Bioinspired Contrast Vision Computation for Robust Motion Estimation Against Natural Signals

Qinbing Fu
Guangzhou University; University of Lincoln

Jigen Peng
Guangzhou University

Shigang Yue
Guangzhou University; University of Lincoln

2021 International Joint Conference on Neural Networks (IJCNN) | 978-1-6654-3900-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/IJCNN52387.2021.9533680

Abstract—**This paper aims at addressing a challenging problem on reliably estimating image motion against highly variable natural signals which artificial dynamic vision systems are faced with. Previously, the visual system response always represents fluctuation and high variance influenced by spatial contrast, the local difference between neighbouring luminance values. Effective contrast computation is therefore a prerequisite for robust motion vision. In this regard, sighted animals such as flies are remarkably adept at estimating image motion regardless of image statistics by rapidly adjusting contrast sensitivity. Current artificial visual systems, however, cannot account for this capability. Learning from neuroscience, here we propose contrast vision computation to improve a state-of-the-art, bio-inspired neural network model for background motion estimation. This includes mainly two neural computation schemes of (1) an instantaneous, feedback divisive contrast normalisation prior to motion correlation in the ON and OFF pathways to reduce local contrast sensitivity, (2) parallel contrast pathways influencing ON/OFF motion signals, negatively, at the pooling output layer to suppress high-contrast optic flows. We created a dataset of many shifting natural images with high input variability to investigate the proposed method. The experiments have demonstrated the effectiveness and robustness of the proposed contrast vision computation to reduce response fluctuation and variance against natural signals. The fidelity of motion perception thus has been significantly increased. The proposed methods could be generic to other motion vision models dealing with high-contrast visual scenes.**

I. Introduction

Motion vision, the perception of surrounding optic flows, provides an indispensable source of sensory feedback for sighted animals and intelligent machines to carry out survival-critical tasks in various challenging environments. For reacting effectively and timely to motion-induced events like collision avoidance and target tracking, the dynamic vision systems need to respond reliably and independently of the environmental statistics which nevertheless poses huge challenges in natural scenes with high input variability.

Spatial contrast represents the local difference between adjacent luminance values that is an essential element in natural image processing [1]. When considering image movements, such complexity is extended to the temporal domain in which the models need to reliably extract information from highly variable natural signals. In addition, since non-linearity is a basic attribute of motion detection, correlation, and prediction, the natural signals with high input variability would be liable to amplify the motion detector response on larger variation; the response thus becomes fluctuating and shows high variance, which is unexpected to accurately estimate image motion.

As a result of millions of years of evolutionary development, biological motion vision systems are capable of quickly modulating the contrast sensitivity, accordingly, maintaining robust performance despite the variability of natural visual inputs. Flies, for instance, are remarkably adept at estimating the angular velocity (AV) of moving natural scenes regardless of image statistics [2]. The underlying circuits and mechanisms, however, have not been well understood. Recently, a few studies have progressed this field by looking into the internal structure of fly visual pathways accompanied by mathematical modelling, and demonstrated the contrast vision computation would be critical to explain such capability against natural signals [2], [3]. Little computational modelling and experimentation, however, have been done on this circuit mechanism.

Regarding bio-inspired motion detector models particularly in natural scenes, some effective methods have been proposed to improve such robustness including the widely used computations on spatial lateral inhibition for shaping the selectivity to diversity of motion patterns [4] including small target motion [5]–[7], looming (approaching) [8]–[10], foreground translating [11], and rotational (or spiral) motion [12]. To improve the adaptivity in natural scene processing, there are also dynamic temporal mechanisms proposed to reduce impact by environmental irregular movements [5], [13], and neural facilitation to enhance time-varying detection performance [14], [15]. A recent work by Wang et al. implemented a contrast pathway to better discriminate between the movement features of small targets and background details [16]. Wang et al. proposed an effective texture estimation pathway parallel to the motion correlation pathway for improving AV decoding of image motion [17]. Nevertheless, we have noticed that the current circuit models are still susceptible to natural signals with high input variability, which always represent high variance and fluctuation; sometimes this misleads estimation on motion direction and magnitude.

To close the performance gap between model and organism, we herein propose contrast vision computation with respect to a few prominent biological findings [2], [3]. Building upon the latest fly motion vision neural network model [11], a more comprehensive signal tuning map is shown in this paper by incorporating the proposed contrast vision including primarily

This research has received funding from the National Natural Science Foundation of China (Grant No. 12031003), the China Postdoctoral Science Foundation (Grant No. 2020M682651), and the EU Horizon 2020 project ULTRACEPT (Grant No. 778602). Correspondence with Dr Qinbing Fu (qinbingfu87@gmail.com)

978-0-7381-3366-9/21/$31.00 ©2021 IEEE

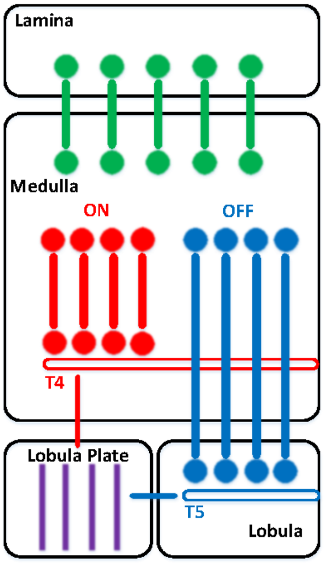


Fig. 1. Schematic illustration of the fly motion vision neural pathways with interneurons participating in motion and contrast computations: the Lamina layer conveys motion information to the Medulla layer which is then split into parallel ON and OFF channels; the motion signals in four cardinal directions are generated at T4 (in Medulla) and T5 (in Lobula) cells, and finally converge with contrast signals at the final quad-stratified Lobula Plate layer.

two neural computation schemes: (1) an instantaneous, feedback divisive contrast normalisation before motion correlation in the ON and OFF pathways to rapidly reduce local contrast sensitivity, (2) parallel contrast pathways combining both spatial lateral inhibition and temporal variation to affect negatively on ON/OFF motion signals at the final pooling layer, in order to suppress high-contrast optic flows. To investigate the proposed methods, we generated input natural signals including a hundred, shifting images. We also compared its performance with models lacking the contrast vision computation, evaluated its internal structure and parameter. The comparative experiments have demonstrated the effectiveness and robustness of the proposed method against input variability. The fidelity of motion perception has been significantly increased indicating the contrast vision computation is critical to dynamical visual processing.

The rest of this paper is structured as the following: Section II briefly reviews related works. Section III elaborates on formulation of the proposed network. Section IV articulates the experimental setting. Section V reports on the experimental results. Section VI discusses future works. Section VII summarises this research.

II. Related Work

Within this section, we review the relevant research on (1) fly (*Drosophila*) physiology with emphasis laid on contrast vision, and (2) normalisation as a sensory circuit mechanism, both of which form the basis of this paper.

A. Fly Contrast Vision

Flies are the most famous model system to study biological motion detection strategies, for reviews see [18]–[20]. The fly visual systems have been broadly applied for various real-world machine navigation applications including micro-aerial vehicles, unmanned aerial vehicles, and ground robotics, for reviews see [4], [21]–[23].

Specifically, the *Drosophila* motion vision pathways have been studied, extensively, which facilitates the understanding of neural computation for motion perception. The signal bifurcation into ON and OFF pathways is the most prominent finding which indicates the early visual motion is split and processed in parallel channels encoding brightness increments (ON) and decrements (OFF) [24], [25]. As illustrated in Fig. 1, there are several neuropile layers behind the eyes of flies to process both motion and contrast vision [2], [3]. Two computational studies have articulated the visual processing throughout this neural structure [4], [11], which is not reiterated here.

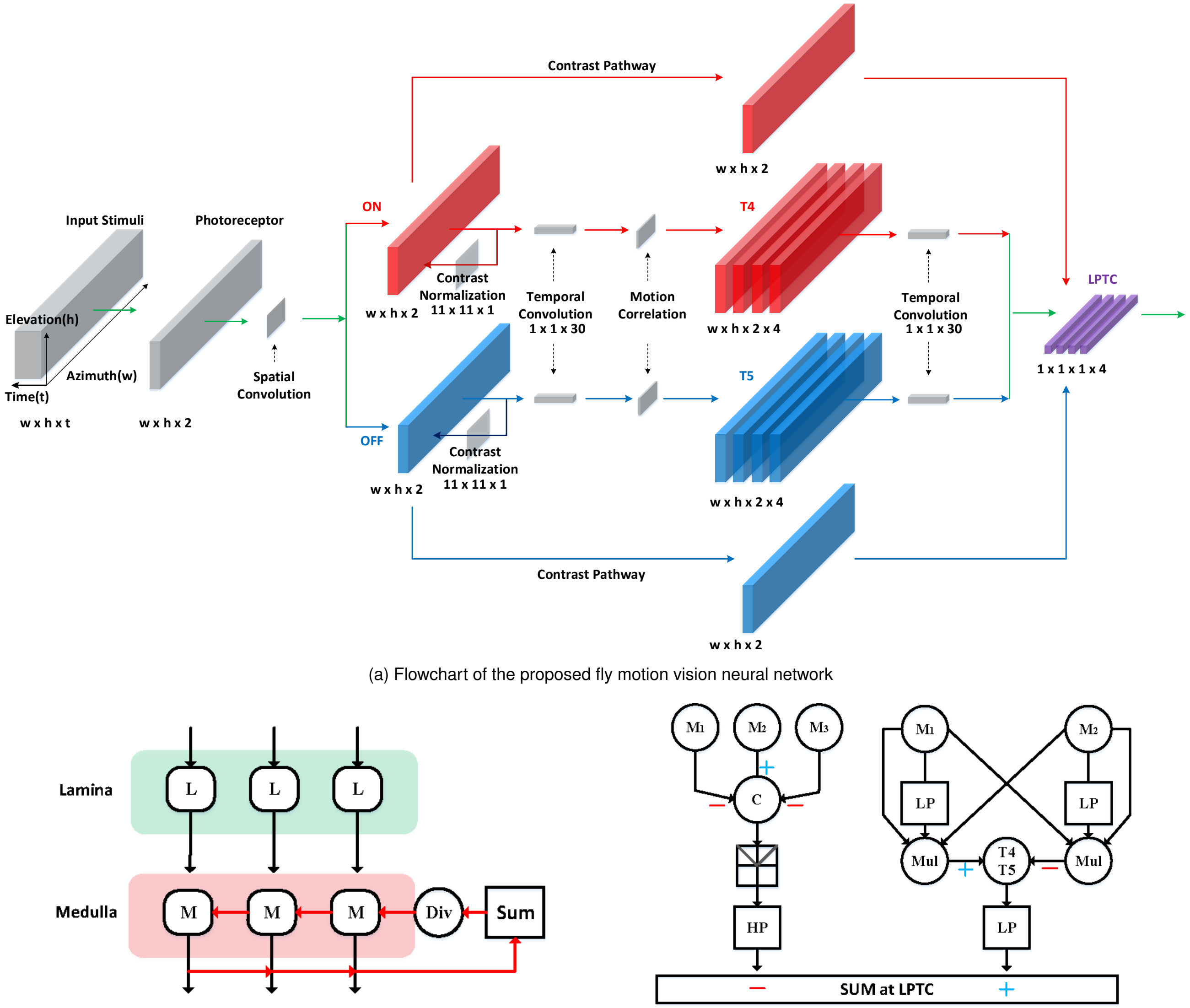
This paper highlights the fly contrast vision with emphasis laid upon two recent, representative discoveries. Bahl et al. pointed out that beginning from the Medulla, contrast and motion computations are carried out in parallel pathways; the signals finally converge in the Lobula Plate where the contrast signal provides negative effect to suppress the local motion signal at the lobula plate tangential cells (LPTCs) [3]. The authors employed optical illusions normally in human psychophysics to investigate such neural mechanisms which is novel in insect's behavioural experiments.

Another latest research by Drews et al. demonstrated the critical role of contrast normalisation based on rapid spatial integration of neural feedback, prior to motion correlation in the fly visual system [2]. To figure out the underlying circuit mechanisms, the authors made a few comparisons on (1) normalisation methods (linear, static or dynamic), (2) feedback or feed-forward signal processing; they also trained a batch of elementary motion detectors (EMDs) in a network, with their proposed contrast computation, to compare to response from the *Drosophila*. Finally, an instantaneous, feedback divisive contrast adaptation was proposed to account for robust motion vision in flies against diverse natural signals. The proposed network modelling will reflect both above neural computations for improving image motion estimation.

B. Normalisation as a Circuit Mechanism

Normalisation has been described as a widespread, generic circuit mechanism for removing higher-order correlations from natural signals [2], [26], which appears to be of general computational advantage for sensory processing across species (mammals [27], invertebrates [2]), and even modalities including olfactory [28], auditory [29], and visual sensing, for review see [30]. Moreover, it has also been combined with cognitive attention [31].

In the fly visual systems, the authors in [2] reported evidence for a non-linear, divisive normalisation which emerges at the level of the Medulla in Fig. 1. Concretely, each foreground, local Medulla interneuron value is divided into a neighbouring background field by spatially integrating surrounding feedback signals which represents dynamic, non-linear properties in normalisation as a sensory circuit mechanism. The proposed



(a) Flowchart of the proposed fly motion vision neural network

(b) Feedback contrast normalisation

(c) Parallel contrast and motion computation

Fig. 2. Illustrations of (a) the proposed network with multiple layers and parallel ON/OFF/contrast pathways, (b) the instant feedback normalisation before motion correlation, (c) the parallel contrast/motion computation and convergence: ‘L’ and ‘M’ are short for Lamina and Medulla interneurons; ‘C’, ‘HP’, ‘LP’, ‘Mul’, ‘Div’, ‘Sum’ are short for contrast, high-pass, low-pass, multiplication, division, summation.

network will for the first time implement it in the intact ON and OFF pathways processing.

III. Visual Neural Network Model

The proposed network will be elaborated in this section with formulation on multiple computational neuropile layers including the Retina, Lamina, Medulla, Lobula, and Lobula Plate, as well as network parameter configuration. Fig. 2a depicts the signal processing flowchart of the proposed network. The emphasis herein is laid upon the contrast vision computation including the feedback contrast normalisation in Fig. 2b, and the contrast-motion competition in Fig. 2c.

A. Network Formulation

Generally speaking, the proposed network consists of several layers in which the motion information bifurcates into ON and OFF pathways each corresponding to a parallel contrast pathway, as illustrated in Fig. 2. All the pathway signals converge at the final quad-stratified LPTCs jointly forming the output to two, i.e., horizontal sensitive (HS) and vertical sensitive (VS) systems for the estimation on direction and strength of image motion.

1) Computational Retina Layer: The input visual stimuli to the network is with $w \times h \times t$ in dimensions, as shown in Fig. 2a. The first Retina layer processes images at every two frames, $w \times h \times 2$, which consists of photoreceptors retrieving luminance values (green-channel or grey-scale $[0, 255]$ in our case). The computation can be defined as

$$P(x, y, t) = (L(x, y, t) - L(x, y, t-1) + P(x, y, t-1)) \cdot \alpha_1, \tag{1}$$

where $L(x, y, t) \in \mathbb{R}^3$ denote the input image streams, x, y and t are spatial and temporal positions. α_1 is a coefficient calculated by a time constant τ_1 and the time interval τ_i between discrete frames both in milliseconds, that is, $\alpha_1 = \tau_1 / (\tau_1 + \tau_i)$. The computation represents a first-order high-pass filtering.

2) Computational Lamina Layer: In the Lamina layer, the information is firstly filtered via a spatial convolution which is often represented by either Gaussian blur or Difference of Gaussians (DoGs) algorithm in bio-inspired modelling of preliminary motion processing [10], [11]. In this paper, we apply the DoGs to achieve the revealed edge selectivity in motion detection circuits, in order to mimic the functionality of Lamina interneurons. In addition to that, it is a variant of DoGs with polarity (ON or OFF) selectivity to fit with the subsequent neural computation on signal bifurcation. The whole calculation depicts a centre-surround antagonism with centre-positive ($3\times 3\times 1$) and surround-negative ($5\times 5\times 1$) Gaussians representing excitatory and inhibitory fields in space. That is,

$$P_e(x,y,t)=\sum_{u=-1}^{1}\sum_{v=-1}^{1}P(x+u,y+v,t)\cdot G_{\sigma_e}(u+1,v+1),$$

$$P_i(x,y,t)=\sum_{u=-2}^{2}\sum_{v=-2}^{2}P(x+u,y+v,t)\cdot G_{\sigma_i}(u+2,v+2), \tag{2}$$

$$G_\sigma(u,v)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{u^2+v^2}{2\sigma^2}\right), \tag{3}$$

where σ stands for σ_e and σ_i, i.e., the excitatory and inhibitory standard deviations. Note that if $x+u>w$ or $x+u<1$, the edge value is used in the convolution, and the same to another dimension. Accordingly, the initial Lamina neuron value is calculated by the following subtraction:

$$LA(x,y,t)=\begin{cases}|P_e(x,y,t)-P_i(x,y,t)|, \text{if} P_e\geq 0\&P_i\geq 0\\ -|P_e(x,y,t)-P_i(x,y,t)|, \text{if} P_e<0\&P_i<0\end{cases} \tag{4}$$

After that, as illustrated in Fig. 2a, the motion is split into two parallel ON/OFF motion pathways, which is represented by the half-wave rectifying. The calculations are expressed as follows:

$$\begin{aligned}ON(x,y,t)&=[LA(x,y,t)]^{+}+ON(x,y,t-1)\cdot\alpha_2,\\ OFF(x,y,t)&=-[LA(x,y,t)]^{-}+OFF(x,y,t-1)\cdot\alpha_2.\end{aligned} \tag{5}$$

$[x]^{+}$ and $[x]^{-}$ denote $\max(0,x)$ and $\min(x,0)$. A small fraction of previous signals is permitted to pass through.

3) Computational Medulla and Lobula Layers: In comparison with all previous works, here the Medulla layer is the place where contrast vision is initiated. This layer splits again the signals into parallel contrast and motion pathways, as illustrated in Fig. 2a and 2c. Before that, there is contrast normalisation in both ON/OFF channels, an instantaneous divisive operation by spatial integration of feedback neural signals (see Fig. 2b). Let M stands for both ON and OFF interneurons, the calculation can be defined as follows:

$$M(x,y,t)=\tanh\left(\frac{M(x,y,t)}{\alpha_3+\hat{M}(x,y,t)}\right), \tag{6}$$

$$\hat{M}(x,y,t)=\sum_{u=-5}^{5}\sum_{v=-5}^{5}M(x+u,y+v,t)\cdot G_{\sigma_c}(u+5,v+5), \tag{7}$$

$$G_{\sigma_c}(u,v)=\frac{1}{2\pi\sigma_c^2}\exp\left(-\frac{u^2+v^2}{2\sigma_c^2}\right). \tag{8}$$

Mathematically speaking, each Medulla interneuron value is divided into its background field via a spatial convolution ($11\times 11\times 1$) on immediate neural feedback signals. The normalised response is activated by a hyperbolic tangent tanh-function and then split into contrast and motion parallel computation.

Firstly, regarding the contrast computation in Fig. 2c, the Medulla cells form a centre-surround antagonism where the spatial contrast is calculated by competing with lateral inhibitions. The calculation can be defined as the following:

$$C(x,y,t)=|M(x,y,t)-\frac{1}{8}\cdot\sum_{u=-1}^{1}\sum_{v=-1}^{1}M(x+u,y+v,t)|,\quad v\neq 0, \text{if } u=0. \tag{9}$$

$$\hat{C}(x,y,t)=(C(x,y,t)-C(x,y,t-1)+\hat{C}(x,y,t-1))\cdot\alpha_1. \tag{10}$$

Secondly, the ON and OFF motion computation conforms to a theme of triple correlation illustrated in Fig. 2c, which could be a common combination form of motion detection across species including humans and flies for robust motion vision against natural signals [32]. As shown in Fig. 2a, the normalised ON or OFF signal experiences a temporal convolution ($1\times 1\times 30$), i.e., a delay defined as

$$\begin{aligned}D(x,y,t)=&\alpha_4 M(x,y,t)+(1-\alpha_4)M(x,y,t-1),\\ &\alpha_4=\tau_i/(\tau_i+\tau_2).\end{aligned} \tag{11}$$

τ_2 is a delay in milliseconds. D stands for both ON and OFF delayed signals. Accordingly, the proposed motion estimation strategy can be expressed as

$$\begin{aligned}R(x,y,t)=&D(x,y,t)\cdot M(x,y,t)\cdot M(x+sd,y,t)\\ &-M(x,y,t)\cdot D(x+sd,y,t)\cdot M(x+sd,y,t),\end{aligned} \tag{12}$$

where sd denotes the sampling distance in every pairwise detectors. We only show the non-linear correlation of neural response on rightward direction where others can be referred to this symmetric computational structure. Importantly, such directionally selective (DS) signals are received by T4 interneurons in the ON channels, T5 interneurons in the OFF channels where R accounts for both. As illustrated in Fig. 2a, there is an additional dimension for both T4 and T5 cells ($w\times h\times 2\times 4$) indicating the signals are correlated and mapped into four cardinal directions in the Medulla and Lobula layers.

Next, the polarity DS information is delayed by another temporal convolution, the computation of which is consistent with Eq. 11. Notably, the distinct DS responses are all generated in a feed-forward manner when arriving the T4 and T5 neurons, each group of which demonstrates the specific direction selectivity [11].

4) Computational Lobula Plate Layer: Finally, there are LPTCs in this pooling layer consisting of four stratified sub-layers where the same DS signals converge together. Importantly, the motion and contrast signals compete at the

TABLE I
NETWORK PARAMETERS

Parameter	Description	Value
τ_1	time constant in high-pass filtering	500(ms)
τ_i	time interval constant between frames	1000/30(ms)
$\{\sigma_e, \sigma_i\}$	standard deviation in spatial convolution	$\{1, 2\}$
α_2	small fraction in half-wave rectifying	0.1
sd	sampling distance in motion correlation	1
τ_2	time constant in temporal convolution	30(ms)
α_3	baseline contrast sensitivity	20
σ_c	standard deviation in contrast normalisation	5
$\{\alpha_5, \alpha_6\}$	gain factors	$\{1, 1\}$
$\{\beta_1, \beta_2\}$	exponent on ON/OFF signals	$\{0.5, 0.5\}$

LPTCs where the contrast signal influences the motion in every cardinal direction, negatively. That is,

$$S_{on}(x, y, t, d) = [T4(x, y, t, d) \cdot \alpha_5 - \hat{C}_{on}(x, y, t) \cdot \alpha_6]^+,$$
$$S_{off}(x, y, t, d) = [T5(x, y, t, d) \cdot \alpha_5 - \hat{C}_{off}(x, y, t) \cdot \alpha_6]^+, \quad (13)$$

where α_5, α_6 are two gain factors. d indicates each of the four cardinal directions.

The LPTCs at each stratified sub-layer integrates both polarity signals as the following:

$$LP(t, d) = \sum_{x=1}^{w} \sum_{y=1}^{h} g((S_{on}(x, y, t, d))^{\beta_1} + (S_{off}(x, y, t, d))^{\beta_2}). \quad (14)$$

Specifically, the response is activated by a Gaussian Error Linear Unit (GELU), g-function [33]. β_1 and β_2 can be set as any positive real number as the exponent to implement the non-linearity.

Furthermore, to fulfil the inhibition between adjacent Lobula Plate layers led by the Lobula Plate-intrinsic (LPi) interneurons [11], there are two output systems in this network, i.e., HS and VS. Let $d = 1$, $d = 2$, $d = 3$, $d = 4$ denote rightward, leftward, downward, upward motion sensitive stratified sub-layers, respectively. The calculations can be expressed as

$$HS(t) = [LP(t, 1) - LPi(t, 1)]^+ - [LP(t, 2) - LPi(t, 2)]^+,$$
$$VS(t) = [LP(t, 3) - LPi(t, 3)]^+ - [LP(t, 4) - LPi(t, 4)]^+, \quad (15)$$

$$\text{where } LPi(t, 1) = LP(t, 2),\ LPi(t, 2) = LP(t, 1),$$
$$LPi(t, 3) = LP(t, 4),\ LPi(t, 4) = LP(t, 3). \quad (16)$$

Accordingly, rightward and downward image motions lead to positive response whilst leftward and upward ones bring about negative response.

B. Network Parameters

The network parameters in this section are given in Table I. The proposed neural network model mainly processes signals in a feed-forward structure, currently has not been trained according to the following three perspectives. (1) There is prior knowledge from neuroscience when constructing this network which mimics the *Drosophila* motion vision pathways, largely simplifies network connections and parameters. (2) The connection between layers is achieved by convolution in either spatial or temporal domain which can learn from previous, related modelling experience on fly visual systems [11], [13], [34]. (3) The parameters of proposed new contrast vision computation are adapted from relevant physiological research including mathematical modelling and parameter optimisation, for instance, the baseline contrast sensitivity (α_3 in Eq. 6), which will be investigated in the experiments. In spite of those, the network training will be involved in our future work which, however, is not the focus of this paper.

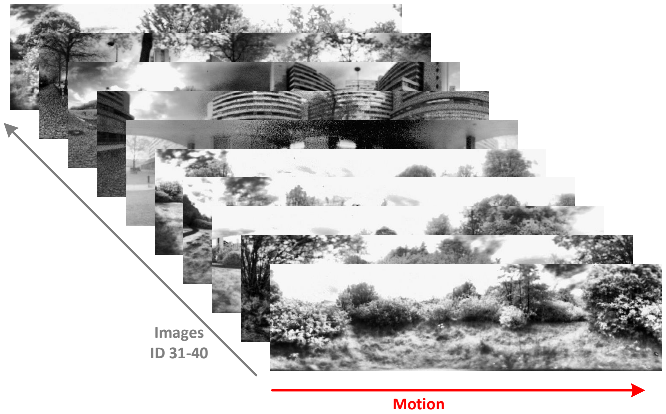


Fig. 3. Samples of natural images (ID 31-40) amongst the dataset including one hundred testing visual scenes. The images shift rightward at constant AV.

IV. EXPERIMENTAL SETTING

The proposed neural network model was set up in Visual Studio (Microsoft Corporation). Data analysis and visualisations were implemented in MATLAB (The MathWorks, Inc., Natick, MA, USA). The input natural image motions were generated by a Python open-source library.

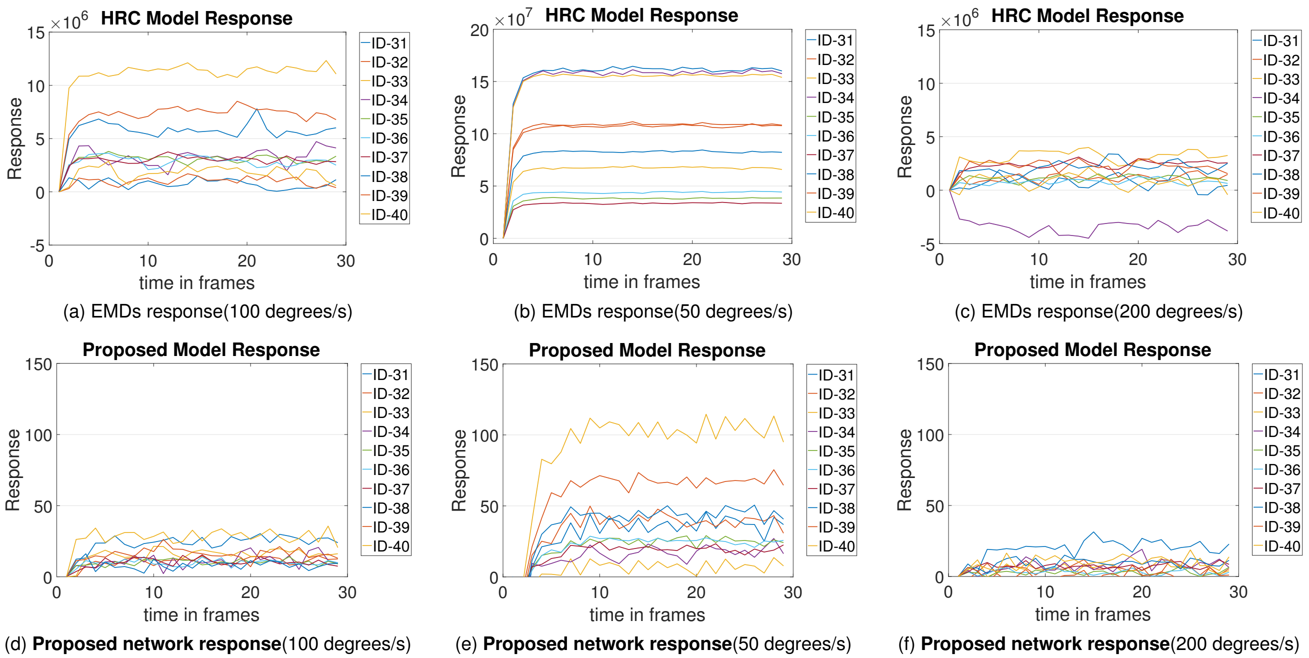
As illustrated in Fig. 3, the input visual stimuli videos were recreated upon a dataset of natural images from [35] showing high input variability, each at 30 frames per second. Each sample image has 927 × 251 pixels, and shifts rightward at the AV of 100 degrees/s. We compared the performance with a classic EMDs (or HRC) network adapted from [36], as well as the proposed network by removing all the contrast vision computations. We also used a batch of images with ID 31-40 in Fig. 3 to investigate the dynamic influence on model performance by setting a lower AV at 50 degrees/s, and a much higher one at 200 degrees/s. The source code algorithm and dataset can be found as open source in https://github.com/fuqinbing/Data-for-Contrast-Vision-Computation-IJCNN2021-paper-.

V. RESULTS

Within this section, the experimental results are illustrated. All the experiments can be categorised into two types of tests: (1) comparative tests on image motion estimation, (2) investigation on network parameter and structure.

A. Comparative Experiments

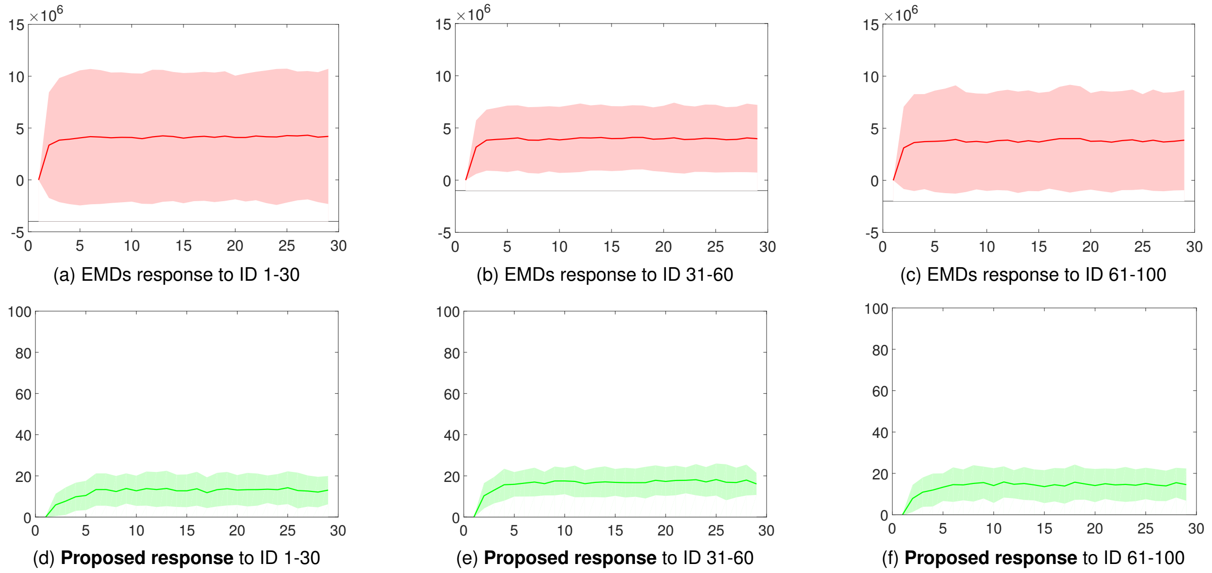
In the first type of experiments, we firstly show the network response to a batch of images with comparison to the HRC model. Fig. 4 compares the results. It can be clearly seen from the plots that the proposed network responds to image motions more consistently, compared to the response of EMDs which represents larger fluctuations and very high variations. Tested by lower-velocity motion, the proposed network fluctuates relatively more greatly, but still within an acceptable range. Such



(a) EMDs response(100 degrees/s) (b) EMDs response(50 degrees/s) (c) EMDs response(200 degrees/s)

(d) **Proposed network response**(100 degrees/s) (e) **Proposed network response**(50 degrees/s) (f) **Proposed network response**(200 degrees/s)

Fig. 4. Comparative response to natural image motion ID 31-40: the images shift rightward at three different velocities.

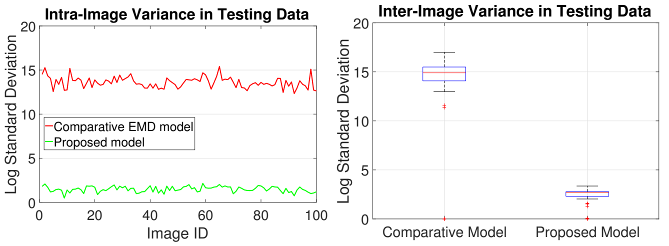


(a) EMDs response to ID 1-30 (b) EMDs response to ID 31-60 (c) EMDs response to ID 61-100

(d) **Proposed response** to ID 1-30 (e) **Proposed response** to ID 31-60 (f) **Proposed response** to ID 61-100

Fig. 5. Comparative statistical response to natural image motion ID 1-100 (100 degrees/s), divided into three subsets: X and Y axes denote response time in frames, and strength; line and shadow indicate mean and variance of response.

a difference is mainly caused by the temporal convolution in triple motion correlation; shortening the delay could be adaptable to slower image motion. On the other hand, the comparative model shows even greater variance of response. Tested by very fast image motion, the proposed network still maintains such robustness; the EMDs though could estimate the image motion, incorrectly, to make wrong prediction on direction (negative response to rightward image motion ID-34 caused by symmetric HRC detectors). For all image motion ID 1-100, Fig. 5 gives more intuitive statistical results on average and variance of response. Furthermore, Fig. 6 compares both intra- and inter-image deviations of response in support of the aforementioned results.



(a) **Intra**-image variance (b) **Inter**-image variance

Fig. 6. Statistical results of response variance on image motion ID 1-100 (100 degrees/s) between the comparative EMDs model and the proposed network.

Some achievements can be abstracted from the experiments so far: (1) the response variance and fluctuation have been significantly reduced by the proposed network; (2) the proposed network estimates the motion direction very robustly; (3) the proposed network is no longer affected greatly by input dynamics and variability; (4) the fidelity of motion estimation against natural signals has been enhanced by the proposed contrast vision computation.

B. Network Investigation

Next, we look deeper into two factors of the proposed network in order to illustrate the merits of contrast vision computation. In another word, we make two comparisons within this network including (1) removing all contrast neural computations, and (2) changing the baseline contrast sensitivity in contrast normalisation.

Fig. 7 compares the response in the above two situations, with average and variance information. Fig. 8 also depicts the intra-image variance along different natural images. Together with the response in Fig. 5, we can draw the following summations. First of all, the network with full contrast vision outperforms other cases showing smaller fluctuation and variance. Secondly, the baseline sensitivity could influence the network performance; here the lower value of the parameter in contrast normalisation could give rise to larger variance in response, but in acceptable increase indicated by Fig. 8a. This factor will be soon investigated with network training to work out its relation to both internal (network) and external (stimuli) parameters. Very interestingly, Fig. 8b also shows that the network gradually improves the robustness, from sieving all contrast-relevant neural computations, to implementing merely contrast normalisation, eventually to possessing the intact contrast vision computation proposed in this paper.

VI. Further Discussion

Through this research, we have demonstrated the efficacy of contrast vision computation in motion estimation of natural images with high input variability. The fidelity of motion perception has been significantly increased compared to previous related methods, worthy of attention from the community. Here we have the following observations for future work.

We have noticed that although the proposed network already shows expected robustness, there are still some crucial parameters, for instance, the baseline sensitivity in contrast normalisation determining the performance. We are setting up a more comprehensive collection of natural signals to train the network, and figure out the relations.

The objective of this work is estimating background image motion which is adapted from our recent modelling of fly motion vision pathways for foreground translating perception [11]. The two networks have differences in network structure and functionality. We will integrate both neural networks in order to give a more comprehensive signal tuning map of fly visual systems. In addition, there are also visual projection neurons, the lobula plate/lobula columnar (LPLC2), that show ultra-selectivity to looming objects only from the centre view [37]. The computational modelling of fly LPLC2 could share the most parts of the proposed network structure including the contrast vision computation, which also deserves investigation.

Lastly, we will continue to investigate the proposed contrast vision computation in other dynamic vision models or neural networks, for instance, the collision detection systems to alleviate the impact by spatial contrast.

VII. Concluding Remarks

Fly visual system is always the prominent paradigm to learn how motion perception becomes reliable and efficient. This paper presents effective neural computation on contrast vision to improve image motion estimation. The main contribution is the computational modelling of dynamic divisive normalisation, and parallel contrast vision pathways to reduce response fluctuation and variance against natural signals. The effectiveness and robustness have been proved in natural scenes with high input variability. This research may inspire relevant modelling on dealing with high-contrast visual scene.

Acknowledgement

We sincerely thank the three anonymous reviewers for their constructive comments on improving this paper.

References

[1] R. A. Frazor and W. S. Geisler, "Local luminance and contrast in natural images," *Vision Research*, vol. 46, pp. 1585–1598, 2006.

[2] M. S. Drews, A. Leonhardt, N. Pirogova, F. G. Richter, A. Schuetzenberger, L. Braun, E. Serbe, and A. Borst, "Dynamic signal compression for robust motion vision in flies," *Current Biology*, vol. 30, pp. 209–221, 2020.

[3] A. Bahl, E. Serbe, M. Meier, G. Ammer, and A. Borst, "Neural mechanisms for drosophila contrast vision," *Neuron*, vol. 88, pp. 1240–1252, 2015.

[4] Q. Fu, H. Wang, C. Hu, and S. Yue, "Towards computational models and applications of insect visual systems for motion perception: A review," *Artificial Life*, vol. 25, no. 3, pp. 263–311, 2019.

[5] S. D. Wiederman, P. A. Shoemaker, and D. C. O'Carroll, "A model for the detection of moving targets in visual clutter inspired by insect physiology," *PLoS One*, vol. 3, no. 7, pp. 1–11, 2008.

[6] K. J. Halupka, S. D. Wiederman, B. S. Cazzolato, and D. C. O'Carroll, "Discrete implementation of biologically inspired image processing for target detection," in *2011 Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, 2011, Conference Proceedings, pp. 143–148.

[7] S. D. Wiederman, R. S. A. Brinkworth, and D. C. O'Carroll, "Performance of a bio-inspired model for the robust detection of moving targets in high dynamic range natural scenes," *Journal of Computational and Theoretical Nanoscience*, vol. 7, no. 5, pp. 911–920, 2010.

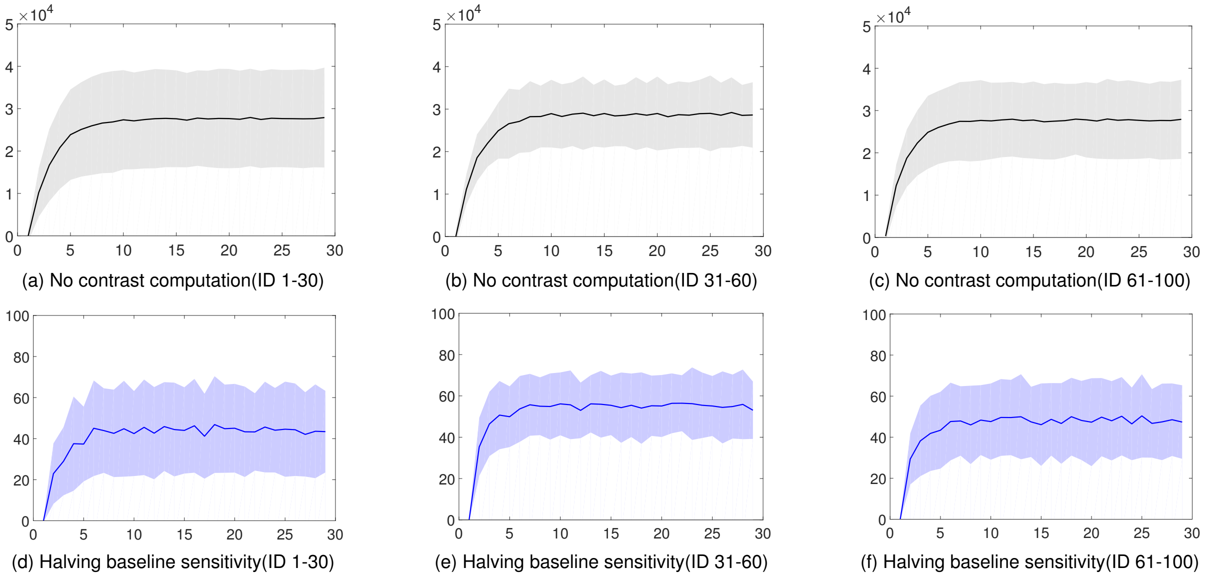
[8] Q. Fu, C. Hu, J. Peng, and S. Yue, "Shaping the collision selectivity in a looming sensitive neuron model with parallel ON and OFF pathways and spike frequency adaptation," *Neural Networks*, vol. 106, pp. 127–143, 2018.

[9] Q. Fu, C. Hu, J. Peng, F. C. Rind, and S. Yue, "A robust collision perception visual neural network with specific selectivity to darker objects," *IEEE Transactions on Cybernetics*, vol. 5, no. 12, pp. 5074–5088, 2019.

[10] Q. Fu, H. Wang, J. Peng, and S. Yue, "Improved collision perception neuronal system model with adaptive inhibition mechanism and evolutionary learning," *IEEE Access*, vol. 8, pp. 108 896–108 912, 2020.

[11] Q. Fu and S. Yue, "Modelling *Drosophila* motion vision pathways for decoding the direction of translating objects against cluttered moving backgrounds," *Biological Cybernetics*, vol. 114, no. 4, pp. 443–460, 2020.

[12] B. Hu, S. Yue, and Z. Zhang, "A rotational motion perception neural network based on asymmetric spatiotemporal visual information processing," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 11, pp. 2803–2821, 2017.



(a) No contrast computation(ID 1-30)

(b) No contrast computation(ID 31-60)

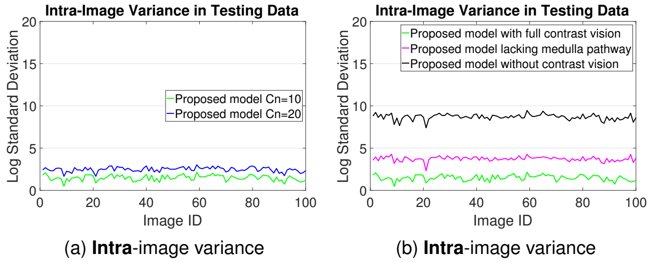
(c) No contrast computation(ID 61-100)

(d) Halving baseline sensitivity(ID 1-30)

(e) Halving baseline sensitivity(ID 31-60)

(f) Halving baseline sensitivity(ID 61-100)

Fig. 7. Comparative statistical response of the proposed network to natural image motion ID 1-100 (100 degrees/s): (a)-(c) response of the proposed network being removed all the contrast computations, (d)-(f) response of the network with lower baseline contrast sensitivity ($\alpha_3 = 10$ in Eq. 6).



(a) **Intra**-image variance

(b) **Intra**-image variance

Fig. 8. Statistical results of response variance on image motion ID 1-100 (100 degrees/s): (a) Comparison between different baseline sensitivity in contrast normalisation (Cn stands for α_3); (b) Comparison between networks with full, partial, and none of contrast vision.

[13] Q. Fu and S. Yue, "Modeling direction selective visual neural network with on and off pathways for extracting motion cues from cluttered background," in *Proceedings of the 2017 international joint conference on neural networks (IJCNN)*. IEEE, 2017, Conference Proceedings, pp. 831–838.

[14] Z. M. Bagheri, B. S. Cazzolato, S. Grainger, D. C. O'Carroll, and S. D. Wiederman, "An autonomous robot inspired by insect neurophysiology pursues moving features in natural environments," *Journal of Neural Engineering*, vol. 14, no. 4, p. 046030, Jul. 2017.

[15] Z. M. Bagheri, S. D. Wiederman, B. S. Cazzolato, S. Grainger, and D. C. O'Carroll, "Performance of an insect-inspired target tracker in natural conditions," *Bioinspiration & Biomimetics*, vol. 12, no. 2, p. 025006, Feb. 2017.

[16] H. Wang, J. Peng, X. Zheng, and S. Yue, "A robust visual system for small target motion detection against cluttered moving backgrounds," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 3, pp. 839–853, 2020.

[17] H. Wang, Q. Fu, H. Wang, P. Baxter, J. Peng, and S. Yue, "A bioinspired angular velocity decoding neural network model for visually guided flights," *Neural Networks*, vol. 136, pp. 180–193, 2021.

[18] A. Borst and T. Euler, "Seeing things in motion: Models, circuits, and mechanisms," *Neuron*, vol. 71, no. 6, pp. 974–994, 2011.

[19] A. Borst and M. Helmstaedter, "Common circuit design in fly and mammalian motion vision," *Nature Neuroscience*, vol. 18, no. 8, pp. 1067–1076, 2015.

[20] A. Borst, J. Haag, and A. S. Mauss, "How fly neurons compute the direction of visual motion," *Journal of Comparative Physiology A*, vol. 206, pp. 109–124, 2020.

[21] N. Franceschini, "Small brains, smart machines: From fly vision to robot vision and back again," *Proceedings of the IEEE*, vol. 102, pp. 751–781, 2014.

[22] J. R. Serres and F. Ruffier, "Optic flow-based collision-free strategies: From insects to robots," *Arthropod Structure and Development*, vol. 46, no. 5, pp. 703–717, 2017.

[23] Q. Fu, C. Hu, P. Liu, and S. Yue, "Towards computational models of insect motion detectors for robot vision," in *Towards autonomous robotic systems conference*, 2018, pp. 465–467.

[24] M. Joesch, B. Schnell, S. V. Raghu, D. F. Reiff, and A. Borst, "ON and OFF pathways in drosophila motion vision," *Nature*, vol. 468, no. 7321, pp. 300–304, 2010.

[25] M. Joesch, F. Weber, H. Eichner, and A. Borst, "Functional specialization of parallel motion detection circuits in the fly," *Journal of Neuroscience*, vol. 33, no. 3, pp. 902–905, 2013.

[26] P. D. Barnett, K. Nordstrom, and D. C. O'Carroll, "Motion adaptation and the velocity coding of natural scenes," *Current Biology*, vol. 20, pp. 994–999, 2010.

[27] D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Visual Neuroscience*, vol. 9, pp. 181–197, 1992.

[28] S. R. Olsen, V. Bhandawat, and R. I. Wilson, "Divisive normalization in olfactory population codes," *Neuron*, vol. 66, pp. 287–299, 2010.

[29] N. C. Rabinowitz, B. D. B. Willmore, J. W. H. Schnupp, and A. J. King, "Contrast gain control in auditory cortex," *Neuron*, vol. 70, no. 6, pp. 1178–1191, 2011.

[30] M. Carandini and D. J. Heeger, "Normalization as a canonical neural computation," *Nature Reviews neuroscience*, vol. 13, pp. 51–62, 2011.

[31] J. H. Reynolds and D. J. Heeger, "The normalization model of attention," *Neuron*, vol. 61, pp. 168–185, 2009.

[32] D. A. Clark, J. E. Fitzgerald, J. M. Ales, D. M. Gohl, M. A. Silies, A. M. Norcia, and T. R. Clandinin, "Flies and humans share a motion estimation strategy that exploits natural scene statistics," *Nature Neuroscience*, vol. 17, no. 2, pp. 296–303, 2014.

[33] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)," *arXiv:1606.08415v4*, 2020.

[34] Q. Fu and S. Yue, "Mimicking fly motion tracking and fixation behaviors with a hybrid visual neural network," in *Proceedings of the 2017 IEEE international conference on robotics and biomimetics (ROBIO)*. IEEE, 2017, Conference Proceedings, pp. 1636–1641.

[35] R. S. A. Brinkworth and D. C. O'Carroll, "Robust models for optic flow coding in natural scenes inspired by insect biology," *PLoS Computational Biology*, vol. 5, no. 11, 2009.

[36] F. Iida and D. Lambrinos, "Navigation in an autonomous flying robot by using a biologically inspired visual odometer," *Sensor Fusion and Decentralized Control in RoboticSystem III Photonics East*, vol. 4196, pp. 86–97, 2000.

[37] N. C. Klapoetke, A. Nern, M. Y. Peek, E. M. Rogers, P. Breads, G. M. Rubin, M. B. Reiser, and G. M. Card, "Ultra-selective looming detection from radial motion opponency," *Nature*, vol. 551, pp. 237–241, 2017.