Chapter 1

Introduction

DESPITE their tiny brains, insects display excellent navigational skills and are able to detect motion and estimate velocity reliably. The efficiency of this system has stimulated the interest of engineers seeking inspiration for robots and other seeing machines. Several biologically inspired models of motion detection have been proposed.

These models of the visual systems of insects, with their relative simplicity and efficiency to computer vision based models, have become the building blocks for improving the various techniques used in motion detection and velocity estimation. In this chapter, the visual system of insects and how they perform visual motion detection is explored. The various models of motion detection are also discussed.

1.1 Introduction

It is really amazing to see how insects navigate with seemingly great ease and how they are quite adept in maneuvering their way through this complex and unpredictable world. Despite possessing a tiny brain that contains far fewer nerve cells than our own (insects have only about 10^5 to 10^6 neurons as compared to 10^{12} neurons in humans [Burrows, 1996]), insects are capable of performing a variety of tasks ranging from collision avoidance [Badia et al., 2007] to posture maintenance [Franceschini et al., 2007]. Recent research, observing freely flying insects reveals a number of computational shortcuts that insects use in successfully perceiving and navigating their visual world—this has inspired implementation of simpler systems to detect motion [Yakovleff et al., 1993; Abbott et al., 1994]. Many insects seem to rely mainly on information provided by an array of motion-sensitive detectors in order to navigate [Lehrer et al., 1998; Beare et al., 1995].

The exquisite performance with which insects chase mates or prey, or land and takeoff to escape capture indicates many clever mechanisms that are still yet to be discovered. The neural network of insects plays a predominant role in assisting them to perform such amazing feats. Much work is being carried out not only in the area of insect motion detection, but also in many areas of neuroscience for reconciling neural properties with natural behaviour [Eggermont, 2001; Touryan and Dan, 2001; Egelhaaf et al., 2002; Egelhaaf et al., 2003; Simoncelli, 2003; Karmeier et al., 2006; Nemenman et al., 2007]. Unlike humans, insects have compound eyes, which are two dimensional arrays of tiny eyes called *ommatidia*. Each *ommatidia* looks in a different direction and has its receptive field overlapping with those of its neighbours, thus increasing the sensitivity to a diffuse source. Importantly, behind each of these *ommatidia*, there exists a detecting channel, which may interact with neighbouring channels at different processing levels—each of the levels has a distinct neuronal layer. In other words, the insect visual system is architecturally structured into parallel processing layers and columns [Jarvilehto, 1985].

The visual information from the compound eye of the insect is sent to extensions of the brain known as the optic lobes and is processed mainly in four regions, namely the *lamina*, the *medulla* and two parts of the *lobula complex*. The lamina, which lies under the layer of receptor cells of the eye, receives direct input from the photoreceptors. Axons from the lamina form the thick optic nerve; they invert the image from front to back as they project to the medulla. Individual neurons in the medulla show sensitivity to directional motion between adjacent receptors. Neurons in the lobula complex are sensitive to wide field directed motion, and to stimuli of particular size or shape [Delcomyn, 1998; Kern et

al., 2005; Cuntz et al., 2007]. Compared to other animals, insects have a visual system of intermediate complexity. The secret of the simplicity of insect's visual system is that, in insects, however, unlike other animals, many of the tasks, like the recognition of objects, have either been eliminated or simplified. So in the visual system of insects, motion detection plays a predominant role [Moini et al., 1996].

1.2 Visual motion detection by the fly visual system

Flying insects depend mainly on retinal motion patterns to control and stabilise their course of motion. This pattern of retinal motion, termed as optic flow, is a rich source of information about both the outside world as well as the path and the speed of locomotion. The motor activity of flying insects is still very complex and far from being understood. The visual information obtained by the photoreceptors is conveyed by the receptor axons in the optic lobe, which consists of successive visual neuropils namely the lamina, medulla, lobula and lobula plate as shown in Figure 1.1.

Each visual neuropil is composed of retinotopically arranged columns and superimposed layers. Columns are built up by parallel centripetal and centrifugal small field neurons. The centripetal output cells of the lobula plate project into the brain and the centrifugal cells project from the brain into the lobula plate [Hausen and Egelhaaf, 1989]. Output connections from the optic lobe to the brain are established by both columnar and tangential cells [Hausen, 1982a]. Columnar cells leave the lobula and the lobula plate as dense bundles and terminate in visual centres of the ventrolateral brain termed optic foci [Hengstenberg et al., 1982]. Tangential output neurons of the optic lobe originate from all the neuropils except the lamina. Some of them connect both the optic lobes (heterolateral elements) and others project also into the optic foci. The optic foci can be regarded as the major sensory integration areas for the motor control in the brain [Hausen and Egelhaaf, 1989; Grewe et al., 2006]. Apart from the optic lobes, they also receive information from the ocelli, and mechanosensory inputs from the antennae and the halteres [Hengstenberg, 1998]. The output elements of these are descending neurons, which pass through the cervical connective and terminate in the motor neuropils of the thoracic ganglion [Hausen and Egelhaaf, 1989; Kimmerle and Egelehaaf, 2000]. There is enough evidence to believe that there are connections between output neurons of the optic lobe, descending neurons and motor neurons of the neck muscles [Strausfeld and Seyan, 1985; Strausfeld et al., 1987], the leg musculature [Strausfeld and Bassemir, 1985] and the indirect flight muscles, which

play a major role in torque generation during steering maneuvers [Hausen and Hengstenberg, 1987]. Hence the major visuo-motor pathways between the compound eye and the motor system consist of a sequence of retinotopic visual neuropils—the output elements of which converge together with other sensory tracts onto descending neurons and in the thoracic ganglion these neurons establish divergent connections to the motorneurons of various groups of muscles.

The main motion computation centre in the whole pathway is the lobula plate [Hausen and Egelhaaf, 1989; Farrow et al., 2006; Haag et al., 2007; Beckers et al., 2007; Borst and Haag, 2007]. The neurons of the lobula plate have large dendritic fields [Pierantoni, 1976] and receive retinotopic local motion input from a 2D array of local motion detectors [Egelhaaf et al., 1989; Borst and Egelhaaf, 1992]. Typical output elements of the lobula plate are two classes of giant neurons termed the horizontal system (HS) and vertical system (VS) [Hengstenberg et al., 1982]. The horizontal system consists of three elements, the dendritic arbours of which are located near the anterior surface of the neuropil. According to these dendritic locations, the three cells are termed north, equatorial and south horizontal cells (HSN, HSE and HSS). The three HS neurons (HSN, HSE and HSS) together occupy the whole retinotopic area of the lobula plate. Each of them covers roughly one third of the dorsoventral extent of the neuropil with some overlap. Correspondingly, their receptive field covers the dorsal, equatorial and ventral areas of the ipsilateral visual hemisphere. The dendrites of the HS neurons arborize in the anterior layers of the lobula plate [Hausen, 1982b; Hausen, 1982c]. The vertical cells are a class of 11 output neurons (VS1-11), also occupying the whole retinotopic area of the lobula plate [Hengstenberg, 1982]. Their dendritic fields are more strip-like and oriented dorsoventrally. The dendrites are stacked from distal to the proximal side of the lobula plate and overlap considerably [Krapp et al., 1998; Franz and Krapp, 2000]. The HS and the VS cells are examples of the output neurons of the lobula plate terminating in the ipsilateral part of the brain and do not fire axonal action potentials. Additional horizontal motion cells including H1, H2 and feature detecting (FD) cells [Egelhaaf, 1985; Kimmerle and Egelehaaf, 2000] project to the contralateral hemisphere and fire spikes and are called heterolateral elements as they connect the left and the right lobula plates. There are also centrifugal cells which project from the brain into the lobula plate, namely the dorsal and ventral centrifugal horizontal cells, which project into the dorsal and ventral half of the ipsilateral lobula plate respectively [Hausen, 1982a; Hausen and Egelhaaf, 1989].

Dipteran compound eyes with retinular cells 1 to 6 have a twist in the rhabdom orientation along its length that weakens their sensitivity to polarized light. Although all retinular cells 1 to 6 exhibited a preferred polarization plane, sustaining and on-off units did not. Likewise, motion detection and optomotor responses were insensitive to the polarization plane for long spatial wavelength stimuli; however, sensitivity to select polarization planes was observed for short spatial wavelengths [McCann and Arnett, 1972]. As such I have not considered the role of polarized light in this Thesis.

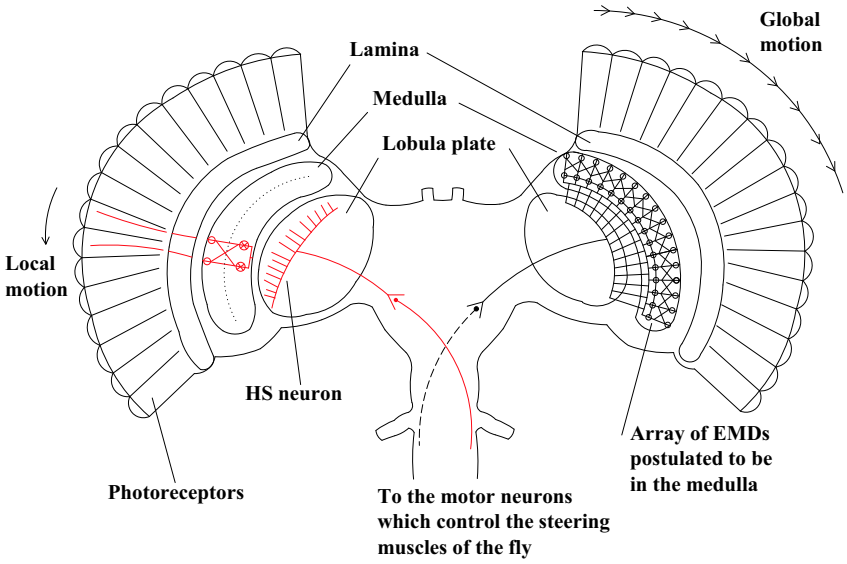

Figure 1.1. Visual system of insects. The visual system of insects showing the early visual pathways projecting from the photoreceptors to the visual neuropils namely the lamina, medulla, lobula and the lobula plate. The visual information is passed from the photoreceptors to the lamina and motion is detected by the elementary motion detectors (EMDs) in the medulla. The various motion detecting interneurons in the lobula plate (the red line in the figure shows the example of a HS neuron) passes on the information through the optic foci to the brain which then send the information through the descending interneurons to the motor neurons, that control the steering muscles thereby enabling flight control in insects.

1.3 Yaw response and flight control steering muscles

Optomotor yaw torque responses are selectively induced by horizontal motion [Wehrhahn, 1986; Hausen and Hengstenberg, 1987] and hence must be controlled by elementary motion detectors (EMDs) having sampling bases parallel to the horizontal axis of the eye lattice or by pairs of EMDs having symmetrically arranged sampling bases with respect to the axis, the outputs of which are added together. Monocular stimulation with progressive (front to back) motion leads to a simultaneous decrease of the ipsilateral and an increase of the contralateral wing beat amplitude and thus to the generation of yaw torque turning the animal in the same direction as the perceived motion. Regressive (back to front) motion is less effective but leads to syndirectional torque responses. Horizontal motion activates two motion sensitive systems behind each eye, which integrate the output of the arrays of the EMDs and that are specifically tuned to large field motion (global motion) and small field motion or object motion (local motion). The large field system is activated by ipsilateral front to back and contralateral back to front motion, and induces syndirectional yaw torque responses of the fly by simultaneous excitation and inhibition of the contralateral and ipsilateral flight motor respectively. The small field system is activated by the ipsilateral horizontal motion of small objects in both directions and induces turning towards the stimulus. The output channels of the small field and large field systems contain different frequency filters. The small field and large field systems dominate in torque-control under stimulation with high frequency and low frequency oscillatory motion respectively [Egelhaaf and Reichardt, 1987]. Since under natural conditions, the rotatory retinal motion patterns arise from self-rotations of the animal in space, HS system appears to act as a visual yaw-monitor, specifically designed for the control of course-stabilising yaw torque generation by the motor system [Eckert, 1981; Hausen and Egelhaaf, 1989].

The large field cells of the lobula plate exhibit synaptic interaction, which is of major significance for their functional role in the control of motion-induced behaviour. These interactions are based on the direct synaptic contacts between tangential cells of both lobula plates and lead to enhanced sensitivities of the cells to particular binocular motion stimuli. The HS cells are post synaptic to the H2 neuron of the contra-lateral lobula plate, which responds selectively to regressive motion in its receptive field [Eckert, 1981]. Due to this excitory synaptic input, the HS cells gain binocular sensitivity and respond selectively to horizontal rotatory motion. Since under natural conditions, the rotatory retinal motion patterns arise from self-rotations of the animal in space [Fermi and Reichardt,

1963; Heisenberg and Wolf, 1988], HS system appear to act as a visual yaw-monitor, specifically designed for the control of course-stabilising yaw torque generation by the motor system [Hausen and Egelhaaf, 1989].

The motion sensitive cells of the lobula plate supply input to the flight control system. The visual system, however, is not the only sensory modality that detects self-motion during flight. Mechanosensory fields at the base of the beating halteres detect forces that result when the fly rotates in space [Nalbach and Hengstenberg, 1994]. These gyroscopic sensors are more effective in detecting high speed body rotations and mediate the reflexive changes in the wing kinematics that stabilize the animal against mechanical perturbations during flight, while the slower rotations or drift during flight is more easily detected by the visual system. Thus these two sensory modalities encode different ranges of oscillation frequencies over which the animal operates. The halteres, like aerodynamically active forewings, are equipped with their own set of steering muscles, which receives strong excitory input from descending visual interneuron. Some haltere afferents form monosynaptic electrical synapses with steering muscle motor neurons. This control loop of visual feedback to haltere muscles, and haltere feedback to wing muscles allows the visual system to initiate responses in wing steering muscles indirectly or to control the gain of haltere mediated reflexes, thus enabling flight control in insects [Fyre and Dickinson, 2001; Hengstenberg, 1998].

1.4 Receptive field organisation of insect neurons

The three HS neurons (HSN, HSE and HSS) together occupy the whole retinotopic area of the lobula plate. Each of them covers roughly one third of the dorsoventral extent of the neuropil with some overlap. Correspondingly, their receptive field covers the dorsal, equatorial and ventral areas of the ipsilateral visual hemisphere. The dendrites of the HS neurons arborize in the anterior layers of the lobula plate. The HS neurons are excited by front to back motion and inhibited in the reverse direction. The dorsal (HSN) and equatorial neurons (HSE) are also excited by contralateral back to front movement [Hausen, 1982c]. The group of 11 VS neurons also occupies the whole retinotopic area of the lobula plate. Their dendritic fields are more strip like and oriented dorsoventrally [Hengstenberg et al., 1982]. The dendrites are stacked from distal to the proximal side of the lobula plate and overlap considerably. The VS neurons are excited by downward motion (preferred direction) and inhibited by upward motion (null direction) in the ipsilateral hemisphere. The receptive field organisation of the VS neurons strongly suggests that each of these

neurons specifically extracts the rotatory component of the optic flow around a particular horizontal axis. Both the HS and the VS neurons play a very significant role in the control of self-motion [Krapp and Hengstenberg, 1996]. The global structure of the response fields strongly suggests that different tangential neurons in the fly's visual system act as sensory filters, which extract specific components of self-motion from the momentary optic flow [Krapp et al., 1998; Franz and Krapp, 2000].

A male specific visual neuron called the male specific lobula giant, MLG1 has been identified as a key player in visual identification of potential mates [Gilbert and Strausfeld, 1991]. A recent study looking at responses to targets against naturalistic background motion actually found that blowfly HS neurons responded to target motion with a higher signal to noise ratio than MLG1 [Trischler et al., 2007]. When this was tested with the hoverfly HSN neuron, it is found that the male HSN neurons are used by hoverflies primarily as a fronto-dorsal target detector, at least where targets are moving against moving clutter. It has recently been argued that blowfly HSE and HSS are not primary yaw detectors as originally proposed, but their main role is in detection of translational optic flow [Karmeier et al., 2006; Kern et al., 2005]. Findings by Boedekker *et al* show that HS cells not only encode information about self rotation, but are also sensitive to translational optic flow and thus indirectly signal information about the depth structure of the environment [Boeddeker et al., 2005]. These response properties were only discovered after it became possible to replay optic flow to neurons that was generated by flies themselves in free flight. But recent findings by Nordstorm et al. (2007) are strongly in support that physiological properties of these neurons play a major role in yaw rotation detection and are inconsistent with a likely role in forward translation. It is also found that the receptive fields of male HSN are much more compact than the females and those described in blowflies and the sexual dimorphism is restricted to the most dorsal neuron and the small fronto-dorsal receptive field could have evolved to separate forward translation from yaw rotation as experienced during flight [Nordstorm et al., 2007].

1.5 Models of motion detection

The study of the insect visual system has offered a solution to the computational bottleneck faced by conventional video-based visual systems. The bottleneck arises because algorithms that are written to detect edges, estimate motion flow fields, find discontinuities in depth, motion, colour and intensity, have to be iterated several hundreds or even thousands of times before a solution is reached. Furthermore, the amount of computer

memory it takes and the computational cost is high. Consequently, the study of simple low-level motion cues, used by insects, has inspired many models of motion detection.

The relevance of the mechanisms and algorithms underlying biological motion detection extends to biology, computation and the noise community [Harmer and Abbott, Dec 2001]. Motion detection is an essential component of visual systems, which relates to many aspects of animal behaviour. The motion signals computed can be very useful as they can be used to encode a third dimension (motion parallax), to calculate time to collision, image segmentation, to obtain information about self-motion, driving eye movements, pattern vision and perception of real moving objects [Nakayama, 1985]. A number of algorithms for motion detection have been proposed in the biological and computer vision literature.

Exner (1894) was the first person to discuss the requirements necessary for generating a motion signal from neural circuitry. However the first mathematical and probably the most famous model of motion detection inspired by biological systems was developed by Hassenstein and Reichardt (1956) after a series of behavioural experiments examining the optomotor response of insects. This so called Reichardt or correlation motion detector possesses a highly parallel architecture. Each elementary motion detector (EMD) detects motion in a preferred direction by comparing a signal from one receptor with a delayed signal from the other receptor. The comparison is performed using a nonlinear, multiplicative, interaction between the two channels. Two EMDs tuned to opposite directions are combined to form a bidirectional motion detector.

While the Reichardt correlator model employed multiplicative interaction as an excitory mechanism, Barlow and Lewick (1965) pointed out that an inhibitory mechanism was also capable of providing directionally selective motion detection. They conducted experiments on rabbit retinal ganglion cells that were stimulated, not by a smooth motion but by a sequence of discrete illumination steps in two neighbouring locations, in either the preferred or null direction of the cell. Barlow and Levick found that the response to the null direction sequence was significantly reduced compared to the sum of the individual responses, whereas responses to the preferred direction sequence was roughly equal to the sum of individual responses. The authors proposed a veto mechanism or 'null-direction inhibition' as the basis for direction selectivity.

Alternative mechanisms of inhibition have also been proposed. One of the mechanisms known as *lateral inhibition* which was developed as a result of studies on horseshoe crabs, was proposed by Hartline and Ratliff as a result of experiments on the compound eye of

limulus [Ratliff and Hartline, 1974]. However, in order to be able to model directional selectivity for motion detection, nonlinear interactions are required. Hence a non-linear version of lateral inhibition, known as shunting inhibition was developed by Pinter (1984) after studies on the neurochemistry of visual cells in some detail. Shunting inhibition has been used to construct motion detection systems that are successful in describing the same behavioural phenomena as the Reichardt detector [Bouzerdoum and Pinter, 1993]. This phenomena also describes observations regarding the variation of the shape and size of the receptive fields with changing light intensity. The structure of the shunting inhibitory motion detector is the same as the EMD cell.

These 'delay and compare' models were derived from behavioural and physiological studies of the biological systems. A second class of systems that is also useful in understanding biological systems were derived to measure spatio-temporal energy characteristics of moving images. Some systems used Fourier domain descriptions of moving images [Watson and Ahumada, 1985; Fleet et al., 1996], while others considered time as another spatial dimension [Adelson and Bergen, 1985].

Watson and Ahumada (1985) designed a scalar motion sensor that exploits the Fourier description of motion. The motivation for the design of the sensor was physiological evidence suggesting the existence of spatial frequency tuned channels. The sensor responds to a sine wave grating moving in the preferred direction by producing an output of the same frequency. The output is zero if the grating is moving in the non-preferred direction. This property is provided by Hilbert filters [Daugman, 1988] in the quadrature path. The scalar motion sensor measures only the motion energy, not the velocity.

Fleet and Jepson (1985) also exploited the Fourier description of moving images using a hierarchical parallel processing scheme. A similar approach was taken by Adelson and Bergen (1985). They used quadrature pairs of spatio-temporal filters, whose outputs were squared and summed to give a measure of spatio-temporal energy. The starting point of Adelson and Bergen's analysis was not the Fourier description just mentioned, but instead considered time as another spatial dimension, so finding velocity amounted to determining a slope. Heeger (1987) has also proposed spatio-temporal energy systems based on three dimensional Gabor filters [Bracewell, 1999].

Another model, the gradient scheme of motion detection computes image velocity as the ratio of the local temporal gradient to the local spatial gradient [Horn and Schunck, 1981; Marr and Hildreth, 1980; Buchner, 1984; Borst, 2007]. Srinivasan modified the gradient model and came up with a generalized gradient scheme, which has the advantage

that it avoids the mathematical operation of division. Division is avoided by parsing the image into edges and standardizing the spatial gradient to a constant, finite value at each edge [Srinivasan, 1990; Srinivasan et al., 1999].

Several authors have proposed elaborations of the 'basic' Reichardt detector by either adding temporal, spatial or spatio-temporal filters before applying the input to the EMD cell [van Santen and Sperling, 1985b; Courellis and Marmarelis, 1990]. In one of the most recent elaborations of this model, Dror (1998), suggests that the inclusion of additional system components to perform pre-filtering, response compression, integration and adaptation to a basic Reichardt correlator can make it less sensitive to contrast and spatial structure thereby providing a more robust estimate of local image velocity. Dror found from his experiments that when the data obtained via intracellular recordings of the steady-state responses of wide field neurons in the hoverfly *Volucella* is compared to his model, the shape of the curves obtained agreed well with the theoretical predictions made by Dror (1998).

In 1990, after conducting many experimental studies on the behaviour of the insect visual system, Horridge proposed a simplified model of the insect visual system called the template model [Horridge, 1990; Sobey, 1990]. This empirical model compares the contrast between two adjacent receptors, at two sampling instances, to form a 2×2 template. These templates can give simple directional information. Operations are performed mainly between adjacent receptors, which make this model easy to implement in a parallel architecture. Several generations of insect vision chips based on the template model have been developed by Moini *et al* [Moini et al., 1996; Moini and Bouzerdoum, 1997; Moini et al., 1997]. When two different objects, converge together, crossover and then diverge again or when they move together at the same speed for a short time, while their paths are crossing over, it was found that the use of chrominance information can assist in tracking the object more accurately than with simple luminance information. Consequently, Chin *et al.* extended the Horridge template model to incorporate colour information. The resulting colour templates were found to detect more moving edges than the gray-scale model [Chin and Abbott, 1999].

In the software implementation of the template model, the array of on-chip photodetectors is replaced by a CMOS camera. The filtering operation in the lamina layer was implemented using temporal differentiation in the custom chips; whilst in our CMOS camera prototype this is performed using simple frame differencing. The CMOS camera is interfaced to a personal computer and the captured image is stored in the buffers. The

difference between the current image in the current frame buffer and the previous image in the previous frame buffer will give changes in the motion, which are thresholded and stored as templates. Template generation here is performed in software whereas some custom chips have template generation elements in the hardware. Thus by comparing the templates obtained from frame differencing with directionally sensitive motion templates, the direction of motion of the moving object can be obtained. These templates can be used to measure the velocity of moving objects. Experimental results show that a moving object or edge causes the same template to occur consistently at subsequent time steps. The angular velocity may be estimated by evaluating the ratio of the displacement of a motion sensitive template, to the time between the template's occurrences [Yakovleff et al., 1994].

The research in this Thesis mainly focusses on the Reichardt correlator model and the template model.

1.6 Velocity estimation in insects

Most of the spatiotemporal energy models—currently the dominant models for motion detection in vertebrates—are mathematically equivalent to correlator models[Adelson and Bergen, 1985]. Correlator models have been applied to explain motion detection in organisms as diverse as insects [Borst and Egelhaaf, 1989; Hausen and Egelhaaf, 1989; Srinivasan et al., 1999; Egelhaaf et al., 2003], humans [van Santen and Sperling, 1985b], birds [Wolf-Oberhollenzer and Kirschfeld, 1994], cats [Emerson et al., 1987] and mammals like wallabies [Ibbotson et al., 1994]. Though insects and humans appear to be capable of estimating image velocities [Srinivasan et al., 1996; McKee et al., 1986], the basic correlator model does not function as a velocity estimator. It reliably indicates directional motion of sinusoidal gratings, but the response depends on contrast (brightness) and spatial frequency (shape) as well as velocity [Egelhaaf et al., 1989].

Despite this problem, the work of Srinivasan and colleagues provides convincing evidence that honeybees integrate optical flow over time to compute distance travelled to a food reward [Lehrer et al., 1998]. The bees can locate the target reliably in subsequent visits even if the spatial content of the surrounding scene is altered. This contradicts the notion that the EMD is an unreliable velocity estimator with an output that depends on spatial structure of the scene and has led Srinivasan to propose that insects might use an alternative motion detector mechanism, a 'generalised gradient detector' [Srinivasan,

1990] that does not suffer from some of the Reichardt detector's problems. The rationale for considering these devices to be based on insect vision is primarily based on the behavioural evidence that bees can detect absolute velocity. But experiments conducted by Haag et al. (2004) demonstrate findings more in favour of motion detection circuits that employ a Reichardt correlator model than a gradient detection scheme. Hence the Reichardt correlator model seems to be the most efficient biological model as it emulates and explains most of the functionalities that are present in the biological systems.

The correlator is a nonlinear system and its response to a moving broad band image such as a natural scene, varies erratically as a function of time. In the absence of additional system components or assumptions, the raw output of a basic Reichardt correlator provides an inaccurate, ambiguous indication of image velocity. But it has been found that the use of various forms of spatial filtering, temporal filtering, saturation, integration and adaptation within the motion detection system may improve the performance of the correlator based system in response to many stimuli, including complex natural images. Experimental results on the fly as a model organism has been used to confirm these predictions [Dror et al., 2000]. The fly is chosen as a model organism here for both experiments as well as computational simulations due to the abundance of behavioural, anatomical and electrophysiological data available for its motion detection. It has also been found that the extended model of the Reichardt correlator works better with natural images, than with sinusoidal gratings, as natural images have predictable statistics and because biological systems are optimised to take advantage of these statistics [Dror et al., 2000]. Based on these studies, we further extend the Reichardt model to investigate and improve its performance as a velocity estimator.

The work in this thesis discusses the accuracy of velocity estimation using both the template model and the Reichardt correlator model suggesting various elaborations to both of these model to improve their performance.

1.7 Thesis overview

Biologically inspired motion detection models are bound to replace the conventional machine vision technology because of their simplicity and significant advantages in many applications. This chapter provides an overview of the history and characteristics of these models with particular emphasis to the Reichardt correlator model and the template model.

The template model and extension of the model to include colour information will be given in Chapter 2. Filtering techniques employed to remove noise in the model are also discussed.

Since a basic Reichardt correlator model does not give accurate estimate of velocity, additional components are added in an attempt to make the correlator model independent of spatial frequency and contrast. Chapter 3 reviews the research carried out by Dror in this area and the experimental verification of his elaborated model shows particular promise for the model to achieve velocity constancy as a function of contrast and spatial frequency.

The velocity estimation using the Reichardt correlator model is compared with the template model and the results of the experiments are presented in Chapter 4.

The ultimate goal of this research is to develop a simple and robust velocity estimator by elaborating the Reichardt correlator model. Chapter 5 discusses motion adaptation present in the fly visual system and the extension of the Dror's model to include adaptation. Some preliminary results obtained from the modelling of contrast feedback adaptation is also presented.

Velocity estimation by our elaborated model can be largely improved if we can decrease *pattern noise*, which is the dependance of the correlator response to structure of the visual stimulus. Chapter 6 explores the effect of pattern noise in insect based motion detection with various sampling methods discussed to reduce it.

Chapter 7 describes the effect of saturation on pattern noise and shows how the implementation of this compressive nonlinearity, at different positions in the model, has an impact on the the pattern noise.

This study is then carried out with different images and different speeds and the correlation of the biological data with modelling data is performed to understand the performance of the different models in Chapter 8.

Chapter 9 gives a conclusion based on the analysis and discussion on the different models and the future physiological and modelling work that is needed to model accurately the complex and still unknown non-linearities present in the visual system of insects.

1.8 Original contributions

The significant contributions described in this Thesis, to the field of bio-inspired engineering are as follows.

Velocity estimation using the template model is performed by extending the Horridge model to include various error checking methods by implementing both chrominance and luminance templates in an attempt to get more reliable results in Chapter 2. The use of colour templates here is a novel approach that offers more flexibility when using the various noise reduction techniques.

Then in Chapter 3 the performance of the template model is compared to the Reichardt correlator model and the physiological experiments carried out on the fly neuron using the same stimuli in each case. Furthermore it is found that Dror's elaborated Reichardt model performance is closer to that of the fly neuron than the template model. The original Reichardt correlator does not act as good velocity estimator because its response depends upon the stimulus contrast and the structure of the stimulus. So in order to reduce the dependance of the response to contrast, we have further extended the Reichardt correlator model to include contrast adaptation by a feedback adaptive mechanism and the results demonstrates much improved performance.

Deviation of the correlator response depending on the structure of the visual scene is termed as *pattern noise* in this Thesis. An important goal for achieving accurate velocity performance is to reduce the pattern noise. The effect of various receptive field shapes on pattern noise is studied and another prominent finding here is that circular sampled arrays reduce pattern noise, thereby improving velocity estimation performance. Based on these results we have also built a 16 pixel yaw sensor model, and its velocity performance reveals that such a small sensor is suited for many diverse applications such as in ultra miniature UAVs and other defence systems.

Another significant contribution, reported in this Thesis, is that it is found that implementation of compressive non-linearity or saturation in the model at different positions proved to change the shape of the pattern noise—indicating saturation has a major affect on pattern noise. Finally, several experimental studies are conducted with different models with saturation implemented at different stages of the model to verify the effect of saturation on pattern noise by comparing the modelling results with the physiological fly neuron data using the same stimuli.

These contributions serve to advance the goal of developing an accurate velocity estimation system using biologically inspired motion detection models. They provide improvements over the existing models and extend its applications to new and varied areas.

Chapter 2

Velocity Estimation Using The Template Model

INSECTS employ a very simple mechanism for motion detection. The template model proposed by Horridge is a biologically inspired motion detection model that permits simple tracking techniques and lends itself easily to both hardware and software implementation. Using the template model, Nguyen *et al* developed simple tracking algorithms, which are capable of efficiently tracking coherent movement of objects in a simple environment. Chin *et al* further extended the template model to include colour templates or chrominance templates, which have been successful in tracking objects moving in close proximity by separating objects of different colours.

In this chapter, the above algorithms are further improved and extended to implement various error checking methods to come up with a low cost, insect vision prototype, which is capable of detecting motion in noisy environment using a colour CMOS camera. Several error checking schemes are used before and after the formation of templates to reduce noise and randomness. This enables the detection of moving objects in real-time even under the influence of noise, thereby enabling many real life applications.

2.1 The Template Model

The template model proposed by Horridge [Horridge, 1990; Sobey, 1990] models the function of the small field motion detection neurons in the medulla. In the template model the temporal contrast of the adjacent cells at consecutive instances are used to determine the direction of motion of an object. For simplicity, only temporal contrast at two instants and from two neighbouring cells is considered.

The visual field is sampled spatially. Each sampling channel detects changes in light intensity at two consecutive sampling instances to show either an increase (↑), a decrease (↓) or a no change (–). Adjacent photoreceptors carry out the detection of directional motion in insects—thus by copying this mechanism, the changed states are spatially combined between adjacent channels. Hence for a pair of adjacent sampling channels, there are 9 possible combinations:

(– –), (↑ –), (↑ ↑), (↑ ↓), (↓ –), (↓ ↓), (– ↓), (↓ ↑), (– ↑).

Since in any visual system directional motion is accomplished as a spatio-temporal operation, the temporal domain is included by associating the combinations obtained at two consecutive sampling times t_0 and t_1, thus yielding 81 spatio-temporal combinations or templates. Thus the temporal contrast of two neighbouring cells, at two sampling instances, are combined to give simple motion measures called 'templates.'

Out of these 81 templates, there are only 8 templates, which indicate coherent motion. The templates sensitive to coherent motion are the ones in which one of the four entries is no change in intensity (–) and the other three must identically either increase (↑) or decrease (↓) in intensity. These templates are known as directionally motion sensitive templates (DMSTs) [Yakovleff and Moini, 1997] as shown in Figure 2.1. These templates possess a 3:1 diagonally symmetric structure. The DMSTs are named using the first eight alphabetic characters, from 'A' to 'H'. The templates having one intensity change entry and three no change entries are purely generated by motion—they occur at the concave corners of the staircase motion trajectory[Nguyen et al., 1993]. In other words, they occur simultaneously in pairs with the DMSTs at each sampling instance. They indicate the current position of the moving object and hence are referred to as position conjugate templates (PCTs). The PCTs indicate the location of their DMST counterparts, which occur at the next sampling instant. The pair of DMSTs and corresponding PCTs are referred to as *conjugate pairs of templates*. The PCTs are labelled using numeric characters from '1' to '8'.

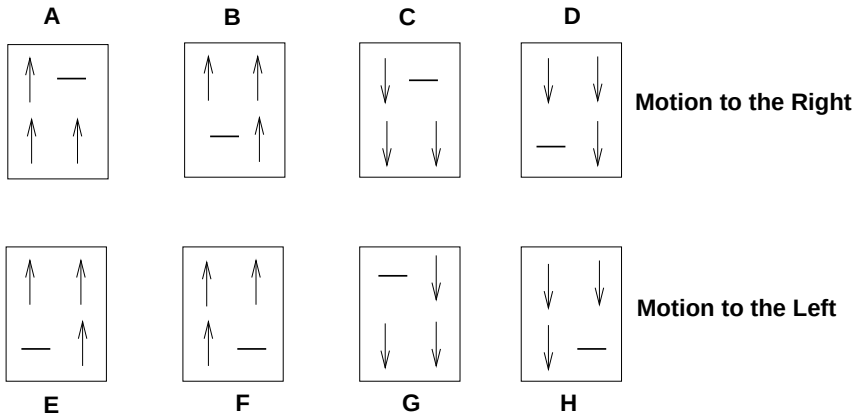

Fig. 2.1. Directionally motion sensitive templates. The DMSTs 'A', 'B', 'C' and 'D' indicate motion to the right and the DMSTs 'E', 'F', 'G' and 'H' indicate motion to the left. This assumes the dark object in front of the light background—for inverse case, simply reverse the arrows. All the DMSTs have a 3:1 diagonally symmetric structure.

The template model is easily implemented in VLSI, because the thresholding and template formation operations are simple and can be carried out in parallel. The presence of only three states reduces the bandwidth requirements compared to other vision systems.

2.2 Prototyping Platform

A series of so-called 'bugeye chips' [Moini et al., 1996; Moini and Bouzerdoum, 1997; Moini et al., 1997] based on the template model has been implemented using grayscale photodetectors. In our previous custom insect vision chips, the retina was modeled using an array of on-chip photodetectors with dynamic range control circuitry. This is replaced by a standard CMOS camera in our present prototype. It should be noted that the CMOS camera does not have Automatic Gain Control (AGC) circuitry at the pixel level, but this feature is not essential for indoor controlled lighting conditions. As our experiments in this Thesis focus on controlled lighting conditions, this simplification is justified.

The filtering operation in the lamina layer was implemented using temporal differentiation in the custom chips; whilst in our CMOS camera prototype this is performed using simple frame differencing. The CMOS camera is interfaced to a personal computer and the captured image is stored in the buffers. The difference between the current image in the current frame buffer and the previous image in the previous frame buffer gives changes in the motion, which are thresholded and stored as templates. Template generation here is performed in software, whereas some custom chips have template generation elements in

the hardware. Thus by comparing the templates obtained from frame differencing with directionally sensitive motion templates, the direction of motion of the moving object can be obtained. The time step size between the frames or the frame interval can be adjusted. But it is found that by keeping the frame interval low, more frames can be captured giving us more template information. When the frame interval is high, frames and information are lost—this is a limitation of the specific hardware and not fundamental problem. In this case, a low frame interval of 60 ms is chosen as it was found optimal for our specific hardware platform.

2.3 Colour Templates

Since the CMOS camera uses the RGB colour scheme, in order to increase the flexibility of implementing noise reduction techniques, the colour components are used to form chrominance templates in the model even though such colour templates do not emulate the insect visual system. This is an engineering extension to help improve the performance of the model. The RGB colour scheme works on the principle that each colour is comprised of different combinations of the three primary colours red, green and blue. The RGB scheme can be converted into the luminance-chrominance model using the formulae below:

$$Y = 0.299R + 0.587G + 0.114B \tag{2.1}$$

$$C_{\mathrm{r}} = R \text{ - } Y$$

$$C_{\mathrm{b}} = B \text{ - } Y$$

$$C_{\mathrm{g}} = G \text{ - } Y,$$

where, Y is called the Luminance, C_{r} is the Red Chrominance, C_{b} is the Blue Chrominance and C_{g} is the Green Chrominance. Also, R is the Red, G is the Green and B is the Blue signal level of an RGB pixel.

The luminance and the chrominance values of each pixel or channel are put through the same differentiation, thresholding and template generation stages as the original grayscale templates. Through the chrominance values, templates can be formed. The same eight directionally motion sensitive elements can be detected for the luminance and each of

the three-colour chrominance parameters. The hypothesis is that some components of noise may be uncorrelated between colour bands and some noise may be correlated, thus the use of the three chrominance levels offers the flexibility for implementing various noise reduction techniques. For example, temporal averaging reduces uncorrelated noise whereas differencing reduces correlated noise.

2.4 Pre-Template Filtering

The signal captured by the CMOS camera is very noisy. Noise does not just come from the environment, but it also originates from the camera when the signal is converted from analog mode to digital mode. Here two algorithms are implemented to pre-filter noise before the signal is passed to the template formation stage. The first algorithm, using the spatial averaging method, is a very simple algorithm that helps to smooth the response and increase the sensitivity of the program without compromising speed. The second algorithm is Multiplicative Noise Cancelation (MNC), which was introduced by Moini et al. (1997).

2.4.1 Spatial Averaging

Spatial smoothing of the input signal is one of the filtering operations carried out by many front-end processing stages for reducing the input noise. Input signal captured by the CMOS camera is stored in a buffer. Before this data is passed to the differentiation stage the values stored in the buffer are then averaged. Spatial averaging is performed by taking the average over a certain number of channels as shown in Figure 2.2. The resulting values are then stored in the original buffer replacing the previous values.

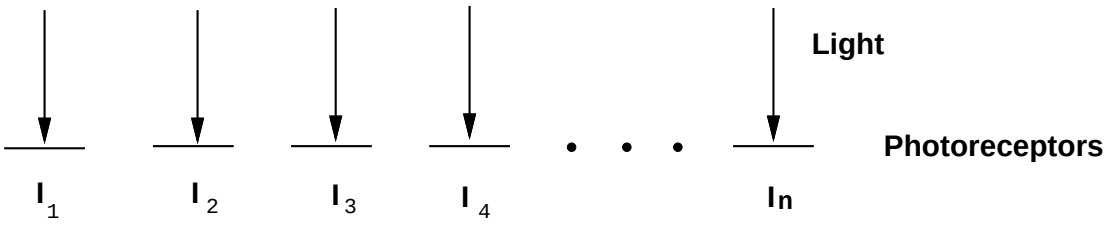

Fig. 2.2. Nomenclature for photoreceptor channel input intensities.

If I_1, I_2, I_3 are the input intensities, then the average $\bar{I}$ over three channels is

$$\bar{I} = \frac{I_1 + I_2 + I_3}{3}.$$

Averaging helps smoothing the edges of the signal hence improving the template formation. Furthermore averaging also helps in detection over a given range of channels instead of just between two adjacent channels. Here, in this software, the spatial average over 3 channels is implemented. Averaging over 5 channels or more is not efficient because of the blurring effect of the camera when the object is moving fast.

2.4.2 Multiplicative Noise Cancelation

Multiplicative Noise Cancellation is an extension of spatial averaging, which also is applied to the input signal before the differentiation stage. Experimental results indicated that the 100 or 120 Hz frequency component of light sources, operating with 50 or 60 Hz mains power, severely affect the detection of motion. In this stage noise is sometimes amplified to a point where it dominates the signals induced by the motion of objects. As this noise is usually common among several neighbouring photodetectors and has a multiplicative nature, spatial averaging is ineffective, but dividing the signal in one channel by the spatial average of the input signals effectively reduces the noise [Moini et al., 1995; Moini et al., 1993]. The MNC consists of spatial averaging, and division-by-average stages. By doing the division by average calculation, the illuminance values that are given by E, as illustrated in Figure 2.3, get cancelled thus reducing the multiplicative noise caused by the incident light.

Applying MNC to photodetector 2 in Figure 2.3, where ρ is the reflectance, the normalised signal Γ_2 is

$$\Gamma_2 = \frac{L_2}{L_1 + L_2 + L_3}$$

$$= \frac{\rho_2 E}{\rho_1 E + \rho_2 E + \rho_3 E}$$

$$= \frac{\rho_2}{\rho_1 + \rho_2 + \rho_3}$$

, which is directly proportional to contrast ratio. The key point here is that the cancellation of the E terms removes the correlated noise due to the 100 or 120 Hz flicker present in indoor lighting conditions. Edge enhancement also results as the denominator reduces at the boundary of a light region [Moini et al., 1995]. Another advantage is that the above expression results in numbers less than one, and thus data compression is also achieved.

MNC has been successfully implemented in our previous custom chips. In the present experiment, MNC is replicated using the CMOS camera to improve the noise performance of the system. Our software was designed such that the averages over 3, 5 or 7 channels can be selected.

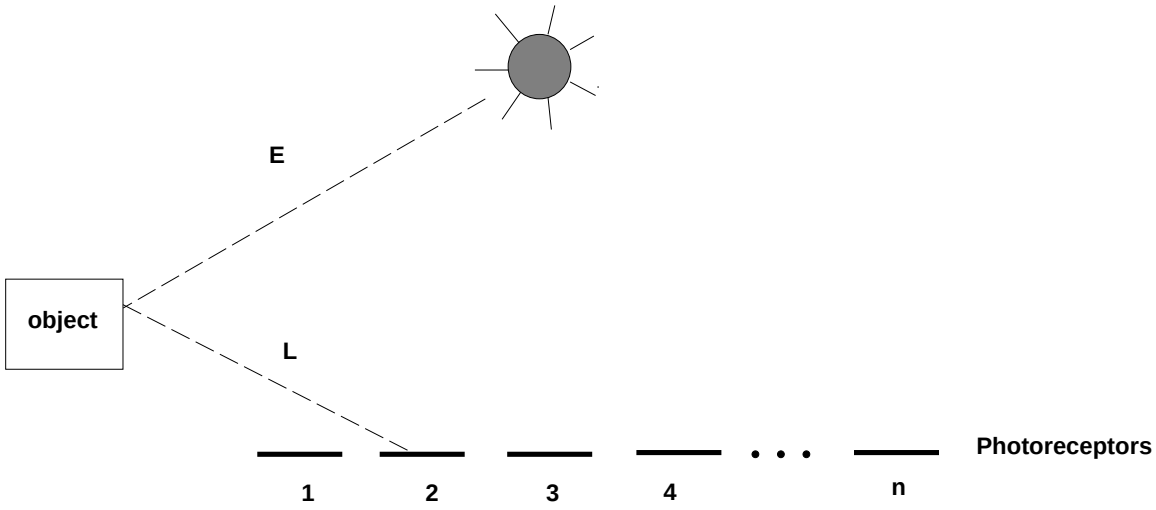

Fig. 2.3. Diagram illustrating the source of the multiplicative noise. For an opaque object with a *Lambertian* surface, the luminance L is E (illuminance) times the wavelength dependent coefficient of reflectance (ρ). If 100 Hz or 120 Hz ac noise modulates the incident illuminance, the MNC method successfully cancels out the E terms thereby removing the multiplicative ac noise.

2.4.3 Experimental Results

In this experiment, a test sequence is generated by moving a dark object at a close range in front of the camera. The system then captures the motion of the object at a frame interval of 60 ms. In Figure 2.4the result of the custom chips is replicated in the grayscale domain without applying any pre-filtering techniques. Spatial averaging and MNC algorithms is then applied to the same test sequence. Spatial averaging results are shown in Figure 2.5, and MNC results are shown in Figure 2.6.

The results show that the spatial averaging algorithm provides a better result than the one obtained without using any de-noising technique. The MNC technique is also found to be successful in removing the unwanted ac noise. However the MNC technique seemed to introduce some noise of its own into the system. This is due to the fact that the division

operation using digital computer is not optimised to have the level of precision needed. The errors that are created in this operation are carried on to the template formation stage and these may later be filtered out by the post-template filtering methods.

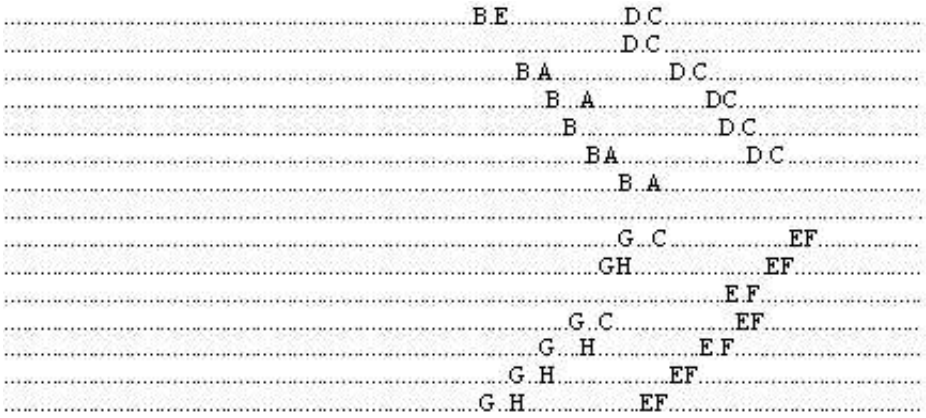

Fig. 2.4. Example templates with no pre-filtering. Grayscale templates for small object moving across our prototype camera (See Section 2.7) at close range. Note that only 8 directionally motion sensitive templates ('A' to 'G') have been included in this result. The other 73 templates were represented by '.' which represents 'no motion'. The horizontal axis represents the angular position of the templates and the vertical axis represents the time in steps of 60 ms. The templates BA and DC indicate motion to the right and the templates GH and EF indicate motion to the left.

2.5 Post Template Filtering

Though the pre-filtering technique filters much of the noise, it does not remove all the noise that is present. These remaining errors are carried over during the template formation and the post-filtering methods aim at removing them. Two algorithms for filtering out noise from the data, provided in real-time by the template formation stage are considered. Both algorithms work based on the principle that certain templates occur together in coherent motion, but differ in that the first is based on the continuity of the motion, while the second uses the association of certain template pairs to detect noise.

2.5.1 Windowing Operation

In coherent motion, where the object is moving with constant speed, a moving edge causes directionally sensitive templates to occur close together in space and time. Experimental

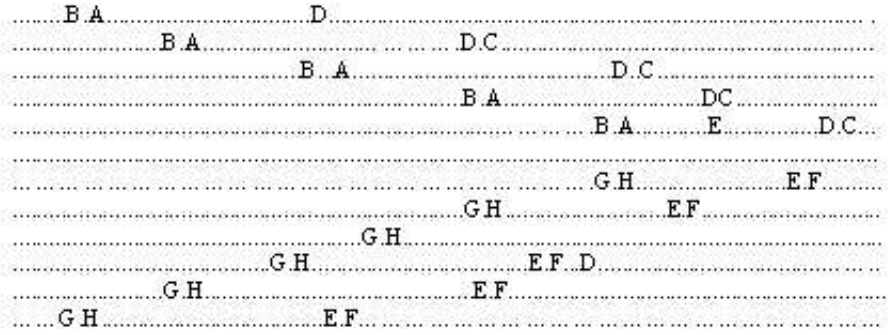

Fig. 2.5. Example templates with spatial averaging. Templates of an object moving at close range applying spatial averaging to the signal. Here average over three channels is chosen. Note that only 8 directionally motion sensitive templates ('A' to 'G') have been included in this result. The other 73 templates were represented by '.' which represents 'no motion'. The horizontal axis represents the angular position of the templates and the vertical axis represents the time in steps of 60 ms. The templates BA and DC indicate motion to the right and the templates GH and EF indicate motion to the left.

and theoretical results [Nguyen et al., 1993] have proven that motion sensitive templates occur in groups. Further experiments using our camera show that motion templates occur

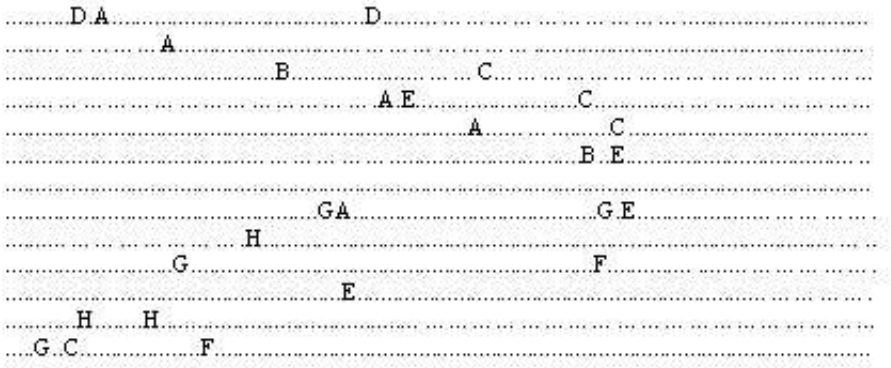

Fig. 2.6. Example templates with MNC pre-filtering. Templates of an object moving at close range applying MNC to the signal. Here average over 5 channels is chosen and used. Note that only 8 directionally motion sensitive templates ('A' to 'G') have been included in this result. The other 73 templates were represented by '.' that represents 'no motion'. The horizontal axis represents the angular position of the templates and the vertical axis represents the time in steps of 60 ms. The templates BA and DC indicate motion to the right and the templates GH and EF indicate motion to the left.

in a group of at least 2 templates. This result indicates that isolated templates are likely to be noise.

To utilise this observation, a window operation is applied to the templates in both time and spatial dimensions. Each 3×3 window of templates (3 in the time dimension and 3 in the spatial dimension) is examined. If there are less than 2 templates indicating motion of one kind (either to the left or to the right) in that window then these templates are considered as noise and are discarded.

The window size might vary depending on the speed of the object. Faster objects that can jump passed several pixels in one period of time might require a bigger window size. Our program was implemented such that different window sizes can be chosen externally.

2.5.2 Template Pairs

Further results from Nguyen (1995) demonstrated that motion templates occur in pairs. There are two kinds of template pairs. The first one is the *motion conjugate* pair, where DMSTs and PCTs occur together in any coherent motion. Furthermore, theoretical results from Nguyen et al. (1993) have shown that a moving edge always generates 2 associated DMSTs: (A, B) and (C, D). These pairs are associated because of their nature of one corresponding to the edge entering a region and the other corresponding to the edge leaving that region. Our template-pair algorithm is based on these results. In this algorithm, each row of templates is examined to look for and locate the template pairs. If a template occurs at one instant, but its partner template (DMST or PCT) does not occur at that same instant, then this template is treated as noise and is discarded.

2.5.3 Experimental Results

Several experiments are carried out to evaluate the effectiveness of the two algorithms. The results from the windowing algorithm and the template pair algorithm explained in the Sections 2.5.1 and 2.5.2 are shown in Figure 2.7 and Figure 2.8 respectively.

In terms of performance, when you see Figure 2.7 and Figure 2.8 and from the results seen in the next section, the windowing algorithm produces less consistent results compared to the template-pair algorithm. The windowing algorithm seems to leave noise in the result wherever the noisy templates are surrounded by other templates. The conjugate-pair approach does not have this problem because it does not just look at the templates locally

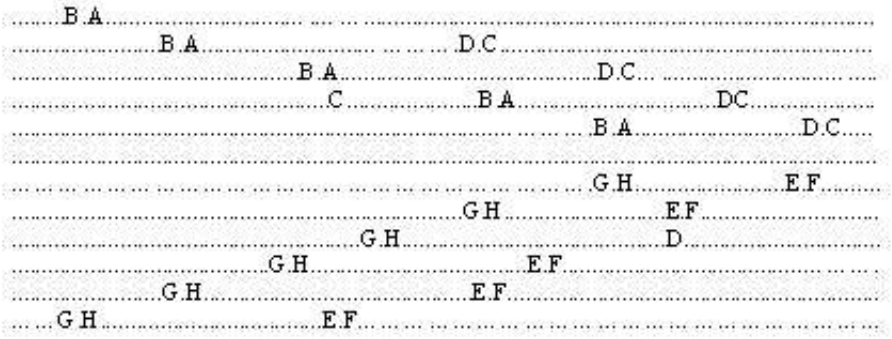

Fig. 2.7. Applying the windowing algorithm to the templates. The windowing technique is based on the principal that a moving edge causes templates to occur close together in space and time. Experimental results show that motion templates occur in a group of at least 2 templates. In this technique, windows of size 3×3 of templates is examined. If there are less than two templates indicating motion in one direction in that window, then it is considered as noise and discarded.

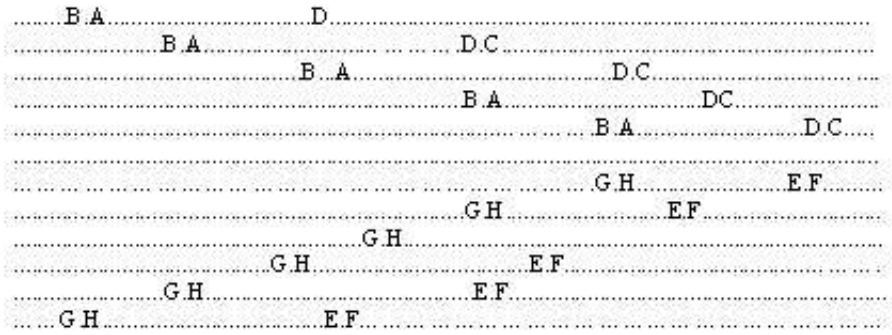

Fig. 2.8. Applying the template pair algorithm to the output. This technique is based on the fact that motion templates always occur in pairs [Nguyen, 1995]. A moving edge always generates 2 associated DMSTs: (A, B) and (C, D). These pairs are associated because of their nature of one corresponding to the edge entering a region and the other corresponding to the edge leaving that region. Our template-pair algorithm is based on these results. In this algorithm, each row of templates is examined to find the template pairs. If a template occurs at one instant, but its partner template (DMST or PCT) does not occur at that same instant, then this template is treated as noise and is discarded.

but looks at templates in a whole row. However, for fast moving objects, where blurring [Yakovleff et al., 1994] is introduced, template pairs normally do not occur correctly.

In these cases the windowing algorithm, with appropriate size windows, provides better performance.

2.6 De-noising the Colour Templates

Colour templates were first introduced by Chin and Abbott (1999). In this experiment, the de-noising techniques are applied to colour templates. Here, colour templates were created, by extracting colour information from the CMOS camera. In order to reduce the processor load, here we only implement blue and red templates (we obtain red and blue signals directly from the camera). However using the formula given in Eq. 2.1, in Section 5, we can easily create green templates. However this is ignored in terms of noise removal as the green signal is not independent of red and blue. Figure 2.9 shows colour templates without any de-noising operations

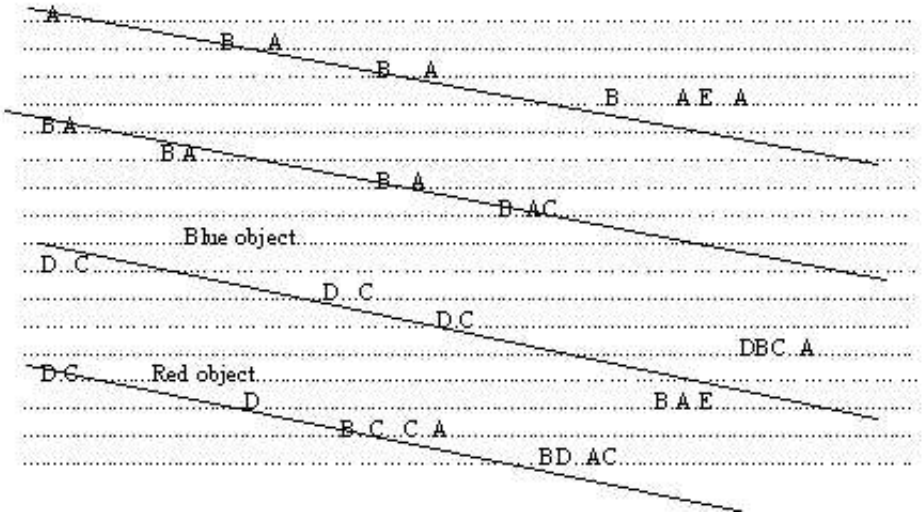

Fig. 2.9. Colour templates before denoising. Colour templates from the multi object test sequence from our prototype. This sequence contains 1 red and 1 blue object. The objects are moving across the camera next to each other with relatively the same speed. Here, colour templates were created, by extracting colour information from the CMOS camera. In order to reduce the processor load, here only blue and red templates (as we can obtain red and blue signals straight from the camera) are used. However using the formula given in Eq. 2.1, in Section 5, green templates can be easily created.

Applying spatial averaging over 3 channels and conjugate pair algorithms explained in the Sections 2.4.1 and 2.5.2 to the above result helped get rid of most of the noise as shown in Figure 2.10.

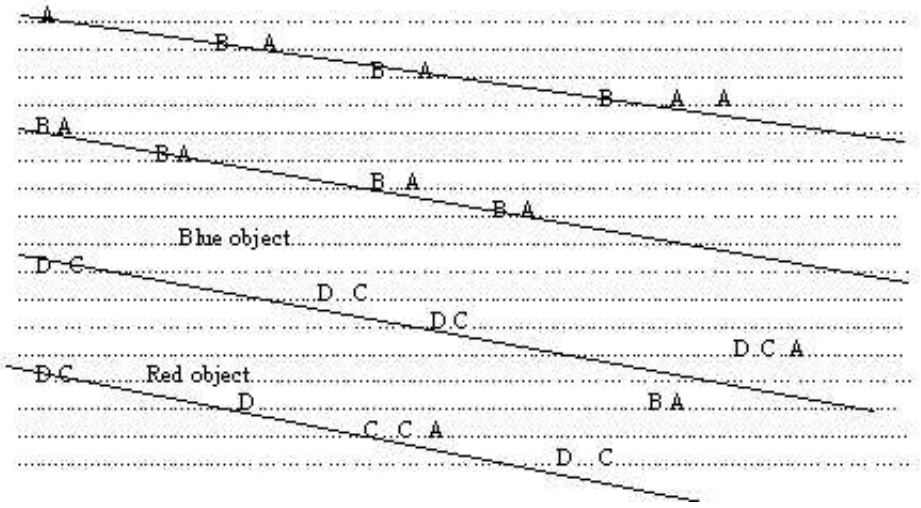

Fig. 2.10. Colour templates after de-noising.. Colour templates from the multi object test sequence from our prototype. This sequence contains 1 red and 1 blue object. The objects are moving across the camera next to each other with relatively the same speed. This figure shows the colour templates after applying the spatial averaging and conjugate pair techniques, explained in the Sections 2.4.1 and 2.5.2 to Figure 2.9. It is seen that these filters help to reduce noise.

2.7 Measuring Angular Velocity

An experiment is set up to test the effectiveness of different de-noising algorithms. In this experiment, the camera was placed in the center of a white hollow cylinder with a vertical black paper bar inside. The cylinder is motor controlled and the angular speed of the cylinder can be adjusted by changing the voltage supply to the motor. Our program then tried to measure the angular speed of the cylinder by detecting the motion of the dark paper. The black paper is used to measure the angular velocity of the luminance templates. The experiment is repeated with red and blue paper stripes to measure the angular velocities of the red and blue chrominance templates respectively.

Experimental results show that a moving object (or edge) consistently causes the same motion sensitive template to occur at subsequent time steps, and at positions corresponding to the displacement of the edge relative to the detector [Yakovleff et al., 1994]. The angular velocity may be estimated by evaluating the ratio of the displacement of a motion sensitive template, to the time between the template's occurrences (i.e., in Figure 2.11, the angular velocity is angular displacement/ΔT).

NOTE:
This figure is included on page 30
of the print copy of the thesis held in
the University of Adelaide Library.

Fig. 2.11. Template response (dark object in front of light background). The response of the template model to the motion of a dark object on a lighter background is illustrated this figure. In between sampling instants, templates are read sequentially from the save memory and printed on a single line as hexadecimal numbers (except for the no motion template, which is printed as a dot). The vertical and horizontal axes therefore represent sampling instants and angular position respectively, and hence the response is shown as a spatio-temporal image. The template encoding is arbitrary. Thus, templates 7 and E, for instance, only occur when the motion is to the left and there is a bright to dark contrast change, while templates 3 and B indicate motion to the right. The angular velocity may be estimated by evaluating the ratio of the displacement of a motion sensitive template, to the time between the templates occurrences (i.e., as shown in this figure, the angular velocity is angular displacement/ΔT. After Yakovleff et al. (1995).

There are two algorithms for estimating velocity in real-time that have been developed and tested. The first algorithm is forward tracking [Yakovleff et al., 1995], and the second algorithm is stair-step tracking [Nguyen et al., 1993]. In our experiment, we used the forward tracking algorithm for velocity measurement. In the forward tracking algorithm, we kept track of certain motion templates within a fixed time "window" of previous (small) displacements, and of the time steps at which they occurred, the velocity is provided by the ratio of the sum of the displacements, to the size of the window.

Results from our experiments are shown in Figures A.1 to 2.20. The rotating speed was increased from 10 rpm to 45 rpm in steps of 5 rpm. The horizontal axis represents real speed in rpm (varying from 10 rpm to 45 rpm), which was measured by using a tachometer. The vertical axis represents speed in rpm (varying from 10rpm to 45 rpm) measured using our program.

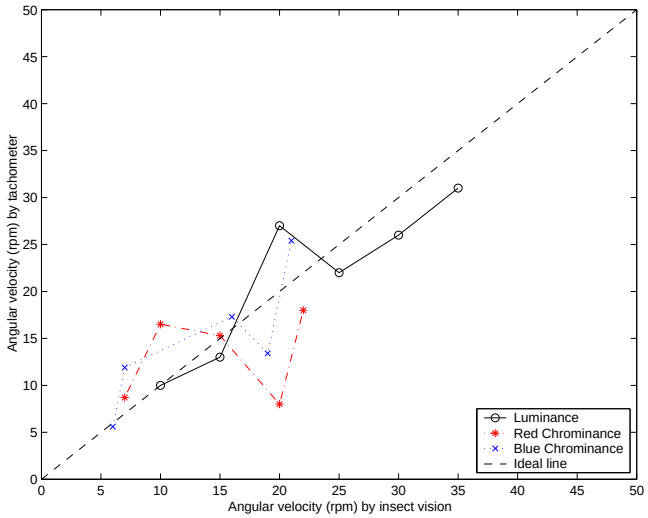

Fig. 2.12. Comparison of angular velocity curves with no pre-filtering. Benchmark angular velocity measured by a tachometer versus angular velocity determined by insect vision system. Case 1: Without using any noise removal algorithms. Frame rate was 60 ms.

Figure A.1 shows that without applying any filtering process the response scatters randomly around the ideal response. While results from Figure A.3 show that the averaging process helps in smoothing the response, the response still scatters around the ideal response but is much smoother. Figure 2.3 indicates that the measured speeds are shifting up from the ideal response. This can be explained by the calculation errors (division) of the digital computer. Figure A.13 indicates that applying the windowing operation to the templates helps in eliminating the constantly varying nature of the response. This is due

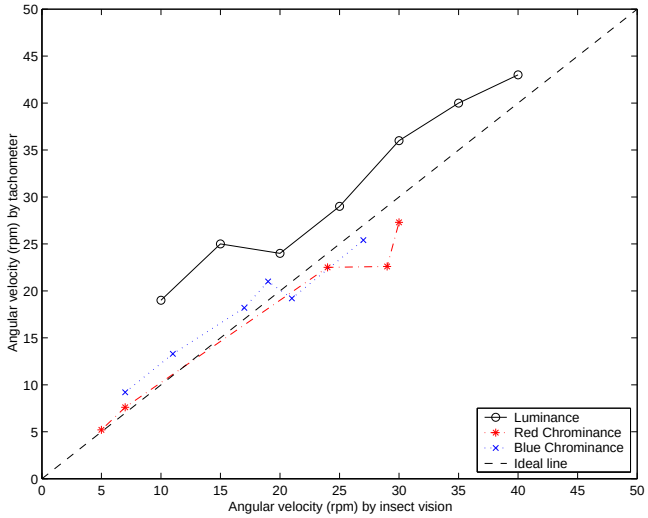

Fig. 2.13. Comparison of angular velocity curves with spatial averaging. Benchmark angular velocity measured by a tachometer versus angular velocity determined by insect vision system. Case 2: Using the spatial averaging algorithm. Frame rate was 60 ms.

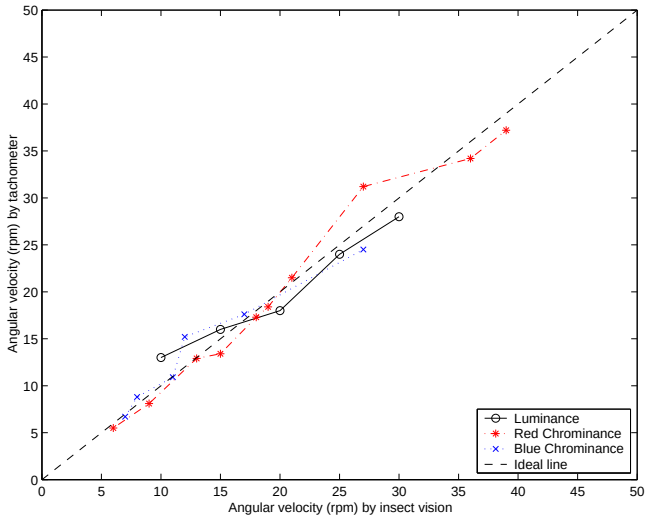

Fig. 2.14. Comparison of angular velocity curves with MNC filtering. Benchmark angular velocity measured by a tachometer versus angular velocity determined by insect vision system. Case 3: Using the MNC method. Frame rate was 60 ms.

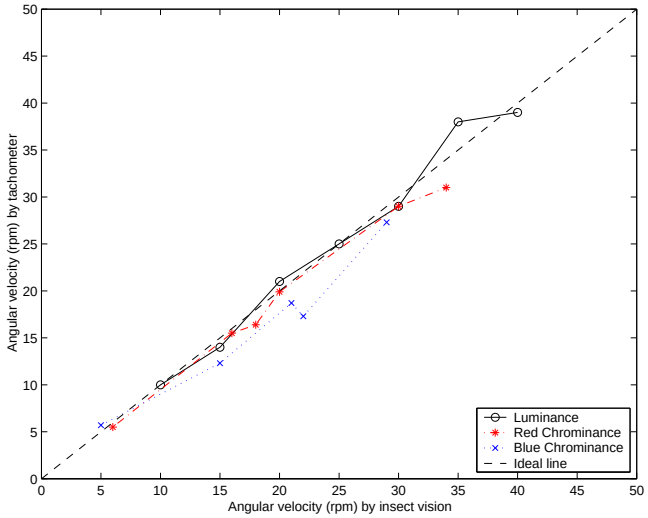

Fig. 2.15. Comparison of angular velocity curves with template pair filtering. Benchmark angular velocity measured by a tachometer versus angular velocity determined by insect vision system. Case 4: Using the template pair method. Frame rate was 60 ms.

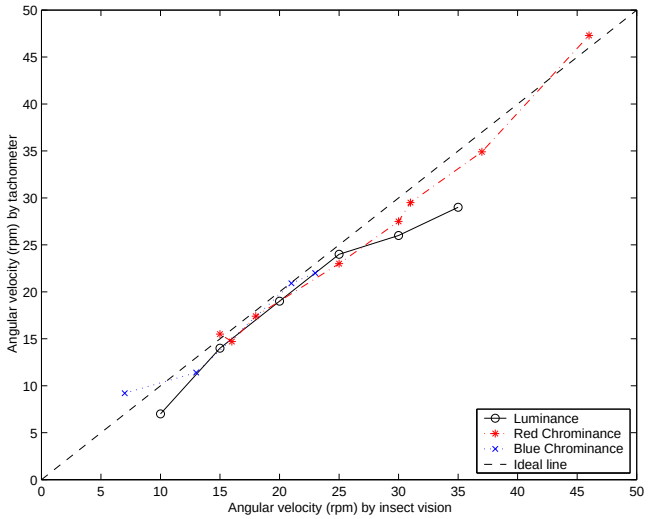

Fig. 2.16. Comparison of angular velocity curves with windowing method. Benchmark angular velocity measured by a tachometer versus angular velocity determined by insect vision system. Case 5: Using the windowing method. Frame rate was 60 ms.

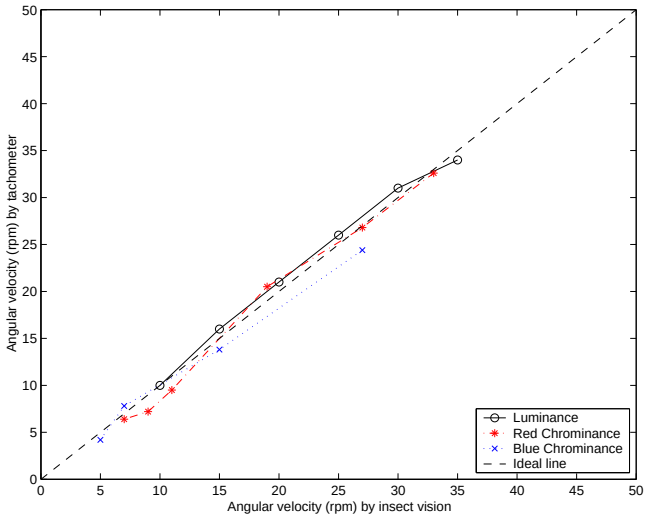

Fig. 2.17. Comparison of angular velocity curves with averaging and template pair method. Benchmark angular velocity measured by a tachometer versus angular velocity determined by insect vision system. Case 6: Using the averaging and template pair technique. Frame rate was 60 ms.

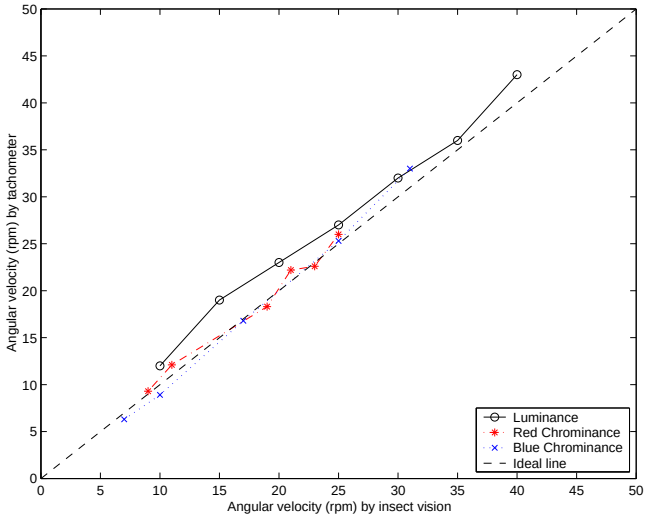

Fig. 2.18. Comparison of angular velocity curves with averaging and windowing method. Benchmark angular velocity measured by a tachometer versus angular velocity determined by insect vision system. Case 7: Using the averaging and windowing technique. Frame rate was 60 ms.

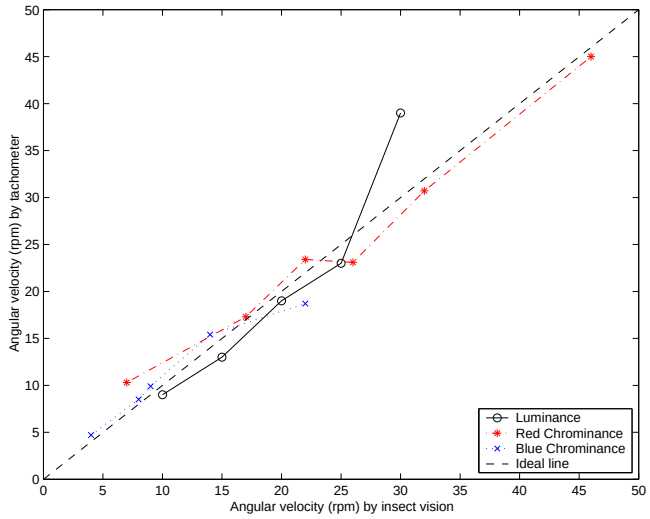

Fig. 2.19. Comparison of angular velocity curves with MNC and template pair technique. Benchmark angular velocity measured by a tachometer versus angular velocity determined by insect vision system. Case 8: Using the MNC and template pair technique. Frame rate was 60 ms.

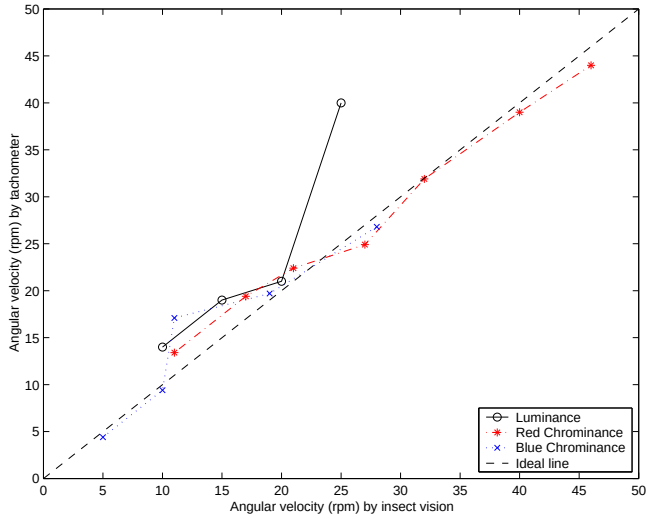

Fig. 2.20. Comparison of angular velocity curves with MNC and windowing method. Benchmark angular velocity measured by a tachometer versus angular velocity determined by insect vision system. Case 9: Using the MNC and windowing technique. Frame rate was 60 ms.

to the local nature of the windowing algorithm. The template-pairs algorithm produces similar performance to the windowing algorithm, but the result is much more consistent.

Figure A.9, Figure 2.18, Figure A.11 and Figure 2.20 represent responses with different combinations of pre-template filtering and post-template filtering. These figures indicate significant improvement in performance compared to using each individual technique or not using any technique at all. This result demonstrates that the pre-filtering techniques and the post filtering techniques do actually combine well in eliminating noise. From these experiments it can be seen that combining averaging, during pre-template filtering, and conjugate-pair techniques, during post-template filtering give the best overall response [Nguyen et al., 2001]. However many more experiments with different conditions need to be carried out before such general conclusion can be reached.

The above experiments are repeated using the stimuli generated by the Vision Egg software and the results are compared with the rotating drum experiments and are presented in the Appendix A. It can bee seen that the Vision egg stimuli give a much better result than the rotating drum as the inconsistencies due to non-uniform speed control using the rotating drum is not present with the Vision Egg experiments (See Appendix A).

2.8 Conclusions

A brief introduction to the fundamental concepts of template model and how they apply to the real time motion detection is provided. Some de-noising techniques were introduced before and after the template formation stage. Several experiments have been carried out and preliminary results from these experiments indicate certain improvement to the template model. Furthermore it also demonstrates that combining pre-template filtering and post-template filtering does provide improved performance as one can expect.

In order to understand more about the template model, the performance of the template model is compared with the predominant model of motion detection, namely, the Reichardt correlator model. So in the next chapter, a study about another prominent model of motion detection, namely, the Hassenstein and Reichardt correlator model is conducted. Since a basic Reichardt correlator model does not give accurate estimate of velocity, additional components are added in an attempt to make the correlator model independent of spatial frequency and contrast. And then in Chapter 4, the velocity performance of the template model is compared to both the Reichardt correlator model and

physiological experiments carried out on the fly neuron using the same stimuli in each case.

Chapter 3

Reichardt Correlator Model

ALTHOUGH motion processing in insects has been extensively studied for over almost 40 years, velocity detection in insects and how the insect brain computes the velocity of a moving feature independent of its size or contrast is a major enigma, which remains unsolved. The dominant model for insect motion detection, first proposed by Hassenstein and Reichardt has gained wide-spread acceptance in the invertebrate vision community. While most of the previous works relate that Reichardt correlators do not signal true velocity, this review chapter examines the accuracy with which the realistic Reichardt correlators can provide velocity estimates in an organism's natural visual environment by clearly illustrating it with experimental recordings, confirming the predictions made. Analysis and simulations suggest that the processes commonly found in visual systems, such as pre-filtering, response compression, integration, and adaptation, improve the reliability of velocity estimation and expand the range of velocities coded. To prove this experimentally, intracellular recordings of steady-state responses of wide field neurons in the hoverfly *Volucella* to the motion of broadband images are used. With the data obtained the mean response level of wide field neuron is plotted as a function of velocity. The shapes of these curves and their dependence on image statistics are found to agree with the theoretical predictions. Thus by applying certain motion adaptation techniques, the elaborated model of the Reichardt correlator can be made less sensitive to contrast and spatial structure of natural scenes thereby providing a more robust estimate of local image velocity under natural operating conditions.

3.1 A Simple Reichardt Correlator

Both behavioural and physiological experiments on the insect visual system indicate that movement detection occurs between neighbouring points of the sampling lattice of the eye. This implies that a motion detector has two input channels and the two channels should possess an asymmetric arrangement. Asymmetry is necessary for the detector to acquire direction selectivity. Furthermore the interaction between two channels must be nonlinear if the detector is to respond selectively to moving gratings. The asymmetrically nonlinear interaction between two input channels is termed the elementary motion detector (EMD) [Buchner, 1976]. Such an EMD will elicit a strong response when a visual stimulus moves in a specified direction (the preferred direction) and weak response when the stimulus moves in the opposite direction [Borst and Egelhaaf, 1989; Srinivasan et al., 1999].

A basic Reichardt correlator is formed by combining two EMDs that are tuned to opposite direction, to indicate bidirectional motion. The nonlinear interaction is assumed to be a multiplication, M, which is the simplest possible nonlinear interaction. The asymmetry is implemented by using a delayed unit, D, which is a first order low pass filter. This filter acts as a delayed element by introducing phase shift of up to 90°.

Figure 3.1 shows a simplified version of the Reichardt correlator model. Receptors A and B are separated by an angular distance $\Delta\phi$. The signal from A is temporally delayed by the low-pass filter D before multiplication by the signal from B. This multiplication produces a positive output in response to rightward image motion. To achieve similar sensitivity to leftward motion and to cancel excitation by stationary flickering stimuli, a parallel delay-and-multiply operation takes place on the opposite arm. The outputs of the two multiplications are subtracted to give a single time-dependent correlator output R.

Although the correlator is nonlinear, its response to sinusoidal stimuli is of interest. If the input is a sinusoidal grating that contains only a single frequency component, the oscillations of the two subunits cancel and the correlator produces a constant output. For any linear delay filter, the output level depends separably on spatial and temporal frequency [Egelhaaf et al., 1989]. If the delay filter D is a first order low pass with time constant τ, a sinusoid of amplitude C and spatial frequency f_s travelling to the right at velocity v produces an output $R(t)$ as described in Equation 3.1.

$$R(t) = \frac{C^2 f_t}{2\pi\tau f_t^2 + 1/(2\pi\tau)^2} \sin(2\pi f_s \Delta\phi) \tag{3.1}$$

where f_t is the temporal frequency of the input signal [Egelhaaf et al., 1989]. At a given spatial frequency, the magnitude of correlator output increases with temporal frequency up to an optimum $f_{t,\text{opt}} = 1/(2\pi\tau)$ and then decreases monotonically as the velocity continues to increase. The output also varies with the square of C, which specifies grating brightness or, in the presence of preprocessing stages, grating contrast. A physical luminance grating must have positive mean luminance, so it will contain a dc component as well as an oscillatory component. In this case, the output will oscillate about the level given by Equation 3.1.

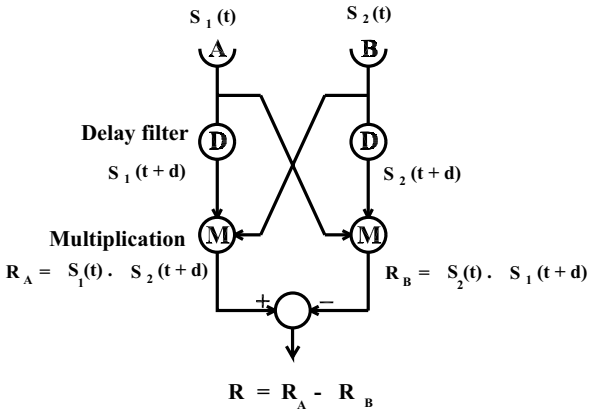

Figure 3.1. The Reichardt correlator has two receptors. A and B that take two input signals s_1 and s_2 with a fixed angular separation $\Delta\phi$. Each of these time dependent inputs passes through a linear delay filter (D) before being multiplied by the other, undelayed signal. The results of the two correlations thus obtained R_A and R_B are subtracted to produce a single output R. An object moving to the right will produce a positive output; an object moving to the left will produce a negative output.

3.2 Correlator response to broadband images

Since the correlator is a nonlinear system, its response to a generic stimulus cannot be represented as a sum of responses to sinusoidal components of the input. In particular, it was found that the response to a broadband image, such as a natural scene, varies dramatically with time, despite the fact that image velocity is constant.

There is a general mathematical relationship between the power spectrum of an image and the mean correlator response, explaining the empirical similarities in shape and differences in magnitude between velocity response curves of different images. Natural images differ from sinusoidal gratings in that they possess energy at multiple non-zero spatial frequencies, so that they produce broadband correlator input signals. As an image moves horizontally across a horizontally oriented correlator, one row of the image moves across the two correlator inputs. One might think of this row as a sum of sinusoids representing the Fourier components of that row. Because of the nonlinearity of the multiplication operation, the correlator output in response to moving image will differ from the sum of the responses to the individual sinusoidal components. The response to a sum of two sinusoids of different frequencies f_1 and f_2 consists of the sum of the constant responses predicted by Equation 3.1 to each sinusoid individually, plus oscillatory components of frequencies $f_1 + f_2$ and $f_1 - f_2$. Sufficient spatial and temporal averaging of the correlator output will eliminate these oscillatory components. The correlator therefore exhibits pseudo-linearity or linearity in the mean, in that the mean output in response to broadband image is equal to the sum of the responses to each individual sinusoidal input component.

This pseudo-linearity property implies that the mean response of a simple Reichardt correlator to a single row of an image depends only on the power spectrum of that row. Using Equation 3.1 for correlator response to a sinusoid and the fact that $f_t = f_s v$, mean correlator output can be written as

$$\overline{R} = \frac{1}{2\pi\tau} \int_0^\infty P(f_s) \frac{f_s v}{(f_s v)^2 + 1/(2\pi\tau)^2} \sin(2\pi f_s \Delta\phi) df_s, \tag{3.2}$$

where $P(f_s)$ represents the power spectral density of one row of the image at spatial frequency f_s.

3.3 Dror's Elaborated Reichardt Correlator

Although the simple correlator model produces more meaningful estimates of velocity for natural images than for arbitrary sinusoids, it suffers from two major shortcomings. First, the standard deviation of the correlator output is huge relative to its mean, with relative error values ranging from 3.3 to 76. Second, the mean correlator response for most natural images peaks at a velocity of $35 - 40°/s$. Because the velocity range below the peak response corresponds to the most probable range of inputs, it is assumed that

NOTE:
This figure is included on page 43
of the print copy of the thesis held in
the University of Adelaide Library.

Fig. 3.2. Block diagram of an elaborated correlator model. T, D, H and M represent temporal filters; S is a spatial filter; ρ and ξ are saturation functions (compressive nonlinearities). The subunit subtraction may be unbalanced, with weights g_{pos} and g_{neg}. The outputs of the various EMDs undergo two-dimensional spatial integration (Σ), which may be non-uniform, with the weights represented by w_i. For simplicity, this figure omits a number of demonstrated nonlinear and adaptive phenomena. After Dror (1998).

in the absence of contradictory information a correlator response is interpreted as the lower of the two putative velocities. Image velocities above the peak will therefore be misinterpreted. A shorter delay filter time constant would raise the peak response velocity, but experimentally described time constants [Harris et al., 1999] are not sufficiently low to account for the fact that insects may turn and track targets at velocities up to hundreds of degrees per second [Land and Eckert, 1985]. In this section it is shown that additional physiological components help to overcome these problems, raising the peak response velocity and lowering the relative error of the correlator output. Figure 3.2 illustrates an elaborate correlator, including spatial and temporal prefiltering, compressive nonlinearities, and output integration [Dror, 1998].

While both spatial and temporal pre-filtering decrease relative error, they have qualitatively opposite effects on the velocity response curves. Low pass spatial filtering increases

the peak response velocity, whereas low pass temporal filtering reduces it. The physiological system exhibits a predominance of spatial low pass filtering and temporal high pass filtering, both of which increase peak response and therefore broaden the range of velocities coded in the monotonically increasing portion of the response curve. This suggests that pre-filters are designed to improve the accuracy of velocity estimation in addition to optimising information transmission in the presence of noise [Laughlin, 1994].

Compressive nonlinearities, which are ubiquitous in neural receptors and synapses, occur at multiple points in the motion detection system. As stimulus contrast increases above a few percent, correlator response amplitude levels off rather than increasing quadratically [Egelhaaf and Borst, 1989], an experimentally observed effect termed as contrast saturation. Following Egelhaaf and Borst (1989), contrast saturation is modeled by including a dominant compressive nonlinearity directly before the multiplication operation on all correlator arms. A hyperbolic tangent function of the form $\rho(C) = \tanh(sC)$ was used. Hence contrast saturation decreases the dependence of the velocity response curve on the overall contrast level of the visual scene, except at very low contrasts. Such a contrast-independent response enables the correlator output to provide a useful indication of image velocity.

Spatial and temporal integration of correlator outputs will not effect the velocity response curve, which by definition is the mean of the correlator response over all space and time. However, integration will decrease the variance of the output signal, as illustrated by Single and Borst (1998). Although integration improves the accuracy of a velocity estimator, it reduces the resolution. Various neural pathways integrate correlator outputs to various extents in space and time, depending on the requirements of a particular task. Wide field neurons specifically measure egomotion as part of the optomotor pathway [Hausen and Egelhaaf, 1989]. In such cases, temporal resolution is more critical than spatial resolution, so the system integrates with respect to space rather than time. Chasing a conspecific, on the other hand requires a local motion estimate, so spatial integration must be limited, although some temporal integration may be permissible.

Maddess and Laughlin (1985) and Harris et al. (1999) described a motion induced decrease in correlator response to motion. Such motion adaptation may allow the correlator to respond sensitively to small changes in motion at low velocities while releasing it from saturation at higher velocities, effectively increasing its dynamic range. Harris *et al* also showed that the delay filter's time constant does not adapt to motion. Therefore delay filter's impulse response is kept constant for all simulations.

3.4 Experimental Verification

In order to test the relationships between image power spectra and velocity response curves, a set of experiments were carried out by David O'Carroll [Dror, 1998] in which the recordings of steady-state responses of wide-field neurons in a hoverfly to motion of broad-band images at different velocities was taken and compared with the analytical and computational predictions. Additional processing may occur subsequent to or in parallel with the wide-field neurons, so measuring of the fly's actual perception of velocity is not necessarily carried out. However, because the wide-field neurons perform extensive spatial integration and because recorded output is averaged over time, the results are effectively velocity response curves for wide-field neurons. It was found that the shapes of these curves and their dependence on image statistics agree with theoretical predictions.

Male specimens of the hoverfly *Volucella* are used for these experiments. The wide field neurons have proven particularly amenable to physiological analysis because their structure and physiological characteristics are nearly identical in different animals [Hausen and Egelhaaf, 1989]. Intracellular recordings from wide field cells indicate that they sum the outputs of local Reichardt correlators in their receptive fields. The HS cells or the Horizontal System typically consists of three horizontal cells termed north (HSN), equatorial (HSE), and south (HSS) horizontal cells because their dendritic trees cover the dorsal, medial and ventral regions of the lobula plate, respectively, with corresponding physiological receptive fields. Experiments with moving periodic gratings indicate that HS cells respond maximally to progressive motion in the horizontal direction and are maximally inhibited by motion in the opposite direction, with a roughly cosine like relationship between angle of motion and response [Hausen and Egelhaaf, 1989; van Hateren, 1990]. As the temporal frequency of the stimulus increases, the response of HS cells rises to a maximum and then declines, following a curve similar to that predicted by Equation 3.2 [Hausen and Egelhaaf, 1989]. The spatial frequency tuning also follows the predictions of the correlator model, with a clear optimum independent of temporal frequency [O'Carroll et al., 1996].

Here recordings from HSN, HSNE, and HSE neurons which are tangential cells of the horizontal system are taken. (In syrphids such as *volucella*, the HS system consists of four neurons—an HSNE neuron in addition to the three HS neurons mentioned). All three of these neurons have dorsal receptive fields. They exhibit graded responses, with horizontal progressive motion eliciting maximum depolarization. Instead of natural images, random texture fields generated by a Picasso Image Synthesizer and a Dual Channel

Velocity Field and Stereogram Generator under control of a Macintosh computer. Modifications to both the Image Synthesizer and the Stereogram Generator allowed the display of moving texture fields at a video frame rate of 300 Hz with a variable raster rate on a Tektronix 608 XYZ display. The random textures consist of horizontal rectangular texture elements ("texels") each of which has an equal probability of being illuminated at the display's maximum intensity or not being illuminated at all. By varying a texture density parameter, texels of 4 different sizes (widths 10.3, 5.2, 2.6, and 1.3 mm/texel) can be used. In addition, by lowering the frame rate to 200 Hz and increasing the raster rate, we can decrease the texel size to 0.86 or 0.67 mm/texel. The apparent size of a texel and the perceived angular velocity depend on the distance of the moving pattern from the fly. The display is generally positioned at about 6 cm from the fly's eyes and the precise distance was recorded for each experiment in order to calculate angular velocity and angular texture density. Texels thus subtended angles ranging from 0.64° (less than the inter-ommatidial angle of the fly) to 9.7° (nearly ten times that angle).

The mean horizontal power spectral density for a row of a random texture image with texel width θ°/texel is given by:

$$P(f_s) = \theta \frac{\sin^2(\pi f_s \theta)}{(\pi f_s \theta)^2}. \tag{3.3}$$

Figure 3.3 shows horizontal power spectral densities $P(f_s)$ for random textures of three different densities. The power spectrum of one natural image is included for comparison. The spectra of random textures are flat at low frequencies but drop as $(f_s)^{-2}$ at high frequencies. As the texel size decreases, image statistics approach those of white noise and the power spectrum becomes flatter.

Figure 3.4 shows velocity response curves for a model correlator predicted analytically from the power spectra of random textures of different densities. This correlator model included spatial blurring by the optics and temporal filtering by light-adapted photoreceptors and LMCs. Saturation effects were not included because the texels themselves provide a binary-valued input signal. The exact shape of the curves and their peak response velocities depend on a number of parameters which are difficult to predict on the basis of available data for these cells, such as the extent of temporal high-pass filtering. Moreover, the predicted curves in this figure do not take account of known nonlinear effects such as gain control. However several important features of the curves can be predicted with confidence. First, they will have the same general shape as the curves predicted for natural images, increasing monotonically up to a peak response velocity and then falling off.

NOTE:
This figure is included on page 47
of the print copy of the thesis held in
the University of Adelaide Library.

Fig. 3.3. Mean horizontal power spectra of random textures used in experiments. The legend indicates the density for each texture. The spectra are flat at low frequencies but drop as $(f_s)^{-2}$ at higher frequencies. As texture density increases and texel size (θ) decreases, the spectrum becomes more flat. The nulls in each spectrum are at multiples of the texel frequency and indicate that the luminance of different texels are uncorrelated. The solid line represents the mean horizontal power spectrum of natural image for comparison. The units of power used for the random texture spectra are consistent for the three curves, but arbitrary in comparison with the natural image. After Dror (1998).

Second, the velocity response curves should shift to the left as texture density increases. Since the luminous intensities of the display could not be adjusted for either illuminated or unilluminated texels, most of the stimuli had a very high contrast of 94%, and a mean luminance of 37 Cd/m^2. In some later experiments, luminance was increased and the stimulus contrast was decreased by adding uniform foreground illumination from one or more light sources. The responses of HS cells to progressive horizontal image motion at 25 approximately logarithmically spaced velocities between 7 and 1800°/s was recorded, with the precise velocity range depending on frame rate and screen distance. Here 200 ms presentations of image motion at each velocity were interleaved with 200 ms adapting stimuli at an angular velocity 400°/s. The last 100 ms of the response at each velocity was averaged to obtain a single response value. To control for effects of adaptation during the experiment, the test stimuli were presented in a monotonically increasing sequence of velocities, followed by a monotonically decreasing sequence. This protocol was designed

NOTE:
This figure is included on page 48
of the print copy of the thesis held in
the University of Adelaide Library.

Fig. 3.4. Velocity response curves for a model correlator for random texture images of various densities, as predicted from the power spectra. As for the models used in previous simulations, the correlator had an inter-receptor angle $\Delta\phi = 1.05°$ and a first-order delay filter with $\tau = 35$ ms. It also included a spatial pre-filter corresponding to optic blurring and a temporal pre-filter corresponding to light-adapted LMCs, with impulse responses and parameters. Each successive curve corresponds to a random texture with double the density of the preceding one; these correspond to five of the texture densities in Fig. 7. These curves were calculated theoretically. While the maximum response levels vary significantly with density, each curve is normalized to a maximum value of 1.0 to facilitate comparison with experimental data. Due to effects such as contrast normalization and output saturation not included in this model, we do not expect these overall variations in magnitude to match those recorded experimentally. After Dror (1998).

to obtain quick measurements while maintaining the cell in a state of uniform adaptation. The protocol was repeated three or more times for a given texture and obtained a single velocity response curve by averaging the responses to motion at each velocity.

Figure 3.5 shows velocity response curves for one HSNE neuron measured at six texture densities. The velocity response curves show the expected shape, rising to a peak response at some optimum velocity and then falling off again. The tuning curves for broadband images have higher optimal velocities than the corresponding tuning curves for sinusoidal gratings of optimal spatial frequency (Figure 3.7 A). As texture density increases, the curves shift to the left, with the optimal velocity decreasing from over 300°/s to 100°/s

NOTE:
This figure is included on page 49
of the print copy of the thesis held in
the University of Adelaide Library.

Fig. 3.5. Velocity response curves measured at six different texture densities for a single HSNE neuron. As density increases, the mean horizontal power spectrum of the texture field becomes flatter and the velocity response curves shift to the left. The legend indicates texture density and the number of measurements averaged to obtain each point in the figure. The magnitude of responses of the cell at different texture densities differed significantly, but the curves presented here have been normalised for comparison. After Dror (1998).

over the range of densities used. As predicted by the model of Figure 3.5, the curves cease to shift left at the very highest densities, as the image power spectrum becomes almost completely flat in the relevant frequency range. The model predicts qualitative aspects of the recorded data surprisingly well, given that model parameters were literature values for typical large hoverflies and were not tuned to the cell in question. The data of Figure 3.6 represents a particularly successful recording session lasting over 2 hours, during which response curves at a wide range of densities were measured and test protocols were repeated twice at most densities. Neurons in other animals gave similar results, but typically with more noise due to shorter recording sessions.

Figure 3.6 indicates the relationship between texture density and optimum velocity for a number of HS cells from several flies, as well as analytical predictions. With one notable anomaly, all cells show the expected decrease of optimum velocity with increasing density. The velocity optima do differ systematically from cell to cell. At any given texture density, some neurons have optimal velocities 50% greater than those of others. These systematic

NOTE:
This figure is included on page 50
of the print copy of the thesis held in
the University of Adelaide Library.

Fig. 3.6. Optimum velocity as a function of texture density for several neurons. The optimal velocity is the velocity of pattern motion for which the neuron gives a maximum steady-state response; this velocity generally decreases as texture density increases, in agreement with theoretical predictions. Neuron 080698 HSNE exhibited anomalous behaviour, while a noisy velocity response curve led to a single anomalous measurement for neuron 120698 HSN at the second lowest texture density. The legends indicate the type of each neuron. The two thick solid lines indicate two sets of recordings from the neuron whose velocity response curves are shown in Fig. 7. In order to reduce the effects of measurement noise, it was estimated the velocity optimum for each cell at each texture density as the velocity optimum of a second-order thin-plate spline fit to the velocity response curve on a logarithmic velocity axis. The amplitude of the curve corresponding to analytical predictions is highly sensitive to the properties of the high-pass temporal pre-filter. A slightly weaker high-pass temporal pre-filter or the addition of a realistic high-pass spatial pre-filter would decrease the predicted response velocities into the range of those observed experimentally. After Dror (1998).

variations between cells may reflect real variations in physiology; for example, temporal prefiltering might vary between organisms or regions of the visual field, which could explain the differences shown. Alternatively, the variations may be due to differences in the recordings, such as the position of the screen relative to the receptive field. Velocity

tuning curves as well as spatial and temporal frequency tuning curves were compared for HSE, HSNE, and HSN neurons, but no significant systematic differences were found.

NOTE:
This figure is included on page 51
of the print copy of the thesis held in
the University of Adelaide Library.

Fig. 3.7. Spatial and temporal frequency tuning for the HSNE neuron of Fig. 3.5. Other cells exhibit similar spatial and temporal tuning. (A) Temporal frequency tuning curve, indicating the response of the cells to a sinusoidal grating of spatial frequency 0.1 cycles/° moving at different velocities. The optimal velocity for this sinusoidal grating is close to 58°/s. (B) Spatial frequency tuning curve, indicating the responses of the cell to sinusoidal gratings of different spatial frequencies moving at velocities chosen to maintain a constant temporal frequency of 5 Hz. After Dror (1998).

Figure 3.7 shows spatial and temporal frequency tuning curves for the HSNE neuron of Figure 3.5, measured using sinusoidal gratings. These confirm that the delay filter time constant and inter-receptor angle parameters used in the model of Figure 3.4 agree

approximately with the values for the correlators feeding this particular wide-field neuron, even though the model parameters were not adjusted to fit our data set.

The experimental velocity response curves do differ significantly from the model predictions in several important respects. First, the experimental curves are noticeably flattened (platykurtic) at high response levels. This probably reflects output saturation or gain control, neither of which were included in the model. Second, the experimental curves are asymmetric on a logarithmic velocity axis, rising more gradually than they fall. The temporal frequency response, effectively a velocity response curve for a sinusoidal grating exhibits a similar effect (Figure 3.7 A). The asymmetry may result from a delay filter more complex than first-order [Mastebroek et al., 1980]. Alternatively, the wide-field neuron may sum outputs of correlators with different delay filters; temporal frequency tuning curves suggestive of such summation have been reported in *Bombylius* [O'Carroll et al., 1997]. These discrepancies may therefore reflect the simplified nature of this particular model rather than a fundamental discrepancy between theory and experiment.

NOTE:
This figure is included on page 52
of the print copy of the thesis held in
the University of Adelaide Library.

Fig. 3.8. Velocity response curves measured for the same HSNE cell at different contrasts. For very low contrasts (2.43%), the curve decreases in magnitude and becomes more symmetric, with the optimal velocity decreasing. Changing the contrast from 5.9% to 94% causes a relatively minor change in the size of the response and the change of the velocity response curve. The contrast is varied by adding foreground illumination from one or more light sources. Legends indicate the contrast and mean luminance for each response curve. Moving images were random textures with density 0.42 texels/°. After Dror (1998).

When the contrast was decreased (Figure 3.8), the curves became much less flattened, suggesting release from saturation. The curves also became more symmetrical and shifted to the left, suggesting that if two delay filters contribute to the output, one may dominate at very low contrast levels. A similar effect is noted in response to sinusoidal gratings in *Bombylius* [O'Carroll et al., 1996].

Perhaps the most interesting feature of Figure 3.8 was that response varies little for contrasts between 6% and 94%. Most real-world contrasts fall in this range.

While the system behaves like a simple Reichardt correlator at low contrasts, these curves support the presence of some form of contrast gain control at higher contrasts. This invariance with contrast was characteristic of the motion-adapted system; the unadapted system exhibits larger variations of the response level with stimulus contrast. From a practical point of view, the invariance of the motion-adapted velocity response curve with overall image contrast implies that mean correlator output may indeed provide an accurate estimate of velocity for a wide range of natural images.

3.5 Conclusion

This study analyzes the important aspects in accurately estimating the local image velocity using a Reichardt correlator and has come up with a conclusion that experimentally supported elaborations of the basic Reichardt correlator enhances its reliability as a velocity estimator. This result has been verified by the data obtained from the experiments done on the wide field neurons of the hoverfly which qualitatively matches the analytical predictions by Dror in several important respects.

It has also been found that many physiological mechanism, such as contrast normalisation and adaptation could further improve the performance of the velocity estimation system. Further work in this area is to include many other experimentally observed phenomena like output gain control, signal rectification and various other forms of adaptation in model simulation and to study their effects on the accuracy of velocity estimation.

In this chapter, Reichardt correlator model and velocity estimation by a Reichardt correlator is studied and in the previous chapter, template model and velocity estimation by the template model was described, in the next chapter a comparison study of velocity performance by both the models is carried out.

Chapter 4

Comparison of the Template and the Reichardt models

BIOLOGICALLY inspired motion detection models are bound to replace the conventional machine vision technology, because of their simplicity and significant advantages in a number of applications. The dominant model for insect motion detection, first proposed by Hassentein and Reichardt in 1956, has gained widespread acceptance in the invertebrate vision community. The template model is another known model proposed later by Horridge in 1990, which permits simple tracking techniques and lends itself easily to both hardware and software. In this paper, this two different motion detecting strategies is compared. It was found from the data obtained from the intracellular recordings of the steady-state responses of wide field neurons in the hoverfly *Volucella*, that the shape of the curves obtained agree with the theoretical predictions made by Dror. In order to compare this with the template model, an experiment is carried out to obtain the velocity response curves of the template model to the same image statistics. The results lead us to believe that the fly motion detector emulates a modified Reichardt correlator.

4.1 Comparison of Reichardt correlator vs Template model

Velocity information of moving objects is widely used in a number of applications. This information can contribute to scene interpretation such as recovering the structure of the environment, estimating time to contact or ranges for obstacle avoidance, segmenting images and predicting future location of moving objects. A basic Reichardt correlator does not act as a velocity estimator because its response depends on contrast and spatial frequency as well as velocity.

Analysis and simulations suggest that the processes commonly found in visual systems, such as pre-filtering, response compression, integration, and adaptation, improve the reliability of velocity estimation by Reichardt correlator and expand the range of velocities coded. As explained and presented in Chapter 3, to prove this experimentally, intracellular recordings of steady-state responses of wide field neurons in the hoverfly *Volucella* to the motion of broadband images were used. With the data obtained the mean response level of wide field neuron was plotted as a function of velocity. Figure 3.4 and Figure3.5 (please refer Chapter 3) shows the velocity response curves for a model correlator predicted analytically from the power spectra of random textures of different densities and the velocity response curves for one HSNE neuron measured at six texture densities respectively. The shapes of these curves and their dependence on image statistics was found to agree with the theoretical predictions.

4.1.1 Template model response

An experiment is set up to compare the velocity response curve of the Reichardt correlator with that of the Template model. In this experiment, the camera was placed in the centre of a hollow cylinder with random horizontal texture elements ("texels") printed on a horizontal roll of paper stuck inside. The cylinder is motor controlled and the angular speed of the cylinder can be adjusted by changing the voltage supply to the motor. Our program then tried to count the number of templates produced by detecting the motion of the texels. By varying the texel density parameter, texels of 4 different sizes can be used. The experiment is repeated with each of the texture density. Since the template models gives us templates as its response, the velocity response curve in this case is the velocity versus template count curve, which is plotted for each texel density.

Figure 4.1 shows the velocity response curves measured at five texture densities using the template model algorithm. The template model detects the motion of objects as edges.

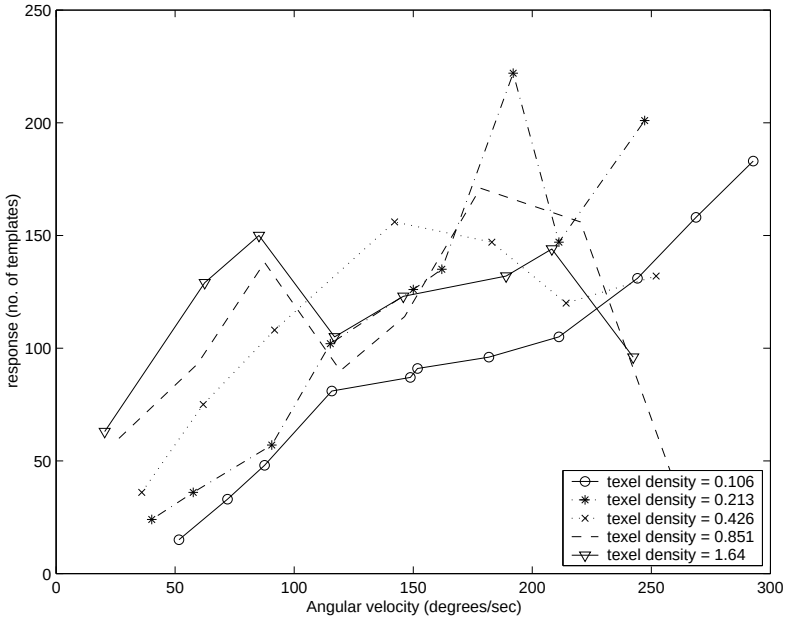

Fig. 4.1. Velocity response curves measured at six different texture densities for the Template model. As density increases, the mean horizontal power spectrum of the texture field becomes flatter and the velocity response curves shift to the left. The legend indicates texture density and the number of measurements averaged to obtain each point in the figure. The magnitude of responses of the cell at different texture densities differed significantly, but the curves presented here have been normalised for comparison.

When the texel density is low, the texels are bigger and fewer in number, hence there are less number of edges and thereby less number of templates at low velocities. but as the texel density increases, the number of edges detected and the number of templates increases resulting in the curve shifting to the left, showing similarity in response to the response of the HSNE neuron. As the texel density increases, it is seen that the response increases then begins to fall. This is due to the blurring effect caused by the fast motion of the texels. The higher the texel density, the blur is caused at lower velocities. But as the velocity is increased, more texels passes in front of the camera at a shorter rate, resulting in increase of the response again and then further increase in velocity causes more motion blur resulting in decrease of response.

The templates are counted with better accuracy at lower velocities than at higher velocities. This ties in well within our expectations because the forward tracking method [Yakovleff et al., 1994] is used in our software to track the templates and to determine their velocity. Using this method, the velocity of a slow moving object is determined and updated at each sampling instant [Nguyen et al., 1996]. In theory, this makes it more useful in

counting templates for objects moving at low velocities. Thus, by using the software to obtain the template counts, we can expect the response of all texture densities to decrease at high velocities. This is due to the higher risk of having lost or missing templates when the texels move at higher speeds. Theoretically, we would expect the curves to be similar in shape to the responses of the model correlator and the wide-field neuron of the hoverfly because the template model is essentially developed from an EMD like the correlator model.

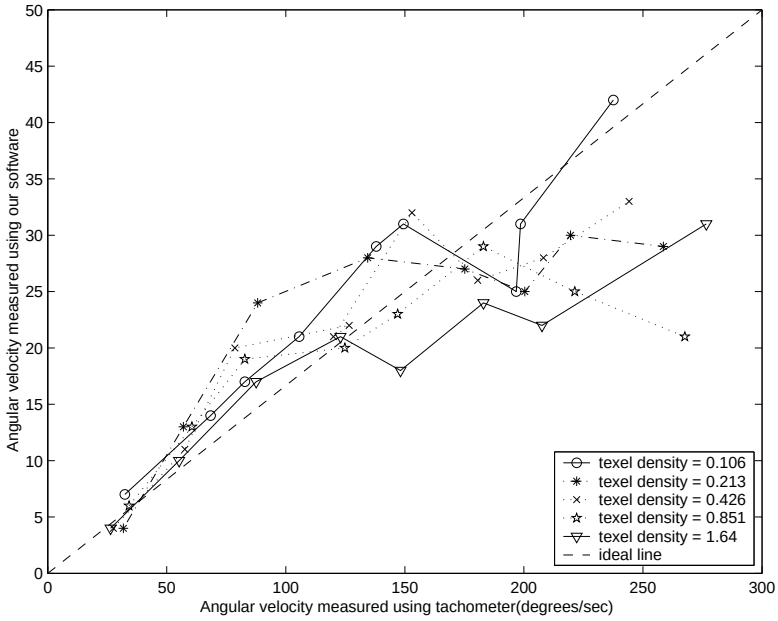

Fig. 4.2. Benchmark angular velocity measured by the tachometer versus benchmark velocity measured using the Template model. The angular velocity measured by the algorithm deviates from the actual measured velocity or the ideal line for all texel densities at higher velocities. As the texel density increases it is found that the velocity from which the deviation from the ideal line starts to occur, decreases.

In the template model,experimental results show that a moving object (or edge) consistently causes the same motion sensitive template to occur at subsequent time steps, and at positions corresponding to the displacement of the edge relative to the detector [Yakovleff et al., 1994]. The angular velocity may be estimated by evaluating the ratio of the displacement of a motion sensitive template, to the time between the template's occurrences. In Figure 4.2, the horizontal axes represents the angular velocity measured using a tachometer and the vertical axis represents the velocity measured, by tracking of the templates, done by the template model algorithm. At lower velocities it is seen that the algorithm gives a correct measure of the velocity as the response coincides with the

ideal response, but as the velocity increases, the response tends to deviate from the ideal line [Rajesh et al., 2003].

This occurs because at higher velocities, the edges are not clearly identified due to the blurring effect caused by the fast motion of texels. Thus the response deviates more from the ideal line at higher velocities. It can also be seen that as the texture density increases, the velocity from which the deviation from the ideal line starts to occur decreases. The lower the texture density, the more accurate the estimation will be at higher speeds. While at high texture densities, the estimation gets worse. This is due to the fact that at higher texture densities, there are more edges to detect and at higher velocities, the texels pass the CMOS camera at a faster rate, resulting in the loss of templates. Motion blurring is also added to this effect and the response gets worse at higher densities.

The above experiments are repeated using the stimuli generated by the Vision Egg software and the results are presented in the Appendix A. The Vision egg software presents the stimuli in a much better way than the rotating drum as the inconsistencies due to non-uniform speed control using the rotating drum is not present with the Vision Egg experiments and hence the output obtained is also much cleaner (See Appendix A).

4.2 Conclusion

This study analyzes the important aspects of accurately estimating the local image velocity using a Reichardt correlator and Template model. It is found that both these models give correct estimate of velocity at low velocities and their efficiency decreases at higher velocities. It also shows that experimentally supported elaborations of the basic Reichardt correlator enhances its reliability as a velocity estimator. Since the basic principle of a template model is the same as the Reichardt correlator, the elaborated Reichardt correlator and the template model produce similar response. This result has been verified by the data obtained from the experiments carried out on the wide field neurons of the hoverfly which qualitatively matches the above two results in several important aspects. But one can find that there is more similarity with the elaborated Reichardt correlator response than with the template model result.

In the next chapter we attempt to further improve the velocity performance of the Reichardt correlator model by implementing contrast adaptation in order to reduce the dependance of the correlator response to contrast.

Chapter 5

Contrast adaptation

MODIFICATIONS are made to the Reichardt detector that improve its performance in velocity detection by reducing its dependance on contrast and image structure. Our recent neurobiological experiments suggest that adaptive mechanisms decrease EMD (elementary motion detector) dependence on pattern contrast and improve reliability. So appropriate modelling of an adaptive feedback mechanism is carried out to normalise contrast of input signals in order to improve the reliability and robustness of velocity estimation. Dror's elaborated Reichardt model [Dror et al., 2000] is expanded to include feedback adaptation and saturation in our model and a comparative study is conducted on the effects of the addition of each elaboration on the performance of the model. The relative error in each case is also studied.

5.1 Elaborated Reichardt Correlator with contrast adaptation

In this chapter, an attempt is made to reduce the dependence of the Reichardt correlator model to contrast so that it can give more accurate velocity information. Firstly, the correlational EMD model for the early stage of motion detection in insects, is elaborated to mimic the properties of the fly visual system. Then based on motion adaptation studies in the insect visual system [Harris et al., 1999; Harris et al., 2000], contrast gain reduction is implemented using a feedback mechanism. It is found that, the resulting adaptive feedback EMD array model is successful in making the response less sensitive to contrast.

5.2 The Insect Visual System

In order to build an elaborated Reichardt model which mimics the properties of the fly visual system, it is very important to first have a clear idea about the different layers of the visual system of insects. So in this section, the visual system of insects is briefly reviewed so as it can be used as a building block for elaborating our model.

Compared to other animals, insects have a visual system of intermediate complexity. The insect visual system is classified into four layers, *retina*, *lamina*, *medulla* and *lobula* as shown in Figure 1.1. Each of these ganglia in the optic lobe are organised into columns and strata.

5.2.1 Retina

The retina has a large array of photoreceptors which sample the visual field. Each of a fly's two compound eyes consists of several thousand facets, or ommatidia, arranged in a vertically oriented hexagonal lattice [Hausen and Egelhaaf, 1989]. The ommatidium is a functional unit comprising a dioptric apparatus and a rhabdom, which accommodates eight photoreceptors. The light captured by each ommatidium is projected onto the eight photoreceptors. Each compound eye of the insect is composed of approximately 3000 to 4000 ommatidia. The ommatidial sampling lattice forms the basis for motion detection. Behavioural experiments indicate that stimulation of two individual photoreceptors in nearby ommatidia is sufficient to produce a turning response [Kirschfeld, 1972].These

results were confirmed by Reihle and Franceschini (1982) in neural recordings from motion sensitive neurons, which suggests that the two arms of the correlator corresponds to nearby ommatidia, with $\Delta\phi$ depending on the inter-ommatidial angle. Studies on photo-transduction in insects reveal that the retinal signal, as sampled by the photoreceptors, is already blurred by diffraction effects of the lens optics as well as the properties of the photoreceptor themselves [Snyder et al., 1977]. Hence the response of the photoreceptor to an input image has the characteristics of spatial low pass filter. Incidentally this spatial filter is matched to the sample 'mosaic', for optimal aliasing. Also it is found that some temporal pre-filtering takes place in the photoreceptors. Photoreceptors, which provide the inputs to the correlator, depend on chemical transduction processes [Laughlin, 1994; Niven et al., 2007] that cannot respond instantly to changes in the luminance signal. They therefore filter their inputs with a temporal impulse response found experimentally to have a log-normal form lasting several tens of milliseconds [Payne and Howard, 1981; Howard et al., 1984].

5.2.2 Lamina

The first visual ganglion, the lamina, has a large number of identical channels. The main output cells of the lamina, the monopolar cells, codes contrast to reduce redundancy caused by large variation in background intensity. Contrast enhancement is thus a vital function of the lamina and this is performed through biological spatial high pass filters (lateral inhibition). Anatomical evidence suggests that the large monopolar cells (LMCs) form the primary input channel to the correlators [Douglass and Strausfeld, 1995]. In addition, there is clear evidence that some kind of temporal high-pass filtering takes place in the lamina [Laughlin, 1994; James, 1990]. Recordings and modelling by James *et al* from the temporal impulse responses of LMCs (in the hoverfly *Eristalis*), reveals that the LMCs have the characteristic of the difference of two log normal functions [James, 1990]. While some LMCs also appear to perform high pass spatial filtering, the phenomena is weaker and more variable than high pass temporal filtering [James, 1990; Laughlin, 1994].

5.2.3 Medulla

The medulla is one the most complex layers of the insect visual system, characterized by numerous classes of local, columnar neurons as well as 'tangential cells', which arborize across large number of medulla columns. Because of their small size, little is known about

the physiology of most medulla neurons. Most of our understanding of properties of such cells comes from studying their target neurons in the lobula and lobula plate, which are large enough for physiological recordings. Nevertheless, some evidence suggests that columnar neurons in the medulla are responsible for key stages of local motion detection. Douglass and Strausfeld (1995) recorded a number of small field retinotopic neurons in the medulla, lobula and lobula plate. They found several motion sensitive cells whose physiological and anatomical characteristics suggest that they serve as components of an array of EMDs.

5.2.4 Lobula

The lobula receives input from the medulla to perform higher level functions. The lobula contains wide field directionally selective motion detection neurons which are responsible for the detection of whole field or whole frame motion. In the lobula, the output of the local retinotopic movement detectors is pooled. This is accomplished in the dendrites of the tangential cells in the lobula plate. Intracellular recordings from the wide field cells indicate that they sum the outputs of local Reichardt correlators in their receptive fields [Egelhaaf et al., 1989].

5.3 Natural stimuli

Naturalistic approaches are very important to understand the neural strategies employed by fly motion detectors during flight control and other complex behaviour during motion [Lewen et al., 2001; Lindemann et al., 2003]. Natural images are not arbitrary. Recent work has shown that certain image statistics are highly predictable in the natural world [Burton and Moorhead, 1987; Tolhurst et al., 1992; Ruderman, 1994] and that the biological visual system is optimized to take advantage of these statistics [Laughlin, 1981; Laughlin, 1994; Laughlin, 1996; Srinivasan et al., 1982; van Hateren, 1992; van Hateren, 1997; van Hateren and Snippe, 2001].

Although the simple correlator model produces more meaningful estimates of velocity for natural images than for arbitrary sinusoids, it suffers from two major shortcomings. First, the standard deviation of the correlator output is huge relative to its mean (3.3 to 76 times). Second, the mean correlator response for most natural images peaks at an angular velocity of $35 - 40°/s$ [Dror, 1998]. Because the velocity range below the peak

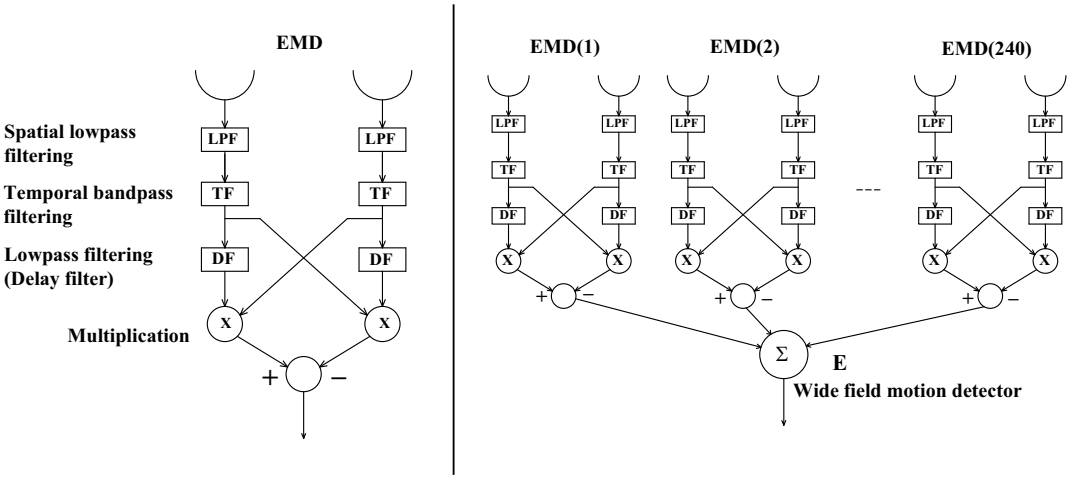

Figure 5.1. The elaborated EMD array. On the left, there is one elementary motion detector elaborated to include spatial low pass filtering (LPF) using a Gaussian filter of half width 2° and a temporal band pass filter (TF) using a log-normal difference filter is shown. This then passed through the delay and compare mechanism of the Reichardt detector using the delay filter (DF) which is a low pass filter. On the right there is the EMD array model, where an array of 37×240 elaborated EMDs are used to detect motion. The input stimulus given is a panoramic image photographed from a natural environment favored by insects, and considering the inter ommatidial angle as 1.5° there will be typically 240 ommatidia (EMDs) in one horizontal slice working together to detect motion. The output of these EMDs is pooled to get an average output (E) to enable wide field motion detection.

response corresponds to the most probable range of inputs, in the absence of contradictory information a correlator response is interpreted as the lower of the two putative velocities. Image velocities above the peak will therefore be misinterpreted. A shorter delay filter time constant would raise the peak response velocity, but experimentally described time constants [Harris et al., 1999] are not sufficiently low to account for the fact that insects may turn and track targets at velocities up to hundreds of degrees per second [Land and Eckert, 1985]. It is found that additional physiological components (temporal high pass filtering and spatial low pass filtering) help to overcome these problems, raising the peak response velocity and lowering the relative error of the correlator output. Analysis and simulations suggest that the processes commonly found in visual systems, such as pre-filtering, response compression, integration, and adaptation, improve the reliability of velocity estimation and expand the range of velocities coded [Dror et al., 2001; Laughlin, 1994; Egelhaaf and Borst, 1989; Single and Borst, 1998; Hausen and Egelhaaf, 1989; Maddess and Laughlin, 1985; Harris et al., 1999].

Following on from this work, an additional elaboration of this model is proposed to take account of major advances in the understanding of non-linear adaptive properties of insect motion detection [Harris et al., 1999; Harris et al., 2000].

Fig. 5.2. The panoramic natural image given as stimulus to the EMD model. A panorama of the image is formed by 'warping' 12 image tiles at 30° intervals to remove lens distortions and then by wrapping its ends together using Apple Quicktime VR software on a Macintosh computer.

5.4 Elaborated motion detector model

In our experiments, natural images photographed from favoured hovering positions of the hoverfly are used. Then the edges of the images are wrapped to form a panoramic image which is given as a stimulus to our fly EMD model. A panorama is formed by 'warping' 12 image 'tiles' at 30° intervals to remove lens distribution and then wrapping its ends together. This is done using Apple Quicktime VR software on a Macintosh computer. The resulting image has a width of 8352 pixels and height of 1264 pixels. Spatial pre-filtering is implemented by two-dimensional convolution of the image with a Gaussian kernel of half width (standard deviation) of 2°, which approximates the acceptance function of typical fly photoreceptor [Hardie, 1985]. Only the luminance (gray scale) information is taken from the image using the green channel since photoreceptors are green sensitive. The spatially low pass filtered image is illustrated in Figure 5.3. The distance between two ommatidia in an insect eye is between 1 and 1.5 degrees. Since the insect is looking at an image of width 8352 pixels in 360 degrees, if the inter-ommatidial angle is considered as 1.5 degrees, there will be an array of typically 240 ommatidia (EMDs) working together to detect motion as shown in Figure 5.1.

This image is temporally filtered with a difference of log-normal filter to copy the response of the lamina monopolar cells [James, 1990; Payne and Howard, 1981; Howard et al., 1984].

The temporally pre-filtered image is then converted to a space-time matrix based on desired velocity, which is then simulated along the height of the image in pixels. The

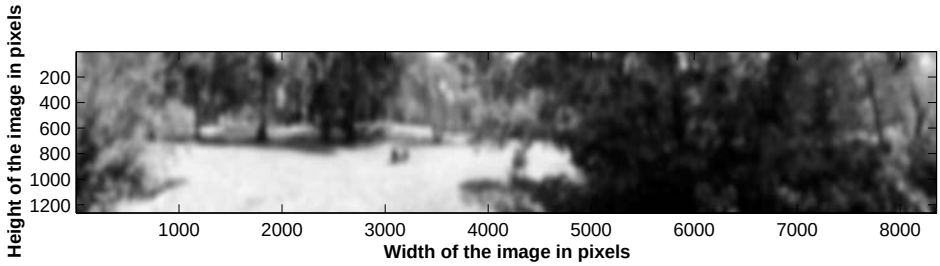

Fig. 5.3. Spatially low pass filtered naturalistic image used as stimuli. The natural image is pre-filtered using a spatial low pass filter, imitating the characteristics of the photoreceptor. The spatial filter used here is a Gaussian filter of half width (standard deviation) two degrees.

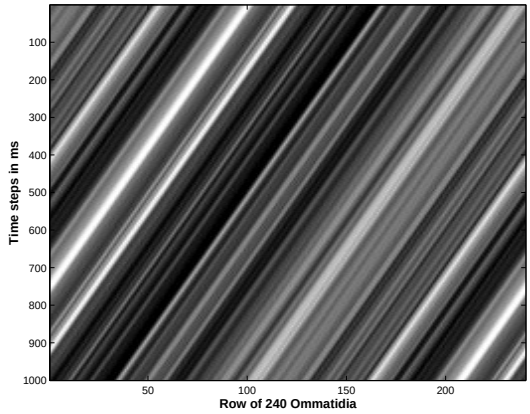

Fig. 5.4. Space time matrix obtained when a single row of EMDs is simulated at a constant velocity. This figure shows a single row of the image animated at a constant velocity of 200°/s and sampled onto an array of 240 ommatidia.

height of the image is also divided into 36 ommatidia by calculating the pixels per degree and keeping the inter-ommatidial angle as 1.5 degrees.

Then this spatio-temporally pre-filtered image is given to the EMD array, which correlates the inputs to give an array of outputs as carried out in the insect eye [Hausen and Egelhaaf, 1989; Douglass and Strausfeld, 1995]. Then the EMD array model copies the lobula by averaging the outputs to produce an average EMD response. The simulations are first performed with a constant velocity of 200 degrees per second and the input output responses are shown in Figure 5.5. Then the simulations are repeated by increasing the velocity in steps and the mean correlator response is shown in Figure 5.6. It is found that at higher velocities, the response decreases a little.

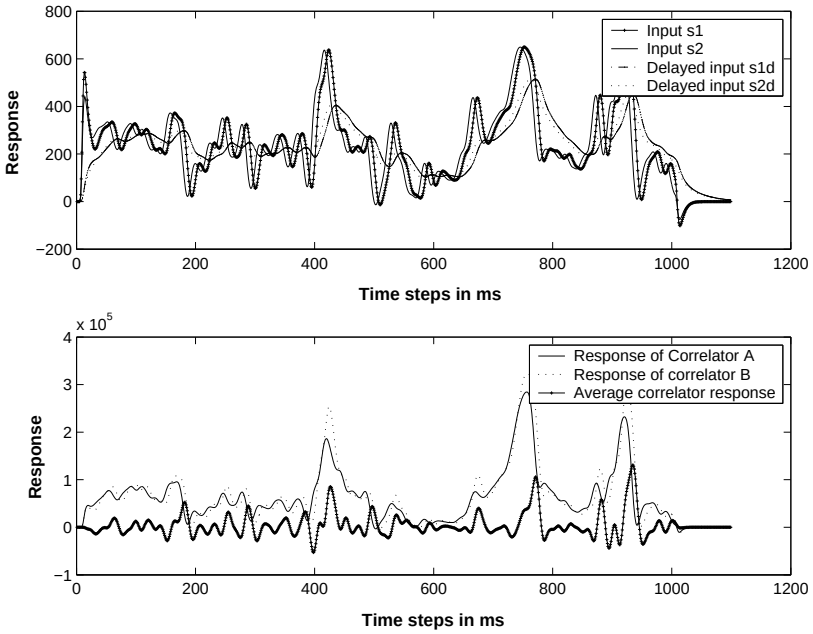

Fig. 5.5. The input and the output responses of a single EMD from the modelled array moving with constant velocity of $200°/s$. Refer to Figure 3.1 for explanation of each component. Note that the EMD output as a function of time is highly variable, and is maximal whenever high contrast features more across the EMD inputs (s_1,s_2). The s1d and s2d are the delayed version of the inputs s1 and s2 after passing through the delay filters, that is, they are outputs of the delay filters in both the arms.

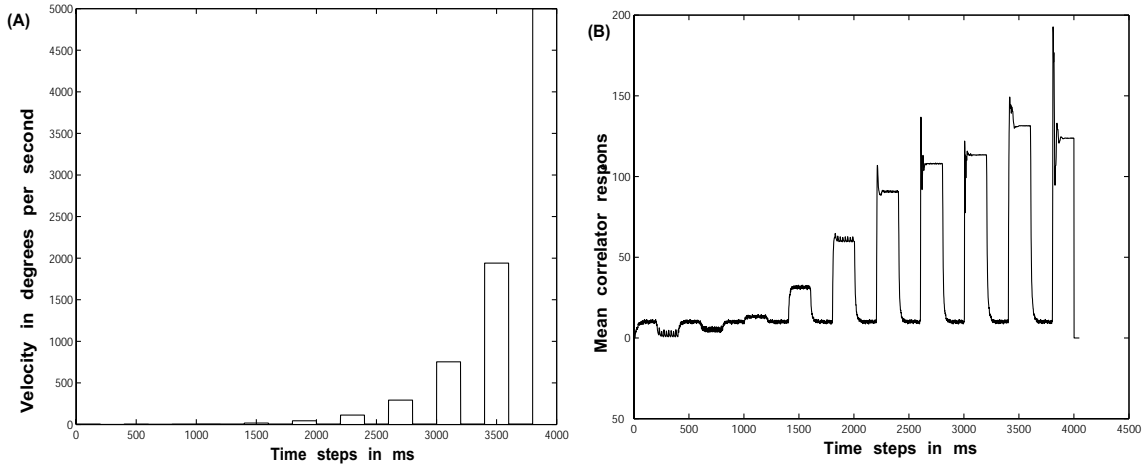

Fig. 5.6. (A)To test motion adapted responses, the velocity is increased step-wise, with interleaved bursts of adapting motion (constant speed). The velocity is increased in steps with time, (B) The mean correlator response of the EMD array model to a natural image moving with velocity increasing in steps across the EMD array.

5.5 Motion adaptation in the fly visual system

When you observe a fly hovering near a flower, you will be struck by the ability of its visual system to estimate self-motion in order to achieve stabilization during hovering, which shows its sensitivity to low velocities. Similarly when you watch these insects engaging in high speed aerial pursuits while chasing mates, it is clear that their visual system adapts easily to higher velocities as well. These activities of insects demand a visual system with a large dynamic range. Although physiological recordings demonstrate that insect motion detectors have such high sensitivity to contrast that they are able to respond over a huge range of velocities, this inherently high sensitivity to motion makes them prone to saturation [Harris et al., 1999; Harris et al., 2000].

Several electrophysiological investigations into the dynamic response properties of fly motion detectors indicate the presence of motion adaptation.

Barlow and Hill (1963) were the first to show that responses of directional motion sensitive neurons decline gradually during prolonged motion stimulation. They suggested that this decline in response and subsequent reduction in background activity that occurred after the stimulus stopped, may be correlated with psychophysically measured motion after effects. Maddess and Laughlin (1985) reported that adaptation does not depend on velocity but rather on temporal frequency of the stimulus, and may alleviate saturation.

Moreover, de Ruyter van Steveninck et al. (1986) found from his studies that the value of the time constant depends on the magnitude of local velocity from which he concludes that the fly visual system uses estimates of local stimulus velocity to tune its filtering operations. Borst and Egelhaaf (1987) disagree with him over the origin of the adapting signal, suggesting that the adaptational state of a motion detector is governed mainly by the temporal frequency of the signal in its input channels.

Clifford and Langley (1995) in their adaptive Reichardt model, proposed that the adapting signal originates from locally integrated responses of a one dimensional array of elementary motion detectors, and is fed back to adapt the time constants of their temporal filters, suggesting that adaptation to higher velocities is achieved by shortening the delay filter in the correlator model [Clifford and Langley, 1997]. Ibbotson et al. (1998) tested this adaptive Reichardt model by recording from neurons in the wallaby visual system, finding little change in the position of temporal frequency optimum before and after motion adaptation, despite showing changes to image step responses similar to those in fly neurons. Furthermore recent studies in the fly by Harris et al. (1999) show little change in the

temporal and spatial tuning properties of fly motion sensitive cells following adaptation, indicating that motion adaptation does not significantly alter the inherent velocity optimum of the EMDs. Instead this work reveals that motion adaptation induces a profound decrease in contrast sensitivity of fly motion sensitive cells via two proposed mechanisms, a local direction sensitive after potential and a local direction insensitive contrast control. This gain reduction may serve to reduce sensitivity to image contrast, as well as reducing tendency of the motion detectors to saturate [Harris et al., 2000].

5.6 Contrast adaptation

In the correlation model of motion detection, each correlator contains an expansive non linearity (multiplication). This would make the correlator output particularly sensitive to the magnitude of input signals and so potentially vulnerable to saturation. The after potential and the gain reduction serve to release the motion pathway from this saturation, allowing to maintain a wide sensitivity across a wide range of stimulus conditions. The after potential acts antagonistically to recent activity in the cell, repositioning the cell's responses within the available signalling range. By analogy with the retina, this type of subtractive mechanism may exploit correlations in continuous signals, reduce redundancy and maintain the operations of synapses in favorable regions of their input output functions.

Similarly the gain reduction component of adaptation scales down the magnitude of signals in the motion pathway. If the correlator nonlinearity is to be protected from saturation the gain control should act on the inputs. Furthermore, since the output of the correlators depends on the spatio-temporal correlation between the input signals as well as their magnitude, the gain control would be best regulated by the magnitude of the correlator output, not the magnitude of the inputs. This is consistent with the observation that adaptation is recruited by motion but not flicker. The gain reduction mechanism will be recruited strongly when either the mean or the variance of the stimulus velocity distribution becomes large.

5.7 Experimental Verification

In order to test the relationships between image power spectra and velocity response curves, a set of experiments were carried out in which the recordings of steady-state

responses of wide-field neurons in a hoverfly to motion of broad-band images at different velocities was taken and compared with the analytical and computational predictions. It was found that the shapes of these curves and their dependence on image statistics agree with theoretical predictions [Dror et al., 2001]. Male specimens of the hoverfly *Volucella* are used for these experiments. The wide field neurons have proven particularly amenable to physiological analysis because their structure and physiological characteristics are nearly identical in different animals [Hausen and Egelhaaf, 1989].

The results of the experiments suggested that while the system behaves like a simple Reichardt correlator at low contrasts, these curves support the presence of some form of contrast gain control at higher contrasts. This invariance with contrast is characteristic of the motion-adapted system; the unadapted system exhibits larger variations of the response level with stimulus contrast. From a practical point of view, the invariance of the motion-adapted velocity response curve with overall image contrast implies that mean correlator output may indeed provide an accurate estimate of velocity for a wide range of natural images [Dror, 1998].

5.8 Contrast Gain Reduction — Feedback Adaptive EMD model

In our elaborated Reichardt correlator array model, in order to remove the dependency of the response to changes in contrast and spatial frequency and to get an accurate estimate of velocity, contrast gain reduction is implemented by a feedback adaptive process as in Figure 5.7. The gain of the EMD inputs is reduced by a signal derived from the rectified and low pass filtered outputs of a local EMD pool with different local preferred direction, fed back to control the gain of the EMD inputs. The model was inspired by observation of contrast dependent gain reduction in the responses of HS neurons following motion stimulation [Harris et al., 1999]. The model captures several aspects of the adaptive phenomena observed in the biological system. In particular, the adaptation is strongest when the local motion detector output is the largest, conferring a robustness in adaptation to motion signals as opposed to static flicker or noise applied to the inputs. This matches data obtained from the electrophysiological experiments, which show that motion stimuli are much more effective at recruiting adaptive gain than other stimuli. Secondly, the adaptive mechanisms remain independent of the direction of local motion, despite the selectivity for motion as the source of adaptation.

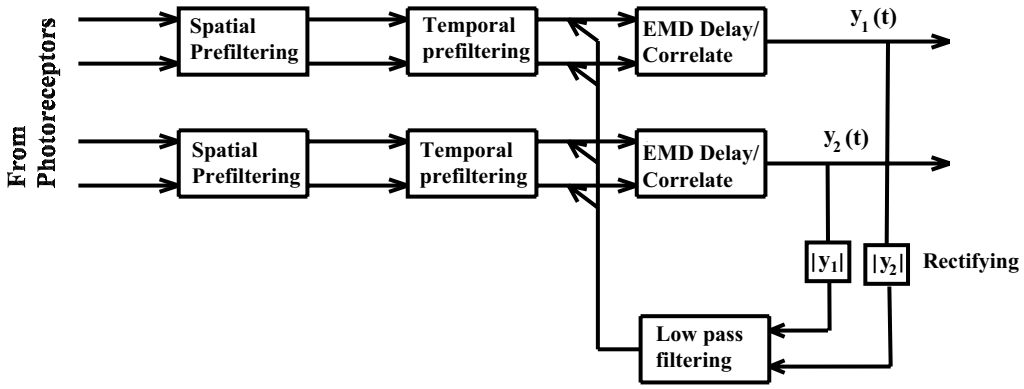

Fig. 5.7. Block diagram of a feed back adaptive EMD model. In order to reduce the dependency of the response to contrast, a further elaboration to this model is implemented that reduces contrast gain by a feedback adaptive mechanism. The output signals $y_1(t)$ and $y_2(t)$ are rectified, low pass filtered and fed back to control the gain of the EMD inputs. So depending upon the outputs of the EMD, the input to the Reichardt model is controlled. The model captures several aspects of the adaptive phenomena observed in the biological system. In particular, the adaptation is strongest when the local motion detector output is the largest, conferring a robustness in adaptation to motion signals as opposed to static flicker or noise applied to the inputs. This matches data obtained from the electrophysiological experiments, which show that motion stimuli are much more effective at recruiting adaptive gain than other stimuli. Secondly, the adaptive mechanisms remain independent of the direction of local motion, despite the selectivity for motion as the source of adaptation.

In order to test the performance of this adaptive EMD array, we have conducted an experiment to compare the responses of the non-adapted elaborated EMD array with the adaptive elaborated EMD array. The experiment is repeated for three different contrasts of 1, 0.5 and 0.1 to understand the influence of contrast on the adapted and unadapted response. From the Figure 5.9 it is seen the response of the unadapted model at contrast 1 is huge and the difference between the response at contrast 1 and contrast 0.5 is also big showing a large variation in the correlator response. But in Figure 5.10 it is seen that when feed back adaptation is implemented, the model response at contrast 1 decreases and is nearly half that of the response seen in Figure 5.9. The difference between the responses at different contrast has also decreased considerably, indicating that the inclusion of adaptation has decreased the variation of the response at different contrasts. Hence it can noted that the adaptive feedback mechanism has resulted in decreasing the contrast sensitivity of the model, compared with an otherwise similar model that lacks adaptive

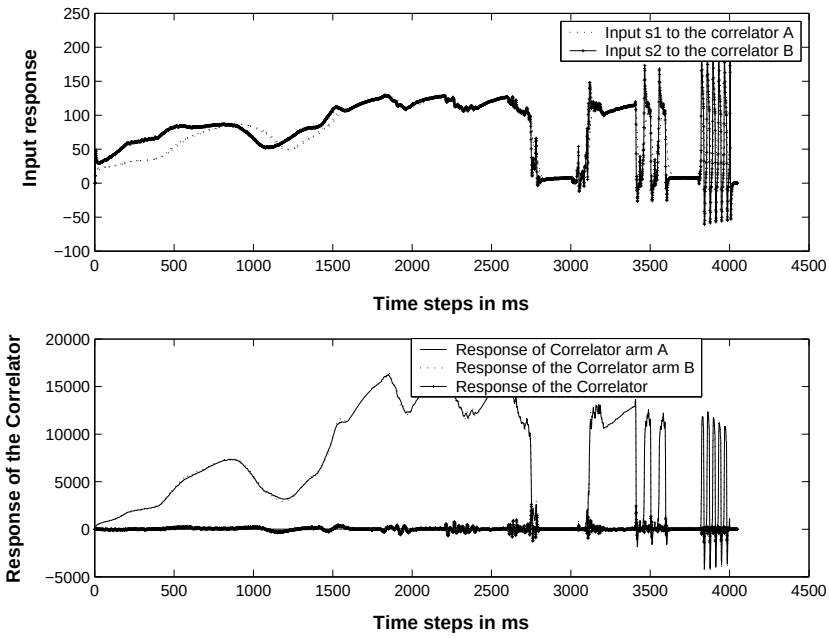

Fig. 5.8. The input and the output responses of one EMD from the adaptive EMD array in response to the test sequence illustrated in Figure 5.6. The input s1 and s2 given to the two correlator arms A and B (A simple Reichardt correlator is illustrated in figure 3.1) are shown in the first figure and the output response from the two correlator arms A and B are shown in the second figure. The response R, after subtraction of the response from two correlator arms is the output response of the correlator.

feedback [Rajesh et al., 2002; Rajesh et al., 2005a]. It is also found that the adaptation is stronger at higher velocities and at higher contrasts.

5.9 Conclusion and future work

A preliminary implementation of an adaptive EMD array model to take account of recent neurobiological research on the fly visual system is presented. This model leads to contrast gain reduction at local EMDs as observed in fly HS neurons [Harris et al., 2000]. In our model, this gain reduction is achieved by direct feedback of local EMD outputs. An alternative mechanism in which the feedback signal gates a feed-forward contrast normalisation [Shoemaker et al., 2001], provides an alternative explanation for the physiological data. Ongoing physiological experiments in our laboratory aim to test predictions of both models and thus to establish whether either of these models are fully consistent with motion adaptation mechanisms in the fly visual systems.

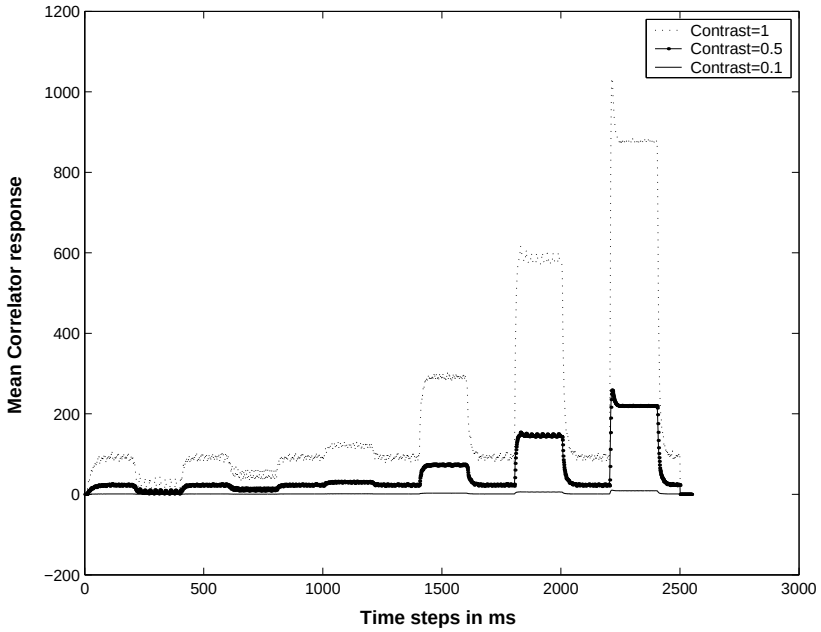

Fig. 5.9. The correlator response of an un-adaptive EMD array at three different contrasts. The response of the elaborated EMD array without including the adaptive feedback loop is noted at three different contrast (1, 0.5 and 0.1). It is clearly seen from the graph that there is huge variation of the response with contrast.

The present model demonstrates that a feedback gain control is capable of reducing the dependence of EMD output on contrast in natural images. Although the model does not in its present form, achieve *velocity constancy* with respect to contrast, as implied by pilot experiments on the fly HS neurons [Harris et al., 2000], this model is deliberately simplified to exclude known additional non-linearities of the fly vision, including logarithmic encoding of luminance and response saturation. Since earlier work already showed that these additional elaborations improve velocity coding by EMDs [Dror, 1998; Dror et al., 2001], inclusion of such components in future models may serve to achieve substantial improvements in performance of an adaptive EMD over earlier versions of the Reichardt correlator.

In the next chapter, in order to further improve Reichardt model velocity performance, a study of the pattern noise is conducted and methods of reducing pattern noise to improve model performance is discussed.

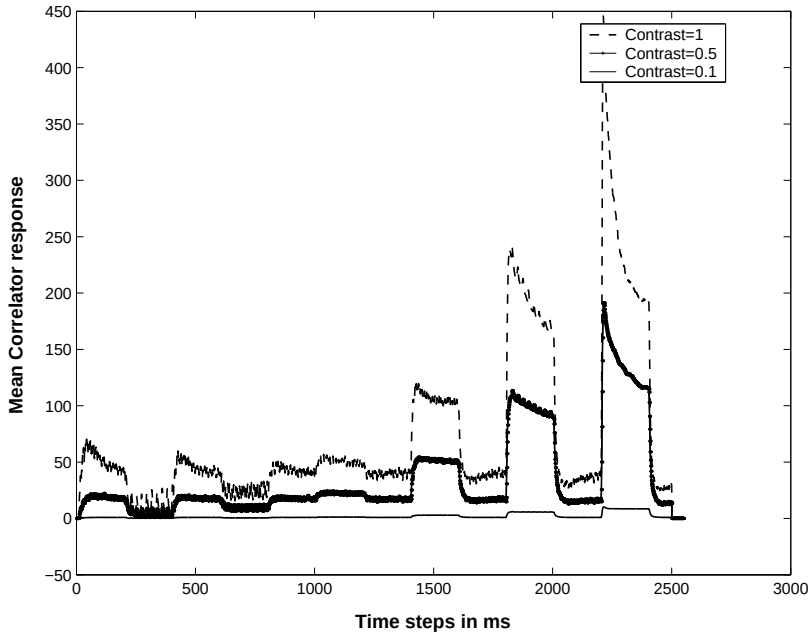

Fig. 5.10. The correlator response of an adaptive EMD array at three different contrasts. The response of the elaborated EMD array including the adaptive feedback loop is noted at three different contrast levels (1, 0.5 and 0.1). The adaptive feedback mechanism helps in decreasing the dependence of the response to contrast.