuit averaged over target intensities versus target velocity and lowpass filter time constant for all four images (45,000 simulations in total). Unsurprisingly, success varies between the images, as expected from their varying degrees of background clutter. However, interestingly, the optimum lowpass filter time constant changes in different images. In more cluttered backgrounds, the pursuit success improves with a higher time constant whilst in the easier images (i.e. Image B) the trend is reversed. One possible explanation for this shift in optimum time constant is that there is frequent camouflaging in the more cluttered backgrounds and the higher time constant leads to enhancement of the area of target disappearance for a longer time. Consequently, the model is able to discriminate the target again and pursue it successfully.

A further 5000 randomized trials were simulated without the facilitation. Figure B.4 shows the success of target pursuits at five target intensities (50 randomized paths at each intensity) for a target velocity of 6 ms^{-1} (3/4 pursuer velocity) at the optimum time constant of each

image. This figure reveals that the ESTMD algorithm is capable of detecting both dark and light contrasting targets.

Expectedly, as target intensity is increased or decreased away from the mean background, pursuit success improves. However, since the target detection depends more on local contrast rather than global contrast, the lowest success pursuit does not happen at the exact mean of the background. In three of the images even in the low contrast scenarios (trough of the curves), the addition of facilitation mechanism improves pursuit success by 2-4%. Nonetheless, as the failure in successful pursuit is almost entirely due to target detection rather than pursuit, this improvement is not significant for these cases. Overall, improvement from facilitation ranges from 0% to 34% in the best-case scenario (for a white target against image A). This figure suggests that the best improvement given by addition of facilitation happens in a more cluttered environment (i.e. Image A) where the model by itself fails to track the target due to its repetitive disappearance.

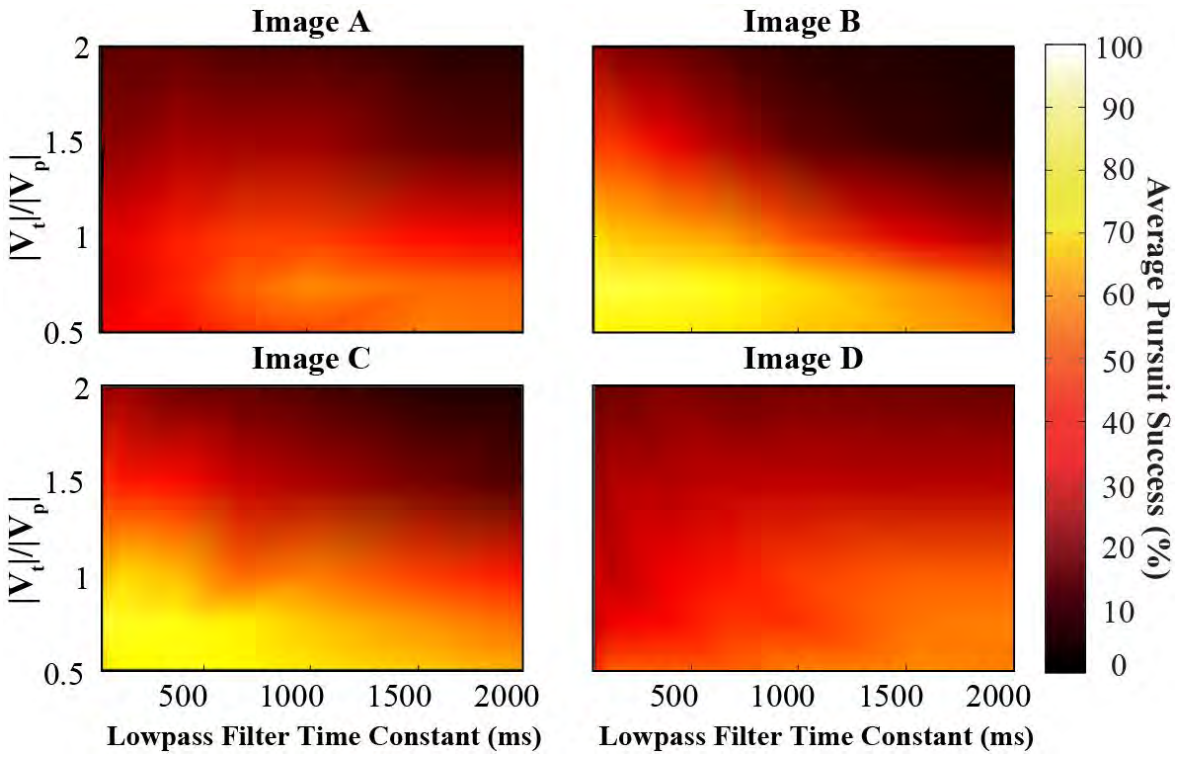

Figure B.3. Average success of target pursuit over target intensities with respect to target velocity and lowpass filter time constant.

Figure B.5 shows three examples of pursuit simulations which the non-facilitated pursuit fails whilst the facilitated case results in successful target capture. In all examples, input target contrast improves throughout the duration of the pursuit since the target moves to the locations that are more optimal with respect to the size tuning inherent in the ESTMD model.

In Example 1 the non-facilitated simulation pursuit fails rapidly, while addition of facilitation reduces pursuit failure and leads to successful target capture. In both Example 2 and 3, the loss of the target in the un-facilitated case is not due to model failure for target detection per se, but rather due to the concurrent detection of distracting (stronger) false positive features. The addition of facilitation boosts the local response in the vicinity of the previously tracked target, maintaining a crude form of 'attentional' focus on the initial target in the presence of such distracters.

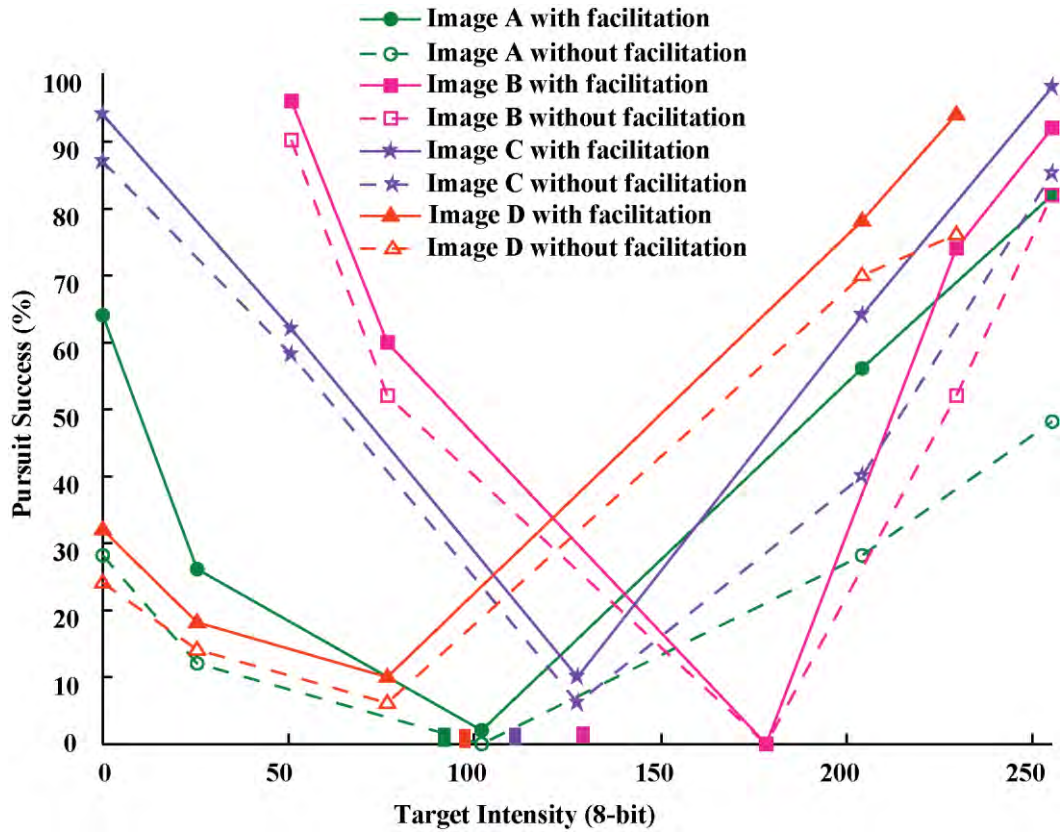

Figure B.4. Pursuit success over a range of target intensities (green channel) from black to white. The small squares on the 'Target Intensity' axis indicate the mean of the background green channel for the image represented with the same colour.

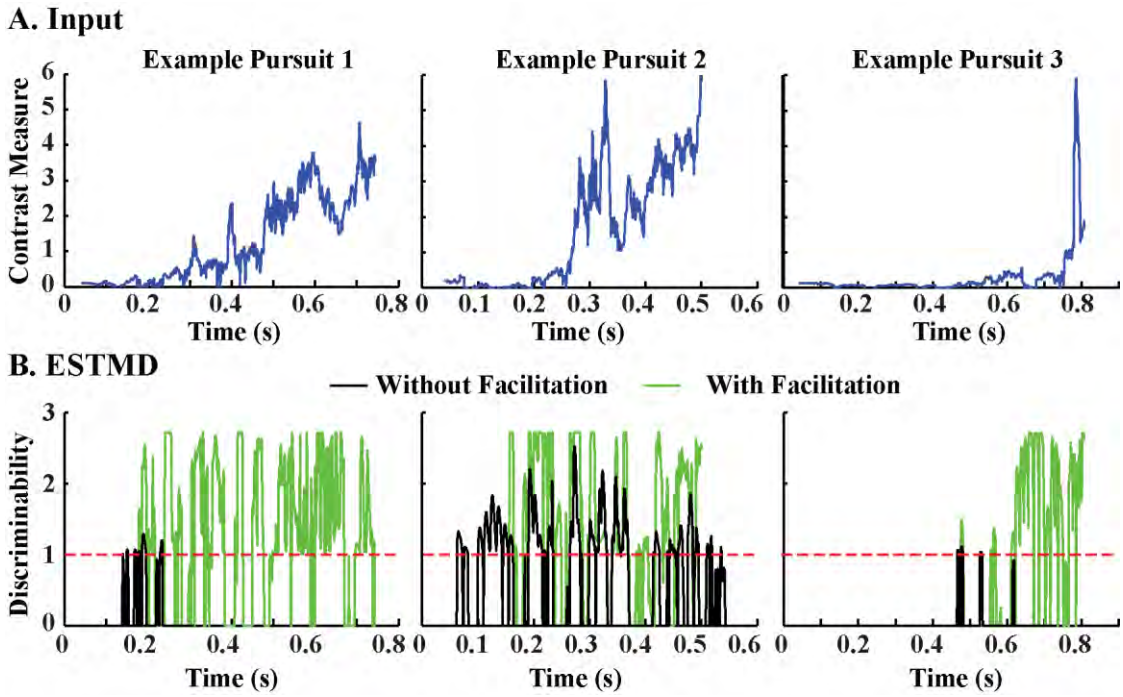

Figure B.5. Examples of target discriminability over the duration of a pursuit at both (a) the input stage (optical blur) and (b) at the ESTMD output. The red dashed line indicates that the target is the current 'winner'.

IV. DISCUSSION

We have demonstrated the efficacy of a target detection and pursuit algorithm inspired directly by insect neurophysiology, with potential future applications for an artificial vision system. Our data show clearly that an elaborated version of the ESTMD model incorporating summation of feature detectors for both dark and light contrasting targets provides robust detection of varying target intensities (of both light and dark) against a wide range of backgrounds, rivalling the remarkable sensitivity for low contrast targets of the insect visual system upon which it is based [29]. Despite this, success inevitably declines as the target approaches the mean intensity of the background.

Although successful in improving target discriminability and successful pursuit, the facilitation mechanism that we have implemented is still a crude approximation to that observed in physiological recordings. Future research will attempt to elucidate the properties of the neuronal architecture that underlies facilitation and we will attempt to

encapsulate this more completely in further model variants. However, even with this 'simple' approximation, we see substantial improvement of both target pursuit success and discriminability over a range of intensities compared with the un-facilitated case.

REFERENCEES

[1] P.S. Corbet, "Dragonflies: Behavior and Ecology of Odonata Ithaca," Cornell University Press, 1999.

[2] S.D. Wiederman and D.C O'Carroll, "Selective Attention in an Insect Visual Neuron," Current Biol, vol.23, pp.156-161, 2013.

[3] Y. Ren, C. Chua and Y. Ho, "Motion Detection with Nonstationary Background," Mach Vision and Appl, vol.13, pp. 332-343, 2003.

[4] N.J. Strausfeld, "Some Quantitative Aspects of the Fly's Brain, in Atlas of an Insect Brain," Springer-Verlag, Heidelberg, Germany, pp.49-56, 1976.

[5] D.C O'Carroll, "Feature-detecting Neurons in Dragonflies," Nature, vol.362, pp.541-543, 1993.

[6] K. Nordström, P.D. Barnett and D.C. O'Carroll, "Insect Detection of Small Targets Moving in Visual Clutter," PLOS Biology, vol.4, pp.378-386, 2006.

[7] K. Nordström and D.C. O'Carroll, "Feature Detection and Hypercomplex Property in Insects," Trends in Neuroscience, vol. 32, pp.383–391, 2009.

[8] S.D. Wiederman, P.A. Shoemaker, and D.C. O'Carroll, "A Model for the Detection of Moving Targets in Visual Clutter Inspired by Insect Physiology," PLoS ONE, vol.3, pp.1-11, 2008.

[9] S.D. Wiederman, R.S. Brinkworth, and D.C. O'Carroll, “Performance of a Bio-Inspired Model for the Robust Detection of Moving Targets in High Dynamic Range Natural Scenes,” J Comput Theor Nanos, vol.7, pp.911-920, 2010.

[10] S.D. Wiederman and D.C. O'Carroll, “Discrimination of Features in Natural Scenes by a Dragonfly Neuron,” J Neurosci, vol.31, pp.7141-7144, 2011.

[11] K.J. Halupka, S.D Wiederman, B.S Cazzolato and D.C O'Carroll, “Discrete Implementation of Biologically Inspired Image Processing for Target Detection,” In IEEE 7th ISSNIP, pp. 143-148, 2011.

[12] K. Nordström, D.M Bolzon and D.C O’Carroll, “Spatial Facilitation by a High-performance Dragonfly Target-detecting Neuron,” Biology Letters, vol.7, pp.588–592, 2011.

[13] J.R Dunbier, S.D Wiederman, P.A Shoemaker and D.C. O’Carroll, “Facilitation of Dragonfly Target-detecting Neurons by Slow Moving Features on Continuous Paths,” Front. Neural Circuits vol.6, 2012.

[14] R.M Olberg , A.H. Worthington and K.R. Venator, “ Prey Pursuit and Interception in Dragonflies” J Comp Physiol A, vol.186, pp.155–162, 2000.

[15] K.J. Halupka, S.D. Wiederman, B.S. Cazzolato and D.C. O’Carroll, “Bio-inspired Feature Extraction and Enhancement of Targets Moving Against Visual Clutter During Closed Loop Pursuit,” In IEEE ICIP, Melbourne, 2013.

[16] S.D Wiederman and D.C. O’Carroll, “Correlation Between OFF & ON Channels Underlies Dark Target Selectivity in an Insect Visual System,” J Neurosci, vol. 33, pp.13225-13232, 2013.

[17] C. Schilstra and J. van Hateren. "Blowfly flight and optic flow. I. Thorax kinematics and flight dynamics," J Exp Biol, vol. 202, pp.1481-1490, 1999.

[18] D. Field, "Relations between the Statistics of Natural Images and the Response Properties of Cortical Cells," OSA. A, vol.4, pp. 2379-2394, 1987.

[19] D.G. Stavenga, "Angular and Spectral Sensitivity of Fly Photoreceptors. I. Integrated Facet Lens and Rhabdomere Optics," J Comp Physiol A, vol. 189, pp. 1-17, 2003.

[20] M.V. Srinivasan and R.G. Guy, "Spectral Properties of Movement Perception in the Dronefly Eristalis," J Comp Physiol A, vol.166, pp. 287-295, 1990.

[21] A.D. Straw, E.J. Warrant and D.C. O'Carroll, "A Bright Zone in Male Hoverfly (Eristalis tenax) Eyes and Associated Faster Motion Detection and Increased Contrast Sensitivity," J Exp Biol, vol. 209, pp.4339-4354, 2006.

[22] S.D. Wiederman, R.S.A. Brinkworth and D.C. O'Carroll, "Bio-inspired Small Target Discrimination in High Dynamic Range Natural Scenes," In IEEE 3rd BIC-TA, pp. 109-116, 2008.

[23] D. Osorio, "Mechanisms of Early Visual Processing in the Medulla of the Locust Optic Lobe-How Self-inhibition, Spatial-pooling, & Signal Rectification Contribute to the Properties of Transient Cells," Visual Neuroscience, vol.7, pp.345-3, 1991.

[24] N. Jansonius and J. van Hateren, "Fast Temporal Adaptation of On-off Units in the First Optic Chiasm of the Blowfly," J Comp Physiol A, vol.168, pp.631–637, 1991.

[25] B. Hassenstein and W. Reichardt, "Analyse der zeit-, reihenfolgen-und vorzeichenauswertung bei der bewegungsperzeption des rüsselkäfers Chlorophanus," Zeitschrift für vergleichende Physiologie, vol.11b, pp.513–524, 1956.

[26] S.D. Wiederman and D.C. O'Carroll, "Biomimetic Target Detection: modeling 2nd order correlation of OFF &ON channels," In IEEE 2013 CIMSVP, Singapore, 2013.

[27] C. Wehrhahn, T. Poggio, L. Bulthoff, "Tracking &Chasing in Houseflies," Biological Cybernetics, vol.45, pp.123-130, 1982.

[28] M.F. Land, T.S. Collett, "Chasing Behaviour of Houseflies (Fannia Canicularis)," J Comp Physiol A,vol.89, pp.331-357, 1974.

[29] D.C O'Carroll, and S.D. Wiederman. "Contrast sensitivity and the detection of moving patterns and features." Philos T R Soc B, vol.369, no 1636, 2014.

Statement of Authorship

Title of Paper	Robustness and Real-Time Performance of an Insect Inspired Target Tracking Algorithm Under Natural Conditions
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	IEEE Symposium Series on Computational Intelligence, 2015, Cape Town, pp. 97-102. DOI: 10.1109/SSCI.2015.24

Principal Author

Name of Principal Author (Candidate)	Zahra Bagheri
Contribution to the Paper	Developed the code, designed the experiments, ran the model simulations, analysed and interpreted the data, wrote the manuscript with editing contributions from the other authors.
Overall percentage (%)	70
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 6/7/2017

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Steven Wiederman
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, assisted with analysis and interpretation of the data, provided editing contribution (10%).
Signature	Date 6/7/2017

Name of Co-Author	Benjamin Cazzolato
Contribution to the Paper	Participated in the initial conceptualisation and experimental and design, assisted with analysis and interpretation of the data, provided editing contribution (10%).
Signature	Date 6/7/17

Name of Co-Author	Steven Grainger
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, read and approved the draft (5%).
Signature	Date 6/12/17

Name of Co-Author	David O'Carroll
Contribution to the Paper	Participated in the initial conceptualisation and experimental design, provided editing contribution (5%).
Signature	Date 15-NOV-2016

Appendix C. Robustness and Real-Time Performance of an Insect Inspired Target Tracking Algorithm Under Natural Conditions

Zahra M. Bagheri[1,2], Steven D. Wiederman[1], Benjamin S. Cazzolato[2], Steven Grainger[2] and David C. O'Carroll[1,3]

[1] School of Medicine, [2] School of Mechanical Engineering, The University of Adelaide, Adelaide, Australia

[3]Department of Biology, Lund University, Lund, Sweden

Abstract— Many computer vision tasks require the implementation of robust and efficient target tracking algorithms. Furthermore, in robotic applications these algorithms must perform whilst on a moving platform (egomotion). Despite the increase in computational processing power, many engineering algorithms are still challenged by real-time applications. In contrast, lightweight and low-power flying insects, such as dragonflies, can readily chase prey and mates within cluttered natural environments, deftly selecting their target amidst distractors (swarms). In our laboratory, we record from 'target-detecting' neurons in the dragonfly brain that underlie this pursuit behavior. We recently developed a closed-loop target detection and tracking algorithm based on key properties of these neurons. Here we test our insect-inspired tracking model in open-loop against a set of naturalistic sequences and compare its efficacy and efficiency with other state-of-the-art engineering models. In terms of tracking robustness, our model performs similarly to many of these trackers, yet is at least 3 times more efficient in terms of processing speed.

I. INTRODUCTION

Many target tracking algorithms have been developed over the last decade for a diverse range of applications, e.g. surveillance, human assistance robots, wildlife monitoring and smart cars. An ideal visual tracker accounts for different problems such as illumination changes, rapid changes in target appearance, non-smooth target trajectories, occlusion and background clutter. Many engineering methods developed for target tracking simplify these scenarios, ensuring a more tractable tracking problem. Moreover, these methods involve complex computation (e.g. particle filters), that require large, high-powered processors. Consequently, these solutions are often impractical in real-time applications, particularly where probability is desirable. These issues highlight the need for an alternative and more efficient approach to solving at least a subset of the target tracking problems.

Studies of insect visual systems suggest there is a solution contained within a 'simple' neuronal architecture (~ 1 million neurons). For example, the dragonfly is a remarkable aerial predator which detects, selects and then chases prey or mates within a visually cluttered surround even in the presence of other distracting stimuli, such as swarms of prey and conspecifics [1], [2]. The dragonfly performs this task despite its light-weight and low-power brain and its low-resolution visual system (acuity of ~1°). The neuronal algorithms behind such a robust and efficient target tracking behaviour (the envy of engineers) is currently being elucidated by our lab and other neuroscientists in the field.

Our approach to engineering a solution to this target tracking problem is to model the neuronal pathway that underlies the dragonfly pursuit behaviour. We record from 'small target motion detector' (STMD) neurons of the insect lobula in response to different visual stimuli. These neurons are size selective, velocity tuned, contrast sensitive, and respond robustly to small moving targets even in the presence of background motion [3-6]. Inspired

by such electrophysiological recordings from STMD neurons, we previously developed an algorithm for local target discrimination [7]. This 'elementary' small target motion detector (ESTMD) model provides nonlinear spatiotemporal matched filtering for small moving targets embedded in natural scenery [7]. Recently, we elaborated this model to include the recent observations of response 'facilitation' [8,9] (a slow build-up of response to targets that move on long, continuous trajectories) [10-14]. We implemented this elaborated model in a closed-loop target tracking system that uses an active saccadic gaze fixation strategy inspired by insect pursuit behaviour [10-14]. Using this closed-loop model we showed that facilitation improves the robustness of pursuit [14]. We also investigated the effect of different environmental variables (background clutter, target contrast, target velocity) and model parameters (spatial and temporal components of facilitation) on pursuit success. Our model predicted an optimal, dynamic behaviour for a temporal component of facilitation that was dependent on background clutter [14].

Although our model showed robust performance in a constrained virtual-reality environment, natural conditions such as illumination changes, local flicker and target occlusion could affect model behaviour. In this paper we test the robustness of our model in open-loop, using videos recorded from natural scenes [15]. This allows us to compare the processing speed and tracking performance of our insect-inspired model with several state-of-the-art engineering algorithms.

II. METHODS

A. Dataset

We used 15 different image sequences downloaded from a publicly available dataset [15]. These sequences had different lengths ranging from 80 to 3000 frames (with an average of

558 frames). Figure C.1 shows a snapshot of these videos at the midpoint of each sequence. All of these videos included camera motion.

Figure C.1. A single frame 'snapshot' of the videos [15] used to test both the performance of our insect-inspired model as well as other previously published tracking models [23-28].

B. Insect-Inspired Target Tracking Model

Figure C.2 shows an overview of the insect inspired target tracking model implemented in MATLAB. The optics of insect compound eyes are limited by diffraction of the facet lenses [16]. We modelled this optical blur with a Gaussian function of full-width at half maximum of 1.4° [16]. We selected only the green channel of the RGB input to simulate the sensitivity of typical insect motion sensitive pathways to green light [17]. Further sub-sampling was applied to the blurred image to model the average inter-receptor angle between photoreceptors [18]. In biological systems, early visual processing by the photoreceptors themselves and 1st order interneurons remove redundant information in space and time, using neuronal adaptation and center-surround antagonism. These properties of visual system were simulated with spatiotemporal bandpass filtering matched to properties observed in insect vision [19].

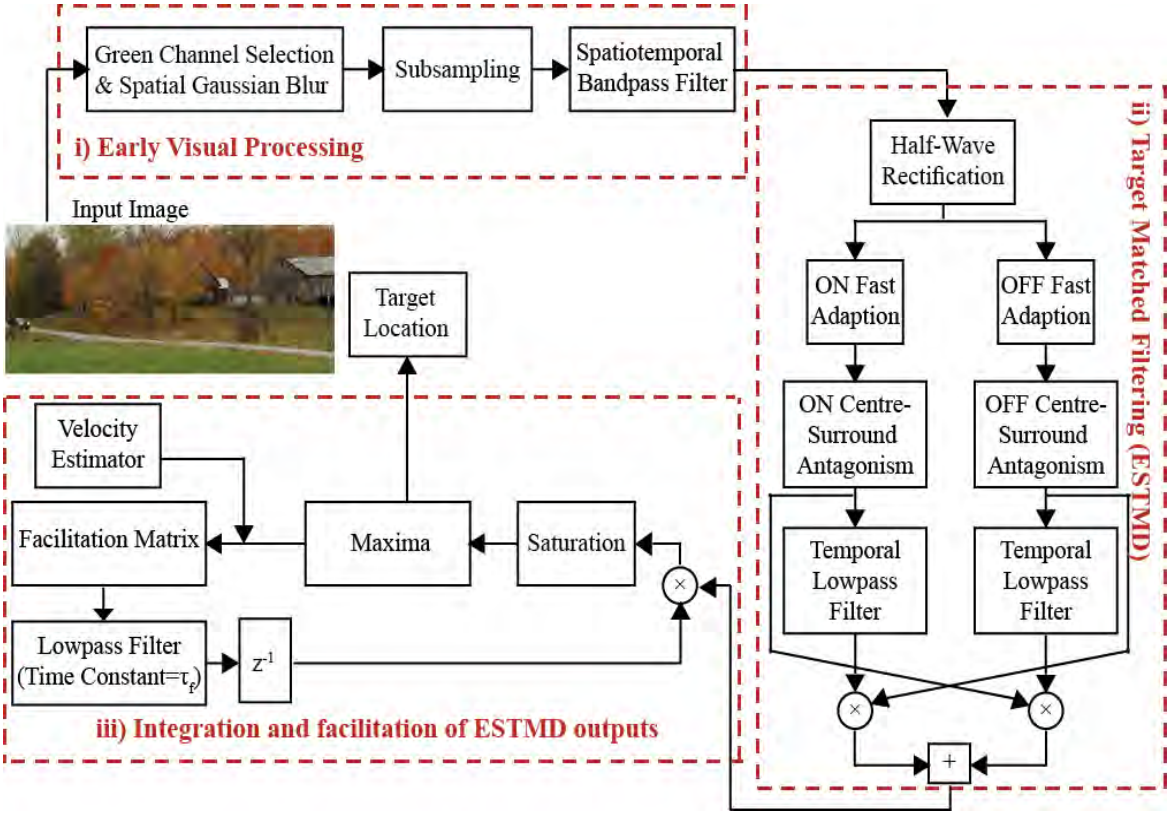

Figure C.2. The overview of the insect-inspired target tracking algorithm. This model is composed of three main subsystems: i) early visual processing, ii) target matched filtering (ESTMD) and iii) integration and facilitation of ESTMD outputs.

The ESTMD subsystem starts with modelling the response properties of rectifying, transient cells as observed in several insect species [20, 21] by separating the ON and OFF contrasts via temporal high pass filtering (τ=40 ms) and half-wave rectification [7]. These independent ON and OFF channels were further processed through a fast adaptive mechanism. The state of adaptation was determined by a nonlinear filter which switches its time constant [7, 11]. Time constants are 'fast' (τ=3 ms) when channel input is increasing and 'slow' (τ=70 ms) when decreasing. This adaptation state causes subtractive inhibition of the unaltered 'pass-through' signal. Additionally, we implemented strong spatial centre-surround antagonism, with each channel surround inhibiting its next-nearest neighbours. This strong surround antagonism conveys selectivity for local edge features. Sensitivity to both dark and light targets was provided by multiplying each contrast channel (ON or OFF) with a delayed

version of the opposite polarity (via a low-pass filter, τ=25 ms) and then summing the outputs [12-14]. The neuron-like soft saturation of each resulting ESTMD was modeled with a non-linear saturation using a hyperbolic tangent function. This serves to ensure all signals lie between 0 and 1. A simple form of the competitive selection observed in dragonflies [2] was modeled by choosing the maximum ESTMD output as the target location.

The slow build-up of facilitation as observed in several dragonfly STMD neurons [8,9] permits the extraction of the target signal from noisy (cluttered) environments. This facilitation mechanism was modeled by building a weighted 'map' dependent on the location of the winning feature but shifted by the target velocity vector [14]. The directional component of this velocity vector was provided using a traditional bio-inspired direction selective model; the Hassenstein-Reichardt elementary motion detector (HR-EMD) [23]. The HR-EMD uses two spatially separated contrast signals and correlates them after a delay (via a low-pass filter). Additionally, the output of the HR-EMD was segmented into three equal intervals to estimate the range of the spatial component of the target velocity [14]. We multiplied the ESTMD model output with a low-pass filtered version of this 'facilitation map' (Figure C.2). The time constant of this filter controls the duration of the enhancement around the predicted location of the winning feature.

C. Benchmarking Algorithms

To establish the computational efficiency of our insect-inspired tracker (IIT) model, we compared its performance with six recent highly-cited algorithms for which code is publicly available. For a fair comparison with respect to processing speed we chose MATLAB implementations of these algorithms.

1- **Incremental visual tracker (IVT)** [24] proposes an adaptive appearance model which stores the latest eigenvectors of the target image and deletes the old observations.

2- **L1-minimization Tracker (L1T)** [25] employs sparse representation by L1 to provide an occlusion insensitive method. This method ignores the target image samples with small probabilities to reduce the cost of computation associated with L1 minimization.

3- **Locally Orderless Tracker (LOT)** [26] proposes a joint spatial-appearance adaptive mechanism to calculate the extent of local disorder in the target. This allows the algorithm to track both rigid and non-rigid targets.

4- **Super Pixel Tracker (SPT)** [27] embeds a discriminative classifier in super pixel (group of pixels which have similar characteristics) clustering to handle changes in scale, motion and occlusion.

5- **Tracking, Learning and Detection (TLD)** [28] is ranked as one of the most resilient available trackers. It combines a discriminative learning method with a detector and an optical flow tracker.

6- **Compressive Tracking (CT)** [29] proposes an appearance model based on features extracted in the compressed domain.

All models were tested in Matlab (R2012b) on the same PC with an Intel i7 3770 CPU (3.4 GHz) and 16 GB RAM. The location of a target bounding box in the initial frame was provided for the benchmark algorithms. Likewise, in the initial frame, we biased our IIT model toward the location of the target by allowing the facilitation to build up in the target region for 40 ms prior to the start of the experiment.

III. RESULTS

Comparing the robustness of different target tracking algorithms is a challenging task since different metrics could be analysed (e.g. scale, shape representation). Here we limit our measure of tracking robustness to correctly locating the target position in each frame. We used different metrics to compare the robustness of the algorithms as well as the processing speed of the trackers.

A. Success Plot

The engineering algorithms represent the target with a bounding box. Therefore, we scored each frame as a successful detection of the target if the centre of the bounding box was within the ground truth box. Similarly, for our IIT algorithm, if the location of the winning feature was within the ground truth box it was considered a successful detection of the target.

Figure C.3 shows the box-and-whiskers plots summarizing the success of all 7 trackers for the 15 different test sequences. On each box, the central mark is the median success rate, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points that are not considered outliers (n=15). The IVT algorithm has the highest median which shows it was capable of correctly locating the target in all frames in half of the sequences. However, the 25 percentile and lowest value are at 30% and 5% respectively, indicating a lack of flexibility of this model under certain circumstances.

Among all algorithms, TLD performs more reliably under different conditions (i.e. it has the highest 25th percentile). Unlike our 'simple' feed-forward computations, these trackers contain several complexities (as described in the method section). Despite this difference, the median of our algorithm (IIT) indicates a performance on par with these other algorithms. Additionally, our model has the lowest inter-quartile range (distance between the 25th and

75th percentiles) showing that our model can perform as robustly as the state-of-the-art engineering algorithms under different natural conditions.

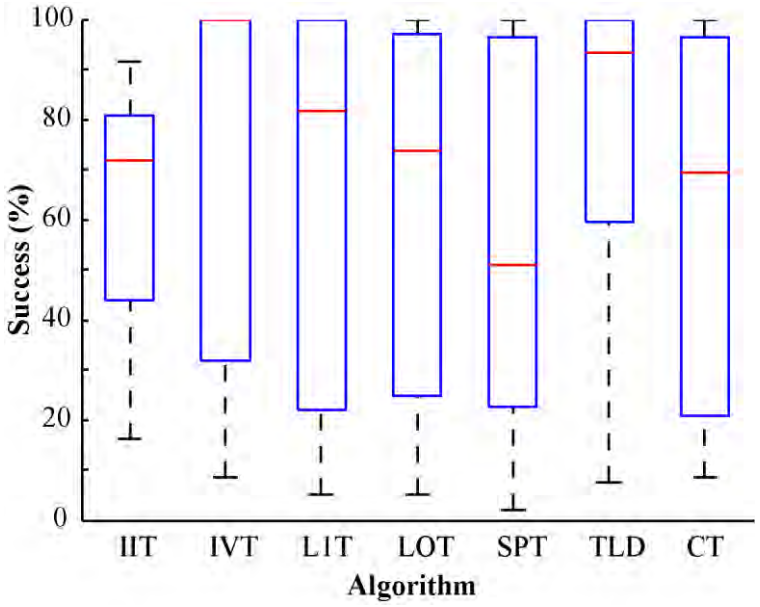

Figure C.3. Box and whiskers plot for successful target tracking of different algorithm for all 15 different image sequences.

B. Precision Plot

The Precision plot is an evaluation method recently adopted to measure the robustness of tracking [15, 30, 31]. It shows the percentage of the frames where the Euclidean distance between the center of the tracked target and the ground truth is within a given 'location error' threshold. Figure C.4 shows the precision plot for all trackers. A higher precision at low thresholds means the tracker is more accurate.

Figure C.4 shows that our algorithm (IIT) has the best precision at the threshold of zero. Between thresholds of 0 and 10 pixels its precision increases rapidly, however is still below the ultimate precision of TLD, L1T and IVT. The main reason behind this behaviour is likely the size selectivity of our model; i.e., it is tuned to small sized objects. Large objects are composed of small parts allowing our model to lock on to these sub-features of the larger object. The result is effective target tracking, but with the location offset from the centre of

the object. Our model's precision increases, catching up to those of other algorithms in the threshold range of 10 and 20 pixels. By a location error threshold of 20, our IIT exceeds the precision of all trackers except TLD. The precisions at the 20 pixel threshold widely used as a performance benchmark in the computer vision literature [15], [30], [31] are given as the representative precision score in Table C.1.

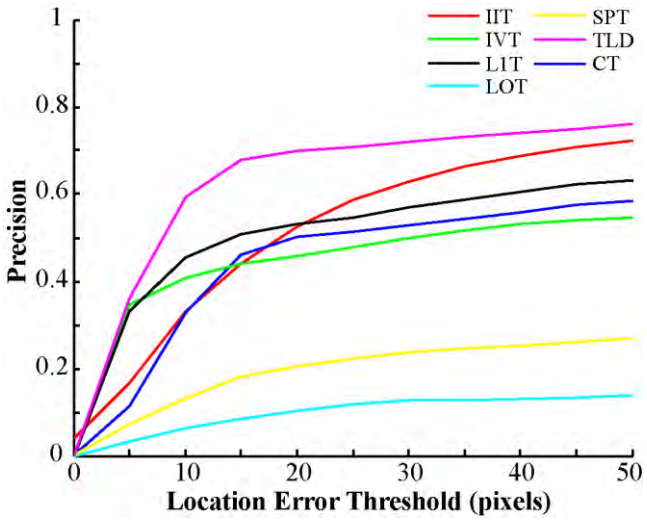

Figure C.4. Precision plot for all 15 sequences.

C. Overall Performance

Table C.1 provides a descriptive summary of performance averaged across all 15 videos. In addition to the average success rate of the 15 sequences, we also calculated the weighted success which shows the percentage of the successful frames out of all the 8374 tested frames. This normalization accounted for the difficulty of 'long term' tracking, where it is easier for the trackers to lock on to the target in a short sequence than a long one.

Table C.1 shows that the average success for our model is below that of TLD and IVT and close to LOT and L1T. However, when it comes to weighted success, our model takes second place, indicating very good long term tracking performance. Our facilitation mechanism (based on the recently observed facilitatory behaviour of target-detecting neurons [8,9])

builds up slowly in response to targets that move in long continuous trajectories, thus improves target detection as tracking progresses.

Table C.1. Summary of experimental results on the 15 video dataset.

Performance Measure	**Algorithm**						
	IIT	***IVT***	***L1T***	***LOT***	***SPT***	***TLD***	***CT***
Average Success (%)	62.7	74.0	63.8	62.2	55.6	74.5	56.6
Weighted Success (%)	73.0	62.3	57.6	34.4	24.6	86.9	48.7
Precision (20 px)	0.53	0.46	0.53	0.01	0.21	0.70	0.50

D. Processing Speed

Although comparable in tracking performance, our model excels in processing efficiency, a critical concern in target tracking applications. Indeed, many trackers are considered impractical in real-time scenarios due to their long processing duration. Figure C.5 shows the processing speed of the tested algorithms, with the IIT exceeding all other trackers (note the logarithmic scale). Our model performs approximately 12 times faster than IVT and TLD and 3 times faster than CT.

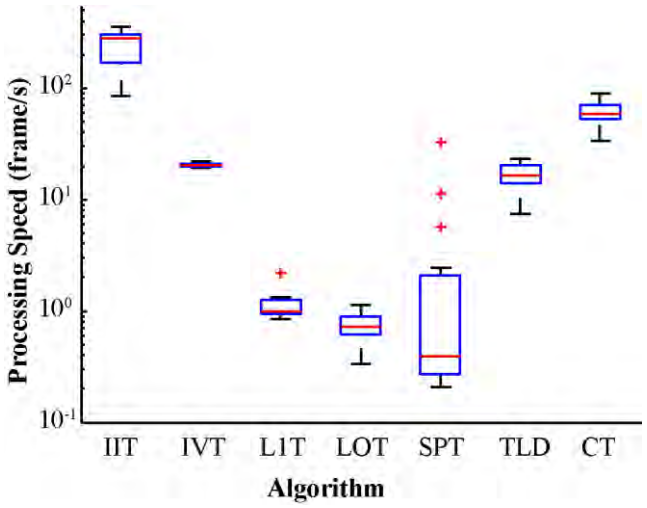

Figure C.5. Processing speed of trackers.

IV. CONCLUSION

We have demonstrated the robustness and efficiency of a target tracking algorithm inspired directly by insect neurophysiology. Our data clearly shows that this model can perform robustly under natural conditions. Despite the relatively simple mechanism we implemented, the robustness of our model can compete with the state-of-the-art engineering trackers. A limitation of our model is that it was primarily designed to detect and track small moving targets. Therefore, it only tracks larger objects composed of smaller moving parts (within the size tuning range of our model). This limits its overall performance robustness compared with the best of the engineered trackers (such as TLD). Nevertheless, in terms of processing speed, our model outperforms all of the engineering trackers, mimicking the remarkable efficiency of the insect visual system upon which it is based. As such, it may be well suited to applications where efficiency is paramount.

Here, we tested our algorithm in open-loop, however, active vision may be a key to exploiting visual information by the simple insect brain for complex tasks such as target tracking. Future research will attempt to implement this model along with insect active

vision system in a robotic platform to test the performance of them together under real-world conditions.

Acknowledgment

This research was supported under Australian Research Council's Discovery Projects (project number DP130104572) and Discovery Early Career Researcher Award (project number DE15010054) funding scheme. We thank the manager of the Botanic Gardens in Adelaide for allowing insect collection and video capture.

References

[1] P. S. Corbet, Dragonflies: Behaviour and ecology of Odonata. Cornell University Press, 1999.

[2] S. D. Wiederman and D. C. O'Carroll, Selective attention in an insect visual neuron. Curr. Biol., vol. 23, pp. 156-161, 2013.

[3] D. C. O'Carroll, Feature-detecting neurons in dragonflies. Nature, vol. 362, pp. 541-543, 1993.

[4] K. Nordström, P. D. Barnett and D. C. O'Carroll (2006) Insect detection of small targets moving in visual clutter. PLoS. Biol., vol. 4, pp. 378-386, 2006.

[5] K. Nordström and D. C. O'Carroll DC, Feature detection and hypercomplex property in insects. Trends Neurosci., vol. 32, pp. 383-391, 2009.

[6] D. C. O'Carroll and S.D. Wiederman, Contrast sensitivity and the detection of moving patterns and features. Philos. Trans. R. Soc. B-Biol. Sci., vol. 369, pp. 20130043, 2014.

[7] S. D. Wiederman, P. A. Shoemaker, and D. C. O'Carroll, A model for the detection of moving targets in visual clutter inspired by insect physiology. PLoS ONE, vol. 3, pp. 1-11, 2008.

[8] J. R. Dunbier, S. D. Wiederman, P. A. Shoemaker and D. C. O'Carroll, Modelling the temporal response properties of an insect small target motion detector. IEEE international conference on Intelligent Sensors, Sensor Networks and Information Processing, pp. 125-130, 2011.

[9] J. R. Dunbier, S. D. Wiederman, P. A. Shoemaker and D.C. O'Carroll, Facilitation of dragonfly target-detecting neurons by slow moving features on continuous paths. Front. Neural Circuits, vol. 6, 79, 2012.

[10] K. J. Halupka, S. D. Wiederman, B. S. Cazzolato and D. C. O'Carroll, Discrete implementation of biologically inspired image processing for target detection. IEEE International Conference on Intelligent Sensors,

Sensor Networks and Information Processing (ISSNIP), pp. 143-148, 2011.

[11] K. J. Halupka, S. D. Wiederman, B. S. Cazzolato BS and D. C. O'Carroll, Bio-inspired feature extraction and enhancement of targets moving against visual clutter during closed loop pursuit. IEEE International Conference on Image Processing (ICIP), pp. 4098-4102, 2013.

[12] Z. M. Bagheri, S. D. Wiederman, B. S. Cazzolato, S. Grainger, and D. C. O'Carroll, A biologically inspired facilitation mechanism enhances the detection and pursuit of targets of varying contrast. International Conference on Digital lmage Computing: Techniques and Applications (DlCTA), pp. 1-5, Sep 2014.

[13] Z. Bagheri, S. D. Wiederman, B. S. Cazzolato, S. Grainger, and D. C. O'Carroll, Performance assessment of an insect-inspired target tracking model in background clutter. International Conference on Control Automation Robotics & Vision (ICARCV), pp. 822-826, Dec 2014.

[14] Z. M. Bagheri, S. D. Wiederman, B. S. Cazzolato, S. Grainger and O'Carroll DC, Properties of neuronal facilitation that improve target tracking in natural pursuit simulations. J. R. Soc. Interface vol. 12, 20150083, 2015.

[15] Y. Wu, J. Lim and M. H. Yang, M, Online object tracking: A benchmark. IEEE International Conference on Computer vision and pattern recognition (CVPR), pp. 2411-2418, 2013.

[16] D. G. Stavenga, Angular and spectral sensitivity of fly photoreceptors. I. Integrated facet lens and Rhabdomere optics. J. Comp. Physiol. A, vol. 189, pp. 1-17, 2003.

[17] M. V. Srinivasan and R. G. Guy, Spectral properties of movement perception in the dronefly Eristalis. J. Comp. Physiol. A, vol. 166, pp. 287-295, 1990.

[18] A. D. Straw, E. J. Warrant and D. C. O'Carroll DC (2006) A bright zone in male hoverfly (Eristalis tenax) eyes and associated faster motion detection and increased contrast sensitivity. J. Exp. Biol., vol. 209, pp. 4339-4354, 2006.

[19] S. D. Wiederman, R. S. A. Brinkworth and D. C. O'Carroll, Bio-inspired small target discrimination in high dynamic range natural scenes. International Conference on Bio-Inspired Computing: Theories and Applications, pp. 109-116, 2008.

[20] D. Osorio, Mechanisms of early visual processing in the medulla of the locust optic lobe-how self-inhibition, spatial-pooling, and signal rectification contribute to the properties of transient cells. Visual Neurosci., vol. 7, pp. 345-355, 1991.

[21] N. Jansonius and J. van Hateren, Fast temporal adaptation of on-off units in the first optic chiasm of the blowfly. J. Comp. Physiol. A, vol. 168, pp. 631–637, 1991.

[22] S. D. Wiederman and D. C. O'Carroll, Correlation between OFF and ON channels underlies dark target selectivity in an insect visual system. The Journal of Neuroscience, vol 33, pp. 13225-13232, 2013.

[23] B. Hassenstein and W. Reichardt, Analyse der zeit-, reihenfolgen-und vorzeichenauswertung bei der bewegungsperzeption des rüsselkäfers Chlorophanus. Z. Naturforsch, vol. 11b, pp. 513–524, 1956.

[24] D. A. Ross, J. Lim, R. S. Lin and M. H. Yang, Incremental learning for robust visual tracking. Int. J. Comput. Vis., vol. 77, pp. 125-141, 2008.

[25] X. Mei, H. Ling, Y. Wu, E. Blasch and L. Bai, Minimum error bounded efficient ℓ1 tracker with occlusion detection. IEEE International Conference on Computer vision and pattern recognition (CVPR), pp. 1257-1264, 2011.

[26] S. Oron, A. Bar-Hillel, D. Levi and S. Avidan, Locally orderless tracking. IEEE International Conference on Computer vision and pattern recognition (CVPR), pp. 1940-1947, 2012.

[27] S. Wang, H. Lu, F. Yang, M. H. Yang, Superpixel tracking. IEEE International Conference on Computer Vision (ICCV), pp. 1323-1330, 2011.

[28] Z. Kalal., K. Mikolajczyk and J. Matas, Tracking-learning-detection. IEEE Trans. Pattern Anal. Mach. Intell.,vol. 34, pp. 1409-1422.

[29] K. Zhang, L. Zhang and M. H. Yang. Real-time compressive tracking. Springer Computer Vision–ECCV, pp. 864-877, 2012.

[30] B. Babenko, M. H. Yang and S. Belongie, Robust object tracking with online multiple instance learning. IEEE Trans. Pattern Anal. Mach. Intell, vol. 33, pp. 1619-1632, 2011.

[31] J. F. Henriques, R. Caseiro, P. Martins and J. Batista, Exploiting the circulant structure of tracking-by-detection with kernels. In Springer Computer Vision–ECCV, pp. 702-715, 2012.

Appendix D. Supplementary Material for Chapter 4

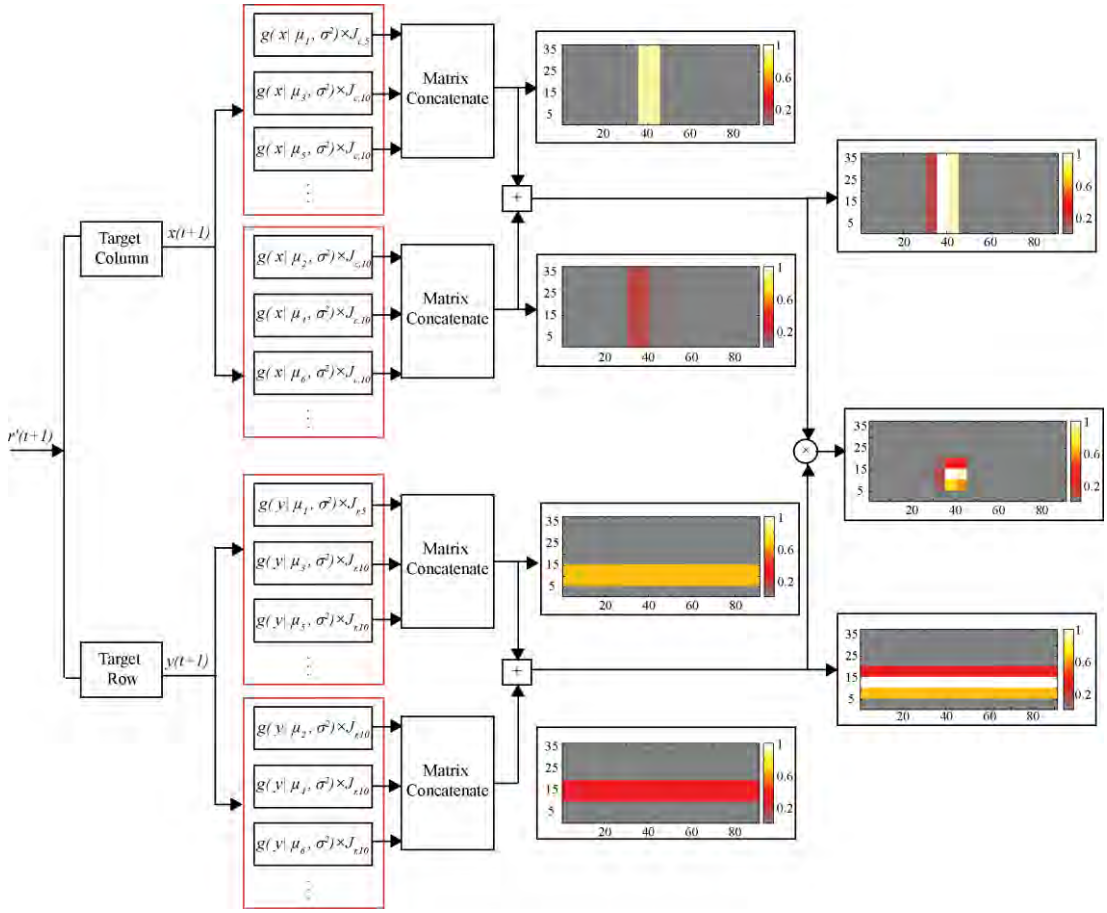

Figure D.1. Process of calculating facilitation matrix (this is referred to as *FG(r')* in Fig.1). The estimated location of the target in the next frame is the input to this process. The weight of each STMD in the facilitation matrix was calculated by Gaussian kernels $g(x|\mu_i,\sigma^2) = e^{(-0.5(x-\mu_i)^2/\sigma^2)}$ in both horizontal and vertical direction. *x* and *y* refer to estimated column and row of the target, μ_i={0, 5, 10, 15, 20, 25, ...} is the mean and σ is the standard deviation. Noting that the full-width at half maximum is desired to be 5°, the standard deviation of the Gaussian is therefore $\sigma = \frac{5}{2\sqrt{2ln2}}$ which provides 50% overlap between neighbouring small-field STMDs. $J_{k,h}$ is a $k \times h$ all-ones matrix. *r* and *c* represent the number of rows and columns in the output of ESTMD, respectively. The output of each stage is shown for *x(t+1)*=39 and *y(t+1)*=12.

Appendix E. Supplementary Material for Chapter 5

Table E.1. Statistical properties of the images which were used both in closed-loop simulations and indoor robotic experiments. The clutter is calculated using the method developed by Silk (1995).

Background	**Mean of Background (green channel, 8-bit)**	**Background Clutter**
Field	130	0.15
House	110	0.21
Library	98	0.27
Botanic	92	0.3
Forecourt	97	0.28

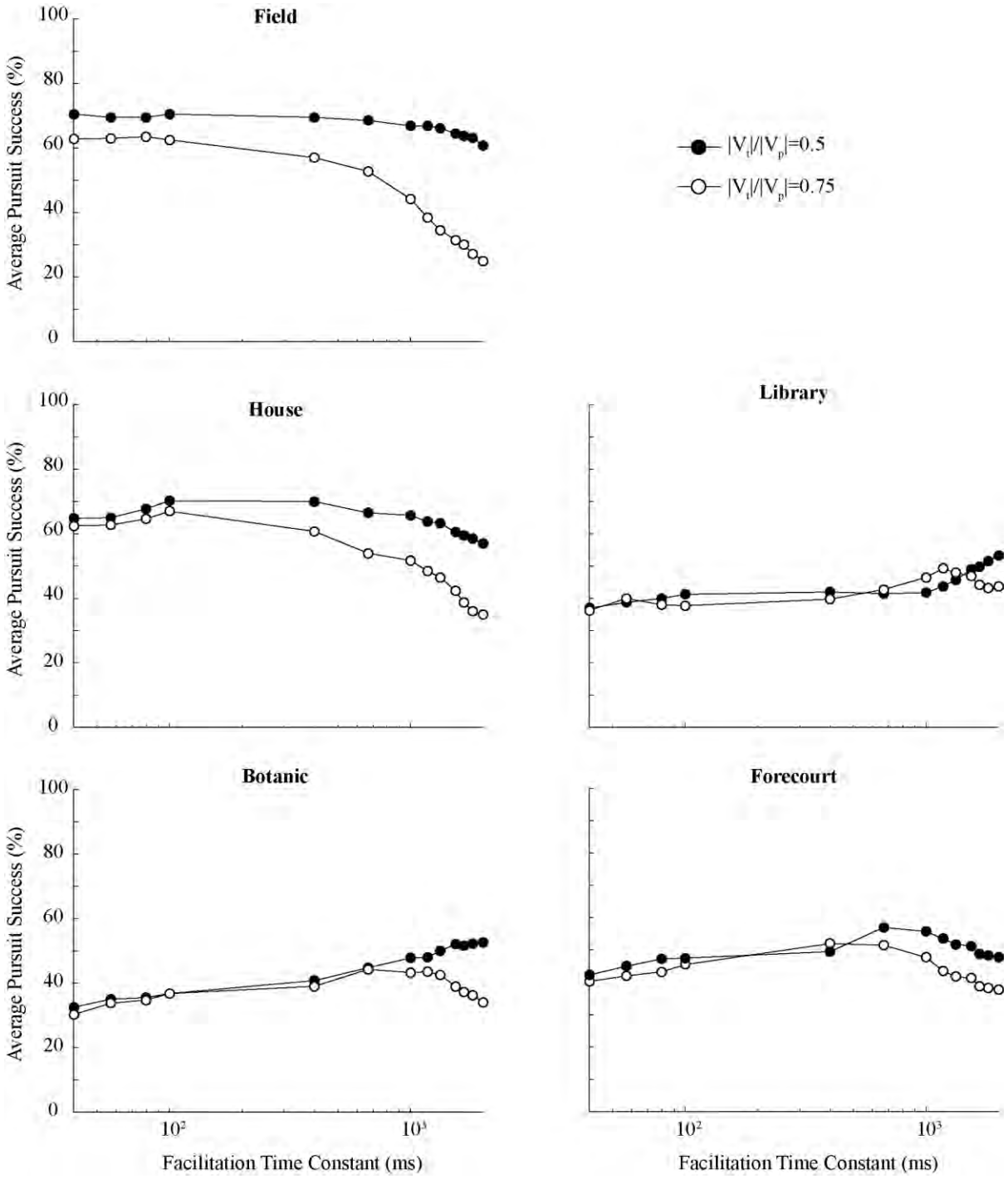

Figure E.1. Results of closed-loop simulations against natural images modified from Figure 4 of Bagheri et al. (2015a). Pursuit success averaged over a range of target intensities for different background images (illustrated in Figure 4a), reveal higher pursuit success (%) for less cluttered scenes (i.e., *Field* and *House*). These results show that the optimum facilitation time constant varies dependent on both target velocity and the background scene.

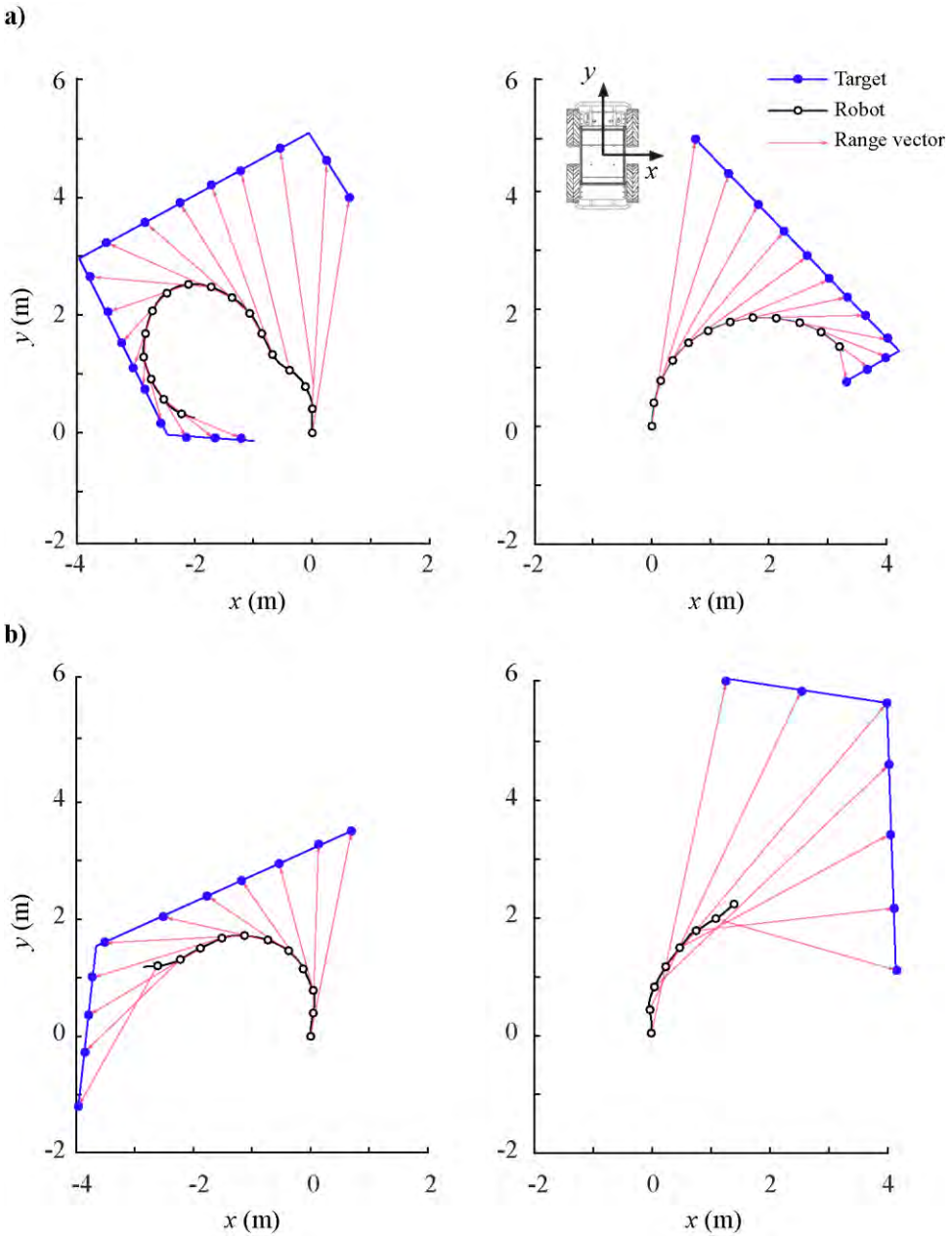

Figure E.2. Top view (xy) of the robot trajectory in a) successful and b) unsuccessful experiments. The y-axis is aligned with the robot wheels at the start of experiment and (x,y)=(0,0) represent the robot position at t=0. The markers show the robot and target positions within 4 s time intervals. In the unsuccessful experiments the target is kept frontally fixated for a while, however, later during the pursuit the target drifts from the centre of field of view as the model cannot detect the target successfully.

Movie 1. '*Backyard*' video which was projected onto the wall in indoor experiments.

Movie 2. '*Park*' video which was projected onto the wall in indoor experiments.

Movie 3. An example of the robot footage robot autonomously tracking target in outdoor environment.

Movie 4. Video output of robot autonomously tracking target in outdoor environment. The purple square marks the location of the winning feature detected by the robot.

Appendix: Mathematical Equations for the Insect-Inspired Target Tracking Model

1) Spatial Gaussian Blur

The optical blur of the compound eye of flying insects is modelled by a Gaussian function (full-width at half maximum of 1.4°) (Stavenga, 2003):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{(-(x-\mu)^2/(2\sigma^2))}, \qquad \text{(E1)}$$

where μ and σ are the mean and standard deviation respectively. Given that the desired full-width at half maximum is 1.4° ($\Delta\rho$=1.4), the standard deviation of the Gaussian is:

$$\sigma = \frac{\Delta\rho}{2\sqrt{2\ln 2}}. \qquad \text{(E2)}$$

2) Temporal Bandpass Filtering of Early Visual Processing

The band-pass temporal properties of early visual processing were simulated with a discrete log-normal transfer function (Halupka et al., 2011):

$$G(Z) = \frac{0.0001z^7 - 0.0011z^6 + 0.0052z^5 - 0.017z^4 + 0.0439z^3 - 0.0574z^2 + 0.1789z - 0.1524}{z^8 - 4.333z^7 + 8.685z^6 - 10.711z^5 + 9.0004z^4 - 5.306z^3 + 2.1448z^2 - 0.5418z + 0.0651} \qquad \text{(E3)}$$

where G(z) is the transfer function of the 8th order temporal filter in the z-domain (sampling time T_s =1 ms).

3) Spatial Highpass Filtering of Early Visual Processing

The weak centre-surround antagonism was modelled by convolving the image with kernel *H*:

$$H = \begin{bmatrix} -1/9 & -1/9 & -1/9 \\ -1/9 & 8/9 & -1/9 \\ -1/9 & -1/9 & -1/9 \end{bmatrix}. \qquad \text{(E4)}$$

4) Half-wave Rectifications of the ESTMD

The separation of the ON and OFF channel was modelled by half-wave rectification (*HWR1*):

$$HWR1 = \begin{cases} ON = \begin{cases} x \; if \; x > 0 \\ 0 \; if \; x \leq 0 \end{cases} \\ OFF = \begin{cases} -x \; if \; x < 0 \\ 0 \; if \; x \geq 0 \end{cases} \end{cases}. \qquad \text{(E5)}$$

A second half-wave rectification was applied to the output of the strong centre-surround antagonism to eliminate the negative values:

$$HWR2 = \begin{cases} x \; if \; x > 0 \\ 0 \; if \; x \leq 0 \end{cases}. \qquad \text{(E6)}$$

5) Centre-surround Antagonism of the ESTMD

Strong spatial centre-surround antagonism was applied by convolving each independent channel with CSA kernel:

$$CSA = \begin{bmatrix} -1 & -1 & -1 & -1 & -1 \\ -1 & 0 & 0 & 0 & -1 \\ -1 & 0 & 2 & 0 & -1 \\ -1 & 0 & 0 & 0 & -1 \\ -1 & -1 & -1 & -1 & -1 \end{bmatrix}. \quad \text{(E7)}$$

6) ESTMD Lowpass Filter:

A discrete first order low-pass filter (τ_{ESTMD}=25 ms, T_s=1 ms) were used to delay each contrast channel (ON or OFF)

$$LP_{ESTMD}(z) = \frac{z+1}{51z-49}. \quad \text{(E8)}$$

7) Saturation:

The neuron-like soft saturation of ESTMD outputs was modeled with a hyperbolic tangent function:

$$s(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad \text{(E9)}$$

8) Facilitation Lowpass Filter:

To mimic the slow build-up of the response of CSTMD1 neurons the ESTMD output was multiplied with a low-pass filtered version of this facilitation map (T_s=1 ms).

$$LP_{Fac}(z) = \frac{z+1}{\left(1+\frac{2\tau_f}{T_s}\right)z + (1-\frac{2\tau_f}{T_s})}, \quad \text{(E10)}$$

we varied the time constant of this discrete low-pass filter (facilitation time constant, τ_f) in the experiments.

Appendix F. Moment of Inertia Metric

In Chapter 3, to provide a metric to represent pursuit success (combining both efficiency and efficacy) at different time constants, was calculated as the first moment of area around the axis perpendicular to x=2 and y=0 (which we refer to this axis as z') in Figure 3.2 and Figure F1. This means that higher frequency of success at shorter capture time yields a higher moment of inertia. For this purpose, the x-axis is normalized by the maximum simulation time (2 s) and the y-axis is divided by the total number of simulations for each time constant (200 simulations), giving a performance metric of

$$I_{z'} = \sum \left(\sqrt{\left(\frac{2-x}{2}\right)^2 + \left(\frac{y}{200}\right)^2}\, \Delta A \right) \qquad \text{(F1)}$$

where $I_{z'}$ is the first moment of area around z' (Θ), and ΔA represents the area of each small element under the curve.

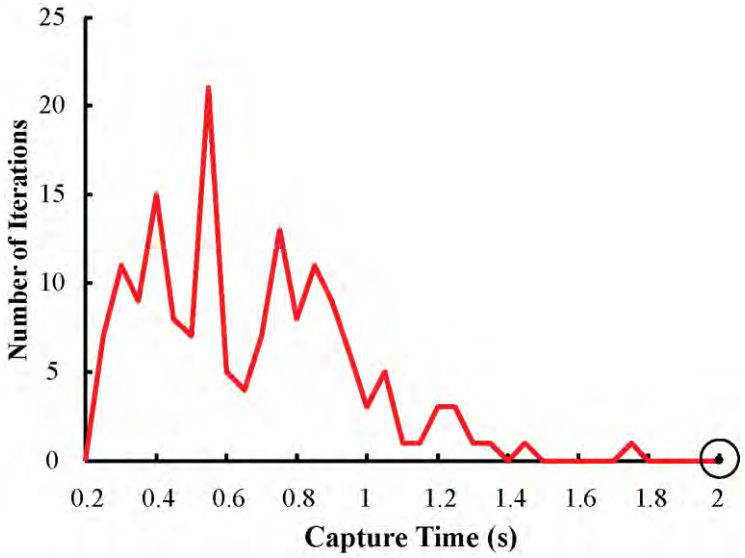

Figure F.1. Distribution of capture time for simulations with target-pursuer velocity ratio (|Vt|/|Vp|) of ¾ (facilitation time constant=133 ms). z' (the dot) is the axis emerging from the page at x=2 and y=0.

Appendix G. STNS Dataset

Figure G1. A single frame ‘snapshot’ of the videos in STNS dataset. The red rectangle shows the target bounding box.

Figure G1. (continued) A single frame ‘snapshot’ of the videos in STNS dataset. The red rectangle shows the target bounding box.

Figure G1. (continued) A single frame ‘snapshot’ of the videos in STNS dataset. The red rectangle shows the target bounding box.